Comparison of Districting Plans for the Virginia House of Delegates

Metric Geometry and Gerrymandering Group

Abstract

At the time of writing, Virginia is in the process of replacing its House of Delegates districting plan after eleven of the districts were ruled unconstitutional by a District Court in June 2018. This report presents a large ensemble of alternative valid districting plans, which we propose to use as a baseline for comparison in the evaluation of newly proposed plans. Our method highlights and quantifies the dilutive effects of packing Black Voting Age Population.

This is a novel application to racial gerrymandering of industry-standard techniques from statistics and computational science.

Contents

1 Summary Report

1.1 Introduction

Using a mathematical sampling technique called a **Markov chain**, we have analyzed numerous alternatives for districting the portion of the Virginia House of Delegates map that is affected by the federal court ruling of June 26, 2018.[1] The results of our trials show the power and flexibility of approaches from mathematics and computing for assessing proposed maps in a redistricting process. A technical report detailing the findings follows below. This section serves as a non-technical overview of findings.

1.2 Scope and Goal for Study

In a June decision of the Eastern District Court of Virginia, 11 House districts were found to be unconstitutional; by expanding to the districts neighboring those, we arrive at a minimum of 33 House districts out of 100 that must be reexamined.[2] The majority of the three-judge panel found that racial considerations had predominated over traditional districting principles in the plan enacted in 2011, in particular, that Black residents were concentrated in a way that diluted their voting strength.

We have analyzed a collection of different House plans: the enacted plan approved by the Legislature in 2011 as House Bill 5005, the Democratic Caucus Plan released in August as HB 7001, the "reform map" independently released by the Princeton Gerrymandering Project on Aug 28, 2018, and a sequence of three plans circulated by Republican legislators in September and October. For each of these, our analysis focuses on the distribution of Black Voting Age Population, or BVAP, across the districts, though we note that this analysis can equally well be extended to focus on partisan performance, preservation of city boundaries, or other quantifiable priorities.

It was confirmed in statements to the court that the 2011 Enacted plan was designed to have $\geq$55% BVAP in 11 districts, which we may call the **packed** districts, without an effort to justify that numerical cutoff on the level of the individual districts. Compliance with the Voting Rights Act in no way requires this or any other numerical BVAP level. The conspicuous elevation of BVAP in the 11 packed districts must necessarily cause depressed BVAP in the 22 neighboring districts. We set

[1] In this decade, the drawing of political boundaries in Virginia was particularly contentious, with a significant number of lawsuits at all levels. A detailed summary of this process can be found at ballotpedia.org.

[2] Bethune–Hill v. Virginia Board of Elections

out to measure the pattern of that dilutive effect.

Without a baseline for the BVAP distribution, it is impossible to assess whether the reduced BVAP occurs in patterns that indicate a cost in additional opportunity districts, given the districting rules and priorities. Our main goal is to observe and quantify the interplay between elevated BVAP in part of the map and broader effects on the districting outcomes. The Markov chain method allows us to construct a baseline for comparison, in order to quantify these tradeoffs.

1.3 Our Main Tool: An Ensemble of Valid Alternative Plans

Using Markov chain techniques, we can apply long sequences of transformations to plans being evaluated, and to neutral "seed" plans, to assemble large collections of districting plans that are constructed only according to the stated rules and principles of a jurisdiction. For this study, our ensemble is built to take into account traditional districting principles in play in Virginia. Our Markov chain sampling process has population equality, contiguity, and compactness built into the steps.[3] In this way, we get a picture of how well the proposed plans comport with the principles found in state and federal law and we gain a sense of how much these principles may be sacrificed if other unstated goals are in play.[4]

1.4 Black Voting Age Population and the Crucial 37-55% Range

The recent lawsuit was brought under the federal Voting Rights Act of 1965, which continues to be a fundamental check on redistricting practices. The range of BVAP values from 37% and 55% is a crucial zone for VRA analysis nationally and in Virginia in particular. Though we emphasize that **BVAP alone is never enough to confirm VRA compliance**, this wide range of BVAP values often triggers the evaluation of racially polarized voting (RPV) patterns as one of several other considerations addressing the totality of circumstances in a possible VRA violation. Nationally, 37% is an empirical bright line for congressional voting: 32 out of 34 current U.S. congressional districts with at least 37% BVAP had Representatives in the 115th Congress who belong to the Congressional Black Caucus, and the ratio drops off precipitously below that level.[5] Furthermore, Virginia-specific data legitimates the significance of bracketing the 37-55% range, which is a large and obvious gap observable in the BVAP of the 2011 Enacted plan. The expert reports in Bethune–Hill v. Virginia Board of Elections show no RPV evidence that any House districts in Virginia that would require numbers over 55% BVAP to comply with the Voting Rights Act, and the 2011 Enacted plan was specifically faulted for aiming districts above the 55% line. Indeed, the report of Maxwell Palmer indicates that a BVAP of 45% would suffice for all but one House district, and 48% would suffice for every House district

[3] The following districting principles are represented in Virginia law and were recognized by the district court to be relevant in this process: population equality; compactness; contiguity; preservation of municipal boundaries; preservation of communities of interest. We performed runs of MCMC while imposing a contiguity requirement and bounds on the allowed compactness and population deviation. Our code allows us to turn on a feature that limits the level of municipal boundary splitting—details available upon request.

[4] Our ensembles are built with software developed and made public by the Voting Rights Data Institute (github.com/mggg/GerryChain). The steps in the chain, the validity checks, and the heuristics that give us increased confidence of representative sampling are described in the technical report attached below.

[5] To see full data on national BVAP percentage and representation in the 115th Congress, visit github.com/gerrymandr/bvap-cbc-notebook/.

in the state.[6] In addition to Palmer's RPV analysis, we note that most of the 33-district affected region is covered by two congressional districts, VA-3 and VA-4, which are comfortably electing Black representatives with BVAP of 45.1% and 41.5%, respectively.

It is important to remember that in some cases, less than 37% BVAP can suffice for a district to provide an opportunity to elect a candidate of choice of the Black community, especially in coalition with other groups; in rare cases (particularly in other parts of the country), more than 55% BVAP may be legitimately required. Still, the figures below highlight the 37-55% range to illustrate the statistical effects of packing Black population: We can measure the cost of pushing districts above that crucial zone by observing whether it causes other districts to be pushed below.

1.5 Methods and Results

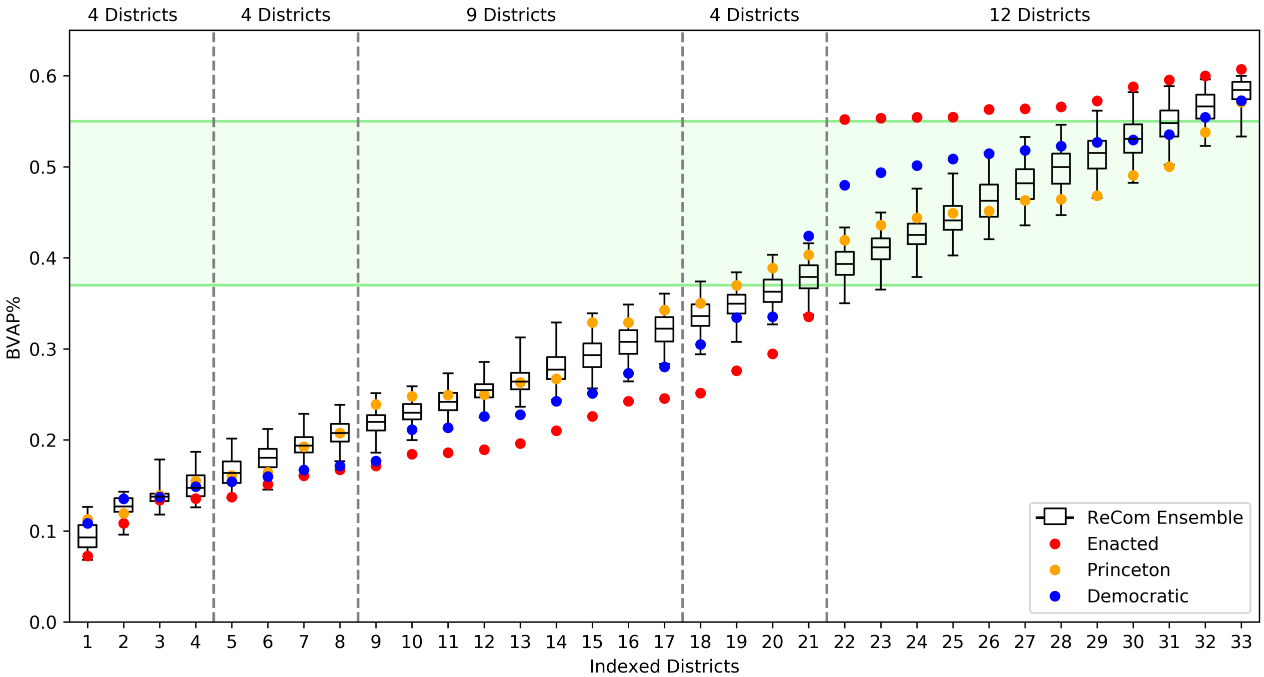
We have constructed large collections of 33-district plans called "ensembles," defined on the region of the state that is directly affected by the court ruling. To compare the BVAP levels among those plans, we have indexed (sorted) the 33 districts from the one with lowest BVAP (1st, appearing leftmost in the figure) to the one with highest BVAP (33rd). That means that in each plan, the twelve districts with the highest levels of Black Voting Age Population are indexed 22-33, and are separated by a dotted line in the figure. We have broken the districts into groups—top 12, next 4, next 9, next 4, final 4—to make it easier to visualize where the effects of elevated BVAP occur.

Our algorithm begins with many different valid starting plans: the plans that have been proposed for adoption and 100 neutral maps, labeled Seed1 through Seed100. We then perform chains of random alterations, collecting a large sample from the resulting maps as our ensemble of comparable plans.

Figure 1 shows one compelling set of results as a box-and-whisker plot. It depicts the ensemble of 20,000 steps from a "recombination" Markov chain, winnowed to show the plans that do not exceed 60% BVAP. (Many more figures with different combinations of hypotheses are described in the technical report, lending robust support to similar conclusions.) The boxes show the 25th-75th percentile of BVAP observed for a district in that position; the median is marked with a horizontal line. The whiskers show the 1st-99th percentile of observations. When a colored dot falls far outside of the whiskers, it means that the plan is an extreme outlier in its racial composition for that district.

The plot gives unmistakable evidence of where and how elevating BVAP in the top 12 districts depresses BVAP in the remainder of the plans. Rather than reducing the BVAP in the areas where it was already very low, we see that the dilutive effects impact districts that were at or nearing the zone in which statistical analysis has indicated opportunities to elect more candidates of choice for the Black community.

