# Èdouard Lucas:

*The theory of recurrent sequences is an inexhaustible mine which contains all the properties of numbers; by calculating the successive terms of such sequences, decomposing them into their prime factors and seeking out by experimentation the laws of appearance and reproduction of the prime numbers, one can advance in a systematic manner the study of the properties of numbers and their application to all branches of mathematics.*

# Computational Approaches for Political Redistricting Part I: Geospatial Data

Daryl DeFord

CSAIL – GDP Group

IAP 2019
Massachusetts Institute of Technology
January 8, 2019

# MORAL:

Computational Redistricting is NOT a solved problem!

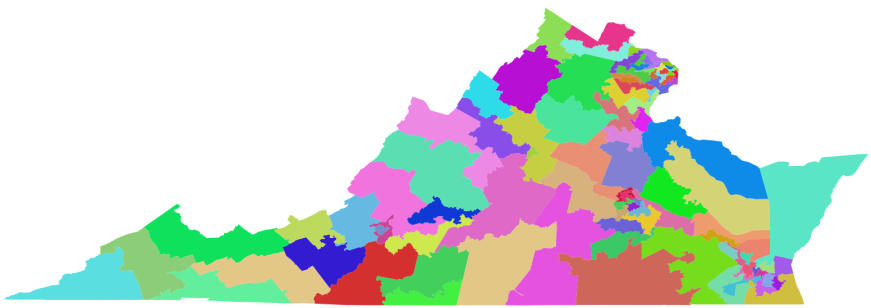
# Advertisements

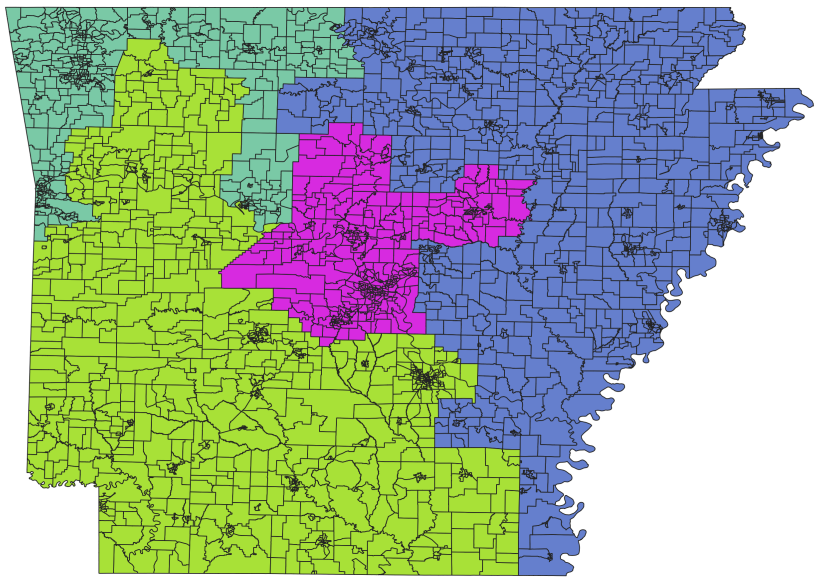
1. VRDI – 6 week summer program for graduate and undergraduate students (Deadline 2/1)
   - Application: tinyurl.com/apply-vrdi-2
   - Information: gerrydata.org
2. Contact:
   - Email: ddeford at mit.edu
   - Website: mggg.org
   - Slack channel: GerryChat.slack.com
3. Research Projects
   - Math Problems: tinyurl.com/gerryprojects
   - Data Problems: tinyurl.com/GerryChainProjects
4. IAP Info:
   - Resources: people.csail.mit.edu/ddeford/CAtPR
   - Today 12-1 More emphasis on building state level data sets
   - Thursday 12-1 MCMC and GerryChain
   - 1/22 12-1 Graph Partitions
   - 1/29 12-1 In–depth state examples

