From:	applications.administrator@capitol.local
To:	Senate Redistricting
Subject:	INETMAIL: Redistricting Public Input
Date:	Tuesday, September 28, 2021 1:15:45 PM
Attachments:	ExplainGraphs_v2.pdf

Date: 2021-09-28
First Name: Andrea
Last Name: Barreiro
Title: N/A
Organization: Math For Unbiased Maps TX (MUM_TX)
Address: [redacted]
City: Dallas
State: TX
Zipcode: [redacted]
Phone: [redacted]

Affirm public info: I agree

Regarding: ALL

Message:
Math For Unbiased Maps TX (MUM_TX): Methods

Now that 2020 Census results have been released, redistricting has begun. Due to the size and complexity of Texas:

1. The number of possible ways to draw voting district plans is enormous.
2. It can be time-consuming to create even a single legally-valid plan that obeys constraints like population-balance and the county-line rule.

It is easier to draw and evaluate valid plans if you have examples to work from. However, it has not been feasible to create a large and representative ensemble of districting plans in prior redistricting cycles with the algorithms and computers available at the time. But it is feasible now. With ensemble sampling, we can generate an unbiased baseline which shows us what we can expect of a “typical map”, and conversely detect when a proposed map is an outlier which shows evidence of gerrymandering.

We are Math For Unbiased Maps TX (MUM_TX), an interdisciplinary, nonpartisan coalition of Texas mathematicians, political scientists and philosophers working to ensure a fair and transparent redistricting process. Our research concerns the development and application of ensemble sampling techniques, and in particular their application to the current TX redistricting cycle. While our research is primarily academic and thus concerned with the production of knowledge, we also believe that knowledge that affects the public, should be made available to the public. That is why we have committed to making our findings available to the general public and legislators in an accessible way, in the service of increasing transparency and public engagement in the process.

We have posted results to the TX Senate comment portal; in these notes, we seek to explain our methods. An attached PDF includes figures, which we will refer to throughout. We have illustrated our methods using results from PLANS2101, which was filed on September 18, 2021. However, we are continuing to update our website with analysis of the many amendments that have been filed since. We will also be analyzing the US Congressional map and any subsequent amendments; and when the time comes, we’ll do the same for the TX House.

You can find us at:
https://www.smu.edu/Dedman/Research/Institutes-and-Centers/DCII/Scholarship/Research-Cluster-on-Political-Decision-Making

What is Ensemble Sampling?

In general, ensemble sampling refers to any process of generating a large number of samples from a large collection; in this case, the set of all possible districting plans. Usually there is an element of randomness, because the collection is too large to look at every member. We can formalize the process of making random transitions between district plans with a mathematical object called a Markov chain. The idea of using transitions between states of a Markov chain to solve high-dimensional problems, or Markov Chain Monte Carlo, was pioneered in statistical physics, but is now used in nearly every area of science and engineering. In the current context, a “state” of the Markov chain will be a districting plan: that is, a division of a political entity into a specified number of subdivisions.

How do we sample?

There are many ways of moving between states in a Markov chain. We use ReCom (short for recombination) in which two adjacent districts are combined into one, and then split into two again, in such a way that legal constraints (such as equal population) are satisfied (D. DeFord, 2021). This is implemented in software called GerryChain, which is a Python package developed by the Metric Geometry and Gerrymandering Group (MGGG), a Boston-based team of mathematicians and computer scientists working on U.S. redistricting, and is available on GitHub (Metric Geometry and Gerrymandering Group, 2018)
(See Figure 1 in the attached PDF for a schematic)

How did we analyze the proposed TX Senate Map?

We used GerryChain to create an ensemble of TX Senate districting plans, using the 2020 Census data released on August 12, 2021. This enabled us to assemble total populations and voting age populations (VAP) in each Census block in different racial groups: for example, Black and Hispanic. We began with a graph that contains the appropriate geographic nodes (all counties that were whole in the proposed plan, VTDs for those counties that were not whole, and split VTDs, for those VTDs that were split in the proposed plan). Population numbers were computed for each geographic unit.

