Rebuttal Report

Moon Duchin
Professor of Mathematics, Tufts University
Senior Fellow, Tisch College of Civic Life

December 28, 2021

1 Background and Introduction

I have previously submitted expert reports in NCLCV vs. Hall. I have been asked by counsel to respond to the report of Dr. Michael Barber, examining his study design and his conclusions.

1.1 Summary of Barber report

In Dr. Barber's report, he uses a new statistical sampling method called Sequential Monte Carlo (SMC) to produce a large collection (called an *ensemble*) of alternative districting plans for both bodies of the North Carolina state legislature—state Senate and state House. SMC is a method based on ideas developed in my research group,[1] but which has not been supported by any peer-reviewed publications.

Dr. Barber proceeds to build ensembles of districting plans for the purposes of comparison, but primarily does so individually on small pieces of the state: groups of counties (often called "county clusters") that correspond to groupings in the Senate and House plans recently enacted in North Carolina (SL-173 and SL-175).

- For legislative redistricting, the Barber report discusses the clusters only on an individual basis, neglecting to assemble them into the big picture for the whole state.
- Dr. Barber omits an ensemble comparison for the enacted Congressional plan, SL-174.

1.2 Summary of findings

- When assembling the statistics from Dr. Barber's own ensembles—completely granting him all methodological choices for algorithm selection and specifications—the enacted House plan is shown to be a major partisan outlier, while the NCLCV alternative plans are not (Figure 6).
- In exactly the same way, the enacted Senate plan is likewise shown to be a major partisan outlier, while the NCLCV alternative plans are not (Figure 5).
- Finally, I was able to run Barber's code to create a Congressional ensemble in the same fashion as his legislative ensembles. Here, too, the enacted plan is a significant outlier in a direction of partisan advantage that is not justified by any good-government goal (Figure 3).

[1] The McCartan–Imai article introducing SMC [5] acknowledges Deford–Duchin–Solomon [3] for "pioneer[ing] the spanning tree-based proposal used in the merge-split algorithm."

2 Ensembles and outliers

Today, the dominant method in computational redistricting analysis is to employ Markov chains to generate ensembles of thousands or millions of alternative valid redistricting plans against which to compare a given proposed plan. When a quantity of interest is measured over the ensemble, it frequently forms a "bell curve" of values, and we can then examine whether the proposed plan falls in the thick of the observed values or whether it is an extreme outlier, falling in one of the tails. If this exercise is carried out with respect to each party's representation, a telltale sign of a partisan gerrymander is when the seat share for a proposed plan falls (a) far from the corresponding vote share, and (b) far to the side of advantage for the party that controlled the line-drawing process. This is particularly problematic in a politically competitive "purple" state like North Carolina.

It is important to note that outlier status is a flag of intentionality, but not necessarily a smoking gun of wrongdoing. Being in a tails of a distribution that was created around certain design principles can often provide persuasive evidence that other principles or agendas were in play. For example, a map might be an outlier as the most compact, or the map that gives minority groups the greatest chance to elect their candidates of choice—these kinds of outlier status would not be marks of a bad plan. But being an outlier can indeed be a sign of problems, as when a plan systematically converts close voting to lopsided seat shares for the party that controls the process.

2.1 Barber methods

The creation and use of districting ensembles in the Barber report can be summarized as follows.

Step 1 *Fix a set of clusters.* Barber focuses on the county clustering found in the enacted plan, not exhaustively considering the dozens of other possibilities.

Step 2 *Partition each cluster.* Split each multi-district cluster into the corresponding number of districts using Sequential Monte Carlo sampling. Create 50,000 partitions (i.e., districting plans) for each cluster.

Step 3 *Winnow.* Selectively discard some of the partitions. Barber uses two statistics from the enacted plan (average Polsby-Popper score and county traversals) as the cutoff for inclusion.

Step 4 *Create an election index.* Barber blends the 11 up-ballot elections since 2014 into a single vote index rather than considering them one at a time. In particular, he sums the votes over all elections before taking shares, which does not control for turnout differences across elections.

Step 5 *Plot histograms and declare outliers.* Barber forms histograms counting "Democratic-leaning districts" for individual clusters, and does not present an overall compilation. His non-standard definition of "outlier" includes a full 50% of the ensemble.

In my opinion, better and more reliable results would have been obtained if several of the choices required in this study design were executed differently.

