COMMONWEALTH OF PENNSYLVANIA
LEGISLATIVE REAPPORTIONMENT COMMISSION

In re: Public Meeting of the Legislative Reapportionment Commission

VOLUME XXIII - Pages 1465-1582

Stenographic report of hearing held in Hearing Room No. 1, North Office Building, Harrisburg, Pennsylvania

Friday
January 14, 2022
2:00 p.m.

MARK A. NORDENBERG, CHAIRMAN

MEMBERS OF LEGISLATIVE REAPPORTIONMENT COMMISSION

Sen. Kim Ward
Sen. Jay Costa
Rep. Kerry Benninghoff
Rep. Joanna McClinton

Also Present:

Robert L. Byer, Esq., Chief Counsel
G. Reynolds Clark, Executive Director
Dr. Jonathan Cervas, Redistricting Consultant
Leah Mintz, Assistant Counsel
G. Carlton Logue, Esq. Deputy Counsel, Senate Majority Leader
Chad Davis, Research Analyst, Senate Republican Policy Office
C.J. Hafner, Esq., Chief Counsel, Senate Democratic Leader
Ronald N. Jumper, Esq. Deputy Chief Counsel, Senate Democratic Leader
Lora S. Schoenberg, Director, Senate Democratic Legislative Services
Rod Corey, Esq., Chief Counsel, House Republican Caucus
James Mann, Esq., Senior Deputy Chief Counsel, House Republican Caucus
Katherine Testa, Esq., Senior Legal Counsel, House Republican Caucus
William R. Schaller, Director, House Republican District Operations
Michael Schwoyer, Esq., Special Counsel, Deputy Chief of Staff for Legislation and Policy, House Democratic Caucus

Reported by:
Ann-Marie P. Sweeney
Official Reporter

Also Present:

Justin Klos, Director, House Democratic Office of Demographic Analysis
David Brogan, Esq., Director, House Democratic Legislation and Policy
Andrew McGinley, Esq., General Counsel, House Democratic Government Oversight Committee

INDEX

CHAIR NORDENBERG: Good afternoon, everyone. My name is Mark Nordenberg, Chair of the Legislative Reapportionment Commission.

I'd like to call this hearing to order and to acknowledge, in particular, the presence of the other Members of the Commission who are here. They include Senator Kim Ward, who is the Majority Leader in the Senate; Representative Kerry Benninghoff, who is the Majority Leader in the House; Representative Joanna McClinton, who is the Democratic Leader in the House; and joining us by Zoom is Senator Jay Costa, who is the Democratic Leader in the Senate.

We will be hearing from four retained experts during this afternoon's hearing. I have been advised by Leader McClinton that she would like to make a statement before we move into the testimony.

Leader McClinton.

REPRESENTATIVE McCLINTON: Thank you, Mr. Chairman.

I did not want to take any time away from all of the citizen and the shareholders who spoke this morning, so thank you for a few moments.

First of all, thank you again, Chairman Nordenberg. Over the last few weeks, we have heard from many different Pennsylvanians regarding the preliminary House plan that was approved by this Commission. The House map, aptly described by our Chairman as a composite map, resulting from months of meaningful collaboration and consultation, is one I am very proud to support. As I said on the day of its adoption, this map is long overdue. It's a long overdue step towards restoring both fairness and equality of representation in the Pennsylvania House of Representatives. It is a significant and meaningful step forward in our shared goal of providing every Pennsylvanian the opportunity for their voice to be heard and for their vote to count.

As many members of the press can and will confirm, I have done my very best to refrain from public comment, and instead I have listened. I've listened to Pennsylvanians, I've listened to interest groups that commented. It was never my goal to mount a public relations campaign or air any grievances. Rather, I wanted to hear what everyone had to say to this point. I have taken the comments to heart, and I am committed to exploring solutions to good faith concerns that have been expressed. As many have noted in their testimony or through op-eds over the last few weeks, as a result of the unprecedented level of transparency and public engagement in this process, again thanks to the leadership of the Chair, the local perspectives that have been brought to this table have been invaluable, and they deserve thoughtful consideration.

There's one area, however, I do feel compelled to personally address today, feedback that the Commission has received concerning minority representation. My Caucus has worked diligently to insure a map that accounts for significant demographic changes and is both fair and representative of the wonderful diversity of this Commonwealth of Pennsylvania, a map that is true to the fundamental principle of our Commonwealth's Constitution, and that is for free and equal elections. Those principles have been our collective North Star from the beginning. They're not only our legal obligation, they are long overdue.

In order to achieve our collective goal, we've thoughtfully and diligently worked to develop a House plan that allows all Pennsylvanians equal opportunity to translate their votes into representation that faithfully adheres to the Pennsylvania constitutional requirements and that complies with the Federal law. Highly regarded scholars support the preliminary House plan, and we have benefitted from their insights and their expertise. We'll hear from many experts this afternoon regarding the metrics by which they judge the preliminary House plan, including on the issues of fairness and minority representation. The data is very important, and it must inform our work.

But I also would like to take a moment and just speak from personal experience. Let's talk for one moment about the real-life implications of this map. My district, and others like it across the Commonwealth, insure that there is in fact diversity in representation, that communities like mine have a voice. Look no further than the legislative district of my partner in this redistricting process, Matt Bradford. Matt's had the privilege of serving the borough of Norristown in Montgomery County in the State House for over a decade. Before Matt, this community was served by a Republican, both of which are white men. For those of you that have not had the opportunity to visit the borough of Norristown, it is, in fact, incredibly diverse and has a vibrant minority population. Over the years, maps have been drawn in a way that diluted the voice of the people of Norristown. Currently, Norristown is in the 70th District. The significant population growth in Montgomery County over the last decade requires an additional seat be located within its borders. The preliminary House plan creates a new Norristown-based seat. This will allow the Norristown community to exercise their political power and elect a candidate of their choice.

During my time in the House, and as a Black woman, to be clear, 1 of only 10 in the entire 253 Member legislative body, and the first who's had the honor to have been elected and serve as Leader of the House Democratic Caucus, I've learned one lesson over and over again: That is, there is strength in coalitions. The Norristown seat is just one example of the coalition-building and the representative nature of this preliminary House map. It's been said many times that this Commission is not only historic in its approach but also its composition. My leadership is informed of course by my own experience, but more importantly, my leadership is proof that opportunity is power. That is a responsibility that I take very seriously. We stand on the shoulders of the soldiers and activists who ensured this right, and it is our responsibility to make sure that the franchise endures. I cannot and I will not support a map that falls short of the constitutional requirement that elections be free and equal -- yes, equal -- or a map that fails to ensure all Pennsylvanians, including people of color, have the equal opportunity to be engaged in their own representation.

The reality is this: The map that we have faithfully fulfills all constitutional reapportionment requirements and finally takes seriously meaningful steps to ensure equal opportunity and representation throughout this entire Commonwealth.

Thank you again, Mr. Chairman, for a few moments.

CHAIR NORDENBERG: Thank you, Leader McClinton.

Does any other Member of the Commission wish to make an opening statement?

(There was no response.)

CHAIR NORDENBERG: If not, let's proceed with the hearing.

Over the course of the months that we have been together, we have heard from dozens of invited witnesses we would put into the category of expert because we asked them to come and share perspectives with us. We're going to hear from four experts this afternoon. These experts, though, have been retained by one Caucus or another. To this point, the experts have come and simply offered their testimony in response to a request from the Commission. Each Caucus was given the opportunity to retain and present expert testimony. Neither of the Senate Caucuses chose to do so. The House Republican Caucus has one expert to present this afternoon. The House Democratic Caucus has three experts to present this afternoon. We will hear from the Republican expert first, and then from the three experts from the Democratic Caucus in succession.

The first of these experts is Dr. Michael Barber. He is an Associate Professor of Political Science and faculty scholar at the Center for the Study of Elections and Democracy at Brigham Young University.

Welcome, Professor Barber. We're glad to have you here.

DR. BARBER: Thank you.

Can you hear me ok?

CHAIR NORDENBERG: Yes, we can hear you well.

DR. BARBER: Excellent. And can you see the screen that I just shared?

CHAIR NORDENBERG: Yes, we can see that, too.

DR. BARBER: Excellent.

Thank you, and good afternoon. My name is Michael Barber, and I'm an Associate Professor of Political Science at Brigham Young University in Provo, Utah. I received my Ph.D. in political science from Princeton University in 2014, with an emphasis in American politics and quantitative methods/ statistical analysis. In my position as Professor of Political Science, I've conducted research on a variety of election and voting-related topics in American politics. Much of my research uses advanced statistical methods for the analysis of quantitative data, and I've worked on a number of research projects that use similar data sets as those required for the redistricting process. I've served as an expert witness in 10 election-related cases in the past 5 years, and a list of these cases and scholarly publications is contained in my CV, which was attached to the report I filed on January 7 to this Commission.

I've been asked by counsel for the Republican Caucus of the Pennsylvania House of Representatives to review the Legislative Reapportionment Commission's proposed redistricting plan for the State House and compare it to a set of simulated redistricting plans that are generated using only criteria outlined in Article II, Section 16, of the Pennsylvania Constitution. This simulation process ignores all partisan and racial considerations when drawing legislative districts. Instead, the computer simulations are programmed to create districting plans that follow only those objectives described in the Pennsylvania Constitution without paying any attention to partisanship, race, or other political considerations. This set of simulated districts is helpful because it provides a representative set of maps to which we can compare the Commission's proposed map to see if it is biased in favor of either party. If the Commission's map produces a similar outcome as the alternative set of simulated maps for which we are certain of the criteria used to generate these maps, we may reasonably conclude that the Commission's plan is also unbiased.

Alternatively, if the Commission's proposed plan significantly diverges from the set of simulated maps, this raises the question of why it diverges, for what reason, and in what direction it is biased. In this study, I created a model that conducted 50,000 simulated State House plans, each containing 203 legislative districts that are of roughly equal population, that are geographically compact, contiguous, and have minimal divisions of counties, cities, townships, and boroughs. These are the nonpartisan criteria outlined in Article II, Section 16, of the Pennsylvania Constitution.

This process of simulating districting plans to provide a comparison set of maps has been widely recognized in the scholarly literature on the topic and has been used in litigation around redistricting plans. Notably, the Pennsylvania Supreme Court accepted and relied upon simulations methodology in the *League of Women Voters* case in 2018. Once the simulated district plans are complete, only then do I compute the partisan composition of each district in the plan by calculating the proportion of votes across all statewide elections conducted between 2012 and 2020 that were won by the Democratic and Republican candidates in each of those districts.

The figure on this screen shows -- or will show in a minute -- the results of this simulation's exercise. Before I show the full distribution of simulation results and where the Commission's proposal falls in relation to this distribution, let me first orient the committee to this figure. The horizontal axis measures the number of Democratic-leaning districts generated by the simulations based on the average of statewide elections from 2012 through 2020. The vertical axis shows the relative frequency of maps in the simulations that generate each of the outcomes displayed on the horizontal axis. The distribution in the middle shows the results of the simulations.

For example, in 17.6 percent of the simulations, there are 97 Democratic-leaning districts. This is the most common outcome in the simulations. As you can see, the distribution roughly follows the bell curve, and approximately 80 percent of the simulations generate between 95 and 100 Democratic-leaning districts. The Commission's proposal, shown as the solid green line on the right side of the figure, is an extreme outlier from the distribution of simulated districts drawn using only the nonpartisan criteria described in the Pennsylvania Constitution. The Commission's plan generates more Democratic-leaning districts than 99.998 of the simulations.

This result is statistically significant. Indeed, the number of Democratic-leaning districts is outside of all but 1 of the 50,000 different simulated maps. This significant deviation suggests that the Commission's map was not drawn using only the nonpartisan criteria in the Pennsylvania Constitution. Had this been the case, we would not expect the Commission's proposal to deviate so dramatically from a set of maps drawn by the computer using only those criteria. Instead, this is very strong evidence that other considerations went into the map-drawing process.

To uncover why there is such a difference between the Commission's plan and the distribution of simulated districts, it is instructive to look at subsets of the State where discrepancies may arise. Given the geographic distribution of voters in Pennsylvania and the clustering and sorting of Democrats within the large and medium-sized cities of the

State, there are only relatively few locations in which Democratic districts can be constructed. Scholarship and political science has noted that the spatial distribution of voters throughout a State can have an impact on the partisan outcomes of elections when a State is, by necessity, divided into a number of legislative districts.

The much-abbreviated summary of how the geographic distribution of voters impacts the redistricting process is that the party whose voters are more geographically concentrated stand at a natural disadvantage when single-member districts are drawn. As this slide illustrates, voters who support the two major parties are not evenly distributed across Pennsylvania. Democratic voters tend to live in the large and medium-sized cities of the State, while Republican voters are more likely to live in the suburban and rural portions of the State. The practical impact of this geographic pattern is that when districts are drawn using the nonpartisan criteria of geographic compactness and minimization of county and municipal divisions, there can be an incidental partisan impact in the number of seats that favor Republicans or Democrats.

The particular distribution of voters in Pennsylvania would not be a problem for Democrats if district boundaries were able to amble about the State and slice up counties and municipalities, thereby allowing those densely concentrated Democratic voters to be spread out more evenly across larger geographic territory to create more Democratic-leaning districts. Jonathan Rodden, one of the foremost political geographers in the nation, notes this by saying: "Democrats would need a redistricting process that intentionally carved up large cities like pizza slices or spokes of a wheel, so as to combine some very Democratic urban neighborhoods with some Republican exurbs in an effort to spread Democrats more efficiently across districts." However, the laws governing redistricting in Pennsylvania run counter to either of these strategies.

With the limited time I have here, I will focus on one such example of how the Commission's proposed map follows this strategy described by Professor Rodden remarkably closely. My report provides more detail and shows the patterns I describe here applied to a number of cities throughout the State. The combined population of Lehigh and Bucks Counties is equal to approximately 16 legislative districts. In the 16 districts that cover this area, the Commission's proposal generates 11 Democratic-leaning districts. The most common outcome in the simulations is nine Democratic districts. The red vertical line at 11 represents the number of Democratic-leaning districts in the Commission's map in this portion of the State. In 99 percent of the simulations, there are fewer than 11 Democratic-leaning districts in these counties. The Commission's plan achieves this in part by dividing the city of Allentown more than is necessary to more evenly distribute the Democratic voters that live in the city across more districts.

Allentown is too large to be completely contained in one district and will need to be divided into two districts in any proposal. However, the Commission's plan divides the city into three districts - 22, 134, and 132 - as shown in the map on the screen. This map shows the distribution of Hispanic voters in and around Allentown. The darker green colors show heavy concentrations of Latino voters in the city itself, while the yellow colors show the areas with lower proportions of Hispanic voters just outside of Allentown. One justification for this division of the city might be to create three majority Hispanic districts, as Allentown has a nearly 50-percent Latino voting age population. However, this is not the case. In fact, exactly the opposite is true, and proceeding district by district shows how.

District 22 has a Hispanic voting age population of 50.8 percent. District 134 has a lower Hispanic voting age population of 38.5 percent. Finally, District 132 has a much lower Hispanic voting age population of 18.1 percent. By taking a piece of Allentown and placing it in District 132, the plan dilutes the voting strength of Hispanic voters in the city by spreading them across Districts 134 and 132, when a proposal that divided the city across only two districts would generate two districts that both approach having a majority Hispanic voting age population.

So if increasing the voting power of minority communities is not accomplished by dividing Allentown into three districts, what is the purpose of this potentially unconstitutional division of the city? Partisan gain stands out as an obvious consideration. The map on the screen shows the partisan distribution of voters across the three districts that intersect Allentown. Dark blue areas are heavily Democratic precincts, and the pink areas are majority Republican precincts. In dividing Allentown across three districts, the Commission's plan creates three districts that all lean Democratic, rather than having two overwhelmingly Democratic districts that were entirely contained within Allentown's city limits.

However, if this division of Allentown was accomplished to increase Democratic representation while also harming the Latino population's ability to elect their preferred candidates of choice by splitting the city more times than is necessary, this choice potentially runs afoul of both the Pennsylvania Constitution's instructions to not split a city unless absolutely necessary, and the Federal directives in the Voting Rights Act to preserve minority communities of interest.

Across the State, the Commission's proposed map follows the same pattern of dividing heavily Democratic cities more than is necessary in order to maximize the number of Democratic-leaning districts that can be drawn. This often has the impact of diluting minority representation in cities. Harrisburg, Lancaster, Reading, Scranton, and State College are all divided into more districts than their respective populations would require. Again and again, we see the Commission's plan following the playbook outlined by Dr. Rodden's academic research: Heavily Democratic cities intentionally divided to be more efficiently spreading Democratic voters across more districts, often to the detriment of minority influence in these cities.

One concern might be that without dividing these cities, a plan may not contain a sufficient number of districts that contain a large share of minority voters. This argument, however, is wrong for two reasons. First, as I showed in the examples of Allentown, dividing these cities into more districts than necessary often has the impact of diluting minority representation, not increasing it. Furthermore, the simulations show that it is entirely possible to generate a similar number of districts that meet a certain threshold of minority voting strength without splitting these cities as often as is done by the Commission's proposal.

In short, the simulations allow us to conclude that, one, the Commission's proposed plan is an extreme partisan outlier. Two, the Commission's plan confers this significant partisan advantage on Democrats by splitting up Pennsylvania's cities and towns. And, three, any justification that this is necessary to generate minority-opportunity districts is false, given that the proposed plan does not generate a significantly larger number of minority-opportunity districts than do the simulations.

With the remainder of my time, I want to spend a few minutes addressing the reports submitted by Drs. Imai and Warshaw. Dr. Imai submitted a report on January 7 in which he asserts that he used simulated districting to compare the Commission's proposal to a set of districts drawn using the nonpartisan criteria outlined in the Constitution. However, this report does not contain any details about the results of his simulations, which makes it virtually impossible to analyze his findings. About 12 hours ago, he submitted an amended report that included some additional details. In the amended report, the appendix contains measures of the simulations with respect to geographic compactness and the number of county and municipal divisions. These results show that Dr. Imai's simulations are dramatically less compact than my simulations or the districts in the proposed plan. Furthermore, Dr. Imai's simulations contain many more municipal splits than my simulations for the proposed plan.

In fact, a close look indicates that the lowest number of municipal splits in his simulations is larger than the largest number of municipal splits in my simulations.

As a result, Dr. Imai's simulations are not creating a valid comparison set as they do not adhere as closely to the districting criteria outlined in the Pennsylvania Constitution. However, despite these significant differences in both Dr. Imai's and my simulations, the Commission's proposal is more Democratic than the simulations in a statistically significant way. Dr. Imai then reports that the difference between the Commission's proposal and the simulations becomes substantially smaller when he includes race as a factor in the simulations. However, even when explicitly incorporating race into the analysis, in three of the six analyses, the Commission's proposed plan is a statistical outlier.

Furthermore, as I showed earlier, even without directly instructing the model to consider race, the simulations generate a very similar number of these districts naturally, given the geographic distribution of minority voters in Pennsylvania's cities and towns. Importantly, they do this without also drawing a Democratic-skewed plan.

Dr. Warshaw provides an analysis of the Commission's proposal using a variety of measures of partisan fairness and finds the map to be, quote, "fair." And yet my results show that this map is an extreme partisan outlier that dramatically tilts towards benefitting the Democratic Party. How can our results be reconciled? The fundamental difference in our results is the definition of the term "fair." When discussing the fairness of the Commission's plan, Dr. Warshaw is referring to what he calls "vote-seat representation," or in other words, does the proportion of votes cast across the State for a particular party yield a similar number of seats in the legislature? This question is tantamount to asking if the Commission's plan generates proportional representation. However, our system of government is not proportional representation. Instead, we elect legislators in individual districts that are composed of specific geographic units. Doing so offers the voters geographic-based representation that many consider to be advantageous.

As a result, the way to generate a fair map by Dr. Warshaw's standard is to engage in partisan gerrymandering or drawing district lines with partisan intent. The benefit of the simulations approach that I present here is that it measures fairness only in terms of the inputs and not the outputs of a map. Thus, a fair map-drawing process does not consider partisanship at any point in the process, neither in the inputs used to draw the map nor in making explicit goals with respect to the partisan balance of the legislature elected using those maps.

This definition of fairness also more closely aligns with the process articulated in the Pennsylvania Constitution, which contains no mention of a desired outcome in the legislature, nor any reference to proportional representation as a favorable objective. In the case of the Commission's proposed map, it clearly fails on this metric of fairness. The plan is drawn to assist Democratic voters in achieving a particular partisan outcome, and in doing so deviates dramatically from a set of maps that we know were drawn using only the neutral criteria outlined in the Pennsylvania Constitution.

Thank you, and I look forward to your questions.

CHAIR NORDENBERG: Thank you, Professor.

Let me open the floor for questions or comments from the Members of the Commission.

Leader McClinton.

REPRESENTATIVE McCLINTON: Thank you, Mr. Chairman, and thank you, Dr. Barber, for participating today and sharing your opinions.

A few specific questions to start here. Your simulation model did not include any constraints based on the requirements in the Voting Rights Act, correct?

DR. BARBER: Yes. As I noted, the simulations only considered the criteria outlined in the Pennsylvania Constitution. However, I also note that as a result of the fact that voters are distributed in a particular way in the State, even a simulation process that does not explicitly include racial considerations ends up generating a number of minority-majority districts, or minority-opportunity districts, that are very similar to the simulations produced by others that explicitly use race at the outset of the process.

REPRESENTATIVE McCLINTON: Your simulation model also did not include any constraints based upon the requirements of Article I, Section 5, of the Pennsylvania Constitution, correct?

DR. BARBER: As I said, the criteria in the model are those that come from Article II of the Pennsylvania Constitution.

REPRESENTATIVE McCLINTON: Just so I'm clear, for the record, Dr. Barber, your model did not include any constraints based upon the requirements of Article I, Section 5, is that correct?

DR. BARBER: That's correct.

REPRESENTATIVE McCLINTON: Thank you.

You are not offering any opinions this afternoon with respect to application of the Voting Rights Act to the preliminary House plan, is that correct?

DR. BARBER: That's correct.

REPRESENTATIVE McCLINTON: You are not offering any opinion as to whether the plan comports with the free and equal guarantee in Article I, Section 5, of the Pennsylvania Constitution, correct?

DR. BARBER: I am offering opinions as to whether the map aligns with a set of districting criteria that are outlined in Article II of the Pennsylvania Constitution. I don't think the simulations -- no simulations are going to take into account anything about free and fair elections, because, you know, we're measuring the outcome of these simulations based on previous elections that have already been conducted.

REPRESENTATIVE McCLINTON: I appreciate your response, but I just want to be clear for this record, you are not offering any opinion as to whether the plan comports with the free and equal guarantee in Article I, Section 5, of the Pennsylvania Constitution, correct?

DR. BARBER: Yes, that's correct.

REPRESENTATIVE McCLINTON: Okay.

The next thing I'd like to talk about is some of the county-by-county analyses you submitted. And just so that my fellow Commissioners can follow along with this line of questioning, I do have copies of the county groupings being passed out right now to all of the Commissioners, and of course their counsels, that are present.

Now, in your initial report that you submitted, you put Lehigh and Schuylkill Counties together, is that correct?

DR. BARBER: I believe so, yes.

REPRESENTATIVE McCLINTON: Do you have access to it?

DR. BARBER: Not immediately.

REPRESENTATIVE McCLINTON: Okay. If you could reference page 9 of your report, just so that you can follow along with me here.

DR. BARBER: Yes.

REPRESENTATIVE McCLINTON: Do you have it up?

DR. BARBER: One moment. Yes, I have it here.

REPRESENTATIVE McCLINTON: Thank you.

Now, in that table, you compare the Commission proposal to your simulations, is that right?

DR. BARBER: Yes, that's correct.

REPRESENTATIVE McCLINTON: And you find that both Lehigh and Schuylkill being compared have four Democratic-leaning districts under the Commission proposal.

DR. BARBER: Yes, I see that.

REPRESENTATIVE McCLINTON: Your simulation range returns anywhere from three to five Democratic-leaning districts.

DR. BARBER: Yes, I see that.

REPRESENTATIVE McCLINTON: So the Commission proposal of four is in the middle of the simulation range.

DR. BARBER: That's correct. In the -- so I think we're looking at the initial report that I submitted, which is a much shorter, abbreviated report compared to the one that I later submitted on January 7. And so in that report, the range simply notes the lowest to the highest number of districts generated by the simulations. That table does not represent the probability or the proportion of simulations that are generated in each of those -- that correspond with each of those numbers, but it is certainly within the range. Yes, that is correct.

REPRESENTATIVE McCLINTON: Now, let's talk about the second table, where you put Lehigh and Bucks County together. Do you have access to that?

DR. BARBER: I do, yes.

REPRESENTATIVE McCLINTON: Okay.

Now, when you change the grouping in your much longer report, you put Lehigh and Bucks together and you produce a very different result that's no longer in the middle of the simulation range, is that right?

DR. BARBER: Yes. The groupings change across the reports as a reflection of the way in which the proposed map draws the districts. And so in the later report, I tried to more closely follow the kind of natural groupings that the Commission's report also does to make a closer apples-to-apples comparison.

REPRESENTATIVE McCLINTON: Now, you're familiar that Lehigh and Bucks are two vastly separate areas in the Commonwealth, is that right?

DR. BARBER: Yes.

REPRESENTATIVE McCLINTON: And you're familiar that they are not comparable areas of the Commonwealth and have very different demographics?

DR. BARBER: Yes.

REPRESENTATIVE McCLINTON: And you're also aware that they are not, in fact, comparable?

DR. BARBER: Yes. I am not trying to make a claim that they are comparable in my report.

REPRESENTATIVE McCLINTON: So in the second table where you compare Bucks County and Allentown, and you recognize they're not in the same region of Pennsylvania and they're not comparable, you never, in fact, compare Lehigh's neighboring Northampton County, do you, in your simulation report?

DR. BARBER: I don't directly draw comparisons between Lehigh County and Northampton, nor do I draw comparisons between Lehigh and any other county. The grouping of the counties is simply, these are counties that are grouped together to align with the way in which the Commission's proposal groups districts, and these counties also, by virtue of their population, yield a round number of districts in terms of their combined populations.

REPRESENTATIVE McCLINTON: Just for clarification, you picked which counties you decided to group together for this simulation, is that right?

DR. BARBER: Yes, based on what I saw in the Commission's proposal.

REPRESENTATIVE McCLINTON: You chose to pair Lehigh and Bucks County?

DR. BARBER: Yes.

REPRESENTATIVE McCLINTON: And you chose not to compare Lehigh and Northampton County?

DR. BARBER: They are not grouped together. I don't make comparisons across counties. So I want to be clear that that's not what these simulations are doing. I'm not comparing one county to another. The counties are simply grouped together because they form a natural, by population, a round number in terms of the number of districts that are composed in those two counties.

REPRESENTATIVE McCLINTON: And you did not do a simulation between Lehigh and Northampton? It's just "yes" or "no."

DR. BARBER: Grouping those two counties together? No, I did not.

REPRESENTATIVE McCLINTON: Okay.

And there's no basis to do that other than trying to manipulate and contradict the conclusion from your initial report, which the Commission's proposal fell in the middle of your simulation range, to what you provided today?

DR. BARBER: I do not think that's an accurate statement. The difference in the groupings is a reflection of the Commission putting forward their proposed map, which then allows for a comparison of, okay, there's a round number of districts in these county groups, and so that allows us to make a comparison in terms of a round number of districts. That's all that's going on there. There's no comparisons across counties. There's nothing like that going on in the report.

REPRESENTATIVE McCLINTON: I don't have any further questions, Mr. Chairman. Thank you.

Thank, you, Dr. Barber.

DR. BARBER: Thank you.

CHAIR NORDENBERG: I've just got a couple of questions. You did indicate that your simulations focused on the quantitative requirements of Article II of the State Constitution, is that correct?

DR. BARBER: Yes, that's correct.

CHAIR NORDENBERG: And in the table on page 7 of your report, it's clear that the Commission's plan performed very well when measured against those requirements of Article

II, isn't that correct?

DR. BARBER: Yes, that's correct.

CHAIR NORDENBERG: And in fact, in certain of the factors, such as total county splits, municipality splits, and the Polsby-Popper measure of compactness, the Commission plan is close to or at the very best performing plans in the calculations that you made?

DR. BARBER: Yes, that's correct. I think that's an important point, for two reasons. One is that it definitely performs very well in terms of the numeric, you know, the number of county divisions and municipal divisions, which I think suggests that it's not necessarily just the number that are split. It's important that that number is low, because that's what the Constitution details, but it's also then important to think about which cities are split and why are they split, if there's some particular reason for the choices that are made in which cities have been split by the Commission's proposal.

The other thing that I think is important to note is that the simulations I provide here are very close to the number of splits that are -- that occur in the Commission's proposal, and other simulations that we're going to see are much further away from that number.

CHAIR NORDENBERG: All right. And to go back to another point that I think you made, and that was reaffirmed through the questions of Leader McClinton, your simulations do not take into account any racial data?

DR. BARBER: That's correct. The simulations are drawn without consideration of partisanship or race. Of course, we can then measure the output or the outcome of those simulations using partisan and racial data, which I presented in the slides earlier.

CHAIR NORDENBERG: You also made reference both to the academic literature in this area and to your own CV, which is impressive. Which of your articles amounts to contributions to the academic literature about the use of the simulations?

DR. BARBER: So, most of my work in redistricting and simulations has been on the litigation side of things. And so as I noted at the beginning of my testimony, I've worked in a number of election-related cases and additionally in some redistricting cases in North Carolina and Ohio prior to working here in Pennsylvania.

CHAIR NORDENBERG: Which means, I gather, that a person who was interested in learning more about your views of simulations could not go to any of your articles and find them to be on-point?

DR. BARBER: I'm not published on the use of redistricting simulations. I note in my report that I'm using a model developed by Professor Imai, who's going to testify later today.

CHAIR NORDENBERG: And have you published in the area of the Federal Voting Rights Act?

DR. BARBER: I have research that discusses race and politics, but I do not have research that specifically discusses provisions of the Voting Rights Act.

CHAIR NORDENBERG: Thank you very much.

DR. BARBER: Thank you.

CHAIR NORDENBERG: Any other questions?

(There was no response.)

CHAIR NORDENBERG: Thank you, again, Professor Barber.

DR. BARBER: Thank you.

CHAIR NORDENBERG: Our second expert witness today will be Professor Kosuke Imai, who is a Professor of Government and Statistics, and an affiliate of the Institute for Quantitative Social Science at Harvard. He previously held a faculty appointment at Princeton University, where he was the founding director of its program in Statistics and Machine Learning.

Welcome, Professor Imai.

DR. IMAI: Thank you very much. Let me share my slides. Are you able to see the slides okay?

CHAIR NORDENBERG: Yes, we can see them.

DR. IMAI: Okay. Thank you very much. Thanks for the Commission for the opportunity to present my redistricting simulation analysis of the preliminary State House reapportionment plan. This is an exciting opportunity for me because the redistricting simulation is one of my primary areas of research, and to be able to share in the actual redistricting process is a tremendous honor.

Let me just briefly introduce myself. As I was introduced, I currently hold a position of Professor in the Department of Government and the Department of Statistics at Harvard University. So this joint appointment is a first in the history of the university. My research area focuses on the intersection between political science, or social science in more general, and the statistics and machine learning. And previously, I was at Princeton University.

In terms of research fields, the first area is causal inference, where I developed statistical methods for determining cause and effect, and particularly I focused on the impact of public policies. The second research field is computational social science, where I developed computational algorithms for improving and evaluating public policies. So my research is really the intersection between statistics, computer science, and political science in general.

For this particular case, my relevant expertise is I've been conducting and developing redistricting simulation analysis over the last 10 years. I've developed several algorithms, actually simulate the redistricting plans, and then I apply them to a variety of cases. I have published several articles doing so.

I would like to also emphasize that I developed the open-source software package called *redist*, which has more than 30,000 downloads, and this package basically makes the cutting-edge algorithms available to other researchers and policymakers, you know, democratize this technology, powerful technology, and this is the one package that Professor Barber used in his own analysis as well. And it's open-source, so that means that the code is publicly available. You can view the code in extent, and it's freely available and you can run it on any computer you might have.

I would like to give an overview of redistricting simulation analysis. So what is the simulation analysis? Well, it's generating a large number of alternative plans that one could have drawn under a specified set of redistricting criteria. And as Professor Barber showed, you then compare these alternative plans with a proposed plan to evaluate its properties.

So what are the benefits? The primary benefits of simulation analysis is its ability to control for State-specific political geography and redistricting laws. So in traditional methods, one would compare the plan from Pennsylvania with the plan from another State, maybe Ohio, maybe New York, and then see which plan is more fair, or less biased. However, the problem of such a comparison is that different States, as you all know, have different political geography, and they may use different redistricting laws. So it's not apples-to-apples comparison. So simulation analysis uses the same political geography and the same set of redistricting laws.

Simulation analysis is also very transparent. You specify a set of inputs and the algorithm will generate alternative plans under a specified set of criteria, which is the inputs that you provide. And as a result, you'll be able to isolate a particular factor you might be interested in. So you could add a factor as an input or you can actually take out the factor and then see how the alternative plans might change.

Finally, it's very important that these algorithms that I've developed and I use have a mathematical guarantee. It generates a representative sample of alternative plans one could have drawn under a set of alternative -- different redistricting criteria. This property is very important, because there are so many ways to draw plans under a set of criteria, so you can never enumerate every single one of them. So instead what you try to do is to get the representative sample of such plans, like the polling in opinion polls. You're not going to get every single person in the U.S., you try to get a representative sample of the voters.

So one thing I want you to take away from my presentation is that input criteria matters. So what you feed the simulation algorithm is critically important, and that has to be carefully chosen not by someone like me, a statistician or political scientist, but by the policymakers.

Just to first summarize some key conclusions, the first conclusion that I'm going to draw for today's presentation is that the consideration of majority-minority districts, in addition to the constitutional constraints, in the simulation algorithm substantially alters the conclusion of simulation analysis. So this is in line with the point I really want to make is that what inputs you put in the algorithm really determines the conclusion. So you have to be careful what you feed into the algorithm.

The second conclusion is related to my finding. When the majority-minority districts are considered, in addition to constitutional constraints, there is no empirical evidence that the preliminary plan is a partisan gerrymander. So given that, let me present the results from my simulations and explain how I reached these conclusions.

The first simulation I ran is something I call race-blind simulation setup. Here I simply used the five constitutional criteria mentioned in Professor Barber's report, so there's 203 geographically contiguous districts, equal population, I chose mine as plus-, minus-5 percent, the districts should be compact, and one should avoid county splits and also municipality splits. So these five criteria are put into the simulation algorithm, that's my package *redist* software package, and then the package will generate, you know, 5,000 in this case, but you can generate more if you want, alternative plans. And these are representative plans that you will get from the population plans one would draw under this criteria.

Now, I would like to mention one thing. I first started with trying to replicate Professor Barber's race-blind simulations, but Professor Barber only didn't use -- only used this criteria, and yet I was not able to duplicate it in part because insufficient information was given in his report. In fact, my package implements several different algorithms. The report did not mention which algorithm was used. So it was difficult for me to duplicate his results, and so what I'm going to present today for the race-blind simulation is my own race-blind simulation. So that's not the same as you see as the analysis Professor Barber conducted.

So here is the race-blind simulation result. So I tried to mimic the presentation using the same format of Professor Barber's report. In fact, I used the same three sets of statewide elections as he used. So he chose 2012 to 2020 statewide election composite, that's on the left; 2014 to

2020 statewide elections, that's in the middle; and 2020 statewide elections alone, that's on the right. Now, there could be other composite of statewide elections or types of elections, and that could, you know, affect the results as well. But I'm simply here, for the sake of comparison, I'm using the same exact set of election results that Professor Barber used.

Now, what do the results say? So, on the X axis, you have the number of Democratic districts. So let's take the middle figure, which is 2014-2020 statewide elections. You see the red vertical line. Maybe you see my cursor right here (indicating). That's basically the number of Democratic districts on the preliminary plan. So that's 105, according to this particular data set. And the gray histogram, these little mountains, is the simulated plans.

What's the number of Democratic districts that would generate on the simulated plan? So model outcome is 101, which is 4 districts less than the preliminary plan. So according to this, the preliminary plan yields four to eight more Democratic districts than the race-blind simulated plans, depending on which election set you use. So if you look at that middle figure again, you see that there's some simulated plan that yields the same number of Democratic districts, but many of them actually yield your number.

Now let's compare this with Professor Barber's.

So it appears that maybe the results are similar, like the preliminary plan seems to be more favorable towards the Democratic Party than the simulated plan. However, the magnitude is quite different. So if you look at the magnitude, Professor Barber concludes it's 8 to 10, which is 2 to 4 districts more than the simulated plan I introduced. It's hard for me to tell where this difference is coming from, but the same race-blind simulation analysis yields much fewer Democratic seats under Professor Barber's analysis, and it ended up overstating the degree to which the preliminary plan is more favorable to the Democratic Party.

Now what I'm going to do is, I'm going to show you the second simulation. So here I want to show how different inputs can make a difference. So in this simulation, I'm going to call it Simulation A, there are five constitutional constraints, as exactly as I stated earlier, the same set. And I just added one constraint which considers eight majority-Black and four majority-Hispanic districts. So everything is the same as the race-blind simulation. I have all the constitutional constraints, and I added one additional constraint. And again, the same algorithm produces a different 5,000 set of alternative plans.

So what did the results look like? So, again, let's look at the middle here, which is between 2014 and 2020 statewide elections. What you see, again, is under the preliminary plan there are 105 Democratic districts. Under the simulated plan, now the model outcome is 103 districts. So the difference now is two districts, which is less than 1 percent of the total number of seats that we have in the State. And this is statistically indistinguishable. So what that means is that this preliminary plan is a typical plan under my Simulation A setup. And remember that Simulation A and the race-blind simulation, the only difference -- the constitutional criteria is the same, and the only difference is the consideration of majority-Black and majority-Hispanic districts.

If you look at the 2012-2020 statewide election, it is a borderline statistical significance. So if you ask a different statistician, he would probably give you different answers. 2020 elections haven't really changed much, but I think this result is much less reliable because it's one election, so it's influenced by specific factors particular to that election year. And that's why most academics would use a composite of multiple elections, like the ones you see on the left two graphs.

So what conclusion do I draw from this? Well, when the majority-Black and majority-Hispanic districts are additionally considered, in addition to the five constitutional criteria, that's very important, the preliminary plan is not a partisan gerrymander, okay. I say, depending on the specific set of elections analyzed, because on the left two, that's what we see.

Now let's look at the last simulation that I'm going to show you. I call it Simulation B. Okay. So here, exactly the same thing, I'm going to use the five constitutional constraints as I've done before. Okay. And then I add an additional constraint for 25 majority-minority districts. Okay. This is not just majority-Black districts or majority-Hispanic districts, it also includes the so-called coalition districts where the minority groups together form a majority. Okay. So what were to happen if I replace the previous, you know, simulation of minority-Black and minority-Hispanic constraint with this constraint? Again, let's look at the middle figure first. Okay.

So 2014-2020 statewide elections, as I indicated many times, on the preliminary plan you get 105 seats. Now the model outcome on the simulated plan is now 106. That's one more than the preliminary plan, which means that the simulated plan is more favorable towards Democrats than the preliminary plan. In other words, the preliminary plan is slightly less favorable towards Democrats than the simulated plan. And in fact, it's in the middle of the distribution, so we cannot statistically distinguish the preliminary plan from a simulated plan. So it's a typical plan under the simulation setup.

Same conclusion holds if you look at the 2012-2020 statewide elections. It's not statistically distinguishable, and it's a typical plan. The preliminary plan is a typical plan under the Simulation B setup in terms of the partisan outcome. So from this I conclude that when the majority-minority districts are additionally considered in addition to the five constitutional constraints, the preliminary plan is not a partisan gerrymander. Using these two composites, the more elections that one used for 2020 statewide election, that's on the right. Okay.

So just to end my presentation by summarizing the findings, my race-blind simulation shows that Professor Barber's race-blind simulation tends to overstate the degree to which the preliminary plan is favorable towards the Democratic party under the race-blind setup. The difference is, it's unclear without more details on how Professor Barber ran the simulation. When the majority-Black and majority-Hispanic districts additionally are considered beyond the constitutional constraints, the preliminary plan is not statistically distinguishable from the simulated plans, depending on the specific set of elections analyzed. And when the majority-minority districts are additionally considered, again, the preliminary plan is not statistically distinguishable from the simulated plans using two sets of composite statewide elections that Professor Barber used.

All together, I conclude that when the majority-minority districts are additionally considered, the preliminary plan is not a partisan gerrymander in terms of the likely number of Democratic districts. So that's the end of my presentation.

Thank you very much.

CHAIR NORDENBERG: Thank you very much, Dr. Imai.

Are there questions from the Members of the Commission?

Majority Leader Benninghoff.

REPRESENTATIVE BENNINGHOFF: Thank you, Mr. Chairman. I was trying to give deference to the Members to the right of you.

Thank you, Dr. Imai. I appreciate your time. It's a lot of information to try to digest. I've been looking through some of your graphs. I just want to make sure I'm hearing exactly what you're saying. If I understand your report, under your blind study, the race-blind analysis, you agree with Dr. Barber that the preliminary LRC map yields a greater number of districts for the Democrats than seen in the 5,000 simulated plans, correct?

DR. IMAI: Yes, except the differences, you know, the difference between the two is much smaller under my simulation.

REPRESENTATIVE BENNINGHOFF: That's all right. I just wanted to -- because I know you progressed through this. I just wanted to make sure I heard you right.

DR. IMAI: Yep. That's correct.

REPRESENTATIVE BENNINGHOFF: Do you consider that plan to be somewhat of an outlier, I think I heard you say, or did I hear that incorrectly?

DR. IMAI: Statistical outlier under that particular simulation.

REPRESENTATIVE BENNINGHOFF: Okay. I appreciate that.

DR. IMAI: Yes.

REPRESENTATIVE BENNINGHOFF: Under the analysis of the race-blind analysis, out of 5,000 simulated plans, how many of those would have predicted a lower number of Democrat seats than the LRC plan?

DR. IMAI: I don't have exact number, but in a vast majority of simulated plans yields fewer number of Democratic districts than the preliminary plan under the, you know, race-blind simulation setup.

REPRESENTATIVE BENNINGHOFF: Fewer of the 5,000 simulated ones than what the LRC plan states?

DR. IMAI: Right. So the 5,000 simulated plans, the vast majority of them yields, you know, fewer Democratic districts than the preliminary plan under race-blind simulation.

REPRESENTATIVE BENNINGHOFF: Okay. I appreciate--

DR. IMAI: I don't have the exact number. So there may be, you know, a few that's similar across. I think we saw a few there in the graph, but I don't have the exact number there.

REPRESENTATIVE BENNINGHOFF: I appreciate your patience there. I mean, you're using a lot of different terminology that has not been spoken much in different--

DR. IMAI: Yes.

REPRESENTATIVE BENNINGHOFF: --hearings today, so I just want to make sure I'm hearing what's being said. I've tried to read through the analysis as well. Unfortunately, it's in very, very small print.

You indicated the race-aware simulation analysis draws districts on the basis of race to guarantee that every simulated plan includes a certain number of majority-Black and majority-Hispanic districts. I have a couple of questions about that.

DR. IMAI: Okay.

REPRESENTATIVE BENNINGHOFF: Again, because I want to make sure I'm hearing it right. I see in your race-conscious simulation you require the computer to draw majority-Black and I believe four majority-Hispanic districts in Philadelphia and 14 other counties near Philadelphia. I'm just curious, how did you select that number of districts?

DR. IMAI: Yeah. So this is consistent with what I observed in the preliminary plan.

REPRESENTATIVE BENNINGHOFF: You were trying to parallel the number to the LRC plan?

Dr. IMAI: Yeah. Number and location.

REPRESENTATIVE BENNINGHOFF: Okay. Then in the race-conscientious Simulation Plan B--

DR. IMAI: Sorry, I should mention that once I, you know, I tell the computer algorithm to find this many majority-Black, majority-Hispanic districts in these areas, once the computer finds them, then I do the rest of the simulations. So I set aside those, the ones the computer finds, and then do the rest of the analysis completely race blind, exactly the same way that I did the race-blind simulation.

REPRESENTATIVE BENNINGHOFF: All right, I'm on this, just keeping track of what you're telling the computer to do.

DR. IMAI: Okay.

REPRESENTATIVE BENNINGHOFF: Again, in your race-conscientious Simulation B, make sure I'm hearing this right, you again told the computer to draw 8 majority-Black districts and 4 majority-Spanish districts, but now also draw 13 coalition districts in Philadelphia, the 14 districts near Philadelphia and also Allegheny County. How did you decide to ask the computer to draw these 13 coalition districts? What was the criteria for that?

DR. IMAI: You'd like to know how, like in general terms, how I instruct the computers to draw a certain number? Is that correct?

REPRESENTATIVE BENNINGHOFF: Well, I mean, the old saying is data in, data out. So you're instructing the computer to do certain things with certain data. I'm just curious as to--

DR. IMAI: Right, so--

REPRESENTATIVE BENNINGHOFF: --how you chose those--

DR. IMAI: Oh.

REPRESENTATIVE BENNINGHOFF: --13 coalition districts--

DR. IMAI: Oh. Okay.

REPRESENTATIVE BENNINGHOFF: --to do so.

DR. IMAI: Oh. Okay. Okay. I see. I got you. I got you.

So the number and the location in the end is what I observed in the preliminary plan. So it came from that. So there are that many majority-minority districts in those areas that I observed in the preliminary plan.

REPRESENTATIVE BENNINGHOFF: Do you know how many of those majority-Black or majority-Hispanic districts that you asked in setup B the computer to draw were actually from Allegheny County?

DR. IMAI: Well, it should be in the report. Sorry about that.

In the appendix -- are you asking about Simulation B or Simulation A?

REPRESENTATIVE BENNINGHOFF: B, sir.

DR. IMAI: Oh, B. So B, yeah, sorry about that. I don't have everything at the top of my -- yeah, I guess I didn't give those particular breakdown. Yes, I have to -- yeah, I don't know that off the top of my head. Sorry. I'm sorry about that.

REPRESENTATIVE BENNINGHOFF: No, that's fine. I mean--

DR. IMAI: Yeah.

REPRESENTATIVE BENNINGHOFF: --in a State of 67 counties, I'm just curious, if we're doing modeling, that we're not just only looking at 1 county out of the 67. So--

DR. IMAI: Right.

REPRESENTATIVE BENNINGHOFF: --it would be helpful for the Commission to know whether any of those areas--

DR. IMAI: Okay.

REPRESENTATIVE BENNINGHOFF: --in Allegheny County would have been represented in--

DR. IMAI: Yeah. I can certainly provide that information.

REPRESENTATIVE BENNINGHOFF: One last quick question, if I could.

You state that race-blind analysis shows a smaller differential between the predicted number of seats in the simulation and that proposed by Dr. Barber's analysis.

DR. IMAI: Yes.

REPRESENTATIVE BENNINGHOFF: But all in all, we're still talking about outliers.

DR. IMAI: That's correct. So the outlier, but, you know, the substantive magnitude is smaller. So the outlier -- the term "outlier" is more like a statistical term, and the difference is the actual difference of the different number of districts.

REPRESENTATIVE BENNINGHOFF: But it also results in a greater number of Democrat districts.

DR. IMAI: Yes. But much smaller.

REPRESENTATIVE BENNINGHOFF: Okay.

DR. IMAI: The difference is smaller.

REPRESENTATIVE BENNINGHOFF: I appreciate your candor on that.

Thank you, Mr. Chairman.

CHAIR NORDENBERG: Leader McClinton.

REPRESENTATIVE McCLINTON: Thank you, Mr. Chairman, and thank you, Dr. Imai.

Is it your position that when a simulation model is constrained to create majority-minority districts in addition to the criteria that's in the Pennsylvania Constitution, such as in Simulation B, there is no statistically significant difference between simulated plans and the preliminary plan?

DR. IMAI: That's correct.

REPRESENTATIVE McCLINTON: Thank you.

Based on those constraints, there's no evidence that the preliminary House plan is, in fact, a gerrymander for partisan advantage, is that correct?

DR. IMAI: That's correct under those settings, under those additional considerations.

REPRESENTATIVE McCLINTON: Thank you, Dr. Imai.

Thank you for your indulgence, Mr. Chairman.

Dr. Imai, one more question, forgive me. Would you be able to respond to the criticism that we just heard from one of your colleagues, Dr. Barber?

DR. IMAI: Sure. First, I would like to, you know, welcome Professor Barber's use of simulation analysis. I've been developing this algorithm that I think is powerful and useful, so I appreciate that.

In terms of his criticism, he mentioned that my simulated plans are less compact than his own. He also mentioned that my simulated plans split more municipalities than his simulated plans. And I want to first emphasize that I included those constraints, as I explained, in all my simulations. So these are part of the constitutional constraints, so they would be more compact and avoiding, you know, these county and also municipality splits is important. And that's part of my algorithm.

Now, one thing I want to first point out is that the preliminary plan is actually very, very compact and avoids -- has fewer municipality plans than any of the simulated plans, whether it's Professor Barber's simulation plan or my own. So in that sense, the preliminary plan, I think, you can say is much more compliant with the constitutional requirements than, you know, than any of the simulated plans that were presented today.

Now, the question is, and this question, I don't know whether legislators care or not, in a simulated plan, why is it that my simulated plan is less compact than Professor Barber's? I cannot really understand that question unless I know exactly which algorithm he used to generate it and how. As I said, in the software package I developed which Professor Barber used, I implement several algorithms, and some algorithms are better suited for certain purposes, and the other algorithms should be used for other cases. So I would like to know what algorithm is being used and how it's used. My own use is based on all the criteria that I specified in my published articles, based on the statistical criteria of what the simulation algorithm is capable of doing.

And then a second point Professor Barber mentioned is that many of my results are still showing that the preliminary plan is a statistical outlier. However, I would like to note that the results are based on the three sets of particular elections that Professor Barber used. There could be other sets of elections that I could have used, but I chose those three sets of elections for the purpose of just a comparison. Because otherwise, you're really comparing apples and oranges, and it's unfair.

As I said in my presentation, the third, you know, right most figure that I presented, that's where the three -- you know, in all analyses, the preliminary plan is the outlier. So when Professor Barber characterizes my results as showing the preliminary plan is statistically an outlier in many of my simulation results, he is counting the last column. I included them because that was among the election sets he used, but as many of you probably agree, and most analytics agree, you want to average many elections so that you get rid of election-specific effects. If you used the 2020 election alone, whatever happened that year is going to greatly influence the results. For that reason, I used, you know, the 2012-2020 and the 2014-2020 election results Professor Barber chose are a more reliable measure, and based on that account under the additional consideration of majority-minority districts beyond the constitutional criteria, that the preliminary plan is not a partisan gerrymander.

Sorry, my answer got very long.

REPRESENTATIVE McCLINTON: No worries.

Thank you, Mr. Chairman. No further questions.

CHAIR NORDENBERG: Professor, just to go over some basic points.

DR. IMAI: Sure.

CHAIR NORDENBERG: You developed the redistricting simulation software that was used by Professor Barber?

DR. IMAI: That's correct. My colleagues and myself developed that package.

CHAIR NORDENBERG: But you were unable to replicate the results that were obtained by Professor Barber in using your software?

DR. IMAI: That's correct. But that's lack of sufficient information provided.

CHAIR NORDENBERG: And, in fact, when you did your own race-blind simulations, you found that his simulations substantially underestimated the likely number of Democratic districts?

DR. IMAI: That's correct. So I did my best to do race-blind analysis, and then that's the finding I obtained.

CHAIR NORDENBERG: You also found that when racial data is included, there is no evidence that the Commission plan is a partisan gerrymander?

DR. IMAI: That's correct.

CHAIR NORDENBERG: Majority Leader Benninghoff used a term that we have all heard - bad in, bad out. Let me modify that term slightly but in ways that I think more suit the circumstances, and that is nothing in, nothing out. If you put in nothing about racial data, then you're going to deviate more significantly from a plan that likely would be developed with those considerations in mind, is that correct?

DR. IMAI: That's correct. That's a possibility, right. So if you -- yeah, you phrased it better than I would have. Yeah.

CHAIR NORDENBERG: And you also indicated that "outlier" is a statistical term. Is "extreme partisan gerrymander" a statistical term?

DR. IMAI: No, not that I know of.

CHAIR NORDENBERG: Professor Barber made reference to the work of Professor Rodden in his presentation. I'd like to, if I can, read a passage from a case study of political geography and representation, a case study of the State of Pennsylvania done by Professor Rodden from Stanford and a coauthor. He begins by asking that we imagine a situation in which a Commissioner or a Special Master is told to choose from among the relevant neutral ensemble a plan for which the anticipated seat share of each party was 50 percent when the vote share was 50 percent. He goes on to say, at the scale of congressional districts or State Senate districts, the range of outcomes in the ensemble is sufficiently large that this could be achieved by selecting one of the most pro-Democratic plans. However, this becomes impossible as districts become smaller and more numerous. The range of outcomes is much narrower at the scale of Pennsylvania House districts, where even the most Democratic plan falls short. To be clear, the lesson is not that a fair plan with 203 districts cannot be drawn in Pennsylvania; rather, such a plan does not emerge from the neutral ensembles.

Do you agree with that statement?

DR. IMAI: Yeah. I respect Professor Rodden greatly. I will just refrain from, you know, saying anything. I'm not against him or anything. Just that, you know, it's an interesting passage from very authoritative figures, and I would just not rather say anything about whether I agree or disagree. Thank you.

CHAIR NORDENBERG: Thank you, very much.

Are there other questions?

Majority Leader Benninghoff.

REPRESENTATIVE BENNINGHOFF: With respect to the Chairman, I believe the record will show when I said about the computer data entry, I said information in, information out.

Not bad in, bad out. I hope that I am correct in that analysis.

DR. IMAI: Some people describe it as garbage in, garbage out.

REPRESENTATIVE BENNINGHOFF: I said neither because I was not trying to be derogative. I just want to say the generality that information in generally meant information out.

DR. IMAI: Very professional.

CHAIR NORDENBERG: Yeah. I learned it garbage in, garbage out, too, but I apologize if I misspoke in paraphrasing you.

Any other questions?

(There was no response.)

CHAIR NORDENBERG: If not, Professor, we thank you very much for your--

DR. IMAI: Thank you.

CHAIR NORDENBERG: --work and for being here today.

DR. IMAI: Thank you for the opportunity.

CHAIR NORDENBERG: I think at this the point I should ask our stenographer if she needs a break.

No, you're ready to go? Okay.

Our third witness this afternoon then is Dr. Matt Barreto. He is a Professor of Political Science and Chicana/o and Central American Studies, the founder of the Latino Policy & Politics Initiative and Voting Rights Project at UCLA. He is the President and Co-Founder of BSP Research, a research and polling firm. He had a previous faculty appointment at the University of Washington.

This is a return appearance for Professor Barreto. He was with us in the fall, and what I remember, Professor, is that it was World Series time and you had a banner for a California team on the wall behind you and expressed your condolences that neither Philadelphia nor Pittsburgh had moved that far. So we enjoyed your last appearance, and we're very glad to have you here again today. Welcome.

DR. BARRETO: Well, thank you, Mr. Chairman. It's a pleasure to be here, and we'll be rooting for the Dodgers again next year to make it a little bit farther. It didn't quite turn out how we anticipated in the post-season.

I'm going to go ahead and turn my screen share on, as the other panelists have done, and pull up some slides that I've prepared for today. If that works, you should see a blue slide up with my name.

CHAIR NORDENBERG: We do.

DR. BARRETO: Okay, great. Well, I'll just go ahead and jump into my analysis, giving only a very short introduction. My name is Matt Barreto. I'm a faculty member at UCLA, where I've been for about 7 years, was at the

University of Washington in Seattle for about 10 years before that, participated in numerous redistricting and voting rights trials, probably close to three dozen. I have published academic work explicitly on the topic of voting rights, racially polarized voting, and while not as proficient as Dr. Imai in his software publications, I've also written and published software that many other people use today to conduct racially polarized voting analysis.

What I'm going to talk about is changing demographics in voting patterns, and I'm going to start out by just doing a little bit of table setting and looking at the current landscape of racial demographics in the State of Pennsylvania.

This table indicates what the population looked like 10 years ago and what the population looked like in the 2020 Census, which, of course, is the data set which informs the redistricting process. Communities of color combined, led by a specially strong growth of the Latino community, but growth numbers for Asian Americans, African Americans, multi-racial populations dramatically outnumbered the close to half-million decline in the white population. And I use this as a starting point because these population numbers must be taken into account when thinking about how to redraw districts 10 years after the previous Census. Just based on the white population decline of 541,000, given the average district size, represents about a loss of 8.4 districts. Given the non-white growth of 841,000 combined, that represents a gain over the status quo of about 13 districts. So this represents this churn, this change represents about 10 percent of the total population of Pennsylvania, and thus we would be encouraged to consider something of around 20 seats moving and shifting their boundaries, given that much change happened in the State of Pennsylvania. It's something that we can't ignore.

In particular, I want to highlight a couple of important regions in the State where this change occurred. The first is in Allegheny County, which has large white non-Hispanic and African American populations. This chart shows within the county which Census tracks had Black population growth and which Census tracks had white population growth. What you can see is that the Black population for Allegheny County as a whole grew, and that this was strongest outside of the city of Pittsburgh, while the pockets where the white population grew happened more likely to be inside the city of Pittsburgh. What this means is that as we consider current existing performing districts, we need to take into account those population shifts, perhaps shift the boundaries around a little bit, in order to continue to create performing districts that abide by the Federal Voting Rights Act.

The second region that I want to focus on is in the central part of the State. This represents about a five-county region or six-county region where there has been very large Hispanic population growth. What you can see here is that in places like Allentown, Reading, Harrisburg, Lancaster, and others, there is an exceptionally large, if not majority, Latino population that has grown dramatically over the last few years. In fact, since 2000, the Latino population has grown from just over 100,000, tripling to about 309,000 today in this region, while the white population combined in this region has declined by about almost 50,000.

What this means is that the Latino population now is large enough to support districts, to support some majority districts, but to also be present in coalition districts and other voting rights compliant districts. And looking at the Census data, this population growth is expected to continue, easily continue for the next 10 years, because it is largely being driven by a younger U.S.-born population who has children coming of age and turning 18, 19, and 20 and becoming eligible voters. And so we would expect these population growth numbers to continue across these regions moving forward.

The third region I focussed on in southeastern Pennsylvania is that of Philadelphia and Delaware Counties. And here it is a very diverse environment today, one that decades ago used to be really only concentrated with large white and Black populations. Today, you can see areas that are shaded in green as being areas where the Latino population is the fastest growing and moving into majority status in these areas. White population growth has continued in this area, mostly in central Philadelphia city, and Black population growth has continued across the entire region, both in pockets of Philadelphia County but also strong population growth in Delaware County. And so as districts again are being considered, these changes across these counties need to be taken into account in order to continue to comply and comport with the Voting Rights Act.

I'll give a very brief overview of the importance of the Voting Rights Act, focusing on Section 2, and then I'm going to jump to some of my analysis and what my conclusions are. But let me just give a little bit of summary. The section of the Voting Rights Act, Section 2(b), that I am interested in, understanding and making sure that we comply with, states that a violation of this Federal act has occurred if members have less opportunity than other members of the electorate to participate in the political process and to elect representatives of choice. It is that second portion of the phrase, to elect representatives of choice, which I think is very important, and many prior court cases have decided it's considered vote dilution if it has been shown to be found that members of racial and ethnic minority groups are not able to elect candidates of their choice due to lines being moved and shaped to dilute their vote.

Specifically, Section 2 prohibits what I just said, racial gerrymandering to dilute minority rights and to have that meaningful opportunity to elect their candidates. This section has been used by racial and ethnic minorities of all different backgrounds, including in the State of Pennsylvania. And the two things that we are most concerned with is, do plans artificially pack and overconcentrate a single group, or do they do the opposite? Do they dilute and crack the group so that they're so small that they can't actually have meaningful influence to elect candidates of choice? These things are tricky to assess and understand, but those of us who have studied it closely and have been observing redistricting plans, you know, are able to find out areas where this packing or cracking occurs and to make sure that we steer clear. We always give advice to say, do not adopt plans that appear to be diluting minority voting opportunities.

There are two considerations that must be taken into account when understanding whether or not there has been this violation of the Voting Rights Act that come out of a famous court decision and are often referred to as the *Gingles* factors, or the *Gingles* test. The first is about coalition and performing districts. If a district is already performing for minority-preferred candidates, its population can change, but it must continue to perform for minority candidates. And so first we want to look at areas where there are performing districts and insure that a new plan that comes along does not dramatically dilute minority opportunities, but that plan might change, as I showed you in one of the earlier slides. There's been considerable change in where the white and Black population in Allegheny County has grown, and if there are performing districts, they need to be maintained.

Related to packing, the courts have regularly found that districts do not need to be supermajority Black or Hispanic. In fact, this can be considered overdoing it and overconcentrating the Black or Hispanic influence in only one district, when in fact two districts were likely to have been drawn. And Professor Barber hit upon one of these examples in his description of advocating for overly packing a Hispanic district so that it was concentrated in only one, not giving it influence in a second district. Courts have allowed, in addition, Black and Hispanic populations together, to be combined. They have recognized those in decisions as forming majority-minority coalitions, if it can be shown that the communities vote together.

So what did we find here in Pennsylvania? First of all, the extensive analysis that I conducted across the entire State, also focusing in on specific regions, demonstrates that minority voters in Pennsylvania are politically cohesive in supporting their candidates of choice. The majority voters, in this case white, usually vote together to defeat minority-preferred candidates, or at least block against minority-preferred candidates. To assess these voting patterns, I conducted court-required ecological inference analysis, often just referred to as EI analysis, using the software package that I co-developed with colleagues called eiCompare.

However, let me keep you on this idea of EI and ecological inference. We don't have to absolutely get that technical and statistical to understand voting patterns in Pennsylvania. Any pundits or analysts of Pennsylvania will not be surprised by the results of our analysis. There's nothing computer driven or technical. We can just look at the CNN exit polls of the last election, which stated that Black and Latino voters combined for an averaged vote of 84 percent for the Democrat to 13 percent for the Republican, and white voters statewide voted in the opposite direction, 57 percent for Mr. Trump, and only 42 percent for Mr. Biden. So even if we use just basic exit polling data that is quite familiar to all of us, we're going to see these same patterns that we demonstrate here in our statistical models.

I'm going to first just give you an illustration and example of how we plot this data and how we eventually arrive at what are called ecological estimates. So here is just a chart that has a Y axis that measures the percent of the vote that might be won by any candidate in a precinct. That goes from 0 to 100. On the X axis along the bottom, we can measure the percent of all the voters who are white or minority in a precinct, again going from 0 to 100. And then each dot that is going to appear on the screen, that's a precinct. So the plot that I'm going to populate for you right now, this is all of the votes, in this case it says, in red, in "Western Pennsylvania." As I started my preliminary analysis, I looked at the western part of the State just to begin. And you get something like this, looking at 2020 State House vote, this is the percent Democrat in the western part of Pennsylvania. And what you can see is that there's a clear pattern in the top left portion of your screen, precincts that have very few white voters are voting overwhelming cohesively Democrat.

As you move to the right side of the screen, you can fit a line to this, that is the best fit regression line, that is the average of the estimate, that shows a considerable downsloping once you get to majority white precincts around here (indicating), falling off dramatically as you get into more heavily white precincts. They say this is called the best fit regression line, and what it estimates for us is what the vote looked like by different racial groups. In this case, considering all of western Pennsylvania, it estimated that the white vote in State House elections was about 29 percent Democrat, and the non-white vote was about 91 percent Democrat, a 62-point gap between the two racial and ethnic communities in this part of the State.

Likewise, we can look at another election. Here is the percent vote for Trump among those same precincts, and you see the opposite pattern. In areas over here where there are very, very few white voters, there is almost no support for Trump. As you move to the right part of your screen, you start to see the line go up and crest at about 70-some percent in the ecological analysis. This is predicting about a 77-percent vote for Trump among whites, but only 11 percent for non-whites. So this basic type of analysis lets us not only illustrate the data to you, but we can also use the ecological inference software to create point estimates, which I have over there in the right. And we can do that not just for any statewide, any region that might be of interest.

So I looked at key regions across the State, and in the report there's many more tables and charts, but here is just a summary of what we found across key regions in the State. From the southwest, which centers Allegheny County, the central part of the State, Lehigh Valley, and then the southeast portion of the State, which includes Philadelphia and Delaware Counties. What you can see in these two examples that I gave you, whether we look at the percent voting Democrat for State House, or the percent voting Republican in the Presidential election, is that there is a dramatic difference between white and minority voters. Minority voters include all non-white voters in this case. And then I have breakout results for the regions that support it for both Black and Latino voters.

It's painfully obvious that there are large discrepancies between how minorities are voting, which are quite cohesively, in supporting strong votes for Democratic candidates of choice, and white voters, who show majority support against Democrats and for Republican candidates of choice in every region, including in the southeast. This can be illustrated in a couple of simple tables. Again, sometimes, you know, just looking at the point estimates, we always want to show our homework, and so I'm just going to show you a couple of similar plots of every single precinct and how they voted with that regression fit line for different regions of the State.

This is the 2020 State House vote in southwest Pennsylvania counties. You can see that in minority areas, extremely cohesive and concentrated. As you move to the right side of the screen, this falls off to falling well below the 50-percent mark. You can see most of the Democratic vote is falling down here below 30 percent even. If we look at another region of the State, in this case the Lehigh Valley, you see a very, very clear pattern. This is a very heavily Latino part of the State where the Democratic vote, again, in areas where there are very few white voters over here, is quite concentrated around 75, falls off sharply as you get into majority white areas, and you see a very large discrepancy where majority white areas in Lehigh are voting very heavily, in many cases well above 75 percent. And also in the central region of the State of Pennsylvania we see the same pattern here, where white precincts are extremely heavily concentrated in central Pennsylvania, even more so, in some cases well above 85 into 90 percent, whereas minority areas over here where there is both a large Black and Hispanic population across these counties are all almost entirely situated well above 75 percent. So we see this not only in the tables and the estimates that come out, but when we actually look at the real data and plot it, the same pattern emerges.

So let me conclude with a summary of what I found when looking at these plans. The first is that the voting analysis is clear. There is a strong finding of racially polarized voting across the State as a whole. In pockets of the State, there are enough white crossover voters, a phrase that the voting rights court decisions have used, to support minority group's candidates of choice in coalition to sustain those minority-performing districts. When looking at the current map, the multiple Black- and Hispanic-performing districts are packed and they exhibit wasted minority votes. Those are areas that do not have to be 80-percent Hispanic to perform for Hispanic candidates of choice. That is overconcentrating those folks in a single district and not giving them an opportunity to have influence in an adjacent district, something that the courts have regularly recognized.

Given that growth of the minority population, it is clear that these existing minority districts can and should be unpacked so that new minority-performing districts are created to comply with the Voting Rights Act. As I started in my presentation, the demographics of the State have changed dramatically over the last 10 years, and the plan should reflect that.

Finally, I conclude with looking at a couple of key districts that I know are of particular interest and looking at how they might perform. Starting with districts that have large and noticeable African American populations, I've created a table here comparing the current minority voting age population, and then what is in the preliminary plan, and how those things have changed over time. And you can see in some instances there are changes to the minority voting age population. In some cases, the African American population is adjusted. However, those districts are, in modeled analysis, looking at real election results, expected to continue to perform quite well for minority candidates of choice. Likewise, when we look at districts in other parts of the State, many of these include heavily Hispanic populations, even in areas where the minority population is downsized a little bit, such as the 22nd, these are areas that are expected to continue to perform easily for minority candidates of choice. Again, evidence that we do not need to overly pack these districts. We need to find the area that complies with the Voting Rights Act and allows minorities to not have their vote diluted. Across many of these tables, our analysis concludes that these will continue to perform for minority candidates of choice.

Thank you. That concludes my slides. I'm happy to take any questions.

CHAIR NORDENBERG: Thank you, Professor Barreto.

Are there questions from Members of the Commission?

Majority Leader Ward. I haven't been able to announce your name since this morning. Thanks for giving me this chance.

SENATOR K. WARD: Thank you very much for your participation, Dr. Barreto.

I have just two simple questions. We talked a little bit about Allegheny County. Do you know what the percentage of the Black votes in Allegheny County are?

DR. BARRETO: The overall percentage of Black voters?

SENATOR K. WARD: Right. Because we talked about the growth of the African American population in Allegheny County. What numbers do you have?

DR. BARRETO: I don't have the raw population totals in front of me. I know that in Allegheny County as a whole, the Black population grew over the past decade, but the growth was mostly concentrated outside of the city.

SENATOR K. WARD: Outside of the city?

DR. BARRETO: The growth.

SENATOR K. WARD: When you were talking about pack or crack, what number are you looking at? What percentage do you consider packing?

DR. BARRETO: So the packing analysis asks whether or not minority votes are being wasted, meaning could a second majority-minority coalition or influence district be drawn? In some cases it can't, and an 80-percent district might be the only option. And so what you have to look at when you assess packing is the adjacent population around that district. Is there enough of a minority population to continue to create at least a minority-performing district in the first place? So to do no harm to that district. But to unpack it in a way that allows minorities to vote in coalition with others to still maintain influence. So there's no bright line. There's no exact number. It goes by each location. I think in some of the voting Rights decisions they say there needs to be an intensely local appraisal.

SENATOR K. WARD: So when you're talking about Latino votes, Black votes, not combined, what is a good number? Is 50 percent too high for a minority or a Latino? Is that too high?

DR. BARRETO: Well, it depends on the circumstances of the district. So if you're trying to draw a majority-minority district, where the group is a majority, then you often are looking at 50 percent. However, if the population is large enough and there are coalition partners, it might be the case that a 40-percent district performs well for minority candidates of choice, but you have to know if there are coalition partners or if there are people block voting, as I showed in some of those charts. Who are the other people that you're going to group with the minority voters? Are you grouping them with people who will block their influence? Are you grouping them with enough people who will vote in coalition?

SENATOR K. WARD: Thank you.

CHAIR NORDENBERG: Other questions from the Commission?

Majority Leader Benninghoff.

REPRESENTATIVE BENNINGHOFF: Thank you, Mr. Chairman.

Dr. Barreto, it's a lot of information. Very well done. I was actually sitting here thinking earlier, how did they do this stuff 30 years ago without all the computers to assist in that? I think your job would have been a lot more difficult.

DR. BARRETO: Absolutely.

REPRESENTATIVE BENNINGHOFF: I've got a lot of pages kind of earmarked here with my fingers. So bear with me as I try to put my thoughts together for the questions I want to ask. Watching the slides, I saw a report a little bit earlier, and you provide a few separate estimates on these cohesions for the Black voters and Latino voters and the Asian voters. I've got a couple of questions along those lines.

Is there any kind of, I think they call them confidence intervals that kind of substantiate the reliability of that data? How are we to determine that?

DR. BARRETO: Yes. I mean, in any estimate you can ask to look at the confidence intervals to determine what is the lower and upper bound of the estimate. In this case, because we showed the actual real precinct results, those have less error in them. They're just the real results. So when you see those groupings all clustered in an area, it gives you much, much more confidence in the estimate.

REPRESENTATIVE BENNINGHOFF: Are they in the report here somewhere where I missed it? And I apologize if I did. I'm briefing it quick.

DR. BARRETO: No. The confidence intervals for the ecological inference estimates are not included in this.

REPRESENTATIVE BENNINGHOFF: Would they routinely be? And again, I'm a little more of a novice to this, so I'm just trying to substantiate all this data and these dots and graphs. Is that something--

DR. BARRETO: What's the question?

REPRESENTATIVE BENNINGHOFF: Is that normally something that would be in a report like this? Or is that something I just need to get subsequently?

DR. BARRETO: Well, it depends on what the task of the report is. As I said, in this case, we presented a lot of actual raw data that are not modeled estimates. For instance, the precinct charts that are in the report, those don't have any error, necessarily. They're just the real actual results of the elections. But depending on the question that someone was asking, if you were publishing an academic paper, you might be asked to put in those confidence intervals.

REPRESENTATIVE BENNINGHOFF: Right. To follow up a little bit on what Senator Ward asked, I think you talked about being able to separately estimate cohesions for the Black and Latino voters. My question would be, then why lump them together in this analysis?

DR. BARRETO: Well, we attempted to do both. I think that I provided estimates for both. On the one hand, I look at the overall minority population, so areas where there are very few white voters at all, to see if there is a cohesive overall minority community. And then also in here I provided estimates for both Blacks and Latinos individually, as you saw on one of those slides. And so depending on the group of interest, there are certainly some regions of the State that are very heavily Black or very heavily Latino, but then there are other regions in central and also in southeast Pennsylvania that have a lot of Black and Latinos living side by side. So I've attempted to provide all of that information.

REPRESENTATIVE BENNINGHOFF: And you talked about some voting data. Does that include primaries as well?

DR. BARRETO: It can include primaries. Of course, again, it depends on the questions that are being asked. In this report, we examined a great number of general elections.

REPRESENTATIVE BENNINGHOFF: Only general elections?

DR. BARRETO: In this analysis, yes.

REPRESENTATIVE BENNINGHOFF: Only because sometimes who turns out is a lot different in those.

You had mentioned a couple districts. I think you mentioned House Districts 22 as where minority candidates have an opportunity to be elected. I was just curious, when you did your analysis, were you aware in the 2020 elections they did have a Hispanic candidate? She actually testified before this committee and did a great job. She lost that Democrat primary by a very slim margin, I think it was 50 or 55 votes, to a white candidate. When I look at the preliminary LRC map, much of that Hispanic base was taken away or redrawn, if you want to call it, to House District 134. That's one of the reasons I want to know, because if you don't consider primaries in this study, how would it be factored into the racial analysis what ended up happening there?

DR. BARRETO: Well, District 22 continues to have a very large Hispanic population, and according to our analysis of candidates of choice, will continue to perform for Latino candidates of choice. That's a topic I have studied and published on extensively for 20 years, and I looked at District 22 very carefully, and it continues to have a very large Hispanic population and voting population. But District 134 also is then allowed to have a Hispanic influence. By overconcentrating Hispanics in one district, that is what courts have called wasted votes. And so in this case, I think 22 is a very strong minority-performing district.

REPRESENTATIVE BENNINGHOFF: Is it as much of a

Hispanic district as it was when Ms. Santiago lost by 55 votes, in your analysis?

DR. BARRETO: It is performing for Hispanic interests at equivalent rates. It will continue to elect Hispanic candidates of choice.

REPRESENTATIVE BENNINGHOFF: How about eligibility of age-appropriate voters of Hispanic/Latino voters? Has that number increased or decreased?

