Using Simulation Algorithms to Detect Gerrymandering: Evaluation of the 2022 Congressional Maps

Kosuke Imai

Harvard University

Quantitative Social Science Colloquium, Princeton University

September 23, 2022

Joint work with Christopher Kenny, Cory McCartan, Tyler Simko, Shiro Kuriwaki, George Garcia, Kevin Wang, and Melissa Wu

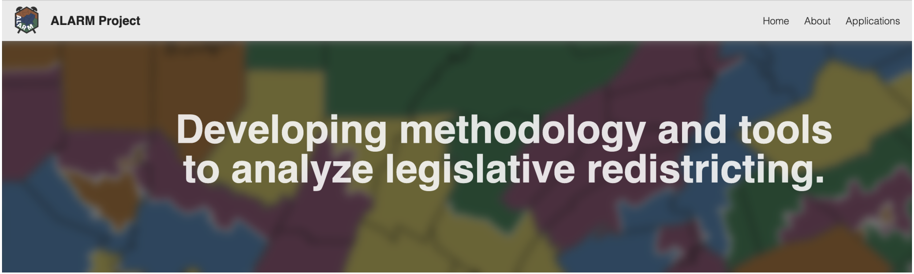
Motivation

- Today's world for quantitative social science:
  1. increasing availability of granular data
  2. rapid methodological advancement
- Social scientists can and should solve problems of the world!

- Redistricting as a major policy decision
- How can we use data and algorithms to evaluate redistricting plans?
  - traditional methods: comparison across states and time periods
  - confounded by state-specific political geography and rules

- Use of simulation algorithms
  1. obtain a representative sample of redistricting plans under constraints
  2. compare the enacted plan with this baseline distribution
- A technological solution to detecting gerrymandering

Algorithm-Assisted Redistricting Methodology (ALARM)

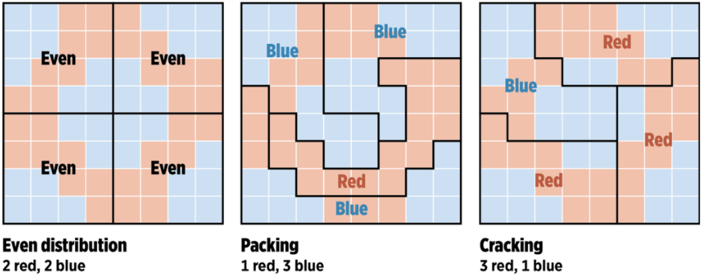


- What we do:
  1. develop efficient and flexible simulation algorithms
  2. build open-source software packages for the entire workflow
  3. evaluate redistricting plans in the United States and elsewhere

- Goal: empower researchers, policy makers, data journalists, and citizen data scientists

Redistricting Basics

- Classic gerrymandering strategies: packing and cracking



- What has changed: availability of granular data and mapping software (e.g., Maptitude)
- US congressional redistricting
  - racial gerrymandering: *Shelby County v. Holder*; *Merrill v. Milligan*
  - partisan gerrymandering: *Rucho v. Common Cause*; *Moore v. Harper*

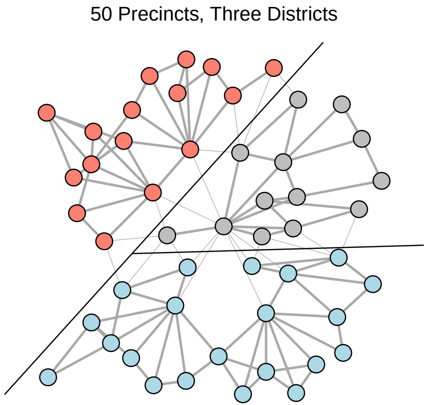
Why Use Simulation Algorithm for Redistricting Evaluation?

