Ensemble Analysis for 2021 Congressional Redistricting in Colorado

Jeanne Clelland[*], Daryl DeFord, Beth Malmskog, and Flavia Sancier-Barbosa

September 10, 2021

Abstract

In this report, we apply techniques of ensemble analysis to establish a baseline context for Congressional redistricting in Colorado following the 2020 Census. We generate a large random sample of redistricting plans that meet the basic legal requirements established by Amendment Y. Using this sample, we establish "reasonable" ranges for what might be expected for county splits, minority population, competitive districts, and partisan seat share for plans generated without explicit consideration of these issues. We also explore how these various priorities interact; in particular, we explore how the constitutional imperative to keep counties whole as much as possible affects the ability to maximize the number of competitive districts and partisan seat share. Finally, we compare the First Staff Plan proposed by the Colorado Independent Congressional Redistricting Commission's nonpartisan staff to our ensemble and comment on its performance relative to the ensemble.

1 Introduction

In the years since the last decennial redistricting cycle, there has been much interest in—and litigation around—quantifying and identifying partisan bias in district plans. Unlike racial gerrymandering, which has historically been limited by the Voting Rights Act of 1965, partisan gerrymandering has largely been unchecked by the courts until fairly recently, primarily due to the difficulty of identifying a quantifiable standard for measuring it.

One recently developed strategy for quantifying partisan bias is the ideal of "ensemble analysis," in which a particular district plan is compared to a large collection of randomly generated, legally valid plans, referred to as an "ensemble" of plans. This idea has been gaining traction in redistricting litigation in the last few years. For instance, Jonathan Mattingly, et. al. performed detailed ensemble analyses of North Carolina's Congressional [9] and state [10] legislative district plans that played key roles in the court cases [3] and [2], and Moon Duchin's ensemble analysis [8] of

[*]The first author was partially supported by a Collaboration Grant for Mathematicians from the Simons Foundation.

Pennsylvania's Congressional Districts played a similar role in [1]. Similar work can be found in Wesley Pegden's expert reports for Pennsylvania [11] and North Carolina [12].

The primary aim of our work is to use ensemble analysis to establish a baseline context for Congressional redistricting in Colorado in 2021, in order to understand what might reasonably be expected for measures such as county splits, minority population, competitive districts, and partisan seat share, based on the state's unique political geography. This baseline may then be applied to evaluate proposed district plans under consideration by the Colorado Independent Congressional Redistricting Commission to ensure that they satisfy the requirements specified by Amendment Y to the Colorado Constitution.

Here and throughout this report, we wish to emphasize that **none of the plans in our ensembles are intended for adoption.** Redistricting is fundamentally a human endeavor, and there are many important considerations that are difficult or impossible to fully incorporate into a computer-generated ensemble. The ensembles that we will discuss here are intended **only** to provide context to which proposed plans may be compared with regard to specific quantitative measures.

2 Introduction to ensemble analysis

In this section we give a brief description of the main ideas and aims of ensemble analysis. For a more detailed treatment of our approach and methodology, please see our paper [5] and Appendix A.

The fundamental goal of ensemble analysis is to model the political geography of a region (in this case, the state of Colorado) in order to better understand what might be expected for a "typical" district plan for the state. Plans may be evaluated with regard to a variety of measures: partisan balance of election results, geographic compactness of districts, competitiveness of district elections, preservation of communities of interest, racial/ethnic population within districts, etc. The main idea is to create a large number of randomly generated, valid plans that satisfy all relevant legal constraints—an "ensemble" of plans. Measures of interest are then computed for each plan in the ensemble using real population and voting data. The result is a statistical range of possible outcomes for each measure, to which any proposed plan may be compared. If a proposed plan appears to be an extreme outlier compared to the ensemble, this may suggest that the plan was deliberately designed to achieve some specific goal, such as partisan gerrymandering.

For this type of analysis, it is natural to build districts from voting precincts, as these are the smallest geographic units for which voting data is readily available. This is one of many reasons why the plans in our ensemble are generally unsuitable for adoption; the final plans will almost certainly divide many precincts in order to achieve their aims—most notably, population equality between districts.

For Congressional districts in particular, it is not practical to require that computer-generated plans built from precincts achieve perfect population equality between districts; instead, we require that the population differential between the largest and smallest districts for any plan in our ensembles be less than 1% of the ideal district population. We have found that this level of flexibility strikes a good balance, allowing our algorithm to generate a wide variety of plans whose statistics remain very close to those of plans with perfect population equality.

Our construction of ensembles begins with a data-rich map of Colorado's voting precincts as of 2020. Details of our processes for data collection and construction of this map are described in Appendix A.1, and details of the algorithm used to build our ensembles are described in Appendix A.2. For this initial analysis, we constructed two ensembles of 200,000 random maps each, incorporating some of the most fundamental constitutional requirements:

- **Contiguity:** The algorithm used to generate district plans automatically guarantees district contiguity; see Appendix A.2 for more details.
- **Population equality:** As mentioned above, we have required that all plans in our district have a population deviation of 1% or less between the least- and most-populous districts.
- **Compactness:** The algorithm used to generate district plans is designed to preferentially sample from more compact plans, and a large body of experimental evidence indicates that it is generally very effective in this endeavor. (See, e.g., [6].) No specific metric for measuring compactness is prescribed by Amendment Y, and we did not explicitly track any quantitative measure of district compactness. However, we have included a few of our randomly generated maps in Appendix A.2 to illustrate that their districts are generally reasonably compact.
- **Preservation of political subdivisions:** Our first ensemble, which we shall refer to as "county-neutral," did not incorporate any information regarding political subdivisions such as cities or counties. Our second ensemble, which we shall refer to as "county-aware," added an algorithm described in Appendix A.2 to minimize the number of county splits. For future ensembles, we may refine this algorithm to attempt to minimize divisions of municipalities and other important communities of interest. **We look forward to input from the Commission and its nonpartisan staff regarding which communities it most strongly prefers to keep wholly within districts so that we can incorporate this direction into future ensembles.**

In Section 3, we will explore how our county-neutral and county-aware ensembles of plans typically perform on the measures of county splits, minority representation, competitive districts, and partisan seat share. For the latter two metrics, we will focus on the composite "election" obtained by averaging partisan outcomes for the 8 statewide elections between 2016 and 2020 that have been identified by the Commission, specifically:

- the elections for President and U.S. Senator in 2016;
- the elections for Attorney General, Governor, Regent At Large, Secretary of State, and Treasurer in 2018;
- the election for U.S. Senator in 2020.

3 Ensemble statistics

The goal of this section is to describe the main statistical properties of our county-neutral and county-aware ensembles in order to establish context for what might reasonably be expected for Congressional district plans in Colorado. In Section 4, we will provide a detailed comparison of the First Staff Plan to both ensembles.

3.1 County splits

The only difference between the algorithms for constructing our county-neutral and county-aware ensembles is that the latter includes a variation that attempts to minimize the number of county splits in each plan. We counted the number of county splits in each plan in two ways:

1. number of “counties split,” which counts the number of counties divided between more than one district;
2. number of “total splits,” which counts the number of times counties are split.

So, e.g., if a county is divided between three districts, this counts as one split towards the “counties split” measure and two splits for the “total splits” measure.

The histograms in Figure 1 describe what percentage of plans in each ensemble exhibited each value for the number of counties split and the number of total splits over the observed ranges. For the county-neutral ensemble, the mean number of counties split was 21.7 and the mean number of total splits was 38.3. For the county-aware ensemble, the mean number of counties split was 10.2 and the mean number of total splits was 15.2. We note that the means for the county-aware ensemble are very close to the values of 10 and 13, respectively, for counties split and total splits in the First Staff Plan. This suggests that this ensemble does a reasonable job of sampling from plans that prioritize keeping counties whole to a similar degree as typical human-drawn plans.

In the next several subsections, we will compare statistics for both ensembles with regard to minority representation, competitive districts, and partisan seat share. By computing statistics for both ensembles, we hope to better understand how the choice to preserve counties (and political subdivisions more generally) affects other priorities.