DR. BARRETO: Well, as Dr. Barber noted, some of the districts did have decreases in populations, but that was because the population has grown so much over the decade that it approached what we call packing, that it was over-concentration, and you don't have to have over-concentration of voters in order for that group to have influence. And so--

REPRESENTATIVE BENNINGHOFF: I'm well aware of that, and I appreciate your candor, but I think it's important in this specific example, especially when you talked about House District 22, that this woman only loses by 55 votes and there's a reduction in the amount of eligible Hispanic voters subsequent of that election cycle. Yes, they draw another district, House District 134, but if the goal was to try to give different types of minority populations the opportunity to get elected and have representation in our Commonwealth, this seems counterintuitive to do that, because earlier in your comments you talked about splitting up these cities. And I just, you know, as a voter, I guess I'd ask you to look at that not just as a scientist, but as a voter, would that make you suspect, if somebody loses an election by 55 votes to an incumbent white candidate, they have probably potential and greater notoriety in the next cycle, when all of a sudden, whoosh, they're cut out of it, and the number of eligible Hispanic voters has been reduced at her next opportunity to run. As a voter, would that give you a moment for pause?

DR. BARRETO: No. As I said, I looked closely at this district. It continues to have a very large Hispanic voting age population, and in coalition with other minority partners, will continue to elect Hispanic candidates of choice. There's no question.

REPRESENTATIVE BENNINGHOFF: Well, I appreciate that, and we'll let that go. I just have personal concerns on that. We can look at that maybe when we reevaluate the maps.

I'm curious, as you looked in your report in the regions that you selected, in the elections of 2020, how many of these cases did minority-preferred candidates lose elections? And if you know that, where was that? Because I think that's important in history.

DR. BARRETO: I don't understand what you're specifically getting at. The number--

REPRESENTATIVE BENNINGHOFF: In any of the 2020 elections, if there was elections with minority candidates in those things, or preferred candidates, I should say, how many of them would have lost their election?

DR. BARRETO: Let me clarify that first point, which I think you just helped for the record, that we're not just looking at whether the candidate themselves is a minority but whether or not it's the minority voting population's candidate of choice.

REPRESENTATIVE BENNINGHOFF: Correct.

DR. BARRETO: In some cases those can be white candidates. But there were a number of elections across the Commonwealth in 2020 for the State legislature where minority-preferred candidates lost.

REPRESENTATIVE BENNINGHOFF: Well, I ask that for the reason because I feel we've had a pretty significant amount of people from within the Latino community and their preferred candidates are generally Latino candidates, they'd like to see more of them elected, and I think that we will do them a disservice if we don't reevaluate some of these districts that numerically are less than what they were before.

I just wanted, to real quick on the whole *Gingles* analysis you talked about, you gave a little background about the specifics of that, and you have to make sure, if I understand what you said, that they show that there's racially polarizing voting in that specific area. You have to show in some case where white voters keep the minority group from electing a preferred candidate. I'm just curious, I assume it also has to be within specific districts. Do you have examples of where in the State that you might contend that the Black majority-minority districts are required to give Black voters an equal opportunity to elect representatives of their choice? And if so, on what basis do we do that?

DR. BARRETO: I think any of the existing Black-majority districts or Black-performing districts, the Voting Rights Act would ask for those to be maintained and not to decrease the influence. And so where we have districts that are performing for Black communities, we would want to see those districts maintained so that the analysis suggests that they continue to perform, whether it's in Allegheny or Philadelphia.

REPRESENTATIVE BENNINGHOFF: Are those the only two regions that you see that happening?

DR. BARRETO: Well, wherever there are large populations of African American voters. There are also some pockets in the central part of the State and across the State. The analysis suggests from a voting rights perspective, are there districts where African American voters are able to elect candidates of choice? If so, those districts should be maintained.

REPRESENTATIVE BENNINGHOFF: And I appreciate that. And just a follow-up to that real quick. In the same vein, where in the State do you contend that Latino majority-minority districts are now required to give Latino voters an equal opportunity to elect representatives of their choice?

DR. BARRETO: Well, I don't have the exact geography in front of me right now, but there are opportunities to draw Latino-performing districts in the Lehigh Valley, in the central part of the State, and in southeast Pennsylvania as well.

REPRESENTATIVE BENNINGHOFF: Do you think that's reflected in the map?

DR. BARRETO: I do. I think it has expanded, dramatically expanded opportunities for Latino representation in this House map.

REPRESENTATIVE BENNINGHOFF: Even though the overall eligible voting Latino of age numbers went down?

DR. BARRETO: What you're referring to is only one single specific district--

REPRESENTATIVE BENNINGHOFF: I am not.

DR. BARRETO: --which is still a majority Hispanic district, and generally the map has higher Latino percentages in many districts, which will allow them to have more influence in electing candidates of choice.

REPRESENTATIVE BENNINGHOFF: And if I could, just so your explanation of a majority Latino community, is that a blend of multiple races or truly purely majority of Latino voters, or eligible voters?

DR. BARRETO: Well, Hispanic is an ethnicity, according to the Census, and those folks can mark any race that they might prefer. I'm not sure I understand exactly what your question is.

REPRESENTATIVE BENNINGHOFF: My question was whether or not there is a majority of Latino eligible age voters in order to try to get a Latino elected versus are we using the word "minority" more universal with multiple races in order to make the numbers say that this is a majority-minority district?

DR. BARRETO: Well, I think both of those things are true. There are a number of examples of majority-Latino districts that are over 50 percent that will perform for Latino candidates of choice. In other areas, there's a large Latino population, perhaps 38- or 42-percent, to where it is the largest group, but when combined with coalition partners crosses over the 50-percent threshold. So Latinos would still be the largest voting bloc and have other voters within that district who support some of their candidates of choice. So that's why we look at the performance of the district, not necessarily a specific magic number threshold.

REPRESENTATIVE BENNINGHOFF: I appreciate your explaining some of your report more thoroughly.

Thank you, Mr. Chairman.

CHAIR NORDENBERG: Other questions?

Leader McClinton.

REPRESENTATIVE McCLINTON: Thank you, Mr. Chairman.

Dr. Barreto, earlier Dr. Barber testified about simulations he did that ignore majority-minority seats in the Commonwealth. Does the Voting Rights Act permit us to ignore majority-minority seats?

DR. BARRETO: Absolutely not. And I believe Dr. Barber used the phrase in his testimony this morning, a Federal directive. You know, my understanding is that that's exactly what should have been entered into the simulations is that we cannot ignore the Voting Rights Act in drawing a districting plan.

REPRESENTATIVE McCLINTON: Thank you.

And when we talk about a minority candidate of choice, does that mean the candidate, himself or herself, must be a minority? Can you explain what specifically that means?

DR. BARRETO: Sure, I'd be glad to. The Voting Rights Act itself, and the subsequent court decisions, have always used the phrase "to elect a representative of choice." It has never established that that representative has to be

Black, Hispanic, Asian, or white. And so what we do is we look at the data, and we let the data tell us who are minority voters supporting? And oftentimes that is another minority, but in many other examples there are white candidates who stand up for minority issues and win an overwhelming percent of the minority vote. So "candidate of choice" or "representative of choice" just simply means who is winning over that minority vote.

REPRESENTATIVE McCLINTON: Thank you.

No further questions.

CHAIR NORDENBERG: In the district that you were discussing with Majority Leader Benninghoff, if I understood correctly, the candidate lost by 55 votes to a white incumbent. Many of the new districts that have been created are districts without an incumbent. Do you have a view as to the extent to which that is an enhancement of the opportunities that exist?

DR. BARRETO: Thank you for the question, Mr. Chairman. There's no question. In fact, I'm just looking at my notes I've scribbled down here. Many of these new Latino-opportunity districts are vacant, meaning there is no incumbent in the Lehigh Valley and Lancaster and in other areas. Incumbents are very difficult to defeat, especially in a primary. So moving these to vacant will create more opportunities for minority-preferred candidates to emerge and run and to win that seat.

CHAIR NORDENBERG: Thank you.

Are there any other questions?

(There was no response.)

CHAIR NORDENBERG: Thank you, Professor Barreto, for your encore performance. It's nice to have you back again with us.

DR. BARRETO: My pleasure. Thank you.

CHAIR NORDENBERG: Ann-Marie, are you still okay? All right.

Our fourth expert for the day has now appeared on the screen. He is Professor Christopher Warshaw, who is an Associate Professor of Political Science at George Washington University. He had a previous faculty appointment at MIT.

Thank you for being with us, Professor Warshaw. The floor is yours.

DR. WARSHAW: Great. Well, I'm honored to be here, Mr. Chair and Members of the Commission. I actually grew up in central Pennsylvania, and I did an internship in Pennsylvania State Government under Governor Ridge's administration while I was in college. So I've followed Pennsylvania politics since then and am deeply honored to be here before you.

So my graduate Ph.D. in political science is from Stanford University. I'm just going to talk a little bit about my background. I also have a law degree from Stanford as well. As the Chair mentioned, I held a faculty position at the Massachusetts Institute of Technology in Cambridge, Massachusetts. I'm currently an Associate Professor of Political Science at George Washington University in Washington, D.C.

So my research focuses on political representation, redistricting, elections, and public opinion. In all, I've written 24 peer-reviewed articles, including two peer-reviewed articles that focus entirely on the consequences of redistricting. I also have a book coming out this summer that's peer-reviewed called *Dynamic Democracy: Public Opinion Elections and Policymaking in the American States*. And one of the central focuses of my book manuscript is on redistricting and gerrymandering in the American States over the last 75 years. So much of my research looks holistically across the redistricting process in all of the States over the last five, six, seven decades to try to assess its causes and consequences. I've also worked recently on several redistricting court cases, including I testified in 2017 in Pennsylvania on a case regarding its congressional districting plan, and my analysis was heavily cited by the Pennsylvania Supreme Court in its final decision that led to the plan being declared a partisan gerrymander.

So what I want to talk to you today about is, first of all, how political scientists think about measuring partisan bias in a districting plan. And the first challenge in measuring partisan bias in a districting plan is, of course, that we have to know what's going to happen on that plan. Now, on a plan that -- if elections have already happened on the plan, then we could use those elections. But on a plan like this, where no elections have actually occurred on this plan yet, we have to have some methodology for projecting how future elections are likely to look. And, of course, there's no one single approach that is guaranteed to work well there. So, in fact, I used three different approaches, which include the approach that Professors Barber and Imai used, as well as other approaches to try to estimate how future elections would look on this plan. And I'll talk about that more in depth in a second.

The second thing I do is once we have in hand the results of how future elections might look on the proposed plan, then we're in a position to evaluate the partisan fairness of a plan. And, once again, there's no one way to measure partisan fairness, but in academic literature, scholars have developed a number of generally accepted metrics. So I'll show you the results of the four most commonly used metrics in the academic literature, which I'll also talk about more in a minute.

I want to just give you a preview of the results that I'm going to show you. Based on three different techniques for projecting future elections and four different metrics for partisan fairness, all of my analyses indicate the plan is fair, with perhaps just a small pro-Republican bias.

So the first thing we have to do in order to evaluate the fairness of this plan, as I mentioned a second ago, is we have to project future elections. And there's no way to know with certainty what's going to happen on future elections on this map. So, as I said, I used a set of different approaches. So the first approach I used, which is very similar to what Professor Barber did and Professor Imai did, is that I used a composite of statewide elections from 2014 to 2020. So this composite includes all 12 of the major statewide and constitutional offices, which I list on my slide, including the Presidential elections, the Governor's elections, and so on.

Now, when I have those elections in hand, there's a number of different ways that we could average across those races to develop some sort of a composite. And I think there's no one way to do this. What I do in my report is that I average results first within year and then across years, because I think that part of what the composite is trying to capture is how might elections vary over the range of a redistricting cycle? And so in the case of Pennsylvania, what that means is that Democrats did reasonably well in these statewide elections, and probably better than they've done in Presidential elections, and certainly better than they've done in State House elections. So it's important to keep that in mind when I go through the analyses to follow.

So across all of these races, Democrats won about 54 percent of the statewide vote the way I averaged them. And I did mention in my report, if I do the averaging approach that Dr. Barber does, then I get about 52 percent Democratic vote share. But my substantive results, it doesn't affect any of my results.

So the second approach that I used to try to project future elections on this plan is the actual 2020 State House elections. So these have the advantage of being the office that we actually care about when we're trying to evaluate this plan. Indeed, what we know from previous research is the best predictor of future legislative elections is past legislative elections. So the 2020 elections are likely to be a really good predictor of what's going to happen in the future on this plan. But there are some downsides, such as the fact that there's no way to know for certain which candidates will run in 2022, and, of course, the State House elections, as Professor Barreto talked about, are somewhat affected by the incumbency advantage and other candidate-specific factors.

Another challenge is that there are some uncontested State House elections in Pennsylvania. So for purposes of my analysis, I impute the uncontested elections based on the 2020 Presidential election results. In my view as a scholar of elections, Presidential elections are generally a very good predictor of State House elections. For instance, we know that nationwide the correlation between the Presidential results and State legislative results is greater than .9 across all States.

And then finally, I use a website called PlanScore.org, which is an open-source website that I'm on the social science advisory board for. And PlanScore uses a statistical model that tries to predict future elections based on the relationship between Presidential election results and legislative election results around the country over the past decade. And the advantage of PlanScore is that, unlike the other two approaches, it produces a probabilistic analysis that fully takes into account uncertainty in the results, so that even in a district where one party might be favored to win, PlanScore would essentially assign them a probability that they would win. So even if a party is unlikely to win a district, perhaps they'd still have a 15- or 20-percent chance of winning, and that would be fully taken into account in all of the downstream analyses.

So once we have these projections of future elections in hand, the next part of my analysis uses four different generally accepted academic approaches to evaluate the partisan fairness of the plan. And when we talk about the partisan fairness, what political scientists and social scientists in general have in mind is they're thinking about the representational process. And, of course, in order for voters to translate their preferences into policy, it has to translate into seats in the legislature. And, of course, we'd all agree that if the majority of voters voted for one party and that party won only a minority of the seats, that that would be an unfair plan. I think this was the kind of intuition that Chairman Nordenberg was alluding to in his remarks a week or two ago. And I think that would be the consensus of political scientists, that in a democracy, the party that wins the majority of the votes should win a majority of the seats. And that's the kind of intuition that these metrics are trying to capture.

I do want to note that none of these metrics hinge on any idea of proportional representation, unlike which I think the way Professor Barber characterized them wasn't quite accurate. And instead, they're all focusing on the relationship between votes and seats and whether one party has an advantage compared to the other party in that relationship.

So I'll talk about four specific metrics, and I think it's important to get a little bit under the hood to understand these metrics, since I'm going to use them to evaluate the proposed plan. So the first metric that I used is perhaps the oldest in political science literature, and perhaps the most widely used, which is the partisan symmetry metric. So the intuition behind the partisan symmetry metric is that if one party gets 60 percent of the seats with 55 percent of the votes, then so too the other party should get 60 percent of the seats with 55 percent of the votes. In other words, the relationship between votes and seats should be symmetric between the parties. Moreover, when one party gets 50 percent of the vote, they should get about half the seats. And if you can win more than half the seats with half the votes, then that indicates a politically biased plan.

So to illustrate the symmetry metric, and I'll do this throughout my illustrious slides for the next few metrics, is that I'm going to show you the gerrymandered 2016 congressional elections on the left-hand panel to just sort of illustrate what a politically biased plan would look like, and then on the right I'm going to show you the proposed Legislative Reapportionment Commission's plan using the reaggregated 2020 State House results. So as I mentioned, you know, there's a number of ways we could project future elections, and these are not meant to be the only way we can do that. So here they're really meant to be illustrative rather than being dispositive, and I'll go through my full analysis of the proposed plan in more detail in a couple of minutes.

So here we can see that the congressional election was entirely asymmetric in that the Republicans won 72 percent of the seats, basically regardless of how many votes they won. Across almost the entire range of the vote distribution between 45 and 55 percent, Republicans won about 72 percent of the seats. In contrast, on the proposed plan, if Democrats get 45 percent of the votes, they get about 45 percent of the seats, and if Republicans get about 55 percent of the vote, they too get about 55 percent of the seats. So the proposed plan looks extremely symmetric, especially compared to the gerrymandered congressional plan from 2016. And one thing I should emphasize, by the way, is the symmetry metric is using counterfactual uniform swings in the vote distribution, and that's how I evaluate this, which, as I talk about in my report, I think has some strengths and some weaknesses.

So the next metric that I used to evaluate the proposed plan is called the mean-median difference. So the intuition behind the mean-median difference is that if a party has an advantage on a map, then the distribution of votes and seats will be essentially skewed to their advantage. And we can see that if they have a greater vote share in the median district than in the average district. And that essentially enables them to translate their votes into seats more efficiently than the disadvantaged party. So in the 2016 congressional map, the Republicans had about a 7.5-percent advantage in terms of the mean-median difference, which meant that they did about 7.5 percent better in the median seat than in the average seat. And we can see that on the left-hand plot here, where you can see the line representing the Democrats' median vote share is far below how they did in the average district. So this indicates the Republicans had a really large and substantial advantage in the translation of votes to seats.

In contrast, on the proposed plan, there's a very small difference between the median district and mean district. It's around 2 or 2.5 percentage points here. I'm going to say very small, a small advantage. And so in this case, this indicates that this plan is reasonably neutral, or perhaps a small pro-Republican advantage.

For the next metric I want to talk through is the efficiency gap. And this is the newer metric, and the intuition behind the efficiency gap is really that one way to think about whether a party has an advantage is whether they're able to translate their votes into seats more efficiently than the other party. And, in fact, in order to maximize your seats share, what a party is really trying to do is to translate their votes into seats as efficiently as possible. So we can see this in this hypothetically example where Democrats win slightly more votes, but Republicans win two out of three seats. In this hypothetical example, it's easy to see that there's far more wasted Democratic votes in the sense that their votes don't actually translate into any seats in the legislature in Districts 2 and 3 than there are in District 1 for the Republicans. And, in fact, in this hypothetical example, there's about a 20-percent efficiency gap advantage for the Republicans, which is almost exactly what we saw in the actual 2016 congressional election, where Republicans had about a 19-percent advantage in the efficiency gap, which was one of the largest in recorded history for a congressional plan.

In contrast, on the proposed LRC plan, based on the reaggregated 2020 State House elections, the efficiency gap is almost -- essentially is very close to zero percent. So the proposed plan looks extremely neutral using the efficiency gap in the reaggregated 2020 State House elections, especially compared to the gerrymandered 2016 congressional plan.

And the fourth academic metric that I'm going to show you today is called the declination. And this is another newer metric, and the idea behind the declination is it's trying to mathematically capture packing and cracking. And what you typically see when voters of one party are packed together, which of course harms their prospects to win a majority in the legislature, and we see this on the 2016 congressional plan, where the slope of this line between the average seat and the average vote share for that party is much steeper than it is for the advantaged party. So the Republicans in this case, in the 2016 congressional election, were again able to translate their votes into seats much more efficiently than the Democrats, and that's what this declination here is showing.

In contrast, on the LRC's plan, the declination is essentially balanced between the parties, so there's a declination of almost zero percent using the reaggregated State House votes. And we can see that because the slope of these lines is almost identically equal to each other, which indicates a neutral declination metric.

The last thing I'm going to tell you about today is the responsiveness of a plan. And in a democracy, if one party wins more votes, they should also win more seats in the legislature. And this is a central metric that academics use to evaluate redistricting plans. So what we see is that, again, the 2016 congressional plan was totally unresponsive to shifts in voter preferences. The purple area on my graph, or the blue area, shows the actual range of votes that the two parties received over the 2014-20 period. And you can see within the range of statewide vote shares the two parties actually received, Republicans would win 72 percent of the congressional delegation across that entire range. In contrast, on the LRC plan, both parties could plausibly win the majority of the seats, and if the votes are split 50-50 between the parties, then so too they would split the legislative seats about 50-50. So in contrast to the congressional plan, the LRC plan looks extremely responsive in this analysis.

Okay. So, finally, now I want to tell you about the analysis that I've done of the proposed plan. And just to remind us, I'm going to use the four metrics that I talked through a second ago, and I'll use the three different approaches for projecting elections that I discussed before that.

So the first way that we project statewide elections is based on the composite of previous statewide elections from 2014-20. And recall that during this period, Democrats did very well statewide. So Republicans got about 46 percent of the statewide vote share during this period, but on the 2014-20 plan, they would have gotten almost half the seats. So this plan had a large and profound bias in favor of Republicans. In contrast on the proposed plan, the Republicans would get 46 percent of the votes and they'd get about 46 percent of the seats.

Now, as I mentioned earlier, none of the metrics that I used require proportional representation. And, in fact, all of the metrics typically assume -- or most of the metrics typically assume some sort of a winner's bonus, where the party that gets more than the majority of the votes usually does a little bit better in the legislature. This relates to the concept that Professor Barber was mentioning about single-member district elections. So typically, if you get 55 percent of the votes, you'd probably get about 60 percent of the seats in the legislature. And on the proposed plan, there is no winner's bonus for Democrats. So they get the same seat share as they do share of the votes. This actually indicates a small pro-Republican bias in the plan.

And we see that across all of the different metrics that I use. So the symmetry metric indicates that if the two parties split the vote 50-50, Republicans would actually get about 52.5 percent of the seats on the proposed plan, based on this composite of past elections. And so, too, in the other metrics, all of them indicate a small pro-Republican advantage. In fact, when we average across all of those metrics, the plan is more pro-Republican, at about 65 percent of previous plans over the last 50 years.

And one of the things I've done in my academic research that actually estimated all of these partisan bias metrics for every congressional and State legislative election and plan since 1972, which 1972 is of course important because that's the first election cycle where every State drew equal populous districts as a result of the Supreme Court decisions in *Baker v. Carr* and the ones that followed that.

So overall, the preliminary plan is relatively neutral, with perhaps a small pro-Republican bias, based on the composite of statewide elections.

Second, based on the 2020 State House results, here I find again that using the actual 2020 results, Republicans got about half the votes in State House elections, but they got 56 percent of the seats, of course. And so this leads to a very large pro-Republican advantage in all of the partisan bias metrics that I estimate. And here, the previous plan was more pro-Republican by about 81 percent of previous plans over the past 50 years.

In contrast, the proposed plan looks extremely neutral, using all of the metrics that I evaluate. The symmetry bias is almost zero. The mean-median, there's still a small pro-Republican mean-median difference, but that's also relatively small, especially compared to the prior plan. The efficiency gap is close to zero, and declination is also close to zero. So overall, I think based on the actual 2020 State House elections, the plan looks extremely politically neutral.

And then finally, I used the PlanScore.org website, which I know it has been mentioned a little bit in prior Commission hearings. So to recall what PlanScore does, it's a statistical model that estimates legislative election results based on all of the legislative election results and

Presidential results around the country over the past decade. And so PlanScore estimates that this plan would have a small pro-Republican advantage really across all four metrics, but you can see on the graphs that that advantage is relatively small. So it would have about a 2.5-percent pro-Republican efficiency gap, which is relatively close to the center of the distribution of prior plans. It is a pro-Republican advantage, but I'd say that's, you know, relatively modest in size, and so, too, on the other metrics. All of them lean a little bit pro-Republican that are close to the center of distribution of prior plans. That's especially true, for instance, of the declination and mean-median metrics. So overall, the PlanScore.org analysis indicates that the proposed plan is relatively neutral, with a small pro-Republican bias.

So next I want to tell you about responsiveness of the proposed plan. So again, as I talked about earlier, my findings indicate that the proposed plan is responsive to shifts in the mass public's preferences, and indeed, the party that gets the majority of the votes would usually get a majority of the seats on this plan, which is distinctly different from the 2014-20 plan, where Republicans always won a majority of seats across the entire range of vote distribution they actually received. And we can see this, for instance, in the 2018 election, where Democrats clearly won a majority of the votes statewide, while Republicans still won a majority of the seats. So too, in 2020, Republicans very narrowly won the statewide vote, but they won a large majority of the seats on the prior plan. In contrast, on the proposed plan, the plan is responsive to shifts in voter preferences, and again, the party that gets the majority of the votes usually gets the majority of the seats.

And the last thing I want to tell you about is the number of competitive districts on this plan, because I know this is something that is, you know, important to many in the public and on the Commission. So there's been some analysis done by the non-partisan Princeton Gerrymandering Project that gave the proposed plan a poor rating in terms of competitiveness. But in my view, that analysis was drawn on just a handful of elections, and it used a very narrow metric for competitiveness, and the elections that it chose actually were like basic -- two of the three were Democratic landslides. So in my view, the better way to look at competitiveness is to use a wider set, as Professor Imai talked about, is to use a wider set of comparison metrics. So here I used the three different ways of projecting future elections that I use elsewhere in my report, then I also use a number of different ways of tallying the number of competitive districts, because, of course, there's no way to know with certainty how many competitive districts there will be on the proposed plan.

And overall what I find is that there's about the same number of competitive districts on the proposed plan as there were on the 2014-20 plan. Moreover, there's about the same number of competitive districts that you see nationwide across other State House elections in 2020. This is based, again, you know, on not just one mode of analysis, but multiple, especially for the comparison between this plan and the prior plan, used on, you know, a wide set of different comparison elections, comparison metrics, ways of estimating the number of competitive districts, you know, overall the number of competitive districts is similar between the two plans. So in my view, the LRC plan is competitive with roughly the same number of competitive seats we've seen previously in Pennsylvania.

So in conclusion, just to briefly sum up some of my findings. You know, I find that this plan is likely to be responsive to shifts in voter preferences, which I think is important from a point of view of democratic theory, that we want to make sure that elected officials are paying attention to what their constituents want, and that the public across Pennsylvania is able to hold their legislature accountable. And indeed, if the vote-seat distribution is unresponsive to shifts in voter preferences, then it's impossible for the public to actually hold their elected officials accountable.

And on this plan, importantly, the party that wins the majority of the votes will usually win the majority of the seats. And I think this is distinctly unlike the prior plan, and I think that's really important, again, from the point of view of democratic theory that, you know, at end of the day, the government should represent the preferences of the majority of the public. And I think this would be the consensus of, you know, dozens of articles and books in political science on representation, and I think it's important for evaluating any plan and its partisan fairness.

And then finally to conclude, based on three methods of projecting future elections in four different generally accepted partisan bias metrics, overall, I find that the plan is fair with just a small pro-Republican bias.

So thank you, again, for giving me the opportunity to testify before you today, and I look forward to any questions that you might have.

CHAIR NORDENBERG: Thank you very much for your testimony, Professor Warshaw, and welcome back to Pennsylvania, in a sense.

Are there questions from any of the other Commissioners?

Majority Leader Benninghoff.

REPRESENTATIVE BENNINGHOFF: Actually, I have a lot of questions in my head, but I'll keep them for now. I was just curious, I do have a few comments I'd like to make, and I assume you want to wait until the closing time period to do that--or closing comments?

CHAIR NORDENBERG: Yes, please.

REPRESENTATIVE BENNINGHOFF: No problem.

Thank you.

CHAIR NORDENBERG: Leader McClinton.

REPRESENTATIVE McCLINTON: Yes, I have some. Thank you, Chairman. I thought my colleague had questions.

Dr. Warshaw, is it your conclusion that according to the five measures commonly used by political scientists to measure partisan bias, the preliminary House plan has a slight Republican bias? In other words, according to these measures, it's generally a fair map, and more fair than the current map?

DR. WARSHAW: Yes. That's exactly right.

REPRESENTATIVE McCLINTON: Thank you.

Based on your analysis of the preliminary House plan, are you confident that if Republicans win a majority of the votes in the next election, that they'll also win a majority of seats in the House of Representatives under the current proposal?

DR. WARSHAW: Yes. Across all the different analyses I did, if Republicans win a majority of the votes, so too they would win a majority of the seats.

REPRESENTATIVE McCLINTON: And based on your analysis of the preliminary House plan, are you confident that if Democrats win a majority of the votes in the next election, that they would also win a majority of the seats in the House of Representatives under this plan?

DR. WARSHAW: Overall, I think that's generally the conclusion I reached. It's very likely that Democrats would also win a majority of the seats if they win a majority of the votes. But when I say the plan has a small pro-Republican bias, one of the manifestations of that is that it's a little bit more likely that Democrats wouldn't get a majority of the seats with a majority of the votes than it is for Republicans.

REPRESENTATIVE McCLINTON: Thank you, Dr. Warshaw.

Thank you, Chairman.

CHAIR NORDENBERG: Professor questions?

Yes, Majority Leader Benninghoff.

REPRESENTATIVE BENNINGHOFF: You're very gracious, Mr. Chairman. I appreciate that.

I don't know if you listened to Dr. Barreto's comments earlier, but I was asking about some of the historical Hispanic communities and what I see as a reduction in the overall population and whether or not some of these districts may have multi-race cohesions that are being counted as a minority district but still, in my opinion, disadvantage a particular sector, the Hispanics. The reason I give that to you as a little bit of background, his answer basically was, well, voters can generally choose their candidate of preference, or something along those lines. I guess my question to you is, in your analysis, you used kind of some matrix nobody else has really talked about, not necessarily ones that we're looking at as constitutional matrixes. But that said, in a general election, people still have the ability to choose who they want to vote for, whoever is on the ballots, R or D, and maybe the past practices and performances that you are using in your analysis are just the result of voters' choices.

DR. WARSHAW: Well, I think that what political scientists have found really across a huge range of academic studies is that the way we draw the districts really has a large effect on the number of Democrats and Republicans, or candidates of whatever sort you want to analyze that actually win seats in the legislature. And so, for instance, if a party, as Democrats were in the previous plan, are packed into a small -- their voters are packed into a small number of districts, then they have a profound disadvantage in the translation of votes to seats. And I think one of the notable things about this plan is that it is neutral in the translation of votes to seats between the two parties. And, again, just a small pro-Republican advantage.

REPRESENTATIVE BENNINGHOFF: I'll reserve my other comments.

Thank you, Mr. Chairman, for allowing that.

CHAIR NORDENBERG: I know that you talked a bit about this, but I hope you'll indulge me and explain what is meant by "proportional representation," and how that differs from the measures that you used.

DR. WARSHAW: Thank you, Mr. Chair.

So proportional representation, as Professor Barber alluded to it, is the idea that if we were electing perhaps 100 representatives statewide, that maybe if the party that wins 53 percent of the vote should get exactly 53 percent of those 100 seats. And if that were true, that would be a proportional representation system. Now, of course, the United States doesn't have, and really has never had, a proportional representation system. And I don't think any theory of democratic representation, which of course is what I study in my academic work, hinges on any notion of proportional representation.

So all four of the partisan bias metrics that I use, you know, use slightly different ways of mathematically translating votes to seats and evaluating bias. But none of them hinge on any kind of assumption that the party that gets 52 or 53 percent of the votes should get exactly that proportion of the seats.

CHAIR NORDENBERG: But to go back to a standard that you've employed on a couple of occasions and that you were you kind enough to connect to me, the basic notion that the party that wins the majority of the votes generally should win a majority of the seats is an accepted proposition amongst political scientists?

DR. WARSHAW: Exactly. I think that would be a consensus view among scholars of political representation and democracy writ large. I think it would be a consensus view that the party that wins a majority of the votes should win enough seats to control the legislature. And I think if that's not true, it sort of calls into question the democratic bona fides of any government.

CHAIR NORDENBERG: Any other questions from my Commission colleagues?

REPRESENTATIVE BENNINGHOFF: No, sir.

CHAIR NORDENBERG: If not, Professor Warshaw, let me thank you, again, for being here, and thank you for the report that you submitted.

DR. WARSHAW: Thank you, Mr. Chair.

CHAIR NORDENBERG: This does bring us almost to the end of today's Session. I will say that it's been a real learning experience for me. I suspect that we are not done discussing and hearing about some of the presentations that have been made.

More particularly, I will say that the House

Republican Caucus has received a report from another retained expert, Jonathan Katz from Caltech, and it is their intention--in fact, we are seeing it right before our eyes--to distribute that report to the other Members of the Commission, and also to provide a copy to our reporter so that it becomes a part of the record.

While I was sitting here, I received a message from Fair Districts that they have a report that they would like to submit in response to the testimony of Professor Barber.

Even though we are not having the chance to question Professor Katz, it seems the best approach is to accept his report and to make it available to all of the Members of the Commission.

Even if it was not your birthday, Mr. Majority Leader, I certainly would give you the opportunity to make a closing statement, if you wish. So let me turn the microphone over to you.

REPRESENTATIVE BENNINGHOFF: Actually, I'll defer to the Minority Leader for a moment. I did want to just assure, the report we're getting from Fair Districts, each of us will get a copy of that as well?

CHAIR NORDENBERG: Yes.

REPRESENTATIVE BENNINGHOFF: And thank you for allowing Dr. Katz's report in there.

I'll defer to the Minority Leader.

CHAIR NORDENBERG: Leader McClinton.

REPRESENTATIVE McCLINTON: Thank you, Mr. Chairman, and thank you, Leader Benninghoff.

Mr. Chairman, I do object to the receipt, submission, and admission of this document that we were just provided at 4:45 p.m. on the first of the last two days of public comment. This report essentially is a criticism of one of our experts whose report has been out for over nine days, publicly accessible. This has just been provided at this late moment in this juncture and proceedings, and it should not be admissible whatsoever.

CHAIR NORDENBERG: Is there a response?

REPRESENTATIVE BENNINGHOFF: Two responses.

Number one, I would be more than glad to allow any of the experts to give their comments on that testimony. I thought from the very beginning all of us agreed that if we didn't agree on some things, that we wanted to have an open and transparent process, allow people to give comments. There was no drop-dead date of providing information. While it was received very abruptly to us as well, we did confer with the Chairman about the best methodology to get that to him, rather than take the chance of mailing it in or just sending it electronically, that we decided to offer to provide that copy in-hand.

The second comment on that, I have no more concern about accepting Fair Districts' report or analysis that they would now like to submit today because more information is more information, and we need to be able to review this, if we truly want to have an open and transparent process. And I would respectfully ask the Chairman to accept both reports.

CHAIR NORDENBERG: You know, let me say, Leader McClinton, that I had some of the same reservations that you expressed, because our Chief Counsel had set forth a clear set of deadlines for the submission of reports by retained experts. In the end, it was my sense that it was better not to keep it out but to keep the process open. I don't know if our Chief Counsel, who is up there on the screen and who has been watching all day, has anything he would like to add.

MR. BYER: Nothing to add to that. I think you've said it well, and I think that the decision to allow both of the reports that have just been tendered is a good decision, as far as I'm concerned.

CHAIR NORDENBERG: Well, that will be my decision, if it is my decision to make, though I appreciate the concerns that you've raised.

Is there anything else to come before the Commission at this time?

(There was no response.)

CHAIR NORDENBERG: Would you like us all to sing

Happy Birthday, or would you consider it a favor if we passed on that?

REPRESENTATIVE BENNINGHOFF: No, sir, but at the appropriate time, I just have a couple of closing comments I'd like to make in reference to a few issues.

CHAIR NORDENBERG: Okay. I think this would be an appropriate time.

REPRESENTATIVE BENNINGHOFF: Thank you, Mr. Chairman.

And as you and I talked early on, we're in a process where we agree to disagree. We don't have to agree on everything, but hopefully we all walk away learning a little bit more than we knew when we first started this several months ago.

I did want to thank the Chairman and the Members of this Commission. From the beginning of this redistricting process, I have looked predominantly to our Constitution and Federal law as my own personal guide. In doing so, I am informed by the two most recent redistricting cases in Pennsylvania, the *Holt* decision in 2011, and the *League of Women Voters* case from 2018, as to what our Constitution demands of us, pretty well laid out. The Supreme Court in *Holt* advised the Commission that it may only consider factors outside of the constitutional mandates so long as they do not do violence to the constitutional restraints regarding equal population, contiguity, compactness, and respect for the integrity of political subdivisions. In addition, in the *Holt* court case, it tells us that "The constitutional reapportionment scheme does not impose a requirement of balancing the representation of the political parties; it does not protect the 'integrity' of any party's political expectation. Rather, the construct speaks of the 'integrity' of political subdivisions, which bespeaks history and geography, not party affiliation or expectations." In other words, the Commission cannot unnecessarily split county or municipal lines to artificially increase the number of a particular party's leaning districts, even if the alleged goal of doing so is to achieve a more proportional seat share relative to the two-party statewide system and vote share, or to negate a natural geographic disadvantage.

Additionally, in the *League of Women Voters of Pennsylvania v. The Commonwealth*, the Pennsylvania Supreme Court interpreted the free and equal election clause to require that "...an individual's electoral power not be diminished through any law which discriminatorily dilutes the power of his or her vote...." Our Supreme Court cautioned that the neutral criteria in Article II, Section 16, like "...compactness, contiguity, and the maintenance of the integrity of the boundaries of political subdivisions..." should take precedence over things such as gerrymandering for unfair partisan political advantage, because those neutral factors "maintain the strength of an individual's vote in electing...representatives." Very key, important points. The court went on and even further noting that adherence to these neutral criteria is the floor and not the ceiling.

Beyond the demands of Article II, Section 16, and Article I, Section 5, I have to say, I'm disappointed that none of the experts advocating for this plan explain questions under Article I, Section 29, of the Pennsylvania Constitution. We may remember in May of last year, the people of Pennsylvania approved a constitutional amendment, Article I, Section 29, which state's "Equality of rights under the law shall not be denied or abridged in the Commonwealth of Pennsylvania because of the race or ethnicity of the individual." Pennsylvania's very own Attorney General, in his plain English statement, described the amendment and said:

"...inclusion of this amendment within the Pennsylvania Constitution signifies that freedom from discrimination based on race or ethnicity is an essential principle in liberty and free government.... This amendment applies to all Pennsylvania state, county and local governmental entities, and guarantees equality of rights under the law.

"This equal right to be free from racial or ethnic discrimination will exist independent from any such rights under the United States Constitution or corresponding federal law."

While I appreciate everything the experts defending this map had to say, I am somewhat concerned on behalf of Pennsylvanians that this new constitutional language was ignored in the development of this plan. Even more importantly, I am convinced that the splits in Harrisburg, Allentown, Lancaster, and Reading creates specific problems under the VRA and this new constitutional language. I know there's a lot of things that this Commission has to address based on numerous complaints that we've received. I believe the Chairman said there was over 3,700 on the portal, and those that came before us by Pennsylvania citizens, and it's good to hear from them. In addressing those complaints and problems in the House map, we must be faithful to our Constitution and Federal law, particularly in light of our new constitutional amendment voted on by the people.

I close in saying I look forward to continuing this dialogue and, hopefully, producing a more constitutionally sound map in the House, one that can garner a unanimous vote by this Commission, an objective I've had from the beginning. I do look forward to continuing this dialogue and, hopefully, getting this done in a timely manner.

Thank you, Mr. Chairman, and thank you to the Commission, for your patience.

CHAIR NORDENBERG: Leader McClinton.

REPRESENTATIVE McCLINTON: Thank you, Mr. Chairman.

Mr. Chairman, I understand and note your ruling in regards to admitting the late-filed Katz report after the deadline that had been provided, and I do ask that I can reserve the right to submit a response to this report.

CHAIR NORDENBERG: You absolutely should have that right.

REPRESENTATIVE McCLINTON: And before we conclude, I certainly want to wish my colleague a very happy birthday.

REPRESENTATIVE BENNINGHOFF: Tough to be 29 again.

REPRESENTATIVE McCLINTON: Twenty-nine? Then I'm 19.

(Laughter.)

REPRESENTATIVE BENNINGHOFF: I will plead the fifth.

CHAIR NORDENBERG: Well, and if spending all day with us was not enough of a treat, he's got tomorrow morning to look forward to as well.

I really do want to thank everyone for what I think has been a very productive day. And as we sit here in Harrisburg, Senator Costa, Chief Counsel Byer, we hope you're both doing well, and we look forward to seeing you back in action in person soon.

SENATOR COSTA: Thank you, Mr. Chairman. I look forward to it as well.

MR. BYER: Thank you very much.

CHAIR NORDENBERG: With that then, today's hearing is adjourned, and we will reconvene for another hearing at 9 o'clock tomorrow morning, right here.

Thank you all.

(Whereupon, the proceedings were concluded at 4:55 p.m.)

I hereby certify that the proceedings and evidence are contained fully and accurately in the notes taken by me during the hearing of the within cause, and that this is a true and correct transcript of the same.

Ann-Marie P. Sweeney

____

ANN-MARIE P. SWEENEY
Official Reporter
Legislative Reapportionment
Commission

THE FOREGOING CERTIFICATION DOES NOT APPLY TO ANY REPRODUCTION OF THE SAME BY ANY MEANS UNLESS UNDER THE DIRECT CONTROL AND/OR SUPERVISION OF THE CERTIFYING REPORTER.

ANN-MARIE P. SWEENEY
Official Reporter
Legislative Reapportionment Commission
P.O. Box 203079
Harrisburg, PA 17120

EXHIBITS

Agenda
Hearing #15
Pennsylvania Legislative Reapportionment Commission

January 14, 2022
2:00 p.m. to 5:00 p.m.
North Office Building, Hearing Room 1

Comments on Preliminary Plan

1. Call to Order and Opening Remarks

2. Comments from Invited Experts

   Dr. Michael Barber: Associate Professor of Political Science and faculty scholar at the Center for the Study of Elections and Democracy, BrighamYoung University

   Dr. Kosuke Imai: Professor of Government and Statistics and affiliate of the Institute for Quantitative Social Science, Harvard University; previous faculty appointment at Princeton University, where he was the founding Director of its Program in Statistics and Machine Learning

   Dr. Matt Barreto: Professor of Political Science and Chicana/o and Central American Studies, founder of the Latino Policy & Politics Initiative and Voting Rights Project, UCLA; President and Co-Founder of BSP Research, a research and polling firm; previous faculty appointment at the University of Washington

   Dr. Christopher Warshaw: Associate Professor of Political Science, George Washington University; previous faculty appointment at MIT

3. Closing Remarks and Adjournment

Report on Proposed Redistricting Plan from the Pennsylvania Legislative Reapportionment Commission

Dr. Michael Barber
Brigham Young University
724 Spencer W. Kimball Tower
Provo, UT 84604
barber@byu.edu

1 Introduction and Qualifications

I have been asked by counsel to review the Legislative Reapportionment Commission's proposed redistricting plan and compare it to a set of simulated redistricting plans across a number of factors commonly considered in the redistricting process and in redistricting litigation.

I am an associate professor of political science at Brigham Young University and faculty fellow at the Center for the Study of Elections and Democracy in Provo, Utah. I received my PhD in political science from Princeton University in 2014 with emphases in American politics and quantitative methods/statistical analyses. My dissertation was awarded the 2014 Carl Albert Award for best dissertation in the area of American Politics by the American Political Science Association.

I teach a number of undergraduate courses in American politics and quantitative research methods.[1] These include classes about political representation, Congressional elections, statistical methods, and research design.

I have worked as an expert witness in a number of cases in which I have been asked to analyze and evaluate various political and elections-related data and statistical methods. Cases in which I have testified at trial or by deposition are listed in my CV, which is attached to the end of this report. I have previously provided expert reports in a number of cases related to voting, redistricting, and election-related issues: *Nancy Carola Jacobson, et al., Plaintiffs, vs. Laurel M. Lee, et al., Defendants. Case No. 4:18-cv-00262 MW-CAS (U.S. District Court for the Northern District of Florida); Common Cause, et al., Plaintiffs, vs. Lewis, et al., Defendants. Case No. 18-CVS-14001 (Wake County, North Carolina); Kelvin Jones, et al., Plaintiffs, v. Ron DeSantis, et al., Defendants, Consolidated Case No. 4:19-cv-300 (U.S. District Court for the Northern District of Florida); Community Success Initiative, et al., Plaintiffs, v. Timothy K. Moore, et al., Defendants, Case No. 19-cv-15941 (Wake County, North Carolina); Richard Rose et al., Plaintiffs, v. Brad Raffensperger,*

[1] The political science department at Brigham Young University does not offer any graduate degrees.

Defendant, Civil Action No. 1:20-cv-02921-SDG (U.S. District Court for the Northern District of Georgia); Georgia Coalition for the People's Agenda, Inc., et. al., Plaintiffs, v. Brad Raffensberger, Defendant. Civil Action No. 1:18-cv-04727-ELR (U.S. District Court for the Northern District of Georgia); Alabama, et al., Plaintiffs, v. United States Department of Commerce; Gina Raimondo, et al., Defendants. Case No. CASE NO. 3:21-cv-00211-RAH-ECM-KCN (U.S. District Court for the Middle District of Alabama Eastern Division); League of Women Voters of Ohio, et al., Relators, v. Ohio Redistricting Commission, et al., Respondents. Case No. 2021-1193 (Supreme Court of Ohio); Harper, et al., Plaintiffs, v. Hall et al., Defendants. Case No. 21-CVS-015426 (Wake County North Carolina)

In my position as a professor of political science, I have conducted research on a variety of election- and voting-related topics in American politics and public opinion. Much of my research uses advanced statistical methods for the analysis of quantitative data. I have worked on a number of research projects that use "big data" that include millions of observations, including a number of state voter files, campaign contribution lists, and data from the US Census. I have also used geographic information systems and other mapping techniques in my work with political data.

Much of this research has been published in peer-reviewed journals. I have published nearly 20 peer-reviewed articles, including in our discipline's flagship journal, *The American Political Science Review* as well as the inter-disciplinary journal, *Science Advances*. My CV, which details my complete publication record, is attached to this report as Appendix A.

The analysis and opinions I provide in this report are consistent with my education, training in statistical analysis, and knowledge of the relevant academic literature. These skills are well-suited for this type of analysis in political science and quantitative analysis more generally. My conclusions stated herein are based upon my review of the information available to me at this time. I reserve the right to alter, amend, or supplement these conclusions based upon further study or based upon the availability of additional information. The opinions in this report are my own, and do not represent the view of Brigham Young

University.

2 Methods

To gauge the degree to which the Commission's proposed map is a partisan gerrymander, I conduct simulated districting analyses to allow me to produce a large number of districting plans that follow traditional redistricting criteria using small geographic units as building blocks for hypothetical legislative districts (election precincts). This simulation process ignores all partisan and racial considerations when drawing districts. Instead, the computer simulations are programmed to create districting plans that follow traditional districting goals without paying attention to partisanship, race, or the location of incumbent legislators. This set of simulated districts is helpful because it provides a set of maps to which we can compare the Commission's proposed map to see if it is biased in favor of either political party. This is because in comparing the Commission's map to the simulated districts, we are comping a map to a set alternative maps that we know to be unbiased. If the Commission's map produces a similar outcome as the alternative set of maps, we may reasonably conclude that the Commission's plan is also unbiased. Alternatively, if the Commission's proposed plan significantly diverges from the set of simulated maps, it may be the case that the proposed plan is biased in favor of one party.

The process of simulating districting plans has been recognized and used in a variety of redistricting cases, including in Pennsylvania.[2] While different people employ slightly different methods, the overall process is much the same. For my simulations, I use a program developed by Fifield et al. (2020).[3]

[2]See League of Women Voters of Ohio v. Ohio Redistricting Commission (2021); Harper v. Hall (2021); Common Cause v. Lewis (2019); Harper v. Lewis (2019); League of Women Voters of Pennsylvania v. Commonwealth of Pennsylvania (2018).

[3]Fifield, Benjamin, , Michael Higgins, Kosuke Imai, and Alexander Tarr. "Automated redistricting simulation using Markov chain Monte Carlo." Journal of Computational and Graphical Statistics 29, no. 4 (2020): 715-728.

Fifield, Benjamin, Kosuke Imai, Jun Kawahara, and Christopher T Kenny. 2020. "The essential role of empirical validation in legislative redistricting simulation." Statistics and Public Policy 7 (1): 52–68.

A significant advantage of the simulation-based approach is the ability to provide a representative sample of possible districting plans that accounts for the unique political geography of a state, such as the spatial distribution of voters or the location and number of administrative boundaries, such a counties. Simulation methods can also to a degree incorporate each state's unique redistricting rules. The simulation-based approach therefore permits us to compare a particular plan to a large number of representative districting plans in Pennsylvania. In the simulations I run, I instruct the model to generate plans that adhere to the redistricting criteria contained in the Pennsylvania Constitution.

Specifically, the model is constrained to conduct 50,000 simulations in which each simulation generates 203 districts that are of roughly equal population (<4.6% deviation above or below the target population of 64,053, which is the same range as in the commission proposal). The algorithm does this by assembling small geographic units — electoral precincts — into larger groups of precincts until a group of precincts is large enough to constitute a new legislative district. The model does this 203 times to create a full redistricting plan containing 203 legislative districts. It then repeats this process 50,000 times, generating a different set of 203 districts with each run of the model. In each of the 50,000 iterations, the model is instructed to generate districts that cross county boundaries as few times as possible. Of course, county populations do not always add up to round units of districts, and so of necessity some county boundaries will be split. The model is further instructed that when a county boundary needs to be crossed, it should avoid splitting the county more times than necessary. The model also includes instructions to generate districts that are geographically compact. The final constraint is an instruction to avoid splitting municipal and township boundaries.

Once the simulated district plans are complete, only then do I compute the partisan composition of each district in each plan. For the partisan composition of each district I rely

---

Kenny, Christopher T., Cory McCartan, Benjamin Fifield, and Kosuke Imai. 2020. redist: Computational Algorithms for Redistricting Simulation. https://CRAN.R-project.org/package= redist.

McCartan, Cory, and Kosuke Imai. 2020. "Sequential Monte Carlo for sampling balanced and compact redistricting plans." arXiv preprint arXiv:2008.06131.

on the election results from statewide elections disaggregated to the level of the precinct. I then reassemble these election results for each of the simulated districts in each of the 50,000 simulations to compute the proportion of votes across all statewide elections conducted between 2012 and 2020 that were won by the Democratic and Republican candidates in those districts.[4] In other words, the partisan index is the average vote share for Democratic candidates in each district for the statewide elections considered between 2012-2020. I choose 2012 as the starting date as this a full set of elections between the decennial census. Furthermore, averages of multiple elections have the benefit of "washing out" the impact of any particular election, since individual elections can vary due to particular candidate features and other idiosyncrasies, and particular years can vary due to national electoral waves (i.e. 2018 was an especially good year for Democrats while 2016 was an especially good year for Republicans nationwide).

[4]The particular races are 2020: President, Auditor, Attorney General, Treasurer; 2018: Governor, US Senate; 2016: President, US Senate, Auditor, Attorney General, Treasurer; 2014: Governor; 2012: President, US Senate, Auditor, Attorney General, Treasurer. I do not include statewide judicial elections in the index. It is uncommon in political science to use judicial elections to measure voters' partisan preferences as research suggests voters treat judicial elections very differently, even when judges run under party labels, than they do partisan elections to legislative and executive positions. Other commonly used measures indices such as Dave's Redistricting and PlanScore.com also omit judicial elections from their partisan indices.

3 Results

3.1 Population, Boundary Splits, and Compactness

Table 1 below compares the Commission proposal to the distribution of simulations for population deviation, boundary splits, and compactness. The Commission proposal and the simulations are within the same range of district population deviations from the target district size. The proposal splits 45 counties 184 times. This is in line with the simulations in terms of the number of counties split. The proposal divides 63 municipalities 102 times. This is also within the range produced by the simulations. On the whole, the proposal appears to perform well at having few municipal splits. However, later in the report I will show how the choice of *which* municipalities to split is informative of why the Commission's proposal is such an extreme partisan outlier compared to the set of simulation results. With regards to district compactness, the Commission proposal is similarly compact and largely in line with the results of the simulations.

Table 1: Commission Proposal and 50,000 Simulations: Population, Splits, and Compactness

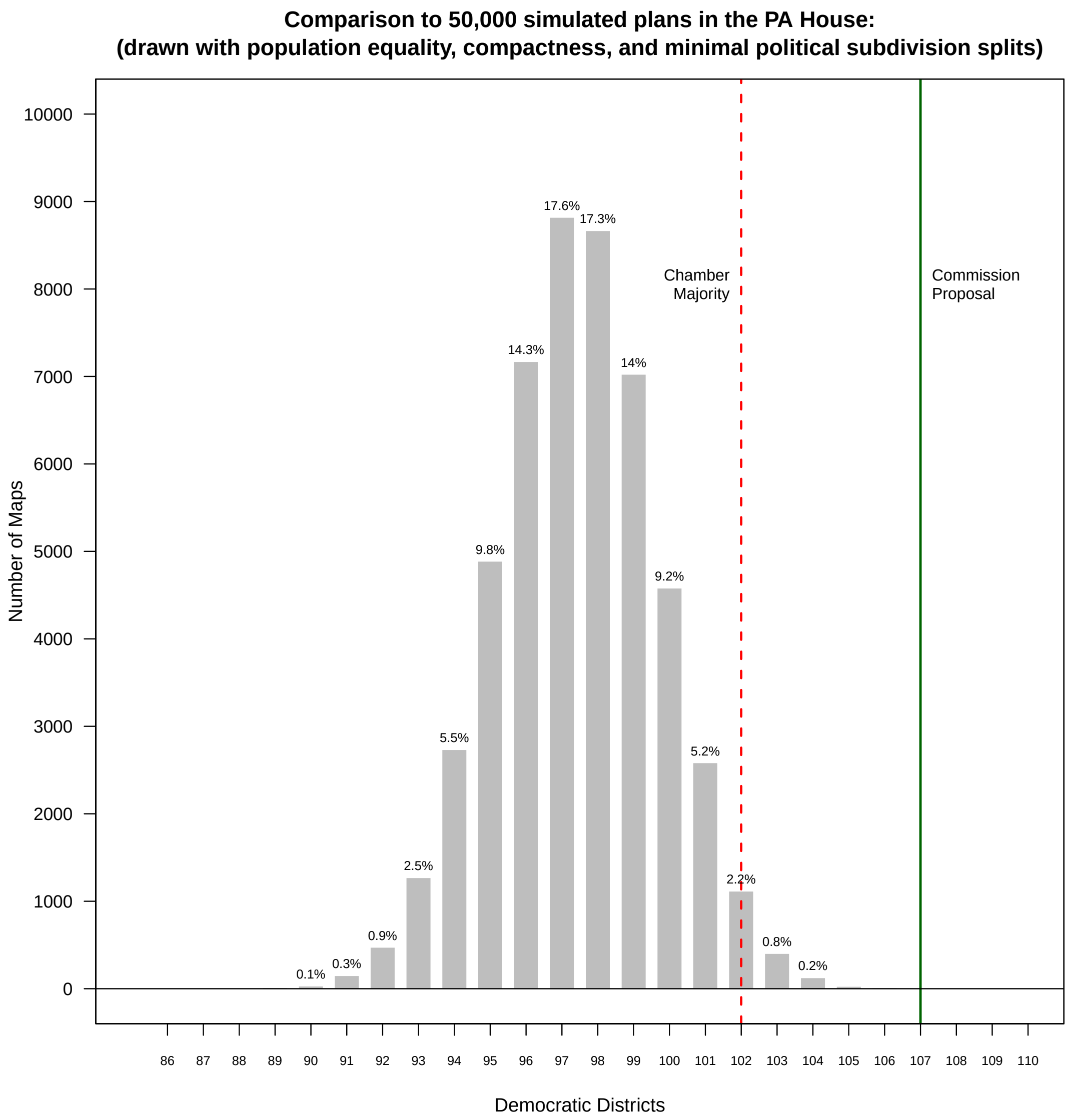
	Commission Proposal	Simulations Median	Simulations Range
Population Deviation			
Smallest District:	-4.62%	-4.61%	[-4.65., -4.25]
Largest District:	4.67%	4.62%	[4.21, 4.65]
Boundary Splits			
Counties Split:	45	46	[42, 52]
Total County Splits:	184	195	[183, 207]
Municipalities Split:	63	81	[60, 102]
Total Municipal Splits:	102	118	[97, 140]
Compactness			
Median Polsby-Popper:	0.34	0.32	[0.28, 0.34]

3.2 Partisanship

Figure 1 displays the distribution of Democratic leaning districts in both the simulations and the Commission's proposal using the partisan index discussed above. For reference the red dashed line in the plot is at 102, the number of seats needed for a majority in the Pennsylvania House of Representatives. The green line shows the results of calculating the partisan index for the Commission proposal. The Commission proposal generates 107 Democratic leaning districts (districts with a partisan index greater than 0.50), which is 10 seats larger than the most common outcome generated by the simulations, 97. The numbers above each bar in the histogram display the relative frequency of each outcome in the simulations. Beginning from the far left side of the figure and adding those numbers up as one moves to the right, we would find that the Commission's plan generates more Democratic leaning districts than 99.998% of the simulations.

Recall that in using the simulations we are comparing the Commission's proposed map to a set of maps drawn by the computer using only those criteria that I instructed the algorithm to follow - namely the pre-specified nonpartisan criteria of equal population, contiguity, geographic compactness and a preference for fewer county and municipal splits. And yet the degree to which the Commission's proposal diverges from the distribution of simulation results is extreme and represents a significant deviation from a fair outcome. Thus, the significant deviation observed here strongly suggests that the Commission's plan was drawn using some other, or additional criteria. This could, of course, include a motivation for Democratic partisan advantage given the incredibly large deviation between the number of Democratic districts generated by the proposal and the range of Democratic-leaning districts generated by the simulations.

Figure 1: **Partisan Composition of Commission Proposal and Simulations**



Note: The grey distribution is the number of Democratic seats generated from the 50,000 simulations. The vertical green line is the number of Democratic leaning seats in the Commission's proposal. The Commission's proposal generates more Democratic leaning districts than 99.998% of the simulations. The red dashed line is placed at 102, the number of seats needed for majority control in the Pennsylvania House of Representatives. The partisan lean of districts in the simulations and the Commission proposal are calculated as the two-party vote share of statewide partisan elections from 2012-2020.

4 Political Geography of Pennsylvania

Where are the discrepancies in partisanship arising? Given the geographic distribution of voters in Pennsylvania and the clustering of Democrats within the large and medium-sized cities of the state, there are only relatively few locations in which Democratic districts can be constructed.

Scholarship in political science has noted that the spatial distribution of voters throughout a state can have an impact on the partisan outcomes of elections when a state is, by necessity, divided into a number of legislative districts. This is largely the case because Democratic-leaning voters tend to cluster in dense, urban areas while Republican-leaning voters tend to be more equally distributed across the remainder of the state.[5] One prominent study of the topic (Chen and Rodden, 2013) finds that "Democrats are highly clustered in dense central city areas, while Republicans are scattered more evenly through the suburban, exurban, and rural periphery...Precincts in which Democrats typically form majorities tend to be more homogenous and extreme than Republican-leaning precincts. When these Democratic precincts are combined with neighboring precincts to form legislative districts, the nearest neighbors of extremely Democratic precincts are more likely to be similarly extreme than is true for Republican precincts. As a result, when districting plans are completed, Democrats tend to be inefficiently packed into homogenous districts."[6]

Rodden (2019) further discusses this with specific reference to Pennsylvania.[7] He

[5]See for example Stephanopoulos, N. O. and McGhee, E. M., Partisan Gerrymandering and the Efficiency Gap, *The University of Chicago Law Review* 82: 831-900, (2015); Chen, J. and Rodden, J., Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures, *Quarterly Journal of Political Science* 8: 239-269, (2013); Nall, C., The Political Consequences of Spatial Policies: How Interstate Highways Facilitated Geographic Polarization, *Journal of Politics*, 77(2): 394-406, (2015); Gimple, J. and Hui, I., . Seeking politically compatible neighbors? The role of neighborhood partisan composition in residential sorting, *Political Geography* 48: 130-142 (2015); Bishop, B., *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart*, Houghton Mifflin Press (2008); and Jacobson, G. C., and Carson, J. L., *The Politics of Congressional Elections*, 9th ed. Lanham, MD: Rowman and Littlefield (2016).

[6]Chen, J. and Rodden, J., Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures, *Quarterly Journal of Political Science* 8: 239-269, (2013)

[7]Rodden, Jonathan A. Why cities lose: The deep roots of the urban-rural political divide. Hachette UK, 2019.. While Rodden is specifically discussing Pennsylvania in this quote, the statement is true of any location with Democrats clustered in urban areas.

states:

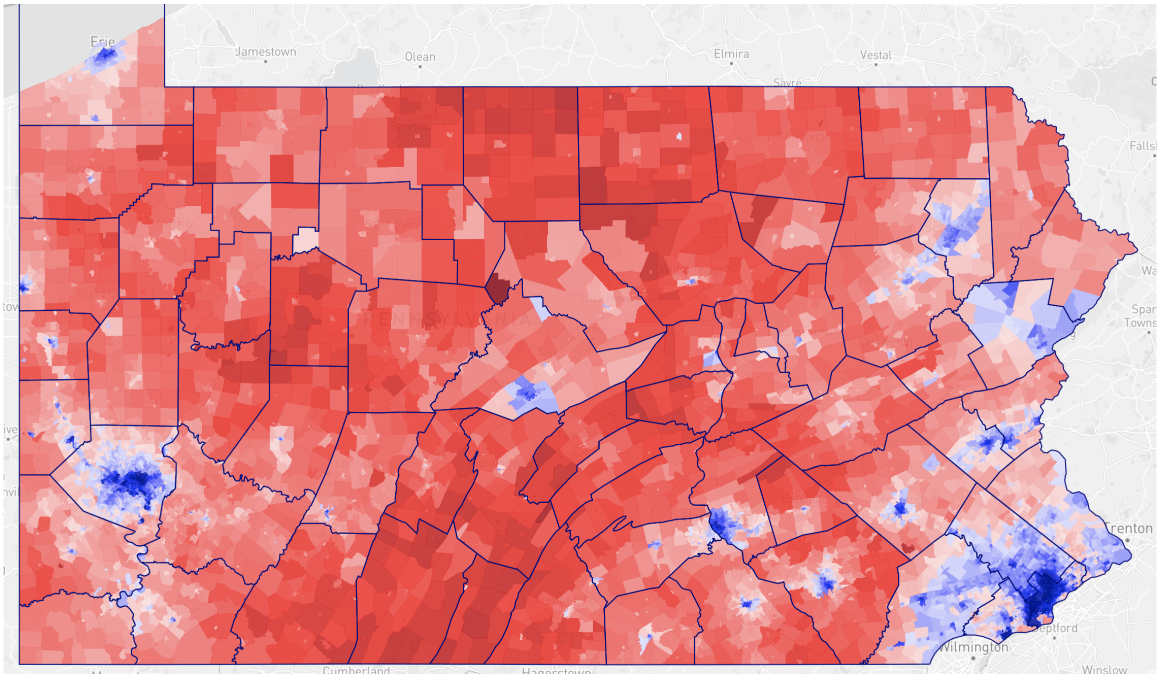
> Then and now, the Democrats have been plagued by a problem with geography. In the years following the New Deal, their supporters became concentrated in the core urban neighborhoods of Pennsylvania's nineteenth-century industrial cities and along the surrounding railroad tracks. They remain so today....Because of the scale and geographic arrangement of Pennsylvania's nineteenth-century cities, the Democrats' problem is severe when districts are very small—as in the state house of representatives—and even worse when they are medium-sized, as in the state senate.

The map below confirms that this is the case in Pennsylvania. We see large Democratic majorities shown in blue in and around Philadelphia and Pittsburgh as well as small pockets of densely populated Democratic voters in the other medium-sized industrial cities of the state. These areas are surrounded by large swaths of the state that are solidly Republican.

The upshot of this pattern is that political parties stand at a disadvantage when their voters are not "efficiently" distributed across the state. To understand what I mean by efficient, imagine two different scenarios. First, imagine a party with a slim majority of voters statewide in which every precinct's vote share perfectly reflected the overall state. In other words, the party has a slight majority in every precinct that adds up to a slight majority statewide. In this case, this party's voters are extremely efficiently distributed in such a way that the party will win every single district despite only a slim majority statewide. Now imagine a different arrangement, a party who still holds a slim majority statewide, but whose voters are heavily concentrated in a few areas and sparsely populated throughout the rest of the state. In this case, despite holding a majority of votes statewide, the party will only win a few seats where their voters are heavily concentrated. The political geography of Pennsylvania closely resembles the second scenario.

The geographic concentration of a party's voters tends to harm that party when single-member districts are drawn by creating districts that favor that party by very large

Figure 2: **Distribution of People and Partisan Preferences in Pennsylvania**



Note: Distribution of Partisan Preferences in Pennsylvania based on the average of statewide partisan elections. Blue = Democratic, Red = Republican

majorities, thus "wasting" many votes in running up large majorities far beyond 50%+1.[8] This occurs in Pennsylvania in the large and medium-sized cities of the state. These overwhelming margins for the party are what drives "wasted votes," which, in turn translate to fewer seats than the statewide proportion of the vote would suggest.[9]

Another way to consider this is to look at a lower level of geography, electoral precincts. Figure 3 shows the distribution of partisan preferences for recent statewide partisan elections for all precincts in Pennsylvania. The top panel notes precincts where there are strong majorities for either party and labels them as "inefficient" precincts (those precincts towards the outer edges of the figure). They are inefficient based on the discussion above

[8]McGhee, E. (2017). Measuring Efficiency in Redistricting. Election Law Journal: Rules, Politics, and Policy, 16(4), 417–442. doi:10.1089/elj.2017.0453

[9]The term "wasted votes" in political science is not to imply that a person's vote is not important or counted, but rather that the vote is less helpful in gaining an additional seat for their preferred party if it is an additional vote in favor of a candidate that has already won a substantial majority of the votes in their district. Technically, all votes beyond 50%+1 would be, as a result, "wasted". However, parties are interested in winning by majorities larger than 50%+1, but not by margins much beyond that point at which their candidate is all but assured to win.

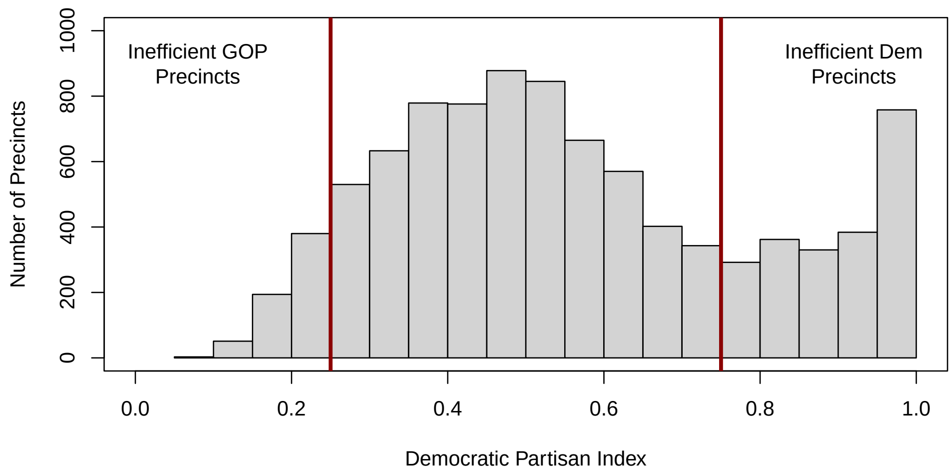
that a party wastes votes if it builds majorities far beyond the needed 50%+1. Note that the distribution is not symmetric and that there are many more precincts with very large Democratic majorities than there are precincts with equally large Republican majorities. The lower panel shows the same distribution but labels "efficient" precincts — those where a party has a majority, but not an overwhelming majority. Note here that there are many more precincts with efficient Republican majorities than there are precincts with efficient Democratic majorities.

This inefficient distribution of votes would not be a problem for Democrats if district boundaries were able to amble about the state and divide municipalities so as to create districts that had less overwhelming Democratic support. Rodden (2019) notes this by saying: "Democrats would need a redistricting process that intentionally carved up large cities like pizza slices or spokes of a wheel, so as to combine some very Democratic urban neighborhoods with some Republican exurbs in an effort to spread Democrats more efficiently across districts" (pg. 155).[10] However, the laws governing redistricting in Pennsylvania run counter to either of these strategies. Pennsylvania's redistricting rules that require districts to be geographically compact and to avoid county and municipal divisions prohibit the type of meandering districts that Rodden describes above. In the end, this means that Republicans begin the redistricting process with a natural geographic advantage due to the combination of laws requiring where and how districts are drawn combined with the particular spatial distribution of their voters.

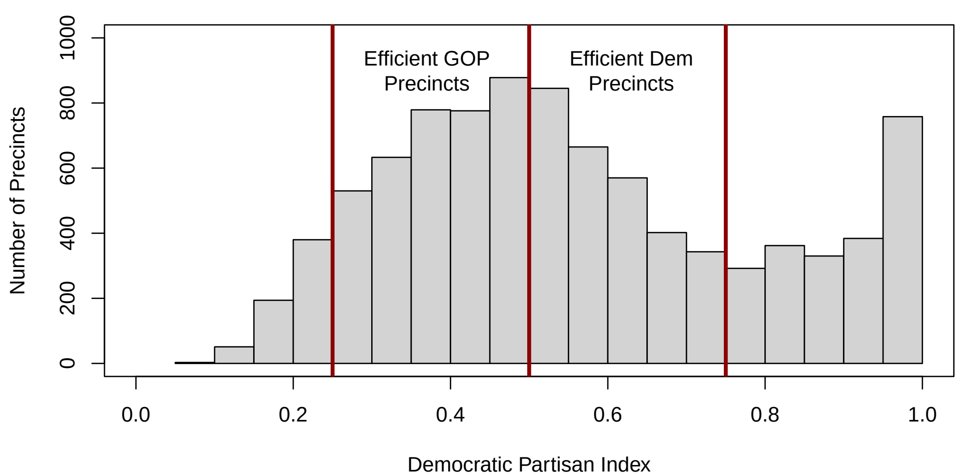
[10]Rodden, Jonathan A. Why cities lose: The deep roots of the urban-rural political divide. Hachette UK, 2019.

Figure 3: **Distribution of Votes Across Precincts in Pennsylvania**

(a) Inefficient precincts



(b) Efficient Precincts



Note: Partisan Index based on the average of statewide partisan races between 2012-2020.

5 Looking at Subsets of Pennsylvania

Given the discussion above, it is instructive to look at locations in the state that have urban clusters of Democratic voters. If the Commission's proposal is attempting to enact a Democratic gerrymander, we should see evidence of what Rodden (2019) discusses above, i.e. the intentional division of Democratic cities that are used to spread Democratic voters out more efficiently to overwhelm Republican votes in the adjacent suburbs and exurbs in order to create more Democratic districts than would otherwise be produced by keeping these municipalities whole.

To do this I focus on a number of counties (or groups of counties) in the state that contain large and medium-sized cities and compare the partisan outcomes in the Commission's proposed plan to the plans generated by the simulations. The table below summarizes these results. Looking at the table shows that the differences we observed between the simulations and the Commission's proposal are due to a systematic overrepresentation of Democrats in these counties with urban cores. Across the 7 groups of counties considered here, in 3 of the 7 cases the Commission's proposal generates one additional Democratic district than the most common outcome in the simulations, and in two regions the Commission's proposal generates 2 more Democratic seats than the most common outcome in the simulations. These deviations add up across the urban areas of the state to a collective deviation of seven seats, which accounts for a significant portion of the difference between the Commission's proposal and the most common outcome in the distribution of Democratic seats generated by the simulations statewide.

How does the Commission's proposed map generate an extra Democratic leaning seat in most of these counties considered in the table above? In the analysis below I show that the Commission's proposal follows exactly the strategy discussed by Rodden (2019) for how the Democratic party would have to work to overcome the disadvantage they face due to the geographic concentration of their voters. Recall the strategy he outlines, "Democrats would need a redistricting process that intentionally carved up large cities like pizza slices or

Table 2: County-by-County Analysis of Commission Proposal and 50,000 Simulations

	Number of Democratic Leaning Districts		
County:	Commission Proposal	Simulations Modal Outcome	% of Simulations Generating Fewer Democratic Seats Than Commission's Map
Philadelphia	25	25	0%
Allegheny	16	16	20.7%
Lehigh and Bucks	11	9	99.3%
Schuylkill, Berks, Lancaster, and Lebanon	5	4	83.5%
Dauphin, and Cumberland	3	2	73.9%
Susquehanna, Lackawanna, and Luzerne	12	10	98.5%
Centre and Clinton	2	1	72.3%

spokes of a wheel, so as to combine some very Democratic urban neighborhoods with some Republican exurbs in an effort to spread Democrats more efficiently across districts" (pg. 155).[11] This is exactly what the Commission's proposed plan does. In many of the largest cities in these counties the Commission unnecessary divides these cities when the population of these cities would not otherwise require them to be divided. The following section proceeds through each of these counties and shows the results of the simulations in the districts in these counties and compares them to the Commission's proposed districts in these counties. I then present maps of the Commission's map's district boundaries in these counties and show how in each case a heavily Democratic city is divided into more districts than its population would otherwise necessitate in order to more efficiently distribute Democratic voters across more districts and produce more districts with Democratic majorities. Furthermore, this is often accomplished by dividing cities that contain substantial minority populations. As a result, many of the districts created using this strategy crack minority populations and dilute their influence in the resulting districts.

[11]Rodden, Jonathan A. Why cities lose: The deep roots of the urban-rural political divide. Hachette UK, 2019.. While Rodden is specifically discussing Pennsylvania in this quote, the statement is true of any location with Democrats clustered in urban areas.

5.1 Lehigh and Bucks Counties

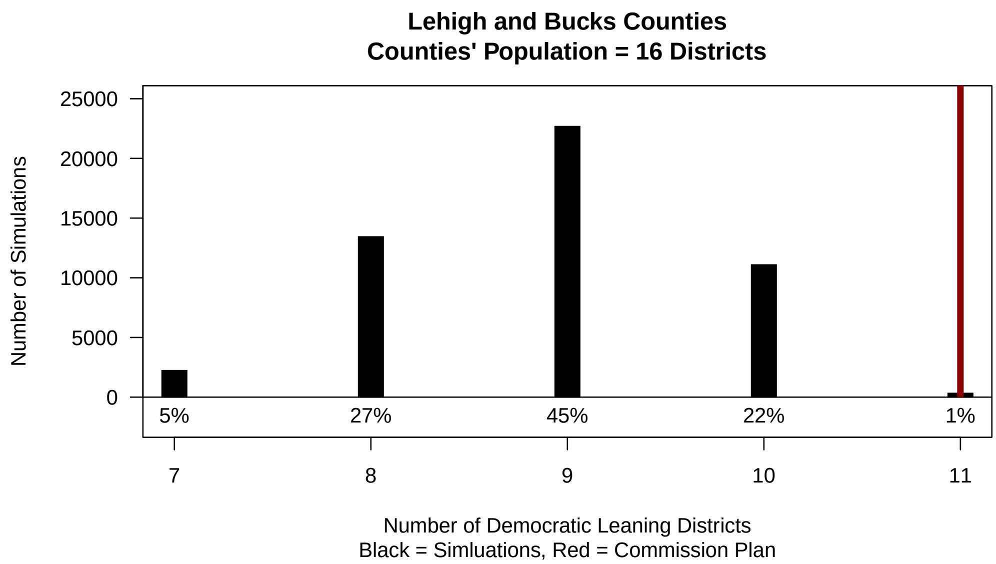
The combined population of Lehigh and Bucks counties is equal to approximately 16 legislative districts. In the 16 districts that cover this area, the Commission's proposal generates 11 Democratic leaning districts. The distribution of Democratic leaning districts based on the statewide partisan elections index calculated for each of the simulation results is shown in Figure 4. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the counties shown below each bar. The most common outcome in the simulations is 9 Democratic districts. The red vertical line at 11 represents the number of Democratic leaning seats in the Commission's map in the portion of the state. In 99% of the simulations there are fewer than 11 Democratic leaning districts in these counties. In only 1% of the simulations are there 11 Democratic leaning districts in these counties, as is the case in the Commission's proposed map.