# Example: Pennsylvania Ensemble

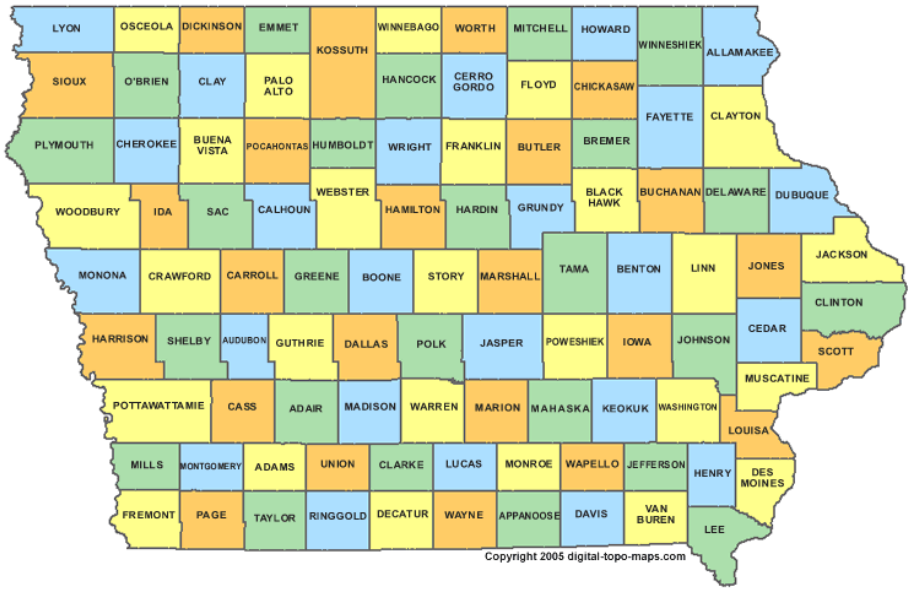
# Outline

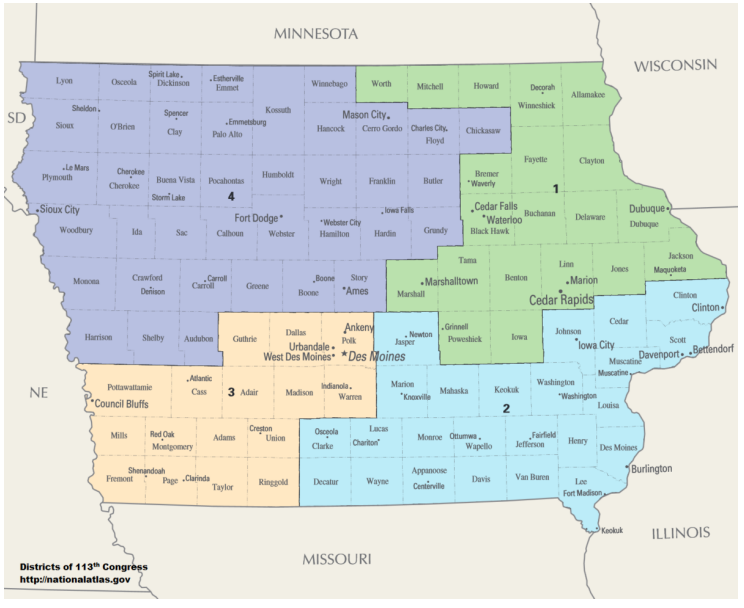
# Electoral Districts



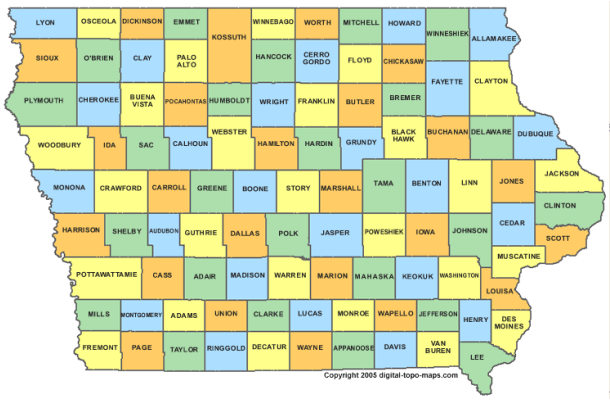


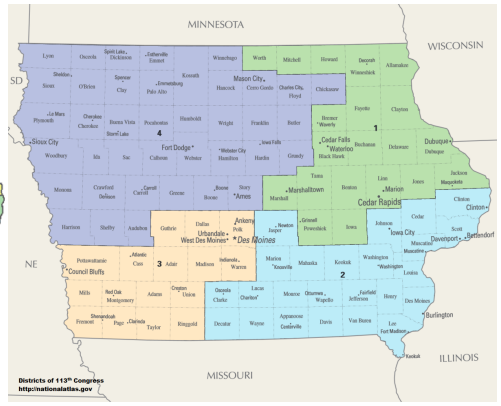
# Example: Iowa



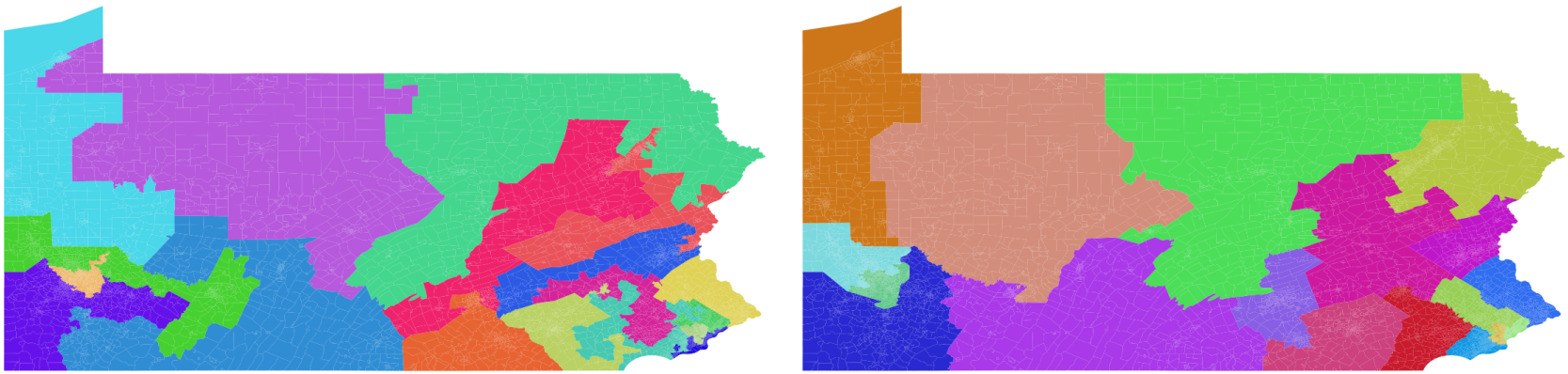
# Example: Iowa




- 4 Congressional Districts, 100 House Districts, 50 Senate Districts
- House districts nest into Senate districts
- Congressional districts made out of counties
- Independent committee with legislative approval
- No partisan data allowed

# Example: Pennsylvania



- 18 Congressional Districts, 203 House Districts, 50 Senate Districts
- Zero–balanced population
- Legislature draws congressional districts committee draws legislative districts
- Partisan behavior allowed

# MORAL:

Computational Redistricting is NOT a solved problem!

# Common Rules

**Example (Common Redistricting Requirements)**

- Population Balance
- Contiguity
- Compactness
- Communities of Interest
- VRA Compliance
- Municipal Boundaries
- ...

# Recent Legislation

- Very Process Focused
- Too specific
