Gerrymandering by the Numbers

Andrea K. Barreiro, Associate Professor of Mathematics,
Southern Methodist University (SMU)

Math For Unbiased Maps TX (MUM_TX)

Faculty from SMU, other Texas universities
and interested community volunteers

MUM_TX is a project of the
Research Cluster on Political Decision-Making,
supported by the SMU Dedman College Interdisciplinary Insitutute

https://www.smu.edu/mumtx

Contact us:
MathForUnbiasedMapsTX@list.smu.edu

Members

Southern Methodist University (SMU)

Andrea K. Barreiro, Associate Professor of Mathematics
Matthew Lockard, Associate Professor of Philosophy
Scott Norris, Associate Professor of Mathematics
Brandilyn Stigler, Associate Professor of Mathematics

Collin College

Dustin Potter, Professor of Mathematics

Community Members/Volunteers

Robert Meyers

Collaborators

Scott Cook, Tarleton State University
Anne Egleston, Tarleton State University
Joaquin Gonzalez, Texas Civil Rights Project

What is "gerrymandering"?

- "political manipulation of electoral district boundaries with the intent of creating undue advantage for a party, group, or socio-economic class within the constituency." (Wikipedia)

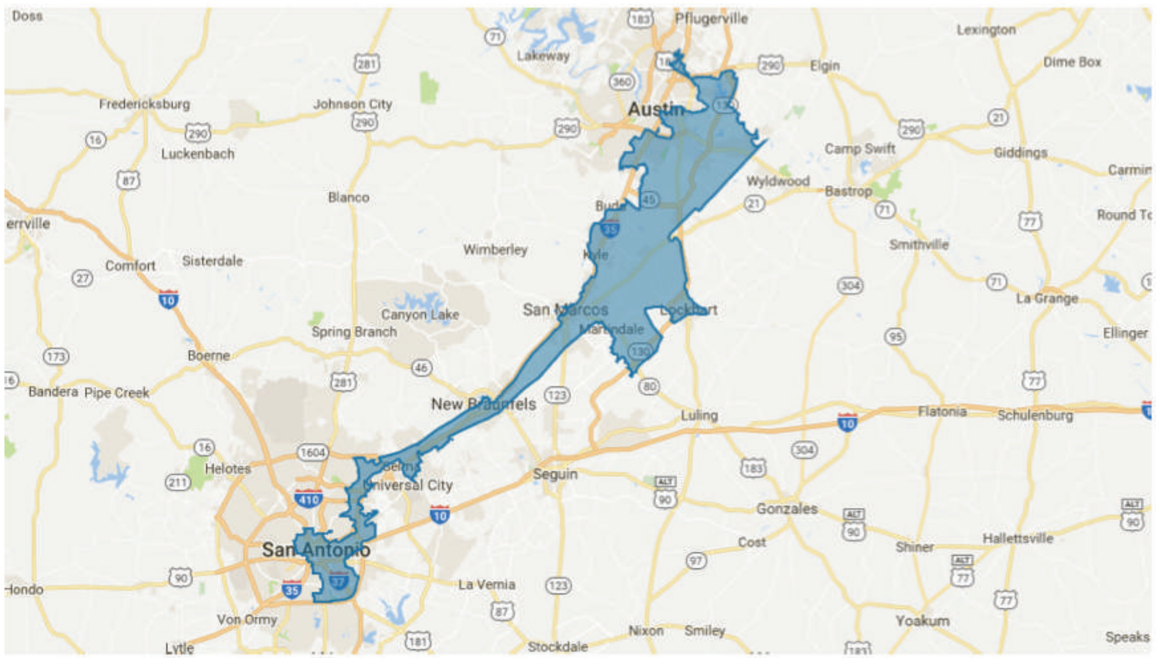
Elkanah Tisdale (1771-1835) (often falsely attributed to Gilbert Stuart)[1] - Originally published in the Boston *Centinel*, 1812.

Two main types of gerrymandering we consider

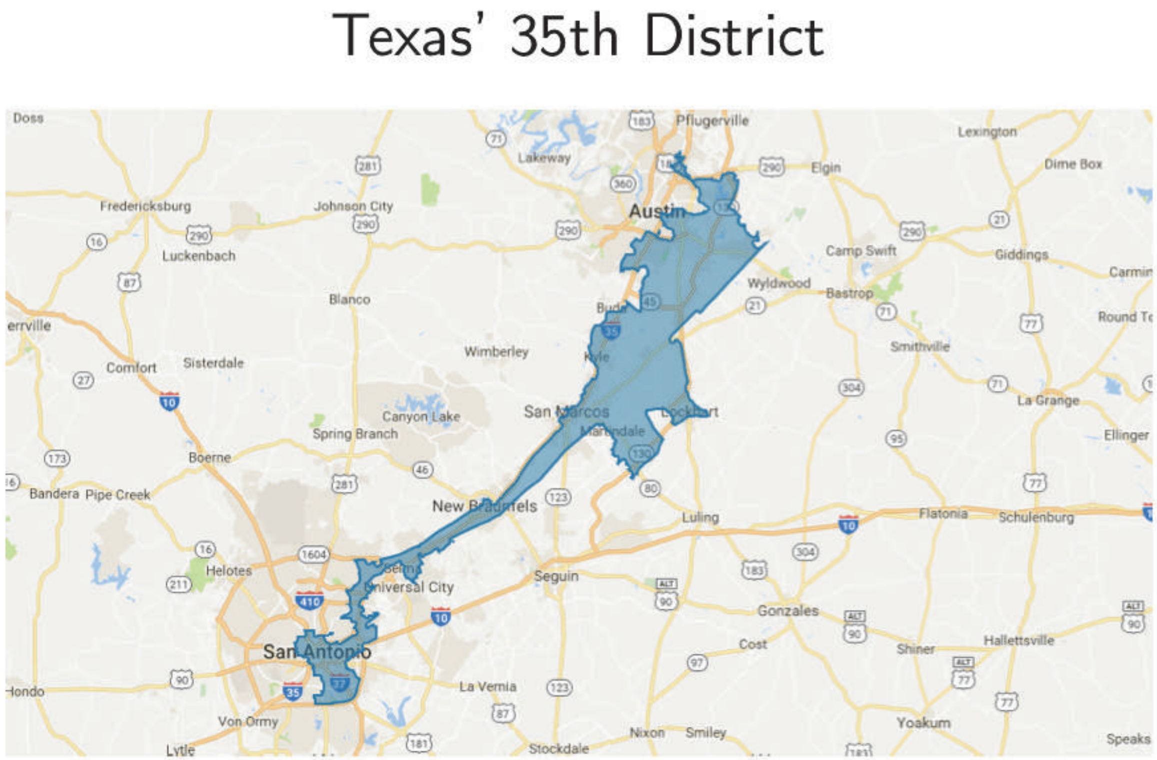
- Partisan
  - Districts are drawn to favor one political party over another.
  - Or, to reduce competitiveness/protect incumbents
  - Unfortunately, “not justiciable” in federal courts (Rucho v. Common Cause, 2019)
- Racial
  - Governed by the Voting Rights Act of 1965
  - Section 2 prohibits “vote dilution”: for example, through “cracking and packing”
  - Unfortunately, Section 5 pre-clearance no longer required (Shelby v. Holder 2013)

Is it gerrymandered?

Texas' 35th District



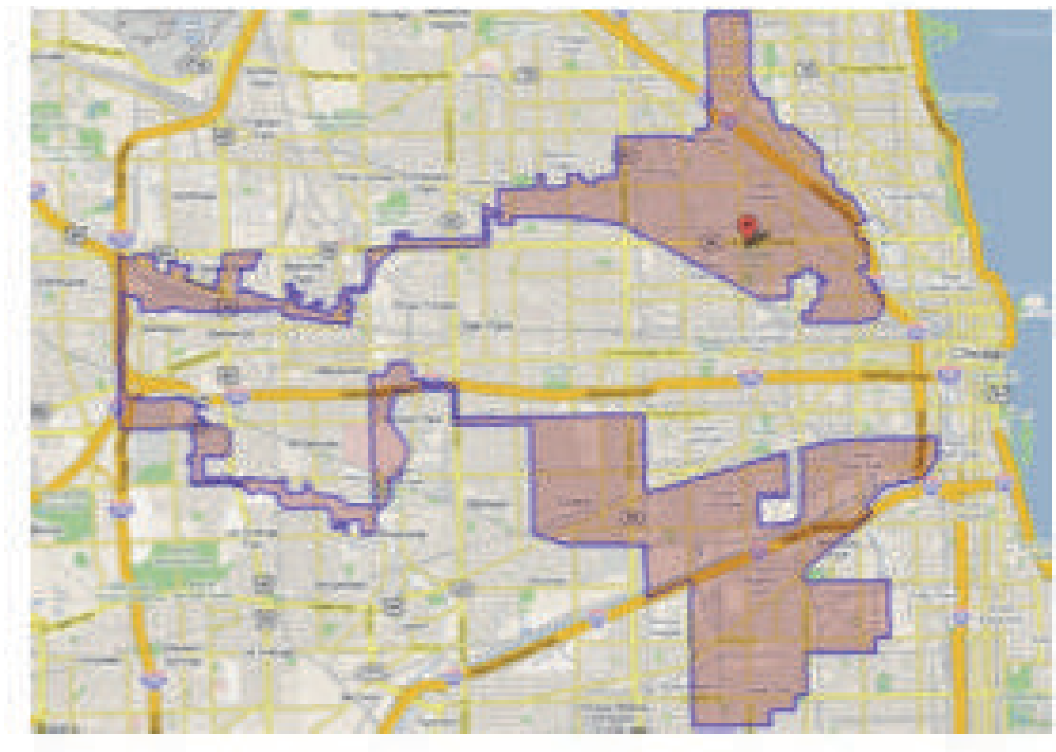
Is it gerrymandered?



- 2017 - “an impermissible racial gerrymander” (3-member panel of federal judges)
- 2018 - overruled by the Supreme Court (Abbott v. Perez)

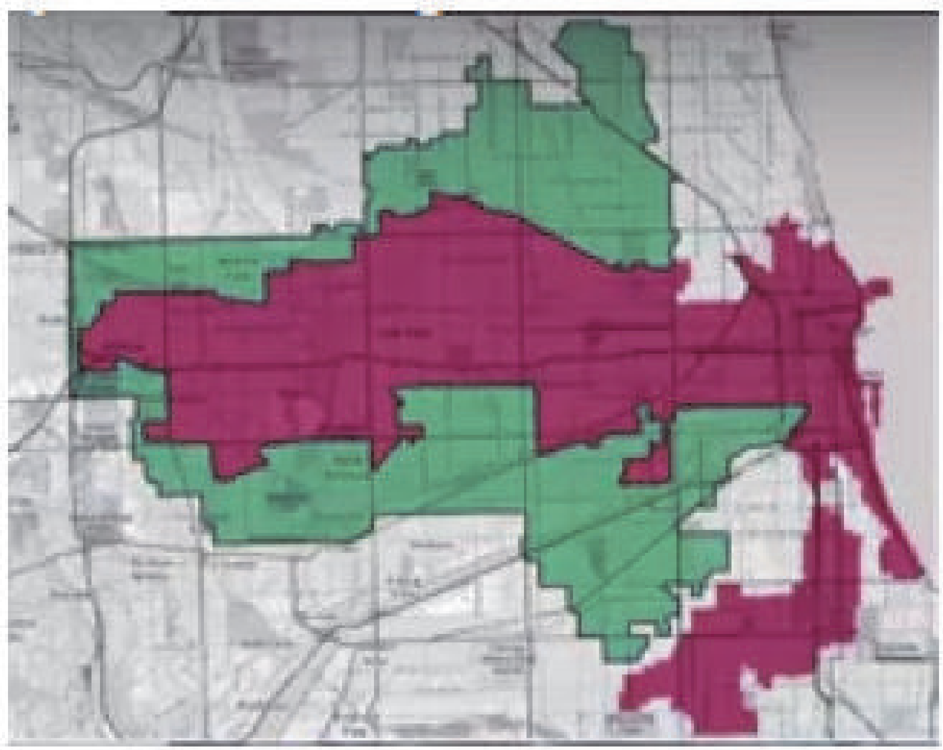
Is it gerrymandered?

Illinois' 4th District



Is it gerrymandered?

Illinois' 4th & 7th Districts



- created after federal courts ordered the creation of a majority-Hispanic district in the Chicago area (1990's)
- links two Hispanic neighborhoods
- Caveat: other litigants have claimed the district is unconstitutional

So, just looking at a district doesn't tell you whether it's gerrymandered. What does?
Today I will show you:

- Statistics we can use to quantify bias (partisan or racial)

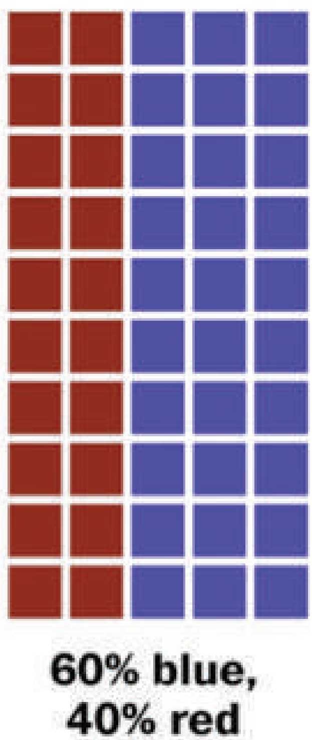
- How we use *ensemble analysis* to gauge what those statistics would look like for an unbiased map
  - i.e. If you could choose a map at random, using only the constraints that each district had equal population, was connected, and was reasonably compact, what would it look like?

- How we have applied these methods to maps created by the Texas Legislature in 2021.
  - Focus on the US Congressional map: subject of a DOJ lawsuit

- …and to Dallas City Council

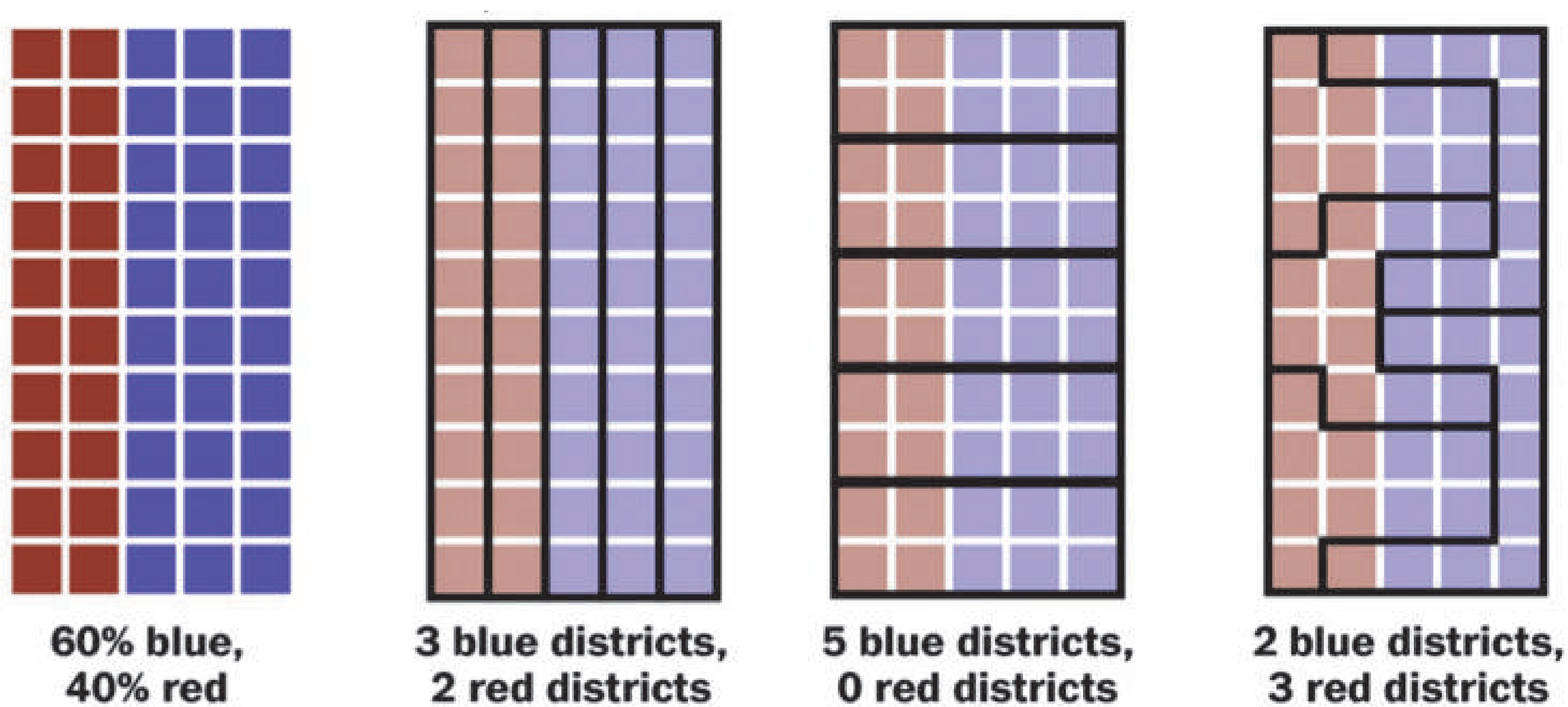
Let's start out with an example of how lawmakers can manipulate a map to their advantage

- 50 people, to be divided into 5 equal sized "districts" (10 people each)



Let's start out with an example of how lawmakers can manipulate a map to their advantage

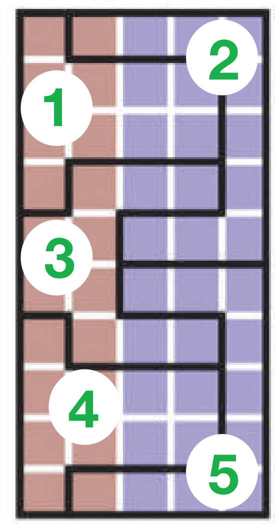
- 50 people, to be divided into 5 equal sized "districts" (10 people each)



Vote share vector:

- Order districts ***by increasing vote share***
- We choose Blue as the point of view (POV) party

District #	#Blue	#Red
1	4	6
2	9	1
3	4	6
4	4	6
5	9	1

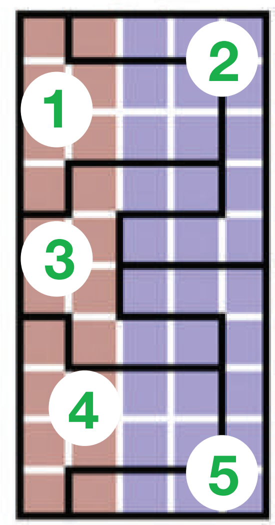


2 blue districts, 3 red districts

Vote share vector:

- Order districts ***by increasing vote share***
- We choose Blue as the point of view (POV) party

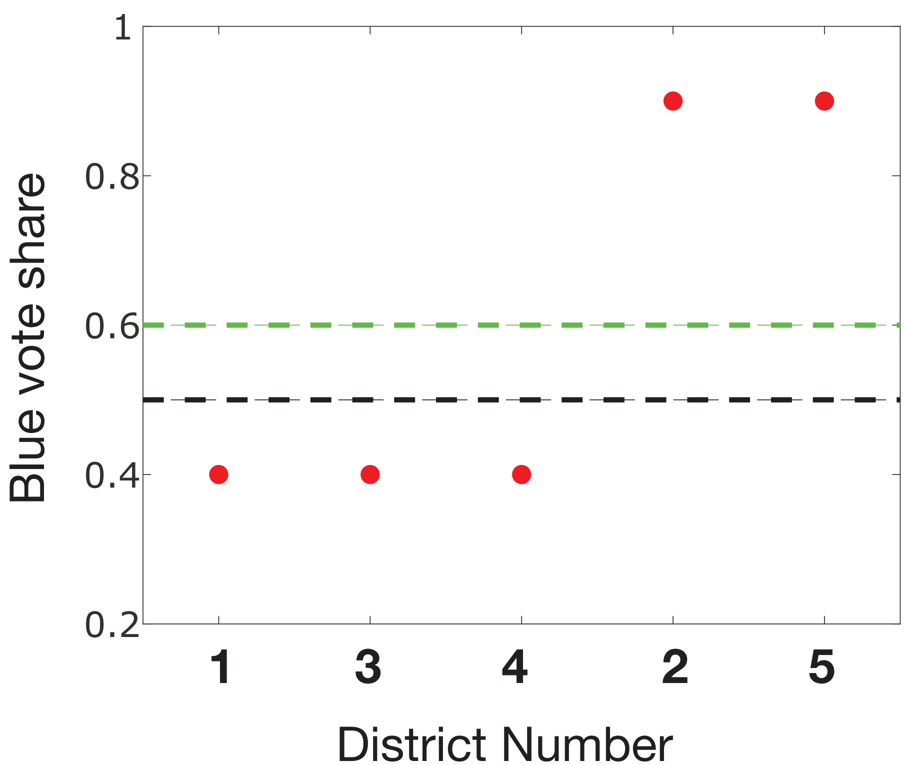
District #	#Blue	#Red	Vote Share
1	4	6	0.4
2	9	1	0.9
3	4	6	0.4
4	4	6	0.4
5	9	1	0.9

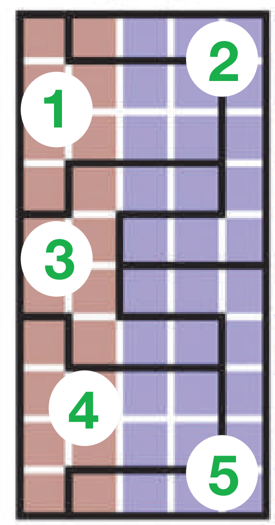


**2 blue districts,
3 red districts**

Vote share vector:

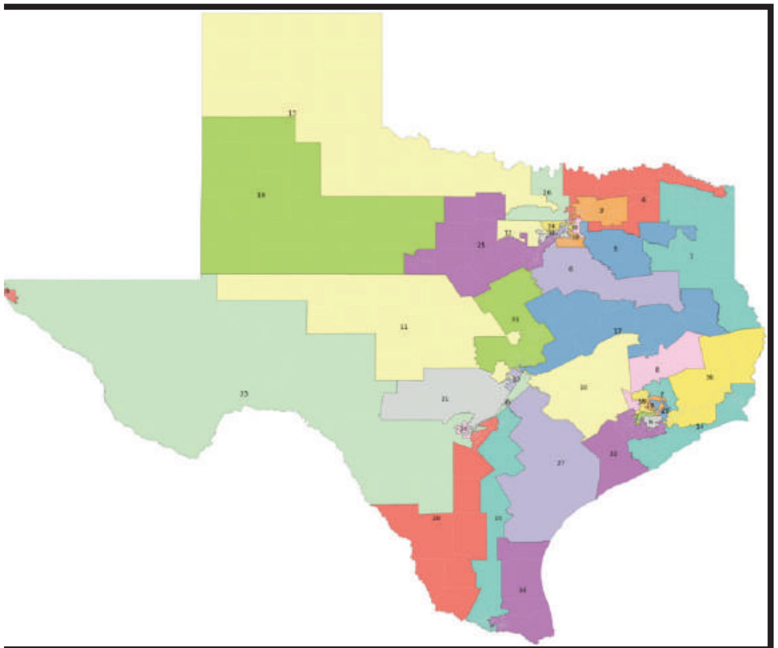
- Order districts ***by increasing vote share***
- We choose Blue as the point of view (POV) party

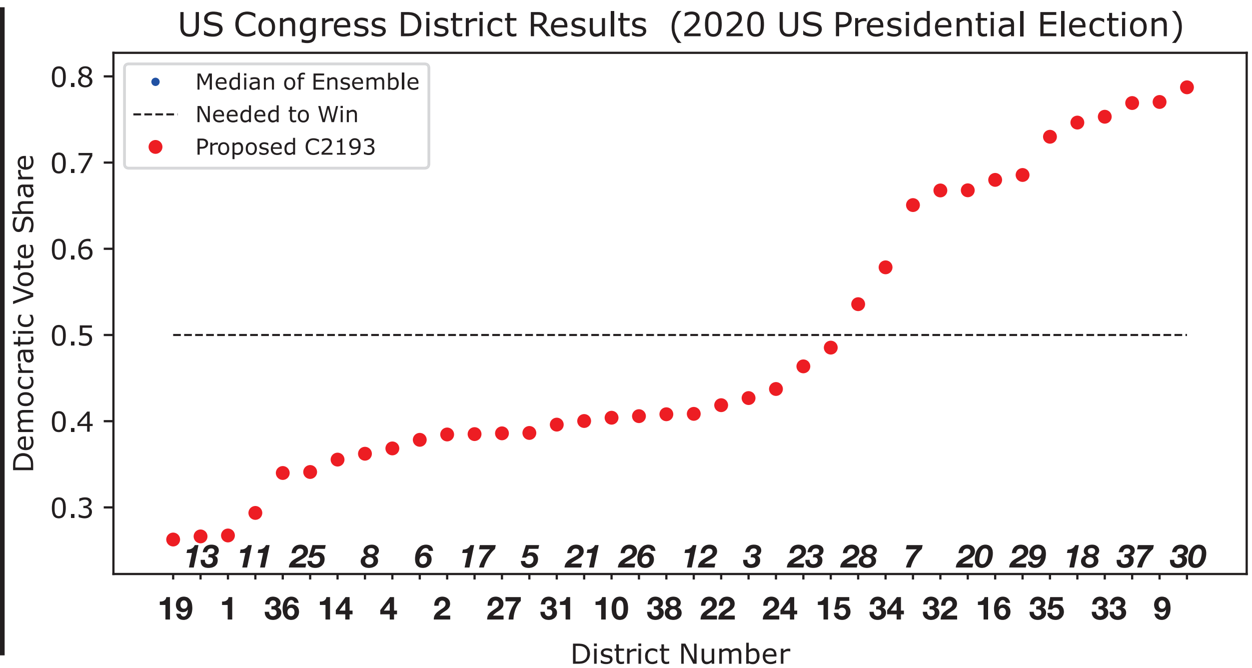




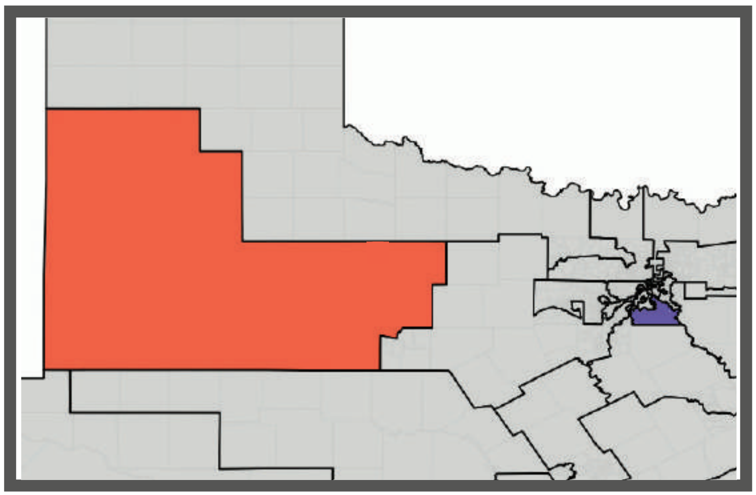
**2 blue districts,
3 red districts**

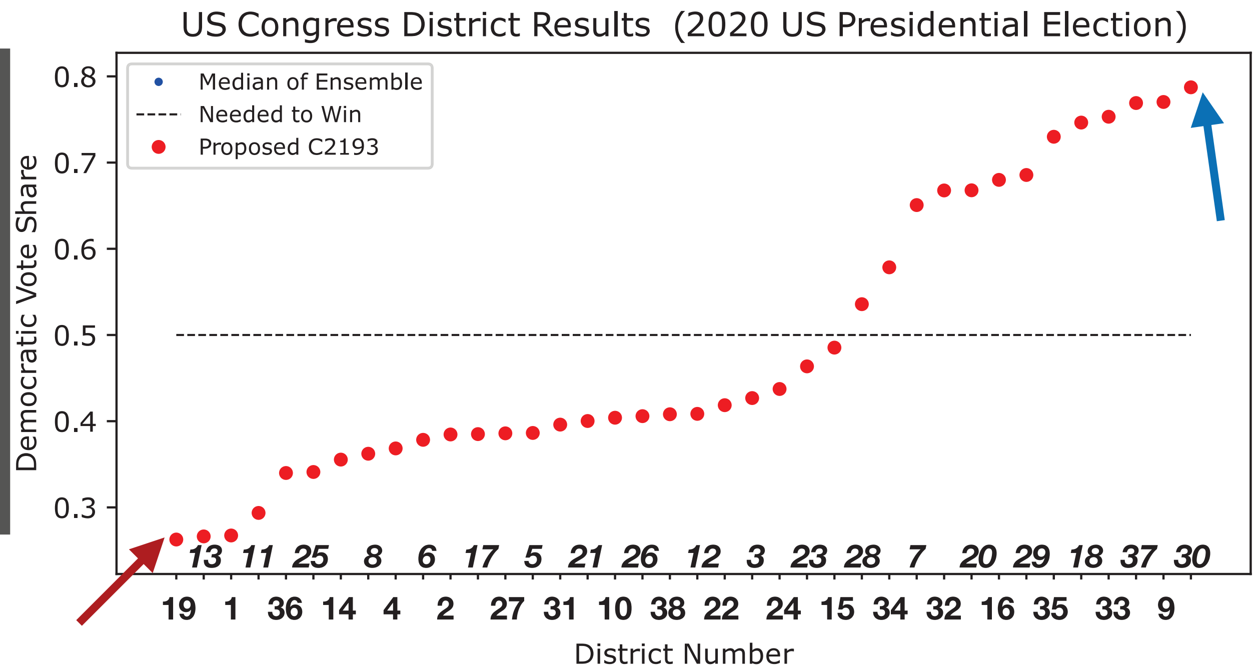
Now let’s look at the new Congressional map enacted by the TX Lege