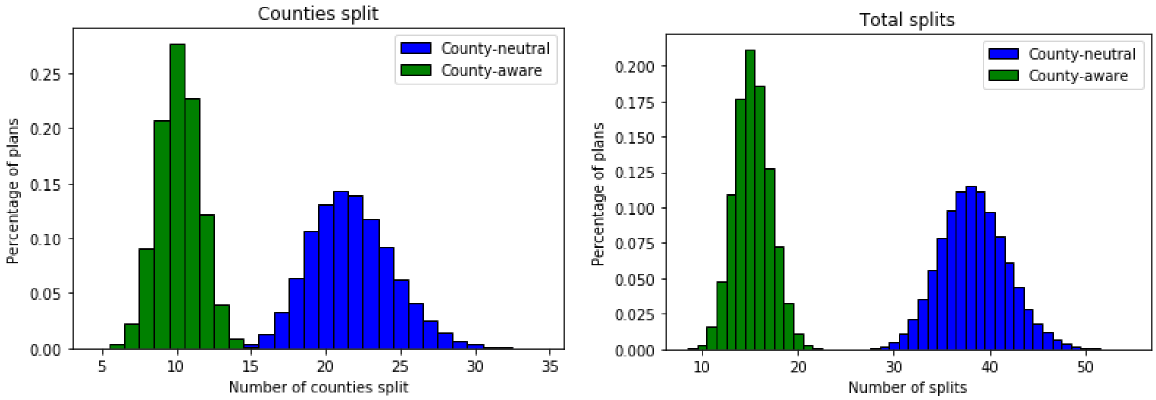

Figure 1: Counties split and total splits for county-neutral and county-aware ensembles

3.2 Minority representation

After contiguity, population equality, and the Voting Rights Act, the next highest priority specified by Amendment Y (co-equal with district compactness and preservation of political subdivisions) is the preservation of communities of interest. This is perhaps the most difficult criterion to model algorithmically, as communities of interest vary widely in nature and in geographic extent, and many different types of communities of interest overlap in complicated ways.

One very significant community of interest—and the only one that we will consider here—is the minority population of the state. Specifically, we will examine the proportions of (1) Hispanic voting age population, and (2) Non-White voting age population within each Congressional district. For context, we note that for the state as a whole, the Hispanic voting age population is approximately 19.2% of the total voting age population, and the Non-White voting age population is approximately 26.6% of the total voting age population. The Commission's nonpartisan staff "does not believe that there is sufficient voting age population to create a majority-minority congressional district within Colorado that complies with the requirements of the Colorado Constitution," but there is still general agreement that districts should be drawn so as to give these communities adequate representation.

For each plan in our ensembles, we compute the percentages of the Hispanic and Non-White voting age populations as a fraction of the total voting age population in each district and record the results. This data is displayed in Figures 2 and 3, organized as follows: For each plan, districts are sorted by Hispanic (resp. Non-White) voting age population percentage, from lowest to highest. The box plots show the ranges of these percentages for the sorted districts (in blue for the county-neutral ensemble and green for the county-aware ensemble)—so, e.g., the second pair of boxes from the left shows the range of Hispanic (resp., Non-White) voting age population percentage in the second-lowest district in each plan. The boxes show the middle 50% of the range, and the whiskers extend from the 1st percentile through the 99th.

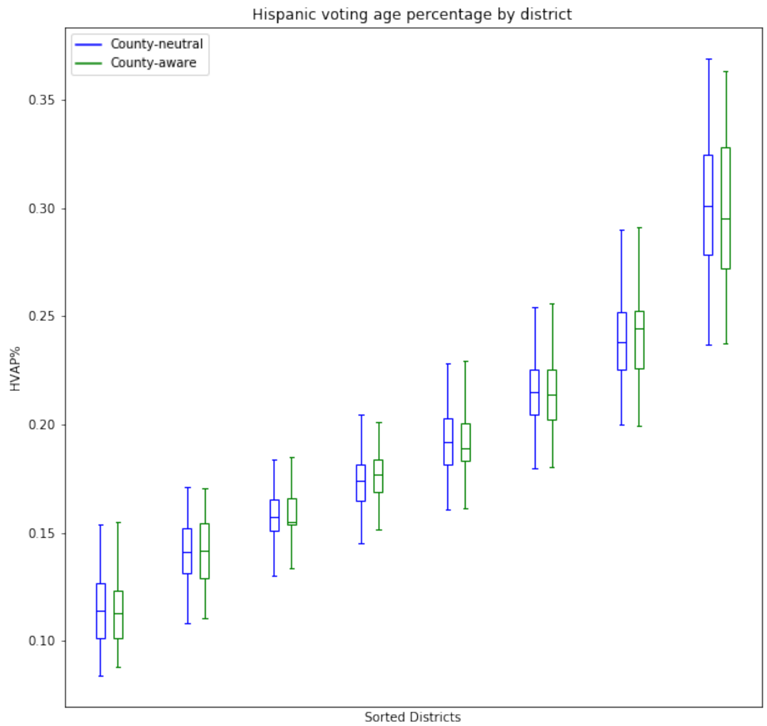

Figure 2: Hispanic voting age percentage by district for county-neutral and county-aware ensembles

One striking feature of these plots is that the preservation of counties does not seem to have much impact on the ability to draw one district with Hispanic voting age population significantly above 30%, or on the ranges of Hispanic voting age population in districts with lower percentages. But the situation is very different for the Non-White voting population; the county-neutral ensemble contains many plans with one district whose Non-White voting population is over 45%, while the expected range for the top district is a few percentage points lower for the county-aware ensemble. Additionally, the county-aware ensemble produces much narrower ranges for the middle 50% of possible values for the 2nd and 4th highest districts than the county-neutral ensemble.

3.3 Competitive districts

Competitive districts are defined in Amendment Y as "having a reasonable potential for the party affiliation of the district's representative to change at least once between federal decennial censuses." The lack of a quantitative standard in this definition has led to much discussion regarding the adoption of a standard for determining which districts will be considered competitive. At the time of this writing, the Commission had not adopted a formal definition, but it had decided to base its measure of competitiveness on an average of partisan outcomes (based only on votes for Democratic and Republican candidates) from 8 statewide elections from 2016 through 2020:

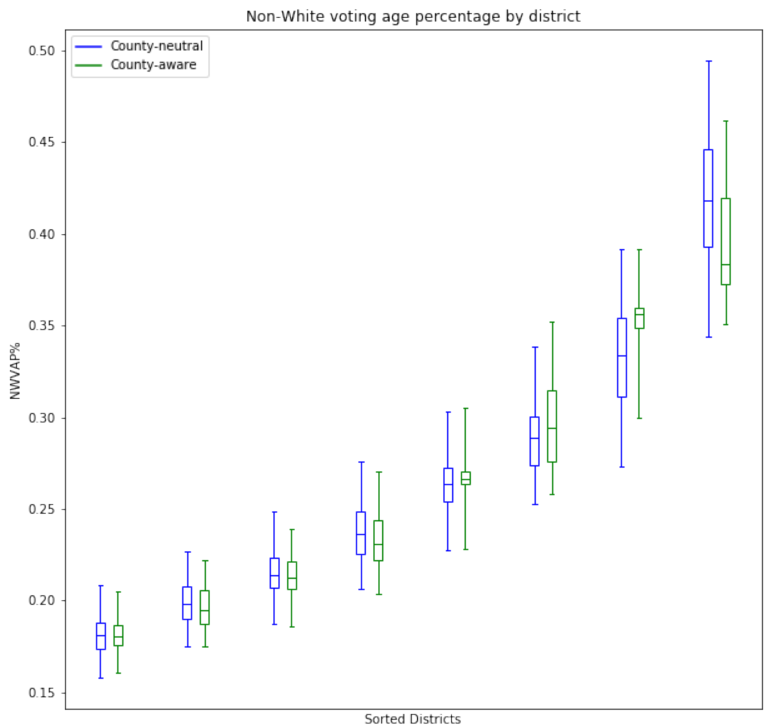

Figure 3: Non-White voting age percentage by district for county-neutral and county-aware ensembles

- the elections for President and U.S. Senator in 2016;
- the elections for Attorney General, Governor, Regent At Large, Secretary of State, and Treasurer in 2018;
- the election for U.S. Senator in 2020.

Each of these elections is given equal weight, creating a "composite election" whose Democratic and Republican vote percentages in each district are equal to the averages of the Democratic and Republican vote percentages, respectively, for these 8 elections in that district.

A typical measure of competitiveness involves prescribing a "vote band" about the 50% mark, and any election whose partisan vote share falls within that band is considered competitive. Since the Commission has not yet adopted a formal definition, we will consider three possible vote bands:

1. An 8.5% band (suggested to us by members of the Colorado Independent Legislative Redistricting Commission), so that partisan vote shares between 45.75% and 54.25% are considered competitive;
2. A 10% band (a common range found in academic literature on competitive elections), so that partisan vote shares between 45% and 55% are considered competitive;

3. An 11.5% band (in case the Commission is interested in the statistics for a wider band), so that partisan vote shares between 44.25% and 55.75% are considered competitive.