[6] Palmer uses a the standard statistical technique, called Ecological Inference (EI), to assess racially polarized voting. An open-source app for using EI on voting data can be found at mggg.org. We remind the reader that, contrary to widespread popular belief, there is no legal requirement that districts be created with majority-minority status to achieve VRA compliance.



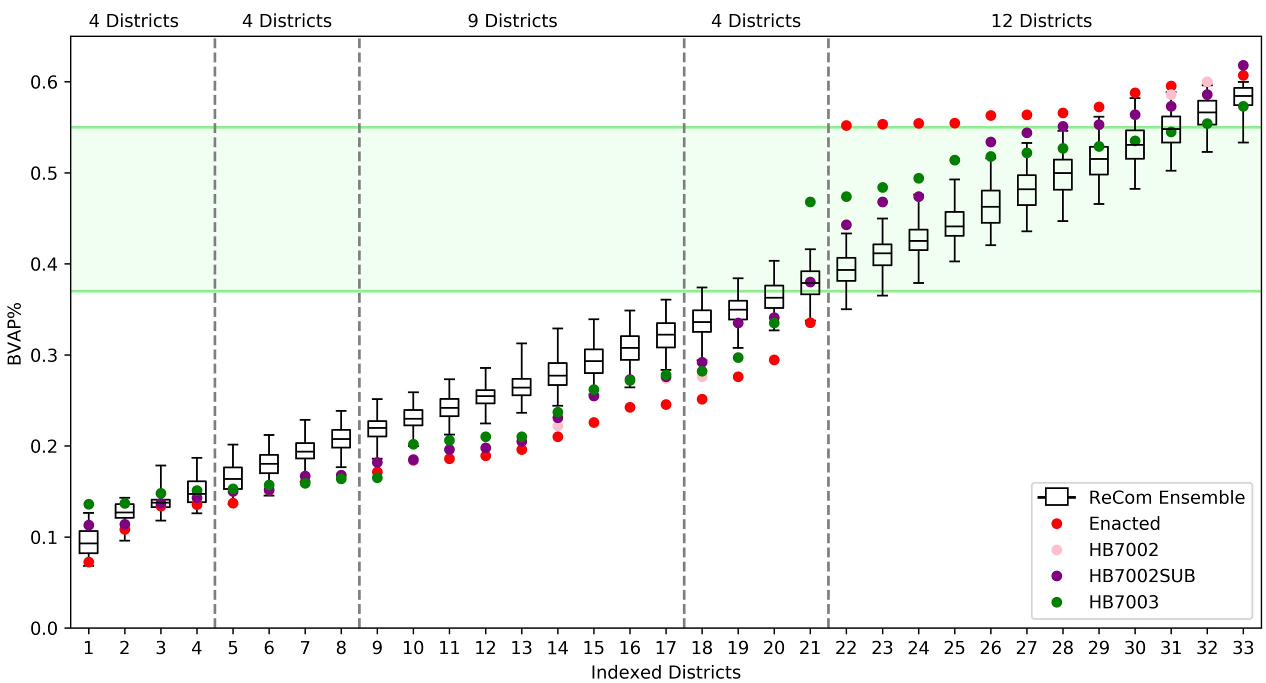


Figure 1 Comparison of the BVAP of the six proposed plans to an ensemble of neutral comparison maps that has been winnowed so each district has $\leq 60\%$ BVAP. The 33 districts of the affected region are arranged on the x axis, and the percentages of Black Voting Age Population are shown on the y axis. The 37-55% BVAP zone is marked in green, and the districts are separated into groups for ease of interpretation.

1.6 Conclusions

- By starkly elevating BVAP in the six districts indexed 22-27, the 2011 Enacted plan causes at least ten and up to 17 other districts to have depressed BVAP levels, far below what would be expected from race-neutral redistricting. The 2011 Enacted plan has no districts at all in the crucial range of 37%-55% BVAP, while neutral redistricting tells us to expect as many as ten.
- The Democratic Caucus plan and all new Republican proposals temper the packing in the top districts but only push one additional district over 37% BVAP.
- The Princeton plan—and hundreds of thousands of race-neutral plans found by Markov chain techniques—push three additional districts over that level without sacrificing population balance, contiguity, or compactness. In fact, our methods suggest that a substantial share of race-neutral plans that comport with traditional districting principles would do so.

We emphasize that there are many local and community-based considerations in play when approving districting plans, and the Markov chain approach only provides data relevant to some of these. It does so by illustrating the range of properties typically observable in the enormous landscape of valid plans. We view this approach as one tool among many in a complex process for evaluating districting plans, and we hope it will incorporated into the analysis of proposed plans to the House of Delegates. We are prepared to analyze new maps as they are proposed.

2 Technical Report

This report describes the technical details associated with our analysis of Virginia House of Delegates districting plans. We describe our methodology for building an ensemble of valid plans to be used as a baseline for assessing the Black Voting Age Percentage in a newly proposed plan. We also discuss the trade–offs between three proposal methods, as the standard approach in the literature does not allow for effective representative sampling.

2.1 Terminology

- **Affected Region:** The region covering approximately one-third of the state of Virginia comprised of the 11 districts from the 2011 Enacted Plan that were found to be unconstitutional in June 2018, together with the 22 more districts that are adjacent to these. (See Figure 2c.)
- **Census Blocks:** Blocks are the smallest geographic units reported by the decennial census and are held constant throughout each decade. There are approximately 300,000 blocks in Virginia and approximately 100,000 blocks in the affected region.
- **Dual Graph:** A network representation of the geography of the state: each node represents a census block and there is an edge between two blocks if they share a border of positive length. At each vertex, we record the population of the block.
- **Boxplot, or Box and Whiskers Plot:** For a dataset consisting of vectors, a standard statistical visualization is to exhibit the data in a plot that uses a box in each component variable to visualize the 25th-75th percentile range in the observed data. "Whiskers" then show a wider range of data, such as the 1st to 99th percentile range, which is the standard used below. The use of boxplot analysis for voting share by party has been used extensively by Jonathan Mattingly and his research team, including in redistricting litigation.[1] Here we use boxplots to understand the racial distribution across districts in proposed plans.
- **BVAP, or Black Voting Age Population:** The BVAP of a geographical region is the number of Black residents whose age is at least 18 as recorded by the 2010 census.

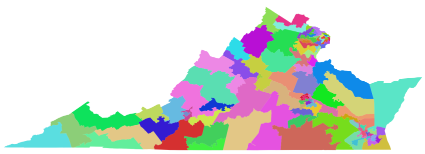
[1] See Mattingly's "Quantifying Gerrymandering in North Carolina" (arxiv.org/pdf/1801.03783.pdf) for a boxplot-based analysis of partisan patterns in North Carolina districts. This evidence was regarded favorably by the District Court for the Middle District of North Carolina in Common Cause v. Rucho (www.commoncause.org/wp-content/uploads/2018/08/MDNC-Opinion.pdf).

- **BVAP Vector:** For a given plan covering the affected region, this is a vector with 33 entries, which are the ratios of BVAP/VAP in the 33 districts, arranged from least to highest. When any statistic is reported in vector form from its lowest to its highest level over the districts in each plan in an ensemble, we say that statistic has been **indexed** or **sorted** over the ensemble.
- **Shapefile:** A shapefile is a data format for storing annotated geographic information. Each block, precinct, and district is represented by a polygon with vertices at geographic coordinates
- **Ensemble:** An ensemble of districting plans is a (large) collection of plans that satisfy the basic constraints of the districting process. Properties of a proposed plan can be compared to the same properties averaged over the entire ensemble.
- **Proposal Method:** In Markov chain Monte Carlo (MCMC), the proposal is a step used to generate a new sample plan based on the current one. We use two main types of proposals, called a flip step ("Flip") and a recombination step ("ReCom"). More details are given in §2.3.2.
- **Proposed Plans:** The following plans are discussed in this report. Figure 2 shows the maps themselves along with the corresponding BVAP% for the impacted districts.
  1. **Enacted:** This plan was approved by the legislature in 2011 as HB5005 and had 11 districts ruled unconstitutional in 2018.
  2. **Dem:** This plan was proposed by the Democratic Caucus as HB 7001 in the first emergency legislative session.
  3. **Princeton:** This plan was constructed by the Princeton Gerrymandering Project as a case study for the Virginia Pilot.
  4. **GOP1:** This plan was proposed by the Republican majority as HB 7002.
  5. **GOP2:** This plan was proposed by the Republican majority as a variant of HB 7002.
  6. **GOP3:** This plan was proposed by the Republican majority as HB 7003 based on the remedial districts proposed in HB 7001.

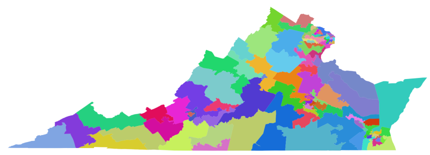
2.2 Data

Collecting the necessary information about geography, demography, and voting results in a compatible fashion is a complex and onerous process. Below we outline the steps we followed to construct data sets for analysis. Cleaned versions of the shapefiles described below can be downloaded from our GitHub repository.[2] For this analysis, we focus on the affected region: 33 affected districts, comprising the 11 found to be unconstitutional and the 22 more districts that are adjacent to those.

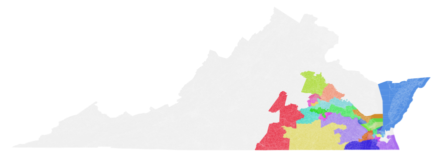
[2](github.com/gerrymandr/VA-Technical-Report)



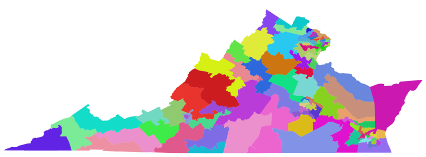
(a) Enacted



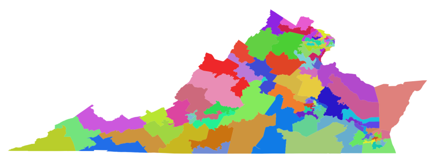
(b) Dem



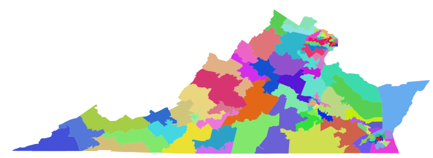
(c) Princeton



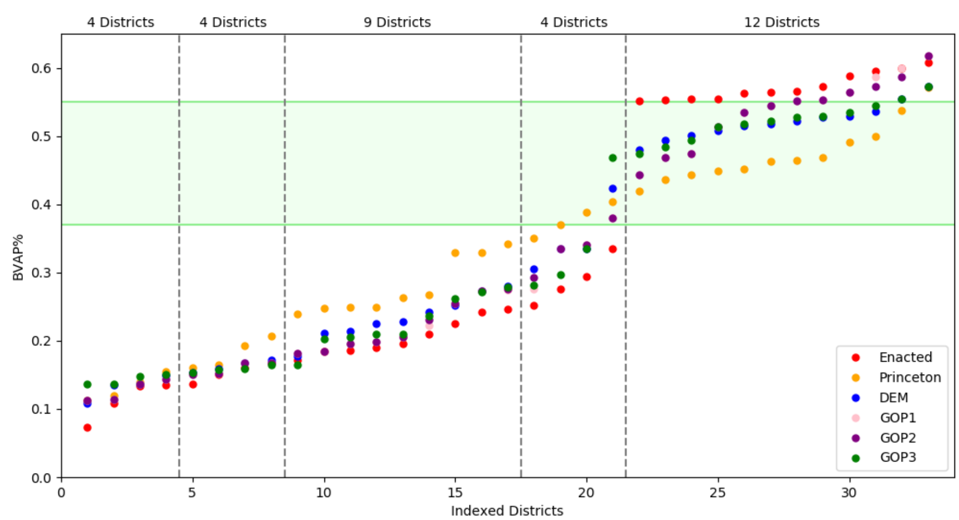
(d) GOP1



(e) GOP2



(f) GOP3



(g) Proposed Plans BVAP

Figure 2 These figures show the six proposed maps analyzed in this report. Note that Princeton only redrew the impacted 33 districts. Figures (g) shows the BVAP% for these proposed plans.