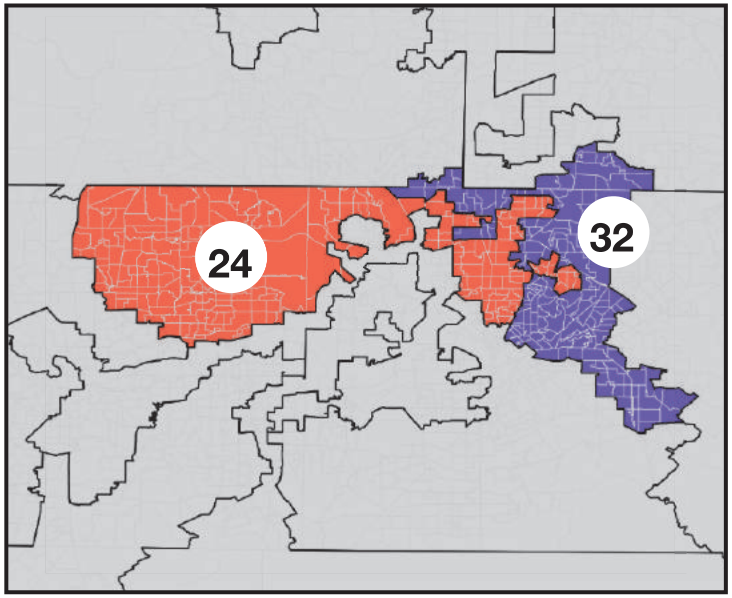
Now let’s look at the new Congressional map enacted by the TX Lege

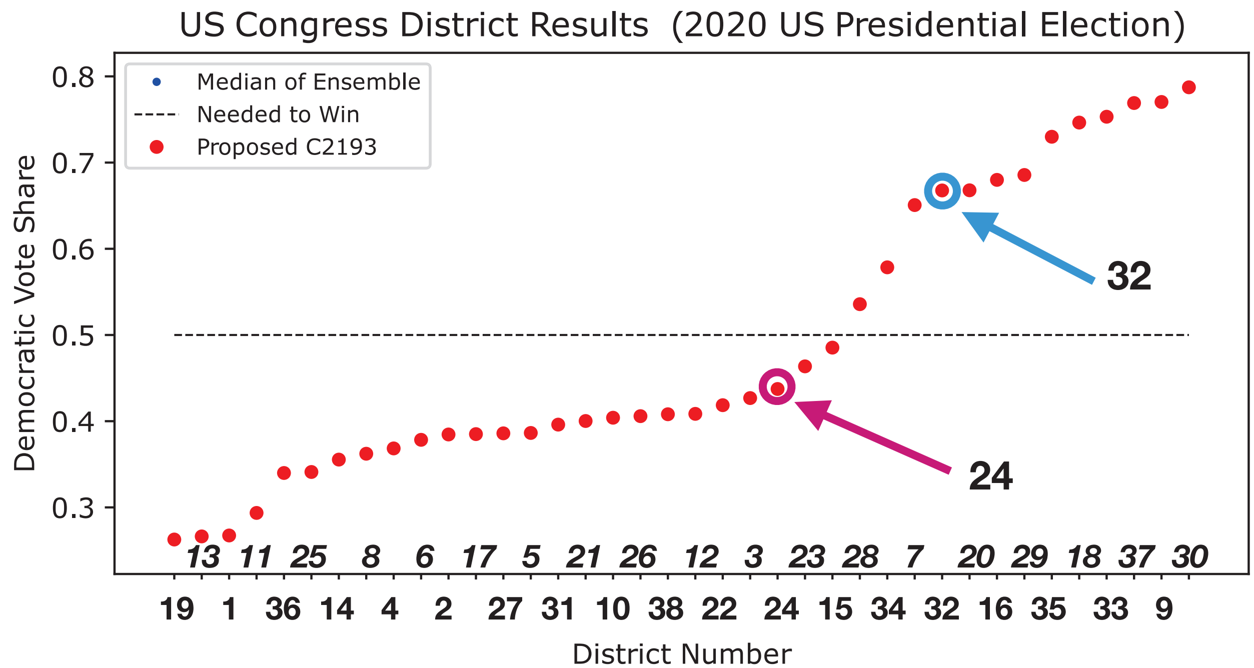




- 38 districts (an increase from 36)
- Districts are ordered by Democratic vote share in the 2020 Presidential election
  - Lowest - **CD19** - 26.3% - Jodey Arrington (R)
  - Highest - **CD30** - 78.7% - Eddie Bernice Johnson (D)

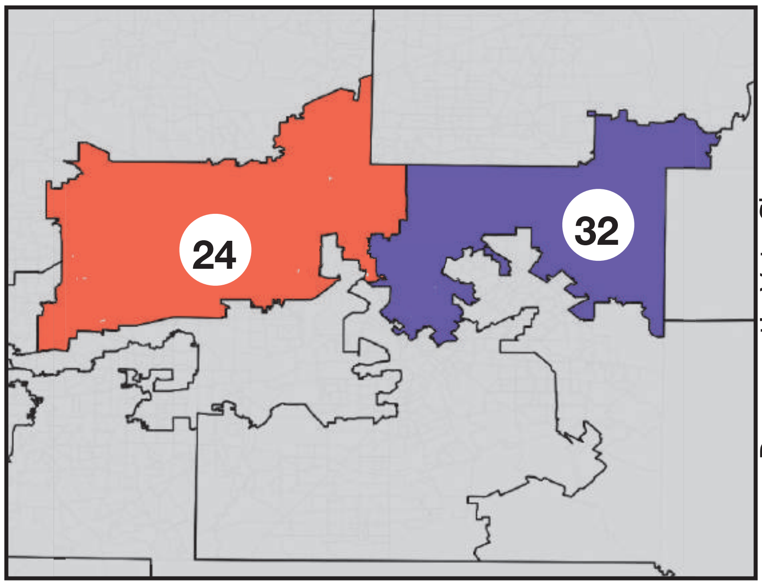
Here are two districts that might be of interest: 24 and 32

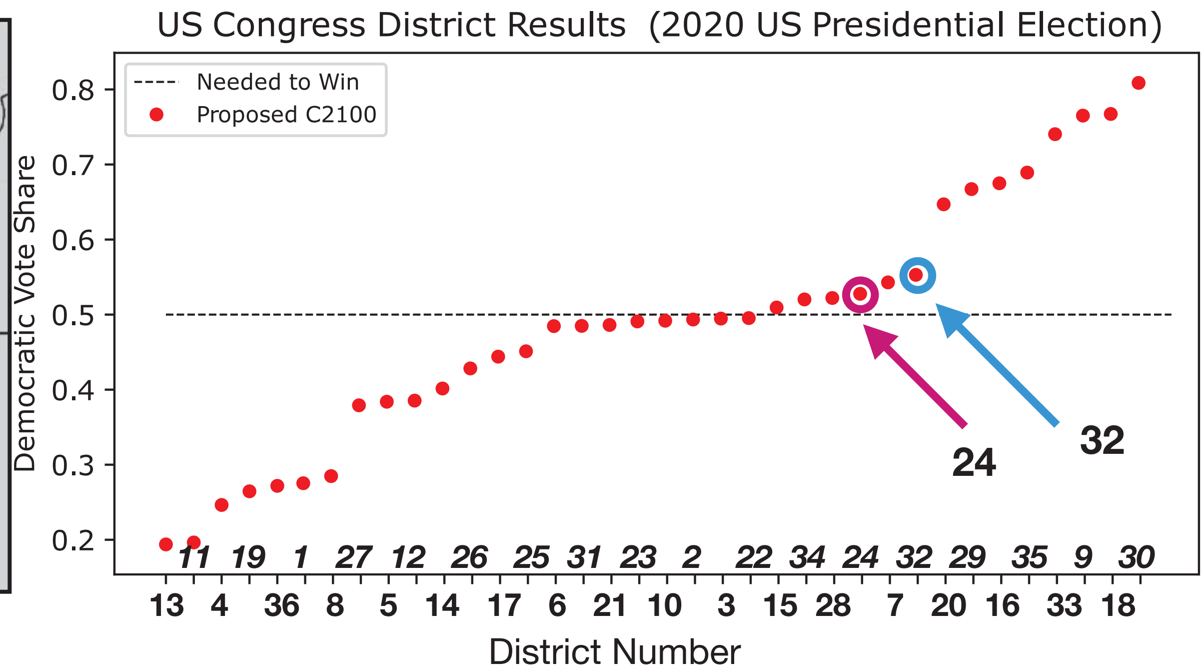




- 38 districts (an increase from 36)
- Districts are ordered by Democratic vote share in the 2020 Presidential election
  - **CD24** - 43.7% - Beth Van Duyne (R)
  - **CD32** - 66.8% - Colin Allred (D)

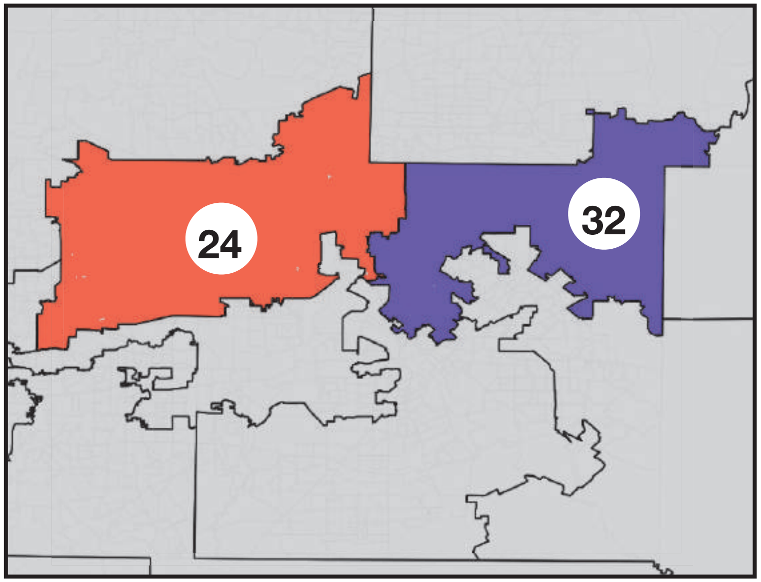
How does this compare with the current districts actually used in the 2020 election?

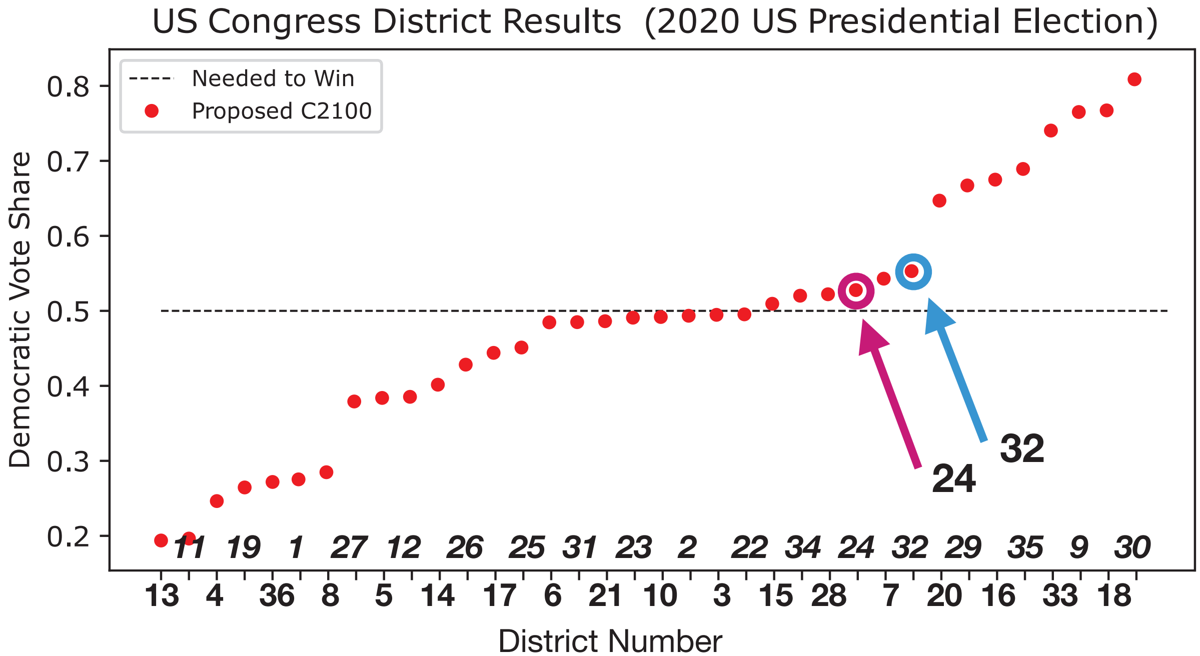


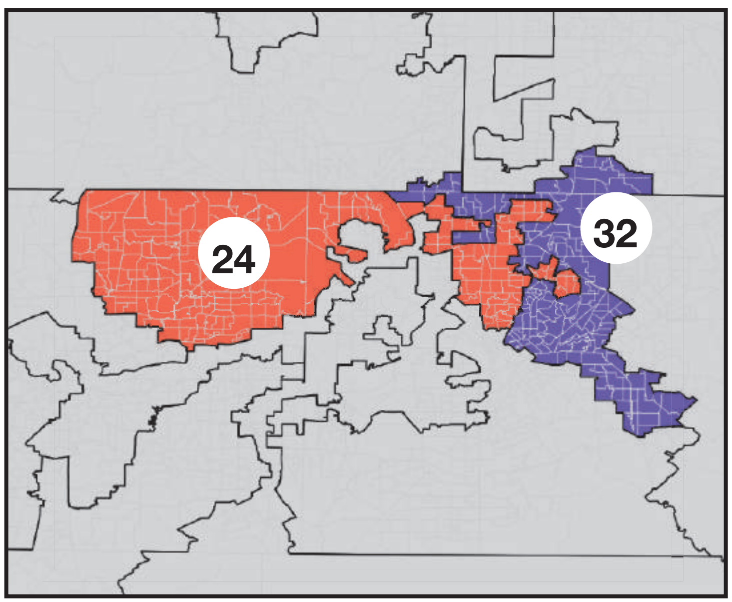


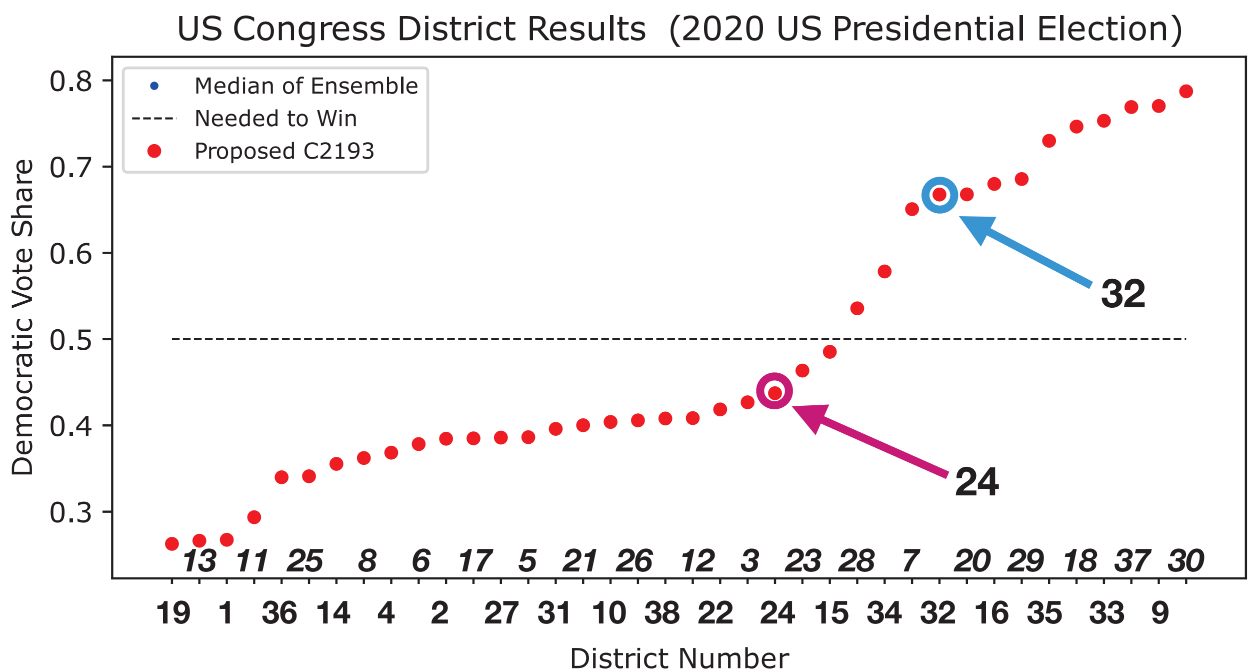
- 36 districts
- Districts are ordered by Democratic vote share in the 2020 Presidential election
  - **CD24** - 52.8% - Beth Van Duyne (R)
  - **CD32** - 55.3% - Colin Allred (D)

Let's compare...

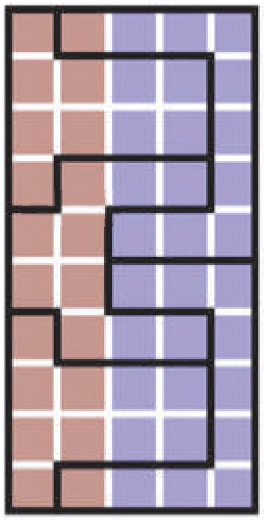




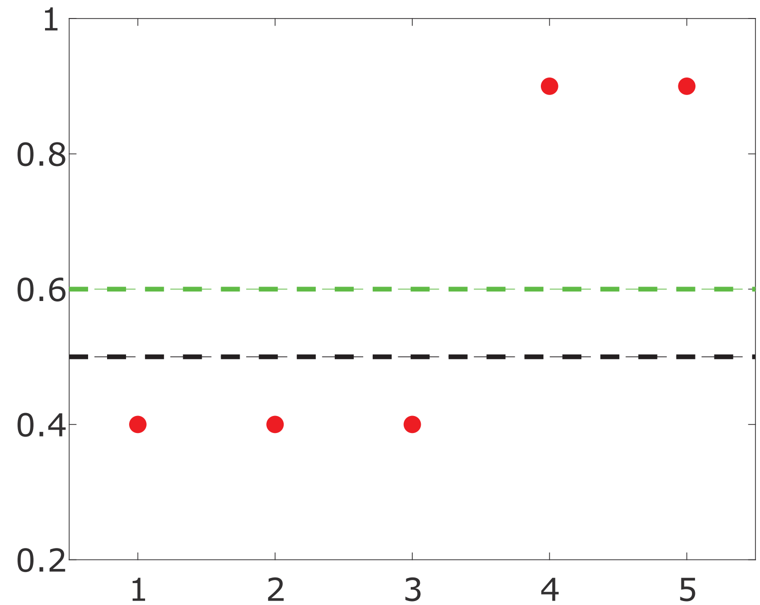


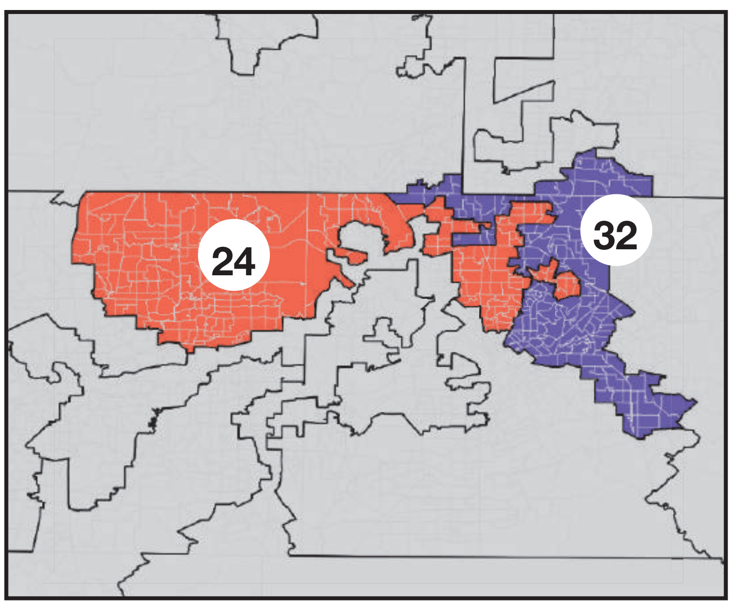


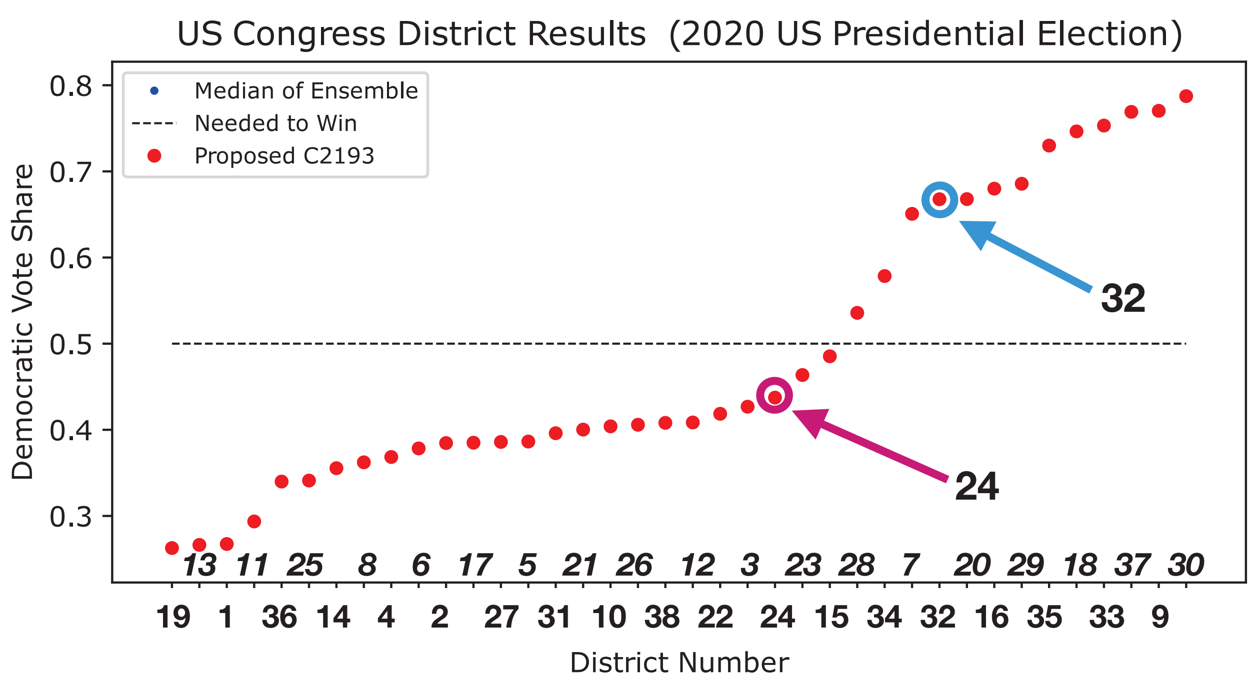
Let's compare...



