Applications of Ensemble Analysis to Gerrymandering in Minnesota and Texas

Eva Airoldi, Oliver Calder, Tom Patterson, Antonia Ritter, Bekka Stein

Fall 2021 & Winter 2022

Outline

I. Introduction to Redistricting and Gerrymandering

II. Legal Requirements for Redistricting Plans

III. Ensemble Analysis with Markov Chain Monte Carlo

IV. Data and Methodology

V. Partisan Metrics for Gerrymandering Across our Ensemble

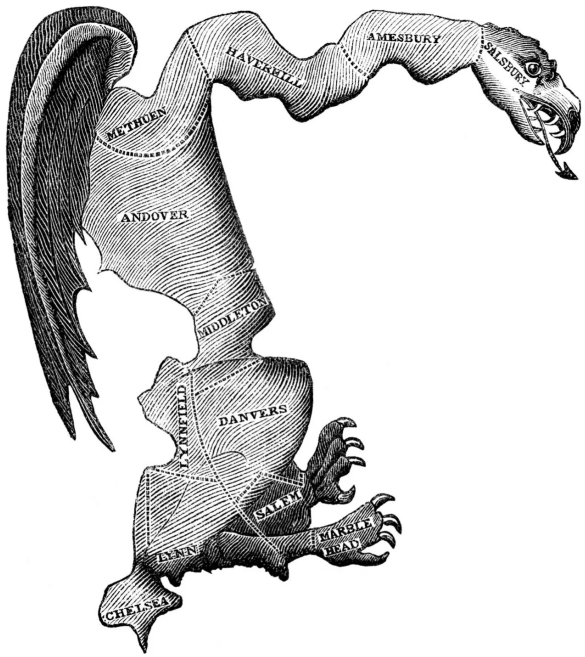
VI. State-Level Conclusions for US House redistricting in MN and TX

What is Gerrymandering?

Every 10 years, every state's congressional district map is re-drawn based on the updated population count from the census.

Definition

Gerrymandering is the practice of drawing electoral districts to benefit one group (political party or racial/ethnic group). [**?**]



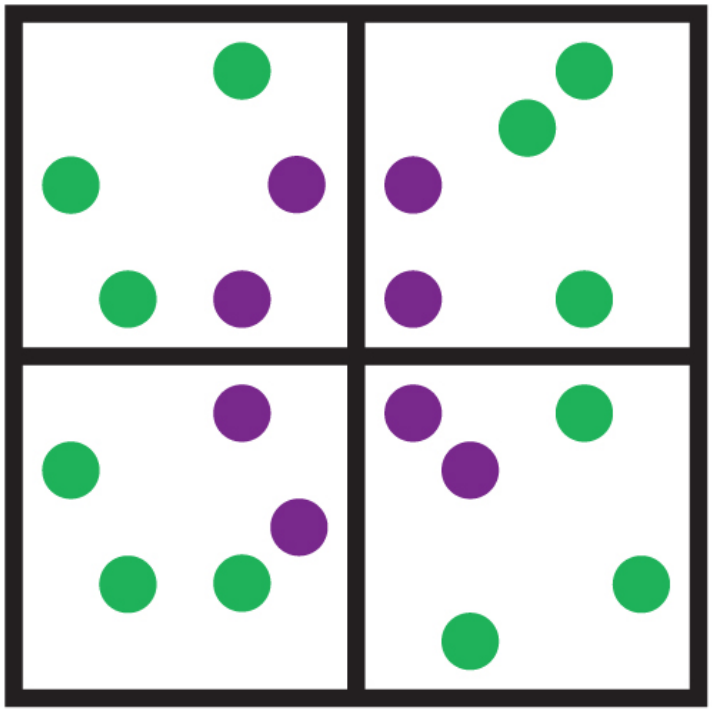
The Big Picture

Through gerrymandering is an issue in local redistricting, we only examine U.S. Congressional districting maps today

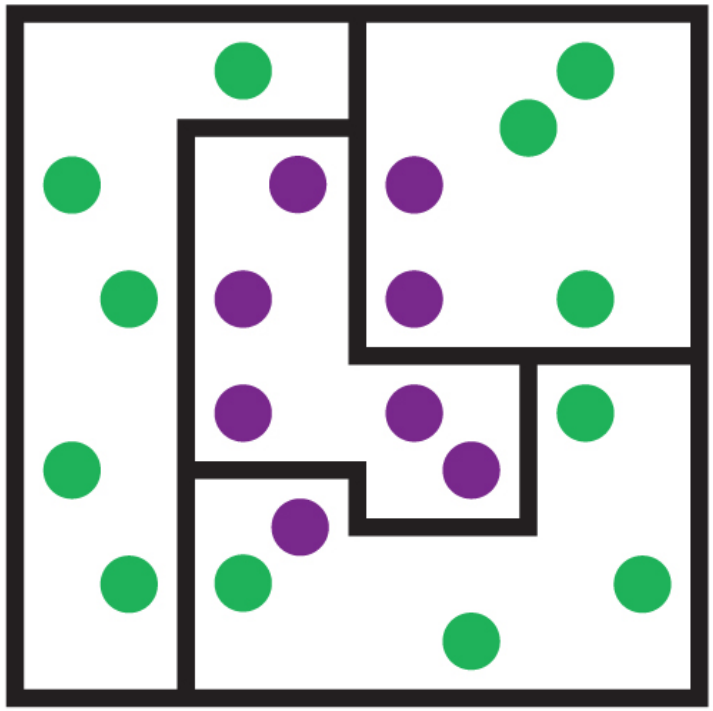
- How can we detect gerrymandering?
- Think about the space of all possible, legal districting maps (it's big)
- We sample from that space – and call the sample our **ensemble**
- If the real map is an outlier in that sample, that's an indication that someone worked to make the map abnormal

Cracking and Packing

Cracking Communities so they can't elect their own representative.



Packing Partisans into one district so the other party wins adjacent districts.



[?]

Redistricting Today

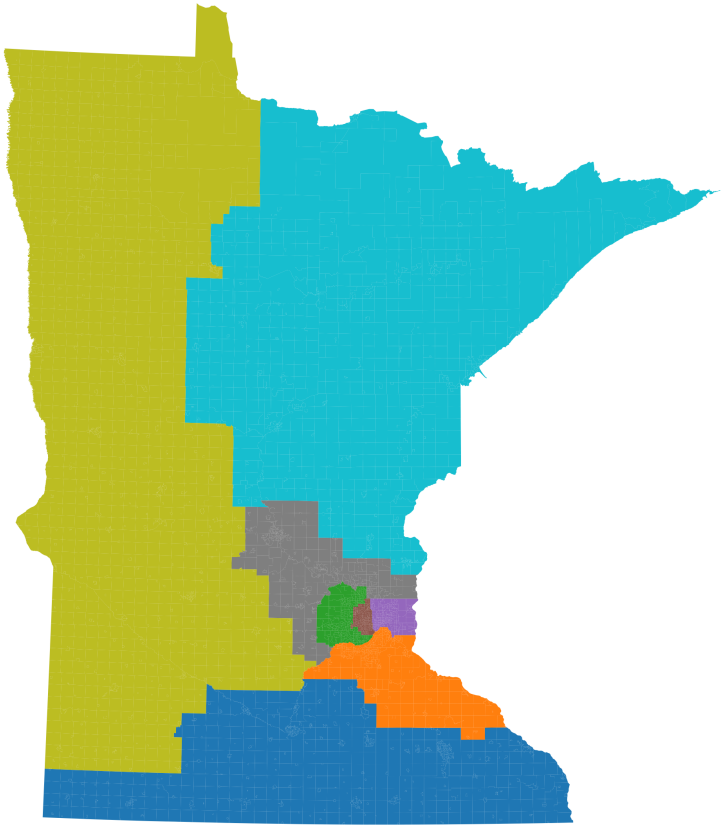
Legal Requirements for Districts:

- Equal population
- As compact as possible
- Does not split existing political boundaries (typically counties)
- Complies with the Voting Rights Act (VRA)

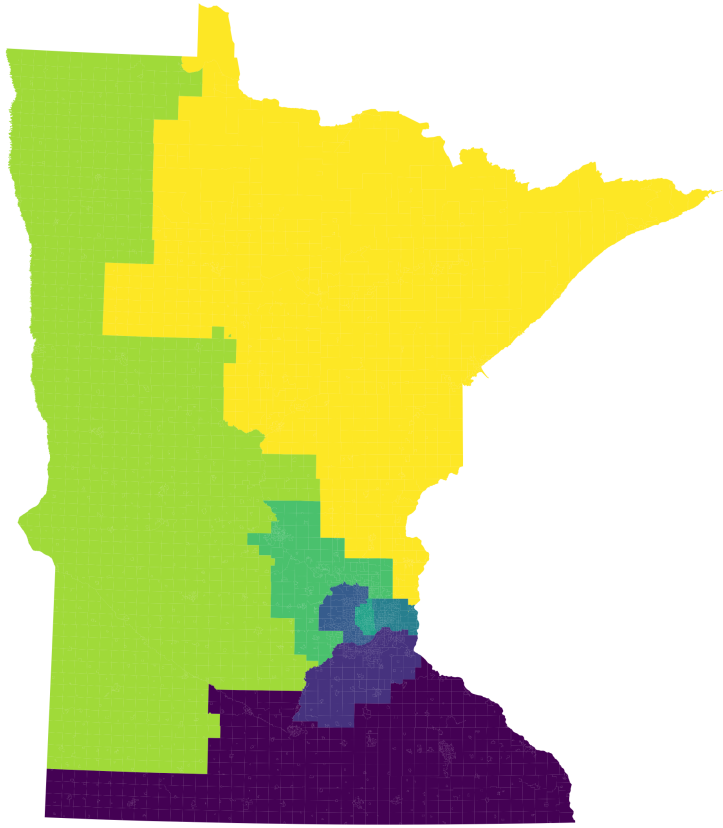
Redistricting Today: Minnesota

- Minnesota legislature proposes maps in the form of bills, they must pass by simple majority
- If the legislature can't pass a bill, the court decides on the new map
- Currently four court-proposed plans that we will analyze
- **New Today:** Minnesota 2020 map released

New MN Map!



Old



New

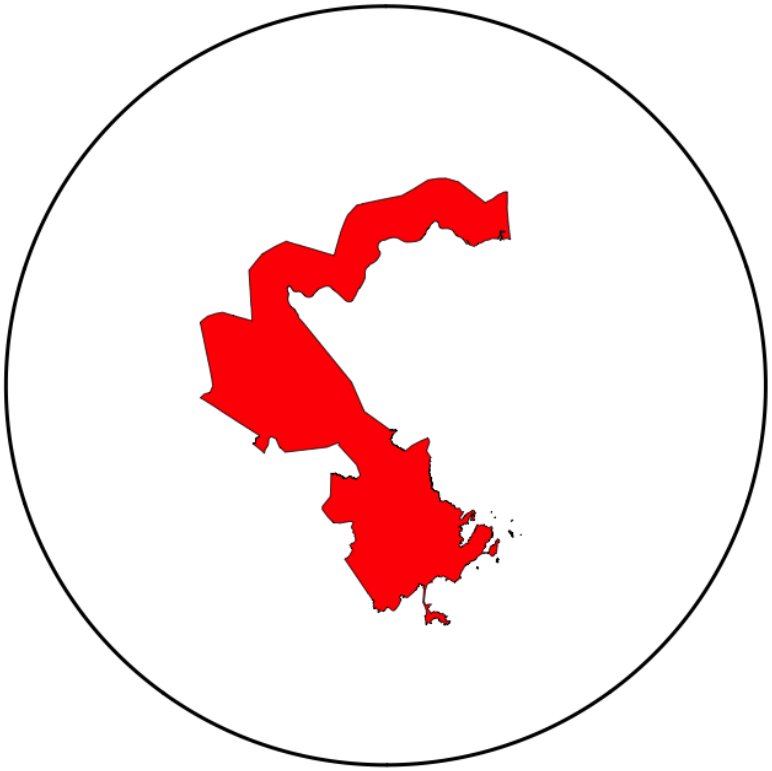
Redistricting Today: Texas

- Texas legislature proposes maps in the form of bills, they must pass by simple majority
- Texas is able to do this successfully
- Texas has already accepted a 2020 map

Compactness: Polsby-Popper

Polsby-Popper is one of many compactness measures. For a district d,

$$\text{PP}(d) = \frac{4\pi A_d}{P_d^2}$$



= ratio of the area of the district to the area of a circle whose circumference is equal to the perimeter of the district.

Motivation for Ensemble Analysis

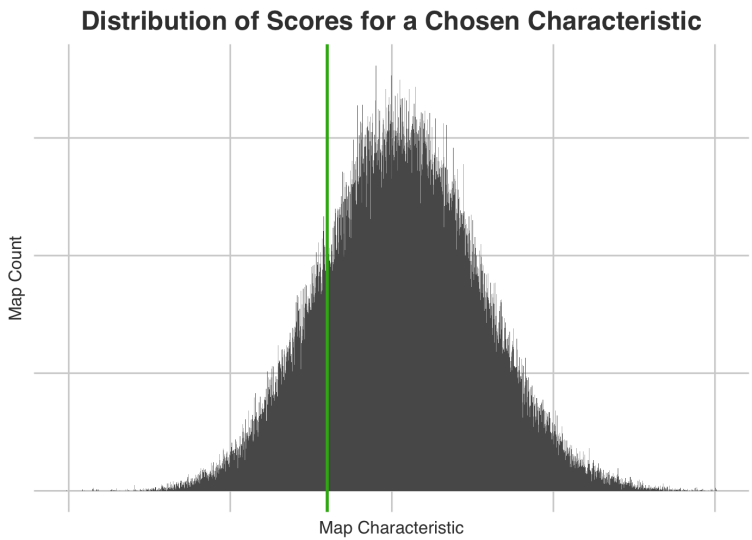
"Gerrymandering is a fundamentally multidimensional problem, so it is manifestly impossible to convert that into a single number without a loss of information that is bound to produce many false positives or false negatives for gerrymandering."

—Moon Duchin

Ensemble Analysis

Definition

An **ensemble** for a given state is a set of random, legal maps which is representative of the set of all possible legal maps in this state.



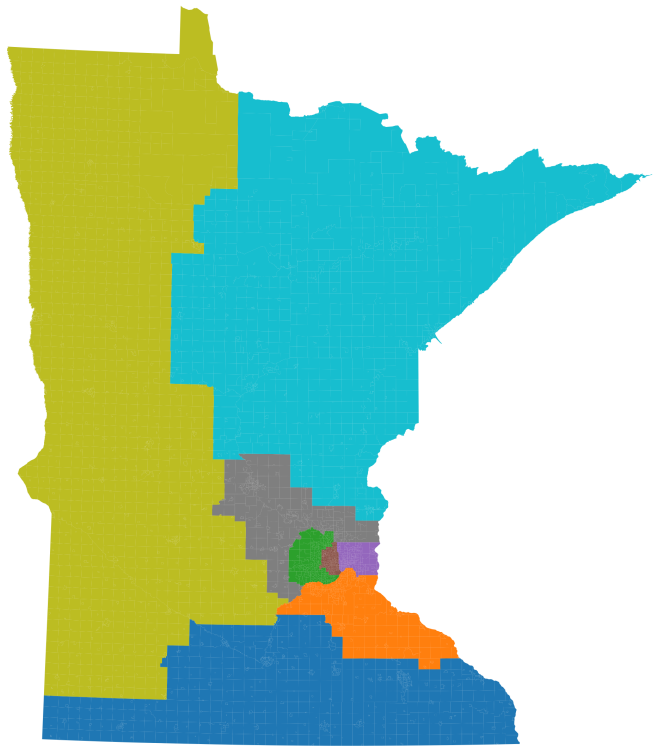
Once an ensemble exists, we can compare characteristics (such as efficiency gap, partisan bias, responsiveness, etc.) of a proposed map to the distribution of values for maps in the ensemble.