2.2.1 Data Collection

The underlying geography and population data was downloaded from the Census.[3] The plans voted on in the legislature are provided by block assignment file.[4] These include the 2011 Enacted plan as well as the plan proposed by the Democratic caucus and plans proposed by the Republican majority. Data for the Princeton plan is available in their github repository.[5]

To report the scores of partisan metrics such as the mean-median gap (§2.4.1), we need to associate partisan voting data to census blocks. Virginia provides historical election data by precinct through the Secretary of State.[6] Not all of these values are easily associated to geographic units, as no official shapefile exists for the precincts. The Princeton Gerrymandering Project contacted each county to obtain the boundaries of the precincts as of 2016 and compiled these electronically.Note that the coding between the state data and the precinct shapefile do not match exactly—there are even some precincts with no votes or even negative votes in some elections. However, aggregate results have been confirmed to match the reported final results of those elections.

2.2.2 Preprocessing

The geographic and census data were merged from shapefiles provided by the census and the state of Virginia using QGIS. First a block-level geographic map of the state was constructed with the population and demographic data. For the plans from the state, the blocks were then assigned to districts using the block assignment files. For the Princeton plan, the map was rounded off onto the blocks using the open source preprocessing software[7] developed at the Voting Rights Data Institute (VRDI), a research initiative of the Metric Geometry and Gerrymandering Group.

There are some slight difference between the demographic data provided by the state of Virginia that we used for our analysis and the federal census data. The data used by Virginia[8] and the relevant census data[9] are available online. For the initial BVAP values, the difference per district is less than 1% for all of the proposed plans.

Once the plans were defined on the entire state, the blocks associated to the eleven unconstitutional districts (63, 69, 70, 71, 74, 77, 80, 89, 90, 92, and 95) and the twenty-two neighboring districts (27, 55, 61, 62, 64, 66, 68, 72, 73, 75, 76, 78, 79, 81, 83, 85, 91, 93, 94, 96, 97, and 100) were extracted into separate shapefiles. Note that the same collection of blocks is not used across all plans for these districts as the exterior boundaries vary slightly in each map.

Starting with the precinct shapefile assembled by the Princeton team, we prorated the votes onto blocks where possible with the VRDI preprocessing software linked above. As this shapefile is only available for the 33 impacted districts, we only provide partisan results for these areas. We used statewide election data from the 2016 presidential, gubernatorial, attorney general, and lieutenant governor races.

[3](www.census.gov/geo/maps-data/data/tiger-data.html)
[4](redistricting.dls.virginia.gov/2010/RedistrictingPlans.aspx)
[5](github.com/PrincetonUniversity/VA-gerrymander)
[6](historical.elections.virginia.gov/)
[7](github.com/gerrymandr/preprocessing)
[8](redistricting.dls.virginia.gov/2010/Census2010.aspx)
[9](www.census.gov/geo/maps-data/data/tiger-data.html)

2.3 Methodology

The core of our methodology is the construction of a Markov chain—a random walk with no memory—on the space of possible districting plans, designed to sample efficiently from the subspace of plans with prescribed properties. The samples constructed by this method are called ensembles of alternatives. Recently, Markov chain analyses of contested plans have been used by experts in the redistricting litigation in North Carolina and Pennsylvania.[10]

To construct an ensemble of plans with a Markov chain, we begin with an initial districting plan, or "seed," and then generate a sequence of new plans by modifying the current plan according to a stochastic (i.e., randomized) rule or proposal distribution. A method for generating initial plans is described in §2.3.1 and the proposals we considered, called Flip and ReCom, are described more fully in §2.3.2 below. We found that high-quality results were only obtained with the ReCom proposal, or with a Mix proposal combining flip and recombination steps.

2.3.1 Starting Points: Generating Seed Plans

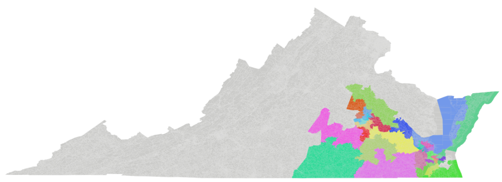
We generated 100 plans balanced to within 0.1% population deviation using a method to recursively select one district at a time. To explain our method, we first recall the idea of a dual graph of the census geography. The roughly 100,000 census blocks of the affected region can be represented using a network or graph with one node for each block and an edge connecting each pair of adjacent blocks. For the affected region in the 2011 Enacted plan, this graph has 91,522 nodes and 222,888 edges. A spanning tree for a graph is a (connected) subgraph in which all of the vertices are maintained, but only some of the edges—by definition, it is a smaller network that has no cycles in its edges but touches every vertex.

Our process of creating seed plans is to start with a randomly chosen spanning tree for the entire affected region, then randomly seek an edge to cut that will leave one of its two resulting connected components within 0.1% of the ideal population size. If such an edge does not exist we randomly generate a new spanning tree. After a district is selected, a new spanning tree is generated for the complement, and the process continues until the entire region has been partitioned into pieces with nearly balanced populations.[11] Example seed plans are shown in Figure 3 and their corresponding BVAP vectors are shown in Figure 6. The 100 seeds already show a range of possible behaviors; some of them have extremely high BVAP in one or even two of the 33 districts.

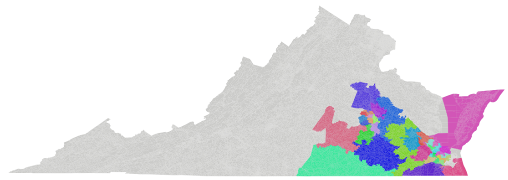
By construction, our seed plans are closely population-balanced. The spanning tree method also tends to produce plans with favorable compactness scores—shapes in a network that have more interior nodes relative to the length of their boundary will have more spanning trees, so will be upweighted by the random spanning tree method. So while we do not claim that the seeds themselves provide representative sample of the space of possibilities, they can be regarded as demonstration plans that show it is possible to get many kinds of BVAP distributions while remaining geographically compact and population-balanced.

[10] In North Carolina, Jonathan Mattingly of Duke testified for the plaintiffs. In Pennsylvania, Wesley Pegden of CMU testified for the plaintiffs, and Moon Duchin of Tufts consulted for the Governor. Reports can be found here:
s10294.pcdn.co/wp-content/uploads/2016/05/Expert-Report-of-Jonathan-Mattingly.pdf
www.pubintlaw.org/wp-content/uploads/2017/06/Expert-Report-Wesley-Pegden.pdf
www.governor.pa.gov/wp-content/uploads/2018/02/md-report.pdf

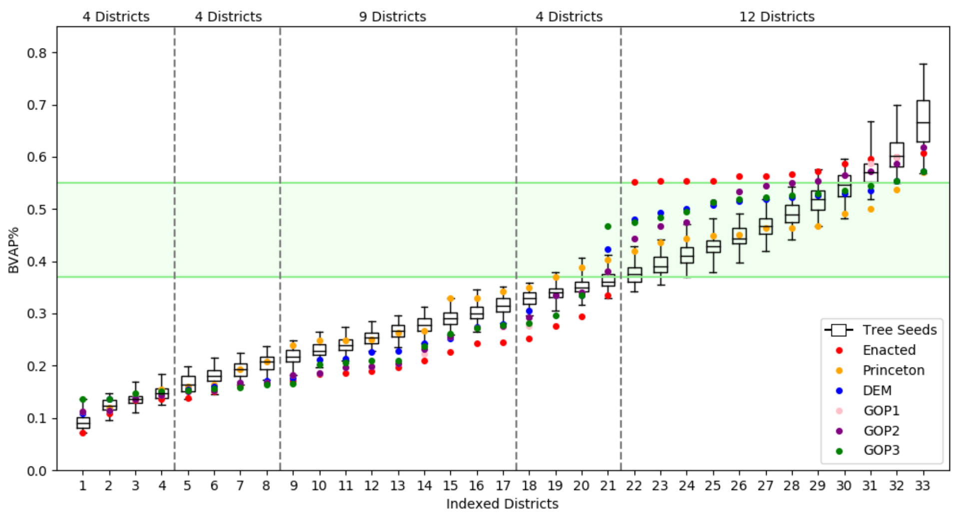
[11] An open-source implementation is available in GerryChain github.com/mggg/GerryChain.



(a) Seed31 Plan



(b) Seed99 Plan



(c) Proposed Plans vs. 100 Seeds

Figure 3 Two examples of the initial seeds made with random spanning trees (a-b) along with a visualization of the full collection of 100 tree seeds (c). No partisan or demographic data is included in this construction, only geography and population.

2.3.2 Altering Existing Plans: Proposal Methods

In our setting, a proposal method is a randomized algorithm for generating a new plan from an existing plan. The Markov chain builds an ensemble iteratively by applying the algorithm to each newly generated plan, starting with the initial plan. The defining property of a Markov chain is that the probability of observing each possible plan next, at a particular iteration, is determined only by the plan at the previous iteration. Our proposal methods all modify an existing plan by randomly choosing a set of its units for which to switch the districting assignment.

There are three designs of Markov chains that are used to build ensembles below.

- **Single edge flip ("Flip").** At each step in the Markov chain, we (uniformly) randomly select a pair of adjacent census blocks that are assigned to different districts, then (uniformly) randomly change the assignment of one of the blocks so that they match.