2 blue districts,
3 red districts





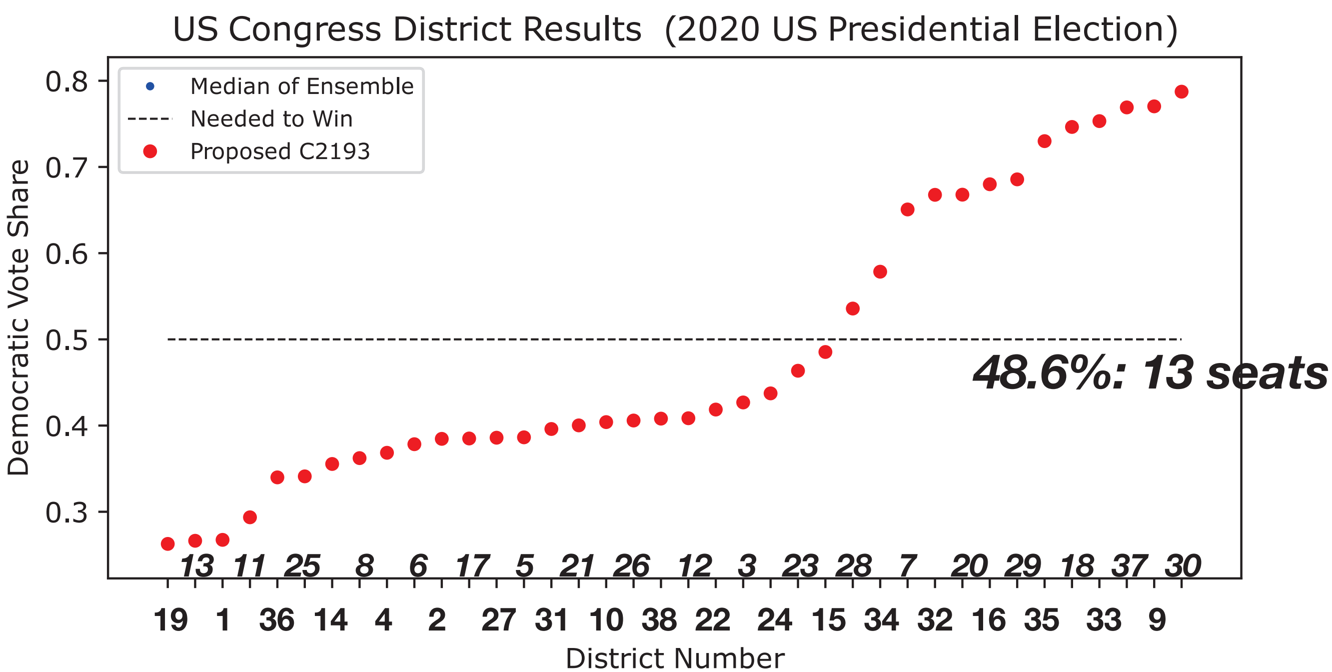


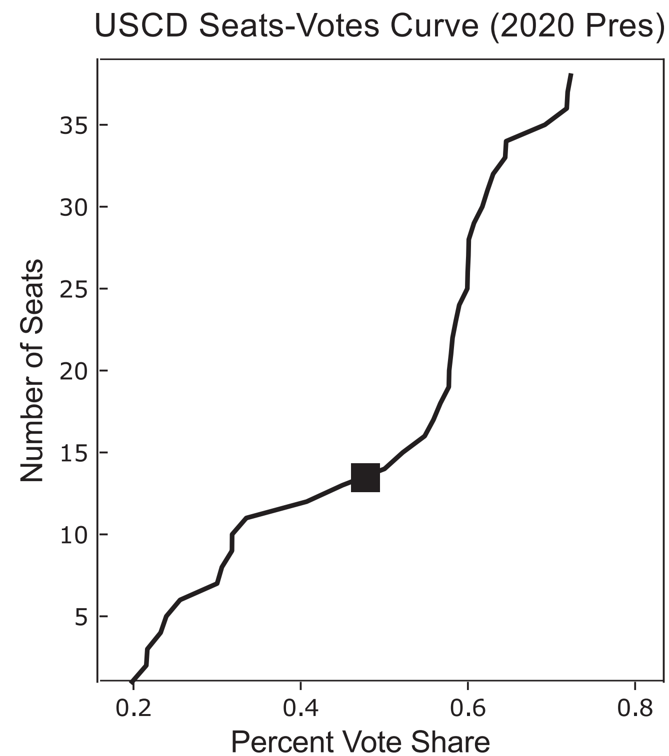
Seats-votes curve

- How many seats would Party A get, if it received 40% of the vote? 50%? 60%?
- We can only speculate how people will vote in a different election.
- A reasonable assumption: a *uniform partisan swing.*
  - i.e. Vote shares in each district will shift uniformly based on overall voter sentiment.
  - e.g.: If the overall share for Party A increases by 5%, its share in each district will increase by 5%
- We use this assumption, plus the actual observed vote share vector, to define a *seats-votes curve*.

Seats-votes curve

- How many seats would Party A get, if it received 40% of the vote? 50%? 60%?
- Use uniform partisan swing assumption to estimate how many seats Party A will win.

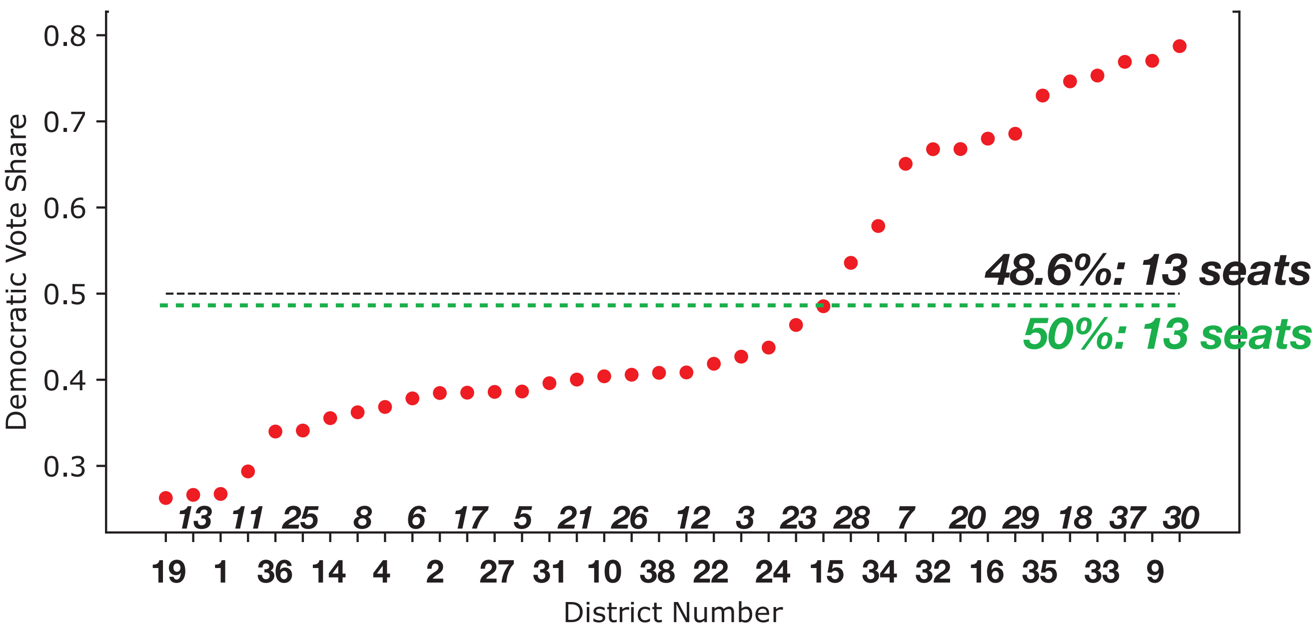


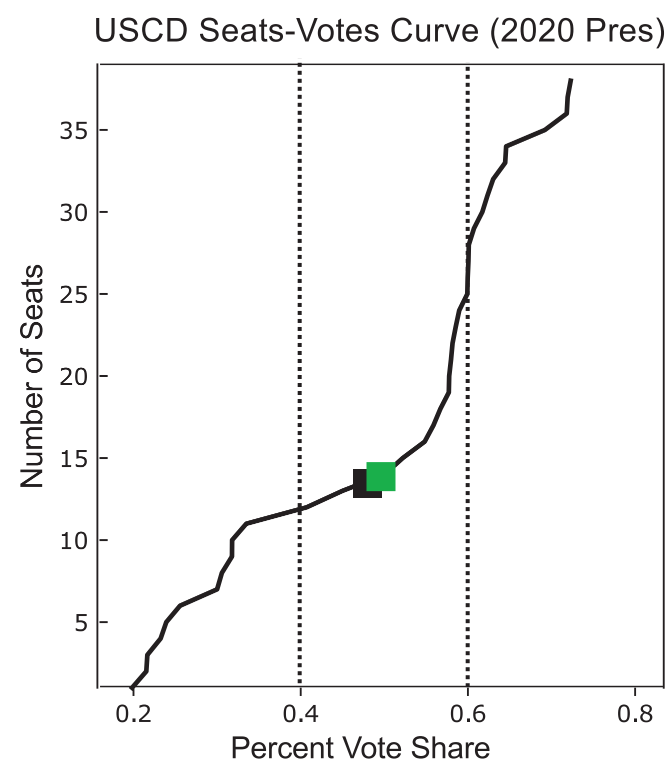


*by convention we use the average of the vote share vector, rather than the statewide vote average

Seats-votes curve

- How many seats would Party A get, if it received 40% of the vote? 50%? 60%?
- Use uniform partisan swing assumption to estimate how many seats Party A will win.

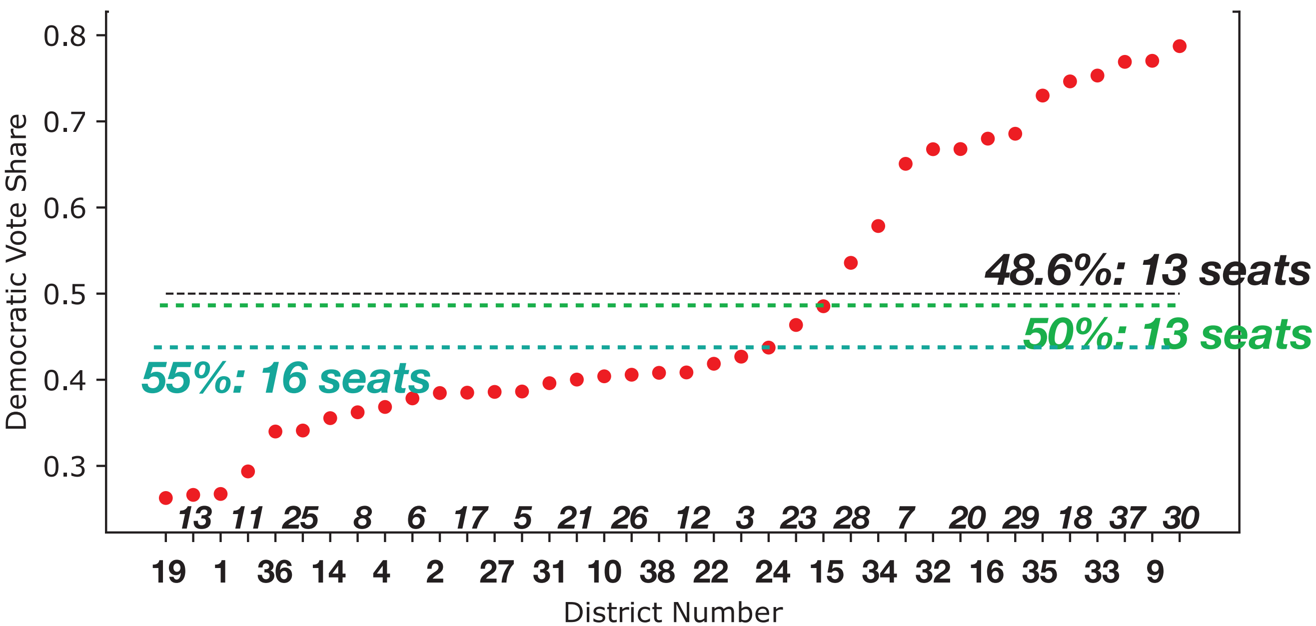


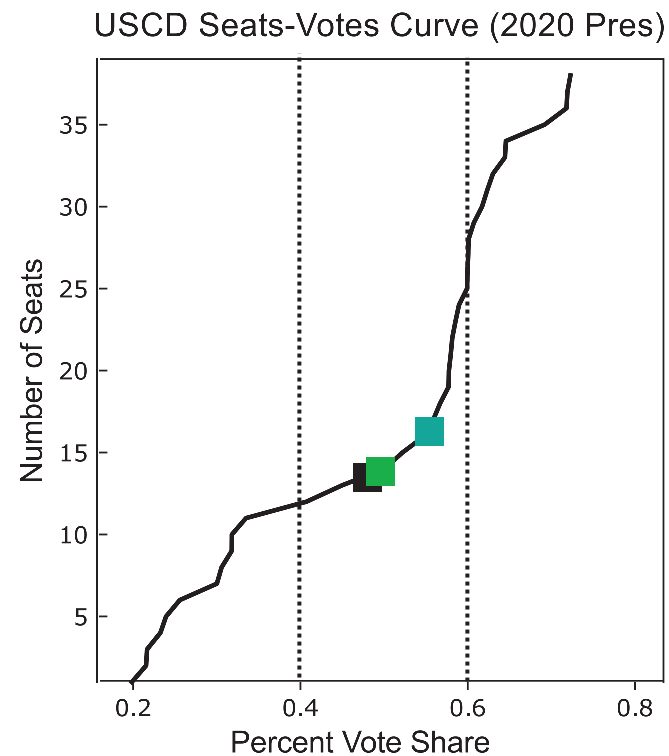


*by convention we use the average of the vote share vector, rather than the statewide vote average

Seats-votes curve

- How many seats would Party A get, if it received 40% of the vote? 50%? 60%?
- Use uniform partisan swing assumption to estimate how many seats Party A will win.

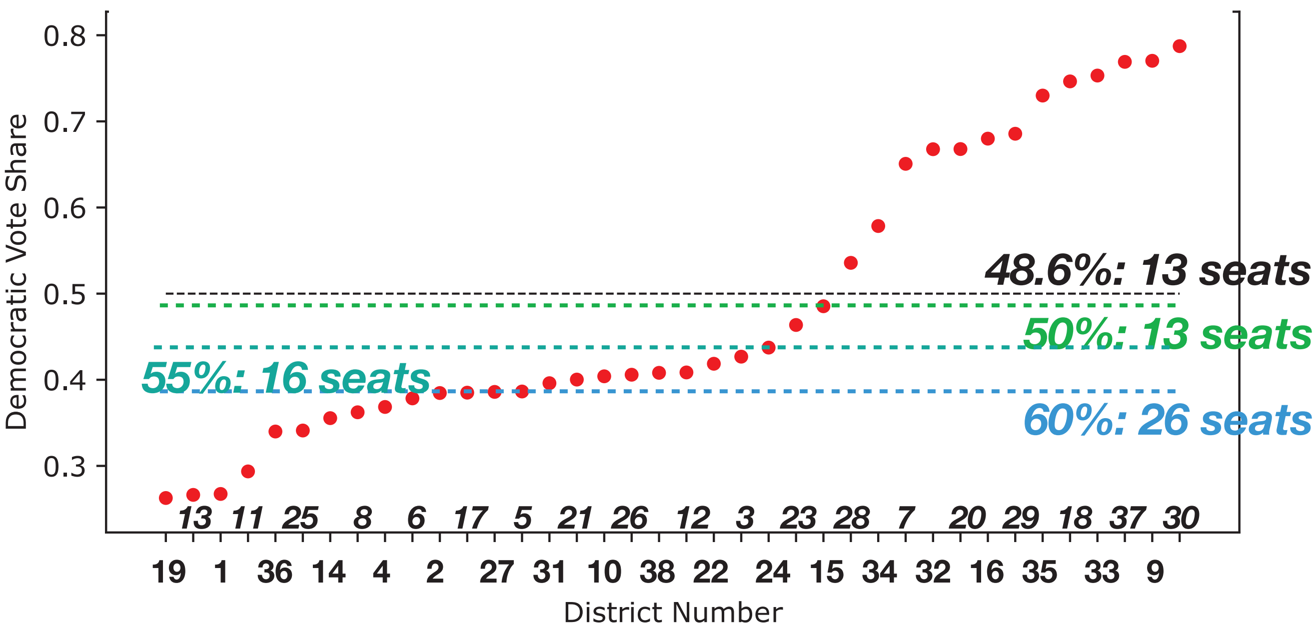


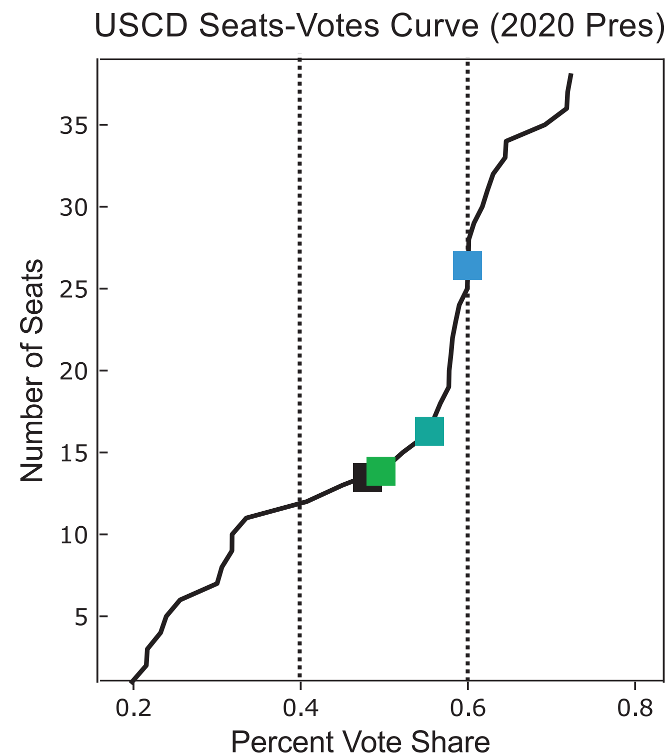


*by convention we use the average of the vote share vector, rather than the statewide vote average

Seats-votes curve

- How many seats would Party A get, if it received 40% of the vote? 50%? 60%?
- Use uniform partisan swing assumption to estimate how many seats Party A will win.

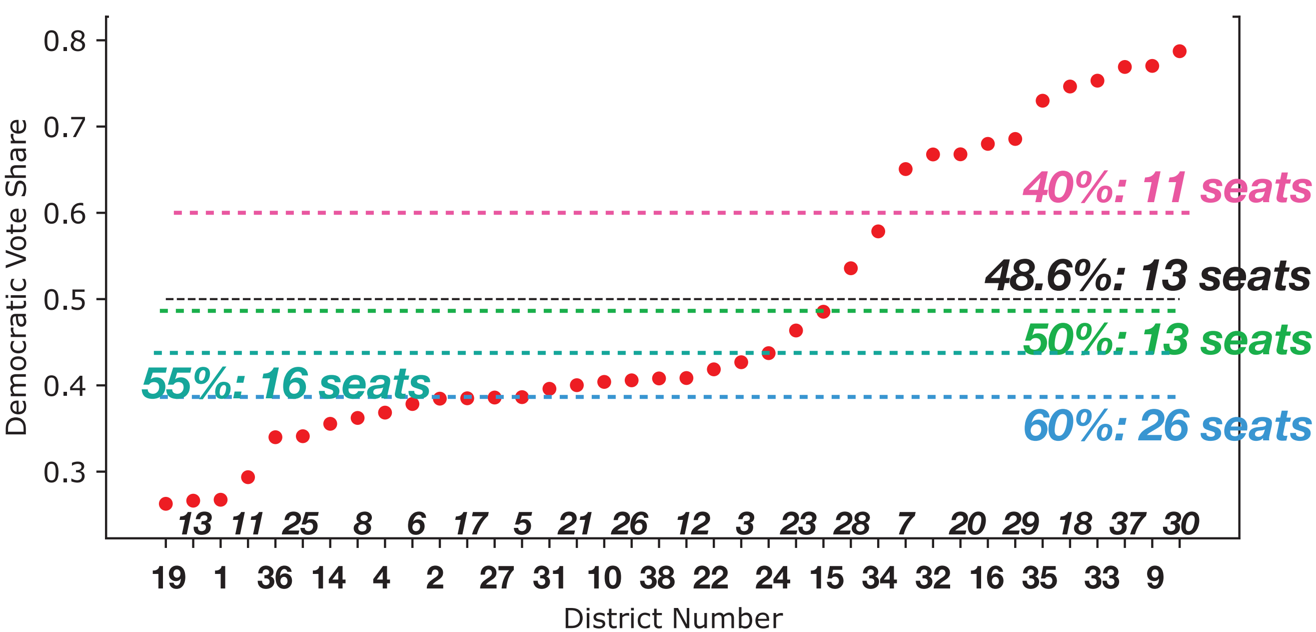


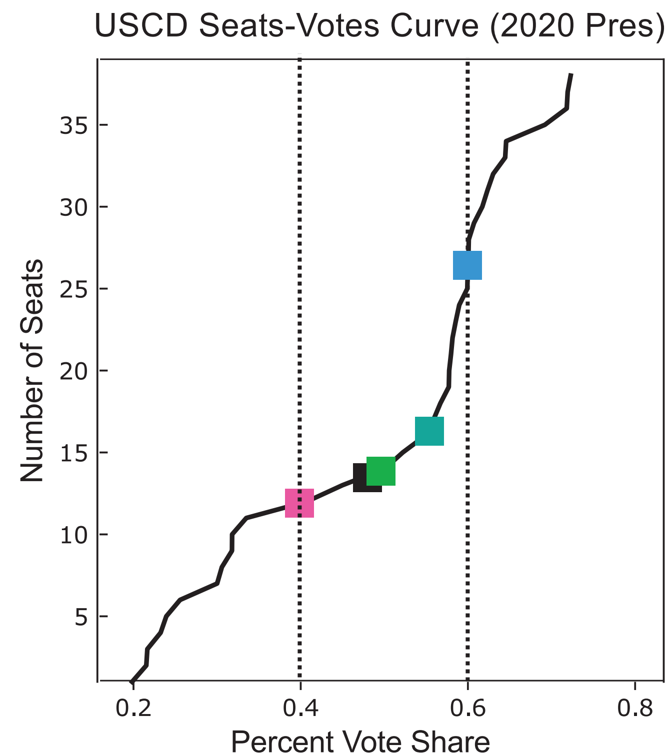


*by convention we use the average of the vote share vector, rather than the statewide vote average

Seats-votes curve

- How many seats would Party A get, if it received 40% of the vote? 50%? 60%?
- Use uniform partisan swing assumption to estimate how many seats Party A will win.

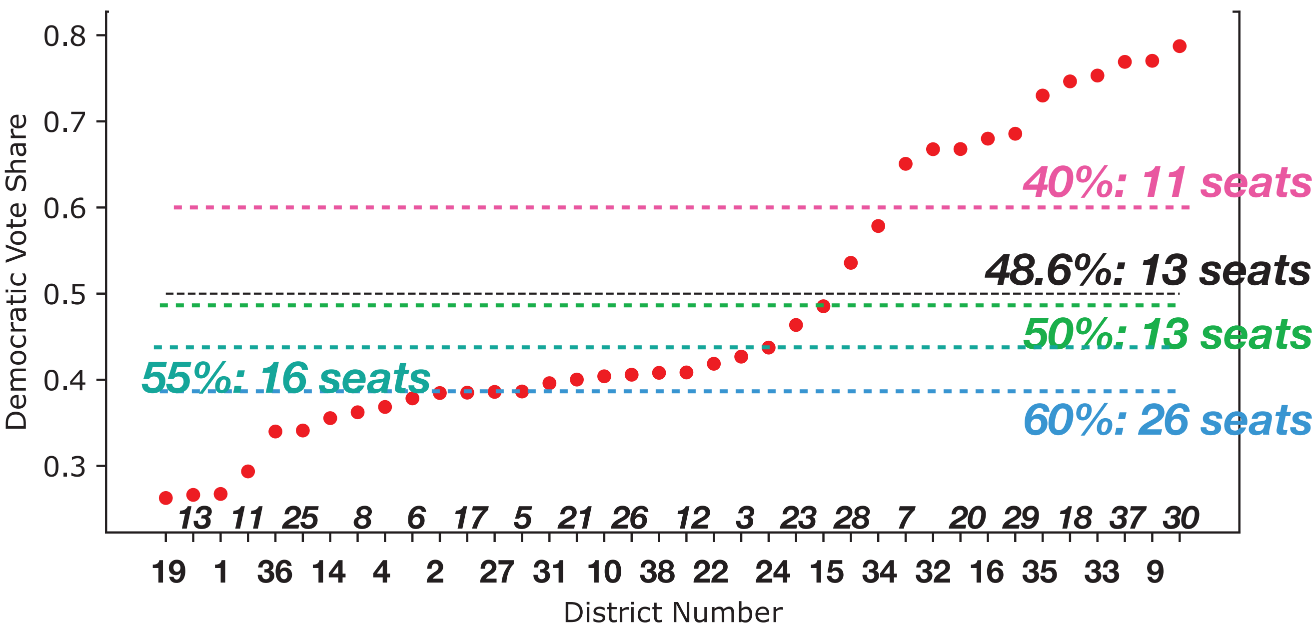


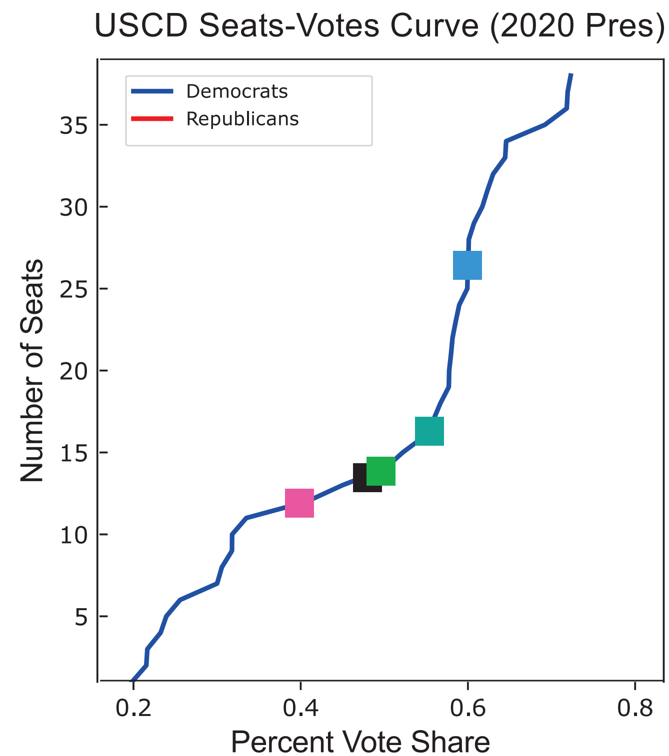


*by convention we use the average of the vote share vector, rather than the statewide vote average

Seats-votes curve

- This is the curve for the Democratic Party. How does the corresponding curve for the Republicans compare?

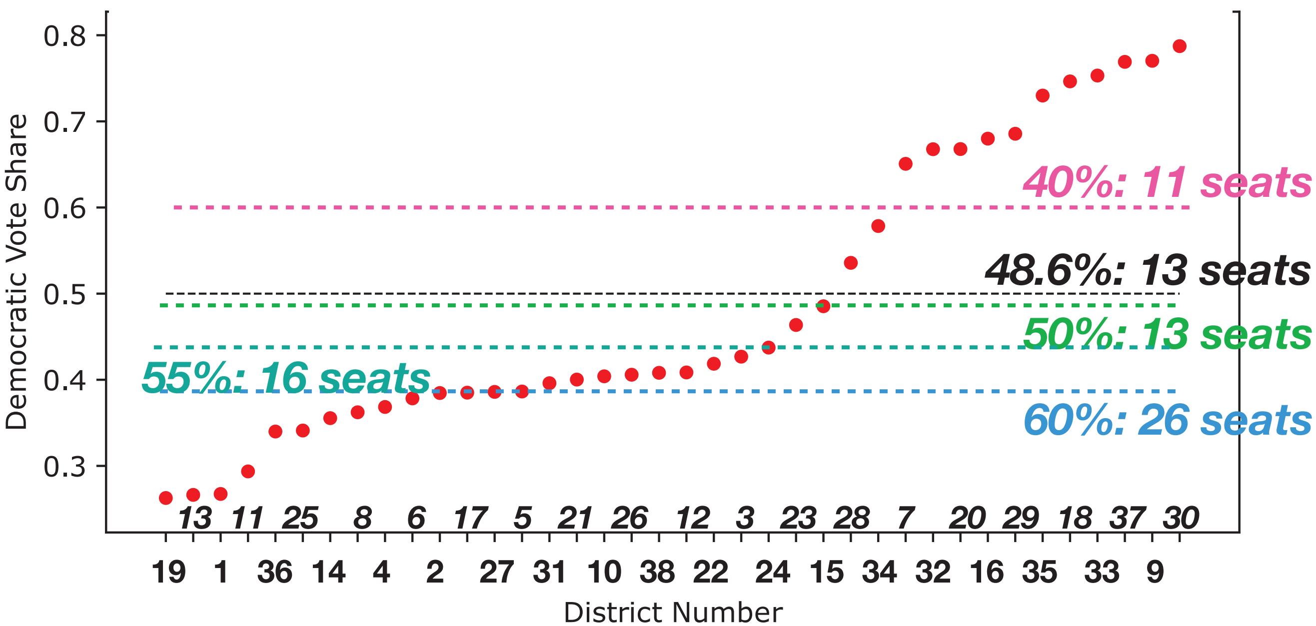


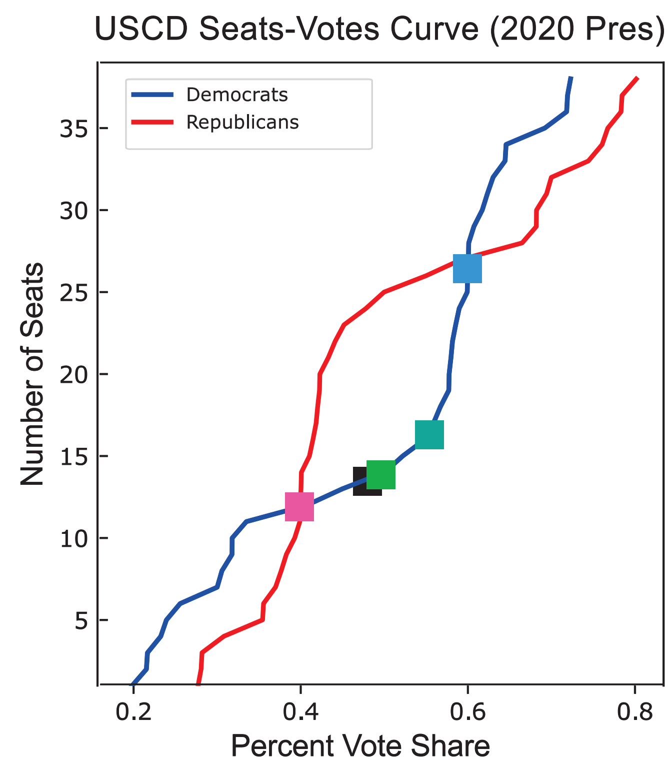


*by convention we use the average of the vote share vector, rather than the statewide vote average

Seats-votes curve

- This is the curve for the Democratic Party. How does the corresponding curve for the Republicans compare?
- Comparing D (blue) vs. R (red), the Republicans have a ***much*** more advantageous S-V curve in the range of feasible outcomes.