Creating the Ensemble

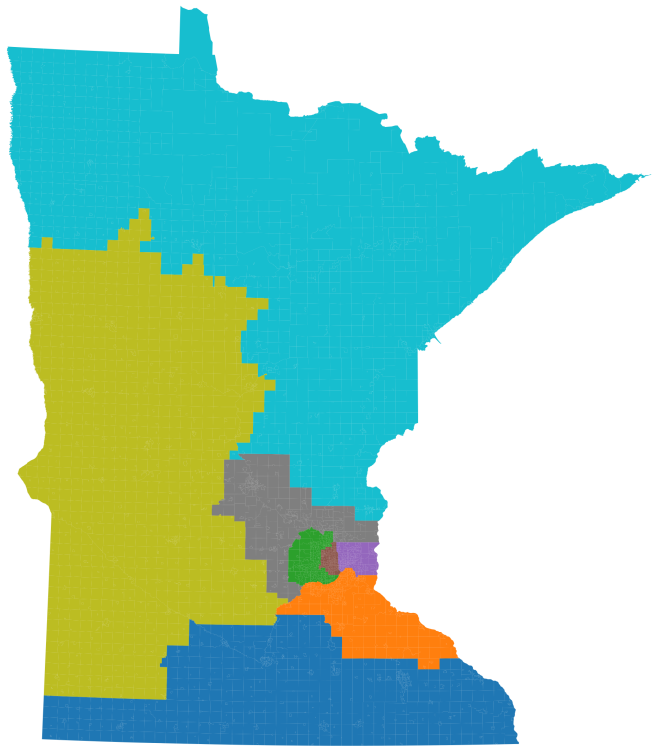
- We create our ensemble using Markov Chain Monte Carlo methods
- Our data comes from the Metric Geometry and Gerrymandering Group (MGGG) Redistricting Lab at Tisch College of Tufts University
- We also used the **Gerrychain** Python library from MGGG to create random maps from this data

Recombination Algorithm: Making Maps from Other Maps

- Each chain starts with a **seed map**; we used the existing (2010) congressional district maps
- Generate the next map in the chain using a **recombination** algorithm

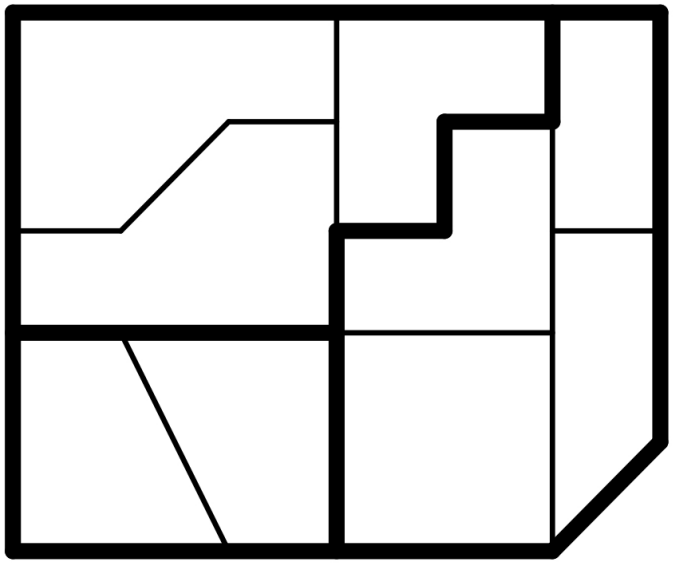


Seed map = map_1

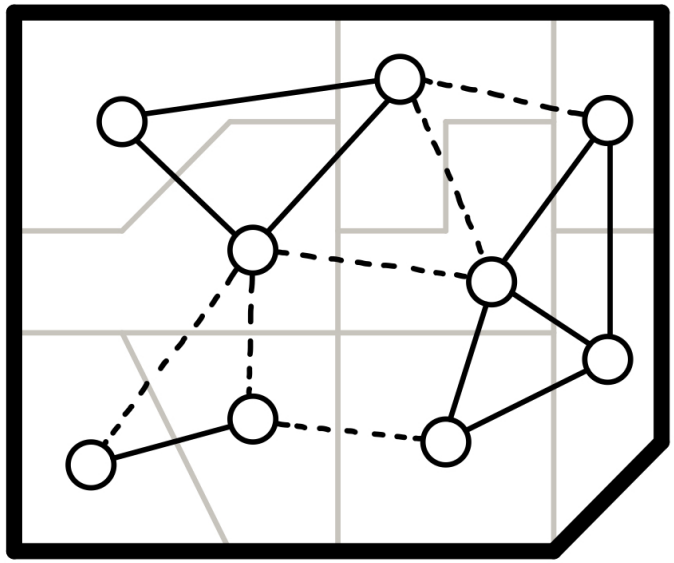


map_2

Dual Graph Representation



This is an imaginary state with 9 precincts and 3 districts



Each precinct is represented by a node, and the nodes are connected if the precincts are adjacent

Markov Chain Monte Carlo Summary

Seed map $=$ map_1

$\downarrow$

map_2

$\downarrow$

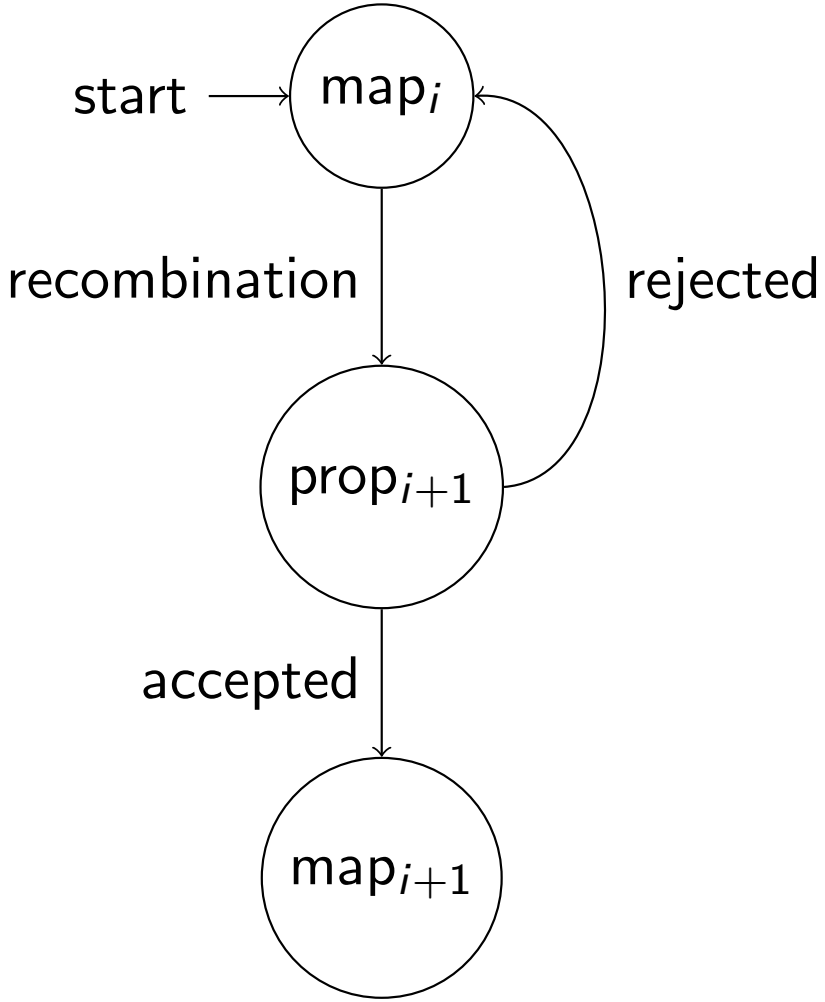
map_3

$\downarrow$

$\vdots$

$\downarrow$

$\text{map}_{150,000}$



Data

Our datasets for MN and TX were cleaned and aggregated by MGGG

- Each state has a shapefile, a format for storing geographic information
- Each shapefile is divided into precincts
- Each precinct contains...
  - Population
  - Demographic information
  - Election results

Score Function

Once we have created a new map using recombination, we have to decide whether it is legal.

For any map ε, we compute the following score for how well this map adheres to legal requirements:

$$J(\varepsilon) = w_P J_P(\varepsilon) + w_C J_C(\varepsilon) + w_S J_S(\varepsilon)$$

where $w_P, w_C, w_S \in \mathbb{R}^+$ are *weights* to each of the individual scores, so each contribute similarly to the total score. [**?**]

Score Function: Population Score

Total Score Function

$$J(\varepsilon) = w_p J_p(\varepsilon) + w_C J_c(\varepsilon) + w_s J_s$$

The population score term is

$$J_P(\varepsilon) = \sqrt{\sum_{i=1}^{n} \left(\frac{\text{pop}(D_i(\varepsilon))}{\text{pop}_{\text{ideal}}} - 1 \right)^2}$$

where

$$\text{pop}_{\text{ideal}} = \sum_{i=1}^{n} \frac{\text{pop}(D_i(\varepsilon))}{n},$$

$D_i(\varepsilon)$ is the ith district in districting plan ε, and ε contains n districts.

Score Function: Compactness

Total Score Function

$$J(\varepsilon) = w_p J_p(\varepsilon) + w_C J_c(\varepsilon) + w_s J_s$$

The compactness score term is

$$J_C(\varepsilon) = \sum_{i=1}^{n} \frac{\text{perimeter}(D_i(\varepsilon))^2}{\text{area}(D_i(\varepsilon))}$$

where $D_i(\varepsilon)$ is the ith district in districting plan ε, and ε contains n districts.

This score is inversely proportional to Polsby-Popper, which was discussed earlier.

Score Function: Split Counties

Total Score Function

$$J(\varepsilon) = w_p J_p(\varepsilon) + w_C J_c(\varepsilon) + w_s J_s$$

The county split score term is

$$J_S(\varepsilon) = \sum_{i=2}^{n} C^{i-2} \cdot |S_i(\varepsilon)| \cdot W_i(\varepsilon)$$

where

$$\begin{aligned} S_i(\varepsilon) &= \{\text{counties in } \varepsilon \text{ split between} \geq i \text{ districts}\}, \\ C &= \text{some large constant}\,(>1), \text{ and} \\ W_i &= \sum_{s \in S_i} \sqrt{1 - F_{i-1}(s)}. \end{aligned}$$

$F_i(s)$ denote the fraction of precincts in county s which fall within the i most dominant districts in the county. As before, n is the number of districts in ε. We used and $C = 2$ as the constant.

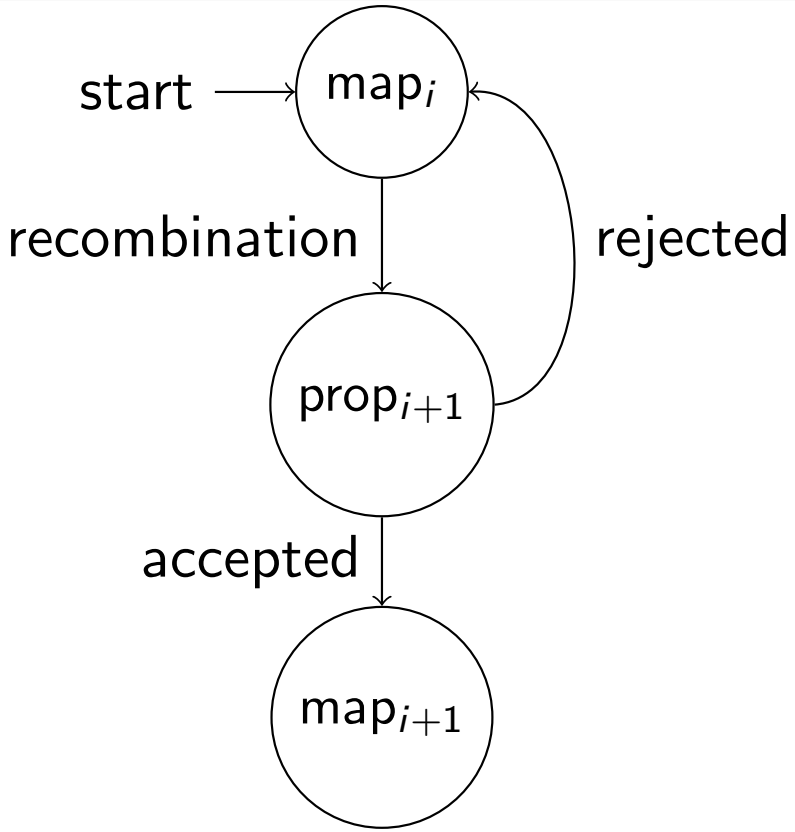
Score Function: VRA

- MGGG linked census data to voting data to determine whether minority groups could effectively elect preferred candidates
- Mattingly et al. (2020) targeted specific demographic distributions within districts
- Shelby County v. Holder (2013) made it significantly harder to challenge redistricting plans on the basis of VRA violations

Acceptance Function

Definition

An **Acceptance Function** is a function which maps from the set of possible maps to $[0, 1]$ and gives a probability for how likely it is that map will be chosen in an MCMC algorithm.



