

Applications of Ensemble Analysis to Gerrymandering in Minnesota and Texas

Eva Airoidi, Oliver Calder, Tom Patterson, Antonia Ritter, Rebekah Stein

Fall 2021, Winter 2022

Abstract

Every ten years, after a national census, all states must redraw a legislative districting map with the apportioned number of districts. Gerrymandering—methods by which political parties may attempt to gain electoral advantages by modifying district boundaries—is becoming more common in these redistricting processes. While districting maps can be, and have been, deemed illegal in court, there is no rigorous process for determining the extent to which gerrymandering is present in a map. In our analysis, we utilize the existing framework of ensemble analysis with some modifications to quantify and contextualize the level of gerrymandering present in 2010, 2020, and proposed 2020 maps for Minnesota and the 2010 map for Texas. We generate a representative sample of all legal districting plans for the given state to identify whether the plan of interest is an outlier on a number of physical, demographic, and partisan metrics of gerrymandering. Additionally, through ensemble analysis, we discover whether underlying characteristics of a state—such as population distribution, geography, etc.—result in a natural bias towards a particular party. We find Minnesota has a slight bias towards benefiting Republicans in our ensemble. For some metrics, the proposed plans appear within the expected range of scores; however, there is slight evidence of cracking and packing of Democrats, as well as of Black and Hispanic voters. Our Texas results show a Republican bias. The 2010 Texas map has multiple measures that lean in favor of the Republican Party. There is some evidence of cracking and packing of Democrats, as well as strong evidence of packing and cracking Black and Hispanic voters in the 2010 map.

1 Introduction

We begin by defining what we mean by gerrymandering.

Definition 1.1. **Gerrymandering** is the practice of drawing electoral districts to benefit one group (political party or racial/ethnic group). **Political or Partisan Gerrymandering** is drawing boundaries that benefit one party, while **Racial Gerrymandering** involves diluting the voting power of members of ethnic or linguistic minority groups. [1]

The media continues to claim that gerrymandering is becoming more present. With the redrawing of every state’s maps after the 2020 census, many sources have claimed that states are accepting maps which are even more intensely gerrymandered than the previous maps, with many unfairly favoring the Republican party. Nonetheless, court challenges to gerrymandered maps have become even more difficult.

In 1986, the Supreme Court ruled that extreme partisan gerrymandering is unconstitutional [2, 3].

That is, districting is illegal if it “consistently degrades a voter’s or a group of voters’ influence in the political process as a whole” [1]. However, the court did not agree on a constitutional standard for evaluating partisan gerrymandering [1, 3], and the Supreme Court has repeatedly failed to strike down districting maps, despite clear use of classic gerrymandering tricks [3]. In the following years, Supreme Court decisions went back and forth regarding whether gerrymandering could be settled by the court itself [2, 3]. In 2004, the Supreme Court rejected all available tests for gerrymandering, leaving no official method for accounting whether a districting map was fair or not [2]. In 2019, the court ruled that “partisan gerrymandering claims present political questions beyond the reach of the federal courts” [1].

We attempt to implement a tool which better identifies and visualizes how gerrymandered a particular map actually is. Many states, due to their population and geographic distribution, have underlying bias towards one major party or the other. We use a method which is able to detect and quantify

how biased a typical legal map is for a given state due to these factors. We are then able to determine how gerrymandered a particular map might be beyond this baseline bias.

2 Background Info and Definitions

2.1 History of Gerrymandering

The term “gerrymander” arose in 1812 when the governor of Massachusetts, Elbridge Gerry, enacted a law that altered the state’s senatorial districts and caused the Federalist Party’s vote to be underrepresented. After a satirical cartoonist, Elkanah Tisdale, noticed one of these newly outlined districts resembled a salamander, he published a cartoon version of the district in the Boston Gazette. He titled the cartoon “The Gerrymander.” [1, 5]

Gerrymandering intensified when Black men earned the right to vote. The Black population was often crammed into one district in each state, leaving the remaining districts predominantly white and the Black population underrepresented in the state’s congressional delegation [6].

2.2 Redistricting Process

Every ten years, the U.S. Census Bureau conducts a national census. Based on the population results, each state is apportioned a certain number of representatives. Then, each state is divided into this number of districts, where each district has close to the same number of residents. Each representative will be elected by the individuals living in their particular district.

The exact process of redistricting varies by state. Some states have apolitical committees to draw their new map, while in other states, the state legislature will draw the map. Since we focus our analysis on two states, we will discuss the particular procedures in these two states.

2.2.1 Minnesota Redistricting

In Minnesota, the State legislature is responsible for the redistricting process. The legislature – with the assistance of the Nonpartisan Legislative Coordination Commission – will create draft congressional and legislative redistricting plans in the form of bills. The plan must then pass as a typical bill would with a simple majority. Finally, the governor must sign the proposed plan into law. The deadline for having

an approved map through this avenue is 25 weeks before the election.

In practice, however, Minnesota’s divided legislature typically cannot pass a redistricting plan. In this situation, once the 25-week deadline is reached, the court is responsible for imposing a new plan. Since 1980, every congressional map in Minnesota has been drawn in this way by the courts. As a result, we will analyze three of the court-proposed plans for Minnesota, including the approved plan. While Republicans and Democrats in the legislature have also proposed maps, it is historically unlikely that these maps would be enacted into law. [7]

In Minnesota, two plans were proposed as bills by Democrats, and one plan was proposed as a bill by the Republicans. We did not analyze any of these maps, however, since they are extremely unlikely to be passed into law. When we began writing this paper, Minnesota had four court-proposed maps—the Wattson, Anderson, Sachs, and Corrie Plaintiff Plans. Just last week, Minnesota approved a map for the new cycle, which is largely a combination of the Wattson and Sachs plans. As a result, we will analyze the Wattson and Sachs Plaintiff Plans as well as the new 2020 map for Minnesota.

2.2.2 Texas Redistricting

As in Minnesota, the Texas State Legislature is responsible for drawing the state’s congressional maps. Unlike Minnesota, however, both bodies of the Texas legislature are controlled by Republicans, so a map has always been approved through this process. [8]

A new 2020 map was approved for Texas before we began analysis. Without access to 2020 census data, however, we determined that analyzing the 2010 Texas map would lead to more valuable results.

2.3 Legal Requirements for Districts

Each state varies in their exact requirements for what makes a district legal. However, in general, there are four main characteristics that districts across all states must follow: 1) districts must have equal population, 2) districts must be as compact as possible (see Section 3.1), 3) districts must not divide existing political subdivisions wherever possible, and 4) districts must be compliant with the Voting Rights Act (see Section 2.3.1).

Many states have recently begun noting that “communities of interest” must not be divided between multiple districts. There is no rigorous definition for what these communities of interest are,

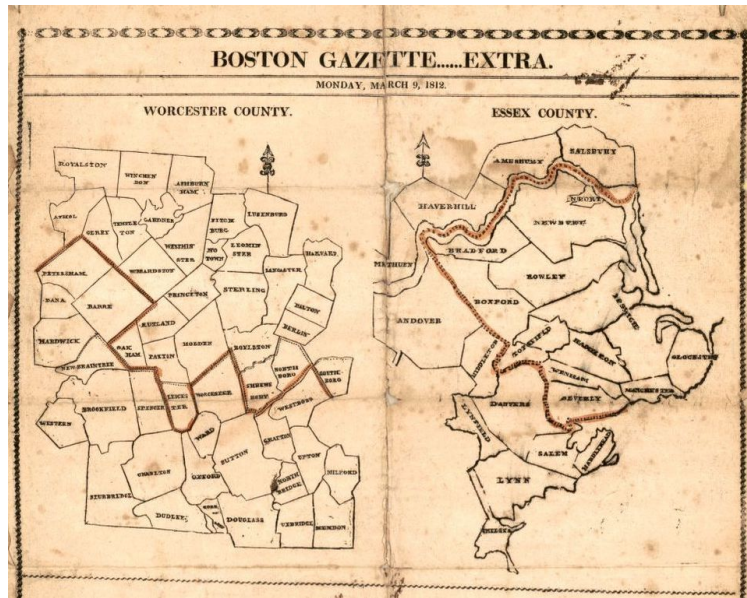
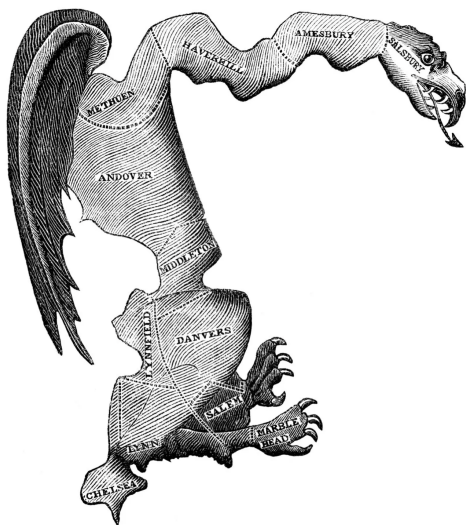


Figure 1: Images from the redistricting plan published in the Boston Gazette on March 9, 1812. [4]



(a) "The Gerry-mander" by Elkanah Tisdale



(b) Actual map of the Gerry-mander

Figure 2: Massachusetts State senate district drawn in 1812.

however. Racial and ethnic groups—such as the Somali population in Minneapolis—are considered communities of interest. Sometimes, smaller political groups or groups with some shared identity are also considered communities of interest to be maintained in one district.

Some states impose additional requirements. This year, Minnesota notes that the reservation lands of federally recognized American Indian tribes could not be split into multiple districts if there was a possible way to keep them together [9].

2.3.1 Voting Rights Act

The Voting Rights Act of 1965 (VRA) aimed to eliminate racial discrimination in voting. Beyond provisions such as outlawing literacy tests as a requirement to vote, the VRA also took several steps to prevent racial gerrymandering in the redistricting process. Specifically, redistricting plans must not result in the “abridgement of the right of any citizen of the United States to vote on account of race or color,” and violations occur if “it is shown that the political processes leading to nomination or election are not equally open to participation by members of a class of citizens 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.” [10] The VRA thus provides an avenue to challenge redistricting plans in court on the basis of racial gerrymandering. However, in *Shelby County v. Holder* (2013) the Supreme Court ruled section 4(b) of the VRA unconstitutional, making it considerably harder to challenge redistricting plans for potential VRA violations. [11]

2.4 Cracking and Packing

Two of the fundamental ideas around gerrymandering are packing and cracking. Packing and cracking are both methods designed to manipulate the distribution of voters in order to minimize the state-wide voting power of the opposing party.

Definition 2.1. We say a political party or racial/ethnic minority group is **packed** into one or more districts when that district is composed largely of that group.

Definition 2.2. We say a political party or racial/ethnic minority group is **cracked** into several districts when each of these several districts contains a small percentage of individuals in this group.

Note that both of these methods can lead to gerrymandering. If one political party is packed into just a few districts, while they will easily win these few districts, the remaining districts may be won by only a small margin by the other political party. Similarly, if one political party is cracked into several districts, they will have just under a majority in each district, and they will be unable to win many seats. Packing and cracking can be used to create maps which give a far higher proportion of seats to one party than their proportion of votes won.

In reality, cracking and packing are often used in tandem. One party may crack a large proportion of the other party into one district which they will win by a large margin. Then, the remaining constituents for the opposing party will be cracked into the remaining districts. As a result, the cracked/packed party will win one district by a large margin and lose several districts by only a small margin. [13]

3 Existing Metrics for Quantifying Gerrymandering

3.1 Compactness

One of the simplest ways to quantify the level of gerrymandering in a redistricting plan is to evaluate the geometric compactness of districts. Here, the term compactness refers not to a particular mathematical definition, but rather, the normal connotation of the word.

Definition 3.1. We consider a district to be **compact** if it is regularly shaped without significant protrusions.

There are numerous ways to calculate the geometric compactness of a shape, each with their own benefits and drawbacks. We now look at three compactness measures, and provide our argument for the use of the Polsby-Popper measure in our ensemble analysis.

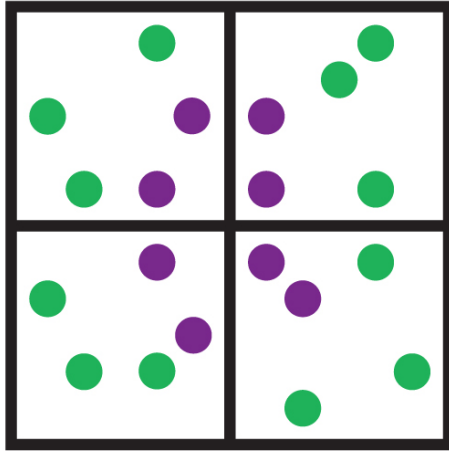
3.1.1 Length-Width Measure of Compactness

Definition 3.2.

$$LW(d) = \frac{\min(l_d, w_d)}{\max(l_d, w_d)}$$

The **Length-Width** measure of compactness is the ratio of the length to the width of the minimum bounding rectangle around the district. [14]

**Cracking Communities
so they can't elect their own
representative.**



**Packing Partisans
into one district so the other
party wins adjacent districts.**

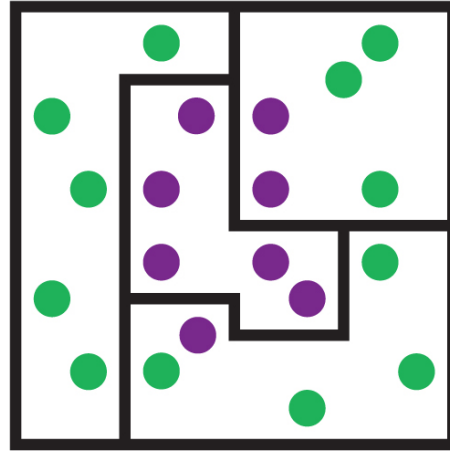


Figure 3: Consider the theoretical state containing four districts and with 12 individuals in the green party and eight in the purple party. In drawing these four districts, the purple party can be cracked between all four districts, winning no districts, or packed into one district, winning just that district. [12]

Scores closer to one mean that the minimum bounding rectangle is closer to being a square, a possible indication that the district is compact. Accordingly, a score closer to zero indicates a non-compact district. The Length-Width measure utilizes the relative compactness of a square in comparison to a rectangle, and has been used during the redistricting process by several Western states that have straight borders. However, the complex borders of the Eastern states demand a more robust measure.

3.1.2 Reock Measure of Compactness

Definition 3.3.

$$R(d) = \frac{A_d}{A_{MBC}}$$

The **Reock** measure of compactness is the ratio of the area of the district to the area of the minimum bounding circle around the district.

Again, scores close to one are an indication of the compactness of the district, while scores close to zero indicate non-compactness. While the Reock measure does utilize the circle's ability to pack more area into a perimeter than any other shape, the measure is susceptible to protrusions angled in different ways producing different compactness scores, depending on whether the angle of the protrusion expands the

diameter of the minimum bounding circle.[15] Since a protrusion in a district to include some population group can indicate gerrymandering, we desire a compactness measure that is unaffected by the angle of the protrusion.

3.1.3 Polsby-Popper Measure of Compactness

Definition 3.4.

$$PP(d) = \frac{4\pi A_d}{P_d^2}$$

The **Polsby-Popper** measure of compactness, sometimes called the isoperimetric quotient, is the ratio of the area of the district to the area of a circle with circumference equal to the perimeter of the district.

Since the Polsby-Popper measure relies only on the perimeter and area of the district itself, all protrusions are treated equally, regardless of the angle they protrude from the district.[15] For this reason, Polsby-Popper is a better measure of compactness than both Reock and Length-Width, and we accordingly use it in the construction of our ensemble.

Compactness measures aim to penalize strangely shaped districts that pack or crack certain political, racial, or socioeconomic groups. However, com-

pactness measures are affected by the oddities of American geography (Maryland’s panhandle, for example), so non-compact districts are not inherently gerrymandered. Conversely, a district can be gerrymandered while still being compact. This brings us to non-geometric measures of gerrymandering.

3.2 Efficiency Gap

One way to measure whether a map has packing and cracking is to compare the number of “wasted votes” between the parties. We define what it means for votes to be “wasted” as follows.

Definition 3.5. A vote is considered to be a **wasted vote** if it is cast for the winning party in a winning district and beyond the 50% threshold required to win a district or if it is cast for the losing party in a district.

Definition 3.6. The **Efficiency Gap** is a measure which describes the difference in quantity of wasted votes between parties (in this case the Democratic and Republican parties). For a given plan ε , the efficiency gap $eg(\varepsilon)$ is defined as

$$eg(\varepsilon) = \frac{\text{Wasted}_R(\varepsilon) - \text{Wasted}_D(\varepsilon)}{\text{Total Votes Cast in } \varepsilon}.$$

Observe that, according to this definition, a negative efficiency gap implies that the Democratic party incurs more wasted votes than the Republican party, while a positive efficiency gap implies the opposite. [16]

3.3 Seats-Votes Curve

3.3.1 Generating a Seats-Votes Curve

The seats-votes curve is another powerful tool used to detect gerrymandering in a redistricting plan. The curve plots the proportion of seats won by a party in a Congressional election against the proportion of the state-wide vote won by that party. To demonstrate how we construct a seats-votes curve, first consider the following definitions:

Definition 3.7. A party’s **Vote Share** in a given election is taken to be the percentage of votes cast for that party. When we simplify an election to consider only votes for two parties (Party A and Party B), the vote share for Party A is equal to

$$\frac{\text{Votes for Party A}}{\text{Votes for Party A} + \text{Votes for Party B}}$$

Definition 3.8. A party’s **Seat Share** in a given election is taken to be the percentage of seats that party won in the election. In all of our considered elections, no seats were won by third party candidates, so we did not need to make simplifications to account for the fact that there are more than two parties.

Definition 3.9. Under **Uniform Party Swing**, we assume that if the vote share for one party increases by one percent, then within each district, the vote share for that party increases by one percent as well.

The observed results from the election of choice (that is, the actual seat share and vote share from this election) provide one point for the seats-votes curve. The rest of the curve must be generated. To do this, we rely on uniform party swing defined above. Beginning at the point of the real election, we either add or take away one percent of the vote from the party we’re drawing the seats-votes curve for. We then check, given our election data and relying on uniform party swing, how many seats that party would have won. We continue in this matter, adding and removing from the vote share for that party, until we have produced a full step-function which shows at which vote shares the party moves from winning x seats to $x + 1$ seats. It is helpful to illustrate this process for a hypothetical election.

Suppose the actual results of a Congressional election in a state with ten districts give party A 60% of the seats with 60% of the vote. Now, uniformly remove votes from party A across all districts until party A has only 59% of the statewide vote. If these changes do not result in any seat flipping, then party A would have still won 60% of the seats with only 59% of the vote. However, if the changes do result in a seat flipping, then party A would have only won 50% of the seats at 59% of the vote, and the seat-votes curve would jump down in a step-wise manner. The rest of the seat-votes curve is filled in using this uniform swing method.

Plots of the seats-votes curve often include one curve for each party; however, these curves must be mirror images of each other. That is, if one party receives 20% of the seat share at 30% of the vote share, then the other party must receive 80% of the seat share at 70% of the vote share.

3.3.2 Metrics Computed from Seats-Votes Curve

The seats-votes curve provides a graphical interpretation of several partisan gerrymandering measures.

Simulated Seats-Votes Curve Using Uniform Swing

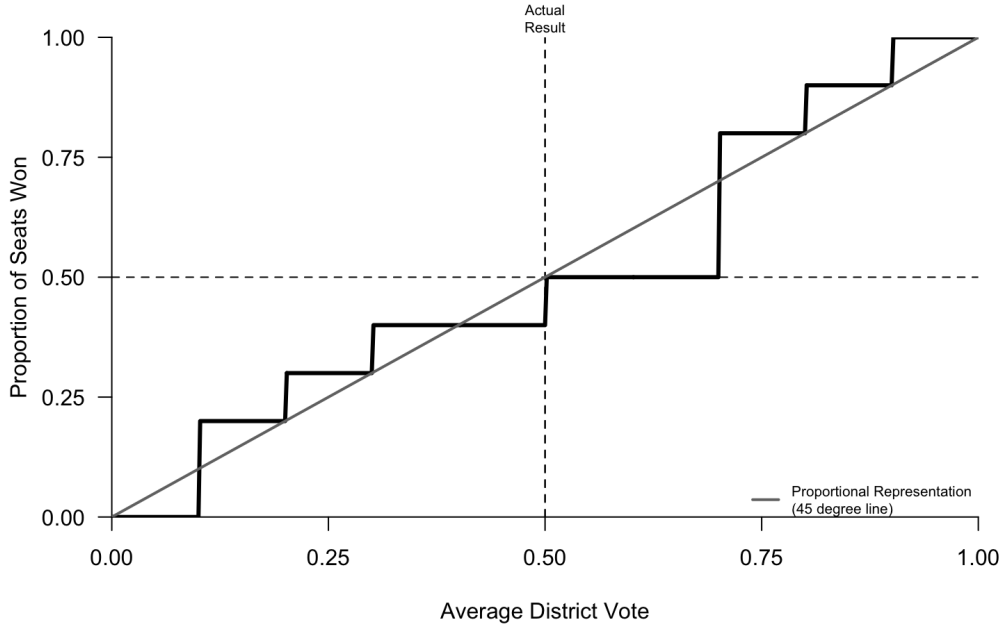


Figure 4: Example seats-votes curve of one party

Definition 3.10. The **Partisan Bias** of a redistricting plan is the difference between a party’s seat share at 50% of the vote and 50%.

Partisan bias can be graphically interpreted as the gap in seat share between the curve at 50% of the vote, and 50% of the seats. While certainly a good initial measure to possibly detect an unfair redistricting plan, partisan bias only compares the curves for both parties at a singular vote share, and indications of political gerrymandering can occur elsewhere.

Definition 3.11. The **Partisan Symmetry** of a redistricting plan is the average distance between the seat-votes curves of each party over a chosen range of vote shares.

Partisan symmetry is most commonly calculated over the interval spanning 45% and 55% of the vote. The seats-votes curve also captures other features of a redistricting plan, beyond just having the potential to indicate an imbalance in seat share relative to vote share. Features like the slope of the seat-votes curve can indicate the competitiveness of the redistricting plan.

Definition 3.12. The **Responsiveness** of the seats-votes curve is the average slope of the curve over a chosen range of vote shares, typically around 45% to 55%.

A steeper slope would indicate that a party could gain significant seat share by only gaining a small amount of the vote share, implying that multiple districts are separated by a narrow vote margin. Conversely, a shallow slope implies that not that many districts are separated by a narrow vote margin, potentially indicating an incumbent protection gerrymander. Therefore, the shape of the seats-votes curve can help reveal the political strategies underlying a redistricting plan.

Going even further, analysis of the end behavior¹ of the seats votes curve can also help identify packing and cracking in a redistricting plan. In a heavily packed district, party A seeks to waste a large number of the party B’s votes by creating a super majority for the opposing party (effectively taking away votes for party B in other districts). The resulting packed district will almost certainly be won by the party B, even at low overall vote shares since most of their votes are being cast in the packed district. Accordingly, the seat-votes curve

¹Behavior near 0% vote share and 100% vote share.

²The curve for party B will jump to seat-share value equal to $\frac{1}{D}$, where D is total number of districts in the state.

for party B will jump up² at a vote share close to zero, leaving evidence of the heavily packed district. Consequently, the seats-votes curve for party A will not reach 100% of the seat share until values of vote share get close to 100%, leaving evidence of party A’s cracking strategy.

