



WELCOME TO THE VOTING RIGHTS DATA INSTITUTE!

June 4, 2018

AMERICAN DEMOCRACY: CHOP AND VOTE



ENUMERATION; APPORTIONMENT; PARTITION; LITIGATION

- Two federal legislative bodies: U.S. Senate and House of Reps
 - ★ Senate: 2 per state
 - ★ House: state's # reps roughly proportional to its share of the U.S. population, adding up to 435
- Every ten years there is a Census to **enumerate** the population
- Then we **apportion** Representatives to the House
- States are required to **redistrict** after each Census



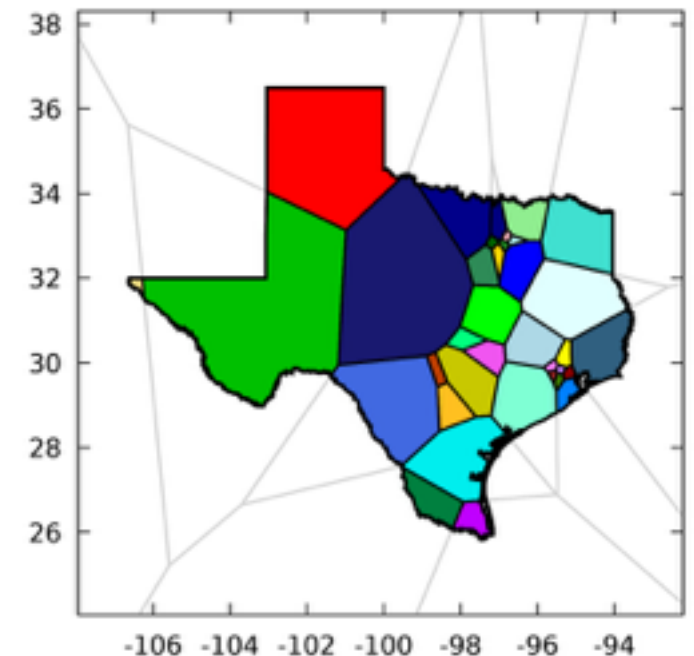
Officials set the boundary lines for their own elections!

SIX “TRADITIONAL DISTRICTING PRINCIPLES,” AND OTHER NORMS

.....

- Equal population - *number*
- Compactness - *shape*
- Contiguity - *shape*

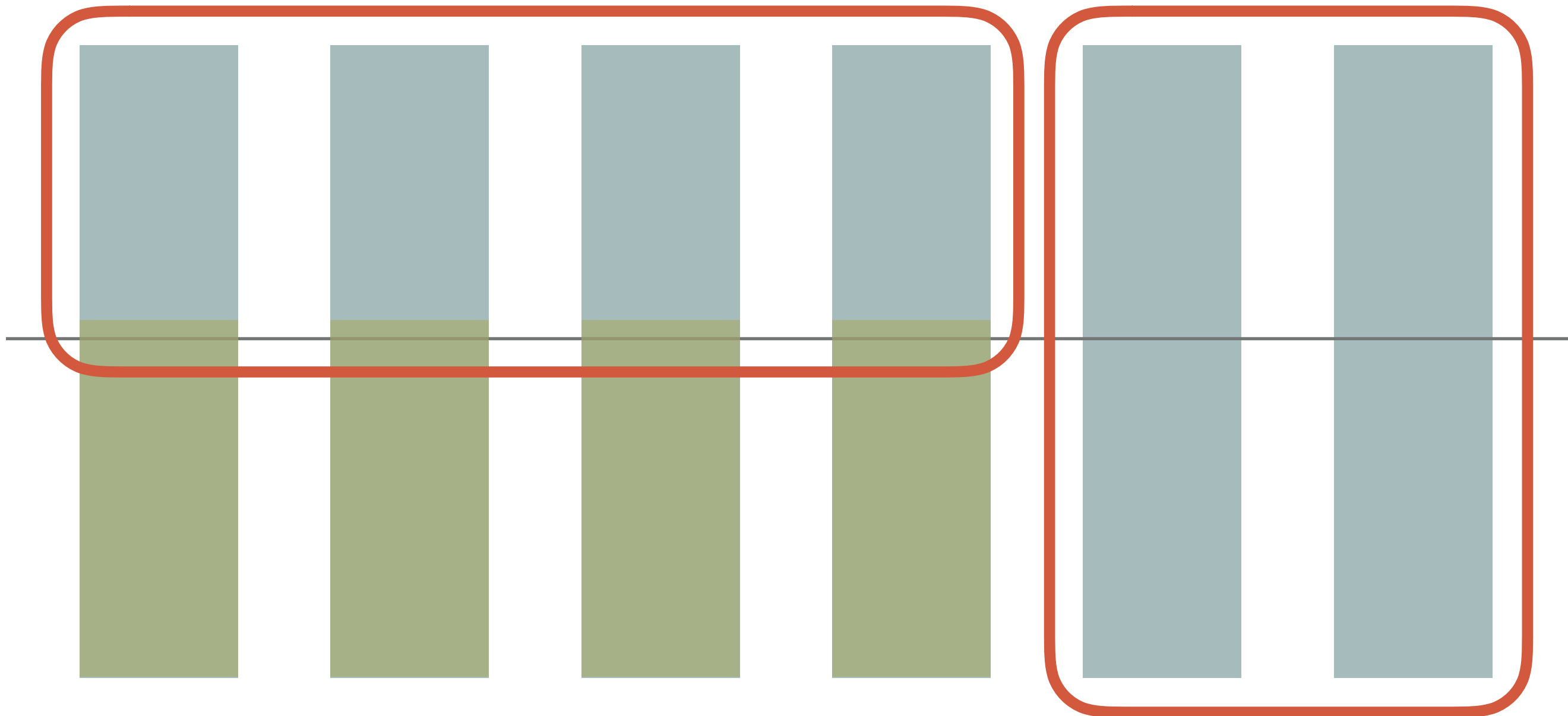
source: Cohen-Addad–
Klein-Young



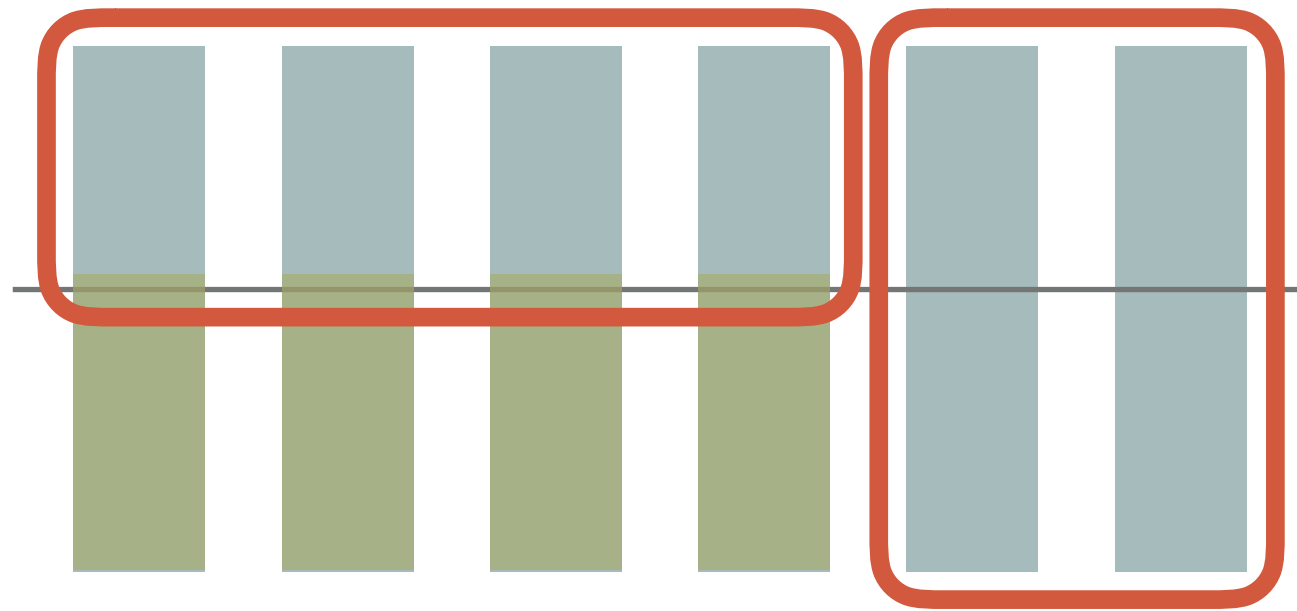
- Respect for county/city boundaries - *political geography*
- Respect for “communities of interest” - *sociology*
- Compliance with Voting Rights Act - *race*
- *Proportionality? Competitiveness? Responsiveness? Stability?*

Cracking

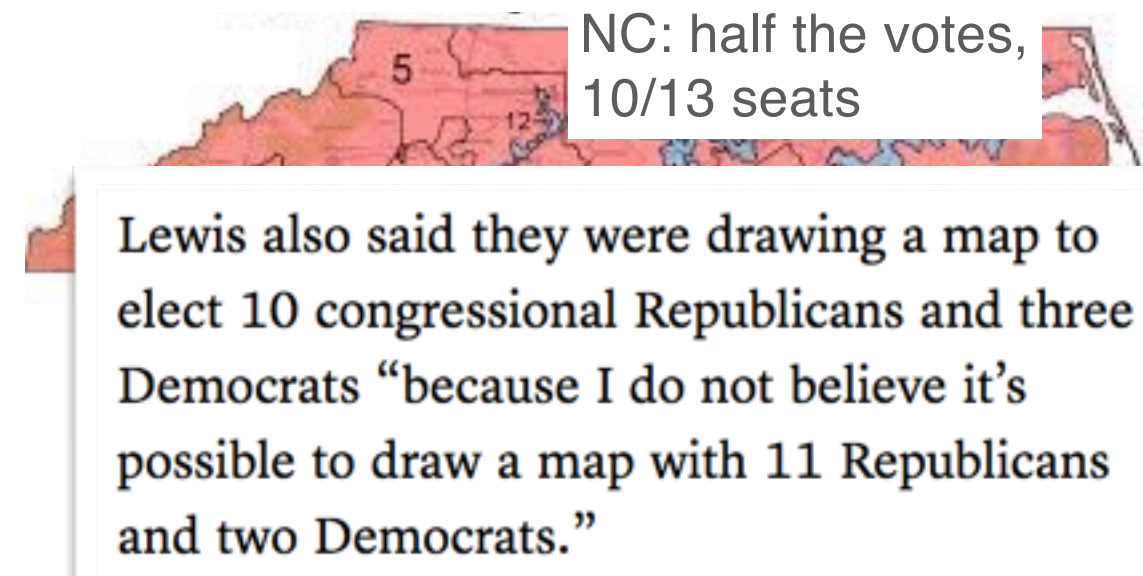
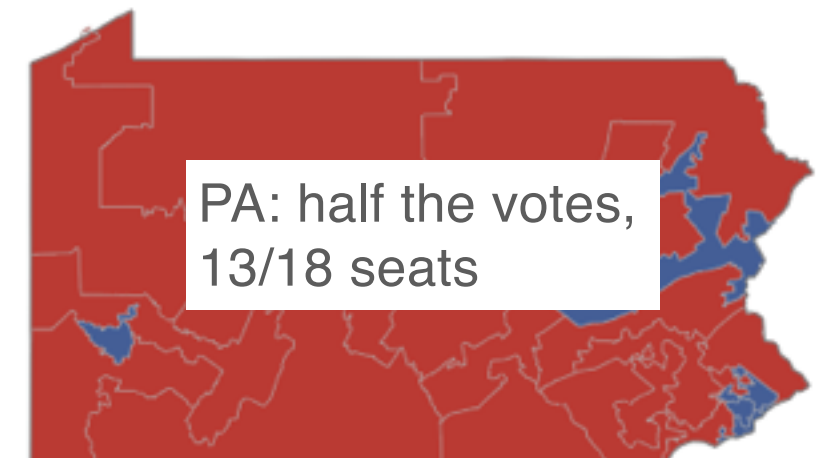
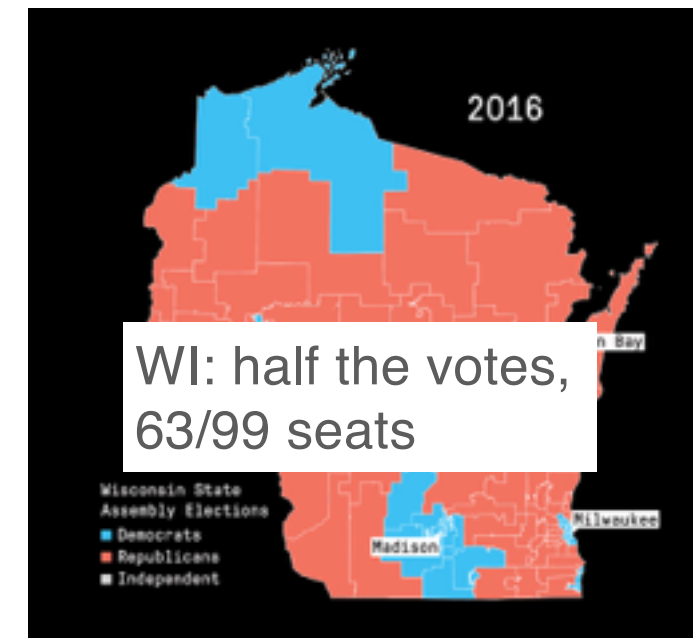
Packing



Efficient majorities for you = Packing and Cracking for your opponents



- So theoretically it's possible to get a seat share that is double your vote share, if you were unconstrained by geography. How does it actually play out?
- Key point: Rs now have 32/50 legislatures, 32/50 governors, and “trifectas” in 26/50 states (vs 8 Dem).
- Was not always so, and both parties gerrymander rampantly when they can!