- Not specific enough
- Exploitable
- Doesn't solve data problems

# Recent Legislation

- Very Process Focused
- Too specific
- Not specific enough
- Exploitable
- Doesn't solve data problems

| | Michigan | Utah | Colorado | Missouri |
|---|---|---|---|---|
| 1 | VRA + POP | VRA + POP | POP ($<$ 5%) + Contiguous | POP |
| 2 | Contiguous | Municipalities | VRA | VRA |
| 3 | Communities of interest | Compactness | COI + Municipal | Representatives of choice |
| 4 | Partisan fairness | Contiguity | Compactness | Partisan fairness + competitiveness |
| 5 | No incumbency protection | Communities of interest | Competitiveness | Contiguity |
| 6 | Municipal boundaries | Natural boundaries | | Municipal boundaries |
| 7 | Compactness | Align boundaries | | Compactness |

# Example: Competitiveness

**Example (What is a competitive district/plan?)**

# Example: Competitiveness

**Example (What is a competitive district/plan?)**

?

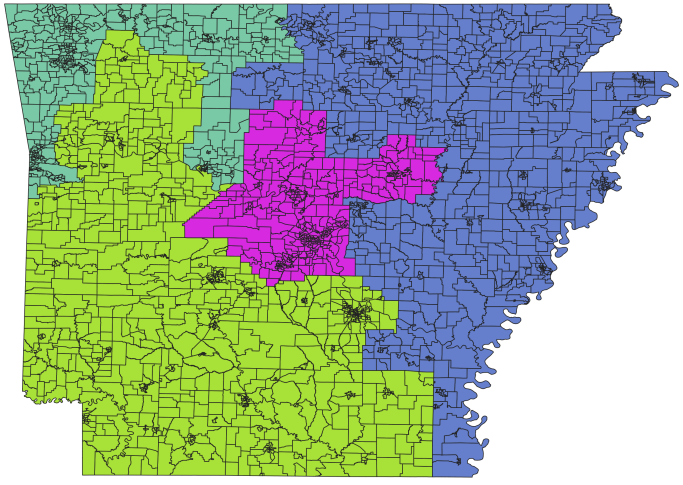
# Example: Competitiveness

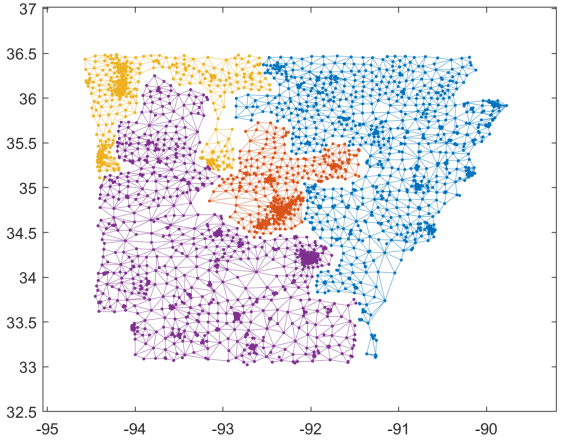
**Example (What is a competitive district/plan?)**

?