One glaring omission from Barber's methods is any consideration of the State's obligations under the Voting Rights Act of 1965, which could impact the partisan bottom line.[2] A non-exhaustive list of other potential flaws in Dr. Barber's methods includes the following.

- *Failure to consider all alternative clusterings.*
  North Carolina law dictates that districts be drawn within groupings or clusters of counties from which several districts will be formed. Sometimes, however, the General Assembly has a choice and can pick multiple groupings consistent with North Carolina law. Dr. Barber only gives cursory attention to alternative clusterings.

- *Use of sampling methodology not vetted by peer review.*
  Even when an idea is promising, peer review is an essential component of vetting. A method may appear promising in concept, but not work in practice. A method may work at small tasks—like the 34-map dataset used for testing in [5]—but not scale well to the enormous sizes needed for realistic problems. Peer review helps surface those issues, which is why the scientific community regards peer review as a mark of reliability.

- *Use of bright-line thresholds for compactness and traversals.*
  Dr. Barber's code already samples with a preference for compactness, and is fully capable of handling traversals in a similar manner.[3] Imposing sharp cutoffs for these at the level of the enacted plan creates highly misleading results.[4]

- *Use of election data in a blended rather than serial fashion.*
  If Barber records a Democratic share of 49% in his outputs, that is likely to reflect a Democratic win in some of the 11 elections and a Republican win in others—this is obscured when the results are blended to a single number. By the same token, a Democratic share of 45% in the blended election index might downplay a map that favors Republicans 11 out of 11 times, which entrenches an advantage.[5]

- *Employing a highly unconventional use of the "outlier" label.*
  As Dr. Barber himself puts it, "I consider a plan to be a partisan outlier if the number of Democratic districts generated by the plan falls outside the middle 50% of simulation results [sic]. This is a conservative definition of an outlier. In the social sciences, medicine, and other disciplines it is traditional to consider something an outlier if it falls outside the middle 95% or 90% of the comparison distribution." As I will show below in my whole-state comparisons, the enacted plans are outliers at any of these levels of significance, while the NCLCV alternative plans are not.

I will discuss the thresholding question further in §2.3. For the remainder of the report, I will set aside the other concerns and will simply assess Dr. Barber's outputs within his own methodological framework.

[2] Robust VRA consideration is fully compatible with computational redistricting, as is shown in [1].

[3] A preference for compactness is coded in the `smc_redist` parameterization in `house_clusters.R`, lines 354–356 and `senate_clusters.R`, lines 349–351.

[4] The imposition of cutoffs, which Dr. Barber calls "culling," occurs in two stages. Stage 1 (country traversals) is found in `house_clusters.R`, lines 531–536 and `senate_clusters.R`, lines 539–544. Stage 2 (average Polsby-Popper) is found in `house_clusters.R`, line 543–564 and `senate_clusters.R`, lines 552–573. An ad hoc adjustment in the Duplin and Wayne House County Grouping is found in lines 566-568 of the House code.

[5] The 49% Democratic lean occurs, for instance, in the NCLCV alternative maps in the Onslow/Pender House cluster. Vote averaging is found in the Barber replication materials in `house_clusters.R` lines 18-28 and `senate_clusters.R` lines 18-29.

2.2 Analysis methods

Reading Dr. Barber's report, it is striking that he only reported that the enacted plan often performed within the middle 50% of each small comparison while never evaluating how the individual choices aggregate at the level of the map as a whole. After all, if moderate partisan advantage is secured over and over again, it may well accrue to extreme advantage overall. In the context of a state legislature, the overall results are crucial: they determine who controls the chamber. Pursuing this in the Barber materials, I found that this is exactly what happens.

First, I was able to extract Dr. Barber's raw statistical outputs for legislative runs from his materials obtained by counsel.[6] With those, I was able to assemble his ensembles for individual clusters into a compiled ensemble for the entire state. The histogram of Senate outcomes can be found in Figure 6 and the histogram of House outcomes can be found in Figure 5. Second, I was able to run Dr. Barber's code to create an ensemble of alternative Congressional plans with exactly the algorithm and with similar specifications to those he used for his legislative demonstrations.[7] A corresponding plot of Congressional outcomes can be found in Figure 3. For all phases of analysis, Dr. Barber pulled electoral data from a free webapp called Dave's Redistricting App (`davesredistricting.org`). In replicating his analysis, I used the same data source in the same manner.