We then generated an ensemble of 500,000 plans, which allows a robust exploration of the space of possible plans. (We can validate with a larger simulation). The constraints we applied at each ReCom step were:

- the population of each district had to be with 3% of the ideal balance. This was the maximum deviation observed in the proposed plan (S2101).
- The number of cut edges (the number of edges in the dual graph whose vertices lie in different districts) had to be less than twice the number in the original plan. This is a discrete proxy for perimeter. In practice, the number of cut edges for the plans we found was less than the number in the enacted plan.
- No additional “county splits” were allowed: if a particular county was entirely contained inside one district in the proposed plan, it must continue to be so for all other alternative plans.

Each step reflected a single districting plan for which we can then compute a list of statistics. For each plan, and each district, we collected the percentage of the voting age population that was Hispanic (HVAP), the percentage of the voting age population that is Black (BVAP), and the percentage of voting age population that is either Black or Hispanic (BHVAP). We also collected the percentage of the vote (among Democratic vs. Republican) from the 2020 Presidential race. In the attached PDF, we show maps illustrating some of the plans we generated.

Visualizations

Next, we will explain some of the visualizations which can enable us to understand how a proposed plan compares to an unbiased baseline.

For each graph, we have included a comparison with the Proposed plan (PLANS2101). For some, we have also included the current plan enacted in 2011 (PLANS2100). See our attached PDF for figures.

- The first figure illustrates the vote share vector, which is just a list of what percentage of the vote was obtained by a party in each district, except that it is sorted by that vote share, instead of by district number. The vote share vectors from the 500,000-map ensemble are illustrated with violin plots (Figure 3). Here the blue "violins" illustrate the distribution of likely vote shares for a district with a given vote share ranking, along with its median (blue dot) and 5% and 95% quantiles (blue bars).

- When we also plot the ordered vote shares for the proposed map (red circles), we obtain a concise visual summary of how gerrymandering is used to distribute votes in order to insulate one party from swings in public opinion. In the range labeled "cracking" Democratic vote shares have been decreased to an improbably low level (compared to the random ensemble), which has the effect of making these seats significantly harder to flip, even if there is a significant shift in public opinion.
- Democratic votes appear to have been concentrated in districts that fall in the range labeled "packing", especially into two overwhelmingly Democratic seats.
- We have identified, in purple, some districts of interest.

- The next figure shows two common numbers that political scientists use to "score" maps. The first such number is called the mean-median score: it is proportional to the difference in statewide vote percentage each party would need to win the majority of the chamber.
- For the proposed map, the Republican Party would need to win only 44.2% of the vote to win 16 seats, while the Democratic Party would need to earn 55.8%; the difference of these numbers gives a (scaled) "mean-median" score of 11.6.
- The second such score is called the partisan bias score: the difference in the number of seats each party wins if each were to earn 50% of the vote.
- For the proposed map, the Republican Party would win 19 seats with 50% of the vote, while the Democratic Party would win only 12 seats; the difference of these numbers gives a "partisan bias" score of –7.

To quantify how gerrymandered this plan is, we compare each score above to the distribution of those scores within the 500,000-map ensemble. Both the "mean-median" and "partisan bias" scores are very far from their typical values within an unbiased ensemble.

- In fact, the mean-median score for the proposed map is more extreme than 99.995% of ensemble values; that is fewer than 1 in 20,000 maps exhibit a similar score.
- The partisan bias score for the proposed map was the *most extreme score* we observed in our 500,000 ensemble; only 1 in 3,000 maps shared this score.
- Finally, not a single map in our 500,000-map ensemble exhibited this level of bias on *both* metrics!

- The same concepts can be used for racial group percentages, such as the HVAP vectors. This is the percentage of the VAP that is Hispanic in each district, except that it is sorted by percentage, instead of by district number. The HVAP vectors from the 500,000-map ensemble are illustrated with violin plots, and the proposed plan is overlaid in red. These show a similar pattern of "cracking and packing" as for the partisan data.
- District numbers for the proposed plan are given along the bottom axis.