Acceptance Function Discussion

Acceptance Function Formula

Given a proposed child map ε' from a parent map ε, let $P(\varepsilon'|\varepsilon)$ be the probability of accepting child ε' from parent ε, defined as follows: [**?**]

$$P(\varepsilon'|\varepsilon) = \min\left(1, \frac{Q(\varepsilon', \varepsilon)}{Q(\varepsilon, \varepsilon')} e^{-\beta\left(\frac{J(\varepsilon')-J(\varepsilon)}{J(\varepsilon)}\right)}\right)$$

- $\frac{Q(\varepsilon',\varepsilon)}{Q(\varepsilon,\varepsilon')}$ is a term that describes how complicated the child map ε' is compared to the parent map ε
  - $\frac{Q(\varepsilon',\varepsilon)}{Q(\varepsilon,\varepsilon')} > 1 \implies \varepsilon'$ is more complicated than ε
- β is the simulated annealing factor which allows us to better explore the whole space of legal maps
  - β starts at 0 and grows to 23.5

Simulated Annealing in Acceptance Function

Acceptance Function Formula

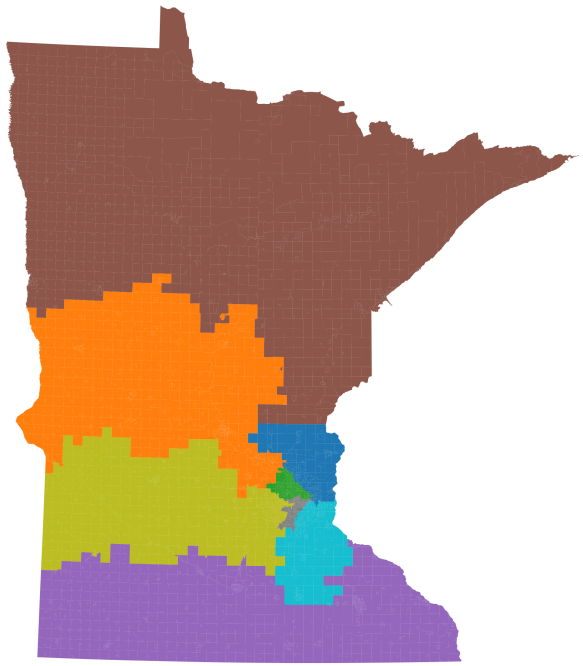
$$P(\varepsilon'|\varepsilon) = \min\left(1, \frac{Q(\varepsilon',\varepsilon)}{Q(\varepsilon,\varepsilon')} e^{-\beta\left(\frac{J(\varepsilon')-J(\varepsilon)}{J(\varepsilon)}\right)}\right)$$

- First 10,000 steps: score function is ignored ($\beta = 0$).
- Between 10,000 and 70,000 steps, β grows linearly from 0 to 23.5.
- After 70,000 steps: $\beta = 23.5$
  - If child score ($J(\varepsilon')$) lower than parent score($J(\varepsilon)$), always accept
  - If child score higher (child is worse than parent), then:

$\frac{J(\varepsilon')-J(\varepsilon)}{J\varepsilon}$ = child % worse	Prob of accepting
0%	100%
1%	79%
2%	63%
5%	31%
10%	10%
20%	1%

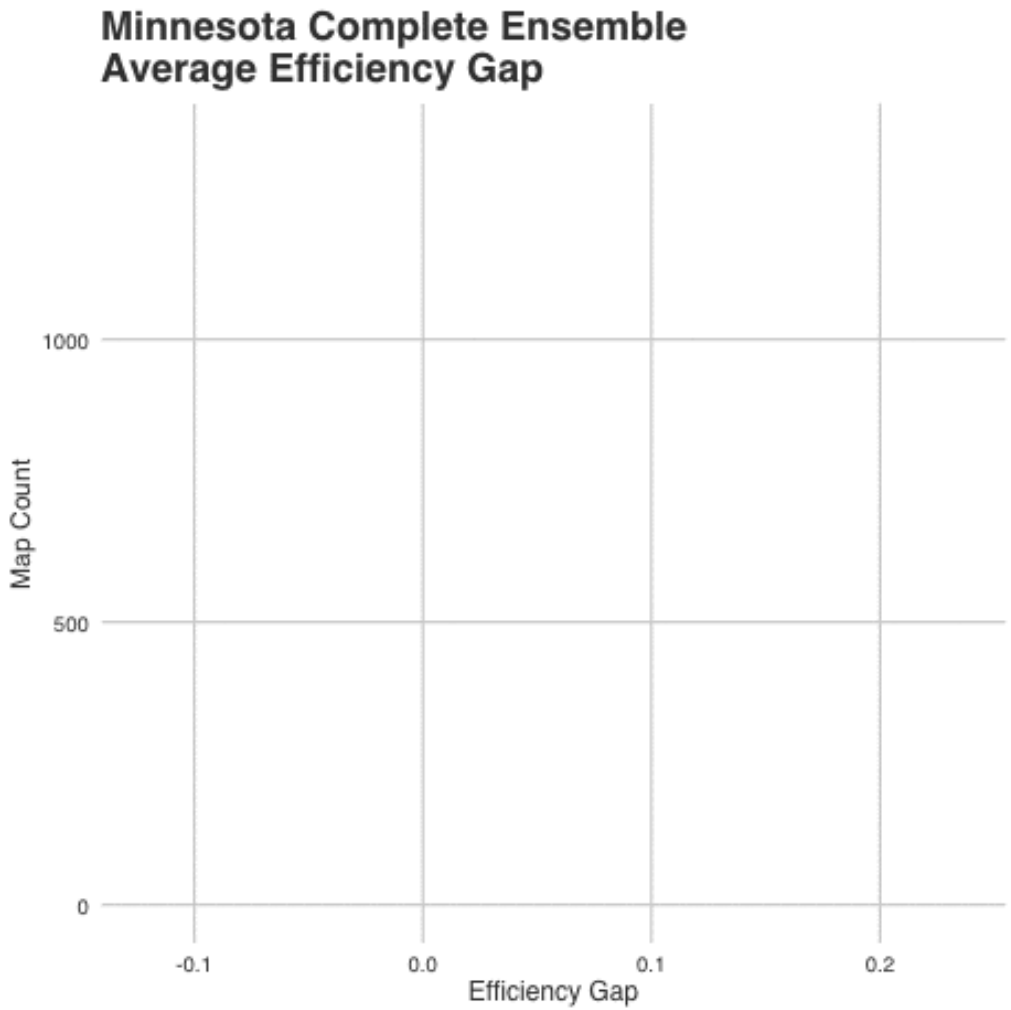
The Ensemble

- We ran 5 chains, each 150,000 accepted maps long
- The first 10,000 maps...
  - The simulated annealing term is $\beta = 0$ so every new map is accepted
  - We don't include these in the analysis
- Our final ensemble consists of 700,000 maps

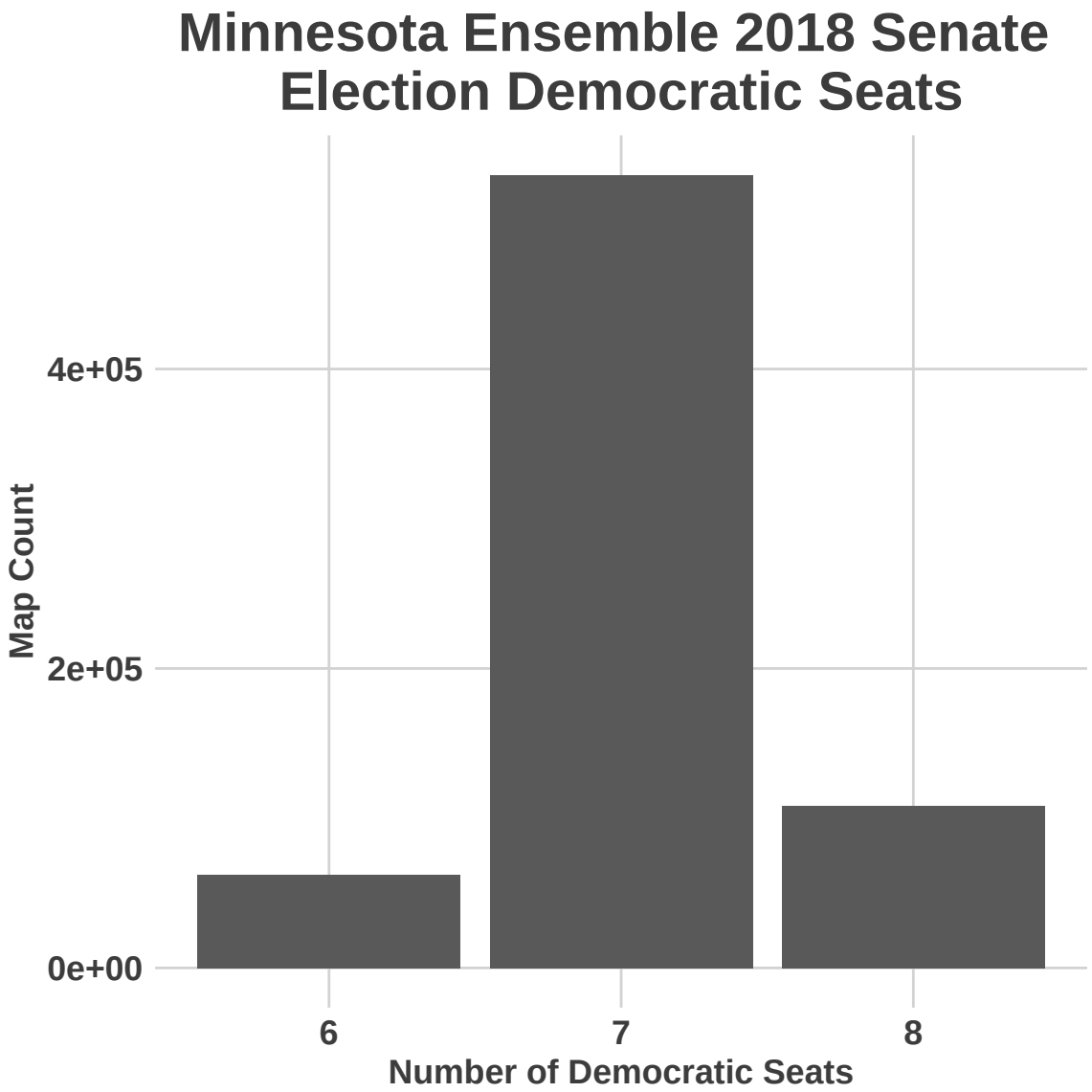


$\text{map}_{100,000}$

Ensemble Growth

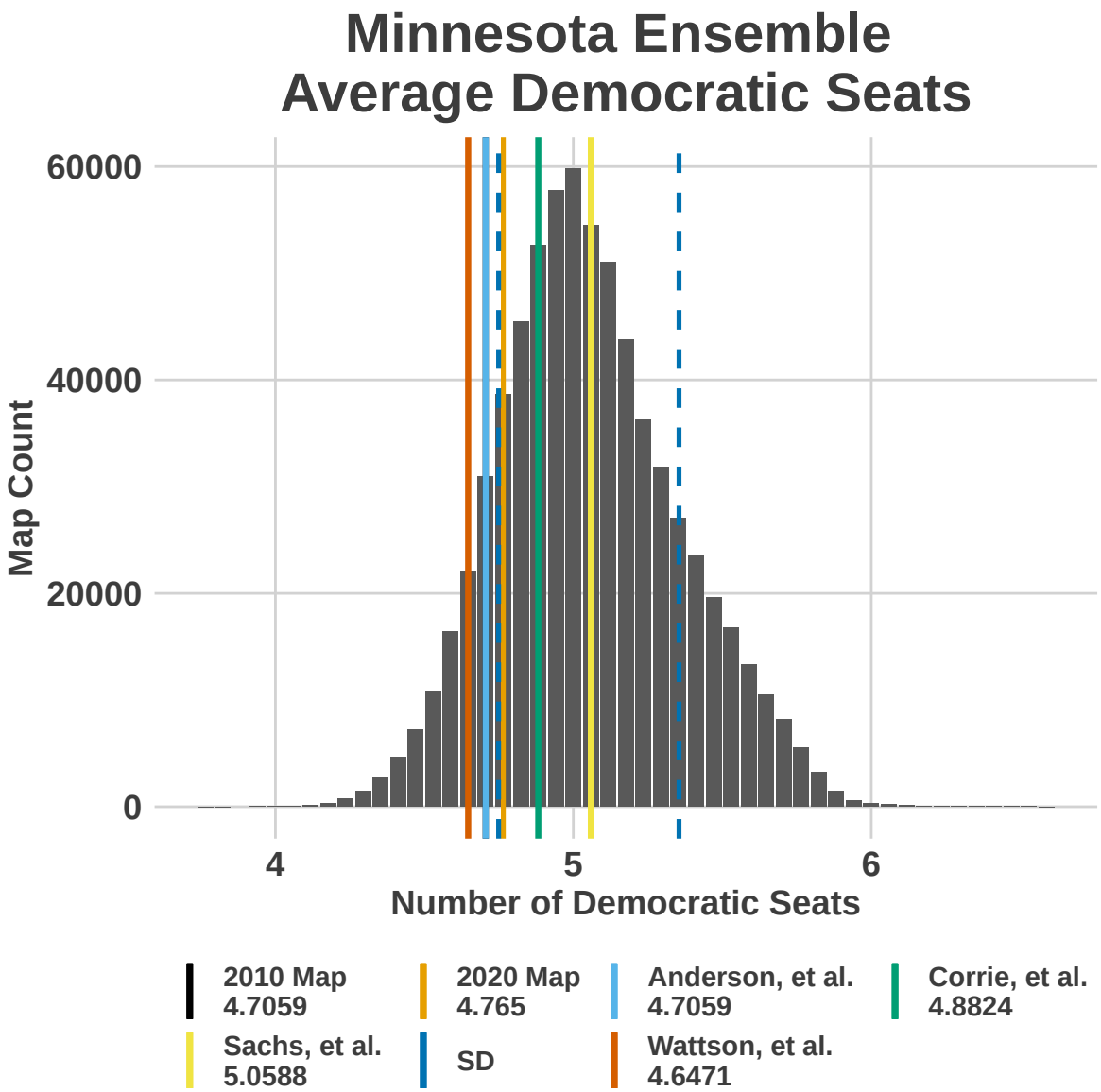


MN Seats Distribution - Senate 2018 Election

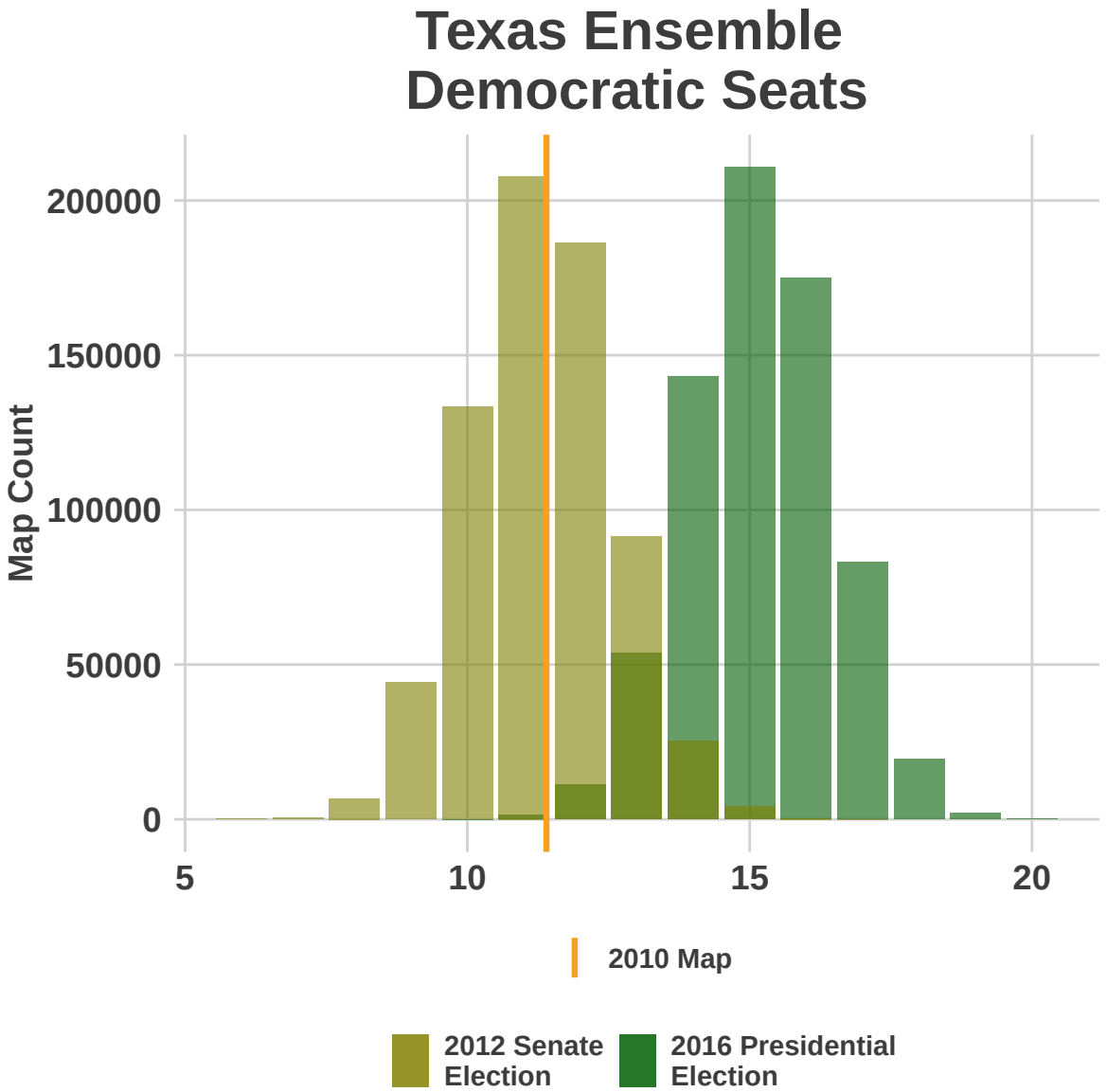


MN Seats Distribution

Mean = 5.052, SD = 0.303



TX Seats Distribution



Efficiency Gap

Definition

A vote is considered to be a **wasted vote** if it is cast for the winning party in a winning district and beyond the 50% threshold required to win a district or if it is cast for the losing party in a losing district.