- Arizona
- Missouri
- New Jersey (almost)
- ...

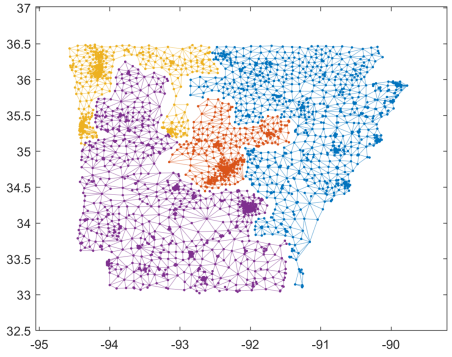
# Mathematical Formalism

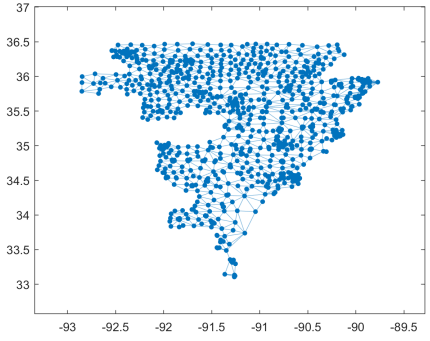


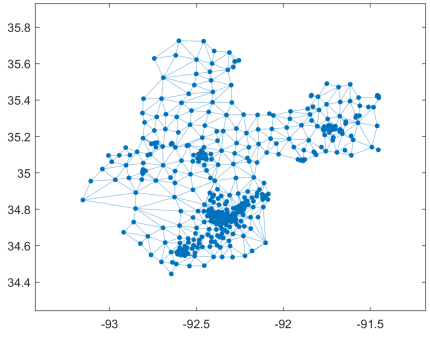


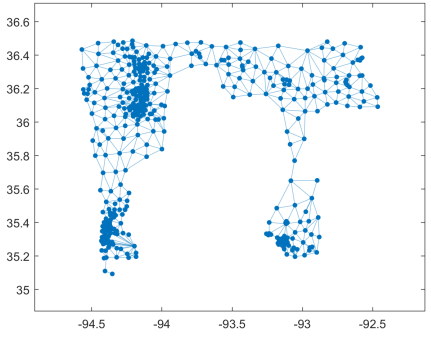

In order to study this problem mathematically we need to abstract the process of districting into the realm of mathematical objects. The first step is to discretize!

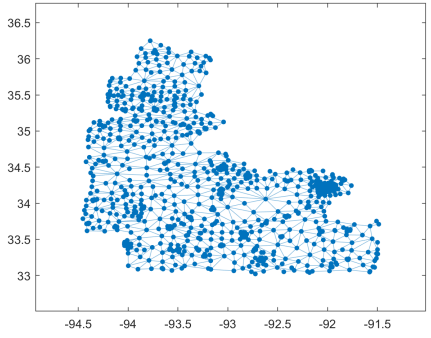
# Graph Partitioning



# Other Measures

- Choice of units
- Population Balance
- Contiguity
- Compactness
- Communities of Interest
- VRA Compliance
- Municipal Boundaries
- ...

# MORAL:

Computational Redistricting is NOT a solved problem!

# What is gerrymandering?

**Example (What is gerrymandering?)**

# What is gerrymandering?

**Example (What is gerrymandering?)**

?

# What is gerrymandering?

**Example (What is gerrymandering?)**

?

**Example (Who is harmed by gerrymandering?)**

# What is gerrymandering?

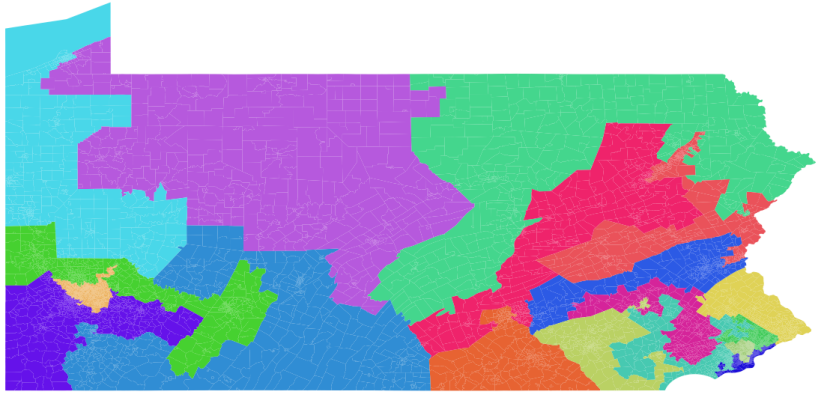
**Example (What is gerrymandering?)**

?