For each plan, we can also ask, "How many Districts have an BHVAP over 50%?" (or 60%, or 70%, etc..). The BHVAP is the combined Black and Hispanic voting age population. These histograms show the values for the ensemble. The values for the Current (green) and Proposed (red) maps are also shown (if you can only see one vertical line, then they are equal). Above, we show the number of Districts that have BHVAP above 50% and 70%; that is the number of Districts that are majority-minority, vs. The number of Districts that are overwhelmingly majority-minority. Here is what we observe:

- The number of majority-minority districts in both the Current (S2100) and Proposed (S2101) plan are much lower than the typical value in the ensemble (10 or 11, vs. 12 or 13).
- The number of overwhelmingly majority-minority districts is higher than the typical value in the ensemble (8 or 9, vs. 4 or 5).
- This suggest that minority voters have been packed into a small number of districts, at the cost of reducing the total number of Districts in which they may be able to elect representatives of their choice.

- We can do this analysis for any group or coalition of interest. For example, it may be of interest to consider the AAPI community either separately or together with Black and Hispanic voters.

[redacted]

Math For Unbiased Maps TX (MUM_TX): Methods

Now that 2020 Census results have been released, redistricting has begun. Due to the size and complexity of Texas:

1. The number of possible ways to draw voting district plans is enormous.
2. It can be time-consuming to create even a single legally-valid plan that obeys constraints like population-balance and the county-line rule.

It is easier to draw and evaluate valid plans if you have examples to work from. However, it has not been feasible to create a large and representative **ensemble** of districting plans in prior redistricting cycles with the algorithms and computers available at the time. **But it is feasible now.** With ensemble sampling, we can generate an unbiased baseline which shows us what we can expect of a "typical map", and conversely detect when a proposed map is an outlier which shows evidence of gerrymandering.

We are ***Math For Unbiased Maps TX (MUM_TX***), an interdisciplinary, nonpartisan coalition of Texas mathematicians, political scientists and philosophers working to ensure a fair and transparent redistricting process. Our research concerns the development and application of ensemble sampling techniques, and in particular their application to the current TX redistricting cycle. While our research is primarily academic and thus concerned with the production of knowledge, we also believe that **knowledge that affects the public, should be made available to the public**. That is why we have committed to making our findings available to the general public and legislators in an accessible way, in the service of increasing transparency and public engagement in the process.

We have posted results to the TX Senate comment portal; in these notes, we seek to explain our methods. We have illustrated our methods using results from PLANS2101, which was filed on September 18, 2021. However, we are continuing to update our website with analysis of the many amendments that have been filed since. We will also be analyzing the US Congressional map and any subsequent amendments; and when the time comes, we'll do the same for the TX House.

You can find us at:

https://www.smu.edu/Dedman/Research/Institutes-and-Centers/DCII/Scholarship/Research-Cluster-on-Political-Decision-Making

What is Ensemble Sampling?

In general, ensemble sampling refers to any process of generating a large number of samples from a large collection; in this case, the set of all possible districting plans. Usually there is an element of randomness, because the collection is too large to look at every member. We can formalize the process of making random transitions between district plans with a mathematical object called a **Markov chain**. The idea of using transitions between states of a Markov chain to solve high-dimensional problems, or *Markov Chain Monte Carlo*, was pioneered in statistical physics, but is now used in nearly every area of science and engineering. In the current context, a "state" of the Markov chain will be a districting plan: that is, a division of a political entity into a specified number of subdivisions.

How do we sample?

There are many ways of moving between states in a Markov chain. We use *ReCom* (short for *recombination*) in which two adjacent districts are combined into one, and then split into two again, in such a way that legal constraints (such as equal population) are satisfied (D. DeFord, 2021). This is implemented in software called GerryChain, which is a Python package developed by the Metric Geometry and Gerrymandering Group (MGGG), a Boston-based team of mathematicians and computer scientists working on U.S. redistricting, and is available on GitHub (Metric Geometry and Gerrymandering Group, 2018)

