Using Algorithms to Detect Gerrymandering and Improve Legislative Redistricting:

Cases from the United States and Japan

Kosuke Imai

Harvard University

Seminar at University College Dublin

November 5, 2024

Joint work with Christopher Kenny, Cory McCartan, Tyler Simko, Shiro Kuriwaki, George Garcia III, Sho Miyazaki, Kento Yamada, Kevin Wang, and Melissa Wu

Who Is Going To Win?

I Cannot Break a Statistical Tie, But American People Will

UPDATED NOV. 5, 2024, AT 6:00 AM

Who Is Favored To Win The 2024 Presidential Election?

538 uses polling, economic and demographic data to explore likely election outcomes.

Harris wins 50 times out of 100

in our simulations of the 2024 presidential election.

Trump wins 49 times out of 100.

There is a less than 1-in-100 chance of no Electoral College winner.

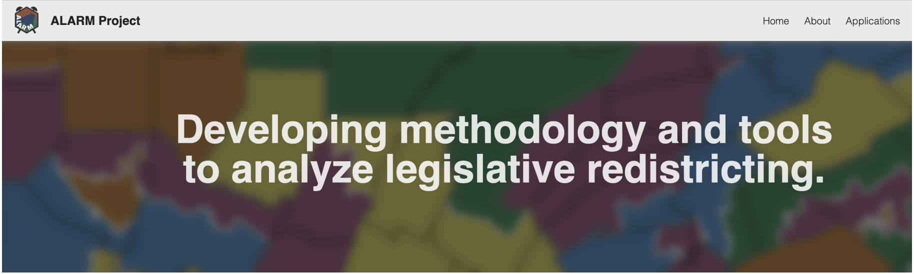
Motivation

- 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 real 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
- Technological solution to detecting gerrymandering
- Tool for analyzing redistricting

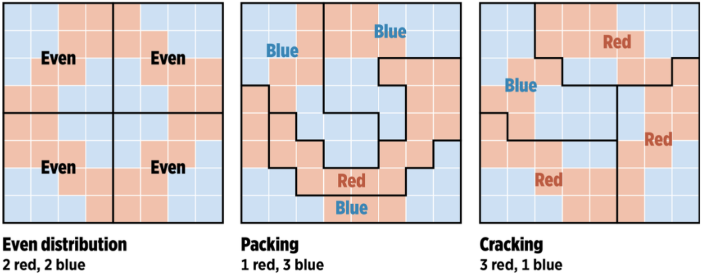
ALgorithm-Assisted Redistricting Methodology (ALARM)



- 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 with powerful tools

Redistricting Basics

- Classic gerrymandering strategies: packing and cracking



- What has changed:
  - availability of granular data
  - mapping software (e.g., Maptitude, Dave's Redistricting app)
- US Congressional redistricting
  - racial gerrymandering: *Allen v. Milligan*
  - partisan gerrymandering: *Rucho v. Common Cause*

Why Use Simulation Algorithm for Redistricting Evaluation?

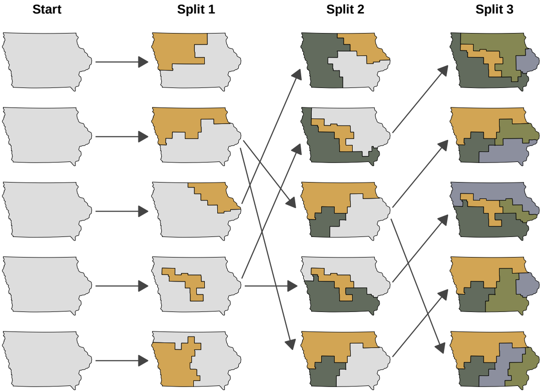
- Traditional redistricting evaluation
  1. compute various fairness metrics
  2. compare them across states and over time
- Confounded by differences in political geography and redistricting rules

- 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

Sequential Monte Carlo (SMC) Algorithm (McCartan and Imai, 2023)

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

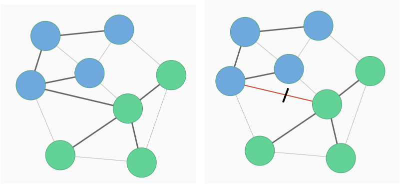


- Advantage: scalable due to parallel structure, less dependent draws
- Limitation: hard to incorporate plan-wide or region-specific constraints

