Rebuttal to report of Michael Barber

Wesley Pegden

December 28, 2021

1 Introduction

In his report, Michael Barber presents the results of simulated district plans as part of an analysis which purports to elicit whether the enacted House and Senate maps of North Carolina are "partisan outliers". Barber makes choices in his analysis that reduce its ability to detect gerrymandering North Carolina clusters; for example, he discusses the partisan bias of the enacted House and Senate maps through the lens of the whole number of "Democratric-lean" districts in one hypothetical election, a lens through which even the effects of extreme gerrymandering in NC county clusters—each with a small number of districts—are made to appear less dramatic.

Nevertheless, his primary analyses (Tables 2 and 32) still find the whole-state House and Senate plans to be partisan outliers compared to his simulated maps, according to the definition he lays out in his report; in particular, he reports the middle-50% of simulated maps to have 46-51 total "Democratic-lean" districts across the House clusters he analyzes, and reports that the enacted map contains 45 such districts. For the Senate he reports a middle-50% range of 19-19 total Democratic-lean districts in his simulations, and that the enacted map contains 16 such districts.

In fact, Barber incorrectly calculated the distribution of Democrat-leaning seats for the whole-state outcomes of his simulation analysis, incorrectly reporting the sums of lower- and upper-quartile seat counts in individual clusters as the lower- and upper-quartile for total statewide seats. When the distribution of "lean Democrat district" counts at the whole-state level are calculated correctly for Barber's simulations (still using the partisan index he defines), one finds that the middle-50% range for Barber's simulated maps in the House is actually 48-50 Democratic-lean districts, not 46-51 as Barber shows, and that **the enacted North Carolina House map lies in the most Republican-biased 00.18% of whole state maps composed of Barber's simulations, and the enacted North Carolina Senate map lies in the most Republican-based 00.39% of whole state maps composed of Barber's simulations.** This computation can be carried out entirely with the figures provided in Barber's report, and uses Barber's simulated maps and Barber's metric of partisan bias (number of lean-Democrat districts), calculated with Barber's own partisan voting index.

Finally, when re-analyzing Barber's simulated maps (as provided in his backup data) to compare their expected performance over a range of electoral outcomes rather than comparing the crude number of "lean Democratic districts" for a fixed election average, the differences between the enacted map and Barber's ensemble of simulated comparison maps becomes more dramatic at the cluster level as well. Through this lens, every cluster which my original analysis found to be optimized for partisanship would qualify as a partisan outlier according to Barber's "middle 50%" criterion, and many are extreme outliers, among the most Republican biased 10%, 1%, or 0.1% of maps, even in clusters where Barber reported that the enacted map was not be a partisan outlier.

2 Barber finds the enacted House and Senate maps to be outliers according to his own definition

On page 29 of his report, in the section on House clusters, Barber writes that he considers a districting plan of North Carolina to be a partisan outlier if it lies outside of the "middle 50%" of simulation results; in Barber's report, the middle 50% are the maps that lie between the 25th and 75th percentiles according to the number of lean-Democrat districts, as measured with the partisan index Barber obtains by averaging election results. He calls this a “conservative definition” of an outlier, noting that “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.”

In both of his whole-state analysis tables (Table 2 and 32), Barber’s own findings report the whole map as falling outside the middle 50% of simulated outcomes for the House and Senate. For example, in the last row, labeled “Total”, of Table 2 on page 31, he reports that in the 26 clusters he analyzed, the enacted map contained 45 statewide “lean-Democrat” districts according to his partisan index, while the middle 50% range of the simulated maps for the total number of seats was $46 - 51$. Similarly, in Table 32 for the Senate, he reports the enacted map scored as having a total of 16 lean-Democrat seats in the 12 clusters used by the enacted map he analyzed, while the middle 50% range for his middle 50% range for the total number of seats in his simulated maps was 19-19. By the definition he chose to offer of a partisan outlier, Barber finds the enacted House and Senate plans are partisan outliers.

