Competitiveness

Article 4 Part 2 Section 1 (14)(F)

"To the extent practicable, competitive districts should be favored where to do so would create no significant detriment to the other goals."

IRC 1.0 Measures

1. Party Registration
2. Arizona Quick & Dirty (AQD)
3. JudgeIt
4. Statewide Election Averages

IRC 1.0 Measures in detail

1. Party Registration
   - Democratic / Republic difference within 7%
   - 3rd-party registration greater than the Dem / Rep difference
2. Arizona Quick & Dirty (AQD)
   - Average of the 2018 and 2020 votes for Corporation Commission
3. JudgeIt
   - Statistical analysis of partisan registration, the vote in the immediately preceding election, and incumbency
   - Closely related to PlanScore's "King/Grofman Partisan Bias" measure
4. Statewide Election Averages
   - 3-election and 4-election averages using 1998 statewide offices

IRC 1.0 Competitive Range

- Competitive range / definition
  - JudgeIt: Statistical Standard Error of +/- 3.5% considered "competitive"
  - That 7% range was used as "competitive" definition for all measures
  - IRC 1.0 discussed, but did not formally count, "bulletproof" districts

IRC 1.0 Discussion

"[Corporation Commission results] were used because they are less likely to be the kinds of races where you have a lot of other variables that come into play, so it would tend to give you an idea of how people would usually vote with respect to Republican and Democratic candidates."

IRC 1.0 Methodology

- AQD and Voter Registration counts calculated "live" within the mapping software
- Later, 1998 election result averages were also "live"
- Maps sent off for JudgeIt analysis
  - Took 4 hours when planned, about a day when unplanned

IRC 2.0 Measures

- Eight averages of 2004 – 2010 statewide election results

IRC 2.0 Measures in detail

- Eight averages of statewide election results:
  2. Average results from 2008 + 2010, each year weighted equally
  3. Average result from 2008 + 2010 plus major party registration
  4. Average result from 2004 + 2006 + 2008 + 2010
  5. Average result from 2004 + 2006+ 2008+ 2010 plus major party registration
  6. Average result from 2004 + 2006+ 2008+2010 plus major party registration; but races where one candidate received more than 60% of the 2-way vote removed
  7. Average result from 2004 + 2006+ 2008+2010 plus major party registration; but races where one candidate received more than 60% of the total vote removed
  8. Weighted average result from 2010 (1/3 weight) + 2008 (1/3 weight) + 2004 (1/6 weight) + 2006 (1/6 weight); races where one candidate received more than 60% of the 2-way vote removed
  9. Weighted average result from 2010 (1/4 weight) + 2008 (1/4 weight) + 2006 (1/8 weight) + major party registration (1/4 weight); races where one candidate received more than 60% of the 2-way vote removed.

IRC 2.0 Range

- We are not aware of any adopted "competitive range" or other "competitive" definition adopted by IRC 2.0

IRC 2.0 Methodology

- Results calculated "live" within the mapping system for all eight measures.

2021: Simple Options

A. Voter Registration

B. Average of a selection of Statewide Election Results

C. Count/Mix of Democratic and Republican election victories in Selected Elections

More Complicated Options

A. Statistical Calculations based on past election data
   A. Seats-Votes Bias / Partisan Swing
   B. Responsiveness / Swing analysis
   C. JudgeIt a.k.a. King/Grofman
   D. Declination
   E. Efficiency Gap
   F. Mean-Median Difference

"Thousands of Maps" Comparisons

- Markov-Chain based programming approach

1. Algorithm attempting to capture other criteria written
2. Thousands of maps generated
3. A competitiveness score based on one measure calculated
4. Resulting number of competitive districts calculated
5. Graph developed showing the range of results
6. Maps under consideration by the Commission compared to that range of results

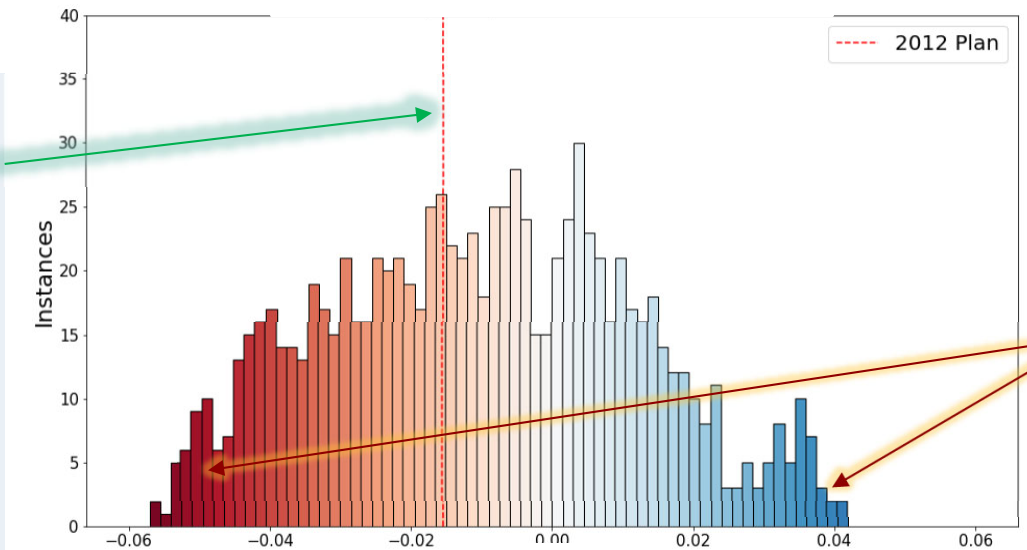
Dr. Duchin Wisconsin presentation, incl. how Ensemble maps are generated: https://youtu.be/ZtwQ1l_WbMU

"Thousands of Maps" Comparisons

- Appealing, but not a complete solution
  - Focusing on the "best" map is letting the algorithm decide
  - Any/all map(s) in a central range are generally considered equally good

Generate this value for each map and compare to some version of the pre-generated chart.

A map in the middle is considered relatively competitive.



A map at either extreme is considered an outlier (in a negative way).

Chart from "Assessing Congressional Districting In Maine And New Hampshire," by Sara Asgari, Quinn Basewitz, Ethan Bergmann, Jackson Brogsol,Nathaniel Cox, Diana Davis, Martina Kampel, Becca Keating,Katie Knox, Angus Lam, Jorge Lopez-Nava, Jennifer Paige,Nathan Pitock, Victoria Song, Dylan Torrance, available online at https://arxiv.org/pdf/2011.06555.pdf

Table of Partisan Statewide Elections

	2020	2018	2016	2014	2012
President	X		X		X
Senate1		X			X
Senate2	X		X		
Governor		X		X	
Secretary of State		X		X	
Attorney General		X		X	
State Treasurer		X		Uncontested	
Superintendent of Public Education		X		X	
State Mine Inspector		X		Uncontested	
Corporation Commissioner	3 seats	2 seats	3 seats	2 seats	3 seats

Partisan Statewide Election Results

	2020Dem	2020Rep	2018Dem	2018Rep	2016Dem	2016Rep	2014Dem	2014Rep	2012Dem	2012Rep
President	1,672,143	1,661,686			1,161,167	1,252,401			1,025,232	1,233,654
Senate1			1,191,100	1,135,200					1,036,542	1,104,457
Senate2	1,716,467	1,637,661			1,031,245	1,359,267				
Governor			994,341	1,330,863			626,921	805,062		
Secretary of State			1,176,384	1,156,532			712,918	779,226		
Attorney General			1,120,726	1,201,398			696,054	782,361		
State Treasurer			1,052,197	1,249,120			Uncontested			
Superintendent of Public Education			1,185,457	1,113,781			724,239	740,273		
State Mine Inspector			1,090,346	1,168,798			Uncontested			
Corporation Commissioner	1,450,194	1,449,963	1,076,800	1,053,862	1,024,501	1,208,002	576,482	766,864	868,726	979,034
		1,434,236	1,006,654	1,049,394	988,666	1,122,849	557,963	761,915	776,472	943,157
		1,379,804				1,061,094			862,876	935,573

	2020Dem	2020Rep	2018Dem	2018Rep	2016Dem	2016Rep	2014Dem	2014Rep	2012Dem	2012Rep
President	50.2%	49.8%			48.1%	51.9%			45.4%	54.6%
Senate1			51.2%	48.8%					48.4%	51.6%
Senate2	51.2%	48.8%			43.1%	56.9%				
Governor			42.8%	57.2%			43.8%	56.2%		
Secretary of State			50.4%	49.6%			47.8%	52.2%		
Attorney General			48.3%	51.7%			47.1%	52.9%		
State Treasurer			45.7%	54.3%						
Superintendent of Public Education			51.6%	48.4%			49.5%	50.5%		
State Mine Inspector			48.3%	51.7%						
Corporation Commissioner	50.0%	50.0%	50.5%	49.5%	45.9%	54.1%	42.9%	57.1%	47.0%	53.0%
(top R & top D Corp. Com. candidates only)										

Data hand-entered for this presentation as the official database is being compiled.
Some numbers may have been misread or transposed.

Disclaimer

The following slides covering the information provided by Drs. McGhee, Duchin and Wang are NDC summaries.

For their exact comments, please see the videos of their presentations and the letters and presentations they provided.

General Summary: Dr. McGhee

- Agrees with the importance of explainability.
- Overall, the Presidential vote in a district is "far and away" the best predictor of Congressional District results.
- In these days of partisan extremes, using just about any statewide election is likely to generate similar results.

General Summary: Dr. McGhee

- Worst measure: party registration
- Better: averages of past election results
- Best: statistical regression analyzing the accuracy of past statewide election result predictions by-district results

Measures Summary: Dr. McGhee I

- PlanScore
  - Generates three Statistical Calculations:
    1. Efficiency Gap
    2. Partisan Bias (King/Grofman measure)
    3. Mean-Median Gap
  - Heavily, but not entirely, based on the previous Presidential election results.
  - As do all statistical analysis based on previous congressional or legislative election results, relies on some "fill" data or other adjustments for uncontested elections and other mathematical challenges.

Measures Summary: Dr. McGhee II

Proposed an alternative approach:

A. Generate a Competitive Range of "reasonable swing"

1. Calculate the actual average U.S. house, state senate, and state house vote share in each of the last five election years (3 averages for each year).
2. Calculate the overall average **across** the last 5 elections for U.S. house, state senate, and state house (3 averages total).
3. Calculate the differences between #1 and #2 for each chamber.
4. Identify the largest absolute difference (i.e., either positive or negative) across all chambers and years.
   - The Legislative or Congressional body with the largest "swing" (difference between single-year and average-over-five-year vote share) is the largest "reasonably expected swing," a.k.a. competitive margin.

See also the August 5th letter from Dr. McGhee for more details.

Measures Summary: Dr. McGhee II

B. Calculate the # of competitive districts in each map:

1. Run the map through PlanScore; this will give you a prediction for each seat in the plan in an average election from the last decade.
2. For each district in the map, subtract 50% from PlanScore's predicted vote share.
3. Any seat where the absolute result of this calculation is smaller than the "reasonably expected swing" is a competitive seat.

NDC Comment: This "averages over 5 election cycles" method does not capture the Democratic and Turnout swing in the state from 2012 to 2020.

Add'l General Comments: Dr. McGhee

- "I would caution against using any statewide office with larger partisan swings that you get from A.1 – A.4 above. That's a sign that the office doesn't have strong down-ballot coattails. (Gubernatorial races are especially notable for this)."
  - *The next slide shows **possible** examples circled by NDC on the table of past election results. NDC has not yet calculated the "partisan swings" Dr. McGhee references to determine if the circled elections are actually outside of that swing range.*

Partisan Statewide Election Results

	2020Dem	2020Rep	2018Dem	2018Rep	2016Dem	2016Rep	2014Dem	2014Rep	2012Dem	2012Rep
President	1,672,143	1,661,686			1,161,167	1,252,401			1,025,232	1,233,654
Senate1			1,191,100	1,135,200					1,036,542	1,104,457
Senate2	1,716,467	1,637,661			1,031,245	1,359,267				
Governor			994,341	1,330,863			626,921	805,062		
Secretary of State			1,176,384	1,156,532			712,918	779,226		
Attorney General			1,120,726	1,201,398			696,054	782,361		
State Treasurer			1,052,197	1,249,120			Uncontested			
Superintendent of Public Education			1,185,457	1,113,781			724,239	740,273		
State Mine Inspector			1,090,346	1,168,798			Uncontested			
Corporation Commissioner	1,450,194	1,449,963	1,076,800	1,053,862	1,024,501	1,208,002	576,482	766,864	868,726	979,034
		1,434,236	1,006,654	1,049,394	988,666	1,122,849	557,963	761,915	776,472	943,157
		1,379,804				1,061,094			862,876	935,573

	2020Dem	2020Rep	2018Dem	2018Rep	2016Dem	2016Rep	2014Dem	2014Rep	2012Dem	2012Rep
President	50.2%	49.8%			48.1%	51.9%			45.4%	54.6%
Senate1			51.2%	48.8%					48.4%	51.6%
Senate2	51.2%	48.8%			43.1%	56.9%				
Governor			42.8%	57.2%			43.8%	56.2%		
Secretary of State			50.4%	49.6%			47.8%	52.2%		
Attorney General			48.3%	51.7%			47.1%	52.9%		
State Treasurer			45.7%	54.3%						
Superintendent of Public Education			51.6%	48.4%			49.5%	50.5%		
State Mine Inspector			48.3%	51.7%						
Corporation Commissioner	50.0%	50.0%	50.5%	49.5%	45.9%	54.1%	42.9%	57.1%	47.0%	53.0%
(top R & top D Corp. Com. candidates only)										

Data hand-entered for this presentation as the official database is being compiled.
Some numbers may have been misread or transposed.

General Summary: Dr. Duchin

- Much harder to predict future election results than to analyze current or past election results.
- Using a selection of real-world election results is clearer than uniform partisan swing and related analysis.
- Easy to understand a count of the # of districts that changed party preference.

General Summary: Dr. Duchin

- An algorithmic analysis of thousands of plans could flag outliers and give a sense of what is possible, but should not pick the specific plan.

- Many statistical measures lack clarity for AZ's goal or have other issues.

- The geography of where voters live can lead to districts are naturally "safe." A "swing" or "competitive" district is won by each party in different elections.

General Summary: Dr. Duchin

- In many states, Senate and/or Presidential elections are a decent predictor of future election results, but other statewide elections reflect a district's responsiveness to shifts in voter opinion.
  - NDC's rephrasing: Are you looking to correctly predict the right winner as often as possible, or to correctly predict seats where both parties have a chance to win?

- Don't worry too much about personality and other campaign-specific factors; if you are taking a large enough selection of elections, these will wash out overall.

Measures Summary: Dr. Duchin

- Party registration is not a good measure.

