autoredistrict

FREE AND OPEN SOURCE REDISTRICTING SOFTWARE

The Right to Vote protects all other rights

“The right of voting for representatives is the primary right by which other rights are protected. To take away this right is to reduce a man to slavery, for slavery consists in being subject to the will of another, and he that has not a vote in the election of representatives is in this case.”

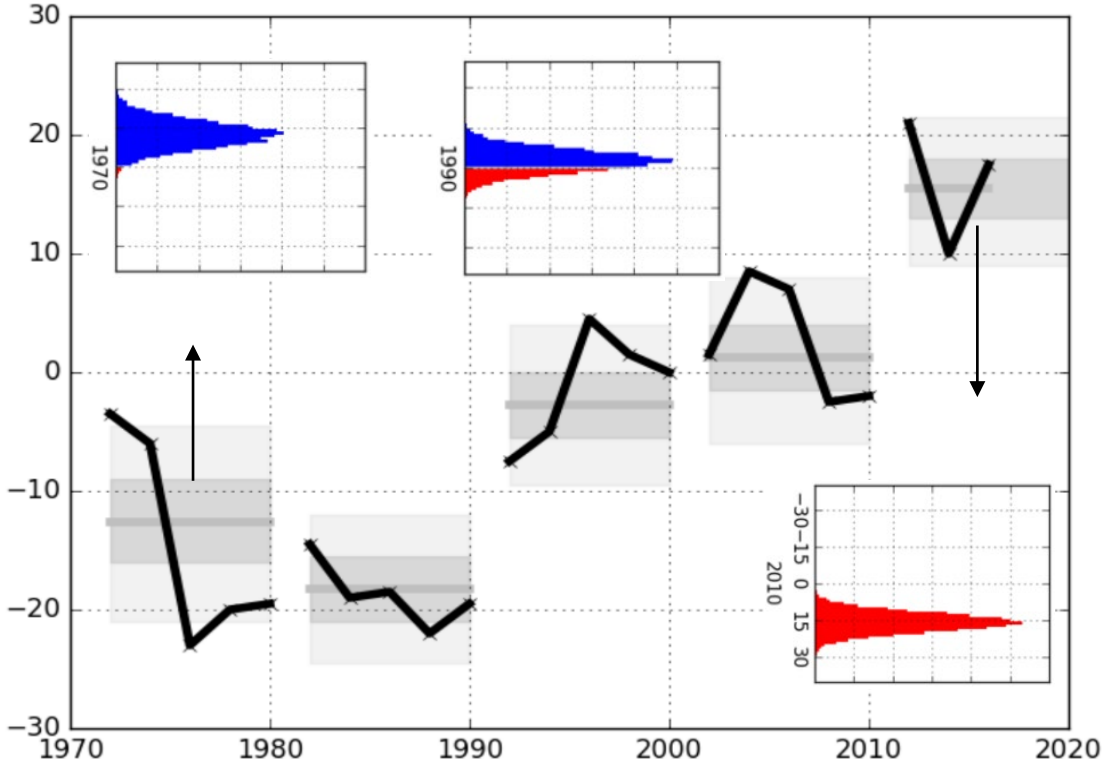
- Thomas Paine, Dissertation on First Principles of Government

Historic Specific Partisan Asymmetry

This research was done by Colin McAuliffe. Thanks to Sam Wang et. al. of Princeton for the vote count data.