The SMC Algorithm

- Splitting off a district using a spanning tree

  1. random generation of spanning trees (Wilson's algorithm)
  2. computing the number of spanning trees



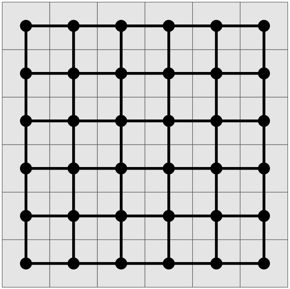
- Target distribution:

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

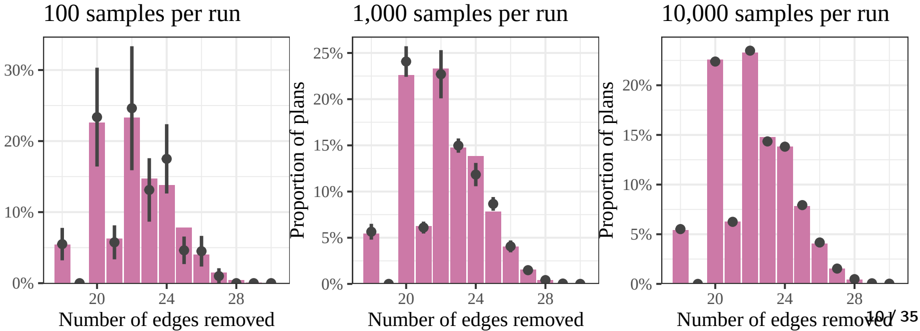
for a given plan ξ where

$$\tau(\xi)^{\rho} = \left[\prod_{i=1}^{n} \tau(G_i(\xi))\right]^{\rho} \approx C_1 \exp(-C_2 \ \rho \quad \underbrace{\text{rem}(\xi)}_{\text{fraction of edges removed}} \quad)$$

Validation



- Divide a 6×6 grid into 6 equal-sized districts
- Enumerate 451,206 plans (out of 356 billion)
- Number of edge removed as a target statistic



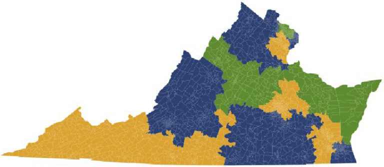
- Like any algorithm, SMC needs to be used with care
- Large redistricting problems and complex constraints are difficult

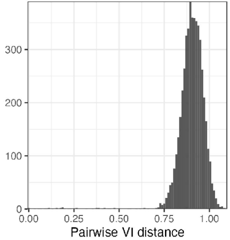
```
SMC: 1,000 sampled plans of 11 districts on 2,465 units
`adapt_k_thresh`=0.985 • `seq_alpha`=0.5
`est_label_mult`=1 • `pop_temper`=0.01

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
```





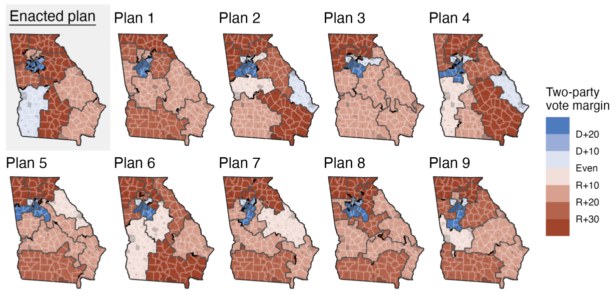
50 State Redistricting Simulations Project

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

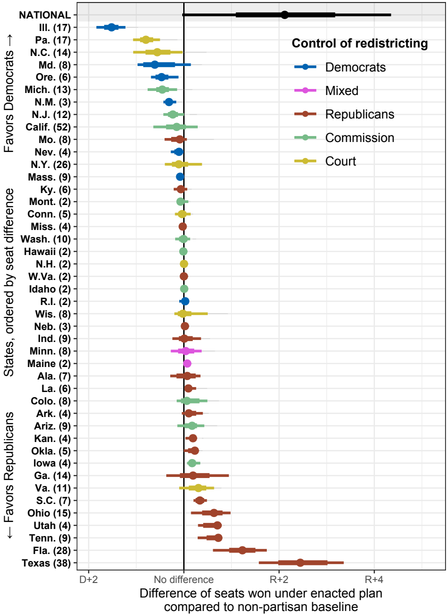
- tidied 2020 Census plus statewide election data from the VEST
- collect state-specific redistricting requirements
- construct algorithmic constraints based on these and traditional redistricting criteria
- 5,000 simulation plans based on SMC
- code and data are available at the Harvard Dataverse

