OPTIMIZED DEMOCRACY…

MOON DUCHIN, NOV 10 2025

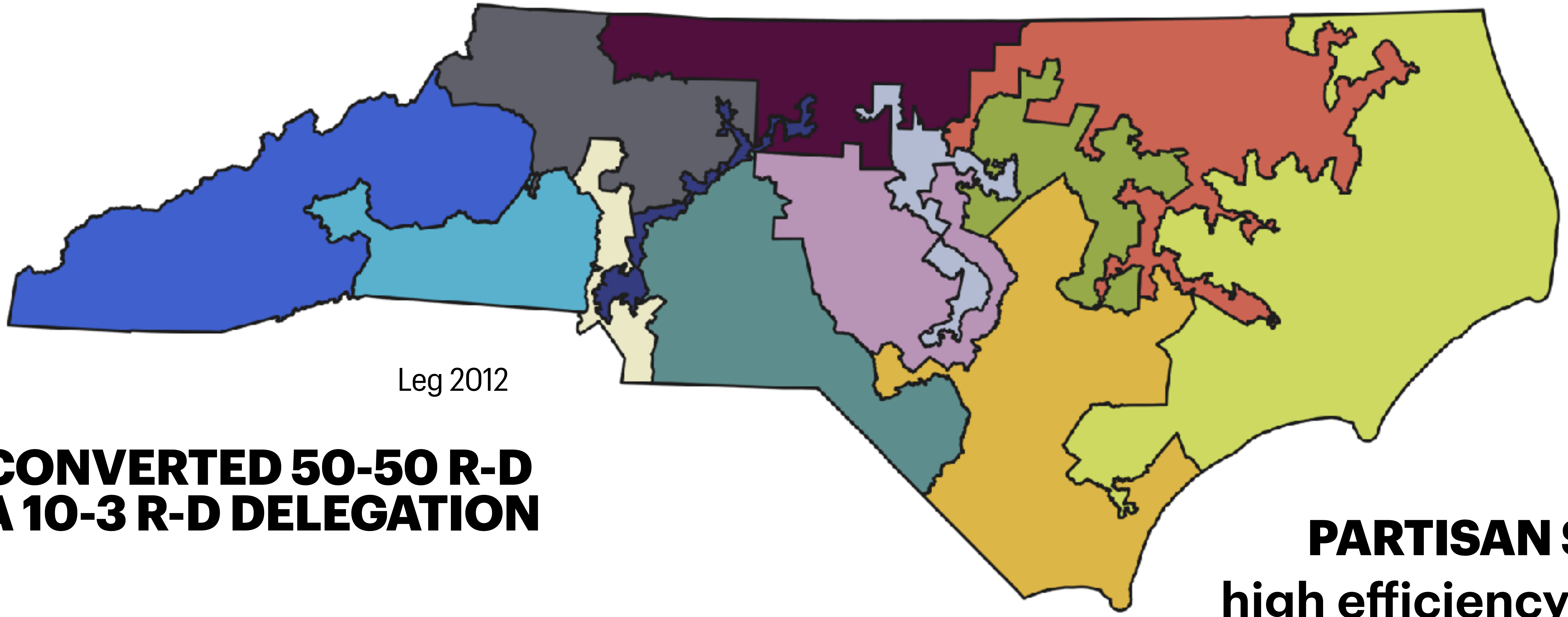
Optimized Democracy

...MODELING DEMOCRACY

MOON DUCHIN, NOV 10 2025

THE PUZZLE OF GERRYMANDERING

WHAT'S WRONG WITH THIS PLAN?



UGLY SHAPES

RELIABLY CONVERTED 50-50 R-D VOTES TO A 10-3 R-D DELEGATION

PARTISAN SCORES ARE BAD

high efficiency gap, mean-median gap, partisan symmetry scores, etc

TARGETED AFRICAN-AMERICANS "WITH ALMOST SURGICAL PRECISION"

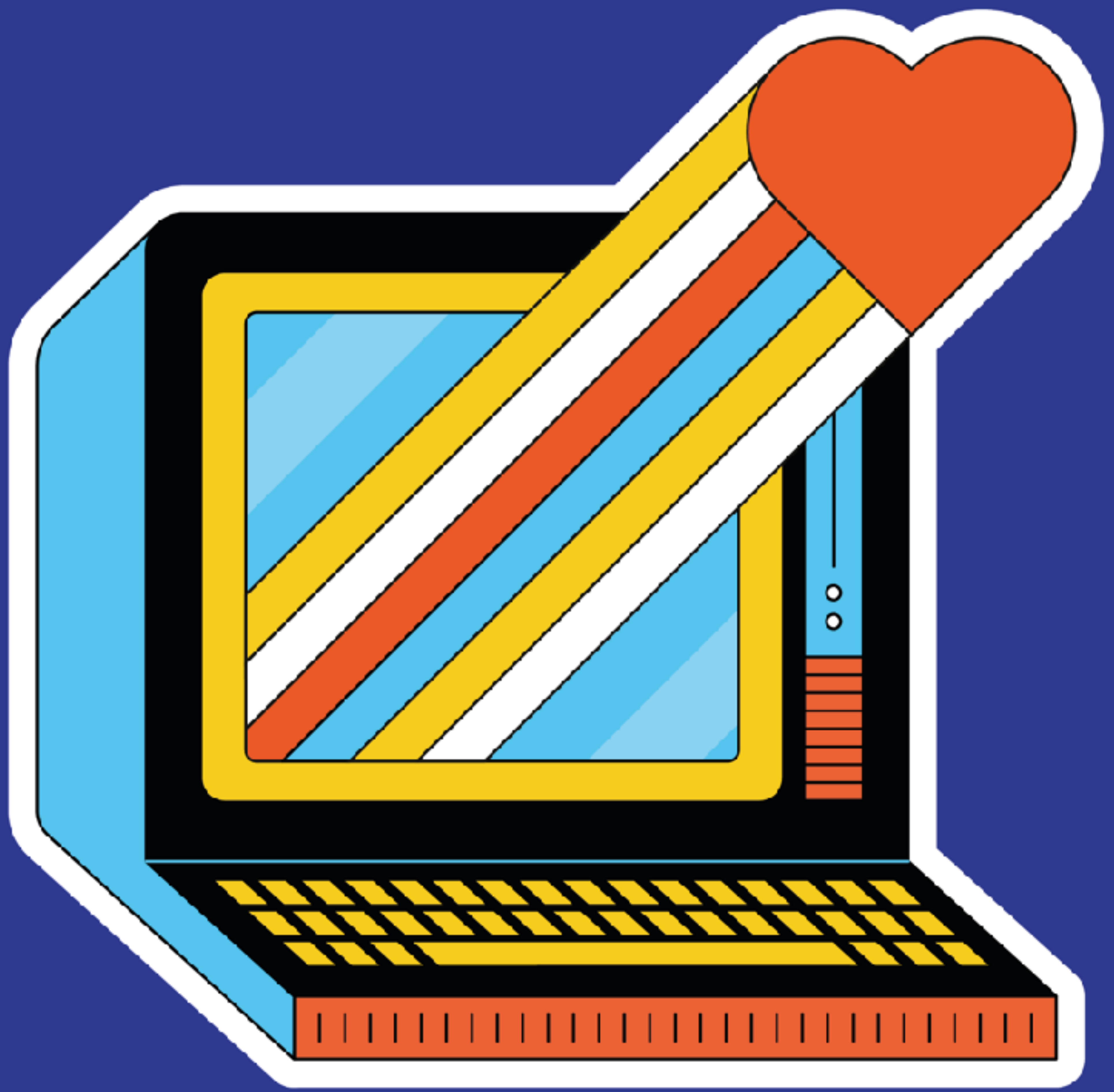
THE DREAM

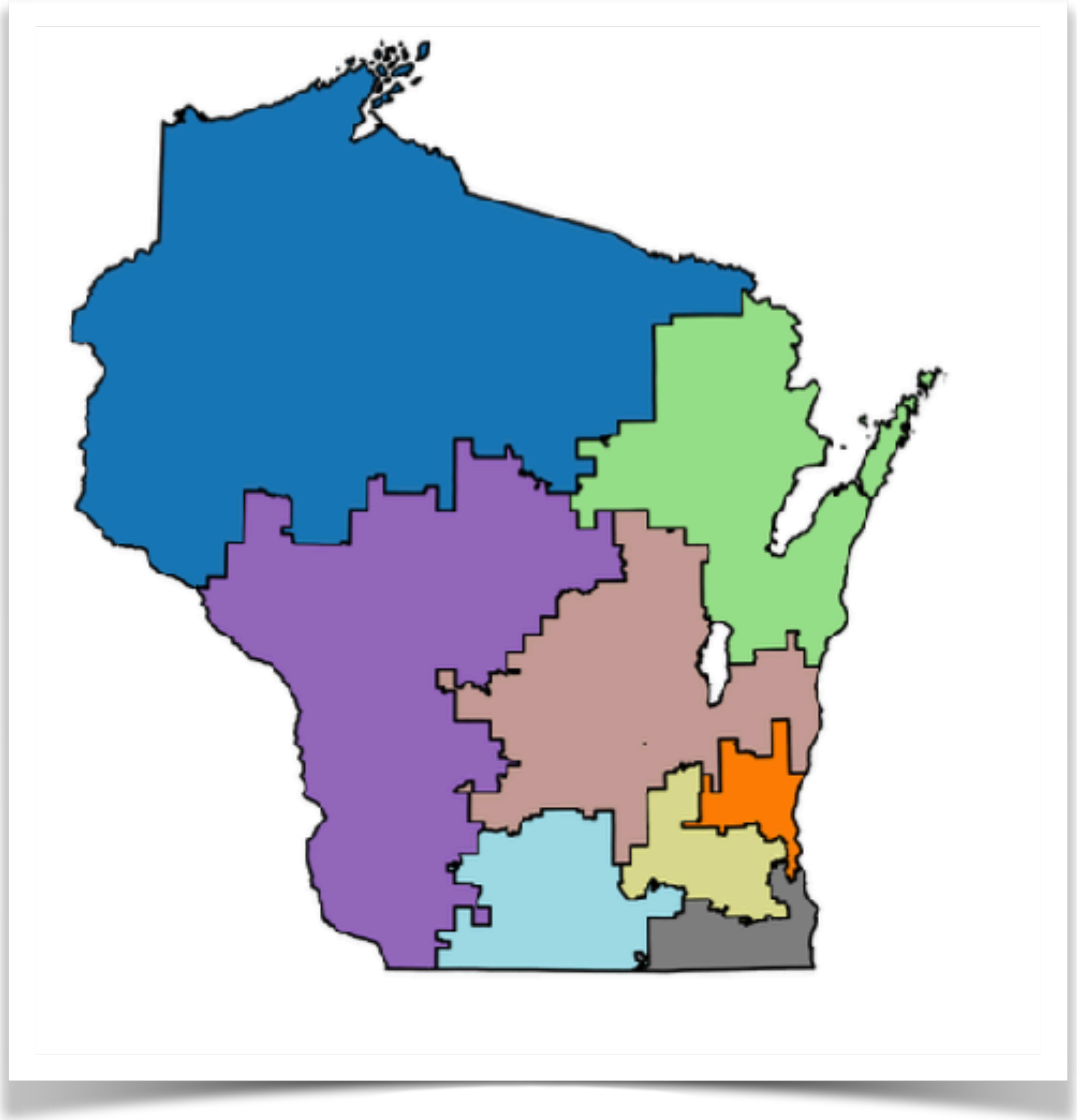
People have aimed to use computers for automated redistricting since at least the 1960s.

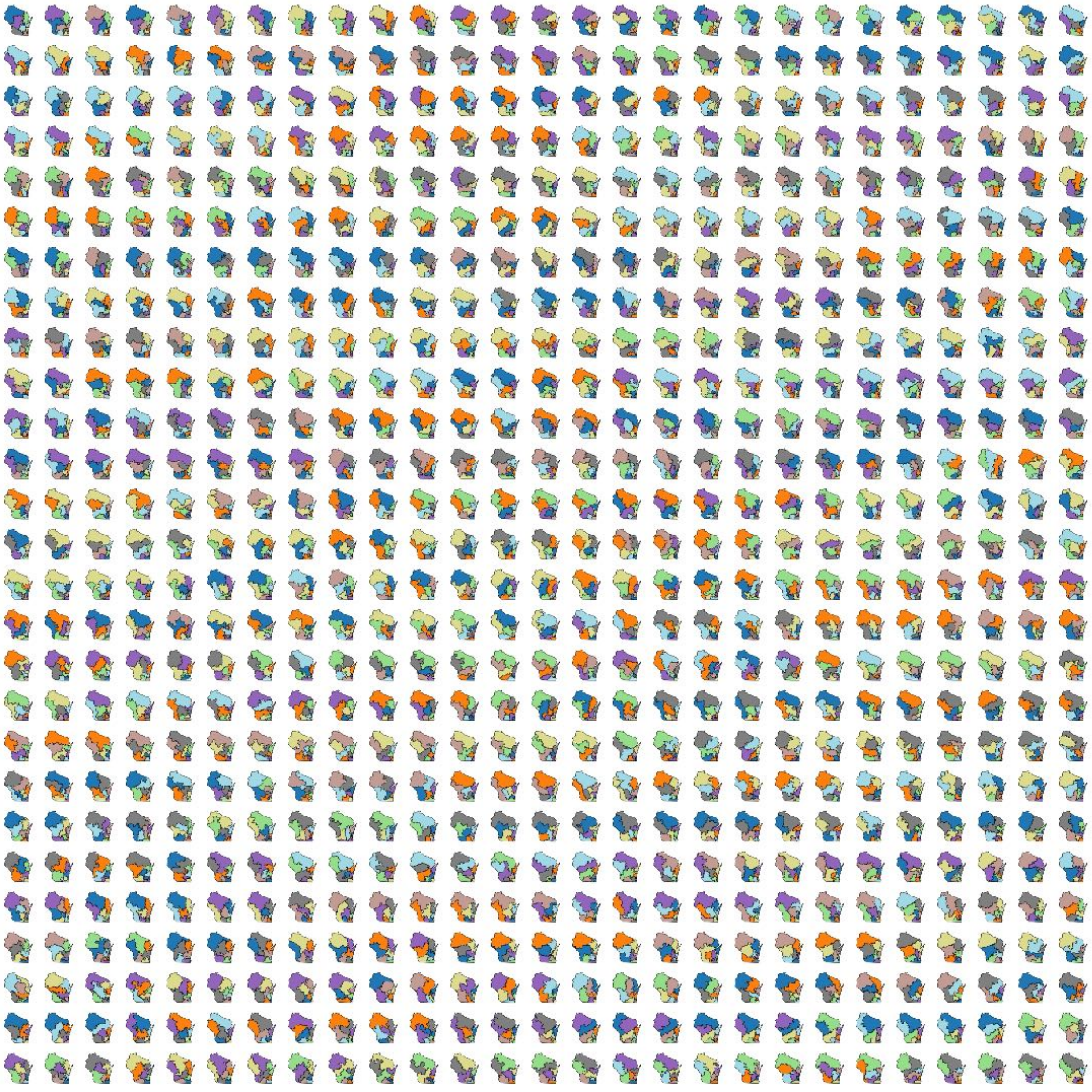
Original Idea: tell a computer the criteria for a good map. Let the computer figure out all the possibilities. Optimize.

Update: optimization stubbornly difficult. Not even clear we know the objective function.

New Idea: SAMPLE — Evaluate your plan in the context of the alternatives.







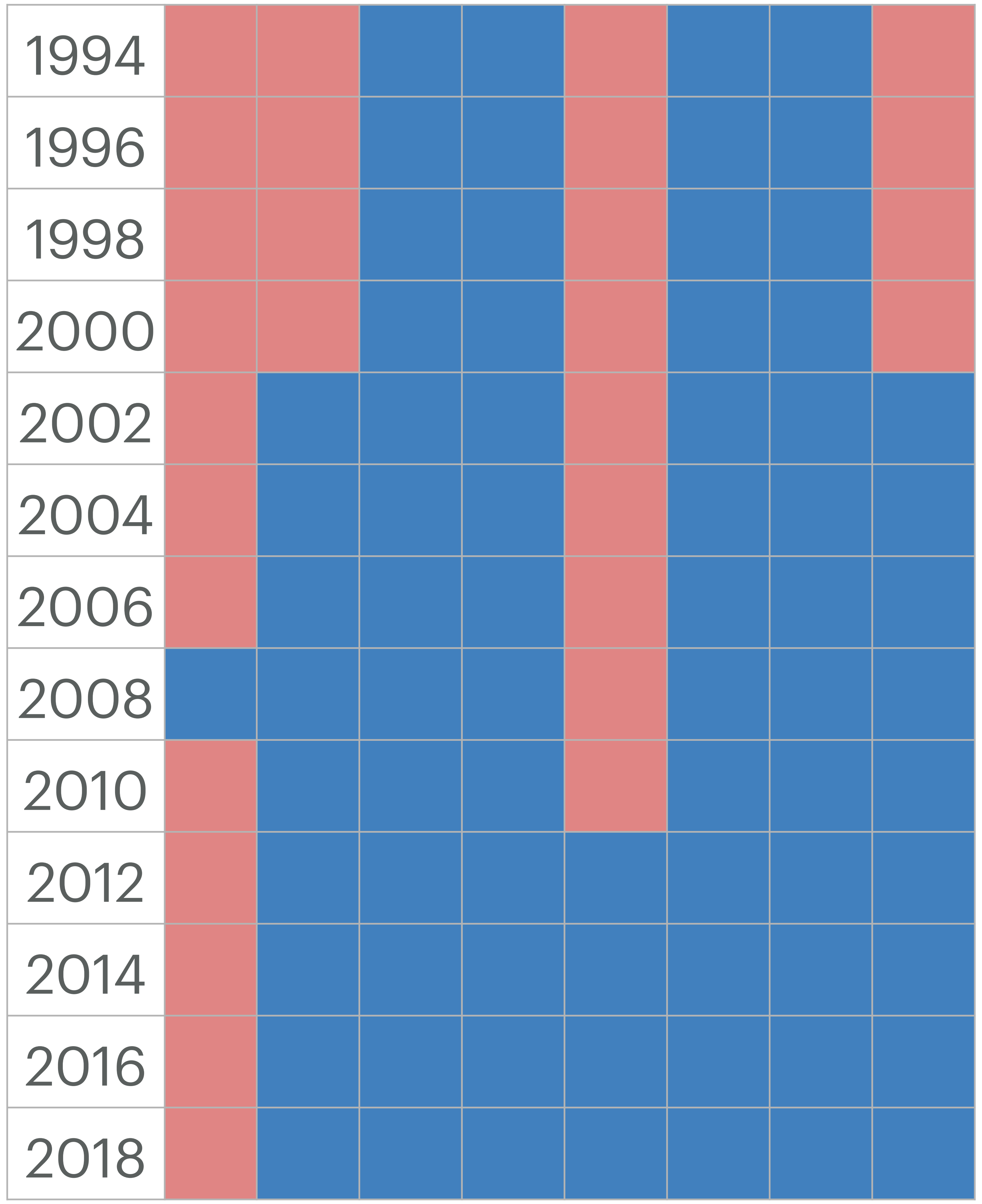
BUT FIRST…

THE SAD STORY OF REPUBLICANS IN MASSACHUSETTS

HOW BLUE IS...