2.3 Filtered and unfiltered results

As I described above, Dr. Barber took his raw districting plan samples (50,000 maps created for each of 12 Senate cluster ensembles and 26 House cluster ensembles) and aggressively filtered them, applying a cutoff that sometimes left under ten maps out of the original set of 50,000. In fact, when Dr. Barber's filtering rule was applied in the Duplin and Wayne House County Grouping (§6.6 on p.58 of Barber Report), *zero* maps were left, because none of the randomly constructed maps had an average compactness score to match the enacted plan in that cluster. Since this is blatantly unworkable for comparison purposes, Dr. Barber made the ad hoc decision to loosen the rule to retain 2704 maps. Other cluster ensembles were filtered down to leave only 4, 6, or 2 out of 50,000 alternatives and did not receive an adjustment. The "outlier" label was then applied to these tiny sets.

To illustrate why this is methodologically unreasonable, consider JaVale McGee, a basketball center who recently signed with the Phoenix Suns of the NBA on a one-year, $5 million contract. If McGee wanted to argue that he is not unusually wealthy, he could choose to restrict the universe of comparison to Americans at least as tall as he is. Since he is 7 feet tall, this would greatly restrict the comparison pool to a relatively tiny group that also includes Mo Bamba (Orlando Magic), Joel Embiid (Philadelphia 76ers), and Brook Lopez (Milwaukee Bucks), all of whom make more money than he does. Not satisfied with this comparison, he could keep increasing the requirements by insisting on comparing to people who don't speak any more languages than he does, are no older than he is, and have lived in at least as many different cities. Eventually he will narrow the pool enough that he doesn't look like an outlier anymore.

Dr. Barber's filtering skews his sample in a similar way, because he effectively insists that maps have a statistic matching or exceeding the enacted map in every cluster—and then uses that pool to compare the enacted map. Overall, this reduces the number of plans under consideration by a factor of over 500 trillion. And it excludes options that may be better than the enacted plan overall but are less compact or have more traversals in a particular cluster.

Generally, if you are trying to argue that you look typical of a range of alternatives, it is obviously unreasonable to first require the alternatives to look like you in dozens of independent ways (i.e., in each cluster individually).

[6] His materials include the numerical outputs from his runs, but as far as I can determine he does not seem to have saved the district assignments for the individual plans in the ensemble.

[7] To be precise, the ensemble was generated at the state level for Congress, since the concept of county clusters is not applicable, and without the compactness and traversal thresholds. I ran the code exactly as Dr. Barber did, except tightening the allowed population deviation to 1% from ideal instead of 5% as in legislative maps. All other choices are identical. My congressional ensemble includes 20,000 maps rather than 50,000 just because of time limitations.

3 Findings

In this section, I will present the full histograms (or "bell curves") of all the results from Dr. Barber's methodology, compiled to the state level and shown without filtering. (Filtered ensembles can be seen in Appendix A, for comparison purposes.)

By Dr. Barber's own constructs, all three levels of districting show that **the enacted plans are partisan outliers and the NCLCV alternative plans are not**.

In the House, the enacted map is in the most extreme 0.00133 fraction of the Barber ensemble—well under 1 percent of sampled House plans are as extreme as SL-175. By contrast, the NCLCV alternative plan is in the upper .2516 share of the ensemble, not an outlier even by the Barber standard.