Georgia Example

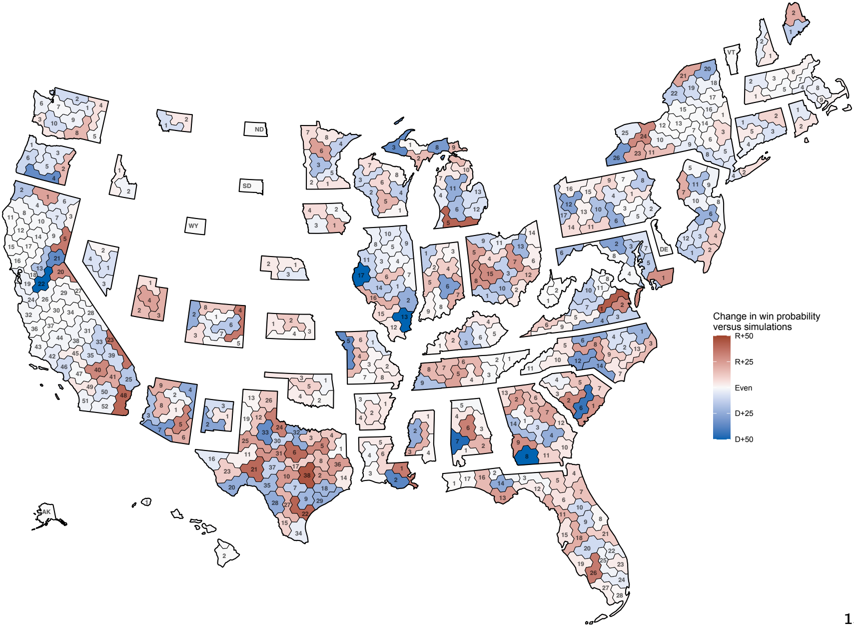
- 14 Congressional districts
- According to Georgia's House Legislative and Congressional Reapportionment Committee, districts must:
  1. be contiguous
  2. have equal populations
  3. be geographically compact
  4. preserve county and municipality boundaries as much as possible
  5. avoid the unnecessary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



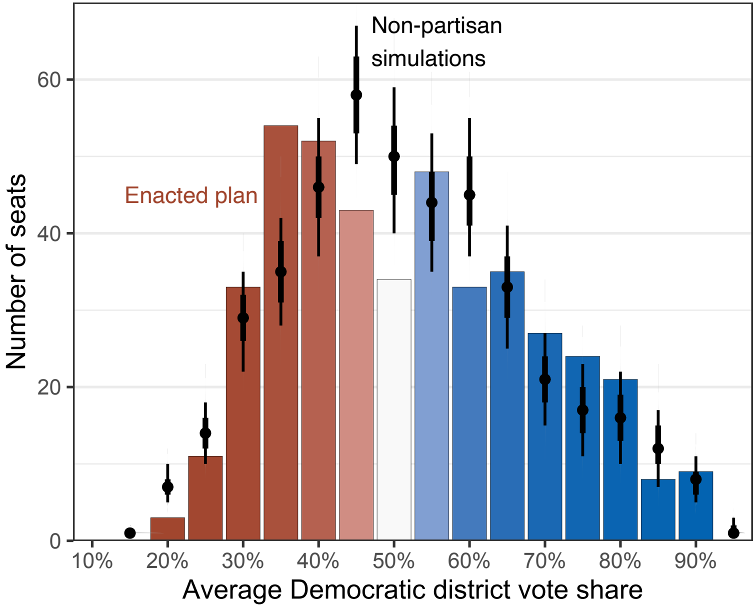
Widespread Partisan Gerrymandering Cancels Nationally



Map of Partisan Gerrymandering



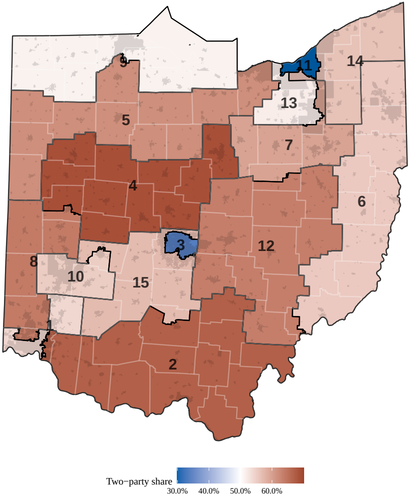
Partisan Gerrymandering Reduces Competitiveness