- One Possibility: "Swing Districts" metric:
  - Look at the actual vote counts in each individual statewide single-seat election of the last 10 years. Count the number of districts that include a party winner "flip."
  - Districts where each party won at least 1 statewide election are "swing districts," and a map with more "swing districts" is a "more competitive" map.

- Second, More Traditional, Possibility: "Vote Band" metric:
  - Use an individual election or select a "vote index" (combination of elections), then calculate an average of a set of statewide single-seat elections over the last 6 or 10 years.
  - Count a district as "competitive" if the vote is 53 – 47% or closer.
    - Could alternatively use 55 – 45% range, or any range you choose.

General Summary: Dr. Wang

- Voter behavior is easier to predict now than historically.

- There is a solid consensus on how to look at how competitive maps were in the past. It is much harder to predict how competitive the districts will be in the future.

- Last decade of Arizona elections gives lots of swing elections for useful analysis.

See also the August 5th letter from Dr. Wang for more details.

General Summary: Dr. Wang

- More intensive statistical analysis may be more appropriate for academic and retroactive analysis than during mapping.

- Be cautious about assumptions of statistical models and wary of the difficulty of explaining their complexity.

- FiveThirtyEight.com model gives a sense of competitive possibilities.
  - *NDC Note: MGGG, Dave's Redistricting, and other sites have similar 'test / demo / evaluate' capabilities.*

Measures Summary: Dr. Wang

- Use a "market basket" of statewide partisan elections from the past ten years (not by-district elections).

Use either or both of the following:

1. "Average Performance" metric
   - Identical to Dr. Duchin's "Vote Band" metric
   - But Dr. Wang typically uses a 46.5% to 53.5% "competitive" band

2. "Responsive Districts" metric
   - Essentially identical to Dr. Duchin's "Swing Districts" metric

Goals for Today's Meeting

A. Select the method(s) you wish to use.

B. Select the election(s) to be include in those calculation.

C. Select whether to establish ranges / rankings and/or competitiveness "points."

Some particulars of each selection may be left for future mathematical calculation.

Decision A: Methods I

Described By	Description	Type	Competitive Range	Understanding	Implementation
McGhee	PlanScore + "Reasonable Swing"	Statistical	Defined by "Reasonable Swing"	Complicated	Short Delay or Send for Quick Review
Duchin	"Swing Districts"	Actual Past Election Results	# of Districts that "swing"	Simple	Live
Wang	"Responsive Districts"				
Duchin	"Vote Band"	Actual Past Election Results	53 – 47%	Simple	Live
Wang	"Average Performance"		53.5 – 46.5%		
IRC 2.0	Simple Average		Undefined?		
IRC 1.0	Down-Ballot Election Average (i.e. AQD)	Averaged Past Election Results	IRC 1.0 used 53.5 – 46.5%	Simple	Live

Decision A: Methods II

Source	Description	Type	Competitive Range	Under-standing	Implementation
Literature	Presidential Election Result	2020 or a weighted or unweighted average of '20, '16 & '12	To be determined	Simple	Live
Literature	Partisan Swing / Symmetry *	Statistical	Standard Error	Complicated	Send for Quick Review
Literature	Mean-Median Difference *	Statistical	Standard Error	Complicated	Live
Literature	Proportionality *	Averaged Election Results	Standard Error	Medium	Generate Report
Literature	Declination	Actual Past Election Results	Unclear	Complicated	Send for review
Literature	Efficiency Gap *	Statistical	Complicated	Complicated	Send for Quick Review
Duchin	Extreme Outlier analysis	Thousands of Plans	Complicated	Complicated	Initial Setup, then Generate Report

* NDC Note: the focus of this tool is measuring fairness / partisan gerrymandering, not district competitiveness.

Decision B: Elections

- Which past statewide elections should be included in the analysis?
  - All 22? Just down-ballot? Just top-of-ballot? Include Corp. Commission?
  - By Year: just those from 2018 – 2020? From 2016 – 2020?
  - Exclude: 56%+ "outliers"? Governor? Corporation Commission?
  - Some other selection?

	2020Dem	2020Rep	2018Dem	2018Rep	2016Dem	2016Rep	2014Dem	2014Rep	2012Dem	2012Rep
President	50.2%	49.8%			48.1%	51.9%			45.4%	54.6%
Senate1			51.2%	48.8%					48.4%	51.6%
Senate2	51.2%	48.8%			43.1%	56.9%				
Governor			42.8%	57.2%			43.8%	56.2%		
Secretary of State			50.4%	49.6%			47.8%	52.2%		
Attorney General			48.3%	51.7%			47.1%	52.9%		
State Treasurer			45.7%	54.3%						
Superintendent of Public Education			51.6%	48.4%			49.5%	50.5%		
State Mine Inspector			48.3%	51.7%						
Corporation Commissioner	50.0%	50.0%	50.5%	49.5%	45.9%	54.1%	42.9%	57.1%	47.0%	53.0%
(top R & top D Corp. Com. candidates only)										

Decision C: Range/Rankings

- Are districts either "competitive" or "not competitive"
  - 53 – 47%? 53.5 – 46.5%? 55 – 45%? Other?
- Or is there a range such as "highly competitive," "somewhat competitive" and "bulletproof" districts?
- Define the range(s)
- Should "points" be awarded by category, or decisions evaluated plan by plan?

Recommendation

1. Select the method(s) you wish to use.

2. Select the election(s) to be include in those calculation.

3. Select whether to establish ranges / rankings and/or "points."

Samples

1. Select the method(s) you wish to use.
   - i.e. Plan Score, Swing Districts and Simple Average?
2. Select the election(s) to be include in those calculation.
   - i.e. all 10 years; 2016-2020 elections; some other limitation?
3. Select whether to establish ranges.
   - i.e. Standard Deviation for Plan Score or past elections average;
   - Some count of district "swings";
   - A 53 – 47% or some other range for averages.
4. Select whether to establish rankings or "points."
   - i.e. no "points," or something like 4 points for "highly" competitive, 2 points for "somewhat" and -1 for "bulletproof."

Final Thoughts

- Direction needed for proper setup of the redistricting software.
  - "API" connection to PlanScore or similar systems are complicated.
  - "Send-off for analysis" options require setting up such systems.

- There is some flexibility going forward: this decision is guidance and direction, but somewhat open to future revision.

arXiv:2005.12731v1 [cs.CY] 26 May 2020

A Computational Approach to Measuring Vote Elasticity and Competitiveness

Daryl DeFord*
CSAIL, Massachusetts Institute of Technology
Moon Duchin
Department of Mathematics, Tufts University
Justin Solomon
CSAIL, Massachusetts Institute of Technology

May 27, 2020

Abstract

The recent wave of attention to partisan gerrymandering has come with a push to refine or replace the laws that govern political redistricting around the country. A common element in several states' reform efforts has been the inclusion of *competitiveness metrics*, or scores that evaluate a districting plan based on the extent to which district-level outcomes are in play or are likely to be closely contested.

In this paper, we examine several classes of competitiveness metrics motivated by recent reform proposals and then evaluate their potential outcomes across large ensembles of districting plans at the Congressional and state Senate levels. This is part of a growing literature using MCMC techniques from applied statistics to situate plans and criteria in the context of valid redistricting alternatives. Our empirical analysis focuses on five states—Utah, Georgia, Wisconsin, Virginia, and Massachusetts—chosen to represent a range of partisan attributes. We highlight situation-specific difficulties in creating good competitiveness metrics and show that optimizing competitiveness can produce unintended consequences on other partisan metrics. These results demonstrate the importance of (1) avoiding writing detailed metric constraints into long-lasting constitutional reform and (2) carrying out careful mathematical modeling on real geo-electoral data in each redistricting cycle.

Keywords: Redistricting, Gerrymandering, Markov Chains, Competitiveness

*The authors gratefully acknowledge the generous support of the Prof. Amar G. Bose Research Grant and the Jonathan M. Tisch College of Civic Life.

Contents

1 Introduction **3**
- 1.1 Contributions . . . 5
- 1.2 Related work . . . 6
- 1.3 Review of competitiveness language in legal criteria . . . 7
  - 1.3.1 Washington, New York, Arizona . . . 7
  - 1.3.2 Colorado . . . 7
  - 1.3.3 Missouri . . . 8
  - 1.3.4 New Jersey . . . 8

2 Competitiveness metrics **9**
- 2.1 Plans and votes . . . 9
- 2.2 Evenness and typicality . . . 10
- 2.3 Vote-band metrics . . . 11
- 2.4 EG/swing zeroing . . . 15

3 Data and methods for ensemble analysis **17**
- 3.1 State data . . . 17
- 3.2 Partisan metrics . . . 18
- 3.3 Ensemble generation . . . 19

4 Ensemble results **20**
- 4.1 Partisan baselines . . . 21
- 4.2 Satisfiable vote-band rules . . . 23
- 4.3 Partisan impacts of vote-band winnowing . . . 24
- 4.4 Partisan impacts of heuristic optimization . . . 25

5 Conclusion **27**

1 Introduction

In 2018 alone, five states saw redistricting reform measures passed by voters at the ballot box, signaling a larger movement to revisit redistricting rules in state laws and constitutions in the run-up to the 2020 Census. A common feature in this surge of reform efforts is redoubled scrutiny of the rules, priorities, and criteria for new districting plans. In addition to reconsidering the treatment of traditional districting principles like population balance, compactness, and preservation of municipal boundaries, most of these state reforms explicitly address the use of partisan data and metrics in the redistricting process.

One reform approach in the recent trend has been to bar mapmakers from partisan considerations. Several states broadly prohibit favoring political parties or incumbents, while others go further and restrict mapmakers from directly using any partisan data while drawing the lines. At the opposite extreme, other reform approaches mandate that certain types of partisan symmetry or fairness metrics should be important decision criteria, effectively requiring that partisan data be taken into account during map construction. Competitiveness rules fall into the second category.[1]

This new attention to redistricting criteria is taking place in the context of an explosion of data and computational power in redistricting. One consequence of this new era of computational redistricting is that methods for locking in partisan advantage—and data proxies that make this possible without explicitly optimizing for election outcomes—create major opportunities for abuse. At the same time, computationally sophisticated ensemble techniques are improving dramatically at detecting partisan outliers. In this paper we devote our attention to a second, complementary application of algorithmically-generated ensembles of plans: we use them to study the effects of enacting different sets of rules and priorities. In particular, we will develop a point of view that ensembles can reveal some districting tasks to be more "elastic" than others: control of line-drawing always confers

[1]Some reform measures explicitly provide for a firewall between partisan-blind and partisan-aware parts of the process. For instance, Utah's successful ballot initiative in 2018 requires the maps to be drawn without considering partisan data but requires the expected partisan behavior to be measured before a plan can be approved [12]. Arizona requires similar behavior from its redistricting commission; the relevant portion of the redistricting law says that "Party registration and voting history data shall be excluded from the initial phase of the mapping process but may be used to test maps for compliance..." [34]. In these cases, the process is designed to prevent mapmakers from directly optimizing for the metrics and instead sets up a failsafe against intended or unintended partisan skew.

some control of outcomes, but some situations prove to be more manipulable than others.

Our main focus here is on rules favoring electoral competitiveness. Competitiveness is frequently regarded as a healthy attribute of elections that promotes voter engagement, responsiveness, and accountability of representatives, among other benefits.[2] During an election, competitiveness of a race might be measured by close polling or by candidate spending; after the fact, it is often assessed by a close outcome. It is much harder to capture the competitiveness level of districting plans at earlier stages, since they are created before there are outcomes, polling, or even a choice of candidates for the coming cycle. Accordingly, there is no agreed-upon means of measuring competitiveness, and states have operationalized it in different ways when they have defined it at all. It is far from clear that a high competitiveness score can be achieved by a plan without other partisan consequences or unintended costs to other districting criteria, or that it will bring about outcomes that increase legitimacy and public trust. With these questions in mind, in this paper we develop statistical techniques to probe the consequences of various ways to quantify competitiveness of districting plans.

The primary reason that it is difficult to evaluate the potential consequences of altering or reordering redistricting criteria is that each set of rules and constraints constructs a different universe of valid plans, and those spaces are vast and complex. Since the set of valid plans typically is too large to enumerate, Markov chain Monte Carlo (MCMC) methods are standard techniques in applied mathematics and statistics for sampling in these situations, used to gather representative samples from the space of interest. Markov chain sampling is currently used by several groups of researchers to evaluate the partisan nature of proposed or enacted districting plans in court challenges [17, 10, 11, 19, 23, 24]. These methods work by constructing large ensembles of plans that satisfy the rules set forth in each state and comparing the statistics of enacted plans to aggregate statistics of the distribution. These techniques, however, require the choice of mathematical specifications of the legal constraints, which are rarely written with enough precision to determine an unambiguous

[2]For instance, consider this language in Colorado's recent redistricting reform legislation: "Competitive elections for members of the United States House of Representatives provide voters with a meaningful choice among candidates, promote a healthy democracy, help ensure that constituents receive fair and effective representation, and contribute to the political well-being of key communities of interest and political subdivisions" [35].

formulation. Understanding the interactions between various districting criteria is fundamental for drawing meaningful quantitative conclusions from ensemble analysis, and we hope that this study provides an invitation for more research in this direction.

1.1 Contributions

We analyze several variants of competitiveness metrics and compare the effects of prioritizing them in different ways. We use both *neutral ensembles* of alternative districting plans (i.e., generated without considering partisan data) and *heuristically-optimized* plans that seek to extremize the metrics for Congressional and state Senate plans in five states.[3] We aim this discussion at two primary audiences. The first is legislators and reform groups who are actively working to formulate new redistricting rules, particularly those considering imposing competitiveness criteria. The second audience is researchers and quantitative analysts who are interested in assessing how well proposed plans uphold the stated criteria from which they were constructed. A message that emerges for both groups is the major impact of seemingly innocuous modeling choices against the idiosyncratic political geography of individual states. We argue that no universal metric of competitiveness should be adopted across the country, but rather that there is a fresh choice to be made in each circumstance. On the other hand, we find that certain metrics are unreasonable and should *not* be adopted.

For instance, packing and cracking are the tools of the trade for gerrymandering, so it is notable that in certain circumstances, a competitiveness metric can approve or even incentivize packed and cracked plans. For instance, in a state with 68-32 voting for Party A, a plan that arranges A voters into some districts at 48% share and compensates that with other districts at 88% share could be textbook vote dilution while looking excellent under a metric that merely counts nearly-balanced districts. If an arrangement like that were the only or the predominant way to satisfy a competitiveness priority, then this would be important to know in advance of enactment. A principal contribution of our ensemble

[3]The only reason not to consider state House plans is that House districts tend to be too small to be made out of whole precincts while maintaining tolerable population deviation. Precincts are the smallest level at which we have authoritatively accurate election outcomes, which are a crucial element of this style of analysis. Vote results can be prorated to smaller units, but this is done at the cost of a potentially significant loss of accuracy.

analysis is that it escapes the tendency to treat votes as though they are geographically unbound and fluidly rearrangeable, instead studying the actual elasticity of the vote by comparing to alternative geographic partitions.