The Commission's plan achieves this by dividing the city of Allentown in Lehigh County more than is necessary so as to more evenly distribute the Democratic voters that live in the city across more districts. Allentown is heavily Democratic and has a population of 126,364, which when divided by the target district size of 64,053 comes to approximately 1.97 districts. Thus, Allentown is too large to be completely contained in one district and will need to be divided into two districts. However, the Commission's plan divides the city into three districts. Figure 5 below shows this using two maps. The top panel shows a map of the Commission's proposed district boundaries in Lehigh County where Allentown is located. The bottom panel focuses exclusively on the city of Allentown and shows how the city is split into three different districts.

The next set of maps shows how this division follows the gerrymandering strategy of dividing Democratic cities into "pinwheel" shapes where Democratic voters in the city can be combined with less Democratic areas outside of the city to make more Democratic districts with comfortable margins, but not the overwhelmingly Democratic margins that would occur if fewer districts were drawn that were more geographically compact and split the city fewer times. In some cases this approach also has the effect of dividing minority communities that live in these cities and diluting their influence by distributing them across multiple legislative districts. Figure 6 shows a map of each of the three districts that intersect Allentown (HD-22, HD-134, HD-132). Each district is colored based on the partisan lean of the precincts in the district. The pattern we see, particularly in Districts 134 and 132, is exactly what I described earlier — the combination of heavily Democratic precincts in the center of the city with more Republican leaning precincts in the suburbs of the city. While Allentown itself is heavily Democratic (its partisan index based on the 2012-2020 statewide elections is 0.72), the inclusion of the more Republican leaning suburbs distributes Democrats more efficiently to create three Democratic leaning districts, two of which (HD-134 and HD-132) have less Democratic support, but are still comfortably Democratic.

The final map shows that this approach also divides the Latino population in the city. Figure 7. As a whole, Allentown has a Hispanic voting age population of 48.9%. While District 22 is majority Latino, Districts 134 and 132 have substantially lower Latino populations (38.5% and 18.1%, respectively) as a result of the districts dividing the city and reaching into more suburban areas with a lower concentration of Latinos.

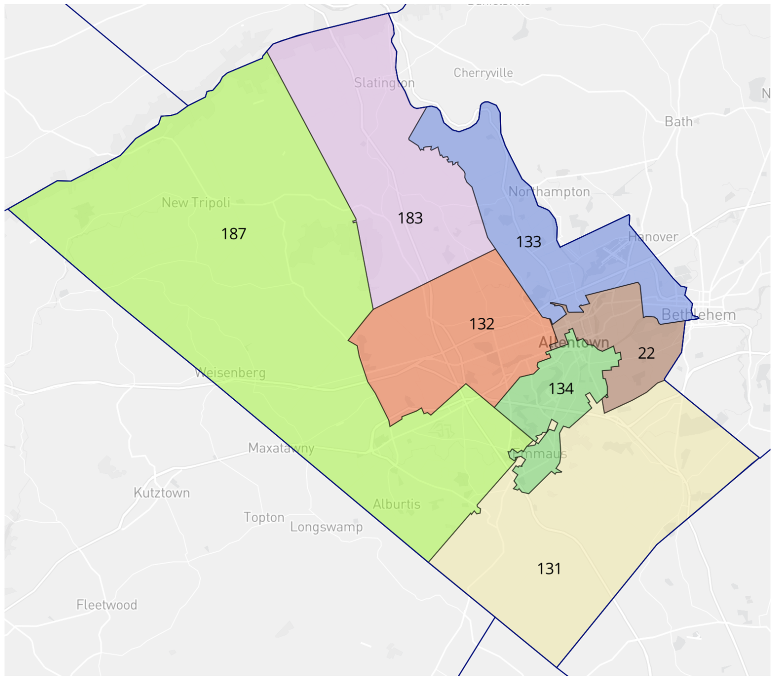
Figure 4: **Distribution of Partisan Districts from Simulations in Lehigh and Bucks Counties**



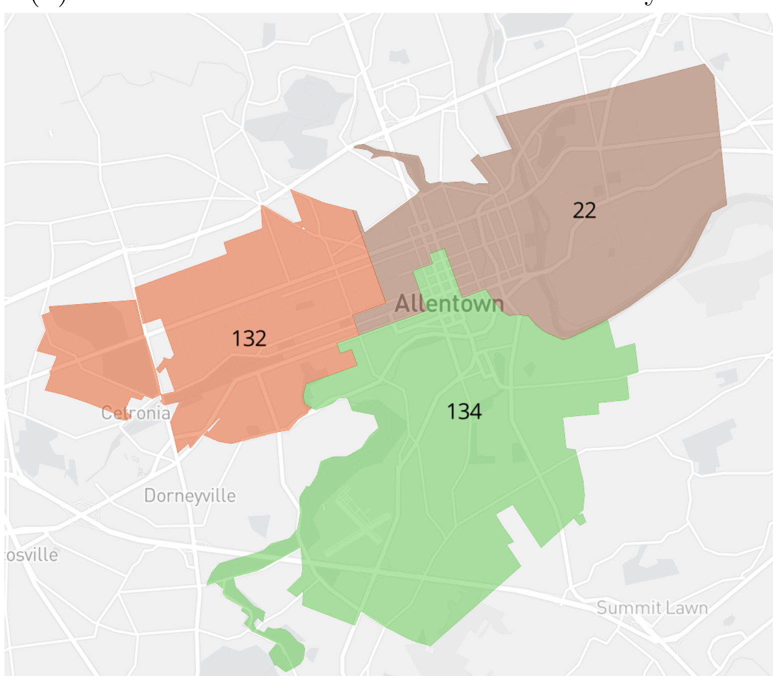
Note: Distribution of likely district partisanship based on the statewide partisan elections index calculated for each of the simulation results. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the cluster shown below each bar. The red vertical line shows the number of Democratic leaning seats in the Commission's proposed map in the same county.

Figure 5: **Commission Proposed Districts in Lehigh County**

(a) Proposal District Boundaries in Lehigh County



(b) District Boundaries within Allentown City Limits



Note: The top figure shows the district boundaries within Lehigh County. The bottom figure shows how the city of Allentown is divided across three districts despite having a population that only requires it to be split into two districts. In each district we see a combination of heavily Democratic urban center with less Democratic suburban areas at the outer edges of the district.

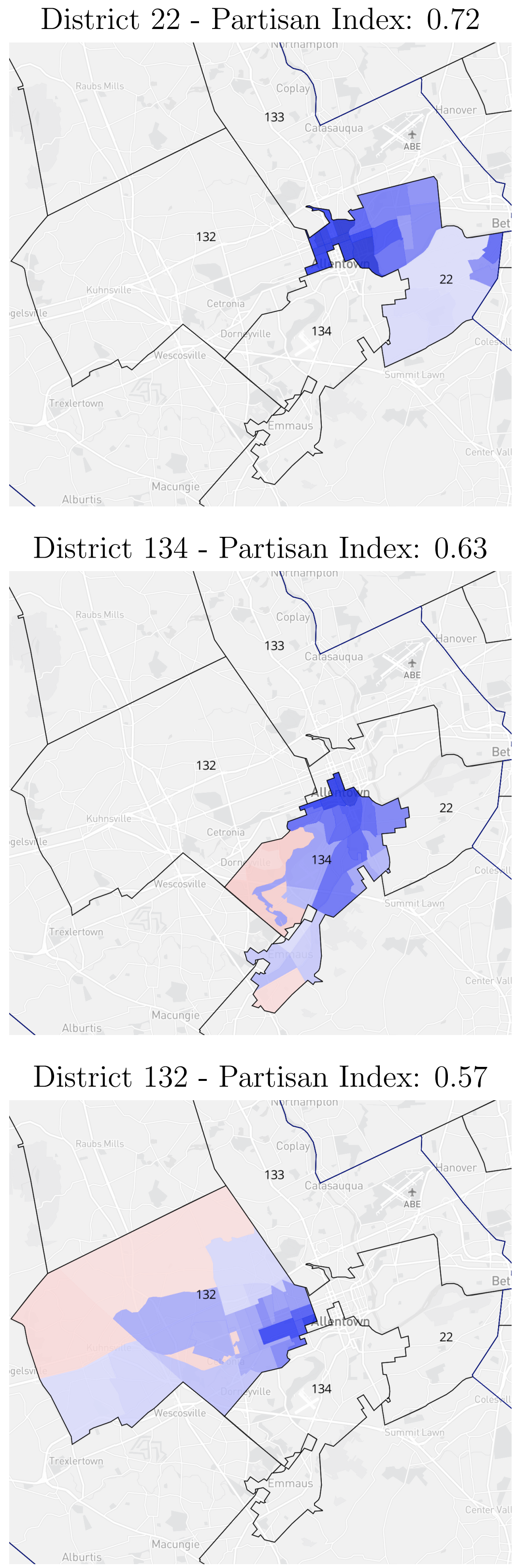


Figure 6: Note: Each panel shows one of the districts that intersect Allentown. The maps are colored according to the partisan composition of precincts in the district.

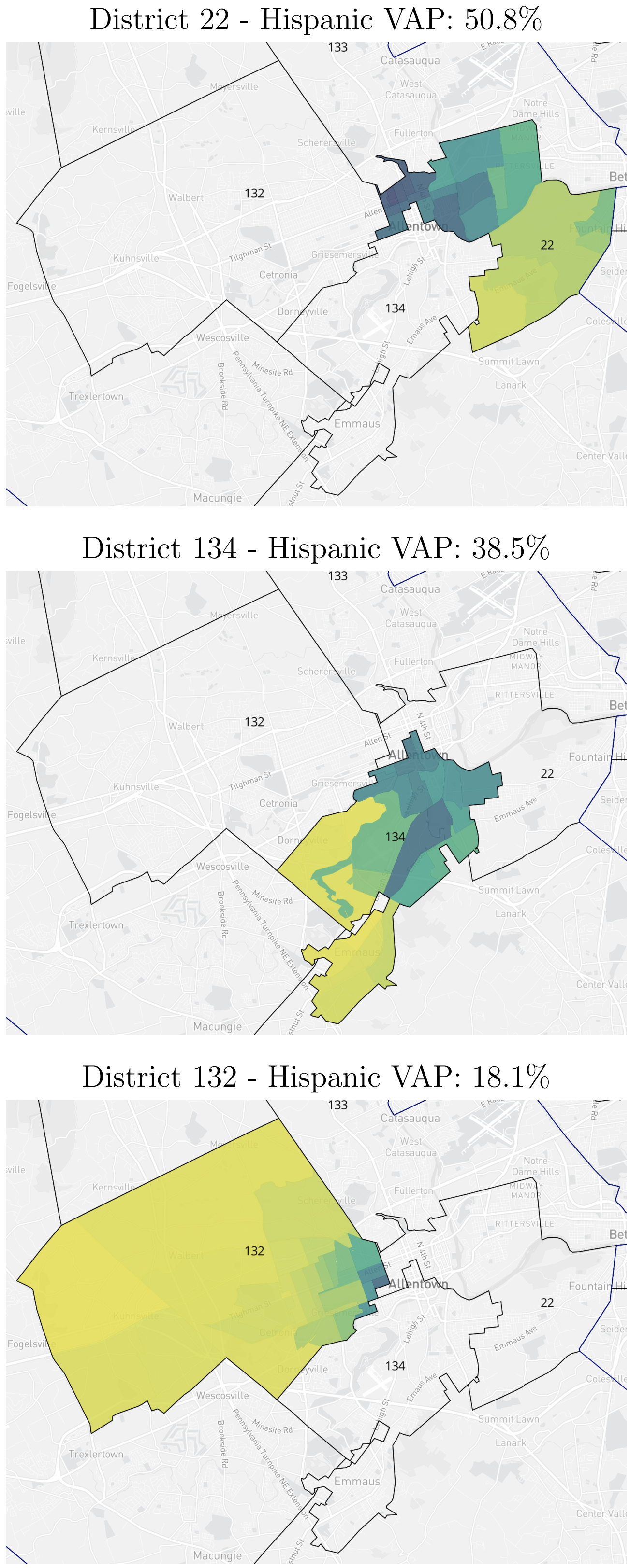


Figure 7: Each panel shows one of the districts that intersect Allentown. The maps are colored according to the Hispanic composition of precincts in the district. Darker shades indicate a greater proportion of Latinos. The city of Allentown has a 48.9% Hispanic voting age population.

5.2 Schuylkill, Berks, Lancaster, and Lebanon Counties

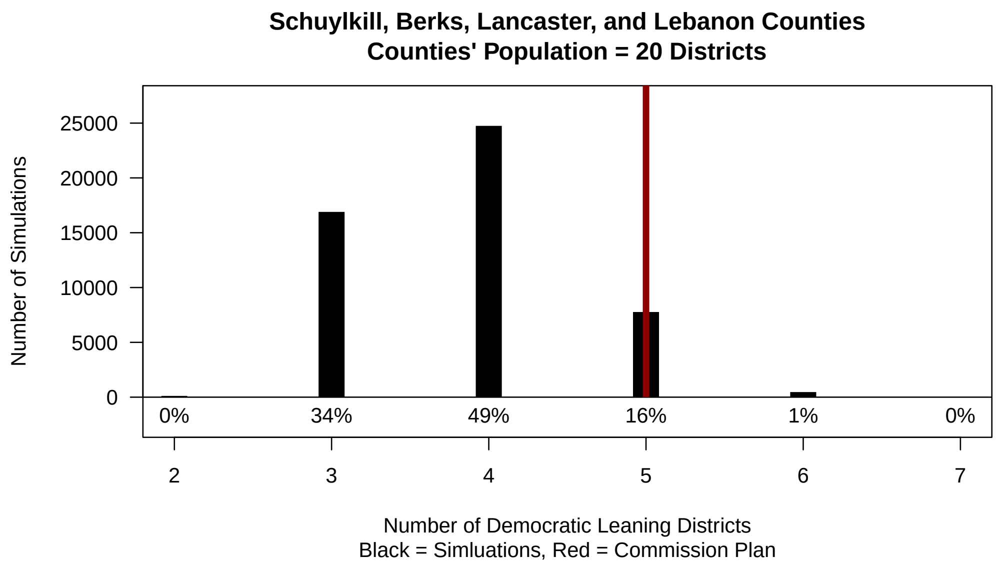
The combined population of Schuylkill, Berks, Lancaster, and Lebanon counties is equal to approximately 20 legislative districts. In the 20 districts that cover this area, the Commission's proposal generates 5 Democratic leaning districts. The distribution of Democratic leaning districts based on the statewide partisan elections index calculated for each of the simulation results is shown in Figure 8. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the counties shown below each bar. The most common outcome in the simulations is 4 Democratic districts. The red vertical line at 5 represents the number of Democratic leaning seats in the Commission's map in the portion of the state. In 83.5% of the simulations there are fewer than 5 Democratic leaning districts in these counties. In only 17% of the simulations are there 5 or more Democratic leaning districts in these counties, as is the case in the Commission's proposed map.

The Commission's plan achieves this by dividing the cities of Lancaster in Lancaster County and Reading in Berks County more than is necessary so as to more evenly distribute the Democratic voters that live in these cities across more districts. Lancaster is heavily Democratic and has a population of 58,431, which when divided by the target district size of 64,053 comes to approximately 0.91 districts. Thus, Lancaster is not larger than the target district population and could be kept whole. However, the Commission's plan divides the city nearly evenly into two districts. Figure 9 below shows this using two maps. The top panel shows a map of the Commission's proposed district boundaries in Lancaster County where the city of Lancaster is located. The bottom panel focuses exclusively on the city of Lancaster and shows how the city is split into two different districts.

The next set of maps shows how this division follows the gerrymandering strategy of dividing heavily Democratic cities and combining them with less Democratic areas outside of the city to make more Democratic districts with comfortable margins, but not the overwhelmingly Democratic margins that would occur if the city were kept whole. In Lancaster this approach also has the effect of dividing and diluting the influence of the Latino community that lives in the city by distributing them across multiple legislative districts. Figure 10 shows a map of each of the two districts that intersect Lancaster (HD-50, HD-96). Each district is colored based on the partisan lean of the precincts in the district. The pattern we see is familiar — the combination of heavily Democratic precincts in the center of the city with more Republican leaning precincts in the suburbs of the city. While Lancaster itself is heavily Democratic (its partisan index based on the 2012-2020 statewide elections is 0.76), the inclusion of the more Republican leaning suburbs distributes Democrats more efficiently to create two Democratic leaning districts rather than one district that is overwhelmingly Democratic.

The final map shows that this approach also divides the Latino population in the city. Figure 11. As a whole, Lancaster has a Latino voting age population of 35.9%. Both Districts 96 and 50 have a lower Latino population (13.7% and 32.8%, respectively) as a result of the districts dividing the city and reaching into more suburban areas with a lower concentration of Latinos.

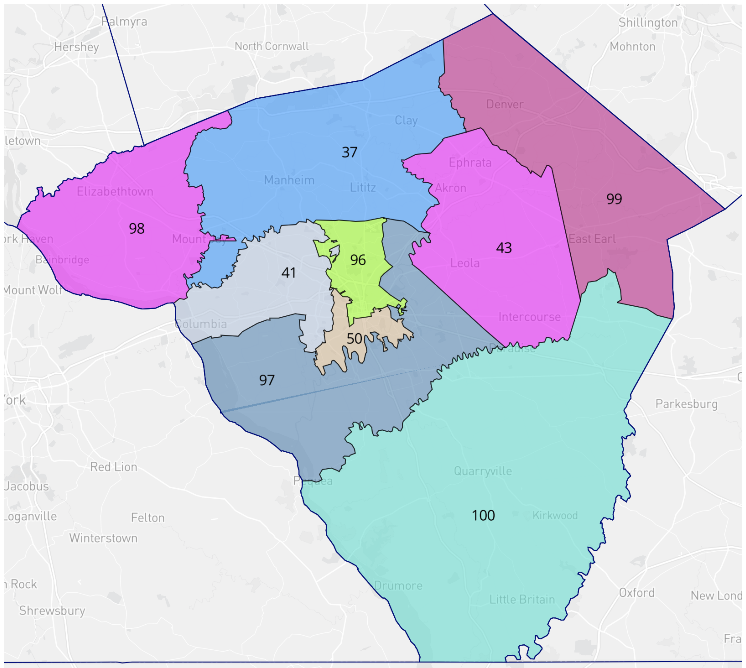
Figure 8: **Distribution of Partisan Districts from Simulations in Schuylkill, Berks, Lancaster, and Lebanon Counties**



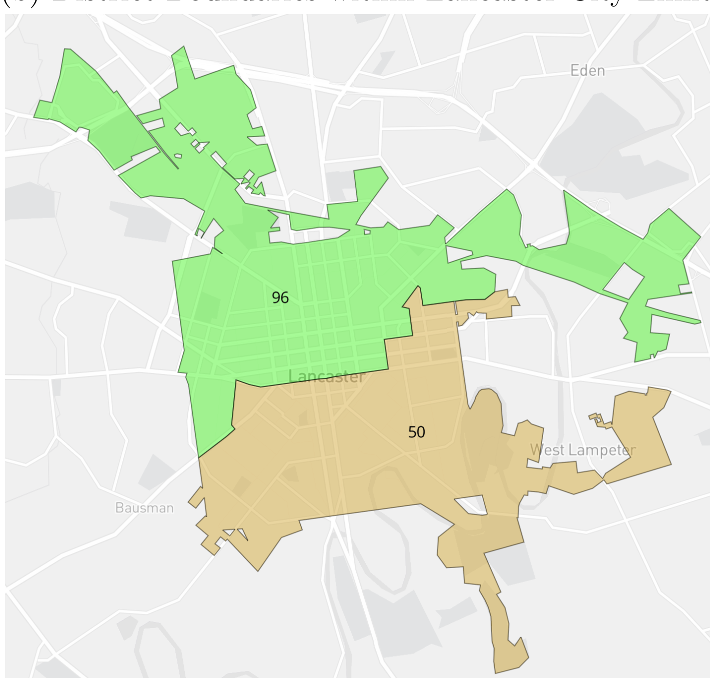
Note: Distribution of likely district partisanship based on the statewide partisan elections index calculated for each of the simulation results. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the cluster shown below each bar. The red vertical line shows the number of Democratic leaning seats in the Commission's proposed map in the same county.

Figure 9: **Commission Proposed Districts in Lancaster County**

(a) Proposal District Boundaries in Lancaster County



(b) District Boundaries within Lancaster City Limits



Note: The top figure shows the district boundaries within Lancaster County. The bottom figure shows how the city of Lancaster is divided nearly equally across two districts despite having a population that would allow the city to be entirely contained in one district.

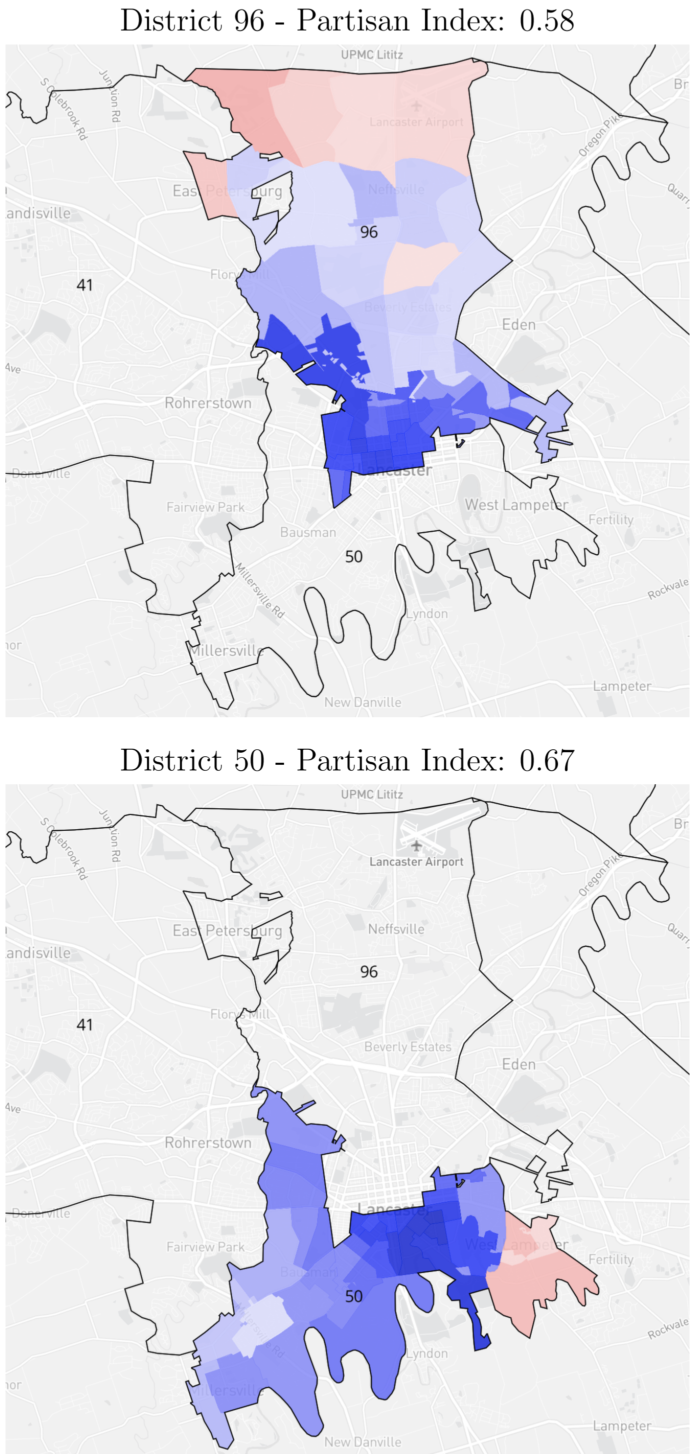


Figure 10: Each panel shows one of the districts that intersect Lancaster. The maps are colored according to the partisan composition of precincts in the district.

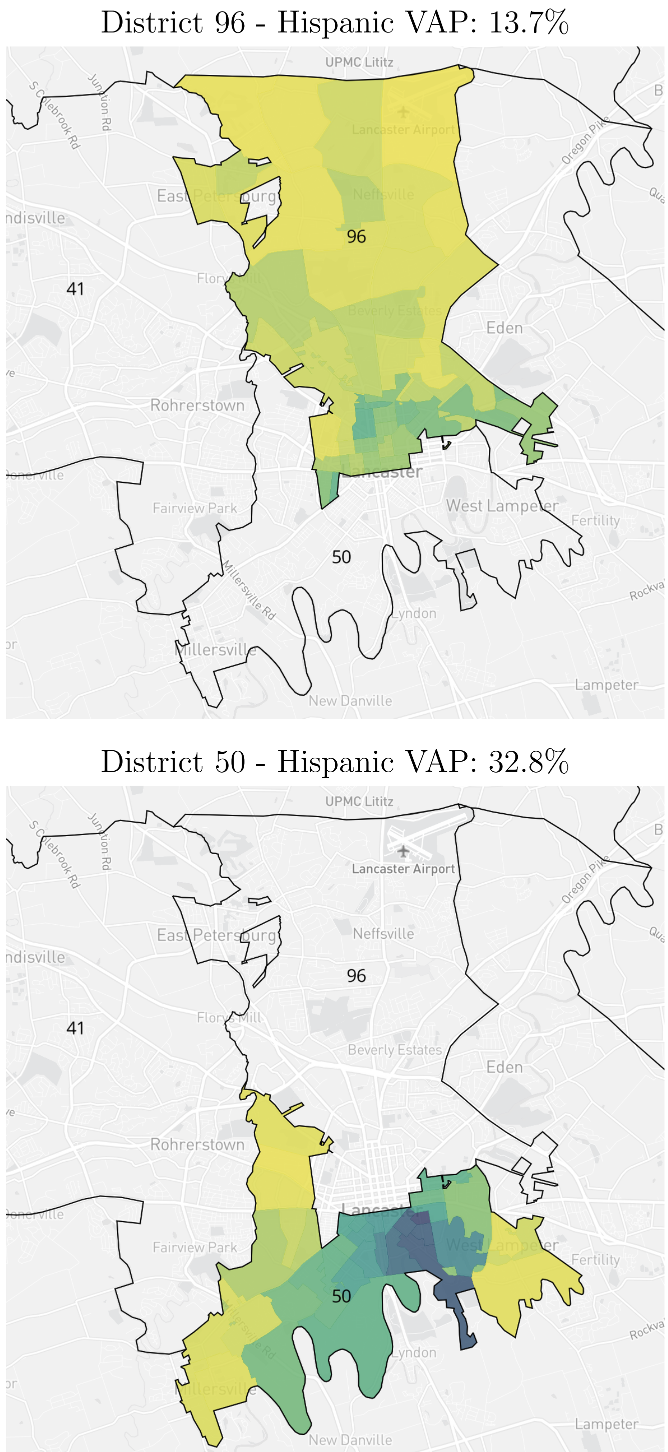


Figure 11: Each panel shows one of the districts that intersect Lancaster. The maps are colored according to the Hispanic composition of precincts in the district. Darker shades indicate a greater proportion of Latinos. The city of Lancaster has a 35.9% Hispanic voting age population.

In Berks County the Commission's plan creates an additional Democratic district by dividing the city of Reading more than is necessary. Reading is heavily Democratic and has a population of 95,719, which when divided by the target district size of 64,053 comes to approximately 1.49 districts. Thus, Reading is too large to be completely contained in one district and will need to be divided into two districts. However, the Commission's plan divides the city four different times into three different districts. Figure 12 below shows this using two maps. The top panel shows a map of the Commission's proposed district boundaries in Berks County where Reading is located. The bottom panel focuses exclusively on the city of Reading and shows how the city is split four times into three different districts.

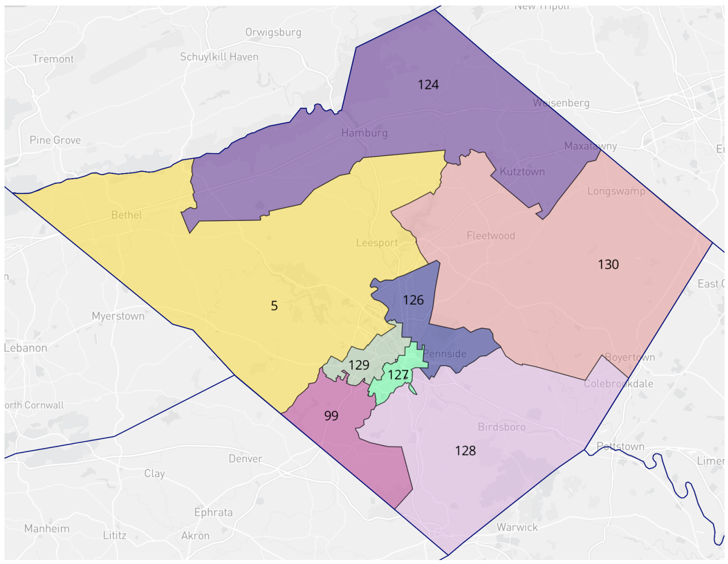
The next set of maps shows how this division follows the gerrymandering strategy of dividing Democratic cities into "pinwheel" shapes where Democratic voters in the city can be combined with less Democratic areas outside of the city to make more Democratic districts with comfortable margins, but not the overwhelmingly Democratic margins that would occur if fewer districts were drawn that were more geographically compact and split the city fewer times. In Reading this approach also has the effect of dividing and diluting the influence of the Latino community that lives in the city by distributing them across multiple legislative districts. Figure 13 shows a map of each of the three districts that intersect Reading (HD-126, HD-127, and HD-129). Each district is colored based on the partisan lean of the precincts in the district. The pattern we see is again repeated — the combination of heavily Democratic precincts in the center of the city with more Republican leaning precincts in the suburbs. While Reading itself is heavily Democratic (its partisan index based on the 2012-2020 statewide elections is 0.79), the inclusion of the more Republican leaning suburbs distributes Democrats more efficiently to create three Democratic leaning districts which all have less Democratic support than the city overall, but are still comfortably Democratic.

The final map shows that this approach also divides the Latino population in the city. Figure 14. As a whole, Reading has a Latino voting age population of 64.0%. All three Districts that intersect Reading have a lower Latino population (35.5% in HD-126, 35.4%

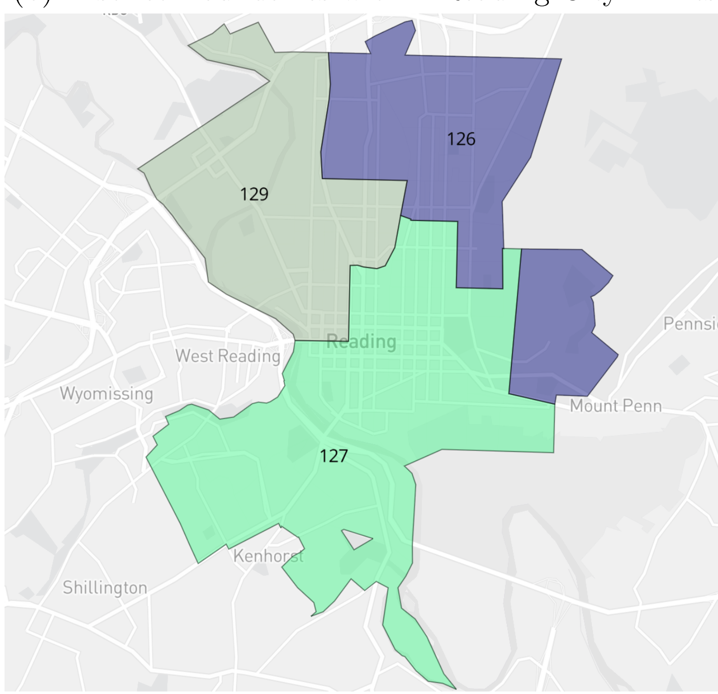
in HD-129, and 51.7% in HD-127) as a result of the districts dividing the city and reaching into more suburban areas with a lower concentration of Latinos.

Figure 12: **Commission Proposed Districts in Berks County**

(a) Proposal District Boundaries in Berks County

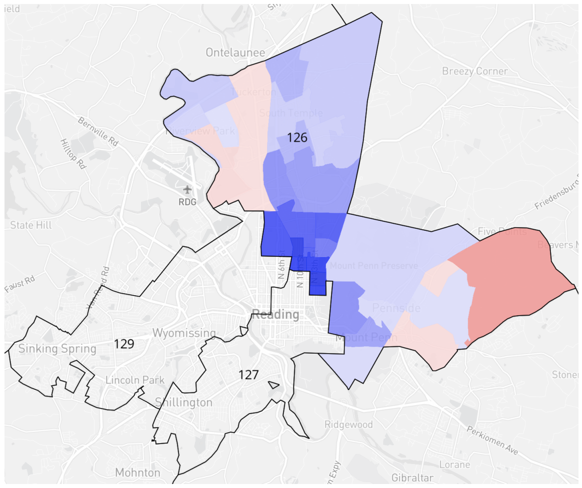


(b) District Boundaries within Reading City Limits

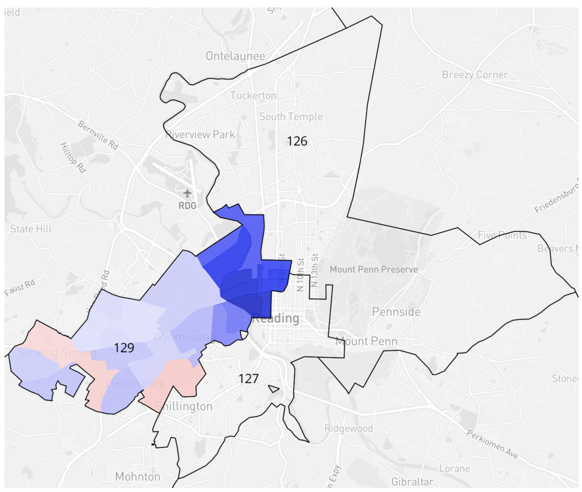


Note: The top figure shows the district boundaries within Berks County. The bottom figure shows how the city of Reading is divided four times into three districts despite having a population that would only require the city to be split into two districts.

District 126 - Partisan Index: 0.60



District 129 - Partisan Index: 0.60



District 127 - Partisan Index: 0.70

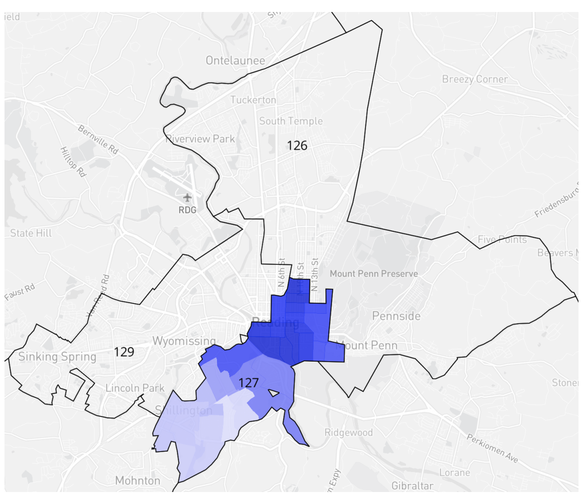


Figure 13: Each panel shows one of the districts that intersect Reading. The maps are colored according to the partisan composition of precincts in the district.

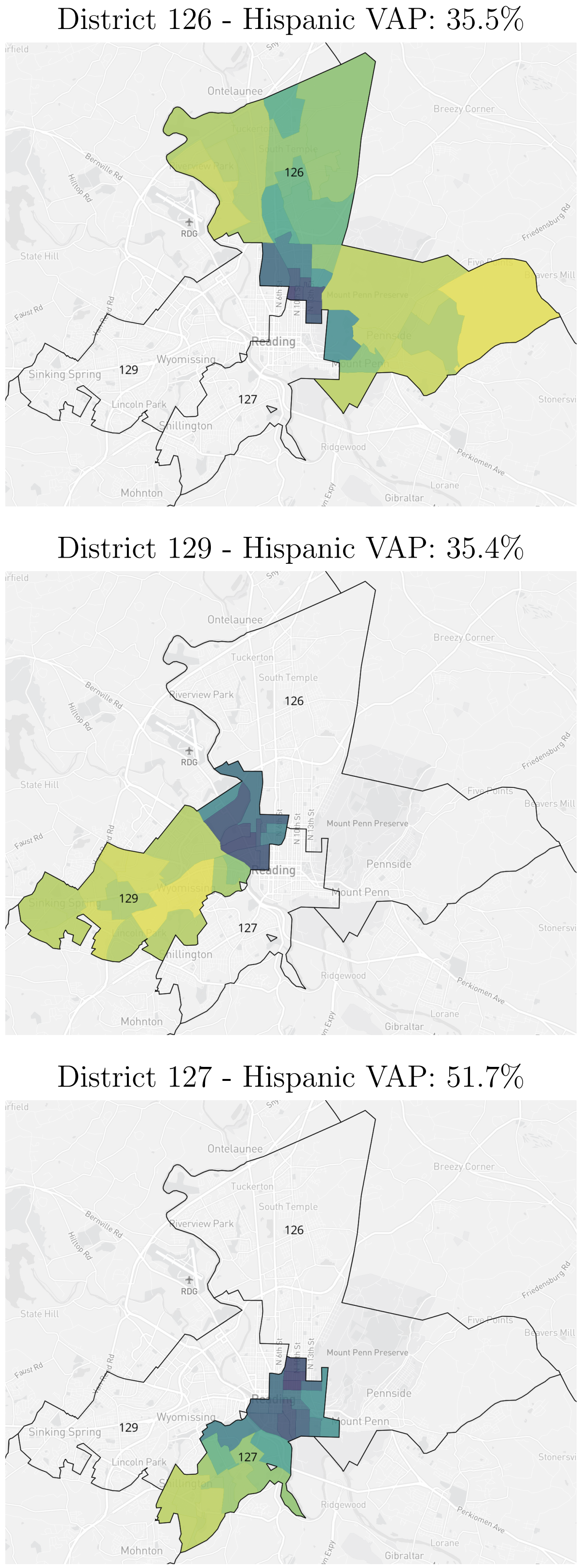


Figure 14: Each panel shows one of the districts that intersect Reading. The maps are colored according to the Hispanic composition of precincts in the district. Darker shades indicate a greater proportion of Latinos. The city of Reading has a 64.0% Hispanic voting age population.

5.3 Dauphin and Cumberland Counties

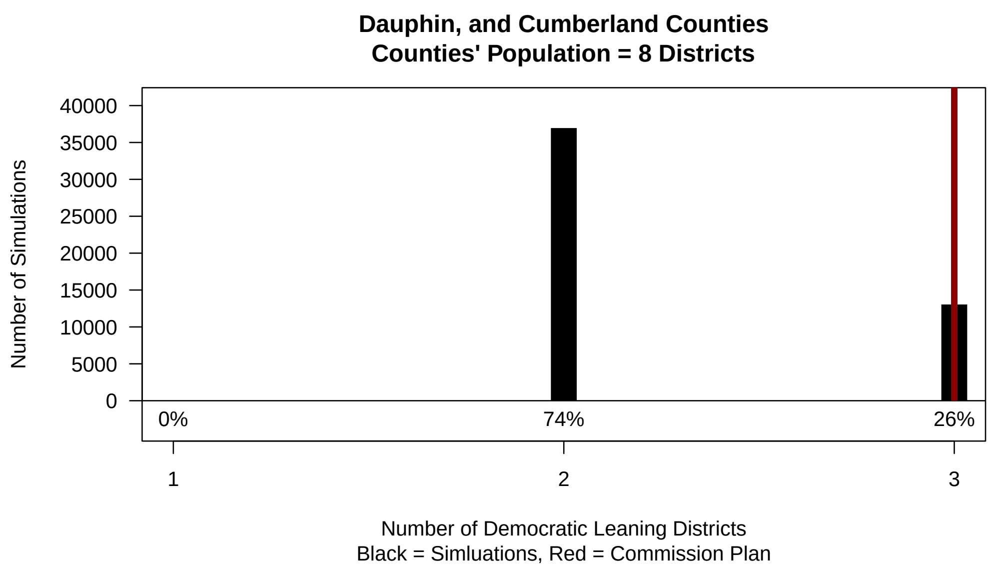
The combined population of Dauphin and Cumberland counties is equal to approximately 8.5 legislative districts. In the 8 complete districts that cover this area, the Commission's proposal generates 3 Democratic leaning districts. The distribution of Democratic leaning districts based on the statewide partisan elections index calculated for each of the simulation results is shown in Figure 15. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the counties shown below each bar. The most common outcome in the simulations is 2 Democratic districts. The red vertical line at 3 represents the number of Democratic leaning seats in the Commission's map in the portion of the state. In 74% of the simulations there are 2 Democratic leaning districts in these counties. There are 3 Democratic leaning districts in only 26% of the simulations in these counties, which is what the Commission's proposed map produces.

The Commission's plan achieves this by dividing the city of Harrisburg in Dauphin County more than is necessary so as to more evenly distribute the Democratic voters that live in Harrisburg across more districts. Harrisburg is heavily Democratic and has a population of 50,679, which when divided by the target district size of 64,053 comes to approximately 0.79 districts. Thus, Harrisburg is not larger than the target district population and could be kept whole. However, the Commission's plan divides the city into two districts. Figure 16 below shows this using two maps. The top panel shows a map of the Commission's proposed district boundaries in Dauphin County where the city of Harrisburg is located. The bottom panel focuses exclusively on the city of Harrisburg and shows how the city is split into two districts.

The next set of maps shows how this division follows the gerrymandering strategy of dividing Democratic cities into "pinwheel" shapes where Democratic voters in the city can be combined with less Democratic areas outside of the city to make more Democratic districts with comfortable margins, but not the overwhelmingly Democratic margins that would occur if fewer districts were drawn that were more geographically compact and split the city fewer times. In Harrisburg this approach also has the effect of dividing the Black community that lives in the city and distributes them across multiple legislative districts. Figure 17 shows a map of each of the two districts that intersect Harrisburg (HD-103, HD-104). Each district is colored based on the partisan lean of the precincts in the district. The pattern we see is again repeated — the combination of heavily Democratic precincts in the center of the city with more Republican leaning precincts in the suburbs. While Harrisburg itself is heavily Democratic (its partisan index based on the 2012-2020 statewide elections is 0.86), the inclusion of the more Republican leaning suburbs distributes Democrats more efficiently to create two Democratic leaning districts that have less Democratic support, but are still comfortably Democratic-leaning.

Figure 18 shows that this approach also divides the Black population in the city. As a whole, Harrisburg has a Black voting age population of 47.3%. Both districts that intersect Harrisburg have a lower Black population (22.2% in HD-103, 31.0% in HD-104) as a result of the districts dividing the city and reaching into more suburban areas with a lower Black population.

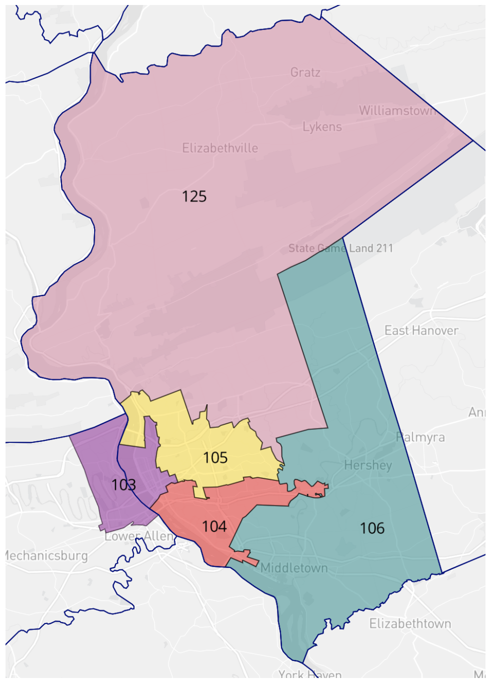
Figure 15: **Distribution of Partisan Districts from Simulations in Dauphin, and Cumberland Counties**



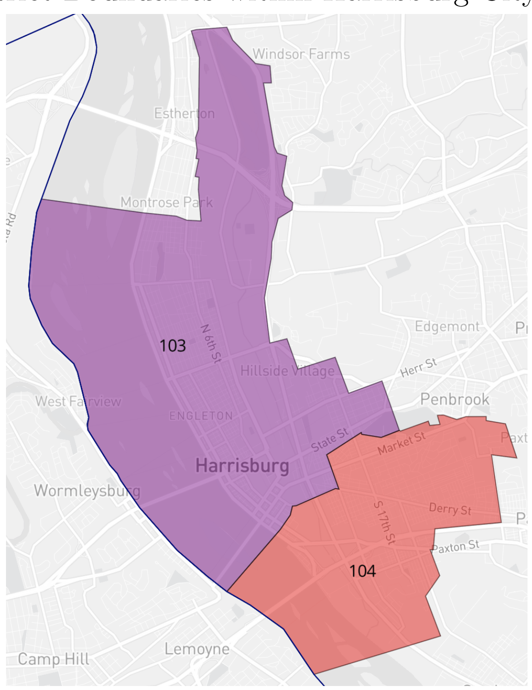
Note: Distribution of likely district partisanship based on the statewide partisan elections index calculated for each of the simulation results. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the cluster shown below each bar. The red vertical line shows the number of Democratic leaning seats in the Commission's proposed map in the same county.

Figure 16: **Commission Proposed Districts in Dauphin County**

(a) Proposal District Boundaries in Dauphin County

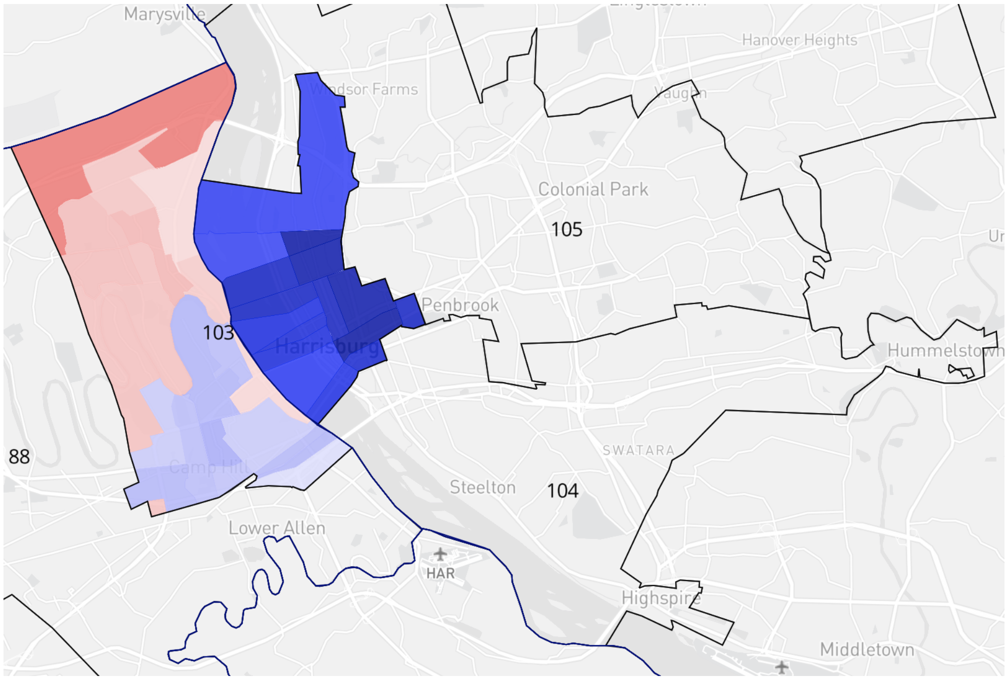


(b) District Boundaries within Harrisburg City Limits



Note: The top figure shows the district boundaries within Dauphin County. The bottom figure shows how the city of Harrisburg is divided across two districts despite having a population that would allow the city to be entirely contained in one district.

District 103 - Partisan Index: 0.62



District 104 - Partisan Index: 0.67

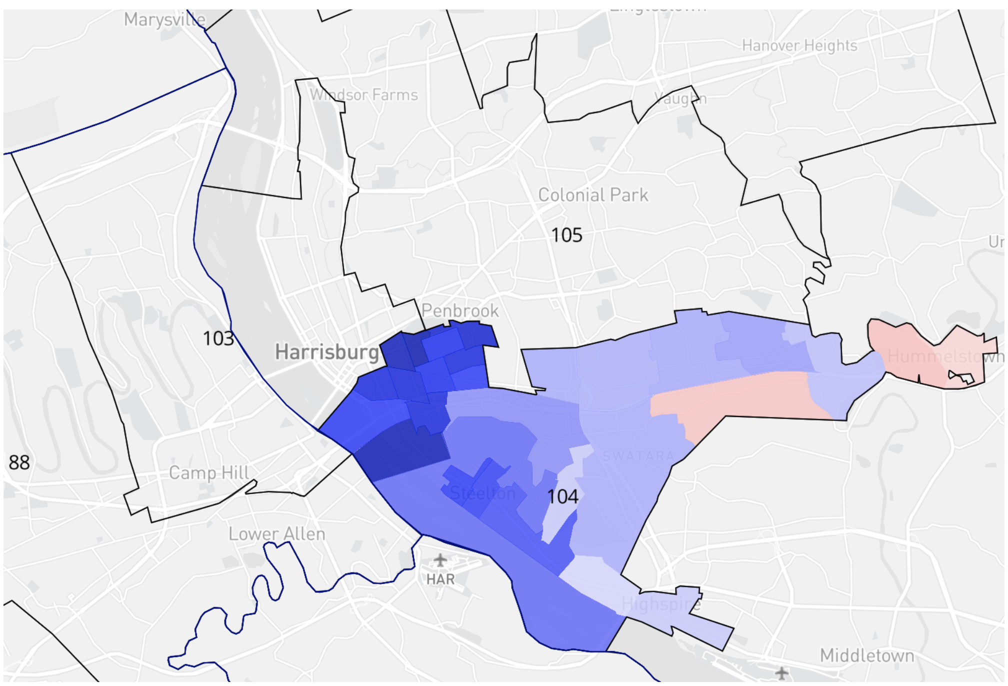
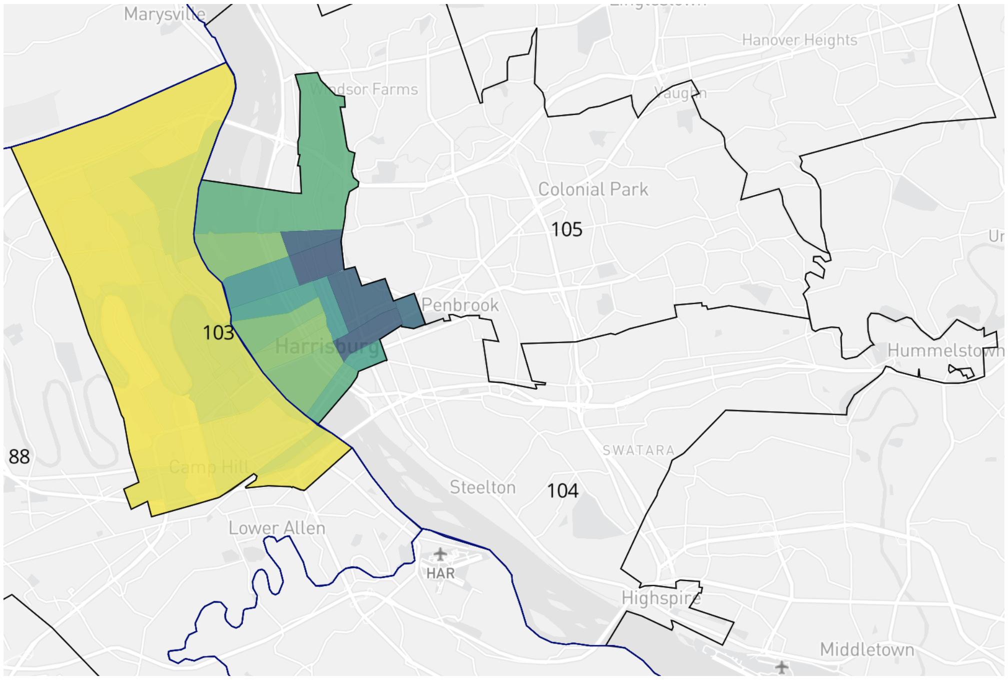


Figure 17: Each panel shows one of the districts that intersect Harrisburg. The maps are colored according to the partisan composition of precincts in the district.

District 103 - Black VAP: 22.2%



District 104 - Black VAP: 31.0%

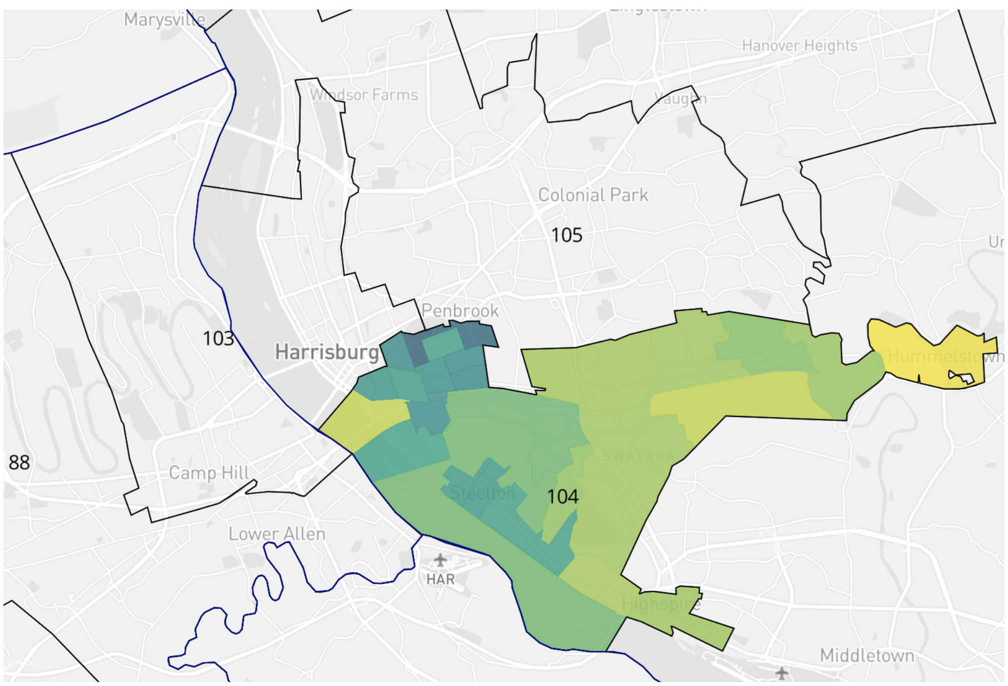


Figure 18: Each panel shows one of the districts that intersect Harrisburg. The maps are colored according to the Black composition of precincts in the district. Darker shades indicate a greater Black population. The city of Harrisburg has a 47.3% Black voting age population.

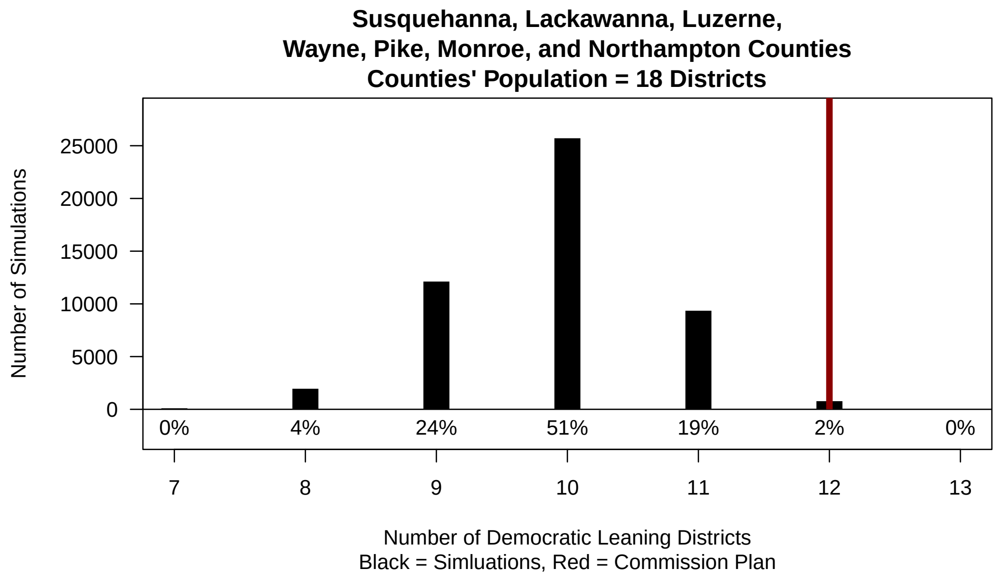
5.4 Northeastern Counties

In this section I consider Susquehanna, Lackawanna, Luzerne, Wayne, Pike, Monroe, and Northampton counties. These counties are grouped together in the northeastern part of the state, and their combined population is equal to approximately 18 legislative districts. In the 18 complete districts that cover this area, the Commission's proposal generates 11 Democratic leaning districts. The distribution of Democratic leaning districts based on the statewide partisan elections index calculated for each of the simulation results is shown in Figure 19. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the counties shown below each bar. The most common outcome in the simulations is 10 Democratic districts. The red vertical line at 11 represents the number of Democratic leaning seats in the Commission's map in the portion of the state. In 98.5% of the simulations there are 10 or fewer Democratic leaning districts in these counties. In only 2% of the simulations are there 11 Democratic leaning districts in these counties, as is the case in the Commission's proposed map.

The Commission's plan achieves this by dividing the city of Scranton in Lackawanna County more than is necessary so as to more evenly distribute the Democratic voters that live in Scranton across more districts. Scranton is heavily Democratic and has a population of 76,627, which when divided by the target district size of 64,053 comes to approximately 1.2 districts. Thus, Scranton is too large to be completely contained in one district and will need to be divided into two districts. However, the Commission's plan divides the city five different times across four different districts. Figure 20 below shows two maps. The top panel shows a map of the Commission's proposed district boundaries in Lackawanna County where Scranton is located. The bottom panel focuses exclusively on the city of Scranton and shows how the city is split five times into four different districts.

The next set of maps shows how this division follows the gerrymandering strategy of dividing Democratic cities into "pinwheel" shapes where Democratic voters in the city can be combined with less Democratic areas outside of the city to make more Democratic districts with comfortable margins, but not the overwhelmingly Democratic margins that would occur if fewer districts were drawn that were more geographically compact and split the city fewer times. Figure 21 shows a map of each of the four districts that intersect Scranton (HD-112, HD-113, HD-114, HD-118). Each district is colored based on the partisan lean of the precincts in the district. The pattern we see is yet again repeated — the combination of heavily Democratic precincts in the center of the city with more Republican leaning precincts in the suburbs around the city. While Scranton itself is heavily Democratic (its partisan index based on the 2012-2020 statewide elections is 0.70), the inclusion of the more Republican leaning suburbs distributes Democrats more efficiently to create four Democratic leaning districts that have less Democratic support, but are still comfortably Democratic-leaning. Scranton does not have a large or geographically concentrated minority population to warrant a specific analysis on how the districts in this county divide specific minority groups in the city (the city has a 71.9% White voting age population, 12.9% Hispanic VAP, 8.5% Black VAP, and 6.1% Asian VAP).

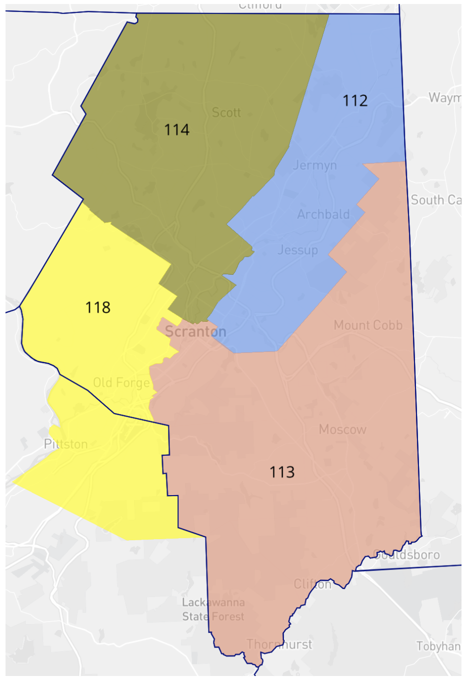
Figure 19: **Distribution of Partisan Districts from Simulations in Susquehanna, Lackawanna, and Luzerne Counties**



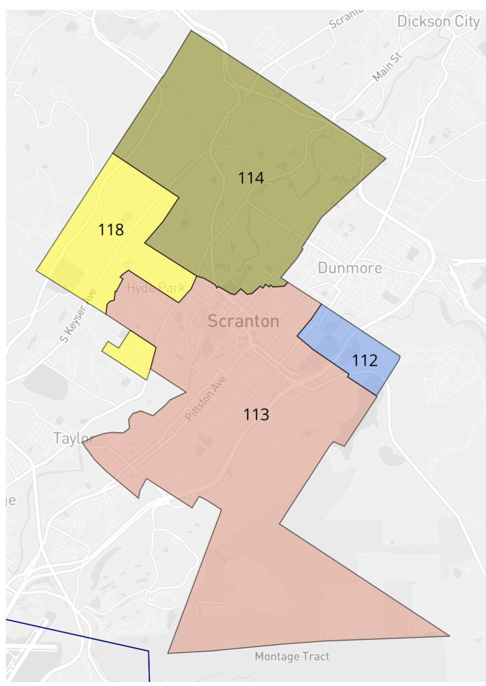
Note: Distribution of likely district partisanship based on the statewide partisan elections index calculated for each of the simulation results. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the cluster shown below each bar. The red vertical line shows the number of Democratic leaning seats in the Commission's proposed map in the same county.

Figure 20: **Commission Proposed Districts in Lackawanna County**

(a) Proposal District Boundaries in Lackawanna County



(b) District Boundaries within Scranton City Limits



Note: The top figure shows the district boundaries within Lackawanna County. The bottom figure shows how the city of Scranton is divided five times across four districts despite having a population that would only require the city to be divided into two districts.

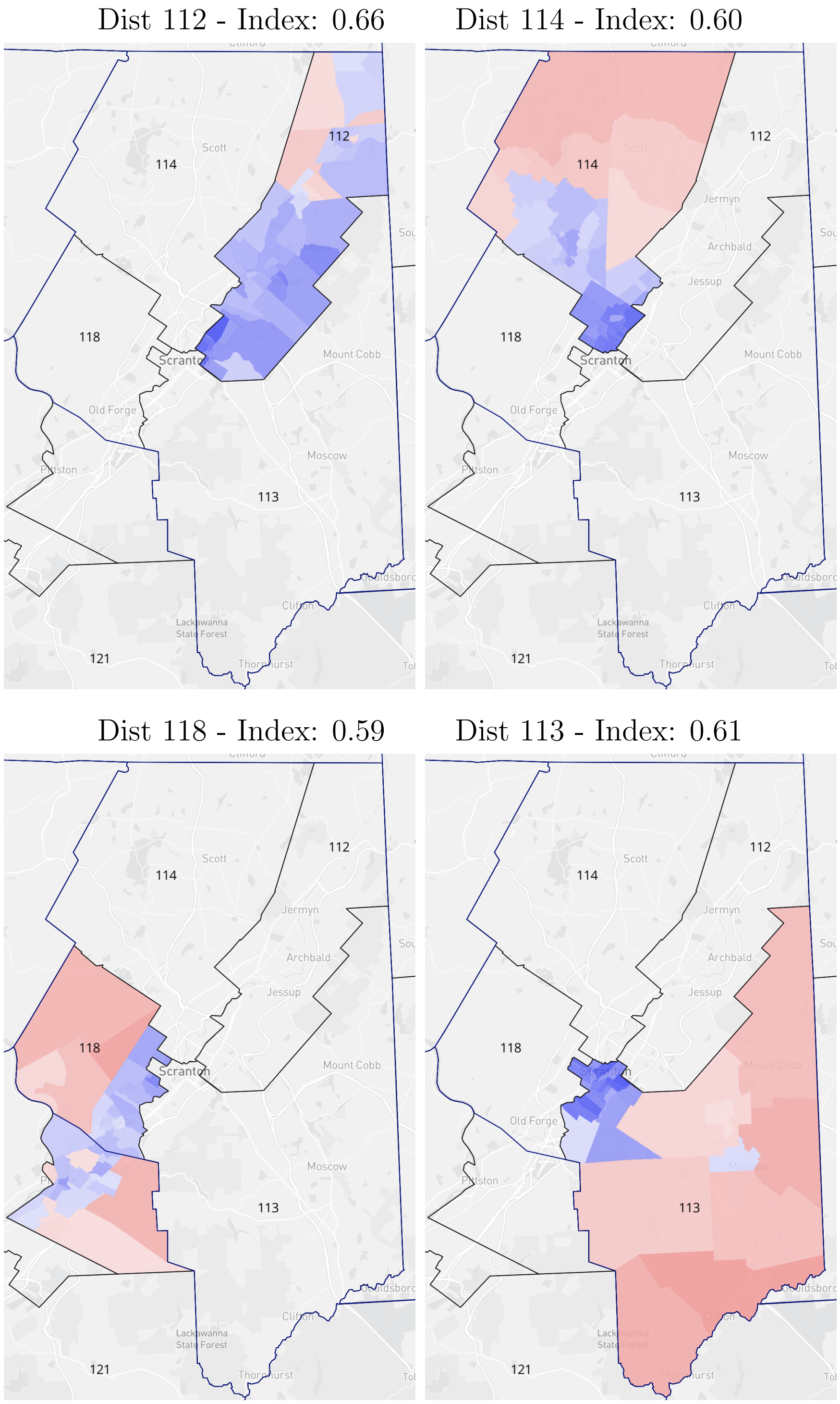


Figure 21: Each panel shows one of the districts that intersect Scranton. The maps are colored according to the partisan composition of precincts in the district.

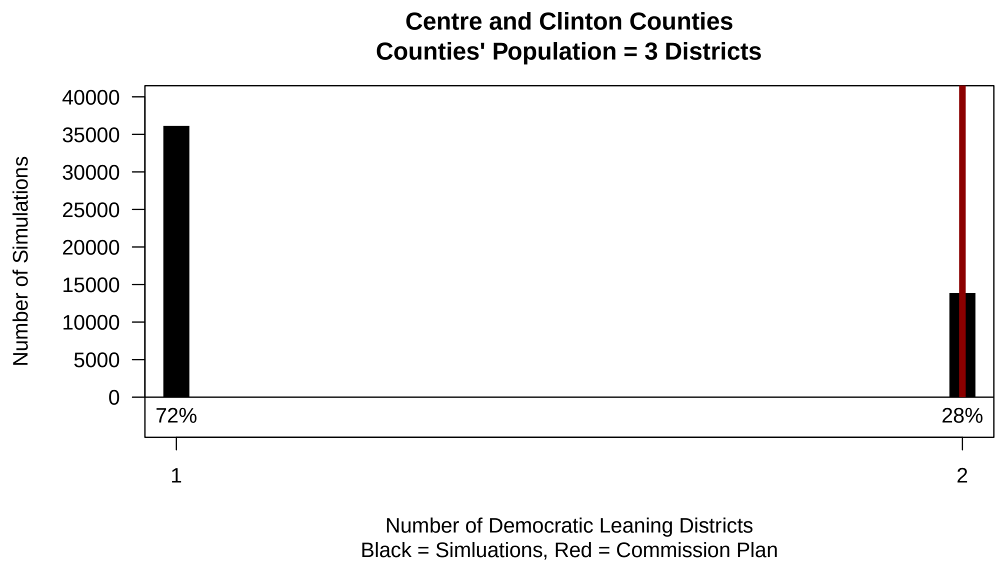
5.5 Centre and Clinton Counties

The final area I consider is the middle of the state in Centre and Clinton counties. The combined population of Centre and Clinton counties is equal to approximately 3 legislative districts. In the 2 complete districts that are included in these counties and the 2 additional districts that are partially in these counties, the Commission's proposal generates 2 Democratic leaning districts. The distribution of Democratic leaning districts based on the statewide partisan elections index calculated for each of the simulation results is shown in Figure 22. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the counties shown below each bar. The most common outcome in the simulations is 1 Democratic district. The red vertical line at 2 represents the number of Democratic leaning seats in the Commission's map in the portion of the state. The simulations generate 1 Democratic leaning district in these counties 72% of the time. There 2 Democratic leaning districts in only 28% of the simulations, as is the case in the Commission's proposed map.

The Commission's plan achieves this by dividing the borough of State College in Centre County more than is necessary so as to more evenly distribute the Democratic voters that live in this city across more districts. State College is heavily Democratic and has a population of 40,508, which when divided by the target district size of 64,053 comes to approximately 0.63 districts. Thus, State College is not larger than the target district population and could be kept whole. However, the Commission's plan divides the city nearly equally into two districts. Figure 23 below shows two maps. The top panel shows a map of the Commission's proposed district boundaries in Centre County where the borough of State College is located. The bottom panel focuses exclusively on the city of State College and shows how the city is split into two different districts. The Commission's plan takes most of the Penn State University campus and combines it with the more rural part of western Centre County as District 77 while the rest of State College is placed in a district with the rural northern and southern portions of the county in District 82.

The next set of maps shows how this division follows the gerrymandering strategy of dividing Democratic cities into "pinwheel" shapes where Democratic voters in the city can be combined with less Democratic areas outside of the city to make more Democratic districts with comfortable margins, but not the overwhelmingly Democratic margins that would occur if fewer districts were drawn that were more geographically compact and split the city fewer times. Figure 24 shows a map of each of the two districts that intersect State College (HD-77, HD-82). Each district is colored based on the partisan lean of the precincts in the district. The pattern we see is yet again repeated — the combination of heavily Democratic precincts in the center of the city with more Republican leaning precincts in the suburbs. While State College itself is heavily Democratic (its partisan index based on the 2012-2020 statewide elections is 0.70), the inclusion of the more Republican leaning suburbs distributes Democrats more efficiently to create two Democratic leaning districts that have less Democratic support, but are still comfortably Democratic-leaning. State College does not have a large or geographically concentrated minority population to warrant a specific analysis on how the districts in this county divide specific minority groups in the city (the city has a 77.6% White voting age population, 5.5% Hispanic VAP, 3.6% Black VAP, and 12.0% Asian VAP).

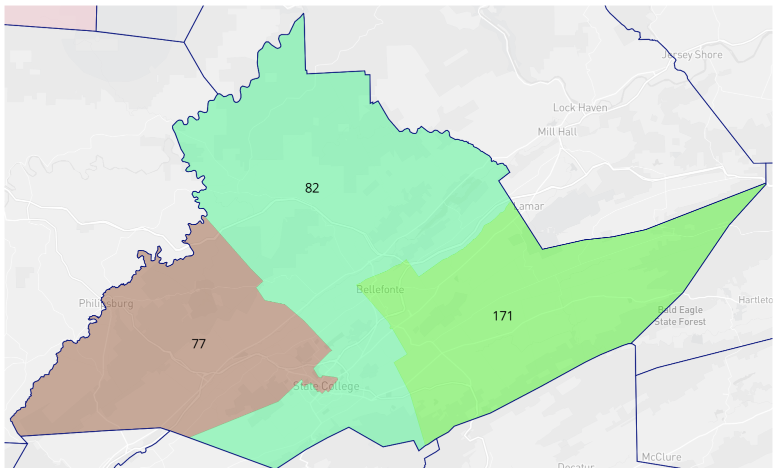
Figure 22: **Distribution of Partisan Districts from Simulations in Centre and Clinton Counties**



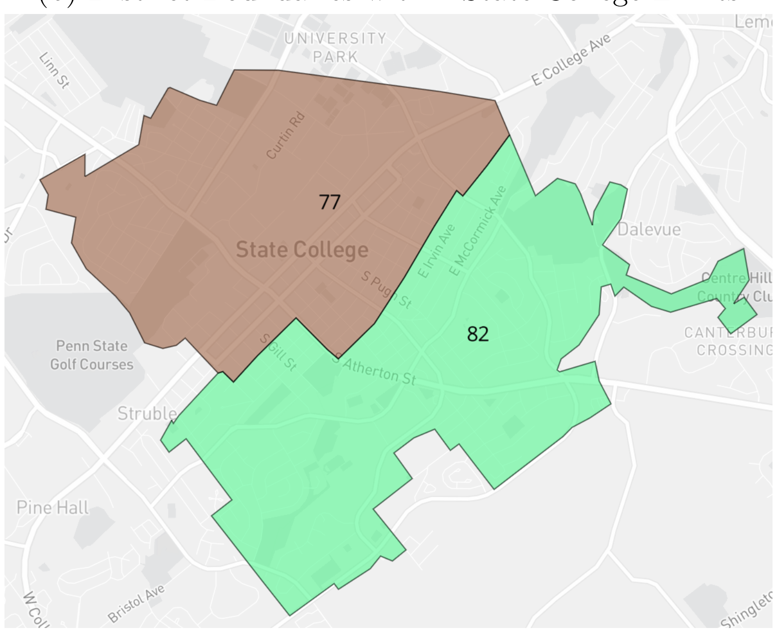
Note: Distribution of likely district partisanship based on the statewide partisan elections index calculated for each of the simulation results. The black bars show the distribution from the simulation results, with the percentage of simulations that generate each of the various possible number of Democratic seats in the cluster shown below each bar. The red vertical line shows the number of Democratic leaning seats in the Commission's proposed map in the same county.

Figure 23: **Commission Proposed Districts in Centre County**

(a) Proposal District Boundaries in Centre County

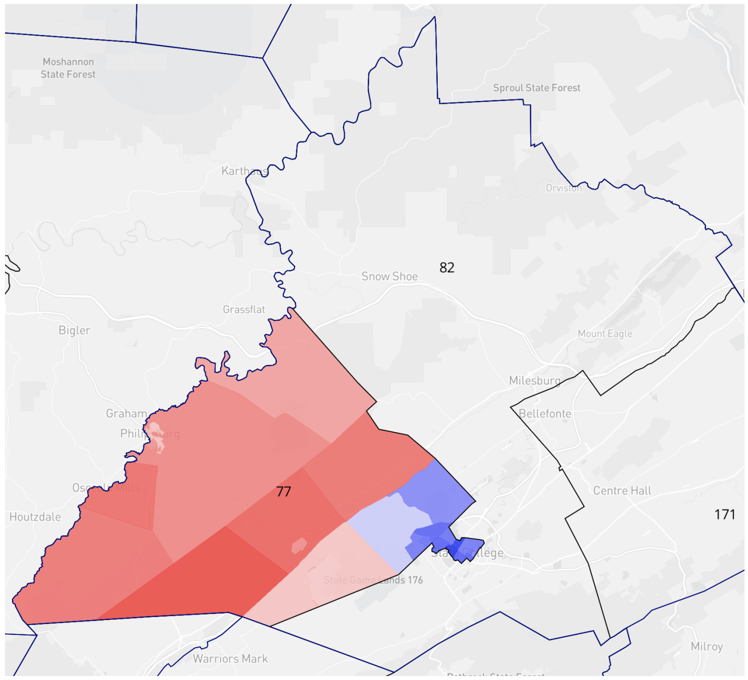


(b) District Boundaries within State College Limits



Note: The top figure shows the district boundaries within Centre County. The bottom figure shows how the city of State College is divided across two districts despite having a population that would allow it to be kept entirely within one district.