3 Barber reports incorrect quartiles for totals across clusters

Recall that in his Table 2, in the last column, Barber reports the range of the “middle 50%” for the number of lean-Democratic districts for his simulations in each cluster, and, at the bottom of the column, for the total across clusters (he reports the range for this total as 46-51). Recall that the bottom of the middle-50% range is the lower quartile of the data, and the top of the range is the upper quartile.

For example, in the House:

- for the Buncombe cluster in the House map, Barber reports in Figure 45 that 28% of his simulated maps contained 2 lean-Democrat districts, while 72% contained 3.
- for the Cumberland cluster in the House map, Barber reports in Figure 55 that 82% of his simulated maps contained 3 districts, while 18% contained 4.

I summarize this information in my Table 1, below:

Cluster	0	1	2	3	4
Buncombe			28%	72%	
Cumberland				82%	18%

Table 1: Fraction of maps with various lean-Democrat-district counts, as reported by Barber for Buncombe and Cumberland county districtings.

In his Table 2, Barber correctly summarizes the middle 50% ranges for the data in each of these clusters as 2-3 and 3-3, respectively; in each case, the lower end of the range is the smallest value below which 25% of his simulated maps lie, and the upper end is the smallest value below which 75% lie.

Suppose though, just as an example, that we wished to calculate the distribution of the total number of lean-Democrat districts across just these two clusters according the Barber’s simulations; this will also enable us to calculate the middle-50% of outcomes for the total lean-Democrat districts across these two clusters.

Note that for maps of these two clusters composed of maps from Barber’simulations, a total of 5, 6, or 7 lean-Democrat districts are possible. For example, 5 lean-Democrat districts can arise only by having 2 such districts in Buncombe and 3 in Cumberland, and fewer are not possible.

According to Barber’s simulations, as summarized in Table 1, 28% of the maps of these two clusters would have 2 lean-Democrat districts in Buncombe, while 82% would have 3 lean-Democrat districts in Cumberland. As the districtings in each cluster can be chosen independently of each other, a total of

$$28\% \times 82\% = 22.96\%$$

of districtings of these two counties would have a total of 5 lean-Democrat districts. (Note that having fewer than 5 lean-Democrat seats happens 0% of the time, according to Barber’s simulations.)

6 lean-Democrat districts can arise from having 2 lean-Democrat districts in Buncombe and 4 in Cumberland, or having 3 lean-Democrat districts in Buncombe and 3 in Cumberland. Thus according to Barber's simulation results the frequency of this outcome would be

$$28\% \times 18\% + 72\% \times 82\% = 64.08\%.$$

Finally, the likelihood of 7 lean-Democrat seats, which arise just when there are 3 lean-Democrat districts in Buncombe and 4 lean-Democrat districts in Cumberland, would be

$$72\% \times 18\% = 12.96\%,$$

(Note that altogether, 22.96%+64.08%+12.96%=100%.)

Evidently, the middle-50% range for the total of lean-Democrat seats across these two counties would be 6-6; the 6-lean-Democrat-district maps include the middle-50% of simulated maps. (6 is both the 25th percentile and the 75th percentile of the number of Democratic-lean seats in the simulated maps.)

Under Barber's incorrect approach, he would have simply added the bottom and top of the middle-50% ranges for Buncombe and Cumberland (2-3 and 3-3, respectively) to arrive at a middle-50% range for the total number of lean-Democrat-districts across these two counties; that procedure would produce a range of 5-6, which is wider than the true middle-50% range of the total number of districts across the two counties (namely 6-6), as correctly calculated above.

In general, the magnitude of this error grows larger and larger the more independent cluster-specific results are aggregated by incorrectly summing the lower and upper quartiles as a substitute for a correct calculation of the distribution of total statewide lean-Democrat districts. In Barber's report, he aggregrates across 26 clusters in this way. As we will see in the next section, this has the effect of inflating the true middle-50% range of 48-50 to an incorrectly reported range of 46-51.