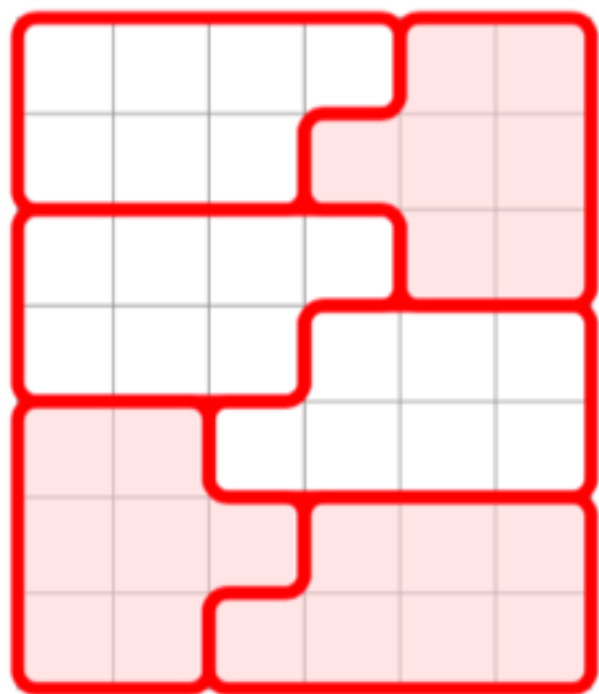


THE LONG CAREER OF THE EYEBALL TEST

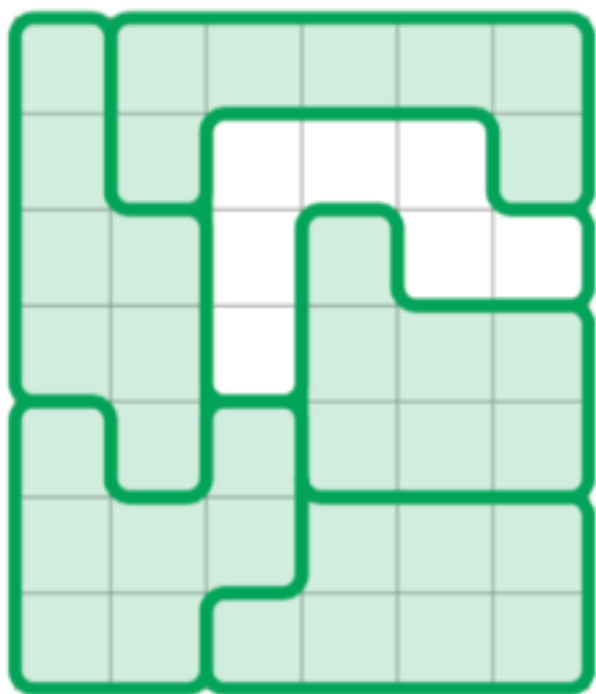


TRADITIONAL PRINCIPLE: DISTRICTS SHOULD BE COMPACT

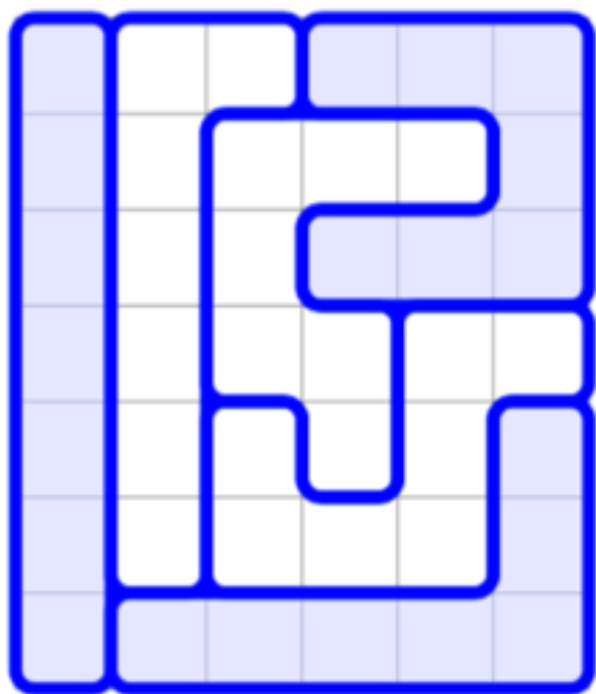
.....



Plump



Slightly Snakey

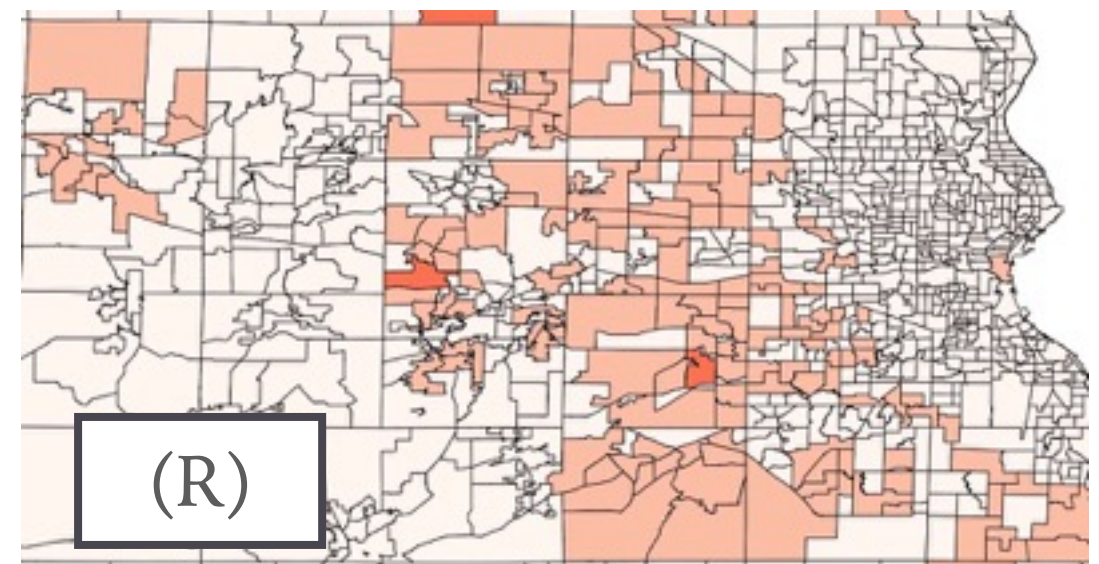
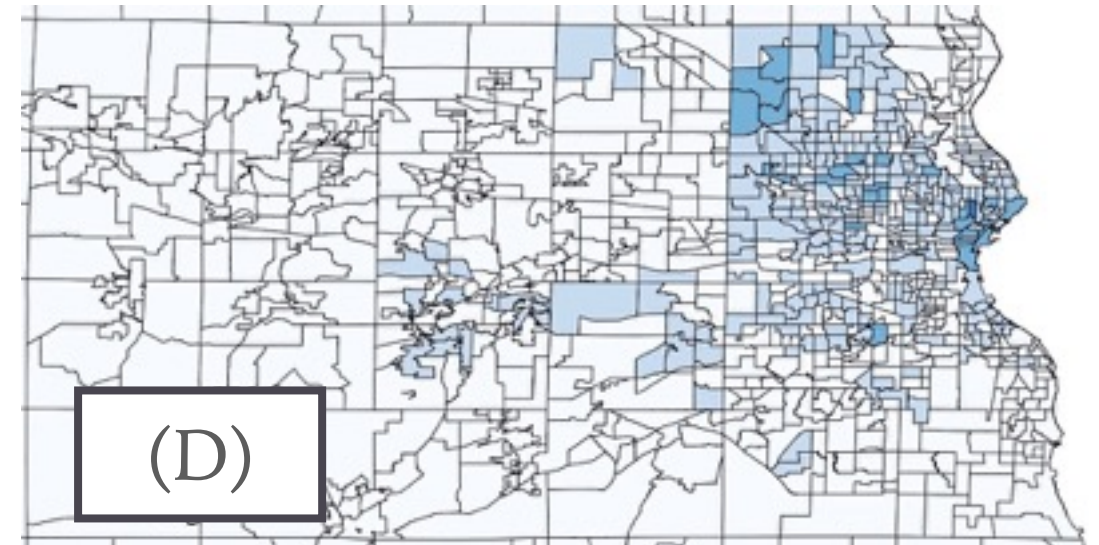


Reptilian



WHY COMPACTNESS?

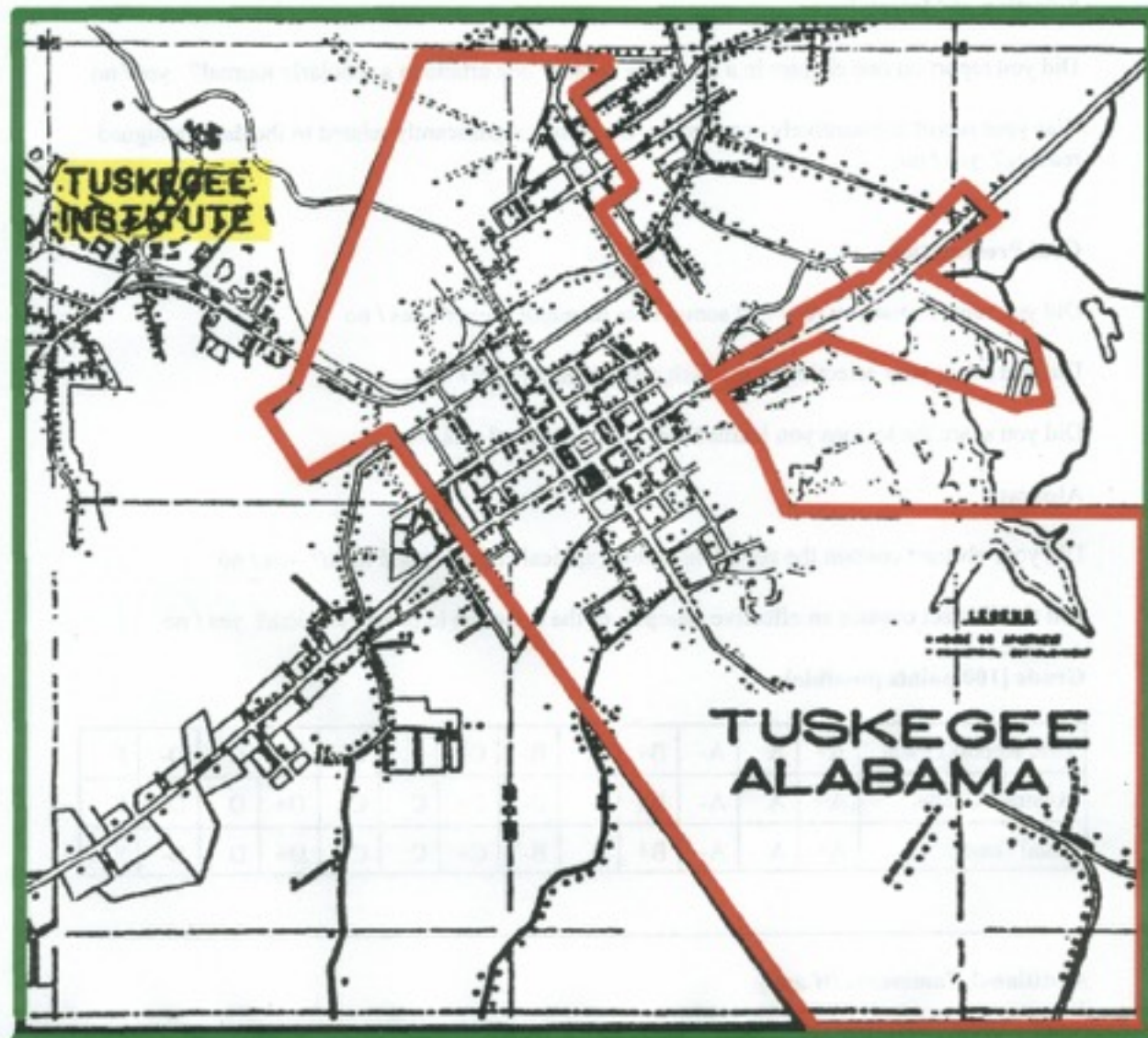
- Packing and cracking with perfect information about voters: districting pen must follow the distribution to be very efficient



EXAMPLE: TUSKEGEE TRANSFORMED

APPENDIX TO OPINION OF THE COURT.

CHART SHOWING TUSKEGEE, ALABAMA, BEFORE AND AFTER ACT 140



(The entire area of the square comprised the City prior to Act 140. The irregular red-bordered figure within the square represents the post-enactment city.)