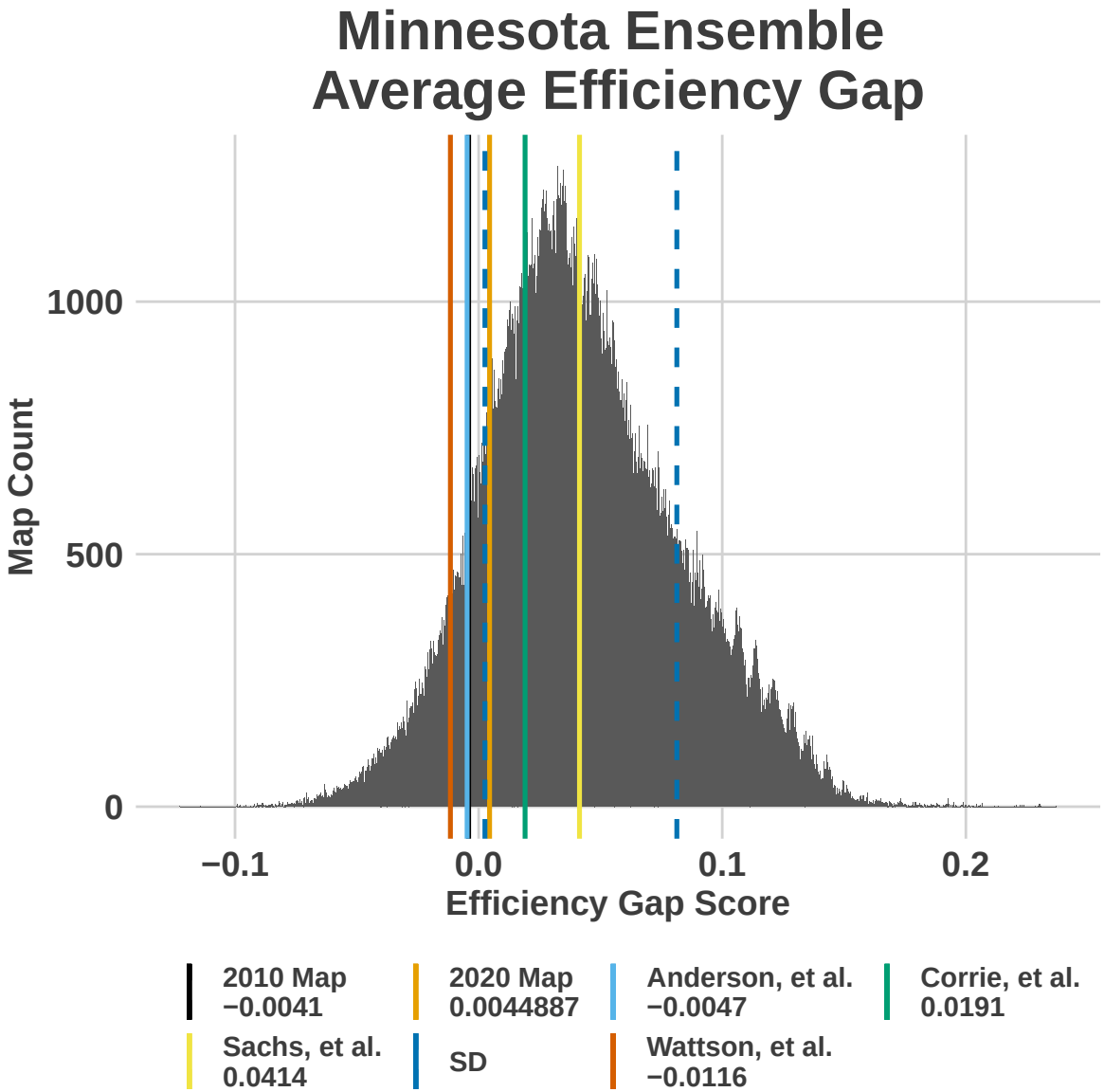
Definition

The **Efficiency Gap** is a measure which describes the difference in quantity of wasted votes between parties (in this case the Democratic and Republican parties) using the following formula:

$$\text{EG} = \frac{\text{Wasted Republican Votes} - \text{Wasted Democratic Votes}}{\text{Total Votes Cast}}$$

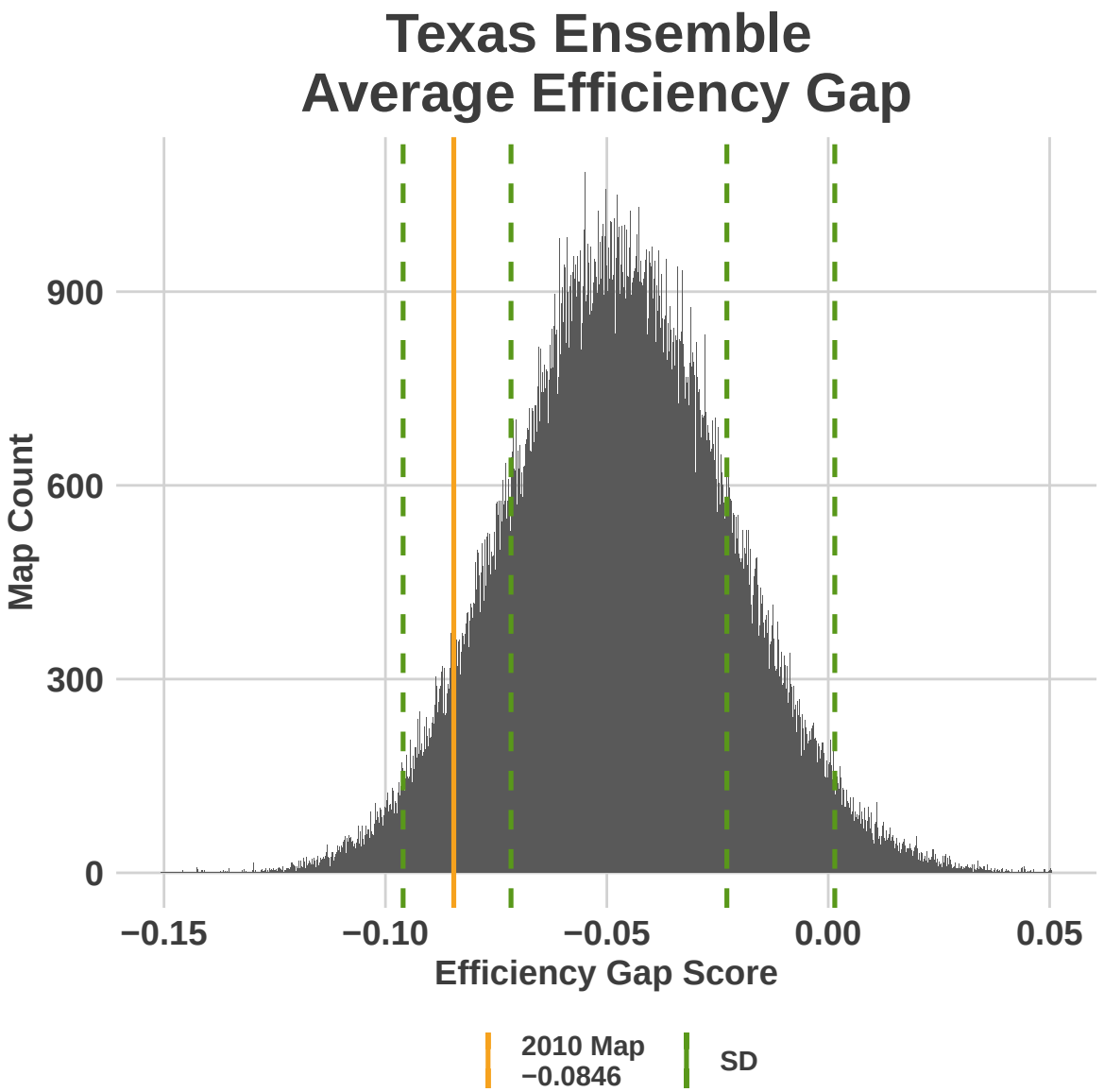
Efficiency Gap in MN

Mean = 0.042, SD = 0.039



Efficiency Gap in Texas

Mean = -0.047, SD = 0.024



Mean-Median

Definition

The **Mean-Median** for a given districting plan compares the state-wide vote share for one party to the median district-wide vote share for that same party. For a given plan ε, the mean-median $\text{mm}(\varepsilon)$ is defined as

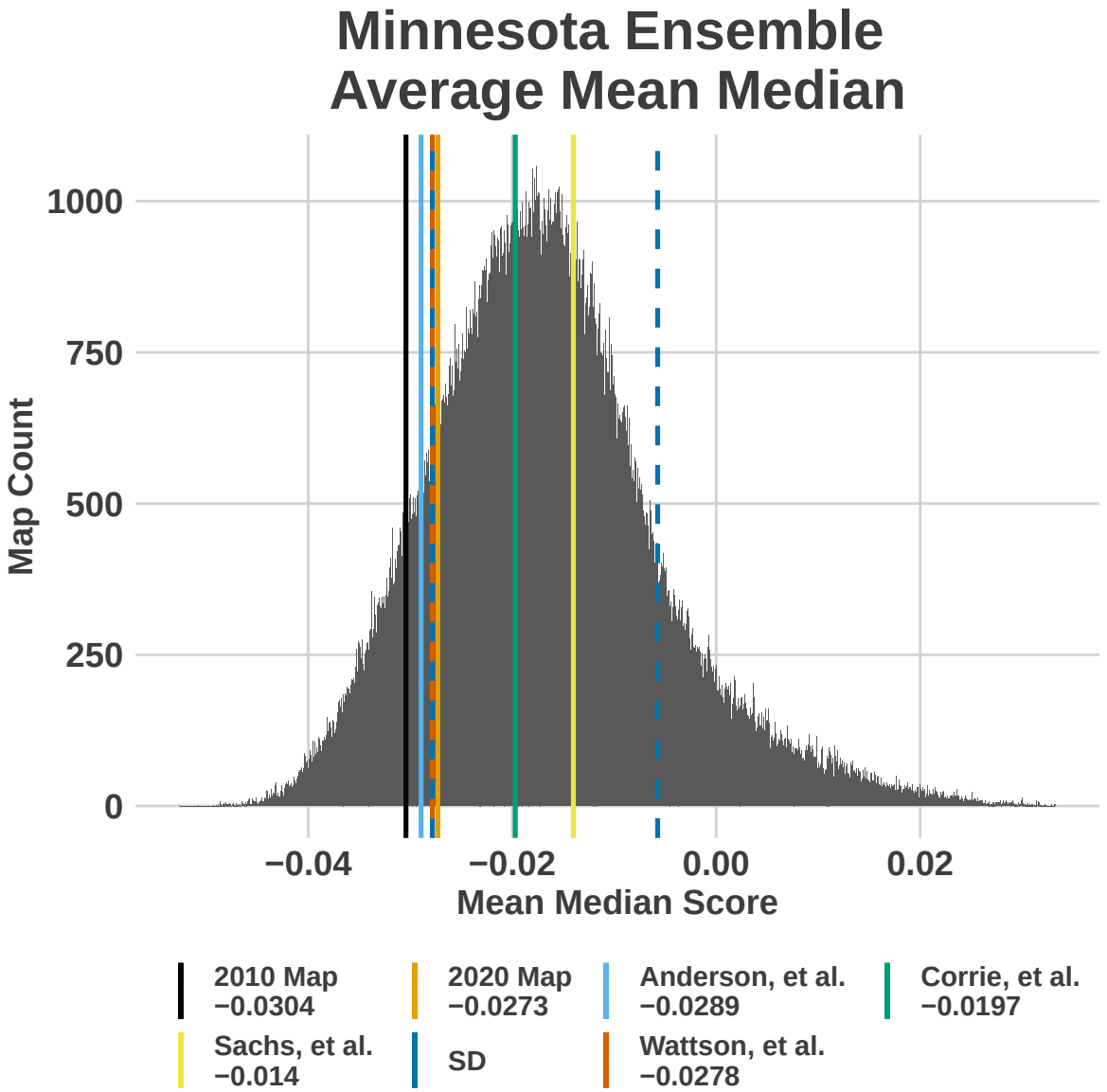
$$\text{mm}(\varepsilon) = \text{state-wide R vote share} - \text{median district-wide R vote share.}$$

If $\text{mm}(\varepsilon) > 0$, then evidence that Republicans are cracked/packed.

If $\text{mm}(\varepsilon) < 0$, then evidence that Democrats are cracked/packed.