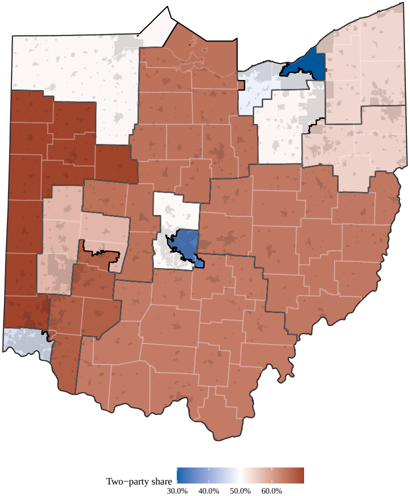


Application in the Court: Ohio Congressional Redistricting

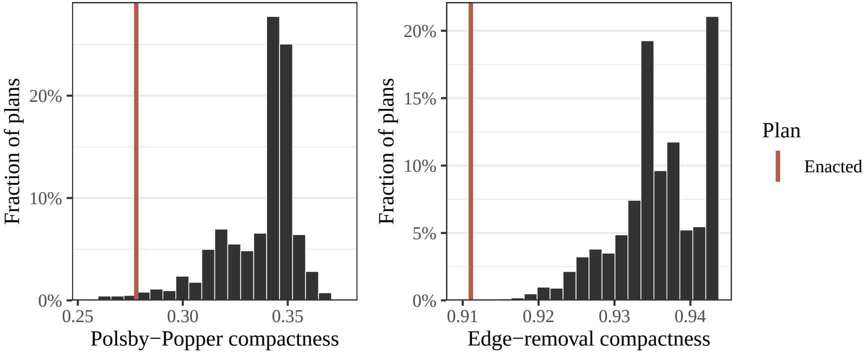
- Currently 16 districts: 4 Democrats and 12 Republicans
- After 2020 Census, the number of seats is reduced to 15 districts
- 2018 Ohio voters passed the constitutional amendment
- I served as an expert witness for Relators: *League of Women Voters of Ohio et al. v. Ohio Redistricting Commission, et al.*
- Simulation analysis
  - 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



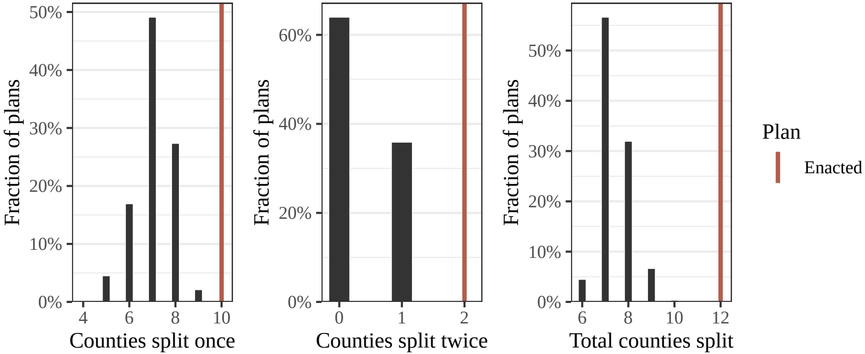


Compactness

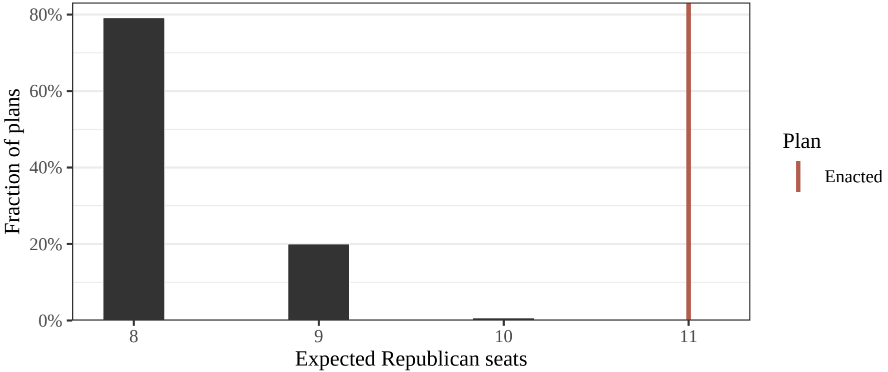


- Polsby-Popper: the ratio of the district area to the area of a circle with the same perimeter
- Edge-removal