3.4 Mean-Median

The mean-median metric also attempts to measure the amount of cracking and packing present in a given map [17]. Whenever we consider a distribution, if the mean and median are different, this implies some skew in our data. If cracking and packing is present in a state, then the median district-wide vote share would be skewed, leading to a nonzero mean-median. This leads to our definition of this measure:

Definition 3.13. The **Mean-Median** for a given districting plan compares the state-wide vote share for one party to the median district-wide vote share for that same party. For a given plan ε , the mean-median $mm(\varepsilon)$ is defined, for a given party A , as the difference between the state-wide vote share of party A and the median district-wide vote share for party A .

A negative value implies that the map favors Party A , while a positive value implies that the map favors Party B [17]. In our analysis, we choose to compute the mean-median using the Republican Party as Party A , meaning a negative mean-median implies that the map favors Republican Party.

A nice interpretation for the mean-median is that $50\% + mm(\varepsilon)$ gives the vote share the Republican Party would need to win to gain 50% of the seats in a given districting plan for a state. Note this relates nicely to the seats-votes curve as well, as it describes the relationship between seat share and vote share at 50% of the vote.

3.5 Partisan Gini

Partisan gini is a score meant to quantify the area between the seats-votes curves of the two parties. As before, let ε denote a particular redistricting map, and let n denote the number of districts in ε . Let $V_{R,i}(\varepsilon)$ denote the Republican vote share of the district with the i^{th} highest Republican vote share, and let $V_{D,i}(\varepsilon)$ denote the Democratic vote share of the district with the i^{th} highest Democratic vote share.

Also, let $V_{R,avg}(\varepsilon)$ denote the average Republican vote share, and let $V_{D,avg}(\varepsilon)$ denote the average Democratic vote share. Note that $V_{R,avg}(\varepsilon) =$

$1 - V_{D,avg}(\varepsilon)$ and, similarly, the Republican vote share of the district with the i^{th} highest Republican vote share is equivalent to 1 minus the Democratic vote share of the district with the $(n-i+1)^{th}$ highest Democratic vote share.

The partisan gini for a given redistricting map ε is computed as follows:

$$pg(\varepsilon) = \sum_{i=1}^n \frac{|(V_{R,i}(\varepsilon) - V_{R,avg}(\varepsilon)) - (V_{D,i}(\varepsilon) - V_{D,avg}(\varepsilon))|}{n}.$$

Since the seats votes curve is a step function, then the partisan gini is essentially the difference between the Republican vote share to win i seats and the Democratic vote share to win i seats, multiplied by the “height” of gaining a seat, for $i \in [1, n]$. The true partisan gini is modified slightly to account for overall vote shares different from 50%, but this description gives an intuitive sense of the metric.

While most metrics produce a negative value if they favor Republicans and a positive value if they favor Democrats, notice that partisan gini is an absolute value of the distance between the curves. That is, a greater distance between the curves indicates a greater difference between seat shares at various vote shares, and thus indicates some sort of inequality between the parties. A partisan gini value of 0 indicates that the seats-votes curves of the two parties are equal, so each party would receive exactly the same number of seats at any given vote share for both parties.

4 Introduction to Ensemble Analysis

All previously mentioned metrics of gerrymandering attempt to measure how gerrymandered a map is in just one number. They do not account for the geographical characteristics of a given state nor the distribution of population throughout the state. In this way, these measures all assume that a value of 0 makes for a “fair” map, ignoring the fact that unintentional bias is nearly always present in several states based on their geography and population distribution alone. Namely, densely-packed cities often have higher concentrations of Democrats, while rural areas tend to be more Republican. This almost always leads to a few districts with a large Democratic majority and several districts with only a slight Republican majority.

This quote from one of the leading mathematicians working on gerrymandering, Moon Duchin, sums up this concept:

Gerrymandering is a fundamentally multidimensional problem, so it is manifestly impossible to convert that into a single number without a loss of information that is bound to produce many false positives or false negatives for gerrymandering.

Ensemble analysis—the technique used by Duchin and in our paper—attempts to combat this principle by comparing the above metrics for a proposed map to the distribution of metrics given for a generated set of random and legal maps. Ensemble analysis doesn’t produce one number that indicates how gerrymandered a proposed map is, but instead, it gives a way to compare whether a proposed map is “unusual” compared to the set of all possible legal maps that could have been drawn in that state. In this way, ensemble analysis no longer reduces gerrymandering to a single-dimensional answer, allowing us to better determine when gerrymandering is and is not present.

To better understand the specifics of ensemble analysis, we must first define what an ensemble is.

Definition 4.1. An **ensemble** for a given state is a set of random, legal maps which is representative of the set of all possible legal maps in this state. The ensemble is created using no partisan data.

The core idea of ensemble analysis, then, is to evaluate whether a map for a state is gerrymandered by comparing it to the maps in the ensemble, which by definition have no intentional bias towards a certain party. Once an ensemble exists, we can compare characteristics (such as efficiency gap, partisan bias, responsiveness, etc.) of a proposed map to the distribution of values for maps in the ensemble. We may then identify whether the proposed map is an outlier or whether it has values comparable to a typical map in the ensemble. While outlier maps need not be gerrymandered, these maps are unusual for a state. This may mean the map is gerrymandered.

Conversely, if a map is suspected to be gerrymandered, we expect it to be an outlier in the ensemble in at least some metrics. This provides stronger supporting evidence of the claim than any of the previously mentioned metrics.

5 Methods

5.1 Drawing Maps with a Markov Chain

5.1.1 Data and Gerrychain Software Package

Our data comes from the Metric Geometry and Gerrymandering Group (MGGG) Redistricting Lab at Tisch College of Tufts University. MGGG aggregated state demographic and geographic data which we downloaded from redistrictingdatahub.org [18]. We also used the Gerrychain Python library from MGGG to create random maps from this data.

5.1.2 Running the Chain

The total number of possible legal redistricting maps for any state is far larger than it would be possible to compute on any existing computer. Instead, we sample from the space of all possible maps. Using Markov Chain Monte Carlo methods we can obtain a random, representative sample. These methods have been proven to produce a true representative sample of the space [19].

While the boundaries of congressional districts can theoretically be drawn anywhere, we restrict them to precinct boundaries. This simplifies our computations, and precincts are small enough that we can still draw districts with a realistic level of granularity.

This allows us to represent redistricting maps as dual graphs. Each node represents a precinct and each edge represents a boundary between two adjacent precincts. Then each redistricting map is a different partition of this graph, with a certain number of edges “cut” to create this partition. The following definition precisely describes what we mean by a cut edge.

Definition 5.1. A **Cut Edge** is an edge of the dual-graph which connects two adjacent precincts in separate districts. That is, an edge which is cut when we partition the dual graph into districts.

Figure 5 demonstrates this concept. Another definition relating to dual-graphs will become important later.

Definition 5.2. Let ε and ε' be two districting maps. A **Cross Edge** is an edge of the dual-graph which is a cut edge in ε but not a cut edge in ε' . That is, an edge which connects two adjacent precincts that are in separate districts in ε but in the same district in ε' .

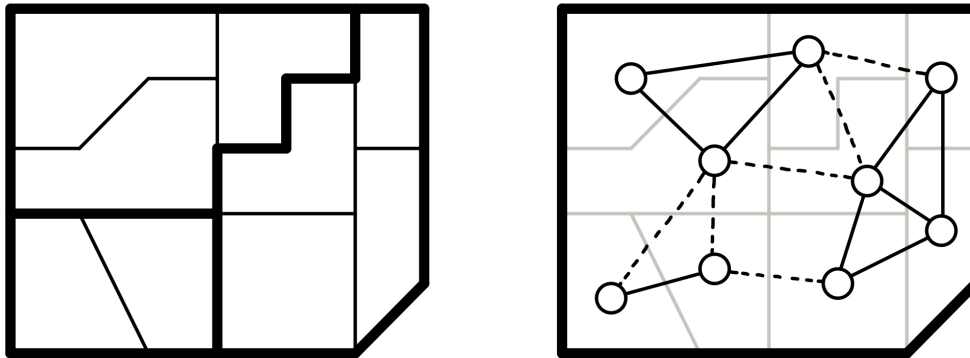


Figure 5: A toy state with 3 districts and 9 precincts. Each circle represents a precinct. The solid lines show edges between adjacent precincts in the same district, and the dashed lines connect adjacent precincts in different districts.

To run the chain, we start with a “seed map,” which has to be some partition of the dual graph into legal districts. We use the states’ previous (2010) congressional districts, but proposed redistricting maps or randomly generated maps are also valid seeds.

Starting from the seed map, we use a recombination algorithm³ to reach the next map in the chain. Recombination fuses two adjacent districts then randomly splits them apart again to form a new map [20] [21]. Figure 6 show the real action of the recombination algorithm in Minnesota.

Once a new map has been generated with recombination, we determine whether to add it to our ensemble using an acceptance function to determine how legal it is and simulated annealing to explore the space of possible maps. Both are discussed in more detail below.

If the map is accepted based on the acceptance function, it is added to the ensemble and is used to create another map using recombination, and the process repeats. If the map is not accepted, we go back to its parent map and re-do the recombination until it generates a map which is accepted.

5.2 Score Function

In order to ensure that all maps in our ensemble are legal, we must write a function which scores the legality of any generated map in our chain. This leads to another definition.

Definition 5.3. Let J be a function from the set of maps in our ensemble to the positive, real numbers. We say J is a **score function** when the output for

³Previous approaches used a “flip” method which reassigns a single node (precinct) to a different district to generate a new map. DeFord et. al. (2021) found that recombination better preserves compactness and increases the speed of convergence.

any map describes how well this map adheres to legal requirements, with lower scores indicating more legal maps.

Previous literature created a score function by writing individual characteristic score functions to evaluate population deviation, compactness, the number of split counties, and Voting Rights Act compliance for each map. The score function is then a weighted sum of the four characteristic score functions. We employ characteristic score functions based on those used by [19] with some slight alterations.

5.2.1 Population Score

In our ensemble, any generated map which has at least two percent population deviation between districts is outright rejected. However, not all maps between zero and two percent population deviation are equally legal. Thus, in this range, we apply the following population score function.

For any map ε with n districts, the population score is

$$J_P(\varepsilon) = \sqrt{\sum_{i=1}^n \left(\frac{\text{pop}(D_i(\varepsilon))}{\text{pop}_{\text{ideal}}} - 1 \right)^2}$$

where $\text{pop}_{\text{ideal}} = \sum_{i=1}^n \frac{\text{pop}(D_i(\varepsilon))}{n}$

with $D_i(\varepsilon)$ denoting the i^{th} district in ε .

Notably, our score function allows districts within a given map to differ by up to 2% from the ideal population, though this does incur a score

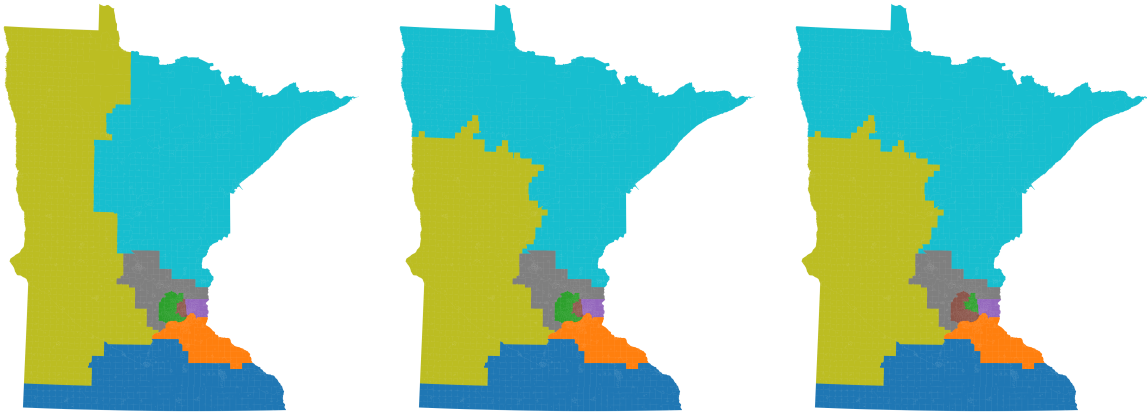


Figure 6: The left map is our seed map, Minnesota’s 2010 congressional district map. In the center is the second map in our chain, generated from the seed map using recombination. The right (and third) map is generated from the second using recombination.

penalty. However, many proposed plans have districts which vary in population by as few as one person, based on the census data. For several reasons, we have chosen to allow plans which vary in population more than these plans.

Firstly, the census does not perfectly capture the complete population, nor does it do so equitably. Secondly, we are using census data from 2010, since equivalent data to that which we are using has not yet been processed and made available using the 2020 census and 2020 election results. Thirdly, we wish to allow a broader exploration of the space of all legal maps, and keeping the population cutoff at 2% accomplishes these goals without getting stuck at some local minimum regarding population distribution.

5.2.2 Compactness Score

For any map ε , our compactness score is calculated as

$$J_C(\varepsilon) = \sum_{i=1}^n \frac{\text{perimeter}(D_i(\varepsilon))^2}{\text{area}(D_i(\varepsilon))}$$

where $\text{perimeter}(D_i(\varepsilon))$ and $\text{area}(D_i(\varepsilon))$ are the perimeter and area of the i^{th} district of the map ε . This is inversely proportional to the Polsby-Popper compactness measure (which gives a score in $[0, 1]$ where 1 is most compact), and similar to Schwartzberg. By flipping the numerator and denominator, and taking out the factor 4π , we map the Polsby-Popper measure onto the interval $[4\pi, \infty]$ (where 4π is most compact).

5.2.3 Split County Score

Keeping counties fully within one district is always preferred to splitting one county into two, or even three districts. As such, we compute a score based on the number of districts into which a county is split.

Despite the fact that many states have laws explicitly restricting county splits, many existing re-districting plans have chosen to split one or more counties to achieve more desirable characteristics in other metrics, such as equal population. Hence, we need a score which penalizes splitting counties into multiple districts, scaled by how “badly” they are split, but which does not prohibit this splitting from occurring.

We implement a split county score that scales linearly with the number of counties that are split and scales exponentially with the number of districts into which a county is split. In order to do so, we must define “district dominance” within a given county, and its relation to the function $F_i(s)$, which is an important component in the split county score function.

Definition 5.4. Given a county s , we define the **most dominant** district in s to be the district which has the highest proportion of its precincts located within s . We may sort all districts which have at least one precinct in s from most to least dominant according to the proportion of precincts within the district which fall within s .

We can now define the function $F_i(s)$, which will be used in the over county split score for a map ε .

Definition 5.5. Let $F_i(s)$ be the proportion of precincts within a given county s that fall within the top i dominant districts of s .

Note, a physically small (few total precincts) district which is fully contained within a given county is more dominant in that county than another district which contains the remainder of the precincts in the county but which has at least one precinct outside the county.

We now define the split county score function, modified from Herschlag, et al. [19], to be $J_S(\varepsilon)$ as follows:

$$J_S(\varepsilon) = \sum_{i=2}^n C^{i-2} \cdot |S_i(\varepsilon)| \cdot W_i(\varepsilon)$$

where $C = \text{some large constant } (> 1)$,

$$S_i(\varepsilon) = \{\text{counties in } \varepsilon \text{ split between } \geq i \text{ districts}\},$$

$$W_i = \sum_{s \in S_i} \sqrt{1 - F_{i-1}(s)}.$$

Here i iterates over the number of districts split into one county, the number of districts split into two counties, ..., the number of districts split into n counties; i does *not* iterate over particular districts (as is common elsewhere in this paper). For example, when $i = 3$, we are considering every county that is split between at least three districts.

Observe that, in the term W_i , we consider $F_{i-1}(s)$ because we want to compute the proportion of precincts in s that are *not* in the first $i - 1$ most dominant districts. That is, if we have a county s which is split into more than three districts, $1 - F_2(s)$ is the proportion of precincts in s which lie outside of the two most dominant districts of s . Thus, W_i takes into account how ‘‘badly’’ a county is split into districts contained in it. For example, if a county s is split into three districts, but one of the districts only has one precinct from s , then the corresponding $F_2(s)$ term will be very large, and thus the contribution of s to the W_3 value in the split county score will be very small.

5.2.4 Voting Rights Act (VRA)

Voting Rights Act (VRA) compliance of a map is very hard to objectively assess. One approach involves combining census and voting data to identify minority preferred candidates in past primary and general elections, and then evaluating a re-districting plan based on the number of districts that would have elected minority preferred candidates [22]. While this approach rigorously evaluates

the electoral opportunity for minorities under a re-districting plan, the amount of data cleaning necessary to link census and voting data is no small task. Another approach, used by Herschlag, et al., seeks to fulfill demographic targets for each district [19]. However, Herschlag, et al.’s demographic targets were based on the precedent set by North Carolina’s previous map, and seeking to fulfill demographic targets generally runs the risk of diluting minority voting power by packing minority voters into one district.

Given the difficulty of assessing VRA compliance in a rigorous manner, coupled with the weakened power of the VRA as a result of Shelby County v. Holder, we decide not to include a VRA component in our scoring function. To do so, we set the weight $w_V = 0$. However, we will still pay attention to the demographic distribution across districts in the analysis of our results, so we can still detect racial and ethnic gerrymandering.

5.3 Weighting the Score Function

Our total score function for a map ε is then merely a weighted sum of each of the above score functions:

$$J(\varepsilon) = w_P J_P + w_C J_C + w_S J_S + w_V J_V.$$

We sought for our composite score J for any given map to have equal contributions from each of the four characteristic score functions above. However, by nature of construction, not all score functions lead to raw scores which are on the same order of magnitude. For example, most population deviation scores in Minnesota were found to be below 1 while several scores for split counties in Minnesota were above 1,000.

To scale all of our scores to have similar mean values, we computed the mean score value for each characteristic in an unweighted ensemble. For ease of comparison, scale this average to 100 for all three characteristic score functions. So we set the weights to be

$$w_i = \frac{100}{\text{mean score } J_i}$$

from a preliminary run of our ensemble. Our weights for Minnesota are as follows:

$$w_p = 3175, w_c = 0.2686, w_s = 0.10893, w_v = 0.$$

Our weights for Texas are as follows:

$$w_p = 1321, w_p = 0.0515, w_s = 0.00063345, w_v = 0.$$

5.4 Acceptance Function

The previously defined score function creates a score for each map of how legal it is, with a lower score indicating a better map. While scores have a lower bound of 0, there is no upper bound on scores. In generating our ensemble, we need a function which turns this score into a probability distribution. This leads to our next definition:

Definition 5.6. An **Acceptance Function** is a function which maps from the set of possible maps to $[0, 1]$ and gives a probability for how likely it is that map will be chosen in an MCMC algorithm. This Acceptance Function is based on the Score Function defined above.

Definition 5.7. Let ε be a map produced by the chain and let ε' be a new map proposed from ε by the chain. We call ε the **Parent Map** and ε' the **Child Map**.

Given a proposed child map ε' from a parent map ε , let $P(\varepsilon'|\varepsilon)$ be the probability of accepting ε' defined as follows:

$$P(\varepsilon'|\varepsilon) = \min \left(1, \frac{Q(\varepsilon', \varepsilon)}{Q(\varepsilon, \varepsilon')} e^{-\beta \left(\frac{J(\varepsilon') - J(\varepsilon)}{J(\varepsilon)} \right)} \right),$$

where β is the simulated annealing factor (see Section 5.4.1 below) and

$$Q(\varepsilon, \varepsilon') = \frac{1}{2} \cdot \frac{\text{cross edges}(\varepsilon, \varepsilon')}{\text{cut edges}(\varepsilon)}.$$

Observe that the term $\frac{Q(\varepsilon', \varepsilon)}{Q(\varepsilon, \varepsilon')}$ is then half the proportion of cut edges in ε which flip to being uncut in ε' . Thus, $\frac{Q(\varepsilon', \varepsilon)}{Q(\varepsilon, \varepsilon')}$ yields a value greater than one when the child map ε' has districts with more complicated borders and more cut edges than the parent map ε . This term of the acceptance function, then, acts to assist the chain in exploring maps which are different and more complicated than the parent map, allowing us to better sample the entire space of legal maps.

This acceptance function is very similar to that defined by Herschlag, et al. However, the function used in Herschlag, et al. does not divide the difference $J(\varepsilon') - J(\varepsilon)$ by $J(\varepsilon)$, as we do in our equation. Note that if the difference is not divided by the score of the parent map, then the absolute magnitude of scores affects the probability of accepting the child map. That is, if the score of the child map and the score of the parent map are scaled linearly while keeping the ratio between them constant, the probability of accepting the child map changes. In

particular, if the magnitudes of the scores increase proportionally, then the probability of accepting a child map which has a higher score than the parent map decreases; if the child map has a lower score than the parent map, it is always accepted anyway.

We wanted the flexibility to change the ratio of the weights between the different score function components without affecting the overall probability accepting. As such, we decided to divide the difference between the score of the child map and the score of the parent map by the score of the parent map.

5.4.1 Simulated Annealing

When using a MCMC algorithm with a score function, we choose more legal maps at a higher probability than less legal maps. Once at a given map, we are more likely to choose a child map which has a better score than the parent map than a child map with a worse score than the parent map. If we follow this method exclusively, however, it has been found that the MCMC chain may get stuck at local maximum. That is, all nearby maps in the space of maps have worse scores (are less legal), and so we will not choose these maps. There may, outside of this locus, be maps with even better scores than the local maximum. In order to avoid getting stuck at these local maxima, we implement an inverse temperature constant known as a simulated annealing constant.

Definition 5.8. A **Simulated Annealing Constant** β serves as a multiplier on the Score Function output for a given child map in calculating the probability of accepting this child map with the Acceptance Function. This parameter β begins at 0 and grows gradually throughout the progress of an MCMC chain to a determined maximum value.

For our analysis, we set $\beta = 0$ for the first 10,000 accepted maps. Then for the next 60,000 maps β grows linearly from 0 to 23.5. The final 80,000 maps are generated with $\beta = 23.5$. As a result of dividing the difference between the child and parent scores by the parent score, we can speak about the general probability of accepting a child map with a score worse than its parent based on the ratio of their scores, for some given β . In particular, note that $\frac{J(\varepsilon') - J(\varepsilon)}{J(\varepsilon)}$ gives a measure of how much worse the child map is than the parent: if $\frac{J(\varepsilon') - J(\varepsilon)}{J(\varepsilon)} = 0.05$, we say that the child map ε' is 5% worse than the parent map ε . We chose $\beta = 23.5$ in order to get an acceptance rate that we desired for worse maps. See below for a table which shows the approximate probabilities of accepting a child map ε' from parent

ε according to how much greater $J(\varepsilon')$ is than $J(\varepsilon)$, assuming $\frac{Q(\varepsilon',\varepsilon)}{Q(\varepsilon,\varepsilon')} \approx 1$:

$\frac{J(\varepsilon')-J(\varepsilon)}{J(\varepsilon)}$	$P(\varepsilon' \varepsilon)$
0.00	1.0
0.01	0.79
0.02	0.63
0.05	0.31
0.10	0.10
0.20	0.02

Using recombination to generate new maps from previous maps, the acceptance function to decide whether these maps are sufficiently legal, and simulated annealing to explore the space of map, we can now generate ensembles by running chains for Minnesota and Texas.

5.5 Analysis Techniques with Ensemble Analysis

For each partisan metric described above, we generate a histogram of all of the values found in our ensemble maps. We can then compare the value of this metric for the maps of interest we are analyzing to the distribution of values in our ensemble to understand how “unusual” the map of interest is for a random, legal map of the state. We consider how many standard deviations the value for the map of interest lies away from the mean of the distribution for all values in the ensemble.

5.5.1 Gathering Data on Proposed Maps

To generate metrics for maps of interest, we run a chain of length one with each map as the seed map. We then are able to determine all of the metric values computed with the exact same data as with our ensemble. Note that the values for efficiency gap and mean-median we compute are different from those reported on 538.com and other sites online. This is due to the fact that we are using relatively old election data and averaging over several elections.