The histograms in Figures 4, 5, and 6 describe what percentage of plans in each ensemble have each possible number of competitive seats according to the given vote band, for both the county-neutral and county-aware ensembles.

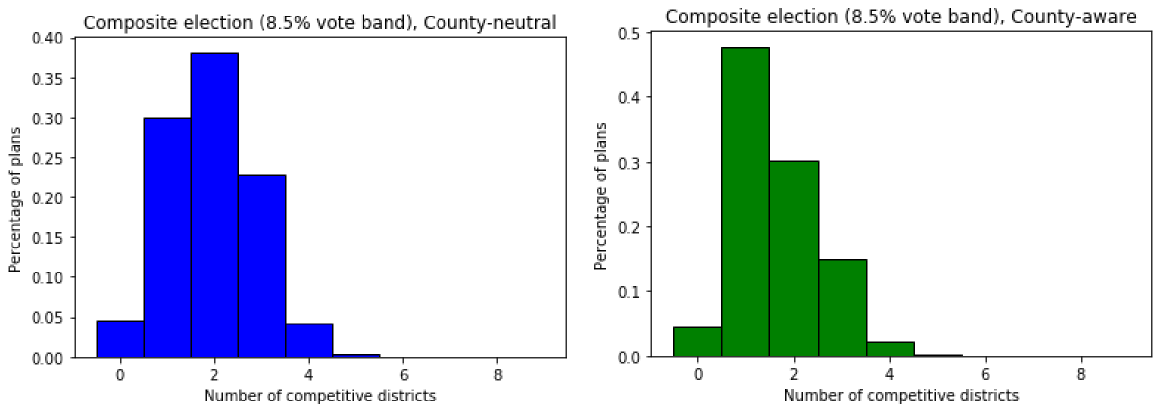

Figure 4: Numbers of competitive seats (8.5% vote band) for county-neutral and county-aware ensembles

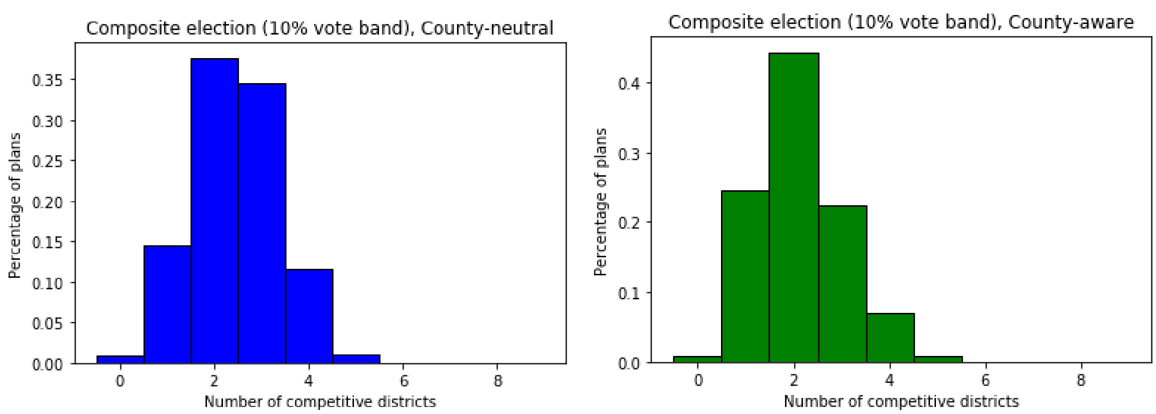

Figure 5: Numbers of competitive seats (10% vote band) for county-neutral and county-aware ensembles

For the county-neutral ensemble, the mean numbers of competitive districts are 1.93, 2.45, and 2.86 for the 8.5%, 10%, and 11.5% vote bands, respectively. For the county-aware ensemble, the mean numbers of competitive districts are 1.63, 2.12, and 2.60 for the 8.5%, 10%, and 11.5% vote bands, respectively.

It is important to observe here that, regardless of the vote band chosen, **constraining the number of county splits reduces the expected number of competitive districts,** with differences

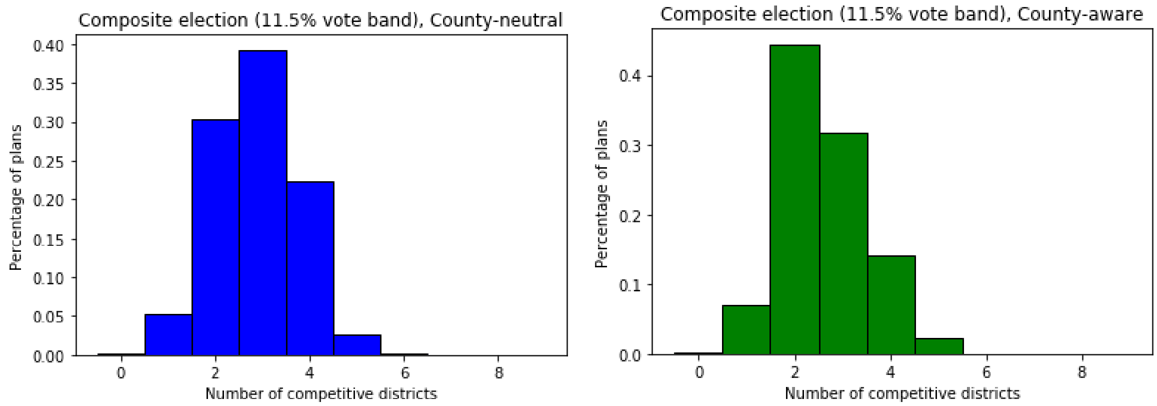

Figure 6: Numbers of competitive seats (11.5% vote band) for county-neutral and county-aware ensembles

ranging from 1/4 to 1/3 of a district on average. It seems likely that the same phenomenon may hold when the number of municipal splits is constrained, and perhaps when keeping other communities of interest whole is prioritized, although our experience suggests that this is a complicated issue that may behave in unexpected ways. (See our paper [5] for examples.)

For a more nuanced view on competitiveness, it is instructive to examine partisan outcomes by district. The box plots in Figure 7 are constructed similarly to those in Figures 2 and 3, except that now the boxes measure the observed ranges of Democratic vote share for each plan in the ensembles, ordered from most Republican to most Democratic. Also included in this plot are horizontal lines at the 50% mark and at the boundaries of each of the three vote bands for reference.

From this figure we can make the following observations:

- The 2 most Democratic districts are essentially never competitive.
- The 3rd most Democratic district is occasionally within the 10% and 11.5% vote bands, but almost never within the 8.5% vote band.
- The 4th most Democratic district is within the 11.5% vote band about half the time for both ensembles, and within the narrower bands somewhat less often. This district is significantly more likely to be within the 8.5% band for the county-neutral ensemble than for the county-aware ensemble.
- The 4th most Republican district is almost always competitive for all three vote bands.
- The 3rd most Republican district is almost always within the 11.5% vote band, and within the narrower bands slightly less often. This district is somewhat more likely to be within the 8.5% band for the county-neutral ensemble than for the county-aware ensemble.

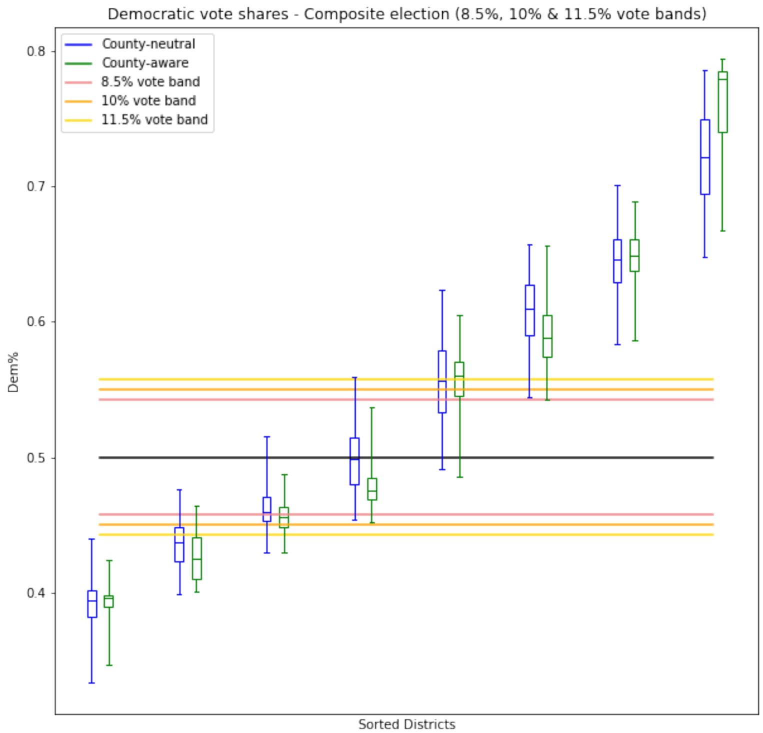

Figure 7: Democratic vote shares by district for county-neutral and county-aware ensembles, with competitiveness vote bands

- the 2nd most Republican district is occasionally competitive, mostly within the 11.5% vote band, and somewhat more likely to be within the narrower bands for the county-neutral ensemble than for the county-aware ensemble.
- The most Republican district is essentially never competitive.