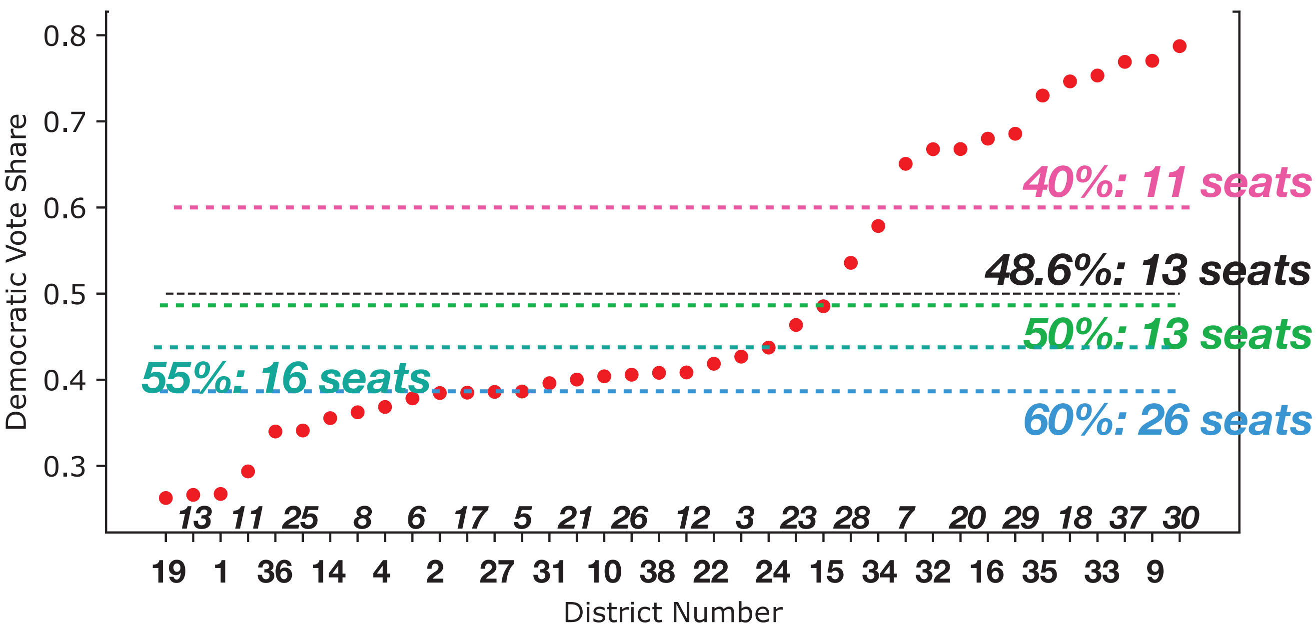


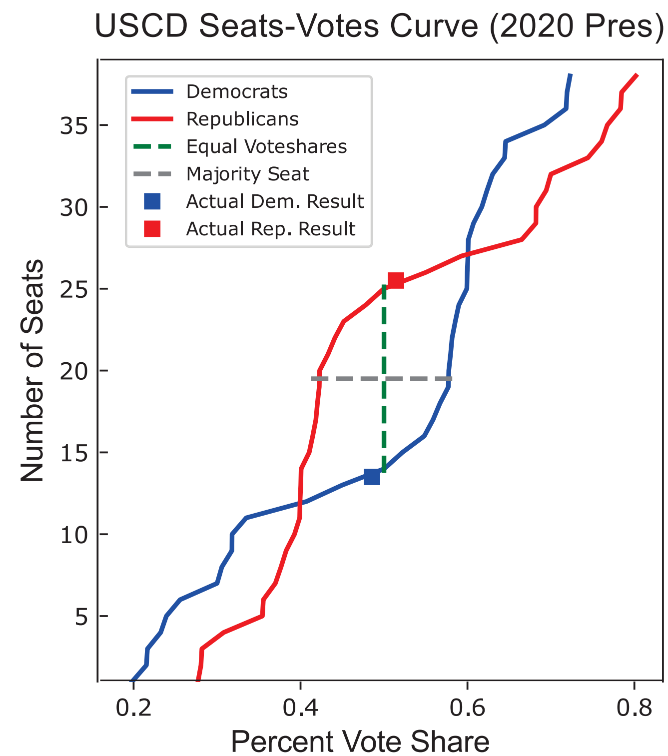


*by convention we use the average of the vote share vector, rather than the statewide vote average

Seats-votes curve

- This is the curve for the Democratic Party. How does the corresponding curve for the Republicans compare?
- Comparing D (blue) vs. R (red), the Republicans have a ***much*** more advantageous S-V curve in the range of feasible outcomes.



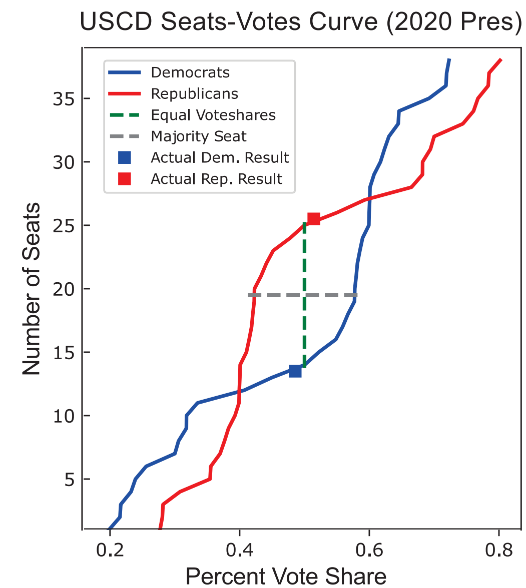


*by convention we use the average of the vote share vector, rather than the statewide vote average

Partisan Bias: a quantitative measure of partisan asymmetry

- How many seats would each party get, if the vote splits 50/50?*

PB = (#Seats for Party A) - (#Seats for Party B)



For C2193 (plan enacted by TX Lege this fall)

PB = (13 Democratic seats) - (25 Republican seats) = ***-12***

For C2100 (plan in use from 2013-2021)

PB = (20 Democratic seats) - (16 Republican seats) = ***+4***

*by convention we use the average of the vote share vector, rather than the statewide vote average

Mean-median score: a quantitative score directly based on the vote share vector

Definitions (same as you learned in your statistics class)

- Let $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ be the vote share vector for a plan with n districts
- The *mean* is the average of these numbers:

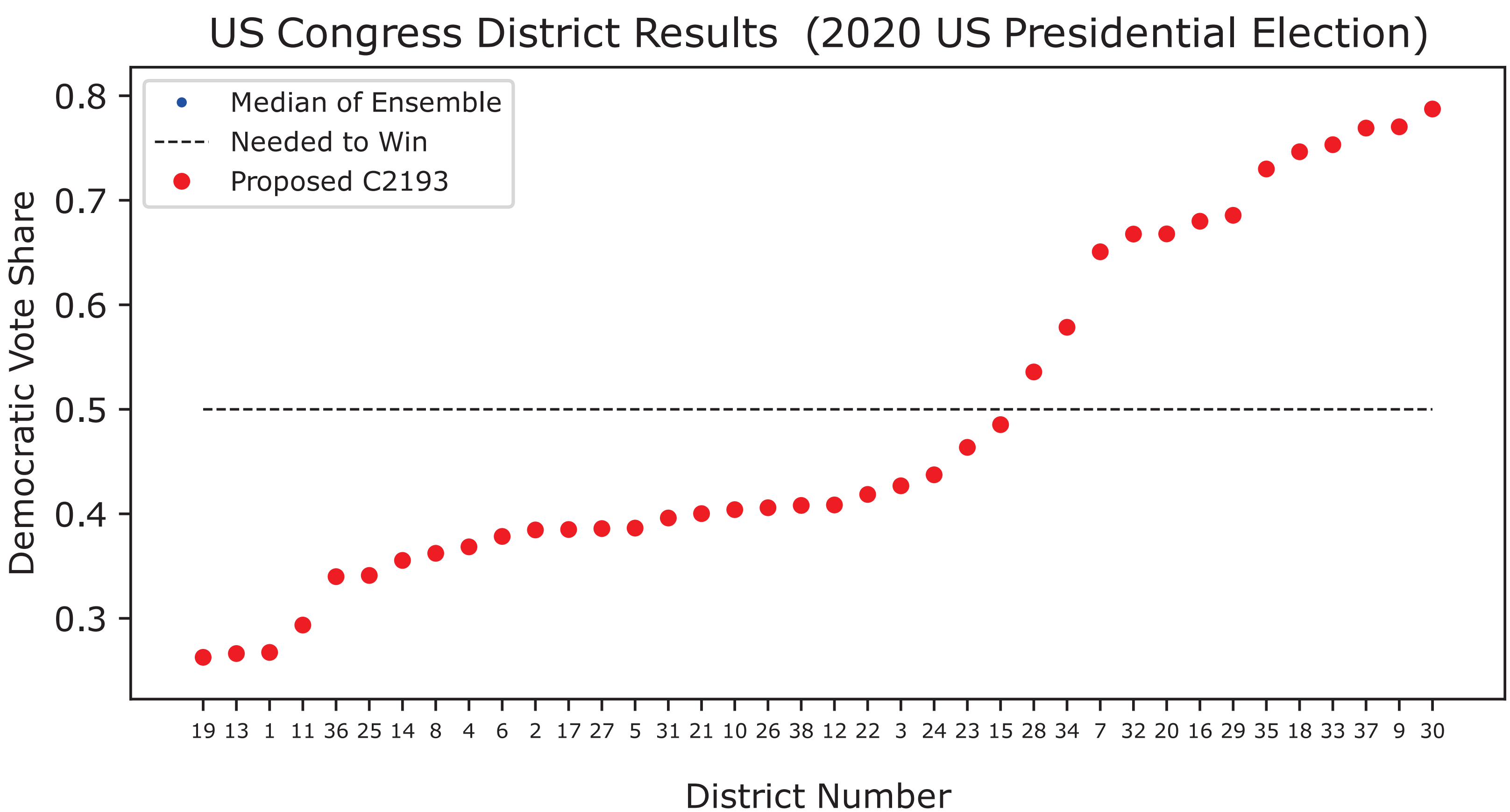
$$\bar{v} = \frac{1}{n}\left(v_1 + v_2 + \ldots + v_n\right)$$

- The *median* is the number which is below exactly half of the v numbers: i.e. the *50% percentile*
  - For US Congress, $n = 38$.
  - If the vote share vector is ordered, then $v_{med} = (v_{19} + v_{20})/2$

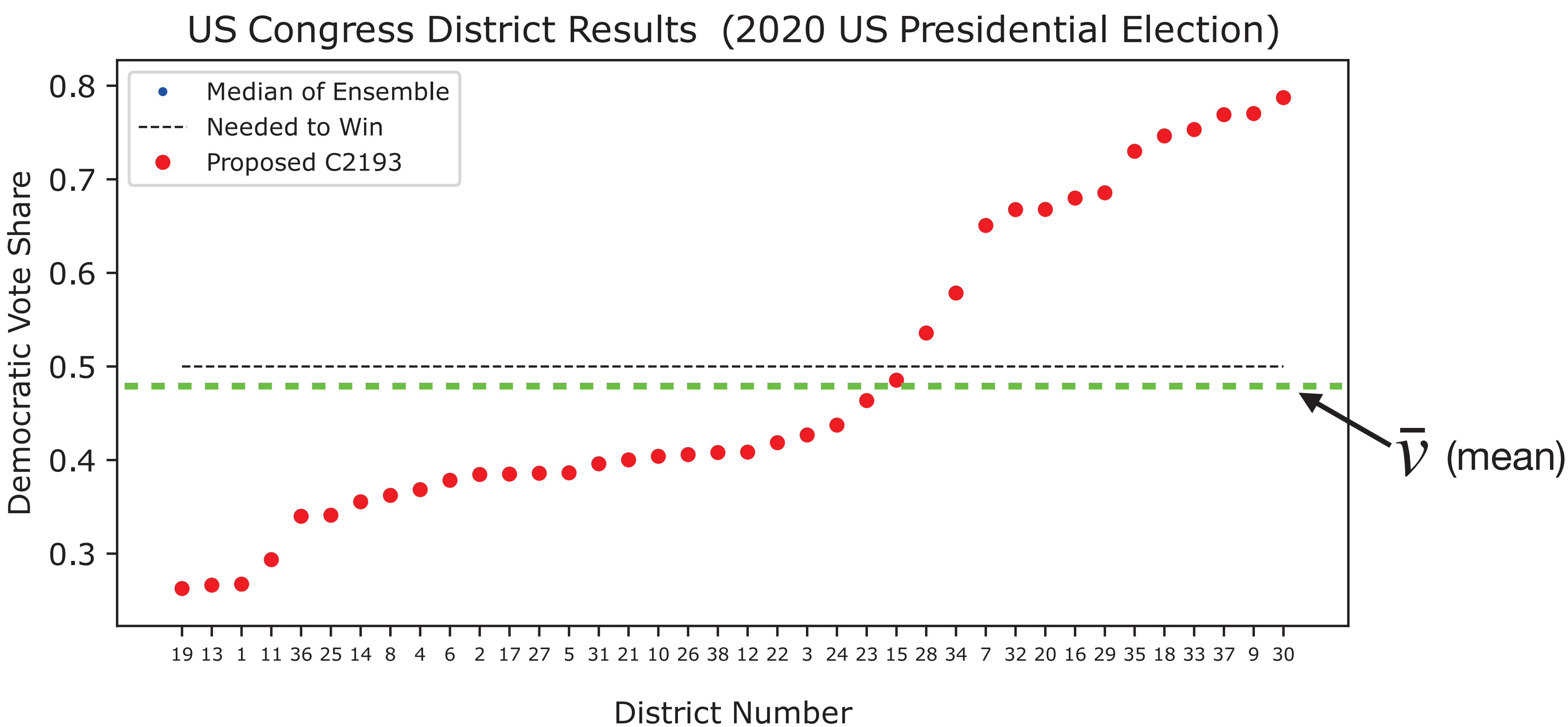
Finally: the mean-median score is the difference between these numbers

$$MM(\mathbf{v}) = v_{med} - \bar{v}$$

$$MM(\mathbf{v}) = \text{ median - mean}$$

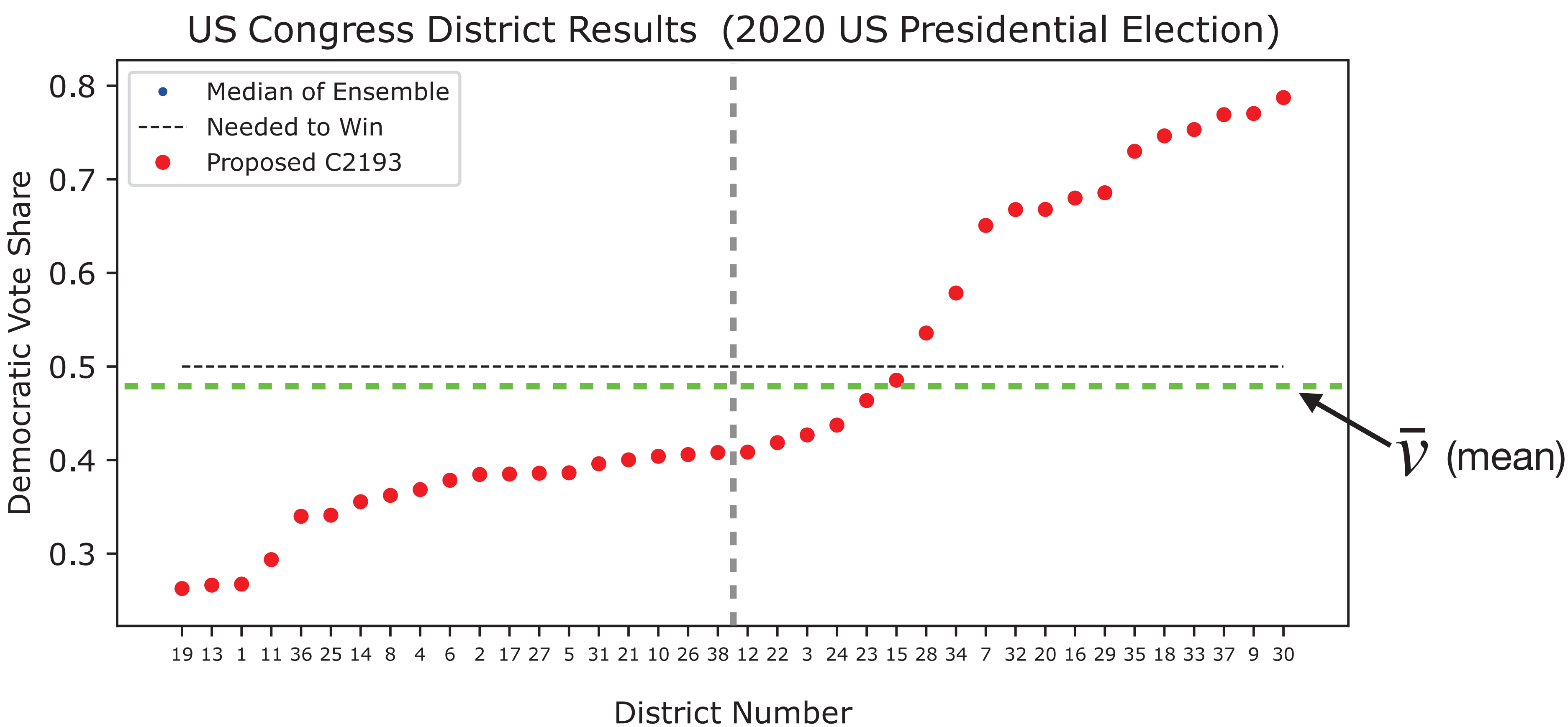


$$MM(\mathbf{v}) = \text{ median - mean}$$



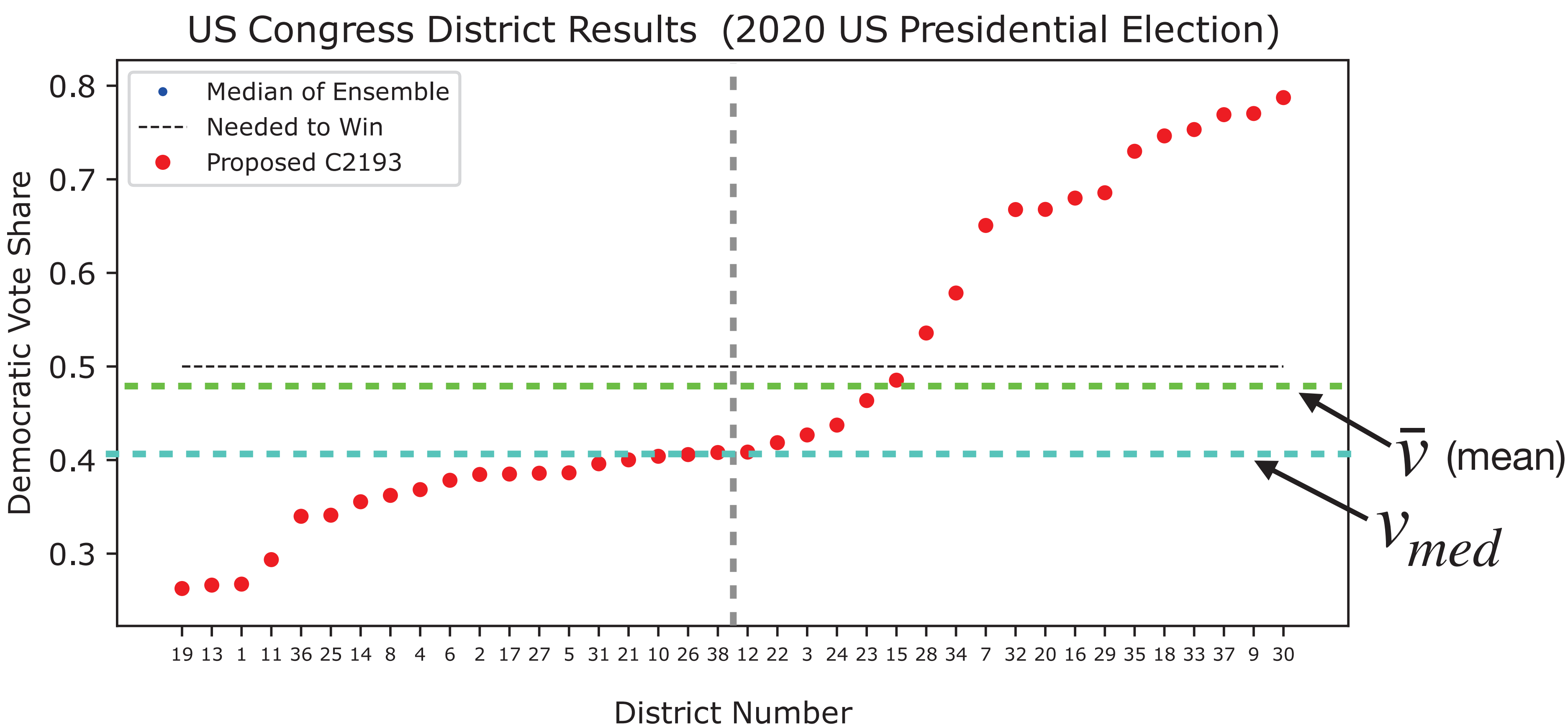
- To find the median:
  Look for the center (19 above, 19 below)

$$MM(\mathbf{v}) = \text{ median - mean}$$



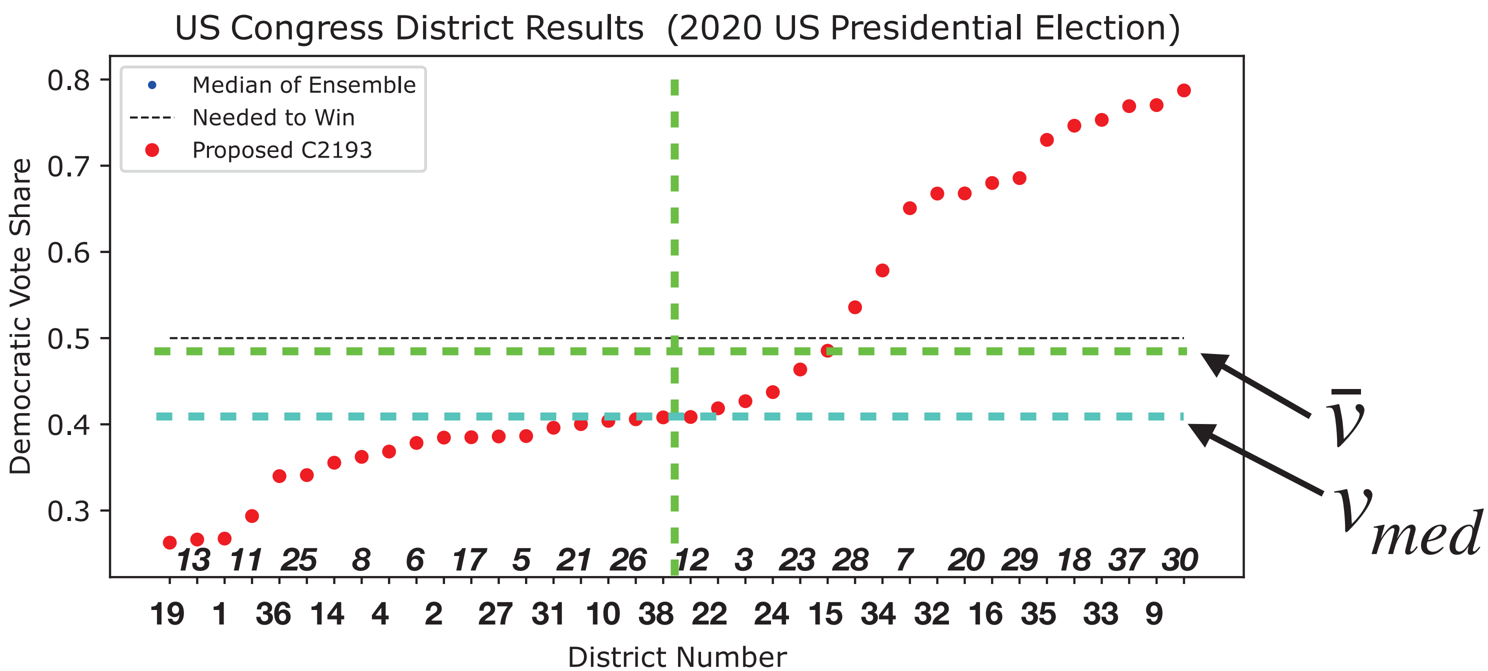
- To find the median: Look for the center (19 above, 19 below)

$$MM(\mathbf{v}) = \text{ median - mean}$$



Mean-median score: a quantitative score directly based on the vote share vector

$$MM(\mathbf{v}) = v_{med} - \bar{v}$$

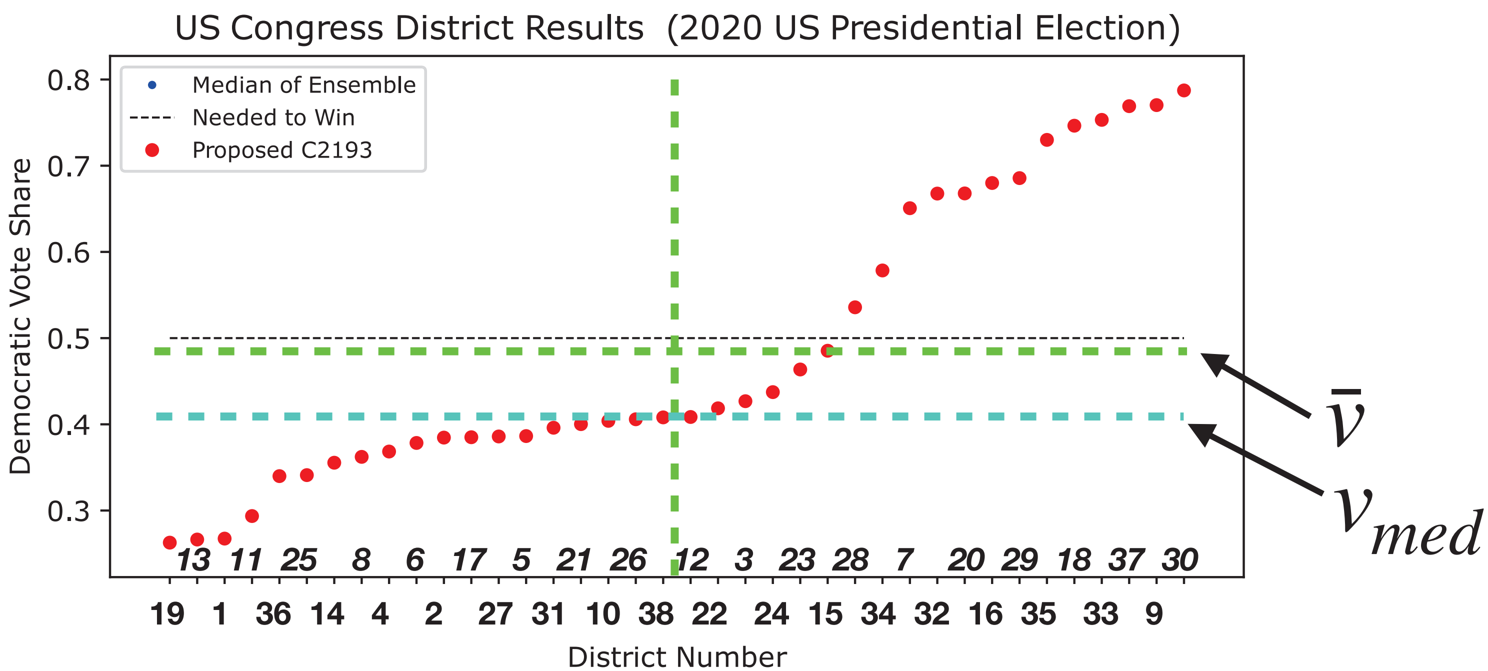


Intuitively: if the median is less than the mean, it means you are disadvantaged in the number of seats you can win.