- **Recombination ("ReCom").** At each step, we (uniformly) randomly select a pair of adjacent districts and merge all of their blocks in to a single unit. Then, we generate a spanning tree for the blocks of the merged unit with the Kruskal/Karger algorithm. Finally, we cut an edge of the tree at random, checking that this separates the region into two new districts that are population balanced to within 1% of ideal district size. Figure 4 shows an illustration of this process.
- **Flip/Recombination mix ("Mix").** We draw a Bernoulli random variable at each step and take a Flip step with probability 99% and a ReCom step with probability 1%.

The Markov analyses used in the court cases discussed in footnote 10 above all used versions of the Flip proposal to generate their ensembles. While this method is efficient, it has some undesirable properties that are exacerbated in this model of Virginia by the choice to work with very small units—census blocks—because of stringent population balance requirements. Shortcomings of the use of Flip ensembles on census blocks will become evident in the results below.

The ReCom proposal is a novel method for the redistricting application, borrowing the name "recombination" from biological terminology for information crossover between two or more entities. In other computational science literature, genetic algorithms sometimes have a similar type of step. What is notable here is that our recombination walk is still a Markov chain, so the theory of MCMC still applies to these methods.

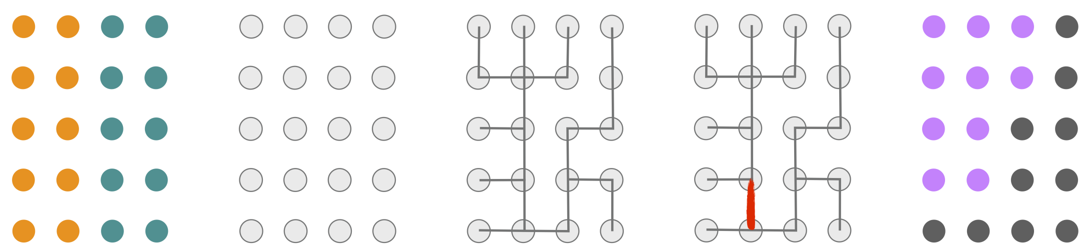
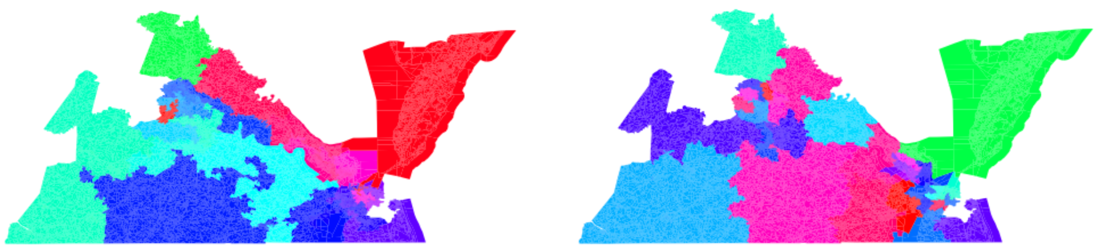


Figure 4 A schematic demonstrating the ReCom step. We begin with an initial plan and select two adjacent districts to merge. A spanning tree is constructed for the merged region, and an edge is identified that will separate the merged area into two equal-sized districts. Removing that edge determines the new plan.

2.3.3 Checking Validity: Rules and Constraints

The House of Delegates in Virginia divides the state into 100 districts, which by law must be contiguous regions with populations balanced to within a maximum of 1% deviation from ideal. Using the 2010 census figures, this means each district has a target of 80,010 people and must lie between a minimum of 79,210 and a maximum of 80,810 people. We required every map in our ensemble to meet this 1% deviation standard.

We can also control compactness in terms of the edges in the dual graph. In redistricting, the term 'compact' has been used for districts that are seen as reasonably shaped, or have tame-looking boundary lines. Such districts will not require very many edges to be cut to separate them from the rest of the state; by contrast, a plan with snaky, winding boundaries will require a large number of cuts. We will use the ratio of total edges to cut edges as a compactness score. Our chain runs limit the number of cut edges in each plan to 12,000, for a worst allowable score of $222888/12000 \approx 18.6$. By contrast, the 2011 Enacted plan has 6110 cut edges and scores roughly 36.5, and the plan in Figure 5 (b) scores over 39.[12] See §2.4.3 for information on the compactness performance of our ensembles.



(a) Plan with 11,996 cut edges

(b) Plan with 5702 cut edges

Figure 5 These examples illustrate that a compactness score based on cut edges tracks well with visual impressions of district shape. The first, drawn from a Flip chain, has a poor compactness score of roughly 18.6, meaning that about one in 19 edges had to be cut from the dual graph to separate the districts. The second plan, drawn from a ReCom chain, scores about 39.

The population and compactness constraints discussed here, in addition to rook-contiguity, were enforced in the Markov procedure by rejecting non-conforming plans.

2.3.4 Collecting Ensembles of Valid Plans

Our results below are derived from Markov chains using varied proposal types, seed plans, and run lengths. We selected Seed31, Seed99, and the 2011 Enacted plan as initial states for long chains. Those particular two of the 100 seeds were chosen because of their qualitatively different BVAP distributions: the BVAP values for Seed31 vary fairly smoothly from lowest to highest, while Seed99 has one district with a significantly higher BVAP (at 76%) than any of the proposed plans. Both seeds differ significantly from the 2011 Enacted plan. Figure 6 (a) shows the comparison of BVAP.

For each of these three initial plans, we ran the following chains:

- 10 million valid Flip steps;
- 20,000 valid ReCom steps;
- 1 million valid Mix steps

The difference in the target sizes we chose for the ensembles is due to the marked difference in the speed of the individual steps. Computing a ReCom step takes approximately 1000 times longer

[12] This type of discrete compactness metric is suggested by the analysis of Duchin–Tenner (arxiv.org/abs/1808.05860).

than a Flip step, due to the computational cost of creating and splitting a spanning tree during recombination. As we will illustrate below, the Flip method does not generate reliable samples from the space of valid districting plans even after ten million steps, and it shows signs that it may not be able to do so even with an ensemble that is several orders of magnitude larger. By contrast, the ReCom method gives output that is consistent with representative sampling.

2.4 Results

The results of the ensemble analysis offer compelling heuristic evidence that the ReCom and Mix chains converge quickly to a stable sampling distribution. To see why, it is useful to compare them against the Flip ensemble. We will show that runs of as few as ten thousand steps in a ReCom chain produce superior results to 10 million steps using the simple Flip chain.

2.4.1 Independence of Seed

One of the characteristic properties of an ensemble drawn from a Markov chain that has mixed, or approached its stationary distribution, is that the samples should be uncorrelated with their initial state. That is, observing the nth step in a well-mixed chain should not allow us to estimate any properties of the initial state. Equivalently, if we observe the distributions of the statistics in Markov chain ensembles from several different starting points, they should eventually be indistinguishable. This is consistent with the observed behavior of our ReCom and Mix ensembles, but not of the Flip ensemble, for both racial and partisan statistical properties.

The consonance between the observed distributions of ReCom and Mix chains, even though there are 50 times as many plans in the Mix ensemble, suggests that in this setting we are not gaining significant additional information about partisan or racial statistics by intermixing the ReCom steps with Flip steps. Nonetheless, the Mix walk may turn out to have value for studying other statistics. Classically, many MCMC researchers seek to "explore and exploit"—large steps are interspersed with runs of small steps, in order to transit the state space while also capturing local variation. The Mix runs achieve this balance.

BVAP Vectors

We begin by studying the distribution of Black Voting Age Population. The BVAP of the three initial seeds can be observed here, while Figure 7 shows the BVAP ranges over the three full ensembles.

Figure 7 demonstrates that the Flip walk has failed to achieve mixing (i.e., sampling from close to the stationary distribution), because it is clearly drawing samples that retain strong memory of the chain's initial state, even after ten million steps. On the other hand, the ReCom and Mix ensembles have converged to qualitatively identical distributions with ensembles of a small fraction of the size. These Markov chains began far apart (at the same points shown in Figure 6), but the observed distributions have quickly become indistinguishable.

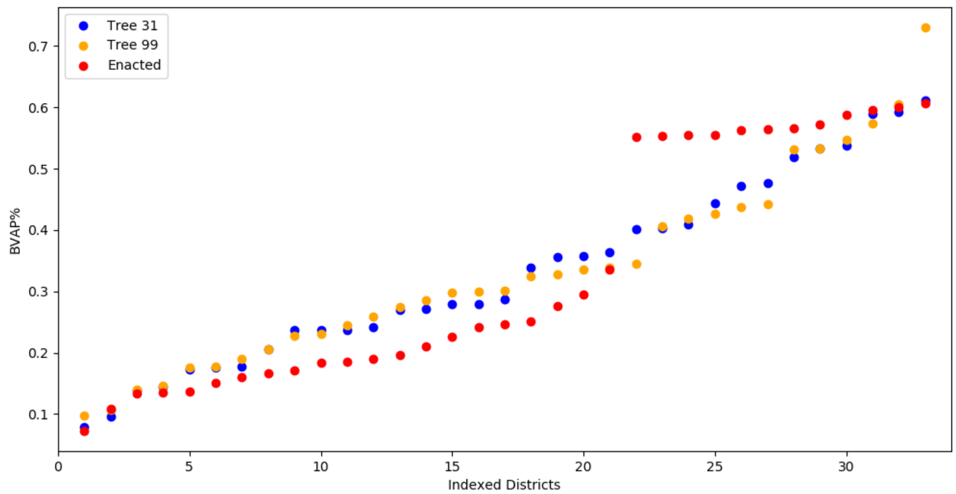


Figure 6 BVAP percentage by district in the three districting plans selected as initial seeds for long chain runs. Recall that the districts are sorted in order of their BVAP.

Mean-Median Gap

Partisan measures evaluated on our ensembles display a similar pattern: the Flip ensemble is heavily determined by the starting point, while the ReCom and Mix ensembles return similar distributions regardless of seed. To illustrate this distinction, we will use one of the most common partisan metrics, the mean-median score, which reports the signed difference between the mean and the median of the set of Republican vote shares across the 33 districts. A positive score suggests a plan that structurally favors Republican outcomes. Political scientists have interpreted this score to measure how far short of a majority of votes the controlling party can fall, while still securing a majority of the seats. We will compute MM scores based on 2016 Presidential votes.[13]

Crucially, there is no reason to believe that plans made without partisan information or intent would tend to have $MM = 0$. The political geography of the state—the locations and varying density of votes—might tilt the MM score toward a nonzero baseline. If we can sample representatively from the space of valid plans, we can understand and control for the effects of political geography.

Figure 8 displays the mean-median distribution for each of our nine ensembles with the same range on the x axis, with zero and the MM value of each initial seed highlighted in the respective plots. The behavior of the Flip distributions has several red flags if the goal is representative sampling. Not only are these distributions different, with very little overlap between the observed values from the three runs but they actually have peaks on opposite sides of zero: the Seed99 Flip ensemble shows a playing field tilted toward Democrats while the Seed31 Flip ensemble is tilted toward Republicans. Neither should be described as providing a baseline for the space of valid plans.