- Gomillion v. Lightfoot (1960)
- Tuskegee redrew its lines in 1957

Before: **square**

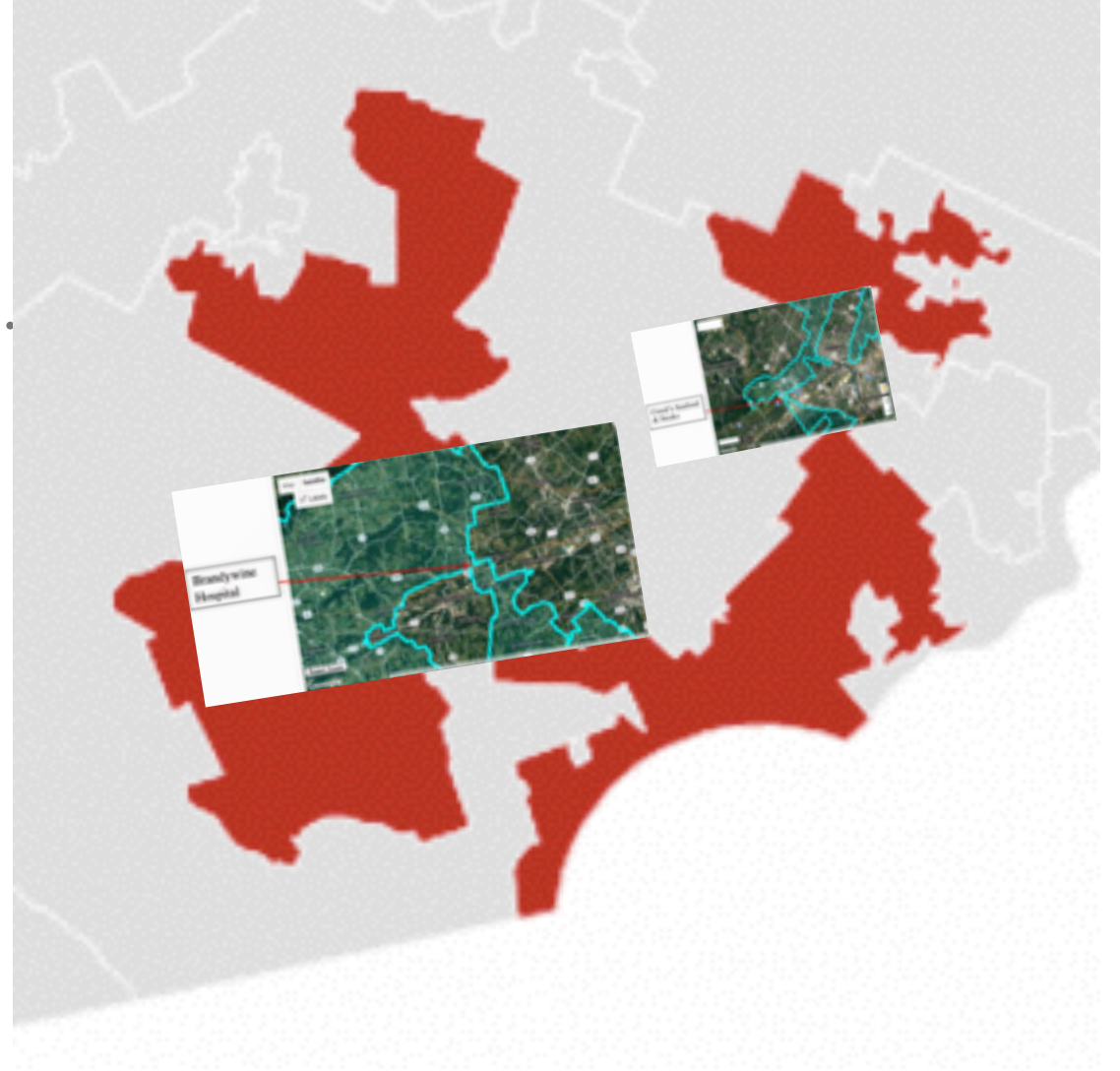
After: **28-sided polygon**

Before: **79% Black**

After: **100% White**

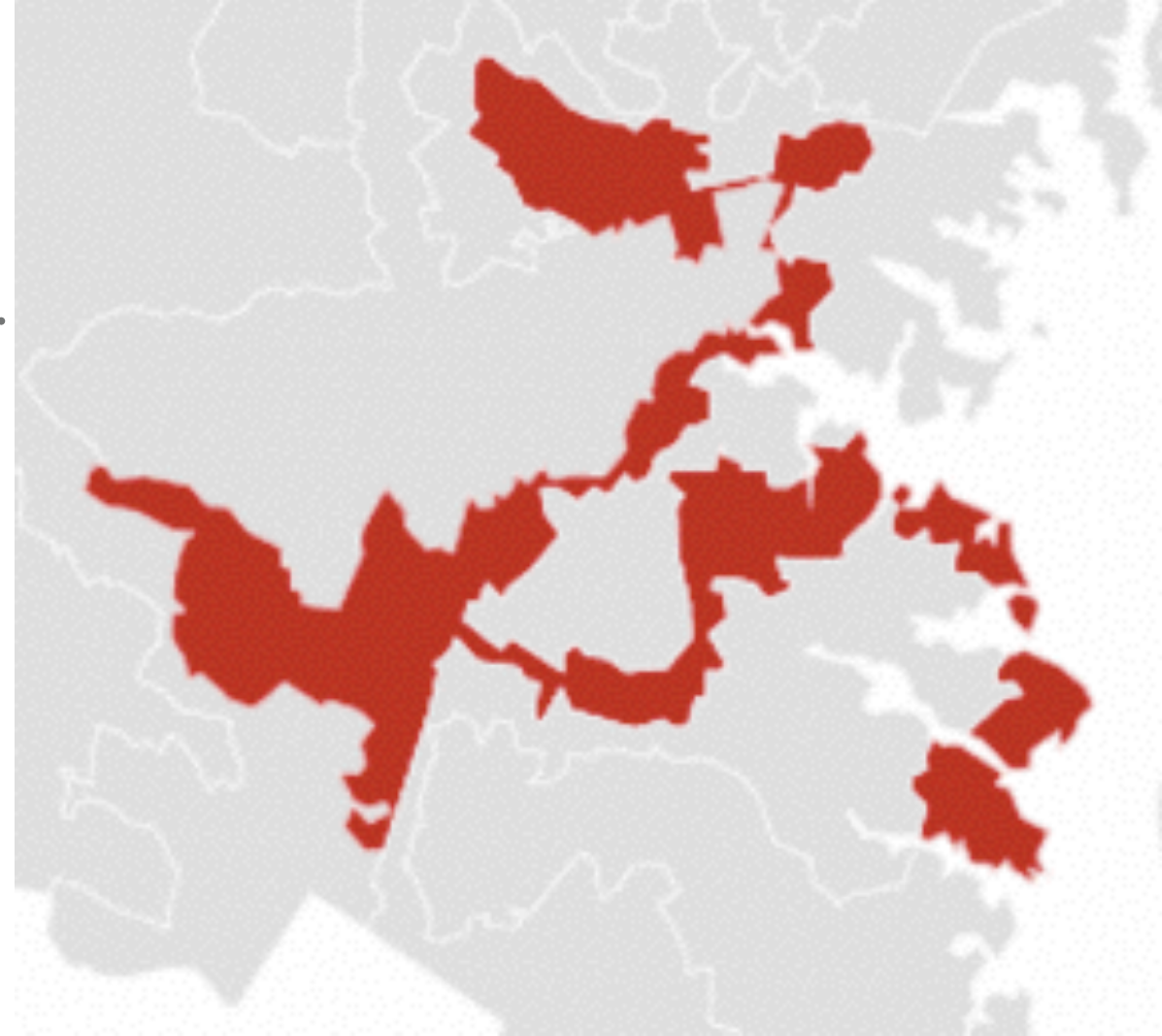
PENNSYLVANIA-7

- Winds through Philly suburbs to make a narrowly R district with a demographic hodgepodge
- 88% white; neatly avoids Black neighborhoods to its East side
- Can be disconnected by the removal of a single building in two different places



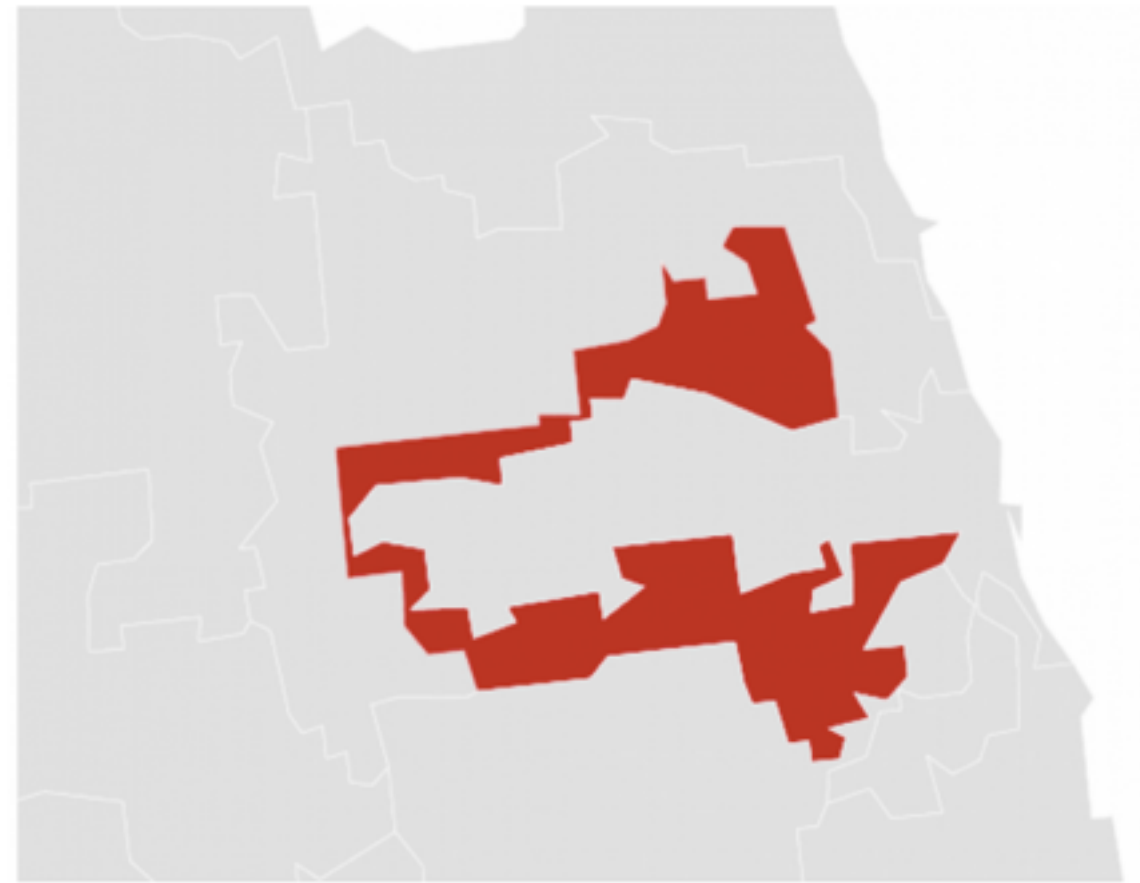
MARYLAND-3

- Plan drawn by Dems, opposed by civil rights and good-governance groups; pushed out 10-term R incumbent with influx of D and minority voters
- “Former Secretary of State John T. Willis, who was in charge of the redistricting as chairman of the Governor's Redistricting Advisory Committee, said the committee did not mean for the 3rd District to look like it does. **That's just how the numbers worked out**, he said. ‘It's a very complex situation, and population is the No. 1 driving characteristic,’ Willis said.... ‘All of our congressional districts don't deviate by more than one person’”



ILLINOIS-4

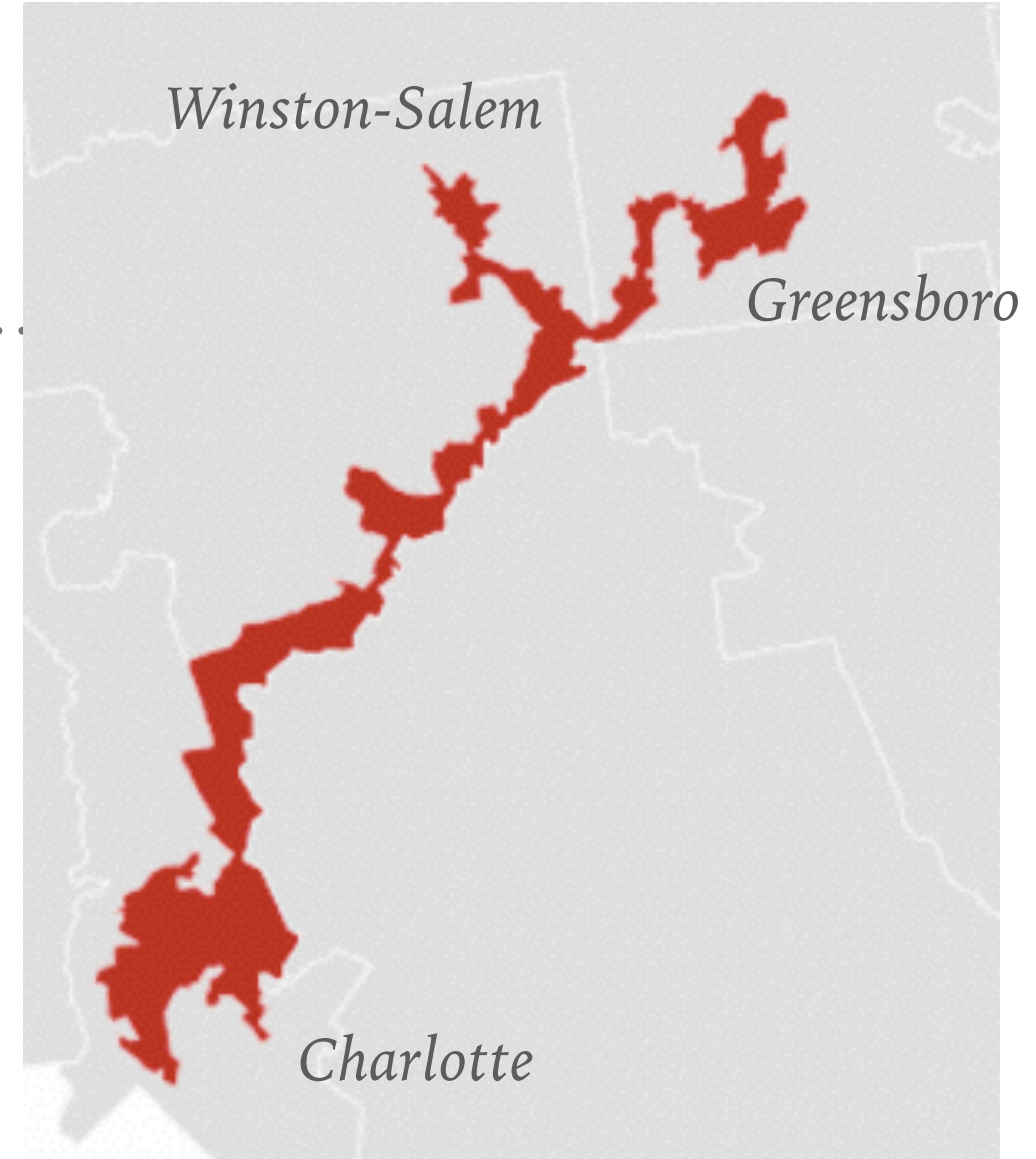
- Chunks of Pilsen (Mexican-American) and Humboldt Park (Puerto Rican) strung together by a highway
- **Luis Gutierrez** has won 13 straight elections, never less than 75% of the vote, often unopposed
- Drawn by civil rights advocates to create a Latino opportunity district fitting together with three majority-Black districts



Accidental packing!

NORTH CAROLINA-12

- The classic, the notorious, the district that launched 1000 lawsuits.
- Shaw v Reno was the first of a long line of Supreme Court basis throwing out districts on the basis of shape...
when race is in the mix
- Created after 1990 census and billed as a voting rights district; slices Black population out of three cities
- Race used as **proxy** for party by both sides: originally a Dem gerrymander, later Repub