5.6 Data Visualization

To generate our graphs we used the Tidyverse package in R [23]. Our graphs are built from the data of 700,005 maps in our ensemble for each state. Metrics for Minnesota and Texas are computed as the averages over a series of past elections.

For Minnesota, we have data for the following elections:

- 2012, 2014, 2016, 2018 State House

- 2012, 2016 State Senate
- 2012, 2014, 2016, 2018 US House
- 2012, 2014, 2018 US Senate
- 2012, 2016 US Presidential
- 2014, 2018 Governor

For Texas, we have data for the following elections:

- 2012, 2014, 2016 US House
- 2012, 2014 US Senate
- 2012, 2016 US Presidential
- 2014 Governor

5.6.1 Score Function Histograms

We created histograms for the distributions of the total score (Figures 31 and 35), population score (Figures 32 and 36), county score (Figures 34 and 38), and compactness score (Figures 33 and 37). This allowed us to easily see the behavior of our chain and relative contribution of each component to the total score. These can be found in the appendices.

5.6.2 Partisan Metric Histograms

Each partisan metric for each state was plotted as a histogram, where the Minnesota 2010, 2020, and two proposed maps as well as the Texas 2010 map were plotted as vertical lines at their respective scores for the measure. We plotted additional lines at the mean, as well as some standard deviation intervals to more easily detect where each of the maps of interest fell in our distribution. See Figures 7,8,9,10, 12,13,19,20,21,22,23,25,and 26.

5.6.3 Ordered Districts Box Plots

In order to have some way to look at relative levels of packing and cracking amongst different demographics, we created a box plot modeled off those seen in Herschlag, et al. [19].

As our chain generated new maps, district borders were continually being reshaped and reordered. Our resulting maps had districts ordered 1 – 8 in Minnesota and 1 – 36 in Texas. However, due to this recombination, the resulting district numbers were arbitrary—the meaning of district 1 for the 100th map in our chain was entirely different than the meaning of district 1 for the 100,000th map in our chain. Therefore, in order to look at district-level results across our chain, we needed to use some other method of ordering our districts.

The first way we chose to order the districts for each map was by Democratic vote share (Figures 14, 15, and 27). To do this, we chose an election of interest, and computed Democratic vote share (DVS_i) to be:

$$DVS_i = \frac{\text{Dem votes in district } i}{\text{Total votes in district } i}$$

where $i \in \{1, \dots, 8\}$, for each district in Minnesota and $i \in \{1, \dots, 36\}$ for Texas. For each map, we then reassigned district labels in increasing order of this democratic vote share. From there we plotted the box plot for each district, where each box plot showed the democratic vote share distribution of the district with i^{th} smallest democratic vote share for each map in our ensemble.

We used the same method when creating the box plots for voting age population (Figures 16 and 28), Black proportion of voting age population (Figures 17 and 29), and Hispanic proportion of voting age population (Figures 18 and 30). Specifically, the voting age population was a metric included for each district, so we reordered districts based on this value.

The black proportion of voting age population for each district (BVS_i) was computed as:

$$BVS_i = \frac{\text{Black voting age pop. in district } i}{\text{Voting age pop. in district } i}$$

The Hispanic proportion of voting age population for each district (HVS_i) was computed as:

$$HVS_i = \frac{\text{Hispanic voting age pop. in district } i}{\text{Voting age pop. in district } i}$$

5.6.4 Seats Votes Curves

To generate seats votes curves for each state, we used the “Political Science Computational Laboratory” (pscl) package in R [24]. This is a package developed by Simon Jackman at the University of Sydney. This package uses the uniform swing method to generate the output seats votes curves.

Each seats votes line output required a string containing the vote share of a given party in each district. We used the vote share numbers calculated in box plot calculations and ordered districts in the same manner. To get a single string, we averaged the value for each district over all the maps in our ensemble. See Figures 11 and 24.

6 Results

6.1 Function of Our Chain

6.1.1 Simulated Annealing

We run five different chains of 150,000 accepted maps for each state (starting with the seed map) to produce a total of 750,000 maps for each state. Since $\beta = 0$ for the first 10,000 accepted maps in the chain, the acceptance function for these 10,000 maps does not take into account the score of each map. Thus, we may be generating maps which do not adhere well to legal criteria during this part of the chain. As a result, in our analysis, we cut out the first 9,999 maps from our analysis, leaving us with 700,005 maps in our ensemble for each state.

6.2 Minnesota

Our results for Minnesota are based on the 700,005 maps included in our ensemble. We overlaid partisan metric scores for the 2010 map, two of the proposed plans (neither of which ended up being accepted), and the 2020 court approved redistricting map.

6.2.1 Democratic and Republican Seats

In Figures 7 and 8 we graph the number of Democratic and Republican seats, respectively, in a given map, averaged across all of our elections. This allows for maps which have a non-integer number of Democratic and Republican seats. Note the mean number of Democratic seats is 5.052 and the standard deviation is 0.303.

The Sachs Plaintiff plan falls almost exactly on the mean for Democratic seats. The 2020 map falls just inside -1 standard deviation, the 2010 map falls just outside -1 standard deviation, and the Wattson Plaintiff plan falls to the left of the 2010 plan (within -2 standard deviations). The Sachs plan is thus the most typical for our ensemble. The remaining three analyzed maps show slightly more Republican bias than the most typical maps in our ensemble; however, the number of Democratic seats is not entirely unusual.

6.2.2 Efficiency Gap

The efficiency gap scores for our ensemble is given in Figure 9. The mean score for efficiency gap in our Minnesota ensemble is 0.042, and the standard deviation is 0.039. This indicates that many of the

maps in our ensemble have a bias towards the Democratic party, with most maps leading to fewer wasted Democratic votes than wasted Republican votes.

The Sachs Plaintiff plan falls almost exactly on the mean for efficiency gap. That is, the Sachs plan is extremely typical for our ensemble. The 2020 map falls just inside -1 standard deviation, the 2010 map falls just outside -1 standard deviation, and the Wattson Plaintiff plan falls farther to the left than the 2010 map. Thus, the 2020 plan, 2010 plan, and Wattson plan all demonstrate more Republican bias than many plans in our ensemble, but they are not entirely unusual.

6.2.3 Mean Median

The mean-median scores for our ensemble are given in Figure 10. The mean score for mean-median in our Minnesota ensemble is -0.017 , and the standard deviation is 0.011 . This indicates that many of the maps in our ensemble have a Republican bias. All of the proposed and accepted maps have similar levels of bias to that of maps in our ensemble.

The Sachs Plaintiff plan falls just to the right of the mean for mean-median. That is, the Sachs plan favors Democrats slightly more than the average plan in Minnesota. The 2020 map and Wattson plan both fall around 1 standard deviation to the left of the mean. That is, these maps demonstrate slight Republican bias, but they are not wholly unusual for the ensemble maps in Minnesota. The 2010 map falls the farthest to the left (indicating more Republican bias), but again isn't entirely unusual for the ensemble maps.

6.2.4 Seats-Votes Curve

Democratic and Republican seats-votes curves are given in Figure 11. The Republican curve is above the Democratic curve at values of vote share close to the observed election, meaning most maps in our Minnesota ensemble have a slight bias in favor of the Republican party.

6.2.5 Partisan Bias

The partisan bias scores for the Minnesota ensemble are given in Figure 12. The mean partisan bias in our Minnesota ensemble is $-.068$, indicating that Minnesota has a slight natural bias that favors Republicans. None of the proposed maps are unusual in the context of our ensemble, each falling within a standard deviation of the mean. While the Wattson plan and 2020 map are further from the mean than

the Sachs plan, both are more fair than the 2010 map.

6.2.6 Partisan Gini

The partisan gini scores for the Minnesota ensemble are given in Figure 13. Recall, unlike all other metrics, partisan gini takes on only positive values, and higher scores indicate not bias for one party in particular, but rather, discrepancies in the vote share necessary to gain a particular seat share. That is, higher partisan gini values only indicate that the districting map contains bias, but not which party the bias favors.

The mean partisan gini score for our Minnesota ensemble is 0.039 , indicating that the maps in our ensemble contain some amount of natural bias. Note that this makes sense because the average Democratic and Republican seats-votes curves for Minnesota (see Figure 11) are not identical.

6.2.7 Democratic Vote Share by District

To create Figure 14, we considered every map in our ensemble with districts ordered by proportion of Democratic voters. For instance, the box on the left shows that throughout our ensemble the least democratic districts voted slightly less than 50% democratic on average.

Minnesota residents might consider this odd, as the state is known to be purple. However, this plot uses the Senate 2018 election to determine who voted Democratic. Due to Amy Klobuchar's popularity from both parties, this paints a picture of Minnesota as mostly Democratic.

Figure 15, on the other hand, uses the same method but with the 2016 presidential election. Here we see that on average in our ensemble, three districts voted more than 50% Democratic. In both these figures, we can see that the proposed and actual maps sometimes fall in the upper or lower quartiles, indicating that they are more or less Democratic than we would expect based on our ensemble. However, none of the maps ever fall in the outliers in our expected range.

6.2.8 Black Voting Age Population by District

Figure 17 is similar but instead of the proportion of Democratic voters, shows the proportion of Black voters. We see that in Minnesota, every plan has roughly the number of Black voters we expect from the ensemble. That said, the districts with the most Black voters in the real and proposed plans all have

more Black voters than most of the ensemble, possibly indicating cracking of these voters.

This figure does also illustrate the distribution of the Black population in Minnesota. Many districts – that is, most of the state – have few Black residents, and a few districts have many times more Black residents.

6.2.9 Hispanic Voting Age Population by District

Similarly, Figure 18 shows the distribution of Hispanic Minnesotan voters. However, there are some districts with noticeably more Hispanic voters than expected, showing possible packing.

6.3 Texas

Our results for Texas are based on the 700,005 maps included in our ensemble. We overlaid partisan metric scores for the 2010 map.

6.3.1 Democratic and Republican Seats

Figures 19 and 20 show the average number of Democratic and Republican seats respectively won across the state averaged across several election results. As with Minnesota, this is why we find maps which take on non-integer values for the number of Democratic and Republican seats.

We see that the mean number of Democratic seats won in Texas averaged across all elections is 11.42. The 2010 map falls exactly on this mean. Given the extensive media discussion of the presence of gerrymandering in Texas’s 2010 map, as well as extreme outlier values for additional partisan metrics below, we further explored how the Texas 2010 map could have the appropriate number of Democratic and Republican seats but such bias in the seats-votes curve and such extensive cracking and packing. Note that all but one Texas election we have data for occurred before 2015, with the most recent election being the 2016 presidential election. Hence, the values averaged across all elections in the Texas ensemble have a bias towards older Texas elections. We wanted to explore whether voter preferences across the state of Texas had, in fact, changed over the course of the decade.

In Figure 21, we graph the number of Democratic seats won in each ensemble map using only 2012 senate election data (light green) and only 2016 presidential election data (darker green). Note the number of Democratic seats won across the entire ensemble is shifted to the right when using the 2016 presidential election data, indicating that a higher

proportion of voters cast Democratic votes in the 2016 presidential election than in the 2012 senate election. The 2010 map then falls on the mean for Democratic seats using 2012 senate election data, but looks to be an outlier when using 2016 presidential election data.

While unique outcomes can occur when looking at just one election, since in reality individuals vote for particular candidates and not just parties, this figure tells an interesting story. In 2012 when the 2010 Texas map was recently put in place, the number of Democratic seats lined up appropriately with what would be expected given our ensemble. However, as the decade went on, voter preferences changed—likely due to growth of cities in Texas—and a higher proportion of Texas voters voted for Democratic candidates. Thus, in 2016, the 2010 map allotted far fewer seats to Democratic candidates than would be expected in a random, legal map. Hence, as the decade went on, the Texas 2010 map became more gerrymandered.

6.3.2 Efficiency Gap

The mean efficiency gap over our Texas Ensemble, seen in Figure 22, is -0.047 with a standard deviation of 0.024 , demonstrating that the maps in our ensemble show a bias towards the Republican party. The 2010 Texas map has an efficiency gap of -0.085 , falling between -1 and -2 standard deviations away from the mean. This means the 2010 Texas map demonstrates substantial Republican bias; however, it does not have more bias than is present in many random, legal maps generated in our ensemble.

6.3.3 Mean-Median

The mean mean-median value over our Texas ensemble, seen in Figure 23, is -0.022 with a standard deviation of 0.013 , indicating again that the maps in our ensemble show bias towards the Republican party. The 2010 Texas map has a mean-median value of -0.042 , falling between -1 and -2 standard deviations away from the mean. This means the 2010 Texas map again demonstrates substantial Republican bias; however, it does not have more bias than is present in many random, legal maps generated in our ensemble.

6.3.4 Seats-Votes Curve

The average Republican seats-votes curve for all maps in our ensemble and the Republican seats-votes curve for the 2010 Texas map are both shown in Figure 24. The curve for the Republican party

is slightly above the line $y = x$ at values close to the observed election result, demonstrating that Texas has a small natural bias that favors Republicans. However, the curve for the 2010 map demonstrates substantially more bias favoring the Republican party.

6.3.5 Partisan Bias

The mean partisan bias for maps in our Texas ensemble, seen in Figure 25, was $-.049$, indicating a slight natural bias that favors the Republican party. However, the partisan bias for the 2010 map falls over three standard deviations away from the mean, providing substantial evidence of a partisan gerrymander.

6.3.6 Partisan Gini

Recall, unlike other measures, the partisan gini takes on only positive values, with higher values indicating more bias in a districting map. The partisan gini score does not indicate which party is favored, however.

Figure 26 shows the partisan gini scores for our ensemble. In our Texas ensemble, the mean partisan gini score is 0.041 , indicating Democrats and Republicans gain different seat shares at the same vote shares throughout our whole ensemble. The 2010 Texas map lies over $+3$ standard deviations away from the mean, demonstrating that the two parties have even more disparate seat shares at the same vote shares under this districting plan. Because this map is such an outlier, this is evidence of gerrymandering.

6.3.7 Democratic Vote Share Distribution by District

Figure 27 shows the Democratic vote share for districts in our ensemble, ordered from the district with lowest Democratic vote share to highest Democratic vote share in each districting plan. In several instances, the 2010 map falls above/below the box, indicating that there is a higher/lower proportion of Democratic voters in that district than in many maps in our ensemble. In districts 32 through 35, however, the 2010 Texas map is an outlier on the upper range of Democratic vote share, meaning these districts contain a higher Democratic vote share than is present in nearly all maps in our ensemble. This shows that the 2010 Texas map packs Democratic voters into a few districts far more heavily than tends to occur in a random, legal map. This

is evidence of gerrymandering in Texas' bluest districts, like in Austin and San Antonio.

6.3.8 Voting Age Population Distribution by District

In Figure 28 we can see that the distribution of voters ranges from below $480,000$ to above $520,000$, and is more varied than the distribution of residents across districts, which is legally controlled.

6.3.9 Black Voting Age Population Distribution by District

Figure 29 shows similar though more extreme results for the Texas Black voting age population. In most districts there are about as many, or fewer, Black voters in each district than expected from our ensemble. However, in the three districts with the most Black voters, there are more than in any map in our ensemble, showing strong evidence of packing in these districts; that is to say, racial gerrymandering.

6.3.10 Hispanic Voting Age Population Distribution by District

Similarly to the previous section, in Figure 30, we can see that in the 28^{th} to 32^{nd} most Hispanic districts, Hispanic voters are packed. In Texas' two most Hispanic districts, the 2010 map has many fewer Hispanic voters than we expect from our ensemble, showing that those districts are cracked.

7 Discussion

7.1 Minnesota Map

In Minnesota, the four analyzed maps largely fell at values which indicate slightly more Republican bias than the mean values for all maps in our ensemble. However, these analyzed maps were never quite outliers, indicating that while they do take on Republican bias, they are not substantially more biased than several random, legal maps. Hence, we are unable to conclude that the 2010, 2020, Wattson Plaintiff, or Corrie Plaintiff plans are gerrymandered.

7.2 Texas Map

In Texas, the 2010 map is an outlier for Democratic, Black, and Hispanic votes share distribution, partisan bias, and partisan gini. In addition, using 2016 presidential data, the 2010 map is an outlier in terms of Democratic seat share, and it may be an outlier

using this data for efficiency gap and mean-median, although we did not compute these values with 2016 election data alone. We can thus conclude that the 2010 Texas map is highly unusual as compared to random, legal maps for the state. Hence, there is strong evidence that both partisan and racial gerrymandering is present in the 2010 Texas map.

Although we cannot make claims directly about the 2020 Texas map, the media has consistently reported that the 2020 Texas map possesses even more partisan and racial gerrymandering than was present in the 2010 map [25, 26]. Since we found strong evidence of both types of gerrymandering in the 2010 map, it is then likely that political and racial gerrymandering are both present in the 2020 map as well.

7.3 General Ensemble Analysis Findings

In general, we found the maps produced in our ensemble in Minnesota took on values closer to 0 for several partisan metrics than the maps produced in our ensemble in Texas. That is, random, legal maps in Texas tend to exhibit greater bias towards the Republican party than random, legal maps in Minnesota. Herein lies the importance of ensemble analysis—all Minnesota maps would appear to be less gerrymandered than Texas maps if they were analyzed only using the values of these partisan metrics. Ensemble analysis, however, allows us to compare just how unusual a particular map is, demonstrating that maps with less extreme partisan metric values in one state may still be more unusual—and thus potentially more gerrymandered—than maps which take on more extreme partisan metric values in another state.

7.4 Limitations and Future Directions

The most significant limitation to our analysis was the quality of data available to us. We relied upon 2010 census data, since the 2020 census data had not been cleaned and merged with voting records yet in a way that was available to us. This means that our population numbers and our population distribution is outdated. When comparing to the 2010 Texas map, this is not hugely significant. However, when looking at newer maps in Minnesota, we split districts into roughly equal population based on old population numbers. Thus, the maps in our ensemble may not be entirely legal maps chosen today, as a result of changing population values.

In order to somewhat mitigate this problem, we placed lower weights on the population score in our scoring function, allowing our produced ensemble maps to have a higher population deviation than is present in any of the approved maps. In this way, we allowed some room for error in where population grew in Minnesota, such that the maps in our ensemble map have a population distribution that is acceptable for the 2020 census results.

Once 2020 census data is available in the necessary format, rerunning our analyses to look at 2020 maps would yield more reliable results. Additionally, this data would allow for direct analysis of the 2020 Texas map with the two added districts in 2020, giving a better method for analyzing the accepted 2020 map.

Additionally, the election data we had access to was limited. While voting patterns don't appear to have substantially changed in Minnesota over the past decade, this was not the case in Texas. We conducted the Democratic seat-share analysis using multiple elections in order to understand the way voting patterns shifted to somewhat mitigate our lack of more recent data in Texas. Nonetheless, our older voting data makes our predictions on how districts would vote under a new districting map less reliable. Once more recent election data is available—especially several elections from 2018 and 2020—rerunning our analyses would provide more reliable results on the level of gerrymandering present in 2020 maps.

Furthermore, we didn't conduct any rigorous statistical tests on our results. While we visually compared the metrics for the plans we analyzed to the distribution, we did not conduct t-tests to determine how unusual these values are. Additional statistical testing on our results would yield a greater understanding of the levels of gerrymandering present in various maps.

In terms of our methodology, there are some places where additional considerations could lead to stronger analyses. Most notably, we did not conduct a rigorous process for determining the weights for score metrics in our overall score function, such as that conducted by Herschlag, et al. Identifying a process to determine how much each score component should contribute to the overall score function, and how this should vary by state, would result in an ensemble with maps that better adhere to the legal criteria in each state. In addition, producing an analysis such as that done by Moon Duchin on VRA compliance would additionally yield an ensemble which contains more legal maps.

Finally, running our analysis on additional states

would lead to broader results about the levels of underlying bias present in various states as well as the presence or absence of partisan and racial gerrymandering in state-wide redistricting maps.

8 Acknowledgements

We would like to acknowledge several individuals who greatly assisted us along the way throughout this project.

First, we would like to thank our advisor, Deanna Haunsperger. We greatly appreciate her guidance and suggestions throughout this project. We would not have been able to accomplish this analysis without her!

We would also like to thank Carlos Parada, a fellow student who began this project with us but unfortunately was unable to stay on for the remainder of the project. His insights and ideas were helpful in developing the framework for our analysis.

We would like to give special thanks to political science professor Christina Farhart and statistics professor Adam Loy for talking through our project with us, providing us with necessary context, and giving us significant suggestions for our methodology.

Lastly, we would like to thank Oliver’s desktop and Eva’s laptop, whose computing power were vital to the generation of our ensembles and analysis of the results, respectively.

References

- [1] Brian Duignan. Gerrymandering. *Encyclopædia Britannica*. <https://www.britannica.com/topic/gerrymandering>.
- [2] Erica Klarreich. How to quantify (and fight) gerrymandering. *Quanta Magazine*, May 2020.
- [3] Redistricting and the supreme court: The most significant cases. *NCSL*, Sep 2021.
- [4] Boston Gazette. Essex county ; worcester county. *The Library of Congress*, 1812.
- [5] “the gerry-mander,” political cartoon. *Encyclopædia Britannica*. <https://www.britannica.com/topic/gerrymandering/media/1/231865/71241>.
- [6] Becky Little. How gerrymandering began in the us. *History.com*, Apr 2021.
- [7] Redistricting in minnesota. *Ballotpedia*.
- [8] Redistricting in texas. *Ballotpedia*.
- [9] Order stating preliminary conclusions, redistricting principles, and requirements for plan submissions.
- [10] Pub. L. 89-110. Voting rights act of 1965, 8 1965.
- [11] Supreme Court of the United States. Shelby county v. holder, 6 2013.
- [12] Michigan’s new independent citizens redistricting commission (micrc). *League of Women’s Voters of Washtenaw County*, Apr 2021.
- [13] Samuel S.-H Wang. Three practical tests for gerrymandering: Application to maryland and wisconsin. *Election Law Journal*, 15(4):367–384, 2016.
- [14] C.C. Harris. A scientific method of redistricting. *Behavioral Science*, 9(3):219–225, 1964.
- [15] Daniel Polsby and Robert Popper. The third criterion: Compactness as a procedural safeguard against partisan gerrymandering. *Yale Law and Policy Review*, 9:301–353, 1991.
- [16] Nicholas O. Stephanopoulos and Eric M. McGhee. Partisan gerrymandering and the efficiency gap. *University of Chicago Law Review*, 82(2):831–900, 2015.
- [17] Princeton gerrymandering project. *Princeton University*.
- [18] Feb 2022. <https://redistrictingdatahub.org/>.
- [19] Gregory Herschlag, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier, and Jonathan C. Mattingly. Quantifying gerrymandering in north carolina. *Statistics and Public Policy*, 7(1):30–38, 2020.
- [20] Daryl DeFord, Moon Duchin, and Justin Solomon. Recombination: A family of markov chains for redistricting. *Harvard Data Science Review*, 3 2021. <https://hdr.mitpress.mit.edu/pub/1ds8ptxu>.
- [21] Daryl DeFord, Moon Duchin, and Justin Solomon. Recombination: A family of markov chains for redistricting. 2019.

- [22] Amariah Becker, Moon Duchin, Dara Gold, , and Sam Hirsch. Computational redistricting and the voting rights act. *Election Law Journal: Rules, Politics, and Policy*, 20(4):407–441, 12 2021.
- [23] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.
- [24] Simon Jackman. *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. United States Studies Centre, University of Sydney, Sydney, New South Wales, Australia, 2020. R package version 1.5.5.
- [25] Nick Corasaniti, Ella Koeze, and Denise Lu. How texas plans to make its house districts even redder. *The New York Times*, Oct 2021.
- [26] Katie Benner, Nick Corasaniti, and Reid J. Epstein. Justice dept. files voting rights suit against texas over new map. *The New York Times*, Dec 2021.