In our treatment, a general discussion of the reasons for preferring competitive districts in terms of broad alignment with societal and democratic goals will remain in the background. These normative questions are extremely important, but they are beyond the scope of the present paper and outside the reach of ensemble techniques, except to the extent that these techniques clarify the desirable and undesirable consequences of the norms.

1.2 Related work

Approaches to competitiveness in the political science literature vary widely. Although most recent reform efforts have focused on making more competitive districts, arguments have been made for instead packing districts with members of a single party to the extent practicable [4]. The opposite situation was considered in [20], where regression analysis was used to detect correlations between traditional districting criteria and competitive districts in the 1990 and 2000 cycles. They found that rules prioritizing population constraints as well as preservation of communities of interest and municipal boundaries were associated with more competitive districts. Additionally, their analysis helped expose the complexity of interactions between multiple redistricting criteria and the tradeoffs that are inherent in the process. Our work extends theirs by using ensembles to consider the potential impacts of directly incentivizing competitiveness, evaluating the properties of many alternative districting plans rather than the currently-enacted maps. Our analysis does not model the impacts of whether the lines are drawn by legislatures, courts, or commissions. Several groups of political scientists have studied the relationship between the identity of the line-drawers and the degree of competitiveness in subsequent elections [7, 14, 21, 22, 31], with mixed conclusions.

A recent article by Cottrell [13] presents an ensemble analysis to evaluate the impact of gerrymandering on the level of political competition around the country, concluding that gerrymandering has only a small cost in competitiveness because neutral processes also produce safe seats at rates similar to what can now be observed. We refer the reader to

his references for a broad discussion of the arguments around competitive standards in the political science literature. This differs from our approach both in goals and methods, but we view our problem domain as complementary to his. In terms of goals, Cottrell seeks evidence of gerrymandering, while we study the impact of rules that favor competitive outcomes. That means he is comparing neutral ensembles to enacted plans, and we are focused on comparing neutral ensembles to methods for favoring competitive outcomes. Another difference is that our analysis uses a newly-developed Markov chain technique with strong evidence of sampling effectiveness.[4]

1.3 Review of competitiveness language in legal criteria

Our research is motivated by several recent examples of redistricting rules introduced at the state level to promote competitive districts.

1.3.1 Washington, New York, Arizona

Some states had competitiveness language on the books before the recent reform wave, but not with quantitative details. Washington requires its commission to "provide fair and effective representation and to encourage electoral competition" and New York's rules require that "districts not be drawn to discourage competition" [37, 43].

Arizona's instructions to its commission include the following language: "To the extent practicable, competitive districts should be favored where to do so would create no significant detriment to the other goals" [34]. In practice, out of 9 districts, this text seems to suggest a conspicuous effort to craft 3 competitive districts (53-47 outcomes or closer in 2012), leaving 4 safe Republican districts and 2 safe Democratic districts (all 60-40 or farther).

1.3.2 Colorado

In 2018, Colorado voters approved a ballot initiative that included a requirement that plans "maximize the number of politically competitive districts," where competitive is defined

[4] With the recombination technique used here, collecting 100,000 plans from a Markov chain is shown to be enough in many circumstances to obtain a consistent distribution of sample statistics, regardless of starting point. See [17] and its references for details.

as "having a reasonable potential for the party affiliation of the district's representative to change at least once between federal decennial censuses" [35]. A literal reading of this language suggests that any district in which both parties have at least a 13% projected chance of winning might match the Colorado law, since $(1 - 0.13)^5 \approx 0.5$. While this interpretation makes several potentially unrealistic assumptions (including that a point prediction would carry through over the full cycle and that there is no impact of incumbency on the elections), the text of the legislation does not provide implementation details.[5]

1.3.3 Missouri

Missouri's 2018 constitutional amendment, passed through a voter referendum known as Clean Missouri, defines competitiveness in terms of a partisan metric called the *efficiency gap*, or *EG* [36]. It is one of the more detailed descriptions of a competitiveness metric in the recent reform measures, requiring the state demographer to create a blended voting index by computing the turnout-weighted averages of the Presidential, Senatorial, and Gubernatorial elections that occurred during the previous census cycle. This pattern of votes is then modified by applying a uniform partisan swing of $\{-5, -4, \ldots, +4, +5\}$ and recomputing the efficiency gap against each modified vote dataset. Proposed districting plans are to be constructed so that each of the corresponding *EG* values is "as close to zero as practicable"—we call this *EG/swing zeroing*. We will show in §2.4 that this definition of competitiveness promotes drastically different qualitative properties of districting plans than the other metrics, and if taken literally it actually puts an *upper limit* on the number of districts that are allowed to fall in the 45-55% range for its voting index.

Missouri's reform is also noteworthy in that it places a higher priority on compliance with the priority on competitiveness than on contiguity or compactness of the plan.

1.3.4 New Jersey

Finally, in 2019, New Jersey's legislature debated a constitutional amendment that defined a competitive district to be one where the expected voting outcome is within 5% of the

[5]This rule compares interestingly to the measure of competitiveness used by the 538 Atlas of Redistricting [6]. The Atlas project defined districts in the range from D+5 to R+5 as competitive, using the CPVI metric described in §2.3. Under their vote modeling, this corresponds to projected probabilities of approximately $18\% - 82\%$ of electing a candidate from a given party.

statewide average over the previous census cycle [25]. A plan was defined to be sufficiently competitive if 25% of its districts were competitive. The bill was widely pilloried (and ultimately defeated) because in a state whose recent voting trends $\approx 57\%$ Democratic, a district must have a Democratic majority to count as competitive under this language.

These examples and others demonstrate a range of efforts intended to quantify and promote "competitiveness" in districting plan design. Our analysis below defines mathematical formalizations in keeping with the legal language reviewed above to study potential impacts of implementation.

2 Competitiveness metrics

We consider several types of competitiveness metrics in this paper, motivated by the recently proposed legislation in several states discussed above.

2.1 Plans and votes

Characterizing a districting plan or individual district as competitive necessarily requires some assumptions about the underlying voter behavior. That is, if we denote a districting plan by $\mathcal{D}$ and a vote pattern by Δ, then all scores C of competitiveness are really functions of both: $C = C(\mathcal{D}, \Delta)$. The choice of Δ that allows the most robust assessment of a Congressional or legislative districting plan is a matter of some debate. Some political scientists insist that it is important to use only *endogenous* data, or vote patterns from the same type of election, but the partisan preference information in these patterns is confounded by other variables such as incumbency, uncontested races, and particular candidate characteristics. Other scholars prefer the use of statewide *exogenous* vote patterns, but then must make a case about the choice of which election or elections to use to create an informative voting index. See for instance [2, 3, 29] for a paper and series of responses on this issue. In all cases, care is required in terms of predictive claims about future voting. We discuss the choice of Δ in more detail in §3.2.

Several states that require partisan analyses of proposed plans have addressed the choice of election data in their legislation. As noted above, Missouri specifies the exact election

data that must be used to evaluate partisan performance [36], while Colorado simply requires [35] that "[c]ompetitiveness may be measured by factors such as a proposed district's past election results, a proposed district's political party registration data, and evidence-based analyses of proposed districts." New Jersey's proposed amendment [25], had it passed, also would have specified exactly which data was to be considered. As with many of the other variables in redistricting legislation, there does not appear to be a consensus around the optimal modeling choice.

The results and visualizations reported below are based on a fixed vote pattern: we use Δ =Pres16, i.e., the precinct-level election returns from the 2016 Presidential election, restricted only to the Democratic and Republican votes cast. No aspect of our method hinges on this choice, and it can be repeated for any other voting pattern in place of Δ.

2.2 Evenness and typicality

Perhaps the most intuitive measure of competitiveness for a district with respect to a vote pattern is the difference between the two-way vote share and 50%. This metric is implicit in Arizona's and Colorado's legal language. Up to scale, this measures the number of votes needed to change the outcome in the district. We might consider this a measurement of *evenness*, and it aligns well with standard ideas about competitiveness.

The Cook Partisan Voting Index (CPVI) is a widely used metric that is constructed by comparing the major-party vote share in each geographical unit—precinct, district, etc.—to the nationwide average across the two most recent Presidential elections [42]. Frequently, a district is described as competitive if its CPVI value is between D+5 and R+5, meaning that the district's vote is within 5% of the nationwide average. Upon inspection, most observers would agree that in the case that the average is skewed away from 50-50, this is not measuring competitiveness but rather *typicality*. Nonetheless, CPVI is reported by the Cook Political Report after each election as a measure of competitiveness across the nation; it also was used by 538's Atlas of Redistricting to count competitive districts [6]. New Jersey's proposed measure described above was based on an idea similar to CPVI, except that districts are compared to the statewide instead of nationwide average. Below, we will denote the statewide share of Democratic voters with D_0.

To measure the effects of quantifying and promoting competitiveness, we compute and compare metrics of evenness and typicality below with a generalized formulation that we call vote-band metrics.

2.3 Vote-band metrics

We start by defining a district to be in the (y, z) **band** with respect to voting pattern Δ, or simply to be a (y, z) district, if its Democratic vote share is within y of a target z. By default we will let y, z be denominated in percentage points, so that for instance a district is in the $(5, 50)$ band with respect to a voting pattern Δ if its vote share is between 45% and 55%.

Aggregating this behavior over the districts in a plan, we define $(\mathcal{D}, \Delta)$ to be (x, y, z) **compressed** if at least x share of the districts are in the (y, z) band. By varying the parameters x, y, and z, we can understand the elasticity of a vote pattern Δ in terms of its divisibility into districts at a certain scale. If voters of one group are very clustered, there might be a great deal of control in the hands of the districters by manipulating how the lines cross the clusters. In case of extreme dispersal, there may be little impact to moving the lines. And in between those extremes, the spatial distribution of votes and the units of data discretization combine to create effects on the range of possible outputs. Examining whether it is possible to compress district-level vote shares into narrow bands of outcomes is one way to examine how much control the districters have to create districting plans with precise properties.

A legal standard that employs this type of metric might require that adopted plans be (x, y, z) compressed, where x, y, z, and the choice of Δ are set by statute. This encompasses both many competitiveness rules (where the target is $z = 50$) and typicality rules (where the target is $z = D_0$ for state typicality). We will consider how plans that satisfy compression properties relate to unconstrained plans, along with studying how varying the parameters affects expected outcomes as a function of political geography.

Figure 1 shows how the currently enacted Congressional and state Senate plans in Utah, Georgia, Wisconsin, Virginia, and Massachusetts relate to competitiveness and state-typicality bands at $y = 5$, using 2016 Presidential vote data. Table 2 records the number

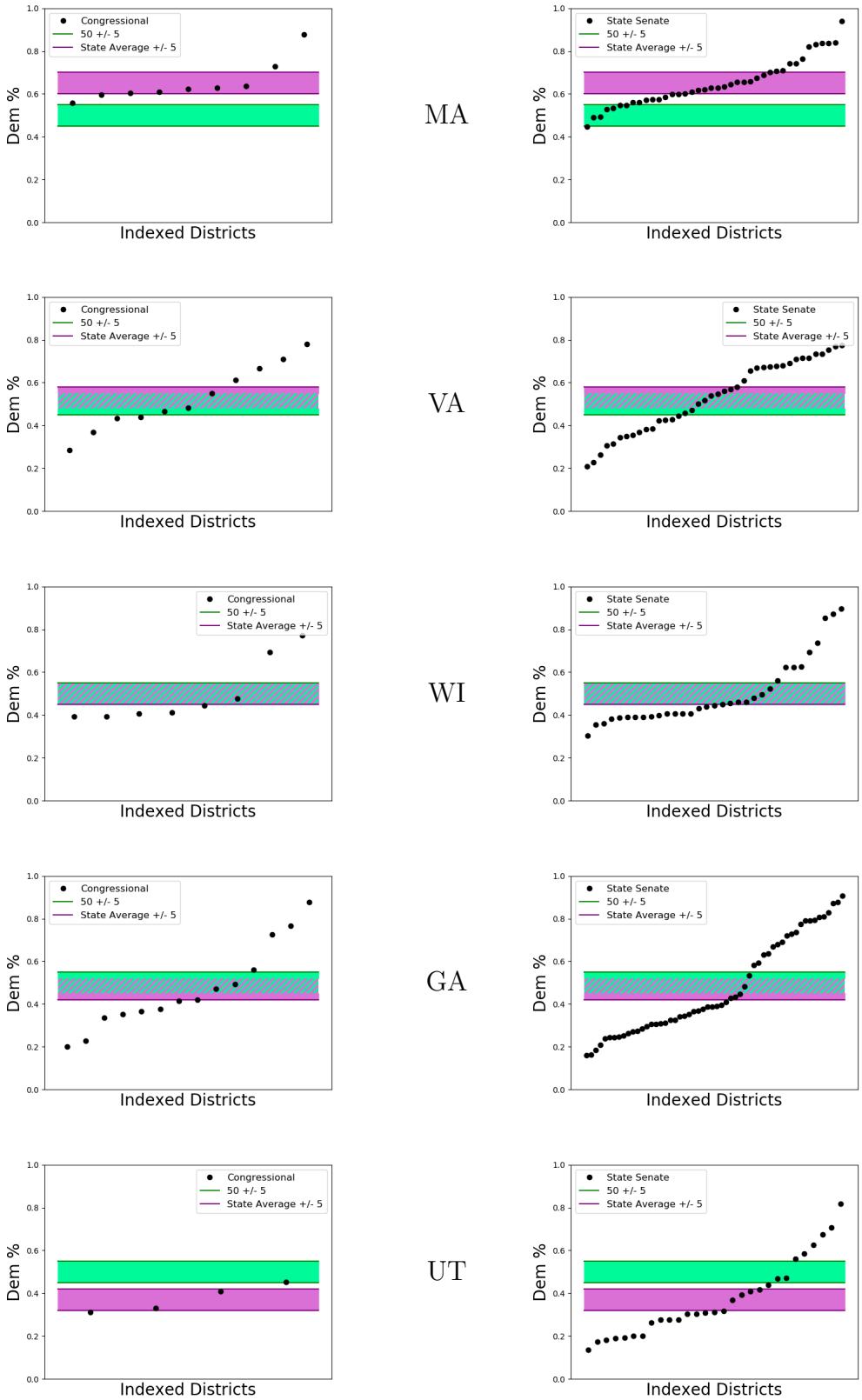


Figure 1: Comparison of the enacted Congressional and state Senate plans to $(5, 50)$ bands (green, competitive) and $(5, D_0)$ bands (purple, state-typical). Here and in most plots below, the districts are ordered from smallest D share to highest to facilitate comparisons across plans.

of districts in those plans that fall into each band. Although none of these states had requirements governing competitiveness when the plans were drawn, they provide a starting example for considering the consequences of vote-band rules.