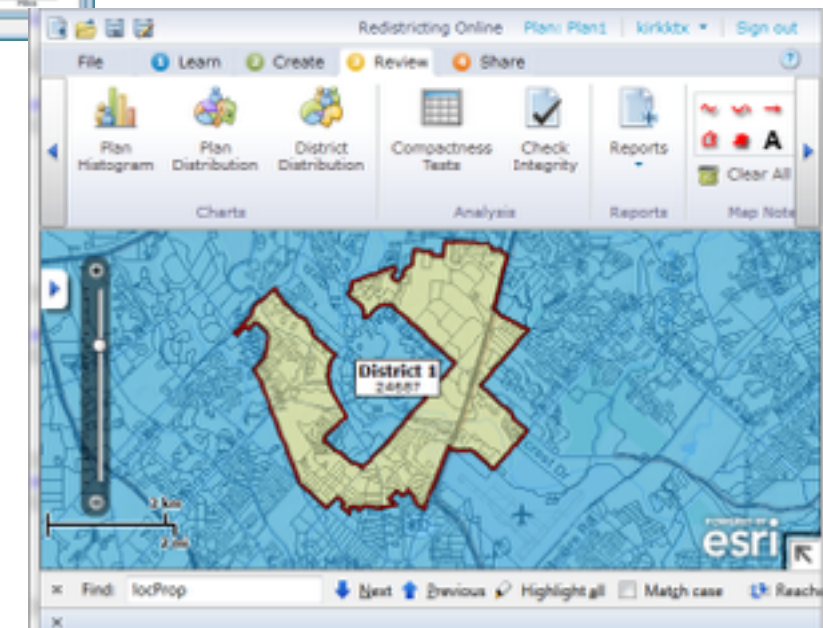
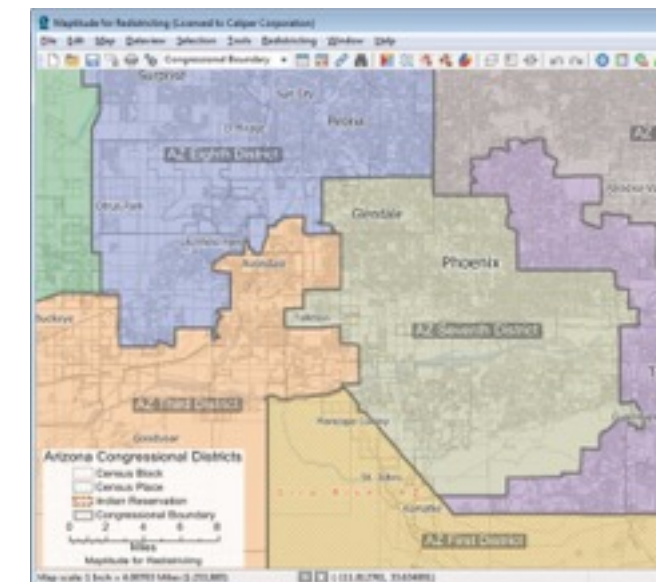
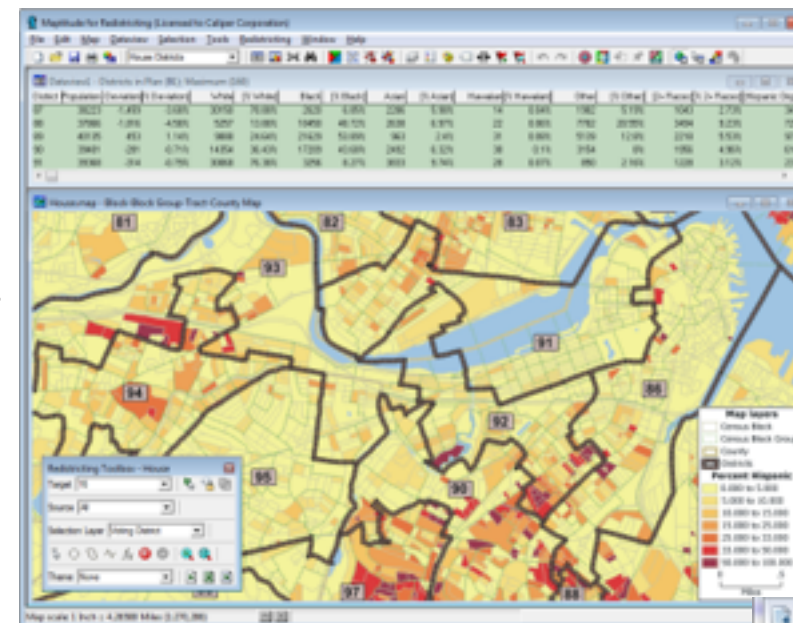
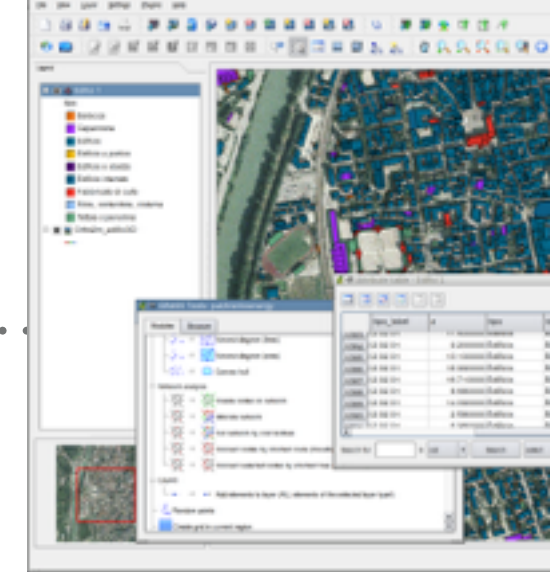
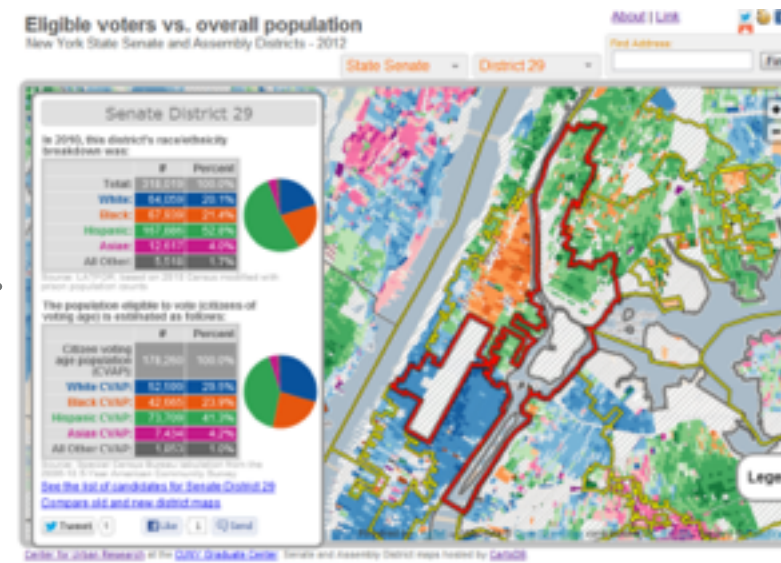
R.I.P.

EYEBALL TEST

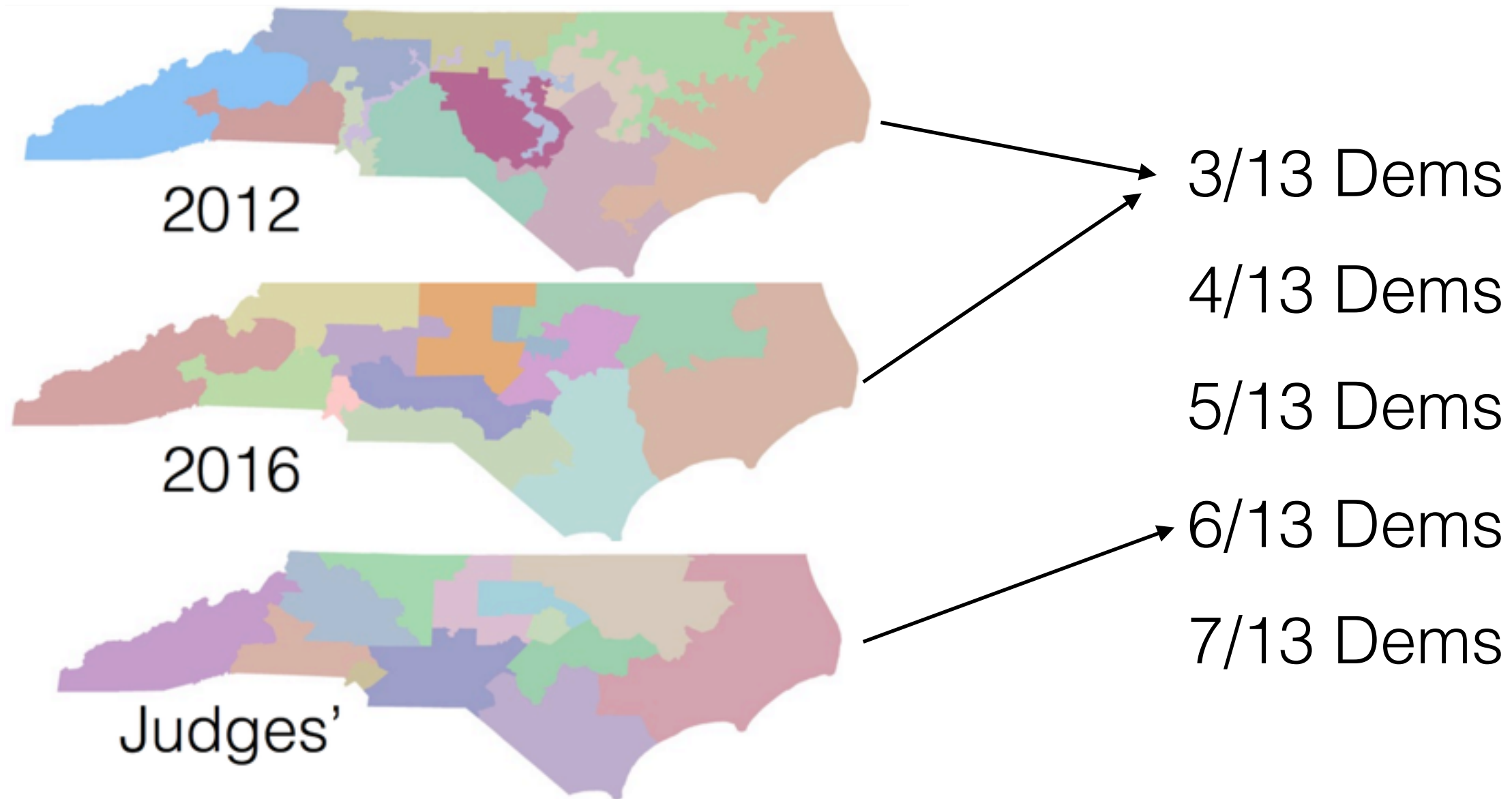


RICH DATA, AD HOC METHODS

- We have incredible descriptive and predictive data, plus the ability to overlay it on spatial “shapefiles”
- Maps still built by hand (??), but can now get extreme performance with contours that don’t tear the eye



AND IN NORTH CAROLINA....



Credit: Mattingly et al


EYEBALL-COMPACTNESS ALONE WON'T SAVE US

.....

- In the last cycle, when legislatures have been ordered to redraw their districts, they have come up with new maps that are less offensive-looking but perform just as extremely.



Share The Facts



Robin Hayes
chairman of the NC GOP



"A 'gerrymander' is by definition and common understanding, a strange looking 'monster' drawing."

North Carolina – Wednesday, January 10, 2018

[SHARE](#) [READ MORE](#)

POLITIFACT

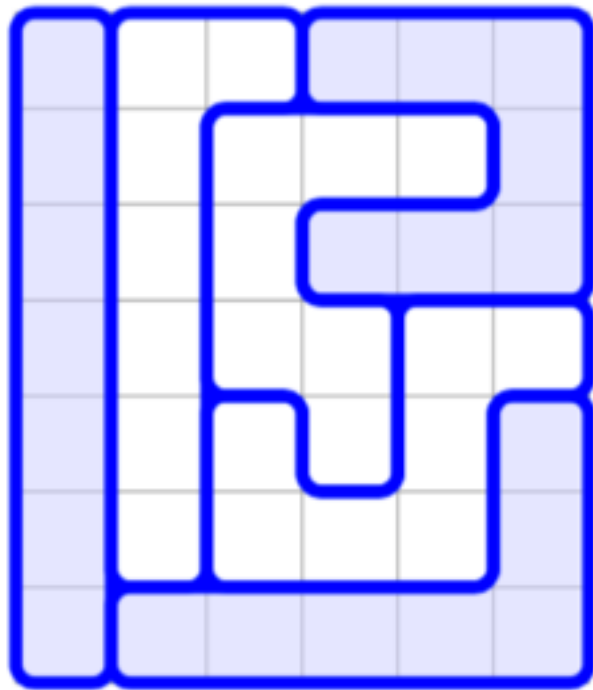
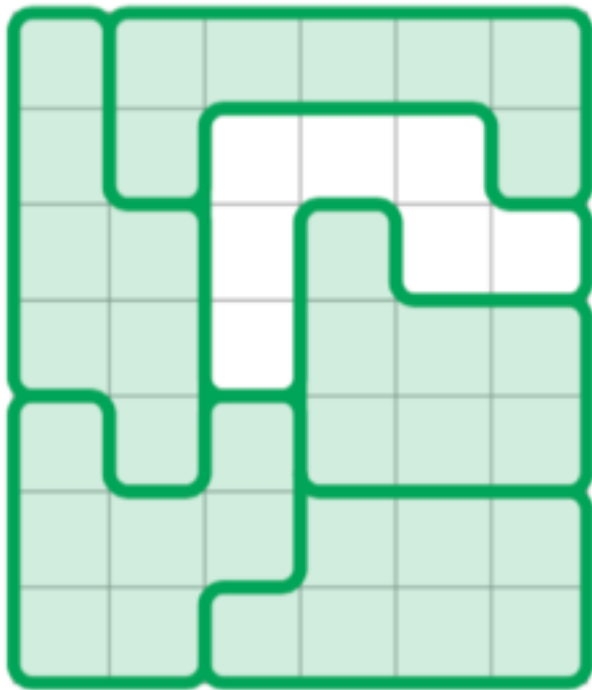
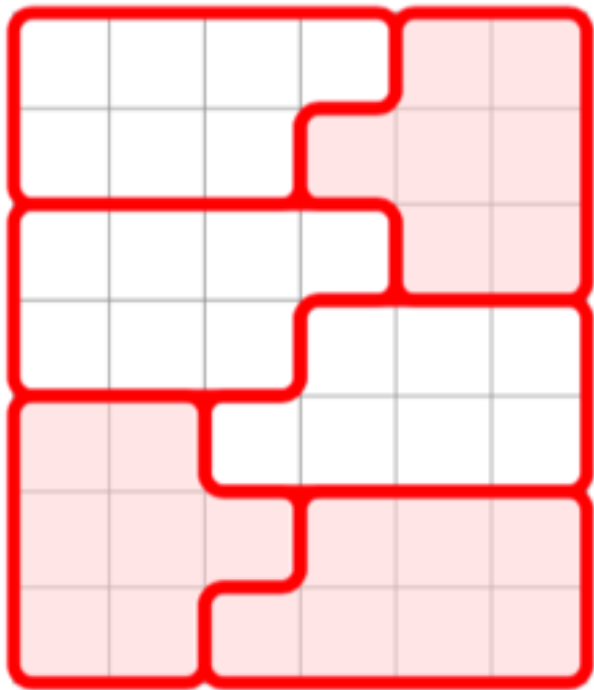
- “These days, with incredibly rich data available to the line-drawers, very skewed districts no longer have to look like exotic reptiles with fangs and claws” —me, in PolitiFact

**IF NOT (JUST) SHAPE,
THEN WHAT?**

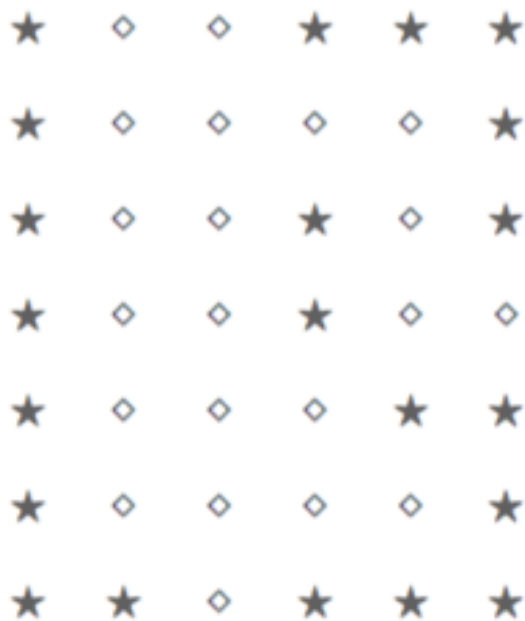
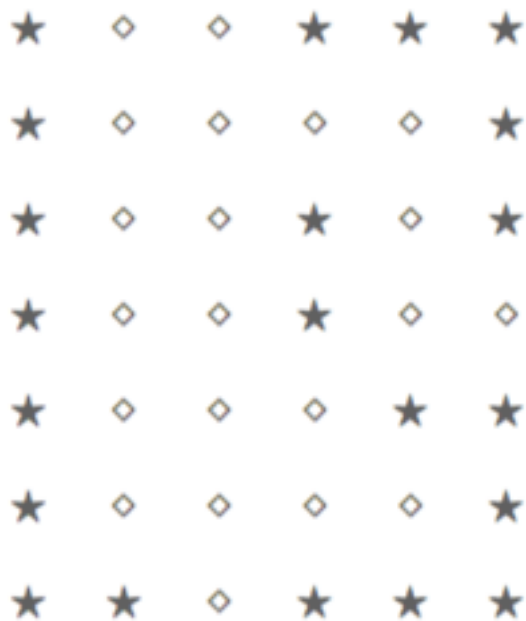
A DISTRICTING PLAN MEETS A DISTRIBUTION OF VOTERS

.....

\mathcal{D}



Δ



A DISTRICTING PLAN MEETS A DISTRIBUTION OF VOTERS

.....



(\mathcal{D}, Δ)

GETTING INFO OUT OF (\mathcal{D}, Δ)

- Δ alone determines V , the **vote share** for a party
- \mathcal{D} alone gives “**compactness scores**” (shape)
- (\mathcal{D}, Δ) pair gives (V_1, \dots, V_n) , the **vote share by district**
- (V_1, \dots, V_n) gives S , the **seat share** for a party (how many $V_i > 1/2$?)