Note 1: The name sounds like you should subtract the median from the mean, but it's the other way around

Mean-median score: a quantitative score directly based on the vote share vector

$$MM(\mathbf{v}) = v_{med} - \bar{v}$$



Intuitively: if the median is less than the mean, it means you are disadvantaged in the number of seats you can win.

Note 2: We multiply the MM score by 200.
The result exactly coincides with the difference in vote share needed to earn 1/2 of the seats.

Mean-median score: a quantitative score directly based on the vote share vector

$$MM(\mathbf{v}) = 200 \times (v_{med} - \bar{v})$$

MM(**v**) = (% Vote share needed by Party B)- (% Vote share needed by Party A)

For C2193 (plan enacted by TX Lege this fall)

MM = (42.3% needed by R) - (57.7% needed by D) = ***-15.4***

For C2100 (plan in use from 2013-2021)

MM = (50.6% needed by R) - (49.4% needed by D) = ***1.2***

Ok, so C2193 hugely advantages Republicans. But how do we know it's "gerrymandering"?

- After all, legislators had to change maps to re-balance population
- Insert quote from Rep. Hunter about how this was "unavoidable"
- But was it?

Ensemble analysis

- We need to know "normal" before we know "not normal".
- What does a "normal" map look like?
- If we could generate a large number of legal, unbiased maps, this gives us a picture of "normal"

Markov Chain Monte Carlo (MCMC) in a nutshell

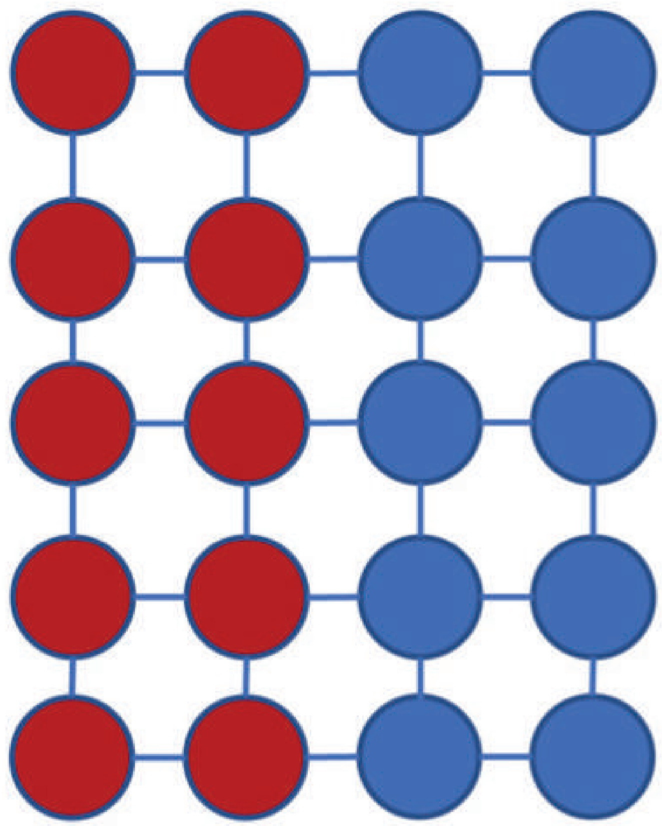
- Guided random search
- Ingredients
  - Space of possible solutions
  - Way to walk around in it
  - Metrics
    - Validator - condition that must be met
    - Objective - property to optimize
- Metropolis-Hastings (1953), developed for the hydrogen bomb (1953): the OG MCMC method

Markov Chain Monte Carlo (MCMC), for redistricting

- Guided random search
- Ingredients
  - Space of possible solutions: **all possible ways of dividing up Texas Census blocks into 38 groups**
  - Way to walk around in it: ***at each step we merge two adjacent districts together, and then split them up*** *(ReCom, DeFord et al. 2020)*
  - Metrics
    - Validator - ***Each of the 38 groups must be physically connected. Each must have equal population***
    - Objective - none used here
- We use *Gerrychain* - Python library developed by the Metric Geometry and Gerrymandering Group (MGGG) - **pypi.org/project/gerrychain**

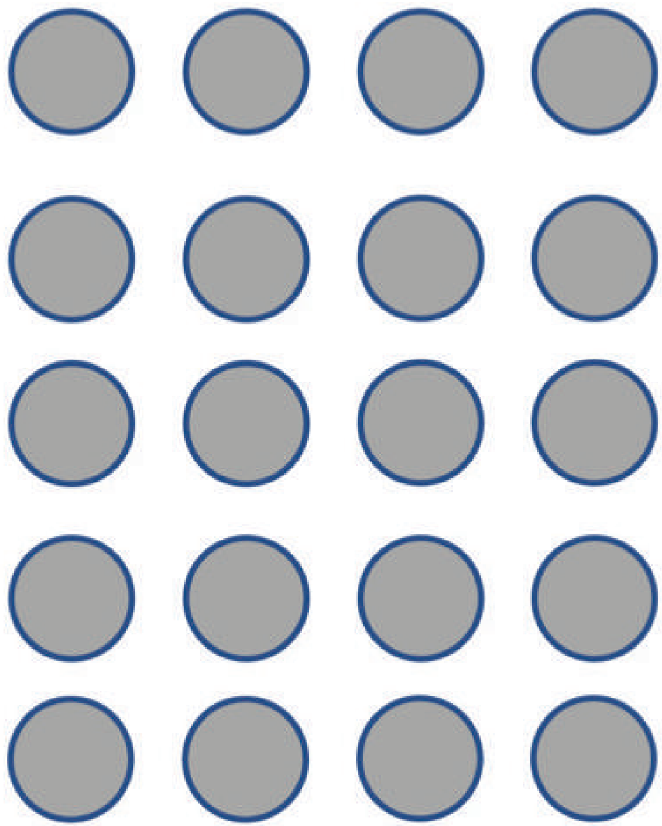
How Recombination (ReCom) works:

1. Select pair of adjacent districts uniformly at random



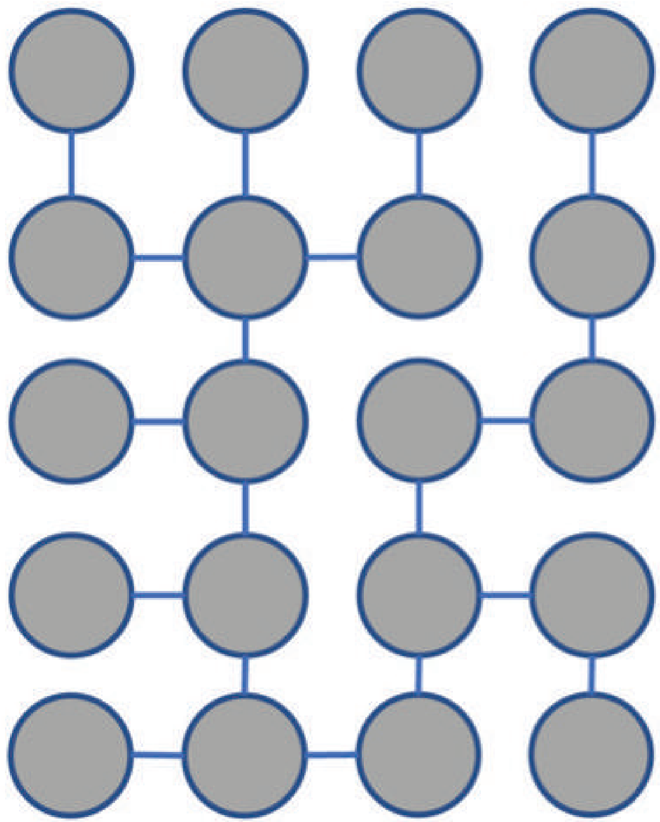
How Recombination (ReCom) works:

2. Merge into a single district



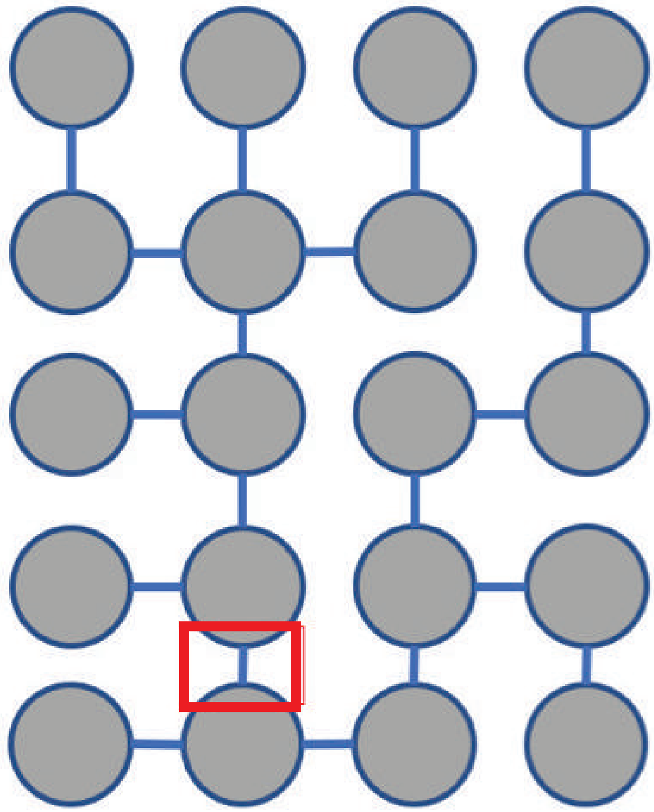
How Recombination (ReCom) works:

3. Find a spanning tree



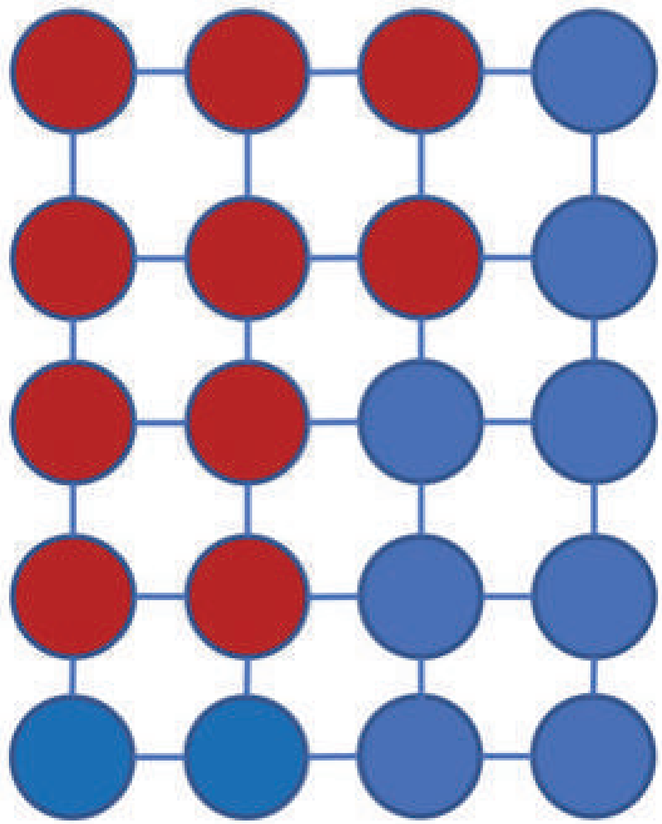
How Recombination (ReCom) works:

4. Cut one edge uniformly at random

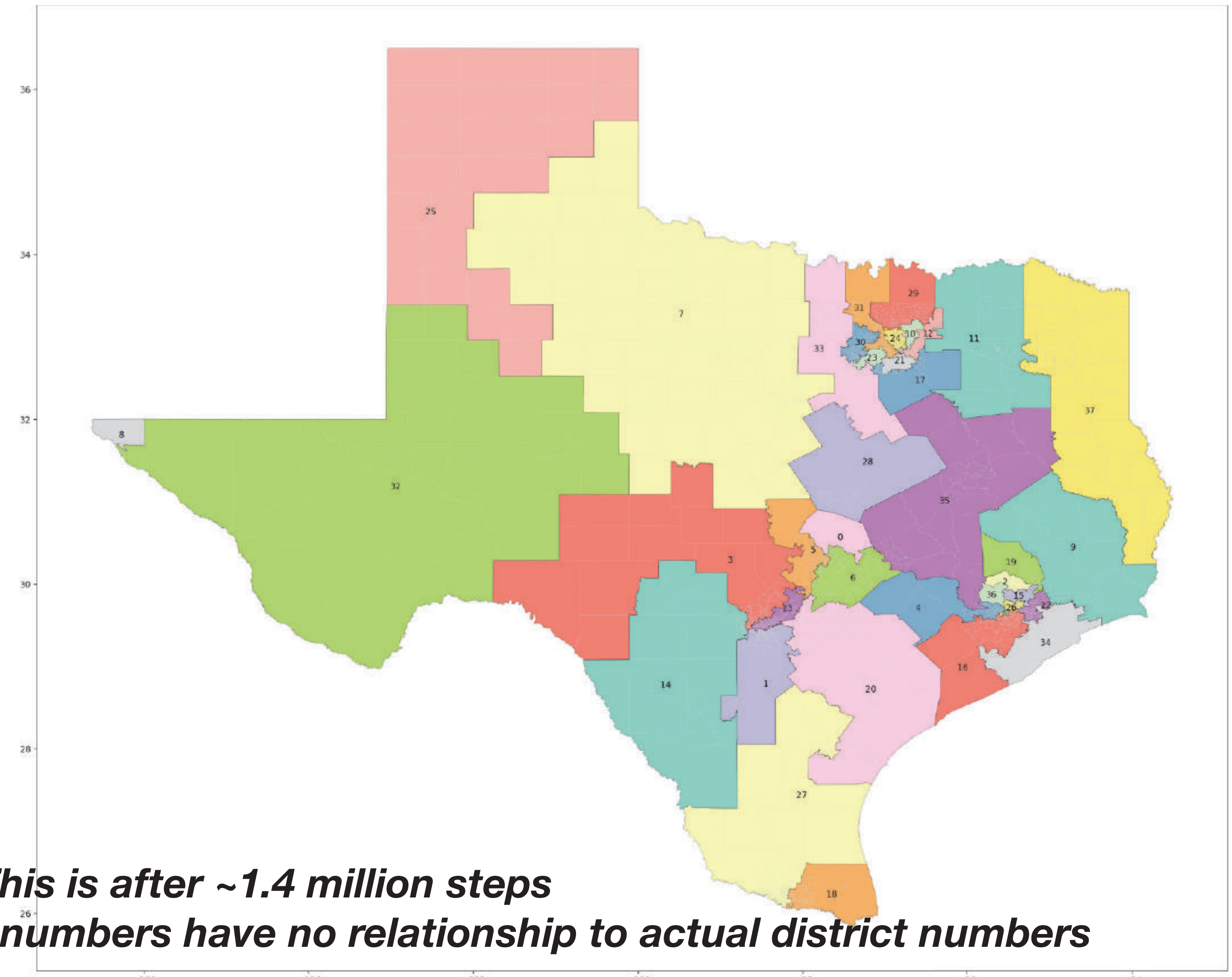


How Recombination (ReCom) works:

5. This generates two “mixed-up" districts: check validators

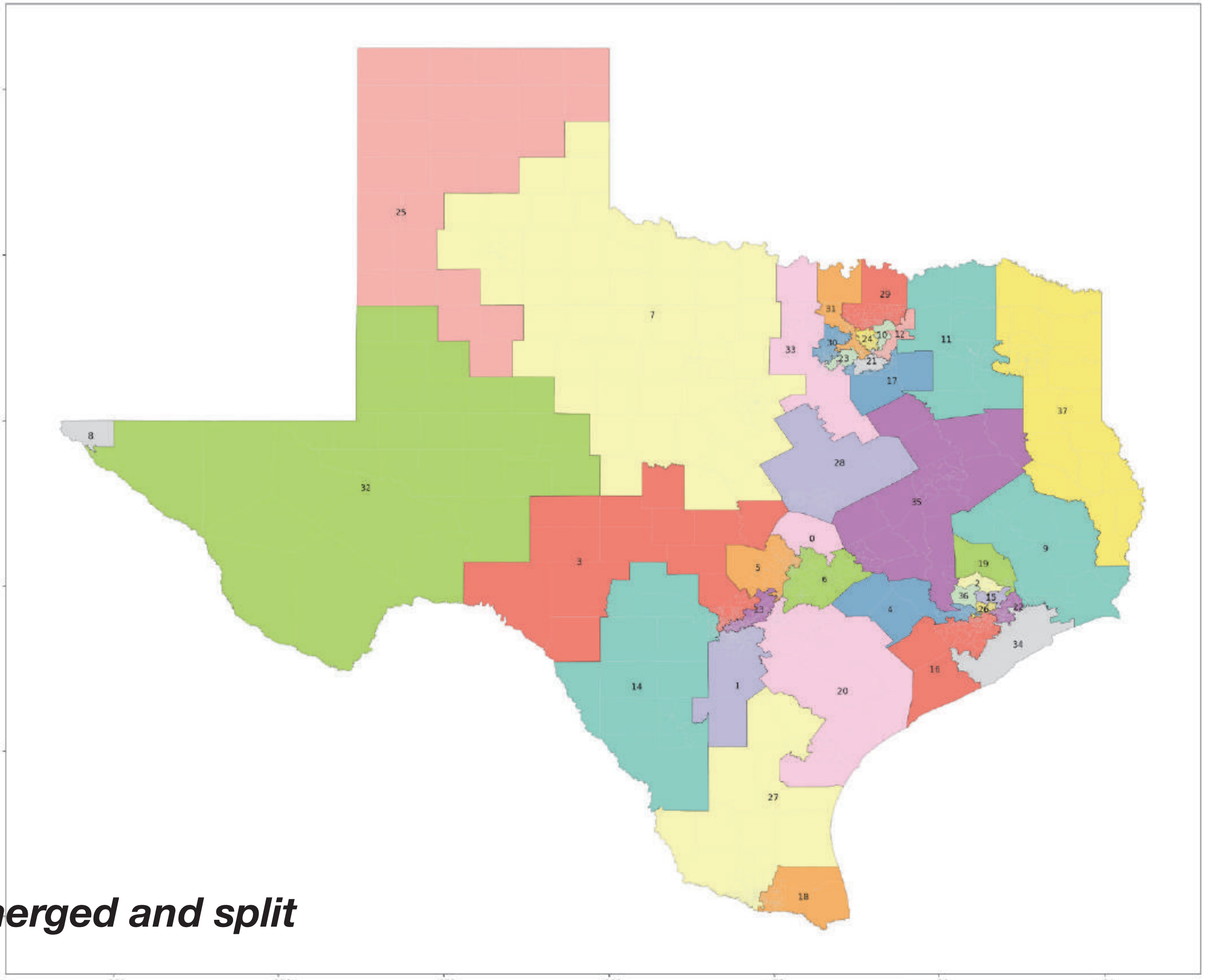


Let's see it work on Texas



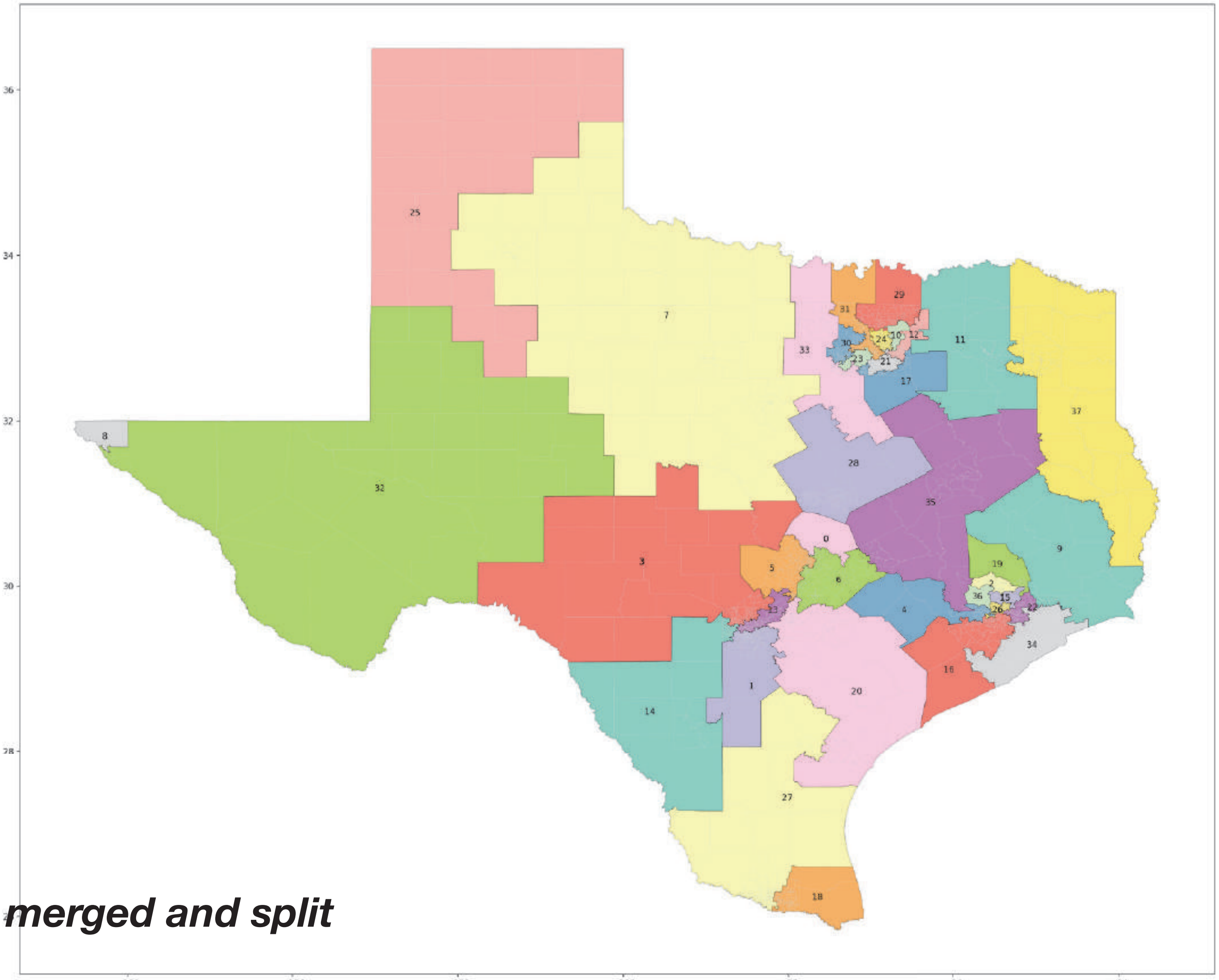
Note: This is after ~1.4 million steps
District numbers have no relationship to actual district numbers