Technical Remark. Probability generating functions can be used to allow larger calculations of the same type as the one above to be performed using publicly web-based computer algebra systems instead of by programming or using statistical software. Note that precisely the same three calculations above would have been performed if expanding the algebraic expression

$$\begin{aligned}(.28x^2 + .72x^3)(.82x^3 + .18x^4) &= (.28 \times .82)x^5 + (.28 \times .18 + .72 \times .82)x^6 + (.72 \times .18)x^7 \\ &= .2296x^5 + .6408x^6 + .1296x^7.\end{aligned}$$

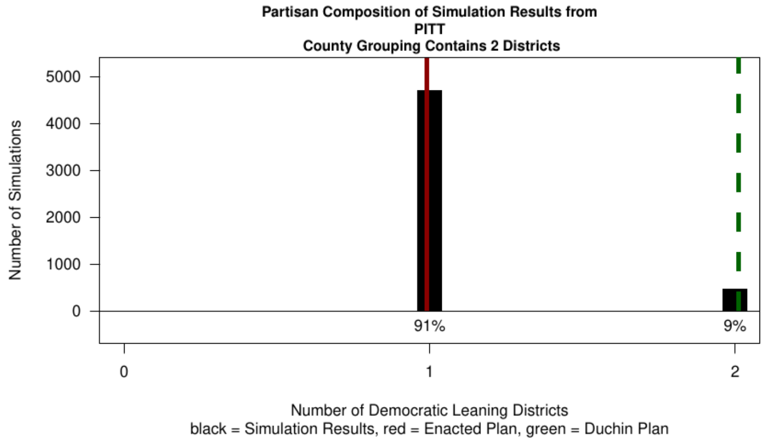
Observe that the polynomial $.28x^2 + .72x^3$ here can be seen as representing the fact that two seats occur in 28% of the maps for Buncombe, while 3 seats occur in 72% of the maps. (Similarly, then, for Cumberland and the polynomial $.82x^3 + .18x^4$.) The same answers that we found above for the fraction of simulated plans with a total of 5, 6, and 7 lean-Democrat districts, respectively, can be read off as the coefficients of x^5, x^6, and x^7, in the resulting expansion.

In the technical remark in the next section, I will point out a similar polynomial expansion which can verify the next section's calculations using public web applications, making the main findings of this rebuttal report easy to independently verify.

4 Correcting Barber's calculations

In my Table 2 on page 13 of this rebuttal report, I report the results of Barber's Figures 11, 14, 17, 20, 25, 28, 31, 34, 37, 45, 48, 51, 55, 58, 61, 64, 67, 70, 73, 76, 79, 82, 85, and 88. Each of these figures reports, for one of the clusters Barber analyzes, the fraction of his simulated maps which achieve different numbers of "lean Democrat" districts according to the partisan index he uses. For example, in Figure 14 on page 44, Barber reports that 91% of his simulated maps had one lean-Democrat district, while the remaining 9% had 2, as seen in this reproduction below:

Figure 14: **Distribution of Partisan Districts from Simulations in Pitt House County Cluster**



This information is then reproduced in my Table 2 on page 13, as the following row:

Cluster	0	1	2	3	4	5	6	7	8	9	10	11	12
Pitt		91%	9%										

In particular, everything in my Table 2 (and the corresponding Table 3 for the Senate) is taken directly from Barber's report itself.

The data in Table 2 can then be used to calculate the distribution of the total number of lean-Democrat seats based on Barber's simulations across the 26 clusters, exactly in the same way as we did above for just 2 clusters from the data in Table 1. The result of the same calculation is the histogram shown in Figure 1. In particular, according to Barber's own simulated map set, and using his own measure of the number of lean-Democrat districts under his own partisan index, **the enacted House map exhibits more Republican bias than 99.82% of maps** composed of Barber's simulations, over the clusters Barber analyzes.