*Isoperimetry,
Convex hull,
Dispersion,
etc*

*Symmetry,
Efficiency gap,
etc*

But I will argue that the best way to understand a plan is by keeping **all** the info in \mathcal{D} , Δ and varying them one at a time.

Fix \mathcal{D} , vary Δ :

Performance of plan under various voting scenarios

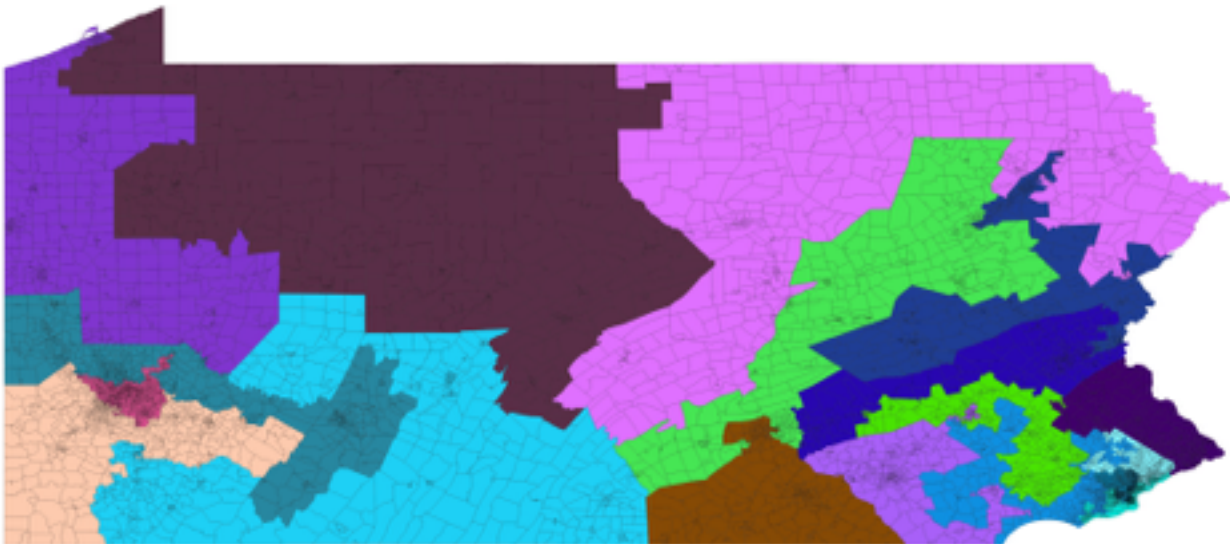
Fix Δ , vary \mathcal{D} a LOT:

How does a plan perform compared to **available alternatives**?