District 77 - Partisan Index: 0.59



District 82 - Partisan Index: 0.53

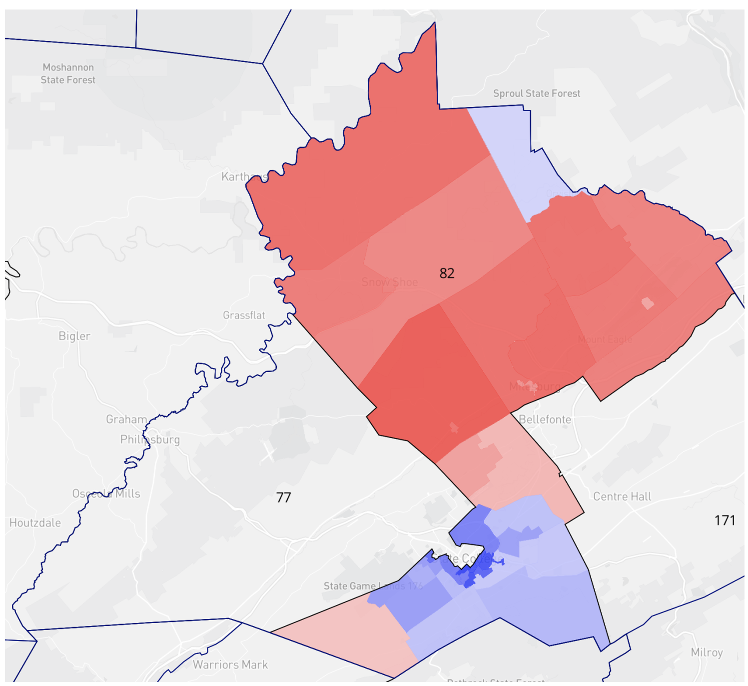


Figure 24: Each panel shows one of the districts that intersect State College. The maps are colored according to the partisan composition of precincts in the district.

6 Comparison to Other District Scoring Programs

To validate the predicted seat shares produced by my analysis, I upload the proposed plan into a commonly used redistricting program - Dave's Redistricting (DRA).[12] This program has been used extensively in redistricting and in redistricting litigation. After uploading the plans, I compare the number of seats the program predicts will lean Democratic to the predictions produced by my analysis. There is perfect agreement when the same elections are used. Table 3 shows the results. In each case I take the proportion of the total two-party vote cast in the elections being included for each district. I then classify each district as a Democratic-leaning district if the Democratic two-party vote share is larger than 0.50.

The DRA uses an index of elections to generate predictions, in a similar way to the indices I described using above. As I noted above, the benefit of an index is that it helps to "wash out" the idiosyncratic features of any particular election, the specific issues in that race, the candidate's qualities (for better or worse), and other factors of the electoral environment. However, the DRA program uses a different combination of elections. The DRA index uses a combination of the 2020 and 2016 presidential elections, the 2018 and 2016 US Senate elections, the 2020 attorney general election, and the 2018 gubernatorial election. When I compute partisan measures that match the DRA index, I get the same results as they do. The DRA index predicts 106 Democratic leaning seats.

Because the choice of elections can have an impact on the predicted seat share for a party, my preferred method is to include all available elections. As discussed above, the main results I present throughout this report use all statewide elections between 2012-2020.[13] I choose 2012 as a starting point because that range incorporates an entire decade, or one decennial census period in which population enumeration and reapportionment take place.

[12]https://davesredistricting.org

[13]I do not include statewide judicial elections in the index. It is uncommon in political science to use judicial elections to measure voters' partisan preferences as research suggests voters treat judicial elections very differently, even when judges run under party labels, than they do partisan elections to legislative and executive positions. Other commonly used measures indices such as Dave's Redistricting and PlanScore.com also omit judicial elections from their partisan indices.

For completeness, I also present the results of the Commission's plan and the distribution of simulations using two alternative indices of statewide elections. First, I recompute an average for all statewide races between 2014-2020 to start after the Holt case in which districts in Pennsylvania were altered as a result of litigation. Finally, I consider an index of statewide elections held in 2020. This measure gives weight to more recent elections and does not include elections from cycles prior to 2020. However, it has the drawback of being heavily influenced by the national political environment of a single election year. Using these indices the Commission's plan contains between 104-107 Democratic leaning districts.

I note that these predictions are independent of the simulations discussed earlier. The predicted seat shares shown below are only a function of different election results and the map put forward by the Legislative Reapportionment Commission. The simulations discussed above provide a comparison of alternative maps that are drawn without consideration of any criteria other than population equality, compactness, and minimizing splits of political subdivisions. They are helpful because they provide a benchmark by which to make an "apples-to-apples" comparison to other districts that are drawn using the same geographic distribution of voters in the state.

Table 3: Comparison of Seat Composition Under Different Elections/Indices

	Commission Plan		% of Simulations Generating Fewer Democratic Seats Than Commission's Map
Election Indices:	Number D Districts	Number R Districts	
DRA index	105	98	
Barber Replication of DRA Index	105	98	
Barber 2012-2020 index	107	96	99.998%
Barber 2014-2020 index	105	98	99.932%
Barber 2020 index	104	99	99.996%

Michael Jay Barber

Appendix A: Curriculum Vitae

Michael Jay Barber

Contact Information

Brigham Young University
Department of Political Science
724 KMBL
Provo, UT 84602

barber@byu.edu
http://michaeljaybarber.com
Ph: (801) 422-7492

Academic Appointments

Brigham Young University, Provo, UT

August 2020 - present Associate Professor, Department of Political Science
2014 - July 2020 Assistant Professor, Department of Political Science
2014 - present Faculty Scholar, Center for the Study of Elections and Democracy

Education

Princeton University Department of Politics, Princeton, NJ

Ph.D., Politics, July 2014

- Advisors: Brandice Canes-Wrone, Nolan McCarty, and Kosuke Imai
- Dissertation: "Buying Representation: the Incentives, Ideology, and Influence of Campaign Contributions on American Politics"
- 2015 Carl Albert Award for Best Dissertation, Legislative Studies Section, American Political Science Association (APSA)

M.A., Politics, December 2011

Brigham Young University, Provo, UT

B.A., International Relations - Political Economy Focus, April, 2008

- *Cum Laude*

Research Interests

American politics, congressional polarization, political ideology, campaign finance, survey research

Publications

19. **"Ideological Disagreement and Pre-emption in Municipal Policymaking"** with Adam Dynes
Forthcoming at *American Journal of Political Science*
18. **"Comparing Campaign Finance and Vote Based Measures of Ideology"**
Forthcoming at *Journal of Politics*
17. **"The Participatory and Partisan Impacts of Mandatory Vote-by-Mail"**, with John Holbein
Science Advances, 2020. Vol. 6, no. 35, DOI: 10.1126/sciadv.abc7685
16. **"Issue Politicization and Interest Group Campaign Contribution Strategies"**, with Mandi Eatough
Journal of Politics, 2020. Vol. 82: No. 3, pp. 1008-1025

15. **“Campaign Contributions and Donors’ Policy Agreement with Presidential Candidates”**, with Brandice Canes-Wrone and Sharece Thrower
Presidential Studies Quarterly, 2019, 49 (4) 770–797
14. **“Conservatism in the Era of Trump”**, with Jeremy Pope
Perspectives on Politics, 2019, 17 (3) 719–736
13. **“Legislative Constraints on Executive Unilateralism in Separation of Powers Systems”**, with Alex Bolton and Sharece Thrower
Legislative Studies Quarterly, 2019, 44 (3) 515–548
Awarded the Jewell-Loewenberg Award for best article in the area of subnational politics published in *Legislative Studies Quarterly* in 2019
12. **“Electoral Competitiveness and Legislative Productivity”**, with Soren Schmidt
American Politics Research, 2019, 47 (4) 683–708
11. **“Does Party Trump Ideology? Disentangling Party and Ideology in America”**, with Jeremy Pope
American Political Science Review, 2019, 113 (1) 38–54
10. **“The Evolution of National Constitutions”**, with Scott Abramson
Quarterly Journal of Political Science, 2019, 14 (1) 89–114
9. **“Who is Ideological? Measuring Ideological Responses to Policy Questions in the American Public”**, with Jeremy Pope
The Forum: A Journal of Applied Research in Contemporary Politics, 2018, 16 (1) 97–122
8. **“Status Quo Bias in Ballot Wording”**, with David Gordon, Ryan Hill, and Joe Price
The Journal of Experimental Political Science, 2017, 4 (2) 151–160.
7. **“Ideologically Sophisticated Donors: Which Candidates Do Individual Contributors Finance?”**, with Brandice Canes-Wrone and Sharece Thrower
American Journal of Political Science, 2017, 61 (2) 271–288.
6. **“Gender Inequalities in Campaign Finance: A Regression Discontinuity Design”**, with Daniel Butler and Jessica Preece
Quarterly Journal of Political Science, 2016, Vol. 11, No. 2: 219–248.
5. **“Representing the Preferences of Donors, Partisans, and Voters in the U.S. Senate”**
Public Opinion Quarterly, 2016, 80: 225–249.
4. **“Donation Motivations: Testing Theories of Access and Ideology”**
Political Research Quarterly, 2016, 69 (1) 148–160.
3. **“Ideological Donors, Contribution Limits, and the Polarization of State Legislatures”**
Journal of Politics, 2016, 78 (1) 296–310.
2. **“Online Polls and Registration Based Sampling: A New Method for Pre-Election Polling”** with Quin Monson, Kelly Patterson and Chris Mann.
Political Analysis 2014, 22 (3) 321–335.
1. **“Causes and Consequences of Political Polarization”** In *Negotiating Agreement in Politics*. Jane Mansbridge and Cathie Jo Martin, eds., Washington, DC: American Political Science Association: 19–53. with Nolan McCarty. 2013.

    - Reprinted in *Solutions to Political Polarization in America*, Cambridge University Press. Nate Persily, eds. 2015
    - Reprinted in *Political Negotiation: A Handbook*, Brookings Institution Press. Jane Mansbridge and Cathie Jo Martin, eds. 2015

Available Working Papers

"Misclassification and Bias in Predictions of Individual Ethnicity from Administrative Records" (Revise and Resubmit at *American Political Science Review*)

"Taking Cues When You Don't Care: Issue Importance and Partisan Cue Taking"
with Jeremy Pope (Revise and Resubmit)

"A Revolution of Rights in American Founding Documents"
with Scott Abramson and Jeremy Pope (Conditionally Accepted)

"410 Million Voting Records Show the Distribution of Turnout in America Today"
with John Holbein (Revise and Resubmit)

"Partisanship and Trolleyology"
with Ryan Davis (Under Review)

"Who's the Partisan: Are Issues or Groups More Important to Partisanship?"
with Jeremy Pope (Revise and Resubmit)

"Race and Realignment in American Politics"
with Jeremy Pope (Revise and Resubmit)

"The Policy Preferences of Donors and Voters"

"Estimating Neighborhood Effects on Turnout from Geocoded Voter Registration Records."
with Kosuke Imai

"Super PAC Contributions in Congressional Elections"

Works in Progress

"Collaborative Study of Democracy and Politics"
with Brandice Canes-Wrone, Gregory Huber, and Joshua Clinton

"Preferences for Representational Styles in the American Public"
with Ryan Davis and Adam Dynes

"Representation and Issue Congruence in Congress"
with Taylor Petersen

"Education, Income, and the Vote for Trump"
with Edie Ellison

Invited Presentations

"Are Mormons Breaking Up with Republicanism? The Unique Political Behavior of Mormons in the 2016 Presidential Election"

- Ivy League LDS Student Association Conference - Princeton University, November 2018, Princeton, NJ

"Issue Politicization and Access-Oriented Giving: A Theory of PAC Contribution Behavior"

- Vanderbilt University, May 2017, Nashville, TN

"Lost in Issue Space? Measuring Levels of Ideology in the American Public"

- Yale University, April 2016, New Haven, CT

"The Incentives, Ideology, and Influence of Campaign Donors in American Politics"

- University of Oklahoma, April 2016, Norman, OK

"Lost in Issue Space? Measuring Levels of Ideology in the American Public"

- University of Wisconsin - Madison, February 2016, Madison, WI

"Polarization and Campaign Contributors: Motivations, Ideology, and Policy"

- Hewlett Foundation Conference on Lobbying and Campaign Finance, October 2014, Palo Alto, CA

"Ideological Donors, Contribution Limits, and the Polarization of State Legislatures"

- Bipartisan Policy Center Meeting on Party Polarization and Campaign Finance, September 2014, Washington, DC

"Representing the Preferences of Donors, Partisans, and Voters in the U.S. Senate"

- Yale Center for the Study of American Politics Conference, May 2014, New Haven, CT

Conference Presentations

Washington D.C. Political Economy Conference (PECO):

- 2017 discussant

American Political Science Association (APSA) Annual Meeting:

- 2014 participant and discussant, 2015 participant, 2016 participant, 2017 participant, 2018 participant

Midwest Political Science Association (MPSA) Annual Meeting:

- 2015 participant and discussant, 2016 participant and discussant, 2018 participant

Southern Political Science Association (SPSA) Annual Meeting:

- 2015 participant and discussant, 2016 participant and discussant, 2017 participant

Teaching Experience

Poli 315: Congress and the Legislative Process

- Fall 2014, Winter 2015, Fall 2015, Winter 2016, Summer 2017

Poli 328: Quantitative Analysis

- Winter 2017, Fall 2017, Fall 2019, Winter 2020, Fall 2020, Winter 2021

Poli 410: Undergraduate Research Seminar in American Politics

- Fall 2014, Winter 2015, Fall 2015, Winter 2016, Summer 2017

Awards and Grants

2019 BYU Mentored Environment Grant (MEG), American Ideology Project, $30,000

2017 BYU Political Science Teacher of the Year Award

2017 BYU Mentored Environment Grant (MEG), Funding American Democracy Project, $20,000

2016 BYU Political Science Department, Political Ideology and President Trump (with Jeremy Pope), $7,500

2016 BYU Office of Research and Creative Activities (ORCA) Student Mentored Grant x 3

- Hayden Galloway, Jennica Peterson, Rebecca Shuel

2015 BYU Office of Research and Creative Activities (ORCA) Student Mentored Grant x 3

- Michael-Sean Covey, Hayden Galloway, Sean Stephenson

2015 BYU Student Experiential Learning Grant, American Founding Comparative Constitutions Project (with Jeremy Pope), $9,000

2015 BYU Social Science College Research Grant, $5,000

2014 BYU Political Science Department, 2014 Washington DC Mayoral Pre-Election Poll (with Quin Monson and Kelly Patterson), $3,000

2014 BYU Social Science College Award, 2014 Washington DC Mayoral Pre-Election Poll (with Quin Monson and Kelly Patterson), $3,000

2014 BYU Center for the Study of Elections and Democracy, 2014 Washington DC Mayoral Pre-Election Poll (with Quin Monson and Kelly Patterson), $2,000

2012 Princeton Center for the Study of Democratic Politics Dissertation Improvement Grant, $5,000

2011 Princeton Mamdouha S. Bobst Center for Peace and Justice Dissertation Research Grant, $5,000

2011 Princeton Political Economy Research Grant, $1,500

Other Scholarly Activities

Expert Witness in Nancy Carola Jacobson, et al., Plaintiffs, vs. Laurel M. Lee, et al., Defendants. Case No. 4:18-cv-00262 MW-CAS (U.S. District Court for the Northern District of Florida)

Expert Witness in Common Cause, et al., Plaintiffs, vs. LEWIS, et al., Defendants. Case No. 18-CVS-14001 (Wake County, North Carolina)

Expert Witness in Kelvin Jones, et al., Plaintiffs, v. Ron DeSantis, et al., Defendants, Consolidated Case No. 4:19-cv-300 (U.S. District Court for the Northern District of Florida)

Expert Witness in Community Success Initiative, et al., Plaintiffs, v. Timothy K. Moore, et al., Defendants, Case No. 19-cv-15941 (Wake County, North Carolina)

Expert Witness in Richard Rose et al., Plaintiffs, v. Brad Raffensperger, Defendant, Civil Action No. 1:20-cv-02921-SDG (U.S. District Court for the Northern District of Georgia)

Georgia Coalition for the People's Agenda, Inc., et. al., Plaintiffs, v. Brad Raffensberger, Defendant. Civil Action No. 1:18-cv-04727-ELR (U.S. District Court for the Northern District of Georgia)

Expert Witness in Alabama, et al., Plaintiffs, v. United States Department of Commerce; Gina Raimondo, et al., Defendants. Case No. CASE No. 3:21-cv-00211-RAH-ECM-KCN (U.S. District Court for the Middle District of Alabama Eastern Division)

Expert Witness in League of Women Voters of Ohio, et al., Relators, v. Ohio Redistricting Commission, et al., Respondents. Case No. 2021-1193 (Supreme Court of Ohio)

Expert Witness in Regina Adams, et al., Relators, v. Governor Mike DeWine, et al., Respondents. Case No. 2021-1428 (Supreme Court of Ohio)

Expert Witness in Rebecca Harper, et al., Plaintiffs, v. Representative Destin Hall, et al., Defendants (Consolidated Case). Case No. 21 CVS 500085 (Wake County, North Carolina)

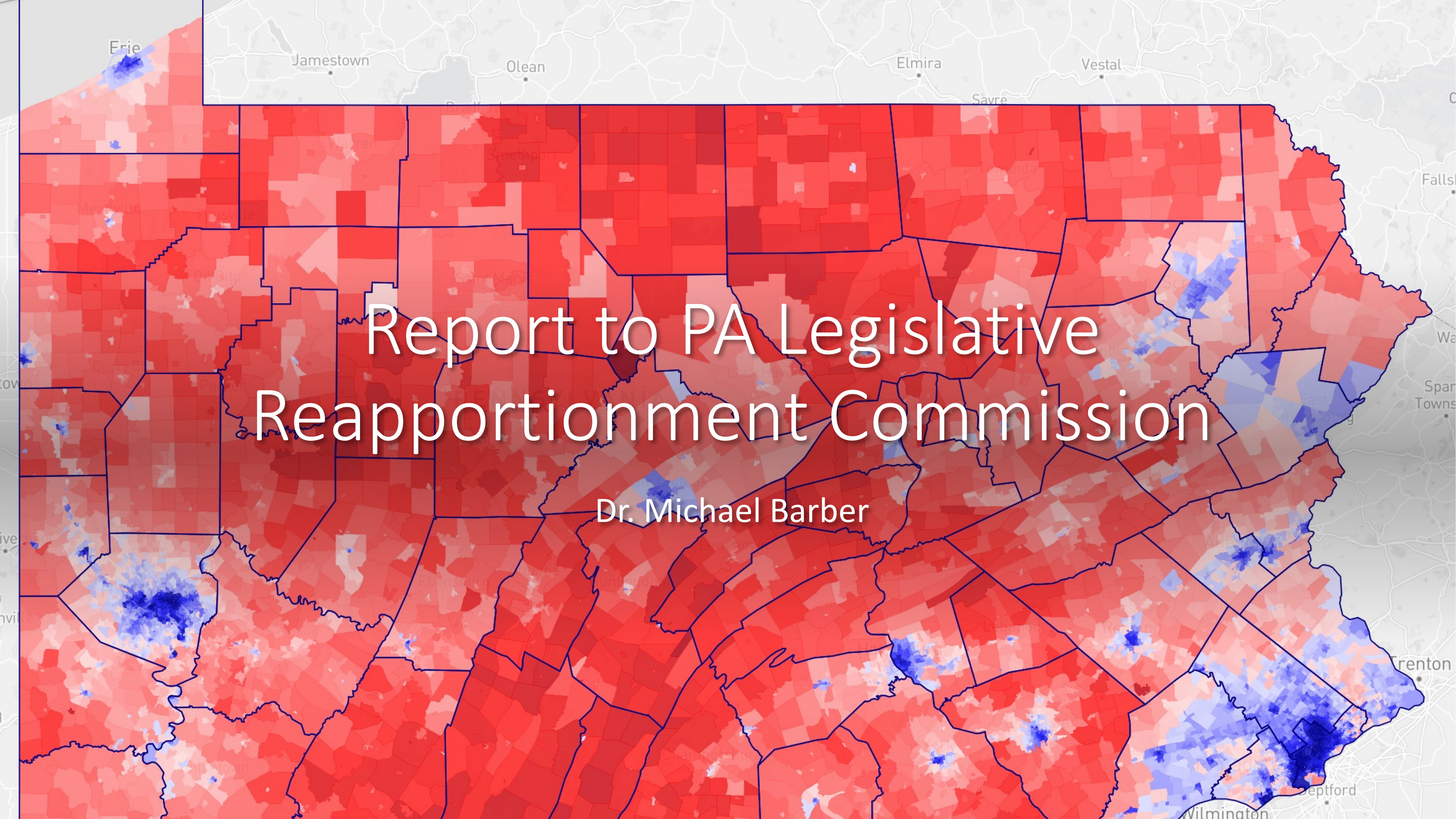
Additional Training EITM 2012 at Princeton University - Participant and Graduate Student Coordinator

Computer Skills Statistical Programs: R, Stata, SPSS, parallel computing

Updated January 7, 2022

Report to PA Legislative Reapportionment Commission

Dr. Michael Barber



Background

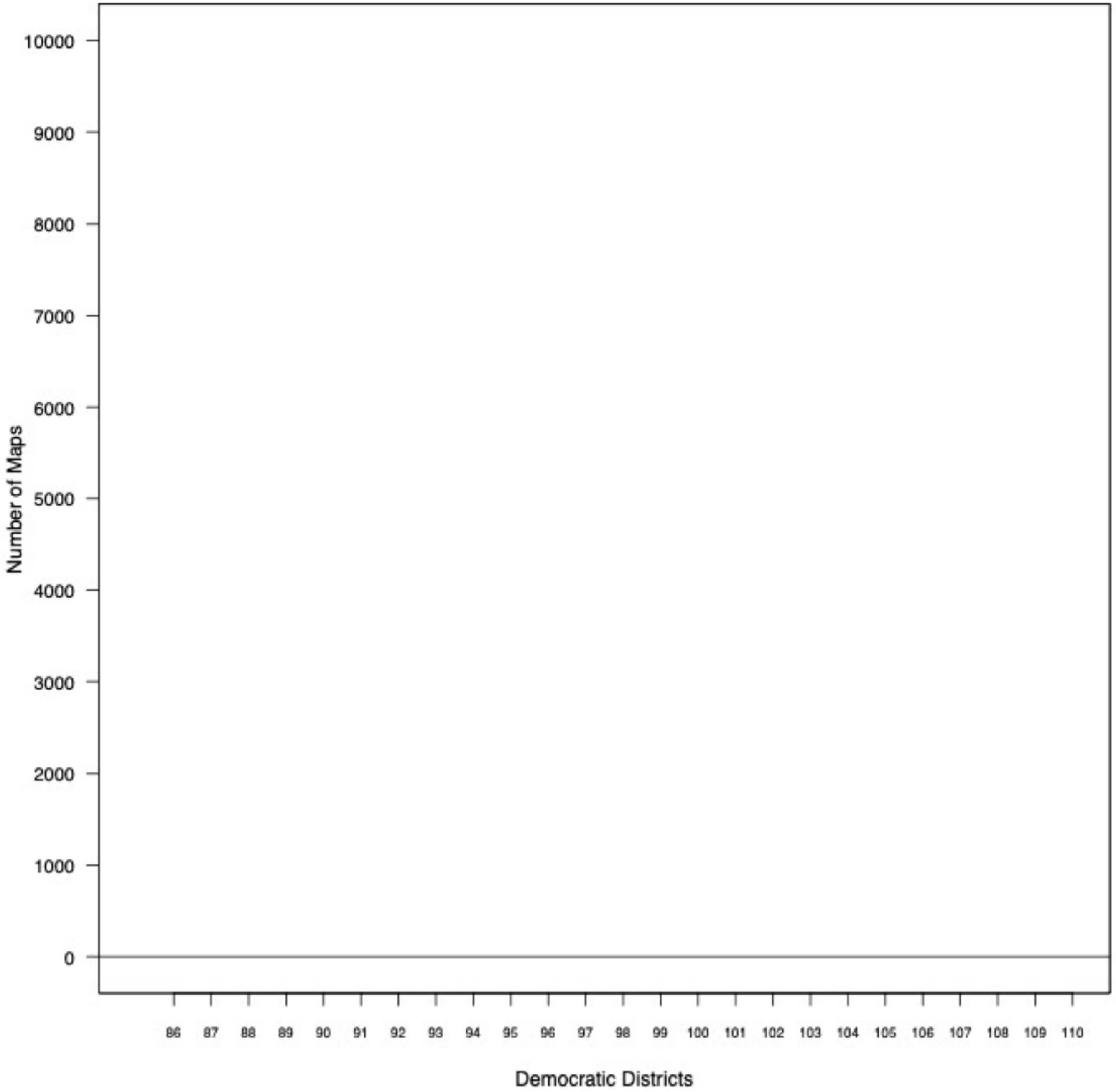
- Associate Professor of Political Science at Brigham Young University
- PhD in American Politics from Princeton University
- Published 20+ articles in American politics
- Expert witness in 10 election-related cases
- Redistricting cases in North Carolina and Ohio

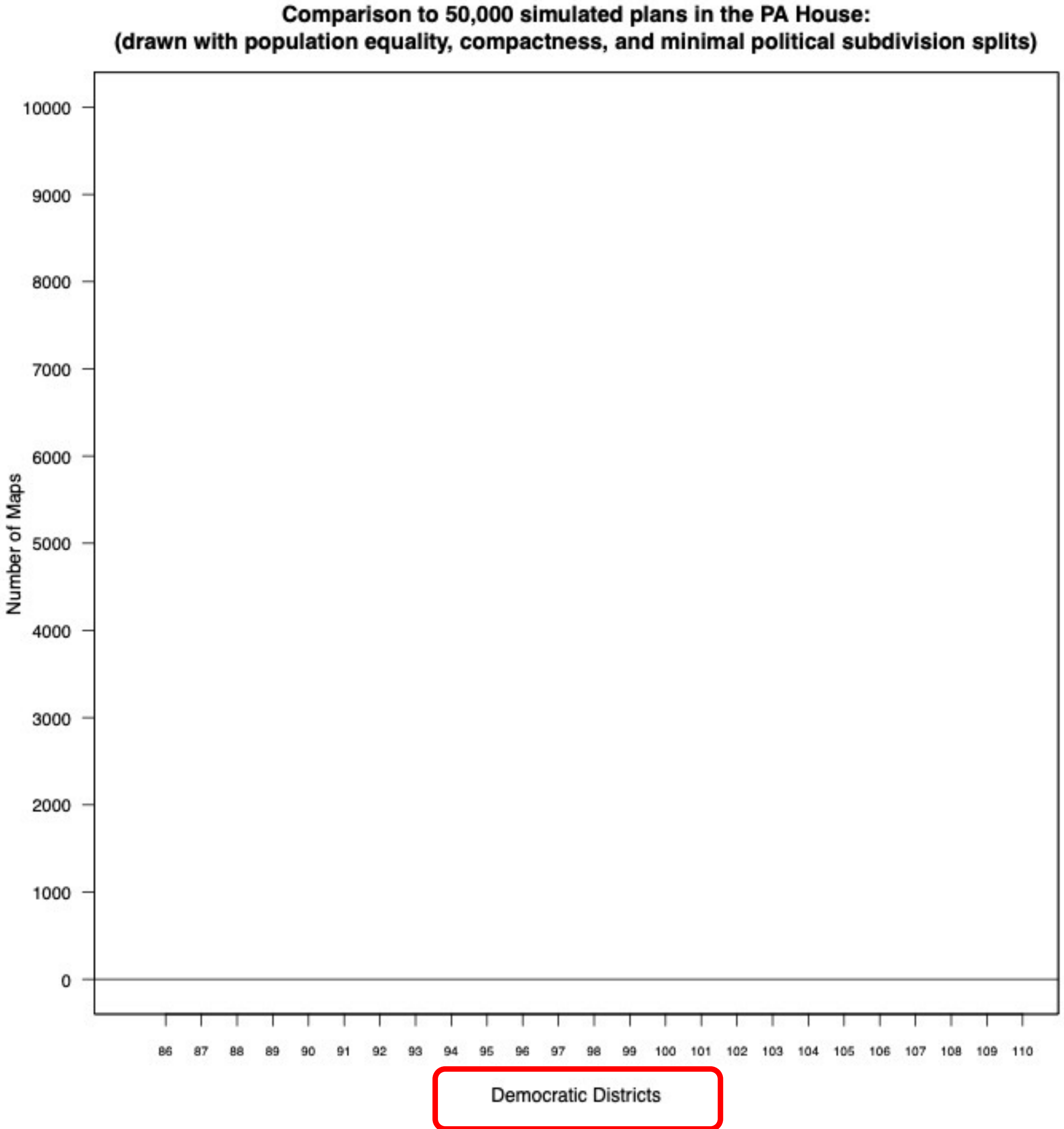
Simulation Analysis

- Inputs
  - 203 districts
  - Criteria in Article II, Section 16 of Pennsylvania Constitution
  - Compactness
  - Contiguity
  - “Unless absolutely necessary” no county, municipal splits

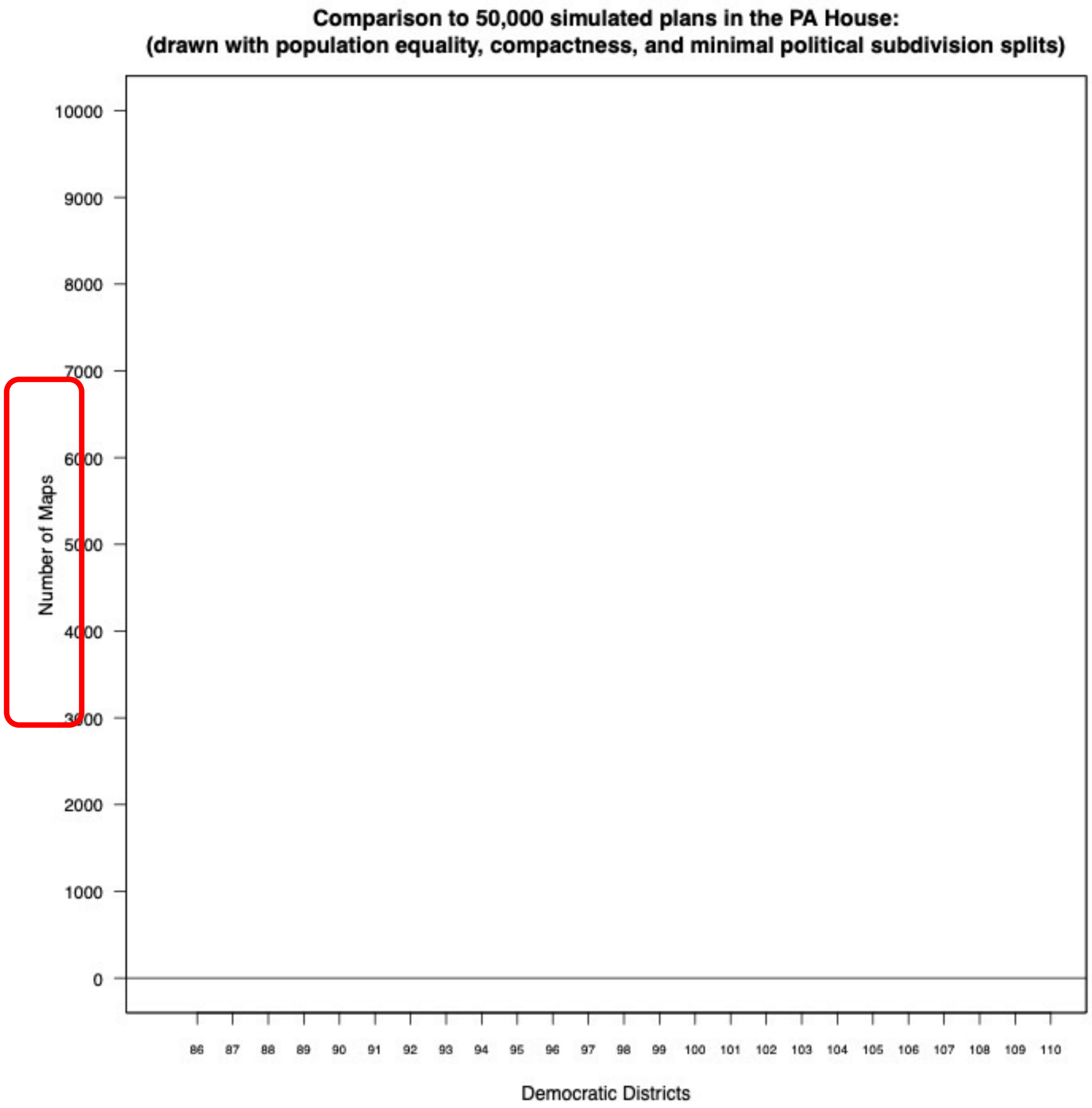
- Outputs
  - 50,000 maps, each with 203 districts
  - Calculate partisan lean of districts using 2012-2020 statewide elections
  - Compare to LRC’s proposed plan

Comparison to 50,000 simulated plans in the PA House:
(drawn with population equality, compactness, and minimal political subdivision splits)

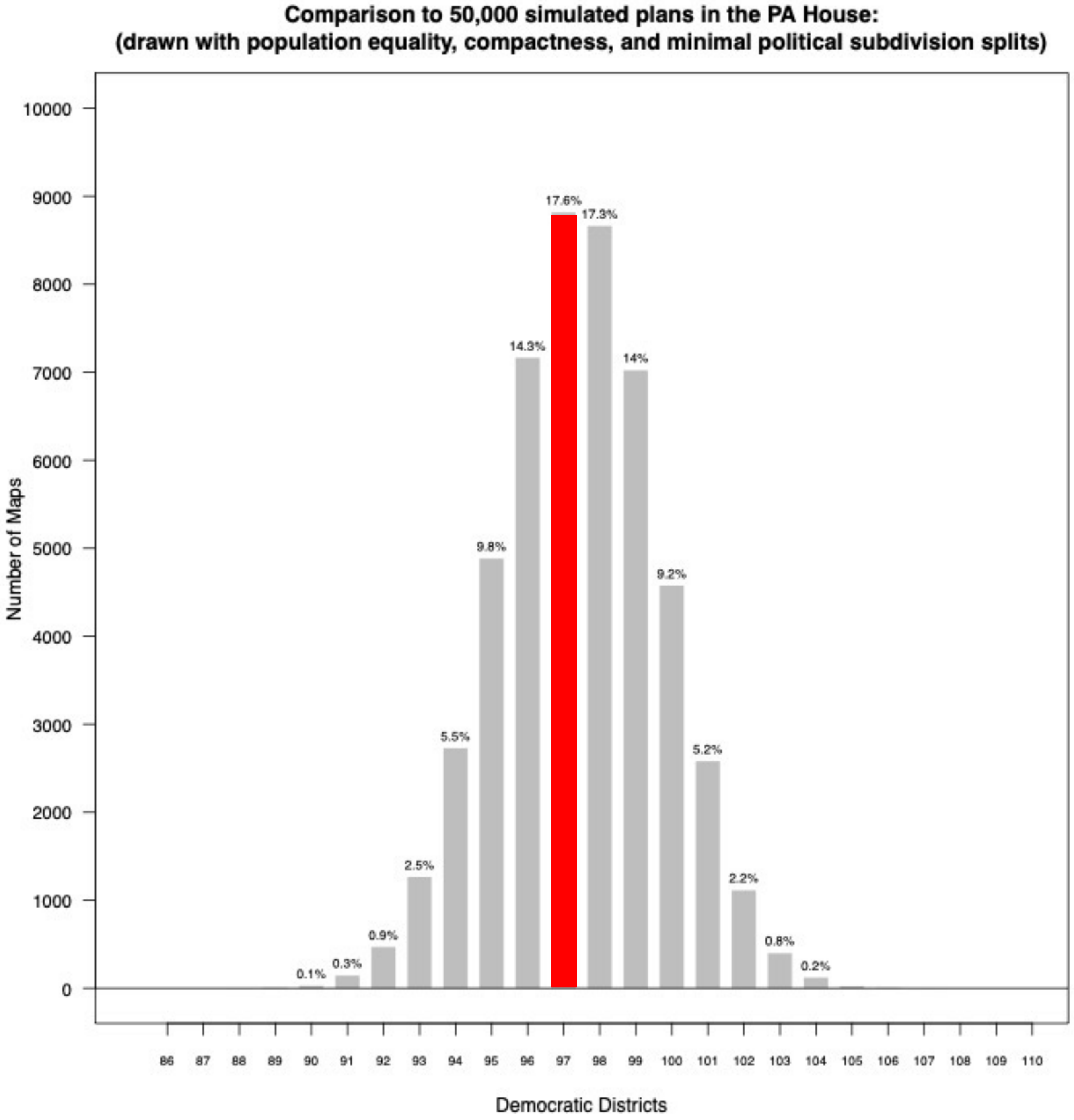




Horizontal Axis: Number of Democratic-Leaning Districts



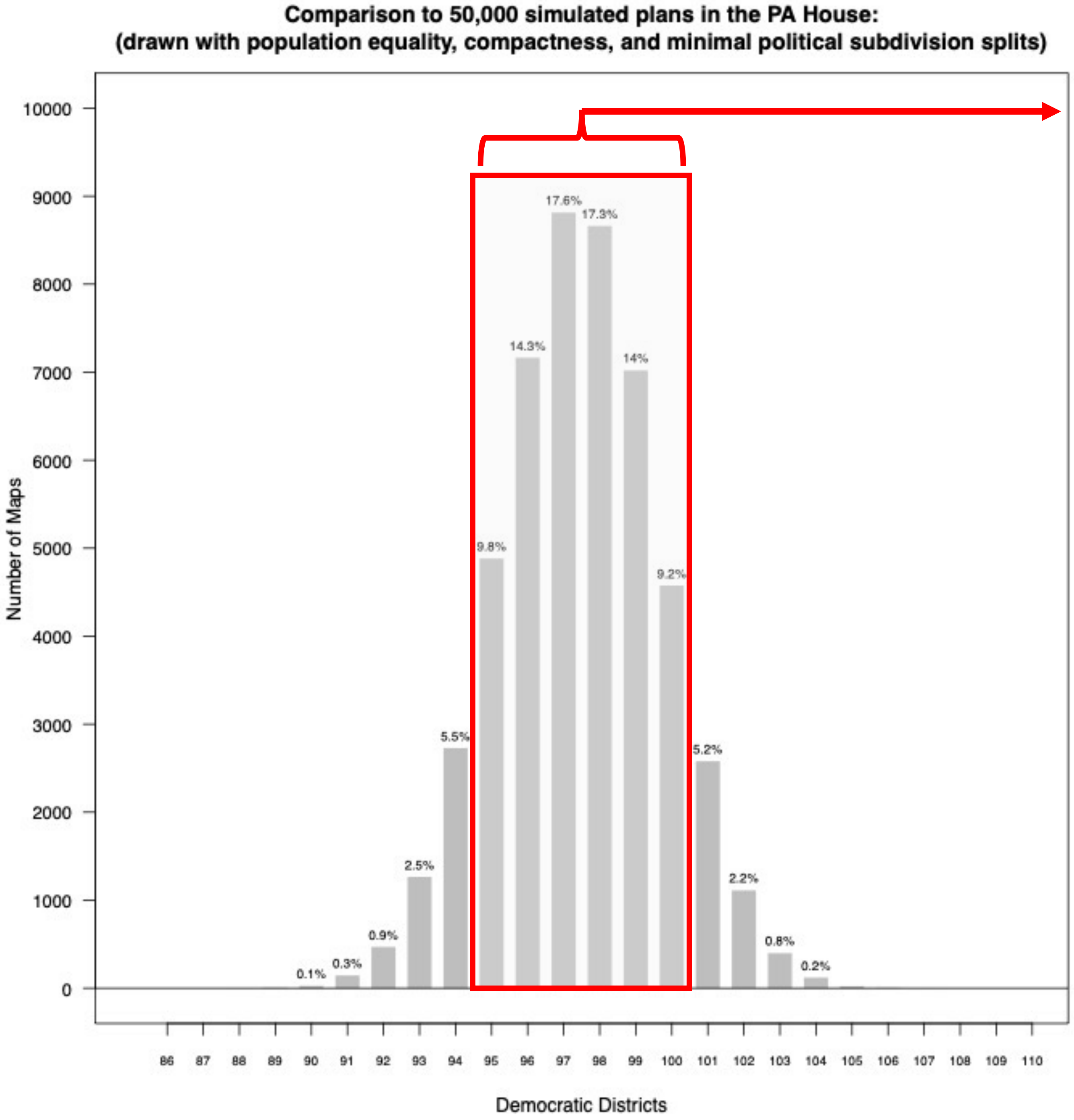
Vertical Axis:
Number of Maps Producing a particular Outcome



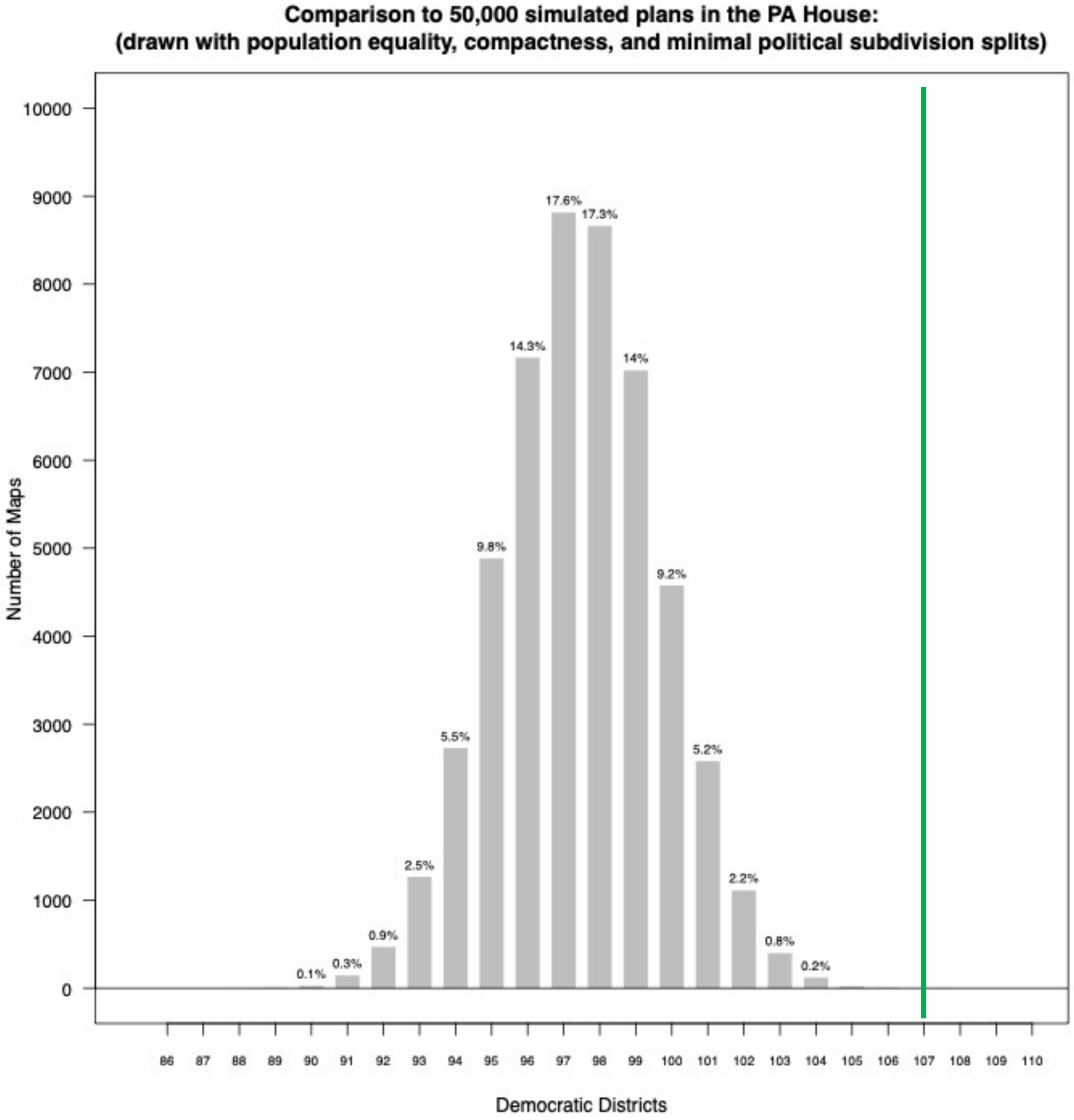
Most Common Outcome:

97 D-Leaning Districts

17.6% of all simulations produce this result.



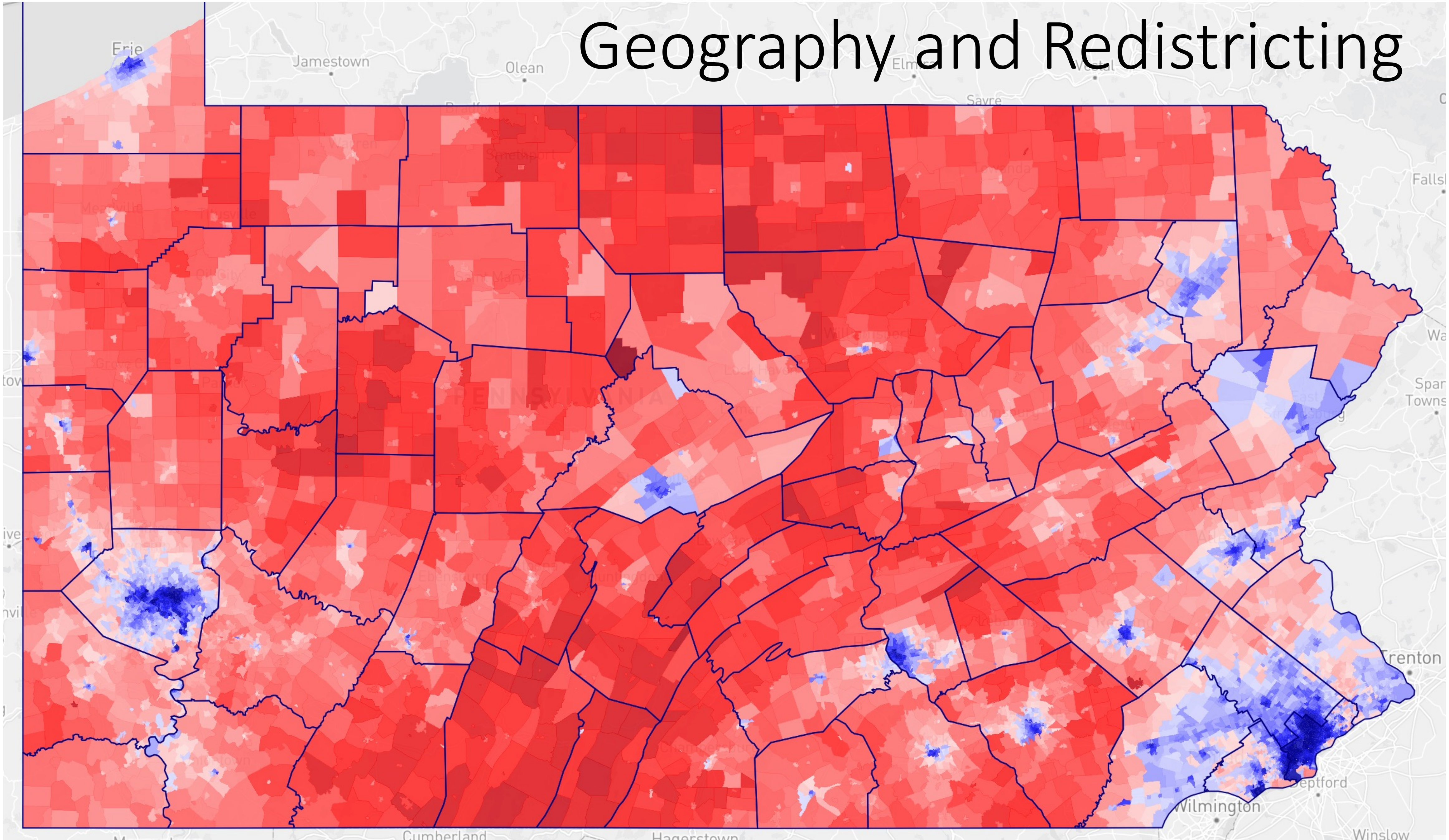
80% of simulations produce 95-100 D-leaning districts.



Commission Proposal:
107 D-leaning Districts

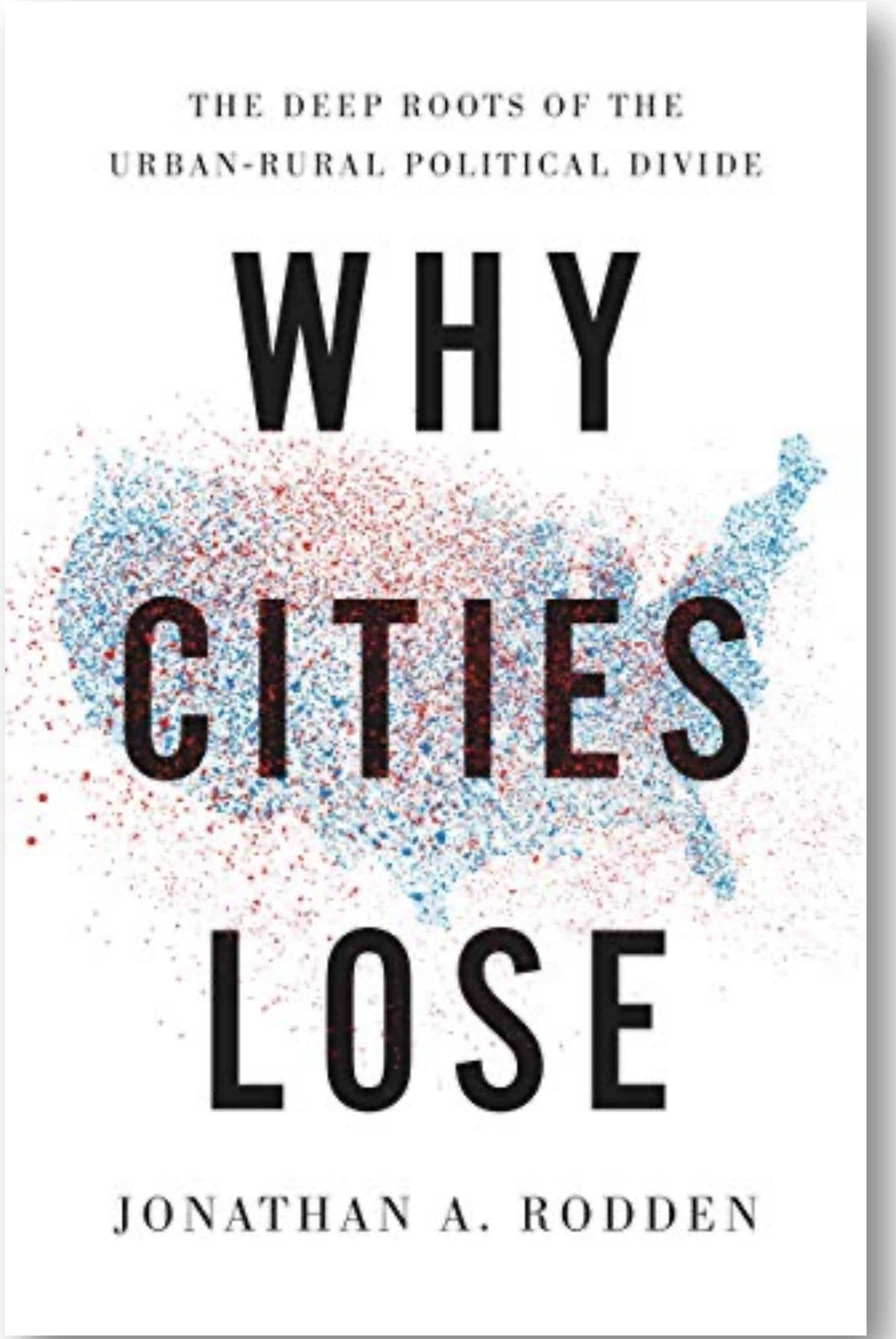
More D-leaning districts than **99.998% of all simulations**

Geography and Redistricting

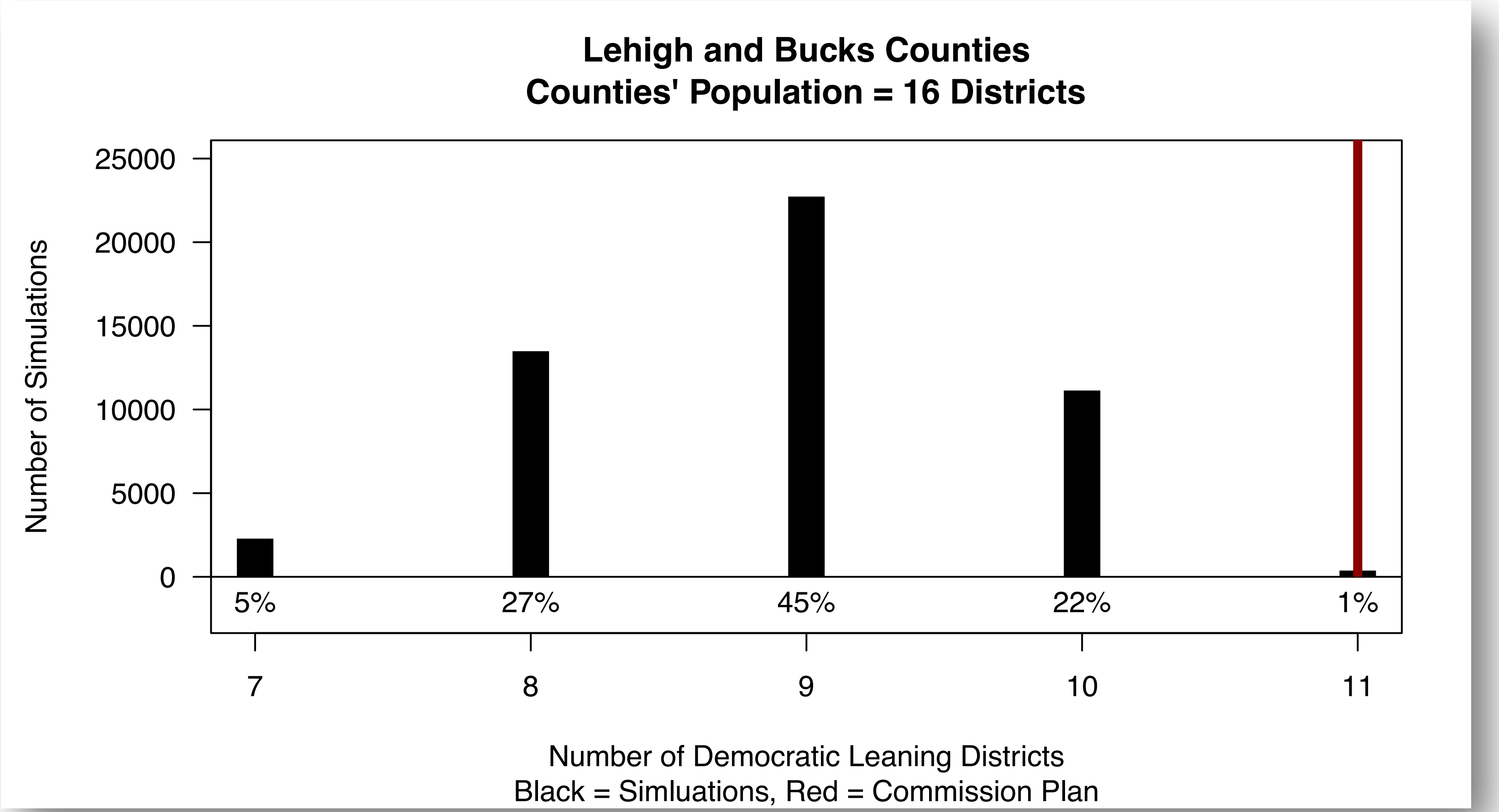


Geography and Redistricting

- “Democrats would need a redistricting process that **intentionally carved up large cities** like pizza slices or spokes of a wheel, so as to combine some very Democratic urban neighborhoods with some Republican exurbs in an effort to spread Democrats more efficiently across districts.”
    - Jonathan Rodden, Why Cities Lose, pg. 155

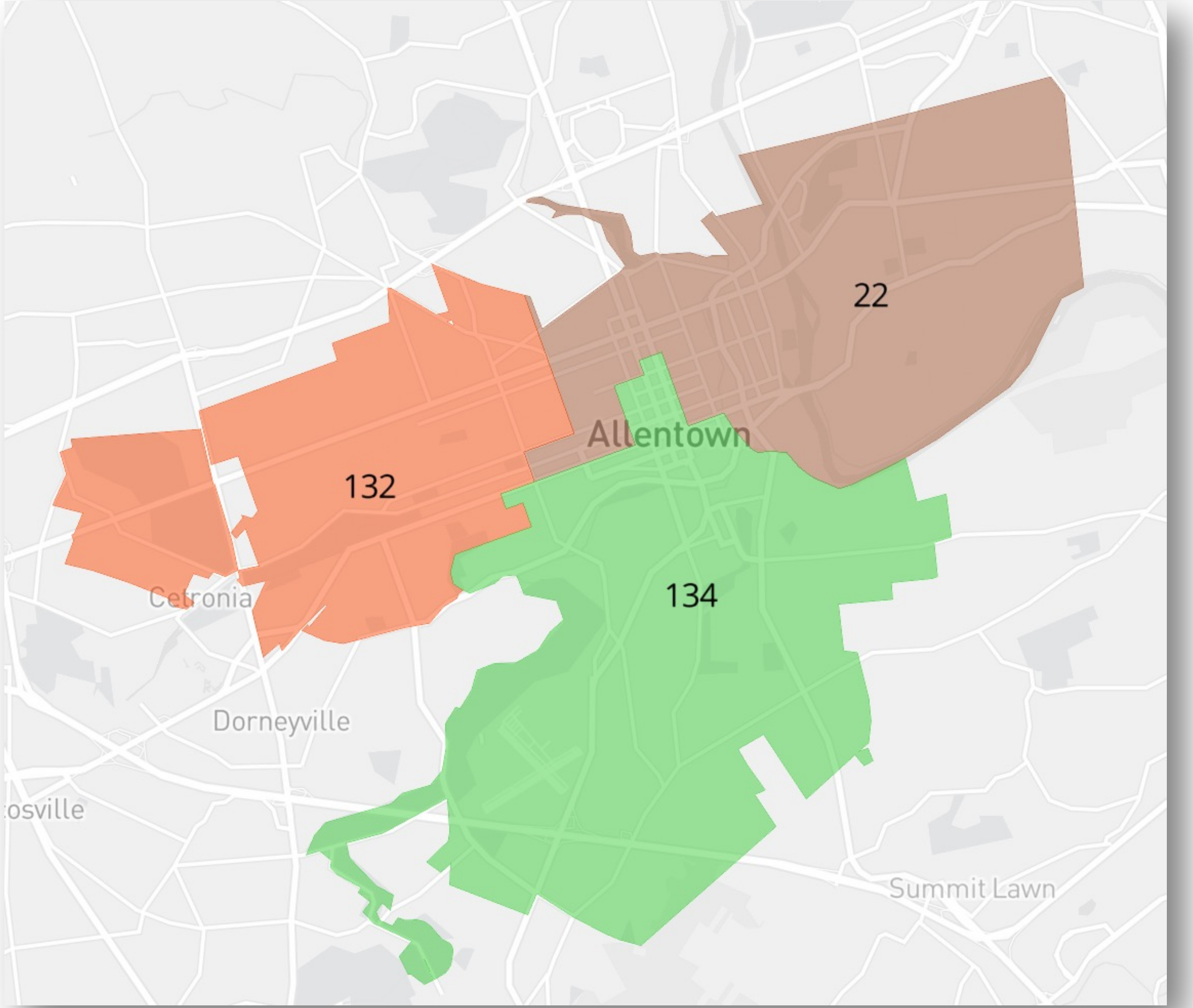


An Example: Allentown



Commission Proposal:
11 D-leaning Districts

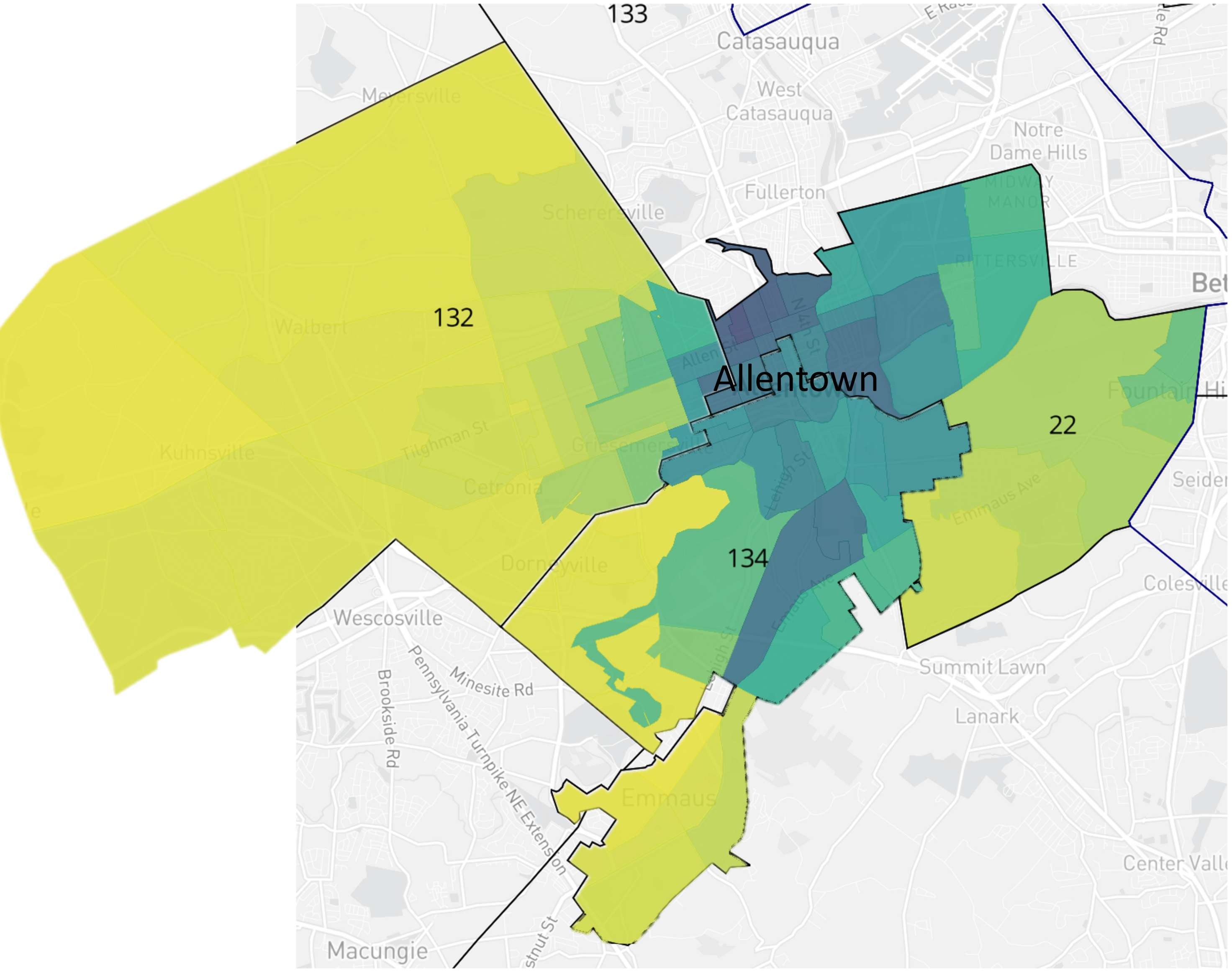
More D-leaning districts than **99% of all simulations**

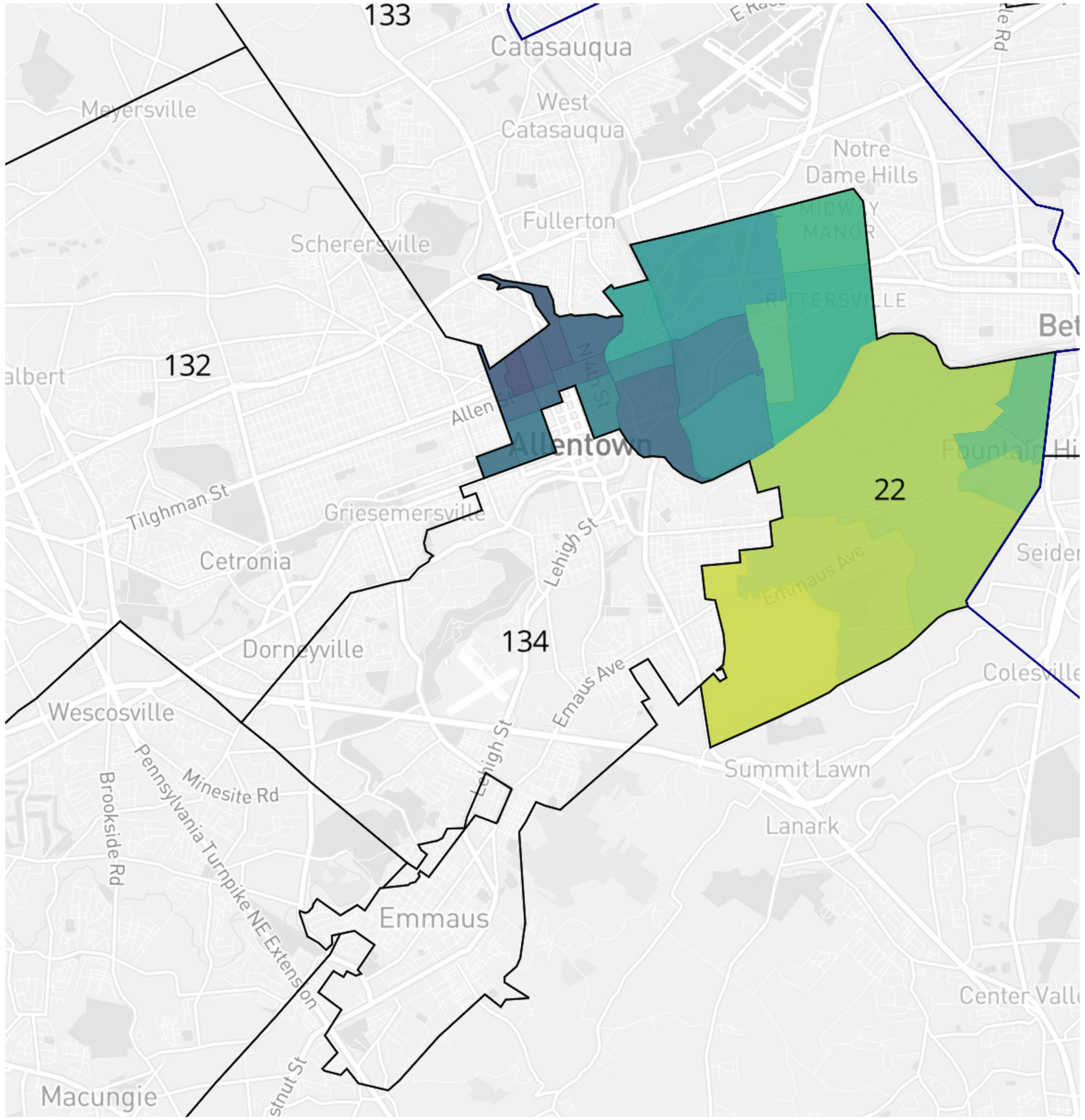


Allentown is divided into 3 districts

- District 22
- District 134
- District 132

Allentown:
49% Hispanic VAP



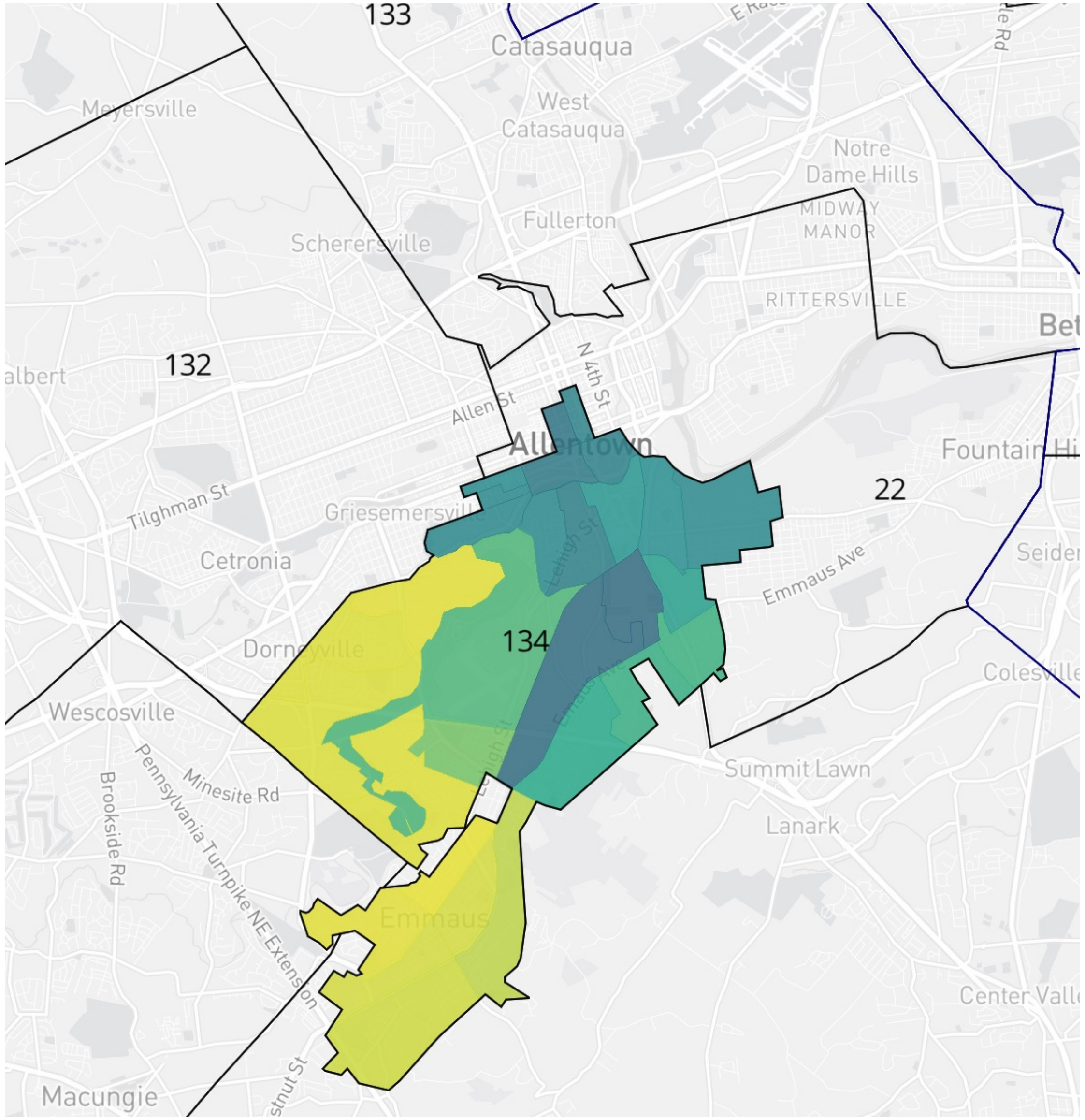


Allentown: 49% Hispanic VAP

- District 22:

  50.8% Hispanic VAP

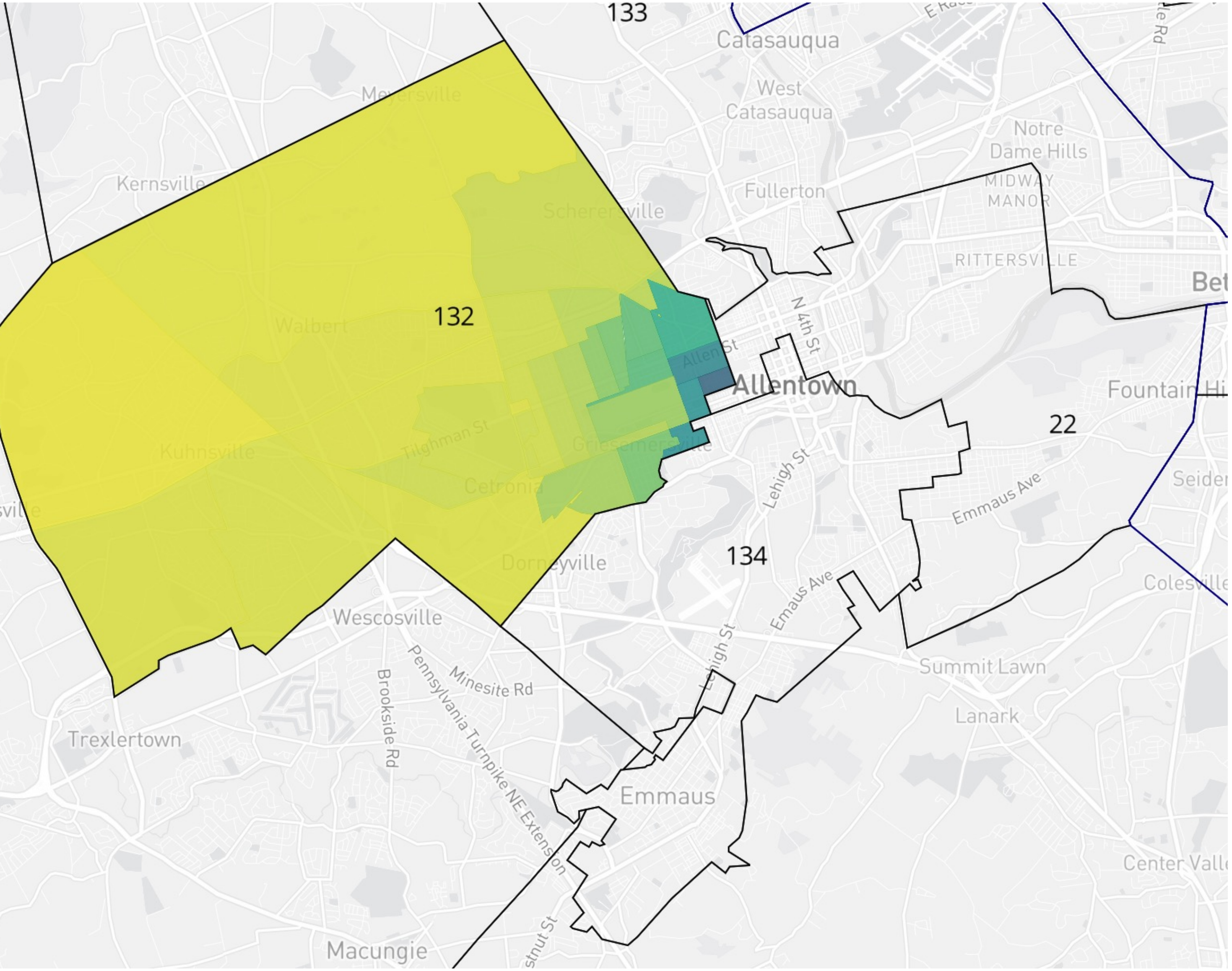
Allentown: 49% Hispanic VAP



- District 134:

  38.5% Hispanic VAP

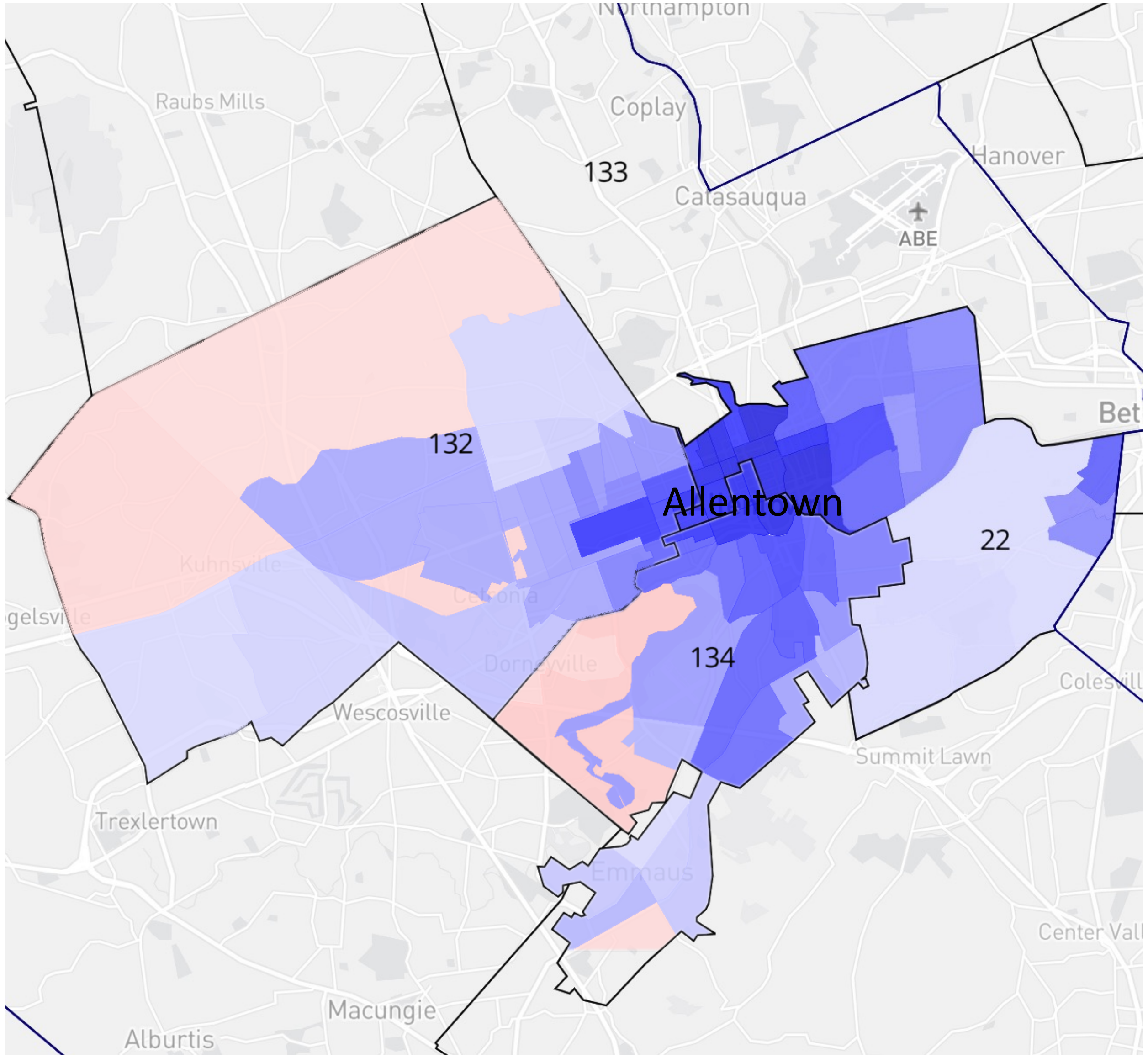
Allentown: 49% Hispanic VAP



- District 132:

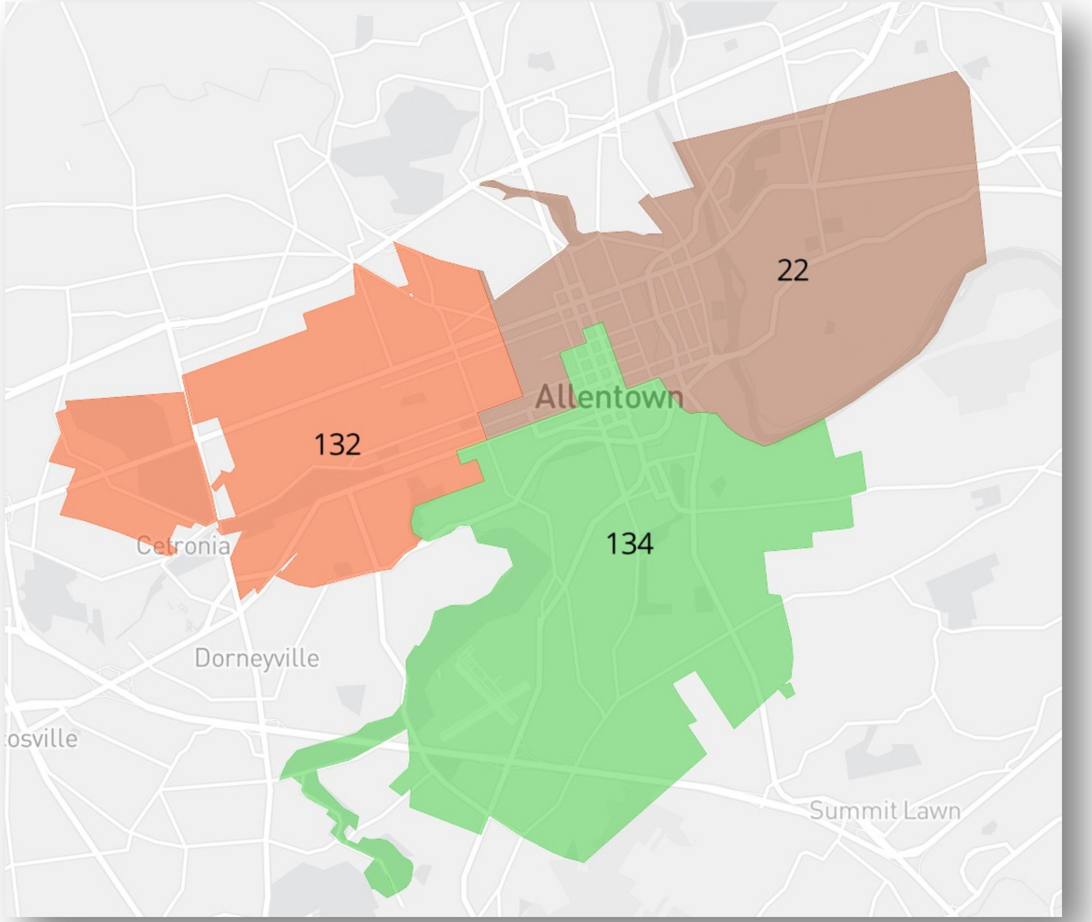
  18.1% Hispanic VAP

Allentown is divided into 3 districts



- District 22: 72% Democratic
- District 134: 63% Democratic
- District 132: 57% Democratic

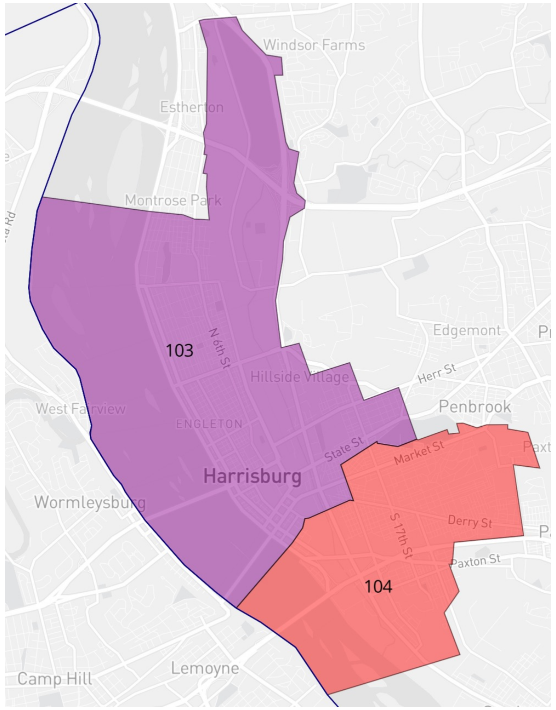
Allentown is divided into 3 districts

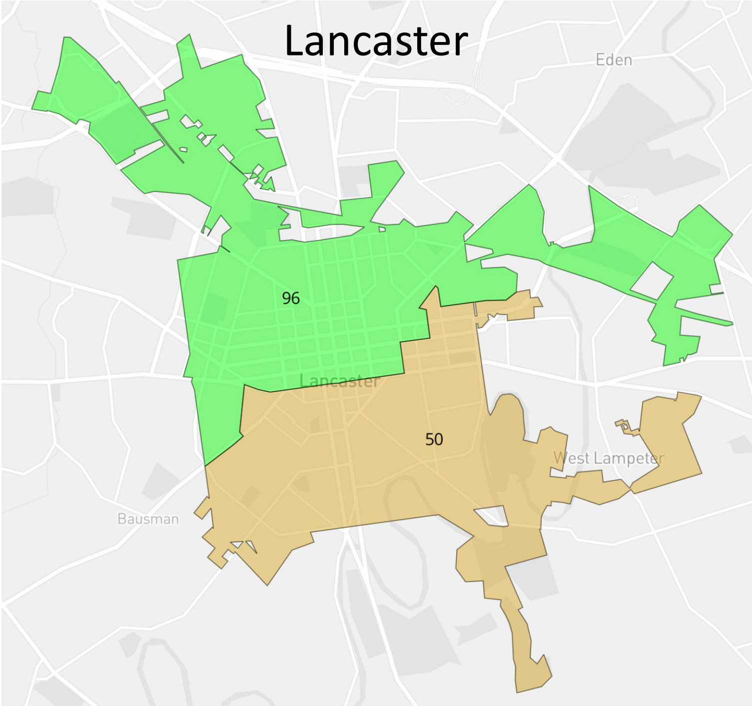


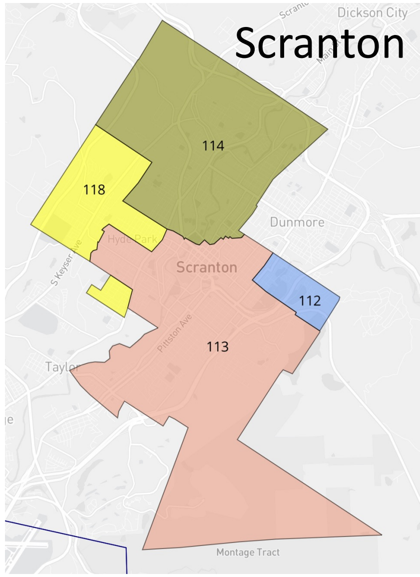
- Dilutes Hispanic population voting power
- Violates PA Constitutional criteria
  - "Unless absolutely necessary" no city shall be divided.

Commission Plan divides Democratic cities more than necessary

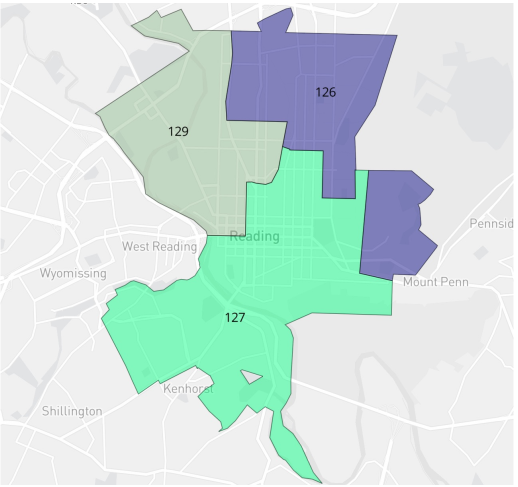
Harrisburg



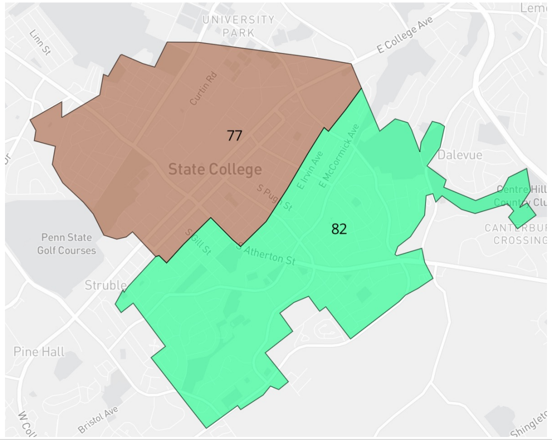


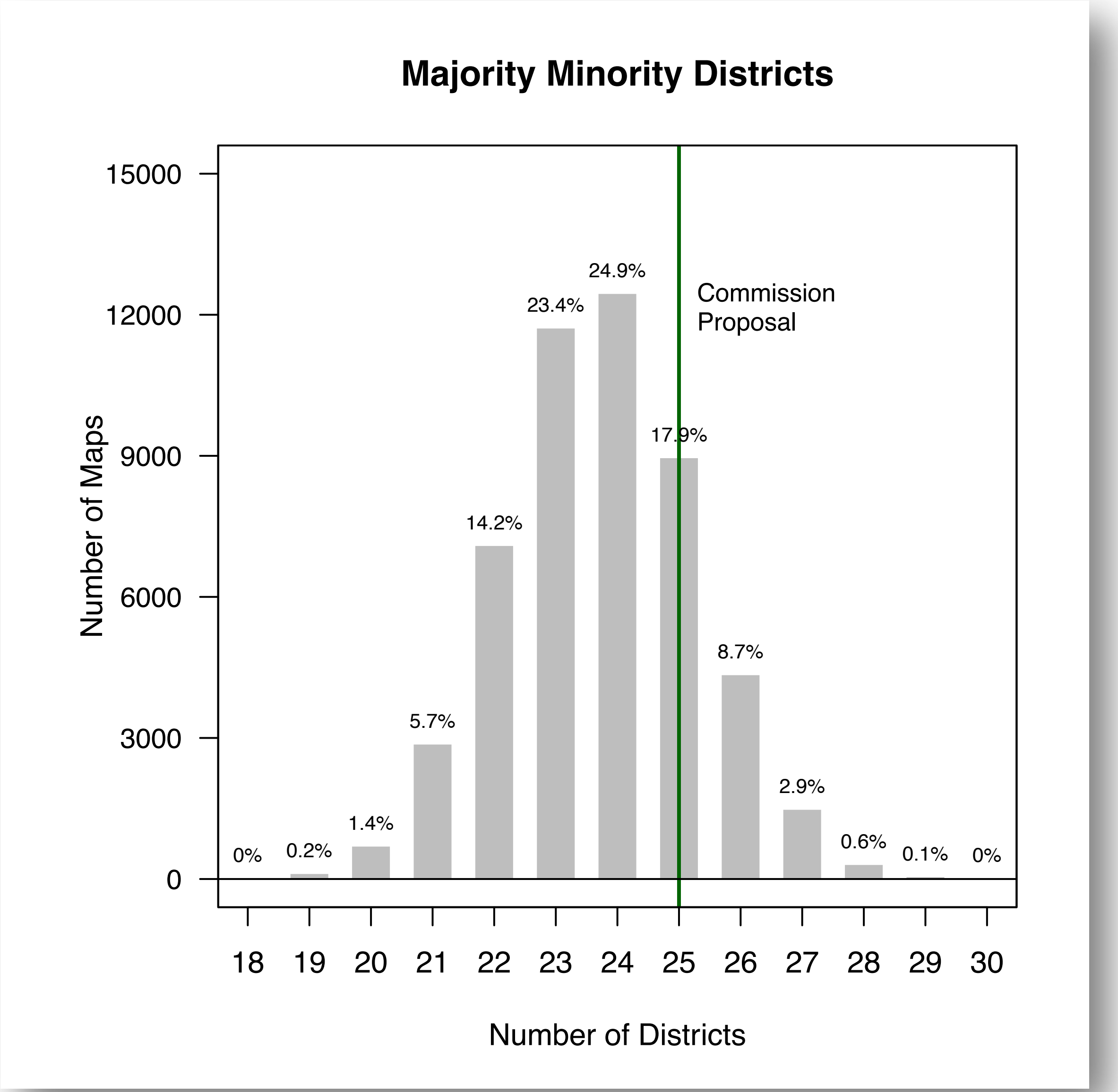


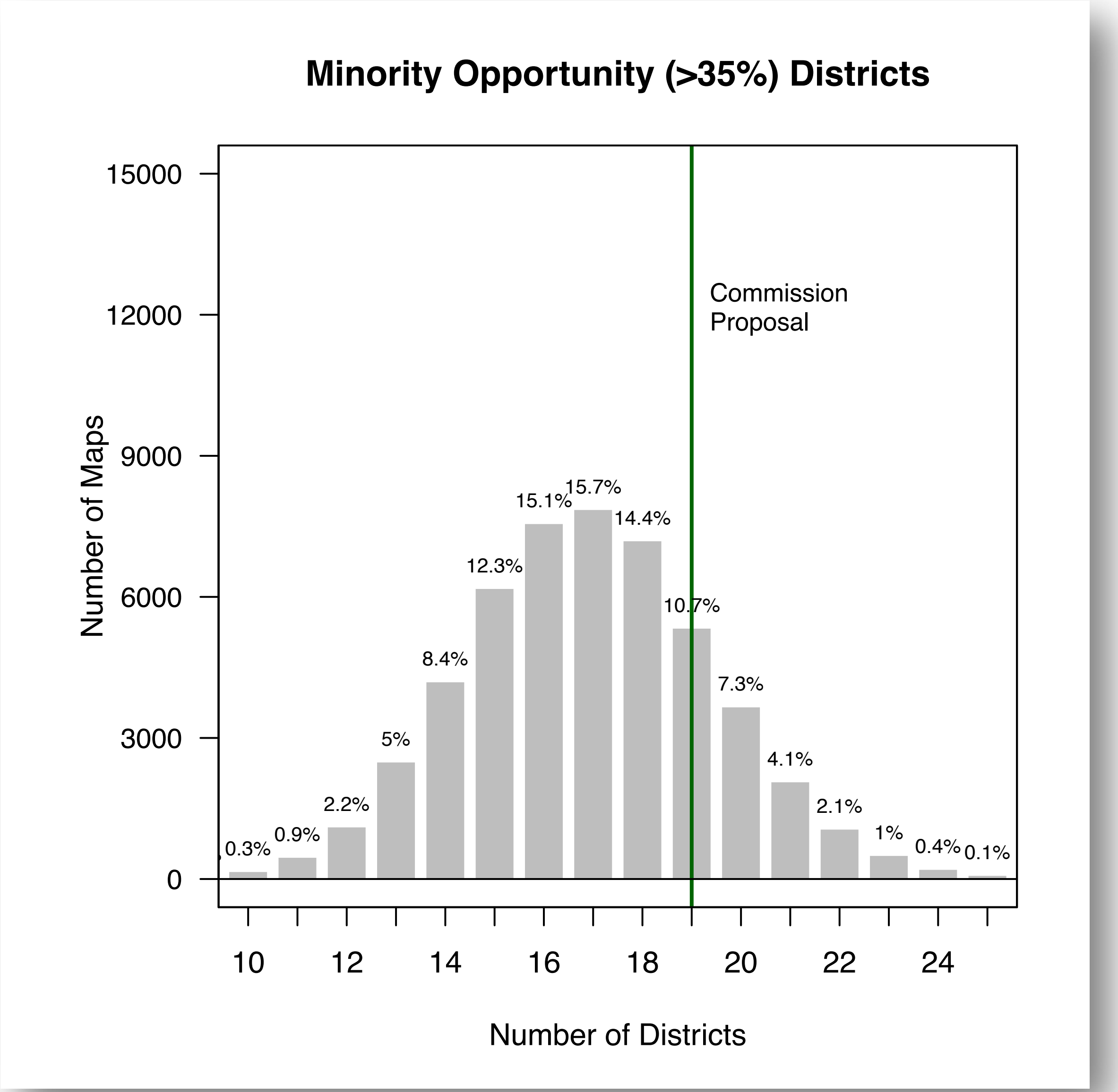
Reading



State College




Majority Minority Districts
Number of Maps
15000
12000
9000
6000
3000
0
Number of Districts
18
19
20
21
22
23
24
25
26
27
28
29
30
0%
0.2%
1.4%
5.7%
14.2%
23.4%
24.9%
17.9%
8.7%
2.9%
0.6%
0.1%
0%
Commission
Proposal


Minority Opportunity (>35%) Districts
Number of Maps
15000
12000
9000
6000
3000
0
Number of Districts
10
12
14
16
18
20
22
24
0.3%
0.9%
2.2%
5%
8.4%
12.3%
15.1%
15.7%
14.4%
10.7%
7.3%
4.1%
2.1%
1%
0.4%
0.1%
Commission
Proposal

Dr. Imai’s Report

- Dr. Imai’s simulations are **less compact** and contain **many more municipal divisions.**
- In both Dr. Imai’s and my simulations, the LRC plan is a **statistically significant outlier**.
- Even when explicitly incorporating race, in **3 of 6 analyses the LRC plan is a statistical outlier.**

Thank You

WRITTEN TESTIMONY REGARDING THE PRELIMINARY STATE HOUSE PLAN FROM THE PENNSYLVANIA LEGISLATIVE REAPPORTIONMENT COMMISSION

Kosuke Imai, Ph.D.

January 14, 2022

Table of Contents

WRITTEN TESTIMONY

I. INTRODUCTION AND SCOPE OF WORK

1. My name is Kosuke Imai, Ph.D., and I am a Professor in the Department of Government and the Department of Statistics at Harvard University. I specialize in the development of statistical methods and computational algorithms and their applications to social science research. I am also affiliated with Harvard's Institute for Quantitative Social Science.

2. I have been engaged to analyze relevant data and provide my expert opinions on Professor Michael Barber's expert report, entitled "Memo on Proposed Redistricting Plan from PA Redistricting Commission." More specifically, I have examined how the consideration of race, in addition to constitutional criteria, may alter the conclusions of the race-blind redistricting simulation analysis Professor Barber conducted regarding the preliminary State House plan approved by the Legislative Reapportionment Commission (hereafter "preliminary plan").

3. Redistricting simulation analysis is a powerful methodology for the empirical evaluation of legislative districting plans. State-of-the-art redistricting simulation algorithms, which are based on Monte Carlo methods, generate a representative sample of all possible plans under a specified set of criteria. This allows analysts to evaluate a proposed plan by comparing it against the simulated alternative plans. Statistical theory lets us quantify the degree to which a proposed plan is extreme relative to the ensemble of simulated plans in terms of partisan outcomes. Statistically significant and substantively large differences in partisan outcomes between a proposed plan and simulated plans provide empirical evidence that the proposed plan may be a partisan gerrymander.

4. A primary advantage of the simulation-based approach over the traditional redistricting evaluation methods is its ability to account for the political and geographic features that are specific to each state, including spatial distribution of voters and configuration of administrative boundaries. Simulation algorithms can also incorporate each state's redistricting rules. These state-specific features limit the types of redistricting plans that can be drawn. This threatens the validity of traditional redistricting evaluation methods because comparison of different plans across states and over different time periods may be confounded by various state and time specific factors.

The simulation-based approach overcomes this problem by comparing a proposed plan against a representative set of alternate districting plans that could be drawn subject to Pennsylvania's constitutional and legal requirements. Given the importance of legislative redistricting in representative democracy, I welcome Professor Barber's efforts to assist the Commission through the application of cutting-edge redistricting simulation methodology. Appendix A provides an introduction to the redistricting simulation methodology.

5. I have examined how the consideration of race, in addition to constitutional criteria, can alter the conclusions of Professor Barber's race-blind simulation analysis regarding the preliminary plan. Professor Barber conducted his simulation analysis without using any information about race. I investigate how the expected partisan outcomes under simulated plans change once race is added as an additional constraint into simulation algorithms. My analysis exploits the ability of redistricting simulation methodology to determine how a specific factor influences the types of redistricting plans one could draw while adhering to other redistricting criteria. The key implication of my analysis is that analysts must carefully choose the inputs to redistricting simulation algorithms based on legal considerations.

6. I first conducted a *race-blind simulation analysis*.[1] Like Professor Barber's analysis, this simulation analysis does not use any information about race but otherwise is designed to be consistent with the reapportionment criteria specified in the Pennsylvania Constitution. Although Professor Barber's analysis is based on the same open-source software package `redist`, which my collaborators and I developed, the report does not provide sufficient information about the choice of simulation algorithm and the exact values of parameters used in his analysis. Unfortunately, this makes it impossible for me to replicate his analysis. Thus, I conducted my own race-blind simulation analysis.

7. I also conducted two simulation analyses that consider race, in addition to consti-

1. For simplicity and convenience, I use "race-blind" to refer to simulations based on criteria in the Pennsylvania Constitution and "alternative" to refer to simulations that incorporate information concerning race as well as the Constitutional criteria. Neither my use of these terms nor the simulations or analyses themselves should be misconstrued as suggesting that race was a predominant factor in the preliminary plan. I did not consider and my testimony does not address whether or how race or any other factor actually influenced the preliminary plan.

tutional criteria, when generating simulated plans. These alternative simulations follow the same set of redistricting criteria, and so the only difference is the additional consideration of race. The first simulation analysis, which is referred to as the *Simulation A analysis*, ensures that, in addition to constitutional criteria, every simulated plan identifies a certain number of majority black and majority Hispanic districts. I also conducted a second simulation analysis, which I refer to as the *Simulation B analysis.* This simulation analysis ensures that every simulated plan includes a certain number of majority-minority districts (MMDs), in addition to the constitutional criteria. These MMDs include coalition districts as well as majority black and majority Hispanic districts.

8. For each of the simulation analyses, I generated a representative set of 5,000 alternative plans that could be drawn under the corresponding set of redistricting criteria. I then compared the likely number of Democratic districts under the preliminary plan with that under each set of 5,000 simulated plans. To make my results comparable with those of Professor Barber's report, I used the same three sets of past statewide elections to compute the likely number of Democratic districts under each plan. Finally, I examined whether and how the consideration of race, in addition to constitutional criteria, alters the evaluation of the preliminary plan by comparing the conclusions of the race-blind simulation analysis with those of the three alternative simulation analyses that incorporate the information about race.

II. SUMMARY OF OPINIONS

9. My analysis shows that the additional consideration of race in simulation algorithms substantially alters the conclusions of redistricting simulation analyses. Under the race-blind simulation analysis, the preliminary plan yields a greater number of Democratic districts than the simulated plans. In comparison to the race-blind analysis, the other simulation analyses, which incorporate race in simulation algorithms, reveal that the difference in the likely number of Democratic districts between the preliminary and simulated plans is much smaller in magnitude and is often not statistically significant. The consideration of race in addition to constitutional criteria in the redistricting simulation analysis, therefore, leads to the conclusion that the proposed plan is not a partisan gerrymander. Below, I summarize my specific findings.

- Consistent with the main finding of Professor Barber's report, my *race-blind simulation analysis* shows that without any consideration of race, the preliminary plan yields a greater number of Democratic districts than the race-blind simulated plans. Although this difference between the preliminary and race-blind simulated plans is statistically significant, the magnitude of the difference is smaller under my race-blind simulation analysis by approximately 2 to 4 districts than Professor Barber's race-blind analysis. This finding contradicts Professor Barber's conclusion that the preliminary plan generates an additional 8 to 10 Democratic districts.

- The *Simulation A analysis* shows that ensuring a certain number of majority black and majority Hispanic districts under each simulated plan as an additional constraint substantially alters the conclusions of race-blind simulation analysis. According to this analysis, the most likely number of Democratic districts under the preliminary plan is greater than that under the simulated plans by 2 to 6 districts, which are 2 to 4 districts fewer than what I found under my race-blind analysis and 4 to 7 districts smaller than the results of Professor Barber's race-blind analysis. These differences between the simulated plans and the preliminary plan are no longer statistically significant, depending on an election set used to measure partisanship.

- My *Simulation B analysis* shows that ensuring a certain number of majority-minority districts (MMDs) under each simulated plan as an additional constraint further narrows the gap between the preliminary and simulated plans. If I use the 2012–2020 statewide elections to measure the partisan outcome, the difference in the most likely number of Democratic districts between the preliminary and simulated plans is only 1 district and is not statistically distinguishable from zero. This contrasts with a statistically significant finding of 6 districts reported in Professor Barber's report. Furthermore, if I instead use the 2014–2020 statewide elections, the preliminary plan most likely yield one *less* Democratic district than the simulated plans. This result contradicts Professor Barber's finding that using the 2014–2020 statewide elections, the preliminary plan most likely gains additional 6 Democratic districts when compared to the simulated plan.

III. QUALIFICATIONS AND EXPERIENCE

10. I am trained as a political scientist (Ph.D. in 2003, Harvard) and a statistician (MA in 2002, Harvard). I have published more than 60 articles in peer reviewed journals, including premier political science journals (e.g., *American Journal of Political Science*, *American Political Science Review*, *Political Science*), statistics journals (e.g., *Biometrika*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society*), and general science journals (e.g., *Lancet*, *Nature Human Behavior*, *Science Advances*). My work has been widely cited across a diverse set of disciplines. For each of the past four years, Clarivate Analytics, which tracks citation counts in academic journals, has named me as a highly cited researcher in the cross-field category for producing "multiple highly cited papers that rank in the top 1% by citations for field and year in Web of Science."

11. I started my academic career at Princeton University, where I played a leading role in building interdisciplinary data science communities and programs on campus. I was the founding director of Princeton's Program in Statistics and Machine Learning from 2013 to 2017. In 2018, I moved to Harvard, where I am Professor jointly appointed in the Department of Government and the Department of Statistics, the first such appointment in the history of the university. Outside of universities, between 2017 and 2019, I served as the president of the Society for Political Methodology, a primary academic organization of more than one thousand researchers worldwide who conduct methodological research in political science. My introductory statistics textbook for social scientists, *Quantitative Social Science: An Introduction* (Princeton University Press, 2017), has been widely adopted at major research universities in the United States and beyond.

12. Computational social science is one of my major research areas. As part of this research agenda, I have developed simulation algorithms for evaluating legislative redistricting since the beginning of this emerging literature. At Harvard, I lead the Algorithm-Assisted Redistricting Methodology (ALARM; https://alarm-redist.github.io/) Project, which studies how algorithms can be used to improve legislative redistricting practice and evaluation.

13. Back in 2014, along with Jonathan Mattingly's team at Duke, my collaborators and I were the first to use Monte Carlo algorithms to generate an ensemble of redistricting plans. Since then, my team has written several methodological articles on redistricting simulation algorithms (Fifield, Higgins, et al. 2020; Fifield, Imai, et al. 2020; McCartan and Imai 2020; Kenny et al. 2021).

14. I have also developed an open-source software package titled `redist` that allows researchers and policy makers to implement the cutting-edge simulation methods developed by us and others (Kenny et al. 2020). This software package can be installed for free on any personal computer with Windows, Mac, or Linux operating system. According to a website that tracks the download statistics of R packages, our software package has been downloaded about 30,000 times since 2016 with an increasing download rate.[2]

IV. METHODOLOGY

15. I conducted *race-blind* and alternative simulation analyses to evaluate the partisan outcomes expected under the proposed House plan. I then compared the conclusions of these simulation analyses to determine whether and how the consideration of race, in addition to constitutional criteria, affects the evaluation of the preliminary plan. The *race-blind* and alternative simulation analyses I conducted critically differ in terms of whether race was used as an additional input to the simulation algorithms. Thus, any differences between my *race-blind* and alternative simulation analyses can be attributed to the additional consideration of race in generating simulated plans.

16. I simulated three sets of possible Pennsylvania state House districting plans that adhere to redistricting considerations. Below, I provide a brief overview of my simulation analysis setup while leaving the details to Appendix B.

A. Race-blind Simulation Setup

17. The first set of 5,000 alternative plans were generated without any consideration of race, mirroring the simulation analysis conducted by Professor Barber. I call them *race-blind*

2. https://ipub.com/dev-corner/apps/r-package-downloads/ (accessed on January 6, 2022)

simulated plans. Given the aforementioned difficulty of replicating Professor Barber's analysis, I designed my own race-blind simulation procedure that generates 5,000 alternative plans under the following five reapportionment criteria based on Article II § 16 of the Pennsylvania Constitution:

- there are a total of 203 geographically contiguous districts
- all districts do not exceed an overall population deviation of $\pm$ 5%
- simulated plans are encouraged to be more compact
- simulated plans are encouraged to split fewer number of counties
- simulated plans are encouraged to split fewer number of municipalities

18. There are several differences in terms of constraints between my race-blind simulation and Professor Barber's. In my analysis, I used a population deviation threshold of $\pm$ 5%. Professor Barber used a slightly different population deviation threshold of $\pm$ 4%. In addition, Professor Barber appears to have included an additional constraint which discourages county multi-splits.[3] I did not include this constraint because Article II § 16 does not distinguish types of county splits and simply states "Unless absolutely necessary no county [. . .] shall be divided in forming [. . .] a [. . .] representative district." The other constraints appear to be qualitatively similar to those described in Professor Barber's report. The lack of detailed information in his report, however, makes it impossible to know how these constraints are imposed.

19. I generated 5,000 race-blind simulated plans by considering the aforementioned five criteria alone, using the Sequential Monte Carlo (SMC) simulation algorithm (McCartan and Imai 2020; Kenny et al. 2021; briefly described in Appendix B). Importantly, my race-blind simulation procedure does not use the information about race at all. One could run the SMC algorithm on the entire state, but I found that doing so results in a poor algorithmic performance indicated by low sampling efficiency and lack of plan diversity. In such instances, it is recommended to divide up the state into several urban and rural regions and independently run the SMC algorithm within each region. I adopted this approach using a total of six regions, which represent major urban and rural

3. Professor Barber writes, "The model is further instructed that when a county boundary needs to be crossed, it should avoid splitting the county more times than necessary." (page 4.)

areas within Pennsylvania (see Figure A5 in Appendix B). They are defined based on the location of VRA-related districts later used in my race-aware simulation analyses. Although other choices are possible, the main advantage of this approach is that it will make my race-blind simulation analysis directly comparable to my race-aware simulation analyses. I found that the resulting race-blind simulated plans are generally less compact and have more administrative boundary splits than the preliminary plan (see Appendices E and F).

B. Alternative Simulation Setups Considering Race

20. I also generated two alternative sets of 5,000 simulated plans using the information about race. Specifically, I instructed my simulation algorithm to create the specified number of majority-minority districts (hereafter "VRA-related districts"; see Appendix C), but otherwise followed the same five redistricting criteria as the race-blind simulation procedure used for the first set. Like my race-blind analysis, these alternative simulation analyses do not use partisan information when generating simulated districts. Similar to the *race-blind* simulation, I found that the resulting alternative simulated plans are generally less compact and have more administrative boundary splits than the preliminary plan (see Appendices E and F).

21. I conducted two alternative simulation analyses that incorporate the consideration of race in addition to constitutional criteria. The *Simulation A analysis* ensures that every simulated plan has a total of 8 majority black districts and 4 majority Hispanic districts. I also conducted the so-called *Simulation B analysis,* which instructs the simulation algorithm to generate a total of 25 majority-minority districts (MMDs) in every simulated plan. These MMDs include 13 coalition districts as well as the same set of 8 majority black and 4 majority Hispanic districts included in the *Simulation A* analysis. Other than the difference in the use of VRA-related districts, these two alternative sets of 5,000 simulated plans were generated under the same set of five redistricting criteria listed above.

22. The two alternative simulation analyses proceed in two steps. First, I independently generated 100 sets of relevant VRA-related districts in each of the regions, which contain at least one such district. Here, I used the merge-split Markov chain Monte Carlo (MCMC) algorithm

(Autry et al. 2020) with the preliminary plan as a starting map. Second, for each set of the simulated VRA-related districts, I used the SMC algorithm to generate 50 sets of the remaining districts in the same way as done in my *race-blind* simulation analysis. Putting these simulated districts together, I obtained a total of 5,000 simulated plans.

C. Partisan Outcome Measure

23. To measure the partisan outcome under a given plan, I exactly follow Professor Barber's approach and compute the likely number of Democratic districts. Although there are other ways to measure partisan outcomes and biases under redistricting plans, this allows me to directly compare the results of my simulation analysis with those presented in Professor Barber's report. Specifically, I first average a set of elections vote totals for each party at the precinct level. Then, under a given redistricting plan, I calculate the number of districts out of the 203 total districts where Democrats have more votes than Republicans. This yields the total number of Democratic districts given the plan. I compute this using the same exact three sets of past statewide elections as those used by Professor Barber: 2012–2020 statewide elections, 2014–2020 statewide elections, and 2020 statewide elections (see footnote 4 of his report for the complete list).[4] I note that the calculation based on the 2020 statewide elections is likely to be unreliable because it relies on the smallest number of elections and may be greatly influenced by the factors specific to the 2020 general election. Appendix D briefly describes data sources used in my analysis.

D. Description of Redistricting Simulation Software

24. In my analysis, I use the aforementioned open-source software package for redistricting analysis `redist` (Kenny et al. 2020), which implements a variety of redistricting simulation algorithms as well as other evaluation methods. My collaborators and I have developed this software package, so that other researchers and the general public can implement these state-of-the-art methods on their own. Our software package is freely available for download and can be used on one's personal computer. Indeed, Professor Barber used this software package, and so our analyses

4. Applying this method to my data, I obtained the same number of Democratic districts under the proposed plan as the one reported in Professor Barber's report. The only exception is the 2012–2020 statewide elections where my calculation yields 106 Democratic districts whereas Professor Barber reports 107 districts.

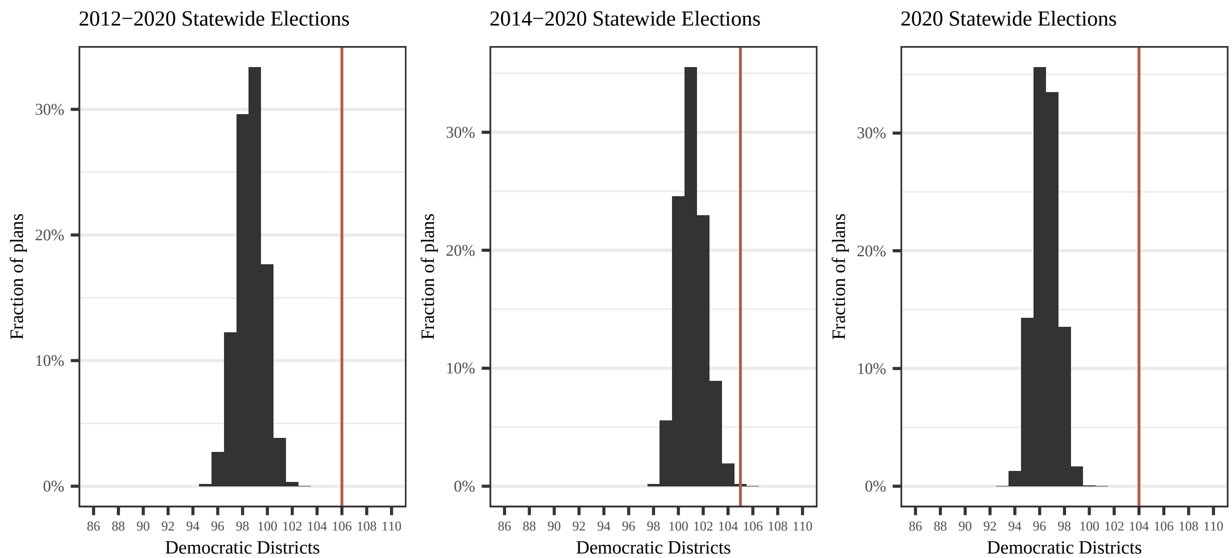


Figure 1: The likely number of Democratic districts across 5,000 *race-blind* simulated plans. Democratic districts are tallied based on an average of statewide elections for the 2012–2020 cycles (left), the 2014–2020 cycles (middle), and 2020 cycle (right). The red vertical lines represent the results under the preliminary plan.

are based on the identical implementation of the same simulation algorithms.

V. EMPIRICAL FINDINGS

25. I now present the results of my simulation analysis. I begin by presenting the results of my *race-blind* simulation analysis and then discuss the findings from my two alternative simulation analyses.

A. Race-blind Simulation Analysis Results

26. Figure 1 presents the likely number of Democratic districts across 5,000 *race-blind* simulated plans, using three different sets of statewide elections. The figure shows that the most likely number of Democratic districts is greater under the preliminary plan than under the *race-blind* simulated plans. The differences are 7, 4, and 8 districts, using the 2012–2020, 2014–2020, and 2020 statewide elections, respectively. These differences are statistically significant though the preliminary plan is within the range of the 5,000 simulated plans if one uses the 2014–2020 statewide elections.

27. For the sake of comparison, Figure 2 reproduces the results of Professor Barber's

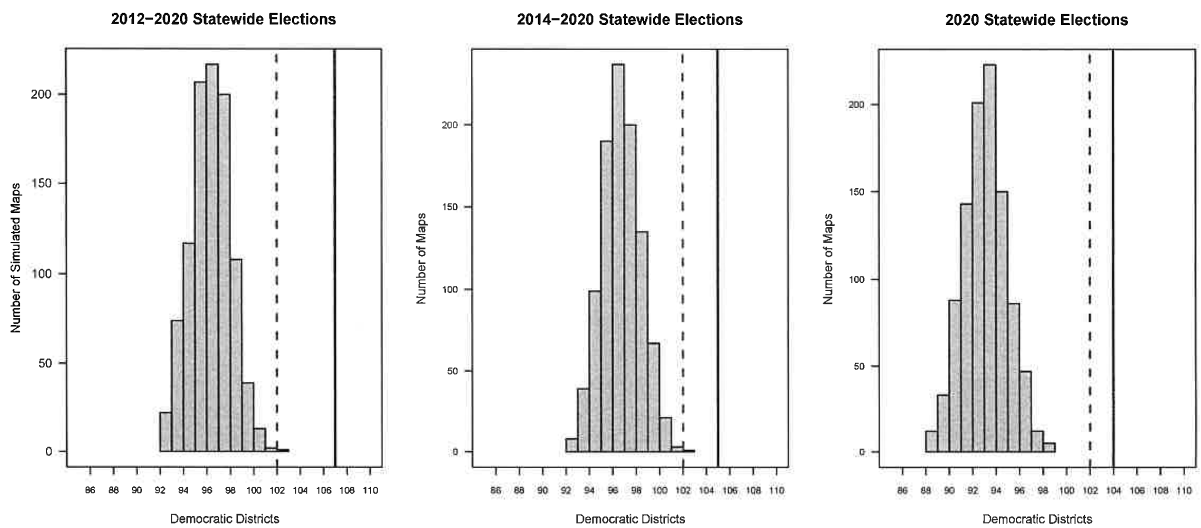


Figure 2: The likely number of Democratic districts across the *race-blind* simulated plans reported by Professor Michael Barber in his report entitled, "Memo on Proposed Redistricting Plan from PA Redistricting Commission." Democratic districts are tallied based on an average of statewide elections for the 2012-2020 cycles (left), the 2014-2020 cycles (middle), and 2020 cycle (right).

race-blind simulation analysis (reported as Figure 1 in his report). Both analyses agree that the preliminary plan has a greater number of Democratic districts than the *race-blind* simulated plans. A careful comparison of the two figures, however, reveals that this difference is smaller under my race-blind simulation analysis than Professor Barber's analysis. The differences between the two analyses are approximately 3, 4, and 3 districts using the 2012–2020, 2014–2020, and 2020 statewide elections, respectively. Professor Barber concludes that the preliminary plan gains about additional 8 to 10 Democratic districts when compared to his race-blind simulated plans. In contrast, my race-blind simulation analysis shows that this difference is most likely about 4 to 8 districts. Given the lack of detailed information about Professor Barber's simulation analysis in his report, I am unable to identify the precise reasons for this difference between my and Professor Barber's race-blind simulation analyses.

B. Simulation A Results

28. Figure 3 presents the results of the *Simulation A* analysis, which incorporates 8 majority black districts and 4 majority Hispanic districts. When compared to the results of my

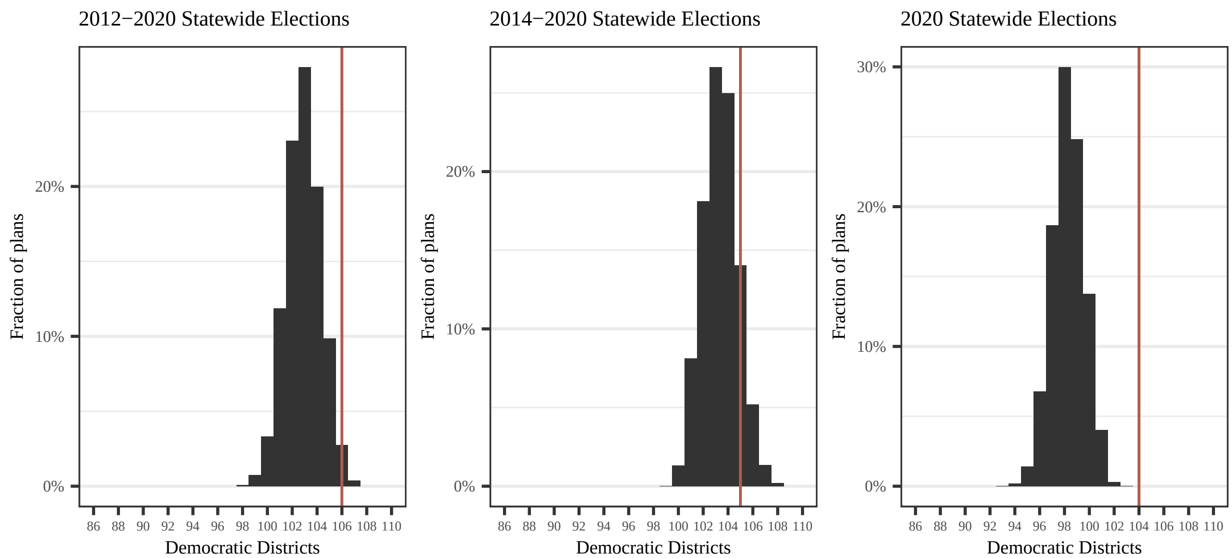


Figure 3: The likely number of Democratic districts across the *Simulation A* plans, each of which has 8 black majority and 4 Hispanic majority districts. Democratic districts are tallied based on an average of statewide elections for the 2012-2020 cycles (left), the 2014-2020 cycles (middle), and 2020 cycle (right). The red vertical lines represent the results under the preliminary plan.

race-blind simulation analysis in Figure 1, the most likely number of Democratic districts under the *Simulation A* plans tends to be greater by about 4, 2, and 2 districts using the 2012–2020, 2014–2020, and 2020 statewide elections, respectively. This shows that the consideration of race based on the identification of majority black and majority Hispanic districts, in addition to constitutional criteria, can substantially alter the conclusion of the *race-blind* simulation analysis.

29. As a result, the difference between the preliminary plan and the *Simulation A* plans is now reduced to 3, 2, and 6 districts using the 2012–2020, 2014–2020, and 2020 statewide elections, respectively. In particular, the difference based on the 2014–2020 statewide elections is no longer statistically significant, whereas the difference using the 2012–2020 statewide elections has a borderline statistical significance. For these sets of elections, the preliminary plan is well within the range of the 5,000 simulated plans. These findings substantially differ from the results based on Professor Barber's race-blind analysis as well as those based on my own race-blind simulation, implying that the consideration of race can alter the conclusions of race-blind simulation analysis. The difference for the 2020 statewide election remains to be statistically significant although the

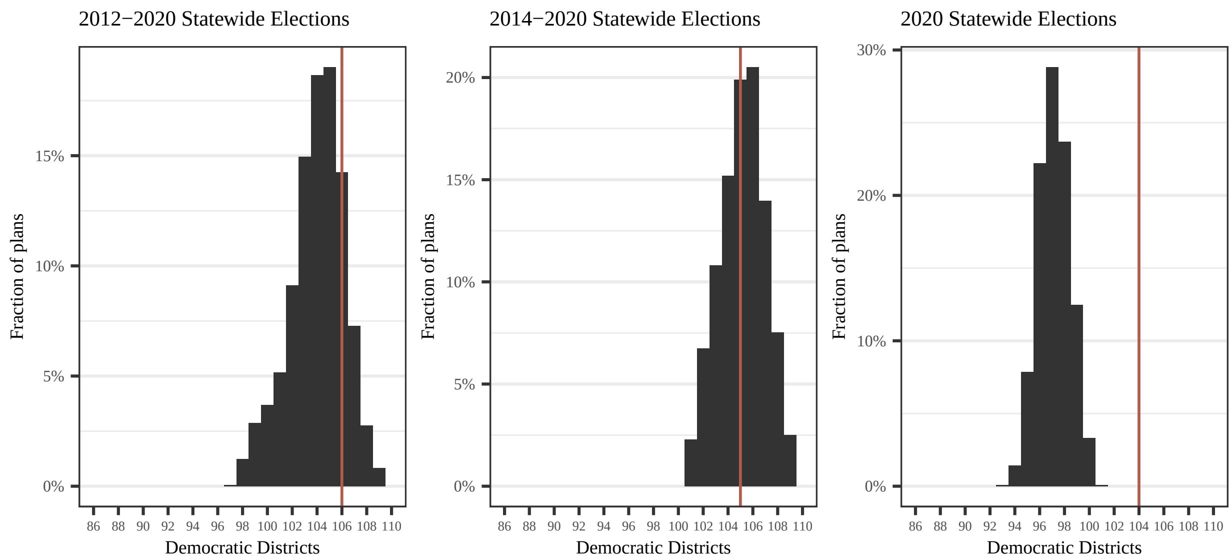


Figure 4: The likely number of Democratic districts across the *Simulation B* plans, each of which has 25 majority-minority districts. Democratic districts are tallied based on an average of statewide elections for the 2012-2020 cycles (left), the 2014-2020 cycles (middle), and 2020 cycle (right). The red vertical lines represent the results under the preliminary plan.

difference in the most likely number of Democratic districts is now reduced to 6 districts rather than 8 districts in my race-blind simulation.

C. Simulation B Results

30. Next, I present the results of my *Simulation B* analysis, which incorporates a total of 25 majority-minority districts (MMDs). Figure 4 shows that the *Simulation B* plans are no longer statistically distinguishable from the preliminary plan using the 2012–2020 and 2014–2020 statewide elections. Indeed, the difference is only 1 district using the 2012–2020 statewide elections. If one uses the 2014–2020 statewide elections, the most likely number of Democratic districts under the *Simulation B* plans is 105, which is one *less* than what would be expected under the preliminary plan. This partisan outcome obtained under the preliminary plan is the second most likely outcome under this simulation analysis (the same outcome is obtained under about 20% of the 5,000 simulated plans). These findings contradict the results of Professor Barber's *race-blind* analysis that the preliminary plan gains additional 8 to 10 Democratic districts. The estimate based on the 2020 statewide elections alone has not substantially changed from that of the *Simulation A*

analysis.

VI. APPENDIX

A. Introduction to Redistricting Simulation

1. In recent years, redistricting simulation algorithms have played an increasingly important role in court cases involving redistricting plans. Simulation evidence has been presented to courts in many states, including Michigan, North Carolina, Ohio, and Pennsylvania.[5]

2. Over the past several years, researchers have made major scientific advances to improve the theoretical properties and empirical performance of redistricting simulation algorithms. All of the state-of-the-art redistricting simulation algorithms belong to the family of Monte Carlo methods. They are based on random generation of spanning trees, which are mathematical objects in graph theory (DeFord, Duchin, and Solomon 2021). The use of these random spanning trees allows these state-of-the-art algorithms to efficiently sample a representative set of plans (Autry et al. 2020; Carter et al. 2019; McCartan and Imai 2020; Kenny et al. 2021). Algorithms developed earlier, which do not use random spanning trees and instead rely on incremental changes to district boundaries, are often not able to do so.

3. These algorithms are designed to sample plans from a specific probability distribution, which means that every legal redistricting plan has certain odds of being generated. The algorithms put as few restrictions as possible on these odds, except to ensure that, on average, the generated plans meet certain criteria. For example, the probabilities are set so that the generated plans reach a certain level of geographic compactness, on average. Other criteria, based on the state in question, may be fed into the algorithm by the researcher. In other words, this target distribution is based on the weakest assumption about the data under the specified constraints.

4. In addition, the algorithms ensure that all of the sampled plans (a) are geographi-

5. Declaration of Dr. Jonathan C. Mattingly, Common Cause v. Lewis (2019); Testimony of Dr. Jowei Chen, Common Cause v. Lewis (2019); Testimony of Dr. Pegden, Common Cause v. Lewis (2019); Expert Report of Jonathan Mattingly on the North Carolina State Legislature, Rucho v. Common Cause (2019); Expert Report of Jowei Chen, Rucho v. Common Cause (2019); Amicus Brief of Mathematicians, Law Professors, and Students in Support of Appellees and Affirmance, Rucho v. Common Cause (2019); Brief of Amici Curaiae Professors Wesley Pegden, Jonathan Rodden, and Samuel S.-H. Wang in Support of Appellees, Rucho v. Common Cause (2019); Intervenor's Memo, Ohio A. Philip Randolph Inst. et al. v. Larry Householder (2019); Expert Report of Jowei Chen, League of Women Voters of Michigan v. Benson (2019).

cally contiguous, and (b) have a population which deviates by no more than a specified amount from a target population.

5. There are two types of general Monte Carlo algorithms which generate redistricting plans with these guarantees and other properties: sequential Monte Carlo (SMC; Doucet, Freitas, and Gordon 2001) and Markov chain Monte Carlo (MCMC; Gilks, Richardson, and Spiegelhalter 1996) algorithms.

6. The SMC algorithm (McCartan and Imai 2020; Kenny et al. 2021) samples many redistricting plans in parallel, starting from a blank map. First, the algorithm draws a random spanning tree and removes an edge from it, creating a "split" in the map, which forms a new district. This process is repeated until the algorithm generates enough plans with just one district drawn. The algorithm calculates a weight for each plan in a specific way so that the algorithm yields a representative sample from the target probability distribution. Next, the algorithm selects one of the drawn plans at random. Plans with greater weights are more likely to be selected. The algorithm then draws another district using the same splitting procedure and calculates a new weight for each updated plan that comports with the target probability distribution. The whole process of random selection and drawing is repeated again and again, each time drawing one additional district on each plan. Once all districts are drawn, the algorithm yields a sample of maps representative of the target probability distribution.

7. The MCMC algorithms (Autry et al. 2020; Carter et al. 2019) also form districts by drawing a random spanning tree and splitting it. Unlike the SMC algorithm, however, these algorithms do not draw redistricting plans from scratch. Instead, the MCMC algorithms start with an existing plan and modify it, merging a random pair of districts and then splitting them a new way.

8. Diagnostic measures exist for both these algorithms which allow users to make sure the algorithms are functioning correctly and accurately. The original papers for these algorithms referenced above provide more detail on the algorithm specifics, empirical validation of their performance, and the appropriateness of the chosen target distribution.

B. Implementation Details

B.1. Race-blind simulation analysis

9. In my *race-blind* simulation analysis, I use the SMC algorithm because, unlike the MCMC algorithms, the SMC algorithm generates nearly independent samples, leading to a diverse set of redistricting plans that satisfy the specified constraints. The *race-blind* simulation analysis proceeds in two steps. First, I divide the state into five clusters and a geographically larger remainder. I do so to ensure proper sampling diversity. Using the SMC algorithm on the full state has difficulties generating a diverse range of plans according to the standard assessment of sampling diversity; see McCartan and Imai 2020 for details. These clusters are both chosen to improve plan diversity and to maintain continuity of analysis between my *race-blind* and alternative simulations. The clusters are as follows, primarily based on counties.

- Region A: Allegheny
- Region B: York, Lancaster, Berks, Lehigh, Northampton, Monroe, Luzerne, Carbon, Schuylkill, Dauphin, Lebanon, Columbia
- Region C: Philadelphia, Delaware, and Chester
- Region D: Montgomery
- Region E: Bucks

10. Regions A and B include small portions of adjacent counties so that the region can comply with population tolerance requirements. In addition, a small portion of Montgomery county is included in Region B; this is because both the enacted and proposed maps break the county boundary in the same place. It should be noted that Region E is separated solely for reasons of contiguity, as Regions B, C, and D separate it from the remainder of the map. A map of these regions can be seen in Figure A5.

11. Article II § 16 of the Pennsylvania Constitution states districts "shall be composed of compact and contiguous territory as nearly equal in population as practicable." The SMC algorithm generates contiguous districts by design. In clusters A, B, C, D, and E, I use a compactness

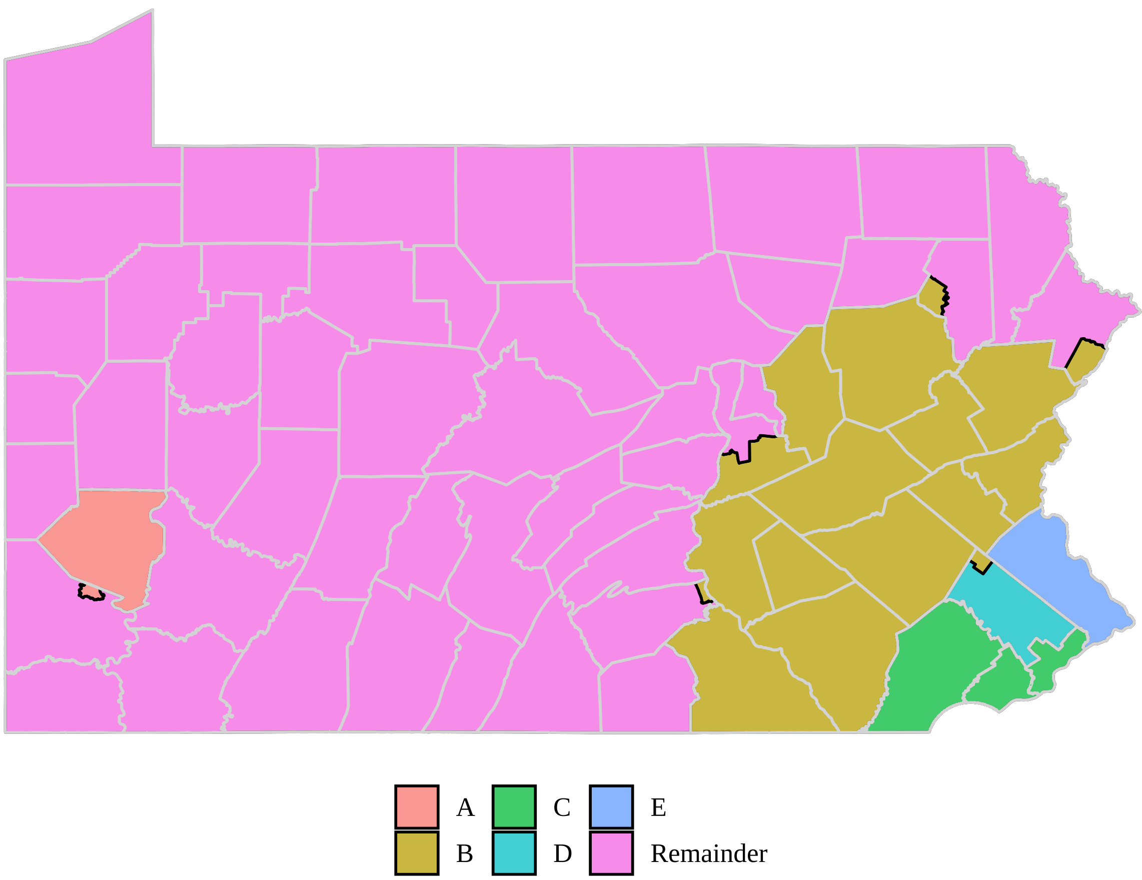


Figure A5: A map of the five clusters and remainder. Light grey lines indicate county lines. Black lines indicate where region boundaries deviate from the county lines.

parameter of $\rho = 1$. Larger values of ρ lead to more compact districts at a cost of sampling efficiency, which must be retained to allow for a diverse sample. In the remainder, I use a compactness parameter of $\rho = 1.001$, which I found to be a maximum value while maintaining proper sampling diversity. I also used a population deviation threshold of $\pm 5\%$.

12. The same article also states "Unless absolutely necessary no county, city, incorporated town, borough, township or ward shall be divided in forming either a senatorial or representative district." To address this, I use two types of constraints: first, the hierarchical county split constraint of the SMC algorithm, and second, an additional constraint on administrative splits of the form $C_{\text{splits}} n_{\text{splits}}$. For this second constraint, C_{splits} is a tuning parameter, and n_{splits} is the number of administrative splits. In each of the five smaller clusters, I apply these constraints to municipalities, with $C_{\text{splits}} = 1$ along with the hierarchical constraint. In the larger remainder cluster, I apply these constraints to counties, using the hierarchical constraint and $C_{\text{splits}} = 0.5$. Values of C_{splits} were selected based on simulation experiments with the data; higher values, which would yield districts with fewer county and municipality splits, diminished the diversity of maps generated.

13. To conduct the simulations, I generated 5,000 simulated maps in each smaller region of the map, and in the remainder. By combining maps, I generate a total of 5,000 plans. Note that this analysis only considers the population tolerance, the number of municipalities and counties split, and the compactness of the districts.

B.2. Alternative simulation analyses that incorporate the consideration of race

14. Using a similar two step approach as the *race-blind* simulations, I sample two alternative sets of simulated plans while incorporating race, in addition to constitutional criteria, into simulation algorithms. I conducted these alternative simulations that consider particular VRA-related districts (see Appendix C). When generating plans in regions which include VRA districts, I used the merge-split MCMC algorithm (Autry et al. 2020; Carter et al. 2019). I initialized the algorithm from the proposed plan within the region. Otherwise, I used the same *race-blind* simulations in that region, generated with the SMC algorithm. I directed the merge-split algorithm so

that it would consider VRA-related districts within each region. I do so by building constraints into the algorithm, to generate maps that include the desired VRA-related districts with higher probabilities.

15. Similar to SMC, the merge-split algorithm generates districts that are contiguous by design. In addition, it only generates plans that meet the specified population tolerance, which is $\pm 5\%$ in this case. For compactness, I again use a compactness parameter of $\rho = 1$ in each region. For administrative splits, note that because merge-split is only used in the individual regions, I apply administrative splits constraints to municipalities. I again set $C_{\text{splits}} = 1$ for the municipality splits constraint. In addition, I incorporate an algorithmic constraint to decrease municipality splits.

16. I use two types of constraints to target VRA-related districts. The first are indicator constraints, which take the form of an indicator function I_K that is equal to 0 if condition K is satisfied, and 1 otherwise. For example, if I intended to sample maps in Region A with 1 district with 50%-60% MVAP population share, K would correspond condition that one such district exists in a given map. Higher constraint values of this type make it less likely for a map to be generated, so maps where K is satisfied are sampled more frequently.

17. The second type of constraint are hinge constraints. There are two forms: a hinge-up and a hinge-down. Hinge-up constraints take the form $\sqrt{\max(x-t,0)}$ where x is the observed population fraction in a particular district (in terms of BVAP, HVAP, or MVAP) and t is the target fraction in that district. Hinge-down constraints take a similar form: $\sqrt{\max(t-x,0)}$, using the same inputs. Again, maps with higher constraint values of this type are less likely to be generated. Hinge-up constraints encourage maps to have districts with population fractions greater than or equal to t, while hinge-down constraints encourage maps to have districts with population fractions less than or equal to t. This form of constraint is a common way to formulate VRA-related constraints (Herschlag et al. 2020).

18. Within each region, I ran this algorithm for 10,000 iterations and obtained 100 sets of simulated VRA-related districts by discarding the first 5,000 draws and storing every 50 draws thereafter. Once I had obtained 100 sets of VRA-related districts with the desirable characteristics,

I ran the race-blind simulations analysis on the remainder of the state. Specifically, I take each set of the simulated VRA-related districts and use the SMC algorithm to generate the remaining districts in the rest of the state without using race information. I use the same constraints as in the earlier race-blind analysis in the SMC algorithm, to increase the compactness of districts and limit the number of county splits. Similarly, I incorporate municipality split constraint into the region based simulations and a county split constraint into the remainder simulations. Combining these together yields a total of 5,000 race-conscious simulated plans.

19. I run two versions of the alternative analyses that incorporate race. The *Simulation A* analysis only imposes VRA-related constraints in regions B and C. The *Simulation B* analysis imposes additional VRA-related constraints in regions B and C, along with new VRA-related constraints in region A.

C. VRA-related Districts

20. The definition of the black voting-age population (BVAP) is based on non-Hispanic blacks who are 18 years and older, the Hispanic voting-age population (HVAP) is defined using Hispanics who are 18 years and older, and the definition of the minority voting-age population (MVAP) is based on the total voting age population minus the voting age population for whites. It should be noted that additional VRA-related districts may appear in the simulations incidentally.

C.1. Simulation A

In the *Simulation A* analysis, I imposed constraints in Regions B and C to consider 8 majority black districts 4 majority Hispanic districts. In Regions A, D, and E and in the remainder, I used the same race-blind simulations generated with SMC.

C.2. Simulation B

In the *Simulation B* analysis, I imposed constraints in Regions A, B, and C to consider 25 majority-minority districts (MMDs) in every simulated plan. The MMD includes 13 coalition districts as well as the same set of 8 majority black and 4 majority Hispanic districts included in the *Simulation A* analysis. In Regions D, E, and in the remainder, I used the same race-blind simulations generated with SMC.

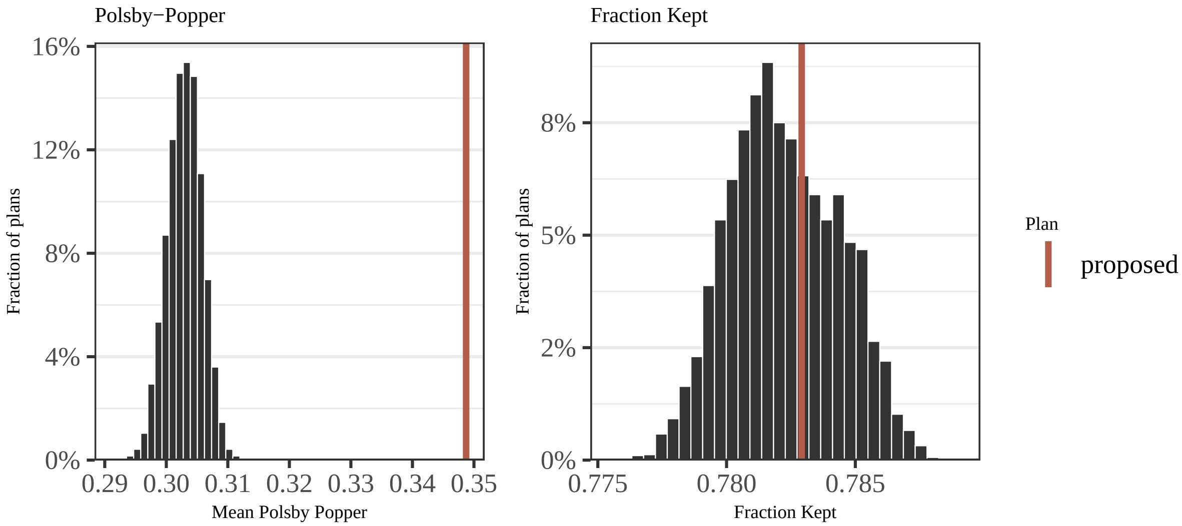


Figure A6: The compactness of the *race-blind* simulated plans according to two measures – the average Polsby-Popper compactness (left) and fraction of edges kept (right). The red vertical line represents the preliminary plan.

D. Data Sources

21. The 2012, 2014, 2016, 2018, and 2020 VTD-level election results were provided by counsel. These were disaggregated proportionally by voting age population down to the 2020 Census block shapefile. The block level data were then aggregated to geographically contiguous components of VTDs. This resulted in the splitting of about 100 VTD to ensure contiguity.

22. The 2020 Census Block shapefiles, total population by race and ethnicity, and voting age population by race and ethnicity were obtained directly from the Census Bureau's Decennial Census API. The VTD and MCD block assignment files block assignment files came from the Census website. The data about the proposed plan were provided by counsel.

E. Compactness of the Simulated Plans

23. I find that the preliminary plan is generally more compact than the simulated plans, and therefore is more compliant with Article II § 16 in this regard. One may be able to modify the simulation analyses so that the generated plans are as compact as the preliminary plan, but this was not possible given the time constraint.

24. I use the average Polsby–Popper (Polsby and Popper 1991) and edge-removal (De-

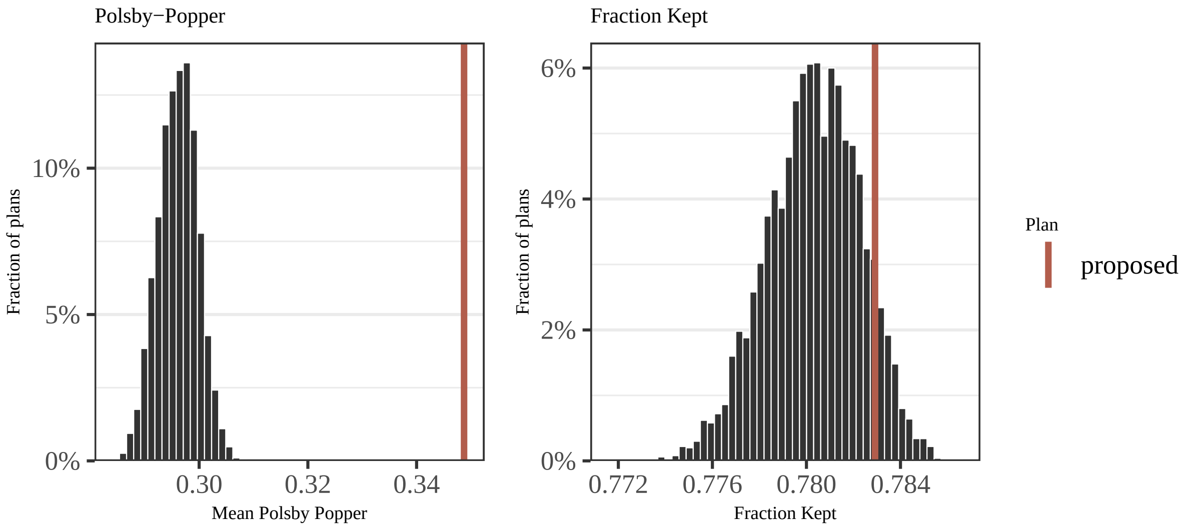


Figure A7: The compactness of the *Simulation A* plans according to two measures – the average Polsby-Popper compactness (left) and fraction of edges kept (right). The red vertical line represents the preliminary plan.

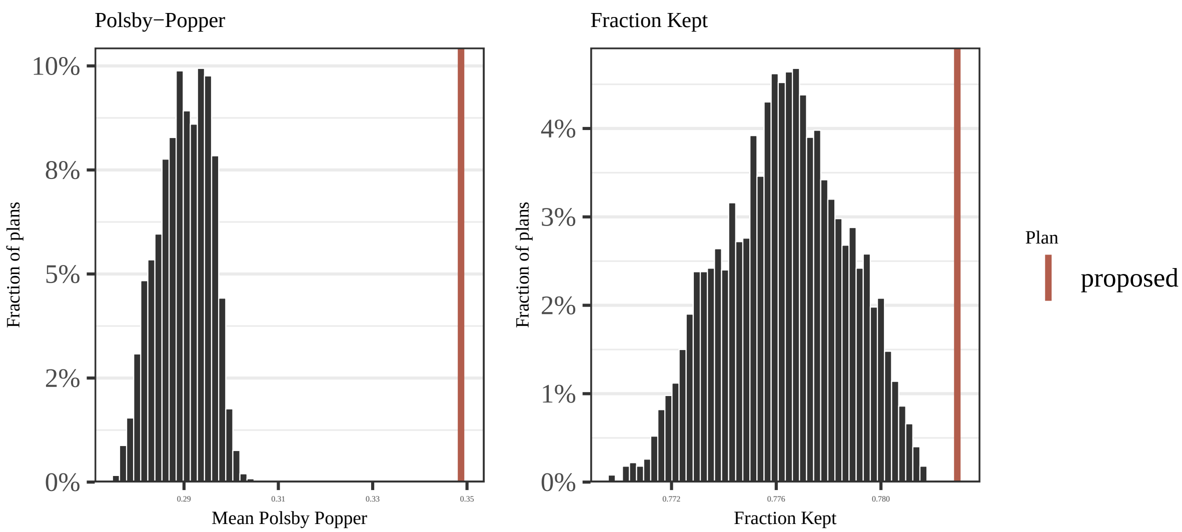


Figure A8: The compactness of the *Simulation B* plans according to two measures – the average Polsby-Popper compactness (left) and fraction of edges kept (right). The red vertical line represents the preliminary plan.

Ford, Duchin, and Solomon 2021; McCartan and Imai 2020) scores, two commonly-used quantitative measures of district compactness. For the edge-removal compactness, I present the fraction of edge kept so that like the Polsby–Popper score, a greater value implies a higher level of compactness. Figure A6 shows that the preliminary plan is similar to the *race-blind* simulated plans in terms of edge-removal compactness, and more compact in terms of the average Polsby–Popper compactness. Figure A7 shows similar results when comparing the preliminary plan to the *Simulation A* plans; the main difference is that in terms of edge-removal compactness, the preliminary plan is more compact than a larger fraction of the *Simulation A* plans than the *race-blind* simulations. Finally, Figure A8 shows that when I compare the preliminary plan to the *Simulation B* plans, the preliminary plan is more compact in terms of both Polsby–Popper and edge-removal compactness than all of the simulated plans.

F. Administrative Splits of the Simulated Plans

25. I now show that the preliminary plan has a fewer number of municipality splits, and a similar number of county splits when compared to all three versions of my simulated plans; therefore, the preliminary plan is more compliant with Article II § 16 in this regard than the simulated plans. Again, one may be able to modify the simulation analyses so that the generated plans split as few municipalities as the preliminary plan, but this was no possible given the time constraint.