State	# Cong Districts	# Sen Districts	D_0 in Pres16	2012-16 CVPI
Massachusetts	9	40	64.7	D+12
Virginia	11	40	52.8	D+1
Wisconsin	8	33	49.6	Even
Georgia	14	56	47.3	R+5
Utah	4	29	37.6	R+20

Table 1: The five states considered in our study, chosen for a wide range of partisan tilts. The Democratic share D_0 is the two-party share in Pres16, so that for instance the sizeable third party vote in Utah is not reflected.

In most states, several districts lie outside both bands. Given fixed (y, z), there is no guarantee that it is even possible to construct a plan with a large number of (y, z) districts—even if z is near the statewide average—while adhering to reasonable compactness and boundary preservation norms. In Virginia, the enacted plan almost fails to have any districts that are competitive under both definitions, even though there is a large overlap between the bands.

One straightforward observation is that there must be some districts at least as far from 50% as the state average. Additionally, as shown in [18], discretization of geographical units can play a large role in determining the range of possible outcomes, since the voting population is not infinitely divisible.

Given a fixed vote share target z, it is natural to ask how the number of (y, z) districts varies as a function of the band width parameter y. Figure 2 shows the number of districts in the enacted plans that lie in competitive and state-typical bands as the width is increased. For most of the current plans, y values of approximately 15-20% would be needed to make half of the districts competitive; that is, about half the districts have outcomes even more skewed than 70-30. Also, as one would expect, it is harder to make even districts in the states that are themselves more partisan-tilted. Comparing Utah and Massachusetts to the other states shows the extent of this effect.

In our experiments, we will explore the interaction of vote-band metrics with political

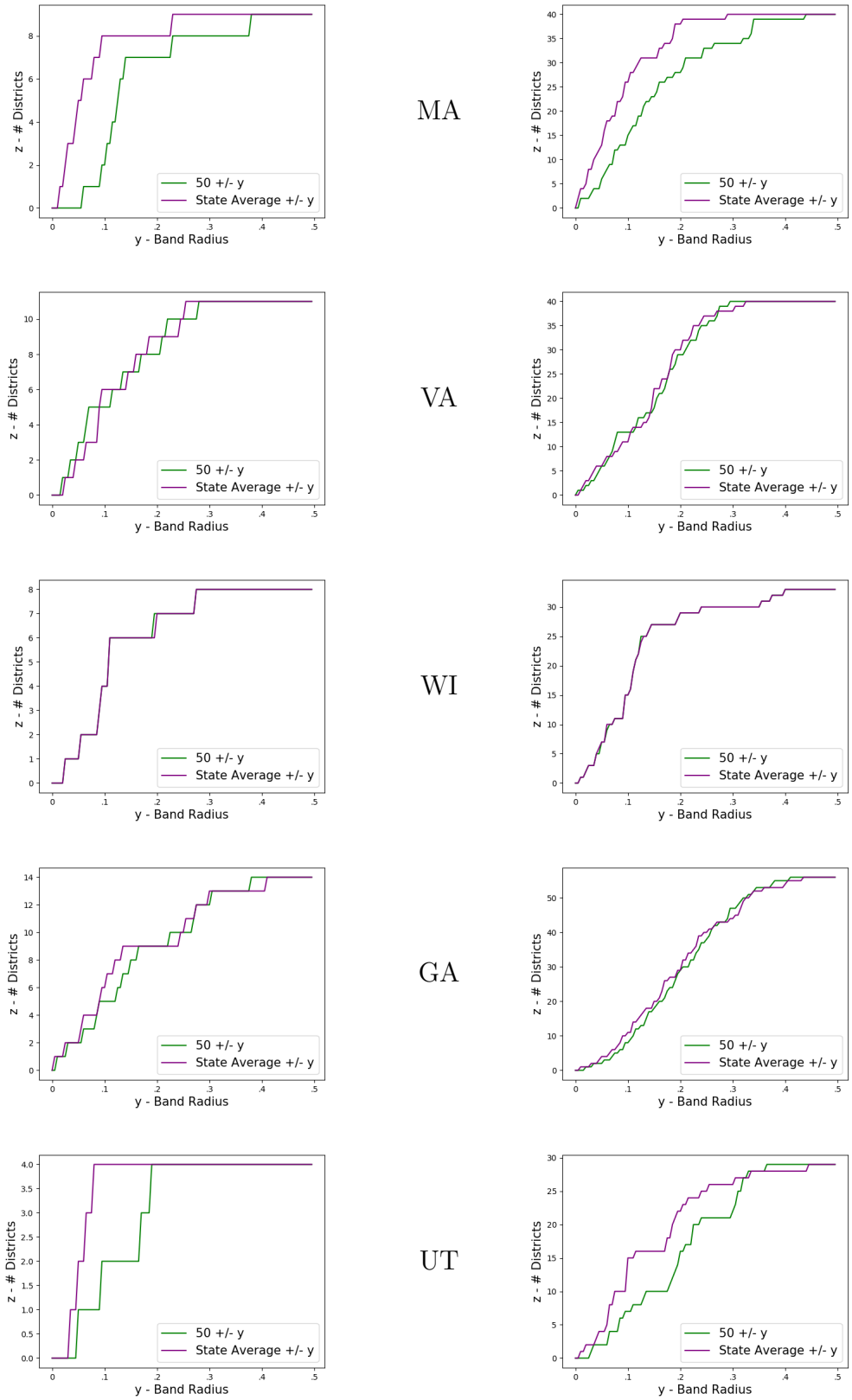


Figure 2: Number of districts in the enacted Congressional and state Senate plans that are $(y, 50)$ (green, competitive) and (y, D_0) (purple, state-typical), as y varies. For instance, all four of Utah's congressional districts are within 10% of the state average.

geography. Although this formulation is simple, we will see that the impacts of enforcing a choice of parameters x, y, and z do not fall uniformly over our test states.

2.4 EG/swing zeroing

As discussed above, Missouri has adopted a definition of competitiveness for redistricting that requires the State Demographer to compute a metric called the *efficiency gap* (described below) for any proposed districting plan, with respect to a particular synthetic vote pattern created by a weighted combination of recent election outcomes. They are also required to consider ten alterations of that vote pattern, obtained by uniformly swinging the district vote shares by $\{\pm 1, \pm 2, \pm 3, \pm 4, \pm 5\}$ and computing the corresponding efficiency gap values. A literal reading of the constitutional text requires that each of those eleven efficiency gap values should be made as small as practicable. Due to the close relationship between efficiency gap and seat outcomes, many authors have noted that $EG \approx 0$ prescribes a "winner's bonus" with slope two in the votes to seats conversion [38, 39, 1].[6] Here, we show that this constraint is directly at odds with vote-band competitiveness.

Let D_0 represent the statewide percentage of Democratic voters and S represent the number of seats won by the Democratic party, out of k total seats. Then, as derived by Veomett in [39], the efficiency gap is

$$EG(\mathcal{D}, \Delta) = 2\left(\frac{D_0}{100}\right) - \frac{S}{k} - \frac{1}{2} + N(\mathcal{D}, \Delta),$$

where $N(\mathcal{D}, \Delta)$ is a noise term depending on S, k, and the average turnout in districts won by each party, which is zero in the case of equal turnout.[7] Noting that D_0 is fixed and neglecting the noise term, we must minimize this expression to satisfy the constitutional requirement, which amounts to securing a number of Democratic seats S as close as possible

[6] That is, having $EG \approx 0$ requires having a tight relationship between overall vote share and overall seat share, with each additional percentage point of votes calling for *two* additional percentage points of seats. Since seats must turn out in a particular way to satisfy the EG test, there is a completely different reason that it is at odds with competitiveness: competitive seats introduce uncertainty and the potential for an unpredictable EG score [1].

[7] Indeed, publications by Stephanopoulos–McGhee sometimes use this full definition of EG and sometimes use the "simplified formula" without the N term. Converting from Veomett's notation to ours and letting $s = S/k$ and ρ be the ratio of average turnout in districts won by Party A to those won by Party B, we have $N = \frac{s(s-1)(1-\rho)}{s(1-\rho)+\rho}$. This vanishes in the equal-turnout case $\rho = 1$.

to $k(2\frac{D_0}{100} - \frac{1}{2})$. Repeating this procedure for each of $\{D_0 - 5, D_0 - 4, \ldots, D_0 + 4, D_0 + 5\}$ gives us a sequence of eleven prescribed seat values, $S_{-5} \leq S_{-4} \leq \cdots S_4 \leq S_5$. But we further observe that prescribing how the seat values change as you shift the votes 5 points up and down is precisely equivalent to prescribing the number and position of seats in the $(5, 50)$ band.

To show the potential impacts of this EG/swing zeroing rule, consider Missouri and Wisconsin. With respect to Pres16, Missouri had roughly $D_0 = 40$.[8] First we work out the consequences of the competitiveness rule on the 8-seat Congressional map, though we note that the Clean Missouri constitutional reform only applies to legislative districting. Continuing to neglect turnout noise, we observe that $EG(\mathcal{D}, \Delta + i) = 2(\frac{D_0+i}{100}) - \frac{S}{8} - \frac{1}{2}$, which prescribes $S = 2$ Democratic seats for shifts between $R + 5$ and $R + 0$ and by $S = 3$ seats for $D + 1$ through $D + 5$. This can *only* be satisfied by having one district with a Democratic share of between 49 and 50 percent and no others between 45 and 55 percent. If the Democratic shares by district were roughly $(22, 22, 22, 22, 22, 49.5, 80, 80)$, the plan would earn perfect marks for "competitiveness" by this EG-swing definition. Having more districts close to even would rate strictly worse.

With a bit more arithmetic, we can extract the pattern. Noting that minimizing EG amounts to solving $\frac{S}{k} \approx 2V - \frac{1}{2}$, we replace V with its extreme values $\frac{D_0-5}{100}$ and $\frac{D_0+5}{100}$. The difference between the largest and smallest number of required D seats is exactly the number required to fall in the 45-55 range. That is, accounting for rounding to the nearest integer, this rule prescribes that the number of $(5, 50)$ seats is equal to $k/5$ or one-fifth of the available seats, plus or minus no more than one, no matter whether D_0 is 35 or 40 or 45 or 50.[9] So, in Missouri's 34-seat Senate, the rule will require precisely 6-7 seats to fall between 45 and 55 percent, with at least 27 safe seats, and in the 163-seat House, it will require 32-33 close outcomes, with at least 130 safe seats.

In Wisconsin, which had roughly $D_0 = 49.6$, the EG-swing standard would require two Congressional districts between 45 and 55 percent and six districts outside that range.

[8] Clean Missouri prescribes that the vote pattern be constructed from the last three races for Governor, Senate, and President. At the time of writing, this amounts to a $D_0 = 50.3$, but when it is computed in 2021 it is likely to drop to the mid-40s. Interestingly, the number of competitive districts required by the rule is barely sensitive to the value of D_0.

[9] For values of D_0 more extreme than 30-70, the allotment for close seats actually goes down as the efficiency gap demands virtually all safe seats for the majority party.

Again, this optimization promotes the creation of many safe districts and is indifferent to the possibility that those are actually made into landslide districts. This effect is more dramatic in Utah ($D_0 = 37$, $k = 4$), where adherence to this rule would require that Democrats are ahead in precisely one Congressional district out of four, for all of the swing values. That is, a Utah plan complying with this definition of competitiveness would not be allowed to have *any* districts within 45 and 55 percent.

This construction is in tension with the plain-English notion, in which a plan is said to be more competitive when more districts are close to even. In some states, demanding that one-fifth of seats (and no more) be competitive might be a reasonable goal and an improvement over the status quo; in other states, it will not be possible or will have major unintended consequences. Considering that this definition also entails a specific number of safe seats for each side, it is not easily separated from a very prescriptive partisan outcome rule and clearly calls for careful partisan tailoring. We will not continue to model more refined consequences of adopting this kind of metric below, but will focus on vote-band definitions.

3 Data and methods for ensemble analysis

We analyze the potential effects of enforcing a vote-band rule as described in §2.3 in two sets of experiments, using data from five states. First, we generate a large neutral ensemble of districting plans for each state and filter the plans according to increasingly strict vote-band constraints, comparing the properties of the compressed plans to those of the full ensemble. Next, we use two heuristic optimization techniques to generate plans that score well on the given competitiveness measure and compare the statistics of the optimized plans to that of the full ensemble.

3.1 State data

In our study, we focus on Georgia, Massachusetts, Utah, Virginia, and Wisconsin, as exemplar states that display a wide range of partisan leans (see Table 1). The CVPI baseline over this time period is approximately 51.5%, as the Democratic party candidate narrowly

won the popular vote in both the 2012 and 2016 elections.

Our methodology uses a discrete formulation of redistricting, viewing a plan as a graph partition for each given state. This is defined with respect to the *precinct dual graph*, which has a node for each precinct and an edge between two nodes if the corresponding precincts are adjacent. ("Dual graph" is a mathematical term for a graph that records adjacency patterns.) Given a fixed number of districts k, a districting plan is a partition of this graph into k connected subgraphs. Each node is decorated with the population of the corresponding precinct and its vote counts with respect to the election Δ, so the population and partisan performance of each district can be computed by summing over the nodes. This discrete approach has enabled a growing literature on sampling techniques in court cases and in scientific literature [9, 11, 17, 23].

To construct the dual graph for each state, we use precinct shapefiles with population and vote data joined.[10]

3.2 Partisan metrics

The choice of Pres16 as the vote pattern Δ for these runs makes it easier to compare results across states and is consonant with the CVPI approach and with several other political science analyses. We note, however, that the resemblance between Presidential and Congressional or Legislative voting patterns varies greatly from state to state and from year to year, and we do not feel that the lessons learned from this analysis hinge on there being a close match.

Nevertheless, we can observe that the number of districts with a Democratic majority in Pres16 matches the actual Congressional 2016 outcome in MA, GA, and UT (with

[10]The underlying geographic and partisan data for each state was originally obtained from the following sources. **Georgia:** Geography from the Georgia General Assembly Legislative and Congressional Reapportionment Office [27]; partisan data from MIT Elections Data and Science Lab (MEDSL) [26]. **Utah:** Geography from the Utah Automated Geographic Reference Center [8]; partisan data from MEDSL. **Wisconsin:** All geographic and partisan data from the Wisconsin Legislative Technology Services Bureau [5]. **Virginia:** All geographic and partisan data from the Princeton Gerrymandering Project OpenPrecincts project [33]. **Massachusetts:** Geography from the Massachusetts Secretary of the Commonwealth and partisan data from the Massachusetts Secretary of the Commonwealth Elections Division [32]. Staff of the Metric Geometry and Gerrymandering Group joined population data from the 2010 US Census to these datasets by aggregating up from census blocks, then used geospatial tools to clean the geographical data. The processed shapefiles and metadata are publicly available for download [30].

9/9, 4/14, and 0/4 Democratic-majority districts, respectively), while assigning one D seat too many in Virginia and one too few in Wisconsin (where the enacted plan laid over Pres16 voting gives 5/11 and 2/8 Democratic-majority districts, respectively). We have also carried out the trials reported below with respect to recent U.S. Senate and Gubernatorial elections where data was available; the findings are broadly similar, with the same big-picture messages about the relative elasticity and the between-state variability of voting patterns.