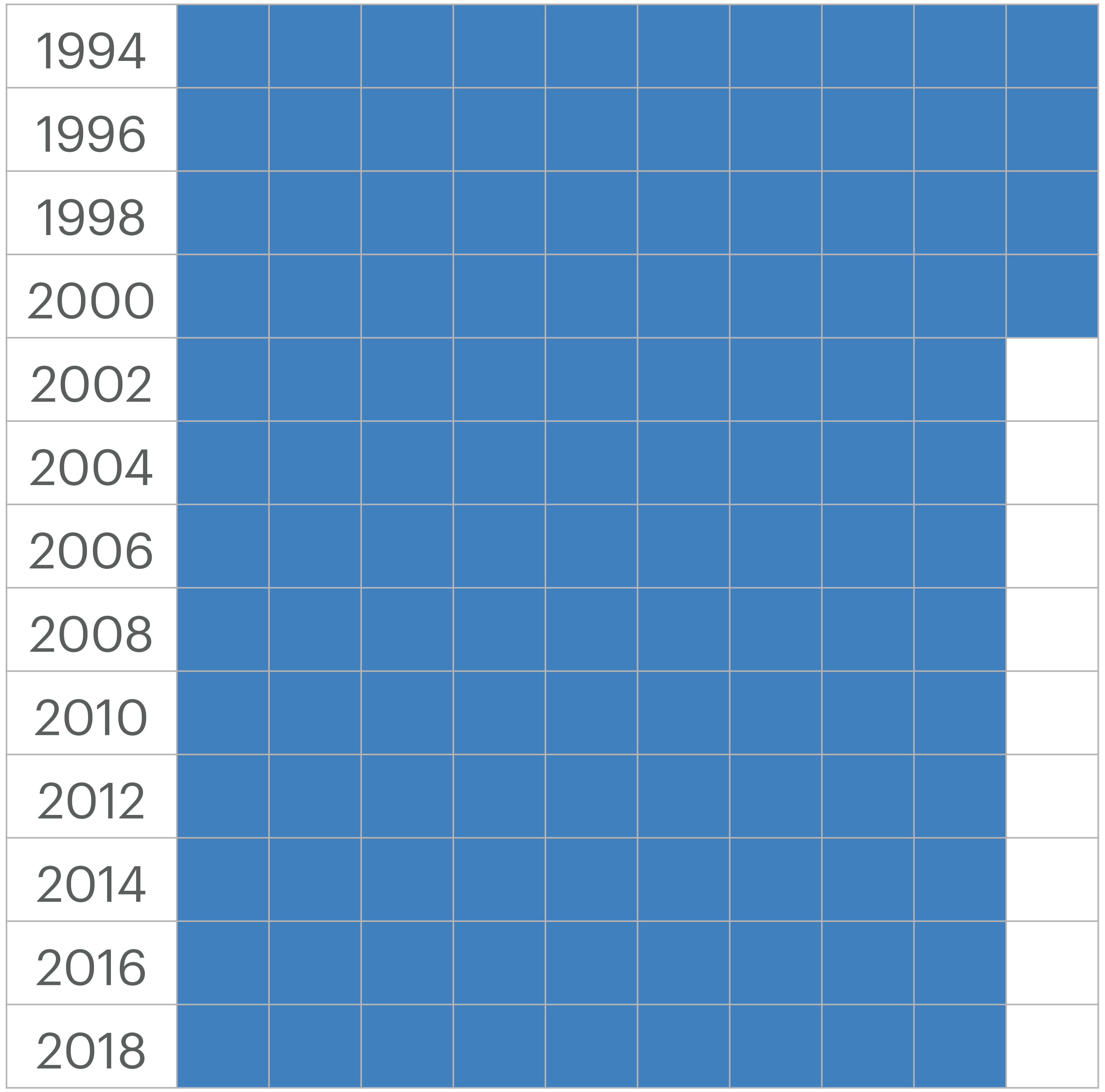
Maryland

36.9% R votes 28% R reps

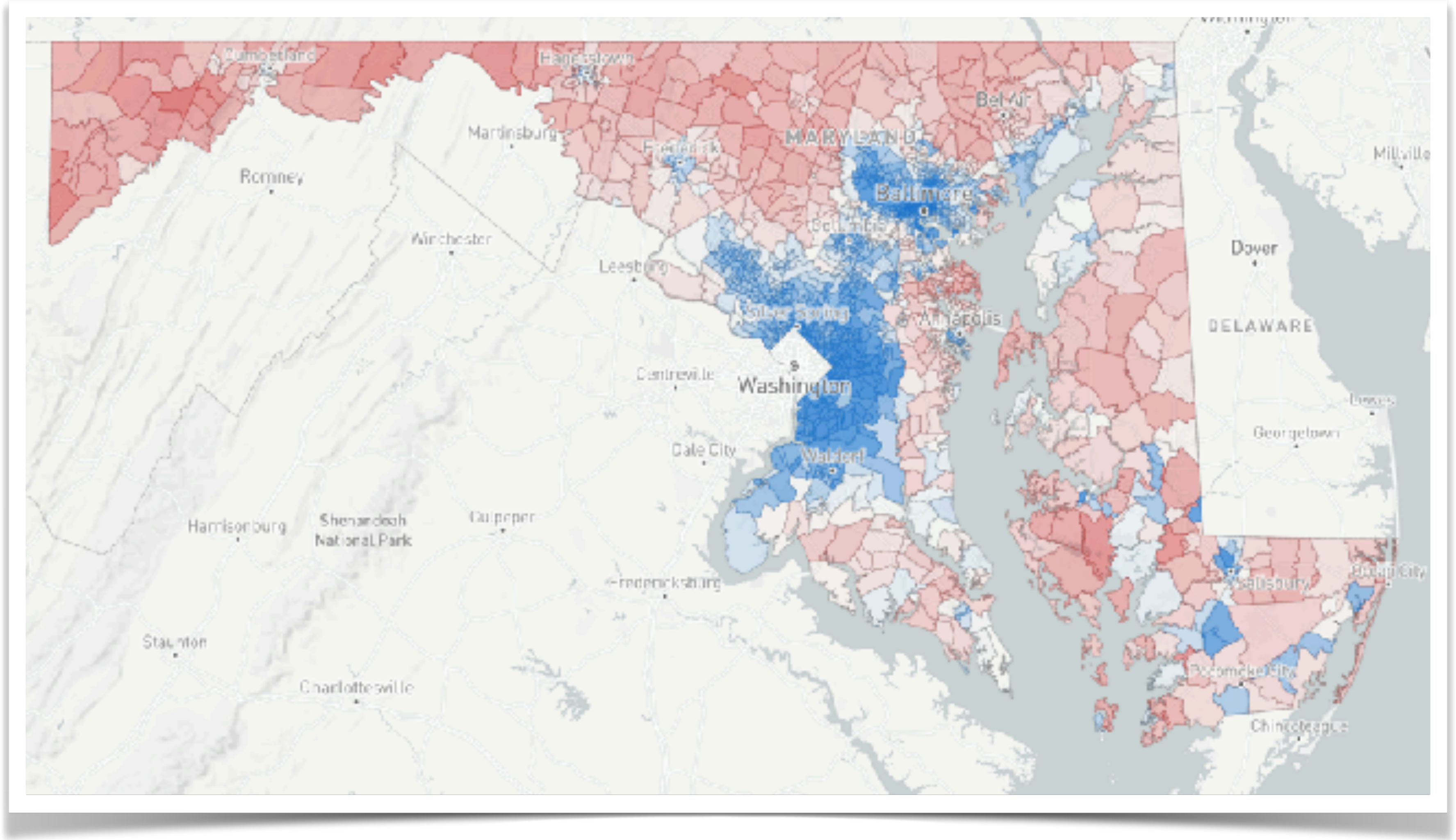


Massachusetts

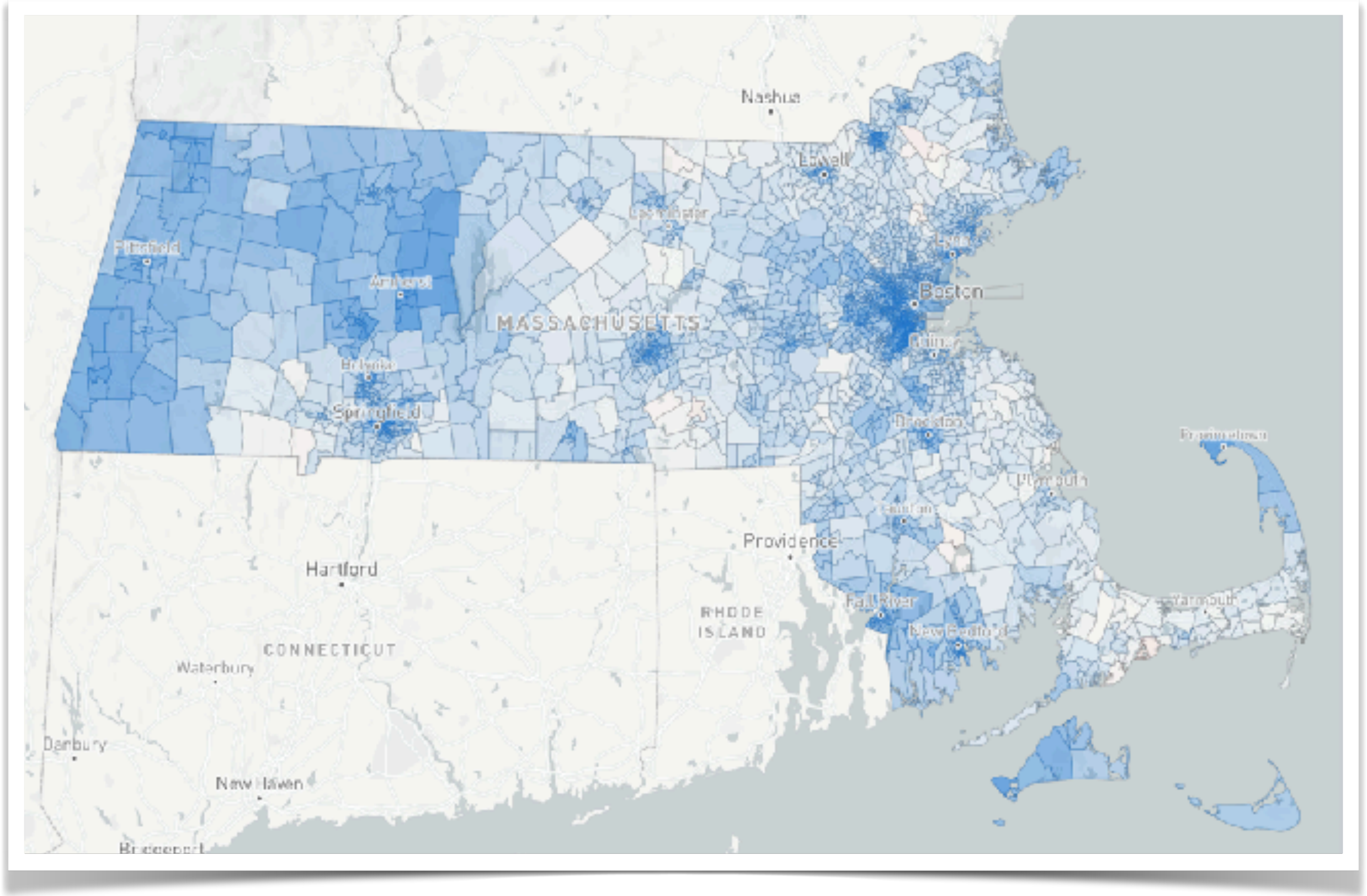
36.7% R votes 0% R reps



GERRYMANDERING? OR GEOMETRY?



this election has 34% R



this election has 32% R

REPUBLICAN FUTILITY THEOREM

There are more ways to redistrict MA than particles in the galaxy… and every possible plan gives a 9-0 Democratic sweep!

see "Locating the Representational Baseline" in Election Law Journal, 2019

Suppose you have a list of units with corresponding populations p_i and R margins $\delta_i = r_i - d_i$; the number of R votes minus the number of D votes. Re-index so that they are ordered from greatest to least by margin per capita:

$$\delta_1/p_1 \geq \delta_2/p_2 \geq \cdots \geq \delta_n/p_n.$$

We will call a collection of units S a *grouping*, and let $p(S)$ and $\delta(S)$ be its population and R margin, found by summing the p_i and δ_i for its units. Let D_k be the grouping indexed by $\{1, \ldots, k\}$. Let K be the smallest integer k for which $\delta(D_k) \leq 0$. This means that D_{K-1} has a collective R majority, but if you add the Kth unit you get a grouping D_K that fails to have an R majority.

Theorem 1. *With the notation above, let M be any positive integer.*

Case 1. $M \leq p(D_{K-1})$. There **exists** an R-majority grouping of size at least M.

Case 2. $p(D_{K-1}) < M \leq p(D_K)$. Inconclusive: such a grouping may or may not exist.

Case 3. $p(D_K) < M$. There **does not exist** an R-majority grouping of size at least M.

Proof. In Case 1, it is clear that a Republican grouping can be created, because D_{K-1} is a Republican-majority grouping of sufficient size.

We present examples to illustrate that Case 2 is inconclusive.

i	r_i	d_i	p_i	δ_i/p_i
1	8	0	8	1
2	1	9	10	$-4/5$
3	0	5	5	-1

i	r_i	d_i	p_i	δ_i/p_i
1	8	0	8	1
2	1	9	10	$-4/5$
3	0	8	8	-1

For both examples, fix $M = 13$. We have $K = 2$ in both examples because $\delta(D_1) = 8 > 0$ and $\delta(D_2) = 0$. Both fall under Case 2 because $p(D_1) = 8$ and $p(D_2) = 18$, while $M = 13$. In the left-hand example there exists an R-majority grouping, made by putting together units 1 and 3 to form a grouping with $\delta = 3$ and population 13. But in the right-hand example there is none, which is easily confirmed by considering all of the combinations.

Finally, in Case 3, we have $p(D_K) < M$.

Claim. *Let $S = D_K$ and suppose that $p(S) < M$. Then for any $S' \subseteq \{1, \ldots, n\}$,*

$$p(S') > p(S) \implies \delta(S') < \delta(S).$$

The claim asserts that D_K has the optimal R margin among all groupings with at least as much population. Since we seek a grouping larger than $p(D_K)$ and since $\delta(D_K) \leq 0$, this implies that a R-majority grouping cannot be formed. So it just remains to prove the claim.

Let $A = S' \setminus S$ and $R = S \setminus S'$ denote the sets of indices added to and removed from S, respectively, to make S'. Since A and R are disjoint, and we have assumed that $p(S') > p(S)$, it follows that $p(A) > p(R)$. Let $\mu = \max\{\frac{\delta_i}{p_i} \mid i \in A\}$ and let $\mu' = \min\{\frac{\delta_i}{p_i} \mid i \in R\}$. Note that, since $R \subseteq S = \{1, \ldots, K\}$ and $A \subseteq S^c = \{K+1, \ldots, n\}$ and the $\frac{\delta_i}{p_i}$ are non-increasing, we have $\mu \leq \mu'$.

Note that every unit $i \notin S$ has a Democratic majority ($\delta_i < 0$). This is because Republican-majority units are added to S in decreasing order of $\frac{\delta_i}{p_i}$ until the overall margin satisfies $\delta \leq 0$, so by construction every unit with a Republican majority is in S. It follows, since $A \subseteq S^c$, that $\mu < 0$.

We have $\mu \cdot p(R) > \mu \cdot p(A)$ because $p(R) < p(A)$ and $\mu < 0$. Also, $\mu' \cdot p(R) \geq \mu \cdot p(R)$. So, transitively, $\mu' \cdot p(R) > \mu \cdot p(A)$.