3.4 Partisan seat share

Partisan seat share—i.e., the number of seats won by each political party in a particular election—is not one of the considerations prescribed by Amendment Y for district plans, but it is perhaps the outcome that is of the greatest interest to the most people. The histograms in Figure 8 describe what percentage of plans in each ensemble result in each possible number of Democratic seats won in the composite election. (The corresponding histograms for the numbers of Republican seats won would be the mirror images of the ones shown here.)

The two ensembles are strikingly different by this measure: For the county-neutral ensemble, outcomes of 4 and 5 Democratic seats are both very common, with 3 and 6 seats each being less common, but still not extreme outliers. But for the county-aware ensemble, over 80% of plans produce 4 Democratic seats, with 3 and 5 seats being much less common, and 6 seats an extreme

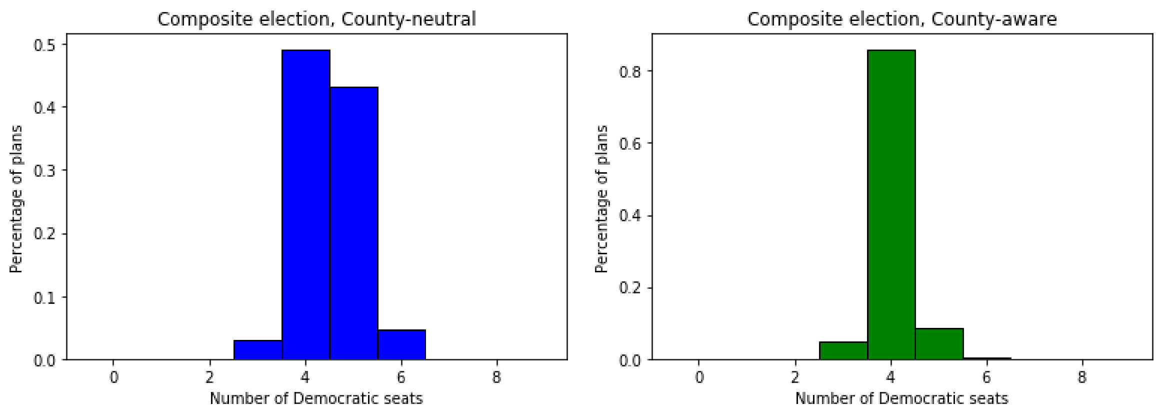

Figure 8: Numbers of Democratic seats won in composite election for county-neutral and county-aware ensembles

outlier which occurs for only 0.2% of plans. The mean numbers of Democratic seats are 4.50 seats for the county-neutral ensemble and 4.04 seats for the county-aware ensemble.

As for competitive districts, we can see a more nuanced picture in the box plots of Figure 7. The key difference is in the 4th most Republican seat, where the box representing the middle 50% of plans for the county-neutral ensemble is almost exactly centered at the 50% vote share line, while the analogous box for the county-aware ensemble is entirely below the 50% vote share line. This difference between the two ensembles explains why this district is about equally likely to be majority-Democrat or majority-Republican in the county-neutral ensemble, while it is usually majority-Republican in the county-aware ensemble.

4 Comparison of First Staff Plan to ensembles

On September 3, 2021, the Commission's nonpartisan staff released the First Staff Plan for Congressional districts. In this section we compare this plan to our ensembles for the measures described in the previous section.

We wish to emphasize that the First Staff Plan is **absolutely not expected to be at or near the mean values** for either ensemble with respect to all the measures that we have computed. Even if the plan were drawn entirely randomly, about half of its computed values would be expected to lie outside the middle 50% range for the ensemble. Furthermore, the Commission and nonpartisan staff are not attempting to draw a completely average plan, but rather to fulfill the Constitutional requirements that dictate that they attempt to preserve communities of interest and attempt to maximize the number of competitive districts. The comparison given here between the First Staff Plan and our ensembles is intended **only** to provide context which may be used by the Commission as just one of many measures to evaluate the First Staff Plan.

4.1 County Splits

The First Staff Plan splits 10 counties and contains 13 total splits. These values are very close to the mean values of 10.2 and 15.2 that we obtained for counties split and total splits, respectively, for our county-aware ensemble.

4.2 Minority representation

In Figures 9 and 10, we add the values for the districts in the First Staff Plan for the Hispanic voting age population and Non-White voting age population, respectively, to the box plots from Figures 2 and 3.

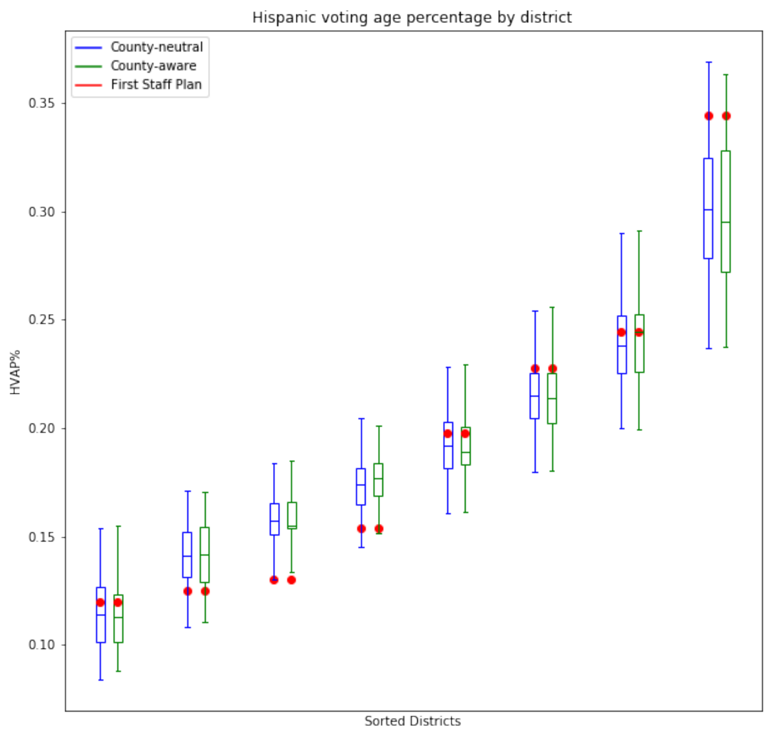

Figure 9: Hispanic voting age percentage by district for ensembles and First Staff Plan

For Hispanic voting age population, the First Staff Plan district with the highest percentage (District 8, with 34.4% HVAP) is somewhat above the middle 50% of both ensembles, but not an extreme outlier. The districts with the 2nd, 3rd, and 4th highest percentages (Districts 1, 3, and 6, with 24.4%, 22.8%, and 19.8% HVAP, respectively) are all slightly above the means of both ensembles, while the districts with the 5th, 6th, and 7th highest percentages (Districts 5, 7, and 2, with 15.4%, 13.0%, and 12.5% HVAP, respectively) are all well below the means of both ensembles. The district with the lowest percentage (District 4, with 12.0% HVAP) is slightly above the mean of both ensembles.