26. The right-hand panels of Figures A9, A10, and A11 demonstrate the preliminary plan splits many fewer municipalities than any version of the simulated plans. The left panel of Figure A9 shows that compared to the *race-blind* simulations, the preliminary plan splits fewer counties. The left panels of Figures A10 and A11 show that compared to the alternative simulations, the preliminary plan splits a similar number of counties.

G. References

Autry, Eric, Daniel Carter, Gregory Herschlag, Zach Hunter, and Jonathan Mattingly. 2020. "Multi-scale merge-split Markov chain Monte Carlo for Redistricting." *arXiv preprint arXiv:2008.08054.*

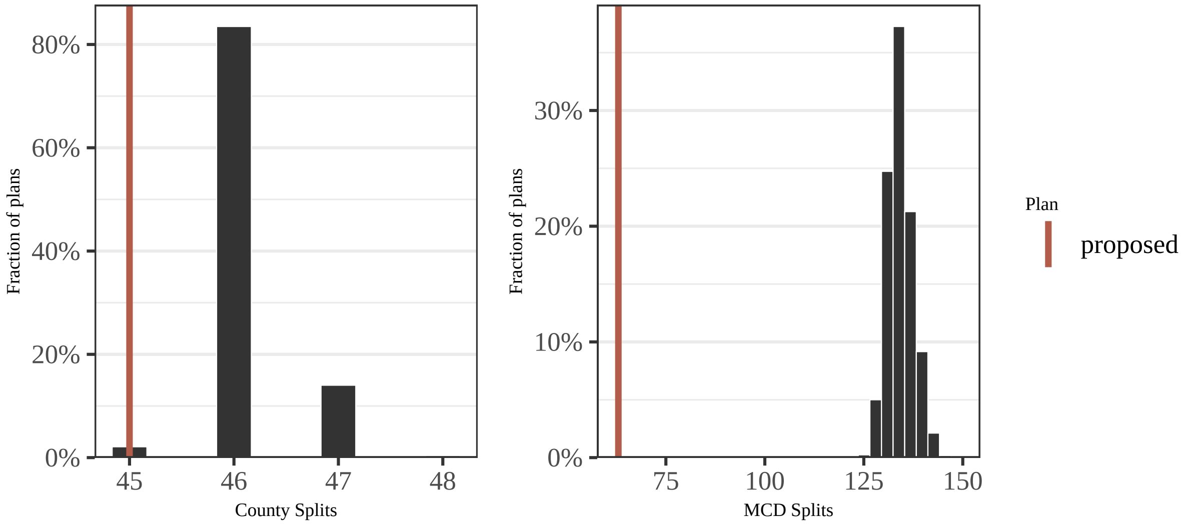


Figure A9: The number of administrative splits in the *race-blind* simulated plans (histogram). An administrative unit is deemed as split if any of its precincts are assigned to different districts. The left plot presents the total number of split counties while the right plot shows the number of minor civil divisions that are split. The red vertical line represents the preliminary plan.

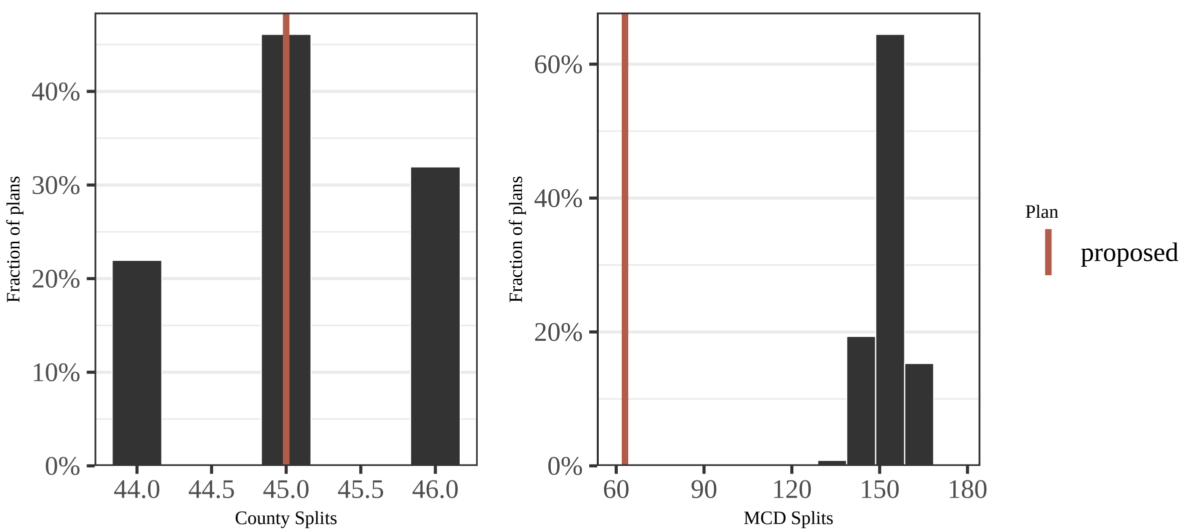


Figure A10: The number of administrative splits in the *Simulation A* plans (histogram). An administrative unit is deemed as split if any of its precincts are assigned to different districts. The left plot presents the total number of split counties while the right plot shows the number of minor civil divisions that are split. The red vertical line represents the preliminary plan.

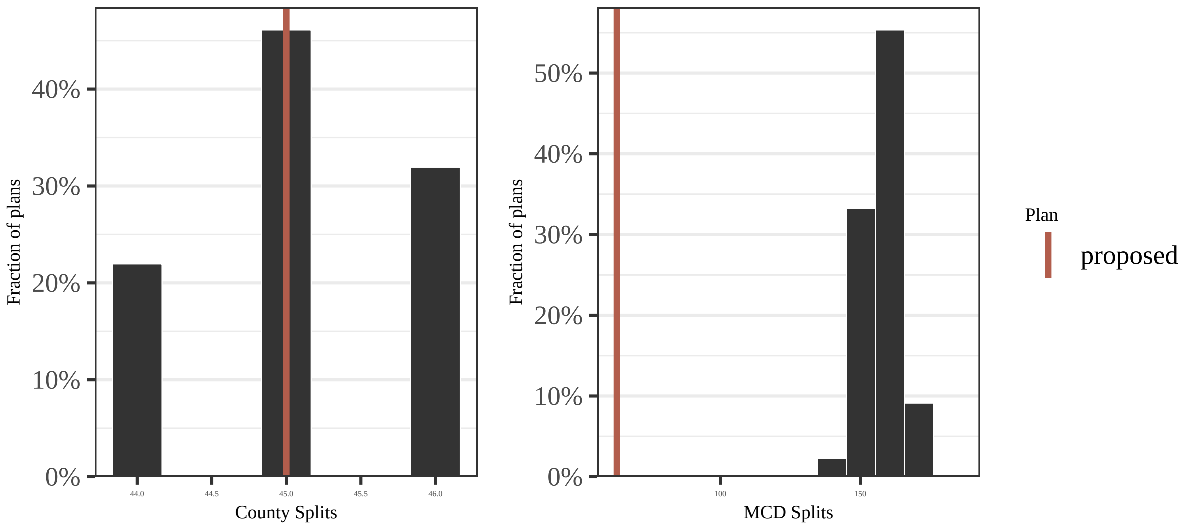


Figure A11: The number of administrative splits in the *Simulation B* plans (histogram). An administrative unit is deemed as split if any of its precincts are assigned to different districts. The left plot presents the total number of split counties while the right plot shows the number of minor civil divisions that are split. The red vertical line represents the preliminary plan.

Carter, Daniel, Gregory Herschlag, Zach Hunter, and Jonathan Mattingly. 2019. "A Merge-Split Proposal for Reversible Monte Carlo Markov Chain Sampling of Redistricting Plans." *arXiv preprint arXiv:1911.01503.*

DeFord, Daryl, Moon Duchin, and Justin Solomon. 2021. "Recombination: A Family of Markov Chains for Redistricting." Https://hdsr.mitpress.mit.edu/pub/1ds8ptxu, *Harvard Data Science Review* (March 31, 2021). https://doi.org/10.1162/99608f92.eb30390f. https://hdsr.mitpress.mit.edu/pub/1ds8ptxu.

Doucet, Arnaud, Nando de Freitas, and Neil Gordon. 2001. *Sequential Monte Carlo methods in practice.* New York: Springer.

Fifield, Benjamin, Michael Higgins, Kosuke Imai, and Alexander Tarr. 2020. "Automated Redistricting Simulation Using Markov Chain Monte Carlo." *Journal of Computational and Graphical Statistics* 29 (4): 715–728.

Fifield, Benjamin, Kosuke Imai, Jun Kawahara, and Christopher T Kenny. 2020. "The essential role of empirical validation in legislative redistricting simulation." *Statistics and Public Policy* 7 (1): 52–68.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter. 1996. *Markov chain Monte Carlo in Practice.* London: Chapman & Hall.

Herschlag, Gregory, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier, and Jonathan C Mattingly. 2020. "Quantifying gerrymandering in North Carolina: Supplemental Appendix." *Statistics and Public Policy* 7 (1): 30–38.

Kenny, Christopher T., Shiro Kuriwaki, Cory McCartan, Evan Rosenman, Tyler Simko, and Kosuke Imai. 2021. "The Use of Differential Privacy for Census Data and its Impact on Redistricting: The Case of the 2020 U.S. Census." *Science Advances* 7, no. 41 (October): 1–17.

Kenny, Christopher T., Cory McCartan, Benjamin Fifield, and Kosuke Imai. 2020. *redist: Computational Algorithms for Redistricting Simulation.* https://CRAN.R-project.org/package=redist.

McCartan, Cory, and Kosuke Imai. 2020. "Sequential Monte Carlo for sampling balanced and compact redistricting plans." *arXiv preprint arXiv:2008.06131.*

Polsby, Daniel D, and Robert D Popper. 1991. "The third criterion: Compactness as a procedural safeguard against partisan gerrymandering." *Yale Law & Policy Review* 9 (2): 301–353.

Redistricting Simulation Analysis of the Preliminary State House Reapportionment Plan

Kosuke Imai

Harvard University

January 14, 2022

Introduction

- Positions
  - Current: Professor in the Department of Government and Department of Statistics, Harvard University
  - Previous: Professor in the Department of Politics and Center for Statistics and Machine Learning, Princeton University

- Research fields
  1. Causal inference
  2. Computational social science

- Relevant expertise
  - Redistricting simulation analysis
  - Development and application of simulation algorithms
  - Open-source software package *redist* (over 30,000 downloads)

Overview of redistricting simulation analysis

- What is redistricting simulation analysis?
  - generate a large number of alternative plans under a specified set of redistricting criteria
  - compare them with a proposed plan to evaluate its properties

- What are the benefits of redistricting simulation analysis?
  1. can control for state-specific political geography and redistricting rules
  2. transparency and ability to isolate a relevant factor
  3. mathematical guarantee $\rightsquigarrow$ representative sample of alternative plans

- Input criteria to simulation algorithms must be carefully chosen

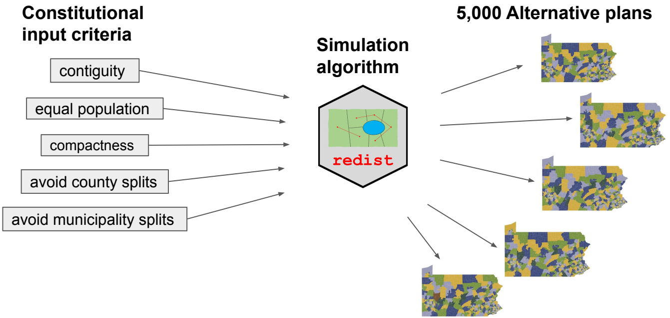
Key conclusions

1. The consideration of majority-minority districts, in addition to constitutional constraints, in simulation algorithms substantially alters the conclusions of simulation analyses

2. When the majority-minority districts are considered, there is no empirical evidence that the preliminary plan is a partisan gerrymander

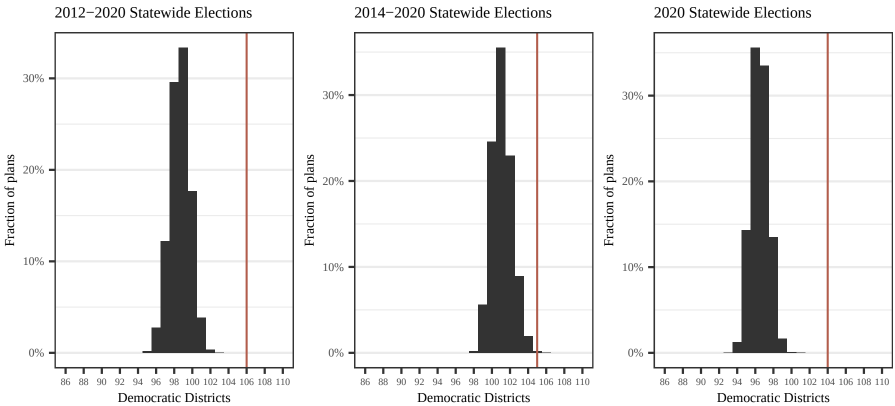
Race-blind simulation setup

- 5 constitutional criteria
  1. 203 geographically contiguous districts
  2. equal population ($\pm$ 5%)
  3. compactness
  4. avoid county splits
  5. avoid municipality splits



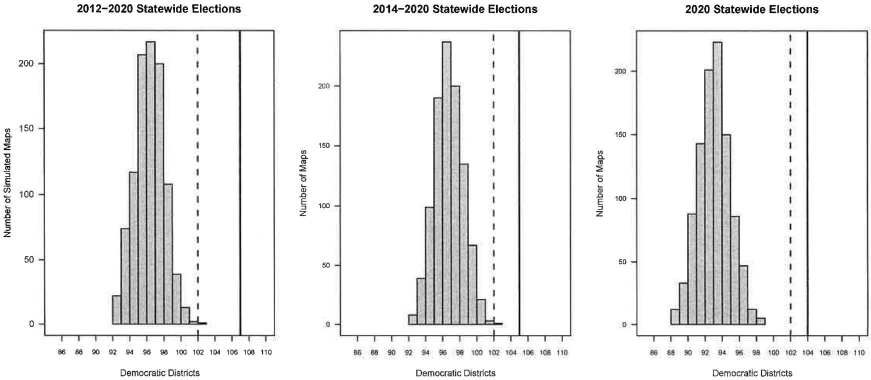
- Could not replicate Prof. Barber's race-blind simulation due to insufficient information

Race-blind simulation results



- I used the same 3 sets of statewide elections as Professor Barber: other composite of statewide elections may produce different results
- The preliminary plan yields 4 to 8 more Democratic districts than the *race-blind* simulated plans

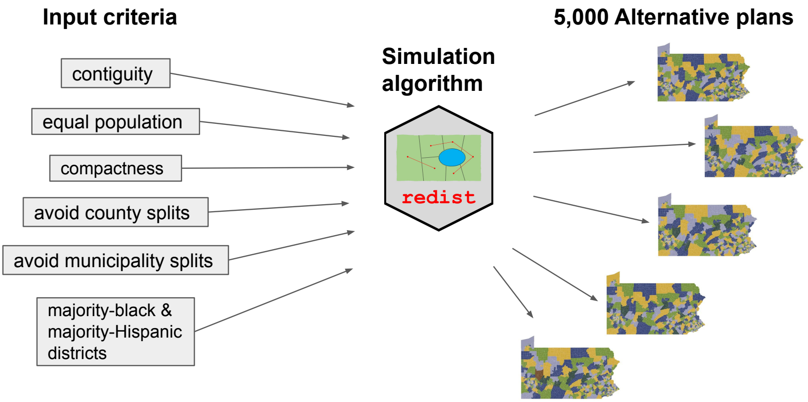
Comparison with Professor Barber's results



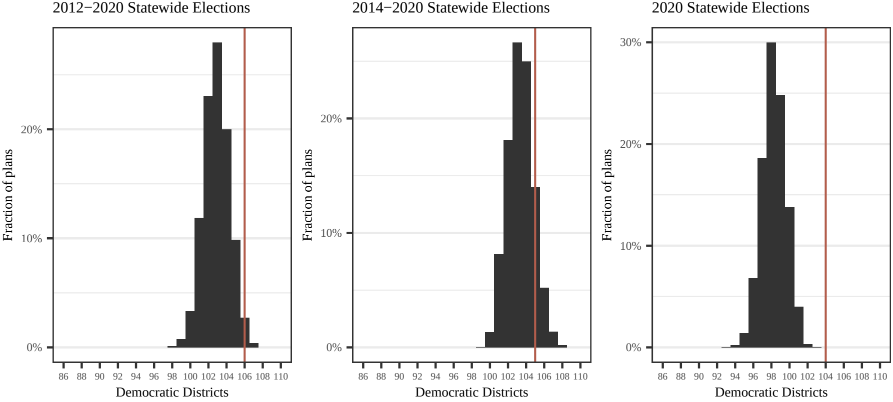
- Professor Barber's race-blind analysis substantially underestimates the likely number of Democratic districts in comparison to my *race-blind* simulation analysis

Simulation A setup

- 5 constitutional constraints are met
- Additional constraint for 8 majority-black and 4 majority-Hispanic districts



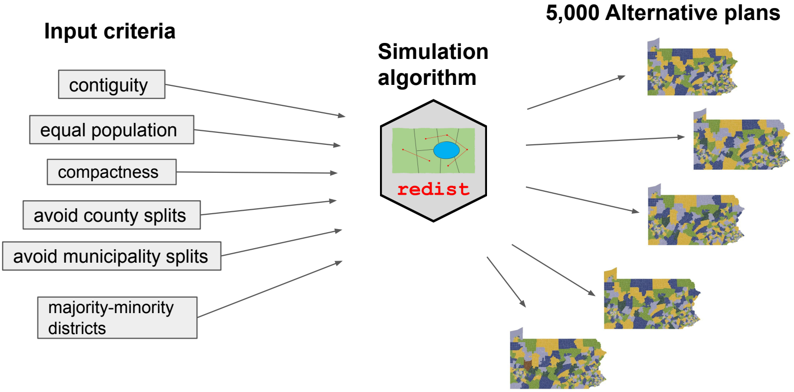
Simulation A results



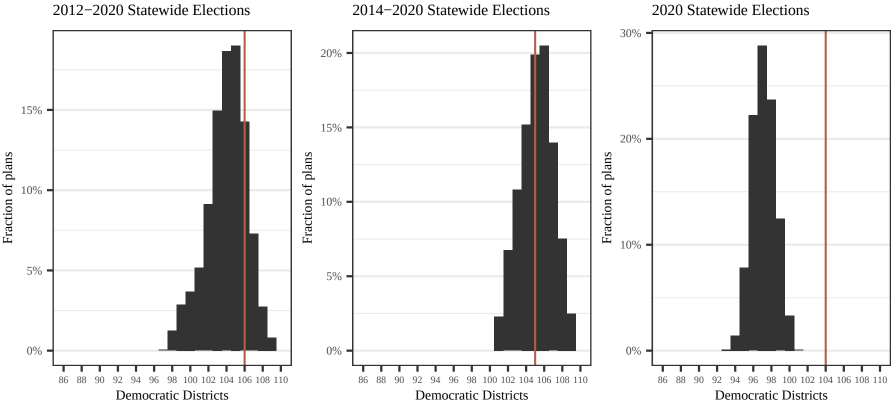
- The preliminary plan is not statistically distinguishable from the simulated plans, depending on the specific set of elections analyzed
- When the majority-black and majority-Hispanic districts are additionally considered, the preliminary plan is not a partisan gerrymander, depending on the specific set of elections analyzed

Simulation B setup

- 5 constitutional constraints are met
- Additional constraint for 25 majority-minority districts including 13 coalition districts



Simulation B results



- The preliminary plan is not statistically distinguishable from the simulated plans, using the 2012–2020 and 2014–2020 elections
- When the majority-minority districts are additionally considered, the preliminary plan is not a partisan gerrymander, using the 2012–2020 and 2014–2020 elections

Summary of findings

1. My race-blind simulation analysis shows that the preliminary plan most likely yields 2 to 4 fewer Democratic districts than Prof. Barber's analysis implies
2. When the majority-black and majority-Hispanic districts are additionally considered, the preliminary plan is not statistically distinguishable from the simulated plans, depending on the specific set of elections analyzed
3. When the majority-minority districts are additionally considered, the preliminary plan is not statistically distinguishable from the simulated plans, using the 2012–2020 and 2014–2020 statewide elections
4. When the majority-minority districts are additionally considered, the preliminary plan is not a partisan gerrymander in terms of the likely number of Democratic districts

January 14, 2022

Assessment of Population Change and Voting Patterns in Pennsylvania

Review of Federal Voting Rights Act

Dr. Matt Barreto, UCLA Political Science & Chicana/o Studies
Faculty Director of the UCLA Voting Rights Project

matt@uclavrp.org 909.489.2955

VOTING RIGHTS PROJECT UCLA

Current Landscape in Pennsylvania

- Population change in Pennsylvania was driven by communities of color while the White population declined

	2010	2020	Change
Total	12,702,379	13,002,700	300,321 (2.4%)
White	10,094,652 (80%)	9,553,417 (73%)	-541,235 (-5.4%)
Latino	719,660 (6%)	1,049,615 (8%)	329,955 (45.8%)
Black	1,327,091 (11%)	1,368,978 (11%)	41,887 (3.2%)
Asian	346,288 (3%)	506,674 (4%)	160,386 (46.3%)
Multi-racial	178,595 (1.4%)	451,285 (3.5%)	272,690 (152.7%)

Current Landscape in Pennsylvania

- Each legislative district is about 64,000

	Pop Change	Districts
White	-541,235	-8.4
Non-White	+841,556	+13.1
Total shift	1,382,791	

Total population shift: 1,382,791 from White to non-White

Represents 10.6% shift of the total 2020 population

10.6% of 203 districts is 21.5 seats that could <u>move</u> based on population changes

Figure 1: Population Changes in Allegheny County 2010 to 2020 (White, Black)

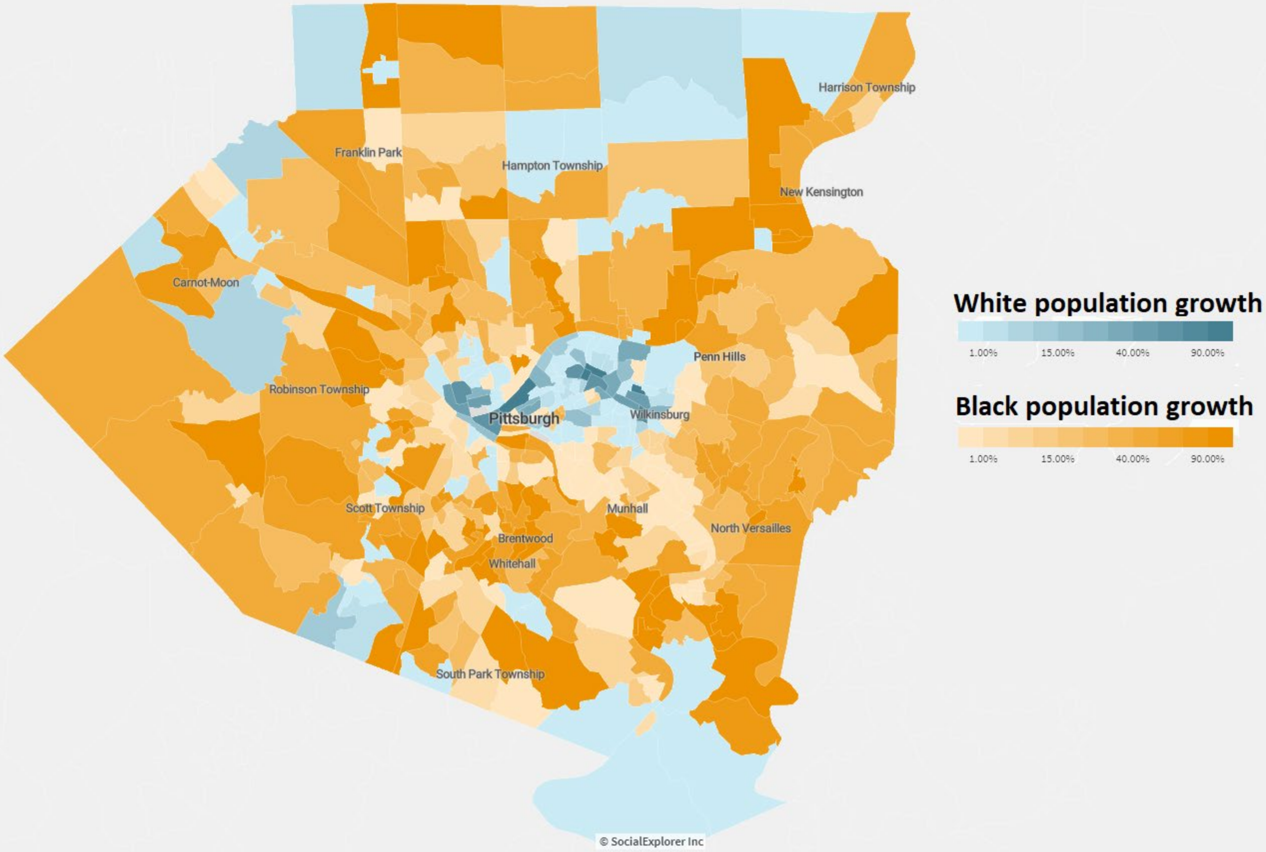
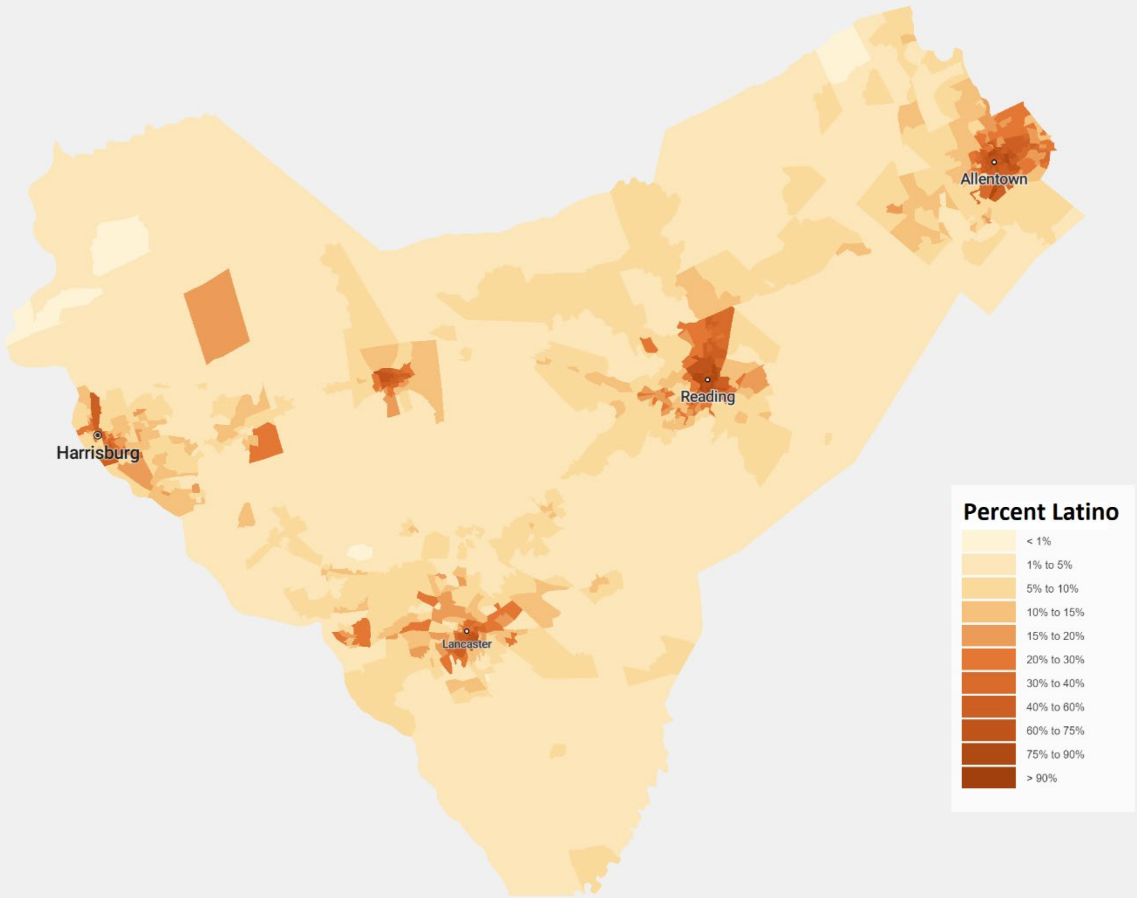
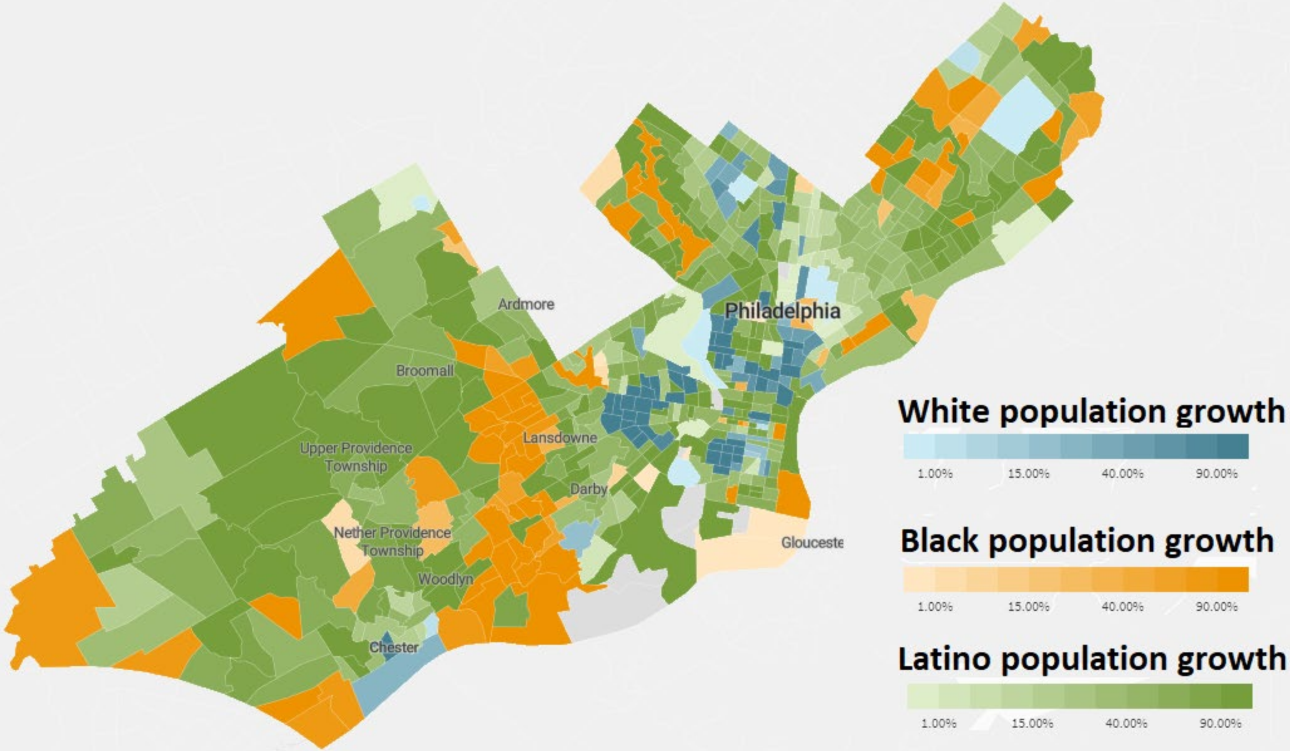


Figure 2: Percent Latino Central Pennsylvania 2020 Census



- Since 2000, Latino population has grown from 111,377 to 309,301 or 178%
- White population has declined by 49,680 (-4%)
- Latino population growth in this region is expected to continue at same rate for next decade

Figure 3: Population Changes in Philadelphia and Delaware Counties 2010 to 2020 (White, Latino, Black)



Section 2 of the Federal VRA

Section 2(b) A violation of subsection (a) is established if, based on the totality of circumstances, it is shown that the political processes leading to nomination or election in the State or political subdivision are not equally open to participation by members of a class of citizens protected by subsection (a) in that its members have less opportunity than other members of the electorate to participate in the political process and to elect representatives of their choice. The extent to which members of a protected class have been elected to office in the State or political subdivision is one circumstance which may be considered: *Provided*, That nothing in this section establishes a right to have members of a protected class elected in numbers equal to their proportion in the population.

Section 2 of the Federal VRA

- Specifically, the VRA Section 2 prohibits districting plans that use racial gerrymandering to dilute minority rights to meaningful opportunity to elect candidates of choice

- Has been used by Black, Latino, AAPI, Native American, White plaintiffs to challenge districting schemes that draw lines in a way that either "pack" or "crack" their population so it does not have meaningful influence

- State redistricting plans must comply with the Federal Voting Rights Act

Gingles: Coalition & Performing Districts

- If a district is already performing for minority-preferred candidates, its population can change, but it must continue performing for minority choices

- Districts do not need to be super-majority Black or Hispanic
  - Can be considered “packing” and likely prevents the minority group from having influence in a second nearby district

- Courts have allowed Black + Hispanic population to be combined in majority-minority coalition districts

Gingles: Minority vote cohesion

- Extensive analysis across Pennsylvania concludes that Minority voters are politically cohesive in supporting their candidates of choice

- Majority voters (White) usually vote together to defeat minority preferred candidates

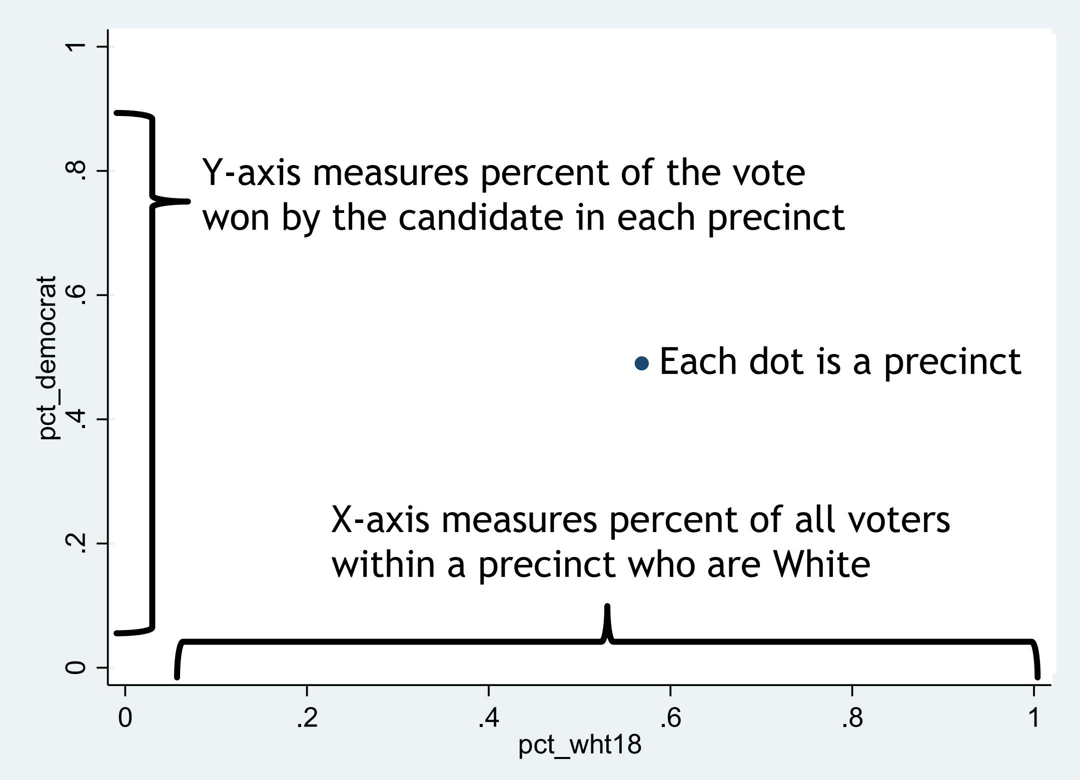
- To assess voting patterns, we conducted court-required ecological inference (EI) analysis

Gingles: Minority vote cohesion

- To assess voting patterns, we conducted court-required ecological inference (EI) analysis
  - However, our data are easily confirmed by major exit polls for recent elections which show minority voters are cohesive
    - CNN 2020: Black/Latino combined 84% Biden to 13% Trump
    - CNN 2020: White voters 42% Biden to 57% Trump
  - So our statistical analysis should come as no surprise to anyone who follows voting trends in Pennsylvania

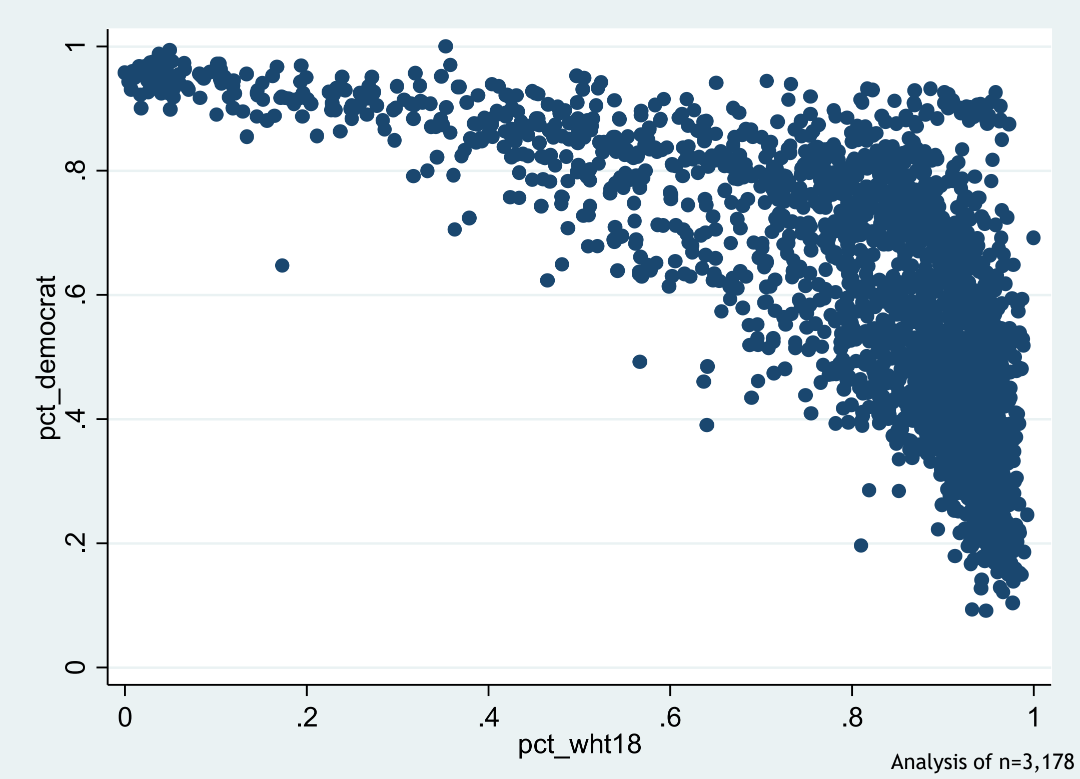
Measuring Racially Polarized Voting

2020 State House – Percent Voting Democrat by Race **Western PA**



Measuring Racially Polarized Voting

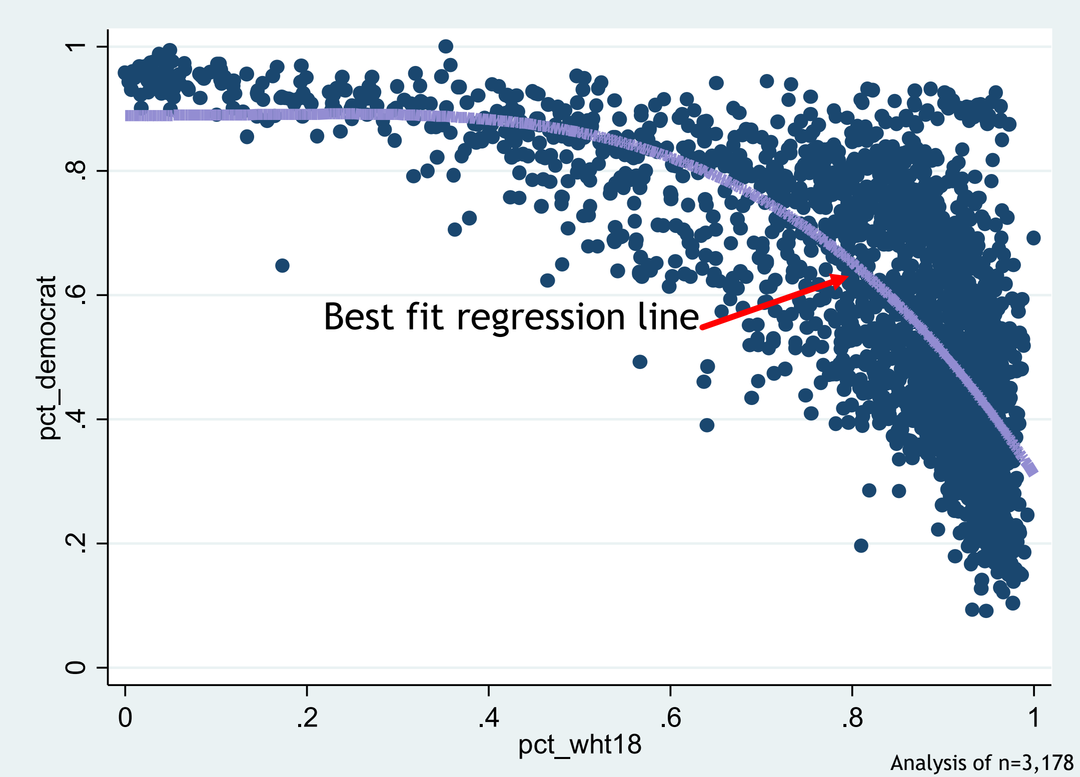
2020 State House – Percent Voting Democrat by Race **Western PA**



Analysis of n=3,178 precincts in Western PA

Measuring Racially Polarized Voting

2020 State House – Percent Voting Democrat by Race **Western PA**



Ecological inference estimates:

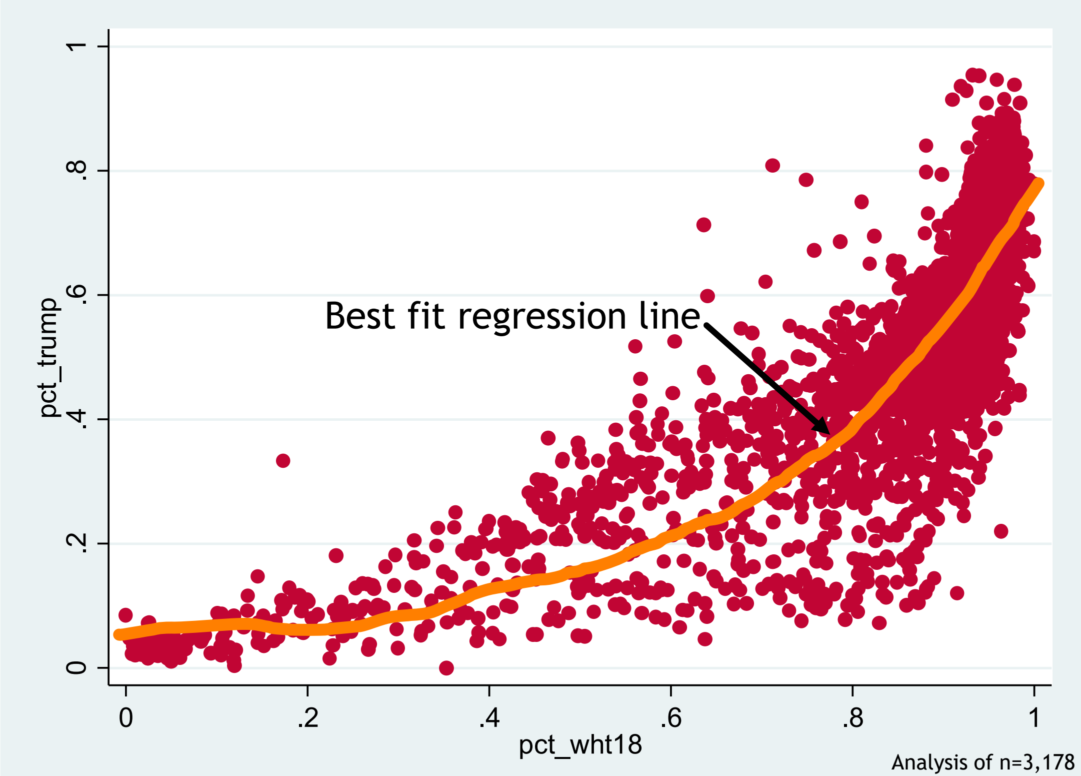
White vote: 29% Dem
Non-White: 91% Dem
Racial polarization: 62

Analysis of n=3,178 precincts in Western PA

Measuring Racially Polarized Voting

2020 President – Percent Voting Trump by Race

Western PA



Ecological inference estimates:

White vote: 77% Trump
Non-White: 11% Trump
Racial polarization: 66

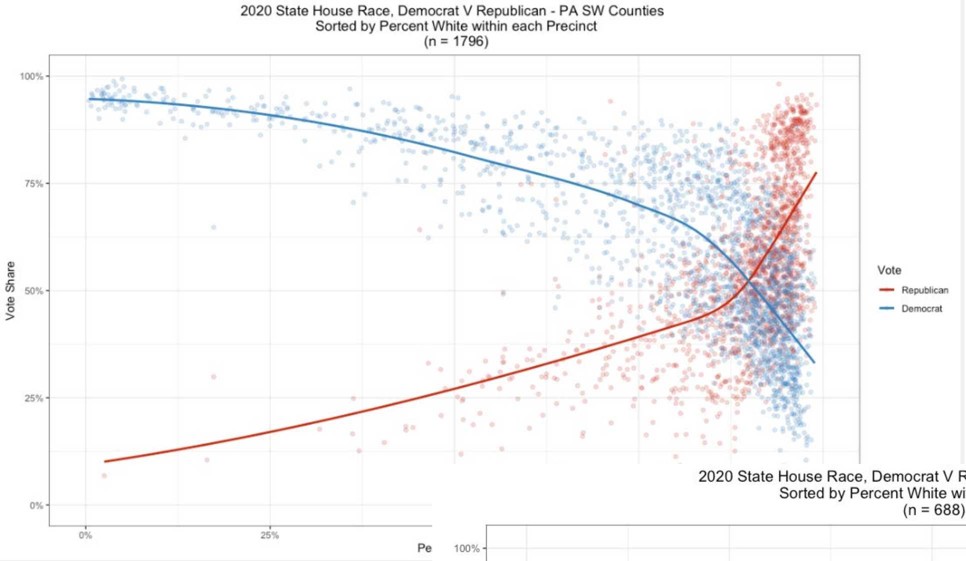
Analysis of n=3,178 precincts in Western PA

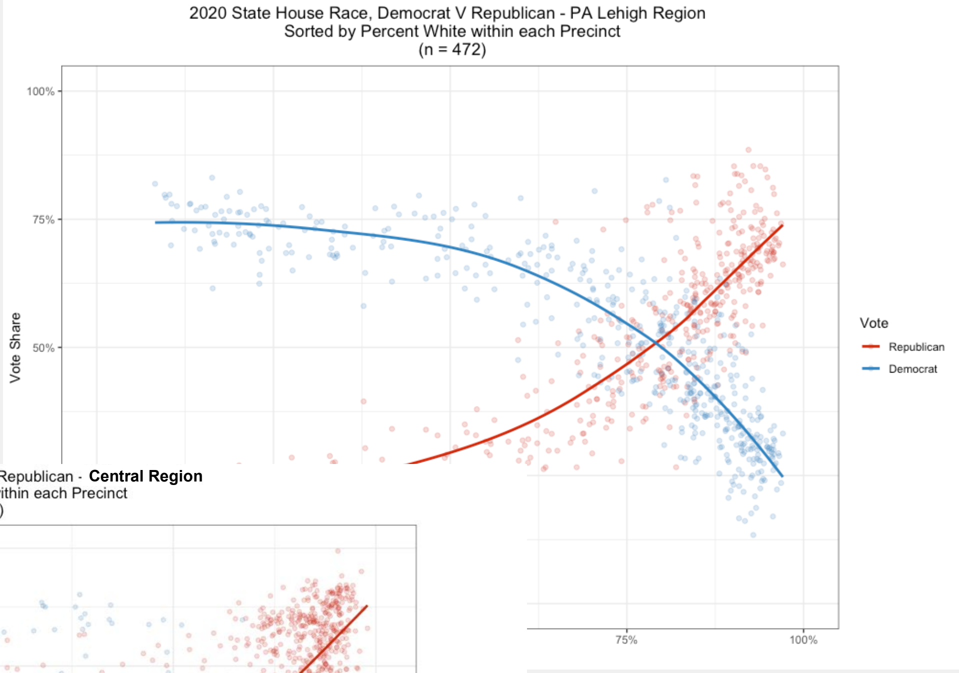
Measuring Racially Polarized Voting

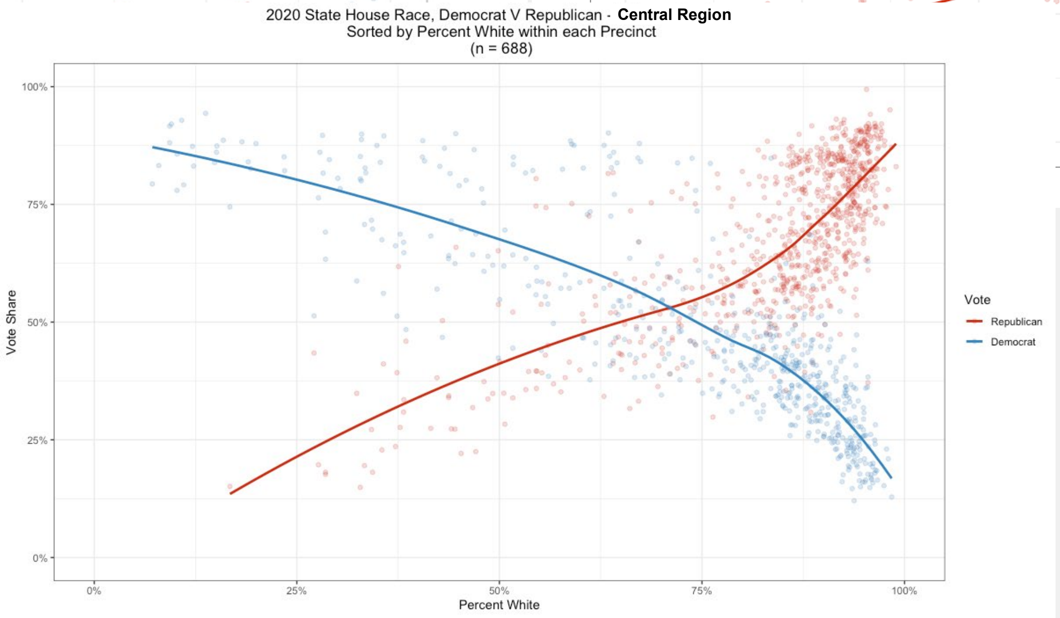
Patterns of Racially Polarized Voting in Pennsylvania 2020 Election

	Southwest		Central		Lehigh Valley		Southeast	
	State House (D)	Donald Trump (R)	State House (D)	Donald Trump (R)	State House (D)	Donald Trump (R)	State House (D)	Donald Trump (R)
White	23.9	75.8	17.9	84.4	21.0	73.9	42.4	52.2
Minority	93.4	5.8	88.4	15.1	75.3	22.5	88.9	5.9
Black	95.5	4.3	92.6	12.3	84.7	9.9	94.6	2.1
Latino	--	--	78.3	21.8	74.1	24.3	82.0	12.8

Ecological Inference statistical estimates by region by race


2020 State House Race, Democrat V Republican - PA SW Counties
Sorted by Percent White within each Precinct
(n = 1796)
Vote Share
100%
75%
50%
25%
0%
0%
25%
Pe
Vote
Republican
Democrat


2020 State House Race, Democrat V Republican - PA Lehigh Region
Sorted by Percent White within each Precinct
(n = 472)
Vote Share
100%
75%
50%
75%
100%
Vote
Republican
Democrat


2020 State House Race, Democrat V Republican - **Central Region**
Sorted by Percent White within each Precinct
(n = 688)
Vote Share
100%
75%
50%
25%
0%
0%
25%
50%
75%
100%
Percent White
Vote
Republican
Democrat

Summary of Voting Analysis

- Voting analysis is clear – there is a strong finding of racially polarized voting across the state as a whole
  - In pockets of the state, enough White voters cross-over to support the Minority group's "candidates of choice" in coalition to sustain additional Minority-performing districts

- Analysis of the current map

1. Multiple Black-performing and Latino-performing districts are packed and exhibit wasted Minority votes, which results in vote dilution

2. Given growth of the Minority population in certain regions of the state, it is clear that existing Minority districts should be unpacked and that new Minority-performing districts created to comply with the VRA

Performance Analysis

- Minority-performing districts in the preliminary plan will perform for minority voters

Dist	Current % MVAP	Prelim % MVAP	Expected performance for Minority Cand of Choice
19	42.0	48.2	80.9
24	55.3	51.0	89.2
34	29.5	40.8	79.9
35	26.7	26.5	62.9
54	4.2	43.0	69.5
189	28.3	35.9	58.4

Performance Analysis

- Minority-performing districts in the preliminary plan will perform for minority voters

Dist	Current % MVAP	Prelim % MVAP	Expected performance for Minority Cand of Choice
22	71.0	61.6	71.2
50	5.7	48.1	65.9
116	30.4	40.5	44.2
126	47.4	42.4	55.4
127	75.6	61.3	68.8
129	14.9	45.4	58.9
134	13.1	48.9	61.9

THANK YOU

Dr. Matt Barreto, UCLA Political Science & Chicana/o Studies
Faculty Director of the UCLA Voting Rights Project

matt@uclavrp.org 909.489.2955

An Evaluation of the Partisan Fairness of the Pennsylvania Legislative Reapportionment Commission's Proposed State House Districting Plan

Christopher Warshaw*

January 7, 2022

*Associate Professor, Department of Political Science, George Washington University. warshaw@gwu.edu. Note that the analyses and views in this report are my own, and do not represent the views of George Washington University.

Contents

1 Introduction

My name is Christopher Warshaw. I am an Associate Professor of Political Science at George Washington University. Previously, I was an Associate Professor at the Massachusetts Institute of Technology from July 2016 - July 2017, and an Assistant Professor at MIT from July 2012 - July 2016.

I have been asked by counsel representing the House Democratic Caucus to analyze relevant data and provide my expert opinions to the Legislative Reapportionment Commission (LRC) about its proposed State House districting plan. I look forward to making a presentation to the LRC on January 14th.

2 Qualifications and Publications

My Ph.D. is in Political Science, from Stanford University, where my graduate training included courses in political science and statistics. I also have a J.D. from Stanford Law School. My academic research focuses on public opinion, representation, elections, and polarization in American Politics. I have written over 20 peer reviewed papers on these topics. Moreover, I have written multiple papers that focus on elections and two articles that focus specifically on partisan gerrymandering. I also have a forthcoming book that includes an extensive analysis on the causes and consequences of partisan gerrymandering in state governments.

My curriculum vitae is attached to this report. All publications that I have authored and published appear in my curriculum vitae. My work is published or forthcoming in peer-reviewed journals such as: the *American Political Science Review*, the *American Journal of Political Science*, the *Journal of Politics*, *Political Analysis*, *Political Science Research and Methods*, the *British Journal of Political Science*, the *Annual Review of Political Science*, *Political Behavior*, *Legislative Studies Quarterly*, *Science Advances*, the *Election Law Journal*, *Nature Energy*, *Public Choice*, and edited volumes from Cambridge University Press and Oxford University Press. My book entitled *Dynamic Democracy in the American States* is forthcoming from the University of Chicago Press. My non-academic writing has been published in the *New York Times* and the *Washington Post*. My work has also been discussed in the *Economist* and many other prominent media outlets.

My opinions in this case are based on the knowledge I have amassed over my education, training and experience, including a detailed review of the relevant academic literature. They also follow from statistical analysis of the following data:

- In order to calculate partisan bias in state house elections on the proposed plan in Pennsylvania, I examined:
  - GIS Files with the 2014-2020 Pennsylvania State House plan and the proposed 2022-30 plan): I obtained both plans from the Legislative Reapportionment Commission's website.
  - Precinct-level data on recent statewide Pennsylvania elections: I use precinct-level data on Pennsylvania's statewide elections between 2016-20 from the Voting and Election Science Team (University of Florida, Wichita State University). I obtained these data from the Harvard Dataverse.[1] I obtained precinct-level data on elections from 2012-14 from the MGGG Redistricting Lab.[2] Finally, I obtained data on state legislative election results from the House Democratic Caucus since they were not available from public sources.
  - A large canonical data set on candidacies and results in state legislative elections: I obtained results from 1972-2020 collected by Carl Klarner and a large team of collaborators. The results from 1972-2012 are based on data maintained by the Inter-university Consortium for Political and Social Research (ICPSR) (Klarner et al. 2013). The data from 2013-2020 were collected by Klarner.
  - Data on presidential election returns in state legislative districts: For elections between 1972 and 1991, I used data on county-level presidential election returns from 1972-1988 collected by the Inter-university Consortium for Political and Social Research (ICPSR 2006) and mapped these returns to state legislative districts. For elections between 1992 and 2001, I used data on presidential election returns in the 2000 election collected by McDonald (2014) and Wright et al. (2009). For elections between 2002 and 2011, I used data on the 2004 and 2008 presidential elections collected by Rogers (2017). For elections between 2012 and 2020, I used data on presidential election returns from the DailyKos website and PlanScore.org.
  - The Plan Score website: PlanScore is a project of the nonpartisan Campaign Legal Center (CLC) that enables people to score proposed maps for their partisan, demographic, racial, and geometric features. I am on the social science advisory team for PlanScore.
- In order to compare the maps in Pennsylvania to congressional elections, I examined:

1. See `https://dataverse.harvard.edu/dataverse/electionscience`.
2. See `https://github.com/mggg-states/PA-shapefiles`.

- A large data set on candidacies and results in Congressional elections: I obtained results from 1972-2018 collected by the Constituency-Level Elections Archive (CLEA) (Kollman et al. 2017). The results from 1972-1990 are based on data collected and maintained by the Inter-university Consortium for Political and Social Research (ICPSR) and adjusted by CLEA. The data from 1992-2018 are based on data collected by CLEA from the Office of the Clerk at the House of the Representatives. I supplemented this dataset with recent election results collected by the MIT Election and Data Science Lab (MIT Election and Data Science Lab 2017) and Dave Leip's Atlas of U.S. Presidential Elections.
- Data on presidential election returns and incumbency status in Congressional elections. I used data on elections in congressional districts from 1972-2020 collected by Professor Gary Jacobson (University of California, San Diego). This dataset has been used in many Political Science studies and has canonical status in the political science profession (Jacobson 2015).

I have previously provided expert reports in six redistricting-related cases:

- Between 2017 and 2019, I provided reports for *League of Women Voters of Pennsylvania v. Commonwealth of Pennsylvania*, No. 159 MM 2017, *League of Women Voters of Michigan* v. *Johnson*, 17-14148 (E.D. Mich), and *APRI et al.* v. *Smith et al.*, No. 18-cv-357 (S.D. Ohio). My testimony was found to be credible in each of these cases and was extensively cited by the judges in their decisions.
- In the current redistricting cycle, I have provided reports in *League of Women Voters* v. *Ohio Redistricting Commission*, No. 2021-1193, *League of Women Voters* vs. *Kent County Apportionment Commission*, and *League of Women Voters of Ohio* v. *Ohio Redistricting Commission*, No. 2021-1449.

In addition, I have provided expert testimony and reports in several cases related to the U.S. Census: *State of New York et al.* v. *United States Department of Commerce*, 18-cv-2921 (S.D.N.Y.), *New York* v. *Trump*; *Common Cause* v. *Trump*, 20-cv-2023 (D.D.C.), and *La Union Del Pueblo Entero (LUPE)* v. *Trump*, 19-2710 (D. Md.).

The opinions in this report are my own, and do not represent the views of George Washington University.

3 Summary

The relationship between the distribution of partisan support in the electorate and the partisan composition of the government—what Powell (2004) calls "vote–seat representation"—is a critical link in the longer representational chain between citizens' preferences and governments' policies. If the relationship between votes and seats systematically advantages one party over another, then some citizens will enjoy more influence—more "voice"—over elections and political outcomes than others (Caughey, Tausanovitch, and Warshaw 2017).

I use three complementary methodologies to project future election results in order to evaluate the partisan fairness of Pennsylvania's proposed State House plan. First, I use a composite of previous statewide election results between 2014-2020 to analyze the new map.[3] Second, I analyze the results of the 2020 State House election on the newly proposed map. Third, I complement this approach using the open source PlanScore.org website, which is a project of the Campaign Legal Center.[4] PlanScore uses a statistical model to estimate district-level vote shares for a new map based on the relationship between presidential election results and legislative results between 2014-2020.[5] Based on these three approaches, I characterize the bias in Pennsylvania's plans based on a large set of established metrics of partisan fairness and place the bias in Pennsylvania's plans into historical perspective. I also analyze whether the proposed plan is responsive to shifts in voters' preferences.

All of these analyses indicate that the proposed map is fair with just a small pro-Republican bias. Indeed, one important feature of the proposed plan is that it enables the party that wins the majority of the votes to nearly always win the majority of the seats. In the actual 2020 State House election, Republicans received 50.5% of the two-party vote and Republicans would win 50.2% of the seats in the proposed plan.[6] In the 2020 presidential election, Democrat Joe Biden received about 50.6% of the two-party vote and he would have won 102 out of the 203 (50.2%) of the State House districts.[7] Based on the statewide elections in Pennsylvania between 2014-2020, the Democrats' statewide two-party vote share averaged about 54% of the vote and they would win nearly exactly

3. These include the following elections: 2016 Presidential, 2020 Presidential, 2014 Governor, 2018 Governor, 2016 Attorney General, 2020 Attorney General, 2016 Senate, 2018 Senate, 2016 Treasurer, 2020 Treasurer, 2016 Auditor, and 2020 Auditor election.

4. I am on the social science advisory board of Plan Score, but do not have any role in PlanScore's evaluation of individual maps.

5. See `https://planscore.campaignlegal.org/models/data/2021D/` for more details.

6. I impute uncontested State House elections using the presidential election results.

7. Following standard convention, throughout my analysis I focus on two-party vote shares.

the same proportion of the seats on the proposed plan (54.5%).[8] Historically, there is a winner's bonus where the party that wins 54% of the votes typically receives about 58% of the seats. So recent statewide elections indicate a modest pro-Republican bias in the plan using a wide variety of Political Science metrics for partisan fairness.

I also reach the conclusion that the plan is relatively neutral, with a small pro-Republican bias, using the predictive model on the PlanScore website. PlanScore projects that Republicans would get about 50.3% of the statewide vote, but Republicans are expected to win 53% of the seats in Pennsylvania's proposed State House plan (and Democrats would win 47% of the seats).[9] Across 1000 simulations, PlanScore indicates that the proposed plan favors Republican candidates in 95% of scenarios. Based on generally accepted Political Science metrics for partisan fairness, PlanScore indicates that Pennsylvania's proposed plan would have a modest level of pro-Republican bias.

The remainder of the report proceeds as follows. First, I discuss partisan gerrymandering and how social scientists measure partisan bias in a districting plan. I also discuss how to conceptualize the responsiveness of a districting plan to shifts in voters' preferences. Next, I examine the partisan fairness of the proposed State House plan, and compare it to the fairness of other plans around the country over the past 50 years. Then, I examine the responsiveness of the proposed plan to shifts in voters' preferences and the number of competitive districts in the proposed plans. Finally, I briefly conclude.

4 Background on Partisan Fairness

This section provides background about how social scientists conceptualize partisan fairness in a districting plan. Partisan advantage in a districting plan may arise either intentionally, due to a deliberate effort to benefit the line-drawing party and handicap the opposing party via gerrymandering (Kang 2017; Levitt 2017), or unintentionally as a result of factors such as political geography, candidate appeal, and electoral swings (Chen and Rodden 2013; Goedert 2014; Seabrook 2017). Whether districting bias is purposeful or accidental, it means that one party's voters are more "cracked" and "packed" than the other side's supporters. In cracked districts, voters' preferred candidates lose by relatively narrow margins; in packed districts, their candidates of choice win by enormous margins

8. I weight the composite scores to give each election cycle equal weight in the index. The seat-level projections are based on the 12 statewide elections where I have precinct-level data. If instead I use the approach that Professor Michael Barber references in his report and simply average across contests, Democrats win 52% of the votes and 52% of the seats on the proposed plan.

9. This is a probabilistic estimate based on 1000 simulations of possible elections using a model of the elections between 2014-2020.

(Stephanopoulos and McGhee 2015). Thanks to disproportionate cracking and packing, the disfavored party is less able than the favored party to convert its statewide support among voters into legislative representation. This gives the favored party the ability to shift policies in its direction (Caughey, Tausanovitch, and Warshaw 2017) and build a durable advantage in downstream elections (Stephanopoulos and Warshaw 2020). It can even lead to undemocratic outcomes where the advantaged party wins the majority of the seats and controls the government while only winning a minority of the votes.

There are a number of approaches that have been proposed to measure partisan advantage in a districting plan. These approaches focus on asymmetries in the efficiency of the vote–seat relationships of the two parties. In recent years, at least 10 different approaches have been proposed (McGhee 2017). While no measure is perfect, much of the recent literature has focused on four related approaches that I describe below.

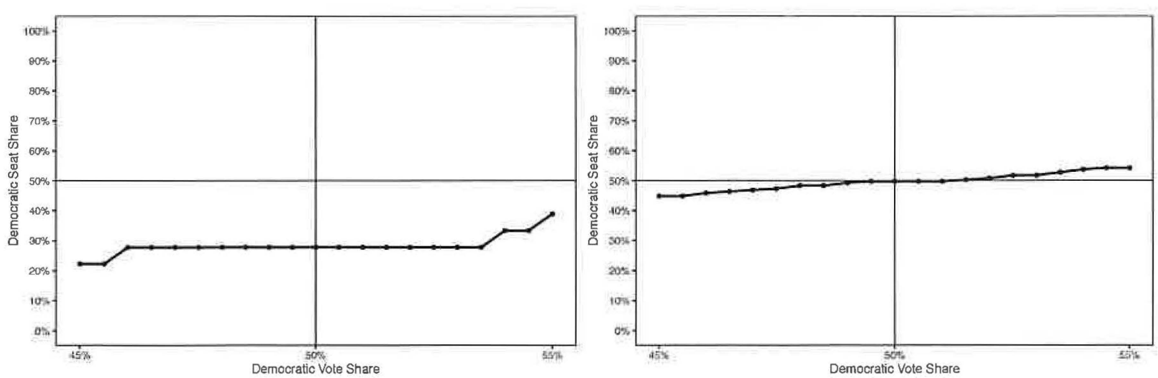
4.1 Symmetry in the Vote-Seat Curve Across Parties

Basic fairness suggests that in a two-party system each party should receive the same share of seats for identical shares of votes. The *symmetry* idea is easiest to understand at an aggregate vote share of 0.5—a party that receives half the vote ought to receive half the seats—but a similar logic can apply across the "seats-votes curve" that traces out how seat shares change as vote shares rise and fall. For example, if a party receives a vote share of 0.57 and a seat share of 0.64, the opposing party should also expect to receive a seat share of 0.64 if it were to receive a vote share of 0.57. An unbiased system means that for V share of the votes a party should receive S share of the seats, and this should be true for all parties and vote percentages (Niemi and Deegan 1978; Gelman and King 1994; McGhee 2014; Katz, King, and Rosenblatt 2020).

Gelman and King (1994, 536) propose two ways to measure partisan bias in the symmetry of the vote-seat curve. First, it can be measured using counter-factual election results in a range of statewide vote shares between .45 and .55. Across this range of vote shares, each party should receive the same number of seats. Symmetry captures any departures from the standard that each party should receive the same seat share across this range of plausible vote shares. For example, if partisan bias is -0.05, this means that the Democrats receive 5% fewer seats in the legislature than they should under the symmetry standard (and the Republicans receive 5% more seats than they should). Second, symmetry can be measured based on the seat share that each party receives when they split the statewide vote 50-50. In an unbiased system, each party should receive 50% of the seats in a tied statewide election. Here, the partisan bias statistic is the "expected proportion of the seats over 0.5 that the Democrats receive when they receive exactly half the average district vote."

To illustrate the symmetry metric, Figure 1 shows what each party's share of the seats would have been across a range of statewide vote shares from 45%-55%. The left-hand panel shows the gerrymandered 2016 US House election. On this plan, Democrats received 22% of the seats when they received 45% of the statewide vote, 28% of the seats when they won half the vote, and just 33% of the seats when they received 55% of the statewide vote. In contrast, Republicans received 66% of the seats when they received 45% of the vote, 72% of the seats when they won half the vote, and 78% of the seats when they received 55% of the vote. This indicates a historically extreme pro-Republican symmetry bias of about -20%.

The right-hand panel of Figure 1 shows the proposed State House plan (using re-aggregated votes in the 2020 State House Elections). On this plan, Democrats would receive about 45% of the seats when they receive 45% of the votes, 49.8% of the seats when they win half the vote, and 54% of the seats when they receive 55% of the votes. Republicans would receive about 46% of the seats when they receive 45% of the votes, 50.2% of the seats when they win half the vote, and 55% of the seats when they receive 55% of the votes. This indicates an almost perfectly fair plan using the symmetry metric with virtually no bias.



(a) Extremely biased symmetry on 2016 US House Map

(b) Relatively neutral symmetry on Proposed State House Map

Figure 1: Plot illustrating an extremely asymmetrical map based on 2016 US House election and the more symmetrical proposed plan using re-aggregated votes in 2020 Pennsylvania State House Elections

A weakness of the symmetry approach is that it requires the analyst to calculate counterfactual elections. This approach has both conceptual and empirical limitations. At a conceptual level, it is not clear that it aligns perfectly with the usual definition of a gerrymander. Indeed, "when observers assert that a district plan is a gerrymander, they usually mean that it systematically benefits a party (and harms its opponent) in actual elections. They do not mean that a plan would advantage a party in the hypothetical event of a tied election, or if the parties' vote shares flipped" (Stephanopoulos and McGhee 2015, 857). At an empirical level, in order to generate symmetry metrics, we need to simulate counter-factual elections by shifting the actual vote share in each district a uniform amount (McGhee 2014).[10] In general, this uniform swing assumption seems reasonable based on past election results (though is probably less reasonable in less competitive states). Moreover, it has been widely used in past studies of redistricting. But there is no way to conclusively validate the uniform swing assumption for any particular election.

An important strength, however, of the symmetry approach is that it is based on the shape of the seats-votes curve and not any particular point on it. As a result, it is relatively immune to shifts in party performance (McGhee 2014). For instance, the bias toward Republicans in Pennsylvania's State House elections was very similar in 2014-2020. Moreover, the symmetry approach has been very widely used in previous studies of gerrymandering and redistricting (Gelman and King 1994; McGhee 2014). Overall, the symmetry approach is useful for assessing partisan advantage in the districting process.

4.2 Mean-median Gap

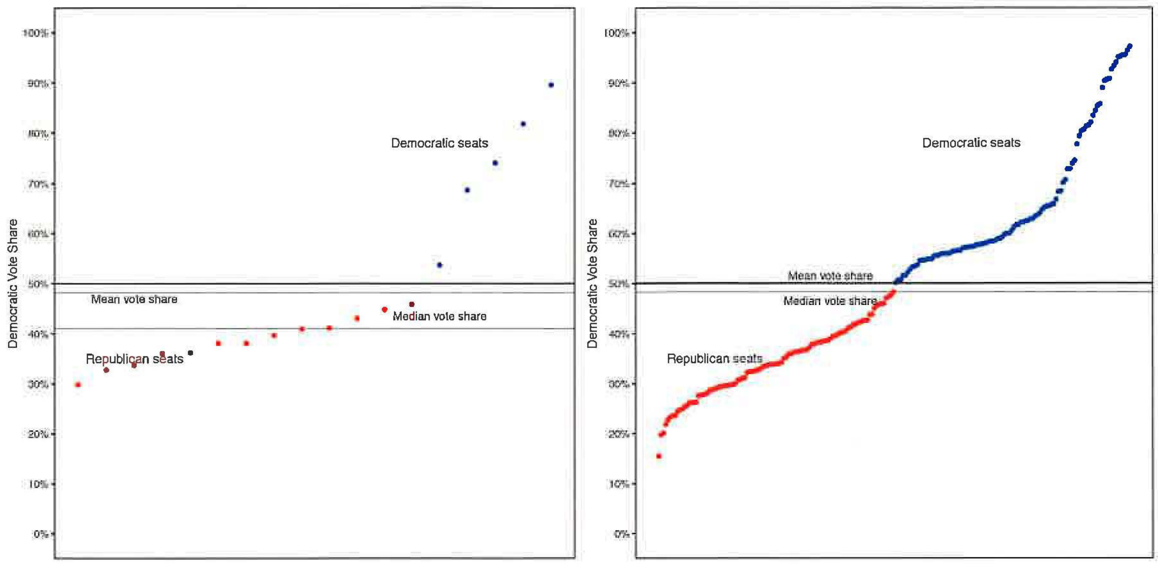
Another metric that some scholars have proposed to measure partisan bias in a districting plan is the *mean-median gap*: the difference between a party's vote share in the median district and their average vote share across all districts. If the party wins more votes in the median district than in the average district, they have an advantage in the translation of votes to seats (Krasno et al. 2018; Best et al. 2017; Wang 2016). In statistics, comparing a dataset's mean and median is a common statistical analysis used to assess skews in the data and detect asymmetries (Brennan Center 2017).

The mean-median difference is very easy to apply (Wang 2016). It is possible, however, for packing and cracking to occur without any change in the mean-median difference (Buzas and Warrington 2021). That is, a party could gain seats in the legislature without the mean-median gap changing (McGhee 2017).[11] It is also sensitive to the outcome in

10. In principle, the uniform swing election could be relaxed, and swings could be estimated on a district-by-district basis. But this is rarely done in practice since it would require a much more complicated statistical model, and probably would not improve estimates of symmetry very much.

11. As McGhee (2017), notes, "If the median equals the win/loss threshold–i.e., a vote share of 0.5–then when a seat changes hands, the median will also change and the median- mean difference will reflect that change. But if the median is anything other than 0.5, seats can change hands without any change in the median and so without any change in the median-mean difference." See also Buzas and Warrington

the median district (Warrington 2018b). In addition, the mean-median difference lacks a straightforward interpretation in terms of the number of seats that a party gains through gerrymandering.