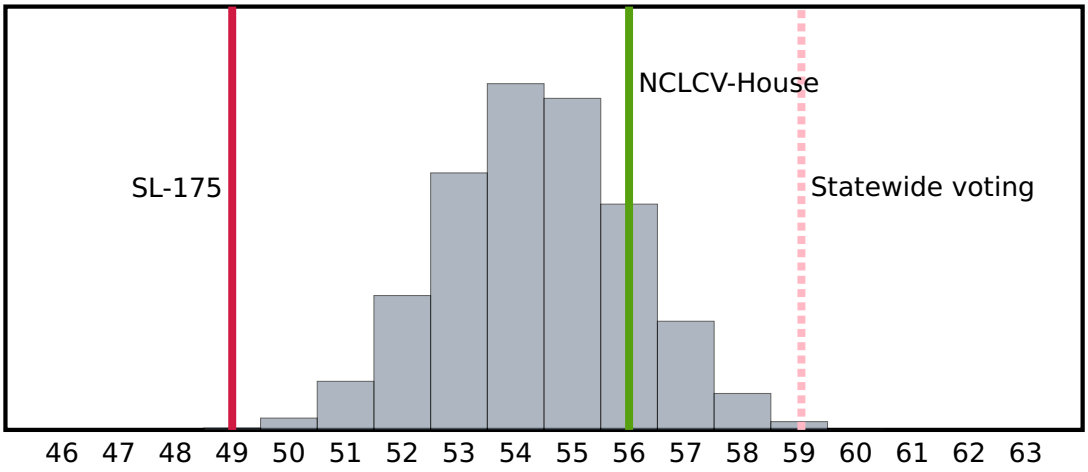

Figure 1: "Democratic-leaning seats" in Dr. Barber's House district ensemble.

At the Senate level, the enacted map is in the most extreme .007 fraction of the Barber ensemble—again, less than 1 percent of sampled plans are as extreme as SL-173. By contrast, the NCLCV alternative map is in the upper .2787 share of ensemble, not an outlier even by the Barber standard.

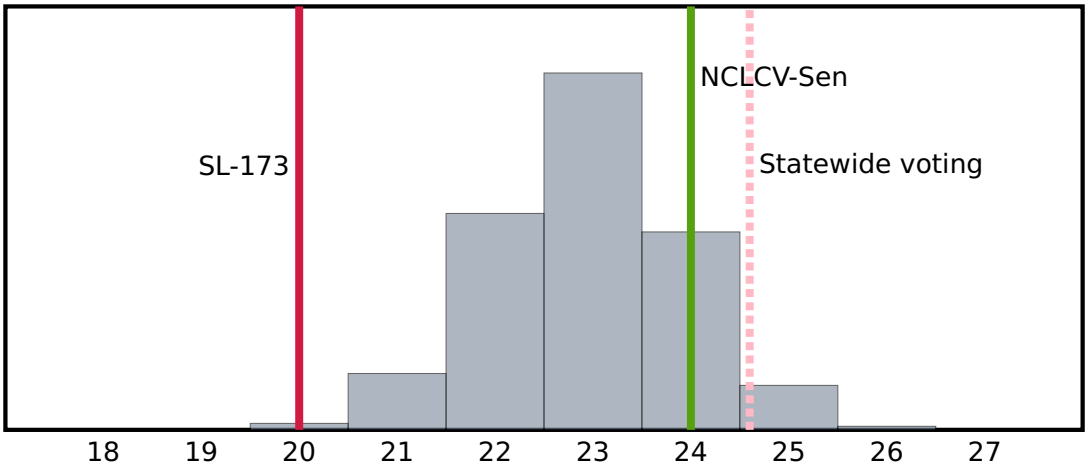

Figure 2: "Democratic-leaning seats" in Dr. Barber's Senate district ensemble.

The Congressional picture, omitted from the Barber report, is likewise crystal clear. The enacted plan is in the most extreme 0.0056 fraction of this Barber-style ensemble, while the NCLCV alternative map is very near the ensemble center—0.5620 share of the ensemble (more than half of randomly constructed maps) has an equal or greater Democratic lean.