- Traditional redistricting evaluation
  1. compute various fairness metrics
  2. compare them across states and over time
- Confounded by differences in political geography and redistricting rules

- Simulation-based redistricting evaluation
  1. generate many alternative plans under a set of redistricting criteria
  2. compare them with a proposed plan to evaluate its properties

- Benefits of simulation approach
  1. can control for state-specific political geography and redistricting rules
  2. transparency and ability to isolate a relevant factor
  3. mathematical properties $\rightsquigarrow$ representative sample of alternative plans

Redistricting as a Balanced Graph Partition Problem



- Efficient enumeration algorithm exists (Fifield *et al.* 2020)
- Only applicable to very small redistricting problems

Existing Algorithms

1. Constructive Monte Carlo (Chen & Rodden, 2013; Magleby & Mosesson, 2018)
   - randomly select "seeds" and grow districts
   - unknown target population

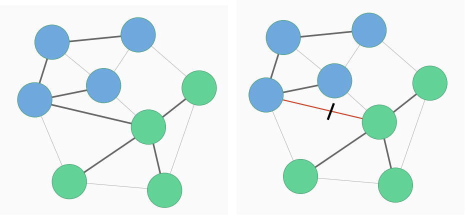
2. Flip algorithms (Fifield *et al.*, 2020; Mattingly & Vaughn, 2014; Chikina *et al.* 2017)
   - start with a valid plan and then reassign units on district boundaries
   - target distribution

$$\pi(\xi) \;\propto\; \underbrace{\exp(-J(\xi))}_{\text{custom constraints}} \;\times\; \underbrace{1_{\xi \text{ connected}}}_{\text{contiguity requirement}} \;\times\; \underbrace{1_{\text{dev}(\xi)\leq D}}_{\text{population balance}}$$

   - incremental changes; applicable for local exploration
   - does not scale; compactness needs to be specified in $J(\cdot)$

② Merge-split algorithms (DeFord *et al.*, 2021; Carter *et al.* 2019)

- randomly choose a pair of adjacent districts, merge them, and split them into two new districts using uniform spanning trees



- target distribution

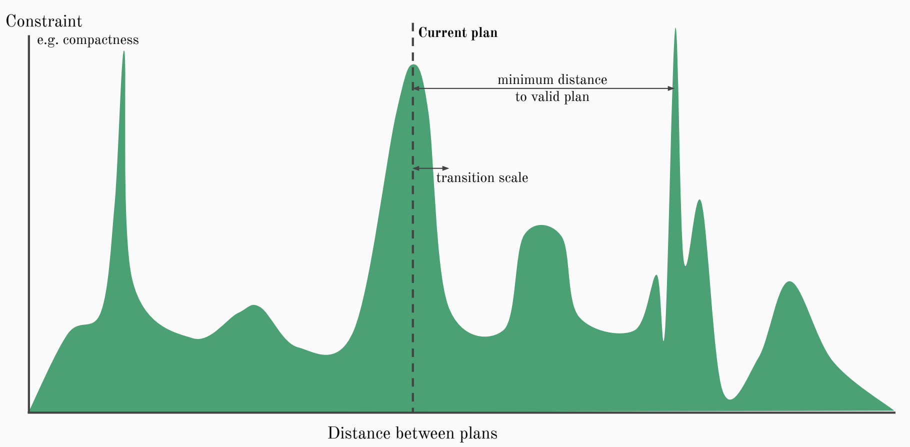
$$\pi(\xi) \propto \underbrace{\tau(\xi)^{\rho}}_{\text{compactness}} \exp(-J(\xi)) \times 1_{\xi \text{ connected}} \times 1_{\text{dev}(\xi) \leq D}$$

where $\tau(\xi)$ counts the product of the number of spanning trees in each district of the plan ξ

- relation with edge removal compactness

$$\tau(\xi)^{\rho} \approx C_1 \exp(-C_2 \rho \text{rem}(\xi)) \quad \text{where} \quad \text{rem}(\xi) = 1 - \frac{\sum_{i=1} |E_i(\xi)|}{|E(G)|}$$

Challenge of MCMC Algorithms



- simulated annealing, parallel tempering ⇝ difficult to apply in practice

Sequential Monte Carlo (SMC) Algorithm (McCartan and Imai, 2020)

- Start with a blank state but in parallel, use the spanning approach to sample a district at a time, resample at each step with weights

- Advantage: unlike MCMC, sampled plans are nearly independent
- Limitation: hard to incorporate plan-wide or region-specific constraints

The Splitting Procedure

1. Generate a uniform spanning tree (Wilson's algorithm)
2. Sort edges by population deviation
3. Sample one edge from top k_i edges and remove it
4. Check population bounds