Note that

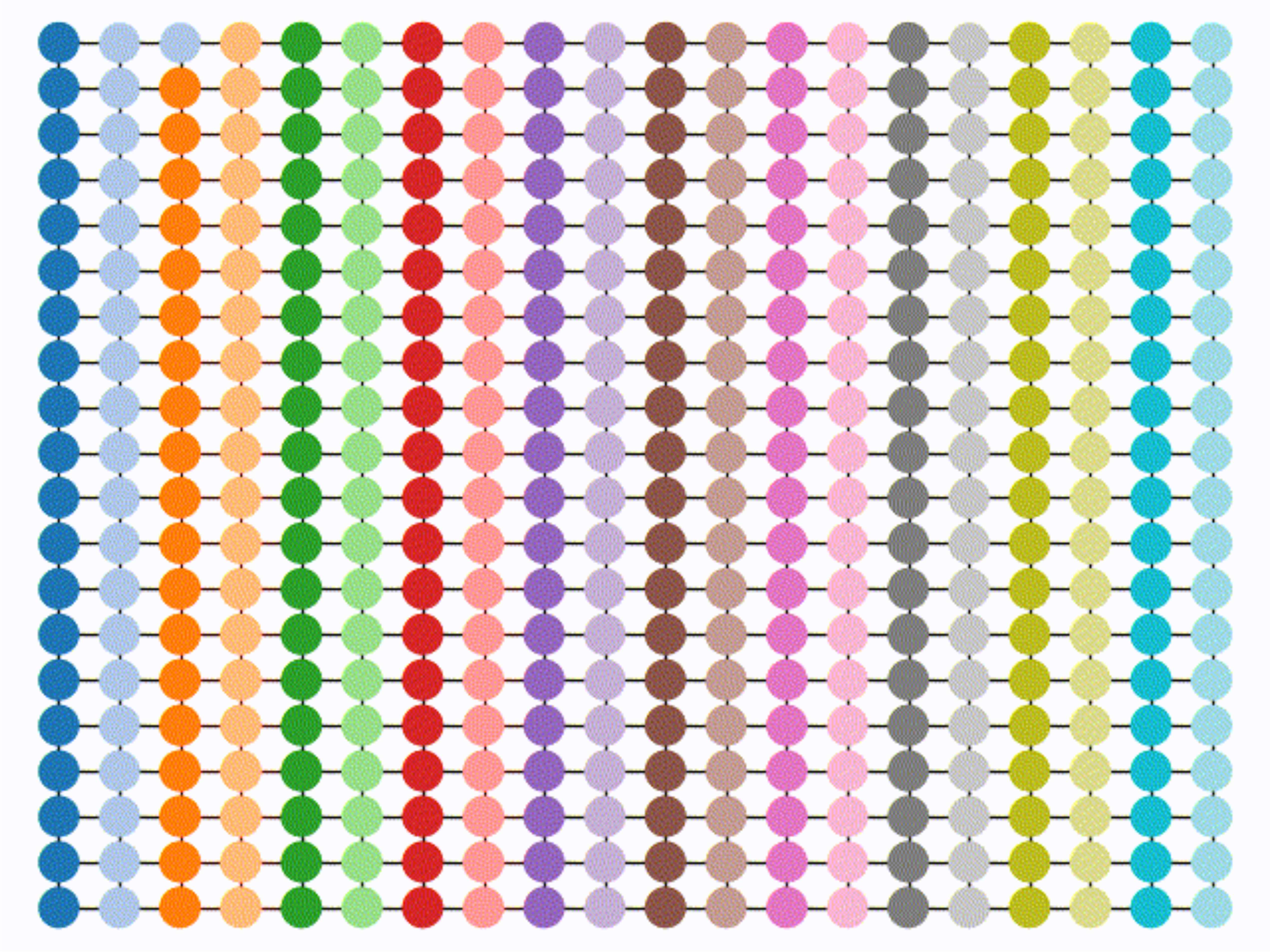
$$\mu' \cdot p(R) = \sum_{i \in R} \mu' \cdot p_i \leq \sum_{i \in R} \frac{\delta_i}{p_i} \cdot p_i = \delta(R).$$

In the absence of such a theorem, the best way to discover whether certain goals are simultaneously achievable is to construct a good sample of plans.

SO THAT'S WHY TO SAMPLE.

NOW,
HOW TO SAMPLE?

first idea: Ising-style MCMC ...not great for our setting

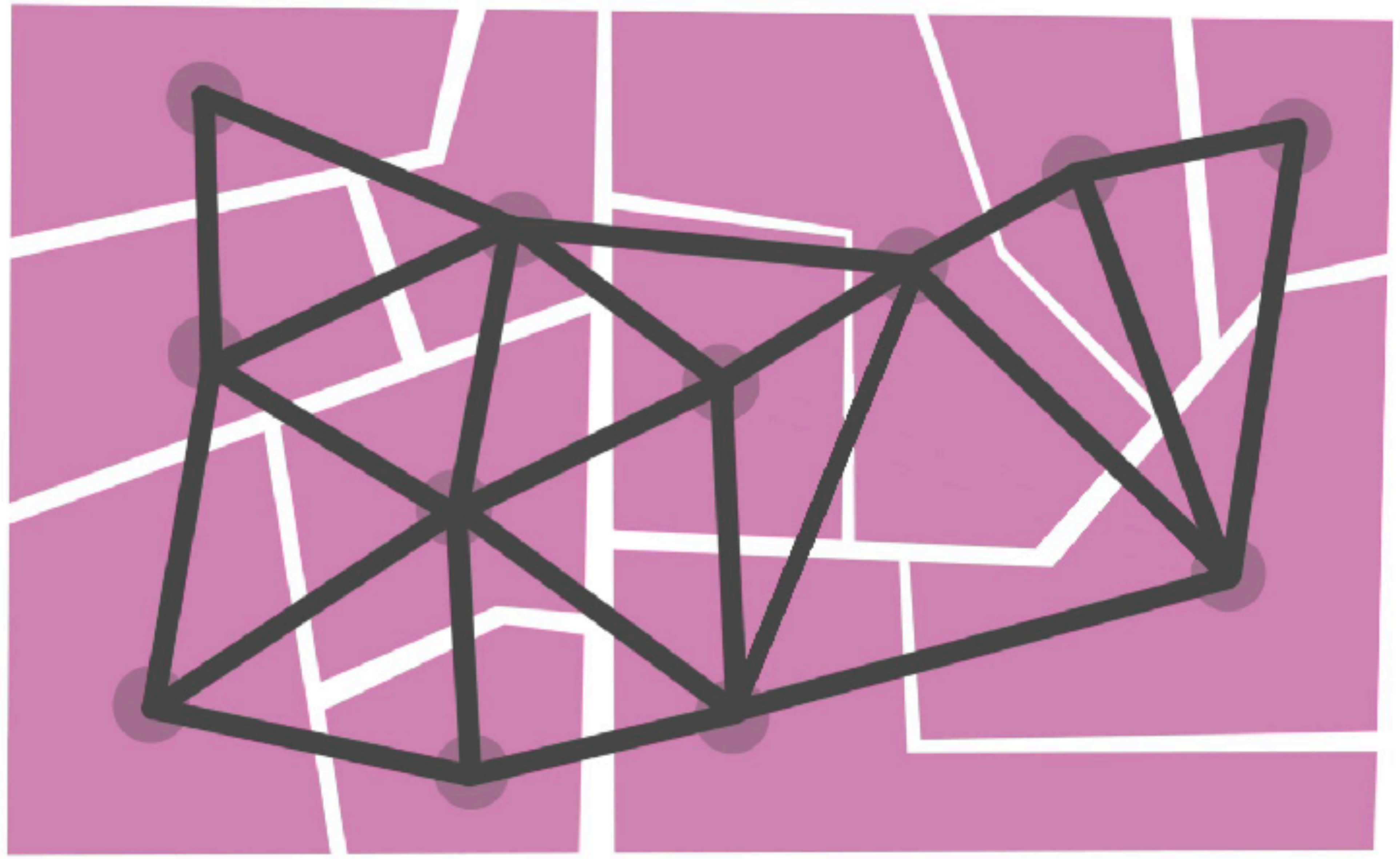


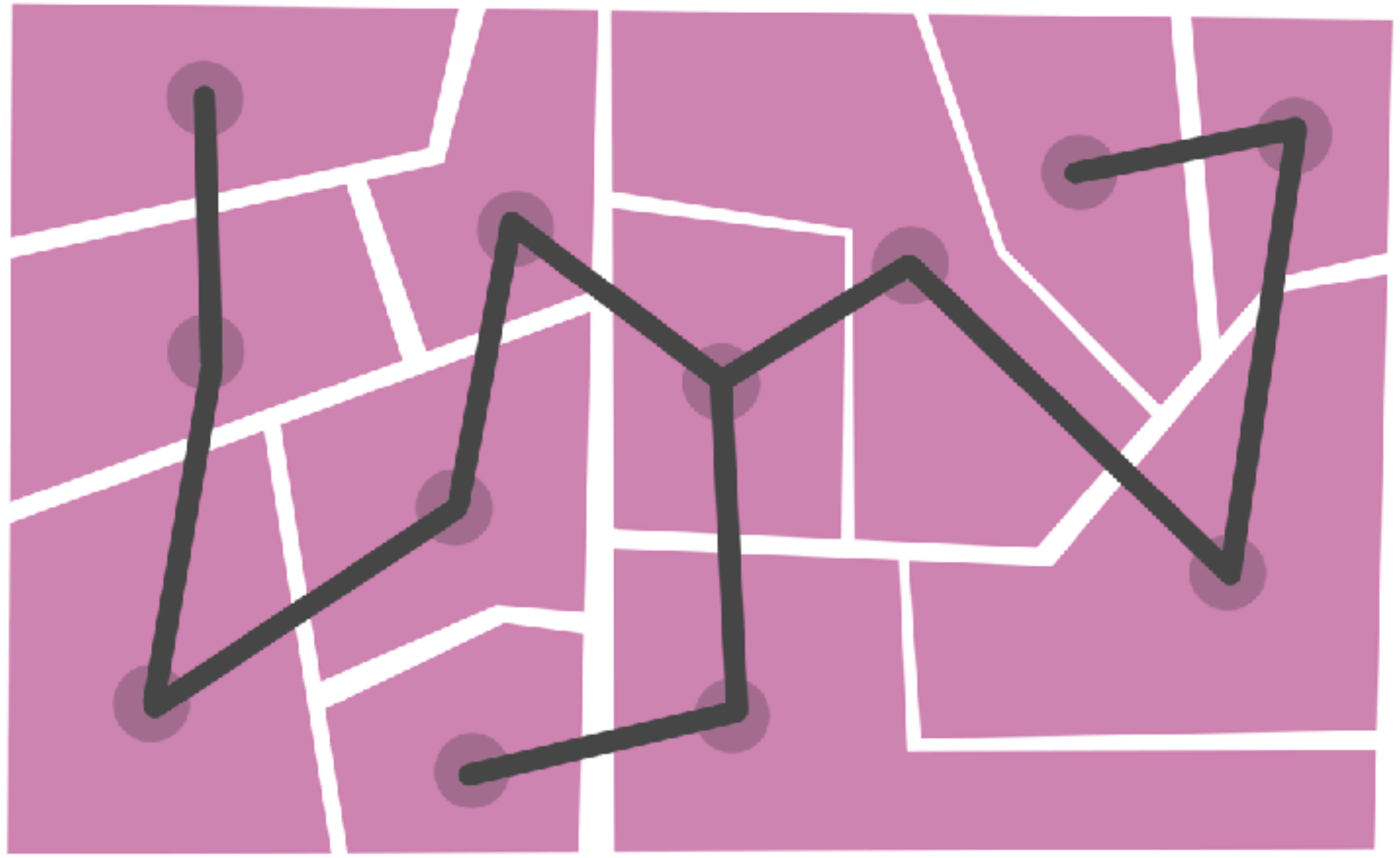
SPANNING TREES

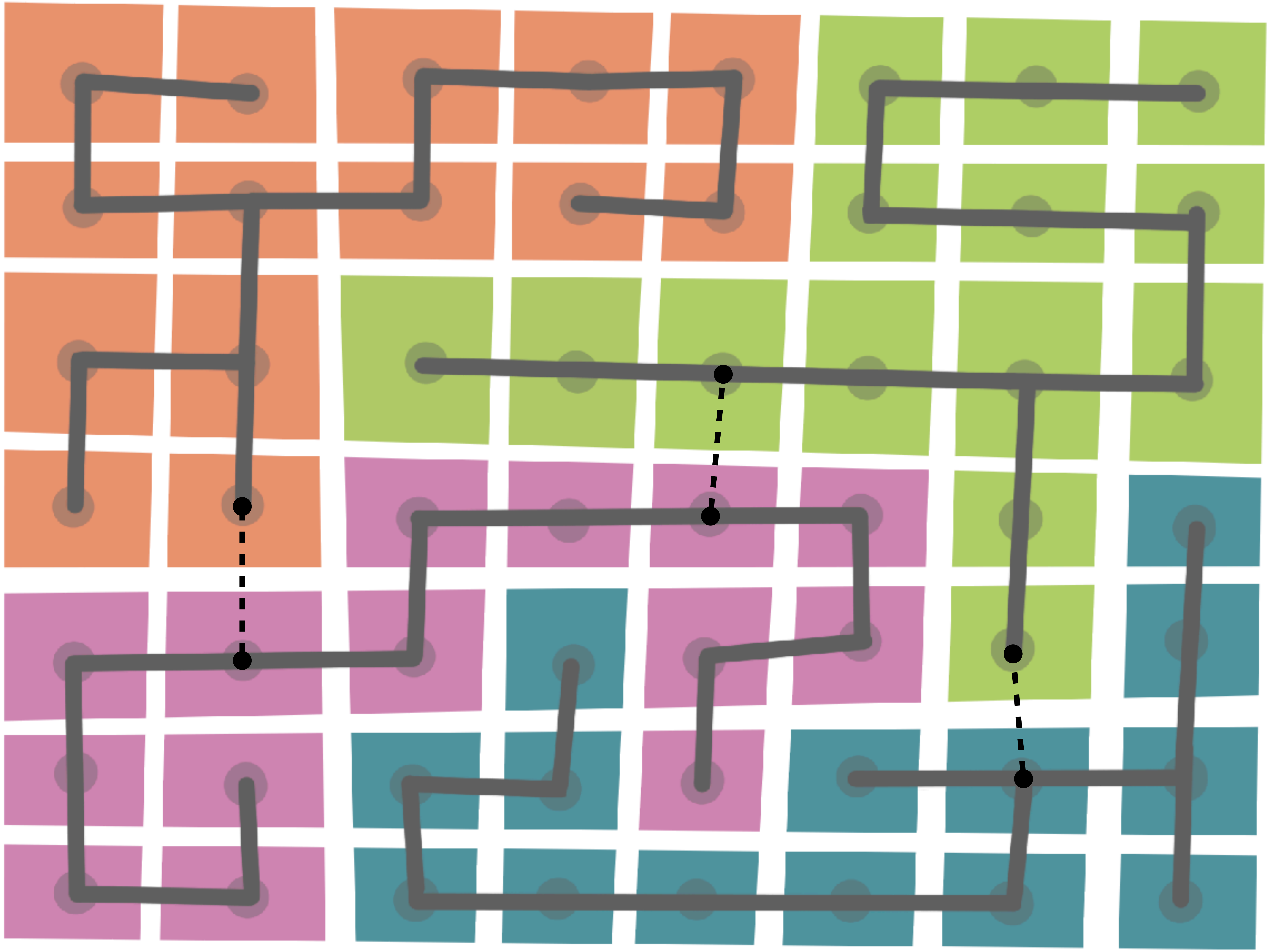
Spanning tree — cycle-free subgraph connecting all vertices of G

Minimally connected “skeleton” of G

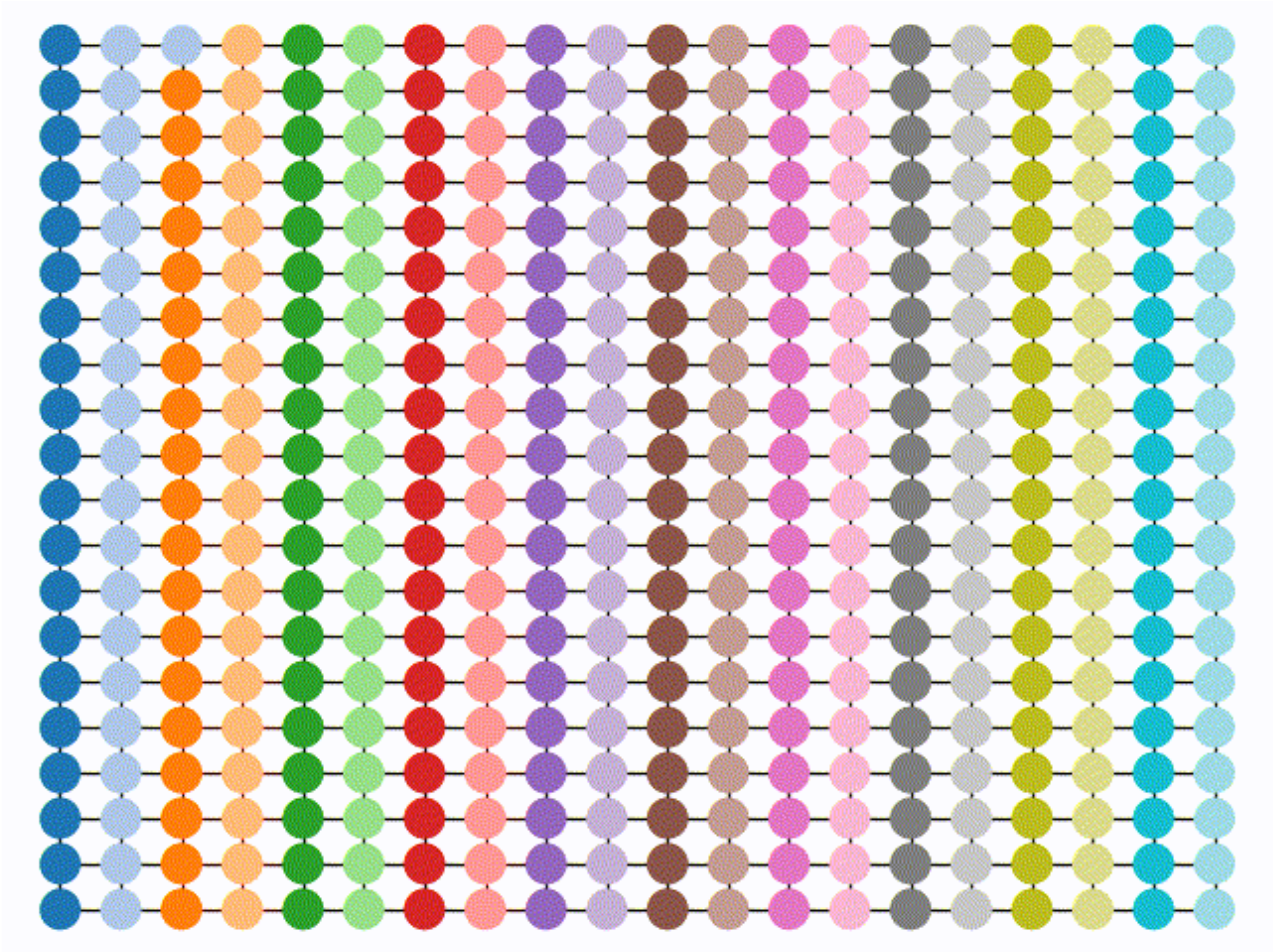
Deleting any single edge leaves exactly two connected components