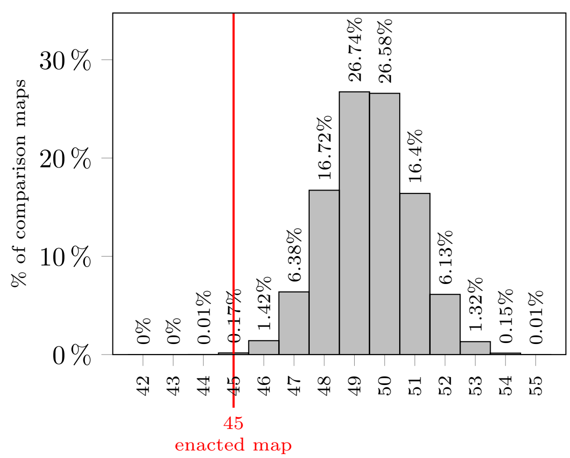


Figure 1: **Total lean-Democrat districts across Barber's House simulations**. This histogram shows the performance of Barber's simulated map set across the total set of House clusters Barber analyzes. It uses Barber's set of simulated maps, Barber's chosen metric (number of lean Democratic seats), calculated using the partisan metric Barber himself calculates in his report. The range 49-50 contains 50% of the simulated maps, the range 48-51 contains 86% of the simulated maps, and the range 47-52 contains more than 98% of the simulated maps. With 45 lean-Democratic districts across these clusters, the enacted map is in the most Republican-biased 0.18% of Barber's simulated maps.

In Table 3 I show Barber's Senate data analogous to the House data I show in Table 2. And in Figure 2, I plot the histogram showing the total of Barber's metric of Democratic-leaning districts across Barber's simulated map set, produced in the same way as I produce Figure 1 for the House. In particular, according to Barber's own simulated map set, and using his own measure of the number of lean-Democrat districts under his own partisan index, **the enacted Senate map exhibits more Republican bias than 99.61% of maps** over the clusters Barber analyzes.

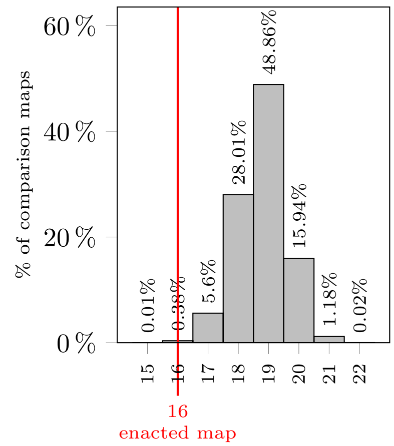


Figure 2: **Total lean-Democrat districts across Barber's Senate simulations**. This histogram shows the performance of Barber's simulated map set across the total set of Senate clusters Barber analyzes. It uses Barber's set of simulated maps, Barber's chosen metric (number of lean Democratic seats), calculated using the partisan metric Barber himself calculates in his report. The range 18-20 contains 93% of the simulated maps, and the range 17-21 contains more than 99% of the simulated maps. With 16 lean-Democrat districts, the enacted map is among the most Republican 00.39% of maps.

Technical Remark. As noted in the earlier Technical Remark, calculating the results of a histogram like Figure 1 is equivalent to expanding a certain polynomial expression. Based on the data in Table 2, (rows with only zero seats possible can be ignored), the polynomial to be expanded is

$$(.91x+.09x^2)(.44+.56x)(x^2)(x^2)(x)(.28x^2+.72x^3)(.82x^3+.18x^4)(x^4)(x)(.33x^2+.5x^3+.17x^4)(.99+.01x^1) \\ \cdots(.18+.82x)(.01x^4+.79x^5+.21x^6)(.01x^{10}+.56x^{11}+.44x^{12})(.02x^{10}+.32x^{11}+.66x^{12})$$

and publicly available tools such as `wolframalpha.com` can be used to verify that this polynomial expands to

$$5.55283\times 10^{-7}x^{56}+0.0000685893x^{55}+0.00147488x^{54}+0.0131615x^{53} \\ +0.0612515x^{52}+0.163979x^{51}+0.265839x^{50}+0.267369x^{49}+0.167218x^{48}+0.0637935x^{47}+0.0141775x^{46} \\ +0.00167669x^{45}+0.000089375x^{44}+1.74341\times 10^{-6}x^{43}+1.08123\times 10^{-8}x^{42}$$

The histogram in Figure 1 can be read off the coefficients in this polynomial. For example, the fact that the coefficient of x^{49} is .267369 corresponds to the fact that Figure 1 reports the fraction of simulated maps with a total of 49 Democrat-leaning districts across the clusters Barber analyzes as 26.74% (rounded to two decimal places).