To evaluate the partisan side effects of competitiveness rules on districting plans, we use three partisan metrics, again without taking a stance on which of these is the best or most useful. Rather, we seek the presence of interrelations. The first partisan metric is simply the number of districts in which the number of Democratic votes exceeds the number of Republican votes (discussed above in this section). A second metric, defined in §2.4, is the efficiency gap. Finally, a third popular partisan metric is the *mean-median (MM) score*, a measure of asymmetry that is defined by subtracting the mean Democratic vote share across the districts of a plan from the median of the same list of values. Drawing on the work of Bernard Grofman, Gary King, and others, this metric is sometimes described as measuring how much one party could fall short of 50% of the statewide votes while still receiving 50% of the seats [28, 41]. Symmetry metrics are explicitly included in Utah's reform language from 2018 [12].

3.3 Ensemble generation

We use open-source software `GerryChain` [40] both for generating the neutral ensembles and for optimizing for competitive districts. `GerryChain` uses a Markov chain to generate districting plans, beginning with an initial state and making iterative changes to the assignments of the nodes. Our ensembles are constructed using the ReCom Markov chain procedure overviewed in [17], which uses a spanning tree method to bipartition pairs of adjacent districts at each step.

For the neutral ensembles, we require the districts to be connected and population-balanced, imposing no further restrictions.[11] We constrain the population to within 2% of

[11] As described in [17], ReCom produces plans with compactness scores in range of human-made plans

ideal for Congressional districts and within 5% for state Senate districts.[12] For each state, we take 100,000 ReCom steps for the Congressional plans and 1,000,000 steps for state Senate plans, beginning at a random seed generated by a recursive tree-based partition; plans with more districts will require more recombination moves to achieve a high-quality sample. See [17] for a workflow describing standard heuristic tests of sample convergence, such as independence of seed and stability under increases in run length.

For the heuristic optimization ensembles, we use two types of hill-climbing proposals to find elevated numbers of $(5, 50)$ districts. We call these methods Opt1 and Opt2 and construct 100 plans using each. We can take far fewer steps here than in constructing ensembles because we are not trying to sample, but only to find extreme examples. To generate each plan, we start with a number of unconstrained ReCom steps (200 for Congress, 1000 for Senate) and then attempt 1,000,000 Flip steps, each proposing to change the assignment of a single node on the boundary to move toward a more competitive plan. The unconstrained steps serve to decrease correlation to the last map before attempting the greedy Flip proposals. In the Opt1 run, proposed plans are rejected if they decrease the number of $(5, 50)$ districts. In the Opt2 run, a proposed plan is accepted if it decreases the sum of the distances between the district Dem shares and the 45-55 percent band. Because Flip steps tend towards poor compactness scores, these steps are additionally constrained to have no more than twice as many cut edges as the partition after the ReCom steps.

4 Ensemble results

In the previous section, we describe how to generate large ensembles of compact, contiguous, population-balanced maps that do not take partisan data into account. It is important to note how we propose to use these neutral ensembles for comparisons. We do not propose that matching summary statistics to, say, the mean or median of a neutral ensemble is

without any need to impose additional compactness constraints. Compactness is measured by a discrete metric called *cut edges*, which counts how many pairs of units were adjacent in the state but are assigned to different districts—these are the edges that would have to be cut to separate the whole graph into its district subgraphs.

[12] Most states balance Congressional districts to within one person, but 2% balanced plans can readily be tuned to tight balance by a skilled human mapmaker. With 1 or 2% deviation, we get efficiently moving Markov chains. In other studies we have confirmed that loosening or tightening population balance does not have substantial impacts on partisan summary statistics [15].

a normative goal in redistricting, and we do not view the neutral ensembles as strictly bounding the range of possibilities. Instead, they provide a baseline range for any scores and metrics we might study, constructed only by the basic rules of redistricting and the political and physical geography of the state.

The key takeaway is that baseline ranges for partisan metrics are sensitively dependent both on the formulation of the competitiveness metric and on the underlying political geography, which controls the elasticity of the vote. For this reason, great care should be taken to evaluate potential impacts of these rules, using data representing the actual distribution of voters across the state, before adoption.

4.1 Partisan baselines

Even for a fixed distribution of votes, the district-by-district vote shares depend, of course, on the district lines. Figure 3 shows the distributions of the vote share across districts for each of the full ensembles. In these plots, the districts in each plan are arranged from the one with the lowest Democratic vote share to the one with the highest. The boxes show the 25th-75th percentile Democratic share, and the whiskers go from 1st-99th percentile. Thus fatter boxes show a more elastic vote, with more control of partisan outcomes in the hands of the line-drawer.

In Table 2 we show the average number of $(5, 50)$ and $(5, D_0)$ districts across the ensemble, compared to the enacted plans. The distribution of values observed across the ensemble is similar to those of the enacted plans, with the exception of Georgia, where the ensemble finds many more districts in both bands.

The baselines reveal the extent to which the various states' political geographies shape the ensemble statistics. With the exception of Virginia, whose MM distribution is centered at zero, all of the states have significantly Republican-favoring average mean–median scores of between two and five percent.[13] For several ensembles, more than 90% of the plans have a Republican-favoring score. This adds to a rapidly growing body of evidence that shows that applying uniform standards across states, without accounting for how voters are distributed throughout each state, can lead to misleading results and unattainable ideals. For example,

[13]The standard narrative around mean–median scores holds that $MM = .05$ means that the Republican party could expect half of the seats with only 45% of the votes.

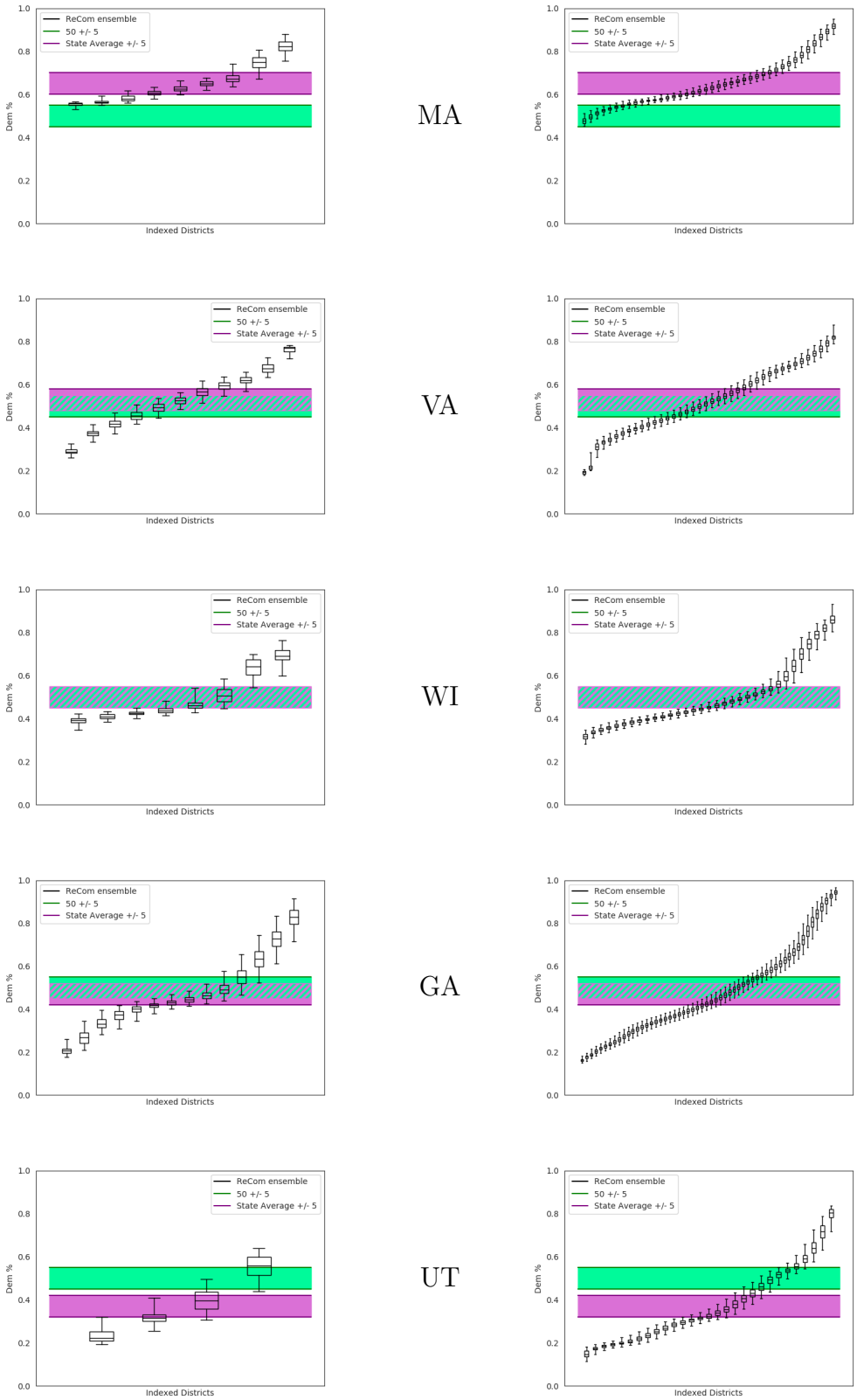


Figure 3: Partisan share boxplots by district for the neutral ensembles in Congress (left) and state Senate (right).

State, Type (#)	Enacted $(x, 5, 50)$	Ensemble $(x, 5, 50)$	D_0	Enacted $(x, 5, D_0)$	Ensemble $(x, 5, D_0)$
MA Cong (9)	0	0.24	64.7	5	3.85
VA Cong (11)	3	2.79	52.8	2	2.82
WI Cong (8)	1	1.76	49.6	1	1.81
GA Cong (14)	2	2.63	47.3	2	4.43
UT Cong (4)	1	0.62	37.6	2	1.05
MA Sen (40)	6	6.85	64.7	13	12.82
VA Sen (40)	6	8.60	52.8	6	8.00
WI Sen (33)	7	8.80	49.6	7	9.10
GA Sen (56)	2	8.57	47.3	4	8.78
UT Sen (29)	2	4.09	37.6	4	4.57

Table 2: Comparison of the number of $(5, 50)$ and $(5, D_0)$ districts—close to 50-50 and close to the state average, respectively—in enacted plans versus ensemble means. Most cases, the enacted plan has slightly fewer $(5, 50)$ districts than the ensemble average—the only examples where the enacted plan has more than the mean are Virginia's and Utah's Congressional plans. Georgia and Utah's Senate numbers stand out from the group in the other direction, with conspicuously fewer close seats than typical plans in the neutral ensemble.

a plan whose MM value is typical for the Wisconsin ensemble would be a significant outlier if evaluated against the Virginia data, even though the difference in statewide average vote share is small.[14]

The seats outcomes let us measure the extent of partisan advantage that is entailed by political geography. For instance, Wisconsin's vote pattern which was very close to 50-50 in Pres16, but the most Democratic-favoring plans found by the state Senate ensemble have only $14/33 \approx 42\%$ Democratic seats.

4.2 Satisfiable vote-band rules

Too loose of a vote-band rule will impose no constraint at all on redistricting; too tight of a rule will be impossible to meet. In this section, we examine the level of constraint imposed by $(y, 50)$ and (y, D_0) rules. This effect is explored in Figures 4-5. Figure 4 shows how often each number of competitive or state-typical districts occurs in a neutral ensemble. This is

[14] In a similar vein, there is growing evidence that the MM score is not performing as advertised to signal meaningful partisan advantage [16]. Nonetheless, we use it here to highlight that competitiveness may not be independent of other popular partisan indicators.

intended to visualize the elasticity of the vote: a skinny and tall histogram indicates that moving the lines does not allow much change to the count of a certain type of district, while a short and fat histogram indicates more freedom to change the outcome by controlling the district lines. In Figure 5, the same data is presented in a new way to facilitate the interpretation of candidate rules. The left column shows the existence and proportion of Congressional plans in the ensemble that are $(x, y, 50)$ compressed, and the right column shows similar plots for the state Senate. The transition regions for each fixed x value are relatively small as a function of y; but it is precisely these transition regions in which a vote-band rule would have any impact. Outside these zones, a vote-band requirement is either vacuous or is so hard to satisfy that a large ensemble finds no examples at all. At a minimum, states must perform modeling of this kind using recent electoral patterns before adopting a numerically specific vote-band rule for competitiveness or typicality.

For all of the states, the neutral ensemble found Congressional plans with no $(5, 50)$ districts—that is, it is always possible to draw a plan with no districts at all between 45 and 55% partisan share. Even within a fixed state, different vote elasticity may be observed at different scales. Comparing the behavior of Georgia's Congressional ensemble to the state Senate ensemble shows that requiring a fixed proportion of seats to be competitive would impose a much more constraining condition on the Congressional districts than the state Senate.

4.3 Partisan impacts of vote-band winnowing

One way that vote-band rules might play out is that among contending plans, those that best satisfy the rules would be selected. To model this scenario, we now extract subsets of plans that are (x, y, z) compressed and compare the aggregate properties of these plans to the full ensemble distributions. Comparing the subset of an ensemble that has a certain property with the full set from which it was drawn is known as *winnowing*. This allows us to observe whether the competitive plans differ from the neutral ensemble in a systematic fashion under the partisan metrics.

We begin by computing the average number of Democratic seats in districting plans that are $(x, 5, 50)$ and $(x, 5, D_0)$ compressed, as a function of x. Figure 6 shows these results

for each state we analyzed. For states with a strong statewide party skew, increasing the number of districts close to the statewide average unsurprisingly tends to shift the seats outcome in favor of the party with the larger statewide percentage. In tilted states, however, vying for competitiveness can actually hurt the minority party. The Massachusetts state Senate results show an example of a potentially surprising consequence of enforcing a competitiveness constraint. For this ensemble, increasing the number of $(5, 50)$ districts increases the expected number of Democratic seats. This is because it is easier to create more seats with a Democratic share of roughly 55% by taking scarce Republican votes from the few Republican-leaning districts.

The case of Wisconsin ($D_0 \approx 50$) is particularly interesting, since a vote-band competitive rule for Congress imposes a Democratic advantage, while a similar rule for state Senate has a Republican lean.

Next, we compare the distributions of mean-median values on the winnowed sub-ensembles, shown in Figure 7. Again, the histograms show a range of different behaviors as the constraints are tightened; several cases are worth highlighting. For the case $z = 50$, both Wisconsin and Georgia, whose full ensemble averages are Republican-favoring, are forced towards the center as the constraint tightens, while Virginia, which was balanced in the full ensemble, becomes more Republican-favoring in both the Congressional and state Senate ensembles under progressively tighter winnowing. Contrast this behavior with Utah, where the winnowed Congressional ensemble leans approximately four points towards Democratic-favoring scores, while the state Senate ensemble looks even more extremely Republican-favoring.