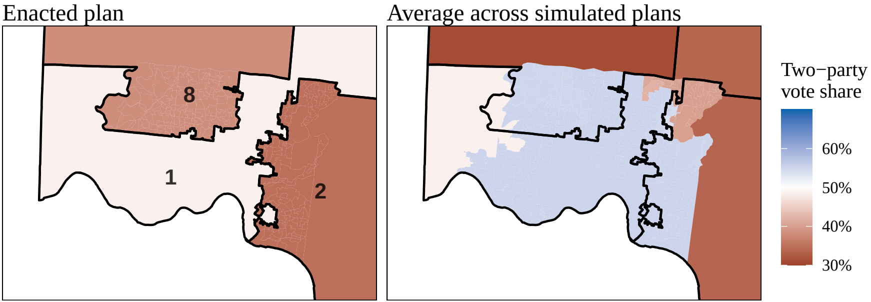
Administrative Boundary Splits



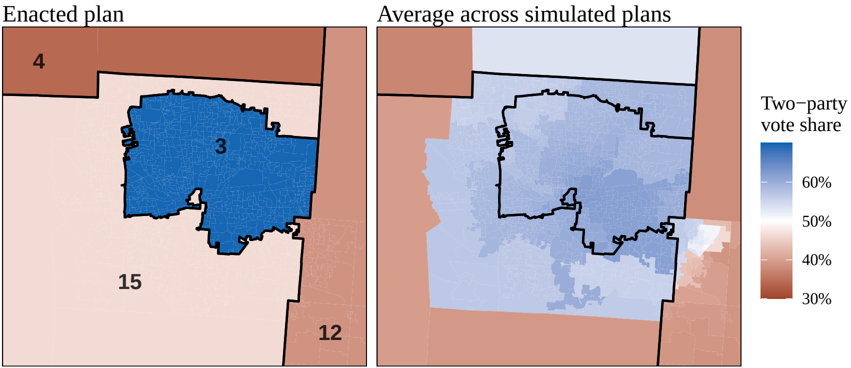
Expected Number of Republican Seats



Cracking: Hamilton County (Cincinnati Area)



Packing: Franklin County (Columbus Area)



Ohio Supreme Court Strikes Down the Enacted Map

The Court Opinion

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.

Supreme Court: *Alexander v. NAACP et al.*

- South Carolina racial gerrymandering case
- Served as an expert witness for the plaintiffs
- Used simulation to provide evidence that a disproportionately large number of Black voters are packed into District 6

- Federal district court ruled in favor of NAACP
- South Carolina directly appealed to the Supreme Court

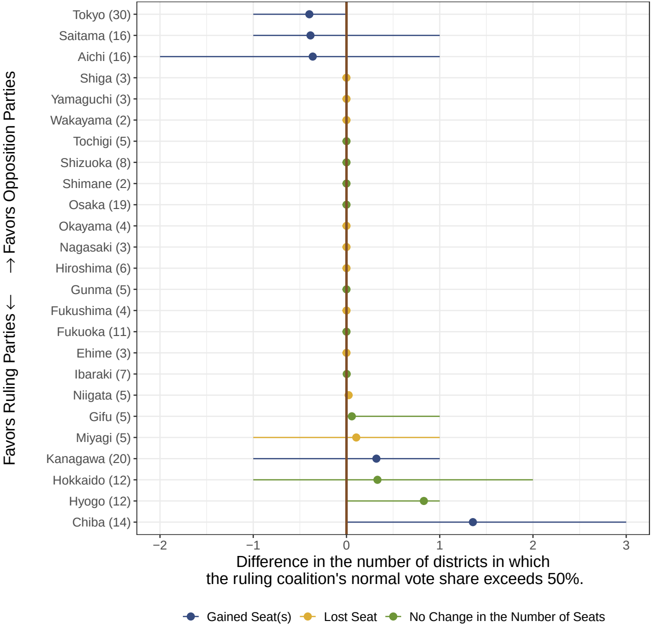
- The Supreme Court reversed the federal district court's decision:

  The Challengers' inference is flawed because Dr. Imai's models failed to consider partisanship. [...] Dr. Imai's algorithm produced maps without requiring that District 1 comply with the legislature's asserted aim of ensuring that District 1 remain a relatively safe Republican seat.