[13] An analysis from Princeton using the same voting data suggests that Clinton would have won 21 of these 33 districts in 2016 (github.com/PrincetonUniversity/VA-gerrymander/). An important caveat follows from the data processing noted above in §2.2.1: vote totals are not reported by census block, so there are real concerns about the accuracy of the process for prorating precinct vote results to the block level without additional cross-validation on the resulting data quality. This section mainly functions as corroborative evidence that our recombination steps converge quickly.

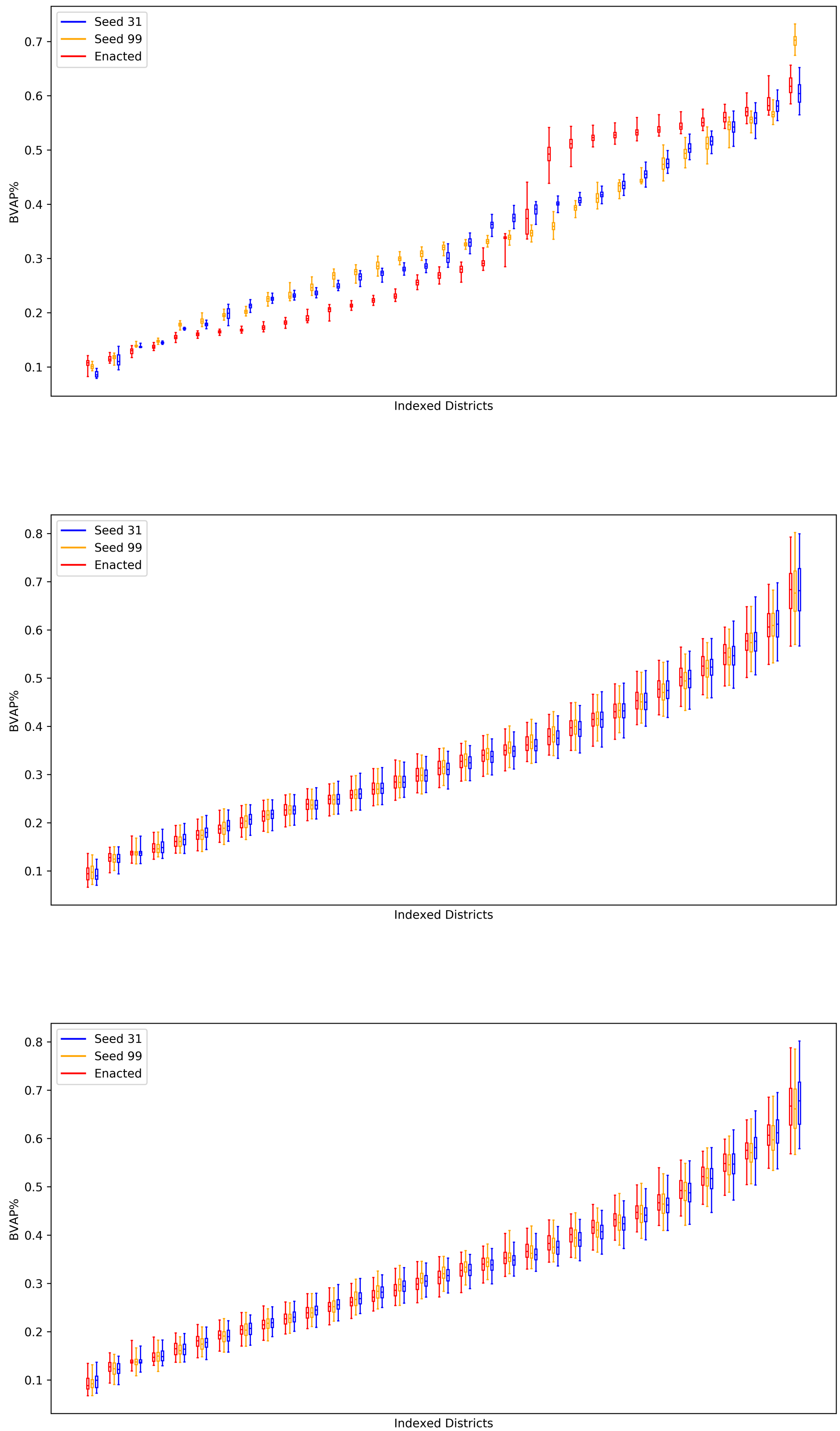
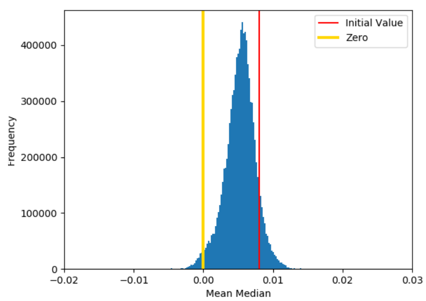
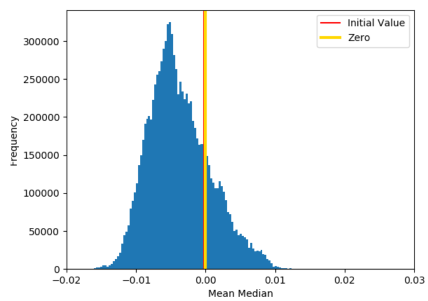


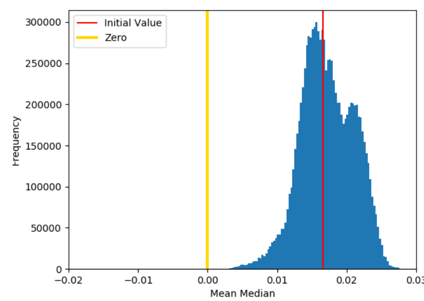
Figure 7 Independence of seed: BVAP by starting point in the Flip ensemble (10 million steps), ReCom ensemble (20,000 steps), and Mix ensemble (1 million steps). The Flip distributions clearly resemble their starting points. By contrast, we observe evidence of convergence in the ReCom and Mix ensembles.



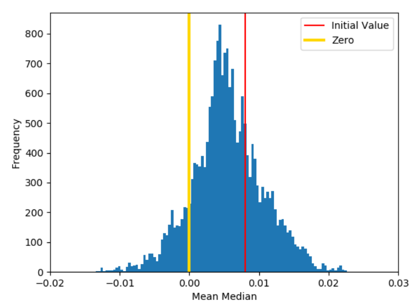
(a) Flip Seed31



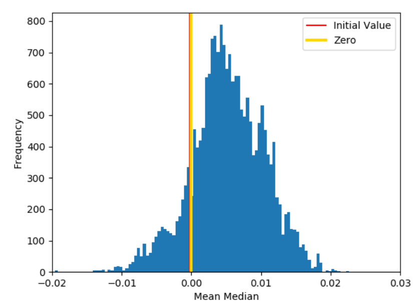
(b) Flip Seed99



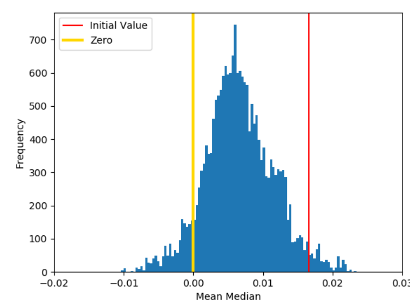
(c) Flip Enacted



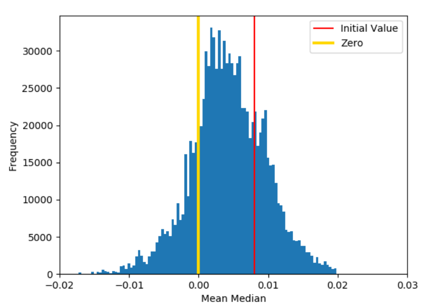
(d) ReCom Seed31



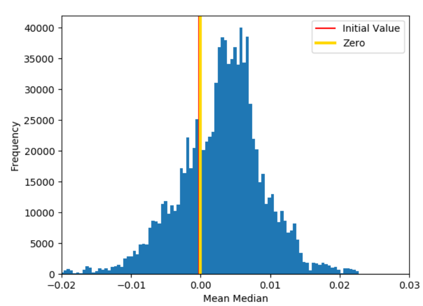
(e) ReCom Seed99



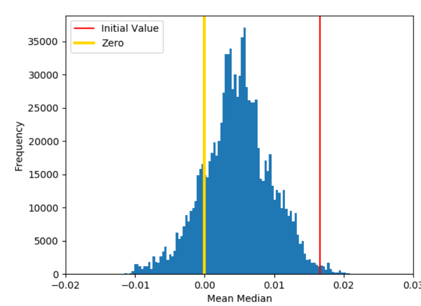
(f) ReCom Enacted



(g) Mix Seed31



(h) Mix Seed99



(i) Mix Enacted

Figure 8 Independence of seed: mean-median distributions by starting point in the Flip ensemble (10 million steps), ReCom ensemble (20,000 steps), and Mix ensemble (1 million steps). The Flip distributions stay clustered around the original value of the plan and have very little overlap, while the ReCom and Mix distributions again appear to have converged.