4.4 Partisan impacts of heuristic optimization

While the previous trials investigated the properties of competitive plans that were observed in a neutral ensemble, laws mandating competitiveness will also incentivize mapmakers to *optimize* for competitive plans. Plans constructed with this aim in mind likely will be significantly different than those drawn without access to partisan information. To expose some of these differences, we use the partisan-aware optimization methods Opt1 and Opt2 (described in §3.3) to construct plans that have significantly more $(5, 50)$ competitive

districts than were found by the neutral ensemble.

To give an extremely crude sense of the success of these short runs with simple optimization heuristics, we compare our optimization runs to those expressly constructed for the 538 Atlas of Redistricting to minimize CPVI values (Table 3). The comparison is not perfect, since the underlying data, models, and evaluation mechanism differ.[15] Also, there are states in which one of the styles of optimization is a bad fit for the task at hand, like Opt2 in Massachusetts. The similarity between our results and theirs, however, suggests that our simple optimization heuristics have some power. An interesting direction for future work would be to use the precinct-sorting methods from [18] to get rigorous upper bounds on the possible number of competitive districts.

State, Type, (Total #)	538 Atlas of Redistricting	Neutral Max	Opt1 Max	Opt1 Mean	Opt2 Max	Opt2 Mean
MA Cong (9)	**4**	2	3	2.11	1	0.05
VA Cong (11)	**9**	7	8	7.33	**9**	6.61
WI Cong (8)	6	5	6	3.99	**8**	4.28
GA Cong (14)	7	8	12	9.41	**14**	10.68
UT Cong (4)	1	**2**	**2**	1.99	**2**	.51
MA Sen (40)	–	12	**17**	15.35	9	5.12
VA Sen (40)	–	17	26	23.63	**28**	20.32
WI Sen (33)	–	17	20	17.33	**25**	15.83
GA Sen (56)	–	19	**37**	32.99	31	20.85
UT Sen (29)	–	9	**14**	11.42	13	6.93

Table 3: Comparison of the number of $(5, 50)$ districts found in the neutral and two styles of hill-climbing ensembles to the 538 Atlas of Redistricting's highly competitive plans. Maximum values highlighted in bold. In most cases, hill-climbing found plans where at least half the districts are $(5, 50)$ competitive, which differs significantly from the results of the neutral ensemble.

Comparing to the box plots of Figure 3 gives a measure of how far these optimized plans are from those in the neutral ensemble. For Georgia Congressional plans, the 1%-99% whiskers of the box plots only hit the $(5, 50)$ band for five out of 14 districts, but the optimization method found plans where all 14 districts are in the $(5, 50)$ band. Once again, we find Georgia to be particularly elastic in its vote distribution and to have enacted relatively uncompetitive plans. Note we have not made an effort to include the full suite

[15] For example, the CVPI values use a slightly Democratic-favoring baseline as a result of the popular vote in the 2016 presidential election, which makes it slightly easier to make competitive districts in Democratic-leaning states and slightly harder in Republican-leaning ones. Additionally, the 538 maps had constraints on the number of majority-minority districts.

of Georgia districting criteria in our model, and the Voting Rights Act in particular might be quite constraining.

Figure 8 shows the number of Democratic seats against the number of $(5, 50)$ districts found by the two optimization methods. As with our previous examples, the effects of promoting plan competitiveness are unpredictable at best, both within states or between Congressional and state Senate plans in the same state. The plot further highlights that different optimization methods targeting competitiveness can drastically change the partisan character of the plans that are found.

For Georgia's Congressional plan, higher numbers of competitive districts correlates with Republican seat advantage. Wisconsin's Senate plans show the opposite correlation, at least for these optimization techniques, tending to find more Democratic seats as the number of competitive districts increases. The optimization ensembles also found seats outcomes that did not appear at all in the significantly larger neutral ensembles, which is a reminder that outcomes that are extremely unlikely under partisan-blind circumstances may nonetheless be relatively easy to construct intentionally.

As with the winnowing results, these optimization runs show that the potentially extreme impacts of enforcing a competitiveness measure depend heavily and unpredictably on political geography and scale.

5 Conclusion

In the national conversation around redistricting reform, competitiveness is only one among numerous priorities, including traditional districting principles such as population balance or compactness as well as compliance with federal constraints like the Voting Rights Act, that are being debated and sometimes hastily written into law. Mapping out the interactions between various values, criteria, and constraints is becoming increasingly important. Data-intensive modeling is indispensable to make informed decisions about these interactions and to identify unintended consequences of quantitatively specific rules. We hope that reform measures for redistricting will call for this kind of modeling to be performed in every census cycle.

Ensemble analysis—and Markov chain sampling in particular—is a useful tool for ex-

ploring these complex spaces and intertwined constraints. The present study sets up a model methodology for surveying the state-specific partisan landscape in advance of setting quantitative thresholds and priorities for state reform. For the specific case of competitiveness, our results show that vote-band rules can be set in a way that does impose nontrivial constraints on redistricters, but sometimes with ancillary effects on other partisan metrics. There is every reason to expect nontrivial effects on the ability to uphold other priorities—from racial statistics to political boundary integrity—which were not considered here. A natural direction for future work is to determine how the operationalizations of state legislative language and other constraints impacts these analyses. Evaluating compliance with the Voting Rights Act of 1965 and the associated Gingles factors, derived from the 1986 Supreme Court ruling in Thornburgh v. Gingles and its successors, provides an example of the complexity of this modeling process and determining interactions between this constraint and competitiveness metrics is an important consideration for future state-specific analyses.

As with any other partisan quantity, securing a high score on a competitiveness metric does not follow effortlessly from the absence of nefarious intent. Democratic deliberation is called for both to clarify the type of competitiveness that is sought—more close outcomes? fewer landslides?—and to weigh this against other priorities. As states continue to experiment with their rules and frameworks for redistricting, we hope that statistically robust ensemble methods become standard tools in the repertoire.

References

[1] Mira Bernstein and Moon Duchin. A formula goes to court: Partisan gerrymandering and the efficiency gap. *Notices of the AMS*, 64(9):1020–1024, 2017.

[2] Robin E. Best, Shawn J. Donahue, Jonathan Krasno, Daniel B. Magleby, and Michael D. McDonald. Authors' Response-Values and Validations: Proper Criteria for Comparing Standards for Packing Gerrymanders. *Election Law Journal: Rules, Politics, and Policy*, 17(1):82–84, November 2017.

[3] Robin E. Best, Shawn J. Donahue, Jonathan Krasno, Daniel B. Magleby, and Michael D. McDonald. Considering the Prospects for Establishing a Packing Gerrymandering Standard. *Election Law Journal: Rules, Politics, and Policy*, 17(1):1–20, September 2017.

[4] Thomas L. Brunell. Rethinking Redistricting: How Drawing Uncompetitive Districts Eliminates Gerrymanders, Enhances Representation, and Improves Attitudes toward Congress. *PS: Political Science & Politics*, 39(1):77–85, January 2006.

[5] Wisconsin Legislative Technology Services Bureau. Open data portal, 2019.

[6] Aaron Bycoffe, Ella Koeze, David Wasserman, and Julia Wolfe. The atlas of redistricting. *FiveThirtyEight*, 2018.

[7] Jamie L. Carson and Michael H. Crespin. The effect of state redistricting methods on electoral competition in United States House of Representatives races. *State Politics & Policy Quarterly*, 4(4):455–469, 2004.

[8] Utah Automated Geographic Reference Center. Voting precincts, 2019.

[9] Jowei Chen and Jonathan Rodden. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8:239–269, 2013.

[10] Maria Chikina, Alan Frieze, Jonathan Mattingly, and Wesley Pegden. Practical tests for significance in Markov Chains. *arXiv:1904.04052 [math, stat]*, April 2019. arXiv: 1904.04052.

[11] Maria Chikina, Alan Frieze, and Wesley Pegden. Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences*, 114(11):2860–2864, March 2017.

[12] Utah Code. Chapter 7 Title 20A Chapter 19 Part 1 Section 103.

[13] David Cottrell. Using Computer Simulations to Measure the Effect of Gerrymandering on Electoral Competition in the U.S. Congress. *Legislative Studies Quarterly*, 44(3):487–514, 2019.

[14] James B. Cottrill. The Effects of Non-Legislative Approaches to Redistricting on Competition in Congressional Elections. *Polity*, 44(1):32–50, January 2012.

[15] Daryl DeFord and Moon Duchin. Redistricting reform in Virginia: Districting criteria in context. *Virginia Policy Review*, 12(2):120–146, 2019.

[16] Daryl DeFord, Moon Duchin, Natasha Dhamankar, Varun Gupta, Mackenzie McPike, Gabe Schoenbach, and Emily Sim. Implementing partisan symmetry: Problems and paradoxes. *Preprint*, 2019.

[17] Daryl DeFord, Moon Duchin, and Justin Solomon. Recombination: A family of Markov chains for redistricting. *ArXiv:1911.05725*, 2019.

[18] Moon Duchin, Taissa Gladkova, Eugene Henninger-Voss, Ben Klingensmith, Heather Newman, and Hannah Wheelen. Locating the representational baseline: Republicans in Massachusetts. *Election Law Journal*, 18, 2019.

[19] Benjamin Fifield, Michael Higgins, Kosuke Imai, and Alexander Tarr. A new automated redistricting simulator using Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics (Forthcoming)*, 2020.

[20] Richard Forgette, Andrew Garner, and John Winkle. Do Redistricting Principles and Practices Affect U. S. State Legislative Electoral Competition? *State Politics & Policy Quarterly*, 9(2):151–175, 2009.

[21] Corbett A. Grainger. Redistricting and Polarization: Who Draws the Lines in California? *The Journal of Law & Economics*, 53(3):545–567, 2010.

[22] John A Henderson, Brian Hamel, and Aaron Goldzimer. Gerrymandering incumbency: Does non-partisan redistricting increase electoral competition? *Journal of Politics*, 80, 2018.

[23] Gregory Herschlag, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier, and Jonathan C. Mattingly. Quantifying Gerrymandering in North Carolina. *arXiv:1801.03783 [physics, stat]*, January 2018. arXiv: 1801.03783.

[24] Gregory Herschlag, Robert Ravier, and Jonathan C. Mattingly. Evaluating Partisan Gerrymandering in Wisconsin. *arXiv:1709.01596 [physics, stat]*, September 2017. arXiv: 1709.01596.

[25] New Jersey Legislature. Senate Concurrent Resolution 43.

[26] MIT Election Science Data Labs. Election data, 2019.

[27] Georgia Legislative and Congressional Reapportionment Office. Statewide voting precincts, 2019.

[28] Michael D. McDonald and Robin E. Best. Unfair Partisan Gerrymanders in Politics and Law: A Diagnostic Applied to Six Cases. *Election Law Journal: Rules, Politics, and Policy*, 14(4):312–330, November 2015.

[29] Eric McGhee. Rejoinder to Considering the Prospects for Establishing a Packing Gerrymandering Standard. *Election Law Journal: Rules, Politics, and Policy*, 17(1):73–82, October 2017.

[30] MGGG. Mggg-States GitHub Repository, 2019.

[31] Peter Miller and Bernard Grofman. Redistricting Commissions in the Western United States. *UC Irvine Law Review*, 3(3):637, August 2013.

[32] Massachusetts Secretary of the Commonwealth Elections Division. Election results, 2019.

[33] Princeton Gerrymandering Project. Openprecincts.org, 2019.

[34] Arizona State Constitution. Article IV Part 2 Section 1.

[35] Colorado State Constitution. Article V section 44.

[36] Missouri State Constitution. Article III Section 3.

[37] Washington State Constitution. Article III Section 43.

[38] Nicholas Stephanopoulos and Eric McGhee. Partisan gerrymandering and the efficiency gap. *University of Chicago Law Review*, page 831900, 2014.

[39] Ellen Veomett. Efficiency gap, voter turnout, and the efficiency principle. *Election Law Journal*, 17.

[40] VRDI. Gerrychain. *GitHub Repository: mggg/gerrychain*, 2018.

[41] Samuel S.-H. Wang. Three Practical Tests for Gerrymandering: Application to Maryland and Wisconsin. *Election Law Journal: Rules, Politics, and Policy*, 15(4):367–384, October 2016.

[42] David Wasserman and Ally Flinn. Introducing the 2017 Cook Political Report Partisan Voter Index. 2013.

[43] New York State Constitution. Article III Sections 4 and 5.

	Democratic seats					Mean-median				
State, Type (Total #)	Full Ensemble	x_1	x_2	x_3	x_4	Full Ensemble	x_1	x_2	x_3	x_4
MA Cong (9)	9.00	9.00	9.00	9.00	—	0.021	0.021	0.015	0.0094	—
VA Cong (11)	6.35	6.37	6.45	6.66	6.67	−0.001	0.001	0.002	0.003	0.005
WI Cong (8)	2.66	2.65	2.67	2.80	3.28	0.044	0.044	0.039	0.026	0.014
GA Cong (14)	4.30	4.30	4.30	4.31	4.30	0.031	0.028	0.024	0.016	0.009
UT Cong (4)	0.85	0.85	0.70	0.70	—	0.018	0.018	−0.004	−0.014	—
MA Sen (40)	38.36	38.38	38.41	38.48	38.59	0.022	0.022	0.021	0.020	0.018
VA Sen (40)	21.39	21.38	21.33	21.24	21.16	0.009	0.010	0.011	0.014	0.016
WI Sen (33)	11.65	11.65	11.63	11.59	11.46	0.044	0.044	0.042	0.037	0.034
GA Sen (56)	21.86	21.86	21.86	21.85	21.82	0.042	0.041	0.038	0.032	0.024
UT Sen (29)	7.21	7.22	7.23	7.23	7.22	0.049	0.049	0.049	0.050	0.051

Table 4: Mean partisan statistics (Dem seats, mean-median) in the successively winnowed ensembles shown in Figure 6-7 for $z = 50$. The x_i values correspond to those in the figures. For example, the table shows the effects of requiring 6, 7, 8, or 9 competitive districts out of 40 in the MA State Senate, which corresponds to the third figure of the top row of Figure 6.