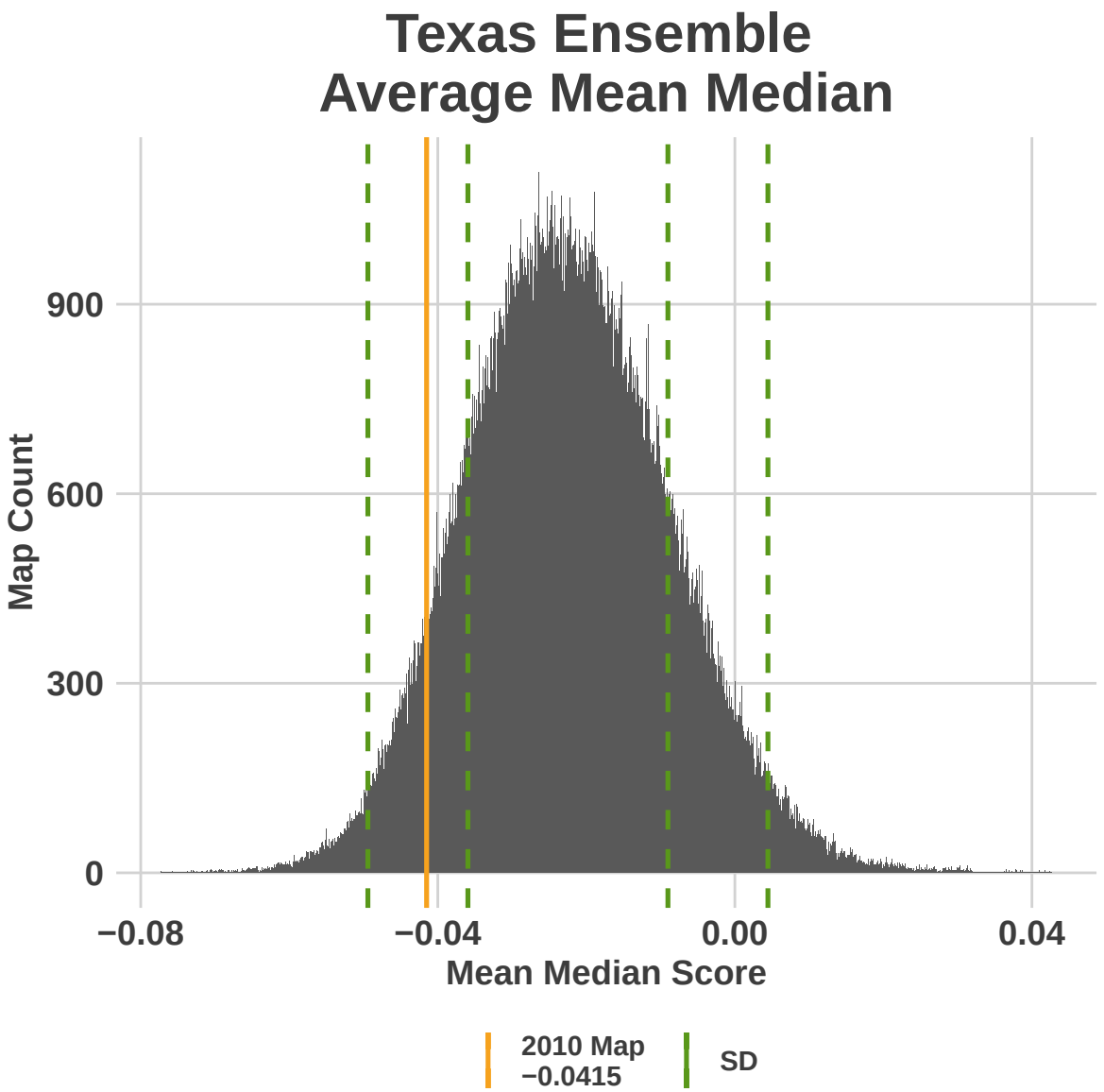
Mean-Median in MN

Mean = -0.017, SD = 0.011

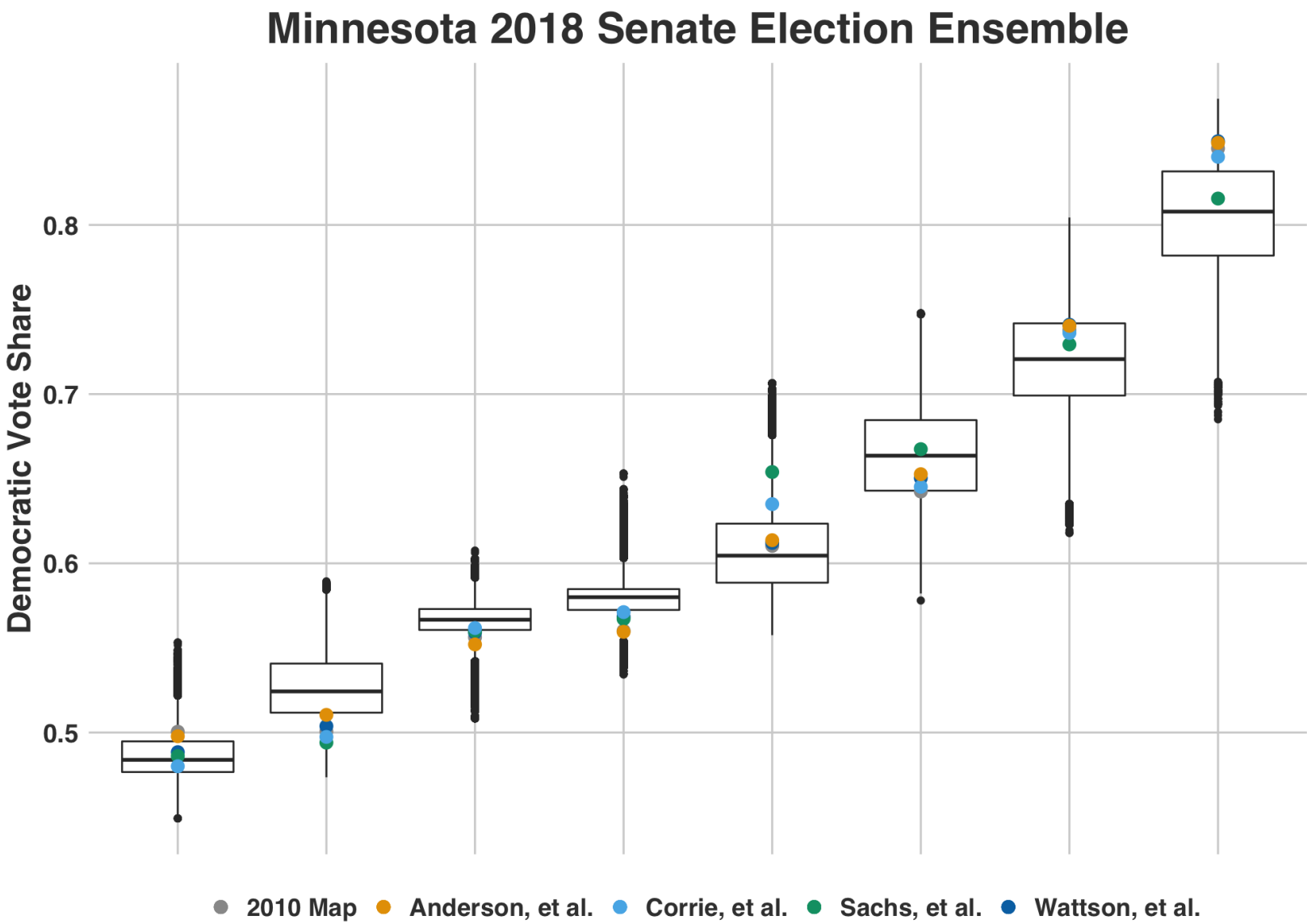


Mean-Median in Texas

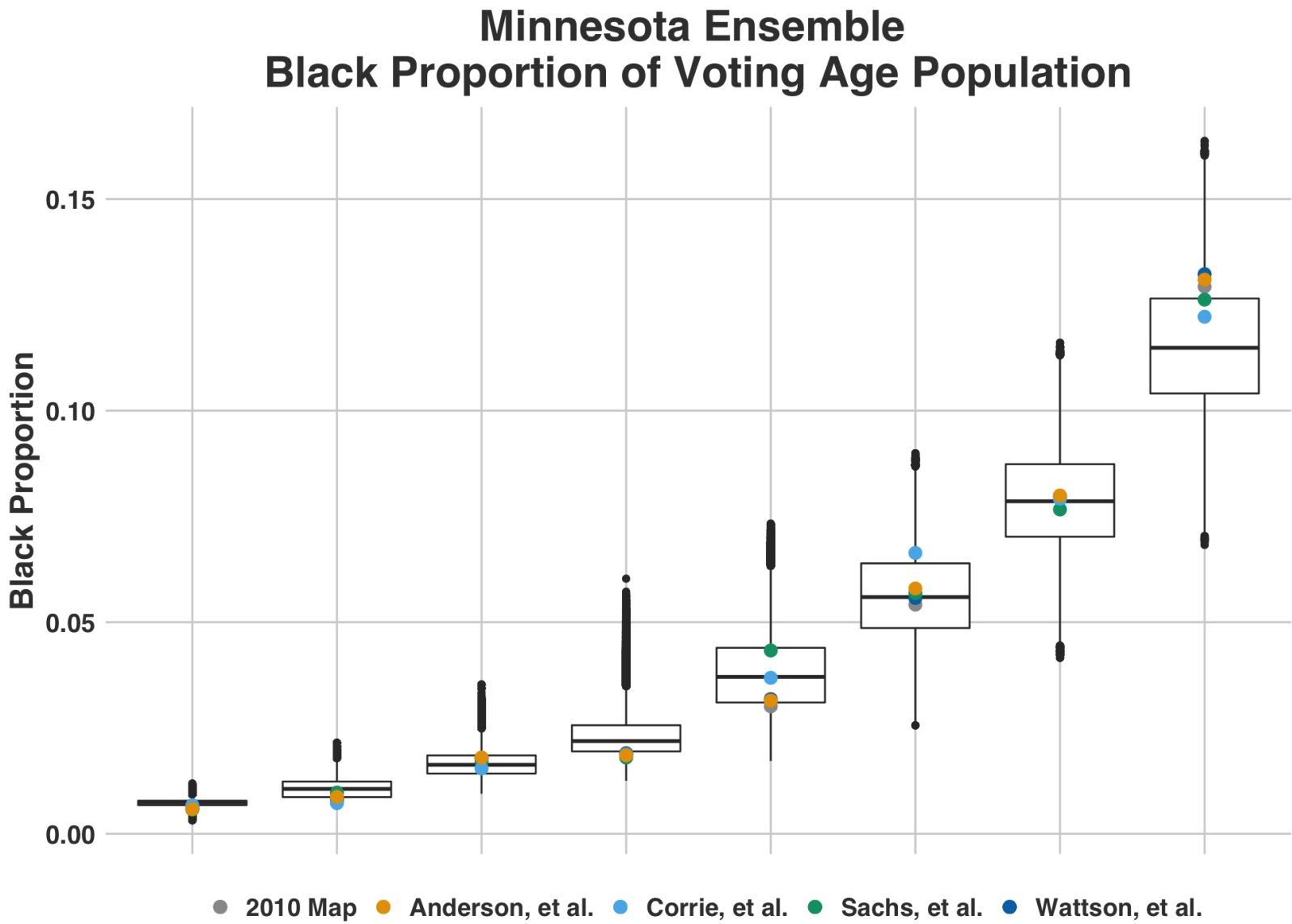
Mean = -0.022, SD = 0.013



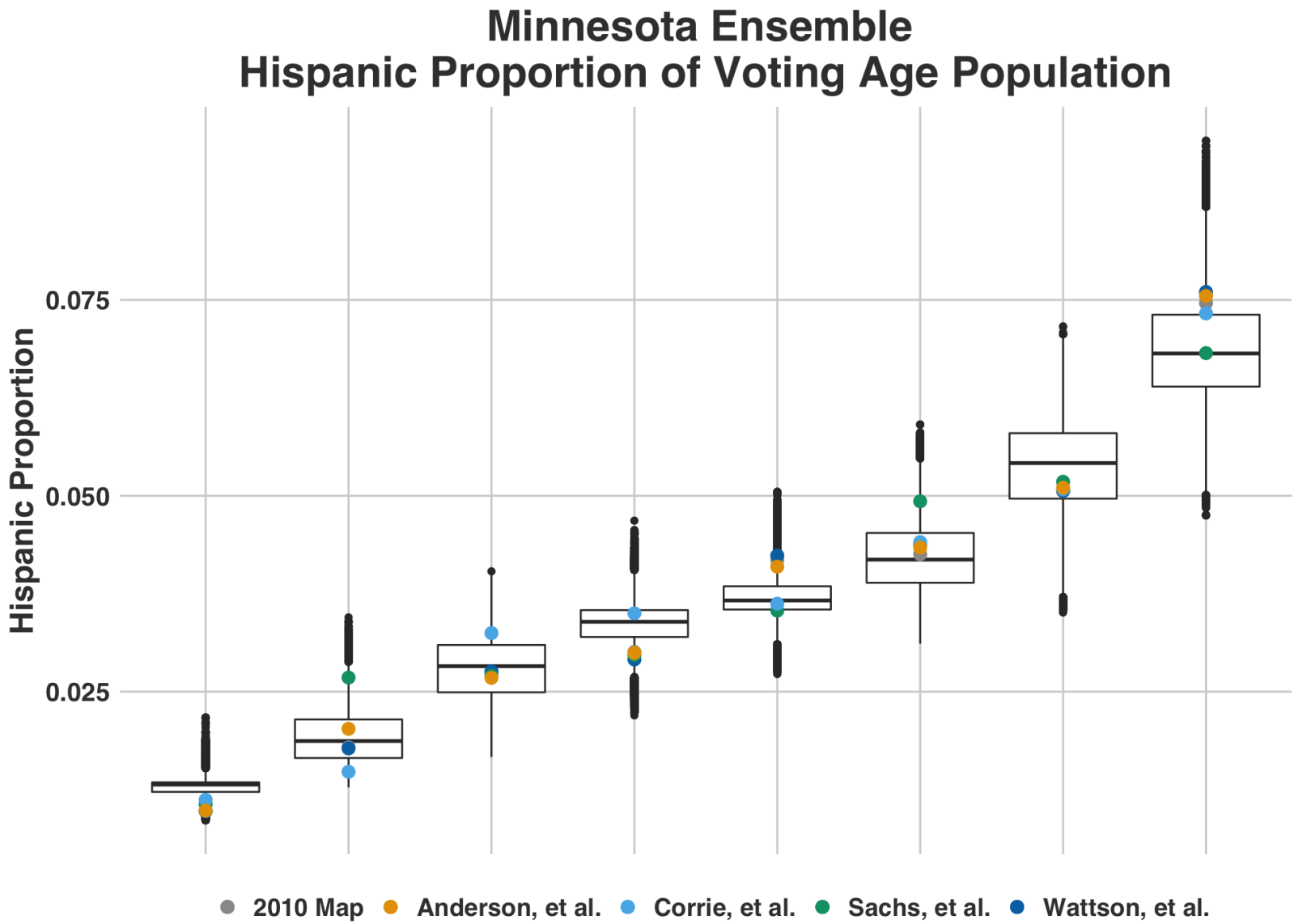
MN Democratic Vote Share Distribution



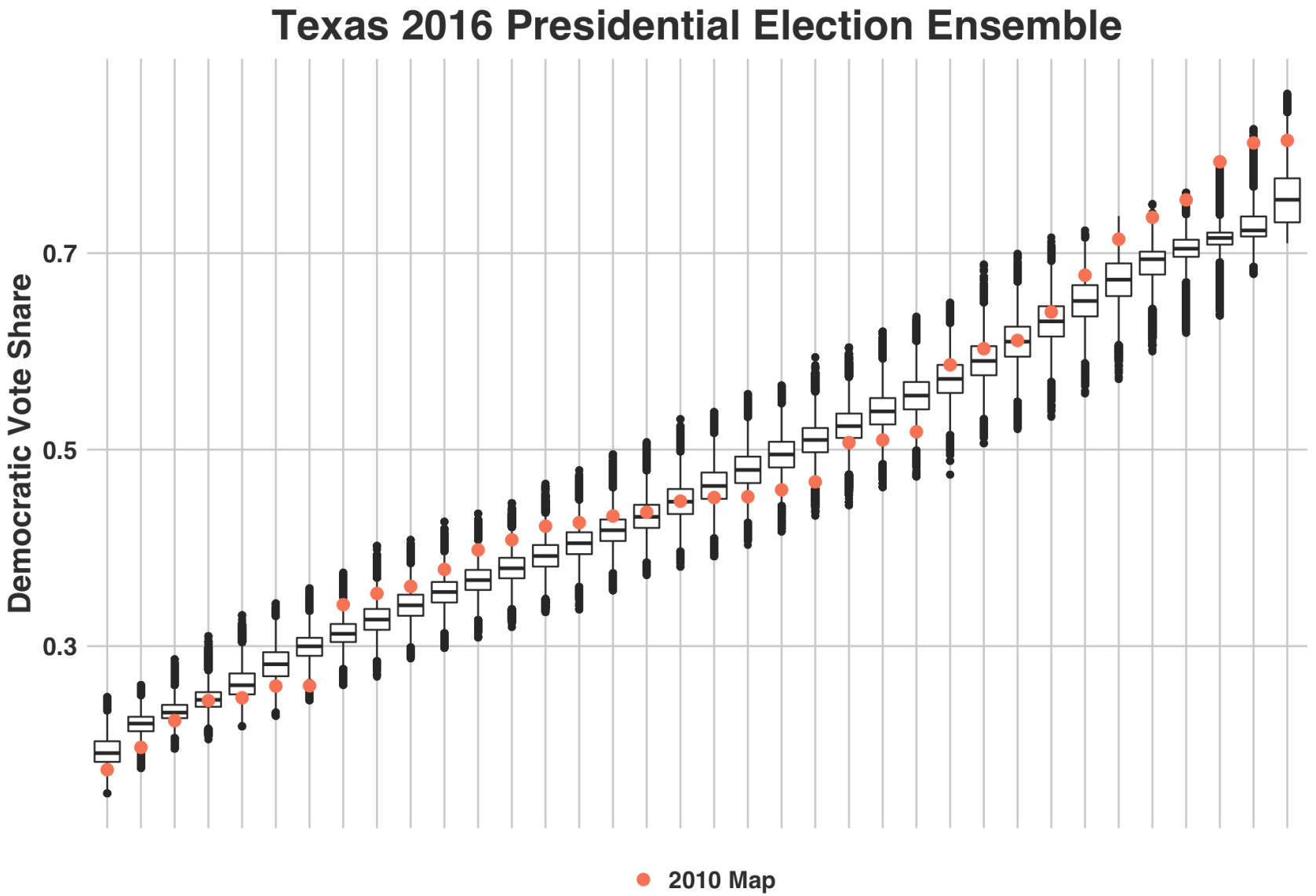
MN Black Vote Share Distribution



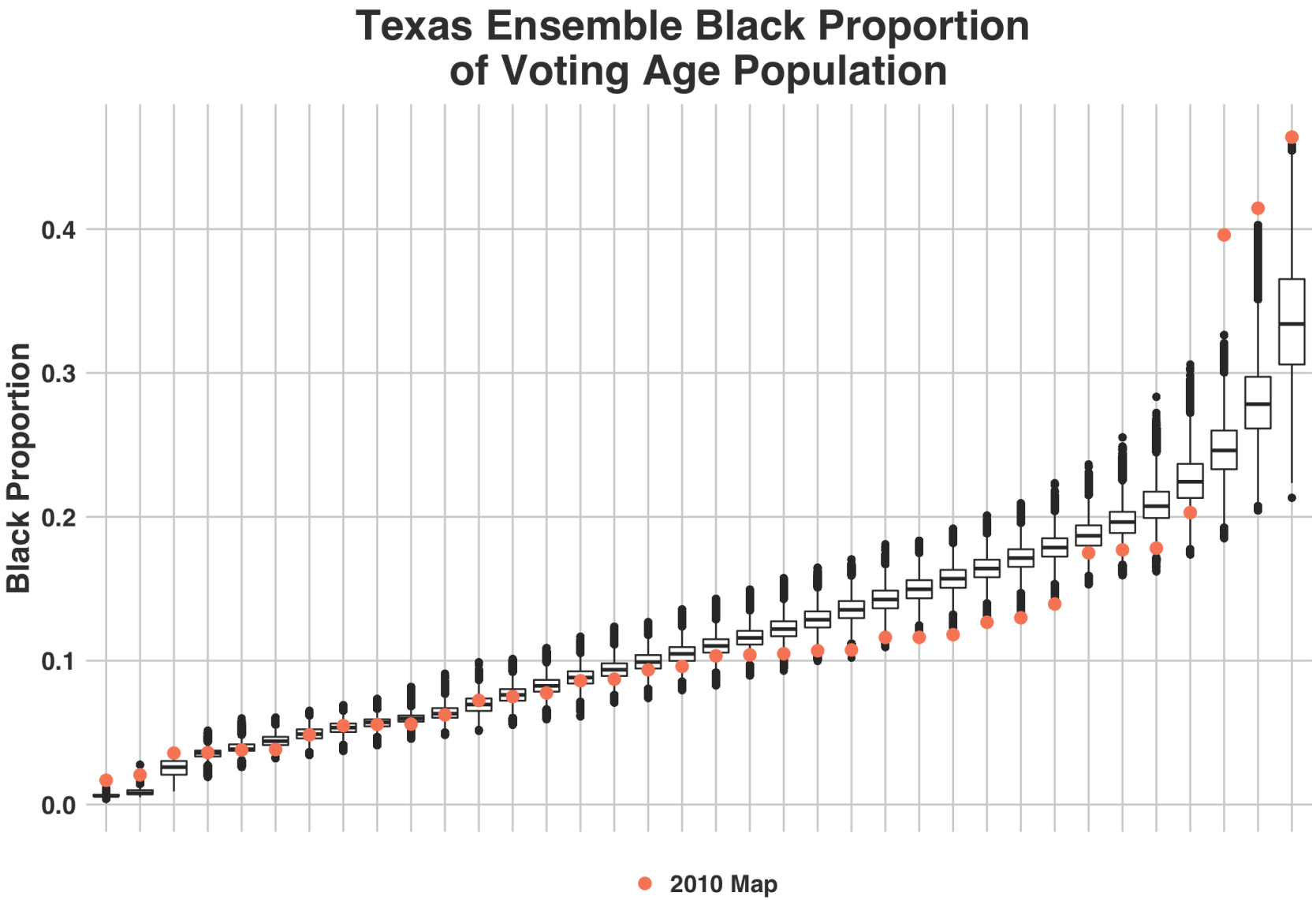