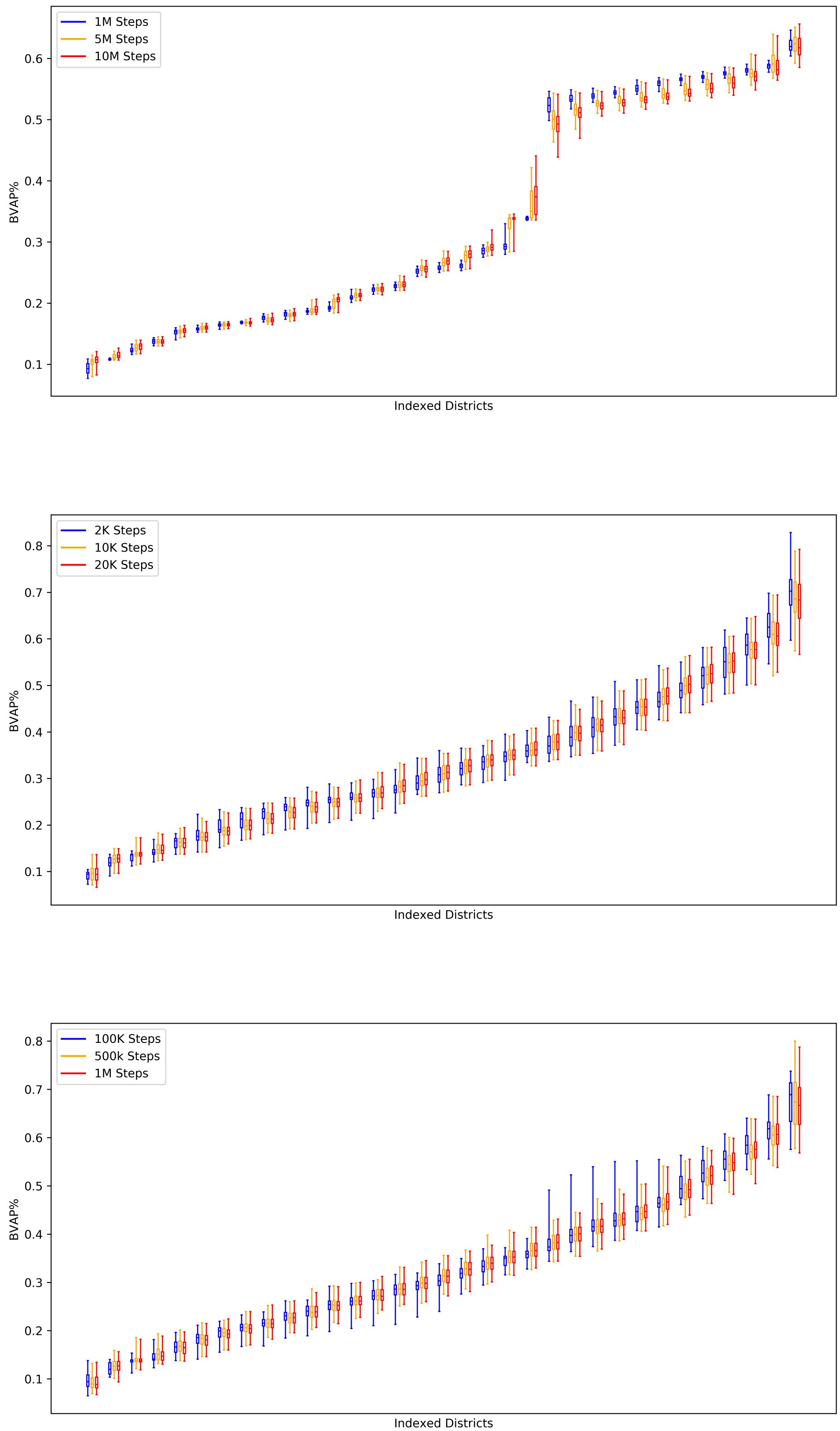


Figure 9 Chain length: BVAP distributions at the 10%, 50%, and complete target lengths for the Flip , ReCom , and Mix ensembles, respectively, based at the 2011 Enacted plan. The Flip ensemble continues to exhibit substantial change, while the ReCom and Mix ensembles have stabilized.

The six histograms showing ReCom and Mix ensembles display encouraging consistency, both between the two proposal variants and across the three initial seeds. They also indicate that a very slight Republican advantage of under one percentage point should be expected from the political geography; it might take 49.5% of the votes for Republicans to secure 50% of the seats in this region.

2.4.2 Chain Length

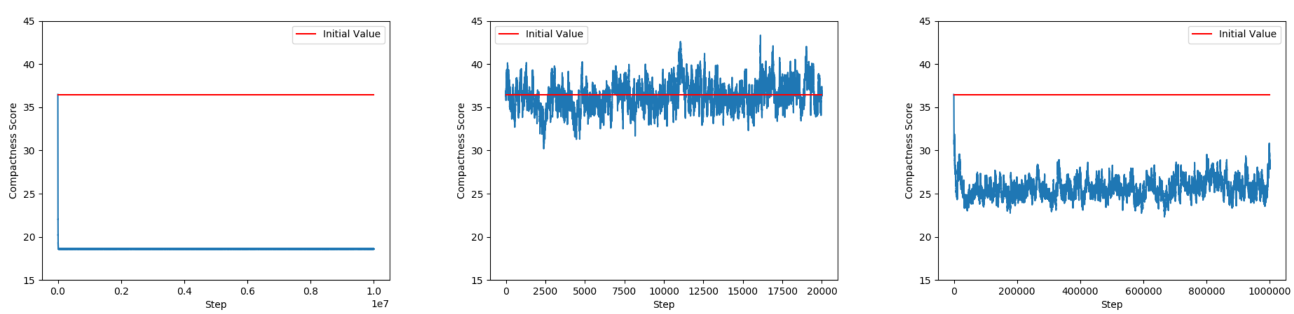
Another characteristic of successfully mixed Markov chains is that the distribution of the statistic being measured should not continue to change significantly as the length of the walk increases. We test this for our chains by comparing the distributions observed over the first 10% and 50% of the run to the full distribution. To date, we have no theoretical guarantees regarding the mixing time of any of these proposal methods, but this adds to the heuristic evidence raising our confidence in effective sampling.

Figure 9 shows the BVAP distributions, beginning at the 2011 Enacted plan and applying each of the three proposals. As in Figure 7, the Flip chain clearly has not converged. On the other hand, the sizes of our ReCom and Mix ensembles seem to be adequate, as there is very little difference between the first 50% of the steps and the full ensembles. In the Mix case, it is interesting to note that even a small proportion of ReCom steps helps to quickly overcome the skewed initial distribution.

2.4.3 Slack in the Constraints

Finally we consider how the different proposals perform with respect to the allowed range of scores that encode the traditional districting principles, highlighting another drawback with the Flip ensembles that is remedied by ReCom steps. We will illustrate this point using compactness; recall from §2.3.3 that we measure the compactness by the ratio of total edges to cut edges in the dual graph of the affected region.

We find that the Flip runs immediately use up any amount of slack that is permitted, and then rarely return to even slightly better scores. This is strikingly illustrated in Figure 10 for runs beginning at the 2011 Enacted plan. By contrast, the ReCom ensemble has compactness scores that are in range of, and often better than, the initial value.



(a) Flip (b) ReCom (c) Mix

Figure 10 Trace plots of compactness for plans encountered in chains initialized at the Enacted plan. The initial score is marked in red. The Flip walk immediately proceeds to the lower bound while the other proposals generate more compact plans.

The compactness scores of the Mix ensemble are intermediate, dragged down by Flip steps and restored to better values by ReCom steps. One consequence of these findings is that, over the scale considered here, all three ensembles are necessarily drawn from very different distributions. It is interesting that the observed differences in compactness between the ReCom and Mix ensembles do not cause a corresponding difference in the racial or the partisan properties of the plans, as shown in Figures 7-8. This suggests that constraining compactness, in itself, may be fairly ineffective at remedying racial or partisan gerrymandering.

These findings also give us grounds to be skeptical that the Flip walk will improve substantially in the convergence of partisan and racial statistics if run for ten, 100, or even 1000 times its current length. Because its plans are so close to the worst allowable compactness score, most new proposed steps are rejected in the Markov process. To achieve a significant qualitative change, the chain would likely have to randomly select a long sequence of lucky flips to get more slack in this bound. If one district becomes very eccentric, it impedes the chain from selecting any substantial changes to the other districts. Ergodicity ensures that representative sampling will eventually occur, but it might take more time than is practicable, possibly even more than the full 10-year census cycle.

2.5 Analysis and Conclusions

Having established a basis for confidence in this sampling method, we turn to analysis of the competing plans.

First, we address one additional possible complaint: our neutral ensembles do find a significant number of plans with extremely high BVAP in one or two districts, and one might wonder if this skews the findings.[14] To address this, we also present winnowed versions of our main ensembles, excluding plans in which any district has BVAP over 60%. This makes no material changes to our findings.

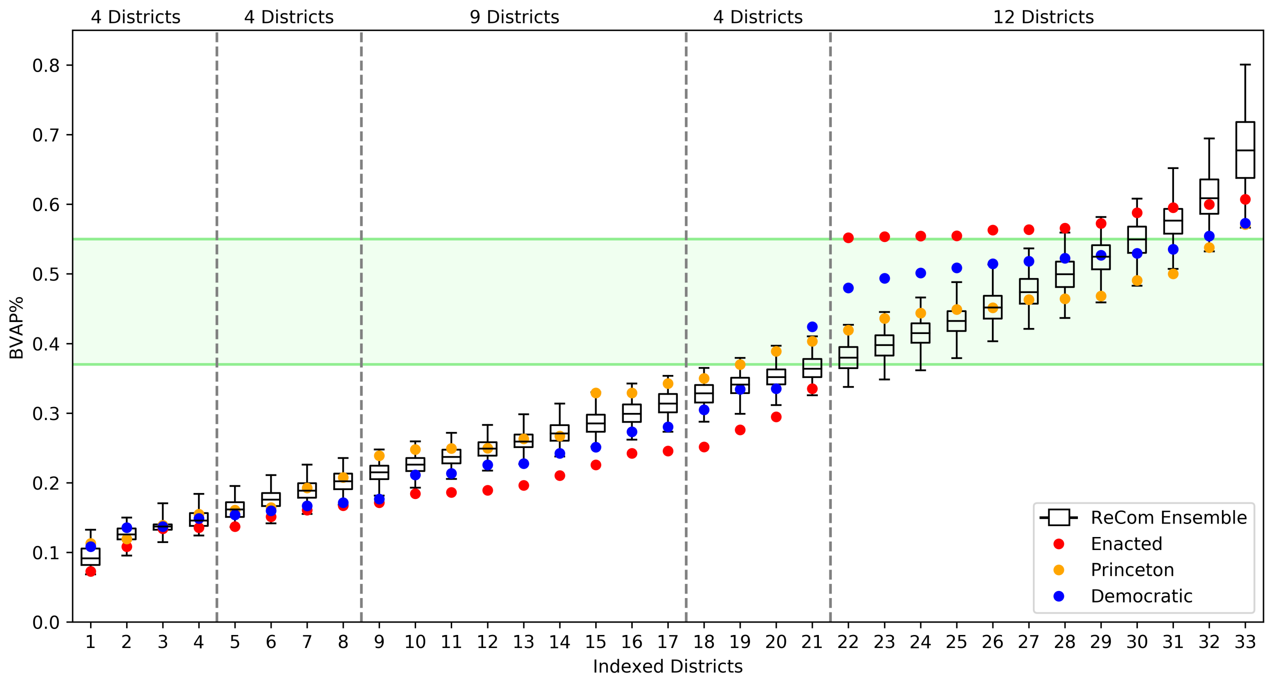
With the ReCom and Mix ensembles and their winnowed variants, we have batches of alternative plans produced by a process that passes numerous tests of quality. With these plans as a baseline for comparison, we are ready to address the motivating questions about the effects of packing in the top 12 indexed districts on the performance of the rest of the plan.

We see that the 2011 Enacted plan has elevated BVAP the top 12 districts at the clear expense of the next four, the following nine, and even the four after that. A substantial portion of the random ensembles is made up of plans with three to four more districts that would merit RPV analysis as potential electable voting rights districts, and nine after that with potential to perform as coalition districts. By contrast, the 2011 Enacted plan exhibits clearly depressed opportunity in all of those, which agrees with the findings of the District Court and the plaintiffs' experts.