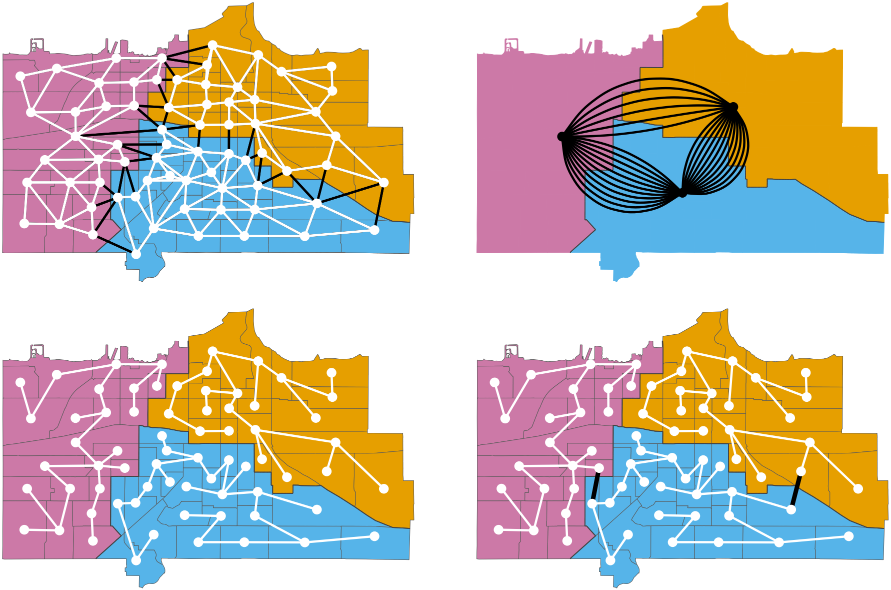
Probability of splitting a new district G_i from $\widetilde{G}_{i-1}$:

$$\frac{\tau(G_i)\tau(\widetilde{G}_i)}{\tau(\widetilde{G}_{i-1})k_i} \underbrace{|\mathcal{C}(G_i, \widetilde{G}_i)|}_{\substack{\text{number of} \\ \text{connecting edges}}}$$

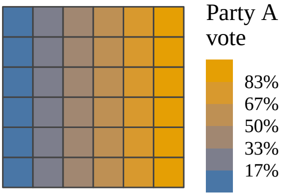
The SMC Algorithm

1. Generate S initial copies of map; set all weights to 1
2. For $i \in \{1, 2, \ldots, n-1\}$:
   a. Until there are S successes
      i. Sample a map according to the weights
      ii. Split off a new district from each sampled map
      iii. Reject if population bounds are not met
   b. Calculate new weights based on splitting probability
3. Output complete plans and compute final weights

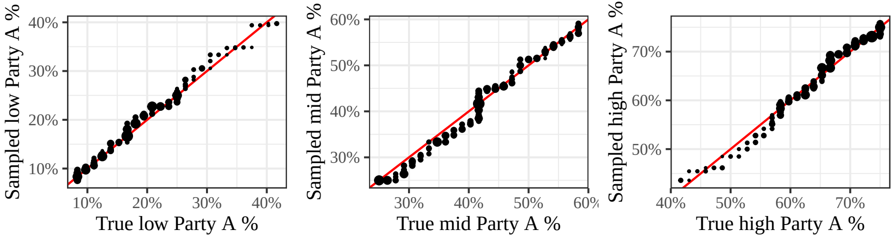
Avoiding County Splits through Quotient Multigraph



Validation



- Divide into 3 districts
- Enumerate all 264,000 possibilities
- Party A vote share: compare simulated with enumerated plans