Let's see it work on Texas



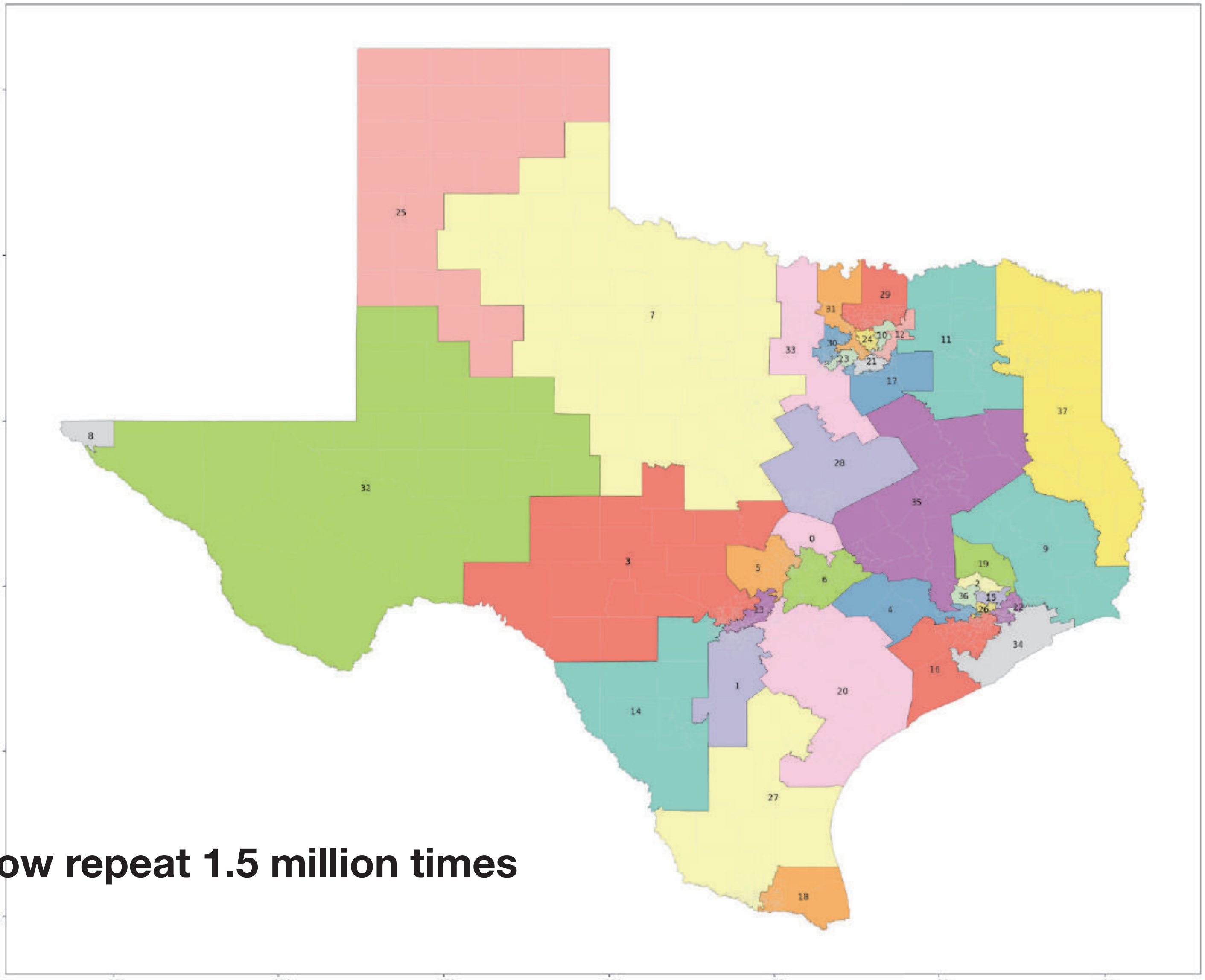
3 and 5 merged and split

Let's see it work on Texas



3 and 14 merged and split

Let's see it work on Texas



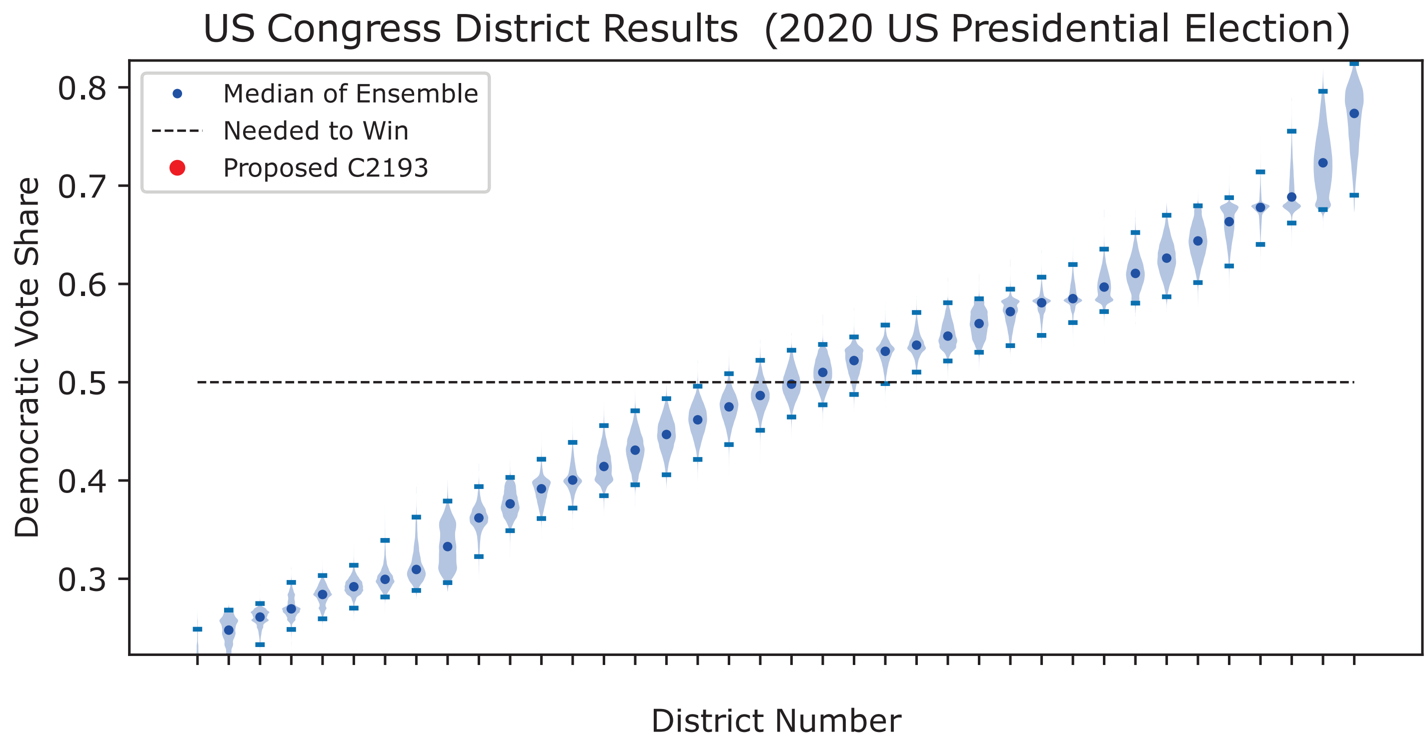
...we now repeat 1.5 million times

For each of our 1.5 million maps, we compute:

- Vote-share vector
- Partisan bias (at 50%)
- Mean-median score
- ...anything else we might compute for an actual plan

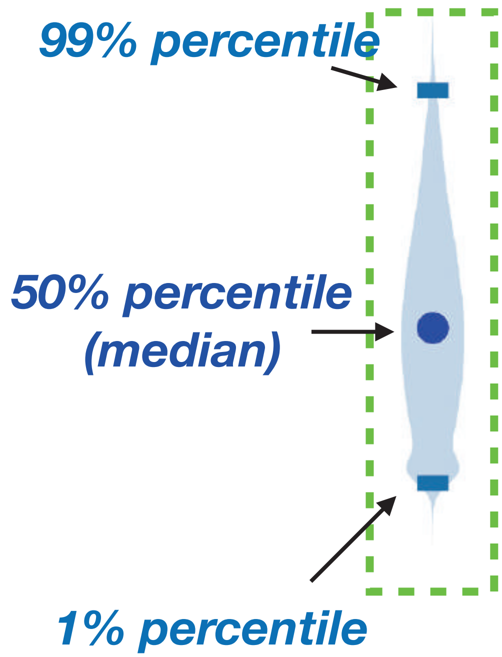
Vote Share Curves

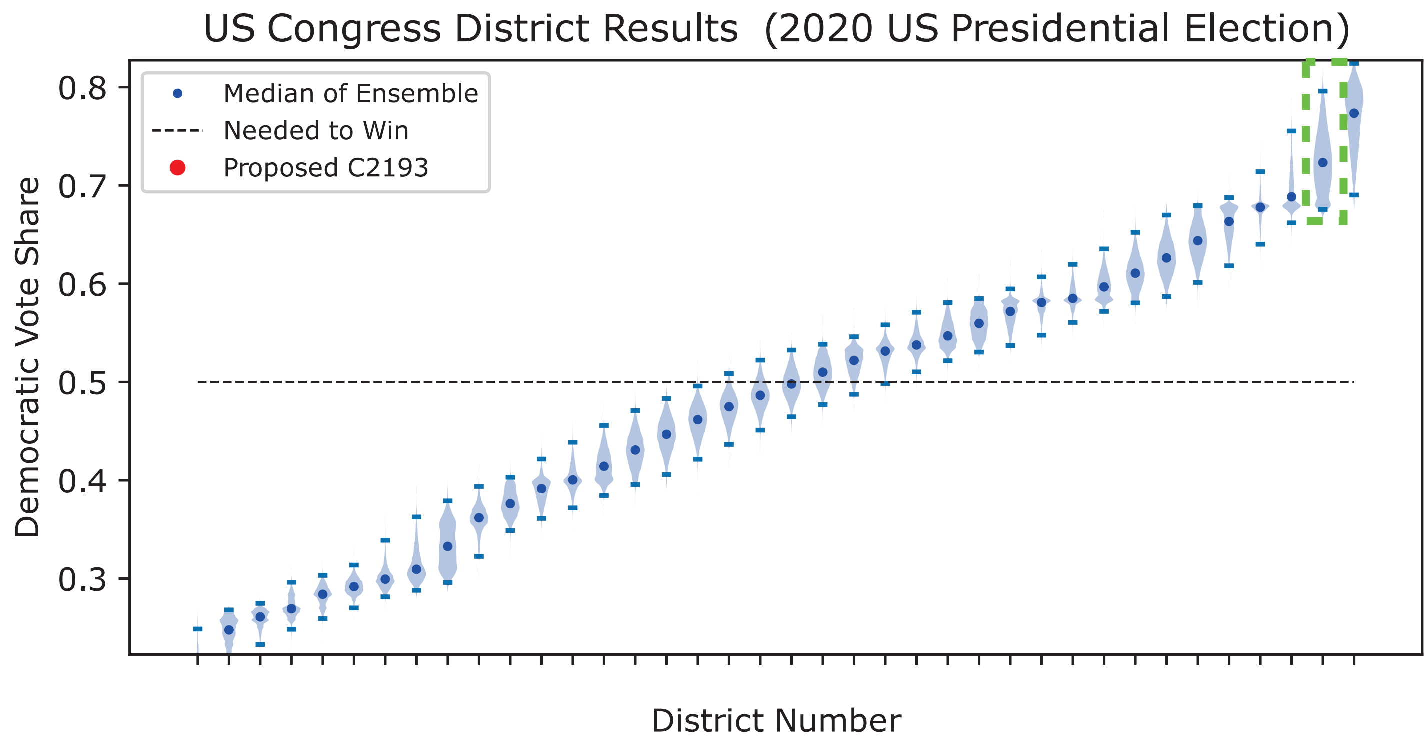
- For each map, order districts by increasing vote share
- 1.5 million legal Congressional maps



Vote Share Curves

- For each map, order districts by increasing vote share
- 1.5 million legal Congressional maps

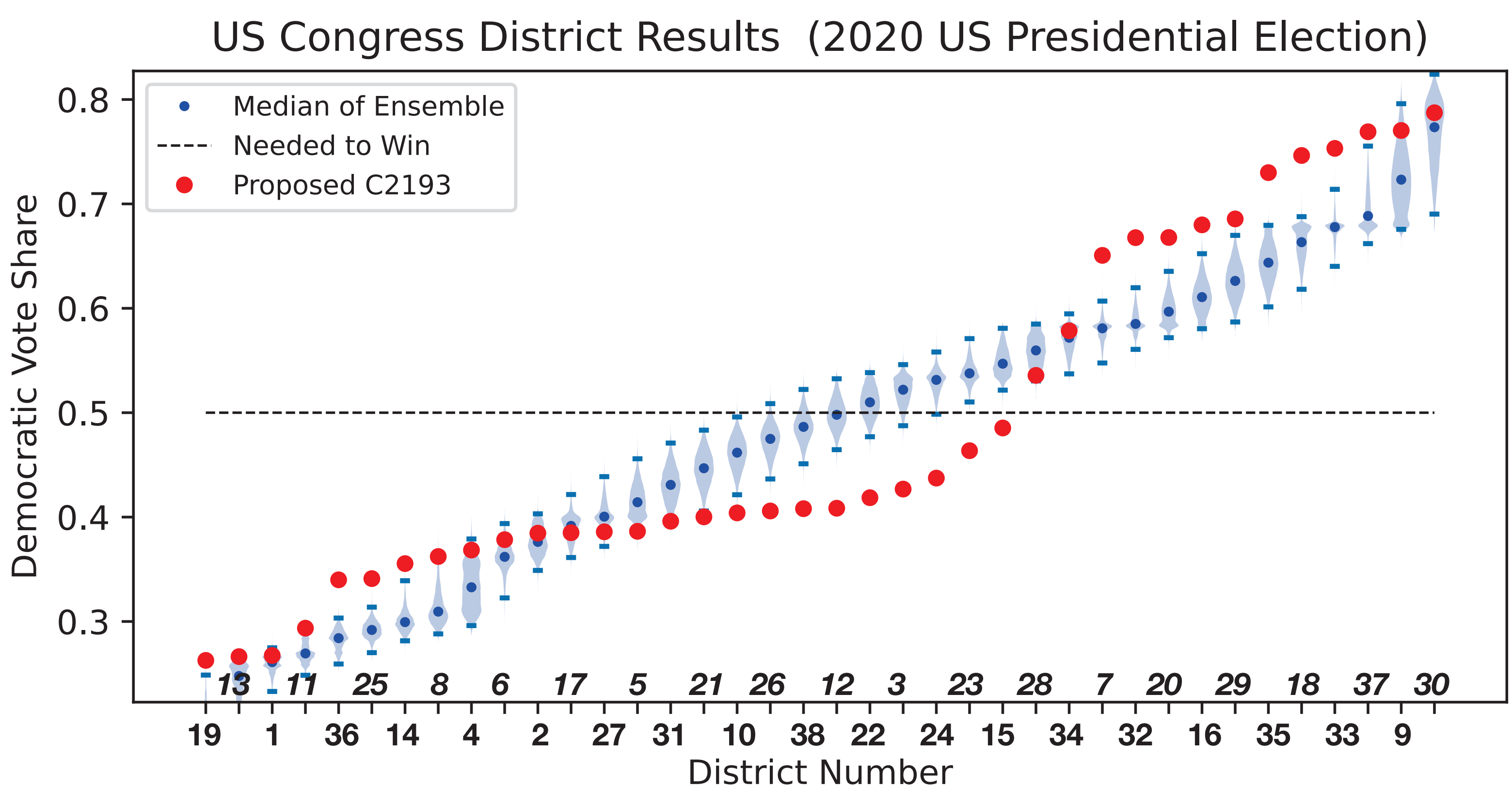




Violin plots give us a convenient way to illustrate the entire distribution of a statistic (here, the *2nd largest Democratic* vote share)

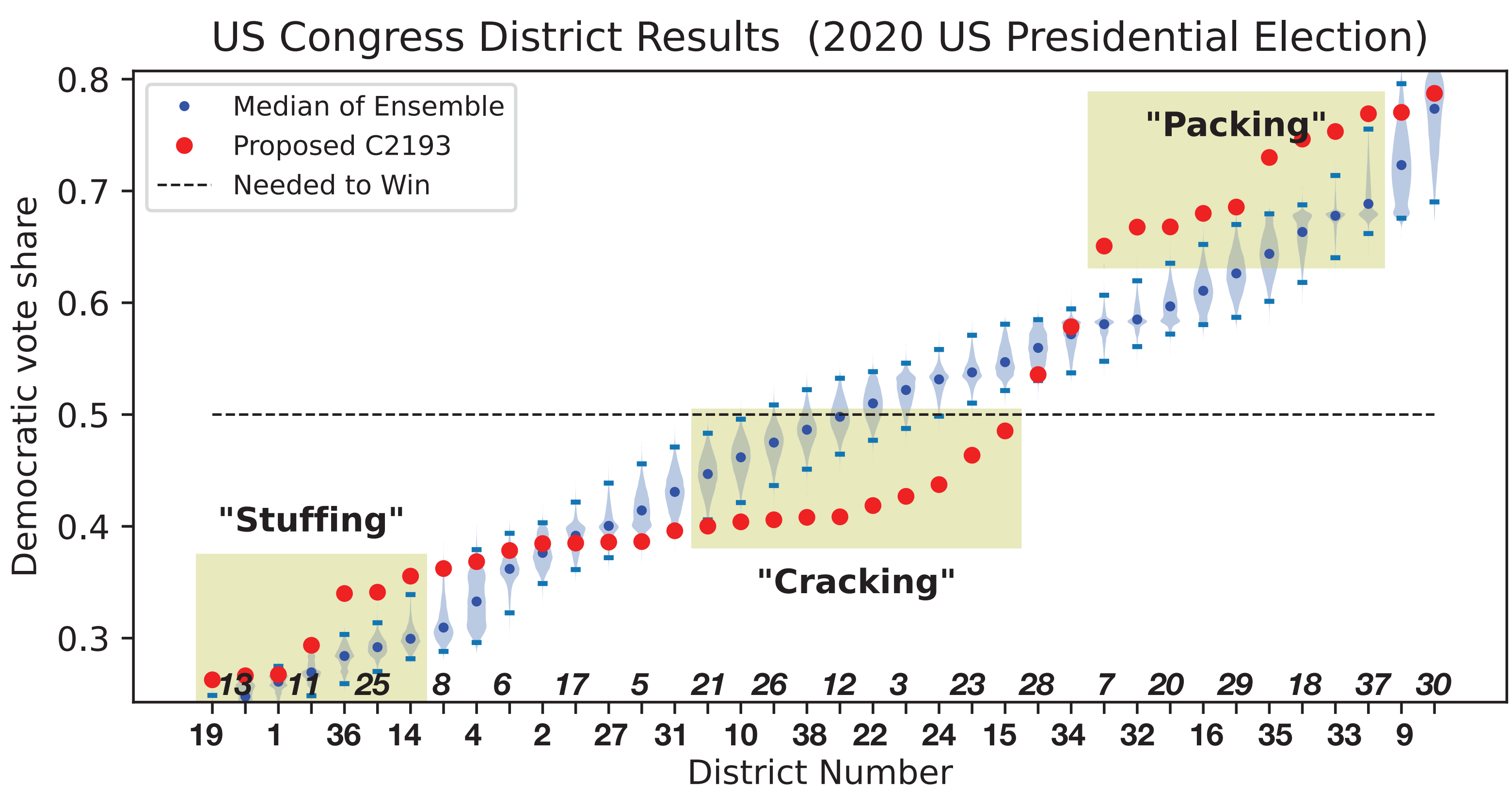
Vote Share Curves

- For each map, order districts by increasing vote share
- 1.5 million legal Congressional maps

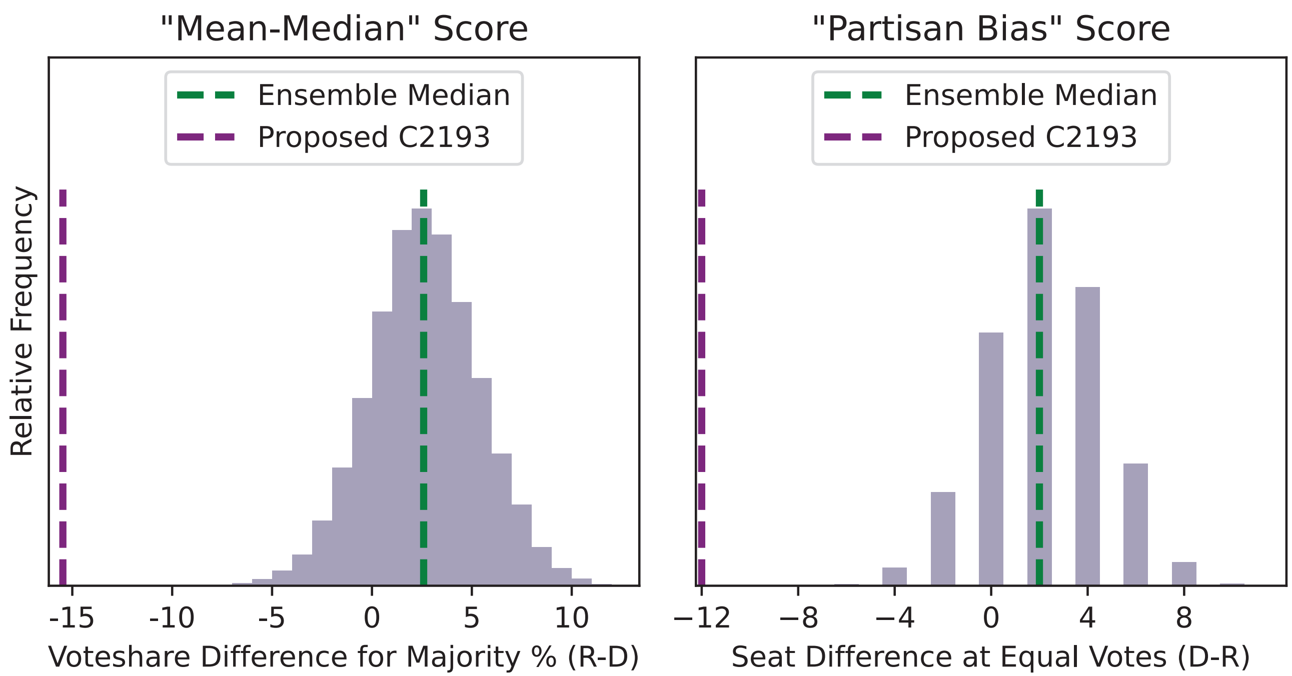


The enacted Congressional map shows distinct signatures of partisan gerrymandering

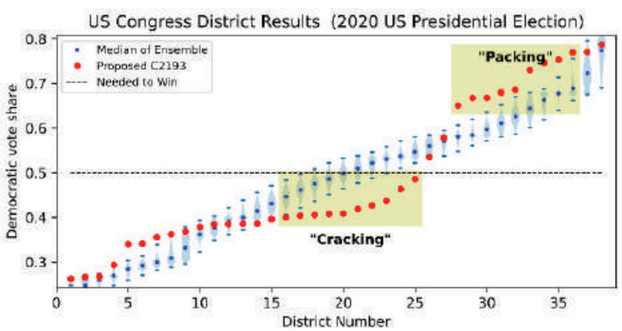
- ***Cracking:*** spreading the opposing party's voters across multiple districts
- ***Packing:*** concentrating the opposing party's voters into a few safe districts

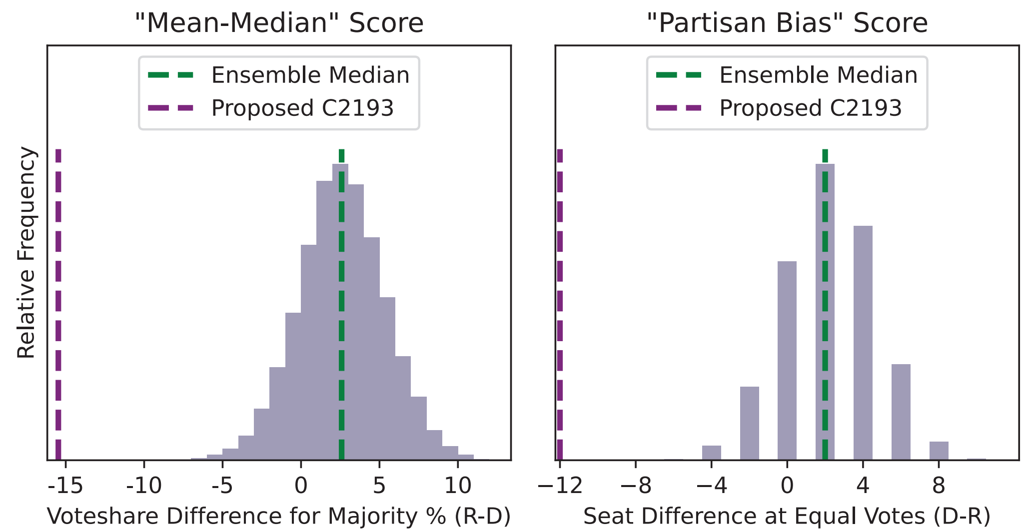


What about summary scores, like MM and PB?

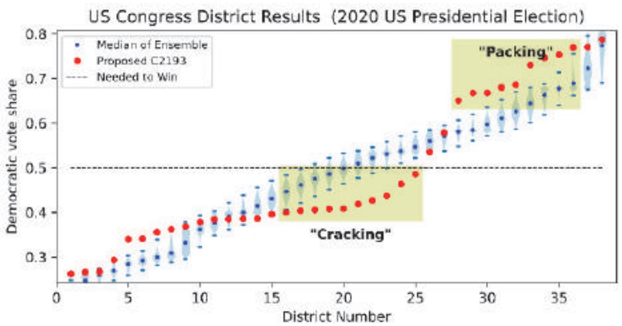


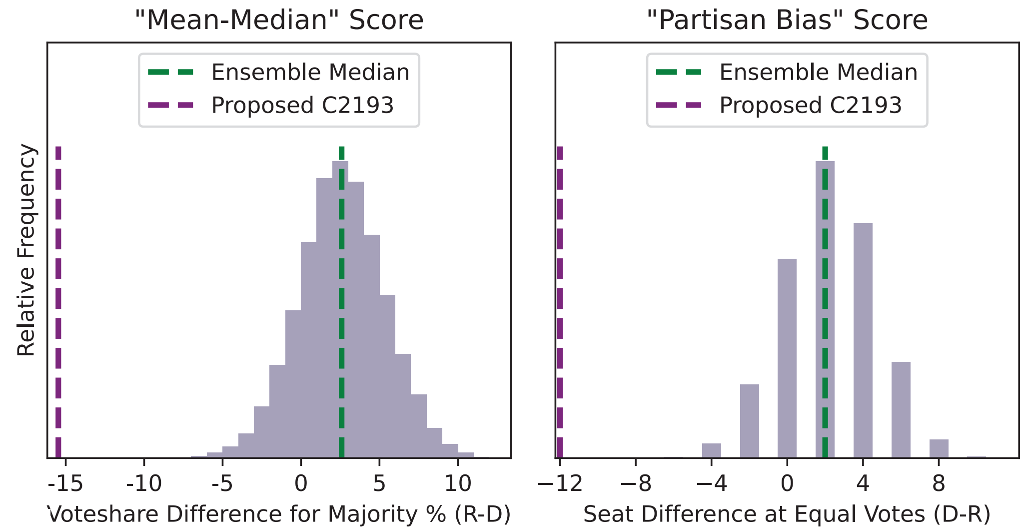
Packing and cracking results in unrepresentative outcomes

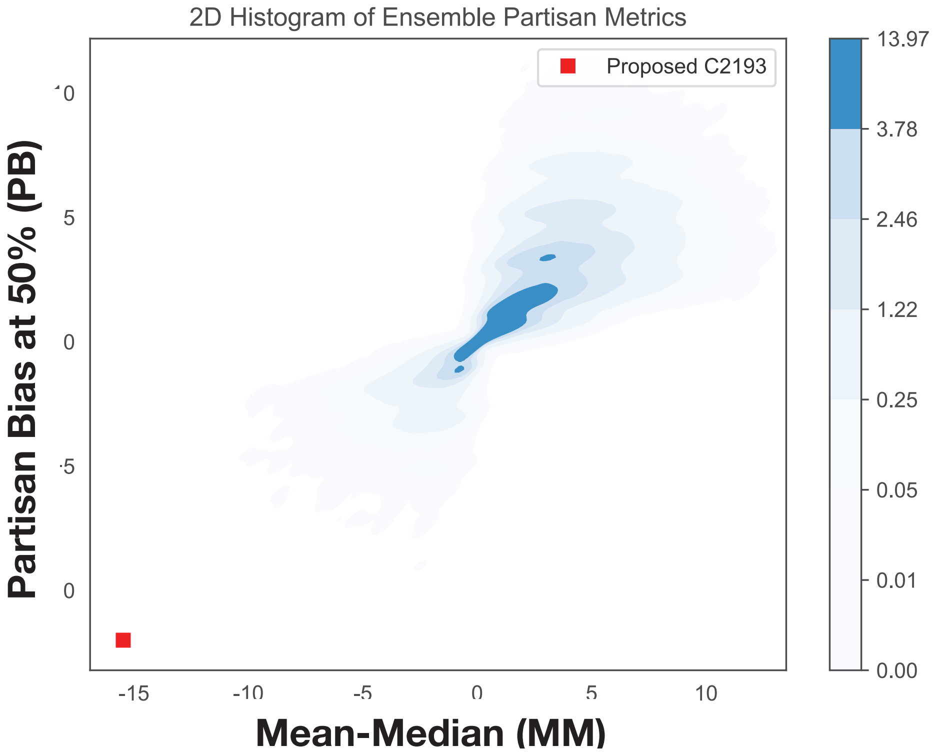




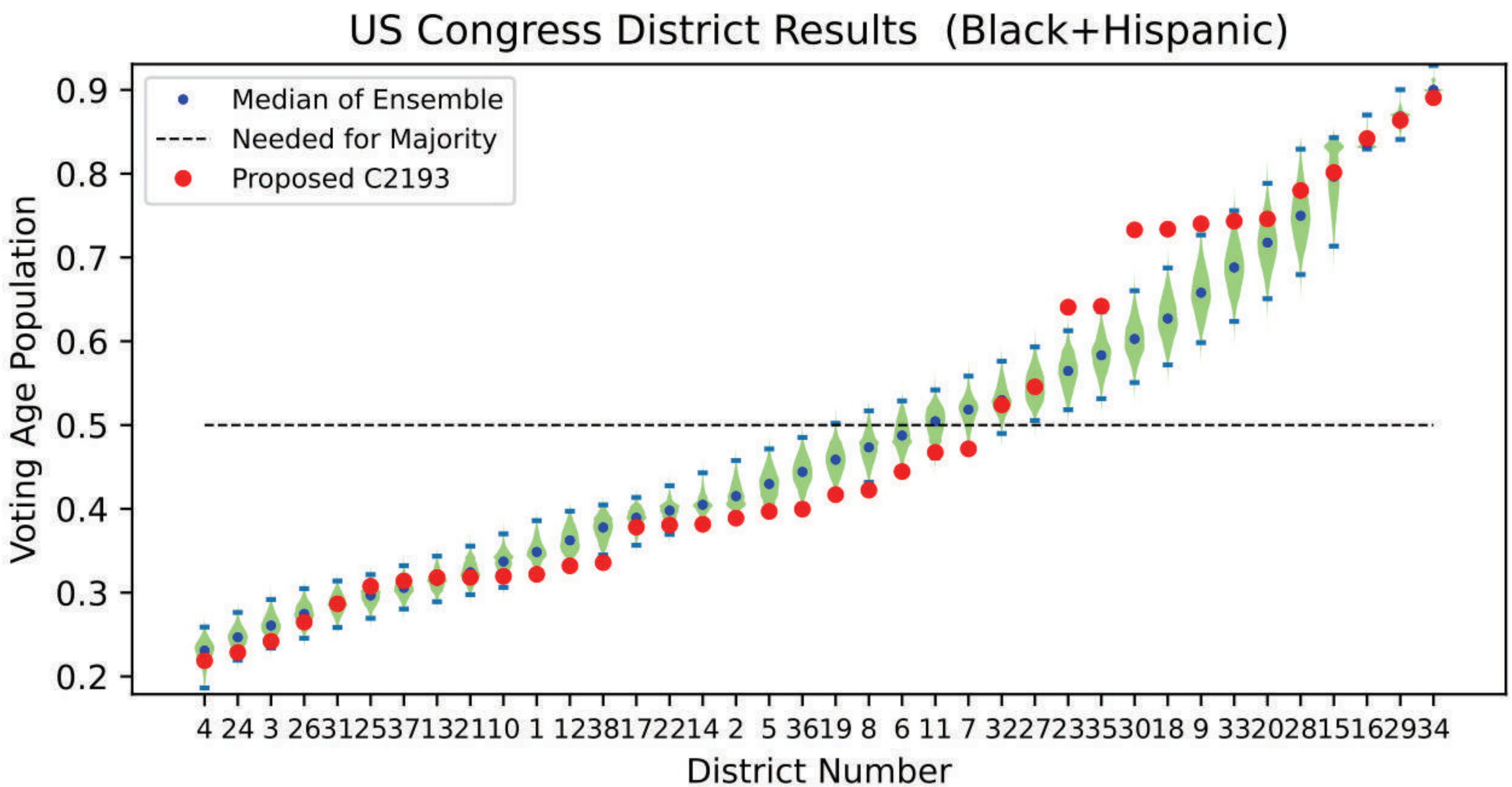
Packing and cracking results in unrepresentative outcomes