The indications of vote dilution are only partially addressed in most of the newly proposed plans. Both the Democratic Caucus plan and the whole range of new Republican plans elevate only one more district above 37% BVAP, while the Princeton plan—and hundreds of thousands of plans that were found by our Markov chains—have three more, without sacrificing traditional districting principles. Winnowing the ensembles to BVAP$\leq$ 60% only strengthens the findings, showing the depressive effect of packing the top districts to be even greater than before.

[14] Figures 11-12 show that some plans in the ensembles can have one or even two districts with extremely high BVAP, even as high as 80%.

In summary, this report contains varied and robust evidence that Markov chains using recombination steps sample effectively and efficiently from the space of valid House districting plans for the region of Virginia affected by the recent court ruling. We emphasize that these ensemble methods should not be used to select a plan for enactment because they are made without local and community-based considerations. Instead, ensemble methods give an effective means of verifying whether a newly proposed plan is an extreme outlier in the universe of valid plans. In concert with other techniques, these methods can give effective means of demonstrating either racial or partisan gerrymandering.



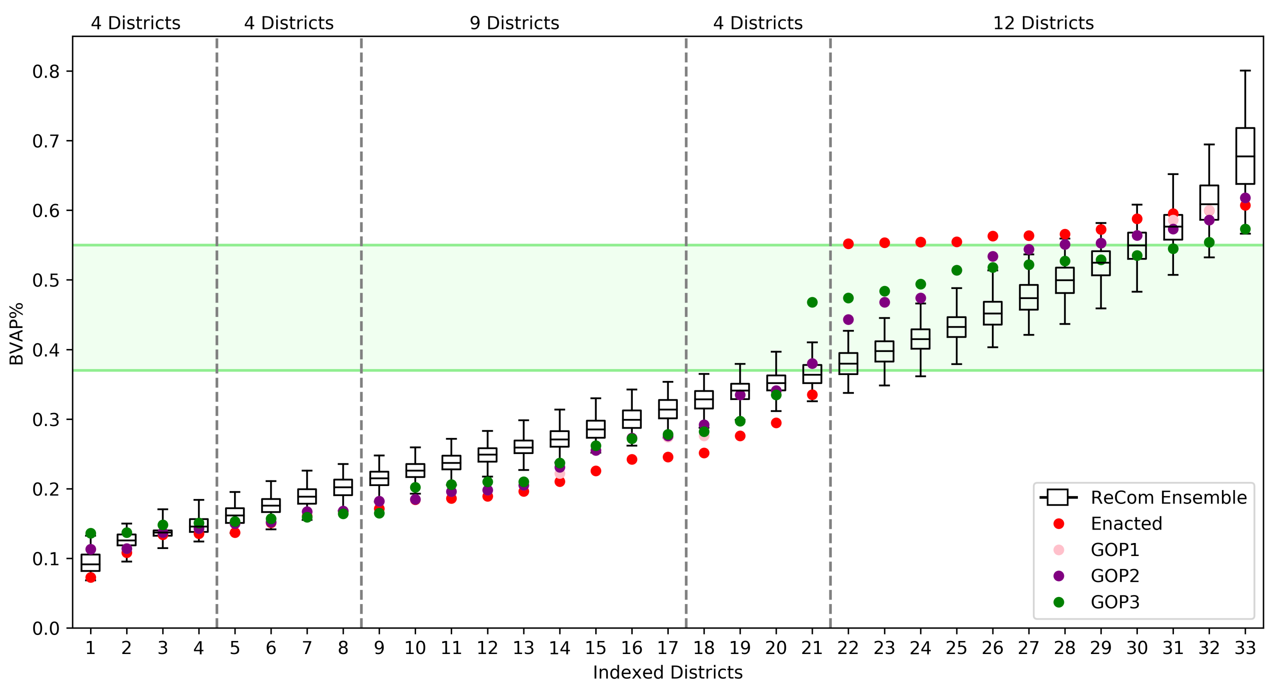
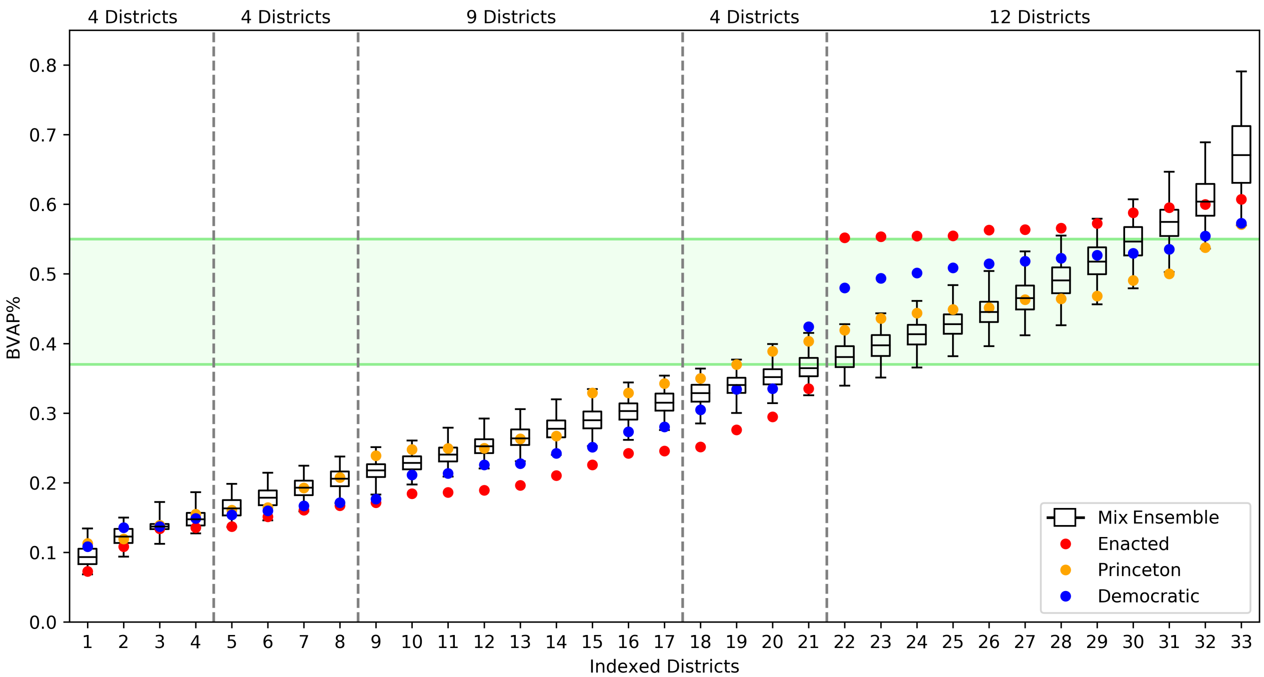


Figure 11 Summary: comparison of the BVAP from the six proposed plans to the full ReCom ensemble. The 37-55% BVAP zone is marked in green, and the districts are separated into groups for ease of interpretation.



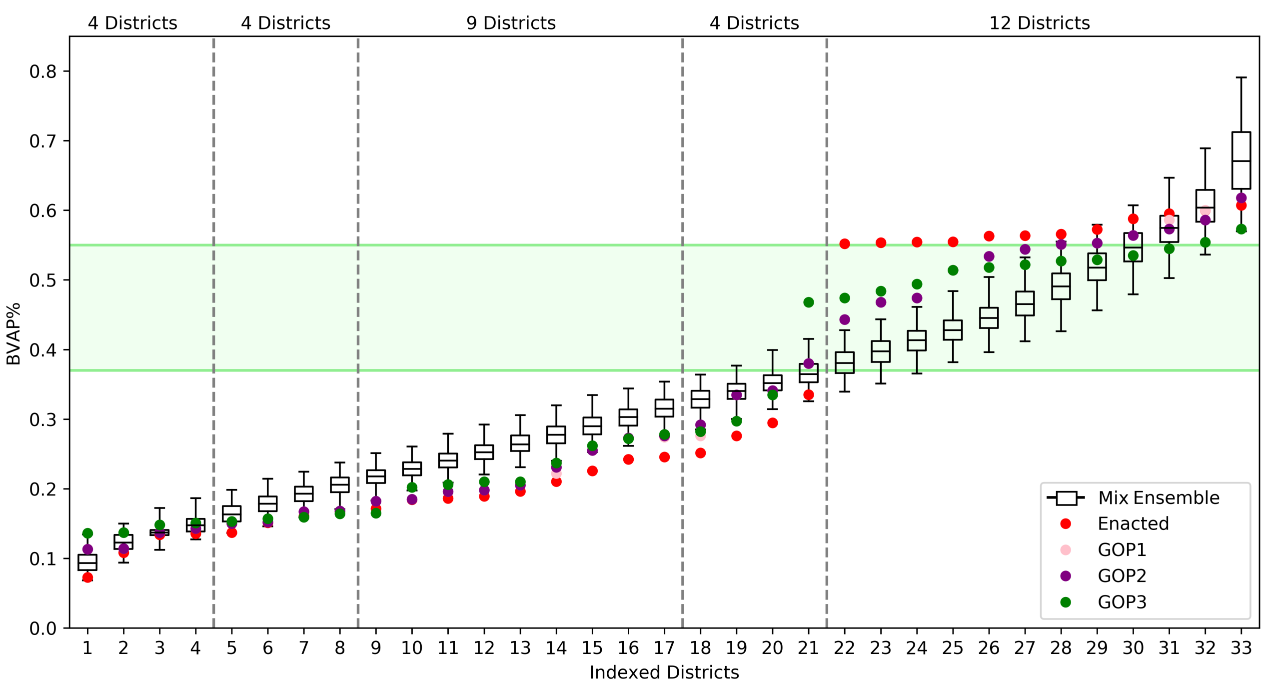
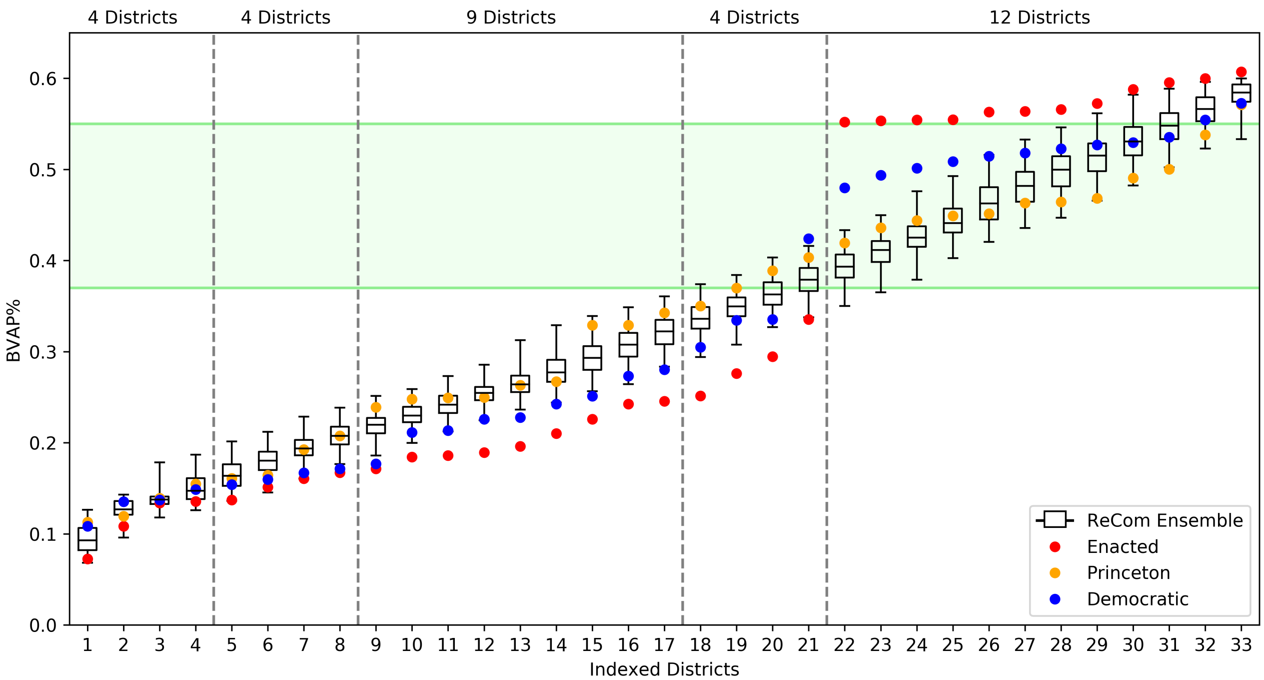