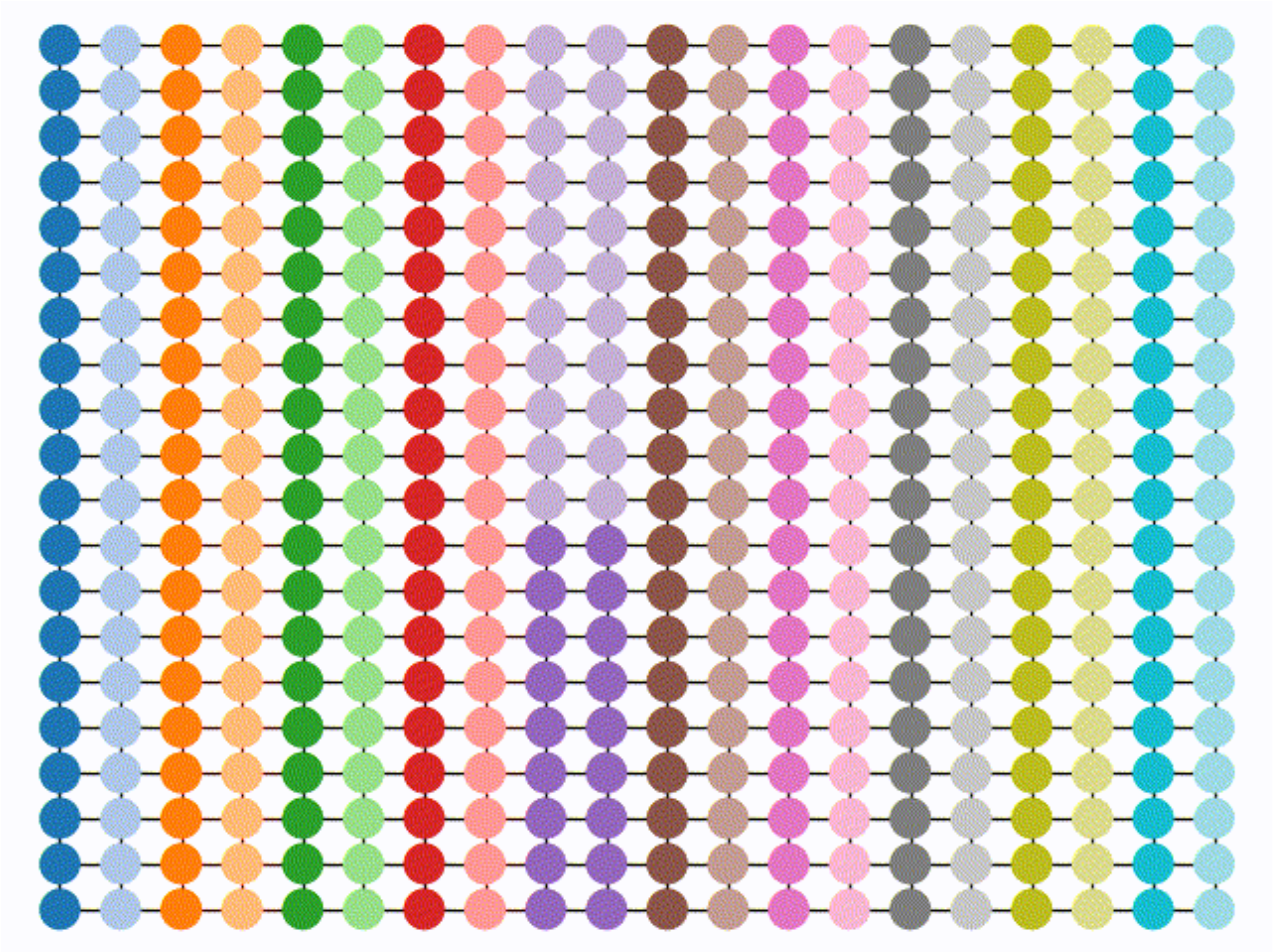




a new chain



flip



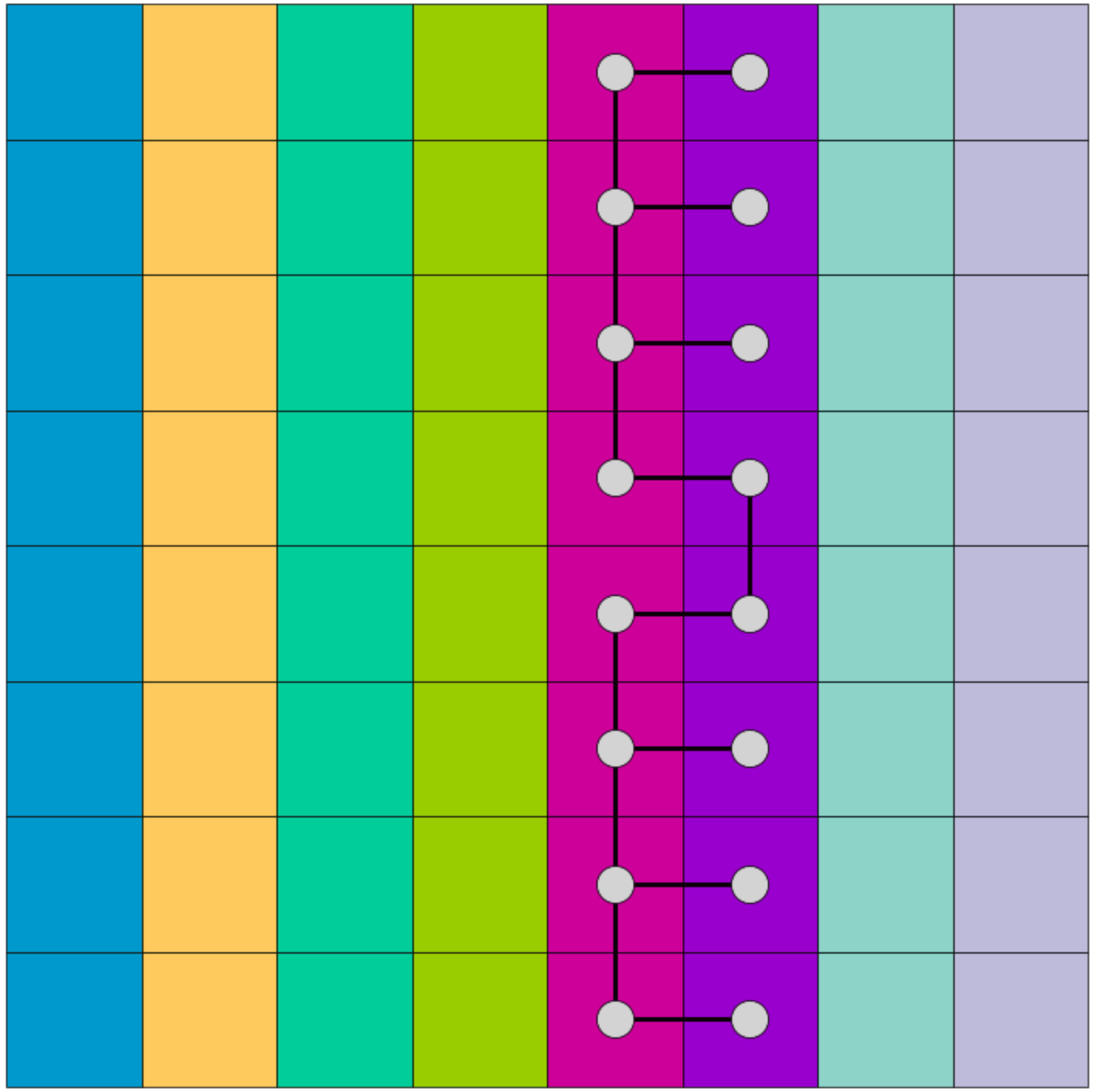
recombination

*with **Najt**, DeFord, Solomon*

RECOMBINATION

"Recombination" is a Markov chain in the space of balanced partitions

- **choose two adjacent districts**
- **fuse them and pick a random spanning tree**
- **look for an edge in the tree with balanced complementary components**
- **cut there to get two new districts**
- **iterate**



A TREE DISTRIBUTION

All recombination methods approximately target the spanning tree distribution :

$$\pi(P) \propto \prod_{P_i} N_{ST}(P_i) \qquad \textbf{for} \quad P = (P_1, \ldots, P_k)$$

Produces districts with many connections within compared to between

cf. "Small-World Networks" and community detection — plus, short boundaries means nice-looking shapes

Now there are reversible recombination chains that can exactly target π (Mattingly et al., Cannon—Duchin—Randall—Rule)

MANY INTERESTING RESEARCH DIRECTIONS

Exciting new theorem:
Trees in grid-graphs are splittable!

(meaning 1/poly fraction of spanning trees have a balance-edge)

resolves conjecture of Charikar et al.

Sampling Balanced Forests of Grids in Polynomial Time

Sarah Cannon[*], Wesley Pegden[†], and Jamie Tucker-Foltz[‡]

January 12, 2024

Abstract

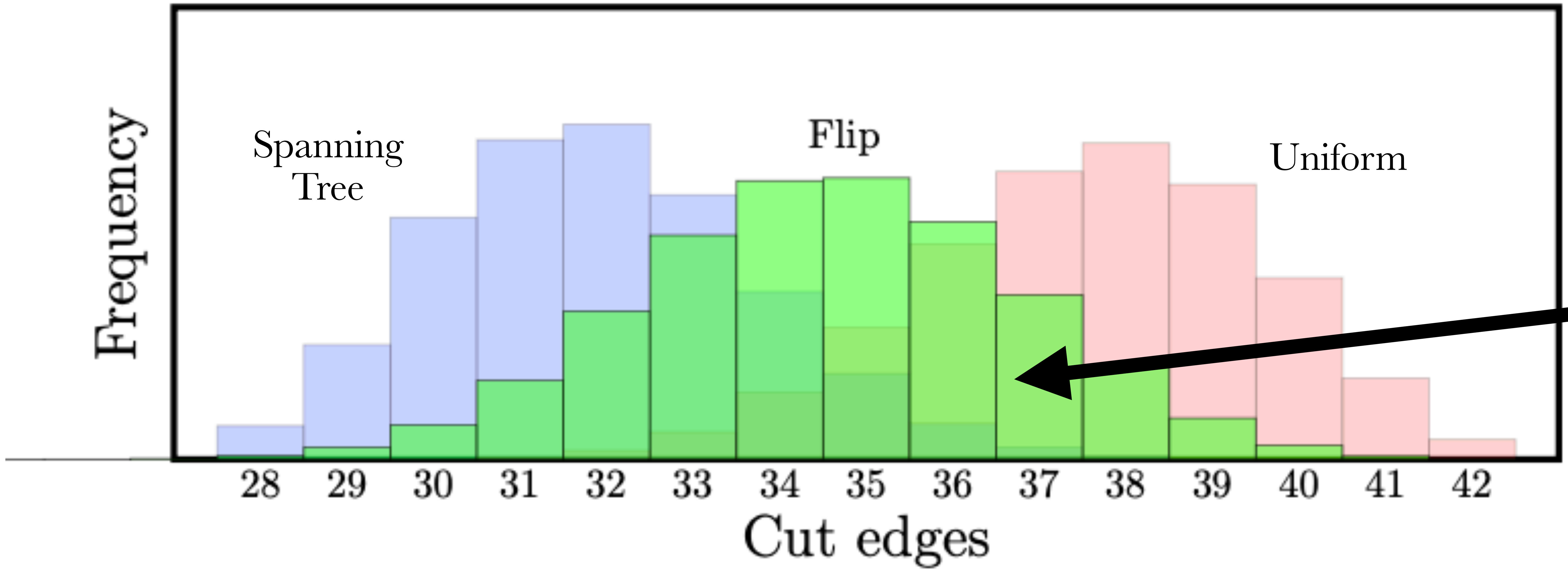
We prove that a polynomial fraction of the set of k-component forests in the $m \times n$ grid graph have equal numbers of vertices in each component, for any constant k. This resolves a conjecture of Charikar, Liu, Liu, and Vuong, and establishes the first provably polynomial-time algorithm for (exactly or approximately) sampling balanced grid graph partitions according to the spanning tree distribution, which weights each k-partition according to the product, across its k pieces, of the number of spanning trees of each piece. Our result follows from a careful analysis of the probability a uniformly random spanning tree of the grid can be cut into balanced pieces.

Beyond grids, we show that for a broad family of lattice-like graphs, we achieve balance up to any multiplicative $(1 \pm \varepsilon)$ constant with constant probability, and up to an additive constant with polynomial probability. More generally, we show that, with constant probability, components derived from uniform spanning trees can approximate any given partition of a planar region specified by Jordan curves. These results imply polynomial time algorithms for sampling approximately balanced tree-weighted partitions for lattice-like graphs.

Our results have applications to understanding political districtings, where there is an underlying graph of indivisible geographic units that must be partitioned into k population-balanced connected subgraphs. In this setting, tree-weighted partitions have interesting geometric properties, and this has stimulated significant effort to develop methods to sample them.

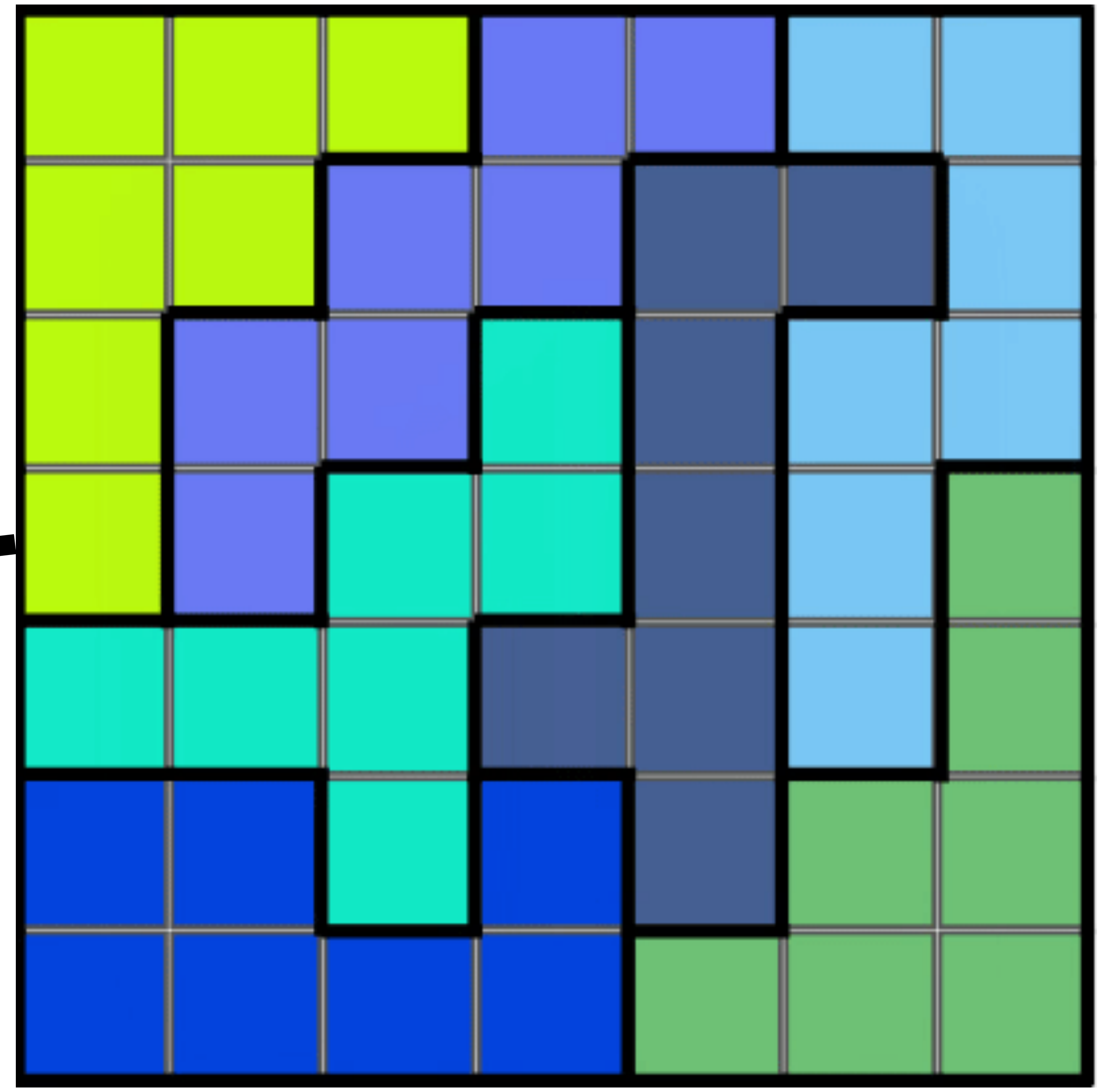
1 Introduction

We consider the following question: given a graph G and an integer constant k, how can one randomly sample partitions of G into k connected pieces, each of equal size? We address this question in the context of the *spanning tree distribution* on partitions, under which the weight of