**LET'S TRY OUR EYEBALLS ON
PENNSYLVANIA**

THE CASE OF PENNSYLVANIA



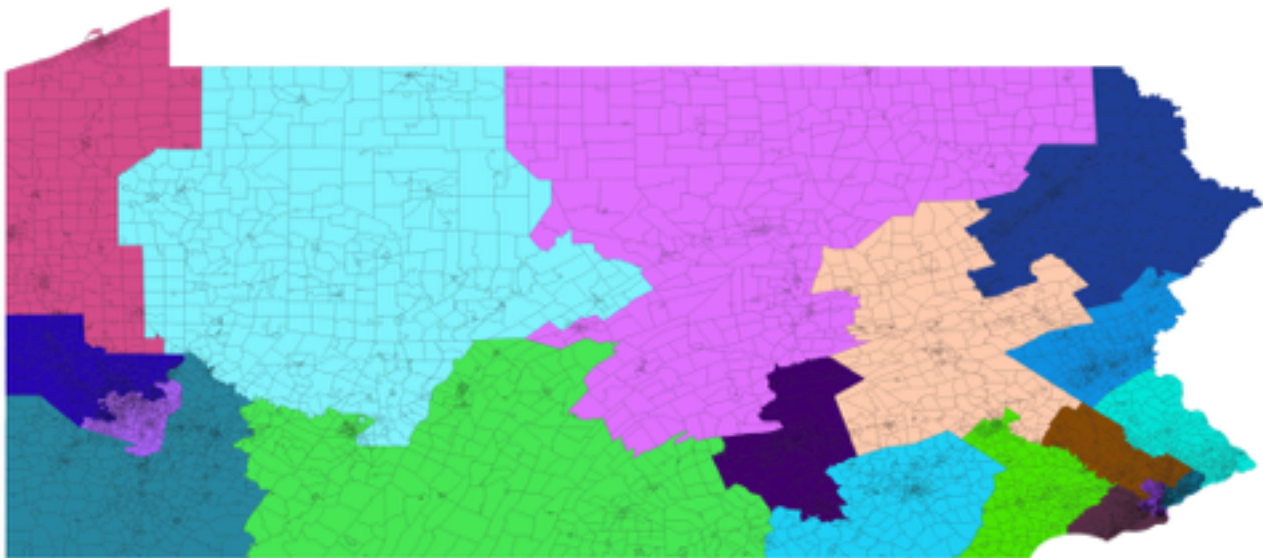
Legislature – Enacted 2011



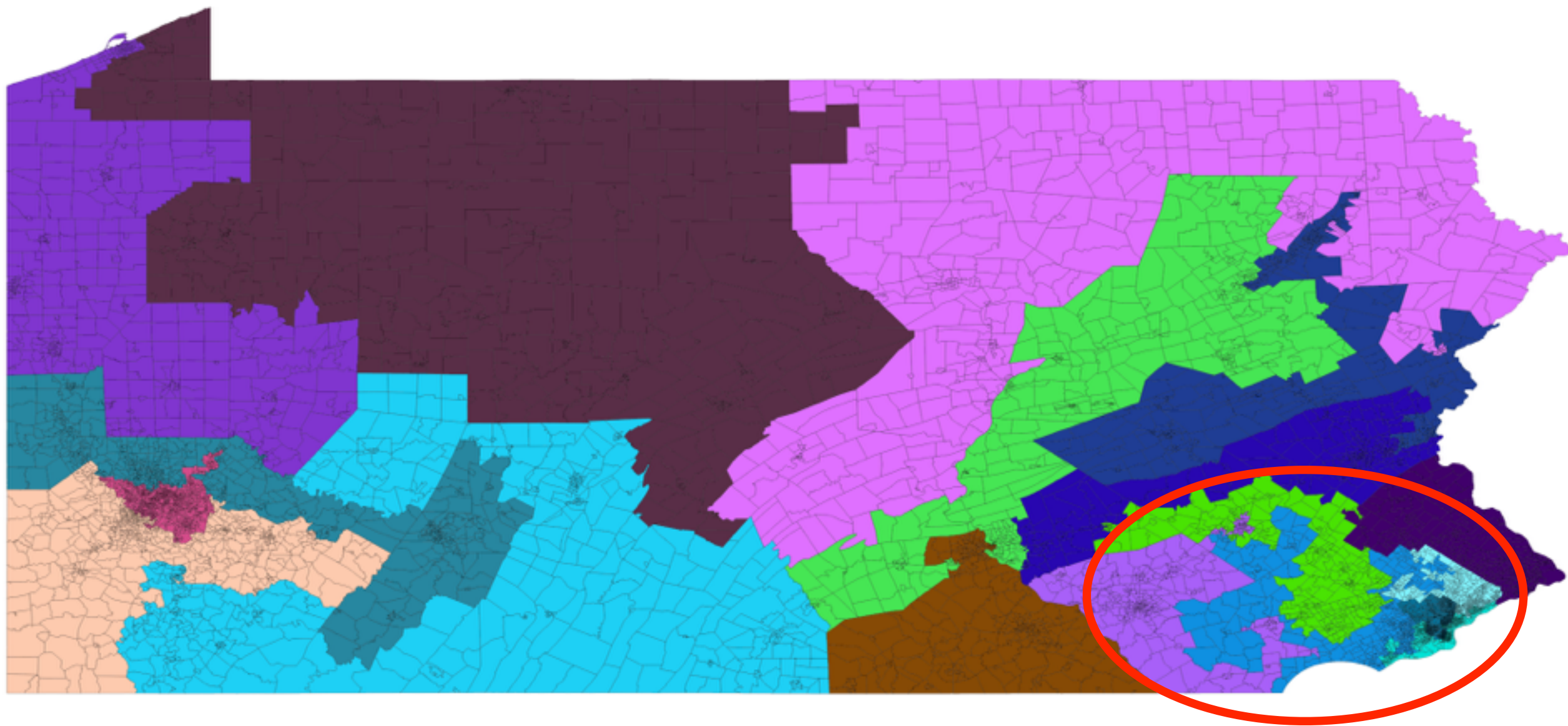
Legislature – Proposed



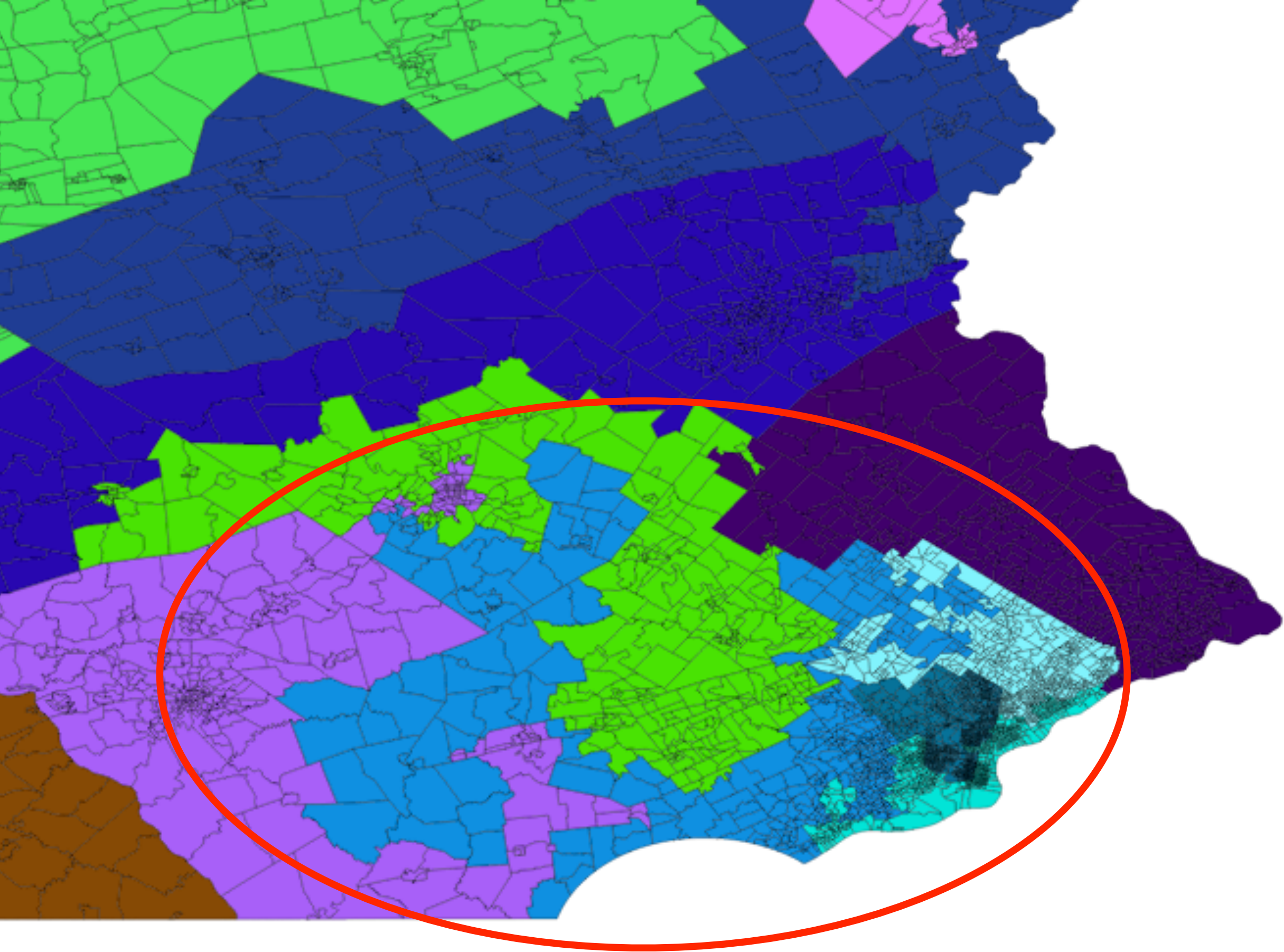
Governor – Proposed



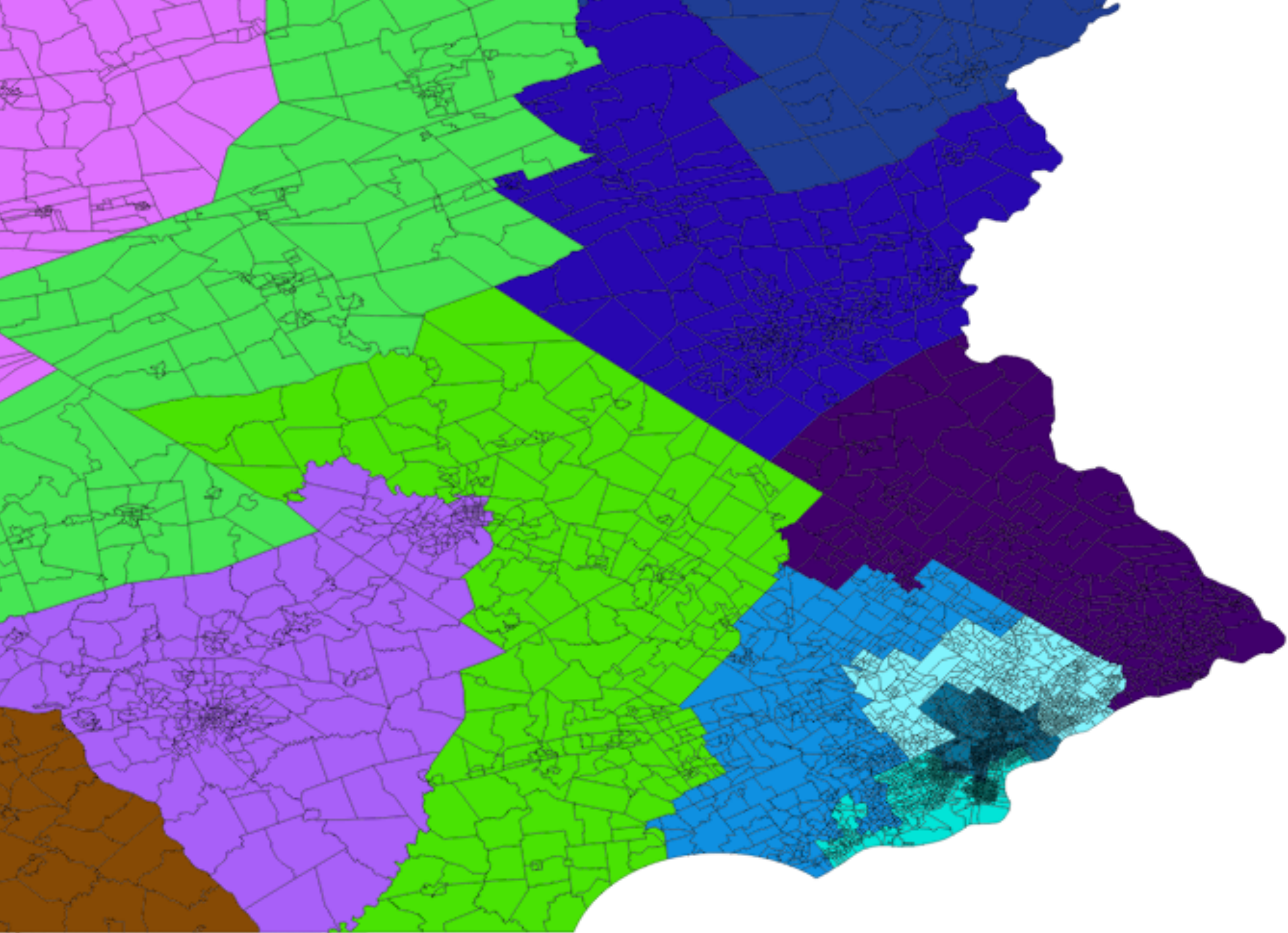
Remedial – Enacted 2018



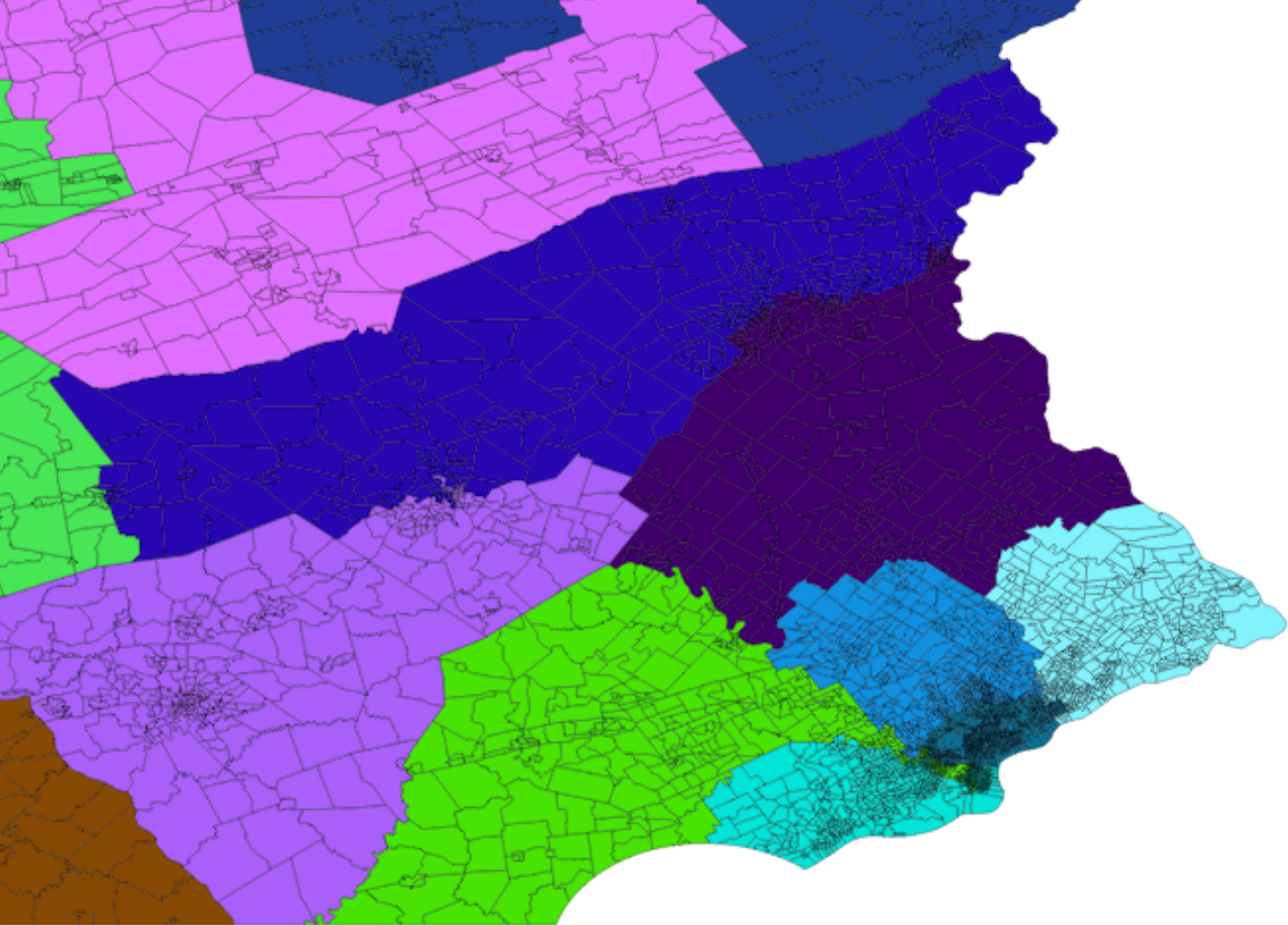
“Goofy kicking Donald Duck”



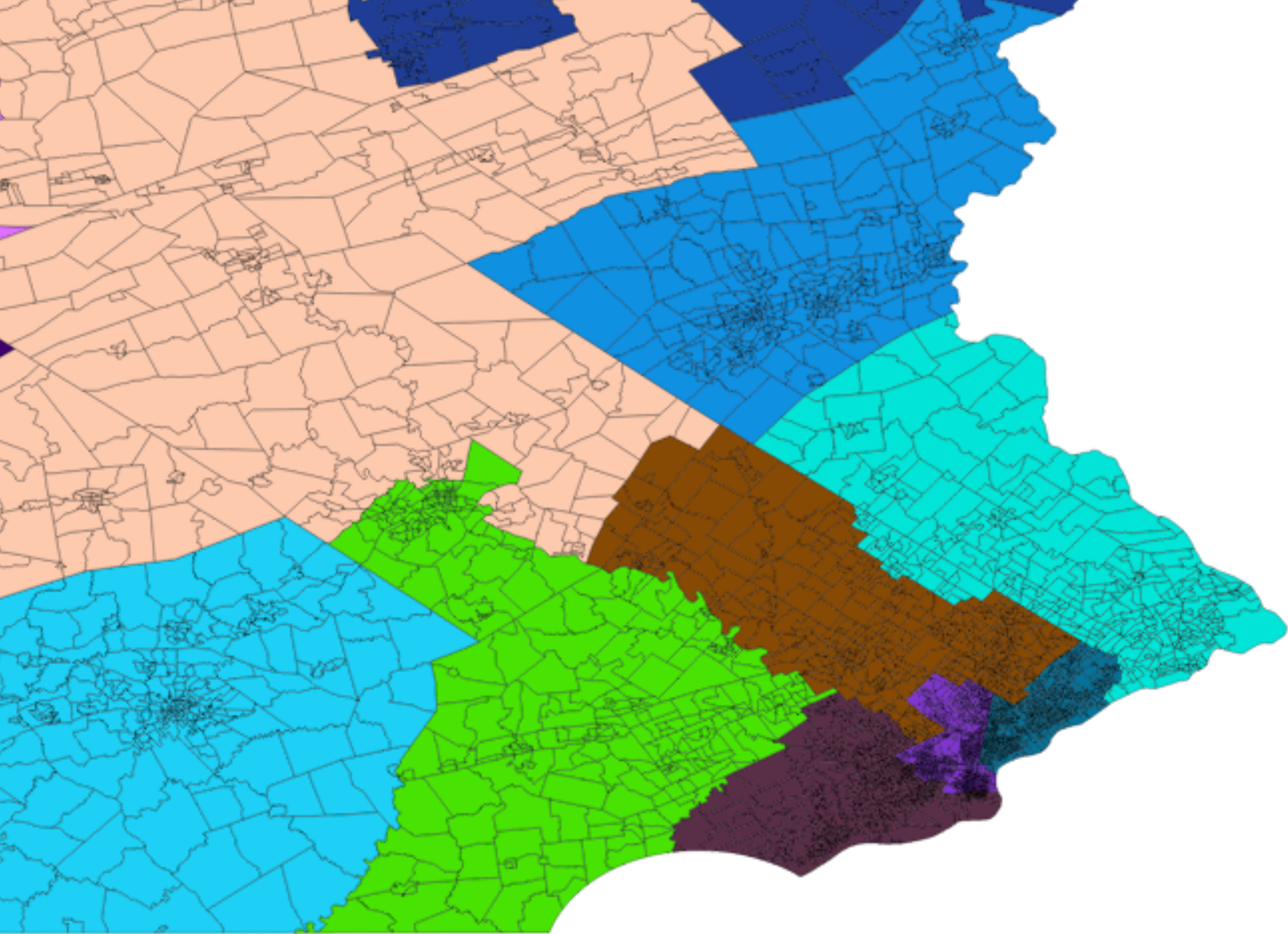
2011 Legislative



2018 Legislative



Governor



Remedial

SO WHAT MAKES A PLAN PERFORM THE WAY IT DOES?



2,970,733



2,926,441



2,951,702



2,865,012

Δ_{Pres}

Δ_{Sen}

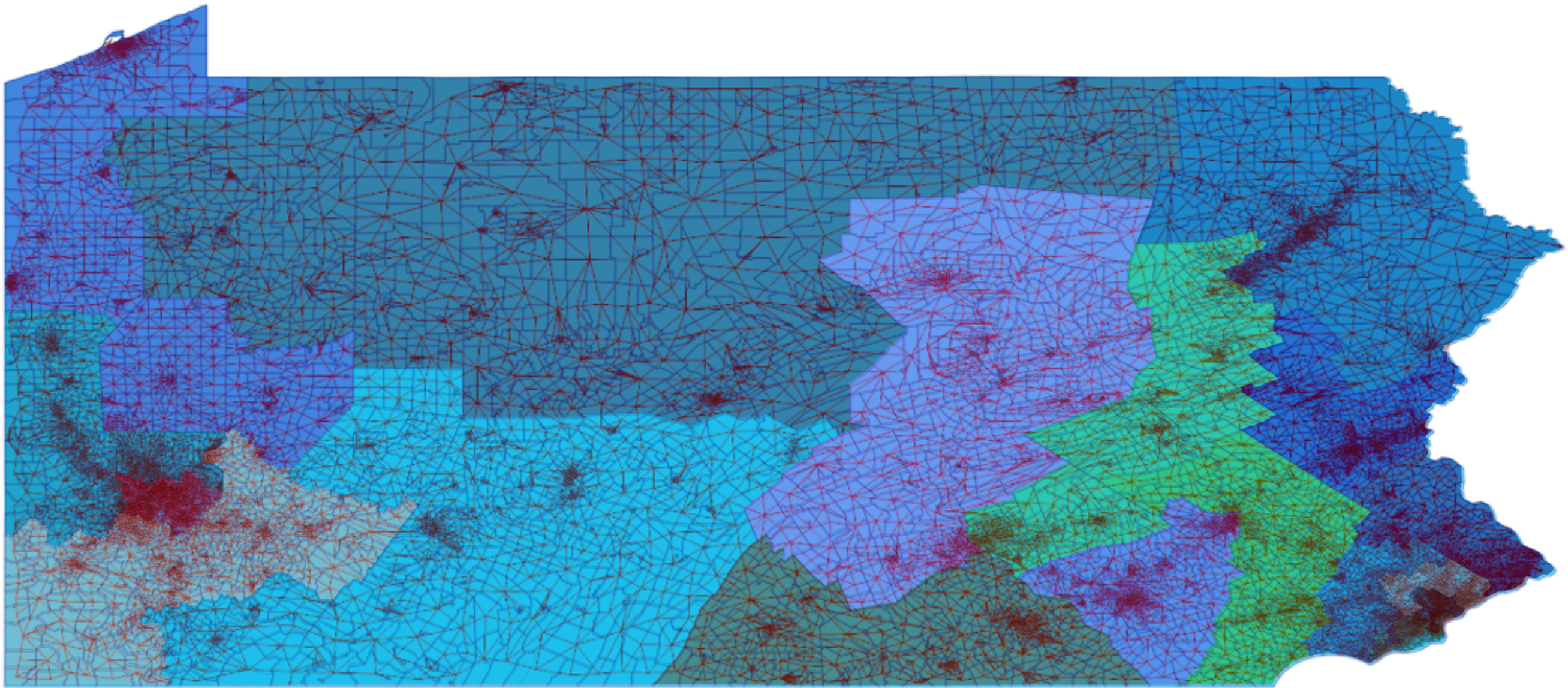
- $(\mathcal{D}_{\text{Rem}}, \Delta_{\text{Pres}}) \mapsto 8 \text{ Dem seats}$
- $(\mathcal{D}_{\text{Rem}}, \Delta_{\text{Sen}}) \mapsto 5 \text{ Dem seats}$

There's geometry here beyond shapes of districts:
spatial distribution of voters matters!

???

THE RANDOM WALK APPROACH

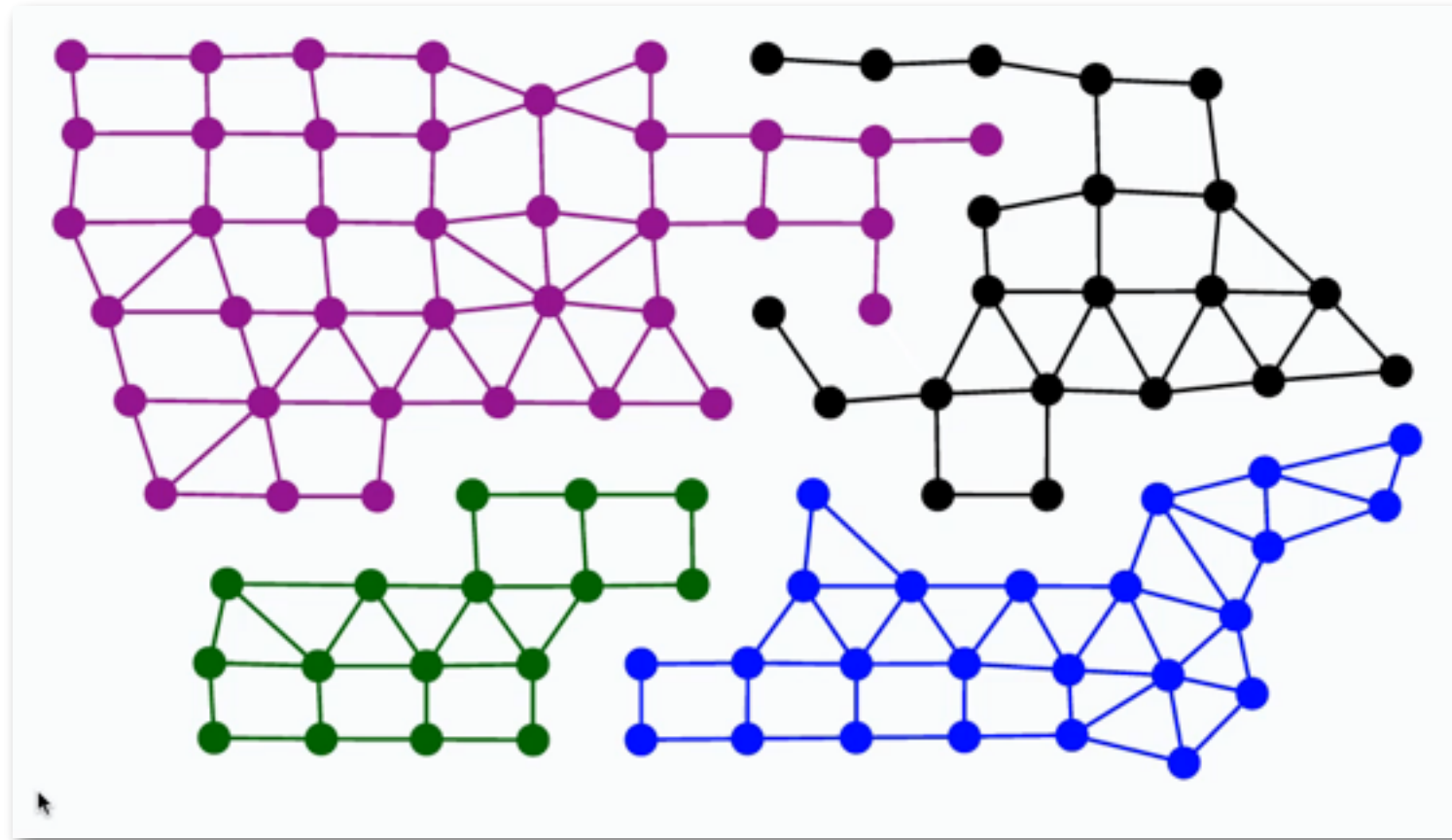
Pennsylvania: 9059 precincts



Start with a plan...