Figure 12 Summary: comparison of the BVAP from the six proposed plans to the full Mix ensemble. The 37-55% BVAP zone is marked in green, and the districts are separated into groups for ease of interpretation.



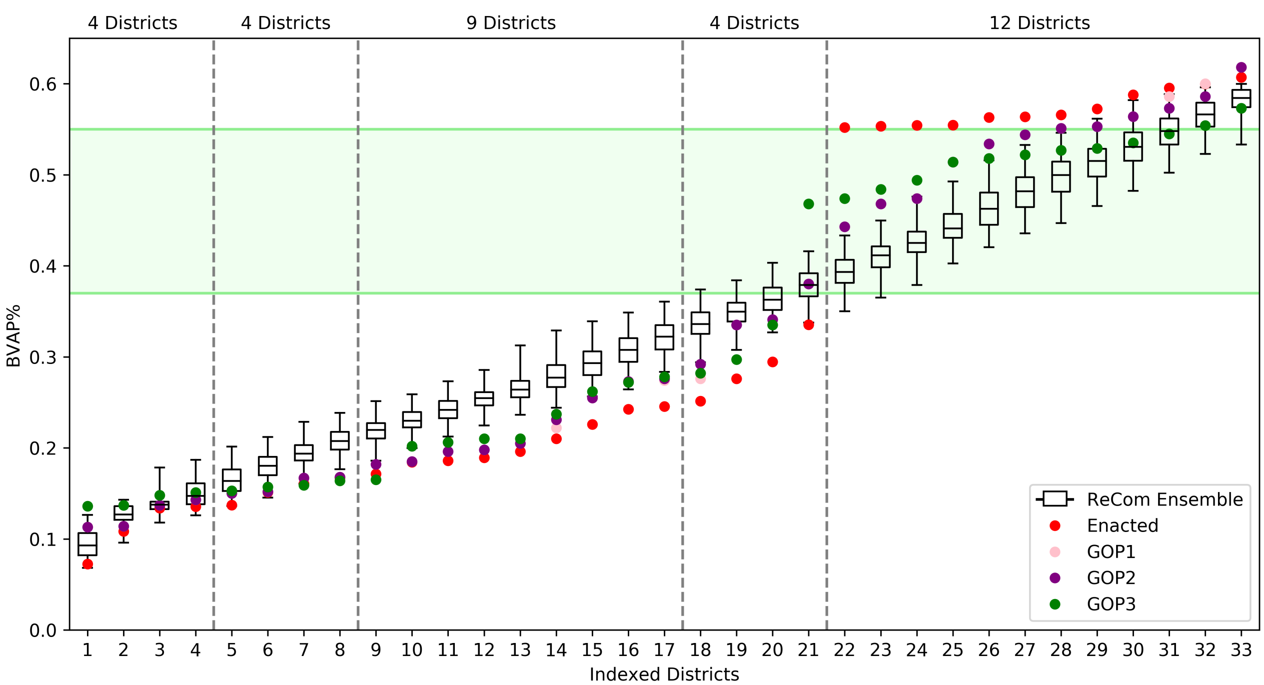
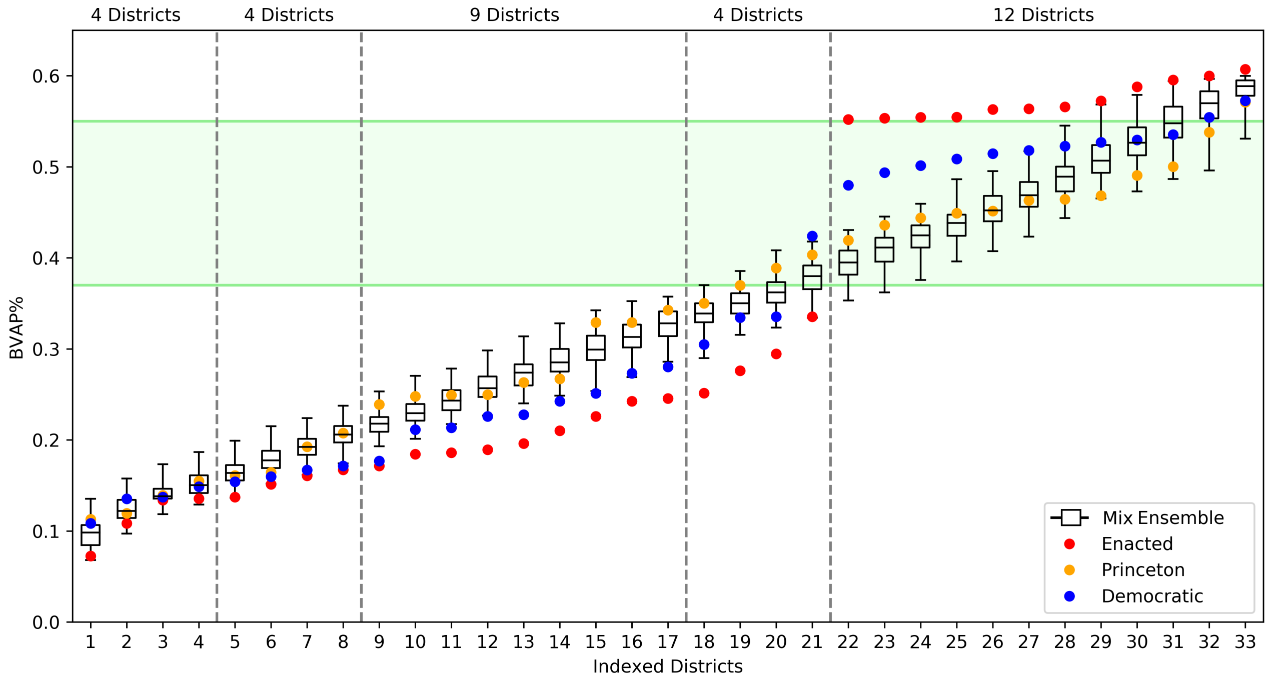


Figure 13 Summary: comparison of the BVAP from the six proposed plans to the winnowed ReCom ensemble (BVAP$\leq$ 60%). The 37-55% BVAP zone is marked in green, and the districts are separated into groups for ease of interpretation.



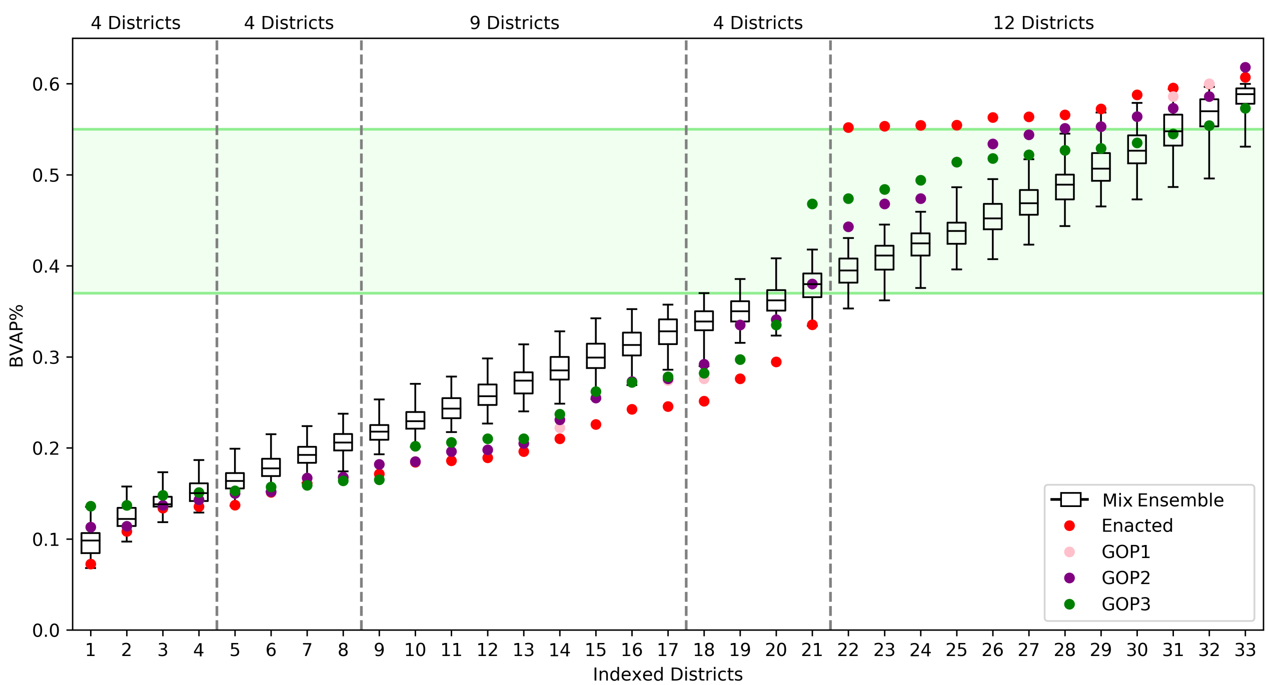


Figure 14 Summary: comparison of the BVAP from the six proposed plans to the winnowed Mix ensemble (BVAP$\leq$ 60%). The 37-55% BVAP zone is marked in green, and the districts are separated into groups for ease of interpretation.