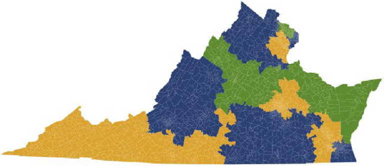
SMC Diagnostics

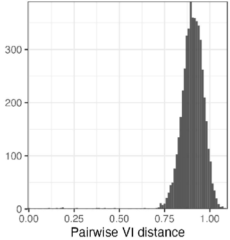
```
SMC: 1,000 sampled plans of 11 districts on 2,465 units
`adapt_k_thresh`=0.985 • `seq_alpha`=0.5
`est_label_mult`=1 • `pop_temper`=0.01

Plan diversity 80% range: 0.82 to 0.98

R-hat values for summary statistics:
pop_overlap        comp         dem       e_dem
     1.0234      1.0112      1.0053      1.0042

Sampling diagnostics for SMC run 1 of 4 (250 samples)
         Eff. samples (%) Acc. rate Log wgt. sd  Max. unique Est. k
Split 1        242 (97.0%)     20.6%        0.36   245 ( 98%)     10
Split 2        240 (95.8%)     31.2%        0.43   193 ( 77%)      6
Split 3        233 (93.4%)     21.8%        0.49   199 ( 80%)      8
Split 4        231 (92.3%)     29.9%        0.56   196 ( 78%)      5
Split 5        219 (87.6%)     36.1%        0.62   195 ( 78%)      3
Split 6        213 (85.0%)     44.9%        0.67   191 ( 76%)      2
Split 7        224 (89.7%)     15.9%        0.59   189 ( 76%)      7
Split 8        227 (90.8%)     24.2%        0.59   192 ( 77%)      4
Split 9        227 (90.9%)     16.9%        0.60   181 ( 72%)      3
Split 10       228 (91.3%)      3.8%        0.58   174 ( 70%)      2
Resample       166 (66.4%)       NA%        0.59   183 ( 73%)     NA
```





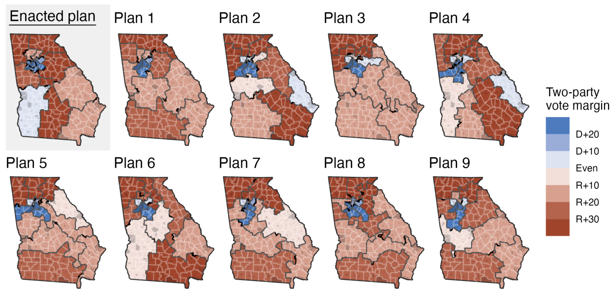
50 State Redistricting Simulations Project

Comprehensive project to simulate alternative congressional redistricting plans for all fifty states.

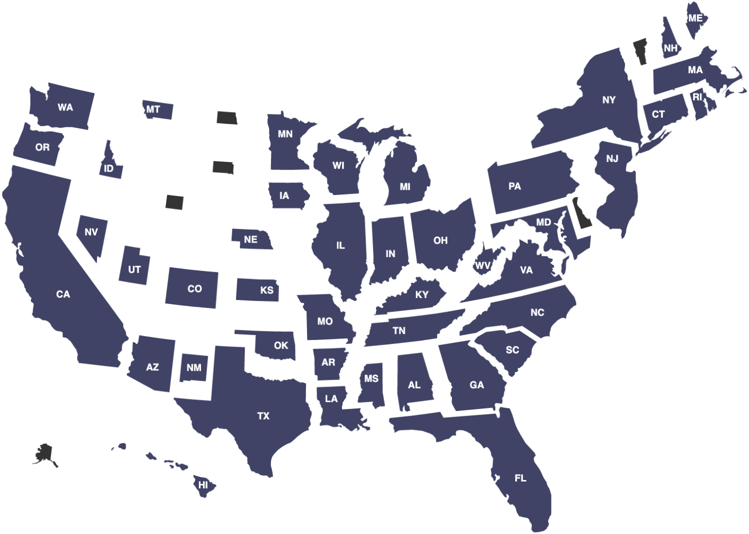
- tidied 2020 Census plus statewide election data from the VEST
- collect state-specific redistricting requirements
- construct algorithmic constraints based on these and traditional redistricting criteria
- 5,000 simulation plans based on SMC
- code and data are available at the Harvard Dataverse

Georgia Example

- 14 Congressional districts
- According to Georgia's House Legislative and Congressional Reapportionment Committee, districts must:
  1. be contiguous
  2. have equal populations
  3. be geographically compact
  4. preserve county and municipality boundaries as much as possible
  5. avoid the unnecessary pairing of incumbents
- We attempted to account for everything except the last one
- Assumption about voting rights act (VRA) compliance



Check out https://alarm-redist.org/fifty-states/



States colored blue have enacted a congressional map and been fully analyzed, states colored gray have enacted a plan but haven't yet been analyzed, or just have a single district (and hence no redistricting), and states colored red haven't enacted a plan yet.