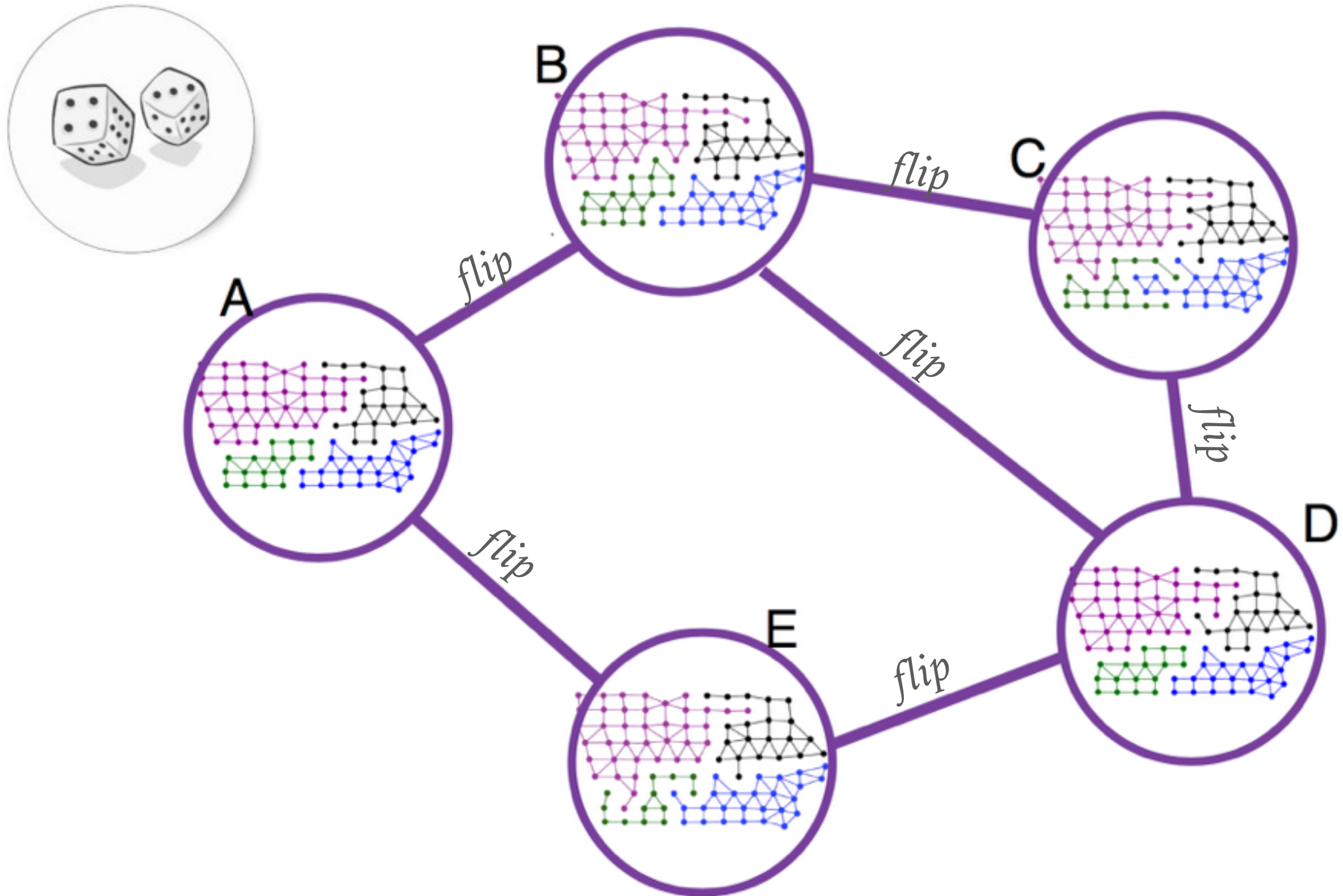
Now make 2^{30} random flips

HOW TO VARY \mathcal{D}



- This process is randomly exploring the space of redistricting plans, which is bigger than we can possibly hope to enumerate completely. Keep exploring till you have seen *billions* or *trillions* of plans.

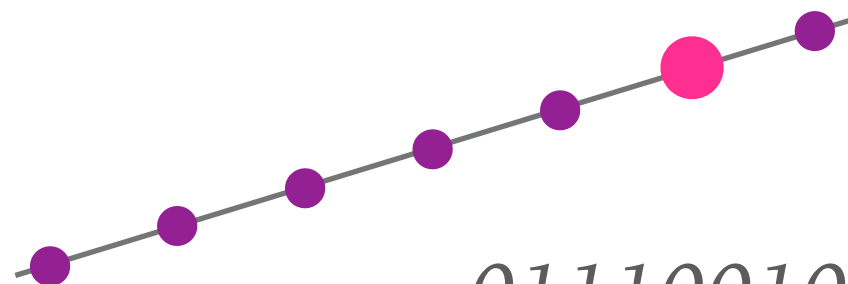
THIS AMOUNTS TO RANDOM WALK ON A “METAGRAPH”



USING THE ENSEMBLE

RANDOM WALKS ON GRAPHS; MIXING TIME DEPENDS ON TOPOLOGY!

- Path of length n ($N=n$ states) – mixing time N^2
- Grid of side n ($N=n^2$ states) – mixing time N
- Binary strings of length n ($N=2^n$ states) – m.t. $\log N \log \log N$



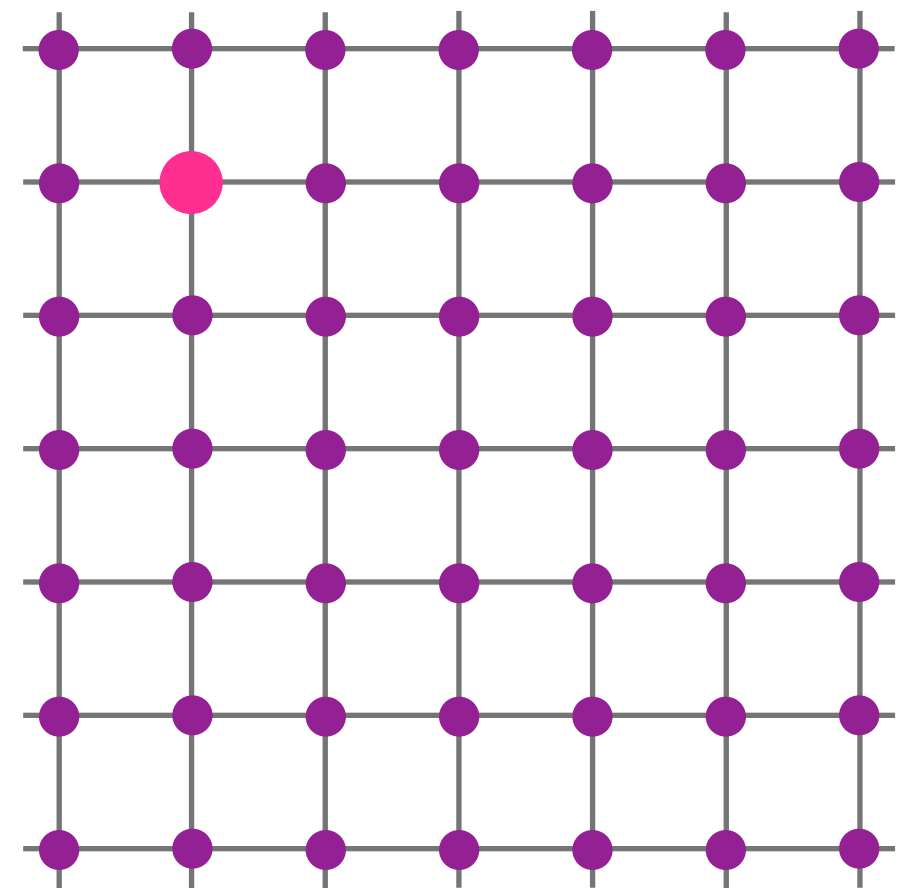
01110010101...110

01111010101...110

01111010001...110

00111010001...110

00111010001...111



- Major research direction: what is the geometry/topology of the redistricting metagraph?

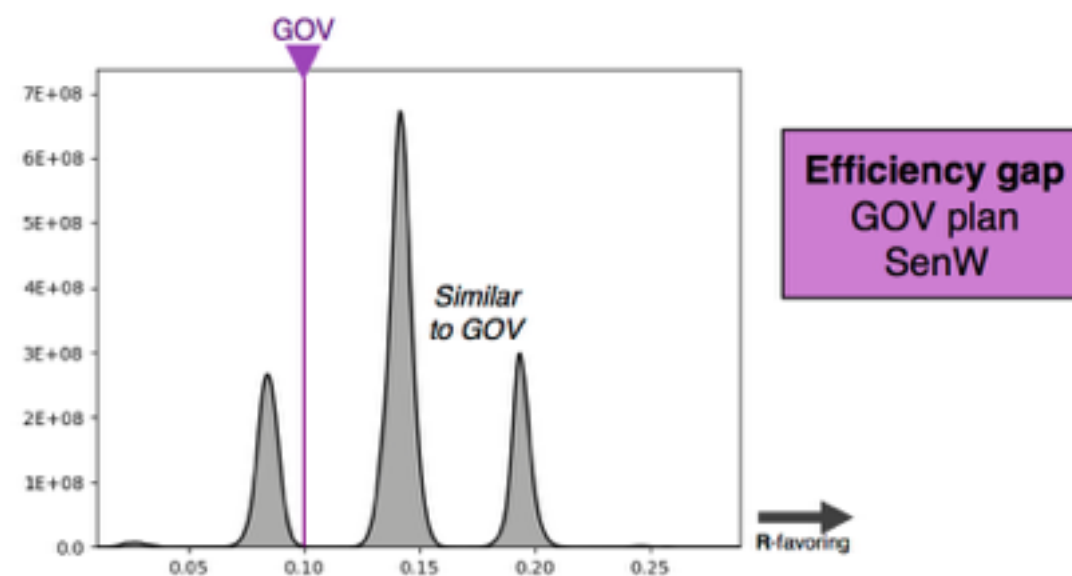
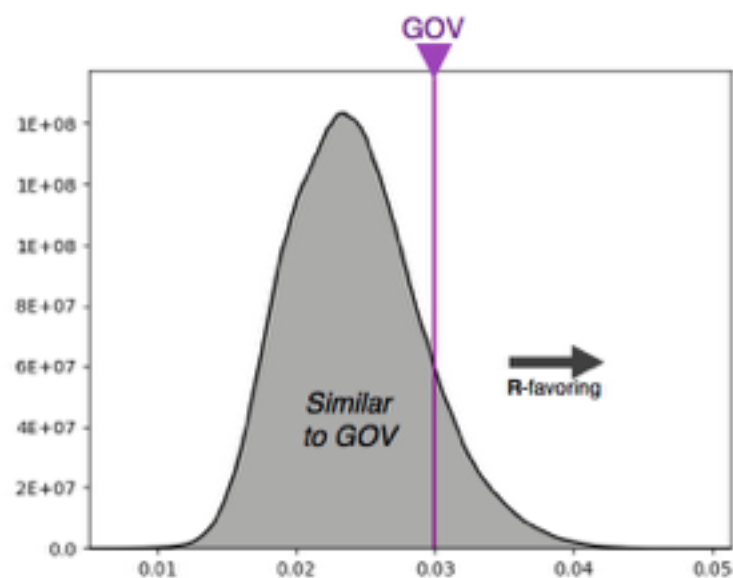
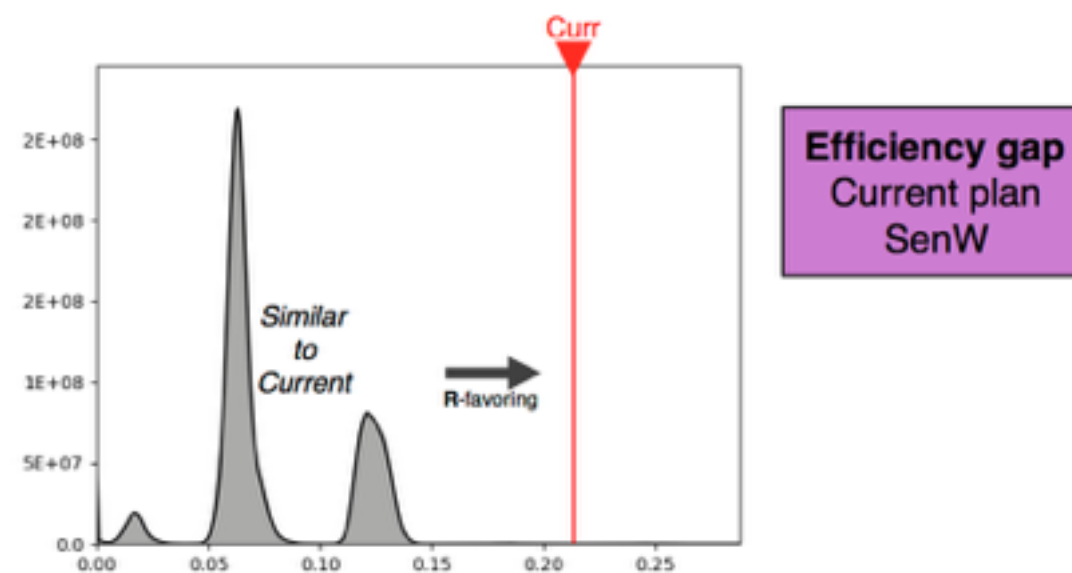
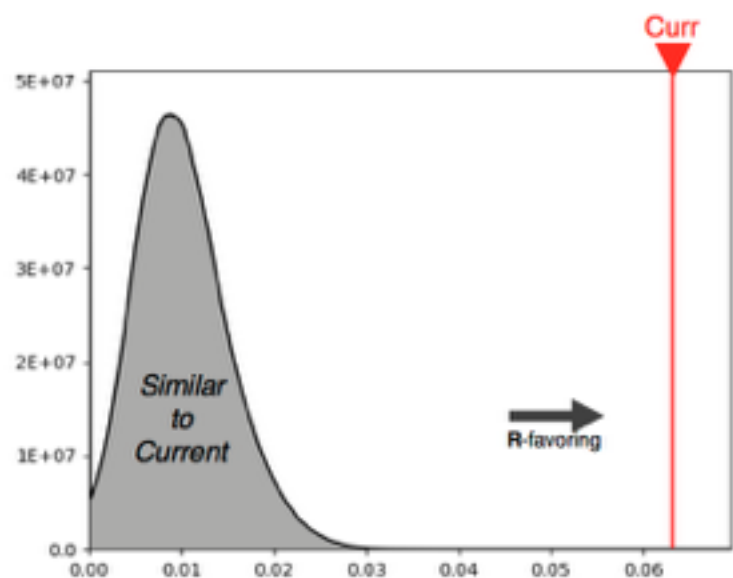
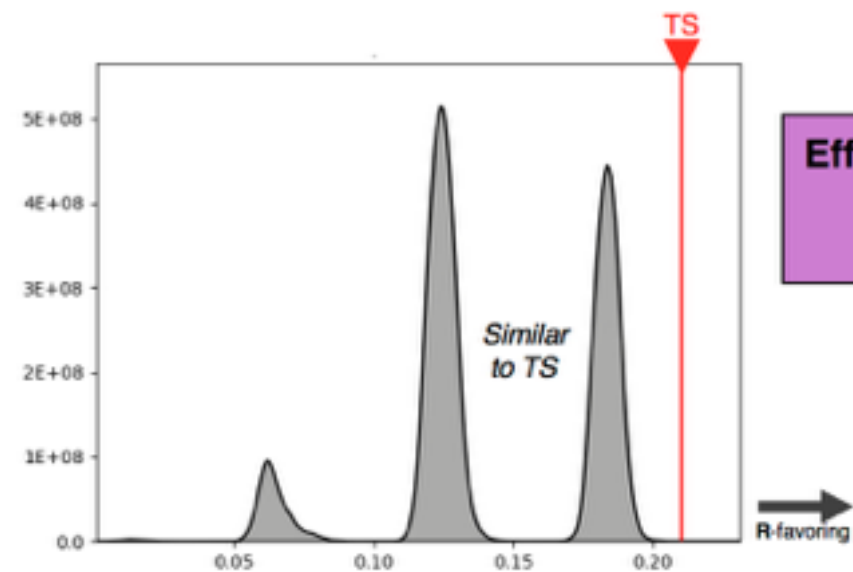
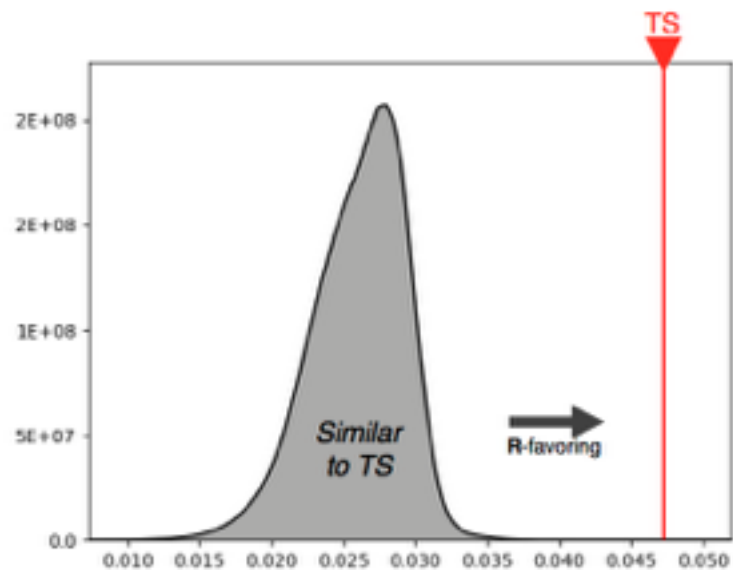
GIVEN MAPS AND VOTER DISTRIBUTIONS, FIX SOME METRICS