Net national

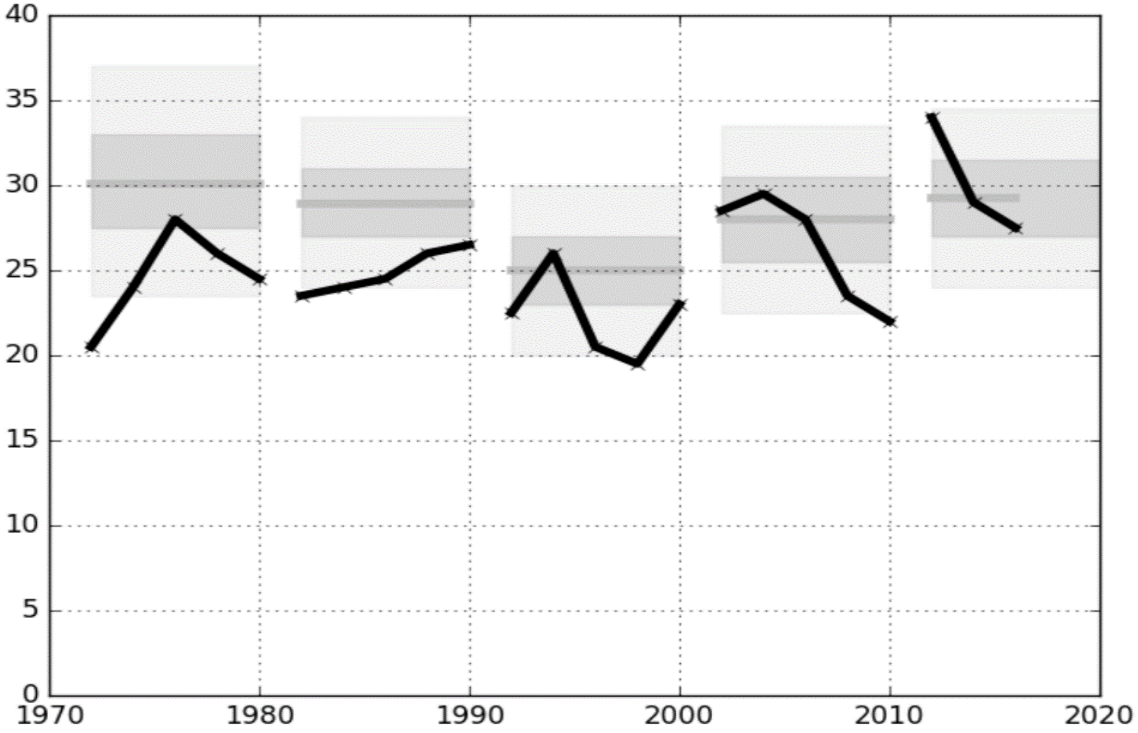
Dems gerrymandering in 1970 and 1980, Repubs in 2010



Total national

For the past 50 years, Gerrymandering has held constant at ~25 seats

That's equivalent to stealing about 20 million votes!



Supreme Court dereliction on Partisan Gerrymandering

- 1986 - Davis v. Bandemer (Indiana) no action taken
- 2004 - Vieth v. Jubelirer (Pennsylvania) no action taken
- 2006 - LULAC v. Perry (Texas) no action taken

- 2018 - Gil v. Whitford (Wisconsin) delayed (filed in 2015)
- 2018 - Benishek v. Lamone (Maryland) delayed (filed in 2013)
- 2018 - Rucho v. Common Cause (North Carolina) delayed (filed in 2016)