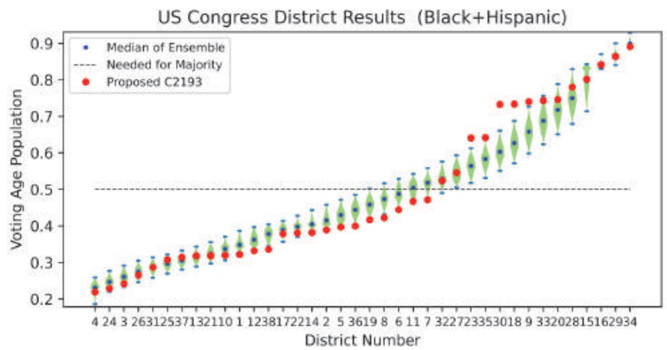




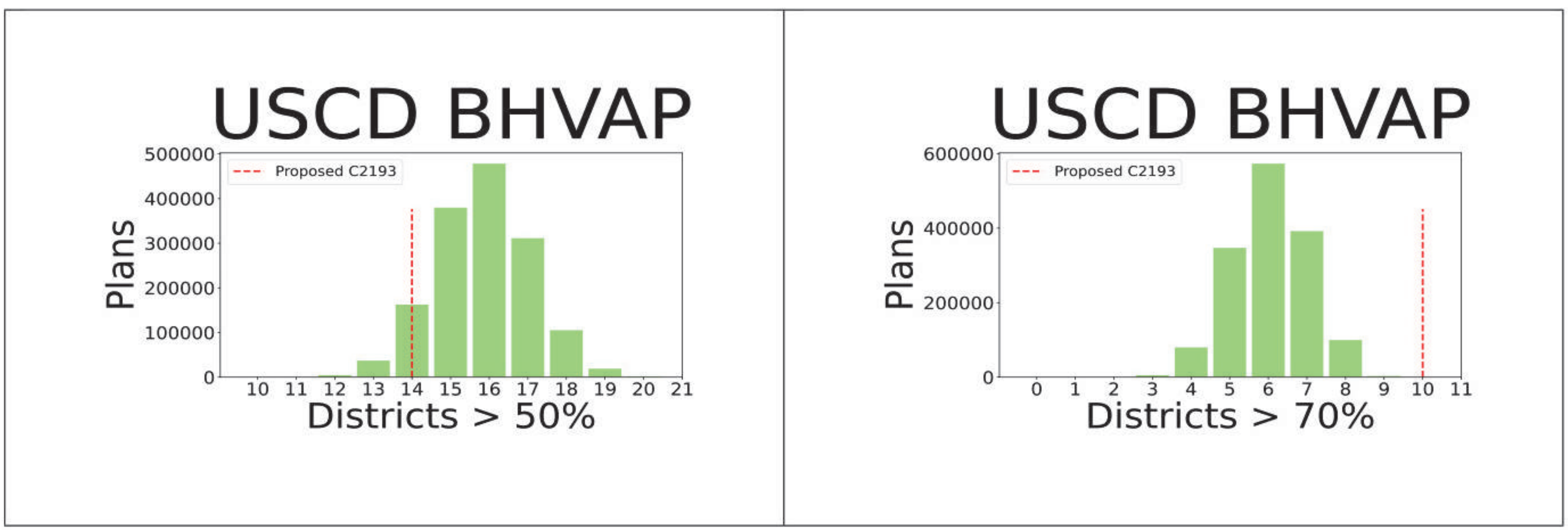
Gerrymandering leaves districts more polarized by race as well as partisanship...



…Resulting in fewer majority non-white districts than we would expect

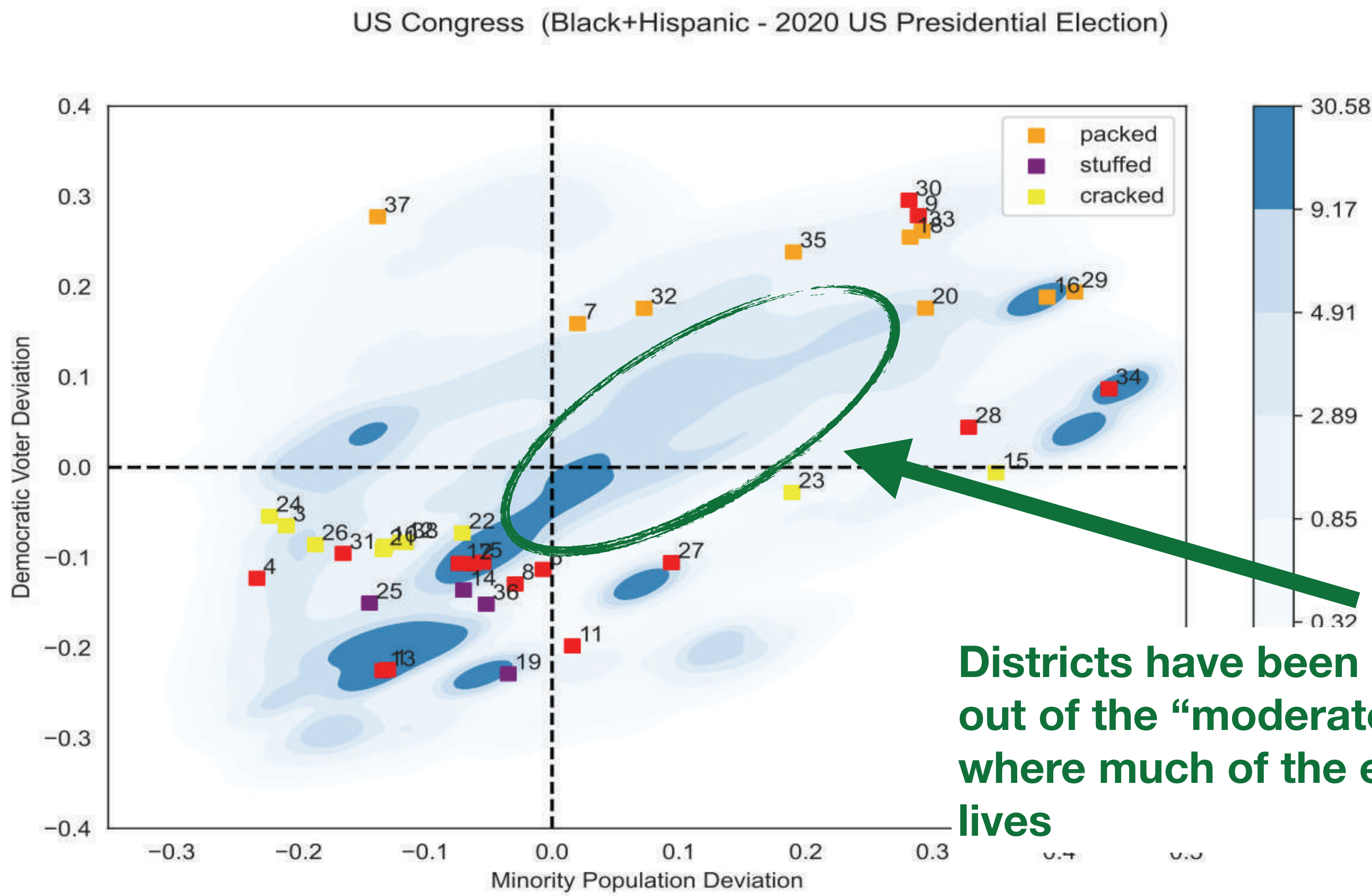


- Ensemble median is 16: actual plan has 14
- Minority voters were “packed” into a few overwhelmingly minority districts



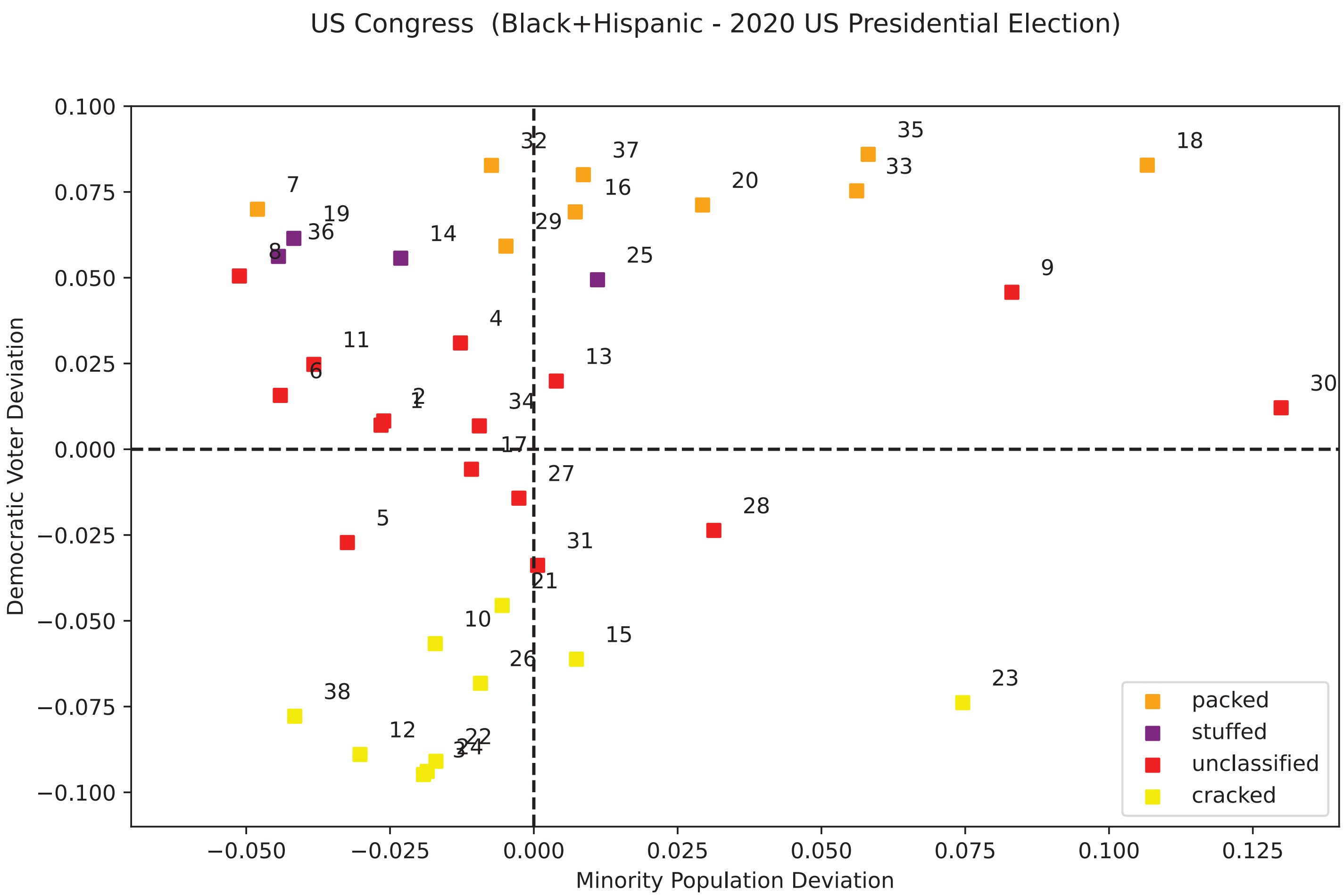
http://www.smu.edu/mumtx/txgerrywatch

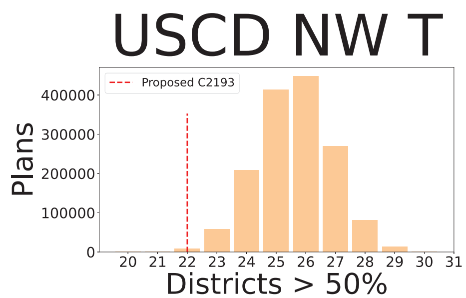
Partisan and racial disparities are *correlated*

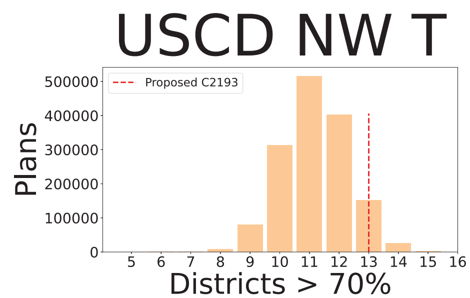


Districts have been pushed out of the “moderate middle” where much of the ensemble lives

Partisan and racial disparities are also correlated when we examine rank order deviations

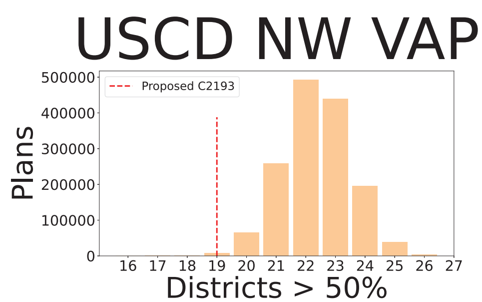


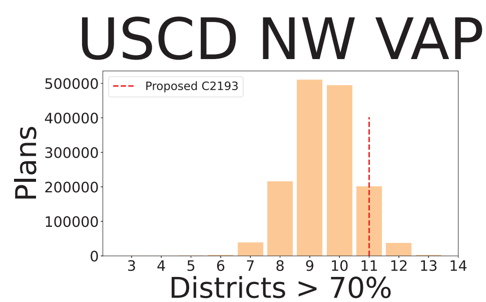




Total population:

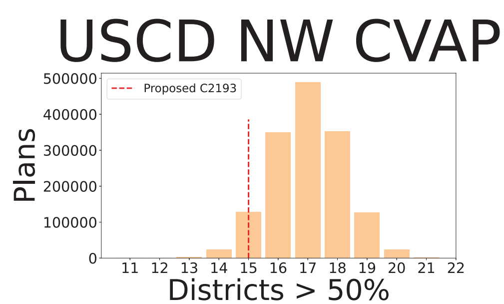
- 22 $\geq 50\,\%$ NW (vs. 26)
- 13 $\geq 70\,\%$ NW (vs. 11)

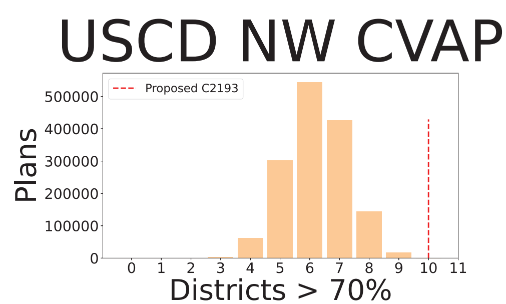




Voting age population (VAP):

- 19 $\geq 50\,\%$ NW (vs. 22)
- 11 $\geq 70\,\%$ NW (vs. 9 or 10)



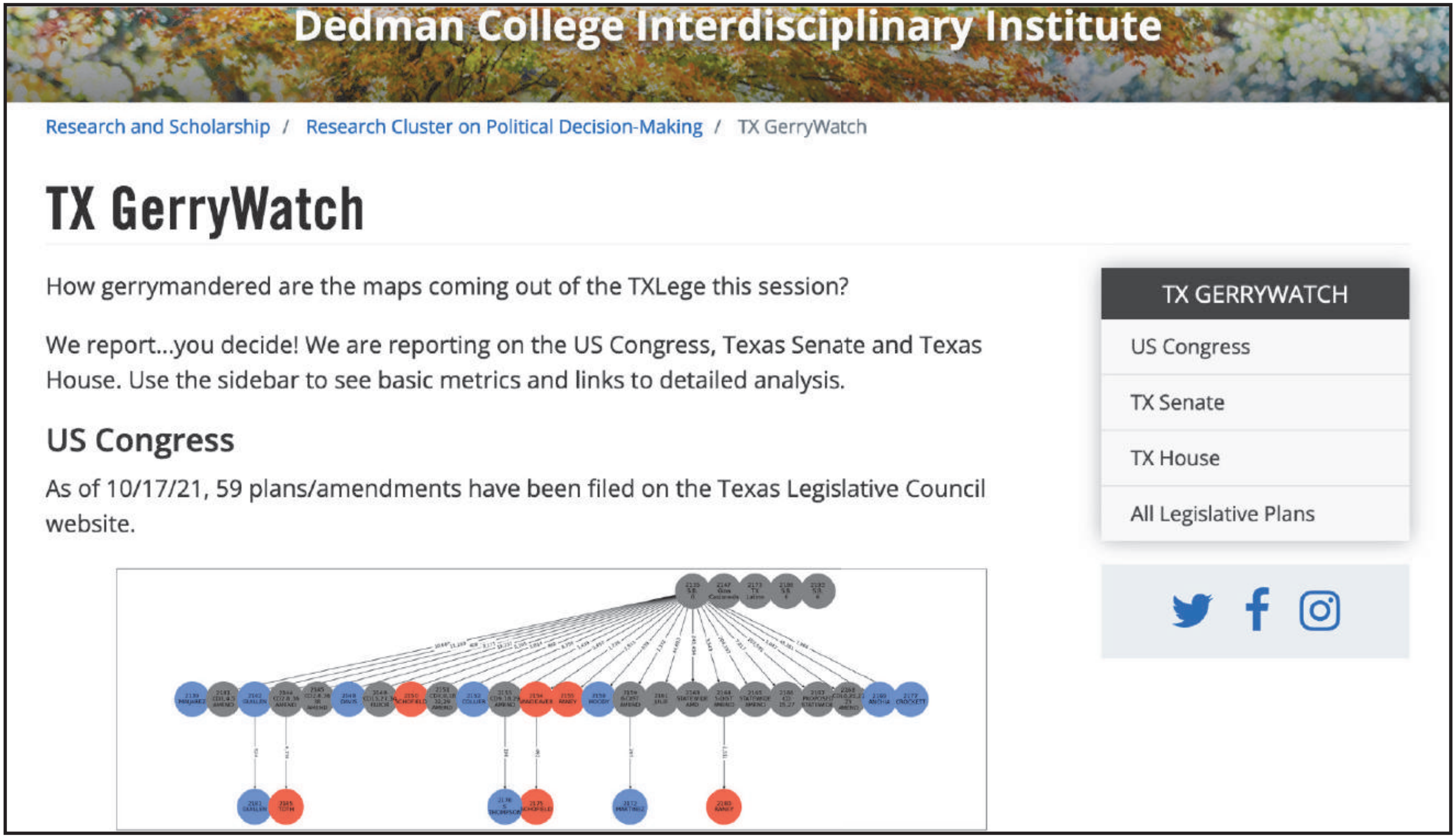


Citizen voting age population (CVAP):

- 15 $\geq 50\,\%$ NW (vs. 17)
- 10 $\geq 70\,\%$ NW (vs. 6 or 7)

http://www.smu.edu/mumtx/txgerrywatch

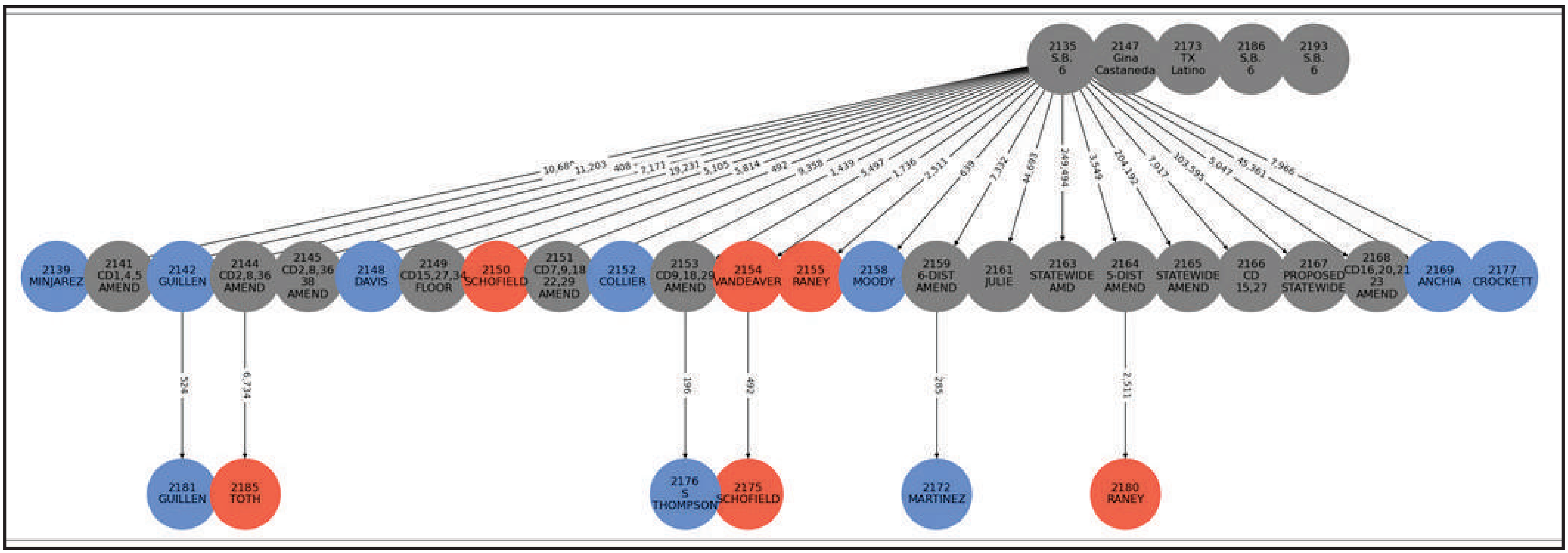
Top level: a graph of plans and relationships



http://www.smu.edu/mumtx/txgerrywatch

Top level: a graph of plans and relationships

- Circle contains ***plan name*** and ***number*** (4-digit numbers are chronological and assigned by Texas Legislative Council)
- Color indicates party of filer (if a legislator)
- Directed arrows indicate target was an amendment to the source
- Numbers on arrows are #Census blocks that changed districts
- You can find all of these plans at: dvr.capitol.texas.gov



Scroll down to find information for each plan

PLAN	Submitted By	MM	MM (%ile)	PB	PB (%ile)	Favors	2D (#of total)
Ensemble median		2.59	50%	2	46.66%	N/A	N/A
C2100 (Current map)		1.25	30.80%	4	75.03%	N	N/A
C2193	S.B. 6	-15.47	0%	-12	0%	R	0 out of 1,500,000
C2186	S.B. 6	-15.48	0%	-12	0%	R	0 out of 1,500,000
C2185	REP TOTH (R)	-15.6	0%	-10	0%	R	0 out of

- MM = mean-median score
- MM (%ile) = percentile score
- PB = partisan bias score
- PB = percentile score
- Favors = which party does it favor? (relative to the ensemble)
- ***More gerrymandered than how many plans in the ensemble?***
- Plan# (here, C2193) links to report (*in progress*)

http://www.smu.edu/mumtx/txgerrywatch

Report on C2193 (enacted US Congressional map)