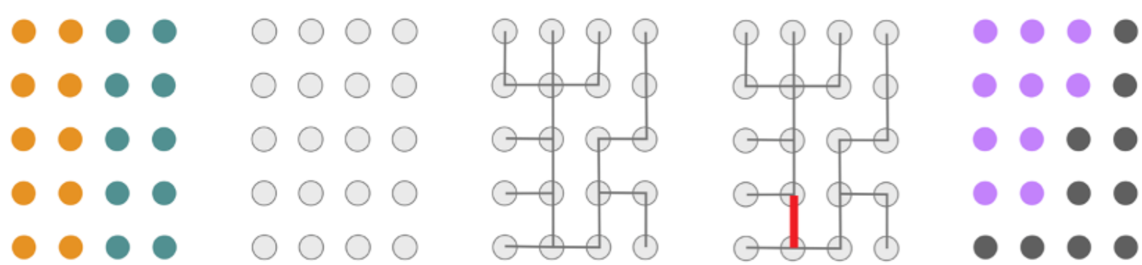


Figure 1: A schematic of a ReCom step. From left: (1) Start with 2 districts that share a boundary (orange and green). (2) Merge the districts into one. (3) Choose a spanning tree on the dual graph. (4) Split the spanning tree into two disjoint graphs, by eliminating a single edge (red). (5) If the resulting districts satisfy the desired constraints (such as population size and compactness), accept the ReCom step with two new districts (purple and gray). (Figure 4 from (D. DeFord, 2021)

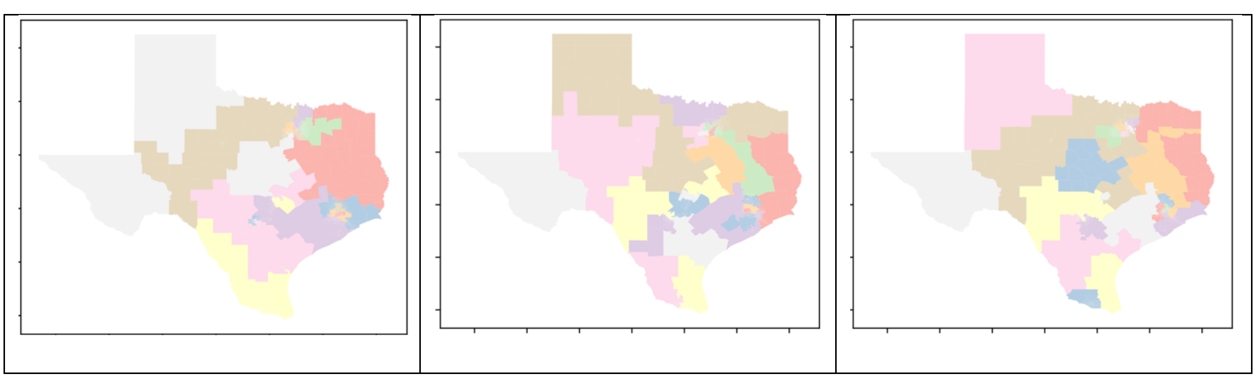
How did we analyze the proposed TX Senate Map?

We used GerryChain to create an ensemble of TX Senate districting plans, using the 2020 Census data released on August 12, 2021. This enabled us to assemble total populations and voting age populations (VAP) in each Census block in different racial groups: for example, Black and Hispanic. We began with a graph that contains the appropriate geographic nodes (all counties that were whole in the proposed plan, VTDs for those counties that were not whole, and *split* VTDs, for those VTDs that were split in the proposed plan). Population numbers were computed for each geographic unit.

We then generated an ensemble of 500,000 plans, which allows a robust exploration of the space of possible plans. (We can validate with a larger simulation). The constraints we applied at each ReCom step were:

- the population of each district had to be with 3% of the ideal balance. This was the maximum deviation observed in the proposed plan (S2101).
- The number of *cut edges* (the number of edges in the dual graph whose vertices lie in different districts) had to be less than twice the number in the original plan. This is a discrete proxy for perimeter. In practice, the number of cut edges for the plans we found was *less* than the number in the enacted plan.
- No additional "county splits" were allowed: if a particular county was entirely contained inside one district in the proposed plan, it must continue to be so for all other alternative plans.