Appendices

A Minnesota Results Figures

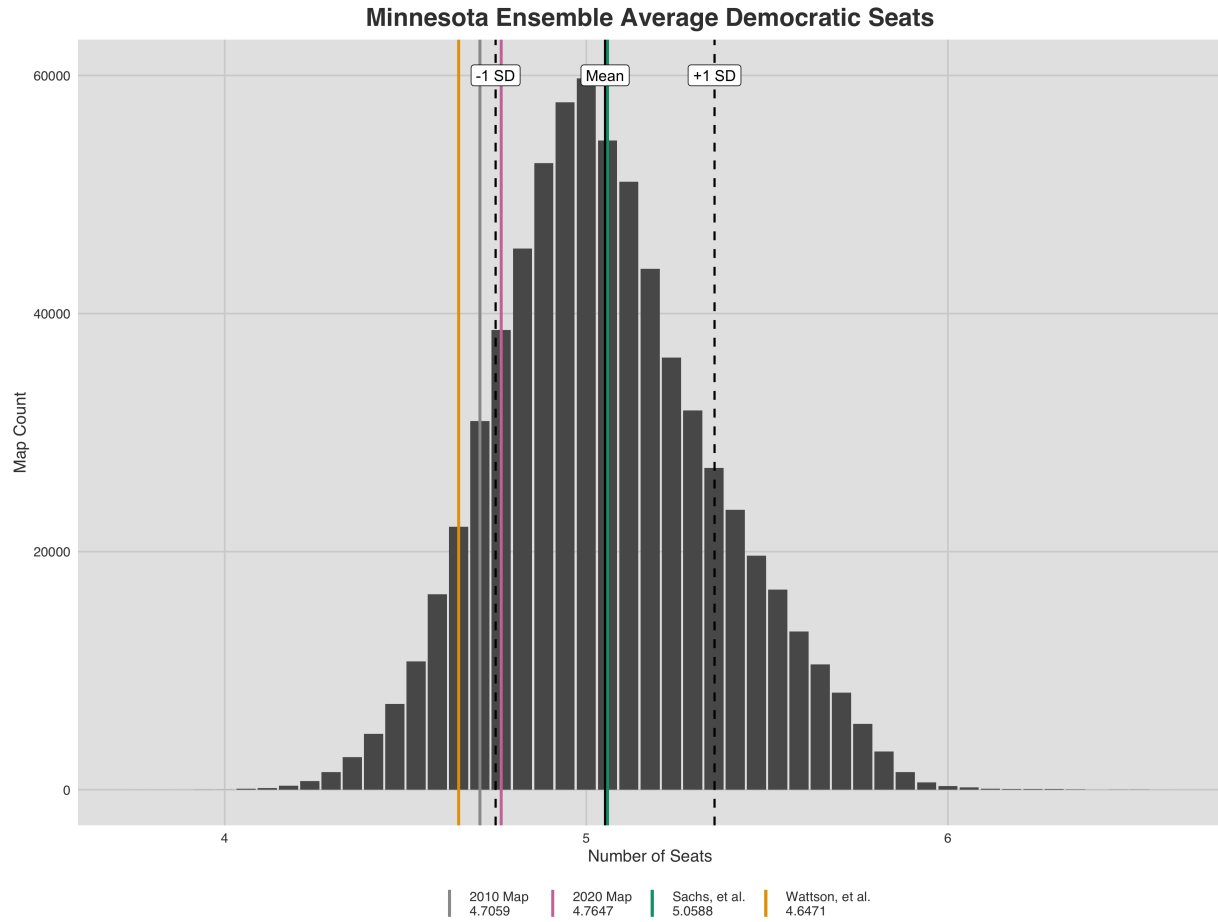


Figure 7: Distribution of average Democratic seats across the Minnesota ensemble. Lines are placed at scores of the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan. The mean falls at 5.052.

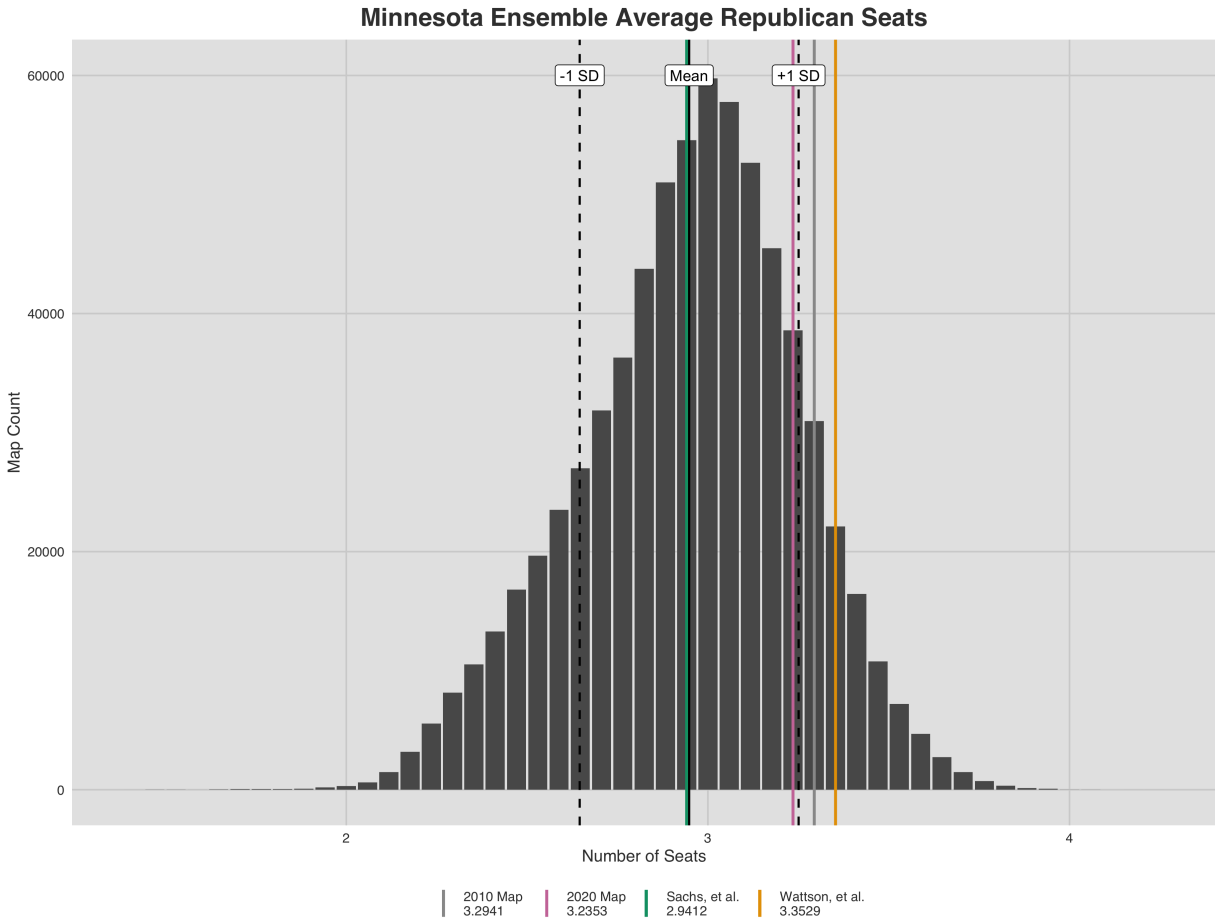


Figure 8: Distribution of average Republican seats across the Minnesota ensemble. Lines are placed at scores of the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan. The mean falls at 2.948.

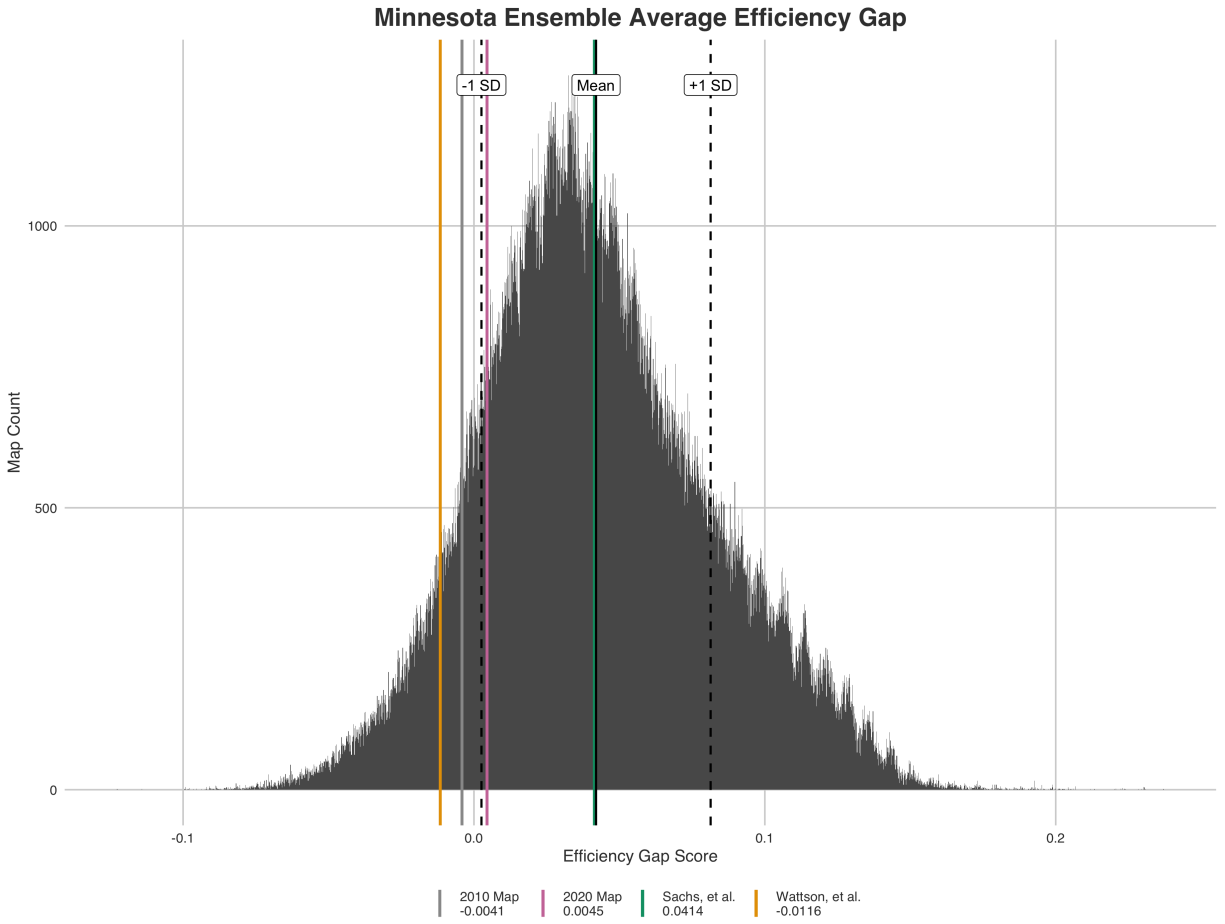


Figure 9: Distribution of Efficiency Gap Scores across the Minnesota ensemble. Lines are placed at scores of the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan. The mean falls at 0.04197.

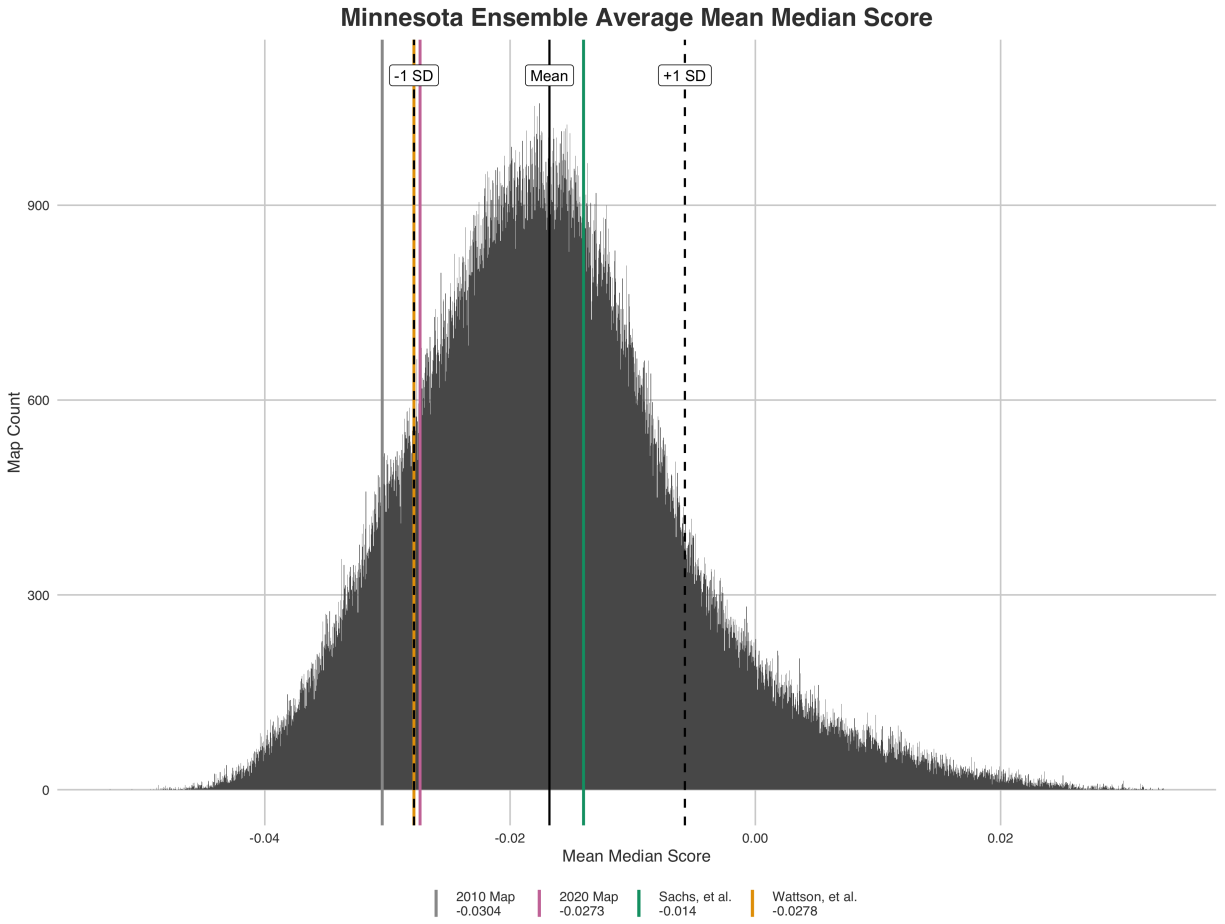


Figure 10: Distribution of Mean Median Scores across the Minnesota ensemble. Lines are placed at scores of the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan. The mean falls at -0.01679.

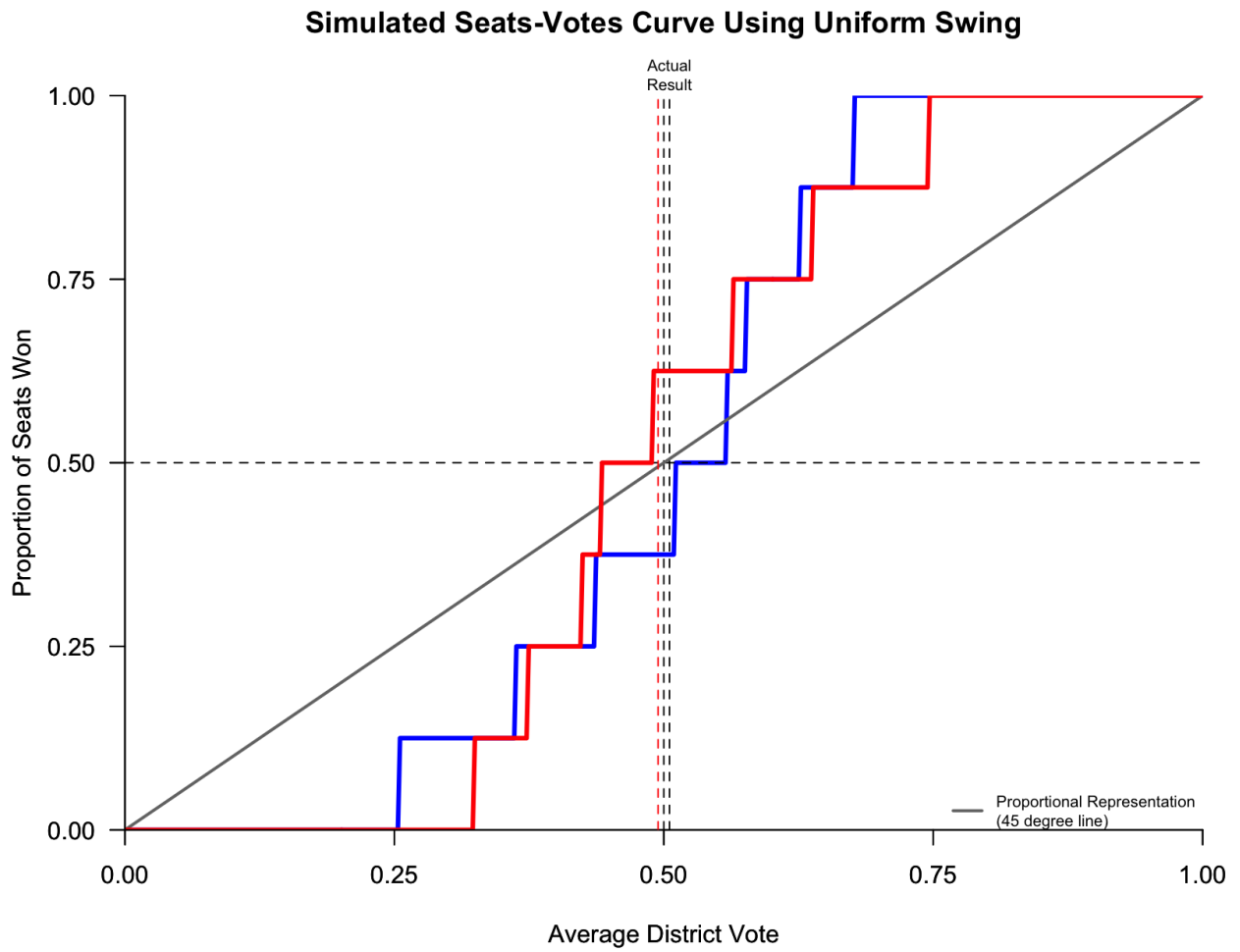


Figure 11: The average Democratic and Republican seats-votes curve of our Minnesota ensemble.

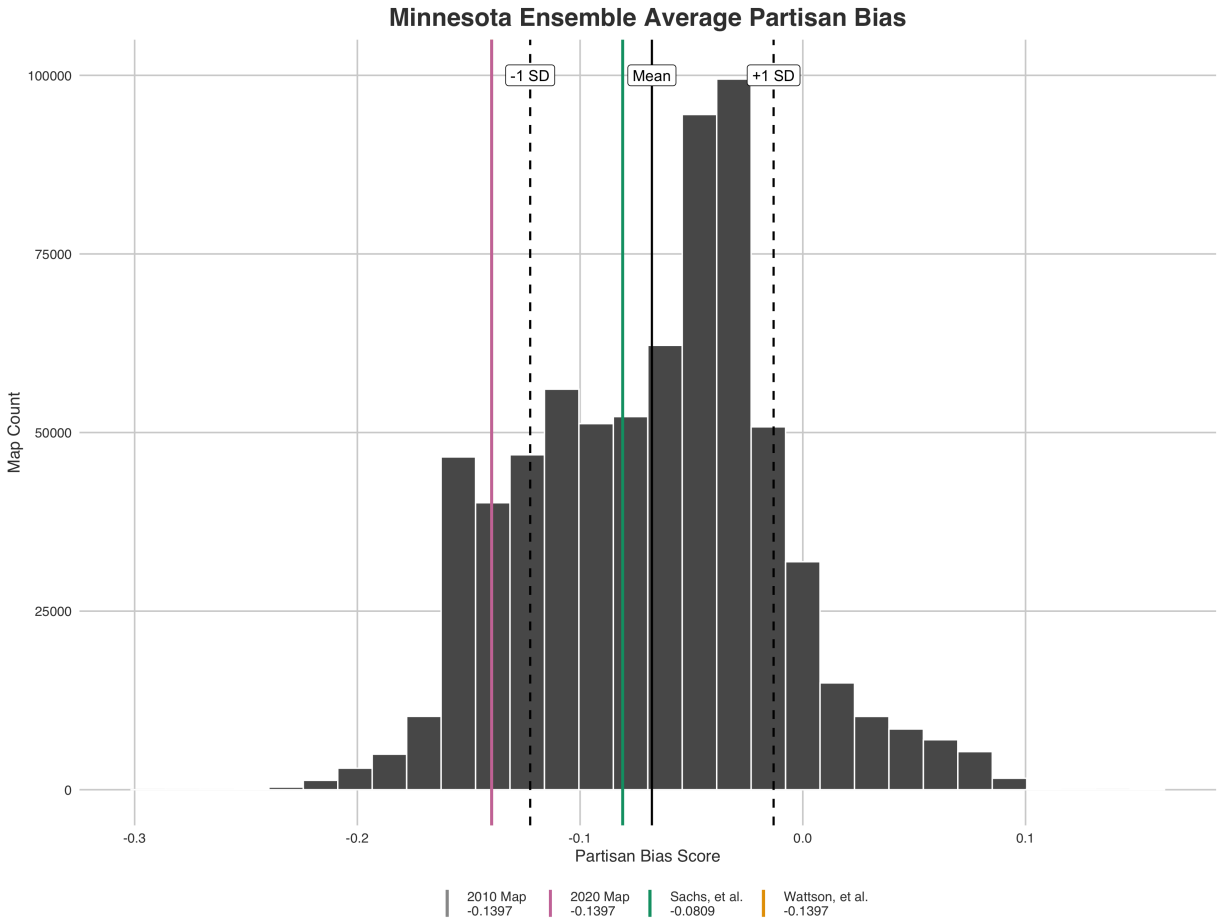


Figure 12: Distribution of Partisan Bias Scores across the Minnesota ensemble. Lines are placed at scores of the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan. Note, the line for the 2010 map lies directly underneath the line for the 2020 map. The mean falls at -0.06776.