	Democratic seats					Mean-median				
State, Type (Total #)	Full Ensemble	x_1	x_2	x_3	x_4	Full Ensemble	x_1	x_2	x_3	x_4
MA Cong (9)	9.00	9.00	9.00	9.00	9.00	0.021	0.021	0.020	0.020	0.015
VA Cong (11)	6.35	6.45	6.61	6.76	6.97	−0.001	−0.002	−0.004	−0.004	−0.003
WI Cong (8)	2.66	2.65	2.65	2.77	3.24	0.044	0.044	0.040	0.027	0.015
GA Cong (14)	4.30	4.18	4.04	3.85	3.61	0.031	0.028	0.026	0.023	0.021
UT Cong (4)	0.85	0.85	0.83	0.86	0.19	0.018	0.018	0.021	0.021	0.016
MA Sen (40)	38.36	38.36	38.33	38.29	38.25	0.022	0.022	0.021	0.020	0.019
VA Sen (40)	21.39	21.44	21.59	21.83	22.12	0.009	0.009	0.009	0.008	0.008
WI Sen (33)	11.65	11.64	11.63	11.57	11.46	0.044	0.044	0.042	0.039	0.035
GA Sen (56)	21.86	21.78	21.59	21.28	20.86	0.042	0.041	0.037	0.033	0.029
UT Sen (29)	7.21	7.20	7.18	7.16	7.11	0.049	0.047	0.045	0.041	0.036

Table 5: Mean partisan statistics (Dem seats, mean-median) in the successively winnowed ensembles shown in Figure 6-7 for $z = D_0$. The x_i values correspond to the successively smaller color-coded plots in the figures. For example, the table shows the effects of requiring 10, 12, 14, or 16 state-typical districts out of 40 in the MA State Senate, which corresponds to the fourth figure of the top row of Figure 6.

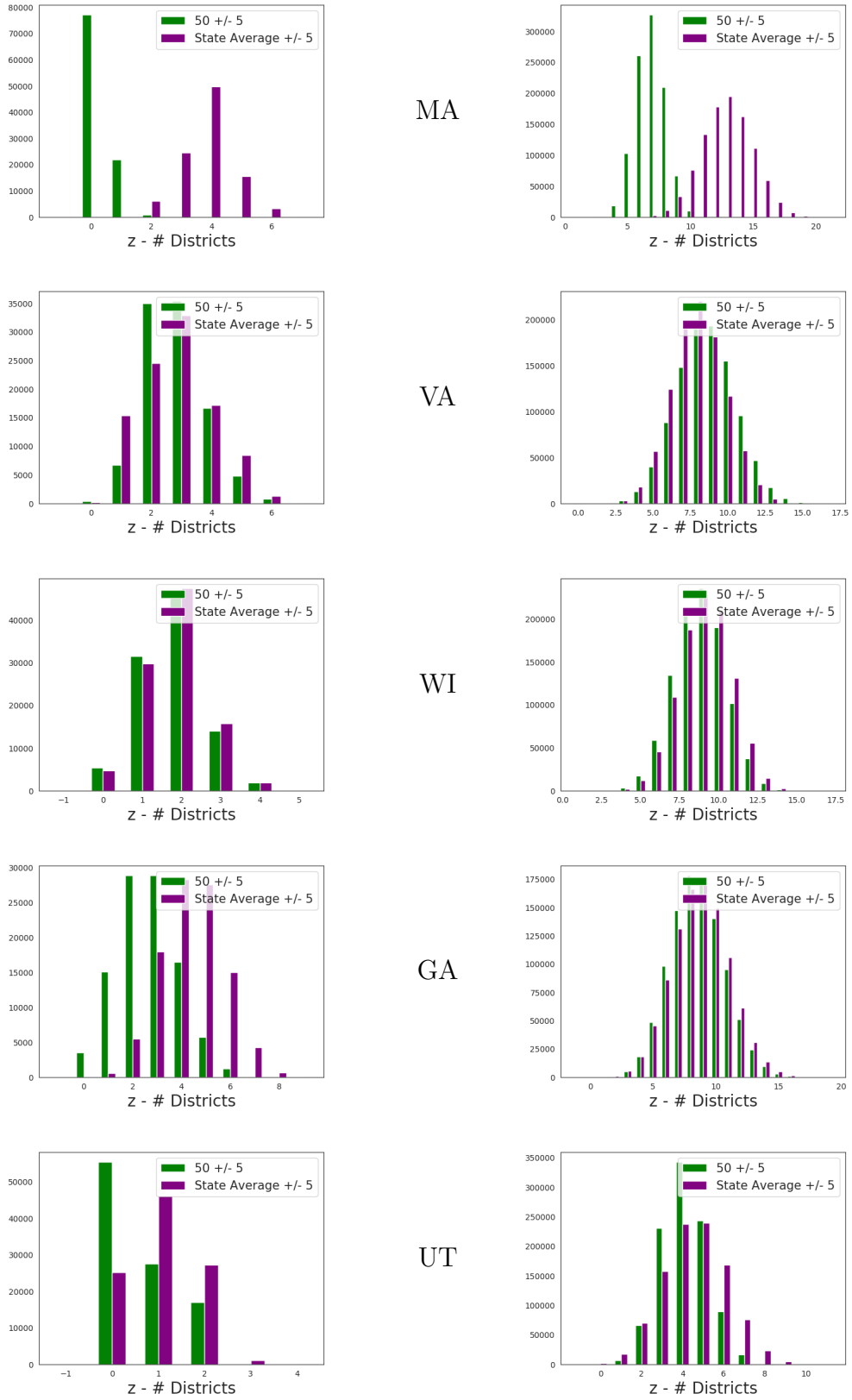


Figure 4: How often each number of $(5, 50)$ and $(5, D_0)$ districts was observed in the neutral ensemble. Scale effects are notable in Georgia especially, where the competitiveness and state-typicality statistics are the same for state Senate but noticeably different for Congress.

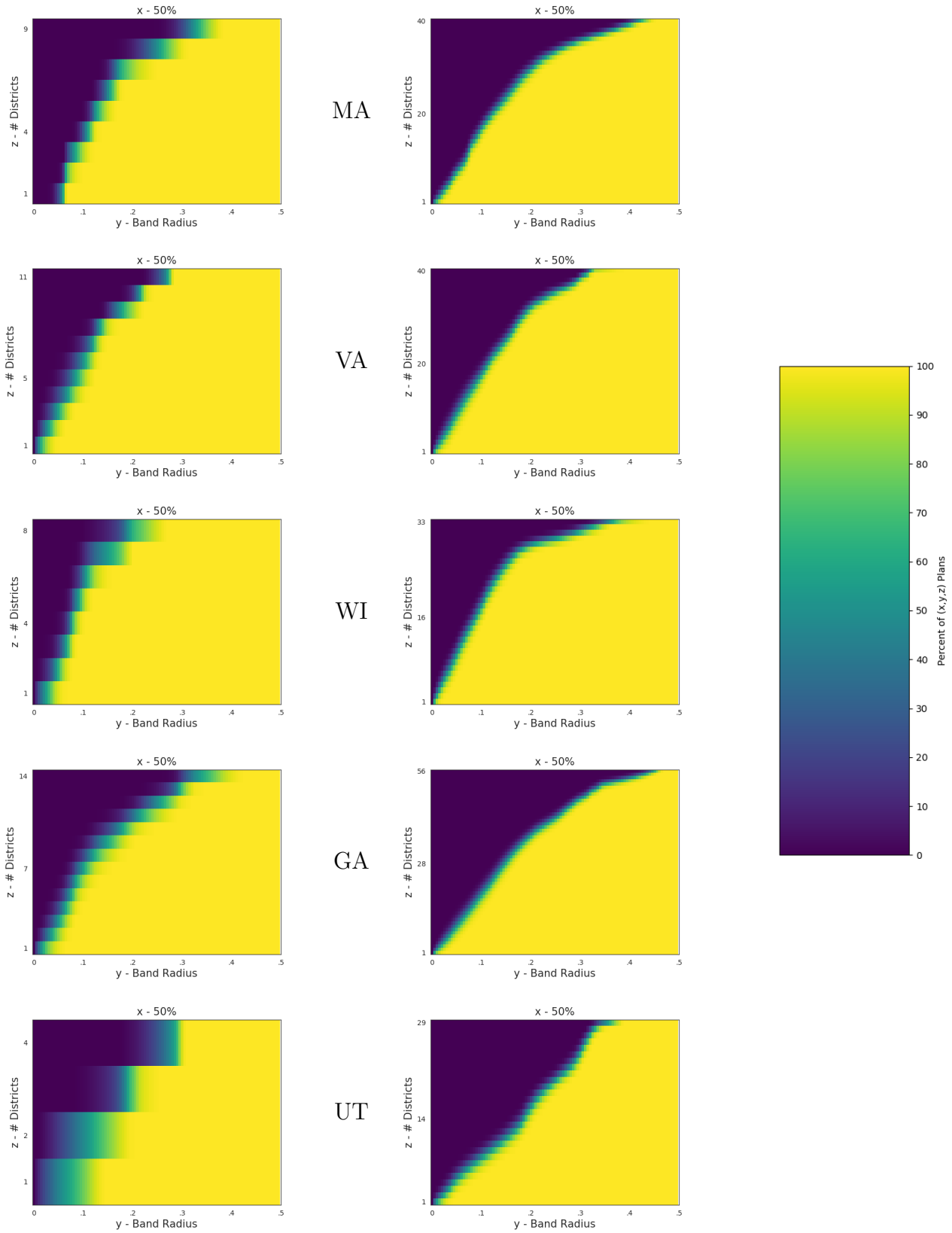


Figure 5: Visualizing the narrow range of vote-band rules that meaningfully constrain a neutral ensemble. Congress (left), state Senate (right).

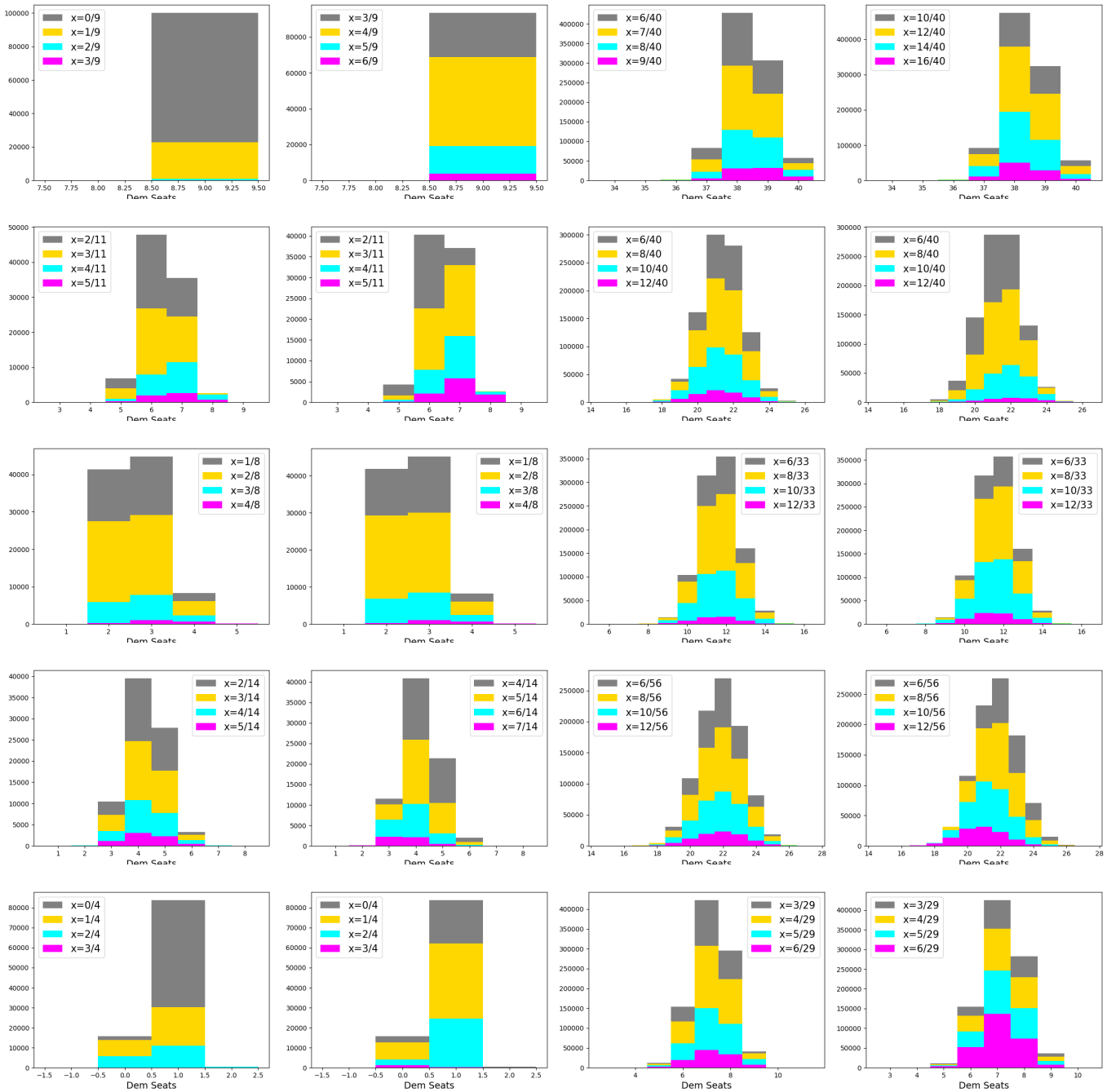


Figure 6: Histograms of Democratic seat outcomes in (x, y, z) compressed plans as a function of increasing x. Columns 1-2: Congress, $z = 50$ and $z = D_0$. Columns 3-4: state Senate, $z = 50$ and $z = D_0$. States, top to bottom: MA, VA, WI, GA, UT.