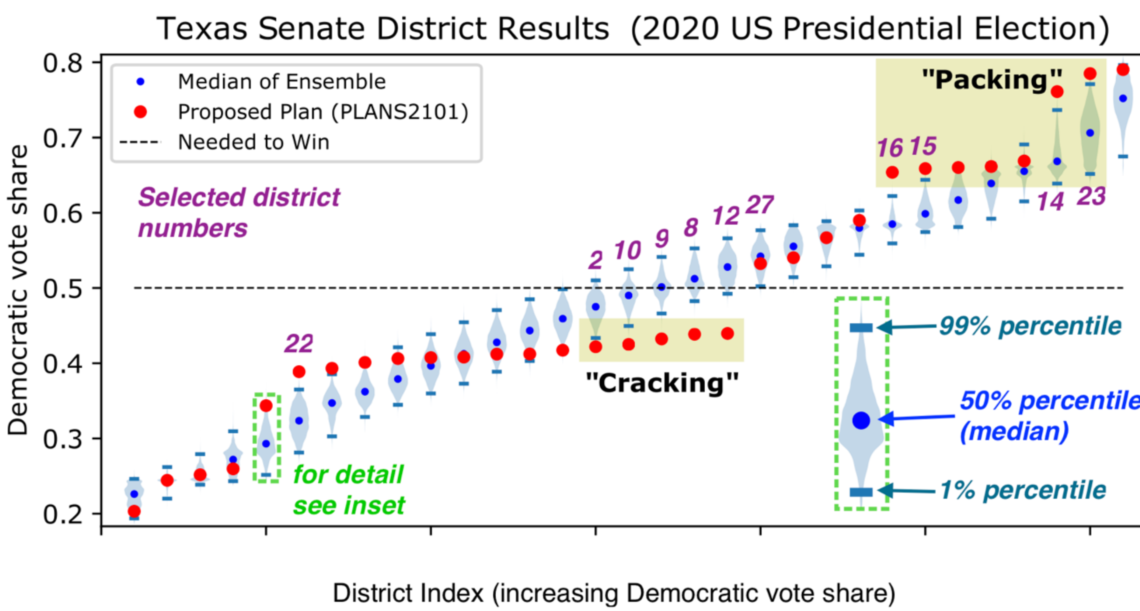
Each step reflected a single districting plan for which we can then compute a list of statistics. For each plan, and each district, we collected the percentage of the voting age population that was Hispanic (HVAP), the percentage of the voting age population that is Black (BVAP), and the percentage of voting age population that is either Black or Hispanic (BHVAP). We also collected the percentage of the vote (among Democratic vs. Republican) from the 2020 Presidential race. Below, we show maps illustrating some of the plans we generated.



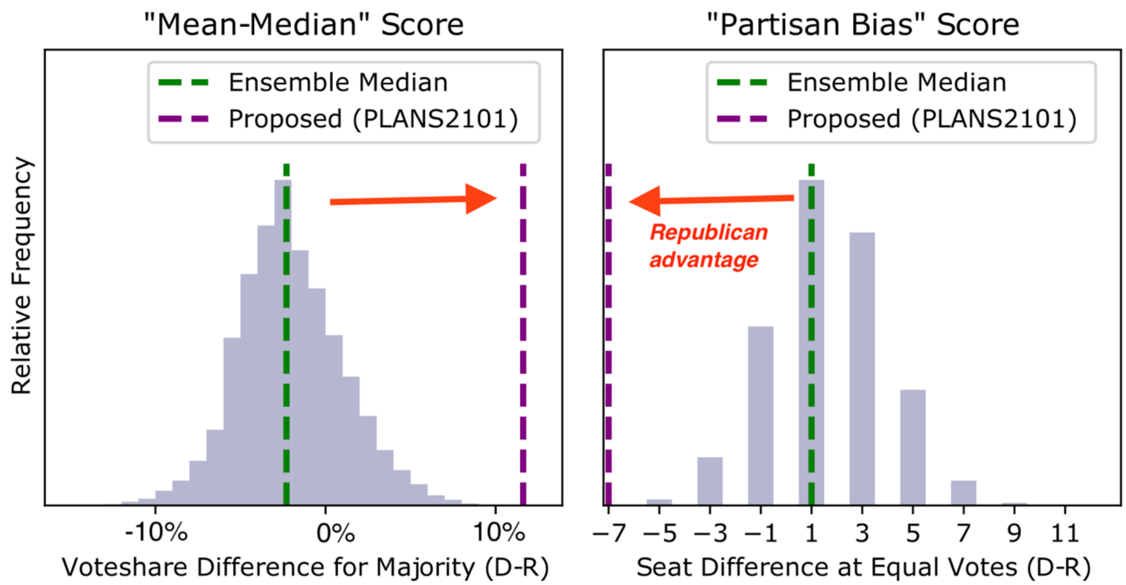
Visualizations

Next, we will explain some of the visualizations which can enable us to understand how a proposed plan compares to an unbiased baseline.