- Recall **mean-median** score measures *how far short of a half of the votes the controlling party can fall, while still having half of the representation*

(e.g., $MM=.03$ means $V=47\%$ will tend to secure $S=50\%$)

- Recall **efficiency gap**, nominally about *which party wastes more votes*, but actually just a seats vs votes score

$EG=2V-S-1/2$ with some random noise added from voter turnout



COMPARING TO THE LOCAL ENSEMBLE

- A p -value bounds the chance that an outcome could be observed by chance under the null hypothesis. (Here, null hypothesis is that the *plan was selected uniformly at random among plans satisfying traditional principles at least as well.*)

A billion districting plans, at least as compact as initial plan, no more county splits

10^{30} steps	D seats	<i>EG</i>	frac with higher <i>EG</i> than plan	pro-R p -value	pro-D p -value	bias	<i>MM</i>	frac with higher <i>MM</i> than plan	pro-R p -value	pro-D p -value	bias
TS-Sen10	5	.212	.000048	.0098	1.41	R	4.7%	.0000005	.00099	1.41	R
TS-Sen16	4	.258	.000004	.0089	1.41	R	4.6%	.00000031	.00078	1.41	R
TS-SenW	5	.21	.000043	.0093	1.41	R	4.6%	.0000004	.0009	1.41	R
current-Sen10	5	.216	.0007	.12	1.41	—	6.2%	.000000014	.00017	1.41	R
current-Sen16	4	.259	.000046	.03	1.41	mild R	4.3%	.000049	.0099	1.41	R
current-SenW	5	.214	.00065	.036	1.41	mild R	6.2%	.00000049	.00099	1.41	R
GOV-Sen10	6	.149	.074	.38	1.36	—	2.5%	.065	.36	1.37	—
GOV-Sen16	7	.095	.998	1.41	.063	—	3.5%	.12	.5	1.32	—
GOV-SenW	7	.099	.78	1.25	.66	—	3%	.12	.49	1.33	—

- $p \leq .05$ is publishable **statistical significance**; $p \leq .01$ is stricter.

THERE'S A THEOREM IN THE BACKGROUND

Chikina-Frieze-Pegden $\sqrt{2\varepsilon}$ test

- For any reversible Markov chain with stationary distribution π , you can examine a plan \mathcal{D}_0 against the null hypothesis $\mathcal{D}_0 \sim \pi$ as follows:
 - ★ fix any real-valued score G
 - ★ take many steps $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \dots$
 - ★ if $G(\mathcal{D}_0)$ is in the most extreme ε fraction of the $\{G(\mathcal{D}_i)\}$, then you can reject the null hypothesis with $p \leq \sqrt{2\varepsilon}$



LOCAL VS GLOBAL

- That is a rigorous approach to a **local** search of the possibility space. **Global** searches are at the research frontier
- How will we know we are searching well?
 - ★ ROBUST - not sensitive to {setup, chain length, interpretive choices}
 - ★ FORGETFUL - different start points give similar results
 - ★ INTUITIVE - give expected answers on toy datasets
 - ★ POWERFUL - fully explores small datasets / doesn't get stuck / finds very different maps

THIS CAN HELP US UNPACK PARTISAN SKEW



2,970,733



2,926,441



2,951,702



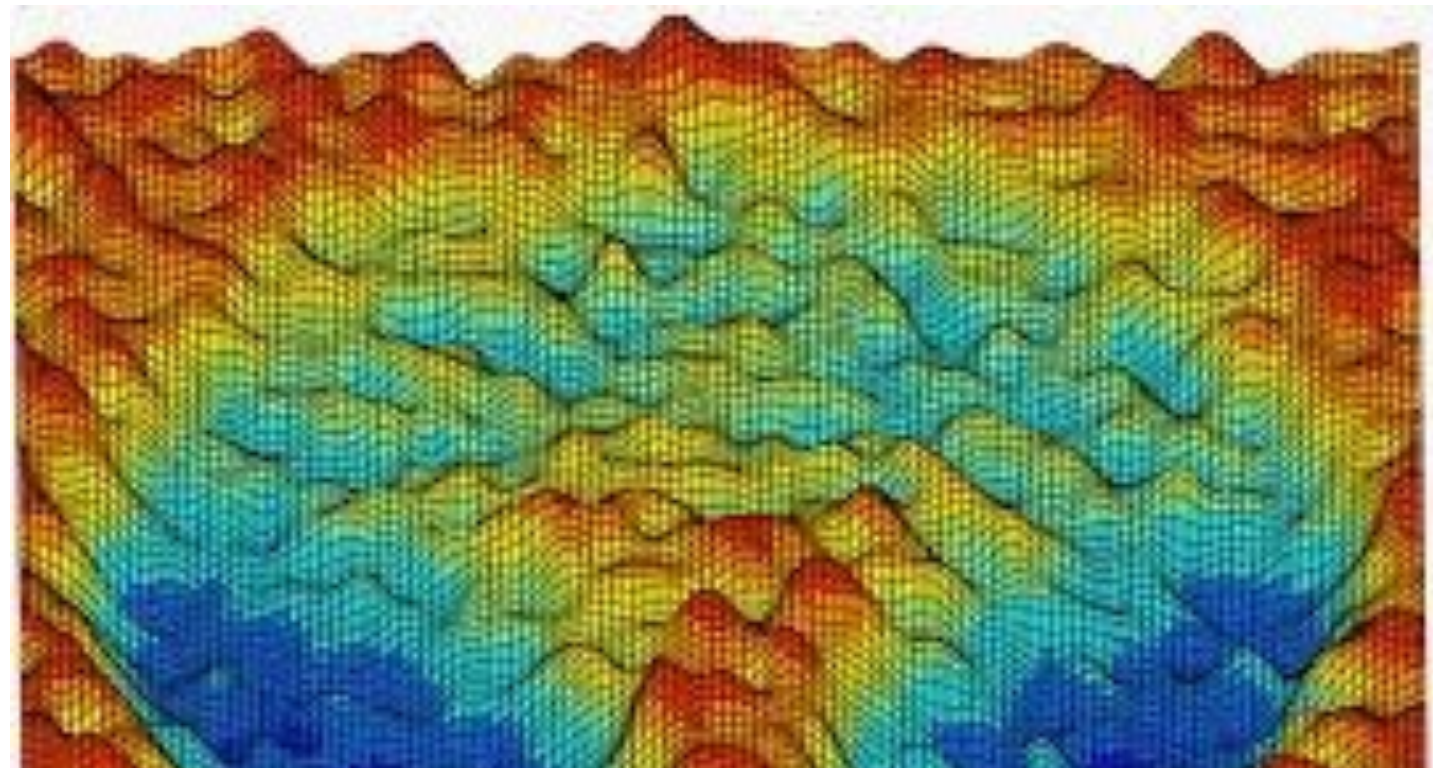
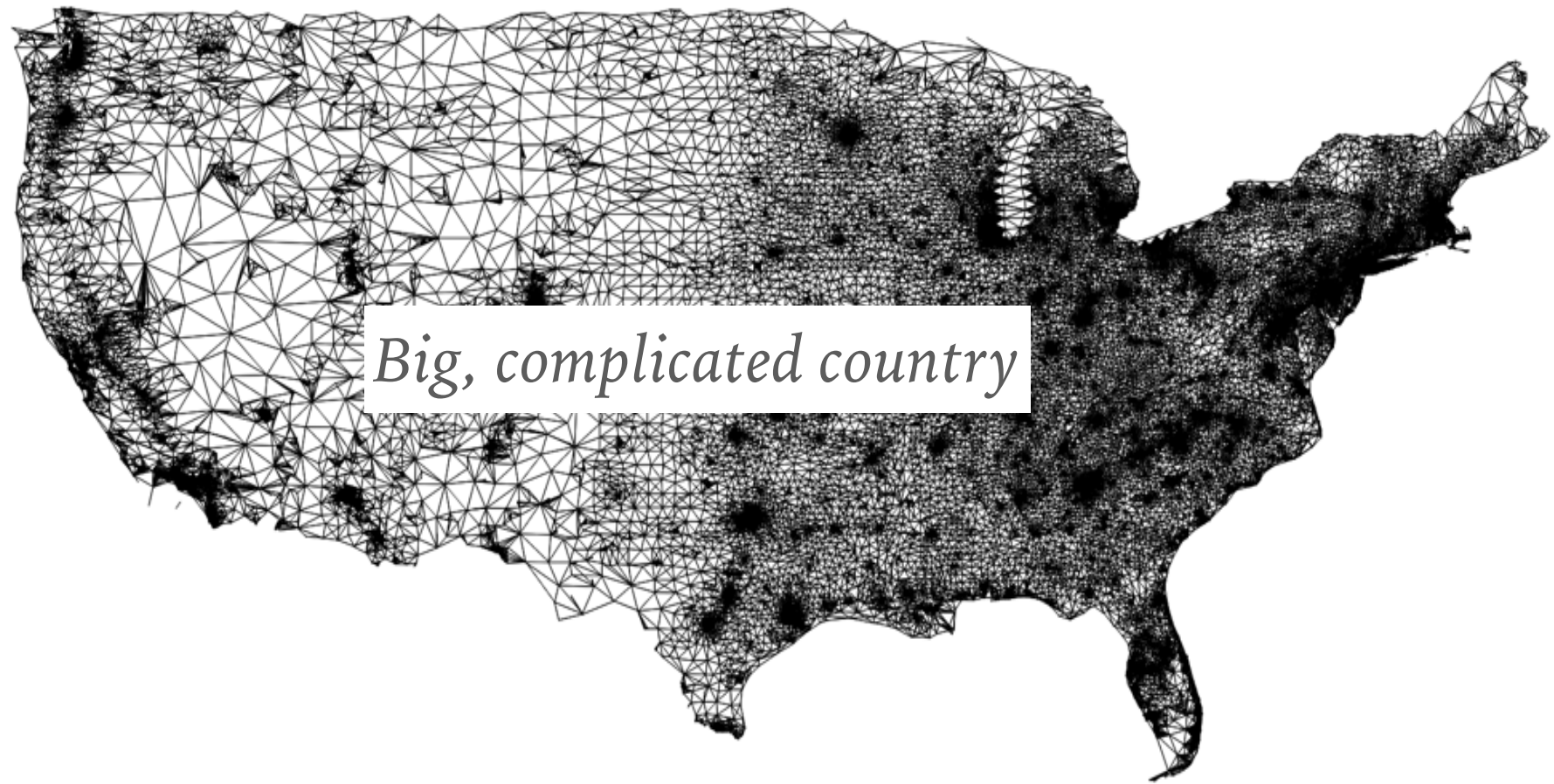
2,865,012

Δ_{Pres}

Δ_{Sen}

- $(\mathcal{D}_{\text{Rem}}, \Delta_{\text{Pres}}) \mapsto 8 \text{ Dem seats}$
- $(\mathcal{D}_{\text{Rem}}, \Delta_{\text{Sen}}) \mapsto 5 \text{ Dem seats}$

- Careful rigging? Political geography? Fluke effects?



Unimaginably vast landscape of ways to cut things up



YOUR JOB THIS SUMMER

- You are 52 students from an incredible array of backgrounds and specialties
- Census 2020 is around the corner, and we want to prepare data and techniques in all 52 states :)
- This week, you'll all stay together and pick up the basic skills (GIS, data formats, markovchain) to do local MCMC yourself
- Over the course of the six weeks, we have a very wide range of projects for you to get involved right at the frontier of research and tool-building for gerrymandering.