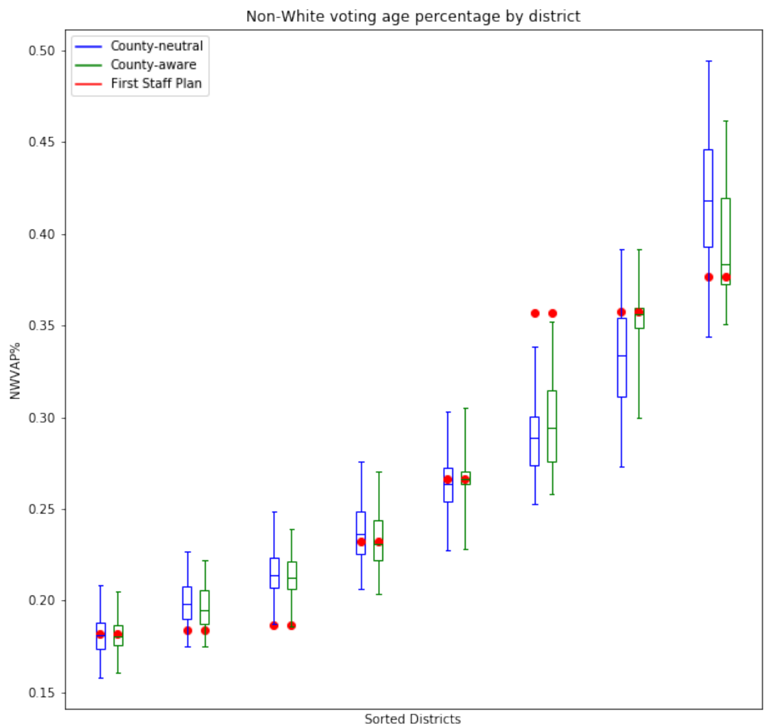

Figure 10: Non-White voting age percentage by district for ensembles and First Staff Plan

For Non-White voting age population, the First Staff Plan district with the highest percentage (District 6, with 37.7% NWVAP) is somewhat below the middle 50% of the county-neutral ensemble and slightly below the mean of the county-aware ensemble. The districts with the 2nd and 3rd highest percentages (Districts 8 and 1, with 35.8% and 35.7% NWVAP, respectively) are above the means of both ensembles, with the Non-White voting age population of District 8 high enough that it might be considered an extreme outlier with respect to both ensembles. The districts with the 4th and 5th highest percentages (Districts 5 and 3, with 26.7% and 23.2% NWVAP, respectively) are close to both ensemble averages. The districts with the 6th and 7th highest percentages (Districts 2 and 4, with 18.7% and 18.4% NWVAP, respectively) are below the means of both ensembles, while the district with the lowest percentage (District 7, with 18.2% NWVAP) is approximately at the mean of both ensemble averages.

4.3 Competitive districts

In Figures 11, 12, and 13, we have added the values for the districts in the First Staff Plan for the number of competitive districts for each of the three vote bands to the histograms from Figures 4, 5, and 6.

The First Staff Plan contains 3 competitive districts, regardless of which of the three vote bands

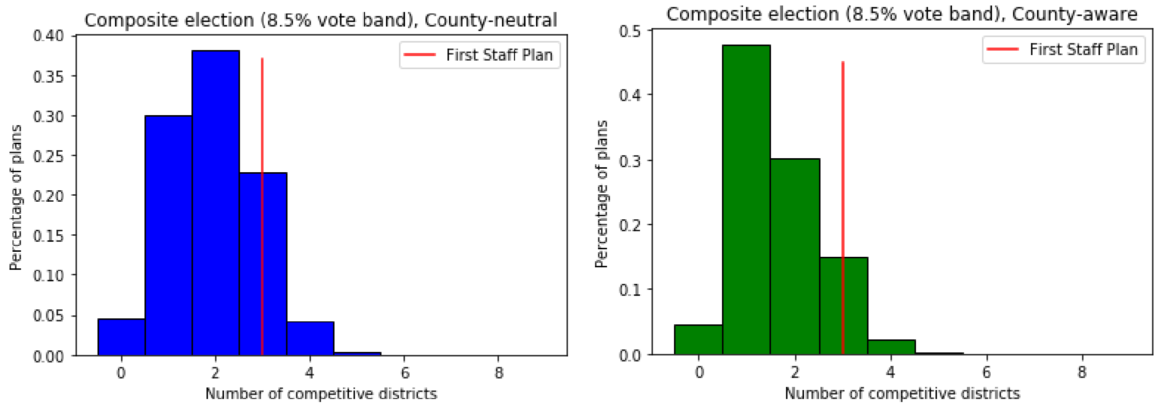

Figure 11: Numbers of competitive seats (8.5% vote band) for ensembles and First Staff Plan

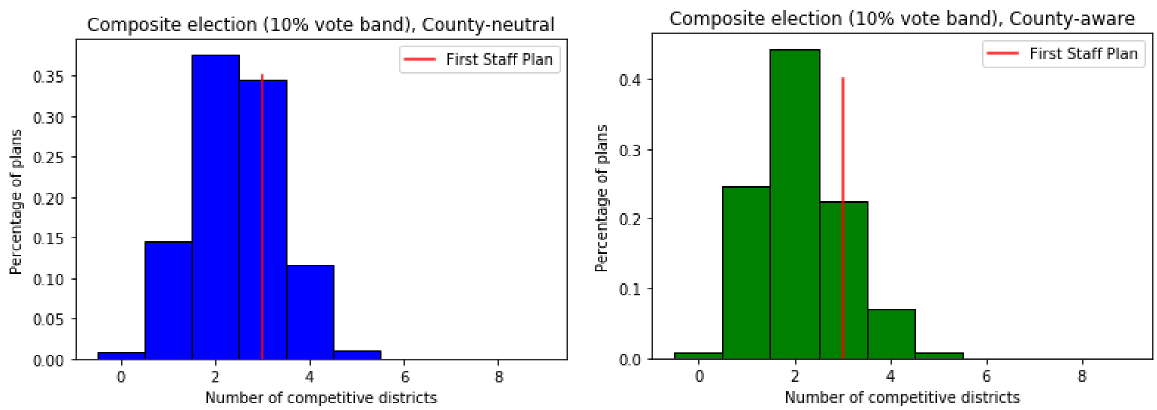

Figure 12: Numbers of competitive seats (10% vote band) for ensembles and First Staff Plan

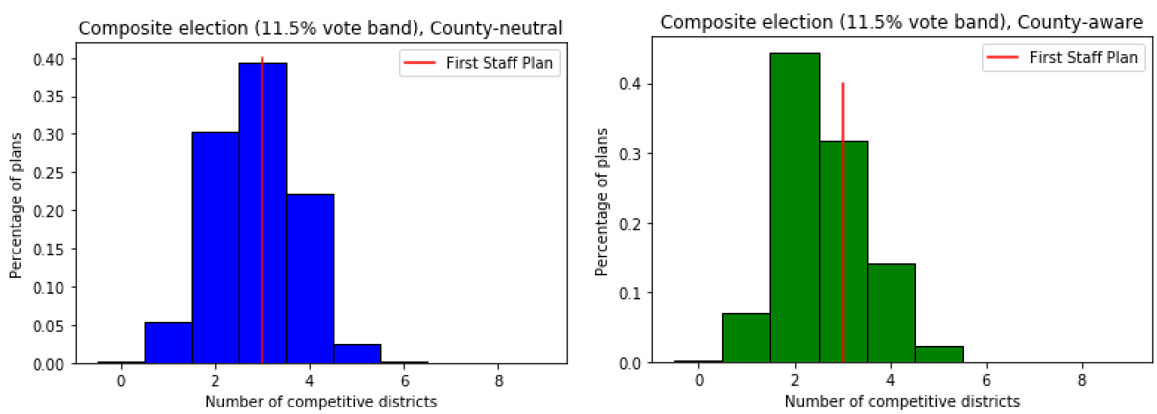

Figure 13: Numbers of competitive seats (11.5% vote band) for =ensembles and First Staff Plan

is used. This is above the mean for both ensembles and all vote bands, and significantly above the mean for the county-aware ensemble with 8.5% vote band. For an additional perspective, we note that:

- 27.3% of plans in our county-neutral ensemble and 17.5% of plans in our county-aware ensemble have 3 or more competitive districts with respect to the 8.5% vote band;
- 47.1% of plans in our county-neutral ensemble and 30.1% of plans in our county-aware ensemble have 3 or more competitive districts with respect to the 10% vote band;
- 64.2% of plans in our county-neutral ensemble and 48.4% of plans in our county-aware ensemble have 3 or more competitive districts with respect to the11.5% vote band.

It appears that the staff placed a high priority on maximizing the number of competitive districts within a fairly narrow vote band. Ensemble results suggest that if the vote band were set at 11.5%, it might be possible to find plans with 4 competitive districts, although this expectation may be affected by other priorities not included in our model.