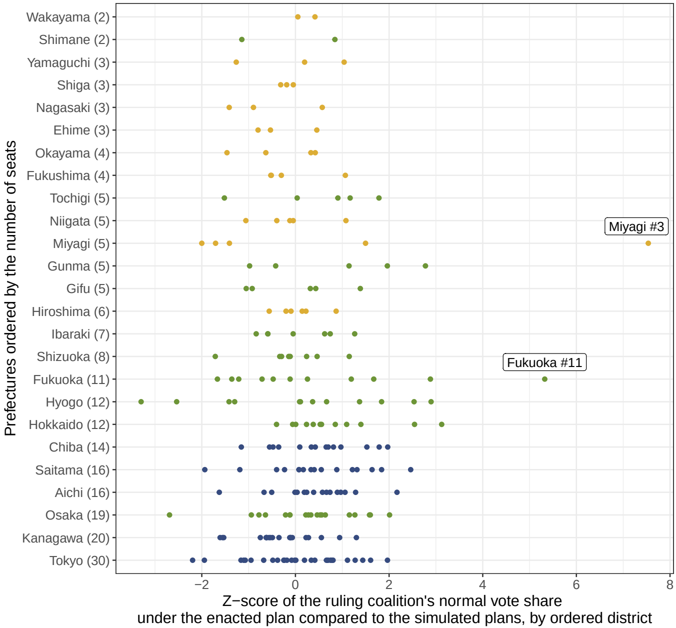
Is There Partisan Bias in Japanese Redistricting?

- Non-partisan commission $\rightsquigarrow$ no partisan bias?
- Potential sources of partisan bias
  - members are appointed by the prime minister and approved by the Diet
  - governors are invited to provide their opinions
- 2020 Japanese redistricting
  - redistricting in 25 prefectures out of 47
  - 10 prefectures lost a seat
  - 5 prefectures gained a seat / seats
  - 5 prefectures redrew districts without changing the number of seats

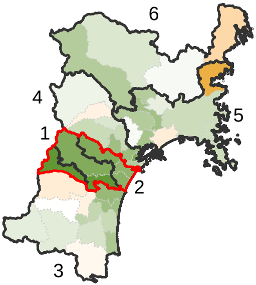
Little Partisan Bias at the Prefecture Level



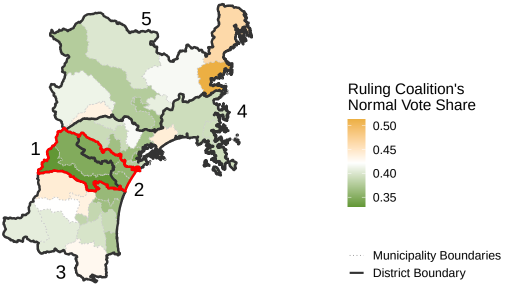
Some but Relatively Little Partisan Bias at the District Level



Redistricting in Miyagi Prefecture



(a) Enacted Plan (2017)

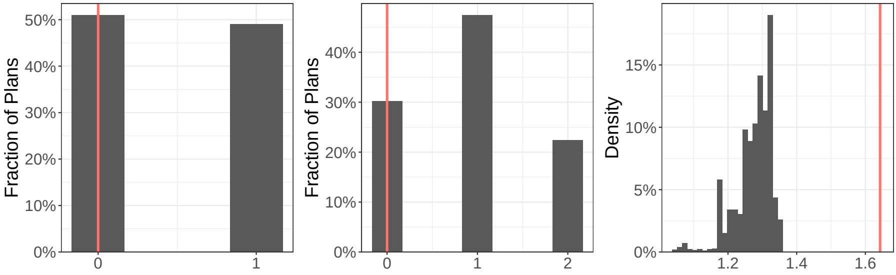


(b) Enacted Plan (2022)

- # of seats: 6 → 5
- # of municipality splits: 2 → 0
- # of county splits: 2 → 0
- Pop. deviation: 1.94 → 1.64