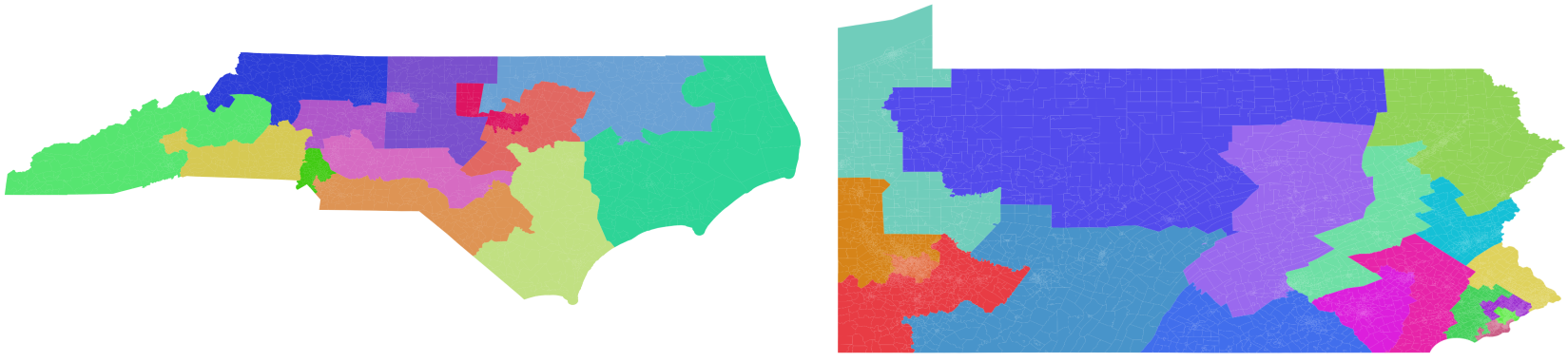
**Example (Who is harmed by gerrymandering?)**

- Partisan
- Racial
- Incumbents
- ...

# Ugly Shapes



# Partisan Imbalance



# Partisan Fairness

- MA
  - Duchin et al. (2018) Locating the representational baseline: Republicans in Massachusetts arXiv:1810.09051
  - Not all partisan outcomes are possible, given discretization
- MD
  - Two recent preprints claiming not gerrymandered
  - Court ruled one district unconstitutional
- NJ
  - Controversial constitutional amendment
  - Competitiveness defined in terms of historical statewide averaging

# Compactness and Partisan Measures

- Compactness
  - Polsby–Popper
  - Reock
  - Total perimeter
  - Convex hull
  - Discrete metrics
  - ...
- Partisan Imbalance
  - Mean–Median
  - Partisan Bias
  - Efficiency Gap
  - Proportionality
  - ...

# MORAL:

Computational Redistricting is **NOT** a solved problem!

# All hope is not lost...

# All hope is not lost...

- The wide variety in rules applied to districting problems (even in the same state) means that any single measure of gerrymandering will be insufficient/exploitable

# All hope is not lost...

- The wide variety in rules applied to districting problems (even in the same state) means that any single measure of gerrymandering will be insufficient/exploitable
- Instead we want to do **outlier analysis** by comparing to large ensembles of other feasible plans.

# All hope is not lost...

- The wide variety in rules applied to districting problems (even in the same state) means that any single measure of gerrymandering will be insufficient/exploitable
- Instead we want to do **outlier analysis** by comparing to large ensembles of other feasible plans.
- This allows us to understand the impacts of the underlying political and demographic geography on a wide collection of metrics.