MN Hispanic Vote Share Distribution



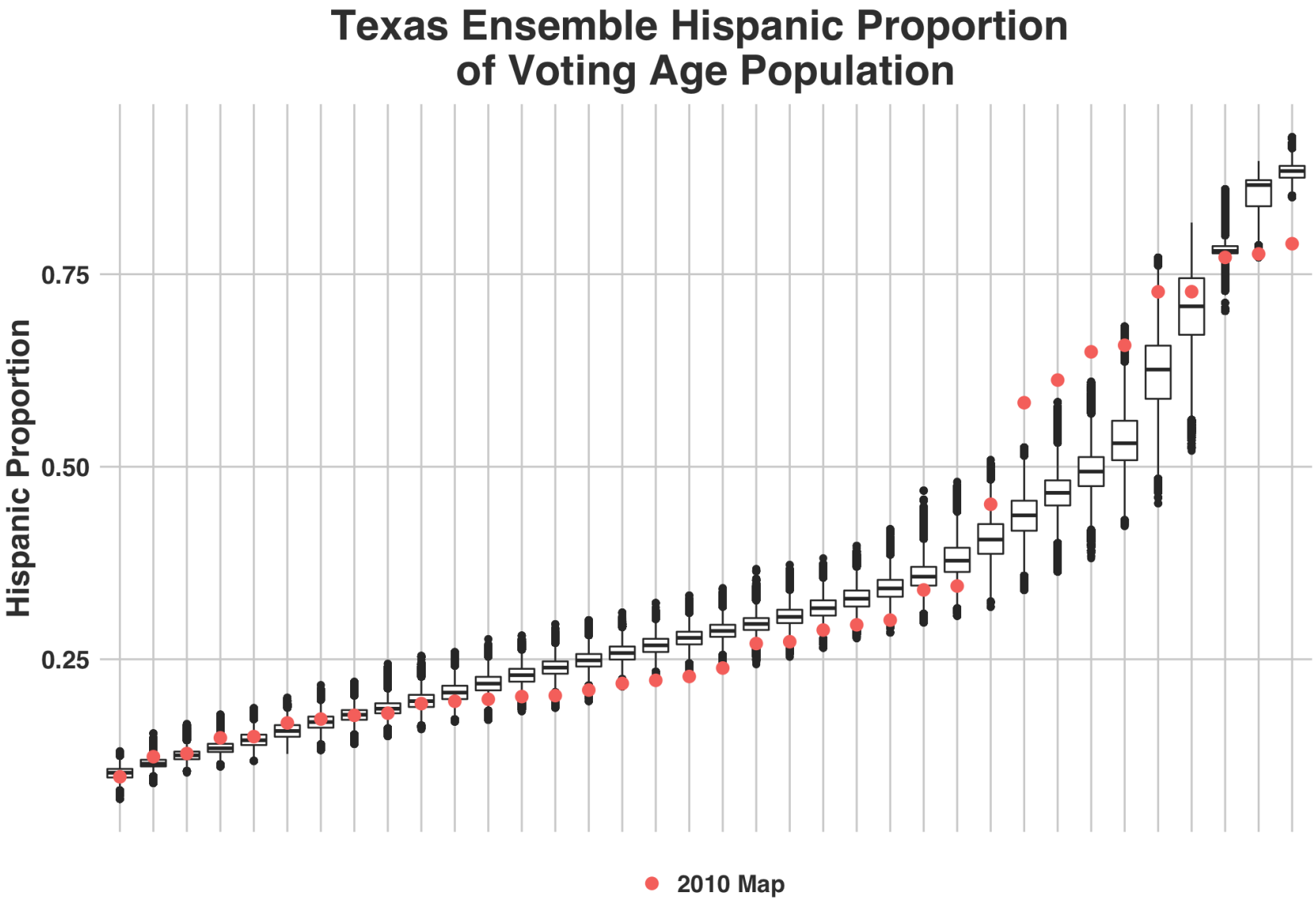
TX Democratic Vote Share Distribution



TX Black Vote Share Distribution



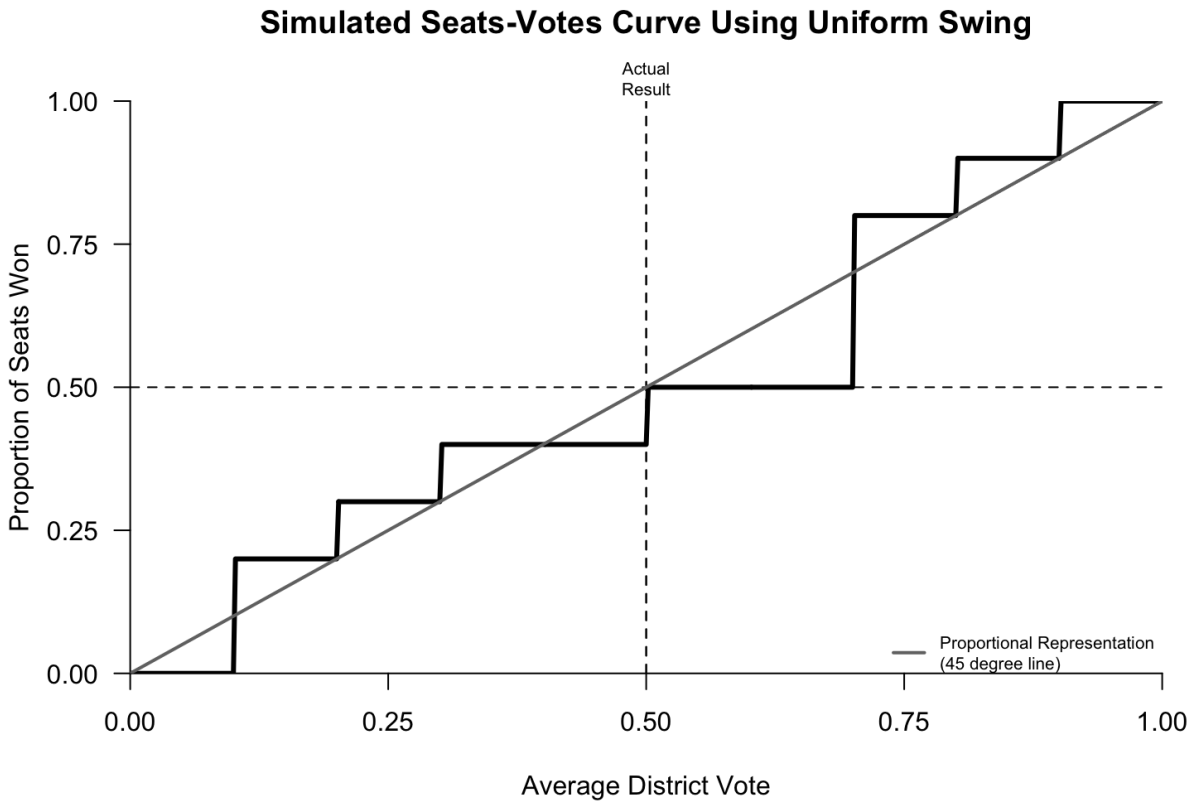
TX Hispanic Vote Share Distribution



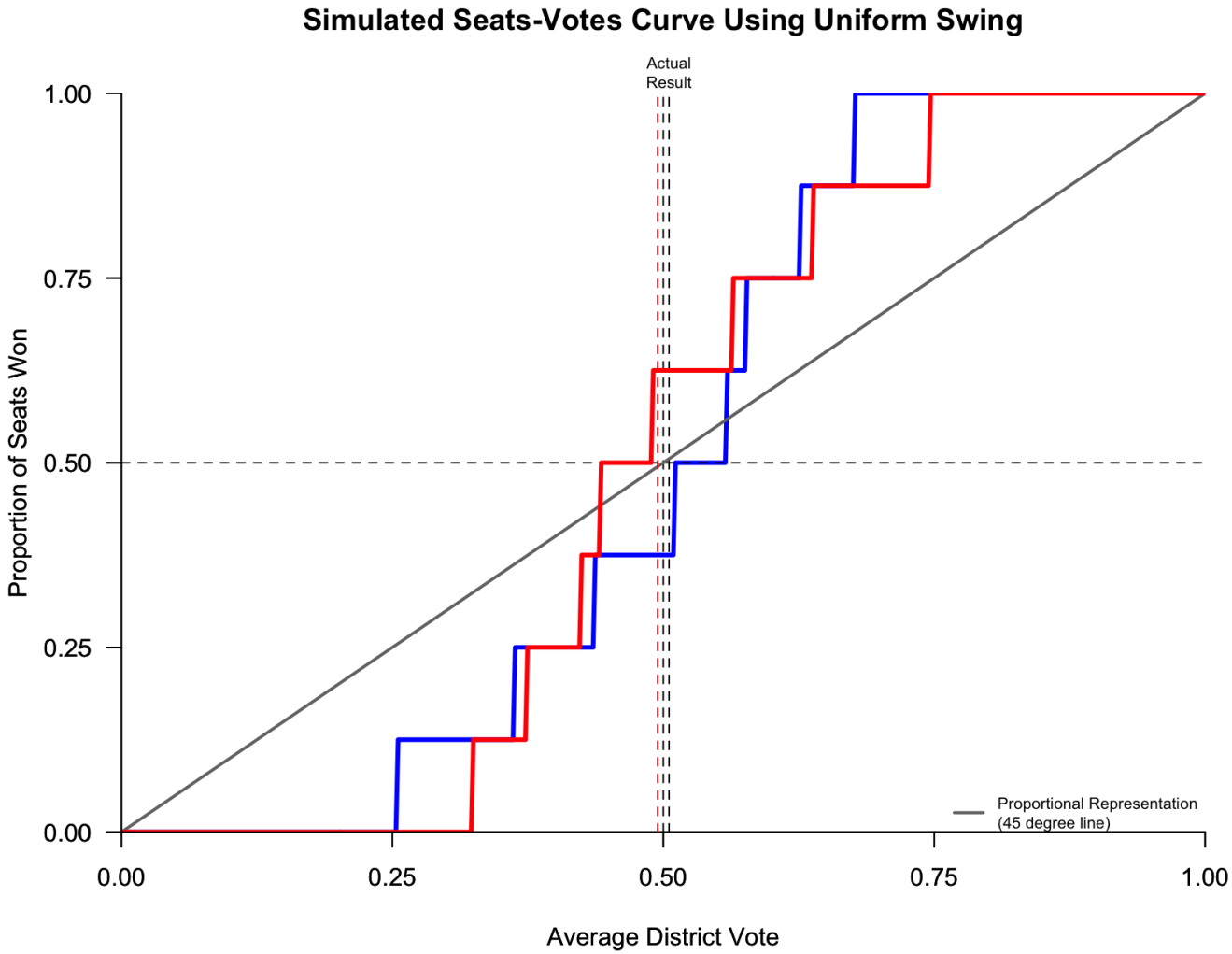
Seats-Votes Curve

Definition

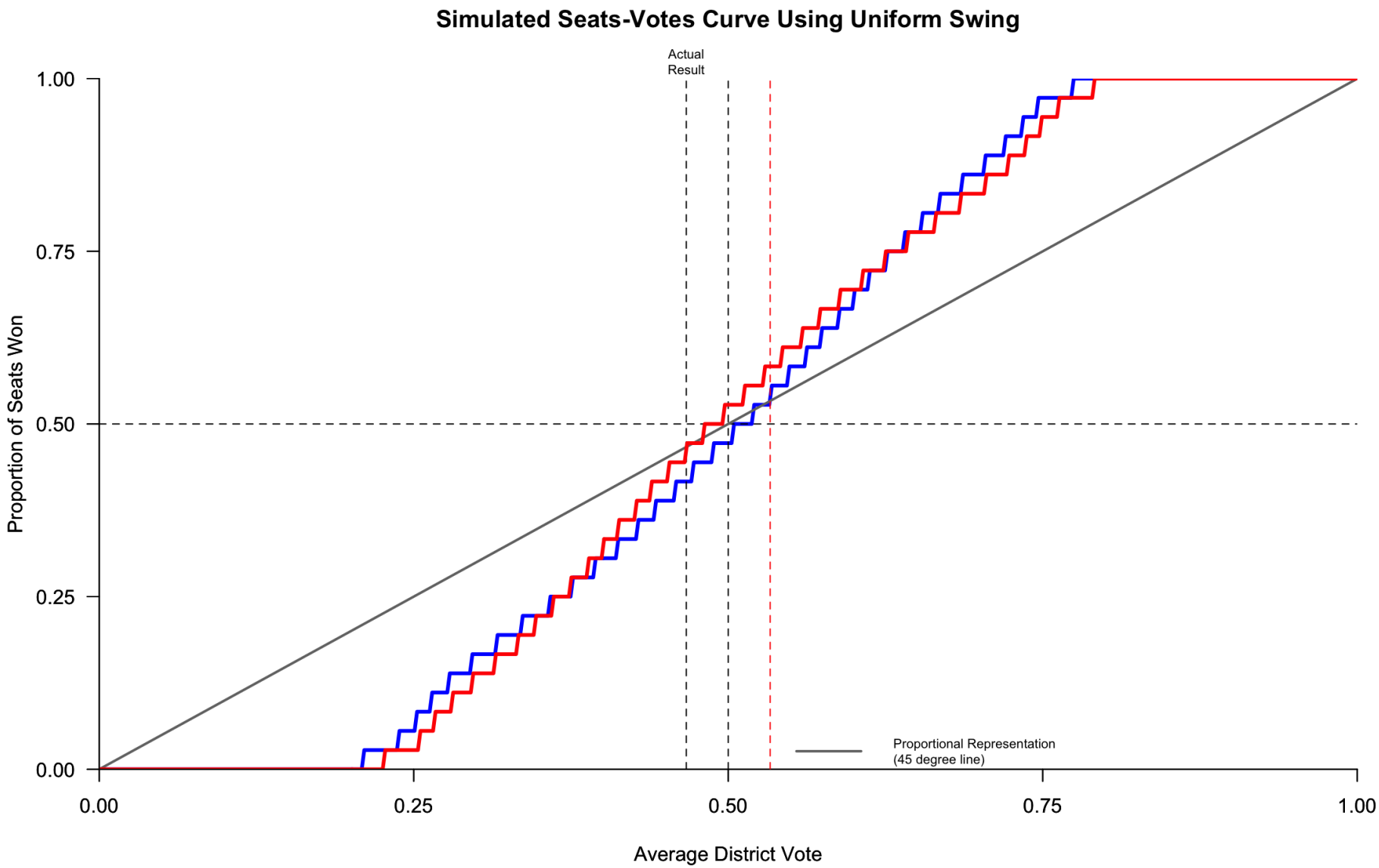