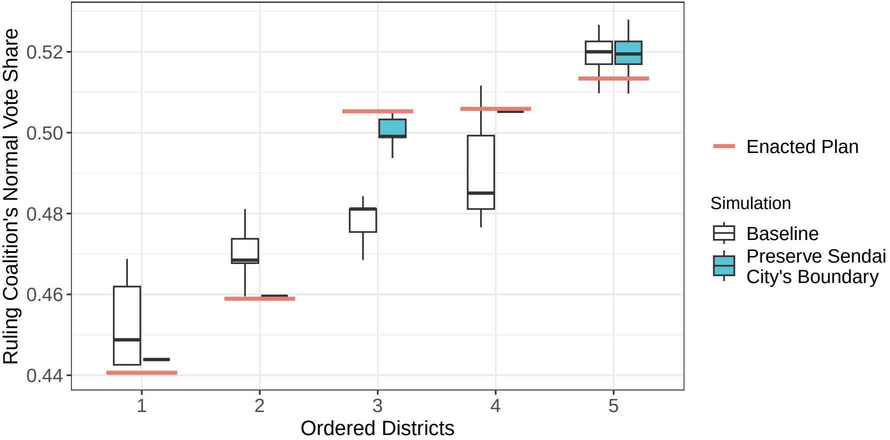
- Lower House Electoral results
  - 2017: 5 LDP, 1 independent
  - 2021: 4 LDP, 2 opposition
- 5,000 simulated plans

Simulated Plans Have Good Properties

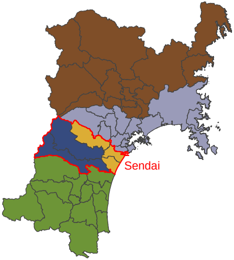


(a) Municipality Splits (b) County Splits (c) Intra-prefecture Max-min Ratio

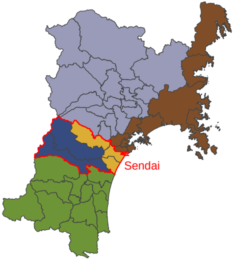
Bias Is Mostly Caused by Preservation of City Boundary



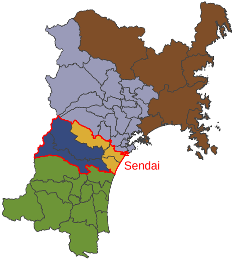
All Plans Considered by the Commission



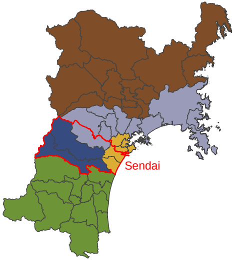
(a) Plan A (Enacted plan)



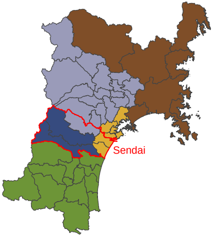
(b) B (MIC official)



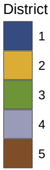
(c) C (MIC official)



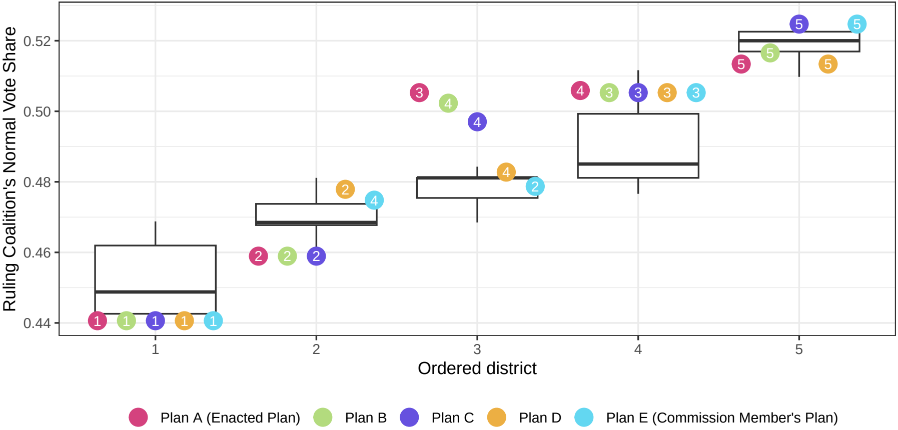
(d) D (MIC official)



(e) E (Comm. member)



Evaluation of Enacted and Draft Plans



Concluding Remarks

- Redistricting matters
  - fair representation and policy outcomes
  - competitiveness of districts and responsiveness
  - political polarization
  - state and local offices, education districts, non-US contexts
- How should we stop gerrymandering?
  - independent commission (e.g., Michigan)
  - use of algorithms to detect gerrymandering
- Role of experts
  - legislative process
  - court testimony
  - work with non-partisan groups and commissions
- Open problems
  - large-scale redistricting problems (e.g., state legislatures)
  - algorithm-generated redistricting plan proposals
  - communities of interest, impact of redistricting rules