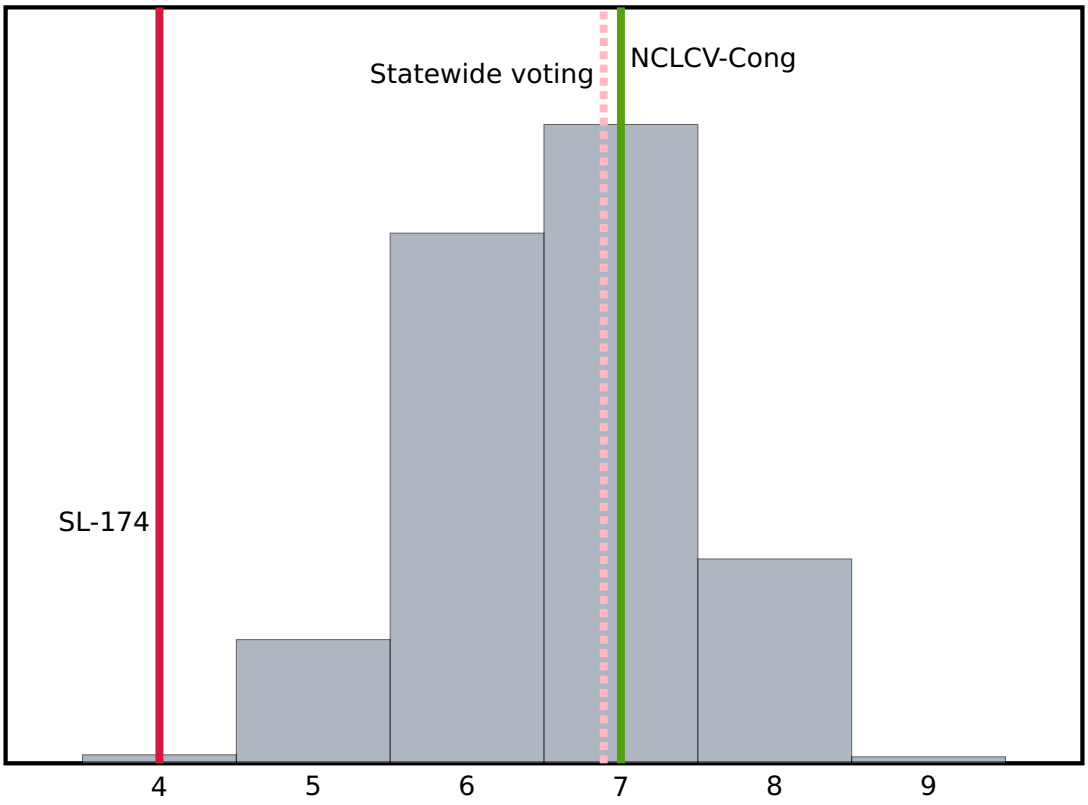

Figure 3: "Democratic-leaning seats" in a Congressional ensemble created with Dr. Barber's code, following his specifications.

4 Conclusion

Granting Dr. Barber all of his methodological choices, the enacted maps are extreme partisan outliers at all three levels, while the NCLCV alternative maps are not.

References

[1] Amariah Becker, Moon Duchin, Dara Gold, and Sam Hirsch, *Computational redistricting and the Voting Rights Act.* Election Law Journal. Available online.

[2] Christopher Cooper, Blake Esselstyn, Gregory Herschlag, Jonathan Mattingly, and Rebecca Tippett, *NC General Assembly County Clusterings from the 2020 Census.* sites.duke.edu/quantifyinggerrymandering/files/2021/08/countyClusters2020.pdf

[3] Daryl DeFord, Moon Duchin, and Justin Solomon, *Recombination: A Family of Markov Chains for Redistricting*, Harvard Data Science Review. Issue 3.1, Winter 2021. Available online.

[4] Moon Duchin, Taissa Gladkova, Eugene Henninger-Voss, Heather Newman, and Hannah Wheelen, *Locating the Representational Baseline: Republicans in Massachusetts.* Election Law Journal, Volume 18, Number 4, 2019, 388–401. Available online.

[5] Cory McCartan and Kosuke Imai, *Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans*, preprint. Available at arxiv.org/abs/2008.06131.

Appendix A: Filtering comparison

To illustrate the skewing effects of the thresholds applied by Dr. Barber, consider a single example: the Pitt House County Cluster, where the number of Democratic-leaning seats in the sample is either 1 or 2. By thresholding compactness and traversals at the level of the enacted map, Dr. Barber is able to drop the frequency of the 2-seats outcome from roughly 25% of the sample to just 9%.

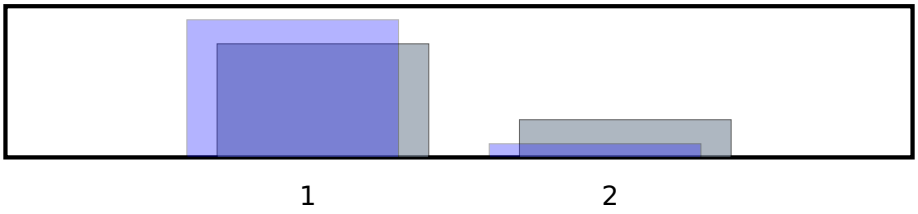

Figure 4: Just focusing on the Pitt House County Cluster (Barber report, p.42), we see that the filtering changes the outcome of 2 "Democratic-leaning seats" from occurring in roughly 25% of the full set of sampled maps (gray) to only occurring in 9% of the reduced sample (blue).