Electoral Modeling

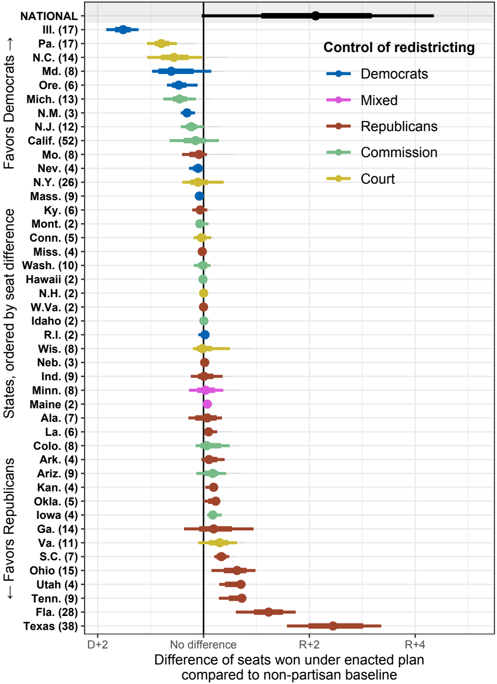
- Precinct-level data from the 2016 and 2020 presidential elections
- Average of the two elections $\rightsquigarrow$ baseline partisanship

$$\hat{D}_j = \underbrace{\frac{1}{2}\left(\frac{D_{16j}}{D_{16j}+R_{16j}} + \frac{D_{20j}}{D_{20j}+R_{20j}}\right)}_{\text{average Democratic vote share}} \times \underbrace{\sqrt{(D_{16j}+R_{16j})(D_{20j}+R_{20j})}}_{\text{(geometric) average turnout}}$$

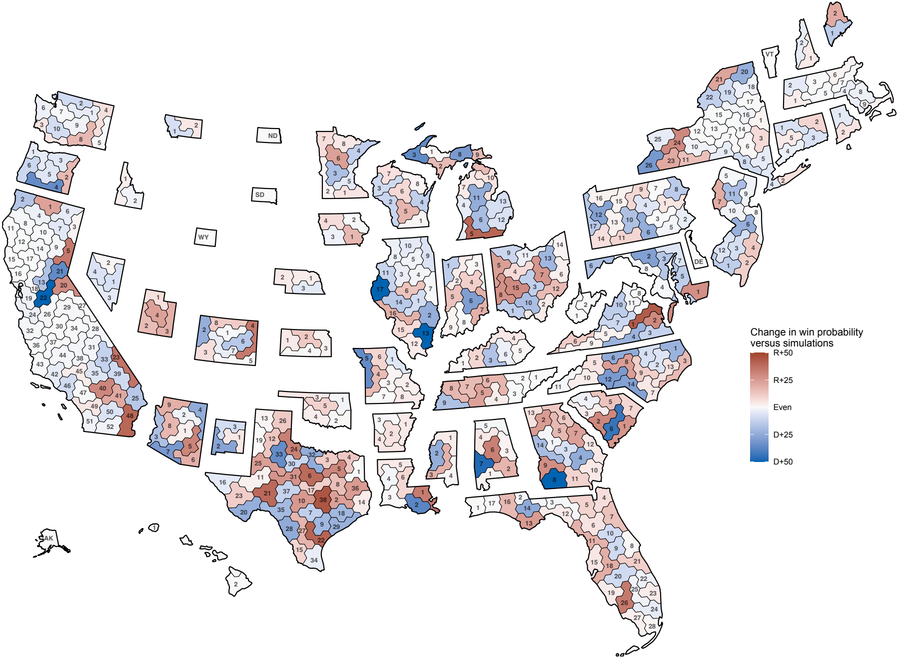
- Model for the Democratic vote share:

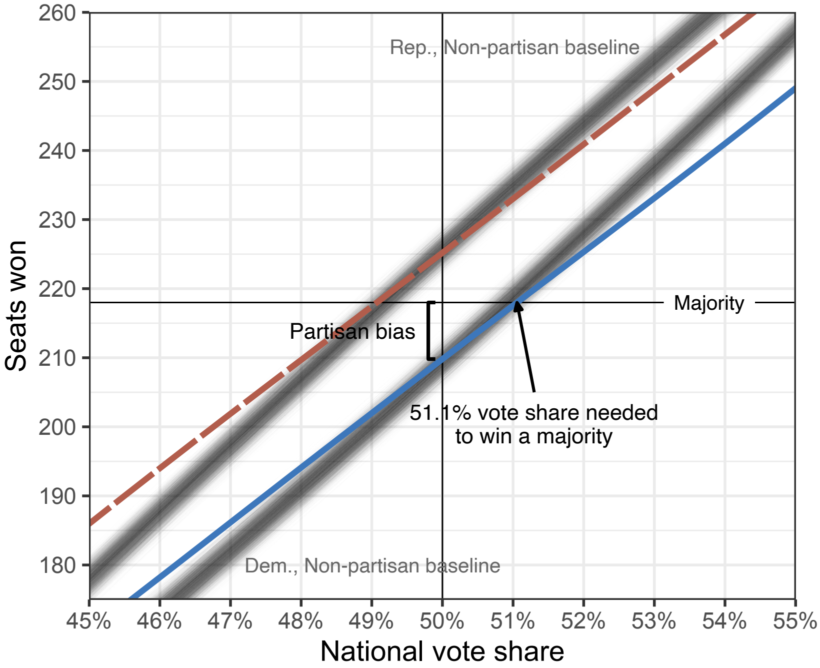
$$\begin{aligned} \operatorname{logit}(y_{it}) &= \alpha_i + \beta_t + \varepsilon_{it}, \\ \beta_t &\overset{iid}{\sim} \mathcal{N}(0, \sigma_\beta^2), \\ \varepsilon_{it} &\overset{iid}{\sim} \mathrm{t}_\nu(0, \sigma_\varepsilon^2), \end{aligned}$$

  where we fix α_i to the baseline Democratic vote share and $(\sigma_\beta^2, \sigma_\epsilon^2, \nu)$ are estimated using the historical House elections since 1976
- We then compute α_i and simulate $(\beta_t, \varepsilon_{it})$ for each district under a given (enacted or simulated) redistricting plan



NATIONAL
Ill. (17)
Pa. (17)
N.C. (14)
Md. (8)
Ore. (6)
Mich. (13)
N.M. (3)
N.J. (12)
Calif. (52)
Mo. (8)
Nev. (4)
N.Y. (26)
Mass. (9)
Ky. (6)
Mont. (2)
Conn. (5)
Miss. (4)
Wash. (10)
Hawaii (2)
N.H. (2)
W.Va. (2)
Idaho (2)
R.I. (2)
Wis. (8)
Neb. (3)
Ind. (9)
Minn. (8)
Maine (2)
Ala. (7)
La. (6)
Colo. (8)
Ark. (4)
Ariz. (9)
Kan. (4)
Okla. (5)
Iowa (4)
Ga. (14)
Va. (11)
S.C. (7)
Ohio (15)
Utah (4)
Tenn. (9)
Fla. (28)
Texas (38)
← Favors Republicans States, ordered by seat difference Favors Democrats →
Control of redistricting
Democrats
Mixed
Republicans
Commission
Court
D+2
No difference
R+2
R+4
Difference of seats won under enacted plan compared to non-partisan baseline


ND
SD
WY
VT
DE
AK
Change in win probability versus simulations
R+50
R+25
Even
D+25
D+50


Seats won
National vote share
260
250
240
230
220
210
200
190
180
45%
46%
47%
48%
49%
50%
51%
52%
53%
54%
55%
Rep., Non-partisan baseline
Dem., Non-partisan baseline
Majority
Partisan bias
51.1% vote share needed
to win a majority

Application: Ohio Congressional Redistricting

- Currently 16 districts: 4 Democrats and 12 Republicans
  - 2020 President: Biden 45%, Trump 53%
  - 2018 Senate: Brown 53%, Renacci 47%