The **seats-votes curve** describes a party's seat share as a function of its vote share under a redistricting plan.



Average Ensemble Seats-Votes Curve MN



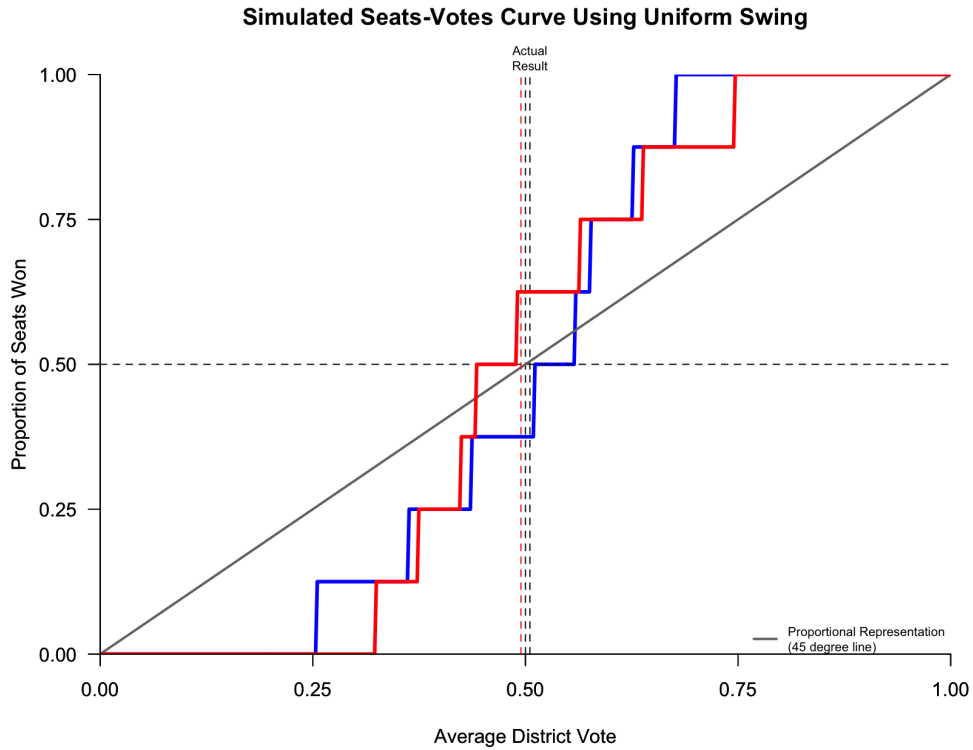
Average Ensemble Seats-Votes Curve TX



Seats-Votes Curve Metrics: Partisan Bias

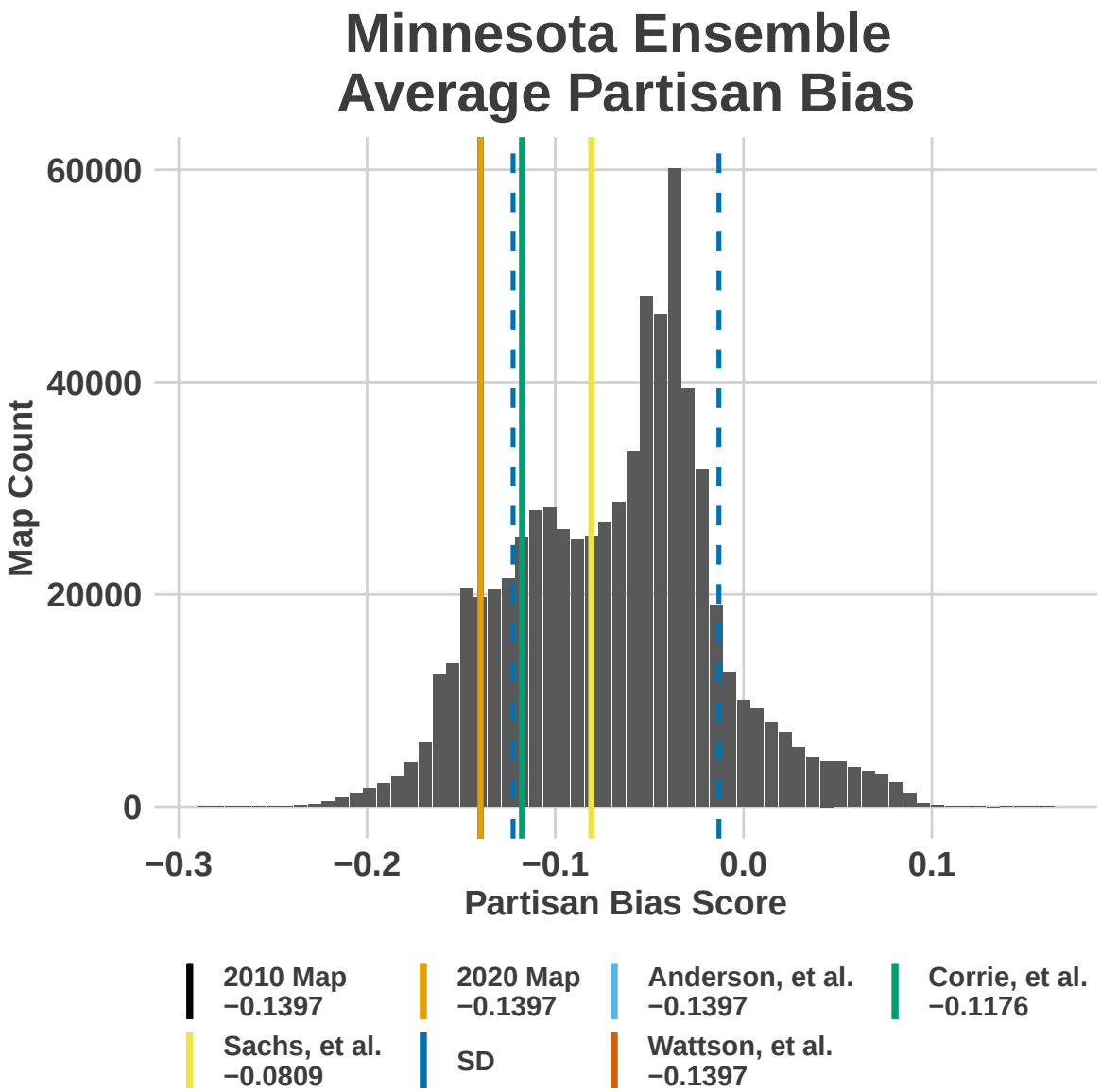
Definition

The **Partisan Bias** of a redistricting plan is the difference between a party's seat share at 50% of the vote and 50%.