The Solution

- **Part 1:** Automated Redistricting
  - Demonstration

- **Part 2:** A sound legal test of Gerrymandering
  - Demonstration

Criteria in → Map out

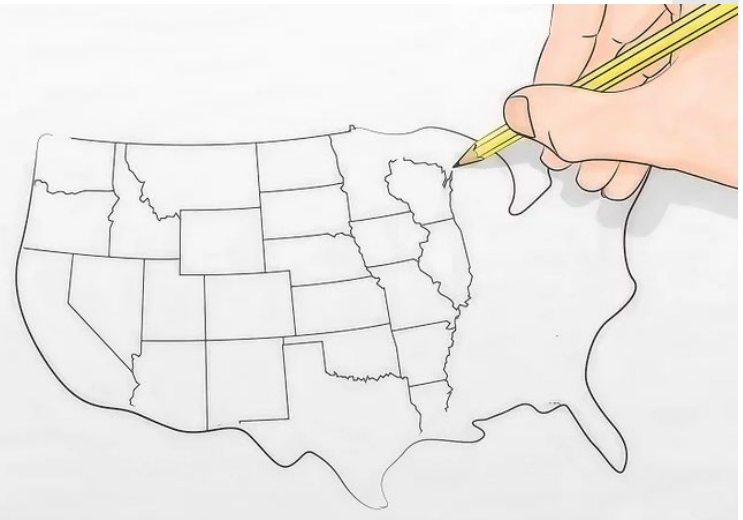
Custom criteria

Custom criteria

Custom criteria

Custom criteria

autoredistrict



Open Source Software

All source code is licensed under “GNU-GPL 3.0”

- Explicitly grants permission to copy, modify, and distribute
- All distributions must include the source code
- All derivative works must inherit this license

Fitness criteria

Geometric

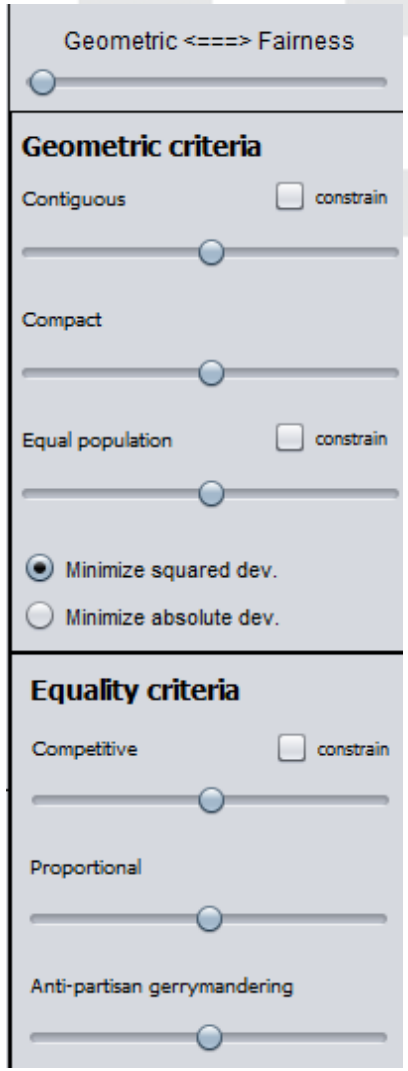
- Connectedness
- Compactness
- Equal population
- County splits

Fairness

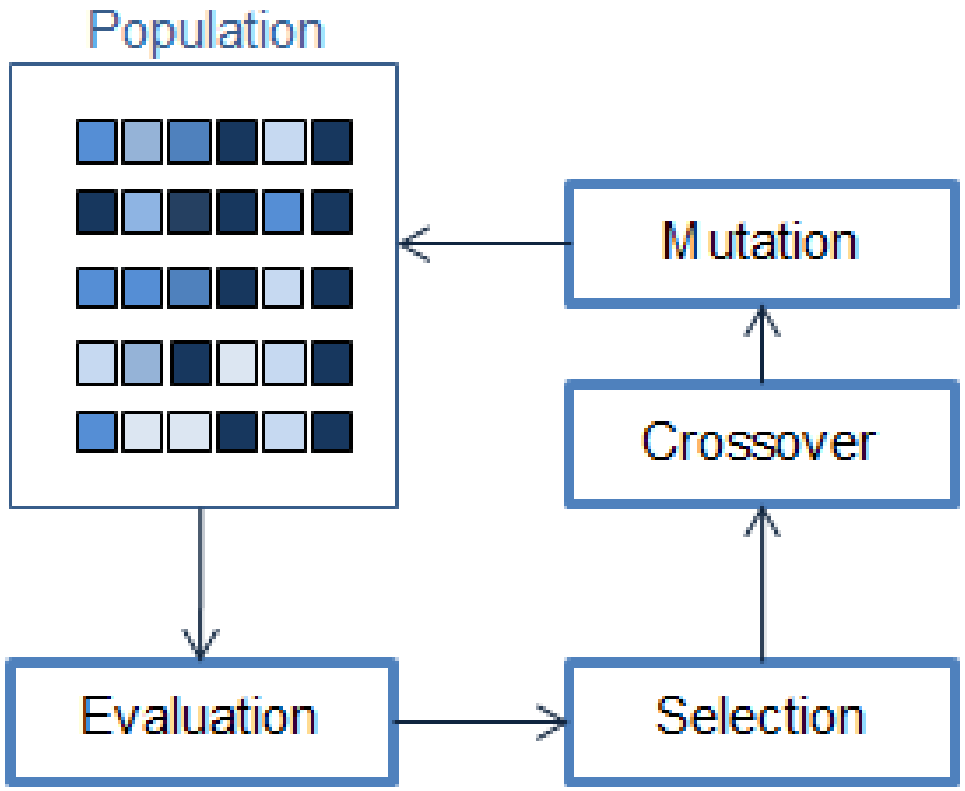
- Competitiveness
- Proportionality
- Partisan Gerrymandering
- Racial Gerrymandering