In Figure 14, we have added the values for the districts in the First Staff Plan to the box plots for the Democratic vote share for the composite election from Figure 7.

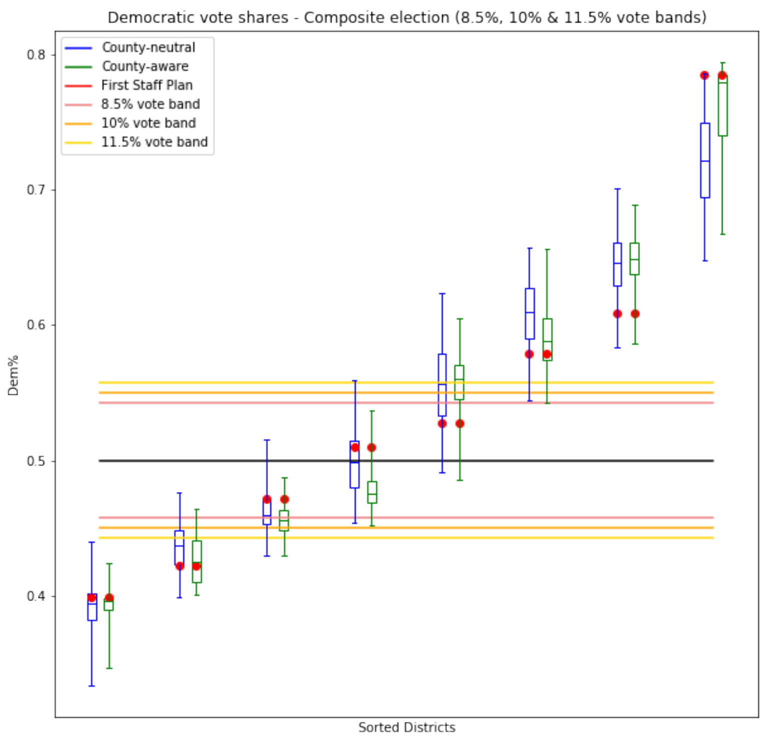

Figure 14: Democratic vote shares by district for ensembles and First Staff Plan, with competitiveness vote bands

Here we can see clearly how the First Staff Plan has been designed to create 3 maximally competitive districts: The 4th most Republican district (District 8, with vote shares of 50.7% D/49.3% R) will almost always be competitive. The 4th most Democratic district (District 7, with vote shares of 52.6% D/47.4% R) and the 3rd most Republican district (District 3, with vote shares of 47.1% D/52.9% R) are both just outside the middle 50% in both ensembles (and slightly further outside this range in the county-aware ensemble), in the direction of greater competitiveness. While neither of these districts look extreme with regard to Democratic/Republican vote shares, it should be noted that plans that achieve this threshold for both of these districts will be more rare than plans that only achieve it for one or the other.

If a wider vote band were chosen and an attempt made to draw a plan with 4 competitive districts, the districts which are next-closest (but not particularly close) to the 11.5% threshold are the 3rd most Democratic district (District 6, with vote shares of 57.8% D/42.2% R) and the 2nd most Republican district (District 4, with vote shares of 42.4% D/57.6% R). While these districts are equally spaced relative to the 50% line in the First Staff Plan, the ensemble suggests that it would be easier to move District 4 into the 11.5% vote band than District 6. But the actual map tells a slightly different story: because Districts 4 and 6 are adjacent, it could theoretically be possible to redraw the boundary between them so as to make both districts more competitive. However, this would come at the cost of splitting Arapahoe County and/or the city of Aurora and thereby breaking up significant communities of interest.

4.4 Partisan seat share

Finally, we compare the First Staff Plan to our ensembles regarding partisan seat share. In Figure 15, we have added the value for the districts in the First Staff Plan for the number of Democratic seats in the composite election to the histograms from Figure 8.

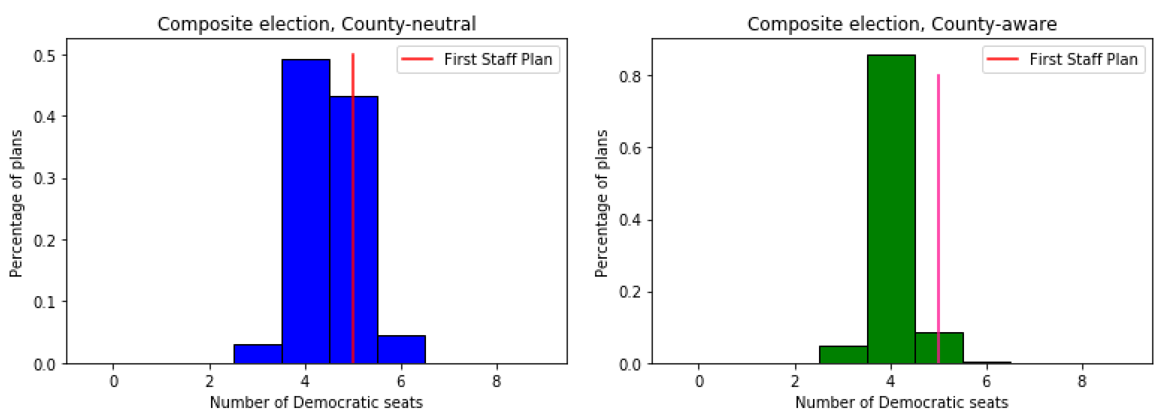

Figure 15: Numbers of Democratic seats won in ensembles and First Staff Plan

The First Staff Plan produces 5 Democratic seats for the composite election. For the county-neutral ensemble, 47.8% of plans produce 5 or more Democratic seats, but for the county-aware ensemble, only 9.0% of plans produce 5 or more Democratic seats. So for the county-aware ensemble, this result might be considered uncommon, but it is not an extreme outlier.

As we can see from Figure 14, the difference between the most common outcome of 4 Democratic and 4 Republican seats for our county-aware ensemble and the outcome of 5 Democratic and 3 Republican seats in the First Staff Plan is the 50.7% D/49.3% R vote share in District 8. This district is extremely competitive and has had both Democratic and Republican majorities among the 8 elections included in the composite election.

Our ensemble is designed to detect partisan bias in the form of extreme outliers, and our results illustrate that unbiased plans can have a range of outcomes, ranging from 3-6 Democratic seats if only one of the second tier priorities (compactness) is built in and from 3-5 Democratic seats if we build in a partial version of another second tier priority (county preservation). Our ensemble does not provide evidence that plans producing seat shares within these ranges display partisan bias. Moreover, building in variations to our ensemble generation algorithm to preserve additional political boundaries and/or communities of interest could produce slightly different ranges and distributions of reasonable outcomes. Consequently, we do not see any evidence of partisan bias in the design of the First Staff Plan.

4.5 Conclusion

The Commission and the nonpartisan staff have clearly put much thought and effort into the design of the First Staff Plan. Our computer-generated ensembles of plans cannot take into account the myriad of considerations that went into its design, or those that the Commission will prioritize for the remaining Staff Plans. For the measures that we did attempt to model—county preservation, minority representation, competitive districts, and partisan seat share—we do not detect any evidence of bias or other problematic features in the First Staff Plan.

5 Acknowledgments

We are very grateful to Moon Duchin and the MGGG Redistricting Lab for their pioneering work in ensemble analysis, for their open source python package GerryChain (available at https://github.com/mggg/GerryChain) which we have used for our computations, and for their ongoing support for our work.

References

[1] *League of Women Voters of Pennsylvania et al. v. Commonwealth of Pennsylvania et al.*, Supreme Court of Pennsylvania, 2018.

[2] *Common Cause v. Lewis*, Wake County, NC Superior Court, 2019.

[3] *Rucho v. Common Cause*, 588 U.S., 2019.

[4] Jeanne Clelland, Nicholas Bossenbroek, Thomas Heckmaster, Adam Nelson, Peter Rock, and Jade VanAusdall, *Compactness statistics for spanning tree recombination*, arXiv e-prints (2021), arXiv:2103.02699.

[5] Jeanne Clelland, Haley Colgate, Daryl DeFord, Beth Malmskog, and Flavia Sancier-Barbosa, *Colorado in Context: Congressional Redistricting and Competing Fairness Criteria in Colorado*, J Comput Soc Sc (2021), `https://doi.org/10.1007/s42001-021-00119-7`.