For each graph, we have included a comparison with the Proposed plan (PLANS2101). For some, we have also included the current plan enacted in 2011 (PLANS2100).



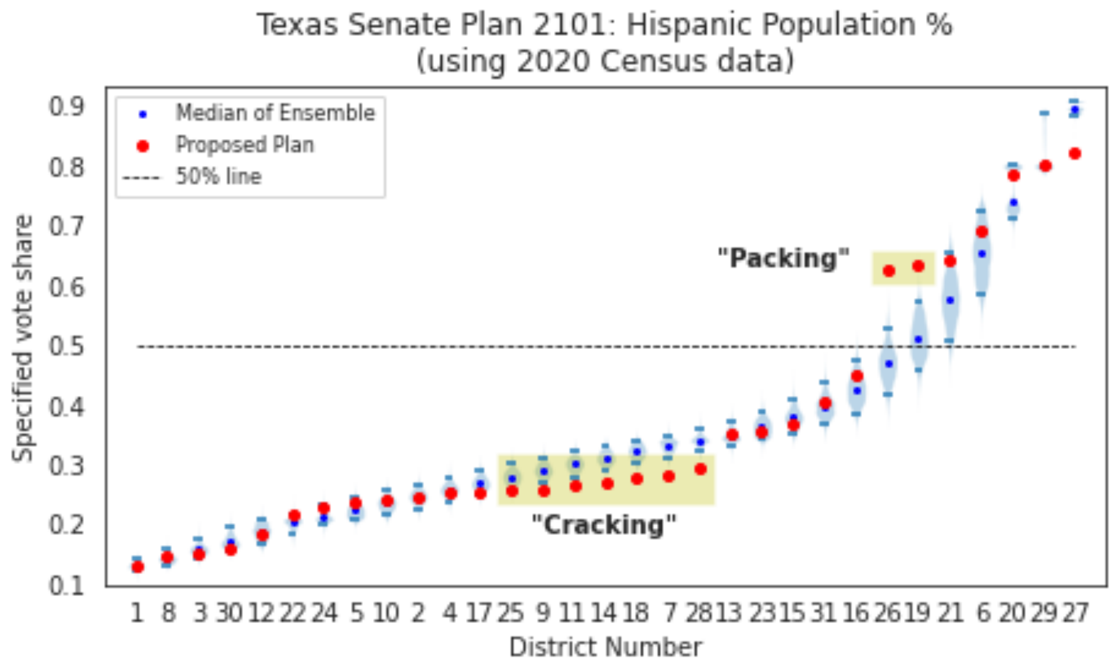
- This is a plot of **vote share vectors**, which is just a list of what percentage of the vote was obtained by a party in each district, except that it is *sorted by that vote share*, instead of by district number. The vote share vectors from the 500,000-map ensemble are illustrated with *violin plots* (Figure 3). Here the blue "violins" illustrate the distribution of likely vote shares for a district with a given vote share ranking, along with its median (blue dot) and 5% and 95% quantiles (blue bars).
- When we also plot the ordered vote shares for the proposed map (red circles), we obtain a concise visual summary of how gerrymandering is used to distribute votes in order to insulate one party from swings in public opinion. In the range labeled "cracking" Democratic vote shares have been decreased to an improbably low level (compared to the random ensemble), which has the effect of making these seats significantly harder to flip, even if there is a significant shift in public opinion.
- Democratic votes have been shoehorned into districts that fall in the range labeled "packing", especially into two overwhelmingly Democratic seats.
- We have identified, in purple, some districts of interest (refer to our statement)