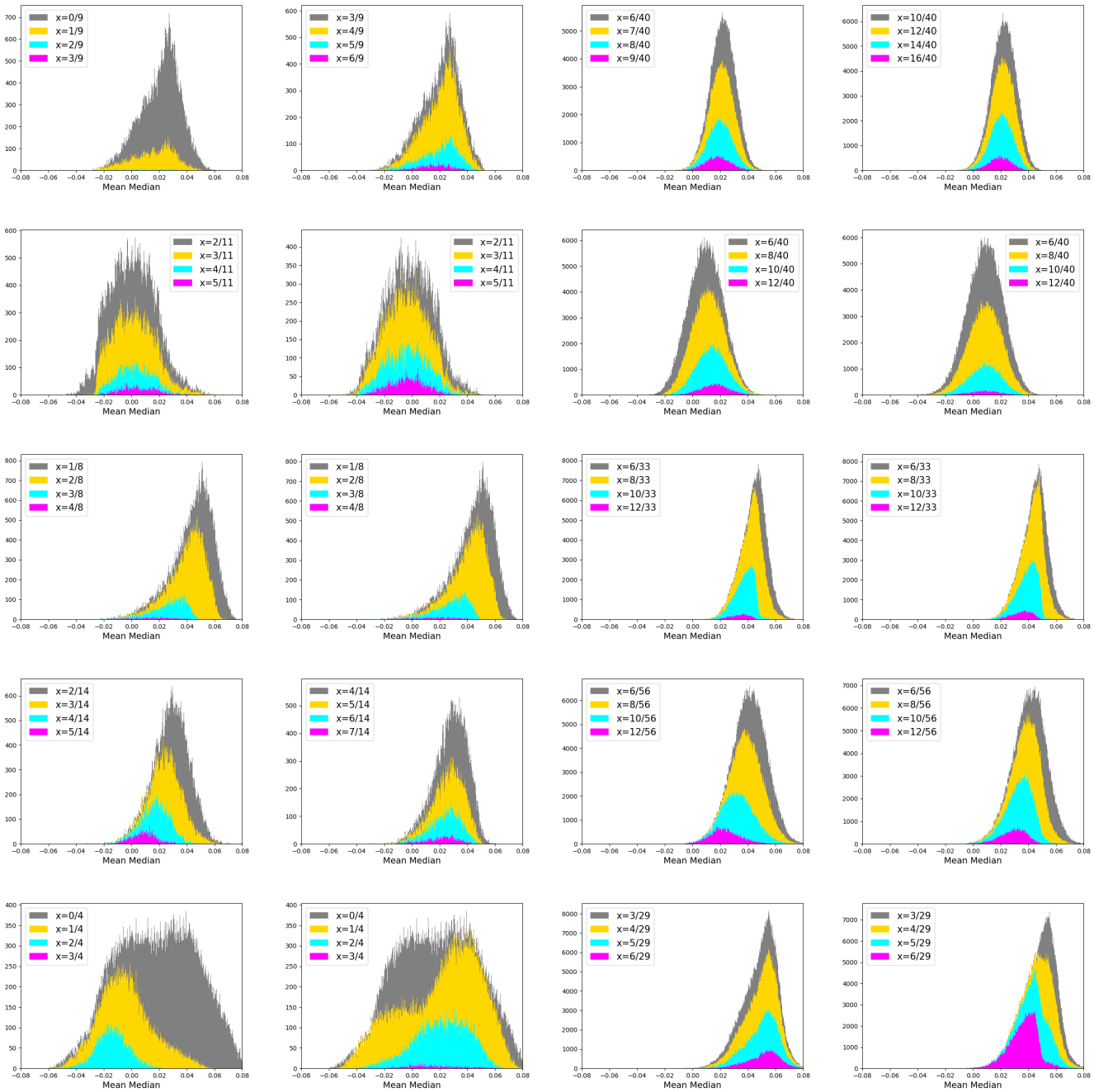


Figure 7: Successively winnowed mean-median distributions for $(x, 5, z)$ compressed plans as a function of increasing x. This figure shows how the MM score reacts as more districts are required to be competitive or state-typical. (The grey histogram is the full ensemble.) Columns 1-2: Congress, $z = 50$ and $z = D_0$. Columns 3-4: state Senate, $z = 50$ and $z = D_0$. States, top to bottom: MA, VA, WI, GA, UT. Recall that $MM = 0$ is thought to signal symmetry between the parties, and positive values are thought to signify Republican advantage. One thing to note is that all states except Virginia show a baseline of pronounced Republican advantage by this measure in the neutral ensemble (i.e., with no partisan gerrymandering).

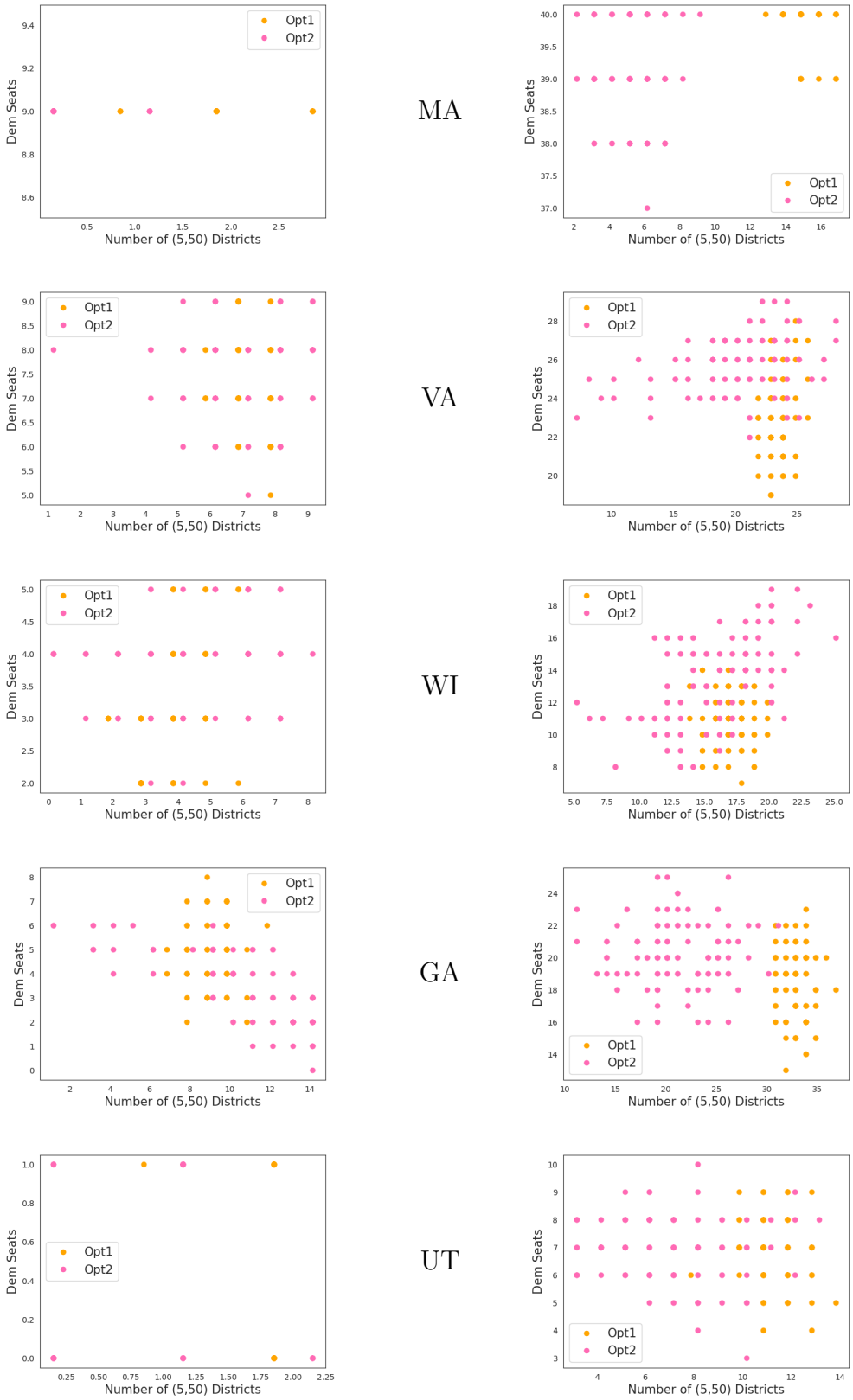


Figure 8: Number of Democratic seats in the plans found by the two optimization methods plotted against the number of $(5, 50)$ districts. The two types of points are slightly displaced so that they are visible in the plot.

SAMUEL S.-H. WANG
Professor
+1 609 258 0388 (tel)
+1 609 258 1028 (fax)
sswang@princeton.edu

Arizona Independent Redistricting Commission
Tucson, AZ

August 5, 2021

To the Arizona IRC:

Thank you for the opportunity to present information regarding ways to measure competitiveness. It was a pleasure to engage with you and with the experts present on August 3rd.

The Arizona Independent Redistricting Commission ("IRC") is charged "[t]o the extent practicable" with crafting "competitive districts . . . where to do so would create no significant detriment to the other goals." Az. Const. Art. 4, Sec. 1(14). Notably, the criteria the IRC is charged with utilizing are not rank-ordered: one does not take priority over another.[1] The IRC must draw competitive districts, if the other criteria will not be significantly harmed. *Ariz. Minority Coalition v. AIRC*, 211 Ariz. 337, 354 (App. 2005). Thus, the IRC must both evaluate competitiveness, and enhance competitiveness where possible, in order to comply with its constitutional mandate.

Within this requirement, however, the IRC has considerable latitude in both defining competitiveness and determining which measures it uses to test compliance. *Ariz. Minority Coalition for Fair Redistricting v. AIRC*, 220 Ariz. 587, 599, n.14. Although meeting the other criteria precludes making every single district competitive, it is nonetheless possible to maximize the number of competitive districts. The question is then how to define "competitive."

[1] Compare Arizona's law, which contains no language suggesting priority, with the criteria lists in California, Cal. Const. Article XXI, Sec. 2(d) ("The commission shall establish single-member districts for the Senate, Assembly, Congress, and State Board of Equalization pursuant to a mapping process using the following criteria **as set forth in the following order of priority**:"); Michigan, Mich. Const. Article 6, Sec. 13 ("The commission shall abide by the following criteria in proposing and adopting each plan, **in order of priority**:"); and Missouri, Mo. Stat. 260.390(b) ("The house independent bipartisan citizens commission shall redistrict the house of representatives using the following methods, listed **in order of priority**:) (emphases added).

When settling on an approach to define competitiveness, we recommend two basic principles: simplicity and transparency. These qualities allow your internal staff to do the calculation quickly, and for the public to understand what you are doing. Transparency will maximize trust and reduce post-redistricting criticism, which can be intense.

Two approaches to defining competitiveness for a hypothetical district are possible: (1) its expected **average performance**, and (2) without averaging, assessing how it would perform in individual past elections, to determine whether it is a **responsive district**.

(1) Average performance. The averaging approach requires evaluating a hypothetical district's average performance for the two major political parties[2] to see if it falls within a "swing" band around 50/50. The IRC may define the size of that band. The Princeton Gerrymandering Project uses a 7% band (from 46.5% to 53.5% for either party) in its fair-districting dashboard to define a competitive district. Some political analysts use a wider band, 5% on either side of the 50/50 line.

To determine the average performance of a district, the IRC should use past voting history, not party registration. Increasingly, people can vote for a party without being registered to that party, as Dr. Eric McGhee said during his August 3rd presentation. Political science research has shown that even nonaffiliated voters are usually reliable in their voting habits. To calculate average performance, then, you would instruct your staff to take the real outcomes of a variety of statewide elections (*e.g.* president, governor, Senate, Corporation commissioner) - a "market basket" to reflect the average behavior of Arizona voters. If you go back 10 years, you can find many elections to fill this basket.

(2) Responsive districts. It is possible to define competitiveness without averaging. Instead, you can ask whether, in any of the market basket of past elections, the district would have ever shifted between parties. This approach is akin to that presented by Moon Duchin at the August 3rd presentation. Prof. Duchin suggested that the IRC look at performance of proposed districts for individual elections over the past decade separately, rather than averaged. This approach aligns neatly with the approach adopted by the Colorado Independent Redistricting Commission, which explicitly adopted a definition of competitiveness that looks to whether districts have "a reasonable potential for the party affiliation of the district's representative to change at least once between federal decennial censuses." Colo. Const. Art. V, Sec. 48.1 3(d). To make this determination, Colorado explicitly recommends looking at "a proposed district's past election results," *id*., among other criteria.

With either approach, the IRC has the option of drawing more or fewer competitive districts. At this point the IRC must balance against other priorities, which will end up leading to some drawn districts being safe for one party or the other. Those other districts might comply with other statutory requirements, such as compactness, avoiding splitting political subdivisions, or ensuring the ability to elect communities or specific ethnic groups. These tradeoffs may

[2] As discussed during the presentation on August 3, 2021, though many political independents are not registered with either political party, their voting patterns suggest a stronger loyalty to the two parties than their independent status would otherwise mandate.

require you to grade yourselves using a "report card" of geographic, partisan, racial, and competitive fairness.

Which elections to use? A market basket approach. In both approaches, it is necessary to decide which past election results to put in the "market basket." We recommend against using district elections, which can depend on specific candidates and issues. Indeed, many Arizona legislative elections even have candidates from only one political party, making it impossible to identify how people would have voted if given a choice. Instead, it is better to look at statewide elections. In Arizona, such elections include the Presidential races in 2012, 2016 and 2020, the Senate races in 2012, 2016, 2018, and 2020, the Governor's races in 2014 and 2018, and 13 Corporation Commissioner races in 2012, 2014, 2016, 2018, and 2020. These 22 elections all have different advantages: Presidential races indicate a broad sense of preference when party affiliation is the overriding factor in a voting decision, Senate and Governor races consider Arizona-specific preferences, and Corporation Commissioner can indicate partisanship when candidates' personal characteristics are less prominent.

Whether using the averaging approach or the responsive-district approach, the calculation of hypothetical outcomes is well within the capabilities of the team at NDC and Timmons. They can count each election once, or they can combine all races for a given year, i.e. 2012, 2014, 2016, 2018, and 2020 would each count once. Other combinations are possible too. Your staff can show you how different past-election data would gauge current district maps.

In summary, we recommend a concept that the other experts, Moon Duchin and Eric McGhee, and we, support: looking to past elections to determine how a proposed district would perform. The averaging approach is simple once you agree on a market basket of races, and a percentage threshold for "competitive." The responsive-district approach is equally simple.

We thank you again for the opportunity to offer our thoughts on redistricting in Arizona. We stand ready to answer any further questions you may have.

Yours sincerely,

Prof. Sam Wang
Director, Electoral Innovation Lab

Adam Podowitz-Thomas
Senior Legal Strategist, Electoral Innovation Lab

From: Eric McGhee
Date: Tue, Aug 3, 2021 at 9:28 PM
Subject: RE: AZ Competitiveness Panel follow up
To: Douglas Johnson

Thanks again for inviting me to share thoughts with the commission. Since the commissioners were looking for something really concrete, here's my recommendation for evaluating any given map:

1) for each of the last five elections (i.e., since the last redistricting) calculate the actual average U.S. house, state senate, and state house vote share (3 averages for each year)
2) calculate the *__overall__* average across the last 5 elections for U.S. house, state senate, and state house (3 averages total)
3) calculate the differences between #1 and #2 for each chamber
4) identify the largest absolute difference (i.e., either positive or negative) across all chambers and years

These first four steps will give a sense of the largest reasonable swing you are likely to get. That's background information that will be used to evaluate any plan. Then:

5) run the map through PlanScore; this will give you a prediction for each seat in the plan in an average election from the last decade
6) for each seat, subtract 50% from its predicted vote share
7) any seat where the absolute result of #6 is smaller than the result of #4 is a competitive seat

The intuition is that a seat is competitive if it can't withstand large partisan swings without flipping.

A few final thoughts:

- While the methodology of PlanScore may seem fancy, at its core it's just a more user-friendly version of JudgeIt. It boils down to predicting outcomes with the correlation between past state legislative and U.S. House votes and the presidential vote.
- I encourage the commissioners to consider lots of different options, but I would caution against using any statewide office with larger partisan swings that you get from #1-#4 above. That's a sign that the office doesn't have strong downballot coattails. (Gubernatorial races are especially notable for this. The governor of LA is a Democrat; the governor of MA is a Republican.) Volatility in a statewide race is only useful if it tells you something about how the districts are likely to perform.

Hope this is helpful! I'll be on vacation starting tomorrow until next Tuesday, but I will have internet access and could respond to something really pressing.

Cheers,
Eric
Board Member, PlanScore