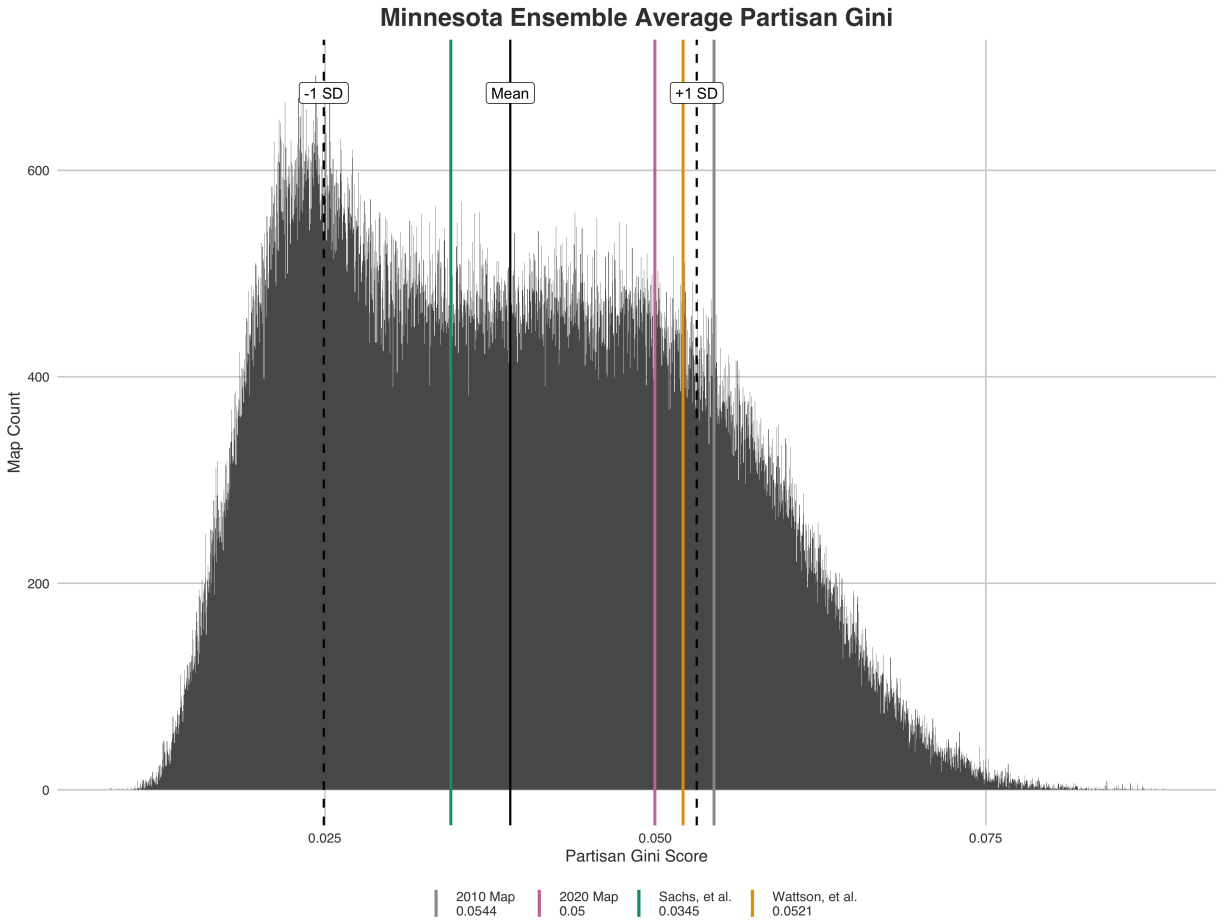


Figure 13: Distribution of Partisan Gini Scores across the Minnesota ensemble. Lines are placed at scores of the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan. The mean falls at 0.03901.

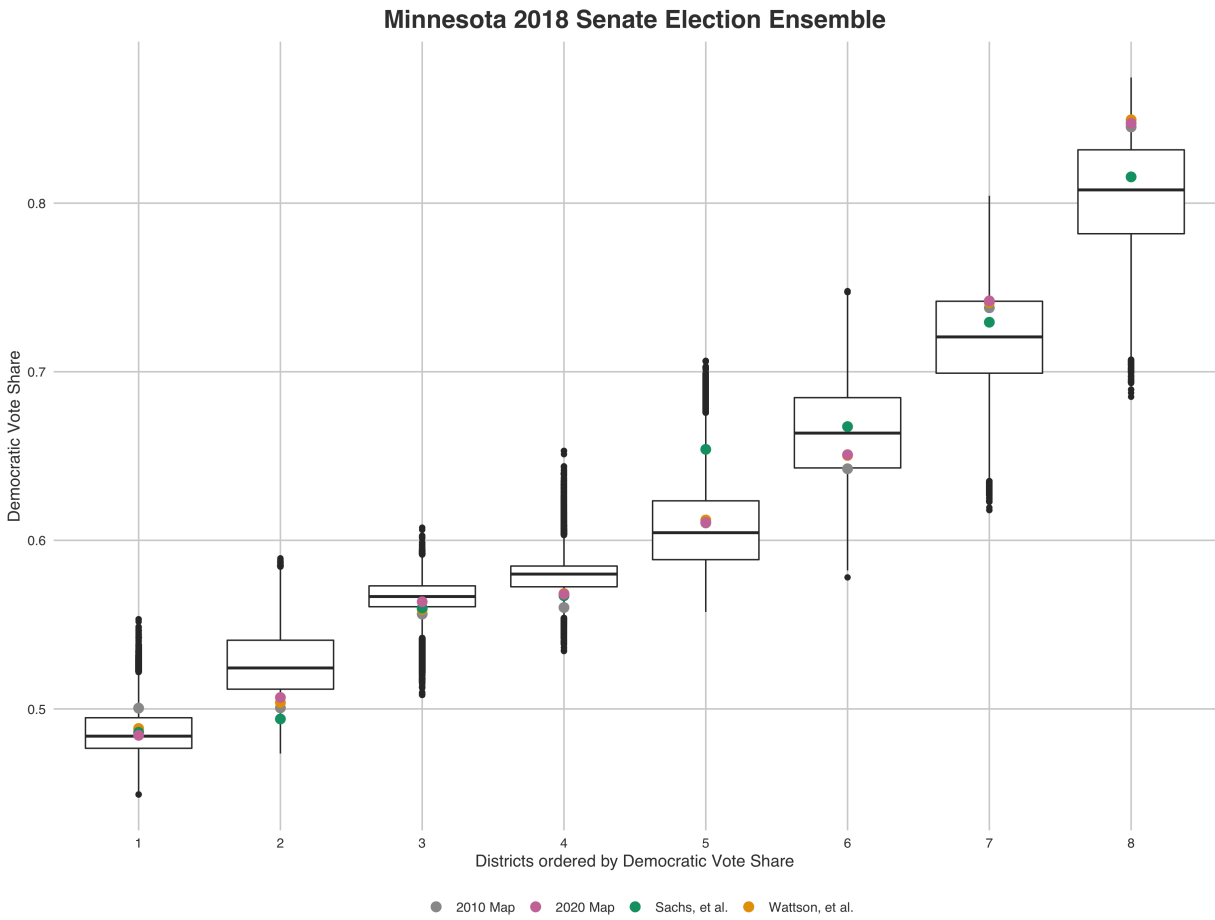


Figure 14: Box plot of Democratic vote share by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2018 Senate Election Results. Dots are placed at Democratic vote shares in each district of the the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan.

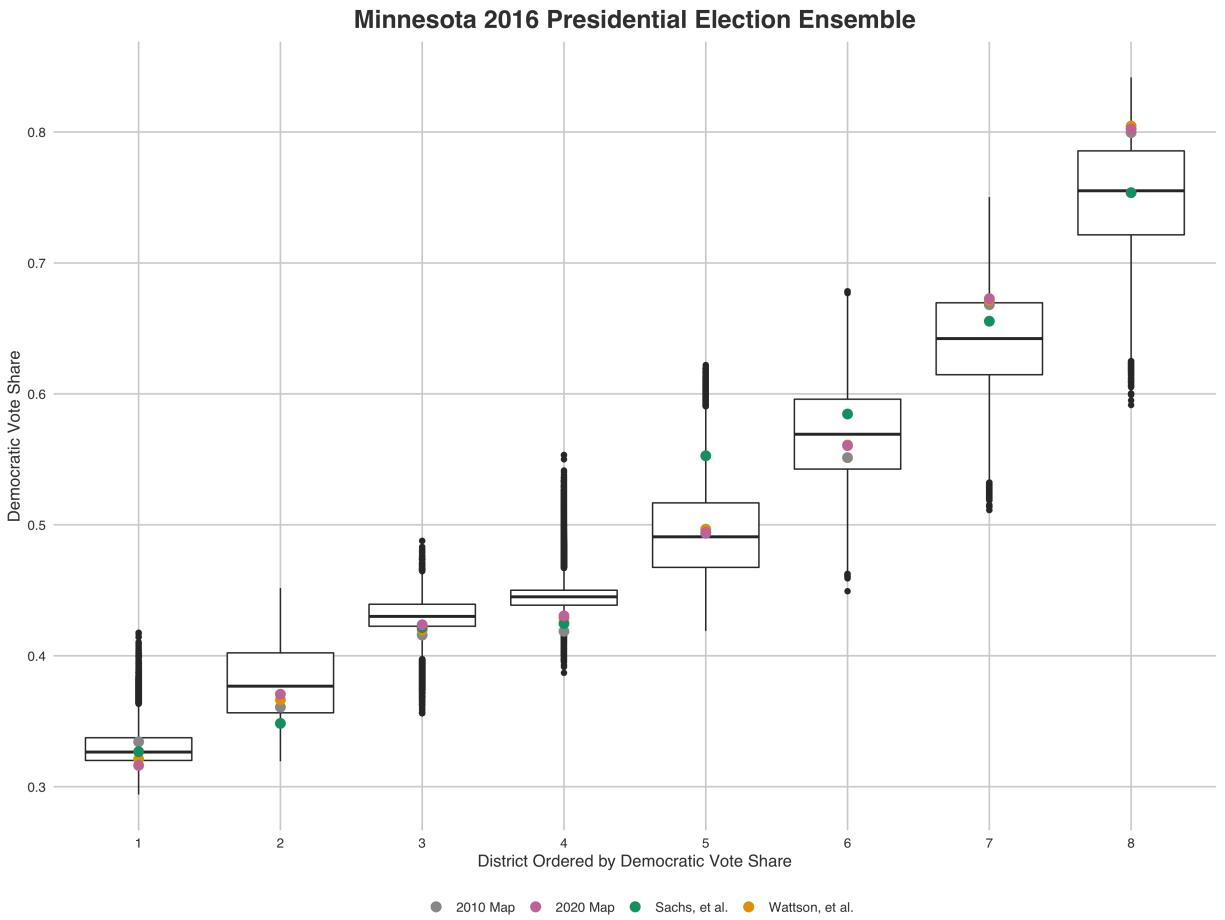


Figure 15: Box plot of Democratic vote share by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2016 Presidential Election Results. Dots are placed at Democratic vote shares in each district of the the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan.

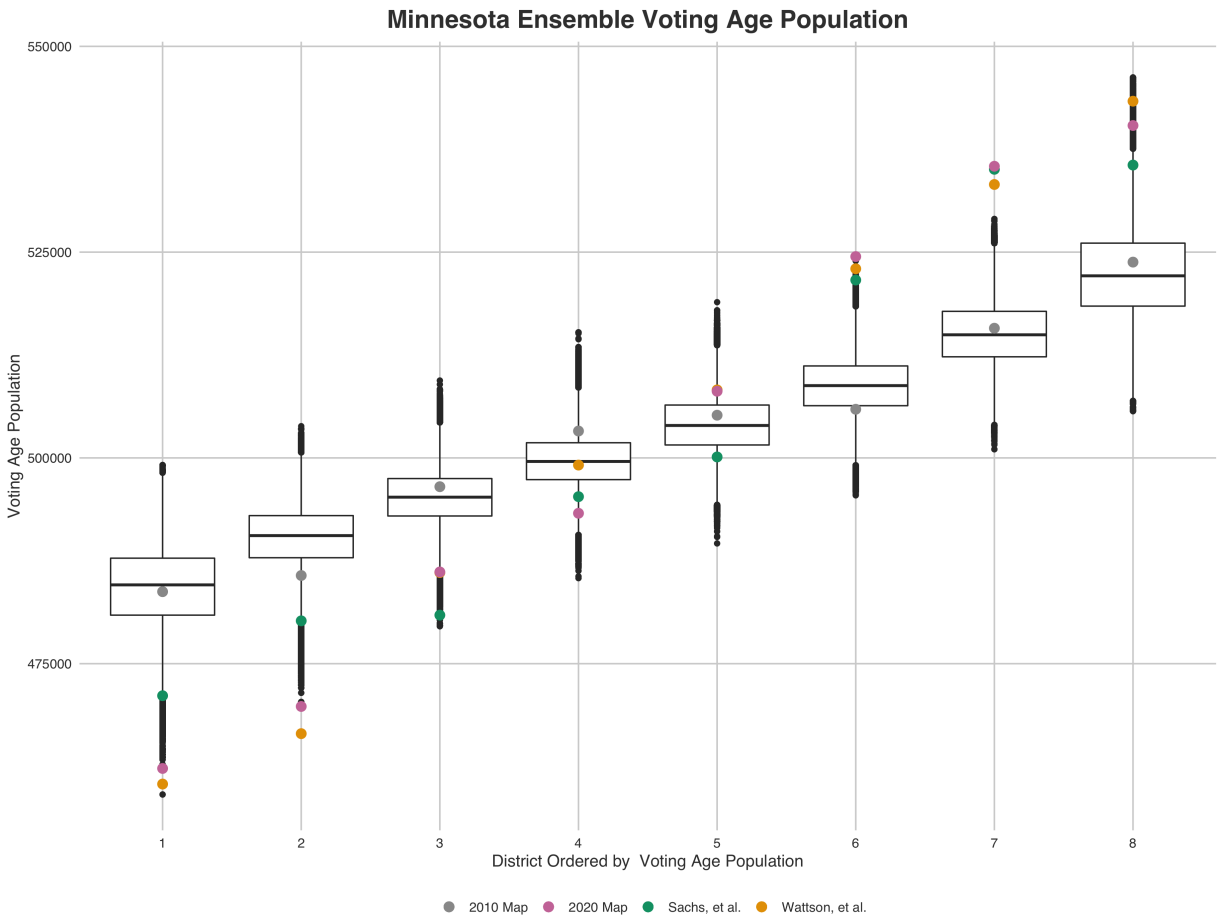


Figure 16: Box plot of voting age population by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2016 Presidential Election Results. Dots are placed at voting age population in each district of the the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan.

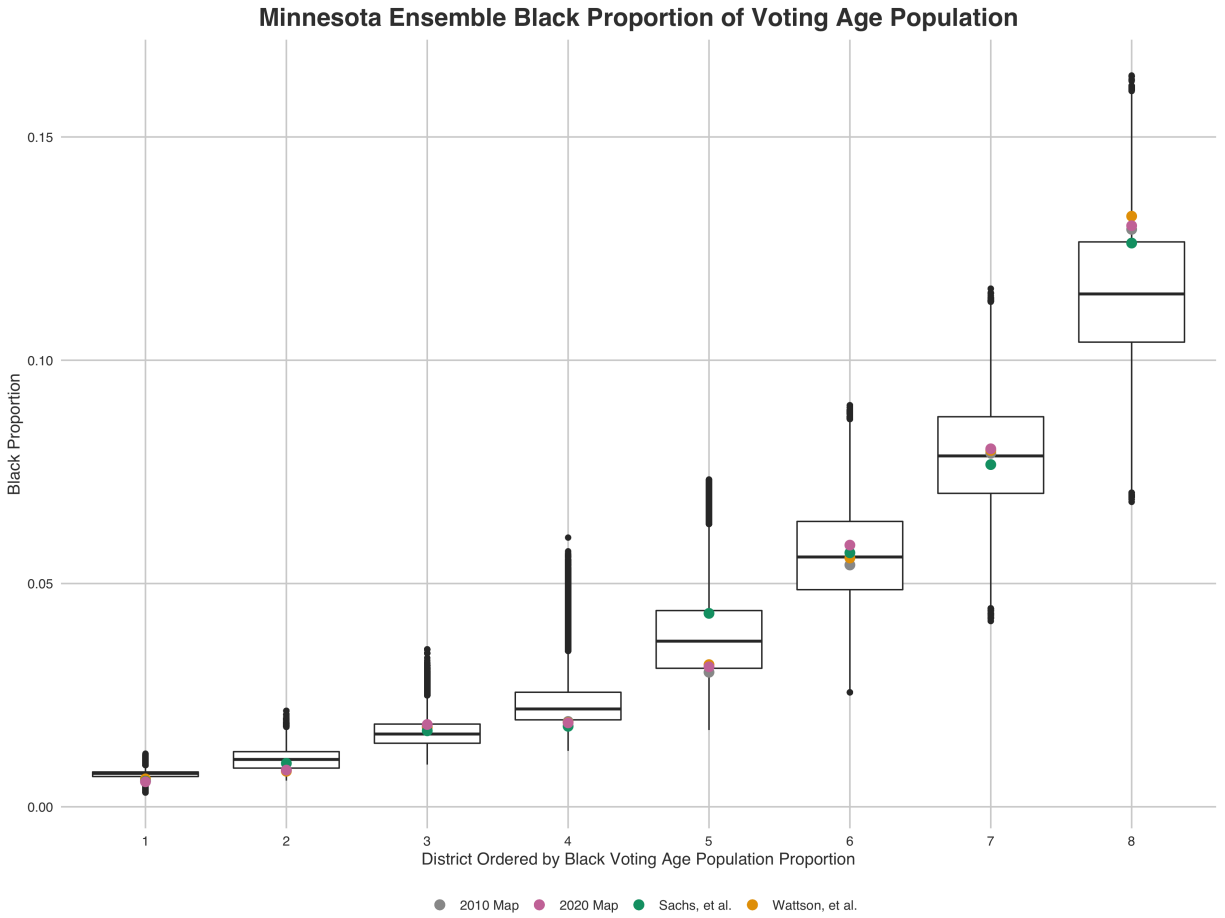


Figure 17: Box plot of Black voting age population by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2018 Senate Election Results. Dots are placed at Black voting age population for each district of the the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan.

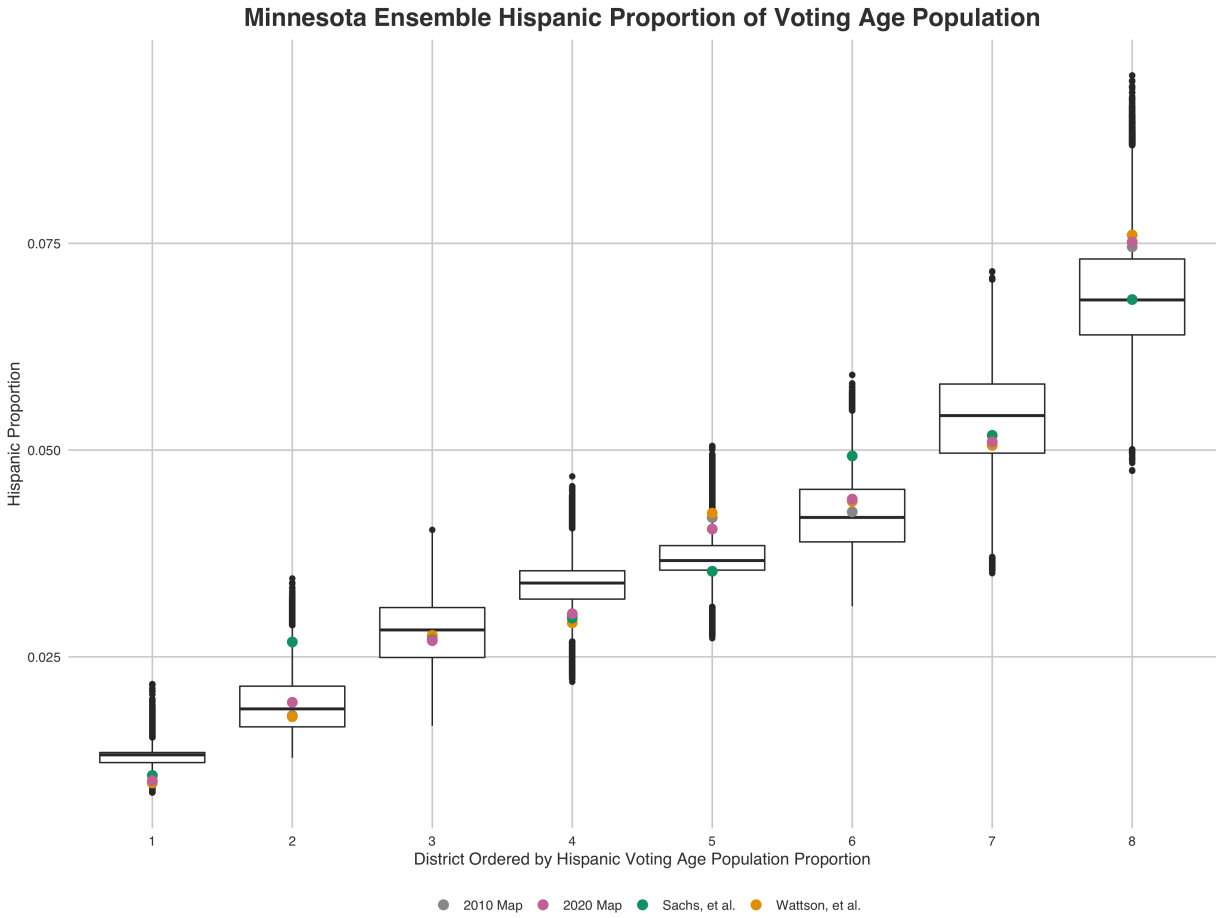


Figure 18: Box plot of Hispanic voting age population by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2018 Senate Election Results. Dots are placed at Hispanic voting age population for each district of the the 2010 map, 2020 map, Sachs Plaintiff plan, and Wattson Plaintiff plan.

B Texas Results Figures

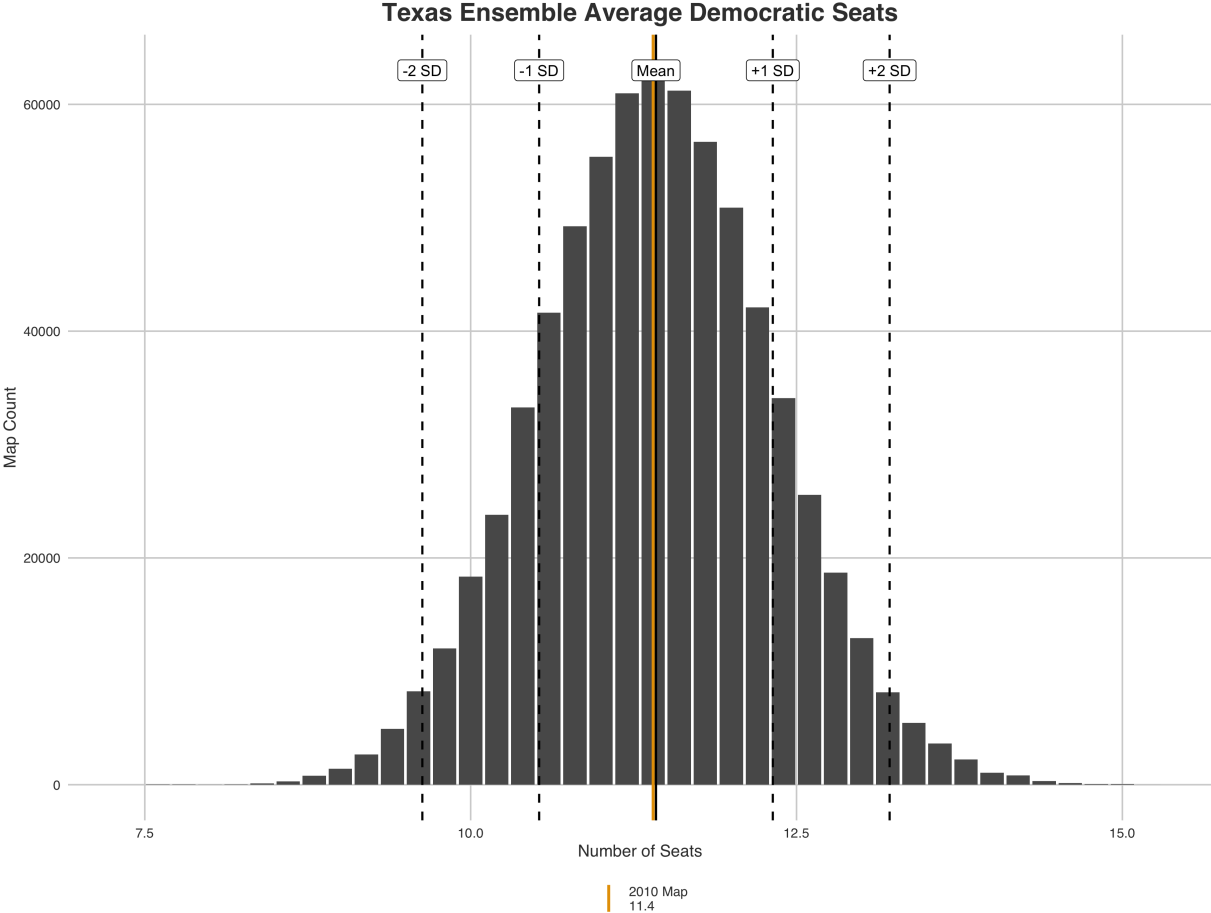


Figure 19: Distribution of average Democratic seats across the Texas ensemble. The mean falls at 11.42, with standard deviation of 0.896.