- After 2020 Census, the number of seats is reduced to 15 districts

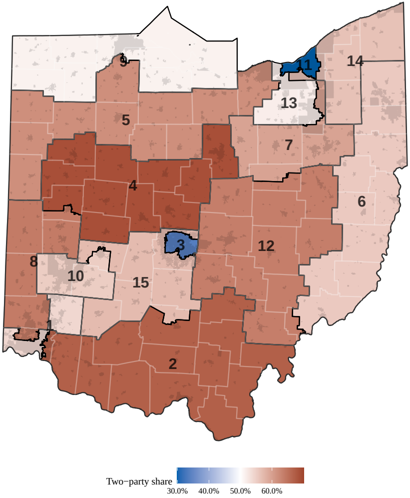
- 2018 Ohio voters passed the constitutional amendment
  - bipartisan support leads to a 10 year map
  - if that fails, it becomes a 4 year map

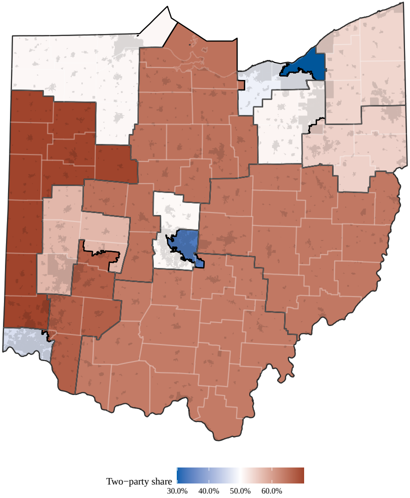
- Redistricting
  - State Senate and House approved the initial map
  - No bipartisan support $\rightsquigarrow$ 4 year map
  - November 20: Governor DeWine signed the map

League of Women Voters of Ohio *et al. v.* Ohio Redistricting Commission, *et al.*

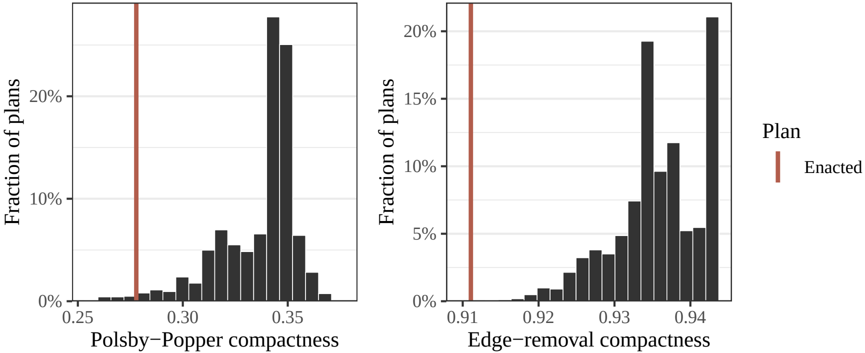
- I served as an expert witness for Relators
- Simulation analysis based on Sequential Monte Carlo algorithm
  - 5,000 alternative plans
  - contiguous and compact districts
  - compliant with the Voting Rights Act (Cleveland)
  - several complicated splitting constraints
  - Section 2(B)(5): out of Ohio's 88 counties,
    - at least 65 counties should not be split
    - no more than 18 counties can be split no more than once
    - no more than 5 counties can be split no more than twice