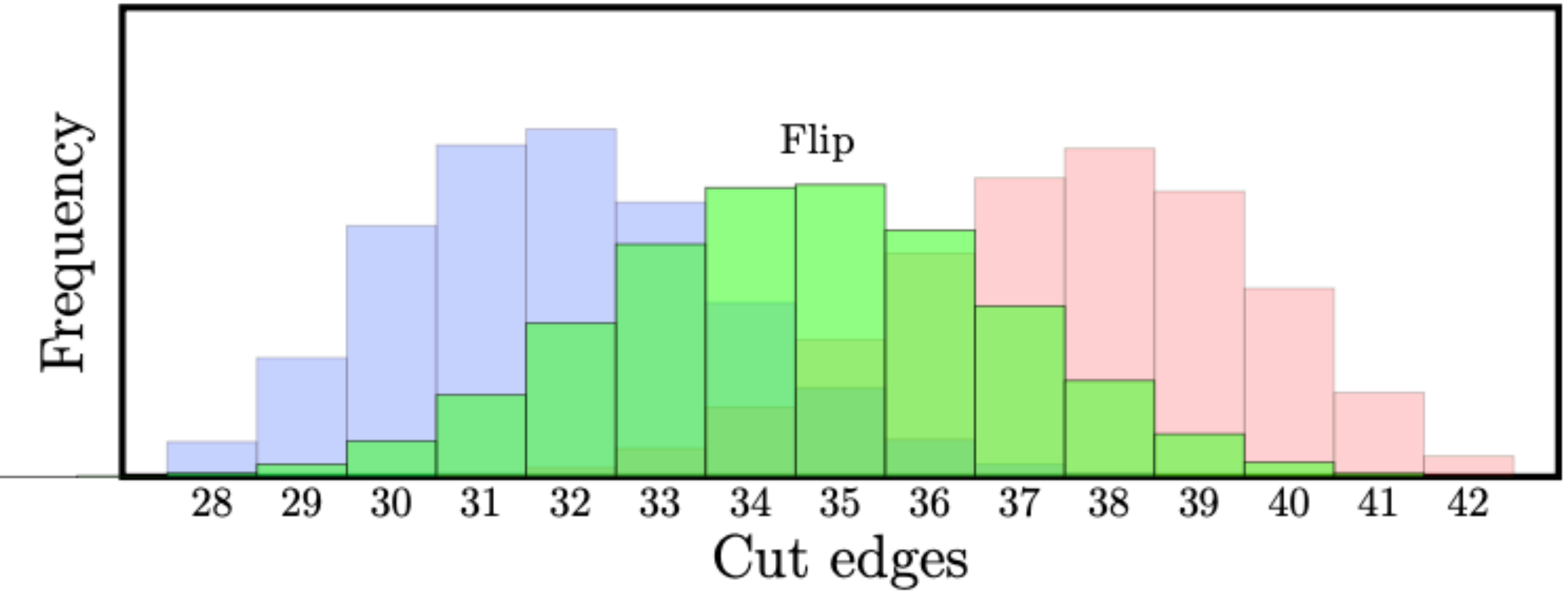


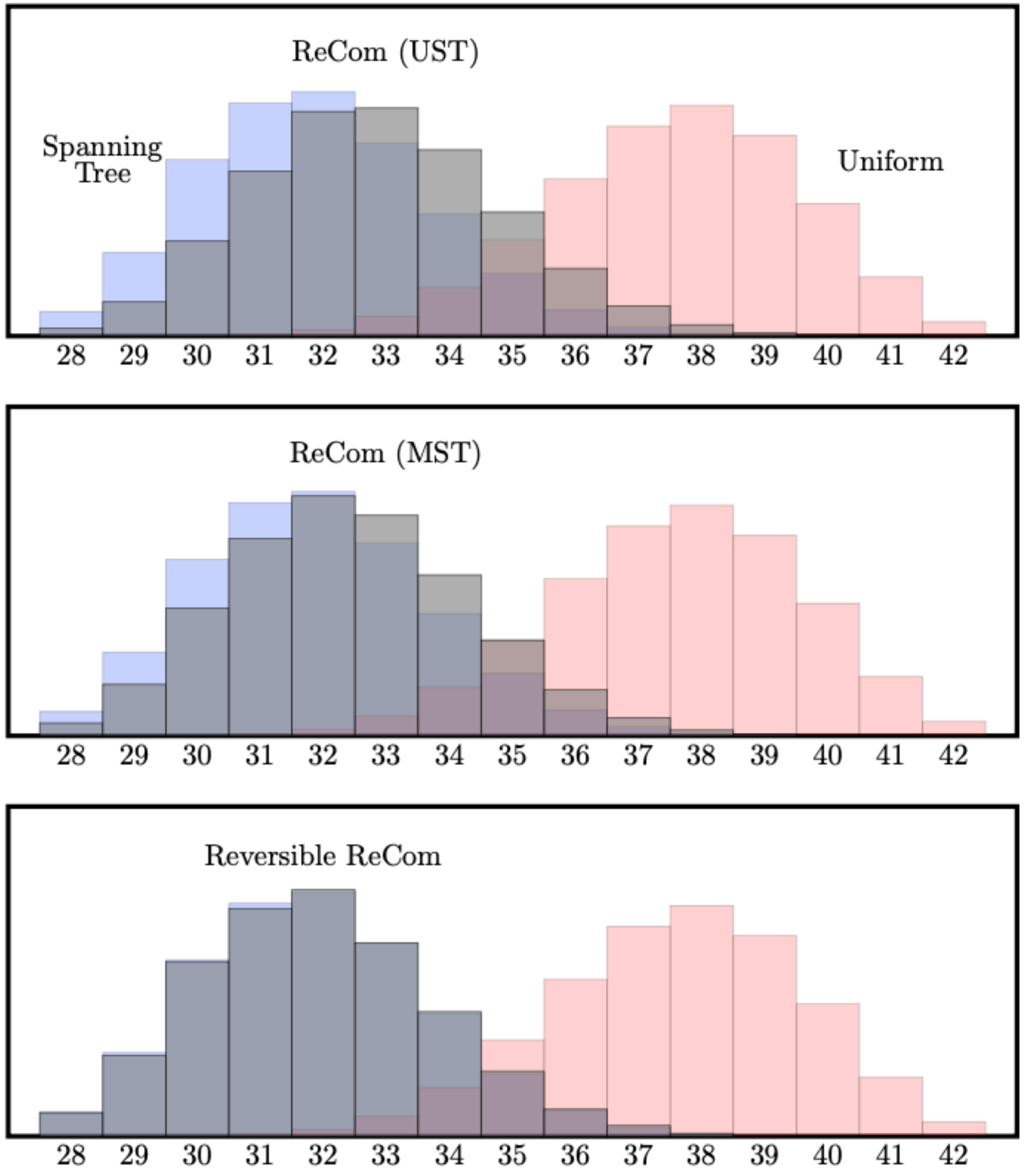
7x7 grid cut into 7-ominoes

(158,753,814 states)



how big is the cut-set?





distributions (can) matter!

HIGH-LEVEL STRATEGIES

MCMC

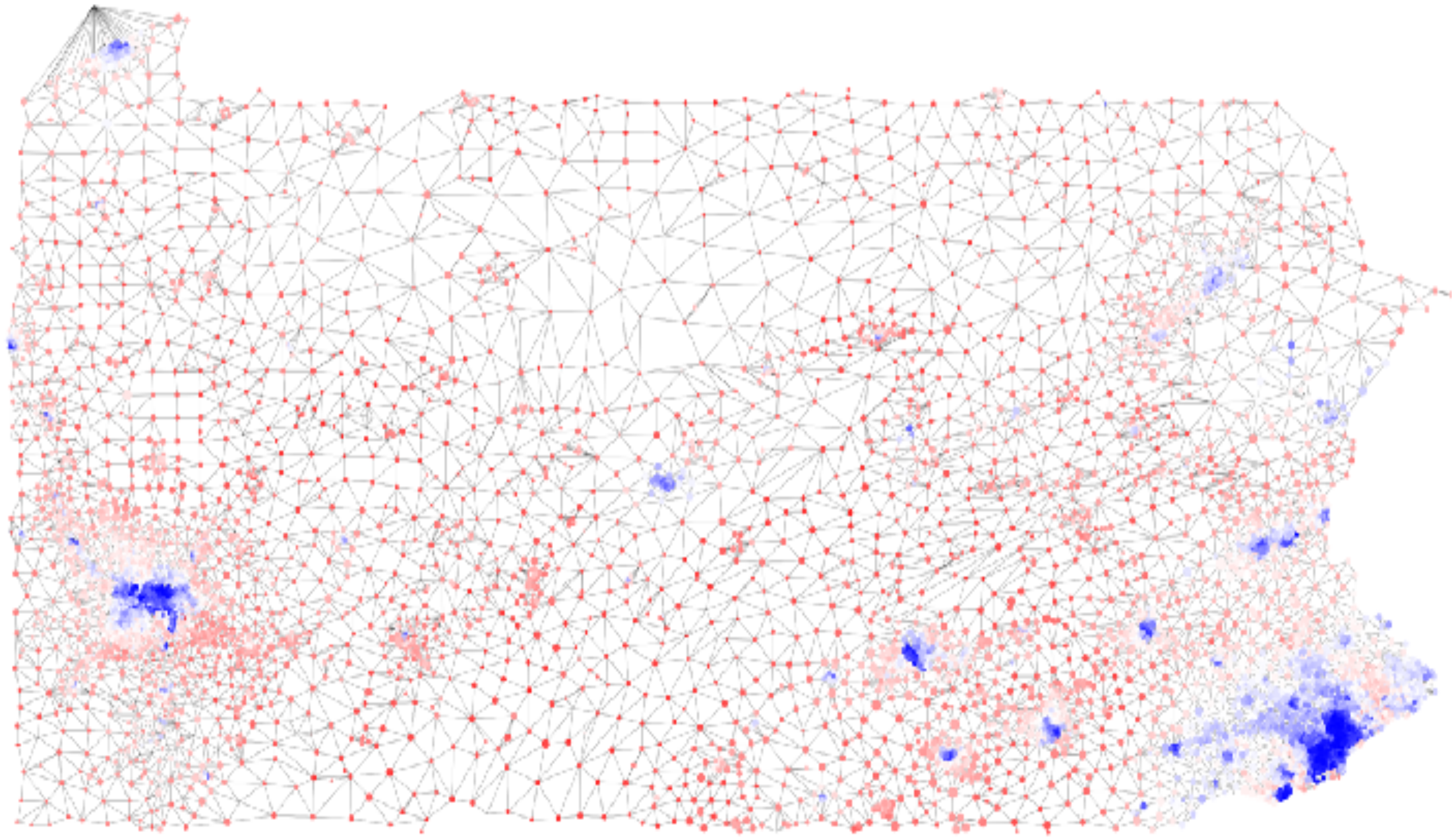
sequential sampling

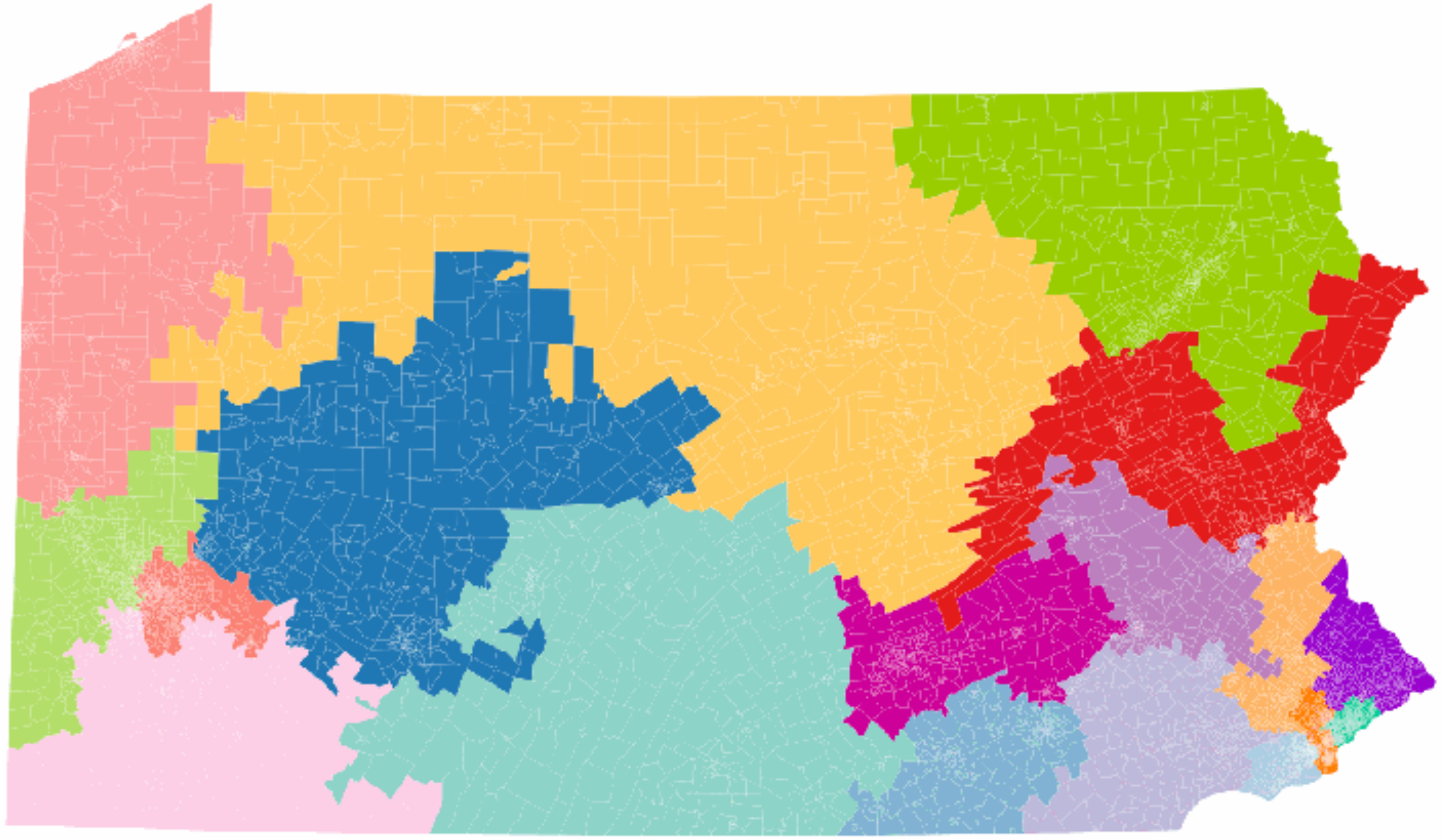
SMC

importance sampling

direct sampling

Let's take a **random sample** of plans



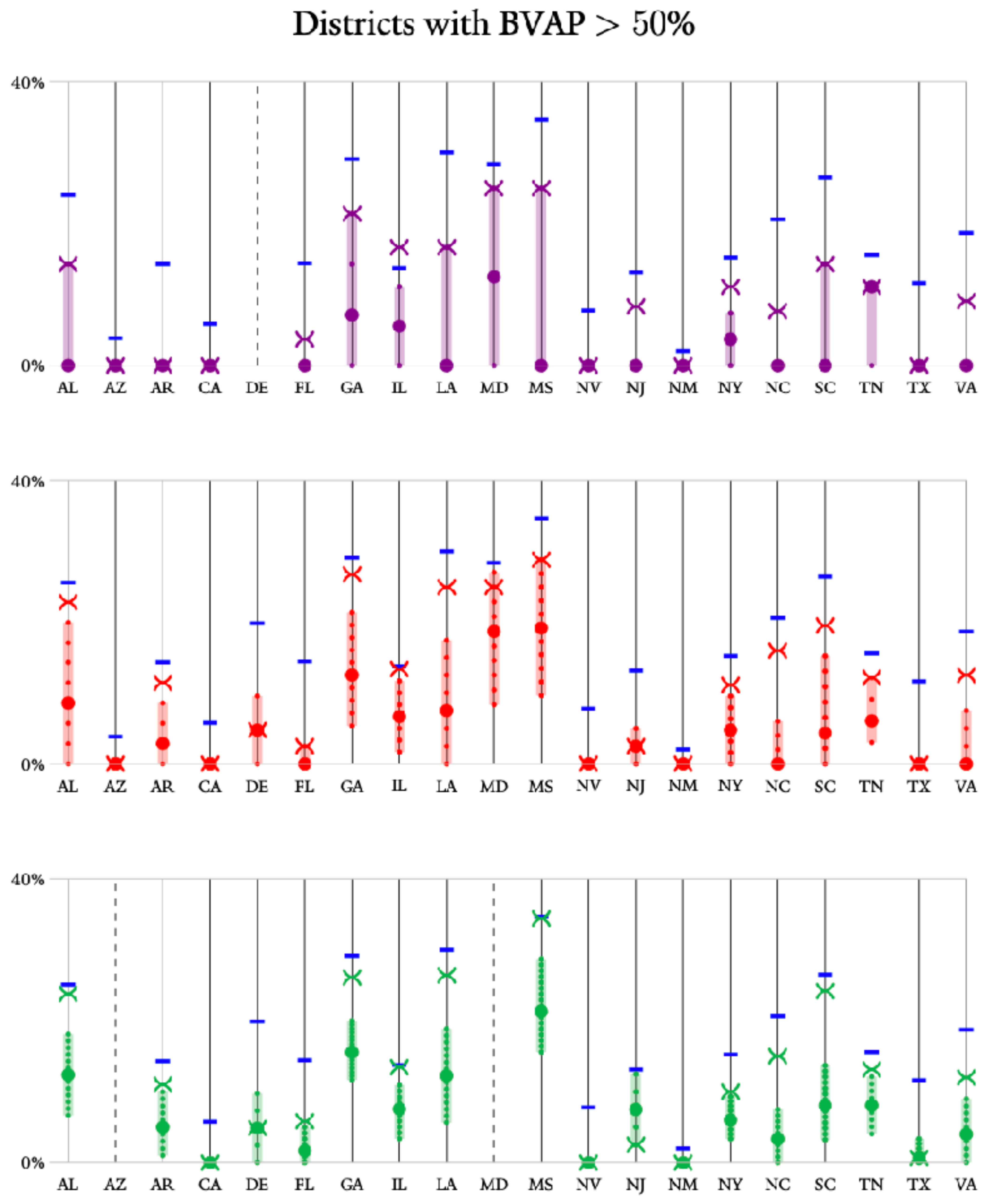


do we tend to get party proportionality from “blind” districts?