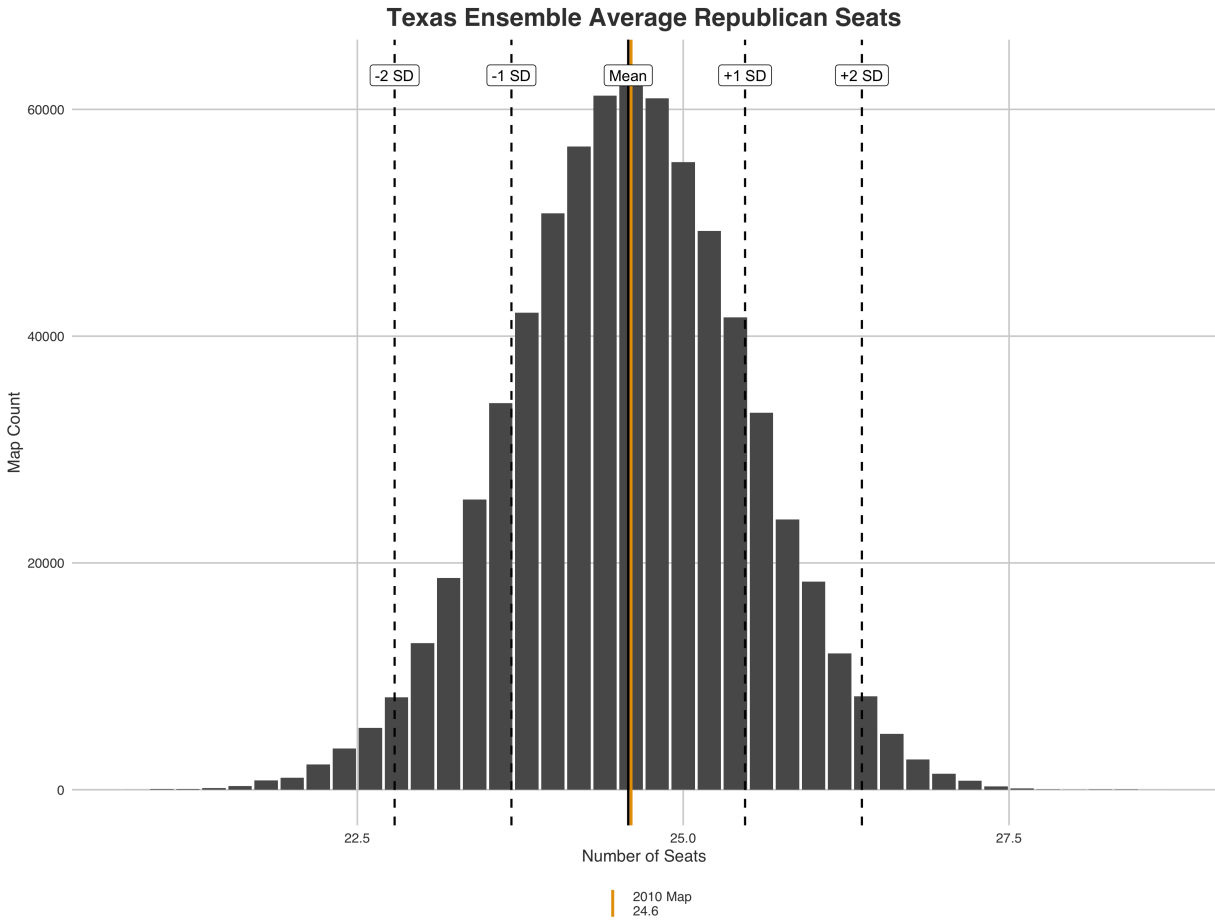


Figure 20: Distribution of average Republican seats across the Texas ensemble. The mean falls at 24.58, with standard deviation 0.896.

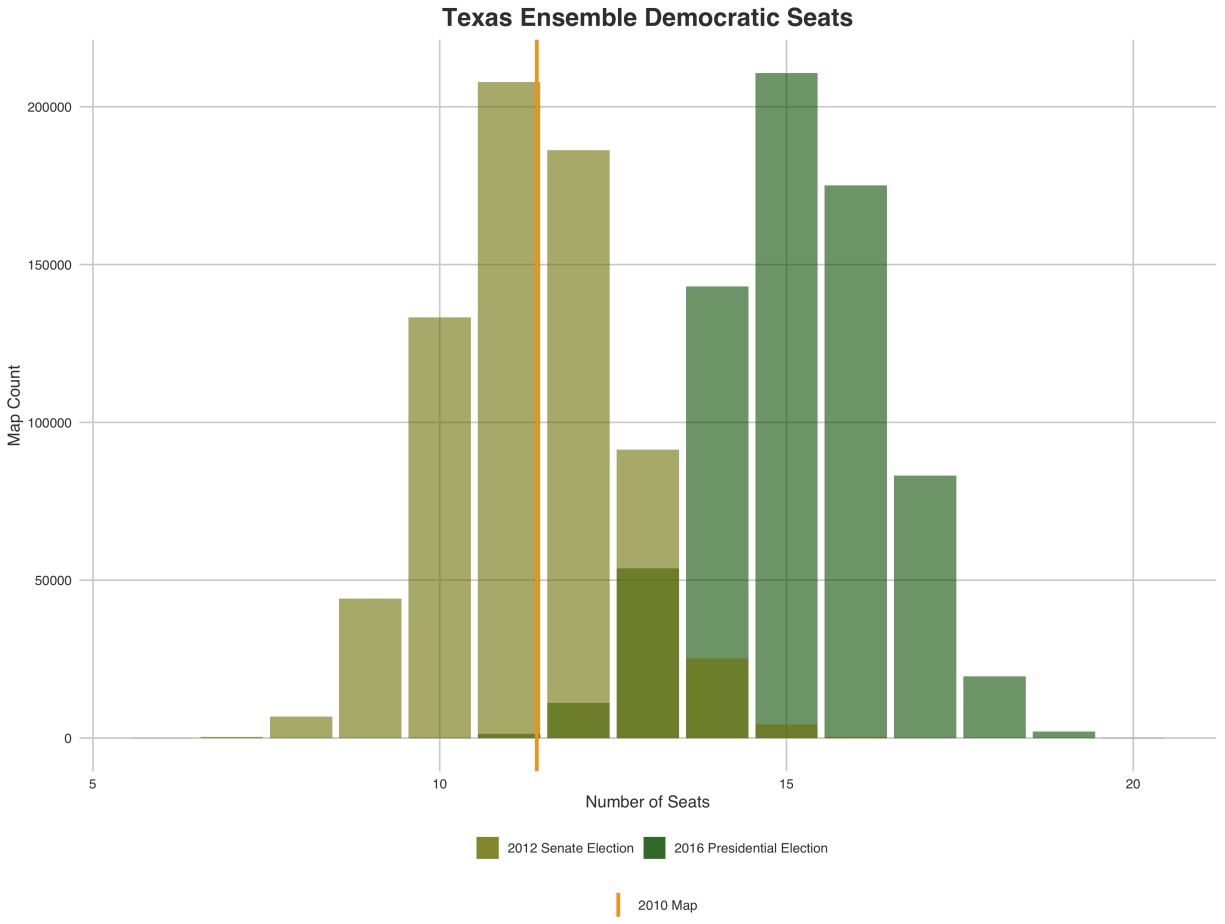


Figure 21: Distribution of Democratic seats across our ensemble for both the 2012 Presidential election and 2016 Presidential election.

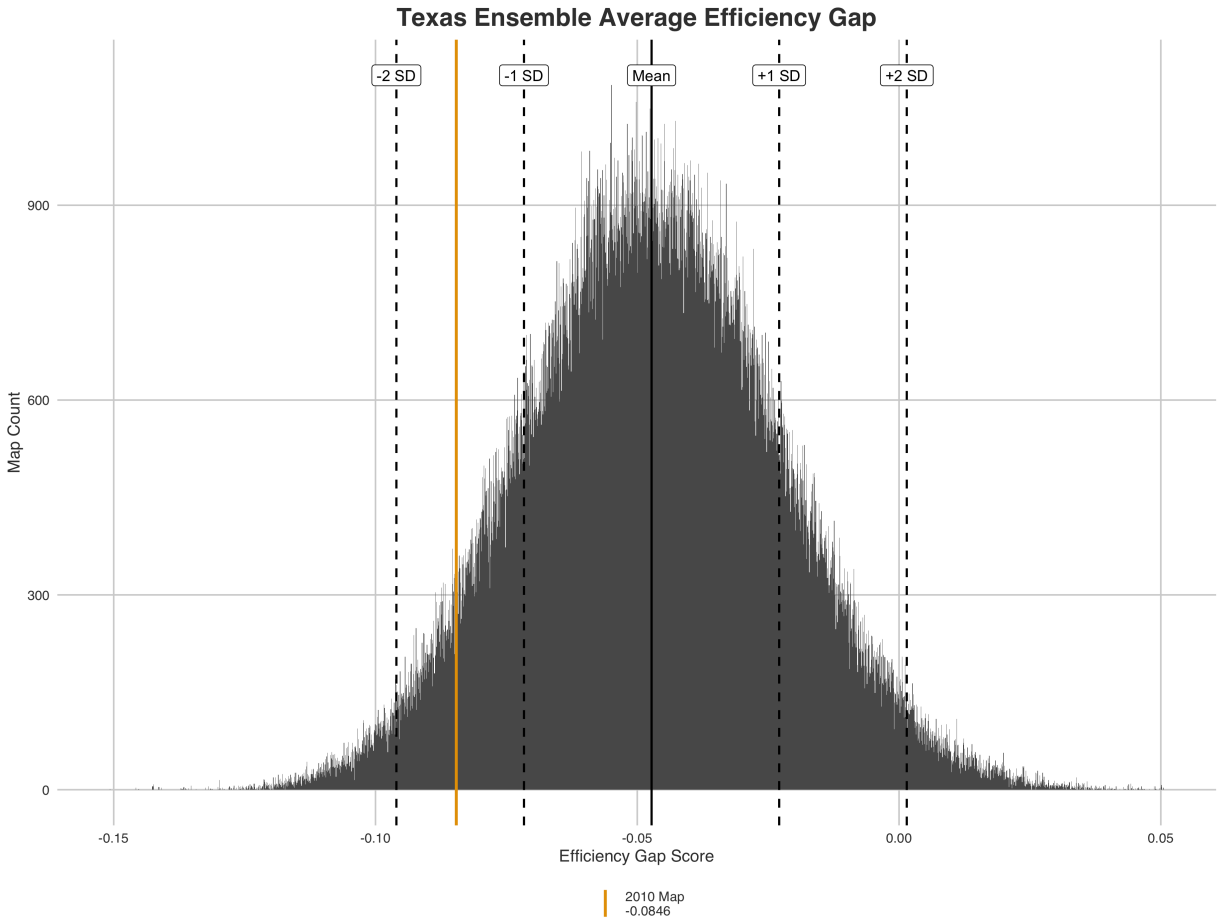


Figure 22: Distribution of Efficiency Gap Scores across the Texas ensemble. The mean falls at $-.047$. The line is placed at the efficiency gap value of the 2010 map.

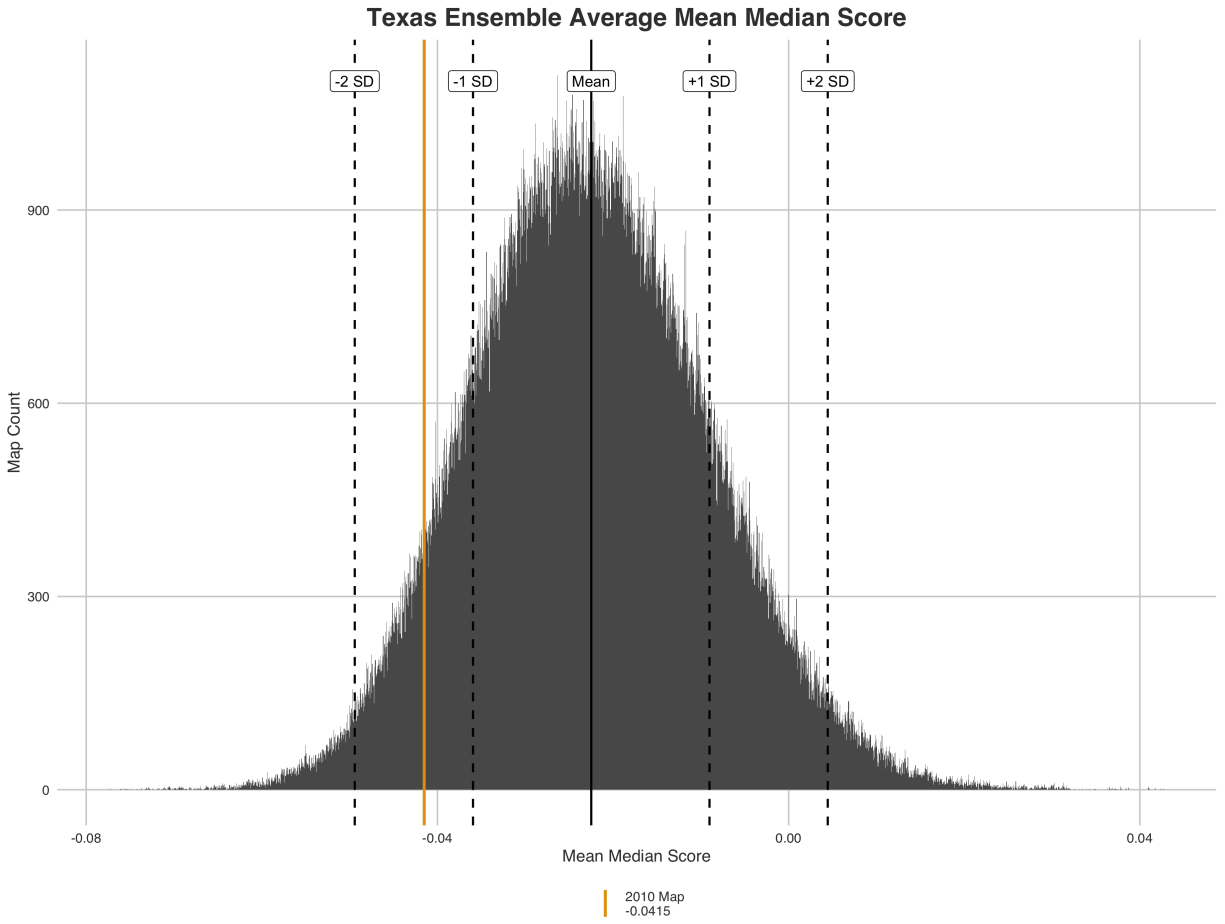


Figure 23: Distribution of Mean Median Scores across the Texas ensemble. The mean falls at -0.022. The line is placed at the mean median value of the 2010 map.

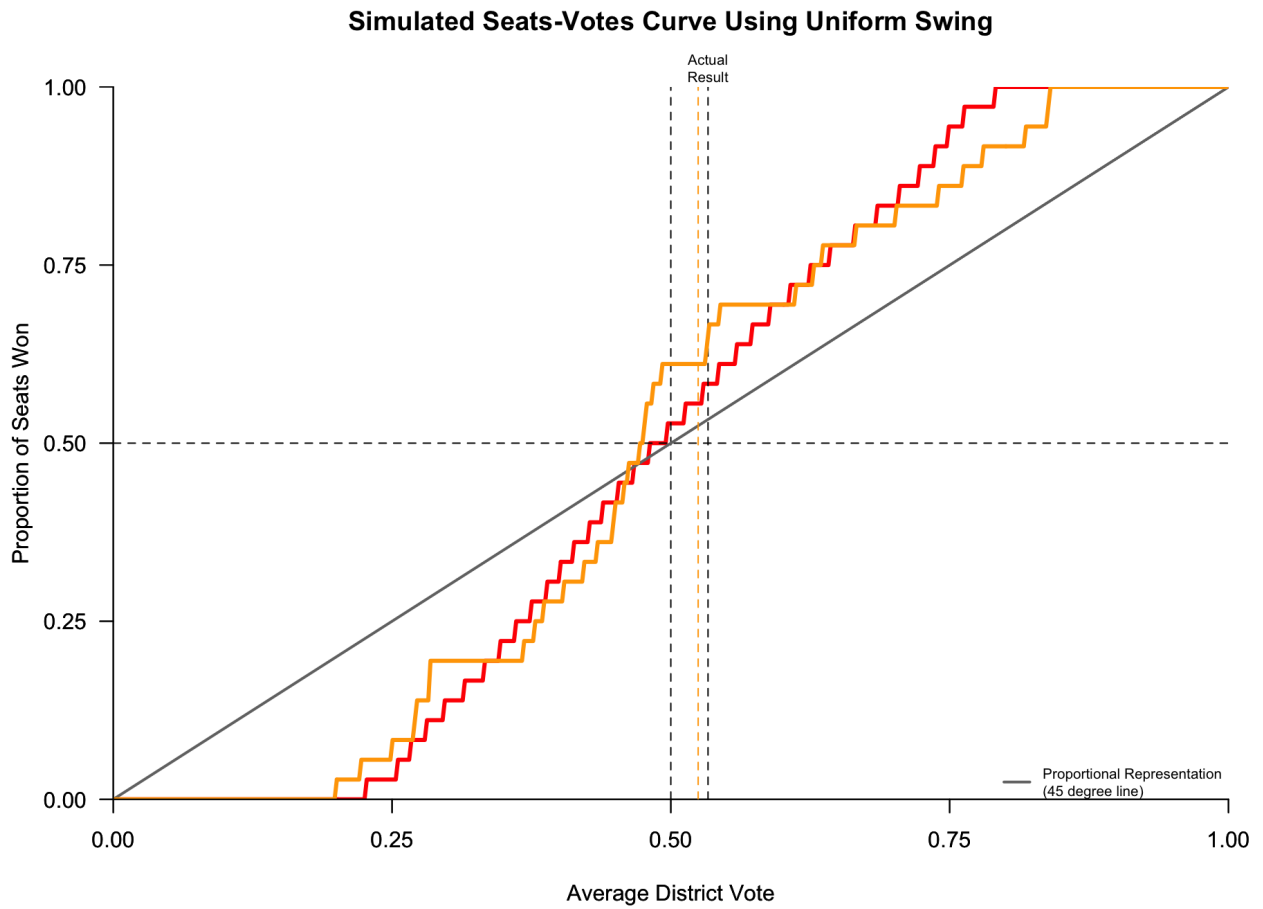


Figure 24: The average Republican seats-votes curve of our Texas ensemble (in red) compared with the Republican seats-votes curve for the 2010 map (in orange).

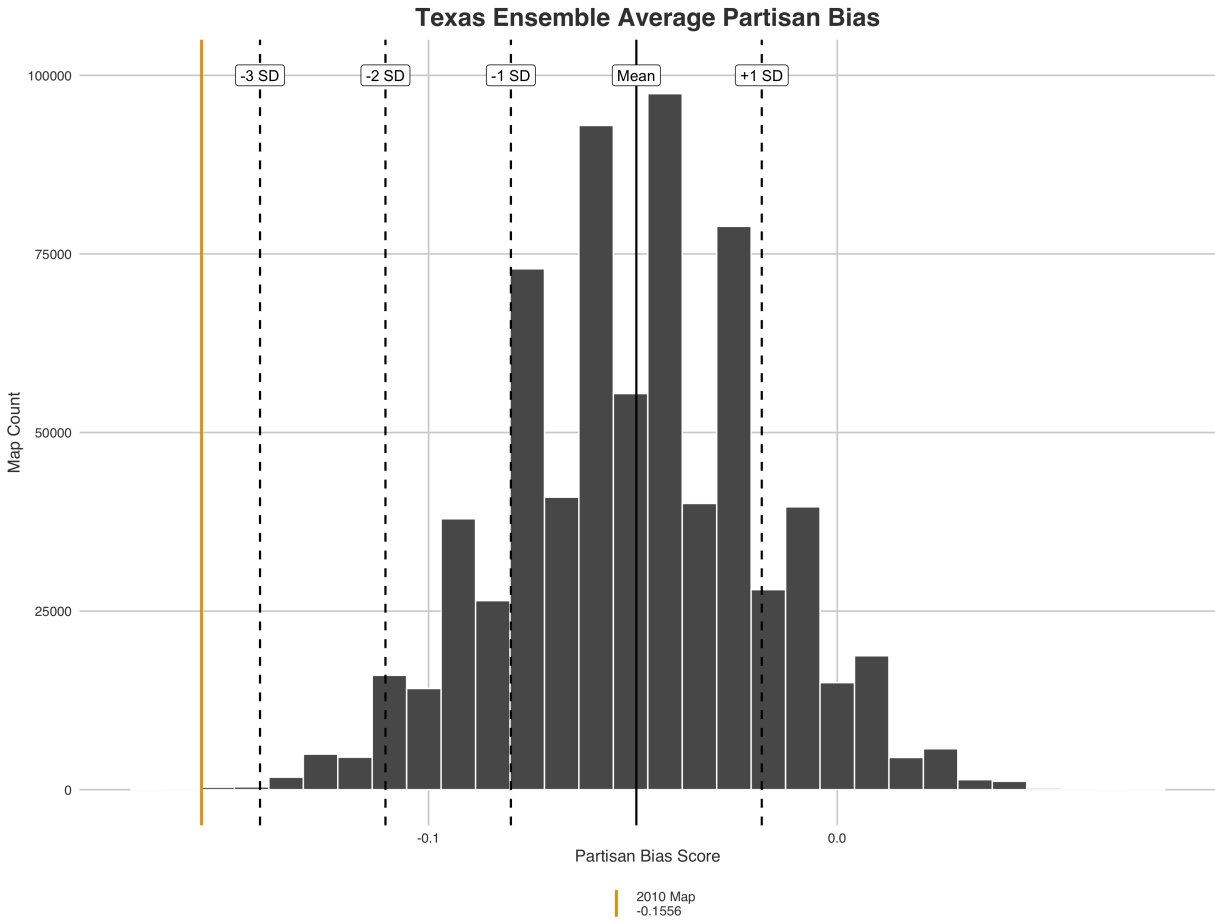


Figure 25: Distribution of Partisan Bias Scores across the Texas ensemble. The mean falls at -0.049 . The line is placed at the partisan bias value of the 2010 map.