(a) Extremely biased mean-median difference on 2016 US House Map

(b) Relatively neutral mean-median diff. on Proposed State House Map

Figure 2: Plot illustrating an extremely biased mean-median difference based on 2016 US House election and a more neutral mean-median difference on the proposed plan using re-aggregated votes in 2020 Pennsylvania State House Elections

Figure 2 illustrates the mean-median difference. The left-hand panel shows the 2016 US House elections. In this election, the mean-median difference was about -7.5%. This means that Republicans did about 7.5% better in the median seat than statewide, which gave them a large advantage in the translation of votes to seats. The right-hand panel shows the proposed State House plan (using re-aggregated votes in the 2020 State House Elections). Across all districts, Democrats won an average of 50.3% of the vote. But they only won 48.3% in the median district. So the mean-median difference here was -1.9%. It still favors Republicans, but much less than on the heavily gerrymandered 2012-16 US House plan.

(2021) who make a similar point using simulated packing and cracking.

4.3 Efficiency Gap

Both cracked and packed districts "waste" more votes of the disadvantaged party than of the advantaged one (McGhee 2014; Stephanopoulos and McGhee 2015).[12] This suggests that gerrymandering can be measured based on asymmetries in the number of wasted votes for each party. The *efficiency gap* (EG) focuses squarely on the number of each party's wasted votes in each election. It is defined as "the difference between the parties' respective wasted votes, divided by the total number of votes cast in the election" (Stephanopoulos and McGhee 2015, 831; see also McGhee 2014, 2017). All of the losing party's votes are wasted if they lose the election. When a party wins an election, the wasted votes are those above the 50%+1 needed to win.

If we adopt the convention that positive values of the efficiency gap imply a Democratic advantage in the districting process and negative ones imply a Republican advantage, the efficiency gap can be written mathematically as:

$$EG = \frac{W_R}{n} - \frac{W_D}{n} \tag{1}$$

where W_R are wasted votes for Republicans, W_D are wasted votes for Democrats, and n is the total number of votes in each state.

Table 1 provides a simple example about how to calculate the efficiency gap with three districts where the same number of people vote in each district. In this example, Democrats win a majority of the statewide vote, but they only win 1/3 seats. In the first district, they win the district with 75/100 votes. This means that they only wasted the 24 votes that were unnecessary to win a majority of the vote in this district. But they lose the other two districts and thus waste all 40 of their votes in those districts. In all, they waste 104 votes. Republicans, on the other hand, waste all 25 of their votes in the first district. But they only waste the 9 votes unnecessary to win a majority in the two districts they win. In all, they only waste 43 votes. This implies a pro-Republican efficiency gap of $\frac{43}{300}$ - $\frac{104}{300}$= -20%.

In order to account for unequal population or turnout across districts, the efficiency gap formula in equation 1 can be rewritten as:

$$EG = S_D^{margin} - 2 * V_D^{margin} \tag{2}$$

where S_D^{margin} is the Democratic Party's seat margin (the seat share minus 0.5) and V_D^{margin}

12. The authors of the efficiency gap use the term "waste" or "wasted" to describe votes for the losing party and votes for the winning party in excess of what is needed to win an election. Since the term is used by the efficiency gap authors, I use it here when discussing the efficiency gap.

Table 1: Illustrative Example of Efficiency Gap

District	Democratic Votes	Republican Votes
1	75	25
2	40	60
3	40	60
Total	155 (52%)	145 (48%)
Wasted	104	43

is is the Democratic Party's vote margin. V_D^{margin} is calculated by aggregating the raw votes for Democratic candidates across all districts, dividing by the total raw vote cast across all districts, and subtracting 0.5 (McGhee 2017, 11-12). In the example above, this equation also provides an efficiency gap of -20% in favor of Republicans. But it could lead to a slightly different estimate of the efficiency gap if districts are malapportioned or there is unequal turnout across districts.[13]

In the case of Pennsylvania's proposed State House map, equation 2 implies there would have been a pro-Democratic efficiency gap of 0.7% using the votes from the 2020 election re-aggregated onto the proposed plan. This is very close to the middle of the distribution of previous Efficiency Gaps in state legislative elections.

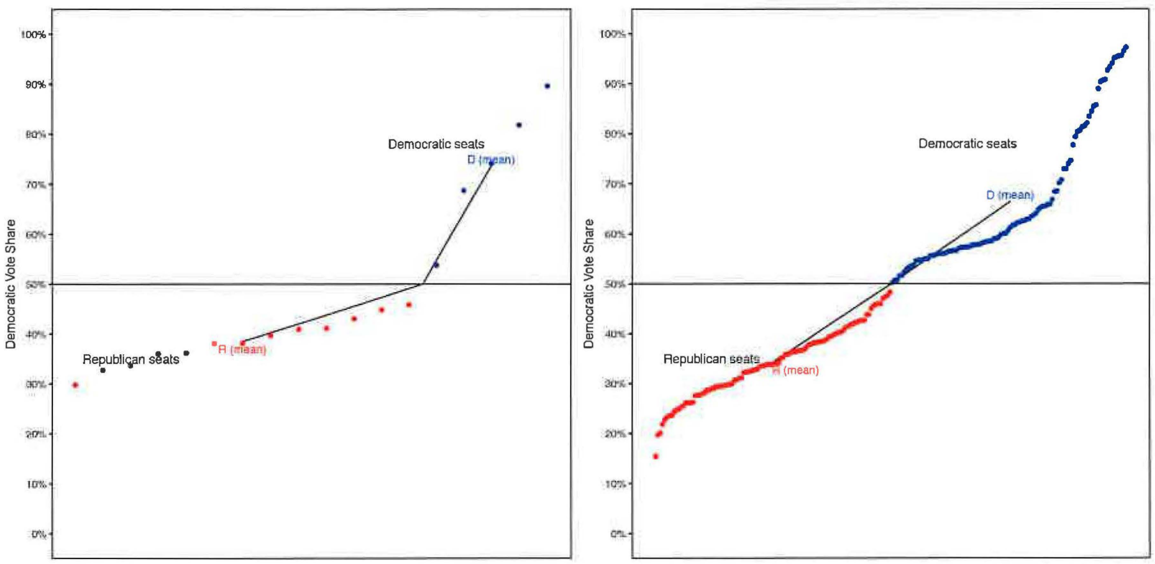
The efficiency gap mathematically captures the packing and cracking that are at the heart of partisan gerrymanders (Buzas and Warrington 2021). It measures the extra seats one party wins over and above what would be expected if neither party were advantaged in the translation of votes to seats (i.e., if they had the same number of wasted votes). A key advantage of the efficiency gap over other measures of partisan bias is that it can be calculated directly from observed election returns even when the parties' statewide vote shares are not equal.

The symmetry metric is closely related to the efficiency gap. In the special case where each party receives half of the statewide vote, the symmetry and the efficiency gap metrics are mathematically identical (Stephanopoulos and McGhee 2015, 856). More generally, the symmetry and efficiency gap yield very similar substantive results when each party's statewide vote share is close to 50% (as is the case in Pennsylvania). When elections are uncompetitive, however, and one party wins a large percentage of the statewide vote, the efficiency gap and these symmetry metrics are less correlated with one another (857).

13. In general, the two formulations of the efficiency gap formula yield very similar results. Because Democrats tend to win lower-turnout districts, however, the turnout adjusted version of the efficiency gap in equation 2 tends to produce results that suggest about a 2% smaller disadvantage for Democrats than the version in Equation 1 (see McGhee 2018).

4.4 Declination

Another measure of asymmetries in redistricting plans is called *declination* (Warrington 2018b, 2018a). The declination metric treats asymmetry in the vote distribution as indicative of partisan bias in a districting plan (Warrington 2018a). If all the districts in a plan are lined up from the least Democratic to the most Democratic, the mid-point of the line formed by one party's seats should be about as far from the 50 percent threshold for victory on average as the other party's (McGhee 2018).



(a) Extremely biased declination on 2016 US House Map

(b) Relatively neutral declination on Proposed State House Map

Figure 3: Plot illustrating an extremely biased declination based on 2016 US House election and a fair declination on proposed plan using re-aggregated votes in 2020 Pennsylvania State House Elections

Declination suggests that when there is no gerrymandering, the angles of the lines (θ_D and θ_R) between the mean across all districts and the point on the 50% line between the mass of points representing each party will be roughly equal. When they deviate from each other, the smaller angle (θ_R in the case of Pennsylvania) will generally identify the favored party. To capture this idea, declination takes the difference between those two angles (θ_D and θ_R) and divides by $\pi/2$ to convert the result from radians to fractions of 90 degrees.[14] This produces a number between -1 and 1. As calculated here, positive values favor Democrats and negative values favor Republicans. Warrington (2018b) suggests a further adjustment to account for differences in the number of seats across legislative

14. This equation is: $\delta = 2^* \ (\theta_R \text{ - } \theta_D) \ / \ \pi$.

chambers. I use this adjusted declination estimate in the analysis that follows.[15]

Figure 3 illustrates the declination metric. The left-hand panel shows the 2016 US House elections, which was an historically extreme pro-Republican gerrymander. Here, it is easy to see that the angle of the line between the x-axis and the average Republican seat is much less steep than the line between the x-axis and the average Democratic seat. The right-hand panel shows the proposed State House plan (using re-aggregated votes in the 2020 State House Elections). In this plot, the slope of the lines to the Democratic and Republican seats are nearly equal. Thus, the declination metric indicates that the plan has a nearly perfectly neutral declination of -.04.

4.5 Comparison of Partisan Bias Measures

All of the measures of partisan advantage discussed in the previous sections are closely related both theoretically and empirically (McGhee 2017; Stephanopoulos and McGhee 2018). Broadly speaking, all of the metrics consider how votes between the two parties are distributed across districts (Warrington 2018a). For example, the efficiency gap is mathematically equivalent to partisan bias in tied statewide elections (Stephanopoulos and McGhee 2018). Also, the median-mean difference is similar to the symmetry metric, since any perfectly symmetric seats-votes curve will also have the same mean and median (McGhee 2017).

Second, each of the concepts are closely related empirically, particularly in states with competitive elections. Figure 4 shows the correlation between each measure. The various measures have high correlations with one another.[16] Moreover, most of the variation in the metrics can be summarized on a single latent dimension (Stephanopoulos and McGhee 2018; Stephanopoulos and Warshaw 2020). So, overall, while there may be occasional cases where the metrics disagree about the amount of bias in a particular plan, the various metrics usually yield similar results for the degree of partisan bias in a districting plan (Nagle 2015). Where none of the metrics is an outlier and they all point in the same direction, we can draw a particularly robust conclusion

15. This adjustment uses this equation: $\hat{\delta} = \delta * \ln(\text{seats}) / 2$

16. While each measure is highly correlated with one another, the efficiency gap and declination measures are particularly closed related and the symmetry and mean-median measures are very closely related. This could be because the efficiency gap and the declination consider the seats actually won by each party, while the symmetry metric and the mean-median difference do not (Stephanopoulos and McGhee 2018, 1557). In addition, the efficiency gap and the declination appear to best capture the packing and cracking that characterize partisan gerrymandering (Buzas and Warrington 2021).

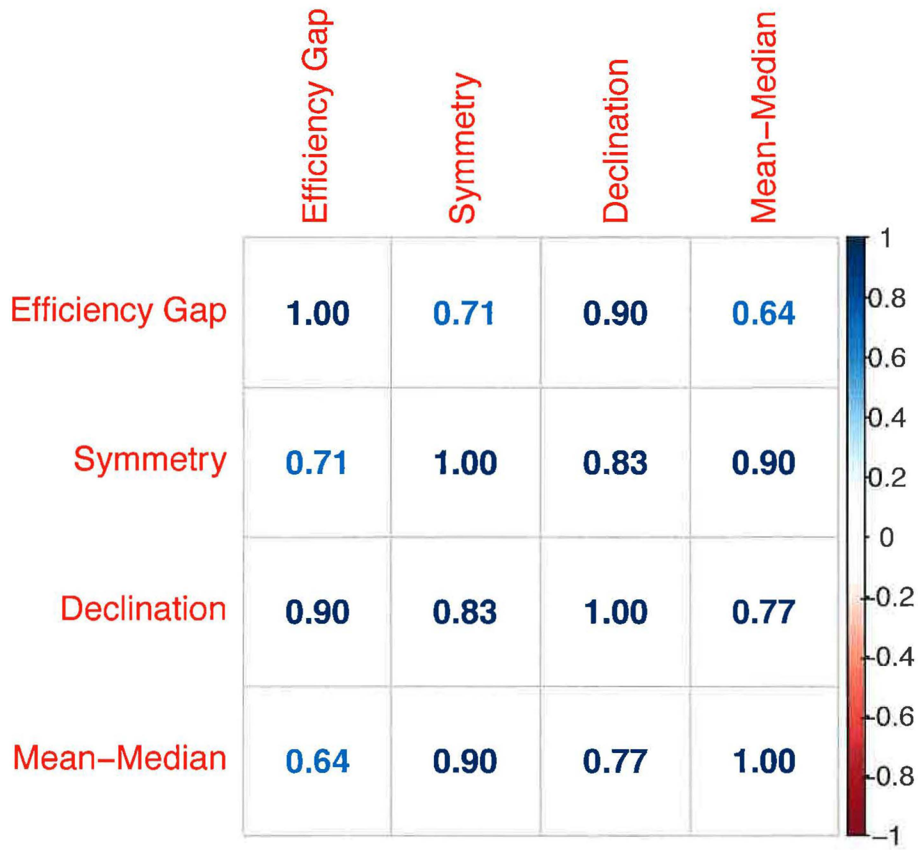


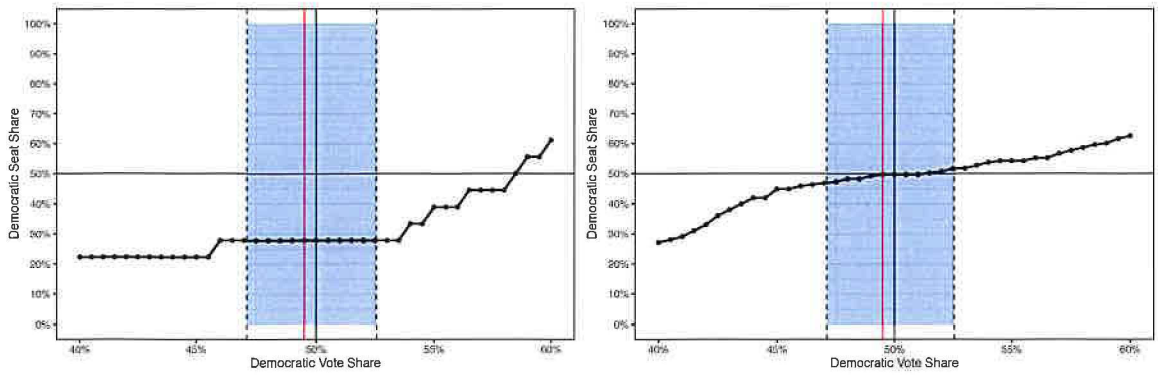
Figure 4: Correlation between measures of partisan bias in competitive states.

4.6 Responsiveness and Competitive Elections

Another benchmark for a districting plan is the percentage of districts likely to have competitive elections under that plan and the responsiveness of the plan to changes in voters' preferences (Cox and Katz 1999). An unresponsive map ensures that the bias in a districting plan toward the advantaged party is insulated against changes in voters' preferences, and thus is durable across multiple election cycles.

To illustrate the concept of responsiveness, Figure 5 shows the vote-seat curve in Pennsylvania for the 2016 US House plan and the proposed State House plan. Similarly to the figure illustrating the symmetry metric, these plots are generated by applying uniform swings to the actual election results.[17] Specifically, I apply a uniform swing in the actual election results until I achieve an average Democratic vote share of 40%. Then

17. The layout of this chart is adapted from charts in Royden, Li, and Rudensky (2018).



(a) Extremely unresponsive 2016 US House Map

(b) Relatively Responsive Proposed State House Map

Figure 5: Vote-seat curve in Pennsylvania using uniform swings in 2020 election results re-aggregated using proposed plan. The shaded area shows the range between the minimum and maximum Democratic statewide vote share in State House elections from 2014-2020. The red line shows the actual Democratic statewide vote share in the 2020 State House elections.

I steadily increase the average Democratic vote share until it reaches 60%. The shaded area shows the range between the minimum and maximum Democratic statewide vote share in State House elections from 2014-2020. The red line shows the actual Democratic statewide vote share in the 2020 State House elections.

The left panel of Figure 5 indicates that Republicans win two thirds or more of the US House seats across all of the range of actual election swings over the past decade. In contrast, the proposed State House plan is relatively responsive to changes in statewide preferences. The Democratic seat share increases by 5 percentage points across the range of actual election results and about 10 percentage points as their statewide vote share goes from 45 to 55 percentage points.

An important factor that affects the overall responsiveness of a plan is the number of competitive districts in a plan. First, this affects the responsiveness of a map as the two parties' statewide vote shares rise and fall. A plan with more competitive elections is likely to be more responsive to changes in voters' preferences than a plan with fewer competitive elections (McGhee 2014). Second, uncompetitive districts tend to protect incumbents from electoral sanctions (Tufte 1973; Gelman and King 1994). This could harm political representation by making legislators less responsive and accountable to their constituents' preferences.

There are a couple of approaches we might use to evaluate whether individual districts on a plan are likely to have competitive elections. We could measure whether a district was competitive in an election based on whether the winning party received less than 55%

of the two-party vote (Fraga and Hersh 2018; Jacobson and Carson 2015, 91).[18] While this definition is sometimes used in the literature, though, it is not clear that a sharp threshold at 55% is the best measure of competitiveness.

Another possible definition of competitiveness might be whether a district is likely to switch parties at least once per decade (Henderson, Hamel, and Goldzimer 2018). This definition is more empirically robust because it is not dependent on any particular electoral threshold for competitiveness. Indeed, in a state with swing voters where the two parties' statewide shares vary substantially over the course of the decade, a district where the winning party normally wins 56% of the vote could be competitive. In another state with few swing voters and very inelastic election results, a district where the winning party normally wins 53% of the vote might not even be competitive.

5 Partisan Fairness of Pennsylvania's proposed State House Map

In this section, I will provide a comprehensive evaluation of the partisan fairness of Pennsylvania's proposed State House districting plan (see Figure 6 for a map of the proposed plan). In order to evaluate the proposed plan, we need to predict future election results on this map. Unfortunately, there is no way to know, with certainty, the results of future elections. Thus, I use three complementary methodologies to predict future State House elections in Pennsylvania and generate the various metrics I discussed earlier.

5.1 Composite of previous statewide elections

First, I use a composite of previous statewide election results between 2014-2020 re-aggregated to the proposed map.[19] For each year, I estimate each party's vote share, seat share, and the average of the partisan bias metrics across races. I then average them together to produce a composite result. This approach implicitly assumes that future voting patterns will look like the average of these recent statewide elections.

When I average across these statewide elections from 2014-2020, Democrats win 54% of the votes and 54% of the seats on the proposed plan (see Table 2).[20] Thus, the plan

18. Fraga and Hersh (2018) justify this definition based on the fact that the Cook Political Report's "median 'leaning' race ended up with a vote margin of 10 percentage points (a 55%–45% race)."

19. These include the following elections: 2016 Presidential, 2020 Presidential, 2014 Governor, 2018 Governor, 2016 Attorney General, 2020 Attorney General, 2016 Senate, 2018 Senate, 2016 Treasurer, 2020 Treasurer, 2016 Auditor, and 2020 Auditor election.

20. I weight the composite scores to give each election cycle equal weight in the index. The seat-level

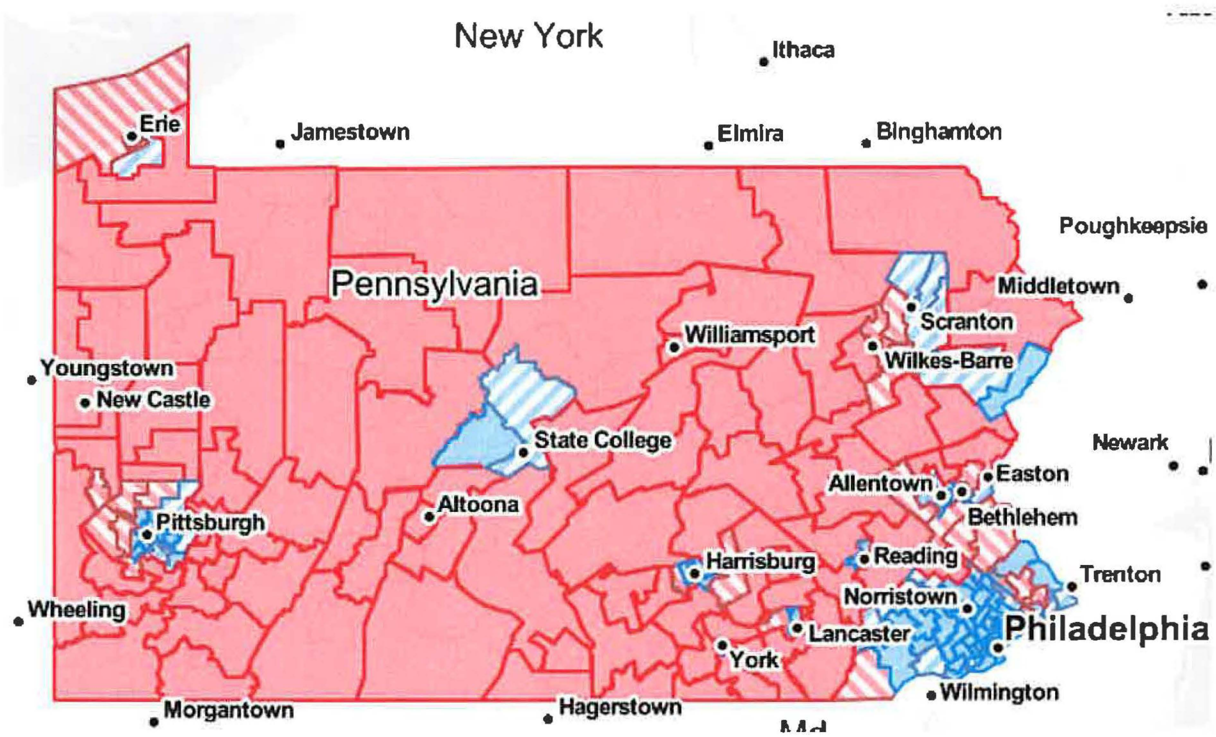


Figure 6: Map of proposed State House Districts from PlanScore.org

Metric	Value	2014-2020 Composite > Biased than this % Elections	> Pro-Rep. than this % Elections
2014-2020 Plan			
Symmetry Bias	-7.7%	77%	85%
Mean-Median	-3.8%	70%	81%
Efficiency Gap	-5.8%	60%	83%
Declination	-.348	66%	82%
Average		68%	83%
Proposed Plan			
Symmetry Bias	-2.5%	29%	61%
Mean-Median	-1.4%	31%	63%
Efficiency Gap	-2.6%	27%	69%
Declination	-.175	38%	65%
Average		31%	65%

Table 2: Composite bias metrics for proposed plan based on statewide elections

satisfies the principal that the party that wins a significant majority of the statewide vote should also win a majority of the seats. However, Democrats did unusually well in these recent statewide elections. In state legislative elections, the two parties typically get closer to 50% of the statewide vote. Thus, another important benchmark is to examine what happens if each party evenly splits the votes. Basic fairness suggests that when the two

projections are based on the 12 statewide elections where I have precinct-level data.

parties split the votes they should also split the seats. But the composite election index indicates that when Democrats win 50% of the votes on the proposed plan, they are likely to only win 47.5% of the seats. This leads to a pro-Republican bias on the symmetry metric of 2.5%.

The plan also has a small pro-Republican bias on the other metrics I evaluate. For instance, Republicans do about 1.4% better in the median district than in the mean district and Republicans have a 2.6% advantage in the Efficiency Gap. Overall, the plan has a larger pro-Republican bias in the translation of votes to seats than 65% of previous plans over the past 50 years.

5.2 2020 State House election results

Next, I use the 2020 precinct-level State House results on both the 2014-20 map and re-aggregated to the proposed map to estimate the various metrics. This approach implicitly assumes that future elections will look like the 2020 election.[21] These endogenous election are likely to be an excellent predictor of future voting patterns in State House elections. But it is important to keep in mind that they could be affected by the individual candidates in each race as well as a host of other factors that wouldn't look exactly the same in future elections.

Metric	Value	More Biased than this % Historical Elections	More Pro-Republican than this % Historical Elections
2014-2020 Plan			
Symmetry Bias	-5.7%	60%	77%
Mean-Median Diff	-4.3%	79%	86%
Efficiency Gap	-4.8%	49%	78%
Declination	-.36	68%	83%
Average		64%	81%
Proposed Plan			
Symmetry Bias	-0.2%	2%	49%
Mean-Median Diff	-1.9%	40%	68%
Efficiency Gap	0.7%	8%	51%
Declination	-.04	9%	50%
Average		15%	55%

Table 3: Partisan bias metrics for State House plan based on 2020 State House election results re-aggregated onto proposed map

21. As is commonly done in the academic literature, I impute uncontested State House elections using the presidential election results. In State House district 7, the Democratic candidate won even though former-President Trump won the majority of the vote. In this district, I adjust the presidential vote so that the Democratic vote share is 51% to ensure that the imputed results yield the correct number of Democratic and Republican seats.

The proposed plan is nearly perfectly unbiased based on the re-aggregated 2020 State House results. Republicans would win 50.5% of the votes and 50.2% of the seats on the proposed plan. Moreover, both parties would receive nearly half the seats when the statewide vote is exactly evenly split. Thus, the symmetry bias is only .2%, which is right in the center of the historical distribution of partisan symmetries. The proposed plan is also nearly perfectly neutral using the other metrics. Only the mean-median difference implies a significant Republican advantage in the translation of votes to seats. When we average across all four metrics, the plan is more extreme than 15% of prior plans, and thus more neutral than 85% of prior plans. When I average across the various metrics, it just has a very small pro-Republican advantage: it is more pro-Republican than 55% of previous plans.

5.3 PlanScore

Third, I evaluate the proposed plan using a predictive model from the PlanScore.org website.[22] PlanScore uses a statistical model of the relationship between districts' latent partisanship and legislative election outcomes. This enables it to estimate district-level vote shares for a new map and the corresponding partisan gerrymandering metrics.[23] It then calculates various partisan bias metrics. Like the earlier approaches, PlanScore indicates that the proposed plan is relatively neutral with a small pro-Republican bias (Table 4).

Metric	Value	Favors Rep's in this % of Scenarios	More Biased than this % Historical Plans	More Pro-Republican than this % Historical Plans
2014-2020 Plan				
Symmetry	-4.5%	99%	50%	72%
Mean-Median Diff.	-2.0%	99%	42%	68%
Efficiency Gap	-4.6%	99%	53%	81%
Declination	-.27	99%	57%	76%
Average		99%	50%	74%
Proposed Plan				
Symmetry	-2.5%	94%	31%	61%
Mean-Median Diff.	-1.2%	94%	27%	61%
Efficiency Gap	-2.5%	95%	32%	70%
Declination	-.15	95%	37%	64%
Average		95%	31%	64%

Table 4: PlanScore partisan bias metrics for proposed State House plan

22. See `https://planscore.campaignlegal.org/plan.html?20211228T165635.851306606Z` for the proposed plan and `https://planscore.campaignlegal.org/plan.html?20220107T194310.216726037Z` for the 2014-2020 plan.

23. See `https://planscore.campaignlegal.org/models/data/2021D/` for more details.

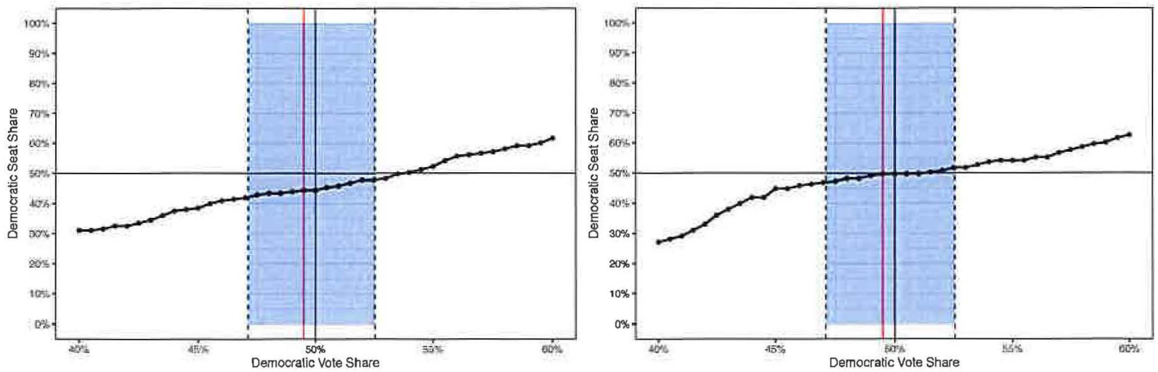
According to PlanScore, the proposed plan has a small pro-Republican symmetry bias of -2.5%. This means that Republicans would win 52.5% of the seats if the two parties evenly split the votes. The proposed plan favors Republicans in 95% of the scenarios estimated by PlanScore. The other metrics look similar to the symmetry metric. Across all the metrics, the proposed plan is more pro-Republican than 64% of prior plans over the past five decades. Figure 7 graphically shows the bias of the proposed plan compared to previous plans from 1972-2020.[24] Overall, the graphs show that the proposed plan is close to the center of the distribution of previous plans over the past 50 years with just a small pro-Republican bias.

Figure 7: Graphs of PlanScore metrics proposed State House plan compared to previous plans from 1972-2020

5.4 Responsiveness of Plan

Another benchmark for a districting plan is the responsiveness of the plan to changes in voters' preferences (Cox and Katz 1999). An unresponsive map ensures that the bias in a districting plan toward the advantaged party is insulated against changes in voters' preferences, and thus is durable across multiple election cycles.

24. Note that the PlanScore graphs are oriented so that pro-Republican scores have a positive value.



(a) 2014-20 State House Map (b) Proposed State House Map

Figure 8: Vote-seat curve in Pennsylvania using uniform swings in 2020 election results on the 2014-20 districts and re-aggregated on the proposed plan. The shaded area shows the range between the minimum and maximum Democratic statewide vote share in State House elections from 2014-2020. The red line shows the actual Democratic statewide vote share in the 2020 State House elections.

Figure 8 compares the responsiveness of the 2014-20 State House plan and the proposed State House plan (using re-aggregated votes in the 2020 State House Elections). It shows the vote-seat curve in Pennsylvania using uniform swings in 2020 election results on the 2014-20 districts and re-aggregated on the proposed plan. The shaded area shows the range between the minimum and maximum Democratic statewide vote share in State House elections from 2014-2020. The red line shows the actual Democratic statewide vote share in the 2020 State House elections.

The graph shows that both the previous plan and the proposed plan are relatively responsive to shifts in voters' preferences. But the 2014-20 plan had a large pro-Republican bias, which is much smaller in the proposed plan. Indeed, the Republican Party won a majority of the seats across all of the plausible range of stateside vote shares in the 2014-20 plan, while both parties could get at least half the seats in the proposed plan.

5.5 Number of Competitive Districts

An important factor that affects the overall responsiveness of a plan is the number of competitive districts in a plan. I use a variety of approaches to estimate the number of competitive districts in both the 2014-20 State House plan and the proposed plan (see Table 5). Overall, my analysis indicates that the previous plan and the proposed plan are very similar in terms of the number of competitive seats. Moreover, both plans do about as well as the average percentage of seats that are competitive across other states' elections for their lower chambers in 2020.[25]

Data:	2020 State House Results	Composite (2014-20)	PlanScore			Mean
Metric:	45-55	45-55	45-55	20%+ Prob. of Each Party Win.	50%+ Prob. Flip in Dec.	
Plan	(1)	(2)	(3)	(4)	(5)	(6)
Average Nationwide in 2020	13%					
2014-20 Plan	13%	24%	23%	20%	25%	21%
Proposed Plan	12%	21%	23%	18%	23%	19%

Table 5: Number of competitive districts using various data sources and metrics.

First, I use the actual 2020 State House results to examine the number of competitive districts. In column 1 of Table 5, I begin by tallying the number of districts where each party's two-party vote share was between 45 and 55%. This approach indicates that 13% of the districts on the 2014-20 plan were competitive and 12% of the districts on the proposed plan were competitive. It is important to note, however, that a sharp threshold at 55% may not be the best measure of competitiveness.

Next, I use a composite of the 2014-2020 statewide election results to estimate the number of competitive districts. Once again, in column 2 of Table 5, I tally the number of districts where each party's two-party vote share was between 45 and 55%. This approach indicates that 24% of the districts on the 2014-20 plan were competitive and 21% of the districts on the proposed plan were competitive.

Lastly, I use PlanScore to estimate the potential competitiveness of individual districts on the proposed plan. In column 3 of Table 5, I show the number of districts where PlanScore estimates that each party's two-party vote share is expected to be between 45 and 55%. This approach indicates that 23% of the districts on the 2014-20 plan were competitive and 23% of the districts on the proposed plan were competitive

It is also possible to use PlanScore to evaluate whether a district is likely to switch parties at least once per decade (Henderson, Hamel, and Goldzimer 2018). PlanScore conducts 1,000 simulations of possible electoral scenarios based on the results of the 2014-2020 congressional and state legislative elections in every state. Using these simulations,

25. The nonpartisan Princeton Gerrymandering Project gives the proposed plan a low grade on competitiveness. However, their analysis has two material flaws as applied to the proposed plan in Pennsylvania. First, it only uses three recent statewide elections to evaluate competitiveness, and Democrats did unusually well in two of those three elections (2018 Senate and 2018 Governor). Overall, Democrats won 55.3% of the two-party vote in those three elections. Second, it uses a single, very narrow vote share range to classify districts as competitive (46.5-53.5%). Combined, these two assumptions mean that the vast majority of the districts that the Princeton Gerrymandering Project classifies as competitive are unlikely to actually be competitive in a close statewide election. Indeed, Republicans would win the vast majority of these districts. Thus, I do not view the Princeton Gerrymandering Project's analysis of the plan's level of competitiveness as a reliable measure of the proposed Pennsylvania State House plan.

PlanScore provides an estimate of the probability that each party will win each seat as well as whether they are likely to have at least a 50% chance of winning each seat once over the course of the decade. In column 4 of Table 5, I estimate the number of districts where each party has at least a 20% chance of winning according to PlanScore. This approach indicates that 20% of the districts on the 2014-20 plan were competitive and 18% of the districts on the proposed plan were competitive. In column 5 of Table 5, I conduct a similar analysis where I tally the number of districts that each party would have at least a 50% chance of winning at least once over the course of the decade. This approach indicates that 25% of the districts on the 2014-20 plan were competitive and 23% of the districts on the proposed plan were competitive

Finally, column 6 of Table 5 averages across all of these approaches. It indicates that 21% of the districts on the 2014-20 plan were competitive and 19% of the districts on the proposed plan were competitive. Thus, the previous plan and the proposed plan are very similar in terms of the number of competitive seats. The proposed plan also has roughly the same percentage of seats that are competitive as other states' elections for their lower chambers in 2020.

6 Conclusion

This report has evaluated the partisan fairness of the Legislative Reapportionment Commission's proposed Pennsylvania State House plan. Based on three methods of projecting future elections and four different, generally accepted partisan bias metrics, I find that the plan is fair, with just a small pro-Republican bias. On this plan, the party that wins the majority of the votes is likely to usually win the majority of the seats. Thus, the plan satisfies a key premise of democratic theory. Moreover, I find that the plan is much more fair than the 2014-2020 State House plan, which had a large and durable pro-Republican bias. The plan is also likely to be responsive to shifts in voters' preferences.

References

Best, Robin E, Shawn J Donahue, Jonathan Krasno, Daniel B Magleby, and Michael D McDonald. 2017. "Considering the Prospects for Establishing a Packing Gerrymandering Standard." *Election Law Journal: Rules, Politics, and Policy.*

Brennan Center. 2017. *Extreme Maps.* `https://www.brennancenter.org/publication/extreme-maps`.

Buzas, Jeffrey S, and Gregory S Warrington. 2021. "Simulated packing and cracking." *Election Law Journal: Rules, Politics, and Policy.*

Caughey, Devin, Chris Tausanovitch, and Christopher Warshaw. 2017. "Partisan Gerrymandering and the Political Process: Effects on Roll-Call Voting and State Policies." *Election Law Journal* 16 (4).

Chen, Jowei, and Jonathan Rodden. 2013. "Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures." *Quarterly Journal of Political Science* 8 (3): 239–269.

Cox, Gary W., and Jonathan N. Katz. 1999. "The reapportionment revolution and bias in US congressional elections." *American Journal of Political Science:* 812–841.

Fraga, Bernard L, and Eitan D Hersh. 2018. "Are Americans stuck in uncompetitive enclaves? An appraisal of US electoral competition." *Quarterly Journal of Political Science* 13 (3): 291–311.

Gelman, Andrew, and Gary King. 1994. "A unified method of evaluating electoral systems and redistricting plans." *American Journal of Political Science* 38 (2): 514–554.

Goedert, Nicholas. 2014. "Gerrymandering or geography? How Democrats won the popular vote but lost the Congress in 2012." *Research & Politics* 1 (1): 2053168014528683.

Henderson, John A, Brian T Hamel, and Aaron M Goldzimer. 2018. "Gerrymandering Incumbency: Does Nonpartisan Redistricting Increase Electoral Competition?" *The Journal of Politics* 80 (3): 1011–1016.

ICPSR. 2006. *State Legislative Election Returns in the United States, 1968-1989.*

Jacobson, Gary C. 2015. "It's nothing personal: The decline of the incumbency advantage in US House elections." *The Journal of Politics* 77 (3): 861–873.

Jacobson, Gary C, and Jamie L Carson. 2015. *The politics of congressional elections.* Rowman & Littlefield.

Kang, Michael S. 2017. "Gerrymandering and the Constitutional Norm Against Government Partisanship." *Mich. L. Rev.* 116:351.

Katz, Jonathan N, Gary King, and Elizabeth Rosenblatt. 2020. "Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies." *American Political Science Review* 114 (1): 164–178.

Klarner, Carl, William Berry, Thomas Carsey, Malcolm Jewell, Richard Niemi, Lynda Powell, and James Snyder. 2013. *State Legislative Election Returns (1967–2010).* ICPSR34297-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-01-11. doi:`10.3886/ICPSR34297.v1`.

Kollman, K., A. Hicken, D. Caramani, D. Backer, and D. Lublin. 2017. *Constituency-level elections archive [data file and codebook].* Ann Arbor, MI: Center for Political Studies, University of Michigan.

Krasno, Jonathan S, Daniel Magleby, Michael D McDonald, Shawn Donahue, and Robin E Best. 2018. "Can Gerrymanders Be Detected? An Examination of Wisconsin's State Assembly." *American Politics Research.*

Levitt, Justin. 2017. "Intent is Enough: Invidious Partisanship in Redistricting." *Wm. & Mary L. Rev.* 59:1993.

McDonald, Michael P. 2014. "Presidential vote within state legislative districts." *State Politics & Policy Quarterly* 14 (2): 196–204.

McGhee, Eric. 2014. "Measuring Partisan Bias in Single-Member District Electoral Systems." *Legislative Studies Quarterly* 39 (1): 55–85.

———. 2017. "Measuring Efficiency in Redistricting." *Election Law Journal: Rules, Politics, and Policy.*

———. 2018. *Assessing California's Redistricting Commission: Effects on Partisan Fairness and Competitiveness.* Report from the Public Policy Institute of California. Available at `http://www.ppic.org/publication/assessing-californias-redistricting-commission-effects-on-partisan-fairness-and-competitiveness/`.

MIT Election and Data Science Lab. 2017. *U.S. House 1976–2016.* Available on the Harvard Dataverse at `http://dx.doi.org/10.7910/DVN/IGOUN2`.

Nagle, John F. 2015. "Measures of partisan bias for legislating fair elections." *Election Law Journal* 14 (4): 346–360.

Niemi, Richard G, and John Deegan. 1978. "A theory of political districting." *American Political Science Review* 72 (4): 1304–1323.

Powell, G. Bingham, Jr. 2004. "Political Representation in Comparative Politics." *Annual Review of Political Science* 7:273–296.

Rogers, Steven. 2017. "Electoral Accountability for State Legislative Roll Calls and Ideological Representation." *American Political Science Review* 111 (3): 555–571.

Royden, Laura, Michael Li, and Yurij Rudensky. 2018. *Extreme Gerrymandering & the 2018 Midterm.*

Seabrook, Nicholas R. 2017. *Drawing the Lines: Constraints on Partisan Gerrymandering in US Politics.* Cornell University Press.

Stephanopoulos, Nicholas O, and Christopher Warshaw. 2020. "The impact of partisan gerrymandering on political parties." *Legislative Studies Quarterly* 45 (4): 609–643.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2015. "Partisan Gerrymandering and the Efficiency Gap." *University of Chicago Law Review* 82 (2): 831–900.

———. 2018. "The measure of a metric: The debate over quantifying partisan gerrymandering." *Stan. L. Rev.* 70:1503.

Tufte, Edward R. 1973. "The relationship between seats and votes in two-party systems." *American Political Science Review* 67 (2): 540–554.

Wang, Samuel. 2016. "Three Tests for Practical Evaluation of Partisan Gerrymandering." *Stan. L. Rev.* 68:1263–1597.

Warrington, Gregory S. 2018a. "Introduction to the declination function for gerrymanders." *arXiv preprint arXiv:1803.04799.*

———. 2018b. "Quantifying Gerrymandering Using the Vote Distribution." *Election Law Journal* 17 (1): 39–57.

Wright, Gerald, Tracy Osborn, Jon Winburn, and Jennifer Hayes Clark. 2009. "Patterns of representation in the American states and Congress." In *annual conference on State Politics and Policy, Chapel Hill, NC.*

Partisan Fairness of the Pennsylvania Legislative Reapportionment Commission's Proposed State House Districting Plan

Christopher Warshaw

George Washington
Department of Political Science

January 14, 2022

My Background

- JD/PhD from Stanford University
- Currently an Associate Professor of Political Science at George Washington University
- Research focuses on political representation, redistricting, elections, and public opinion. In all, I have written 24 peer reviewed articles and I have a book coming out this summer called *Dynamic Democracy: Public Opinion, Elections, and Policy Making in the American States*.

Roadmap

- Methodology
  - Project future elections based on three different approaches
  - Evaluate partisan fairness using four different metrics
- Results
  - All of my analyses indicate that the plan is fair with just a small pro-Republican bias

Projecting Future Elections

I use three different approaches to project future election results on the new map:

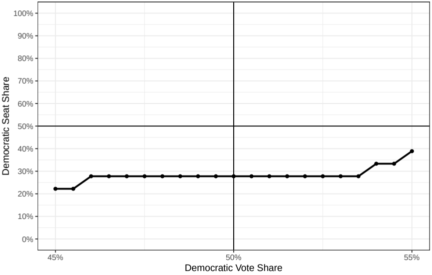
1. Composite of statewide elections from 2014-2020
   - Includes: 2016 Presidential, 2020 Presidential, 2014 Governor, 2018 Governor, 2016 Attorney General, 2020 Attorney General, 2016 Senate, 2018 Senate, 2016 Treasurer, 2020 Treasurer, 2016 Auditor General, and 2020 Auditor General election.
   - I average results within year and then across years.
2. 2020 State House elections
   - Impute uncontested elections based on the 2020 presidential election results.
3. PlanScore.org
   - Statistical model that predicts results on a new plan based on the relationship between presidential election results and legislative election results around the country over the past decade.

Partisan fairness

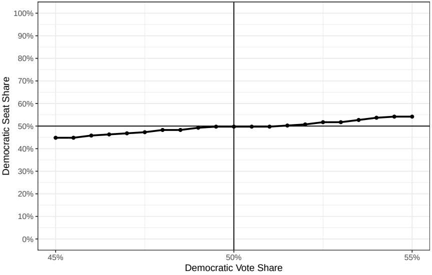
I use four different generally accepted academic approaches to evaluate the partisan fairness of the plan.

1. Symmetry
2. Mean-Median Difference
3. Efficiency Gap
4. Declination

1) Symmetry

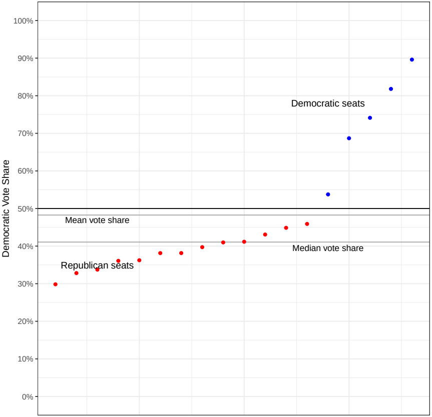


(a) Biased, Pro-Rep. symmetry on 2016 US House Map

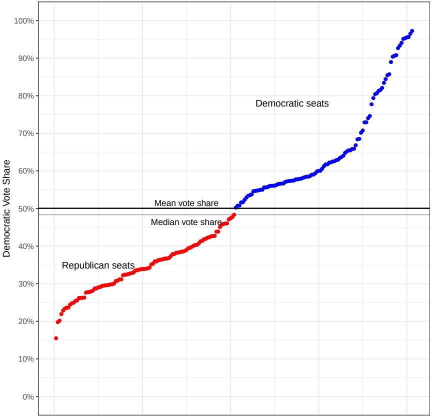


(b) Relatively neutral symmetry on Proposed LRC State House Map

2) Mean-Median Difference



(a) Biased, Pro-Rep. mean-median difference on 2016 US House Map



(b) Relatively neutral mean-median diff. on Proposed LRC State House Map

3) Efficiency Gap

Table: Hypothetical Example of Efficiency Gap

District	Democratic Votes	Republican Votes
1	75	25
2	40	60
3	40	60
Total	155 (52%)	145 (48%)
Wasted	104	43

Plan	Efficiency Gap
Hypothetical Example	-20%

3) Efficiency Gap

Table: Hypothetical Example of Efficiency Gap

District	Democratic Votes	Republican Votes
1	75	25
2	40	60
3	40	60
Total	155 (52%)	145 (48%)
Wasted	104	43

Plan	Efficiency Gap
Hypothetical Example	-20%
2016 Congressional Election	-19%

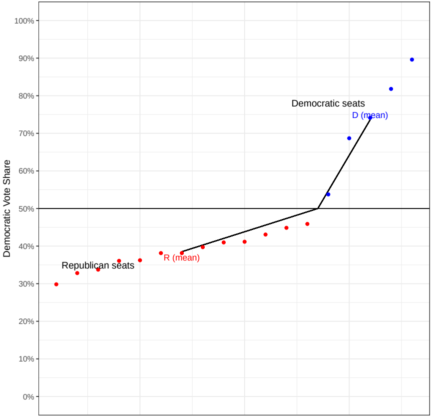
3) Efficiency Gap

Table: Hypothetical Example of Efficiency Gap

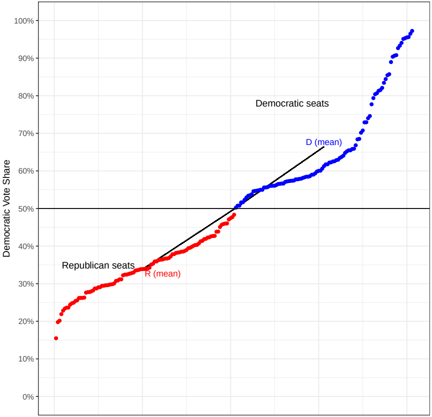
District	Democratic Votes	Republican Votes
1	75	25
2	40	60
3	40	60
Total	155 (52%)	145 (48%)
Wasted	104	43

Plan	Efficiency Gap
Hypothetical Example	-20%
2016 Congressional Election	-19%
LRC Plan based on 2020 State House Election	0.7%

4) Declination

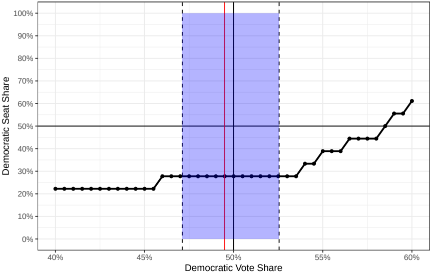


(a) Biased, Pro-Rep. declination on 2016 US House Map

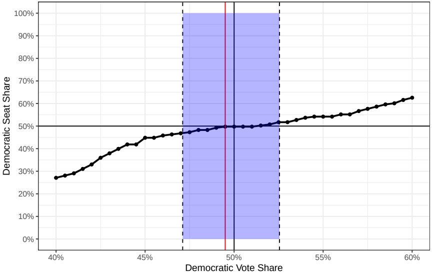


(b) Relatively neutral declination on Proposed LRC State House Map

Responsiveness



(a) Unresponsive
2016 US House Map



(b) Relatively Responsive
Proposed LRC State House Map

Composite of previous statewide elections from 2014-2020

Metric	Value	2014-2020 Composite: > Biased than this % Elections (1972-2020)	> Neutral than this % Elections (1972-2020)	> Pro-Rep. than this % Elections (1972-2020)
2014-2020 Plan				
Republican Vote Share	46%			
Republican Seat Share	49%			
Symmetry Bias	-7.7%	77%	23%	85%
Mean-Median	-3.8%	70%	30%	81%
Efficiency Gap	-5.8%	60%	40%	83%
Declination	-.348	66%	34%	82%
Average		68%	32%	83%
Proposed Plan				
Republican Vote Share	46%			
Republican Seat Share	46%			
Symmetry Bias	-2.5%	29%	71%	61%
Mean-Median	-1.4%	31%	69%	63%
Efficiency Gap	-2.6%	27%	73%	69%
Declination	-.175	38%	62%	65%
Average		31%	69%	65%

The preliminary LRC plan is relatively neutral with a small pro-Republican bias based on the composite of statewide elections.

2020 State House election results

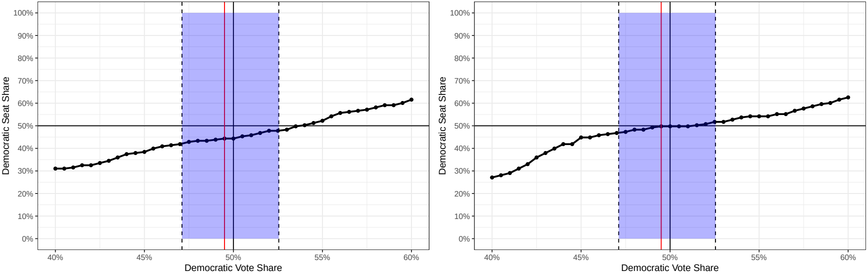
Metric	Value	More Biased than this % Historical Elections (1972-2020)	> Neutral than this % Elections (1972-2020)	More Pro-Rep. than this % Historical Elections (1972-2020)
2014-2020 Plan				
Republican Vote Share	50%			
Republican Seat Share	56%			
Symmetry Bias	-5.7%	60%	40%	77%
Mean-Median Diff	-4.3%	79%	21%	86%
Efficiency Gap	-4.8%	49%	51%	78%
Declination	-.36	68%	32%	83%
Average		64%	36%	81%
Proposed Plan				
Republican Vote Share	50%			
Republican Seat Share	50%			
Symmetry Bias	-0.2%	2%	98%	49%
Mean-Median Diff	-1.9%	40%	60%	68%
Efficiency Gap	0.7%	8%	92%	51%
Declination	-.04	9%	91%	50%
Average		15%	85%	55%

The preliminary LRC plan is politically neutral based on the 2020 State House election results.

PlanScore

The preliminary LRC plan is relatively neutral with a small pro-Republican bias based on the PlanScore.org website.

Responsiveness of Proposed Plan



(a) 2014-20 State House Map

(b) Proposed State House Map

The preliminary LRC plan is responsive to shifts in mass preferences and the party that gets a majority of votes would usually get a majority of the seats.

Number of Competitive Districts

Data:	2020 State House Results	Composite (2014-20)	PlanScore			Mean
Metric:	45-55	45-55	45-55	20%+ Prob. of Each Party Win.	50%+ Prob. Flip in Dec.	
Plan	(1)	(2)	(3)	(4)	(5)	(6)
Average Nationwide in 2020	13%					
2014-20 Plan	13%	24%	23%	20%	25%	21%
Proposed Plan	12%	21%	23%	18%	23%	19%

The preliminary LRC plan and the 2014-2020 plan are very similar in terms of the proportion of competitive seats. The LRC plan also has roughly the same percentage of seats that are competitive as other states' elections for their lower chambers in 2020.

Conclusion

- The plan is likely to be responsive to shifts in voters' preferences.
- On this plan, the party that wins the majority of the votes is likely to usually win the majority of the seats.
- Based on three methods of projecting future elections and four different, generally accepted partisan bias metrics, I find that the plan is fair, with just a small pro-Republican bias.

Pennsylvania Racially Polarized Voting Analysis

Jonathan N. Katz

January 14, 2022

1 Introduction

I was asked by counsel to discuss the statistical issues related to estimating the voting behavior of racial and ethnic groups necessary for conducting a racially polarized voting analysis in Pennsylvania as well as review the analysis of Dr. Matt A. Barreto in his memo of January 7. 2022. In making my findings, I have applied standard statistical methods, which I regularly employ in my research and which have been published in peer-reviewed journals.

A summary of my report and basic findings is as follows:

- All existing statistical methods for ecological inference — i.e., inferring group voting behavior from aggregate data — rely heavily on the problematic *constancy assuption* in the absence of ethnically or racially homogeneous precincts.
- Given that there are no homogeneous Latino/Hispanic voting precincts in the state, any attempt at ecological inference (including both ecological regression and EI) of Hispanic voting behavior is suspect and not scientifically valid. As a result, it is not possible to say whether or not Hispanics typically vote en bloc or in consistent coalition with other ethnic or racial groups.
- Finally, Dr. Barreto's analysis of racially polarized voting is statistically flawed and no scientifically valid inferences can be drawn from it.

In the next section of my report, I review my qualifications. The following section discusses the statistical methods for estimating voting behavior from aggregate data. This is referred to as ecological inference in the statistics and social science literature. The next section then discusses the problem of identifying and making ecological inferences in Pennsylvania given the available data.

2 Qualifications

I am currently the Kay Sugahara Professor of Social Sciences and Statistics at the California Institute of Technology. I previously served for seven years as the Chair of the Division of the Humanities and Social Sciences at Caltech (which is akin to being a dean at other universities). Further, I was also formerly on the faculty at the University of Chicago and a visiting professor at the University of Konstanz (Germany). A complete copy of my curriculum vitae is in attached to this report.

I received my Bachelor of Science degree from the Massachusetts Institute of Technology and my Masters of Arts and Doctor of Philosophy degrees, both in political science, from the University of California, San Diego. I did post-doctoral work at Harvard University and the Harvard-MIT Data Center. I am an elected fellow of both the American Academy of Arts and Sciences and an inaugural fellow of the Society for Political Methodology. I am a former fellow of the Center for Advanced Study in the Behavioral Sciences.

I have written numerous articles published in the leading journals as set forth in my curriculum vitae. I am currently the Deputy Editor for Social Sciences of *Science Advances*, the open access journal of the American Association for the Advancement of Science. I previously served as co-editor of *Political Analysis*, the journal of the Society for Political Methodology, and I was a co-founding editor of the Political Science network (a collection of on-line journals). I have also previously served on the editorial boards of *Electoral Studies*, *Political Research Quarterly* and the *American Journal of Political Science*. I have frequently served as a referee of manuscripts for most of the major journals in my fields of research and the National Science Foundation.

I have done extensive research on American elections and on statistical methods for analyzing social science data. I am a member of the Caltech/MIT Voting Technology Project, serving as the co-director of the project from October 1, 2005 to September 30, 2010.

Over the past two decades, I have been involved in numerous elections cases for both Democratic and Republican clients involving the Federal Voting Rights Act, partisan gerrymandering, the evaluation of voting systems, or the statistical evaluation of electoral data. I have testified or consulted in court cases in both state and Federal courts in the states of Arizona, California, Florida, Georgia, Indiana, Illinois, Maryland, Michigan, Missouri, New Hampshire, New Mexico, Nevada, North Carolina, Ohio, Oklahoma, Oregon, Texas, Virginia, and Washington.

3 Methods for Ecological Inference

The problem of inferring voting behavior from aggregate information is known as ecological inference. That is, we are interested in estimating how groups of voters, say members of a Minority Group and Others (i.e., non-members of the Minority Group), voted in a given election when all we observe are the precinct-level returns and the demographic make-up of the precincts.

3.1 Homogenous Precincts and the Method of Bounds

A common starting point is to consider only homogeneous precincts. That is, we could examine the election results from precincts that are closest to racially/ethnically homogeneous in character. For example, if a precinct were completely homogeneous, say with a population that was 100% of a particular Minority Group, then we know what fraction of that Minority Group that voted for a given candidate in the precinct: it is just the share the given candidate got in the precinct. Besides being simple, this statistical estimate does not require any additional assumptions to be valid. While this might be a useful starting point, as a statistical procedure it is problematic, since it throws out most of the data unless most of the precincts are homogeneous.

However, we can use the intuition from the homogeneous precincts to place bounds on the level of support each group gives a candidate. Consider the following equation, which is true by definition (and without any further statistical assumptions), that relates the vote share of given candidate to the voting behavior of a particular Minority Group and Others:

$$V_i = \lambda_i^M X_i + \lambda_i^O(1 - X_i), \tag{1}$$

where V_i is the share of the vote a given candidate received in precinct i, X_i is the fraction of Minority Group voters in the precinct and therefore $(1 - X_i)$ is the fraction of Other voters, assuming for the moment that there are only two groups in the electorate. λ_i^M is the fraction of the Minority Group voting for the given candidate and similarly λ_i^O is the fraction of Others voting for the given candidate. In other words, the equation states the fact that the total vote share for a candidate must equal the proportion of Minority Group voters who support them multiplied by the proportion of the electorate that is in the Minority Group plus the proportion of the Other voters who support the candidate multiplied by the proportion of the electorate which is Other.

In the case of only two groups — e.g., a particular Minority Group and Others – and only two candidates, then racially polarized voting occurs when λ_i^M and λ_i^O are on opposite sides of 0.5 — e.g., $\lambda_i^M > 0.5$ and $\lambda_i^O < 0.5$. That is, a majority of one group voting for one candidate and the majority of the other group voting for the opposite candidate. If this holds, then the larger the difference between support levels, the greater the level of polarization. Of course, since we are dealing with statistical estimates, this difference must be greater than the statistical uncertainty in the estimates.

Now consider homogeneous Minority Group precincts again. In these precincts, $X_i = 1$, so that the equation simplifies to $V_i = \lambda_i^M$ as we stated above. However, from these precincts we can not say anything about the voting behavior of Others because any proportion of Others voting for a given candidate is consistent with the observed vote shares in these precincts. We can generalize this idea using Equation 1. Consider, for example, a precinct where $X_i = 0.6$, that is 60% of voters are Minority (and, therefore, 40% are Other), and the candidates vote share, V_i, is 0.5.

Since 60% of the voters are part of the Minority Group and the given candidate got 50% of the vote, then at most $\frac{5}{6}$ths of the Minority Group voters could have voted for the candidate. If it were higher than this bound, then the vote share for the candidate in the precinct would have to be higher. On the other hand, even if all of the Others voted for the candidate, then at least $\frac{1}{6}$th of the Minority Group would have had to vote for the candidate as well, otherwise the candidate's vote share would have been less than 0.5. Thus, we know that proportion of Minority Group voting for the candidate, λ_i^M, must be greater than $1/6$ and less than $5/6$ and λ_i^O can take on any value between zero and one. We actually know more than this: we know that the feasible values for this district must lie on the line segment, called a constraint line, defined by the bounds $(\frac{1}{6}, 1)$ and $(\frac{5}{6}, 0)$. Using standard algebra by plugging in $X_i = 0.6$ and $V_i = 0.5$, we find that $\lambda_i^{OW} = -\frac{3}{2}\lambda_i^M + \frac{5}{4}$.

Duncan and Davis (1953) fully developed the method of bounds outlined above to analyze ecological data. Unfortunately, with a large number of precincts, it is difficult to make much direct use of these bounds since we need a way to combine them to understand typical behavior in the district. These bounds do, however, provide important useful information as we will see below.

3.2 Ecological or Goodman's Regression

An alternative approach that examines all precincts simultaneously was developed by Goodman (1959) and is perhaps the most commonly used procedure. It is referred to in the literature as ecological regression or Goodman's regression. Like the method of bounds, it is based on the identity in Equation 1. Suppose that the fraction of support for a given candidate for both Others and a Minority Group members was the same across all precincts in the district. A bit more formally, suppose that $\lambda_i^M = \lambda^M$ and $\lambda_i^O = \lambda^O$ for every precinct i. Then we could estimate these fractions by choosing the best fitting line to the precinct-level data. This is just a standard linear regression, the most commonly used statistical procedure in the social sciences. From these estimates we could then compare the voting behavior between groups.

However, there is no free lunch, ecological regression allows one to identify the estimate across all districts and in any data set by making the heroic assumption of no variability of voting behavior across precincts and individuals, which is usually referred to as the *constancy assumption*. In fact, Goodman himself was extremely cautious in recommending the use of ecological regression to infer individual relationships given this required assumption. He stressed that only "under very special circumstances" should ecological regression be relied upon to produce reasonable estimates (Goodman 1953: 664, see also Robinson 1950).

A more technical critique of ecological regression and its constancy assumption was made by Freedman et al. (1991) (see also Gelman et al. 2001). They develop an alternative model that they called the "neighborhood model". The full argument is highly technical and it is beyond the scope of this report. In brief, they show in aggregate election results this model is mathematically equivalent to Goodman's ecological regression level but has dramatic differences at the individual (or group level). That is, the aggregate data can not identify individual behavior except under untestable assumptions and different such assumptions lead to dramatically different estimates of individual behavior. Finally, King (1997) showed ecological regression can produce widely inaccurate estimates of group voting behavior. Thus the consensus of the statistical literature is to reject analysis based on ecological regression (see, for example, Schuessler 1991 or Flanigan and Zingale 1985).

3.3 Ecological Inference/EI

King (1997) has developed an alternative approach called Ecological Inference or EI. While the technical details are complex, its advantage is that it uses all available information to generate more accurate estimates of voting behavior from aggregate data. EI is basically a way to combine the regression approach of Goodman (1959) with the bounds from Duncan and Davis (1953). Further, it allows the estimates to vary (systematically) across precincts. The idea is we calculate the constraint lines for every precinct. We then choose as our estimate for a given precinct a point on its constraint line near the center of the intersection of all of the other lines. The actual point chosen is based on a standard statistical model. We can then use these precinct estimates to calculate quantities of interest such as the average support level across the district.

It is important to note that since King's method relies heavily on the bounds information, it works best when at least some of these bounds are informative — i.e., narrower than the entire range from 0 to 1. This will happen when more precincts have large proportions of each of the groups whose voting behavior we want to estimate. In other words, we will need some precincts that are relatively homogeneous for each ethnic group we want to study. When this is not the case, EI can go wildly wrong as noted by King himself (1997, see chapter 9).

That is, the EI estimates are not well identified when the bounds are not informative. This is because EI is then just a slight generalization of Goodman's regression in this case. It, therefore, relies on the same problematic constancy assumption for identification that has been rejected in the statistical literature when bounds data is unavailable.

3.4 More than Two Groups or Two Candidates

The above discussion on the development of methods for ecological inference assumed that we only had two groups and two candidates (or vote choices). Accommodating more than two groups is rather straight-forward, although notation and intuition become more complicated, especially for the constraint lines. All that is required is adding the additional group fractions to Equation 1.

Allowing for more than two candidates or vote choices, however, is a bit more complicated. In the special case of only two choices, we only need to model the vote share going to one of them since we then automatically know the fraction going to the other candidate: this is just one minus the first vote share. If, for example, we add a third choice, then we need to model the vote share going to any two of the options and then we get third by subtracting the sum of the other two shares from one. Formally, we need to add an additional equation for each vote choice greater than two. Typically, there will always be more than two vote choices even when there are only two candidates because some individuals will choose not to vote in the election. We need to account for this abstention in order to make proper inferences. However, since what we care about is the share of voters supporting each of the candidates, we need to condition out these non-voters. This is not straight-forward, but can be done once we estimate the full set of options: don't vote or vote for one of the candidates on the ballot.

In the general case of more than two groups and more than two vote choices, racially polarized voting is also a more complicated concept. If we only have two choices, then we get voting cohesion among each group automatically since one of the choices must receive a majority of support from the members (ignoring the unlikely event of an exact tie in the election). However, when we have more than two choices, it is possible that no choice receives majority support of the group. In fact, given the estimation uncertainty, it may not be possible to infer which candidate is preferred by the members of the group.[1] Even if we find that the groups both have a strictly preferred candidates (i.e., they are cohesive), we still need to see if the distribution between the groups is statistically different to find racially polarized voting.[2]

I finally note that adding additional groups and vote choices to King's (1997) EI is not straight-forward. The generalization was first developed by King, Rosen, and Tanner (1999). Unfortunately,

[1]Formally, we can not rule out the null hypothesis that the group equally split their votes across two or more choices.

[2]Formally, we need to reject the null hypothesis that the distribution of vote shares across groups is identical.

their approach was computationally inefficient and was later refined by Rosen, Jiang, King and Tanner (2001).

4 Problem with Ecological Inference in Pennsylvania elections

As discussed above, EI produces reliably estimates only when we have substantial numbers of homogeneous precincts. Unfortunately, this is not the case with regards to Latinos/Hispanics in Pennsylvania. In fact, according to the official 2020 Census data there is exactly **one precinct**, Philadelphia Ward 10, Precinct 06, that has just over 90% Hispanic voting age population. I note that even with this level of homogeneity, the bounds on voting behavior are still rather large. In fact, even if relaxed our criteria for homogeneous precincts to be greater than 80% Hispanic voting age population, there are still only 23 precincts out of almost 9200 statewide that are even modestly homogeneous. These are almost all in the Philadelphia area with the exception of three precincts in Reading. This is just not enough for statistically valid estimates.

Further, given the lack of homogeneous precincts, it is not possible to see if two separate groups of voters, say Hispanics and Blacks, typically vote together in coalition. This statistical analysis is just another version of ecological inference that requires homogeneous precincts to provide reasonable estimates.

4.1 How badly can EI go wrong?

How badly can EI go wrong when there are not a sufficient of homogeneous precincts? It not possible to say in data directly from Pennsylvania elections, since we would need to know the true voting behavior of Latinos.

However, in analysis that I did in a California in the case in Federal district court, *Luna v. Kern County*, we can get some feeling for how badly. Like in Pennsylvania, there are essentially no homogeneous Latino precincts in Kern County. We can get reliable estimates of the number of Latinos who are registered as Democrats by using Census name matching techniques to the voter rolls. This will serve as our true value benchmark. Then we can use EI to estimate the fraction of registered Democrats (in exactly the same way we would estimate the number of Latinos voting for a particular candidate) based on the fraction of Democrats and non-Democrats registered in the precinct (i.e., the vote percentages) and fraction of Latinos and non-Latinos in the precinct.

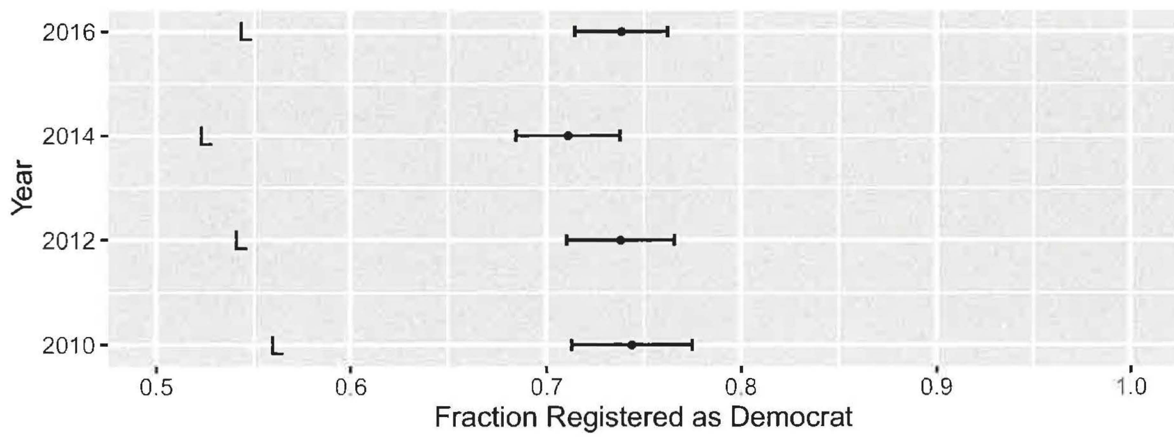


Figure 1: *EI Estimates of the fraction of Latinos who are Registered as Democrats in Kern County in the June elections from 2010 to 2016. The center dot represents the point estimate and the error bars provide the 95% confidence interval for the estimate. The true value from the registration rolls is denoted by the letter "L".*

The EI estimates of the fraction of Latinos who are registered in Kern County as Democrats in the June elections from 2010 to 2016 are displayed graphically in Figure 1 (next page). The center dot on the chart is the point estimate and the bars around it are the "95% confidence intervals." When a statistical estimator is well identified, the true value should be contained in this confidence interval almost always. As you can see, the EI estimates do not perform well at all. The true values for each election are are shown in the figure by the letter "L". As you can see the true fraction of Latinos registered as Democrats is around 55%, but all of the estimates are in the mid 70s. That is, EI is significantly over-estimating the fraction of Latinos who are registered as Democrats. And perhaps a more disconcerting finding in this analysis is that the confidence intervals do not contain the true value in any of the four elections examined.

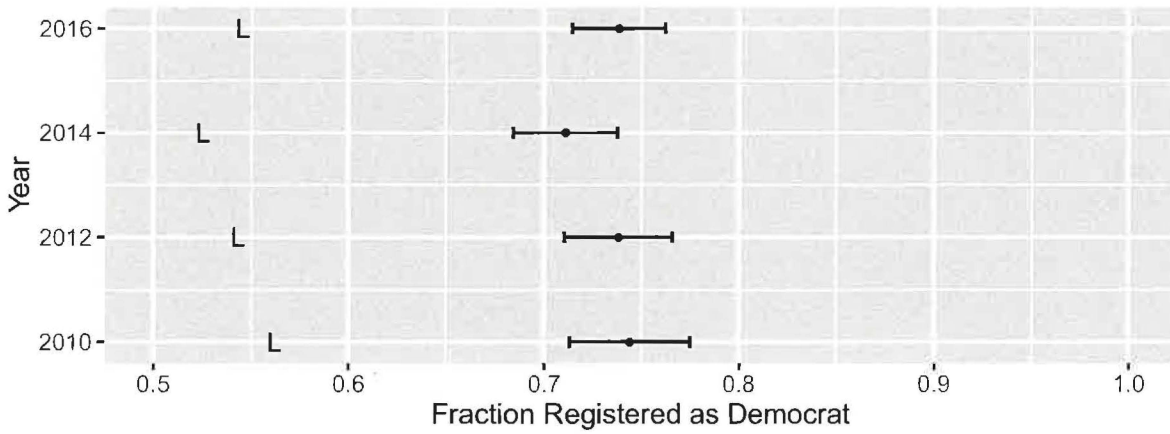


Figure 2: *Ecological Regression Estimates of the fraction of Latinos who are Registered as Democrats in Kern County in the June elections from 2010 to 2016. The center dot represents the point estimate and the error bars provide the 95% confidence interval for the estimate. The true value from the registration rolls is denoted by the letter "L".*

For completeness, I have also include the estimate of the ecological regression results from the same analysis from the Kern County data in Figure 2. Given that there are no homogeneous precincts, the estimates are very similar to the EI estimates above. And as before, the ecological regression results are substantially different from the ground truth estimate.

5 Problems with Dr. Barreto's Analysis

Dr. Barreto's central racially polarized voting analysis is contained in the series of graphs relating Republican vote share in a precinct to the percent of White voting age precinct in various collection of counties in Pennsylvania.[3] The claim is that these graphs show that there is racially polarized voting in the state. However, this analysis contains numerous serious statistical flaws and no valid scientific claims about the presence or absence of racially polarized voting in Pennsylvania may be drawn from it.

The best way to characterize these graphs is that they show that there is an *aggregate*, non-linear relationship between the percentage of White voting age population and Republican candidates vote shares in precincts across various regions in the state. However, this is not even an ecological regression analysis, which as discussed above is not considered a reliable way to estimate group voting behavior (see, for example, Goodman 1953 and 1959, Freedman et al. 1991, and King 1997).

Why is this not ecological regression? Recall that the accounting identity that was used to generate the ecological regression model (Eq. 1, above) was based on *voters*. That is, the total share of a

[3]The figures are unnumbered but they begin on page 7 of his memo.

candidate's vote had to be equal to the the sum of the share of each groups' *voters* who voted for them times the share of the *voters* in that group. However, Dr. Barreto uses Census figures for total voting age population for his analysis. The accounting identity underling ecological regression does not hold in this case because not all eligible citizens vote. Further, the turnout varies systematically by race (see, for example, Ansolabehere, Fraga, and Schaffner Forthcoming). This differential turnout rate further biases the estimates rendering his analysis unreliable. This could be fixed by estimating turnout by group. That is, we would add another vote choice in our accounting identity, not voting or abstaining. However, Dr. Barreto did not do this.

Suppose that we ignored the problem of using total population instead of voters by assuming that there were no differences in turnout by racial group, would Dr. Barreto's analysis be a statistically valid ecological regression analysis? Unfortunately, there are still other fundamental statistical flaws in his analysis. Most prominently, there is strong reason to doubt the constancy assumption that underlies ecological inference in this case. Recall that the constancy assumption is that there is no systematic variation across precincts in how a given group votes. So, for example, White voters in predominately non-White neighborhoods vote the same as Whites in neighborhoods with few other minority voters. In general, this is considered a heroic assumption (see Robinson 1950 and Gelman et at. 2001), but there are two reasons to specifically doubt that it holds in Dr. Barreto's analysis.

First, all of the graphs in Dr. Barreto analysis racially polarized voting analysis, include what he call "regression lines" — to suggest that they are ecological regression analysis discussed above — that are suppose to show the relationship between the Republican vote share and the percent White voting age population the precinct. These are the red and blue lines that summarize the points in the graphs. However, this is not the regression line defined for ecological regression developed by Goodman (1950). In the case of ecological regression analysis the the regression line must be a straight line — i.e., linear. That is, the only difference (on average) in vote share for a candidate in a given precinct can only be driven by its demographic makeup. This comes from the accounting identity decomposing the total percent of the votes as the sum of the votes coming from each constituent group given in Eq. 1. Instead, the lines in Dr. Barreto's graphs are locally weighted regression lines (Cleveland and Devlin 1988), also referred to as LOWESS or LOESS lines. They allow one to visually detect non-linear relationships in scatter plots.

In fact, the graphs and their LOWESS regression lines clearly show non-linear relationship in every election Dr. Barreto examined. In particular, as a precinct becomes closer to homogeneous (100%) White on the right-hand side of graph, the Republican share of the vote drastically increases (or correspondingly the blue line showing the Democratic vote declines dramatically). The only way this can happen is if the probability a White voter in the more White districts vote for Republicans at higher rates than their counterparts in more mixed districts. Yet, this is a direct violation of the constancy assumption.