For the senate, from Table 3, the probability generating function is

$$(.77x+.23x^2)(x^2)(.23+.77x)(.93x^2+.06x^3)(.01x^4+.24x^5+.75x^6)(.05x^4+.95x^5)x(.97x+.03x^2),$$

which expands to

$$0.000227131x^{22}+0.0118152x^{21}+0.159415x^{20}+0.488577x^{19} \\ +0.280141x^{18}+0.0559707x^{17}+0.00377389x^{16}+0.0000807399x^{15} \quad (1)$$

giving the results shown in Figure 2.

5 A more sensitive cluster-by-cluster analysis of Barber's maps

In the previous section, I showed that even against Barber's simulated maps, using the partisan index Barber calculates, and using Barber's preferred metric for partisan bias (the number of lean-Democrat districts using that partisan index), both the enacted House and Senate plans are extreme partisan outliers.

This is true despite the fact that using the number of whole lean-Democrat districts with only a single proxy for partisanship is unlikely to capture the effects even of extreme gerrymandering in North Carolina county clusters, where a small number of seats are at stake in each, and the effects of extreme gerrymandering can be to put one or two seats into play (or take them out of contention), even in cases where districts do not change columns in a single hypothetical election.

In other words, I take Barber's single partisan index (which has a two-party statewide Democratic vote-share of 49%), and analyze what would happen under his simulations, on average, if you swung the election results so that Democrats did better or worse by a normally-distributed swing matched to past statewide North Carolina elections. This is the same metric I used in my initial report.

In this section, I re-analyze Barber's results, still using his simulated maps, and still using his partisan index, but comparing maps in each cluster using the seats-expected metric (calculated with respect to that index), which evaluates how a map would be expected to perform under a range of conditions rather than one fixed hypothetical election.

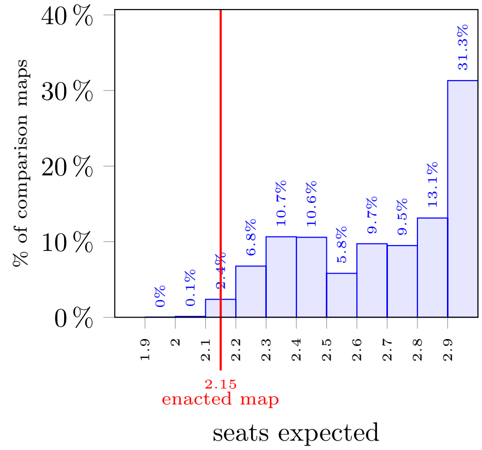
Below, I conduct this analysis for every county cluster I analyzed in my original expert report. In every cluster for which my analysis found the enacted map to be among the most optimized-for-partisanship possible maps (the first six House analyzed in the subsections below, and every Senate cluster analyzed below), Barber finds the map to be a partisan outlier according to the "middle-50%" definition he uses in his report. I summarize the outlier status of these 6+5 House and Senate clusters according to Barber's simulations in the following table:

Cluster	Enacted map among most Republican-biased...
House: Buncombe	00.797%
House: Forsyth-Stokes	00.0805%
House: Guilford	00.00646%
House: Mecklenburg	04.43%
House: Wake	05.78%
House: Pitt	24.2%
Senate: Cumberland-Moore	00.0024%
Senate: Forsyth-Stokes	00.01%
Senate: Granville-Wake	00.035%
Senate: Guilford-Rockingham	00.25%
Senate: Iredell-Mecklenburg	00.1%
	...against Barber's simulations.

Among the four remaining clusters in my report, there are two where the enacted maps are nevertheless extreme outliers against Barber's simulation sets. I summarize the results for these four clusters in the following table:

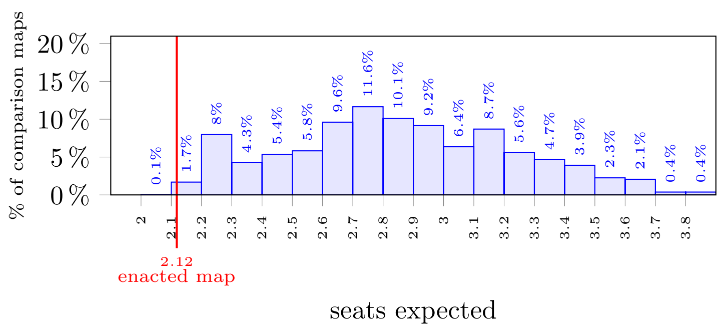
Cluster	Enacted map among most Republican-biased...
House: Alamance	39.4%
House: Brunswick-New Hanover	73.9%
House: Durham-Person	00.00265%
House: Cabarrus-Davie-Rowan-Yadkin	00.352%
	...against Barber's simulations.