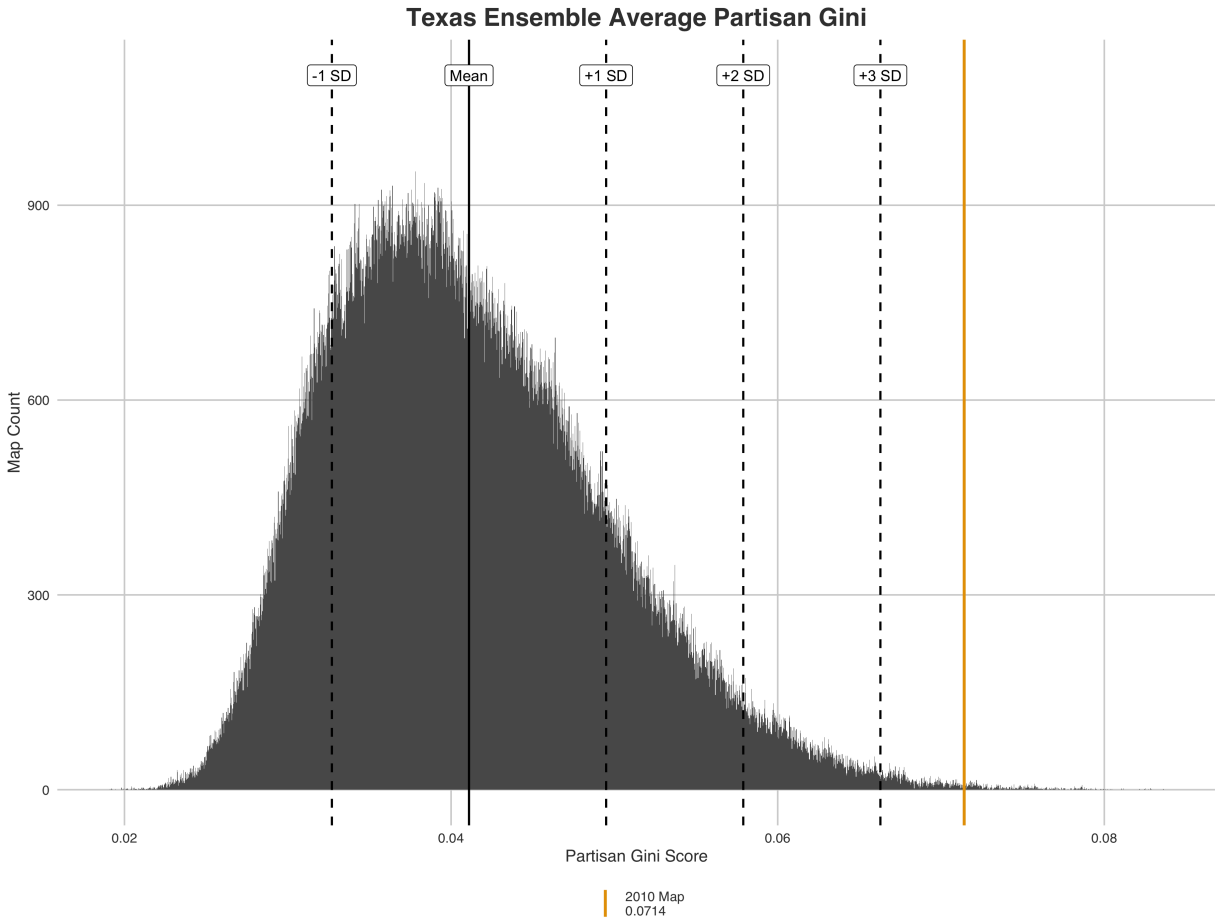


Figure 26: Distribution of Partisan Gini Scores across the Texas ensemble. The mean falls at .041. The line is placed at the partisan gini value of the 2010 map.

Texas 2016 Presidential Election Ensemble

Ordered from district with smallest Democratic vote share in each map to largest

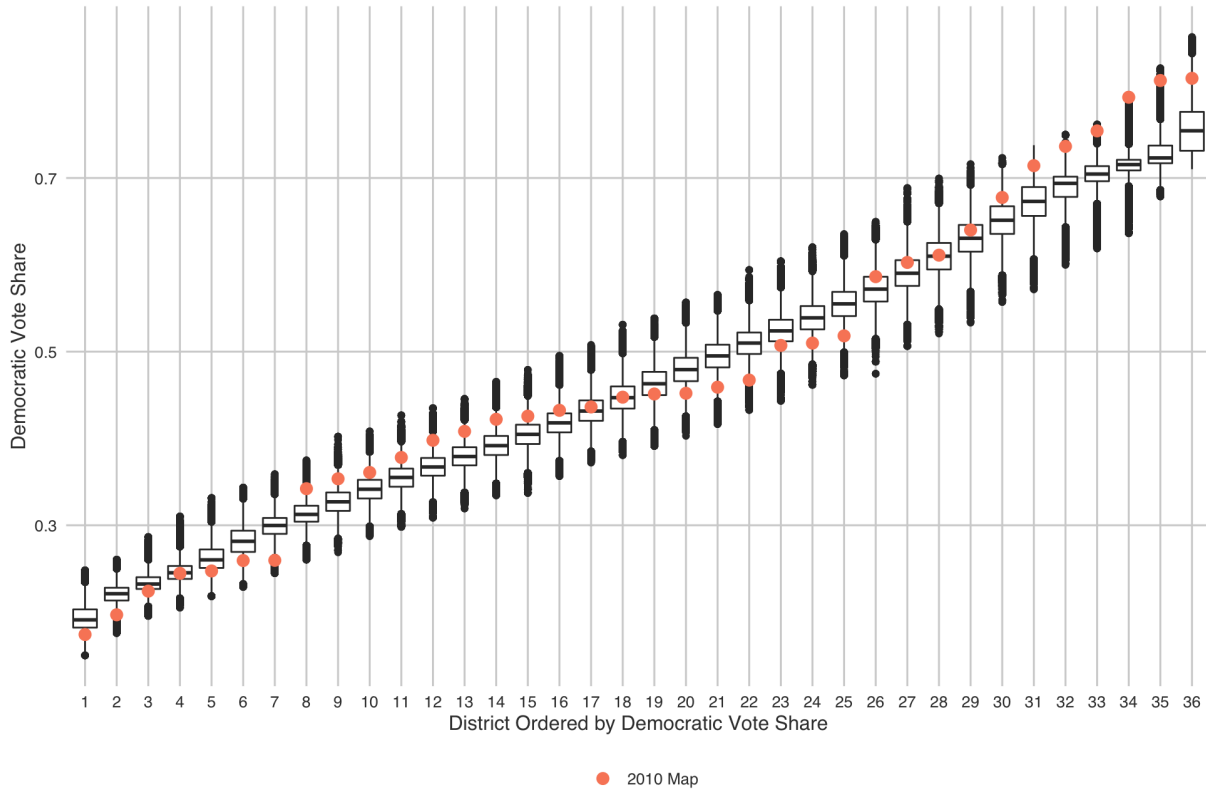


Figure 27: Box plot of Democratic vote share by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2016 Presidential Election Results. Dots are placed at Democratic vote shares in each district of the the 2010 map.

Texas Ensemble Voting Age Population

Ordered from district with smallest voting age population in each map to largest

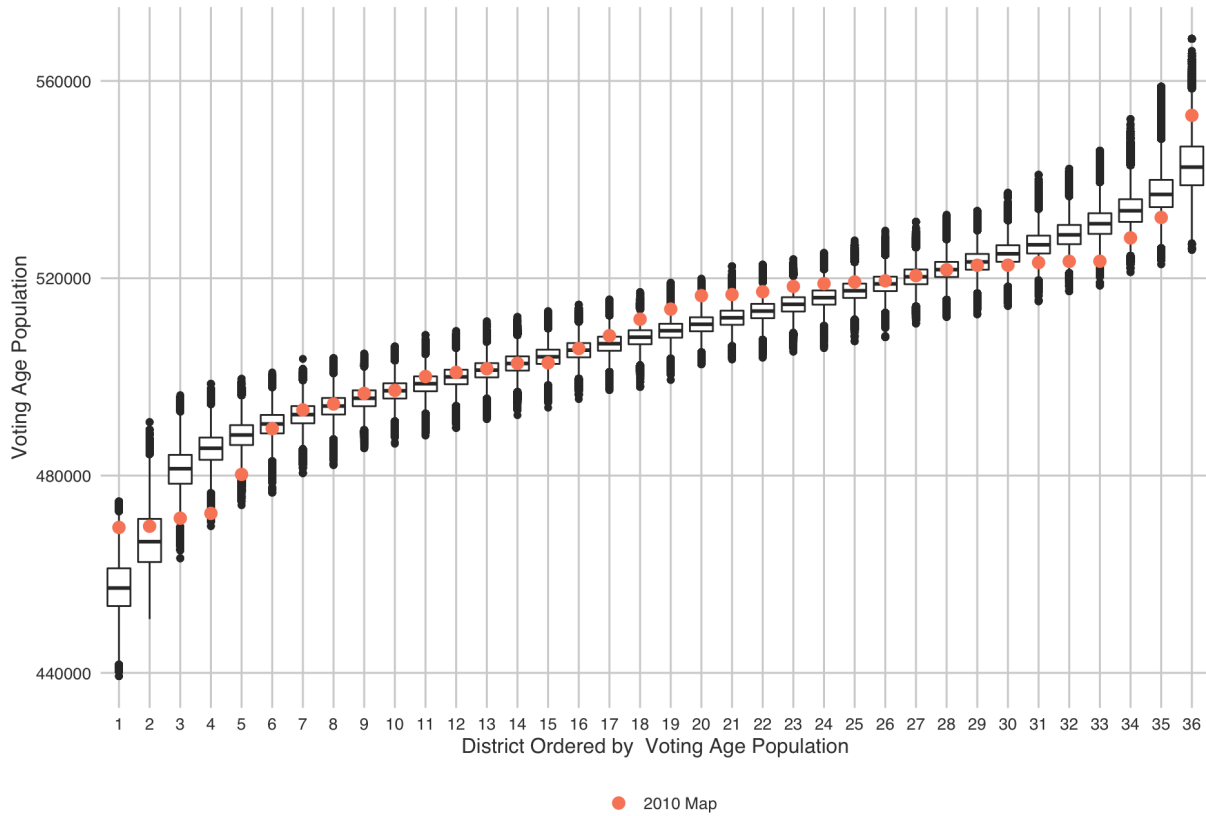


Figure 28: Box plot of voting age population by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2016 Presidential Election Results. Dots are placed at voting age population in each district of the the 2010 map.

Texas Ensemble Black Proportion of Voting Age Population

Ordered from district with smallest proportion of Black voting age population in each map to largest

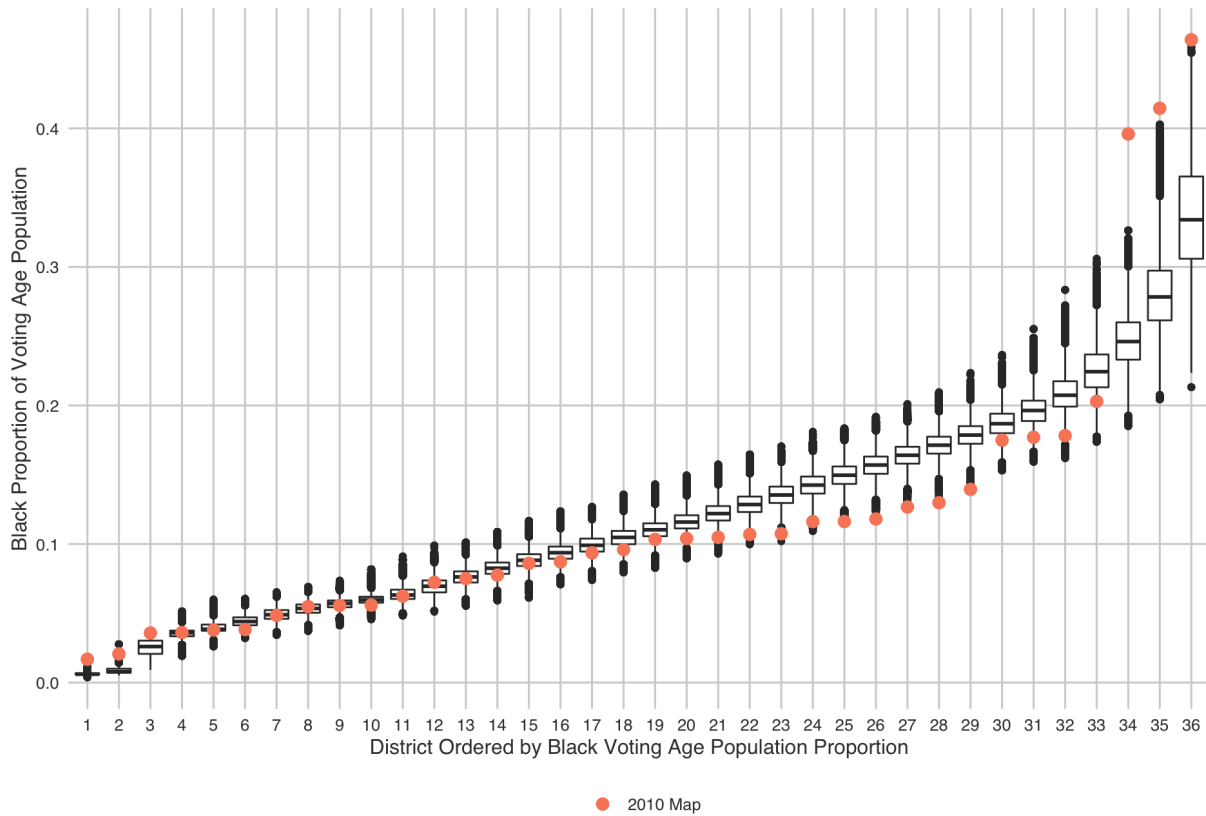


Figure 29: Box plot of Black proportion of voting age population by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2016 Presidential Election Results. Dots are placed at Black proportion of voting age population in each district of the the 2010 map.

Texas Ensemble Hispanic Proportion of Voting Age Population

Ordered from district with smallest proportion of Hispanic voting age population in each map to largest

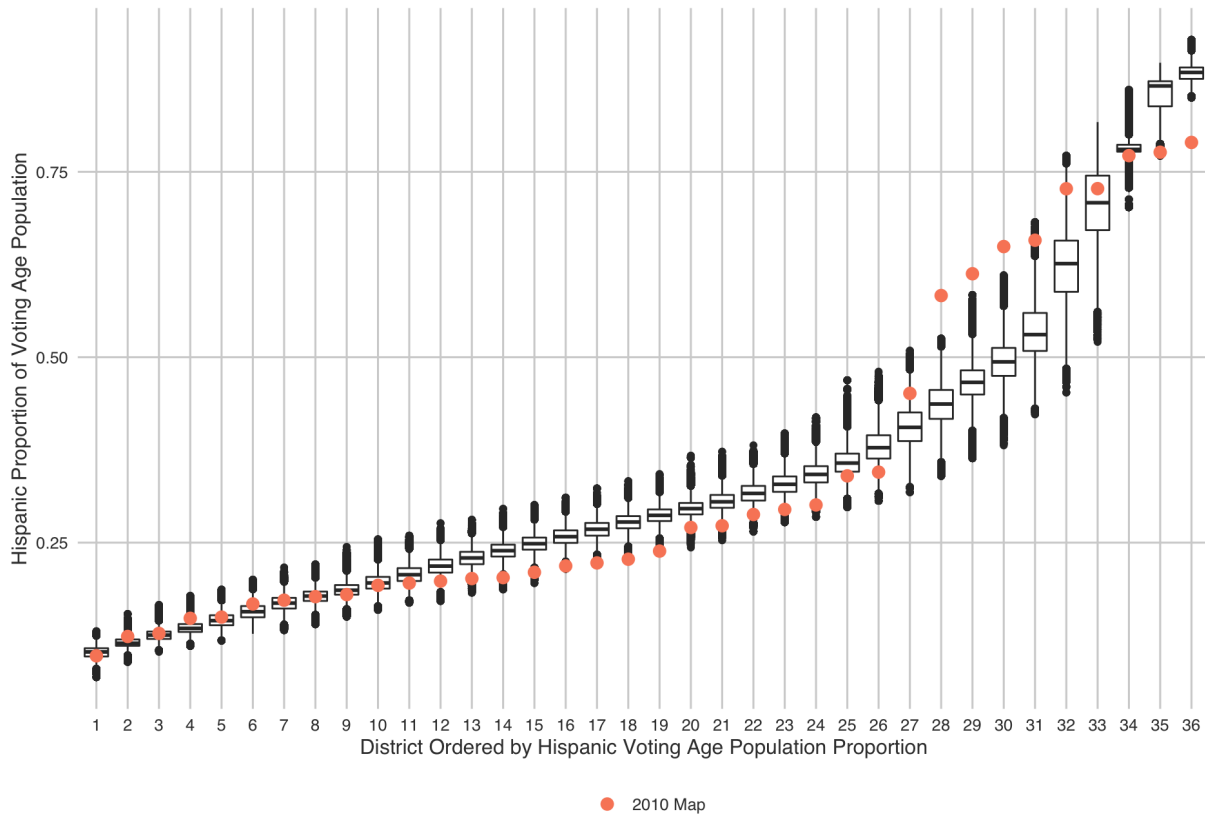


Figure 30: Box plot of Hispanic proportion of voting age population by district for each map, with the districts being ordered from least Democratic to most Democratic. Results are computed using 2016 Presidential Election Results. Dots are placed at Hispanic proportion of voting age population in each district of the the 2010 map.

C Minnesota Scores

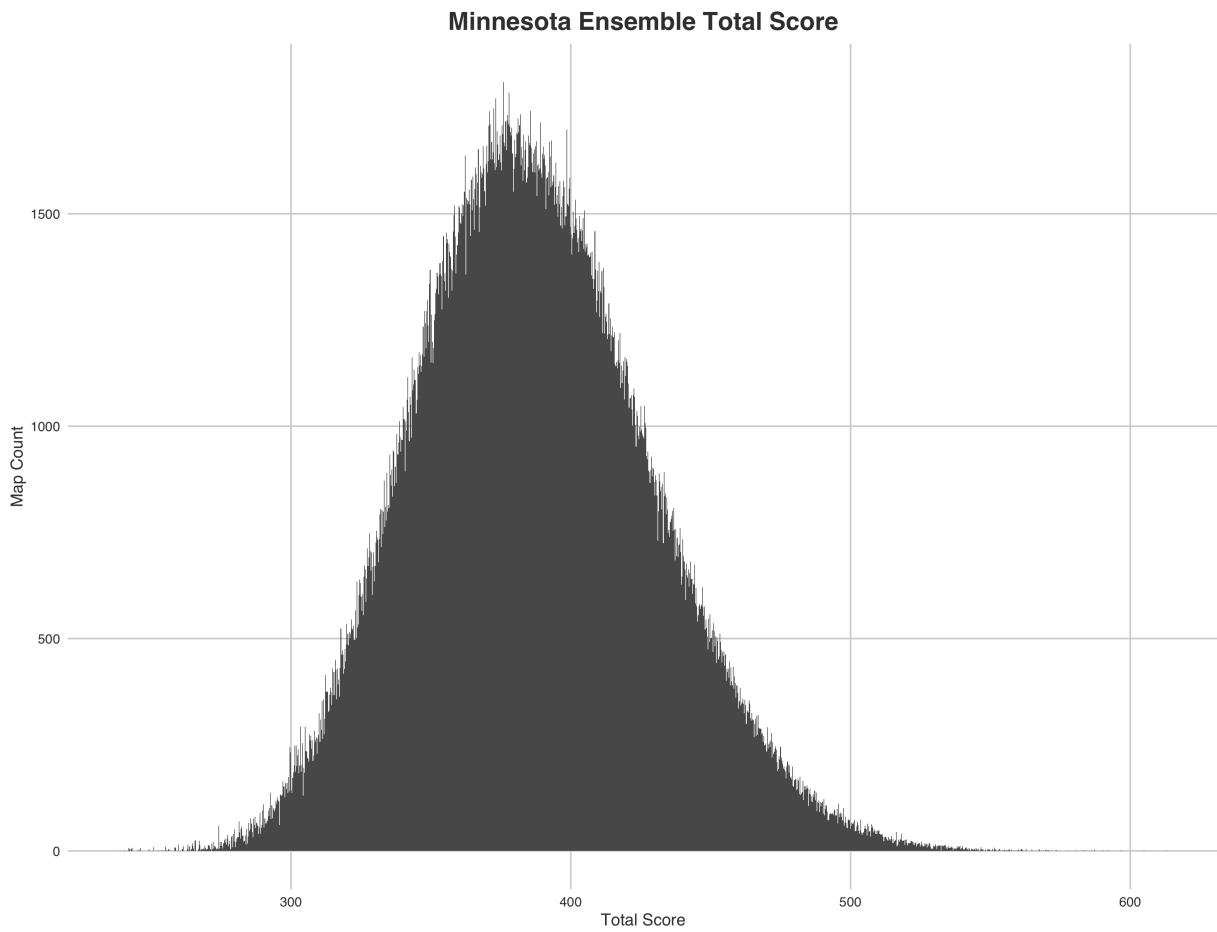


Figure 31: The distribution of total score $J(\varepsilon)$ for 700,005 used maps in our Minnesota ensemble. The mean falls at 386.8.

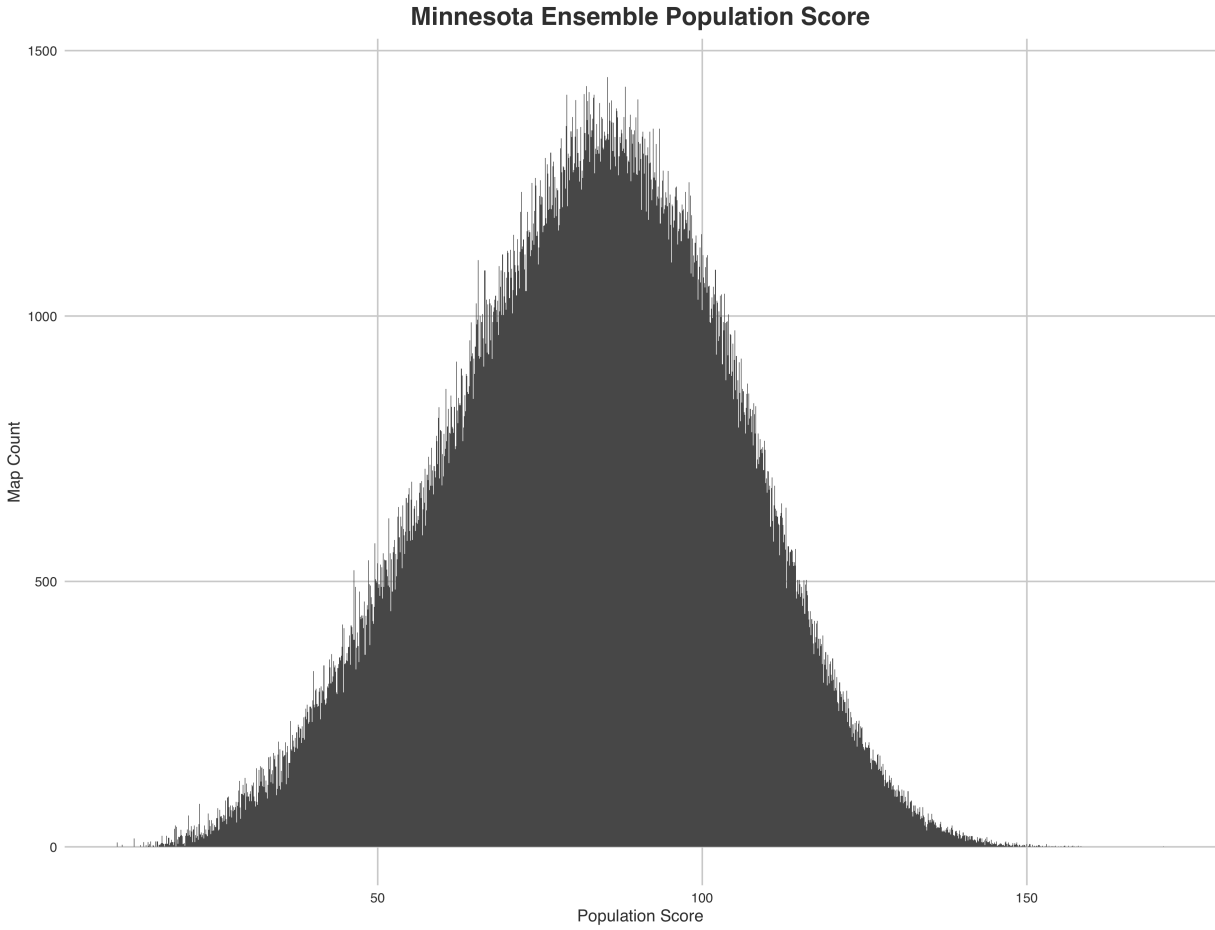


Figure 32: The distribution of population score $J_p(\varepsilon)$ for 700,005 used maps in our Minnesota ensemble. The mean falls at 82.5.