[6] Daryl DeFord, Moon Duchin, and Justin Solomon, *Recombination: A family of Markov chains for redistricting*, arXiv e-prints (2019), arXiv:1911.05725.

[7] Persi Diaconis, *The Markov chain Monte Carlo revolution*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 2, 179–205.

[8] Moon Duchin, *Outlier analysis for Pennsylvania congressional redistricting*, Expert report to Pennsylvania Governor Tom Wolf for League of Women Voters v. Commonwealth of Pennsylvania, 2018, accessed May 2020 at https://www.governor.pa.gov/wp-content/uploads/2018/02/md-report.pdf.

[9] Jonathan Mattingly, *Report on Redistricting: Drawing the Line*, Expert report to U.S. District Court for the Middle District of North Carolina for Common Cause v. Rucho, 2017, available at https://s10294.pcdn.co/wp-content/uploads/2016/05/Expert-Report-of-Jonathan-Mattingly.pdf.

[10] ———, *Expert Report on the North Carolina State Legislature*, Tech. report, 2019, available at https:// sites.duke.edu/quantifyinggerrymandering/files/2019/09/Report.pdf.

[11] Wesley Pegden, *Pennsylvania's Congressional districting is an outlier: Expert Report*, Expert report of Wesley Pegden in League of Women Voters of Pennsylvania v. Commonwealth of Pennsylvania, 2017, available at https://www.pubintlaw.org/wp-content/uploads/2017/06/Expert-Report-Wesley-Pegden.pdf.

[12] ———, Expert report of Wesley Pegden in Common Cause v. Lewis, 2019, available at https://www.commoncause.org/north-carolina/wp-content/uploads/sites/22/2019/09/Common-Cause-v.-Lewis-trial-court-decision-9.3.19.pdf, pp. 94-105.

A Technical details

A.1 Data collection

In order to build the precinct map used to generate ensembles, we obtained data from the following sources:

- A shapefile with the geographic boundaries of all 2020 voting precincts in Colorado, including precinct-level election results for all statewide elections in 2020, was given to us by Louis Pino from the Commission's nonpartisan staff.
- In the summer of 2019, the third author's student Haley Colgate compiled a shapefile with the geographic boundaries of all 2018 voting precincts in Colorado, including precinct-level election results for all statewide elections in 2018, with the assistance of Todd Bleess of the Colorado State Demography Office.
- A shapefile with the geographic boundaries of all 2016 voting precincts in Colorado, including precinct-level election results for all statewide elections in 2016, was obtained from the Voting and Election Science Team's repository on the Harvard Dataverse at https://dataverse.harvard.edu/dataverse/electionscience.
- Population data from the 2020 Census was taken from the 2020 PL 94-171 Data Summary File for Colorado based on the Decennial Census at the Census Block Level, obtained from the Redistricting Data Hub at https://redistrictingdatahub.org.

The open source python package Maup, developed by the MGGG Redistricting Lab and available at https://github.com/mggg/maup, was used to aggregate/disaggregate all population and election data from their original geographies onto the precinct geographies in the 2020 precinct shapefile. The resulting shapefile contains all the data required to compute population and election results for any district composed of 2020 precincts.

A.2 Ensemble generation

In order to generate our ensembles, we used the Recombination ("ReCom") method developed by the MGGG Redistricting Lab in 2018. (See [6] for a thorough treatment of this method.) For this method, the precinct map is modeled by a mathematical object called a **dual graph**, where each precinct is represented by a point called a **vertex**, and two vertices are connected by an **edge** if the precincts that they represent share a geographic boundary of positive length. A map of Colorado's 2020 voting precincts and its dual graph are shown in Figure 16.

A district plan is then represented by a partition of the dual graph into connected subgraphs, one for each district. (See Figure 17.) A partition is **valid** if it represents a legally valid district

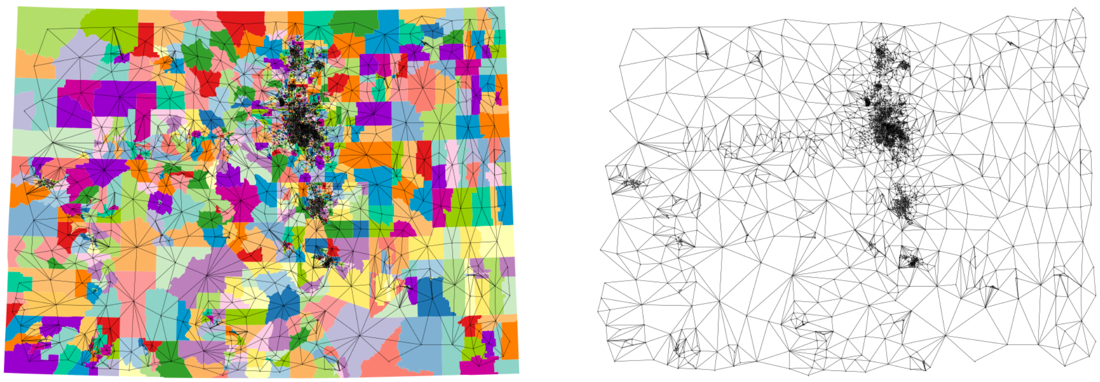

Figure 16: Colorado 2020 precinct map and dual graph

plan; at a minimum, the districts in the plan should be contiguous and have (approximately) equal population.

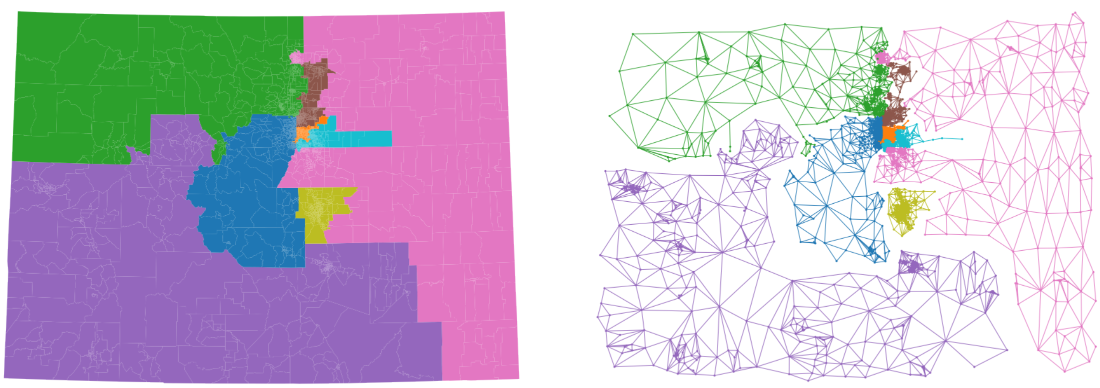

Figure 17: Staff Plan 1 Congressional districts and corresponding dual subgraphs

An ensemble starts with one randomly constructed valid plan, called the "seed plan." The ensemble is then constructed by a mathematical process called a **Markov chain,** in which each new plan is created by applying a random process to modify the previous plan in some way. For the ReCom method used to build our ensembles, this random process works as follows: At each step, the algorithm randomly selects a pair of adjacent districts and merges the two subgraphs corresponding to these districts into a single graph. Next, it generates a **spanning tree** for the merged graph—i.e., a subgraph consisting of all the graph's vertices and a subset of its edges, with the property that this subgraph is contiguous and has no closed loops—chosen randomly and uniformly from the set of all spanning trees of the merged graph. Finally, it looks for an edge to cut in order to create two new districts that each satisfy the population constraint. (District contiguity is automatic with this method.) This process is illustrated in Figure 18.

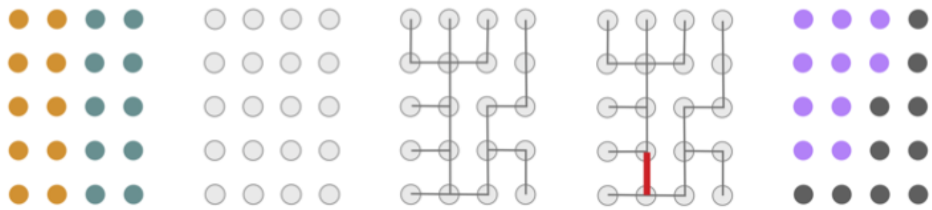

Figure 18: A ReCom step (Figure 4 in [6]; used with permission.)