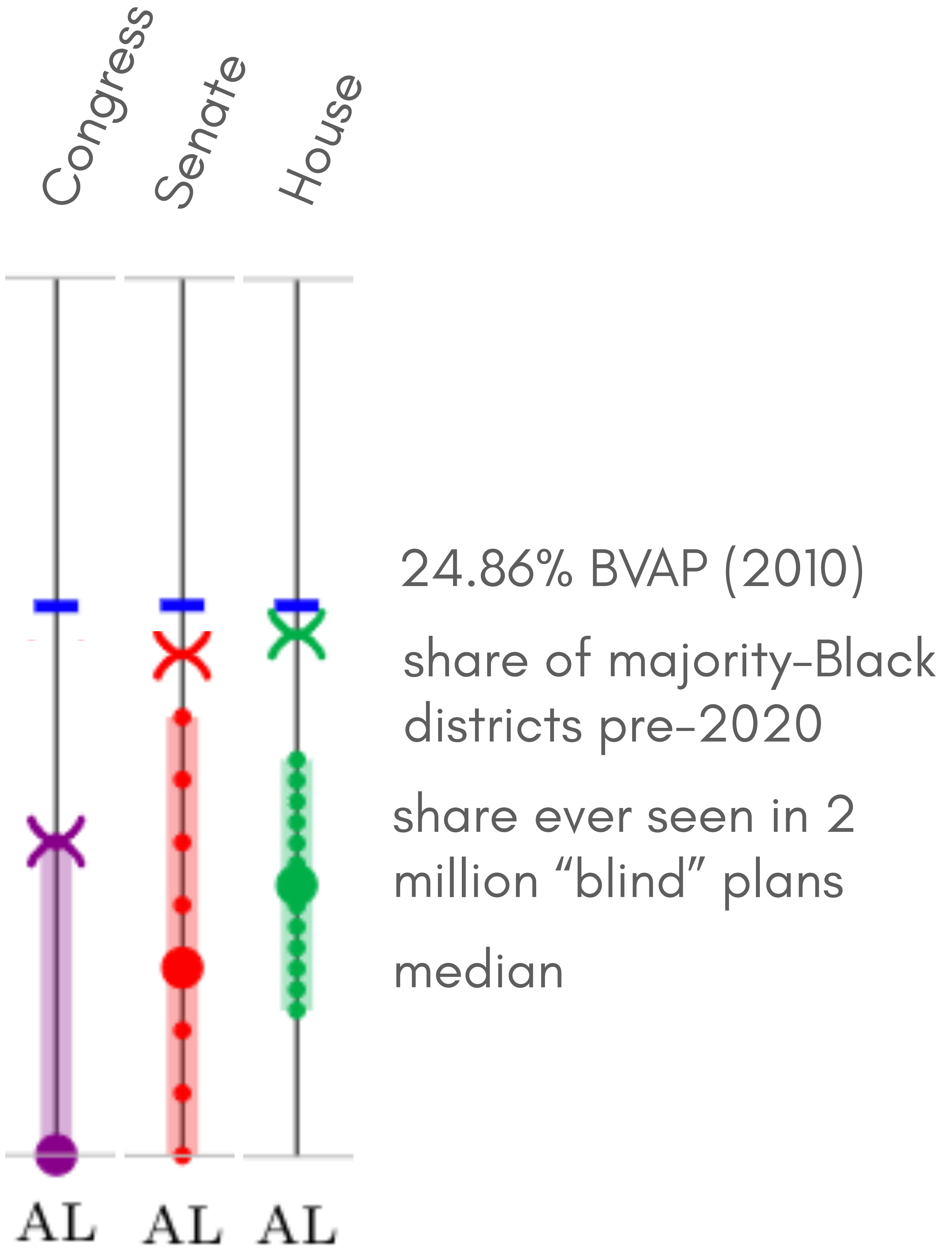
45.74
Ohio
(16 districts)
Pres16
2 3 4 5 6 7 8
39.04
Ohio
(16 districts)
Sen16
1 2 3 4
49.59
Wisconsin
(8 districts)
Pres16
2 3 4 5
48.27
Wisconsin
(8 districts)
Sen16
1 2 3 4 5
57.82
Maryland
(8 districts)
Pres16
6 7
56.96
Maryland
(8 districts)
Sen16
5 6 7
47.33
Georgia
(14 districts)
Pres16
2 3 4 5 6 7
42.82
Georgia
(14 districts)
Sen16
2 3 4 5 6
49.65
Pennsylvania
(18 districts)
Pres16
5 6 7 8 9
49.28
Pennsylvania
(18 districts)
Sen16
3 4 5 6 7 8 9
56.16
Oregon
(5 districts)
Pres16
2 3 4
62.93
Oregon
(5 districts)
Sen16
4 5
41.99
Texas
(36 districts)
Pres12
7 11 15
Democratic seats
41.85
Texas
(36 districts)
Sen12
7 11 15
Democratic seats
48.02
N. Carolina
(13 districts)
Pres16
2 3 4 5 6 7 8
Democratic seats
46.98
N. Carolina
(13 districts)
Sen16
2 3 4 5 6 7 8
Democratic seats
61.79
Massachusetts
(9 districts)
Pres12
8 9
Democratic seats
53.78
Massachusetts
(9 districts)
Sen12
3 4 5 6 7 8
Democratic seats

no we do not
45.74
Ohio
(16 districts)
Pres16
39.04
Ohio
(16 districts)
Sen16
49.59
Wisconsin
(8 districts)
Pres16
48.27
Wisconsin
(8 districts)
Sen16
57.82
Maryland
(8 districts)
Pres16
56.96
Maryland
(8 districts)
Sen16
47.33
Georgia
(14 districts)
Pres16
42.82
Georgia
(14 districts)
Sen16
49.65
56.16
62.93
Oregon
(5 districts)
Sen16
41.99
Texas
(36 districts)
Pres12
41.85
48.02
N. Carolina
(13 districts)
Pres16
46.98
N. Carolina
(13 districts)
Sen16
61.79
Massachusetts
(9 districts)
Pres12
53.78
Massachusetts
(9 districts)
Sen12
Democratic seats
Democratic seats
Democratic seats
Democratic seats
Democratic seats
Democratic seats

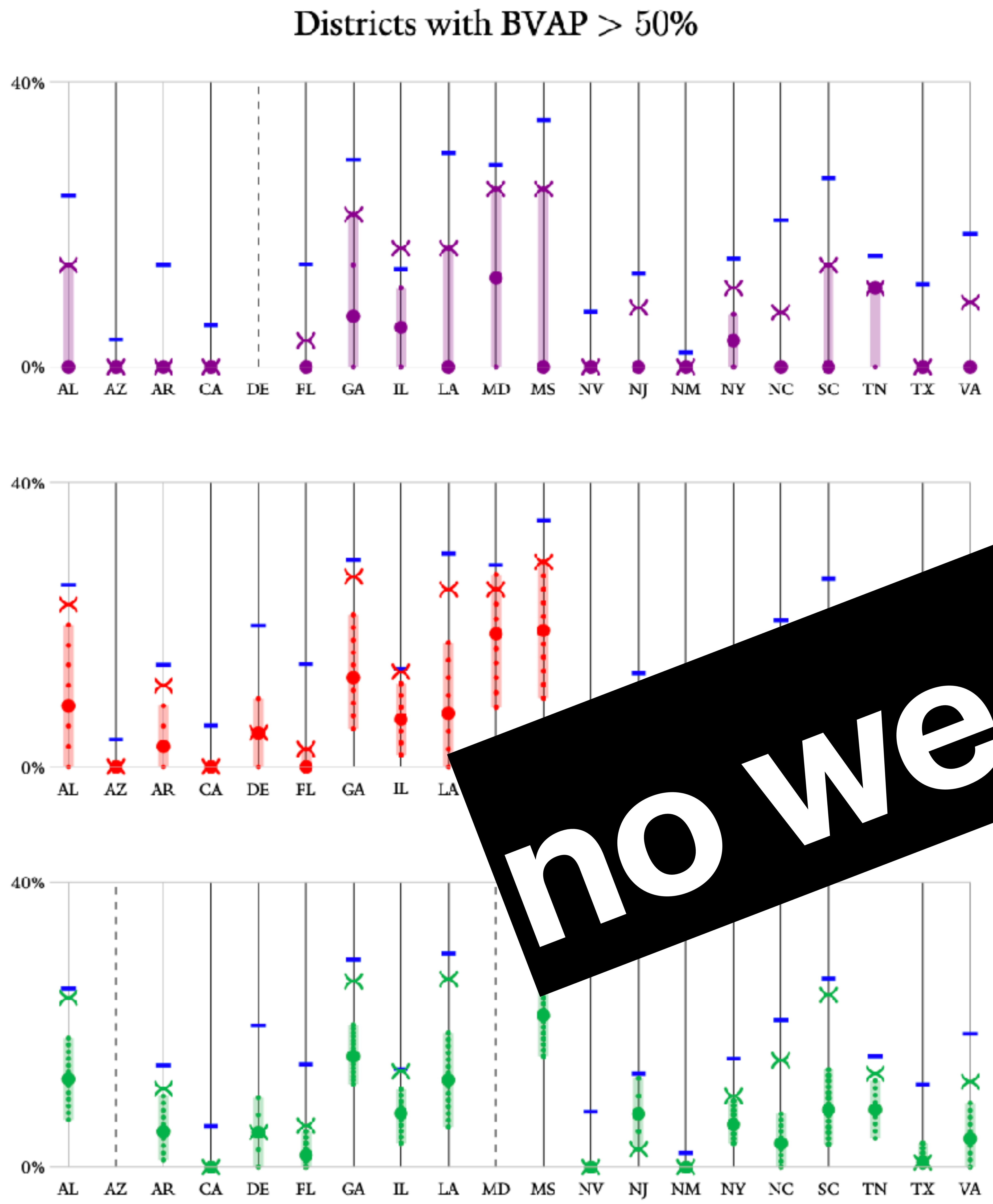
and how about race?



Districts with BVAP > 50%
40%
0%
AL AZ AR CA DE FL GA IL LA MD MS NV NJ NM NY NC SC TN TX VA



Congress
Senate
House
AL AL AL
24.86% BVAP (2010)
share of majority-Black districts pre-2020
share ever seen in 2 million "blind" plans
median



no we do not

24.86% BVAP (2010)

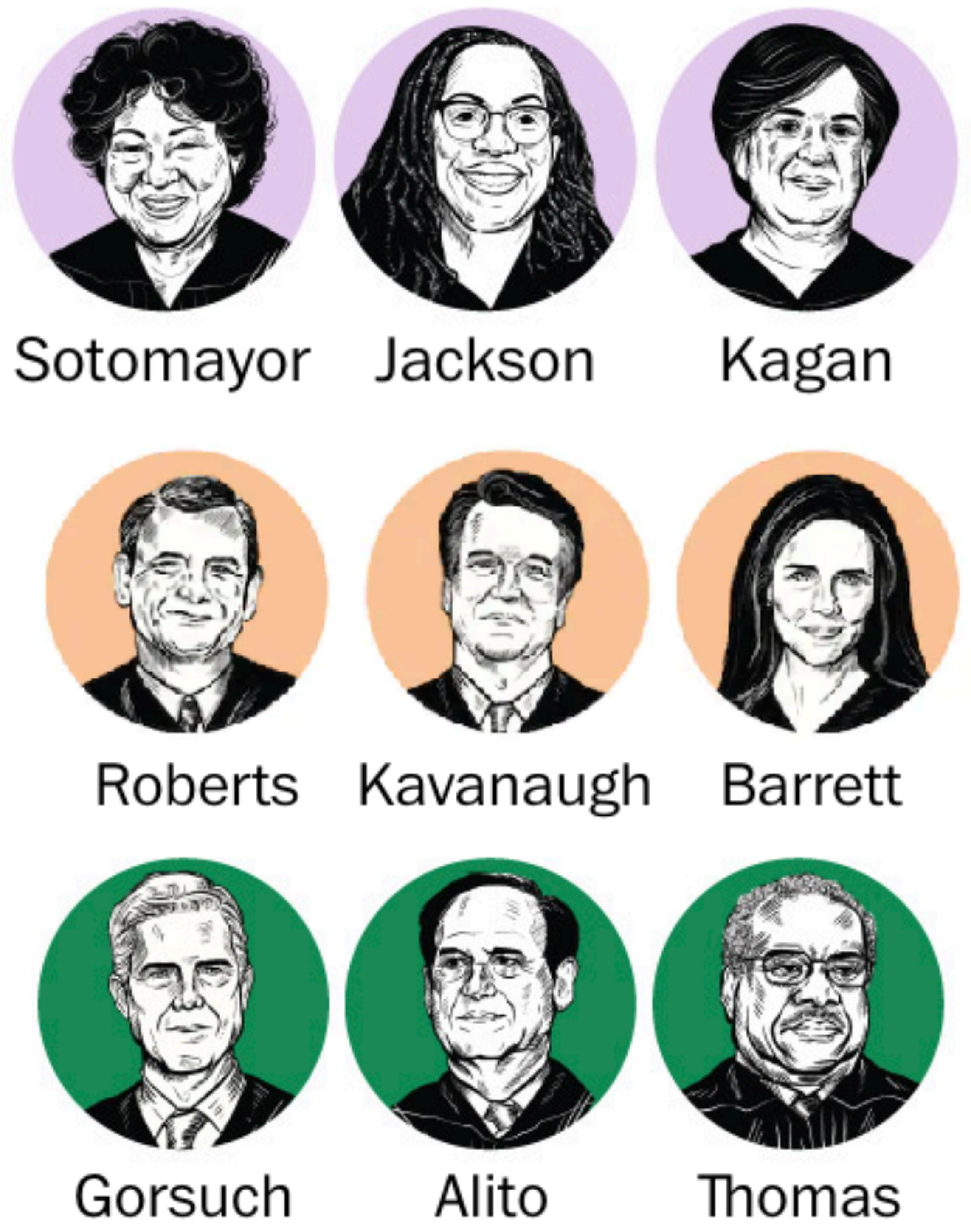
share of majority-Black districts pre-2020

share ever seen in 2 million “blind” plans

median

so sampling helps us understand districts!

but this is risky.



II

Even if §2 applies here, however, Alabama should prevail. The District Court found that Alabama's congressional districting map "dilutes" black residents' votes because, while it is *possible* to draw two majority-black districts, Alabama's map only has one.[5] But the critical question in all vote-dilution cases is: "Diluted relative to what benchmark?" *Gonzalez* v. *Aurora*, 535 F. 3d 594, 598 (CA7 2008) (Easterbrook, C. J.). Neither the District Court nor the majority has any defensible answer. The text of §2 and the logic of vote-dilution claims require a meaningfully race-neutral benchmark, and no race-neutral benchmark can justify the District Court's finding of vote dilution in these cases. The only benchmark that can justify it—and the one that the District Court demonstrably applied—is

Somebody's clerks read my papers (!)

[9]The majority notes that this study used demographic data from the 2010 census, not the 2020 one. That is irrelevant, since the black population share in Alabama changed little (from 26.8% to 27.16%) between the two censuses. To think that this minor increase might have changed Dr. Duchin's results would be to entirely miss her point: that proportional representation for *any* minority, unless achieved "by design," is a statistical anomaly in almost all single-member-districting systems. Duchin & Spencer 764.

He's agreeing with my conclusion that blind districting "fences out" minority representation!

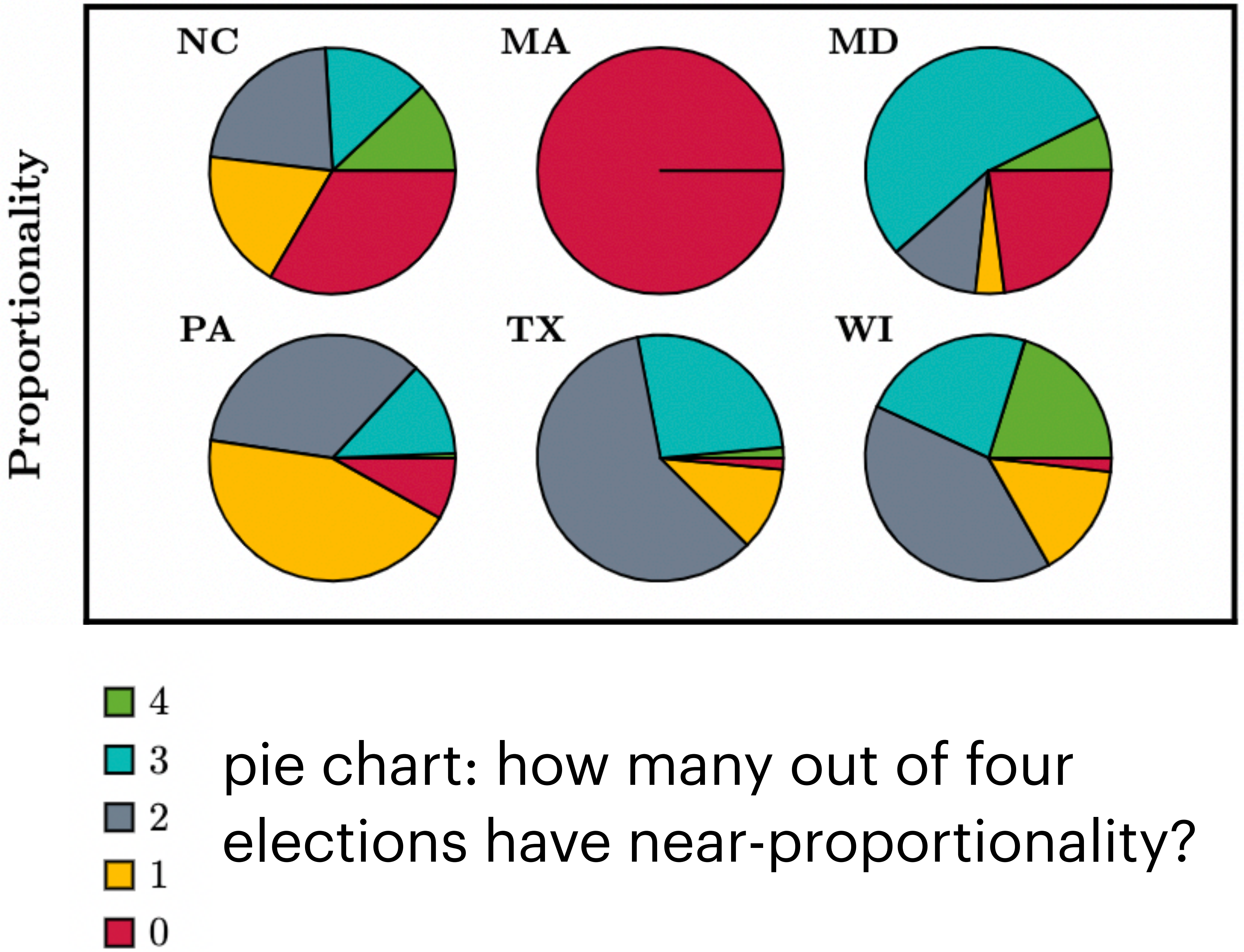
Why is this "risky"? Because he goes on to conclude that having no minority representation must be acceptable.

REFORM PROPOSAL: REDISTRICTING FOR PROPORTIONALITY

Newish paper with Gabe Schoenbach considers whether it is feasible to design/ select districts with proportionality as a goal

tl;dr — yes!