5.1 House: Buncombe



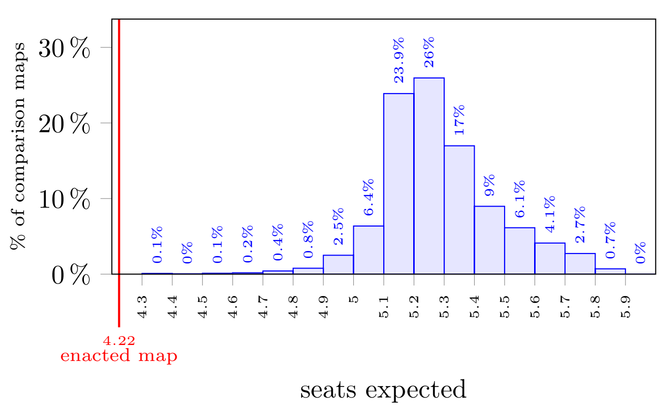
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.797% of maps.

5.2 House: Forsyth-Stokes



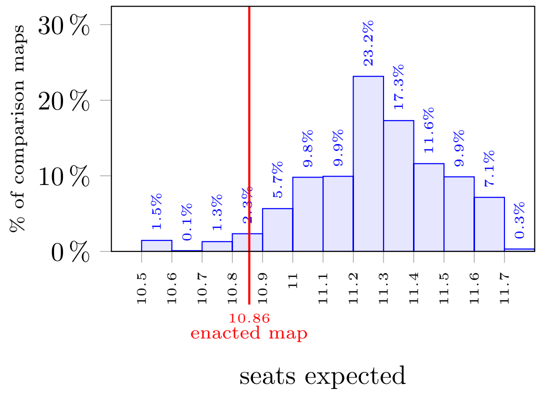
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.0805% of maps.

5.3 House: Guilford



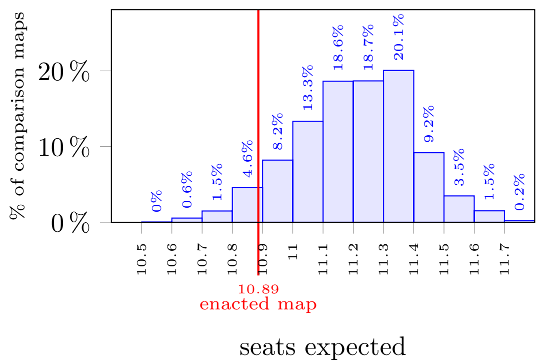
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.00646% of maps.

5.4 House: Mecklenburg



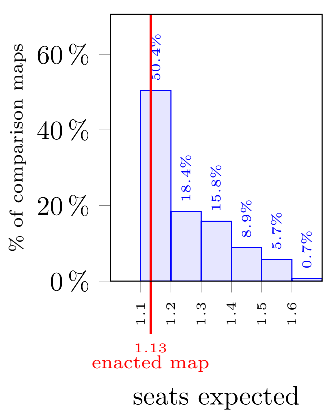
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 4.43% of maps.

5.5 House: Wake



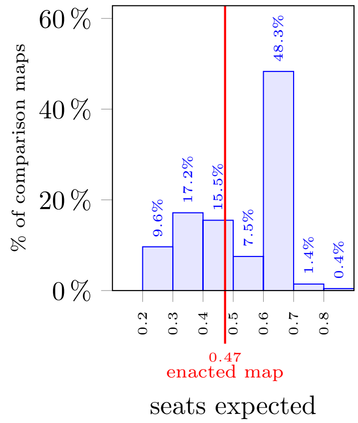
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 5.78% of maps.