Part of the appeal of the Markov chain approach is a well-developed theory and a long history of applications of Markov chain sampling methods (see, e.g., [7]). In particular, a sufficiently long Markov chain is theoretically guaranteed to produce an ensemble that accurately represents a specific probability distribution on the entire space of valid district plans. In general, this probability distribution is difficult to determine explicitly, but for the ReCom method there is good heuristic and experimental evidence indicating that the probability of any particular plan appearing in the ensemble is closely related to a natural discrete measure for district compactness. In practice, this means that this method is strongly biased towards plans with relatively compact districts and has no other detectable bias towards any particular type of plan (see, e.g., [4] and [6]). Some examples of plans produced by this method are shown in Figure 19.

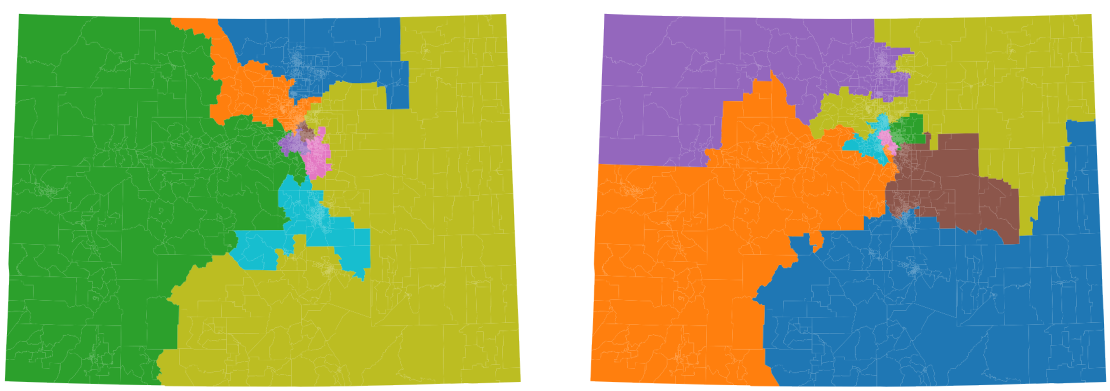

Figure 19: Examples of plans created by the ReCom method for county-neutral ensemble

Our county-neutral ensemble was generated with the basic ReCom method as described above. For our county-aware ensemble, a variation was used in the construction of the spanning tree for the merged graph, in which the random choice of edges to form the spanning tree is more heavily weighted towards intra-county edges, so that the resulting spanning tree contains relatively few edges connecting precincts in different counties. When the tree is cut, it is less likely to produce districts that split counties. As we can see from the histograms in Figure 1, this variation is quite effective in reducing the number of county splits in the resulting plans. Some examples of plans produced by this variation are shown in Figure 20.

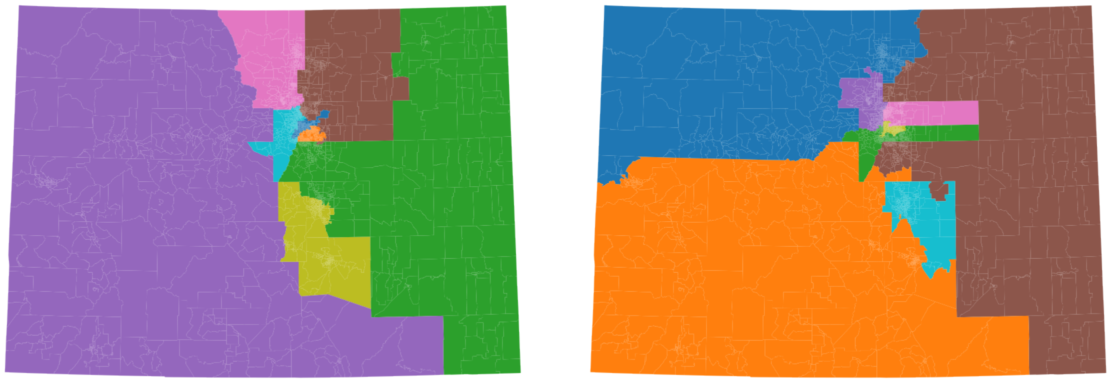

Figure 20: Examples of plans created by the ReCom method for county-aware ensemble

A.3 Ensemble size

Regarding the question of how long is "sufficiently long" for a Markov chain to produce a representative sample of plans, there is unfortunately no good theoretical answer. This question is usually answered heuristically, by running chains until statistics of interest appear to stabilize in a way that is not dependent upon the choice of seed plan. This stabilization is referred to as "convergence" of the statistics being measured.

For our ensembles of Congressional plans, we initially constructed three separate ensembles of 200,000 plans each, starting from three different seed plans with substantially different values for partisan seat share across districts for a variety of elections. (As these chains were intended only for benchmarking, we collected less data for these chains than for the final chains used for our main analysis.)

A typical example of the results of this experiment is shown in Figure 22, which shows the Democratic vote percentages by district for the 2018 Attorney General election for each of the three chains. The values for each of the three seed plans for each district are marked as dots of the corresponding color. As this figure shows, the boxes and whiskers for all three ensembles span essentially the same ranges, despite very different starting values for the seed plans.

For another example, Figure 22 shows the histograms for the Republican vote shares from the most Republican district in this election for each of the three chains, with the values for each of the three seed plans included for comparison. The agreement between the three histograms is not perfect—and in fact it never will be, regardless of how long we run the Markov chains—but we can clearly see the shape of the frequency distribution to understand which values are "typical" and which values might be regarded as extreme outliers.

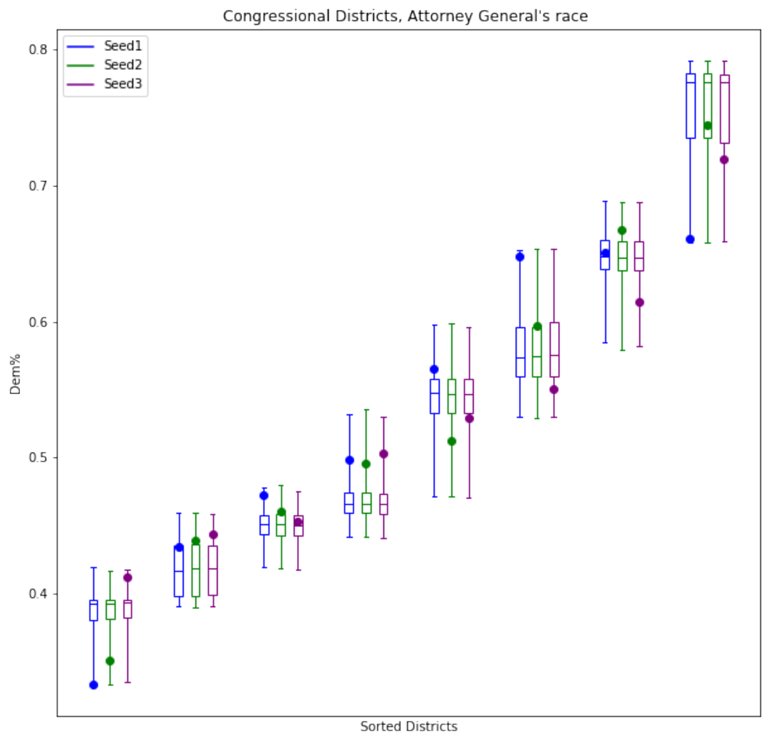

Figure 21: Democratic vote shares by district (AG18 election) for three ReCom ensembles of size 200,000 with different seed plans

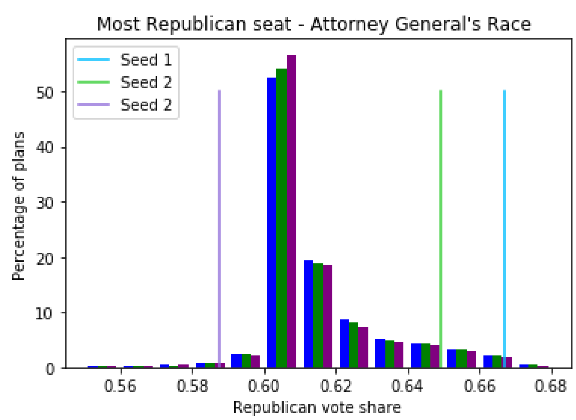

Figure 22: Republican vote shares for the most Republican district (2018 Attorney General election) for three ReCom ensembles of size 200,000 with different seed plans

All statistics that we examined for these three ensembles exhibited similar convergence behavior to those shown above after 200,000 steps, and we concluded that an ensemble of size 200,000 is sufficient to conduct a reliable analysis.