In fact, in nearly every state this occurs in at least 10% of a blind ensemble



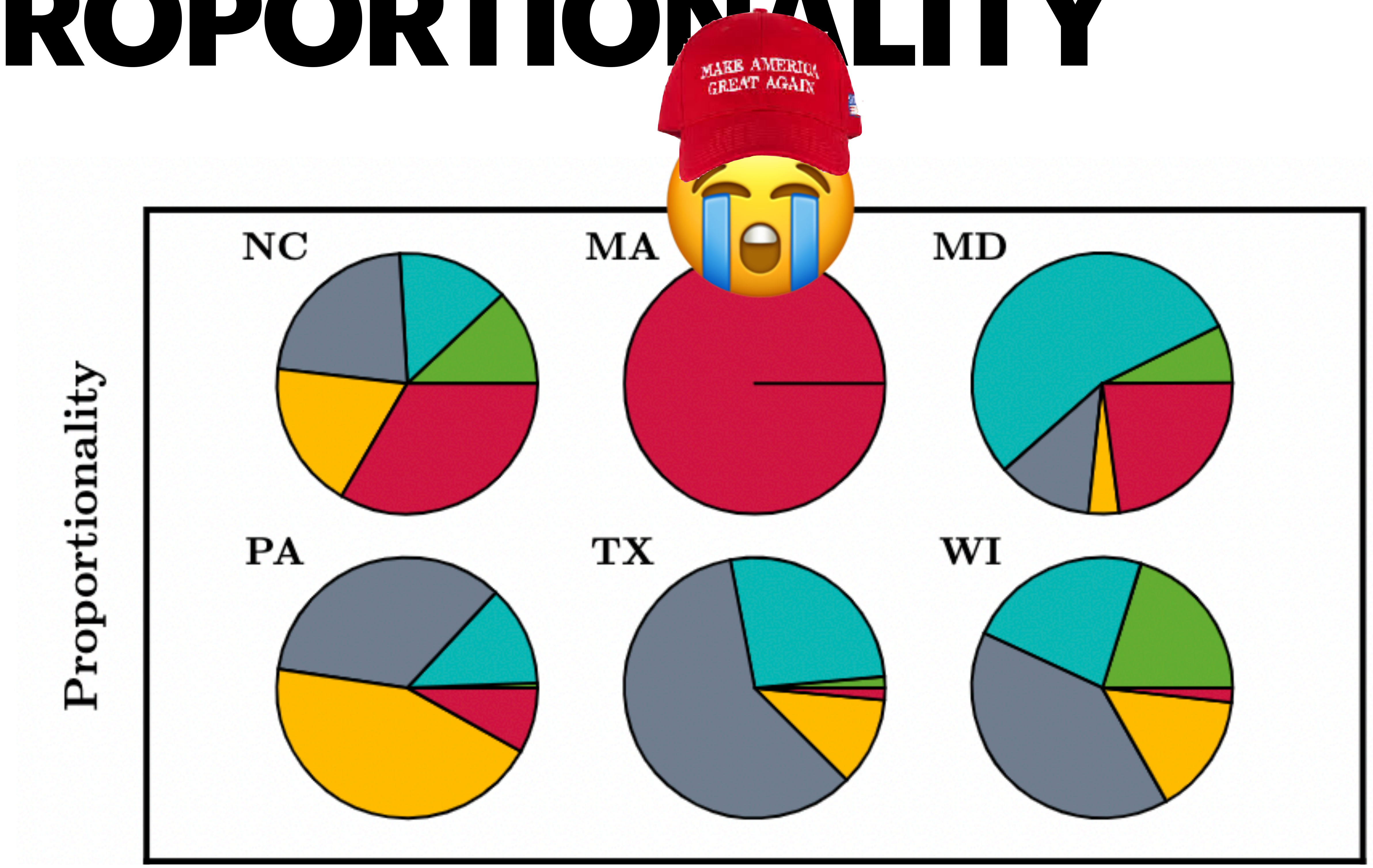
pie chart: how many out of four elections have near-proportionality?

REFORM PROPOSAL: REDISTRICTING FOR PROPORTIONALITY

Newish paper with Gabe Schoenbach considers whether it is feasible to design/ select districts with proportionality as a goal

tl;dr — yes!

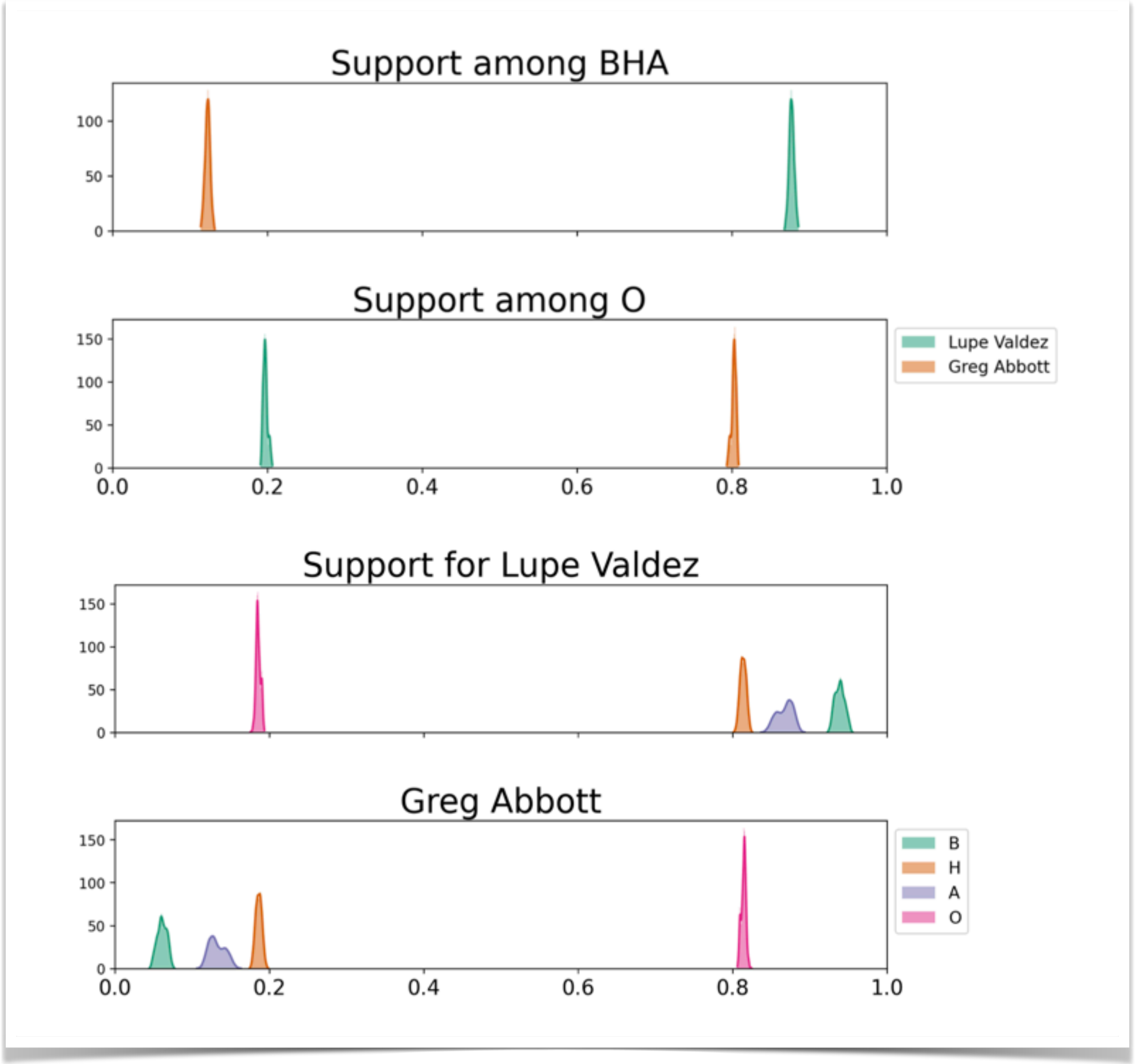
In fact, in nearly every state this occurs in at least 10% of a blind ensemble

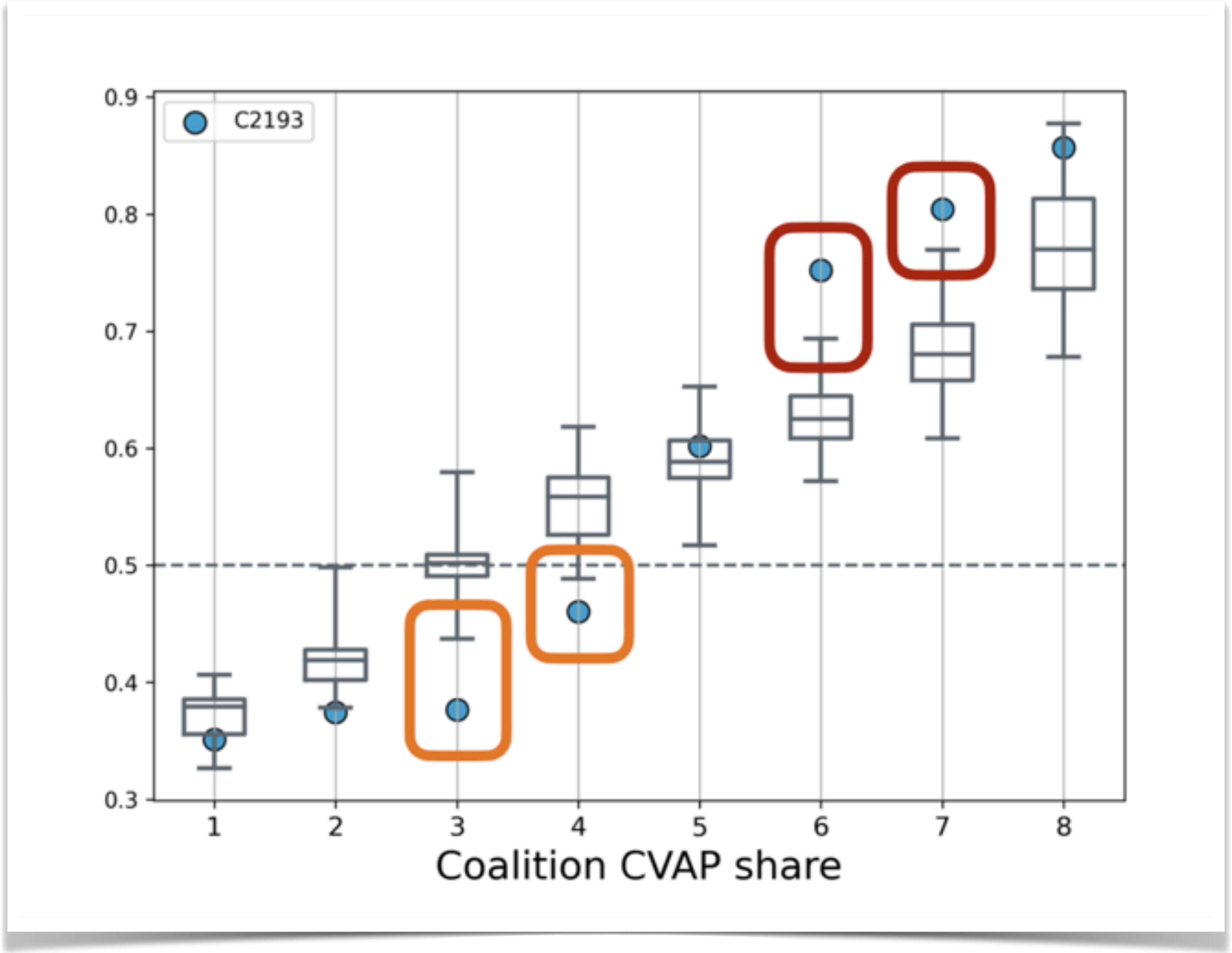


pie chart: how many out of four elections have near-proportionality?

Texas, May 2025

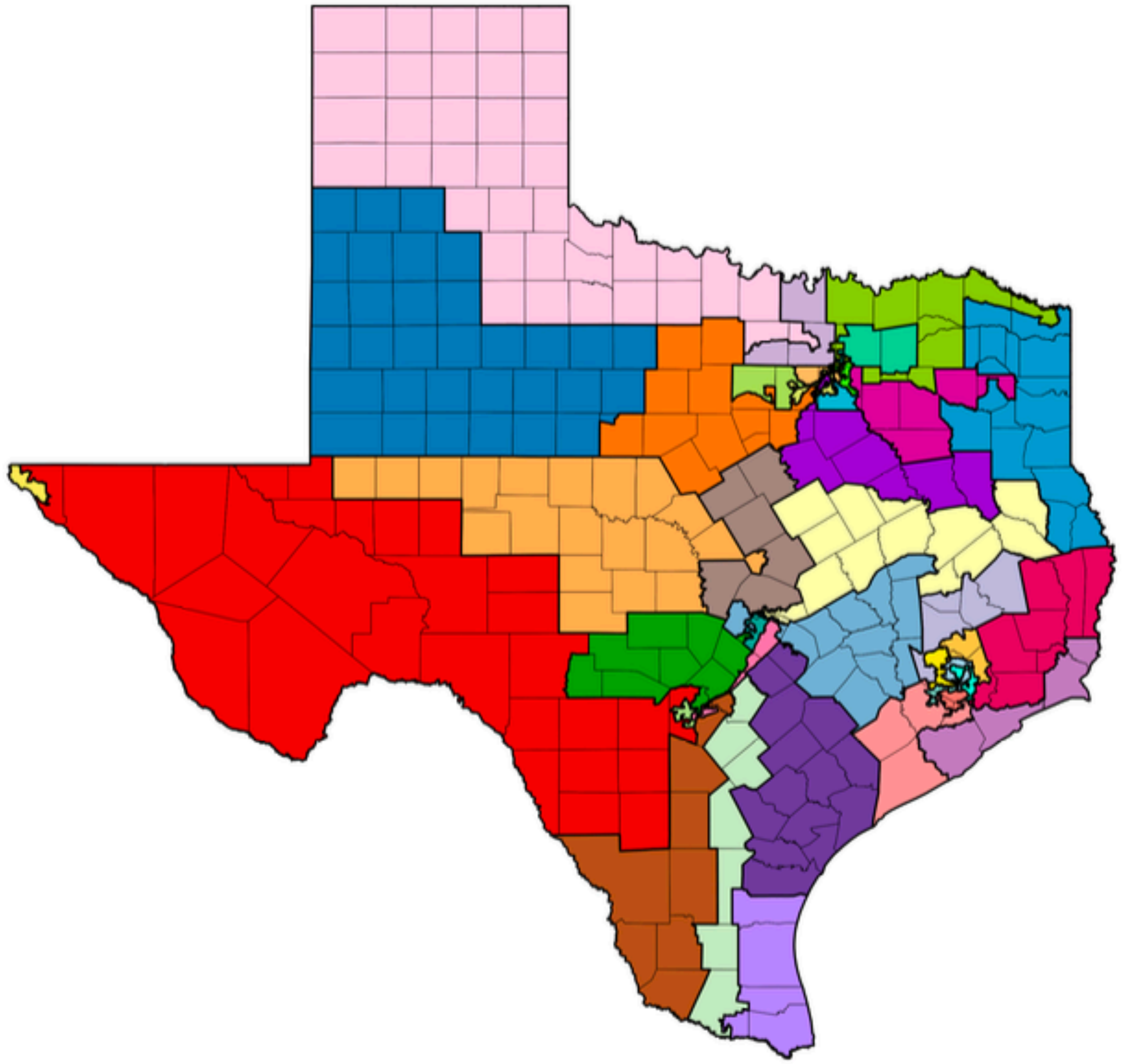
VRA case on behalf of Black-Latino-Asian coalition undisputed racial polarization