5.6 House: Pitt



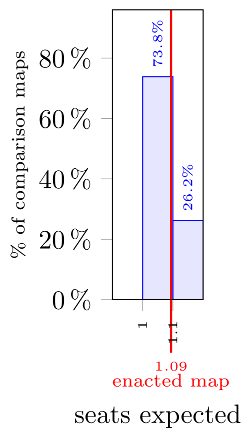
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 24.2% of maps.

5.7 House: Alamance



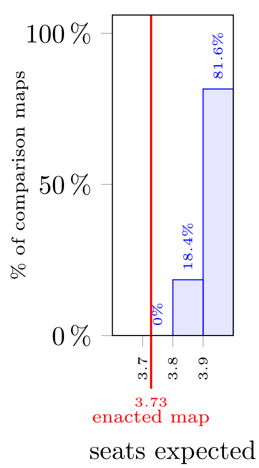
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map is not an outlier.

5.8 House: Brunswick-New Hanover



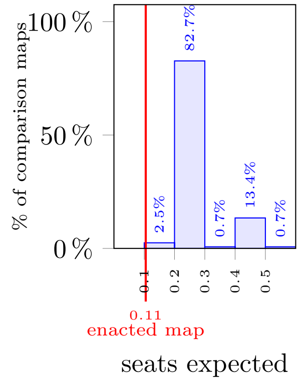
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map is not an outlier.

5.9 House: Durham-Person



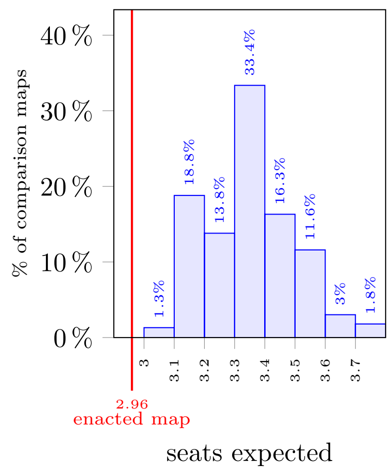
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.00265% of maps.

5.10 House: Cabarrus-Davie-Rowan-Yadkin



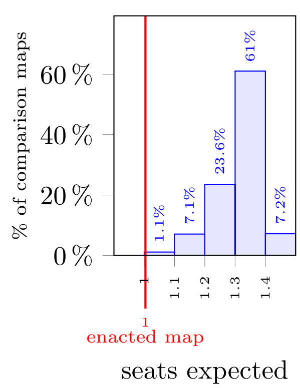
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.352% of maps.

5.11 House: Cumberland



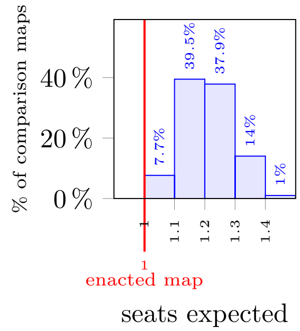
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.0095% of maps.

5.12 Senate: Cumberland-Moore



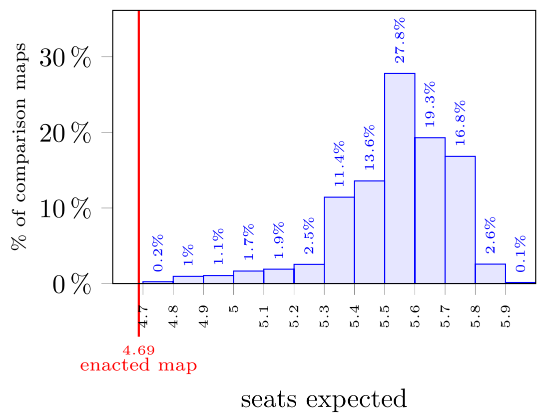
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.00235% of maps.