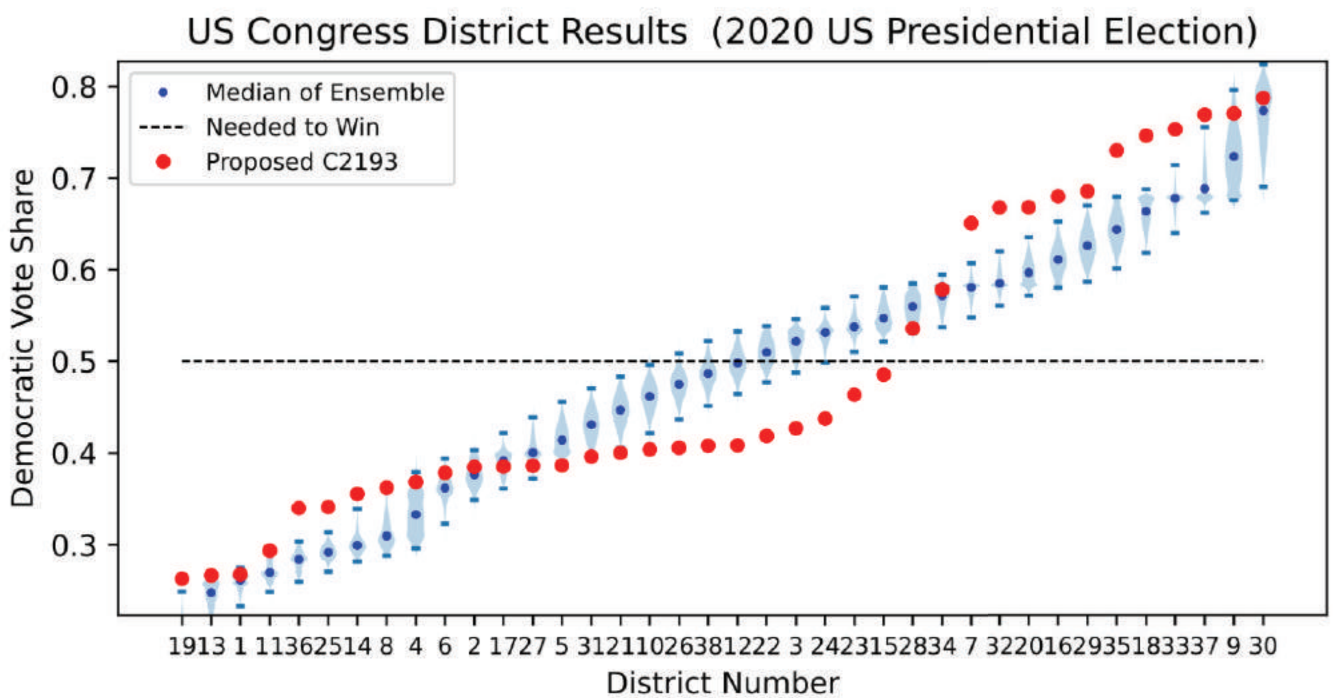


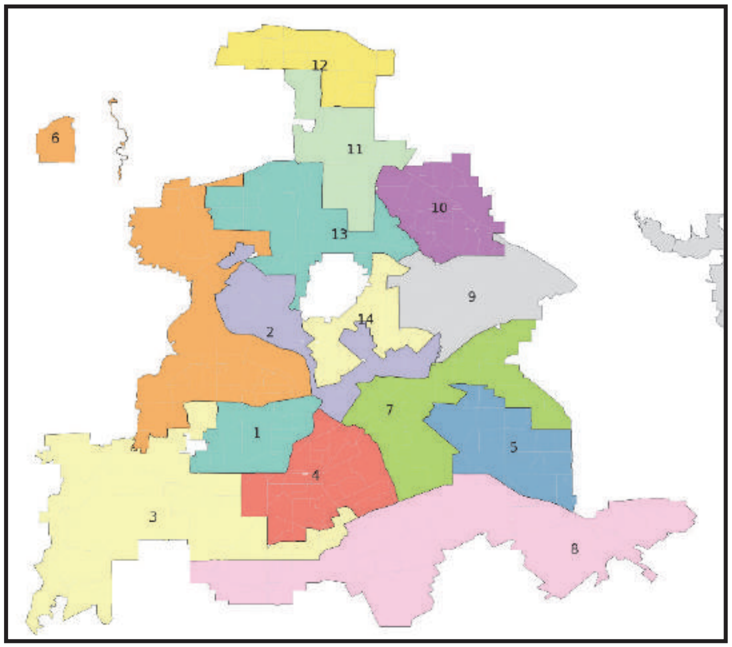
Figure 1

MUM_TX Statement on C2193

Math For Unbiased Maps TX (MUM_TX) is an interdisciplinary, nonpartisan coalition of Texas mathematicians, political scientists and philosophers working to ensure a fair and transparent redistricting process. Our research concerns the development and application of ensemble sampling techniques, and in particular their application to the current TX redistricting cycle. In brief, we use *Markov Chain Monte Carlo* techniques to generate a large number of random, legally valid maps which can then be used as an unbiased baseline to understand what a typical map should look like. Conversely, when a proposed map is

Dallas City Council

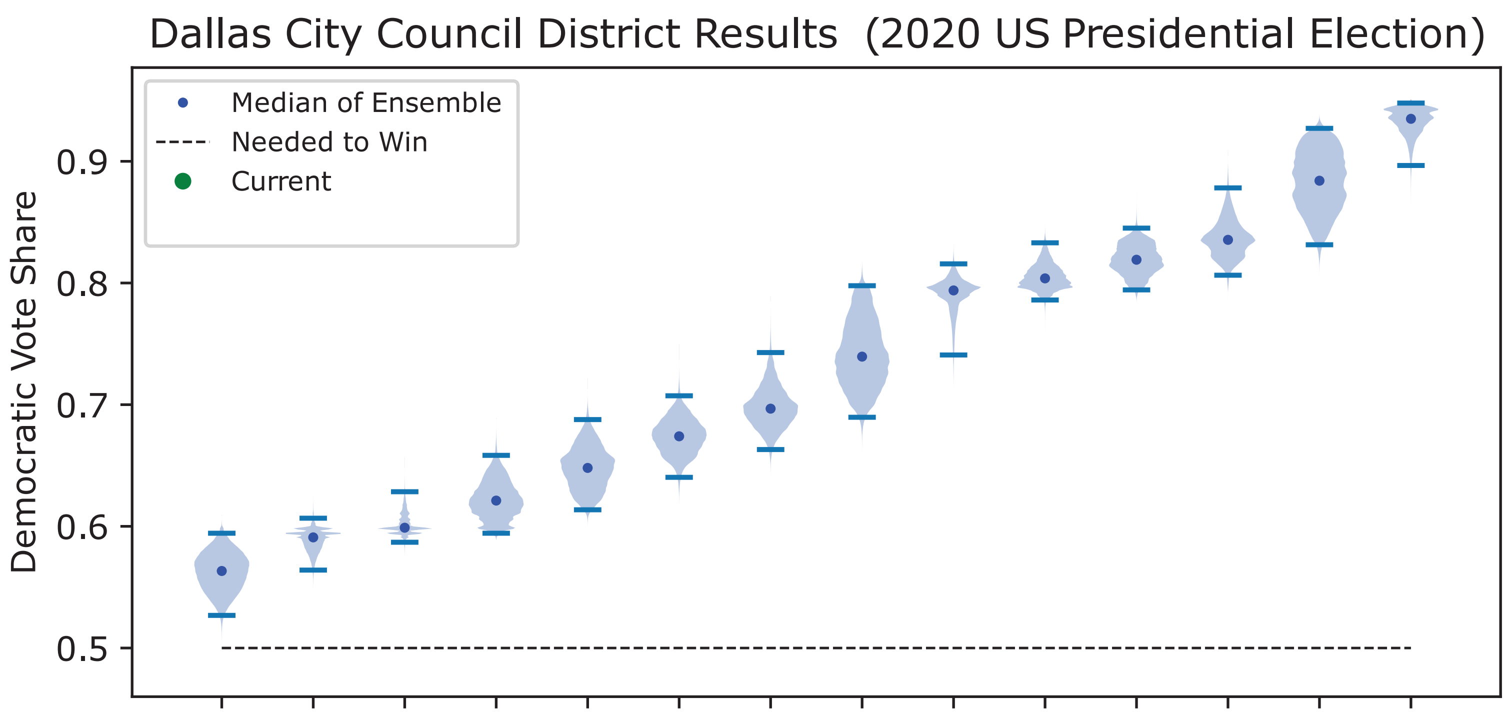
- 14 districts with nonpartisan representatives
  - They will vote on a new plan in March (?)
  - A Redistricting Commission will review and recommend plans
  - Plans are still being solicited https://districtr.org/event/cityofdallas



https://dallasredistricting.com/

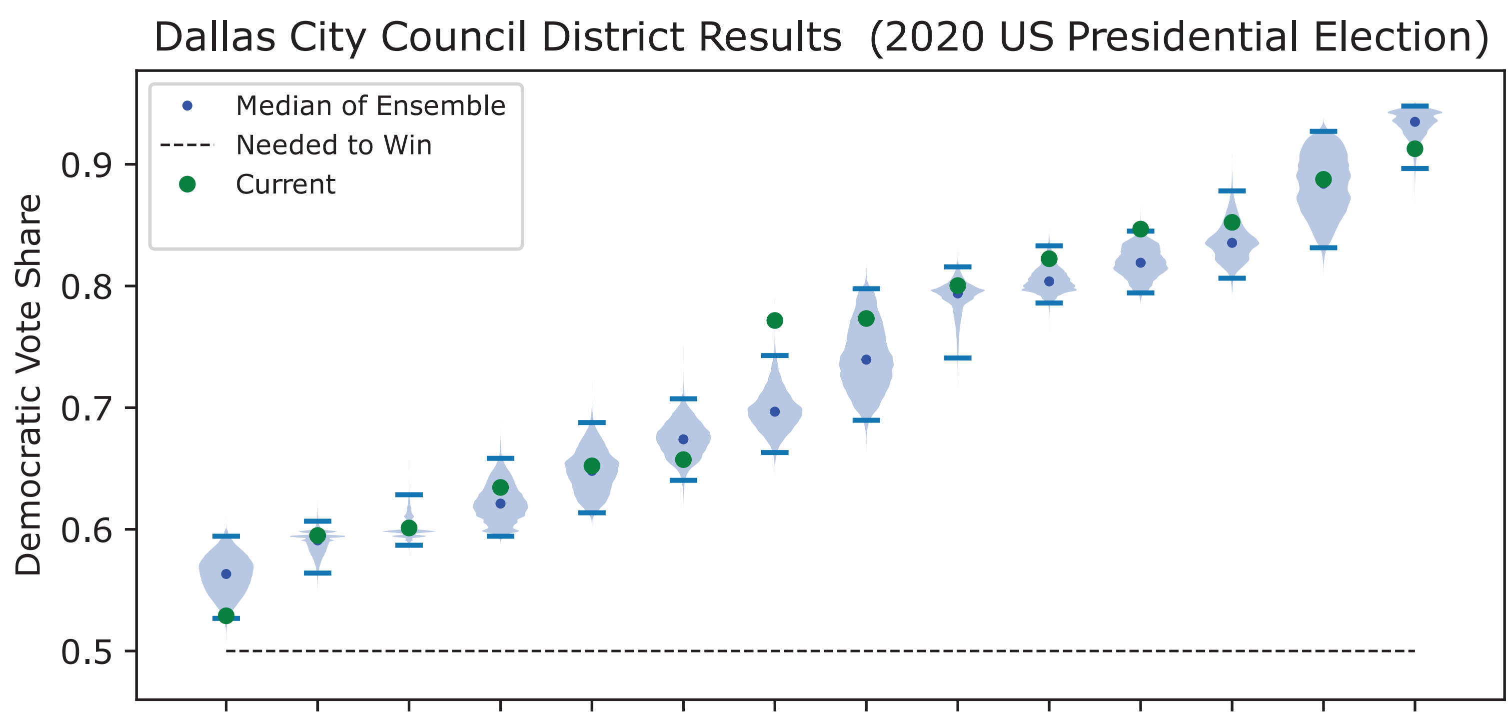
Dallas City Council

- We used Gerrychain to generate ~10 million maps
- We tracked the same statistics as for USCD
- unlike TX Congress, Dallas is strongly Democratic:



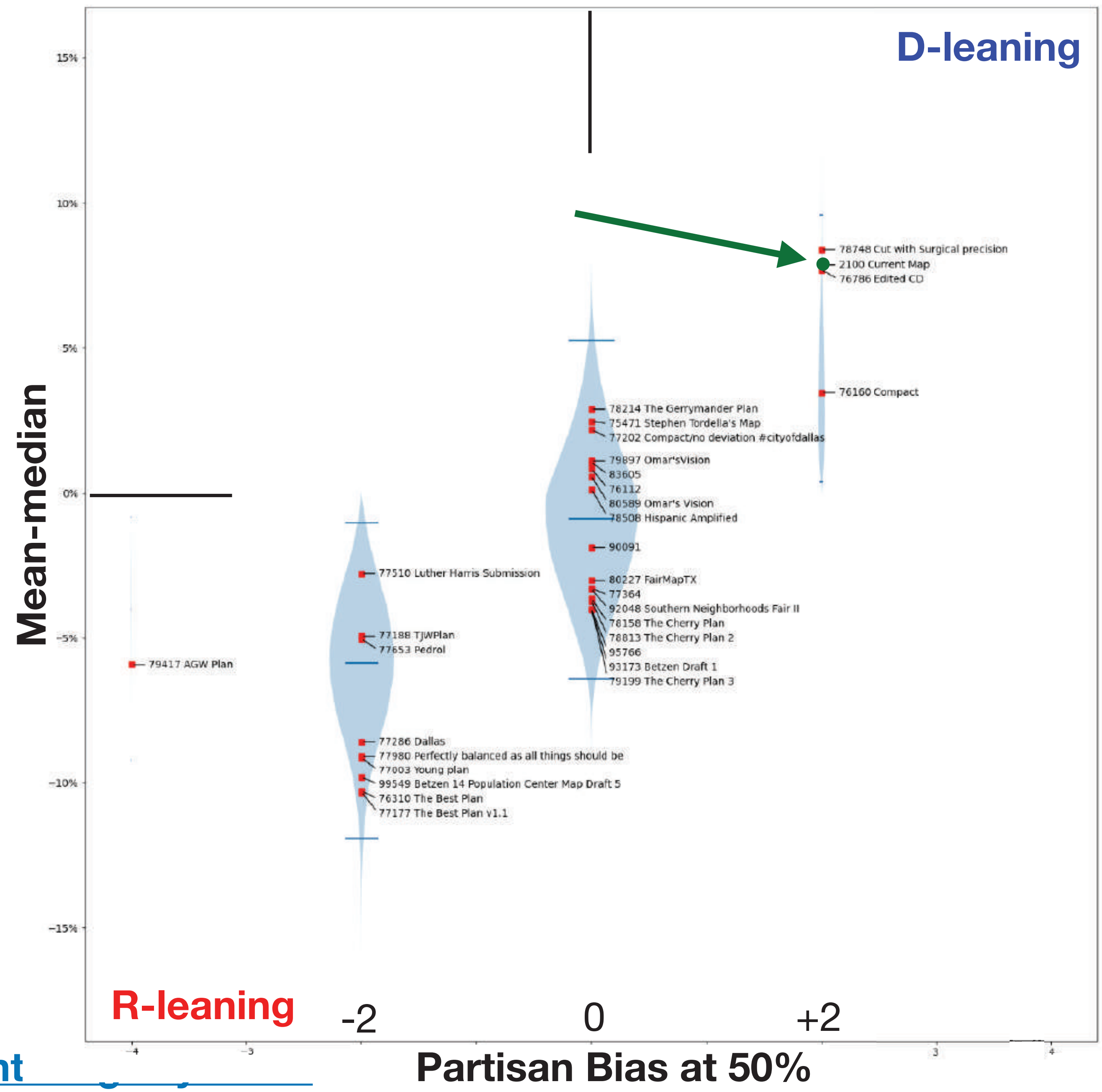
Dallas City Council

- For the current map (**green**)
  - MM = 7.9
  - PB = +2
- Is the current map gerrymandered?



Dallas City Council

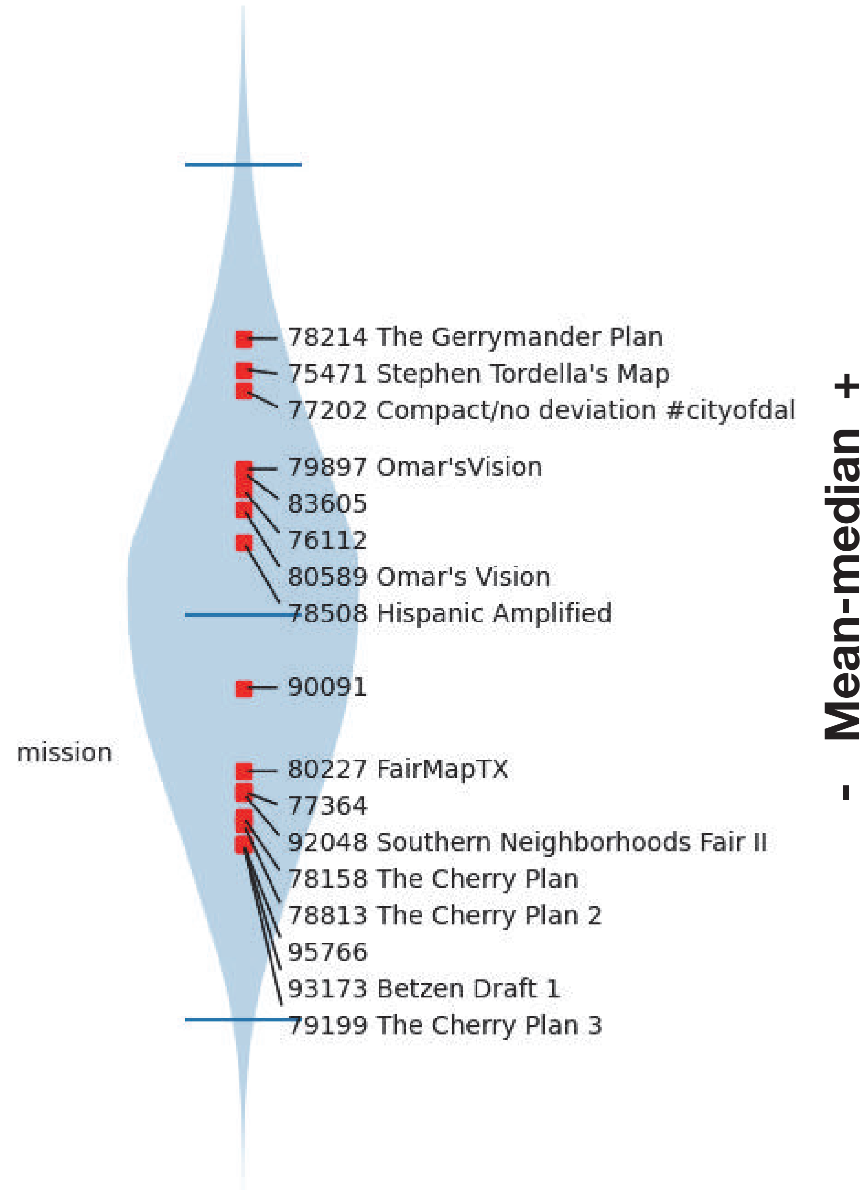
- We analyzed every plan posted to https://districtr.org/event/cityofdallas that met population constraint (max 10% deviation)
- PB only takes a small number of integer values: one “violin” for each value
- Size of violin corresponds to likelihood of PB



http://www.smu.edu/mumt

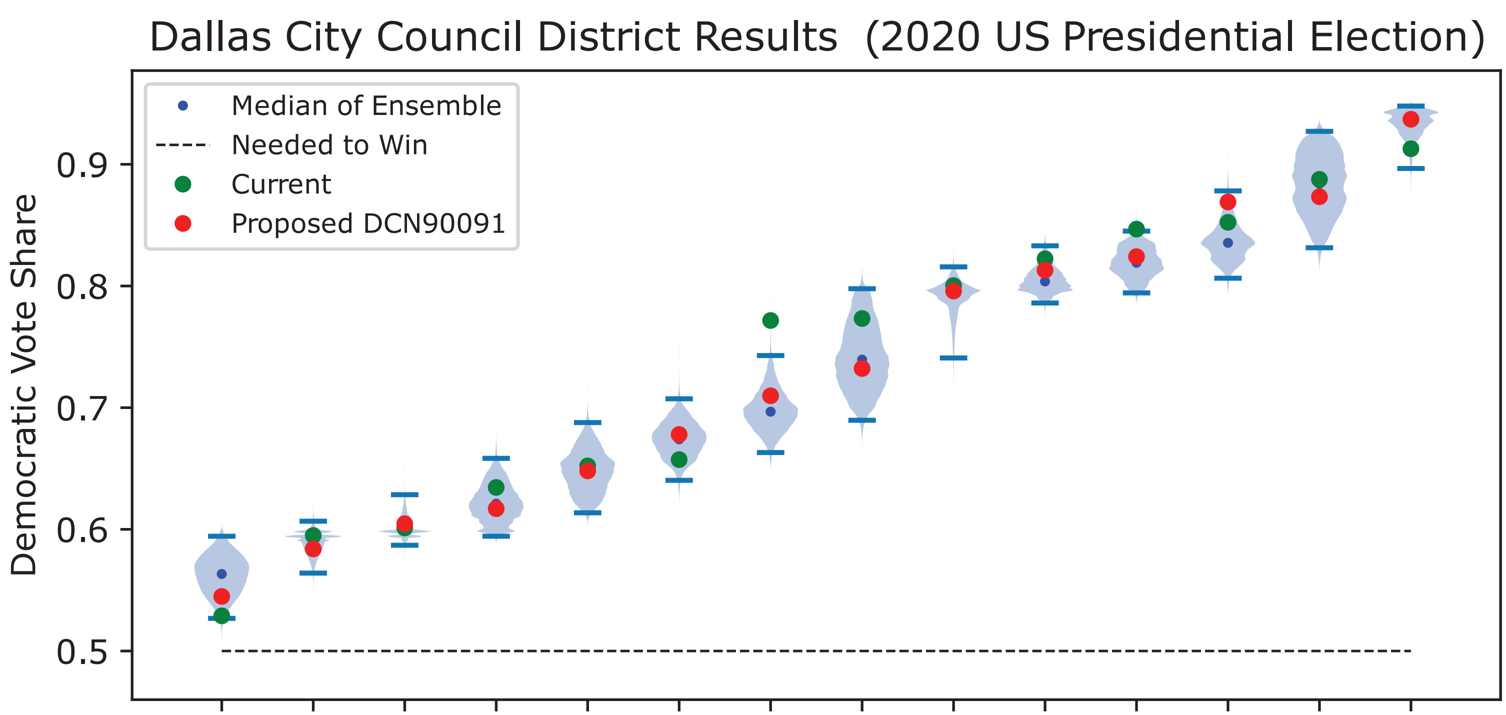
Focusing in on PB = 0

- 90091 has MM closest to the ensemble median



Dallas City Council

- For the current map (**green**)
  - MM = 7.9, PB = +2
- For 90091 (**red**)
  - MM = -1.9, PB = 0



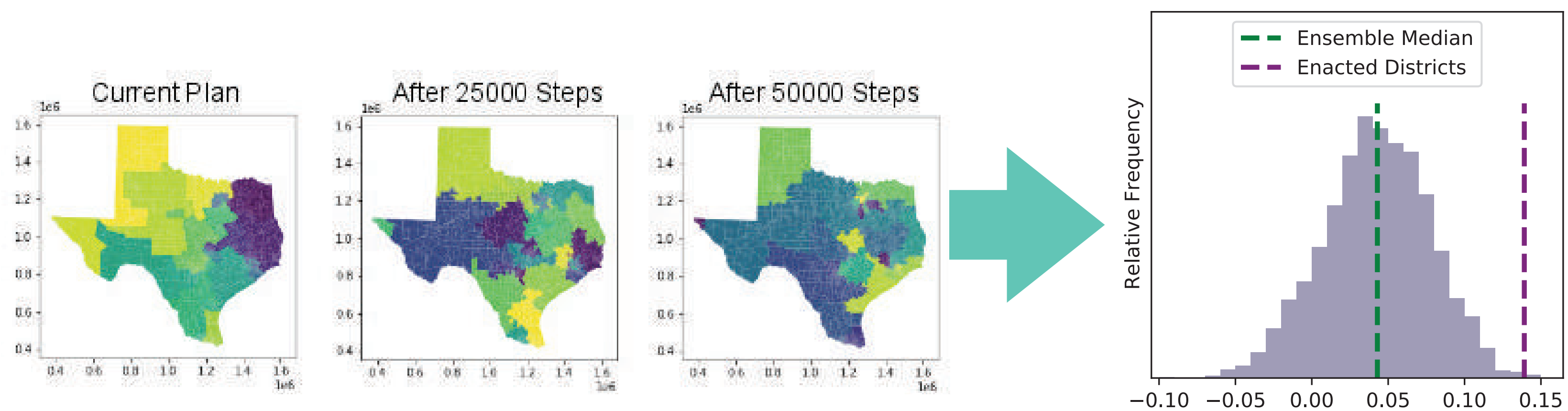
Conclusion

- We used ensemble sampling to analyze every map proposed for US Congress, TX Senate, and Texas House during the recent redistricting session.
- Our results show **the enacted maps are gerrymandered on both partisan and racial dimensions**.
- Results are documented on our website: **www.smu.edu/mumtx**
- We are currently analyzing Dallas City Council maps
- Also see:

github.com/drscook/MathVGerrmandering_CMAT_2021

github.com/scott-norris-math/GerryWrap

pypi.org/project/gerrychain



Ensemble sampling is reliable and replicable

- The algorithms are peer-reviewed and implemented in open source software (**GerryChain**)
- It has been used to develop plans in other states and as evidence in court cases
  - "Mathematicians' Brief" in Rucho vs. Common Cause, 2019 **(Right)**
  - League of Women Voters of Mich. v. Benson, 2019 (MI)
  - Ohio A. Philip Randolph Institute v. Householder, 2019 (OH)
  - League of Women Voters v. Commonwealth, 2018 (PA)
  - Common Cause v. Lewis, 2019 (NC)

Ensemble sampling is fast

- Computations for each set of results shown earlier took < 3 hours on a 2013 MacBook Pro

Nos. 18-422, 18-726

IN THE
Supreme Court of the United States

ROBERT A. RUCHO, ET AL.,
Appellants,
v.
COMMON CAUSE, ET AL.,
Appellees.

On Appeal from the United States District Court for the Middle District of North Carolina

LINDA H. LAMONE, ET AL.,
Appellants,
v.
O. JOHN BENISEK, ET AL.,
Appellees.

On Appeal from the United States District Court for the District of Maryland

AMICUS BRIEF OF MATHEMATICIANS, LAW PROFESSORS, AND STUDENTS IN SUPPORT OF APPELLEES AND AFFIRMANCE

Case Study: Redistricting in Pennsylvania

- **June 2017:** League of Women Voters challenges PA congressional map
- **November 2017:** Wes Pegden (Carnegie Mellon Univ mathematician) develops MCMC (**M**arkov **C**hain **M**onte **C**arlo) techniques that evaluate enacted map against ensemble of many alternate maps. His expert witness testimony is pivotal to the court's decision to strike down the PA map.
- **February 2018:** Moon Duchin (Tufts Univ mathematician) hired by PA Gov. Tom Wolfe to guide redistricting efforts
- **Summer 2018-present:** Mathematicians across US work to improve Pegden's MCMC techniques and make it more widely available

Summary

Mathematicians were key both to evaluating the enacted PA map AND guiding the redistricting. MCMC methods were already highly effective in 2017, and we've significantly improved them since.

- https://ballotpedia.org/League_of_Women_Voters_of_Pennsylvania_v._the_Commonwealth_of_Pennsylvania
- https://www.governor.pa.gov/newsroom/governor-wolf-enlist-non-partisan-mathematician-evaluate-fairness-redistricting-maps/
- https://www.governor.pa.gov/wp-content/uploads/2018/02/md-report.pdf

Ensemble sampling can help you draw fair maps!

References

- Data and software
  - 2020 Population data, geodata from Census Bureau
  - 2020 election results: Texas Legislative Council
  - Software from MGGG (GerryChain): https://gerrychain.readthedocs.io/en/latest/
  - Precinct-level election geodata from MGGG (pre-2020): https://github.com/mggg-states
- Legal cases: see earlier slide
- Media coverage
  - I. Lapowsy, "The Geeks Who Put a Stop to Pennsylvania's Partisan Gerrymandering", Wired, February 2018, https://www.wired.com/story/pennsylvania-partisan-gerrymandering-experts/
  - S. Roberts, https://www.technologyreview.com/2021/08/12/1031567/mathematicians-algorithms-stop-gerrymandering/
- Expert Reports
  - M. Duchin, *Outlier analysis for Pennsylvania congressional redistricting*, available at https://mggg.org/uploads/md-report.pdf
  - J. Mattingly, *Expert Report on the North Carolina State Legislature*, available at https:// sites.duke.edu/quantifyinggerrymandering/files/2019/09/Report.pdf
  - All expert reports prepared by MGGG: https://mggg.org/reports
- Academic centers
  - MGGG Redistricting Lab (Tufts): https://mggg.org
  - Quantifying Gerrymandering (Duke): https://sites.duke.edu/quantifyinggerrymandering/

Contacts

MathForUnbiasedMapsTX@list.smu.edu

https://www.smu.edu/Dedman/Research/Institutes-and-Centers/DCII/Scholarship/Research-Cluster-on-Political-Decision-Making