# Arkansas Outlier Example

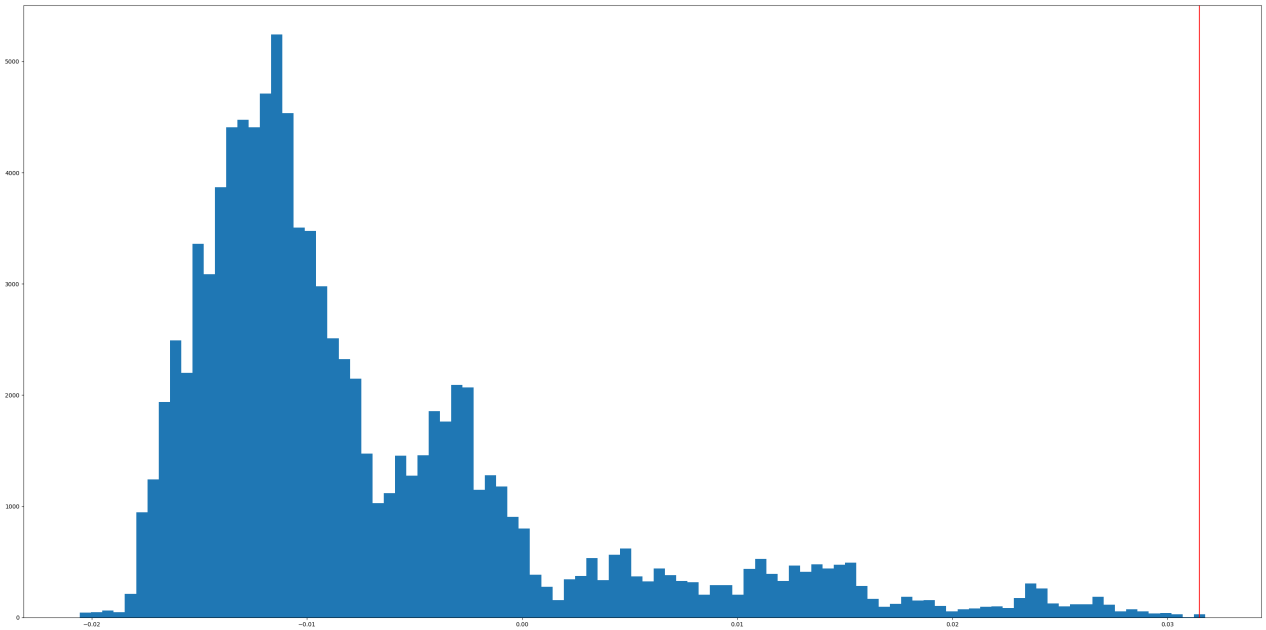
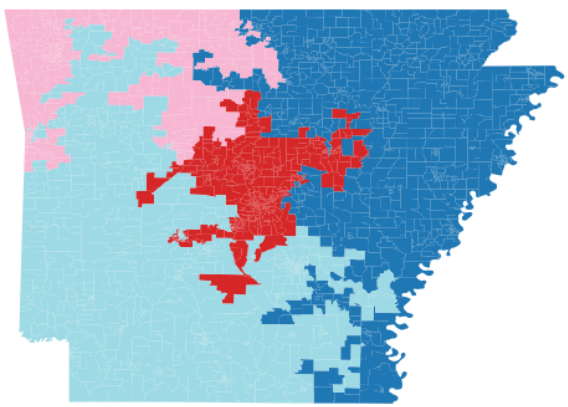
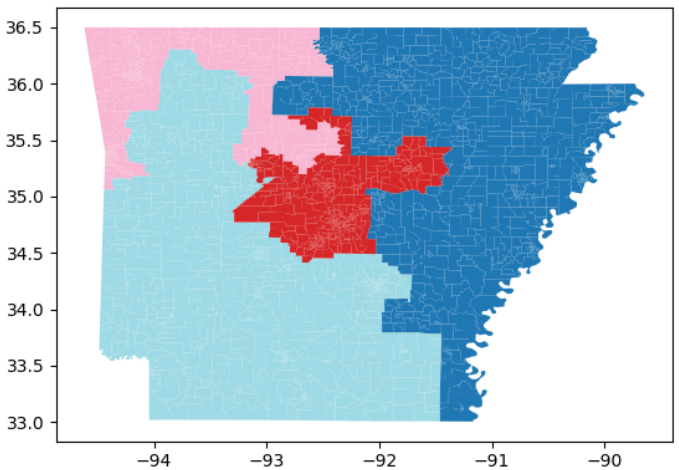


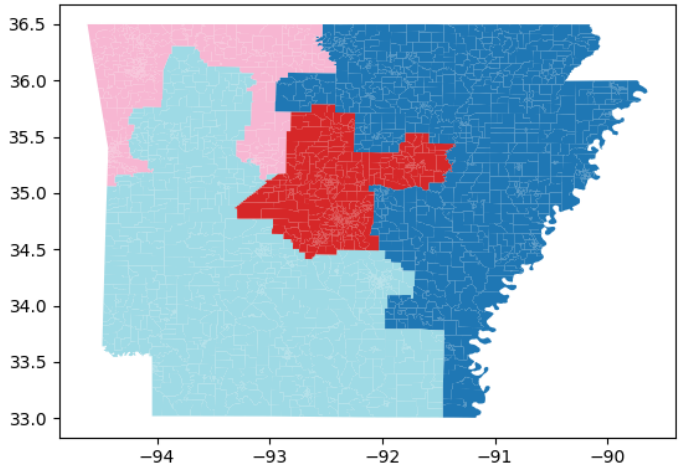

Figure: Mean–Median score using senate 2016 election data on 1,000,000 plans.