The Enacted and Example Simulated Plans



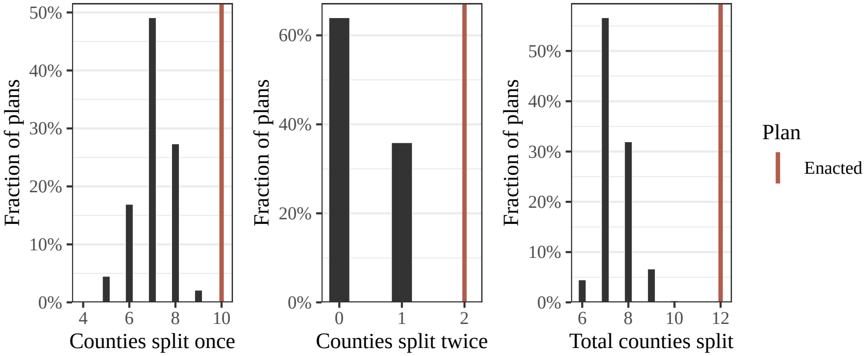


Compactness

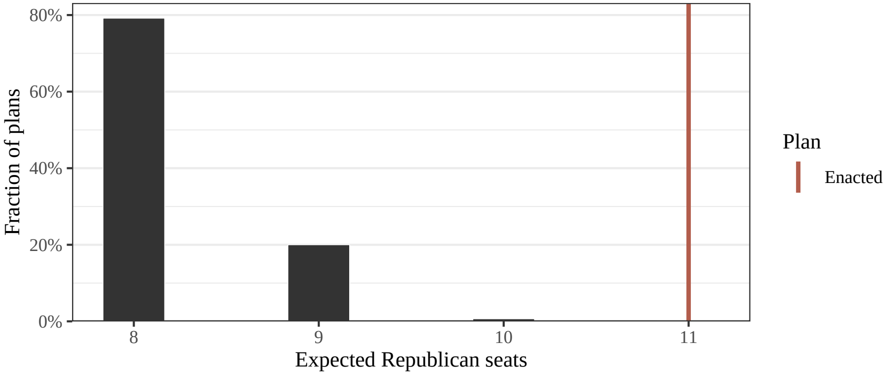


- Polsby-Popper: the ratio of the district area to the area of a circle with the same perimeter
- Edge-removal

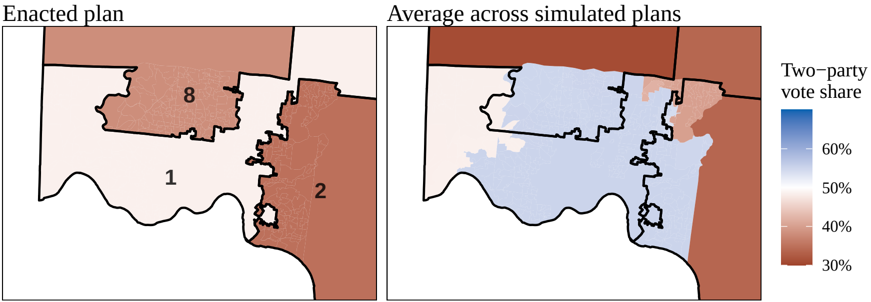
Administrative Boundary Splits



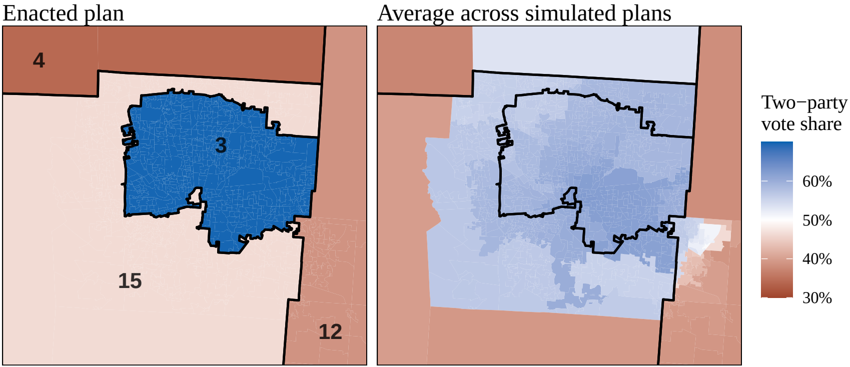
Expected Number of Republican Seats



Hamilton County: Cincinnati Area



Franklin County: Columbus Area



Ohio Supreme Court Strikes Down the Enacted Map

Id. at Section 1(C)(3)(a). The above evidence, particularly Dr. Imai's conclusion that the enacted plan will result in, on average, 2.8 more Republican seats than are warranted, shows that the General Assembly's decision to shift what could have been—under a neutral application of Article XIX—Democratic-leaning areas into competitive districts, i.e., districts that give the Republican Party's candidates a better chance of winning than they would otherwise have had in a more compactly drawn district, resulted in a plan that unduly favors the Republican Party and unduly disfavors the Democratic Party.

Concluding Remarks

- Redistricting matters
  - fair representation and policy outcomes
  - competitiveness of districts and responsiveness
  - political polarization
- How should we stop gerrymandering?
  - independent commission (e.g., Michigan)
  - use of algorithms to detect gerrymandering
- Roles of experts
  - legislative process
  - court testimony
- Open problems
  - large-scale redistricting problems (e.g., state legislatures)
  - redistricting plans based on Census blocks
  - algorithm-generated redistricting plan proposals