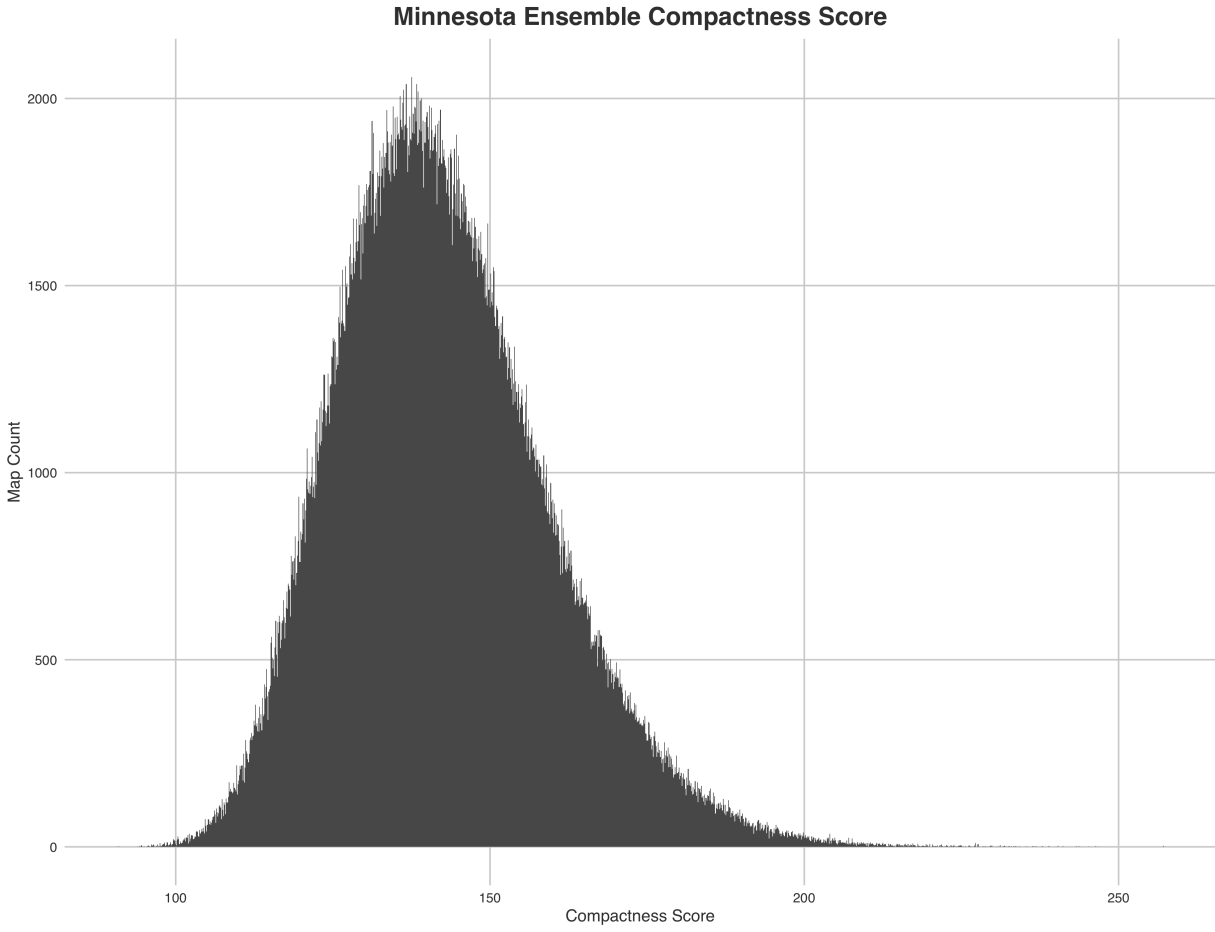


Figure 33: The distribution of compactness score $J_c(\varepsilon)$ for 700,005 used maps in our Minnesota ensemble. The mean falls at 142.6.

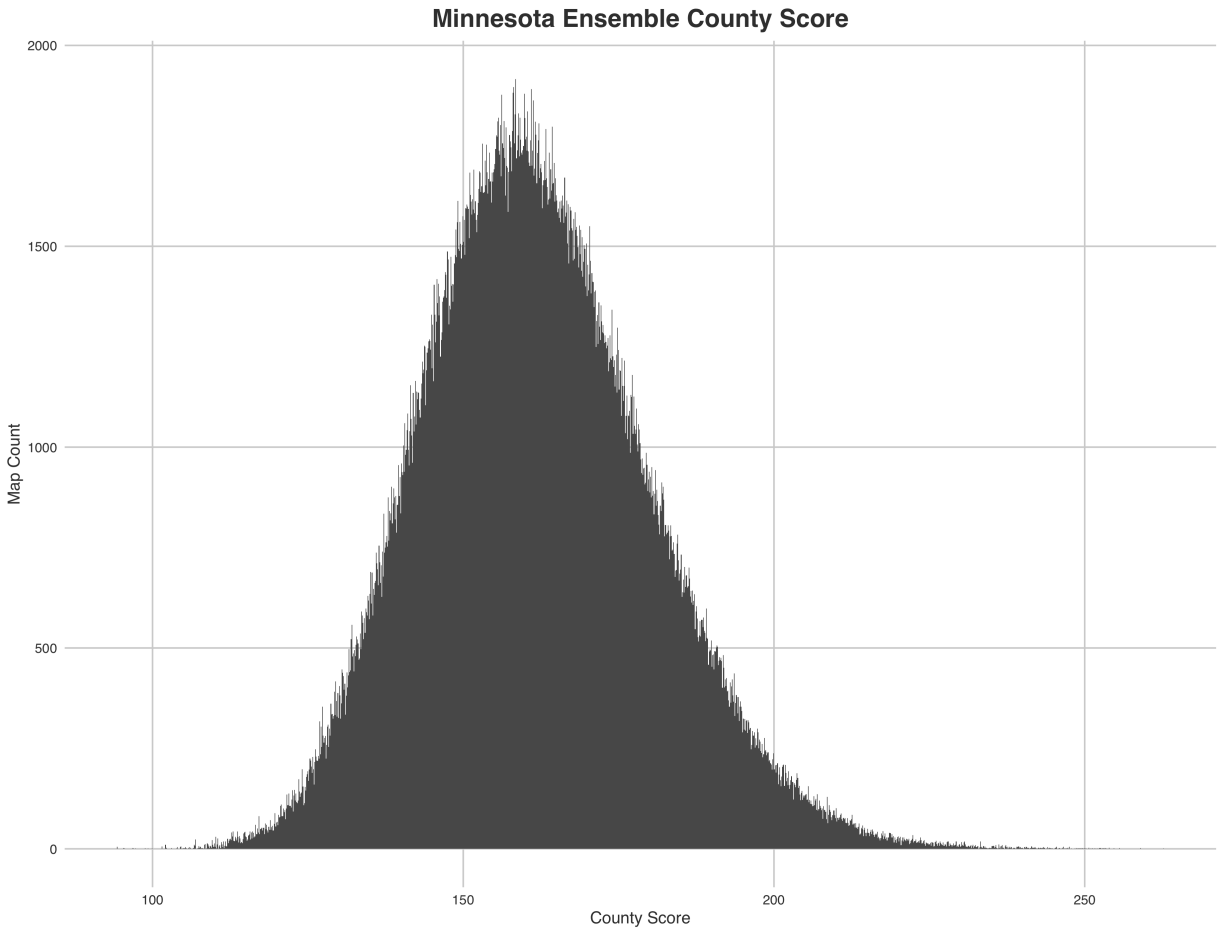


Figure 34: The distribution of county score $J_c(\varepsilon)$ for 700,005 used maps in our Minnesota ensemble. The mean falls at 161.8.

D Texas Scores

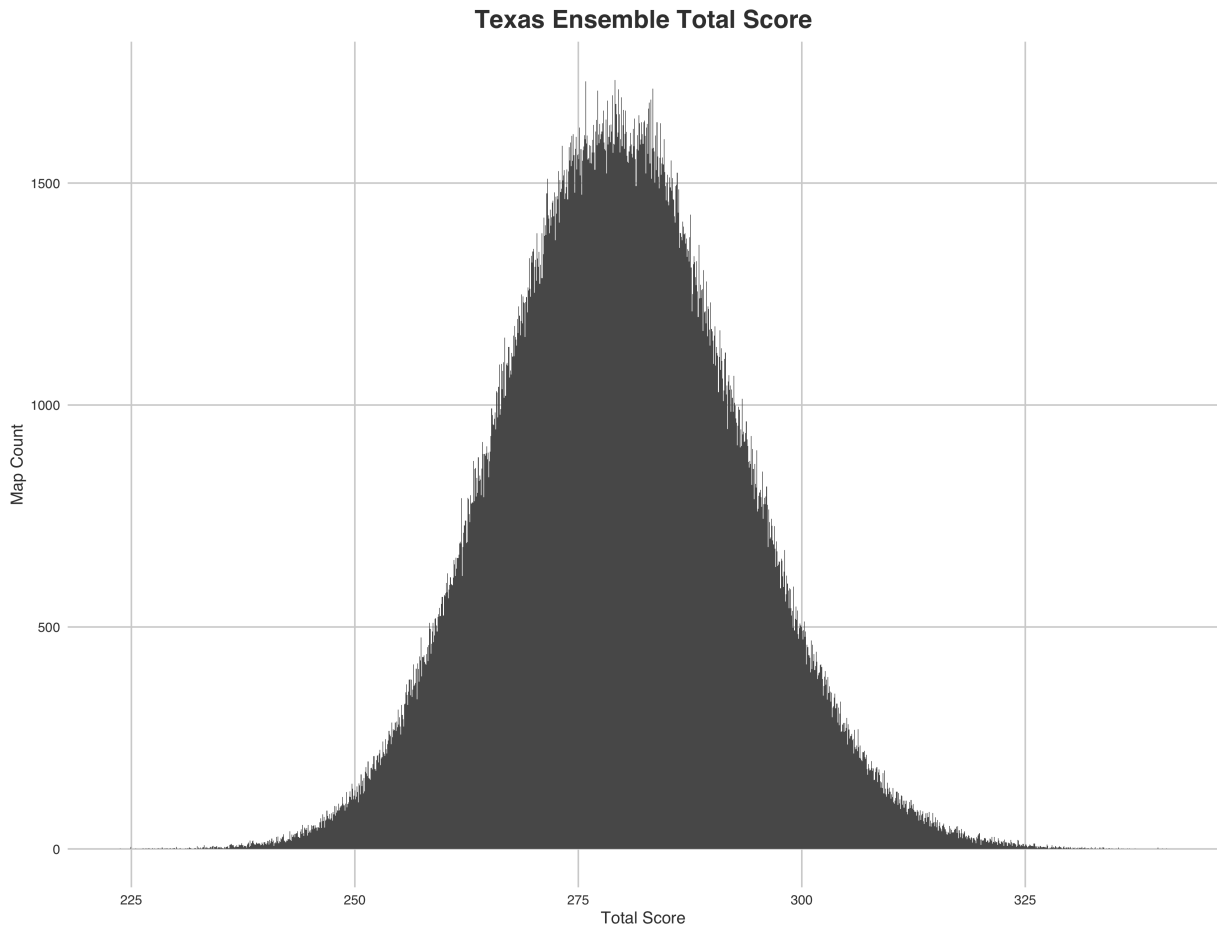


Figure 35: The distribution of total score $J(\varepsilon)$ for 700,005 used maps in our Texas ensemble. The mean falls at 279.6177.

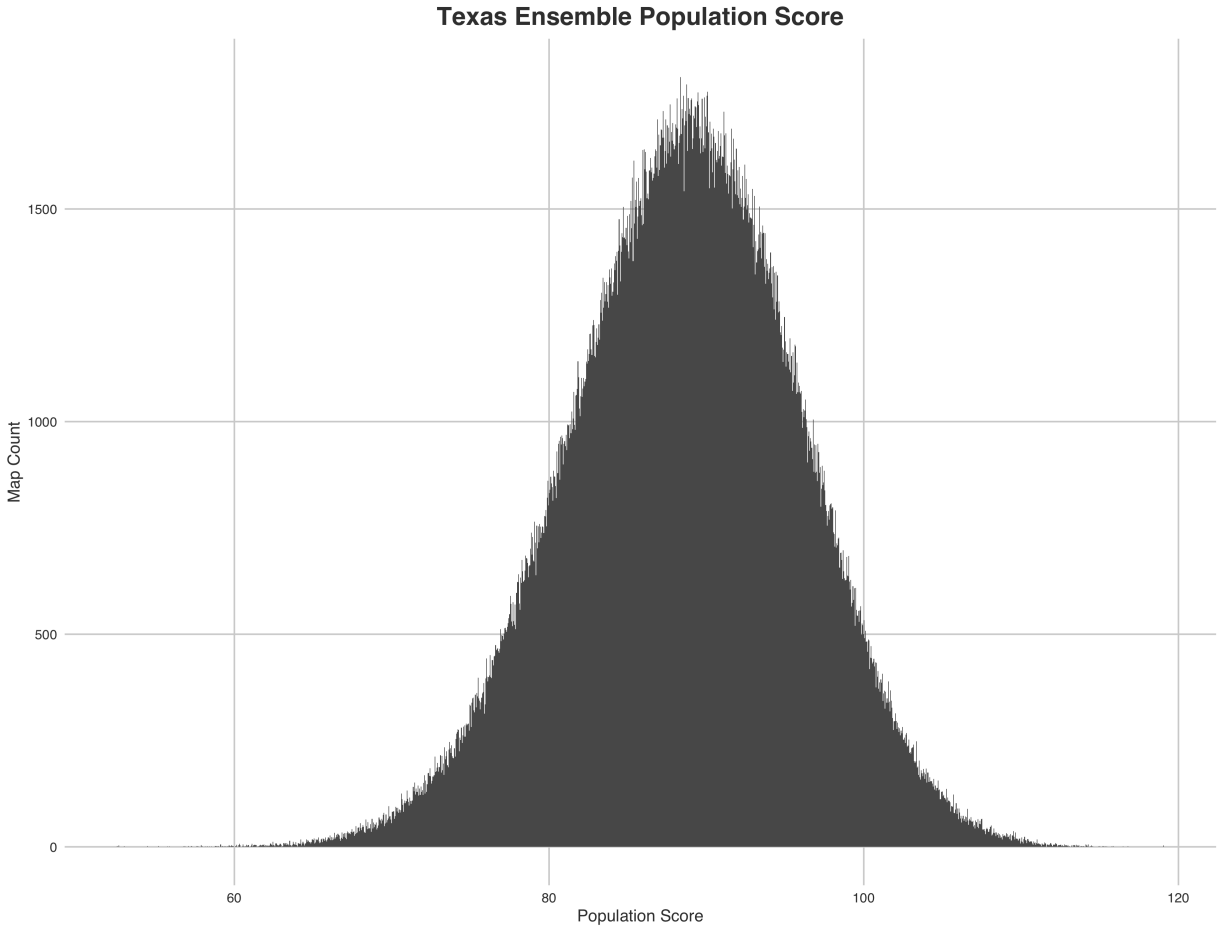


Figure 36: The distribution of population score $J_p(\varepsilon)$ for 700,005 used maps in our Texas ensemble. The mean falls at 88.5613.

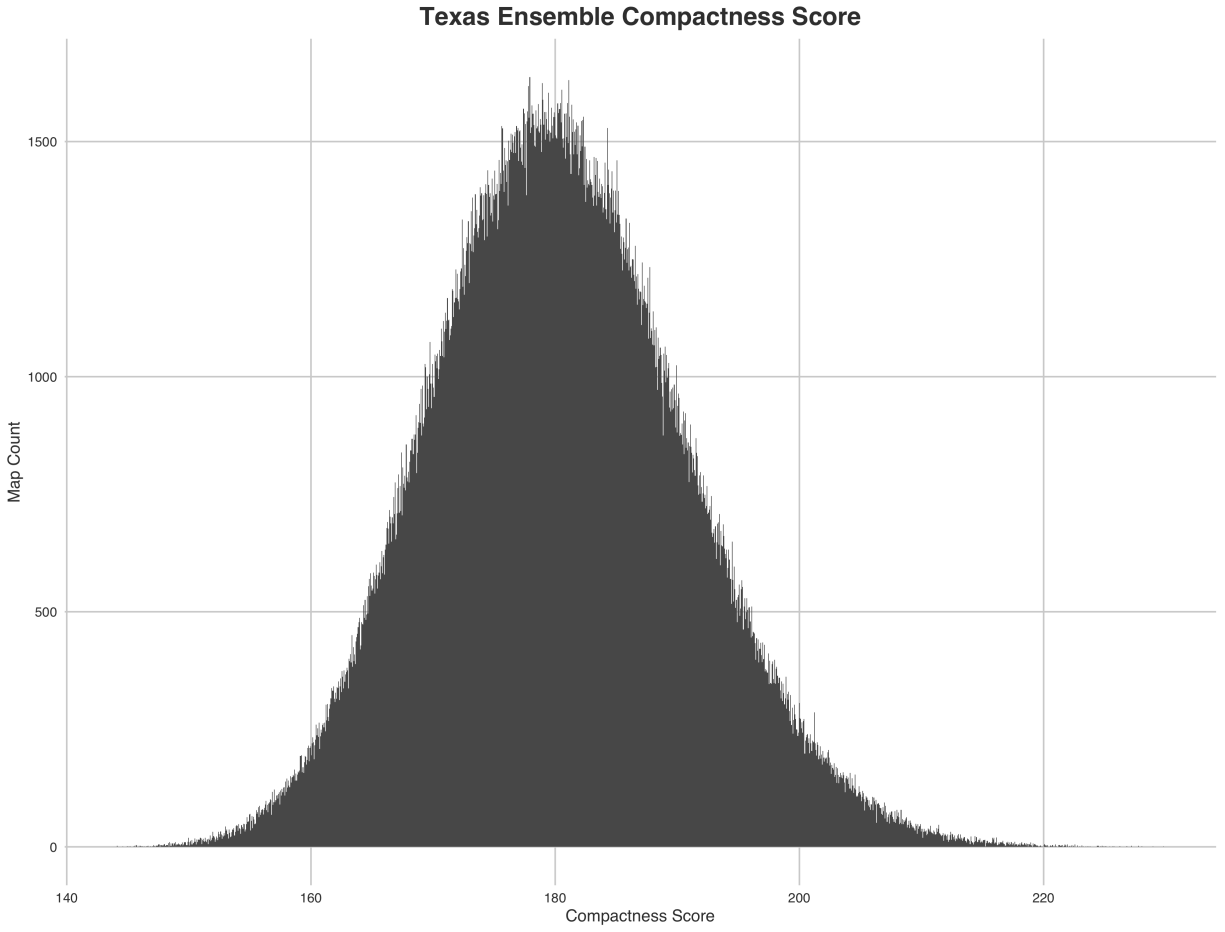


Figure 37: The distribution of compactness score $J_c(\varepsilon)$ for 700,005 used maps in our Texas ensemble. The mean falls at 180.3355.

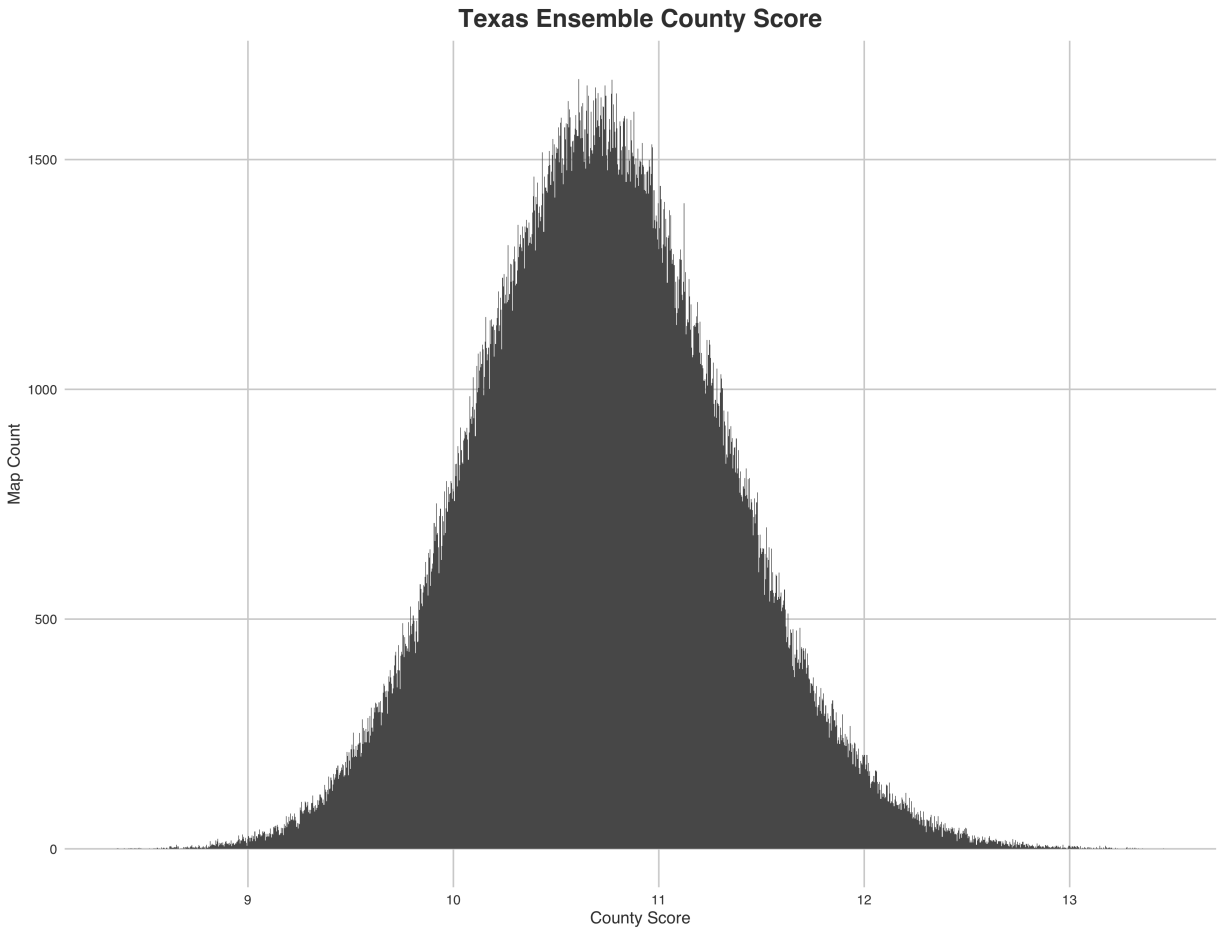


Figure 38: The distribution of split counties score $J_2(\varepsilon)$ for 700,005 used maps in our Texas ensemble. The mean falls at 10.72092.