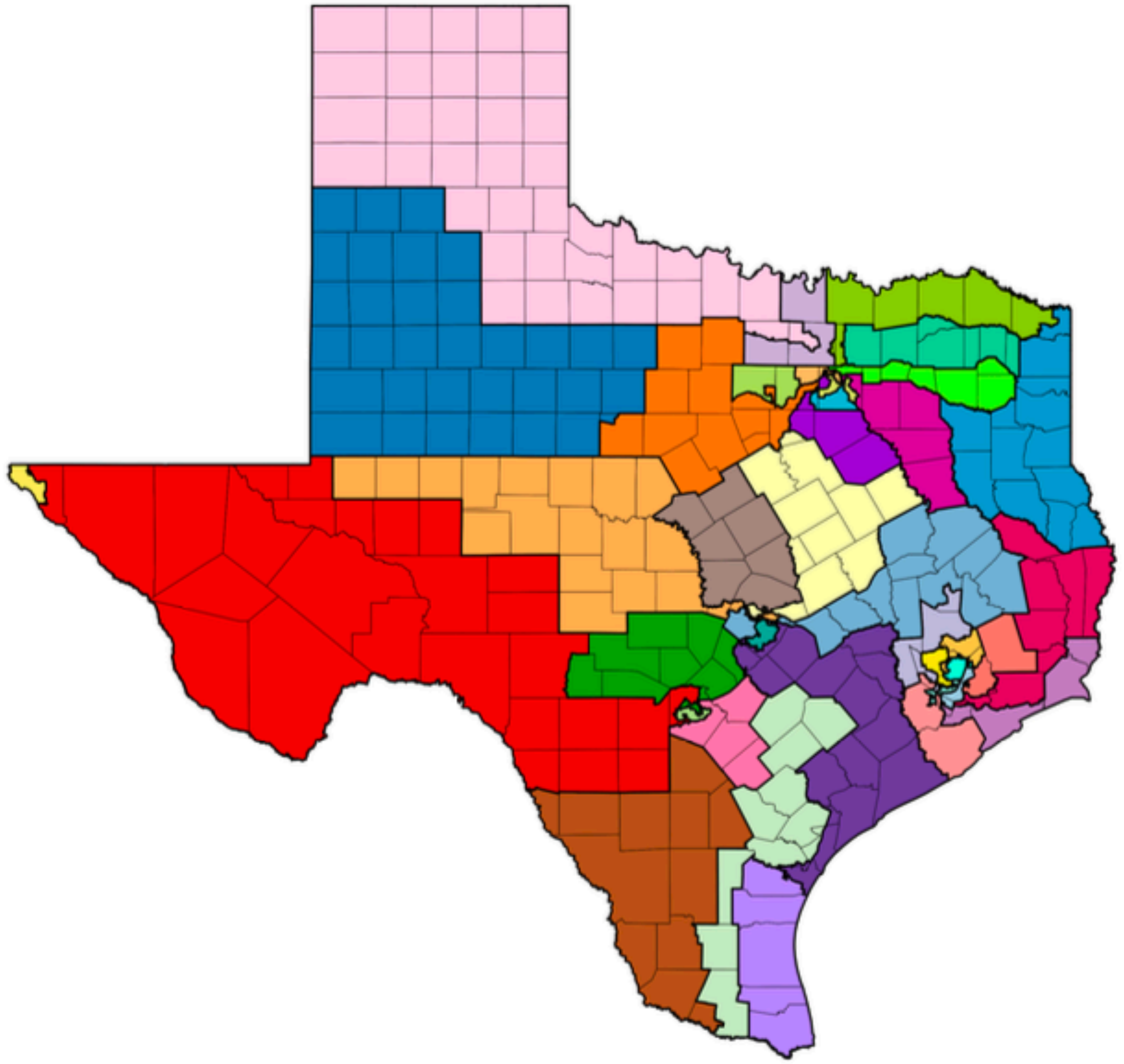


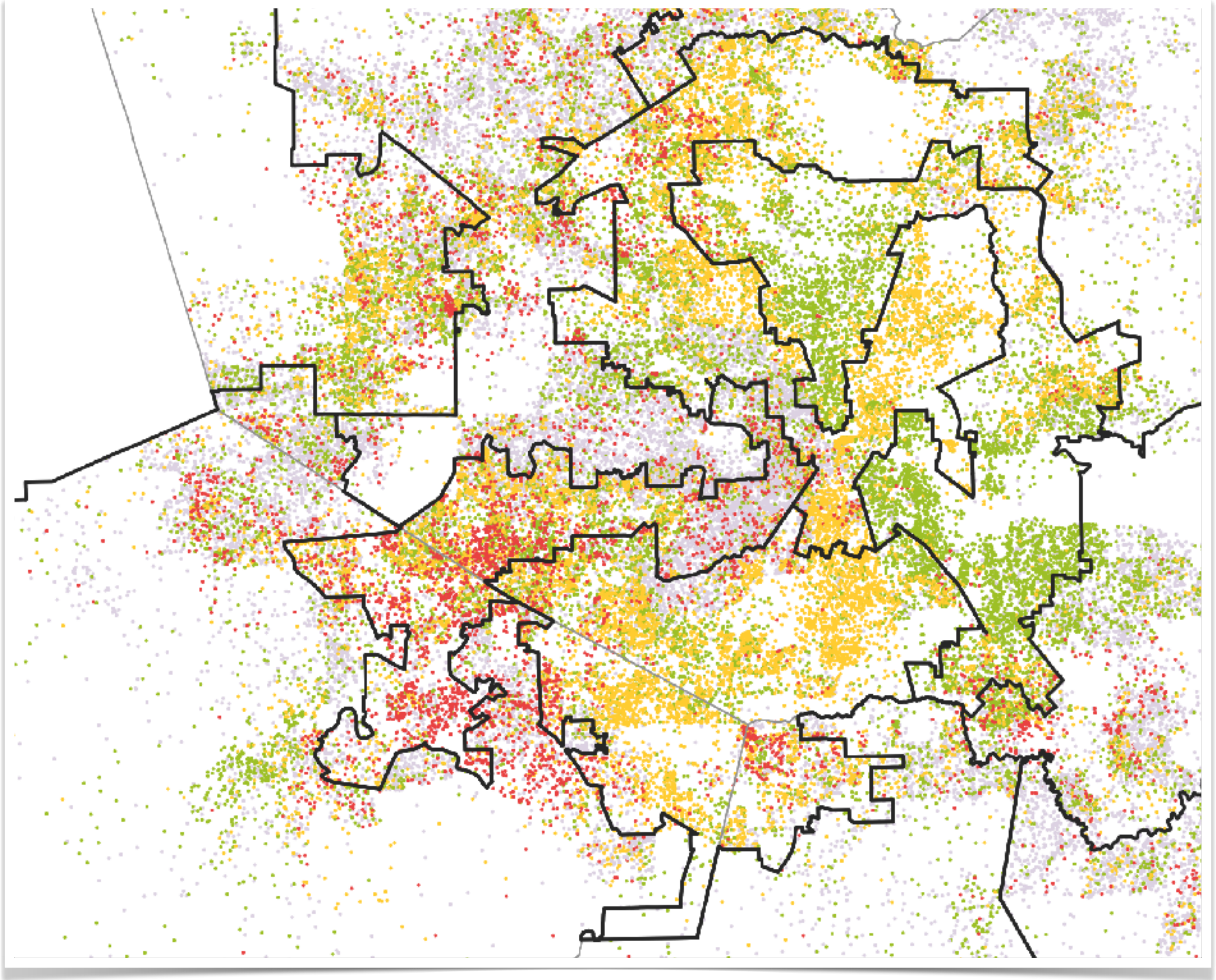


Houston-area districts:
extreme packing and cracking
compared to alternatives

August 29, 2025: Governor Abbott signs aggressive new partisan gerrymander into law

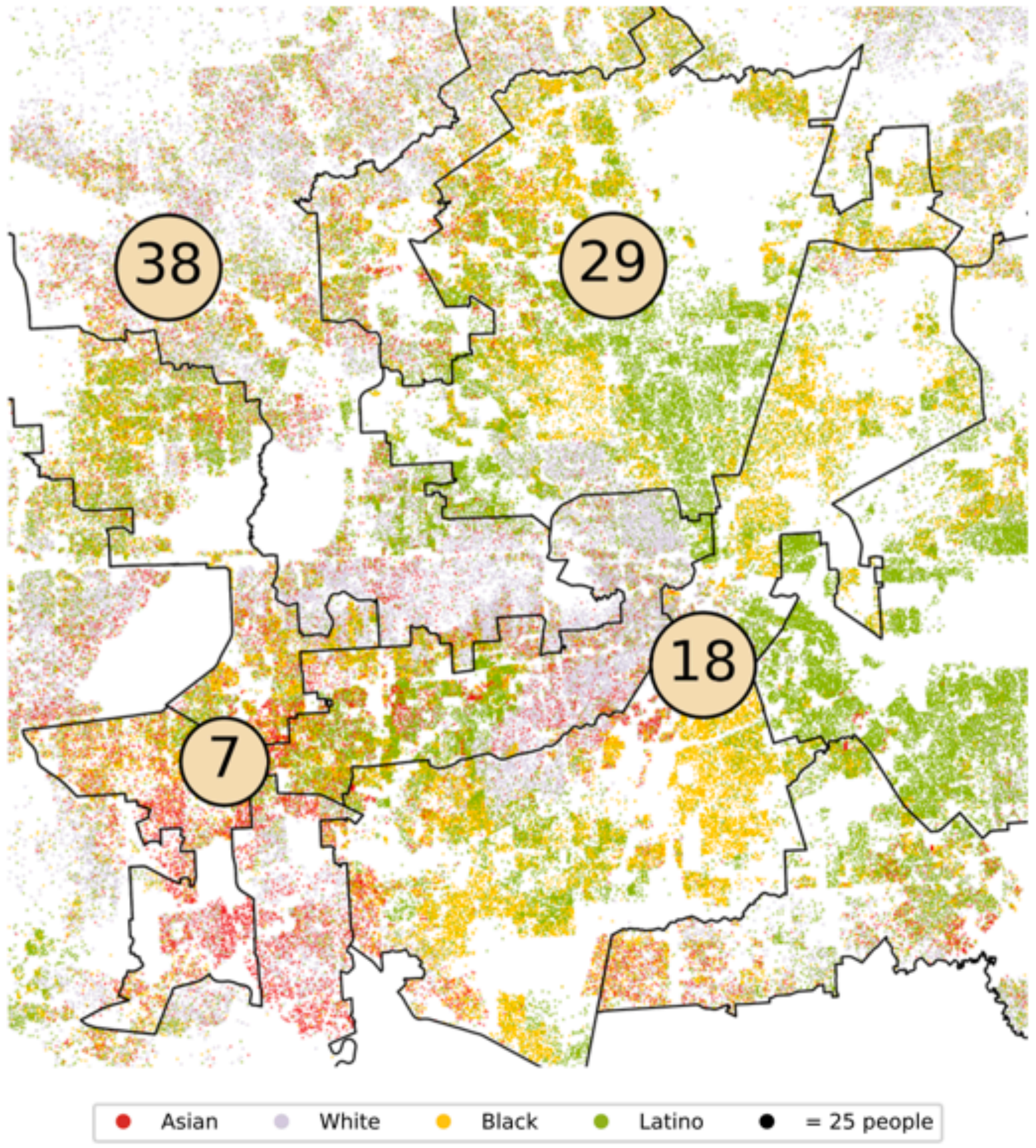






May plan in Houston area

Figure 5: Dot density from Cluster C2 in Harris/Ft Bend shows patterns of sorting by race.



August plan in Houston area

Texas, October 2025

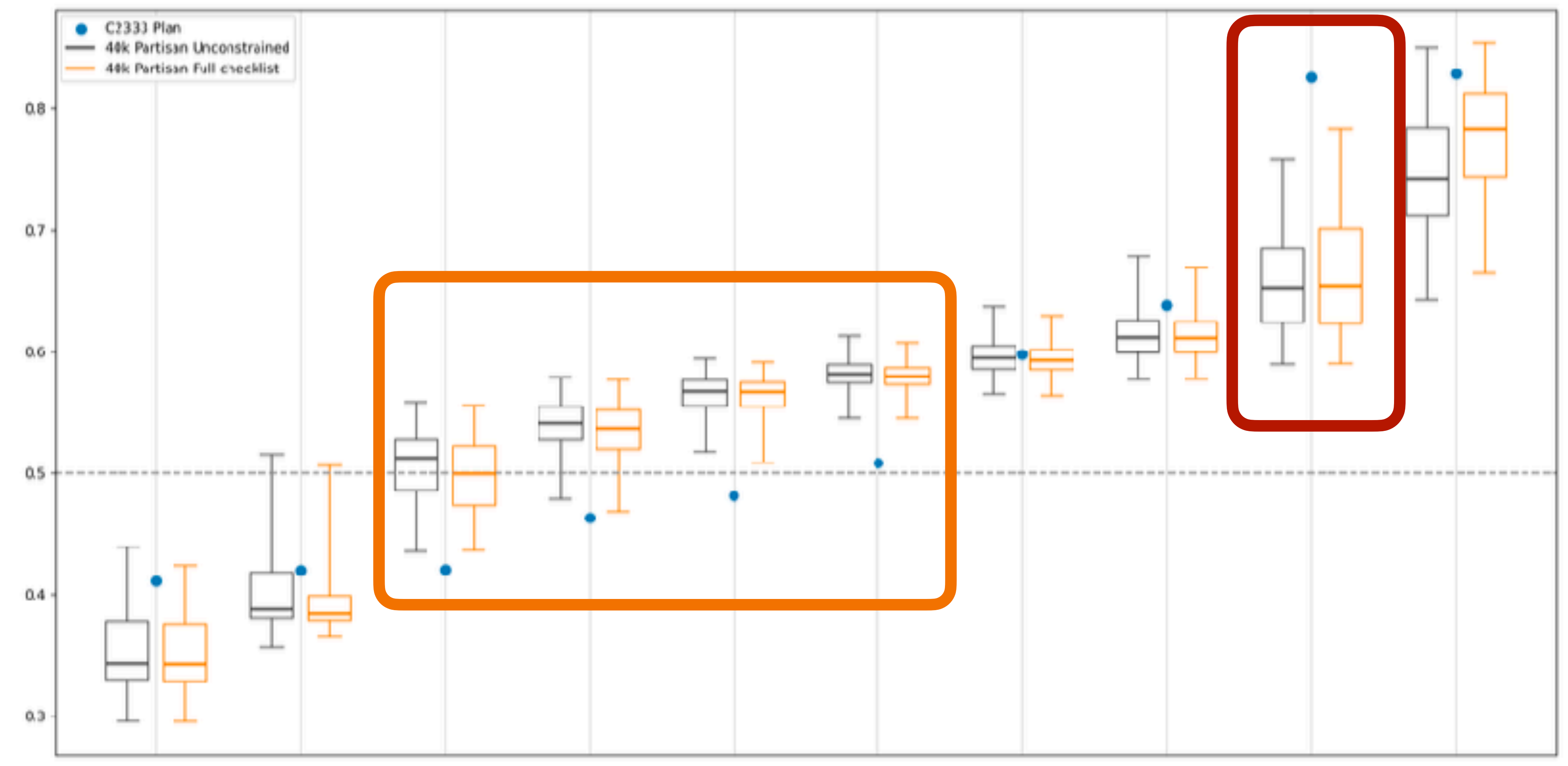
Constitutional case against racial gerrymander

		C2193 (2021)			C2333 (new)		
		Primary	General	Effect	Primary	General	Effect
C1	CD 5	13/14	0/14	Republican	10/14	0/14	Republican
	CD 6	13/14	0/14	Republican	13/14	0/14	Republican
	CD 12	12/14	0/14	Republican	13/14	0/14	Republican
	CD 24	7/14	0/14	Republican	7/14	0/14	Republican
	CD 25	13/14	0/14	Republican	14/14	0/14	Republican
	CD 30	14/14	14/14	POC-preferred D	14/14	14/14	POC-preferred D
	CD 32	8/14	14/14	**White D**	9/14	0/14	**Republican**
	CD 33	13/14	14/14	**POC-preferred D**	9/14	14/14	**White D**
C2	CD 2	10/14	0/14	Republican	9/14	0/14	Republican
	CD 7	7/14	14/14	White D	7/14	14/14	White D
	CD 8	11/14	0/14	Republican	12/14	0/14	Republican
	CD 9	11/14	14/14	**POC-preferred D**	13/14	0/14	**Republican**
	CD 14	11/14	0/14	Republican	11/14	0/14	Republican
	CD 18	11/14	14/14	POC-preferred D	11/14	14/14	POC-preferred D
	CD 22	10/14	0/14	Republican	11/14	0/14	Republican
	CD 29	13/14	14/14	POC-preferred D	12/14	14/14	POC-preferred D
	CD 36	10/14	0/14	Republican	11/14	0/14	Republican
	CD 38	6/14	0/14	Republican	7/14	0/14	Republican
C3	CD 10	10/14	0/14	Republican	8/14	0/14	Republican
	CD 11	12/14	0/14	Republican	11/14	0/14	Republican
	CD 20	13/14	14/14	POC-preferred D	13/14	14/14	POC-preferred D
	CD 21	10/14	0/14	Republican	10/14	0/14	Republican
	CD 23	13/14	0/14	Republican	11/14	0/14	Republican
	CD 27	13/14	0/14	Republican	10/14	0/14	Republican
	CD 35	11/14	14/14	**POC-preferred D**	12/14	0/14	**Republican**
	CD 37	6/14	14/14	White D	7/14	14/14	White D

New modeling challenge: the “Clarence Thomas checklist”!

Take into account partisanship plus counties, cities, transportation networks, urban-rural balance, etc

Figure 9: Cluster C2 (Harris/Ft Bend): This time, four of ten districts—again, all with expected POC CVAP near 50%—have outlyingly low levels of minority citizens, while one district far above 50% is elevated to an outlying degree. Filtering by the full checklist of TDPs (orange) does not change this finding.



WE MIGHT GET A TEXAS DECISION

TODAY

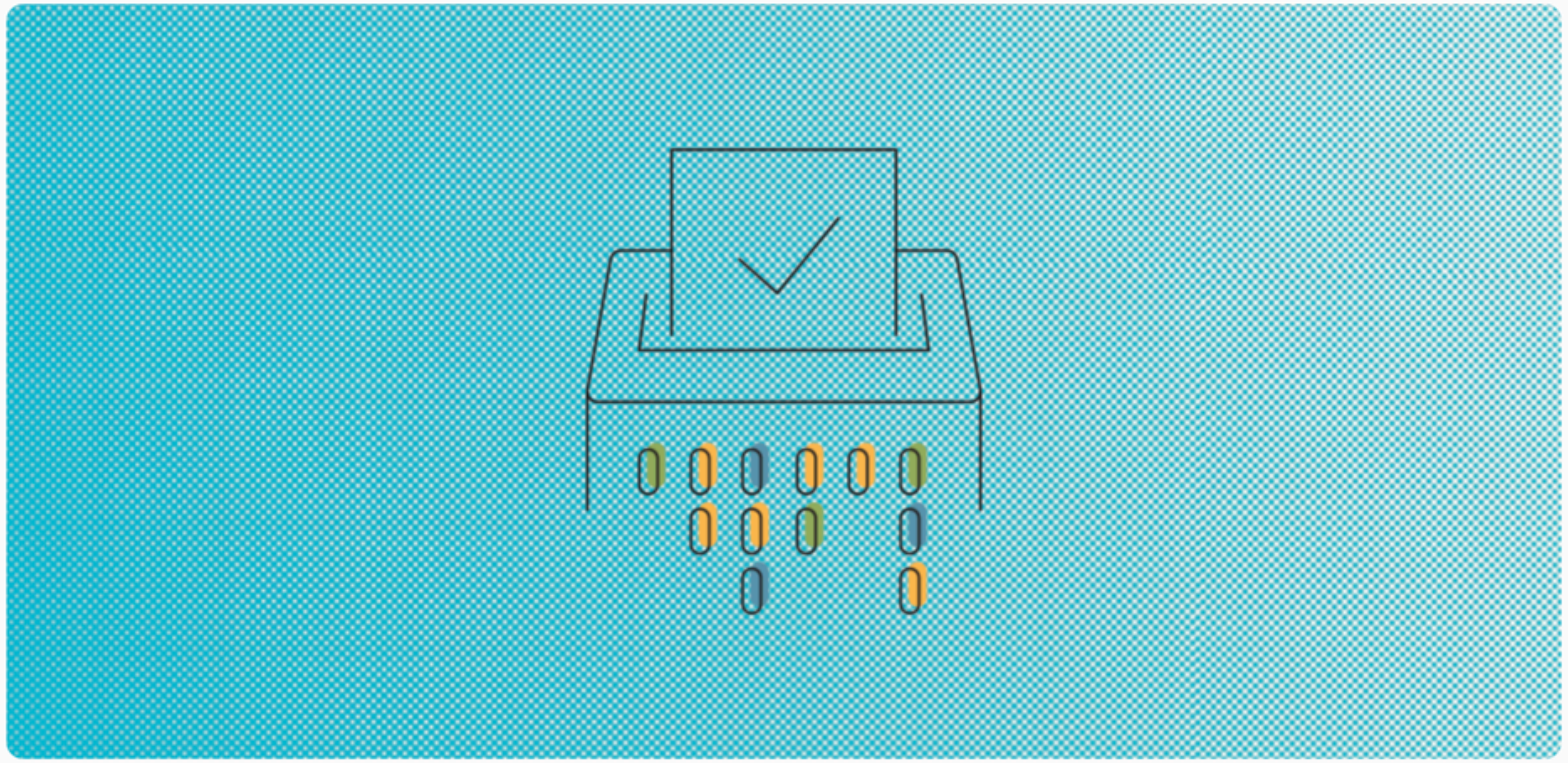
😬

THANK YOU

and I’d be happy to hear from you

(mduchin@uchicago.edu)

THE UNIVERSITY OF CHICAGO
DATA SCIENCE
INSTITUTE



RESEARCH INITIATIVE

Data & Democracy

A collaboration between the DSI and the Center for Effective Government that conducts cross-disciplinary research, convenes key stakeholders, and circulates and amplifies the findings needed to protect democracy in the digital age.