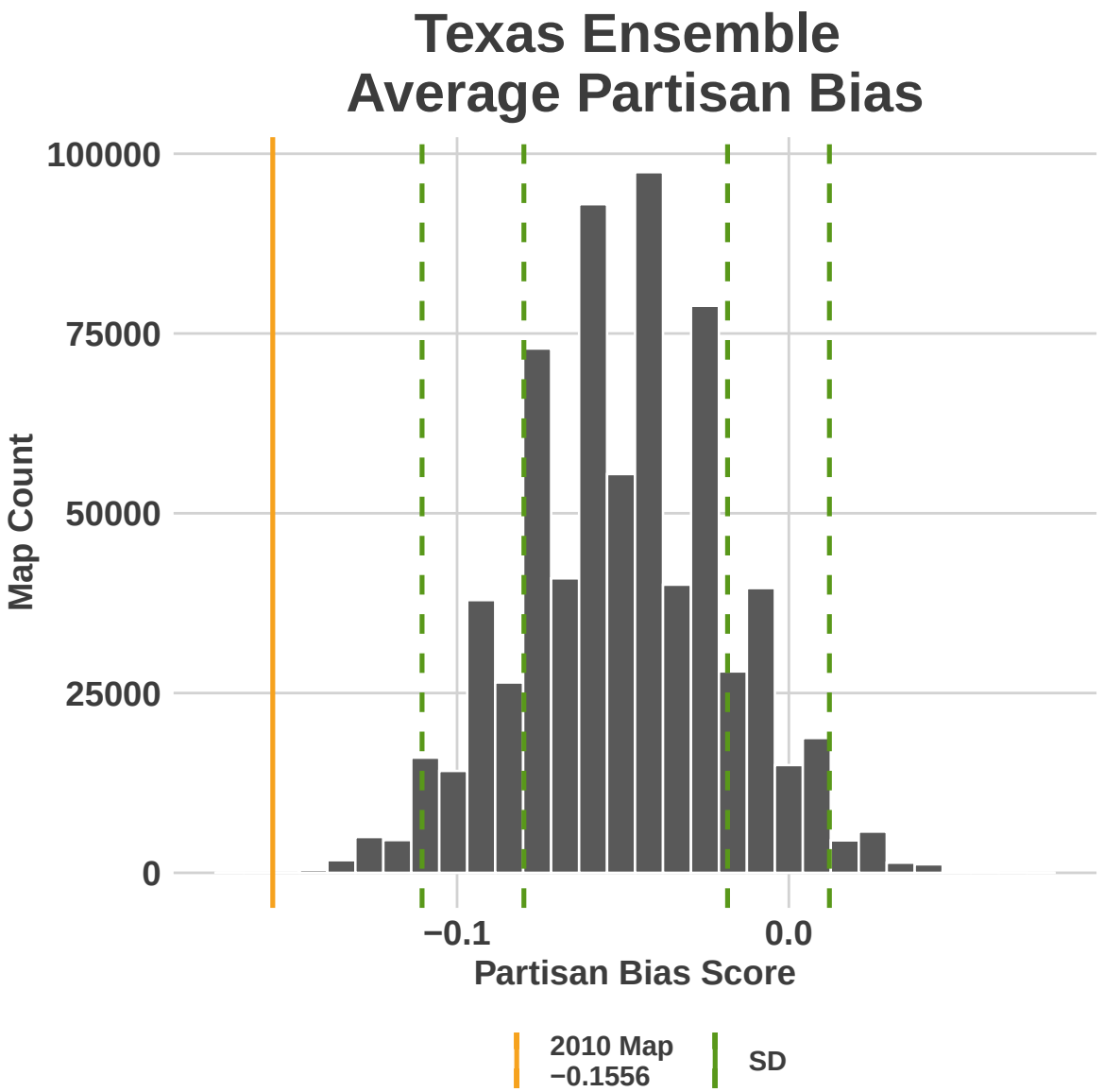
Partisan Bias in MN

Mean = -0.068, SD = 0.055



Partisan Bias in TX

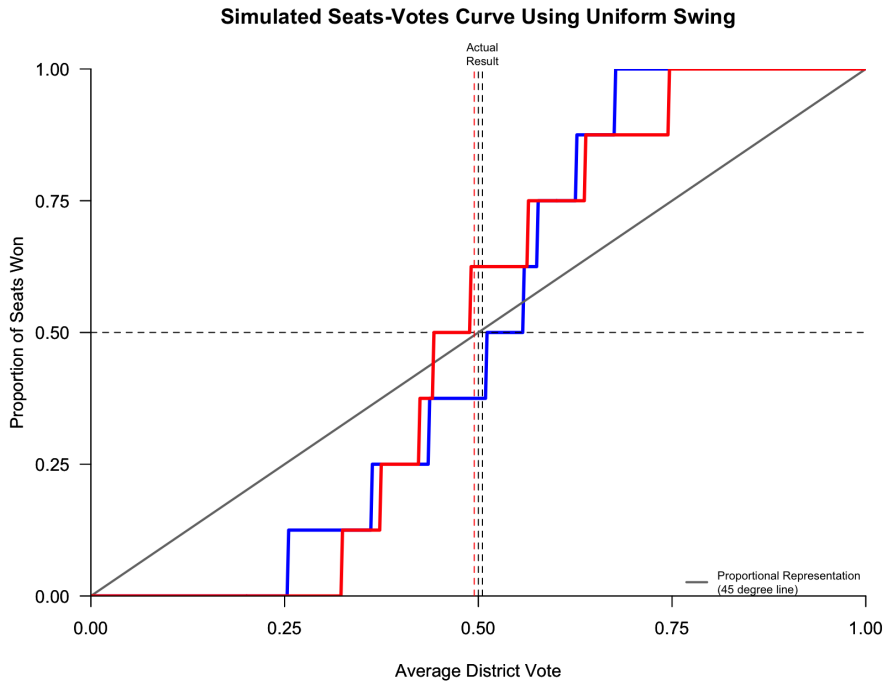
Mean = -0.049, SD = 0.031



Seats-Votes Curve Metrics: Partisan Gini

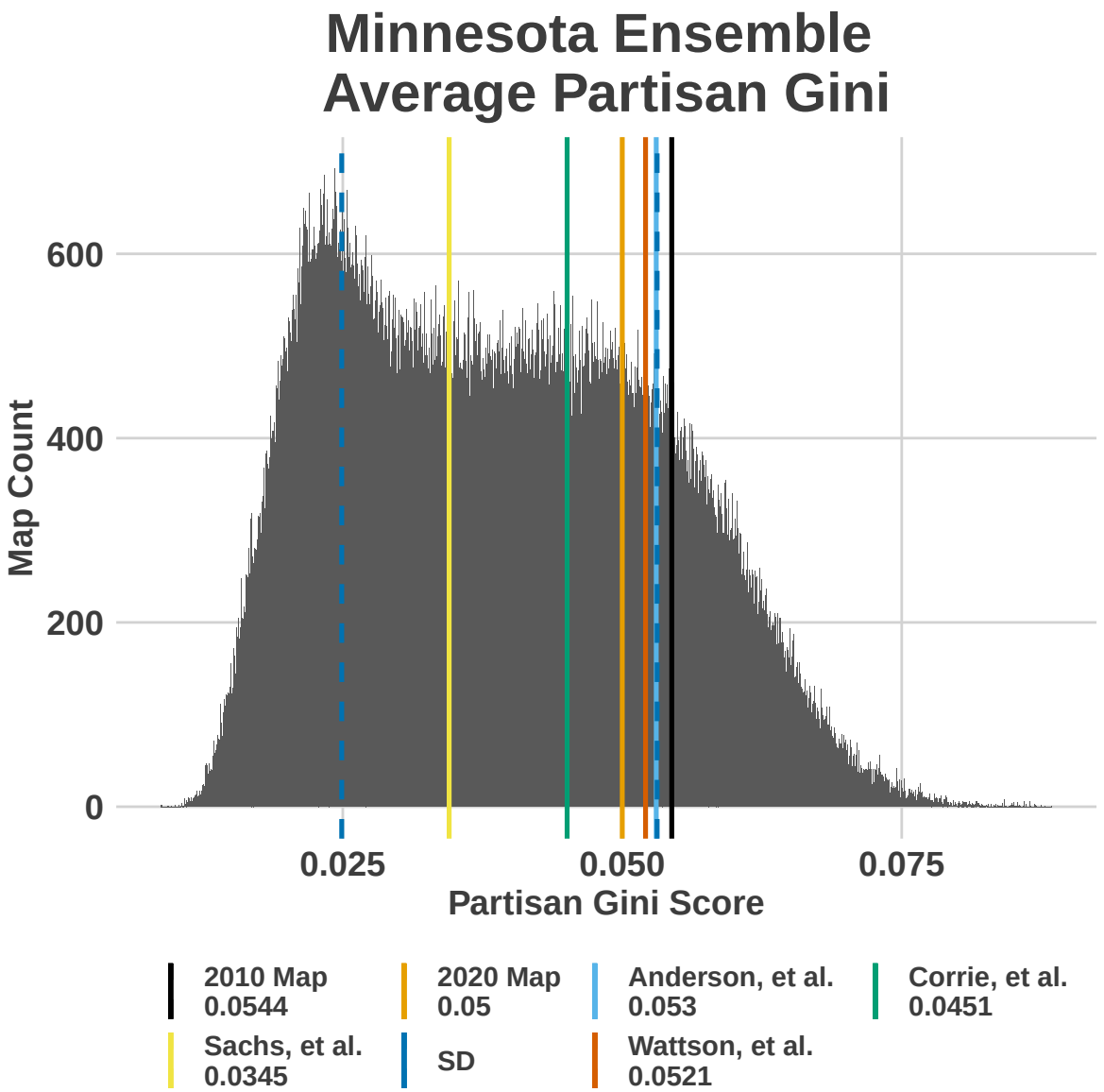
Definition

Partisan Gini is equivalent to the area between the seats-votes curve and its reflection about (50%, 50%). Values map to [0, 1], where a value of 0 means the curves for the two parties are identical.



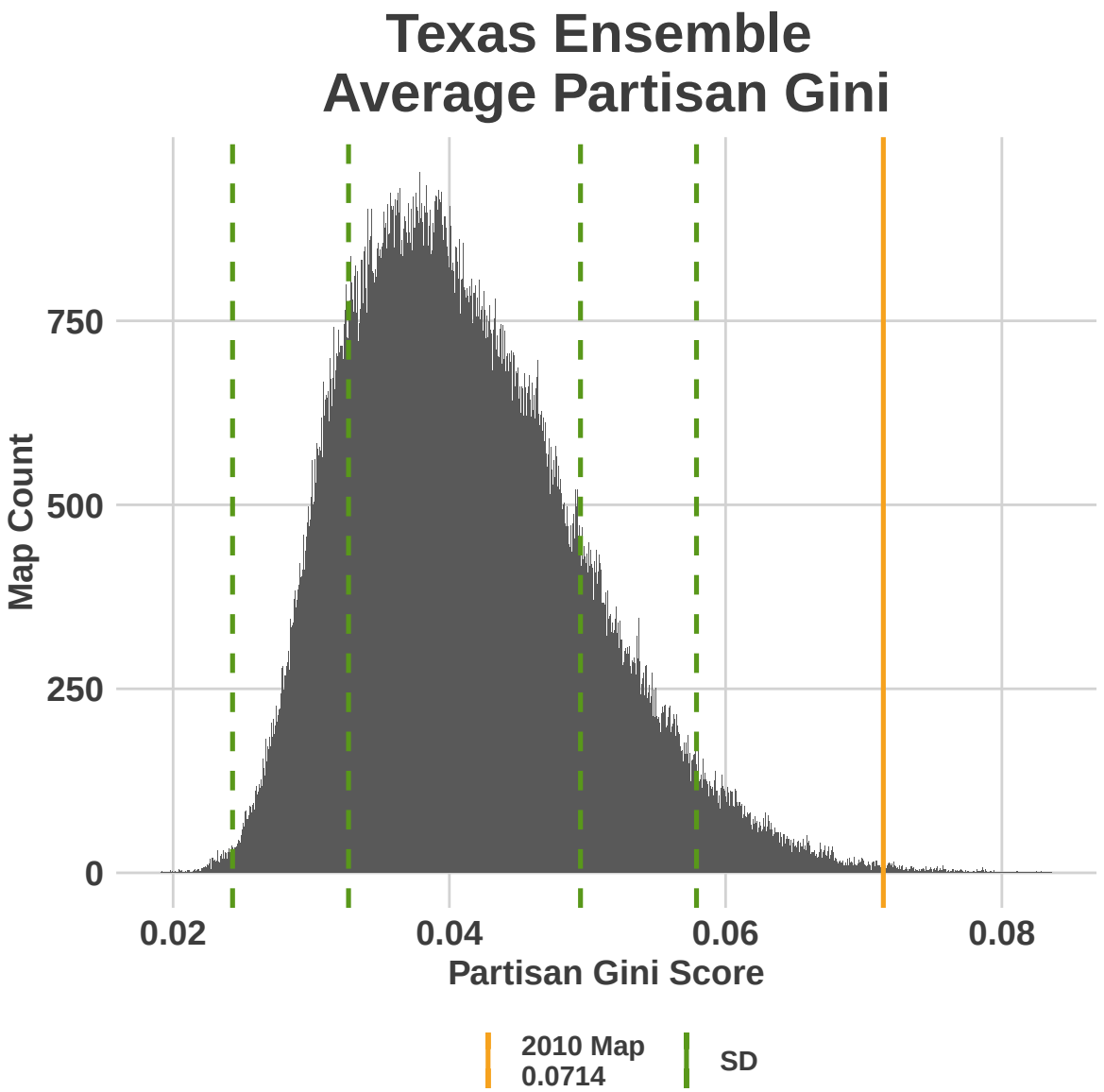
Partisan Gini in MN

Mean = 0.039, SD = 0.014



Partisan Gini in TX

Mean = 0.041, SD = 0.008



Minnesota Conclusions

- Wattson and Anderson Plaintiff plans are the most unusual of the proposed plans
- These two proposals along with the 2010 map are not outliers but are less typical for our ensemble
- The Corrie and Sachs Plaintiff plans fall near the middle of the ensemble on most metrics

Texas Conclusions

- The 2010 map is an outlier for minority vote share distributions, partisan bias, and partisan gini
- Although not quite an outlier, the 2010 map is unusual for efficiency gap and mean-median
- Media reports the approved 2020 Texas map shows even more evidence of gerrymandering than the 2010 map:

How Texas Plans to Make Its House Districts Even Redder

By Nick Corasaniti, Ella Koeze and Denise Lu Oct. 3, 2021

Justice Dept. Files Voting Rights Suit Against Texas Over New Map

The department said the state's redistricting plan would violate the Voting Rights Act by discriminating against minority voters.

Overall Conclusions

- The mean value for many partisan metrics in our ensemble was different in MN and TX
- Pure value of these metrics alone isn't sufficient for identifying an unusual/unfair map
- Ensemble analysis is necessary to understand how unusual a map is for a given state

Thanks to...

Deanna Haunsperger

Carlos Parada

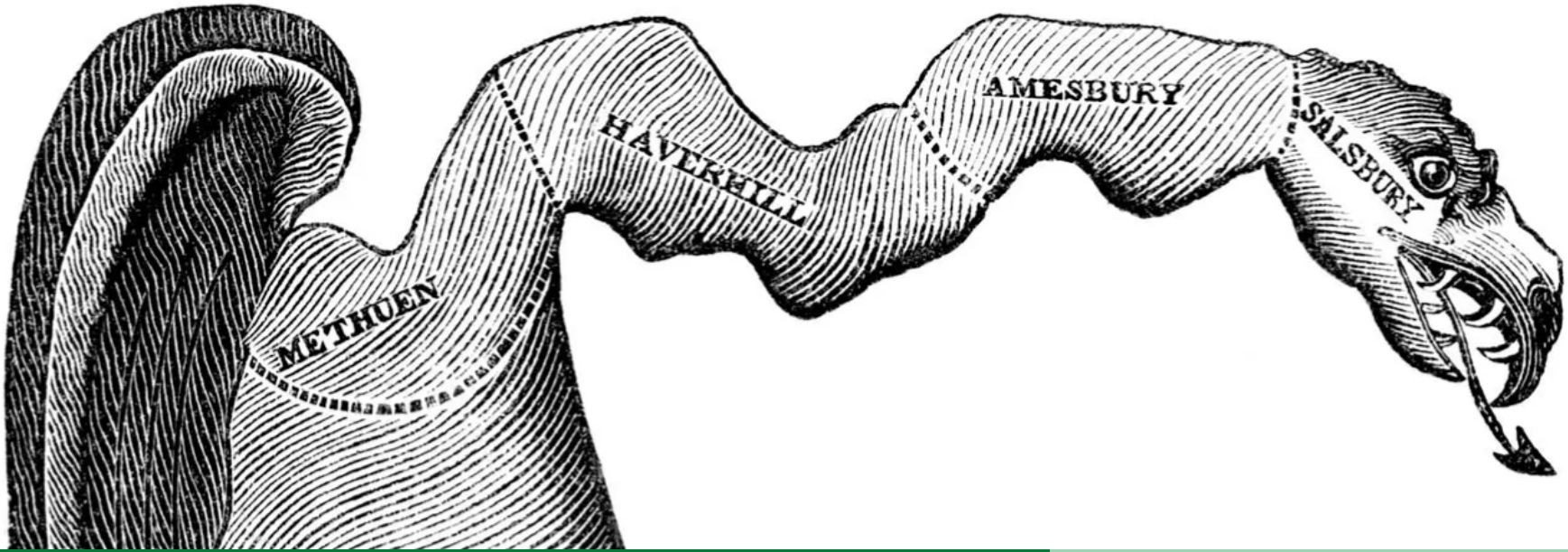
Christina Farhart

Adam Loy

Moon Duchin

Jonathan Mattingly

Q & A



Sources

- Encyclopedia Britannica
- Michigan's New Independent Citizens Redistricting Commission
- *Quantifying Gerrymandering in North Carolina* by Greogry Hershlag, et al.