The second reason to doubt the constancy assumption is that Dr. Barreto's analysis lumps all minority groups, for example, Blacks and Hispanics, into one group, Non-White, in his graphs. This is presumably done to make the graphs easier to read as they would need be three dimension with three groups (Black, Hispanic, and White). Also it solves the problem that I discussed above, that there are no homogeneous Hispanic districts in the state. However, I do note that in many of his analysis there are no homogeneous non-White districts even when you combine all minority voters into a single group.

The justification that Dr. Barreto gives to group Hispanic and Black voters together is that they both overwhelming support Democrats by citing evidence from exit polls available from CNN. However, if we actually examine the exit poll results for the 2020 Presidential election by race on CNN, we see that 92% of Black respondents report voting for Biden whereas only 69% of Hispanics report voting for him (ignoring the statistical uncertainty in these estimates). While it is the case that a majority of both groups supported Biden in the election, it is at very different rates. How does this relate to the constancy assumption? Let us assume for the sake of argument that these estimates are exactly correct for the groups' population level of support. Since Dr. Barreto combines Black and Hispanics into the non-White group, the expected level support for Biden in precincts that are mostly Black will be close to 92% whereas precincts that are more Hispanic will be closer to 69%. Thus the only way the level of a non-White group support for the Democratic candidate to be constant is if the ratio of Black to Hispanic voters is constant across all precincts in his analysis. However, this is demonstrably false and thus implies that the constancy assumption is also false for the non-Whites in his analysis. However, if Dr. Barreto were to separate out Hispanics and Blacks, as he should have done, we run into the problem discussed above that there are no homogeneous Hispanic precincts making the estimates of their voting behavior suspect.

Another concern with Dr. Barreto's analysis is that it focuses almost exclusively on statewide offices, which are referred to as *exogenous* elections in redistricting litigation. These are generally considered less informative about a group's voting behavior than examining elections for which maps are being drawn, which are referred to as *endogenous* elections.[4] Presumably, this is because outside of the Philadelphia area most state legislative elections have relatively small numbers of precincts per district making estimates of voting behavior imprecise.

When Dr. Barreto does examine the endogenous elections, he does so by lumping all elections into one graph. Unfortunately, this is never done in a racially polarized voting analysis. By grouping elections across districts in his analysis, Dr. Barreto is assuming that a vote for the Democratic candidate in one legislative or Congressional district is the same choice as voting for the Democratic candidate in another. However, this is simply not true. In different districts, voters choose between different Democratic and Republican candidates that clearly vary along many qualities, for example, incumbency, race, popularity, etc. This is why political scientist always analyze legislative elections separately or if they do some sort of combined analysis, they control for observable differences across races. Dr. Barreto did not do any such correction for systematic differences in his analysis.

[4] I note this is not the sense the words exogenous and endogenous are used in statistics.

6 References

Ansolabehere, Stephen, Bernard L. Fraga, and Brian F. Schaffner. Forthcoming. "The CPS Voting and Registration Supplement Overstates Minority Turnout." *Jounral of Politics*. Available at: https://www.journals.uchicago.edu/doi/10.1086/717260.

Cleveland, William S. and Susan J. Devlin. 1988. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* 83(403):596–610

Duncan, Dudley and Beverly Davis. 1953. "An Alternative to Ecological Correlation." *American Sociological Review* 64:610–625

Flanigan, William H. and Nancy H. Zingale. 1985. "Alchemist's Gold: Inferring Individual Relationships from Aggregate Data." *Social Science History* 9(1): 71–91.

Freedman, David, Stephen P. Klein, Jerome Sacks, Charles A. Smyth, Charles G. Everett. 1991. "Ecological Regression and Voting Rights." *Evaluation Review* 15(6): 673–711.

Gelman, Andrew, David K. Park, Stephen Ansolabehere, Phillip N. Price, and Lorraine C. Minnite. 2001. "Models, assumptions and model checking in ecological regressions." *Journal of the Royal Statistical Society, Series A*. 164: 101–18.

Goodman, Leo 1959. "Some Alternatives to Ecological Correlation." *American Journal of Sociology* 64:610–625.

Goodman, Leo 1953. "Ecological Regressions and the Behavior of Individuals." *American Sociological Review* 19:663–664.

King, Gary. 1997. *A Solution to the Ecological Inference Problem*. Princeton, NJ: Princeton University Press.

King, Gary, Rober O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Robinson, W.S. 1950. "Ecological correlations and the behavior of individuals." *American Sociological Review*, 15(3):351–357

Schuessler, Alexander A. 1999. "Ecological Inference." *PNAS* 96(19): 10578–10581.

A Curriculum Vitae

Jonathan N. Katz

D.H.S.S. (228-77)
California Institute of Technology
Pasadena, CA 91125
(626)395-4191
e-mail: jkatz@caltech.edu

Education

Ph.D. University of California, San Diego. Political Science, June 1995.

M.A. University of California, San Diego. Political Science, June 1992.

S.B. Massachusetts Institute of Technology. Applied Mathematics June 1990.

Academic Appointments

California Institute of Technology:

Kay Sugahara Professor of Social Sciences and Statistics, January 2012 – Present.
Professor of Social Sciences and Statistics, June 2009 – December 2011.
Professor of Political Science, November 2003 – May 2009.
Associate Professor of Political Science, April 1998 – August 1998 and July 2000 – October 2003.
Assistant Professor of Political Science, July 1995 – March 1998.

University of Chicago:
Assistant Professor of Political Science, September 1998 – June 2000.

Harvard University
Post-Doctoral Fellow in Positive Political Economy, July 1994 – June 1995.

Other Employment

Principal, Katz Statistical Consulting,
January 2000 – Present.

Co-Founder and Chief Data Scientist, Adaptivo Inc,
June 2017 – December 2018.

Scientific Advisor, Global Consequences Inc.,
October 2014 – January 2016.

Statistical Advisor, Dispute Resolution Data, LLC.,
August 2015 – September 2016.

Honors and Awards

Elected Fellow of the American Academy of Arts and Sciences, 2011.

Elected Inaugural Fellow of the Society for Political Methodology, 2008.

Center for the Advanced Study in the Behavioral Sciences Fellowship, 2005–2006.

John M. Olin Foundation Faculty Fellow, 1999–2000.

National Science Foundation Graduate Research Fellow, 1991–1994.

Publications

Books

Elbridge Gerry's Salamander: The Electoral Consequences of the Reapportionment Revolution. (with G. Cox). New York: Cambridge University Press. 2002.

Articles in Refereed Journals

Government Partisanship, Labor Organization and Macroeconomic Performance: A Corrigendum (with N. Beck, R.M. Alvarez, G. Garrett, and P. Lange). *American Political Science Review.* 87(4):945–949. 1993.

What To Do (and Not To Do) with Times-Series–Cross-Section Data in Comparative Politics (with N. Beck). *American Political Science Review.* 89(3):634–647. 1995.

Careerism, Committee Assignments and the Electoral Connection (with B. Sala). *American Political Science Review.* 90(1):21–33. 1996.

Why Did the Incumbency Advantage in U.S. House Elections Grow? (with G. Cox). *American Journal of Political Science.* 40(2):478–497. 1996.

Nuisance vs. Substance: Specifying and Estimating Time-Series–Cross-Section Models (with N. Beck). *Political Analysis.* 6:1–36. 1996.

Taking Time Seriously: Time-Series–Cross-Section Analysis with a Binary Dependent Variable (with N. Beck and R. Tucker). *American Journal of Political Science.* 42(4):1260–1288. 1998.

A Statistical Model for Multiparty Electoral Data (with G. King). *American Political Science Review.* 93(1):15–32. 1999.

The Reapportionment Revolution and Bias in U.S. Congressional Elections (with G.Cox). *American Journal of Political Science.* 43(3):812-840. 1999.

Post-stratification without population level information on the post-stratifying variable, with application to political polling (with C. Reilly and A. Gelman). *Journal of the American Statistical Association.* 96(453):1–11. 2001.

Throwing Out the Baby With the Bath Water: A Comment on Green, Yoon and Kim (with N. Beck). *International Organization.* 55(2):487–498. 2001.

A Fast, Easy, and Efficient Estimator for Multiparty Electoral Data (with J. Honaker and G. King). *Political Analysis.* 10(1):84–100. 2002.

The Mathematics and Statistics of Voting Power (with A. Gelman and F. Tuerlinckx). *Statistical Science.* 17(4): 420-435. 2002.

Standard Voting Power Indexes Don't Work: An Empirical Analysis (with A. Gelman and J. Bafumi). *British Journal of Political Science.* 34: 657–674. 2004.

Indecision Theory: Quality of Information and Voting Behavior (with P. Ghirardato). *Journal of Public Economic Theory.* 8(3): 379–399. 2006

Comment on 'What To Do (and Not To Do) with Times-Series–Cross-Section Data in Comparative Politics' (with N. Beck). *American Political Science Review.* 100(1):676–677.

Gerrymandering Roll-Calls in Congress, 1879-2000 (with G.W. Cox). *American Journal of Political Science.* 51(1):108-119. 2007.

Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments (with N. Beck). *Political Analysis.* 15(2):182–195. 2007.

Comment on 'Estimating incumbency advantage and its variation, as an example of a before-after study'. *Journal of the American Statistical Association.* 103(482):446–448. 2008.

Correcting for Survey Misreports using Auxiliary Information with an Application to Estimating Turnout (with G. Katz). *American Journal of Political Science.* 54(3):815–835. 2010.

An Empirical Bayes Approach to Estimating Ordinal Treatment Effects (with R.M. Alvarez and D. Bailey). *Political Analysis.* 19(1):20–31. 2011.

Implementing Panel Corrected Standard Errors in R: The pcse Package (with D. Bailey). *Journal of Statistical Software.* 42(1):1–11. 2011.

Modeling Dynamics in Time-Series-Cross-Section Political Economy Data (with N. Beck). *Annual Review of Political Science.* 14:331–352. 2011.

Estimating Partisan Bias of the Electoral College Under Proposed Changes in Elector Apportionment (with A.C. Thomas, A. Gelman, and G. King). *Statistics, Politics and Policy.* 0:1–13. 2012.

Of Nickell Bias and Its Cures: Comment on Gaibulloev, Sandler and Su (with N. Beck and U. Mignozzetti). *Political Analysis.* 22:274–278. 2014.

An Audit of Political Behavior Research (with J. Robison, R.T. Stevenson, J.N. Druckman, S. Jackman, L. Vavreck). *SAGE Open.* 2018:1–14. 2018.

Constitutions of Exception: The Constitutional Foundations of the Interruption of Executive and Legislative Function (with M. McCubbins). *Journal of Institutional and Theoretical Economics.* 174(1):77–98. 2018.

How to Evaluate Measures of Partisan Fairness for Legislative Redistricting (with G. King and E. Rosenblatt). *American Political Science Review.* 114(1): 164–178. 2020. `https://doi.org/10.1017/S000305541900056X`

Hidden Donors: Analyzing the Censoring Problem in U.S. Federal Campaign Finance Data (with R.M. Alvarez and S. Kim). *Election Law Journal.* 19(1):. 2020. `https://doi.org/10.1089/elj.2019.0593`

The Essential Role of Statistical Inference in Evaluating Electoral Systems (with G. King and E. Rosenblatt). *Political Analysis.* Forthcoming

Other Articles

Empirically Evaluating the Electoral College (with A. Gelman and G. King) in A. Crigler, et al (editors), *Rethinking the Vote: The Politics and Prospects of American Election Reform.* New York: Oxford University Press. 2004.

Detecting Electoral Fraud: The Case of 2002 General Election in Georgia (with R.M. Alvarez) in R.M. Alvarez, T.E. Hall, and S.D. Hyde (editors), *Election Fraud: Detecting and Deterring Electoral Manipulation.* Washington, DC: Brookings. 2008.

Fraud or Failure? What Incident Reports Reveal about Election Anomalies and Irregularities (with D.R. Kiewiet, T.E. Hall, R.M.Alvarez)in R.M. Alvarez, T.E. Hall, and S.D. Hyde (editors), *Election Fraud: Detecting and Deterring Electoral Manipulation.* Washington, DC: Brookings. 2008.

Machines Versus Humans: The Counting and Recounting of Pre-scored Punchcard Ballots (with R.M. Alvarez, E.K. Hartman, and S. Hill) in R.M. Alvarez, L. Atkeson, and T.E. Hall (editors), *Confirming Elections: Creating Confidence and Integrity through Election Auditing.* Palgrave Macmillan. 2012.

What's Age Got to Do with It? Supreme Court Appointees and the Long Run Location of the Supreme Court Median Justice (with M. Spitzer). *Arizona State Law Journal.* 46(1):41 - 88. 2014.

Other Professional Activities

Deputy Editor for Social Sciences, *Science Advances*
March 2018 – Present.

Co-Editor, *Political Analysis*
January 2010 – December 2017.

Member, Expert Panel on Measles Mortality Estimates, World Health Organization, 2004.

Member, Caltech/MIT Voting Technology Project,
October 2003 – Present.

Recent Expert Witness Cases

Rep. Antonio Maestas et al. v. Diana Duran (2012, New Mexico State District Court)

Rene Romo, et al. v. Ken Detzner, and Pam Bondi (2013, Florida Circuit Court)

Diego v. City of Whittier (2014, Superior Court of the State of California, County of Los Angeles)

Jim Soliz, et al. v. Santa Clarita Community College District (2014, Superior Court of the State of California, County of Los Angeles)

Bethune-Hill, et al. v. Virginia State Board of Elections, et al. (2015 and 2017, U.S. District Court for Eastern District of Virginia)

Luna, et al. v. County of Kern, et al. (2017, U.S. District Court for Eastern District of California)

Bruni v. Huges (2020, U.S. District Court for the Southern District of Texas)

Miller v. Huges (2020, U.S. District Court for the Western District of Texas)

Casey v. Garner (2020, U.S. District Court for the District of New Hampshire)

Clarno, et al. v. Fagan (2021, Oregon Circuit Court, Marion County.)

January 6, 2022

Today the Pennsylvania Legislative Reapportionment Commission heard testimony from redistricting experts on the preliminary maps. Among those experts was Dr. Michael Barber of Brigham Young University, employed by the PA House GOP to provide statistical techniques demonstrating that the preliminary House map is “an extreme partisan gerrymander.”

Fair Districts PA is submitting testimony from Dr. Constantine Gonatas, a consultant focusing on data science and simulations, uses sophisticated computer methods to accelerate predictive analysis. His recent statistical analysis to quantify energy reliability for the US Department of Defense was awarded the DoD “project of the year award.” He has developed software to draw maps without human intervention optimized for partisan balance and minimal county splits.

A Pennsylvania native, Dr. Gonatas has followed the work of Fair Districts PA with great interest. He served as a judge for the FDPA 2021 mapping contest, providing many hours of map analysis and has contributed analyses of the LRC House map (attached).

In his testimony, Dr. Barber, the House GOP witness, suggested that the LRC map is “an extreme partisan gerrymander.” Dr. Gonatas duplicated the same kind of random analysis with a far larger statistical sample than Barber’s. Cross-modeling several different ways, his work repeatedly produces Democratic seat shares consistent with the LRC preliminary plan.

According to Dr. Gonatas’ analysis, “the ensemble data show that the LRC plan is within the normal range of the existing, biased Pennsylvania political geography… Prof Barber’s results showing 93-96 House Democratic seats would be extreme outliers for randomly sampled maps according to multiple methods, including those limiting county splits to feasible plans.” Dr. Barber’s recommendation does not satisfy a basic requirement that with 50% of the votes a party should capture about 50% of the legislative seats.

Furthermore, Dr. Barber’s analysis does not protect minority voters as required by the Voting Rights Act. Dr. Gonatas calls attention to Dr. Barber’s failure to acknowledge the dynamic between voter packing, Voting Rights Act requirements and municipal splits:

“Expert witness Barber makes an extensive statement that numerous municipalities are split unnecessarily, violating the state requirement that districts be compact and not split municipalities without justification. Specifically, he calls into question districts for Allentown, Lancaster, Reading, Harrisburg and Scranton. To follow his recommendations and consolidate these areas into fewer districts would deprive these cities of some seats by packing Democratic voters into fewer districts with supermajorities, potentially at cross purposes with the Constitutional requirement of “free and fair elections.” Moreover, in many cases such consolidation would deprive minority voters an opportunity to elect a representative of their choice, in violation of the Voting Rights Act."

Dr. Gonatas' analysis complements testimony from other witnesses at today's hearing and offers additional detail on some of these subjects. It also complements Campaign Legal Center PlanScore analysis. According to five different PlanScore metrics, the LRC House map falls within the range of balanced plans, although well on the right of exact balance.

Of interest to Fair Districts PA has been the bias trend across the decades of LRC activity. While the proposed LRC House map goes far to unravel two decades of increasing partisan bias, it does not meet the lower bias scores of the 70s, 80s and 90s.

The proposed plan is NOT an outlier, not a partisan gerrymander, and was clearly not drawn to deliver an overwhelming advantage to either political party.

Dr. Gonatas's analysis is attached along with some graphics from his work and from PlanScore results.

This will also be submitted on the LRC Comment site. Thank you!

To: Chairman Mark Nordenberg, Pennsylvania Legislative Reapportionment Commission
From: Dr. Constantine Gonatas, CPG Advisors, *Data Science Consulting*
Re: Voting Rights Act compliance in Pennsylvania
Submitted by Dr. Carol Kuniholm, Chair, Fair Districts PA, January 14, 2022

Summary: I present a partisanship analysis of the preliminary LRC General Assembly plan several different ways. First, I show the inherent bias from "seat/ vote curves" indicating how many republicans and democrats would be elected in the LRC plan at various statewide vote strengths, seeing especially if 50% of the votes translates into 50% of the seats. These indicate the LRC plan contains some bias towards Republicans.

Ensembles containing hundreds of thousands of random district plans are compared to the LRC plan, showing the LRC plan is no outlier but in the middle of the partisanship distribution.

Racial balance for certain urban districts is tabulated, suggesting that these often contain significant minority populations, thus consolidation into fewer districts risks losing minority influence districts.

Introduction

I am a consultant focusing on data science and simulations, with specialized expertise in redistricting, map-analysis and Monte Carlo simulations. I received my PhD in physics from the University of Chicago in 1990 with emphases on data analysis and characterization of incomplete datasets, with a dissertation on astrophysical data.

Recently, maps I submitted in the Princeton Election Group were winners of their Gerrymandering competition. Markov chain statistical analysis I performed for the US Department of Defense to clarify energy reliability were awarded the DoD "project of the year award."

I have submitted testimony and expert witness statements in legal and regulatory proceedings, including before the Federal Energy Regulatory Commission, and in a case before US Court of Appeals for the DC circuit and US Supreme Court in *Federal Energy Regulatory Commission vs. Electric Power Supply Association.*

As a consultant I've developed methods to accelerate software using parallel processing and optimization for political mapping. For example, I've used computers to draw maps without human intervention optimized for partisan balance and low county splitting[1] using the Tufts MGGG Gerrychain software[2] and the "Recom" algorithm for map randomnization[3]. I've performed statistical analyses using voter registration and election data files. I also use mapping software such as QGIS and Maptitude to analyze political data. In earlier positions, I developed technologies using machine learning for solar energy forecasting and optimization algorithms for the most economical use of energy storage given a forecast. As a member of the data analytics group of an energy company, I

---

[1] http://www.cpg-advisors.net/districtsim.php/ shows examples of PA State Senate & PA General Assembly maps with low seat and votes bias

[2] Mggg.org

[3] Recombination: A family of Markov Chains for redistricting (Deford, D., Duchin, M.; Solomon, J.) Data Science Review 3(1), 2021

developed simulations to analyze and manage economic risks of large energy projects and advise on corporate investment decisions.

Pennsylvania Election Data and the LRC Preliminary House Map

Census data was downloaded from the Pennsylvania Redistricting website, selecting data set #2 (with prisoner reallocation)[4] to determine precinct populations. This differs in a moderate degree from analyses eg. "Daves' Redistricting" using unadjusted Census data.

Election data was obtained from the Harvard Dataverse[5], a compendium of precinct-level results covering statewide contests from 2016-2020. I also performed runs using data from the MGGG state data repository[6] for Pennsylvania, including 2012 and 2014 elections. Statewide results are a better predictor of partisanship than district races, which in many cases are skewed by unopposed candidates.

In Table 1 I list the statewide democratic 2-party vote fraction for elections from 2012-2020. Primarily this analysis focuses on the most recent contests (2016-2020) since those are most likely to shape outcomes stretching till 2030 however I tabulate 2012 & 2014 for reference.

Mean democratic vote fraction for 2016-2020 is 52.6%. However, this includes blow-out Democratic re-elections for Governor and Senate (2018), skewing partisanship analysis. Excluding one of these contests [Senate] reduces the mean vote share to 52.0%, excluding both results in 50.6% share. By comparison, expert witness Prof. Barber performed analysis covering all 2012-2020 contests including Auditor & Treasurer, with a mean Democratic voteshare of 52.8%, thus biased slightly more to Democrats than the 2016-2020 analysis.

Table 1: Statewide Partisanship

Statewide Democratic Voteshare	
Atty general 012	57.4%
Senate 2012	54.6%
President 2012	52.7%
Governor 2014	54.8%
President 2016	49.6%
Atty general 016	51.4%
Senate 2016	49.3%
Senate 2018	56.7%
Governor 2018	58.7%
Atty general 020	52.3%
President 2020	50.6%
mean 2012-2020	53.5%
mean 2016-020	52.6%
mean 2016-2020 ***excluding*** Senate 2018	52.0%

[4] https://www.redistricting.state.pa.us/maps/
[5] https://dataverse.harvard.edu/dataverse/electionscience
[6] https://github.com/mggg-states

mean 2012-2020 including auditor, treasurer	52.8%

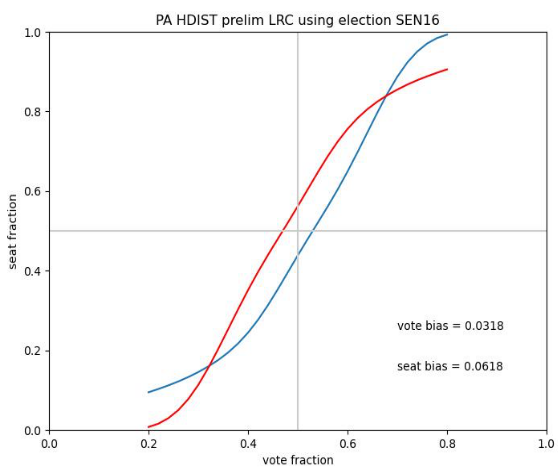
For 2020, average statewide Democratic vote share was 51.5%. Republicans won 113 State House contests (90 for Democrats) thus Republicans obtained a 55.7% State House majority with 48.5% of the statewide voteshare. Prof. Barber refers to the extensive literature indicating that in many states, including Pennsylvania, Democratic voters cluster inefficiently in cities, allowing Republicans to earn more seats despite having equal or even (as in the present case) fewer votes[7]. I have also performed analysis showing that Pennsylvania, along with several other states, has a political geography that is naturally biased towards Republicans, with a particular focus on the effect of reducing county splits on this bias[8].

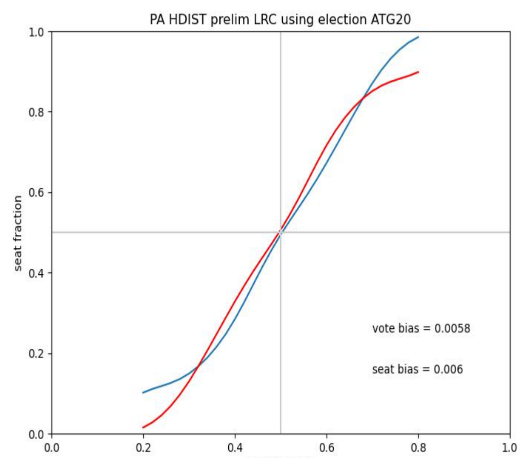
Simulation methods over random ensembles do not take race into account, thus they do not reflect districts protecting minority voters as required by the Voting Rights Act. While the simulation methods used in this submission contain the partisan bias embedded in Pennsylvania's changing political geography, I do not conclude as implied by Barber that this bias *must* be present in an equitable plan.

Measures of Bias and the LRC Preliminary Plan

A fundamental view of the equity in a plan is shown by the "seats/ votes curve." These curves show, for each party, the legislative seatshare obtained if each precinct chooses a legislative candidate in the same proportion as for a historical statewide election. For brevity I show two exemplary charts for only two electoral contests then summarize results for all seat share curves in a table.

Figure 1: seat/ vote curves computed for two proxy elections





The red curve displays seat share for Republicans on the vertical axis as a function of different, hypothetical vote shares in statewide general elections [horizontal axis]. The blue curve shows similar

---

[7] Rodden, Jonathan A. Why cities lose: The deep roots of the urban-rural political divide. Hachette UK, 2019; Stephanopoulos, N. O. and McGhee, E. M., Partisan Gerrymandering and the Efficiency Gap, The University of Chicago Law Review 82: 831-900, (2015); Chen, J. and Rodden, J., Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures, Quarterly Journal of Political Science 8: 239-269, (2013)

[8] Gonatas, C. P. (2021) https://arxiv.org/abs/2103.01735

seatshare data for Democratic legislators, only with the horizontal axis showing hypothetical statewide vote share for Democrats. That is, at a particular point on the horizontal axis, for example 0.5, the curves show the respective share of the General Assembly held by the respective parties following an election where each party received equal votes. Each panel uses a different "proxy" general election to represent precinct electoral result distribution.

If the two curves meet at 0.5 vote share, we have the "democratic ideal that a party attracting about 50% of the popular vote, also ought to be winning about 50% of the contested seats.[9]" The closer the curves are together – providing symmetry in outcomes at different respective vote shares for each party – the less the deck is stacked in favor of one party over another. Numerous analytics can be derived from these seat share curves[10].

The simplest metrics derived visually from the seat share curves are the vote and seat bias . As a convention I compute these for the Democratic party[11], defining the vote bias as the fraction above (or below) 0.5 vote share required for Democrats to achieve a 0.5 seat share, and the seat bias is the additional fractional seat share required by the Democratic party, achieving a 0.5 vote share statewide, to reach 0.5 seat share. Positive values indicate a handicap for Democrats to achieve parity[12].

The above charts show that using the 2016 Senate election as a proxy for precinct voting, the Preliminary LRC plan would tilt heavily towards Republicans but using the 2020 Attorney General's election instead, the Preliminary plan would provide only a slight edge to the Republican party.

Table 2 – Vote & Seat Biases for the LRC Preliminary Plan

LRC Preliminary Plan: seat and vote biases (+ favors republicans)		
Sample Election	Vote Bias	Seat Bias
President 2016	0.026	0.039
Senate 2016	0.031	0.062
AG 2016	0.026	0.050
Governor 2018	0.012	0.019
Senate 2018	0.016	0.026
President 2020	0.009	0.013
AG 2020	0.006	0.006
average	*0.018*	*0.031*

The average vote bias covering 2016 – 2020 elections in the LRC Preliminary plan is 0.018 and the average seat bias is 0.031. With the existing state wide 2016-2020 Democratic vote share (52.6%), the

[9] Mark Nordenberg opening statement at Dec 16 2021 meeting of Legislative Reapportionment Commission

[10] Nagle, J. F and Ramsay, A. (2021) "On Measuring Two-Party Partisan Bias in Unbalanced States," Election Law Journal vol 20 p. 116; Nagle, J. F. (2017) "How Competitive Should a Fair Single Member Districting Plan Be?" Election Law Journal vol 16, p. 196

[11] Choice of convention here considering the Democratic party is the oldest party, founded by Andrew Jackson in 1828, the Republican party being founded in 1854 by Abraham Lincoln and others leaving the fracturing Whig party together with free-soil Democrats

[12] Positive values for bias are analogous to positive values in handicap sports betting, eg. point spreads, to equalize an even betting proposition

Plan indicates Democrats could achieve a bare majority, discounting electoral advantages for incumbents, but excluding the lop-sided 2018 Governor's and Senate races, Democratic vote share (50.6%) indicates the 0.6% excess over 50% would not be sufficient to achieve a majority, other factors being equal. Thus these data on bias show the LRC Plan is favorable to Republicans.

Table 3a – LRC Plan Metrics (2016-2020 elections as proxy)

LRC Plan Metric	Value
Dem Seats won	105.57
efficiency gap	0.0376
mean-median	0.009
Polsby-Popper	0.3382

Table 3b – Democratic House Seats Won in LRC Plan vs. Statewide Proxy Contest

Electoral Proxy	Imputed DEM GA Seats Won
Atty general 2012	129
Senate 2012	111
President 2012	95
Governor 2014	118
President 2016	94
Atty general 2016	101
Senate 2016	84
Governor 2018	130
Senate 2018	119
Atty general 2020	109
President 2020	102

Using the 2016-2020 statewide election precinct results to sample hypothetical legislative returns I obtain the average metrics shown in Tables 3 a& b, to be compared in the memorandum to results from ensembles of random districting plans. The average number of democratic seats won (105.57) is tabulated together with the efficiency gap, mean- median and Polsby-Popper metrics.

The efficiency gap[13] was introduced by political scientists Stephanopoulos and McGhee to quantity the "wasted votes" often accentuated by partisan gerrymandering. Its definition is:

$$Wasted\_votes_i = (vote\ for\ losers)_i + (vote\ for\ winners_i - 0.5 \times all\ votes)$$

$$EG = (Wasted\_votes_D - Wasted\_votes_R)/(total\ votes)$$

That is, all votes for losing candidates are "wasted" as are any extra votes beyond what's needed to win in a districts (0.5 vote share). The efficiency gap is then the proportional difference in wasted votes between the two parties. Positive values indicate more Democrats' votes are "wasted."

[13] Stephanopoulos, N; McGhee, E. (2015) "Partisan Gerrymandering and the Efficiency Gap," University of Chicago Law Review, vol 82, p 831

"Mean – Median" is a simple measure of skew in the vote distribution among all seats in a legislature. I define it as

Mean-median = (average Democratic voteshare statewide) – Democratic voteshare in median seat

where for the 203 seat General Assembly, the median seat is the 102nd most partisan seat. If statewide vote is equal between the parties, mean-median measures the extent voters of one party are packed into a relatively smaller number of districts they win with supermajorities, while losing more broadly. Positive values here indicate Republican bias.

Polsby-Popper measures compactness.

Polsby-Popper = $4\pi \times A/P^2$

where A is the area of a district and P is its perimeter. The larger it is, the larger the district's area is compared to its perimeter, hence more compact. In the unusual case of a district as compact as a circle, Polsby-Popper = 1. The average of all the Polsby-Popper measures over the 203 districts in the LRC plan = 0.3382.

The data in Table 3b show a wide scatter in the expected number of Democratic (Republican) General Assembly wins in the LRC Preliminary Plan, using different statewide elections as proxies for the imputed vote in each district for statehouse candidates, with an average expected number of Democratic House seats of 105.57 using only the 2016-2020 contests. The data in Table 3a show positive values of mean-median and efficiency gap, indicating the LRC Preliminary plan is biased towards Republicans, consistent with the vote and seat biases.

Ensemble Analysis

I further performed simulations of random district plans three different ways to determine if the LRC plan is an outlier vs. a partisan-blind mapper (a 40-core server running massively parallel computations). The purpose of performing ensembles measurements three ways was to cross-check consistency so that a possible deficiency of one method would be backed up by an independent calculcation.

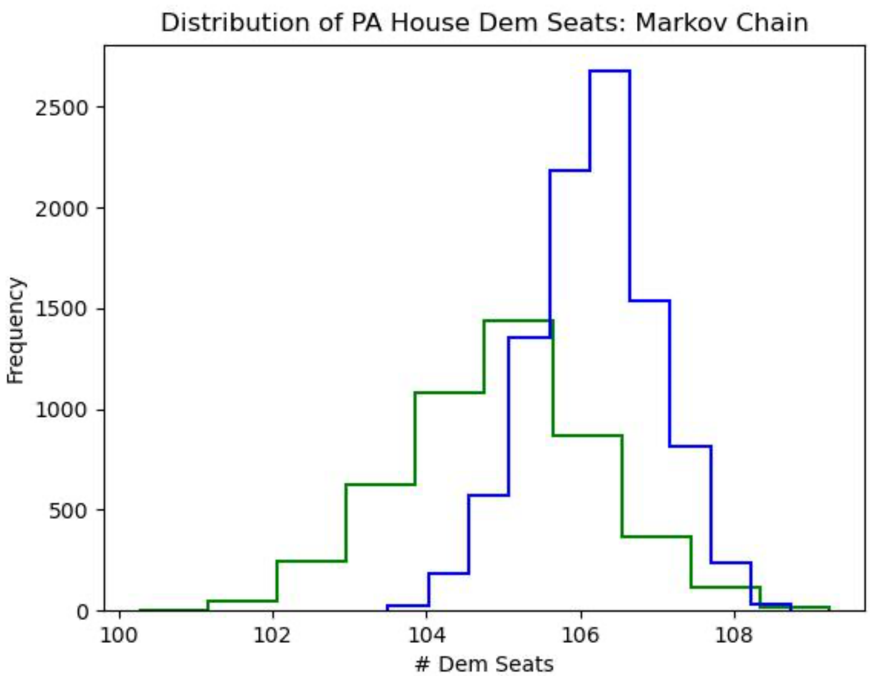
I first used two Markov chain methods taking an initial seed (the LRC preliminary plan) together with the Recom algorithim (DeFord et al. 2021) to propagate a step-wise sequence of alterations where at each step, two adjacent districts are merged then re-split randomly. This is analogous to randomnizing a deck of cards by removing two cards in a single step, then replacing them in the deck at random places. At each step the change from a prior state of the deck of cards to the next state is small, but the deck becomes completely randomnized after a significant number of steps.

Pennsylvania 203 House districts are one of the most difficult plans to analyze because it takes many steps for a chain to evolve from one state to another completely independent step, thus I evolved the chains by hundreds of thousands of steps to obtain snapshots that are independent of each other. Secondly, I constrained this algorithm to only accept states with a fixed number of county and municipal splits. Thirdly, I took a "random unconstrained" approach where instead of starting from an initial state and evolving it step by step, cut a spanning tree covering all of Pennsylvania at random locations into 203 districts with the required population balance[14]. This third method is orders of magnitude slower than the other methods but each state generated is independent of others. This method did not constrain county or municipal splits, however.

[14] "recursive_tree_part" in https://gerrychain.readthedocs.io/en/latest/api.html#spanning-tree-methods

Figure 2: Markov Chain Simulation Histogram of PA House Dem Seats

[blue = constrained 187 county splits, 107 municipal splits; green = no constraints]



In Figure 2 I compare Democratic House seat shares for the two Markov chain methods, with the blue curve tracing the frequency distribution over an ensemble constrained for 187 county and 107 municipal splits, and the green curve covering an ensemble with no splitting constraints but where the ensemble converged to an average of 360 county splits. In both ensembles I select only one out every 50 Markov steps for inclusion in a restricted sample, to allow the chain states to randomnize, as indicated by the correlation function for subsequent values of Democratic seats won. Both chains ran for 480,000 total steps with a population deviation of +/- 5%. I also performed runs with the county split limit set at various higher levels to determine the effect county splits have on other metrics.

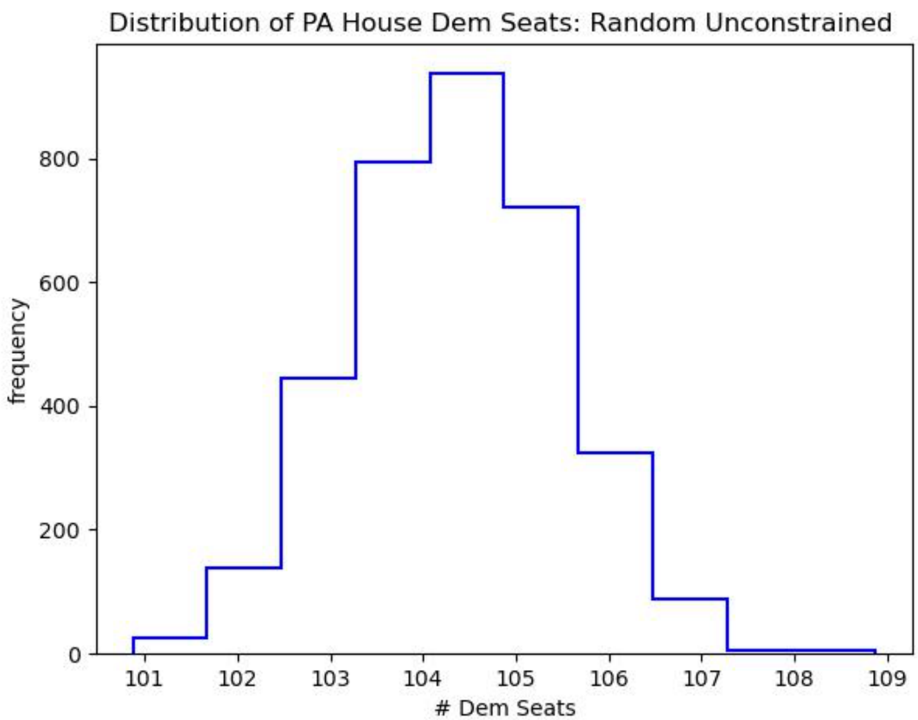
Figure 2 shows that both the county split-constrained and unconstrained Markov chains have a similar distribution only the split constrained distribution has a narrower peak centered on a mean of 106.18 Democratic seats won, and the unconstrained distribution is shifted to a slightly more Republican-favored outcome, with a mean of 104.94 Democratic seats won.

None the expert testimony I are aware of using Markov chains to assess contested plans for outliers constrains county splits, thus scores plans vs a universe of plans that would nearly all be ineligible for consideration as a realistic (or lawful) plan. However, constraining county splits adds the analytic risk that the ensemble could be biased by only exploring a limited region of plans easily accessible by the

Markov chain shuffling. By doing the analysis both way I show this bias is small. I further limit this risk by comparing to a distribution generated by the random tree method[15].

Figure 3: Random Tree Simulation Histogram of PA House Seats

In Figure 3 I show the result for the random tree simulation, covering approximately 2,500 plans.



Without county or municipal split limits, the average number of county splits was 382. Again, the maximum population deviation was +/- 5%. Here the mean number of Democratic seats won was 104.33, only 0.61 seats fewer than its direct comparable, the unconstrained Markov chain.

Table 4: Ensemble Run Metrics

simulation detail	county splits	muni splits	mean seats won	polsby-popper	efficiency gap	mean-median
Markov chain 2016-20	187	107	106.18	0.325	0.040	0.021
Markov chain excluding '18 senate	192	105	104.00	0.323	0.042	0.022
random tree, no county split limits	382	N/A	104.33	0.215	0.045	0.025
Markov chain	195	112	106.30	0.320	0.040	0.019
Markov chain	293	120	105.65	0.300	0.042	0.023
Markov chain	320	N/A	105.60	0.285	0.042	0.023
Markov chain no county split constraints	360	N/A	104.94	0.230	0.044	0.024

Table 4 summarizes metrics measured from the various runs. Notably, the 105.57 Democratic House seats measured in the LRC Preliminary plan using 2016-20 election data is well within the distributions from the ensembles above and biased closer to the Republican party than the county-split constrained

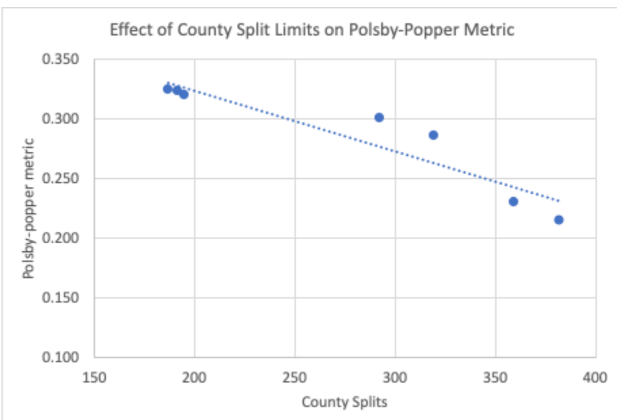
[15] I have developed software that uses the Markov chain method to generate "preferred" plans, for example preferring plans with fewer county splits at each step. Runs of this type have achieved PA House plans with 186 county splits for 2010 demographic data, and as low as 175 county splits when relaxing the population balance requirement to +/- 10% (prior to cleaning maps with a population balance step)

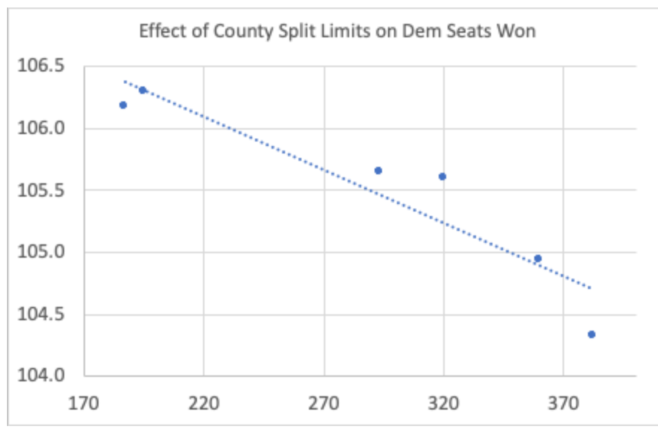
Markov chain run with 187 county splits. The biases of the existing political geography are evident from these runs, which all have similar efficiency gaps, showing Republican bias close to that of the LRC plan. By the standard of Becker et al[16] the ensemble data show that the LRC plan is within the normal range of the existing, biased Pennsylvania political geography.

One case considers 2016-2020 elections *excluding* the 2018 Senate race, reelection of Robert Casey won by a wide margin, thus possibly not representative of true partisanship in the State. Democrats win 104 House seats in this scenario, down by over 2 seats from the base case 2016-2020 election data set. Although I did not perform a run further excluding the Wolf reelection bid in 2018, I extrapolate that would further reduce Democratic seats won, to less than a majority of the General Assembly.

The effect of county splits on compactness and Democratic seats won is shown in Figure 4, where fewer splits unsurprisingly leads to higher Polsby-Popper score (compactness). Surprisingly, fewer county splits leads to to slight increase in Democratic seats won, unexpectedly since intuition suggests that spreading dense Democratic concentrations in Philadelphia and Allegheny counties increases Democratic vote efficiency.

Figure 4 – effect of County Splits on Compactness & Democratic Seats Won





Comparison to other expert witness statements

Prof Barber of BYU has performed ensemble analyses with histograms showing 93-96 House Democratic seats depending on the sample of elections used to generate the ensemble. These would be extreme outliers for the county-split constrained ensemble (as well as for the other ensembles), thus inconsistent with the present analysis. He does not explain how Democratic statewide voteshare of 52.8% (2012-2020) resulting in the 47% seatshare of his analysis may be equitable, even though it conflicts with a basic statement that a majority vote should translate into a majority of seats.

[16] Becker, A., Duchin, M., Gold, D., Hirsch, S. (2021) Election Law Journal vol 20, p. 407: "Normal range, not ideal. We advocate using redistricting ensembles to learn a normal range for metrics and measures under the constraints of a set of stated redistricting rules and priorities. Ensembles allow us to justify statements such as Plan X is an outlier in its partisan lean, taking all relevant rules into account. While talking about normal ranges and outliers, we should avoid the temptation to valorize the top of the bell curve (or its center of mass, or any other value) as an ideal. By analogy, we can talk about people who are unusually tall or short without believing that any height is most desirable or ideal. If the 50th percentile height for American women is 5'4" and the 99th percentile height is 5'10", we can conclude that a woman who is six feet tall is unusual, and we can look for reasons (family history, diet, and so on) to explain her height. But it would be quite strange to decide that a woman who is 5'4" is a better" height than one who is 5'5".

Prof Warshaw assesses various metrics (efficiency gap, declination, mean-median) of the proposed LRC map using PlanScore, concluding the proposed map reduces Republican bias as compared to the existing plan but retains Republican bias. The efficiency gaps and mean-median scores differ from the metrics obtained here but it is not immediately clear when the elections used by PlanScore to compute the scores are the same as in this work, so no direct comparison is possible.

He conducts an “endogenous” analysis, using 2020 votes for State House candidates to assess future elections using the proposed plan. This is distinct from the analysis of this submission, which uses only “exogenous” data, that is statewide election data unbiased by incumbency or the lack of contested elections in many districts. Thus his analysis, while complementary, does not offer a direct comparison to this work.

Municipal Splitting and Racial Balance

Expert witness Barber makes an extensive statement that numerous municipalities are split unnecessarily, violating the state requirement that districts be compact and not split municipalities without justification. Specifically, he calls into question districts for Allentown, Lancaster, Reading, Harrisburg and Scranton. To follow his recommendations and consolidate these areas into fewer districts would deprive these cities of of some seats by packing Democratic voters into fewer districts with supermajorities, potentially at cross purposes with the Constitutional requirement of “free and fair elections.” Moreover, in many cases such consolidation would deprive minority voters an opportunity to elect a representative of their choice, in violation of the Voting Rights Act.

I provide an overview without detailed analysis required for Voting Rights Act litigation. The demographics in Table 5 shows that while Pennsylvania gained population from the 2010 census to the 2020 census, it was not sufficient to retain its 18th Congressional seat. Non-hispanic whites *declined* in absolute numbers and all of Pennsylvania’s growth is due to disproportionate increase in minority voters. Therefore, special consideration is due to ensure their representation in the House.

Table 5: Population Changes 2010 to 2020

Census: 2020	2020	2010	change in population	% pop change
total population	13,002,700	12,702,379	300,321	2.4%
non-hispanic white	9,553,417	10,094,652	(541,235)	-4.3%
minority	3,449,283	2,607,727	841,556	6.6%

Table 6: Population Breakdowns for Urban House Districts:

House District	Total Population	Non Hispanic White	Total VAP	Non Hispanic VAP	Non hispanic white %	Non hispanic white VAP %
Allentown:						
22	62647	19063	46398	16576	30.4%	35.7%
132	63549	39336	50914	33751	61.9%	66.3%
134	63349	27157	48501	23519	42.9%	48.5%
Lancaster:						
50	62727	27149	48625	23760	43.3%	48.9%
96	65891	44763	52589	37879	67.9%	72.0%
Reading:						
126	61746	29800	46467	25594	48.3%	55.1%
127	61291	18784	45077	16310	30.6%	36.2%
129	61096	28988	46248	24093	47.4%	52.1%
Harrisburg:						
103	63950	37143	50957	31527	58.1%	61.9%
104	65021	25065	48832	21620	38.5%	44.3%
105	65356	34966	51817	30182	53.5%	58.2%
Scranton:						
112	62127	54484	50385	45297	87.7%	89.9%
113	61487	43833	49102	37003	71.3%	75.4%
114	61604	51002	49417	42556	82.8%	86.1%
118	62791	54619	51240	45964	87.0%	89.7%

Table 6 breaks down the non-hispanic white populations and voting-age population for the urban districts Barber calls into question for excessive splits. Mostly, the districts in the indicated municipalities have non-hispanic white populations at levels (< 65%) where either minority voters could elect a minority preferred candidate outright, or in coalition with some non-hispanic white voters minorities could influence the outcome.

Expert witness Prof Barreto of UCLA provides testimony showing the presence of racially polarized voting, one of the requirements of the Gingles test under which the Voting Rights Act is adjudicated. However, the Legislative Redistricting Committee is not limited to the bare minimum requirements of the Voting Rights Act; it may draw minority influence districts where population growth and historical under-representation of minority office-holders suggest that would be equitable. Consolidating districts as Barber suggests would likely deprive minority voters of representation.

Dinos Gonatas

(978) 985-4309 cpgonatas@cpg-advisors.net

260 Old Marlboro Road, Concord MA 01742 www.linkedin.com/in/dinosgonatas

Key Skills

- Data science and analysis for actionable results.
- Public policy outreach with high impact filings, journal articles.
- Member of expert witness economics brief at Federal Appeals DC Circuit and Supreme Court.
- Python, Matlab, databases, including machine learning and optimization.

Experience

Principal

CPG Advisors, *Data Science Consulting* (2006-present)

- Finalist in Princeton Election Consortium map competition with a machine-drawn entry for WI State Senate map showing low political bias yet conforming to traditional districting criteria.
- Developed technology for machine-drawn redistricting optimizing partisan bias, county splits. Supervised intern creating features for population balance and geographical smoothing.
- Analyzed Texas election and population data over 8 election cycles to show Gingles minority and anglo bloc voting ecological regression trends by region and year.
- Awarded Department of Defense "Project of the Year" for electric generation simulation covering 5 military bases showing power reliability for solar + storage.
- Intervened in regulatory proceedings for energy. Met with Federal Commissioner. Member of expert witness team. Wrote opinion pieces on mitigating carbon emissions.

Product Manager, promoted to Group Business Development and Intellectual Property Manager

Oxford Instruments, *$250M analytical instrumentation supplier headquartered in Oxford, UK* (2003-2006)

- Conceived patent strategy for analytical instrumentation portfolio. Identified infringers from reverse engineering. Performed due diligence to ensure non-infringement of competitors' patents.
- Delivered presentations to CEO approving litigation in UK, France, Germany, Switzerland. Selected and managed litigation team settled by $30M deal. Developed expertise in comparative legal systems for intellectual property.

Licensing Officer

Massachusetts Institute of Technology (2002-2003)

- Negotiated patent licenses with a LED startup, Luminus Devices, enabling $28M venture round.
- Coached students and post-doctoral fellows on technology commercialization.

Product Manager

Corning, *Lasertron fiber-optic communications components division*, (2001-2002)

- Introduced fiber-optic receiver. Obtained pilot feedback to guide engineering team and product.
- Led team qualifying infrared lasers, complying with requirements and customer acceptance.

Director, Product Development

Panamsat, *$800M satellite communications division of Hughes Electronics* (2000-2001)

- Proposed broadcast service to cellular base stations. Investigated economics and market needs.
- Developed business plans for high bandwidth communications satellites for North America.

Business Development Manager, International (Latin America)

Enron, *international gas pipeline/power plant division, and energy services* (1996-2000)

- Managed & performed power system simulations forecasting price risk in Latin America
- Performed scenario analysis showing manageable power supply risks for $3B acquisition.

Early Career in engineering and finance at ExxonMobil and Tenneco Energy

Education

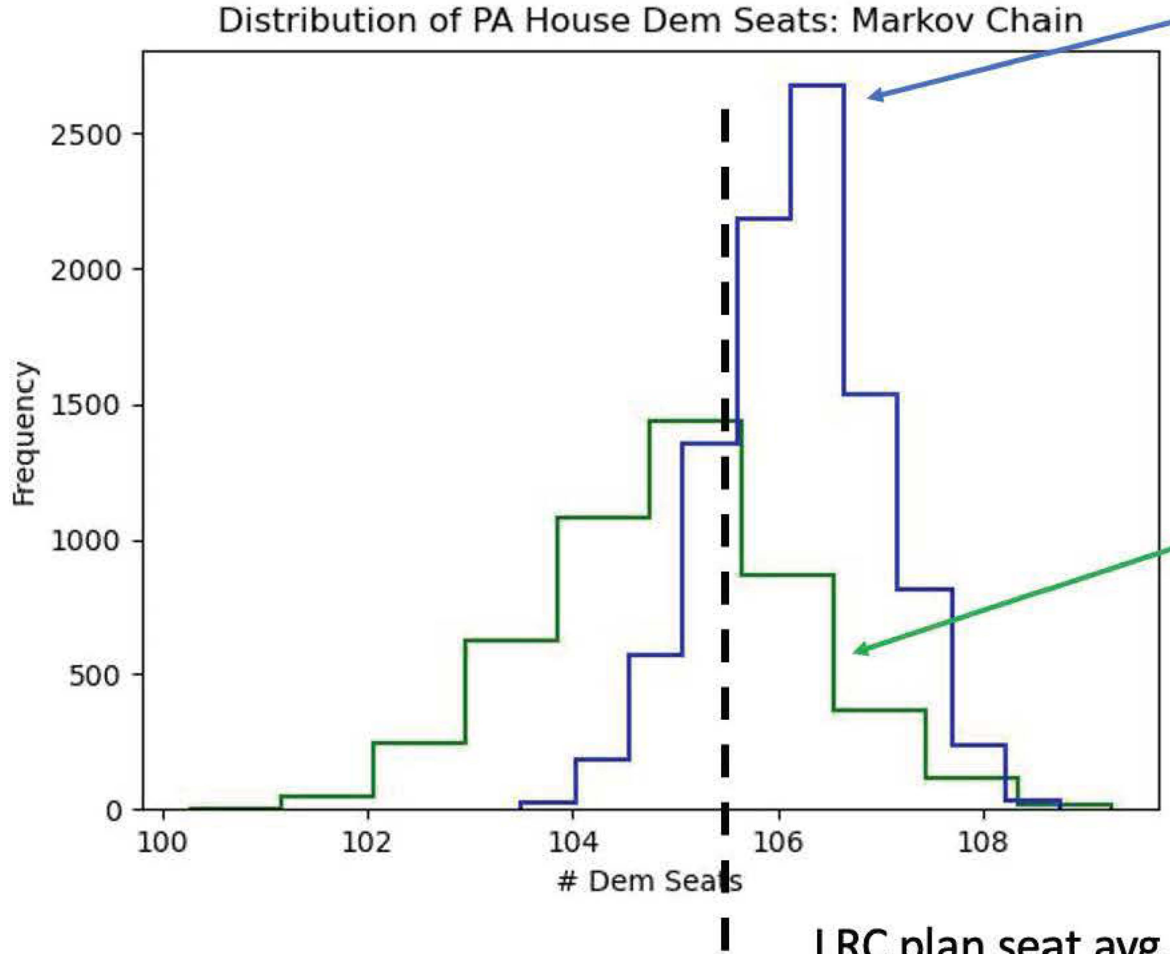
MBA, Babson College
PhD, Physics, University of Chicago
BA, Physics, Princeton University

Biases from all seatshare curves

LRC Preliminary Plan: seat and vote biases (+ favors republicans)		
Sample Election	Vote Bias	Seat Bias
President 2016	2.6%	3.9%
Senate 2016	3.1%	6.2%
AG 2016	2.6%	5.0%
Governor 2018	1.2%	1.9%
Senate 2018	1.6%	2.6%
President 2020	0.9%	1.3%
AG 2020	0.6%	0.6%
average	***1.8%***	***3.1%***

- Vote bias like a point spread, + values =
  extra % of Dem votes needed to achieve 50% seats
- Average vote bias = +1.8%, seat bias = +3.1%
- 👉 R's get 3.1% premium @ 50% vote share
- D's need 51.8% of votes to get 50% of seats

Ensemble Results for Dem Seat Distribution

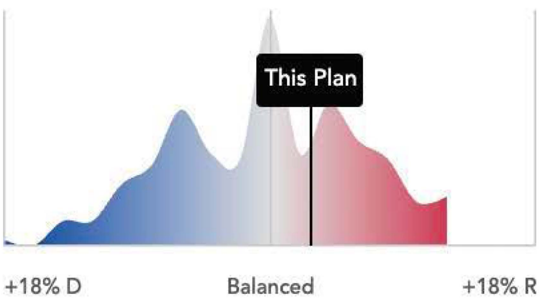


Markov model constrained to 187 & 107 county & muni splits

Markov model w/o split constraints (360 splits on avg)

LRC plan seat avg, using 2016-2020 elections as proxy NOT COUNTING CANDIDATE INCUMBENCY NOR DISCOUNTING STRONG DEM STATEWIDE WINS in 2018

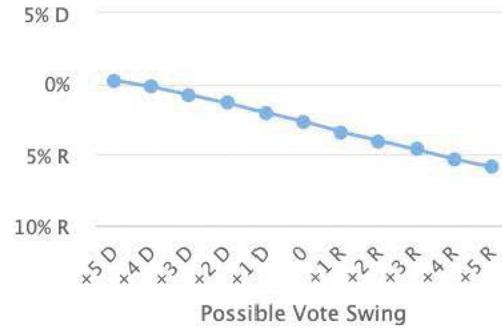
Efficiency Gap: 2.6% R



Votes for Republican candidates are expected to be inefficient at a rate 2.6% R lower than votes for Democratic candidates, favoring Republicans in 88% of predicted scenarios.*

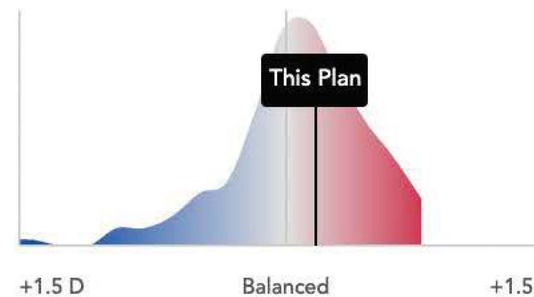
Learn more ›

Sensitivity Testing



Sensitivity testing shows us a plan's expected efficiency gap given a range of possible vote swings. It lets us evaluate the durability of a plan's skew.

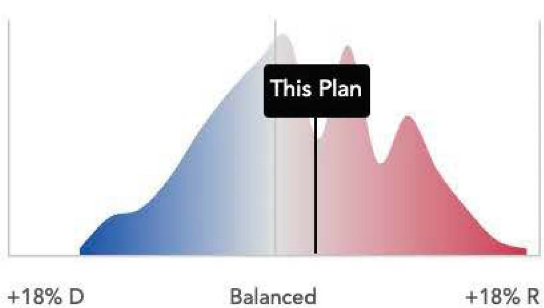
Declination: 0.17 R



The difference between mean Democratic vote share in Democratic districts and mean Republican vote share in Republican districts along with the relative fraction of seats won by each party leads to a declination that favors Republicans in 93% of predicted scenarios.*

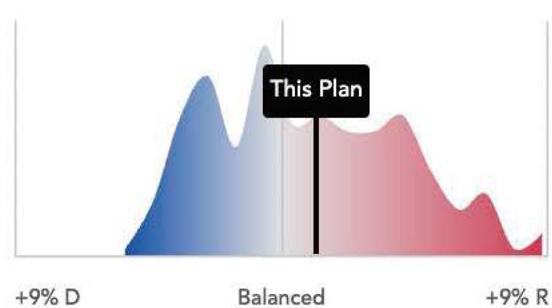
Learn more ›

Partisan Bias: 2.9% R



Republicans would be expected to win 2.9% R extra seats in a hypothetical, perfectly tied election, favoring Republicans in 97% of predicted scenarios.* Learn more ›

Mean-Median Difference: 1.2% R



The median Republican vote share is expected to be 1.2% R higher than the mean Republican vote share, favoring Republicans in 97% of predicted scenarios.* Learn more ›

PlanScore
A PROJECT OF CAMPAIGN LEGAL CENTER

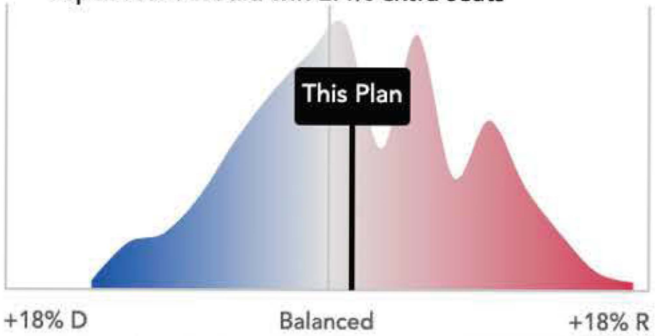
PARTISAN BIAS
IN THE PENNSYLVANIA STATE HOUSE

How many extra seats a party would win in a hypothetical, exactly-even 50-50 election; compiled by Fair Districts PA

1972

+1% Republican

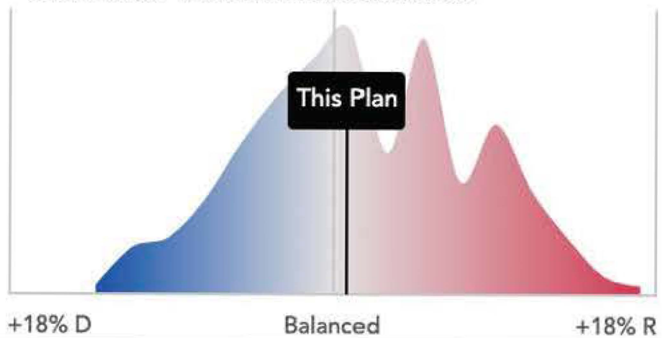
Republicans would win 1.4% extra seats



1982

+1% Republican

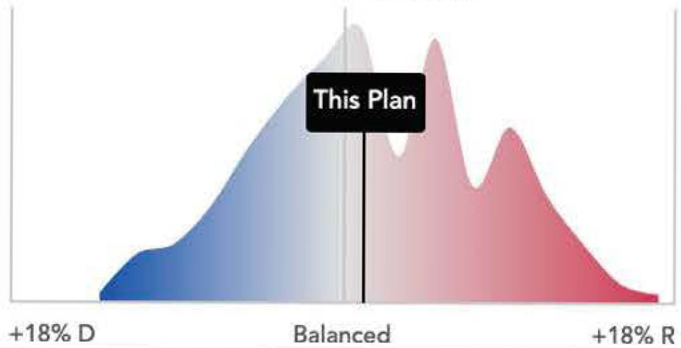
Republicans would win .8% extra seats



1992

+1% Republican

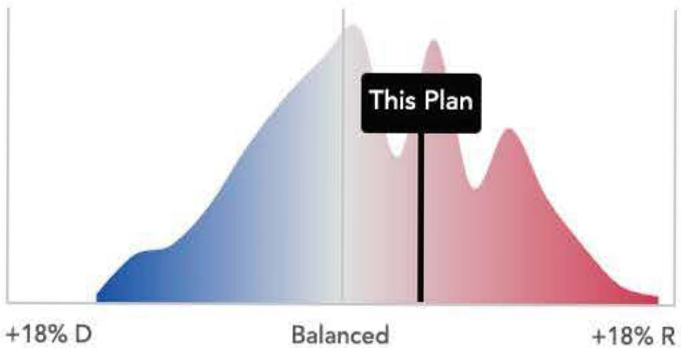
Republicans would win 1.2% extra seats



2002

+4% Republican

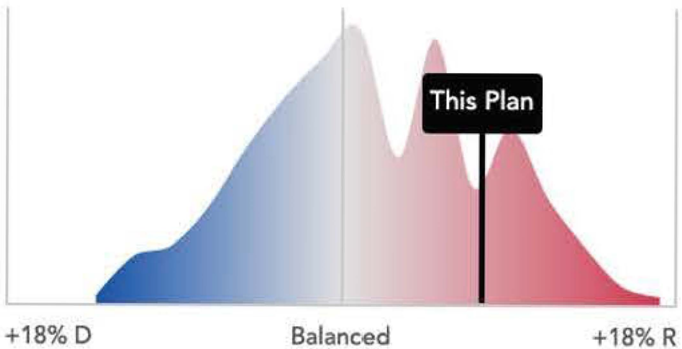
Republicans would win 4.3% extra seats



2012: Current Plan

+8% Republican

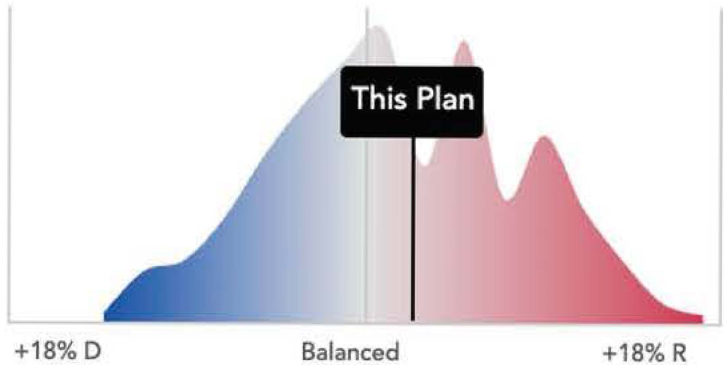
Republicans would win 7.5% extra seats



2022: Preliminary Plan

+2.5% Republican

Republicans would win 2.5% extra seats



Reply to Dr. Katz by Dr. Barreto, January 18, 2022

The report submitted by Dr. Jonathan Katz after the conclusion of expert testimony on January 14, 2022 has no relevance whatsoever to understanding Voting Rights Act issues, or voting patterns in Pennsylvania.

Preliminarily, it is important to underscore what is not in Dr. Katz's report. Dr. Katz does not opine that there is a lack of cohesiveness in minority voting patterns in Pennsylvania or that there is no racially polarized voting in Pennsylvania. In short, he makes no effort to actually use statistical methods to contribute to our understanding of voting patterns in Pennsylvania. His essential premise—that Ecological Inference or EI is not a proper basis for evaluating Latino political cohesiveness in Pennsylvania—is not valid. Ecological inference has been extensively published as a reliable method in political science journals, and has been regularly used, in fact, required, by state and federal courts in adjudicating voting rights lawsuits. That Dr. Katz finds it challenging to use widely accepted scientific methodology to identify and understand racially polarized voting in Pennsylvania is both suspect and unremarkable.

It is also noteworthy that Dr. Katz does not dispute that Black voters in Pennsylvania are very cohesive and clearly meet the second *Gingles* criteria of minority cohesiveness. Further, Dr. Katz does not dispute that White voters in Pennsylvania block vote against minority candidates of choice, meeting the third *Gingles* criteria of bloc-voting. To be clear, there is no meaningful analysis or data in Dr. Katz's report concerning application of the *Gingles* test in Pennsylania.

Dr. Katz's factual assertions with respect to Latinos in Pennsylvania are not accurate. Dr. Katz claims that "there are no homogenous Latino/Hispanic voting precincts in the state." This is not accurate. In fact, there are 23 precincts in Pennsylvania which are over 80% Latino, with an additional 51 precincts which are greater than 70% Latino. Indeed across the state there are 213 precincts which are majority-Latino, and there are a total of 283 precincts in which Latinos are the single largest racial group (i.e. precincts where Latinos make up less than 50% but outnumber Whites, Blacks and all other groups.). I have attached a list of all 283 such precincts with the State House 2020 votes for Democratic and Republican candidates as an appendix.

These heavily Latino precincts provide clear data to support the conclusion that Latino voters in Pennsylvania are cohesive and support Democrats. That data is summarized in Table 1 below.

Table 1: Average State House Democratic Vote 2020 in Heavily Latino Precincts

% Latino Range	Precincts	Avg. Dem vote	Std. Dev	Min	Max
80 – 100%	N=23	97.3%	.0712	.7704	1.00
70 – 100%	N=74	91.8%	.1256	.4588	1.00
50 – 100%	N=213	89.9%	.1613	.3346	1.00
Plurality	N=283	90.3%	.1562	.3356	1.00

In my presentation on January 14, 2022 to the LRC, I reported an estimated Latino vote ranging from 74% to 82% for State House legislative elections across different regions of the state. Dr.

Katz provides no contrary opinions concerning the level of Latino cohesion, instead just opining that ecological inference is complicated and contains uncertainty. It is true that all of social science empirical analysis is complex and there is not perfect certainty. Our job as trained methodologists and social scientists is to use the best known tools and models to reduce uncertainty and provide best available estimates from which we can draw inferences and conclusions. Ecological inference models to determine racial and ethnic voting patterns are a widely accepted and published methodology in political science.

Dr. Katz criticizes the ecological inference method in his report, but he has not published any peer-reviewed academic articles finding that ecological inference is unreliable. His opposition to ecological inference runs counter to an abundance of social science published research[1] that support ecological inference as an appropriate tool for estimating racial and ethnic voting patterns and its accepted use by voting rights experts in evaluating whether or not racially polarized voting exists.

Other critiques made by Dr. Katz are also baseless. He includes a section claiming that King's EI model is generally wrong and does not work well in elections with multiple racial groups or multiple candidates (Sections 3.3 and 3.4). All analyses in my report however are based on two-candidate elections. Dr. Katz criticizes King (1997) and advocates for the use of Rosen et al. (2001), yet the full analysis that I conducted and reported is based on eiCompare software package which uses *both* King and Rosen, and allows the analyst to *compare* how they perform. In published research,[2] we have twice demonstrated that the King and Rosen methods are highly correlated with one another and both provide accurate results. In this case, my analysis relies on both the King and Rosen approaches to ecological inference within the eiCompare software package.

---

[1] Grofman, Bernard. 1991. "Statistics without substance: A critique of Freedman et al." Evaluation Review, 15: 746-769; Lichtman, Alan. 1991. "Passing the Test." Evaluation Review. 15, 770-799.; Tanner, Martin. 1996. Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions, 3rd Ed., Springer, New York; King, Gary. 1997. A Solution to the Ecological Inference Problem. Princeton University Press, King, Gary, Ori Rosen and Martin Tanner. 1999. "Binomial-Beta hierarchical models for ecological inference" Sociological Methods and Research, 28: 61-90; King, Gary. 1999. "The Future of Ecological Inference Research: A Comment on Freedman et al." Journal of the American Statistical Association Vol. 94, No. 445 (Mar., 1999), pp. 352-355; Rosen, Ori, Wenxin Jiang, Gary King, Martin Tanner. 2001. "Bayesian and frequentist inference for ecological inference: the RxC case" Statistica Neerlandica. 55:2; Grofman, Bernard and Samuel Merrill. 2004. "Ecological Regression and Ecological Inference." In Gary King et al., eds. Ecological Inference: New Methodological Strategies. Cambridge University Press.; Barreto, Matt 2007. "Si Se Puede! Latino Candidates and the Mobilization of Latino Voters." American Political Science Review. 101 (August); Grofman, Bernard and Matt Barreto. 2009. "A Reply to Zax's (2002) Critique of Grofman and Migalski (1988): Double Equation Approaches to Ecological Inferences" Sociological Methods and Research. 37 (May); Collingwood, Loren, Kassra Oskooii, Sergio Garcia-Rios, and Matt Barreto. 2016. "eiCompare: Comparing Ecological Inference Estimates across EI and EI: RxC." The R Journal. 8:2; Imai, Kosuke and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records" Political Analysis; Barreto, Matt, Loren Collingwood, Sergio Garcia-Rios and Kassra Oskooii. 2019. "Estimating Candidate Support: Comparing Iterative EI and EI-RxC Methods" Sociological Methods and Research. 48(4).

[2] Collingwood, Loren, Kassra Oskooii, Sergio Garcia-Rios, and Matt Barreto. 2016. "eiCompare: Comparing Ecological Inference Estimates across EI and EI: RxC." The R Journal. 8:2; Barreto, Matt, Loren Collingwood, Sergio Garcia-Rios and Kassra Oskooii. 2019. "Estimating Candidate Support: Comparing Iterative EI and EI-RxC Methods" Sociological Methods and Research. 48(4).

In Section 4 of his report Dr. Katz suggests that ecological inference is not appropriate for Pennsylvania due to lack of homogeneous precincts, but the only evidence he relies on is data from California. He provides no relevant evidence in this section that my conclusions concerning Pennsylvania are inaccurate.

Curiously, to support his argument that ecological inference is not valid for understanding Latino voting patterns in Pennsylvania, Dr. Katz relies primarily on Latino party registration rates in Kern County, California from years ago. Even here, his analysis of Latino party registration in Kern County, California is deeply flawed and did not properly consider voters who had no party registration which is why his model over-estimated Democratic registration. In the case Dr. Katz cited, *Luna v. County of Kern*, 291 F. Supp. 3d 1088 (E.D. Cal. 2018), Dr. Katz's critique of ecological inference was rejected by the court. With respect to "Dr. Katz's critiques," the decision states: "[T]he court is unpersuaded that these criticisms preclude plaintiffs from demonstrating Latino political cohesiveness by a preponderance of the evidence." The court found "no basis to conclude that there is some minimum number of homogenous precincts required before [ecological regression] and [ecological inference] analysis have any probative value" in a VRA case. The court noted that Dr. Katz himself admitted that the political scientist who developed ecological inference (Gary King) "indicated no bright line percentage of homogenous precincts is necessary in order for ecological inference estimates to be reliable." The court in *Luna* further noted that, in addition to the lack of support for Dr. Katz's position "in the field of statistics, numerous cases finding racial polarization have relied on statistical analyses that did not include HPA [homogenous precinct analysis] and made no mention of homogenous precincts whatsoever." Finally, the court found that "Dr. Katz's insistence on 'sufficient' homogenous precincts is undercut by his own work in previous cases, where he performed [ecological regression] and [ecological inference] analyses without any reference to the number of homogenous precincts in the relevant jurisdiction." The *Luna* court held that Dr. Katz's critique did "not raise a doubt sufficient to refute" plaintiffs' expert's conclusion that racial polarization existed.

Given that a federal judge so soundly dismissed Dr. Katz's theory concerning homogenous precincts, the Commission should question why such a debunked theory was offered at the very last moment. The late submission suggests that proponents of Dr. Katz's report held it until the 11th hour to shield both Dr. Katz and his report from fair examination and scrutiny.

Dr. Katz claims on page 10 of his report that my analysis "focuses almost exclusively on statewide offices." However in every single instance I provide vote pattern results and estimates for State House elections, which would be considered the endogenous elections. To provide the Commission with *more* information and evidence, I also include voting patterns in additional elections. As Dr. Warshaw stated in his testimony on January 14, 2022, it is a longstanding conclusion in Political Science that voting in State Legislative elections is highly correlated with Presidential elections and elections for other major partisan statewide offices. If Dr. Katz's argument were correct, we would not observe a strong correlation between statewide elections and district elections, but the correlation is undeniable. Excluding elections in which a candidate ran unopposed, the correlation between Democratic vote for State House and Democratic vote for Attorney General at the precinct level is 0.9448 in 2020. The correlation between

Democratic vote for State House and Democratic vote for U.S. President at the precinct level is 0.9233 in 2020. Thus, as decades of political science literature suggest, these elections are highly consistent.

There was a question at the hearing concerning consideration of primary elections, but neither court precedent nor peer-reviewed political science literature require an evaluation of primary election results to draw conclusions about racially polarized voting. Indeed, in this instance, we are interested in whether or not Whites, Blacks, and Latinos vote for the same or different candidates to represent them and the most probative elections here are general elections where voters are choosing which candidates to send to the state Capitol. While primary election data may be instructive in case specific situations where minorities do not vote in coalition, there is strong evidence in Pennsylvania that minorities do vote in coalition. Moreover, as the Chairman noted, the preliminary plan creates a number of open seats which will provide opportunities for minority candidates in primary elections.

To be clear, the precinct scatterplots are presented because they are illustrative examples of voting patterns that are clear, concise and easy to interpret. However, the full ecological inference models are based on the eiCompare which estimates models for Whites, minorities-combined, Blacks, and Latinos, using both ecological methods advocated by King and Rosen et al.

In summary, the report by Dr. Katz concedes much and does not offer any evidence or data on voting patterns in Pennsylvania. Dr. Katz only raises generic critiques of ecological inference which have been debunked by courts and in the social science literature. Read in full, his report is non-responsive and offers nothing that in any way detracts from the well-supported conclusions and opinions in my report.

Matt A. Barreto
Matt A. Barreto
January 18, 2022
Agoura Hills, California