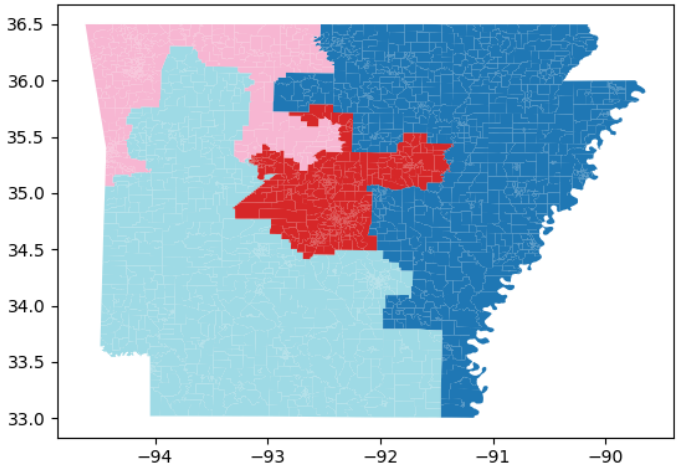
# Which ensembles?



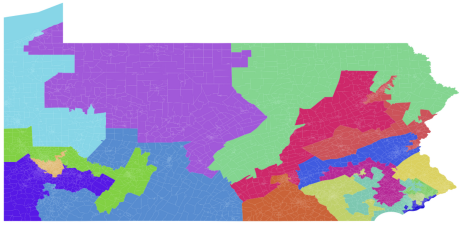


# Arkansas Tree Ensembles

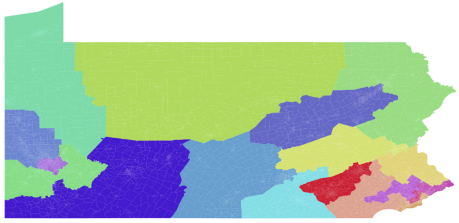


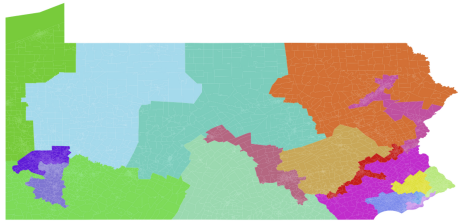
# Pennsylvania Landscapes



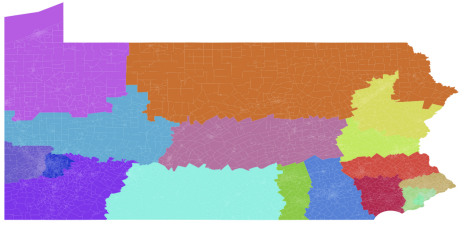
2011



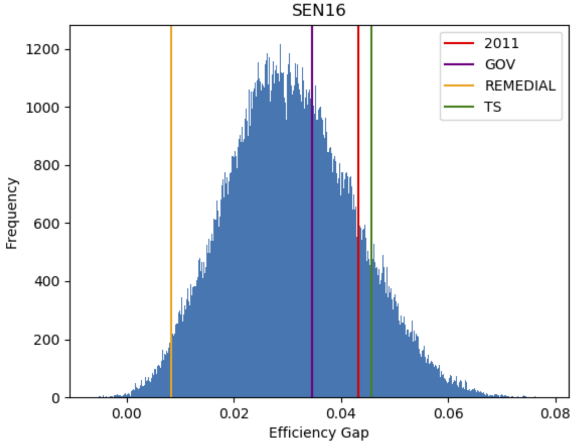
538 GOP

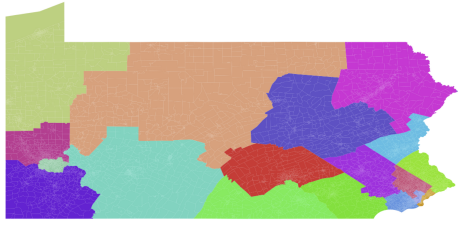


538 Dem



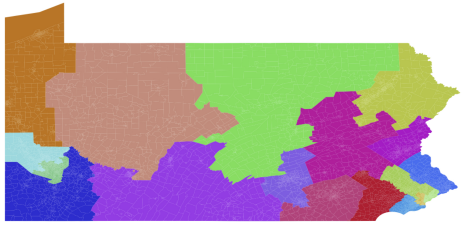
8th Grade



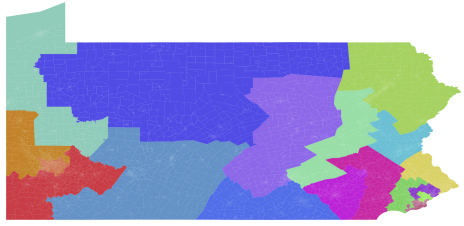



538 Compact

Gov

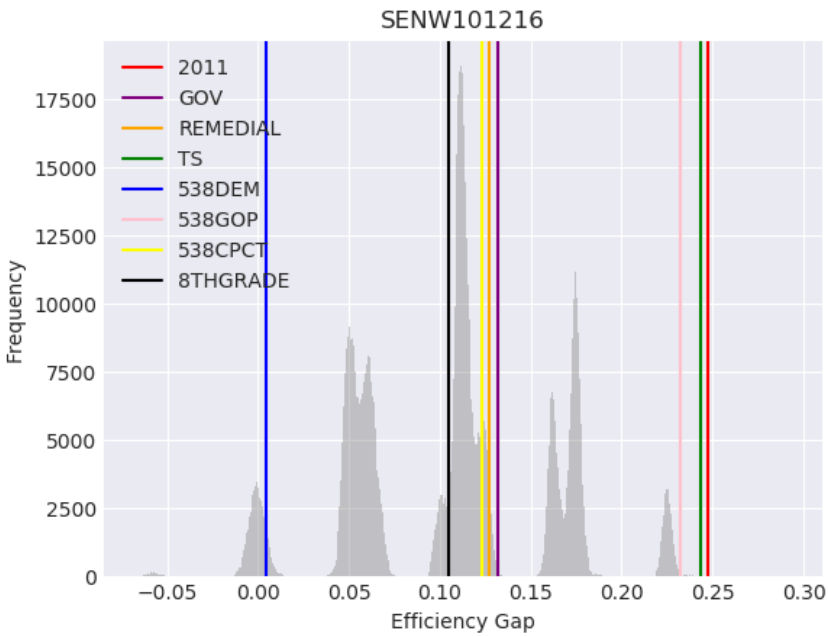


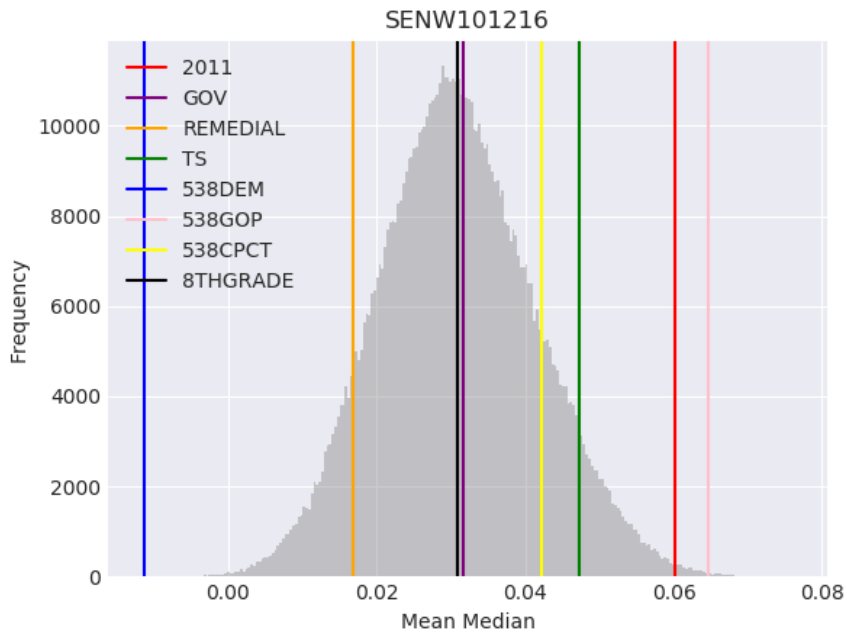
Remedial



TS

# Pennsylvania Landscapes



# Ensembles in Practice

- The appeal of an ensemble method is that you get to control the input data very carefully
- However, just because a particular type of data was not considered doesn't mean that the outcome is necessarily "fair"
- There are lots of "random" methods for constructing districting plans
- Most don't offer any control over the distribution that you are drawing from

# Necessary Inputs

- Geographic Units
  - Census vs. Political
  - Size of problem
- Demographic Data
  - Population
  - Voting Population
  - Race (for VRA)
- Initial Districting Plan
  - Discretization
- Voting Data
  - Aggregated in unusual units