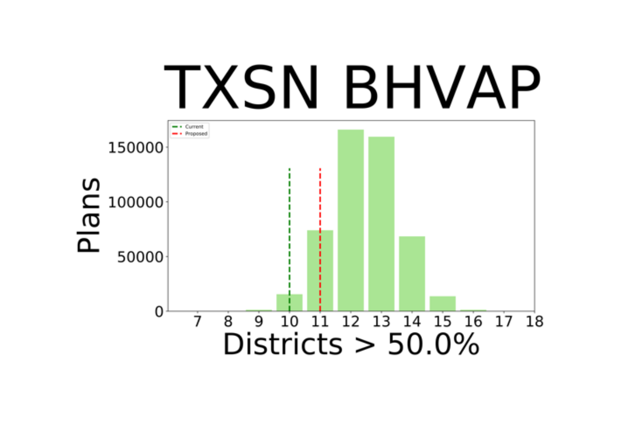
- This plot shows two common numbers that political scientists use to "score" maps. The first such number is called the **mean-median** score: it is proportional to the difference in statewide vote percentage each party would need to win the majority of the chamber.
- For the proposed map, the Republican Party would need to win only **44.2%** of the vote to win 16 seats, while the Democratic Party would need to earn **55.8%**; the difference of these numbers gives a (scaled) "mean-median" score of **11.6**.
- The second such score is called the **partisan bias** score: the difference in the number of seats each party wins if each were to earn 50% of the vote.
- For the proposed map, the Republican Party would win **19 seats** with 50% of the vote, while the Democratic Party would win only **12 seats**; the difference of these numbers gives a "partisan bias" score of **–7**.

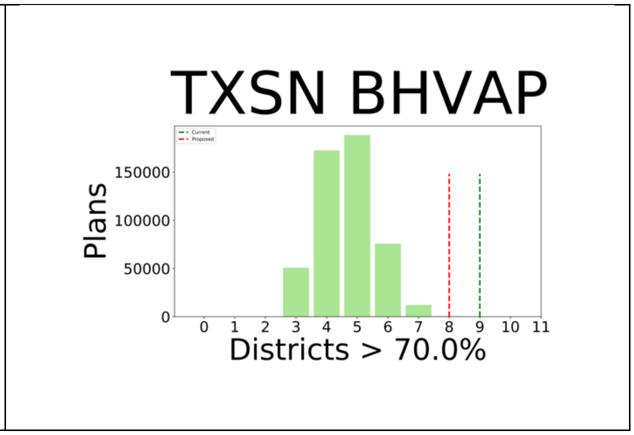
To quantify how gerrymandered this plan is, we compare each score above to the *distribution* of those scores within the 500,000-map ensemble. Both the "mean-median" and "partisan bias" scores are very far from their typical values within an unbiased ensemble.

- In fact, the mean-median score for the proposed map is more extreme than 99.995% of ensemble values; that is fewer than **1 in 20,000** maps exhibit a similar score.
- The partisan bias score for the proposed map was the *most extreme score* we observed in our 500,000 ensemble; only **1 in 3,000** maps shared this score.
- Finally, **not a single map** in our 500,000-map ensemble exhibited this level of bias on *both* metrics!



- The same concepts can be used for racial group percentages, such as the **HVAP vectors.** This is the percentage of the VAP that is Hispanic in each district, except that it is *sorted by percentage*, instead of by district number. The HVAP vectors from the 500,000-map ensemble are illustrated with violin plots, and the proposed plan is overlaid in red. These show a similar pattern of "cracking and packing" as for the partisan data.
- District numbers for the proposed plan are given along the bottom axis.





For each plan, we can also ask, "How many Districts have an BHVAP over 50%?" (or 60%, or 70%, etc..). The BHVAP is the combined Black and Hispanic voting age population. These histograms show the values for the ensemble. The values for the Current (green) and Proposed (red) maps are also shown (if you can only see one vertical line, then they are equal). Above, we show the number of Districts that have BHVAP above 50% and 70%; that is the number of Districts that are majority-minority, vs. The number of Districts that are ***overwhelmingly*** majority-minority. Here is what we observe:

- The number of majority-minority districts in both the Current (S2100) and Proposed (S2101) plan are much lower than the typical value in the ensemble (10 or 11, vs. 12 or 13).
- The number of overwhelmingly majority-minority districts is higher than the typical value in the ensemble (8 or 9, vs. 4 or 5).
- This suggest that minority voters have been packed into a small number of districts, at the cost of reducing the total number of Districts in which they may be able to elect representatives of their choice.
- We can do this analysis for any group or coalition of interest. For example, it may be of interest to consider the AAPI community either separately or together with Black and Hispanic voters.

Bibliography

D. DeFord, M. D. (2021). Recombination: A Family of Markov Chains for Redistricting. *Harvard Data Science Review*.

Metric Geometry and Gerrymandering Group. (2018). *GerryChain User Guide*, Revision c87da7e6. Retrieved March 9, 2021, from https://gerrychain.readthedocs.io/en/latest/index.html