5.13 Senate: Forsyth-Stokes



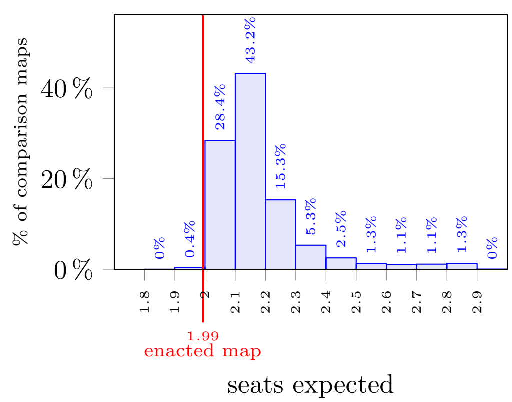
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.0104% of maps.

5.14 Senate: Granville-Wake



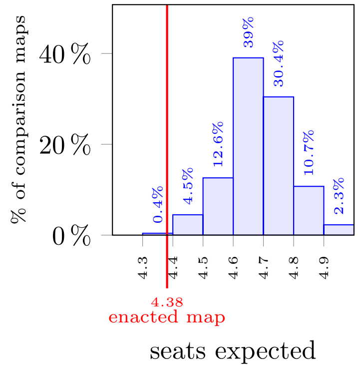
Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.0353% of maps.

5.15 Senate: Guilford-Rockingham



Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.251% of maps.

5.16 Senate: Iredell-Mecklenburg



Against the comparison-set of Barber's simulated maps for this cluster, the enacted map in this cluster is among the most Republican-biased 0.104% of maps.

Cluster	0	1	2	3	4	5	6	7	8	9	10	11	12
Davidson	100%												
Pitt		91%	9%										
Alamance	44%	56%											
Columbus-Robeson	100%												
Carteret-Craven													
Duplin-Wayne	100%												
Nash-Wilson			100%										
Caswell-Orange			100%										
Alexander-Surry-Wilkes	100%												
Franklin-Granville-Vance		100%											
Alleghany-*etc*	100%												
Beaufort-*etc*	100%												
Buncombe			28%	72%									
Anson-Union	100%												
Onslow-Pender	100%												
Cumberland				82%	18%								
Harnett-Johnston	100%												
Catawba-Iredell	100%												
Durham-Person					100%								
Brunswick-New Hanover		100%											
Forsyth-Stokes			33%	50%	17%								
Cabarrus-*etc*	99%	1%											
Chatham-*etc*	18%	82%											
Guilford					1%	79%	21%						
Avery-*etc*	100%												
Mecklenburg											1%	56%	44%
Wake											2%	32%	66%

Table 2: This table collects in one place the fraction of maps in Barber's House simulation sets realizing each number of lean-Democratic seats, as reported by Barber in his Figures 11, 14, 17, 20, 25, 28, 31, 34, 37, 45, 48, 51, 55, 58, 61, 64, 67, 70, 73, 76, 79, 82, 85, and 88. He does not present figures for the clusters in Alleghany-Ashe-Caldwell-Watauga and Beaufort-Chowan-Currituck-Dare-Hyde-Pamlico-Perquimans-Tyrrell-Washington clusters because his 0-Democratic-district results for those clusters are based on a very small number of maps. For Carteret-Craven his method does not produce any maps.

Cluster	0	1	2	3	4	5	6
Cumberland-Moore		77%	23%				
Chatham-Durham			100%				
Alleghany-*etc*	100%						
Brunswick-Columbus-New Hanover	23%	77%					
Bladen-*etc*	100%						
Guilford-Rockingham			94%	6%			
Alamance-*etc*	100%						
Granville-Wake					1%	24%	75%
Iredell-Mecklenburg					5%	95%	
Buncombe-Burke-McDowell		100%					
Cleveland-Gaston-Lincoln	100%						
Forsyth-Stokes		97%	3%				

Table 3: This table collects in one place the fraction of maps in Barber's Senate simulation sets realizing each number of lean-Democratic seats, as reported by Barber in his Figures 95, 98, 103, 106, 110, 113, 117, 120, 123, 128. He does not present figures for the Bladen-Duplin-Harnett-Jones-Lee-Pender-Sampson and Cleveland-Gaston-Lincoln clusters because his 0-district results for these clusters are based on a small number of maps.

I hereby certify that the foregoing statements are true and correct to the best of my knowledge, information, and belief.

Wesley Pegden
12/28/2021