# Data Availability

**Example (What adjective best describes US Electoral data?)**

# Data Availability

**Example (What adjective best describes US Electoral data?)**

Abominable*

# Data Availability

**Example (What adjective best describes US Electoral data?)**

## Abominable*

* Alternatively, any adjective from "You're a mean one, Mr. Grinch!"

# github.com/mggg-states

- Currently complete data for 7 states
- Varying levels of confidence in data fidelity
- Working to collect data for all states
- More importantly, constructing a pipeline for processing 2020 census data

Other Projects:

- OpenElections Project
- NvKelso
- Princeton Gerrymandering Project
- Redistricting Reform Project

# MGGG Software

1. github.com/mggg
   - Districtr
   - GerryChain
2. github.com/gerrymandr
   - Data Preprocessing
   - Compactness measures
   - Segregation Measures
   - State Specific Analyses
   - Jupyter Notebooks
   - ...
3. mggg.org
   - GridLandia
   - MetaGraph Sizes
   - Ecological Inference Apps

# Data Preparation Process

Full example at tinyurl.com/GerryDataGuide

- Select geographic units
- Attach demographic data
- Attach initial plan
- Attach voting data

# Geographic Units

# Demographic Data

# MORAL:

Computational Redistricting is NOT a solved problem!

# The End

Thanks!