The effects of this cluster-by-cluster restriction do not wash out when aggregated to the full state, but instead add up to a noticeable shift toward the enacted plan, as demonstrated in the House and Senate figures below.

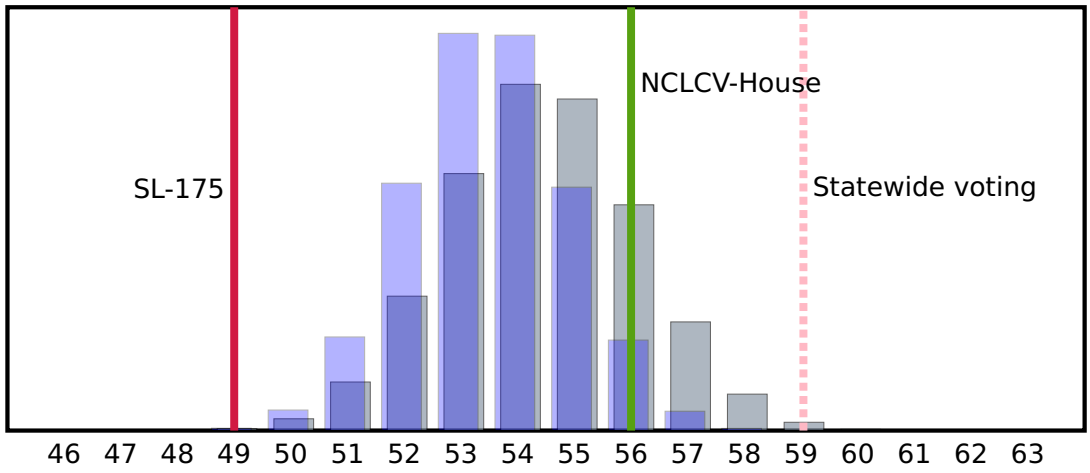

Figure 5: "Democratic-leaning seats" in Dr. Barber's House district ensemble. The unfiltered ensemble (gray) includes $50,000^{26} \approx 1.5 \cdot 10^{122}$ maps; the filtered ensemble (blue) is smaller by a factor of octillions.

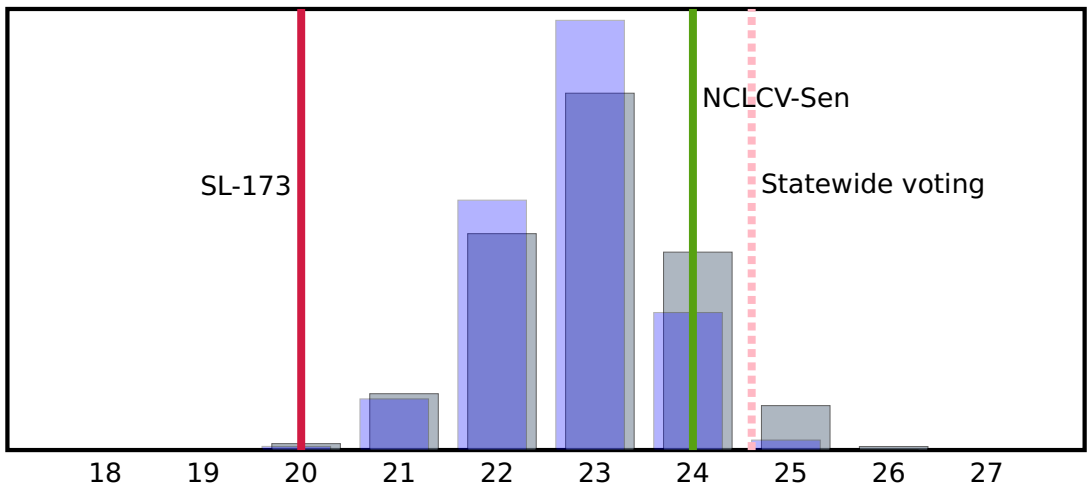

Figure 6: "Democratic-leaning seats" in Dr. Barber's Senate district ensemble. The unfiltered ensemble (gray) includes $50,000^{12} \approx 2.4 \cdot 10^{56}$ maps; the filtered ensemble (blue) is smaller by a factor of trillions.

Significantly, even the subsets of alternative plans that have been heavily limited by the cluster-by-cluster thresholds—that is, the blue bell curves instead of the gray—still show the enacted plans to be extreme outliers, while the NCLCV alternative plans are both far less extreme and comport with statewide voting.