User-selected weights

- Normalized scores are then weighted by the user
  - Shown by the sliders to the right
  - Enables the user to prioritize criteria on-the-fly
  - A master slider for geometry vs fairness criteria
- Criteria scores are then added together to get a grand total



The Genetic Algorithm: Steps



1) Evaluate – score the fitness

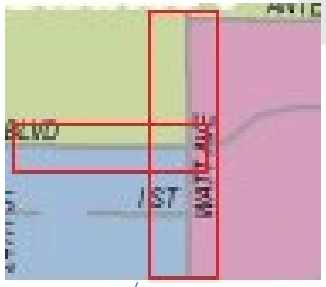
2) Select – pick high-scoring maps to create next generation from

3) Recombine – randomly take genes from each parent, exponentially approaches the best solution *(the key driver of evolution)*

4) Mutate – adds variety

AutoRedistrict starts with large refinements and gradually makes smaller refinements

- Only the perimeters of the districts are mutated
- Rate of mutation is reduced over time
  - On an exponential schedule
- AutoRedistrict is “done” when refinements are negligible

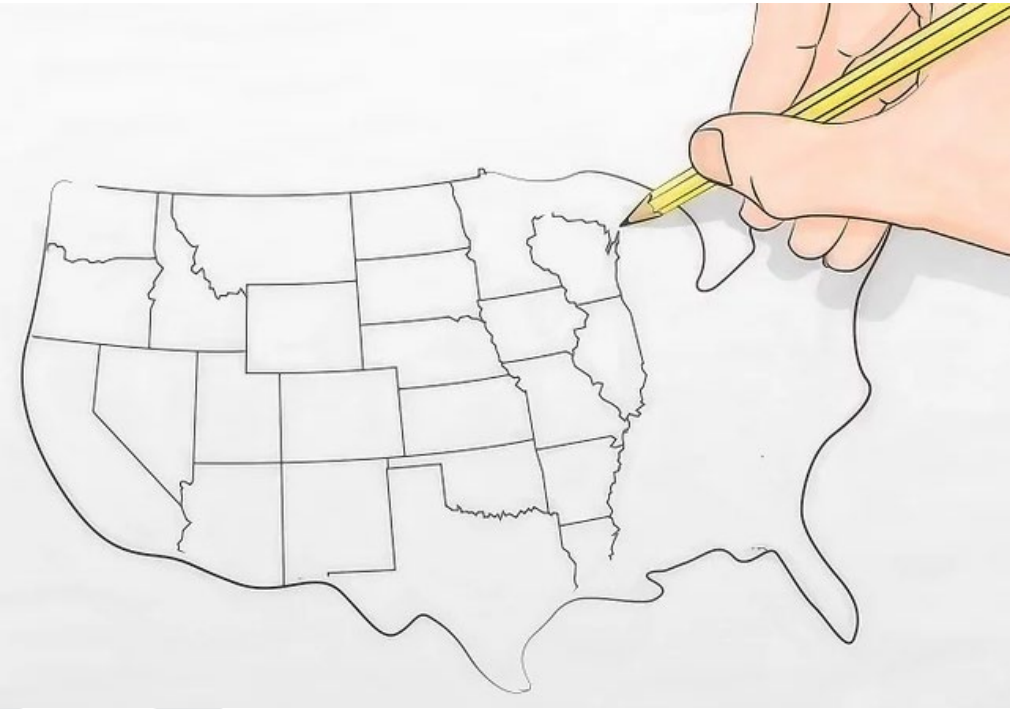


only genes at a border are mutated

In short, it’s a search engine.

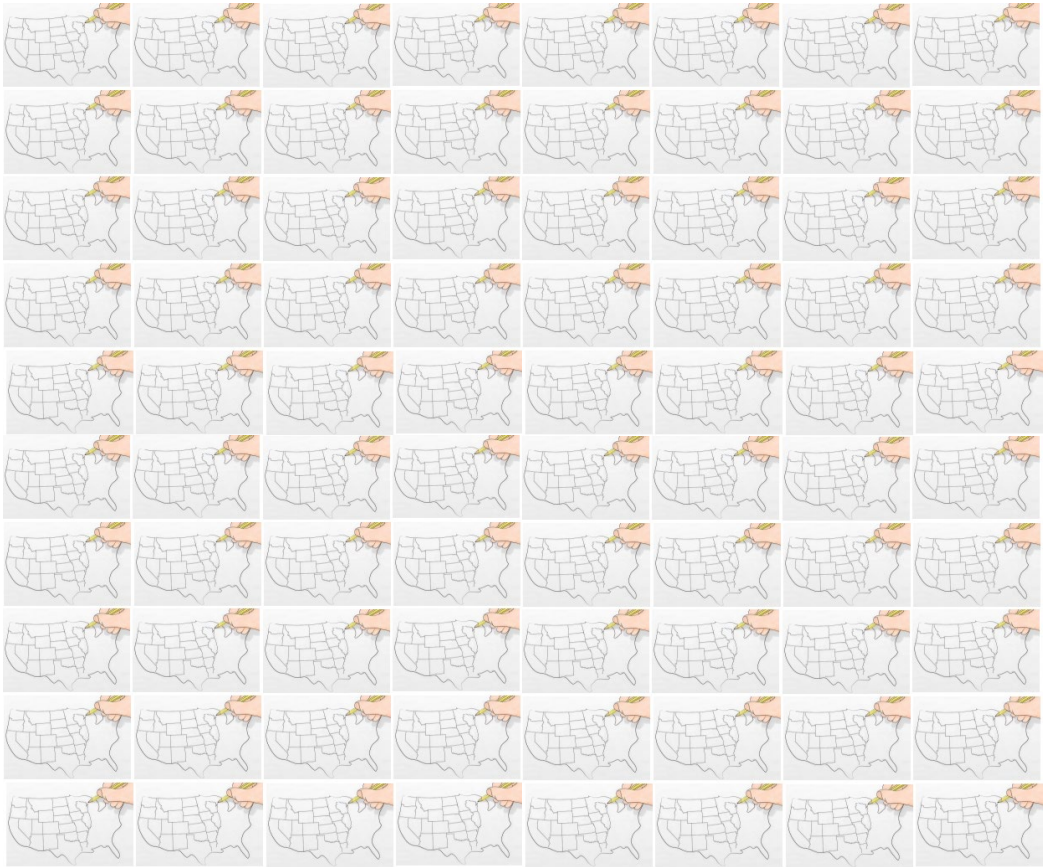
- AutoRedistrict explores almost all possible district arrangements
- On a typical desktop PC, it can evaluate hundreds of maps per second
- **This outperforms any human being by orders of magnitude**
- More evaluations = better results

Man vs. Machine

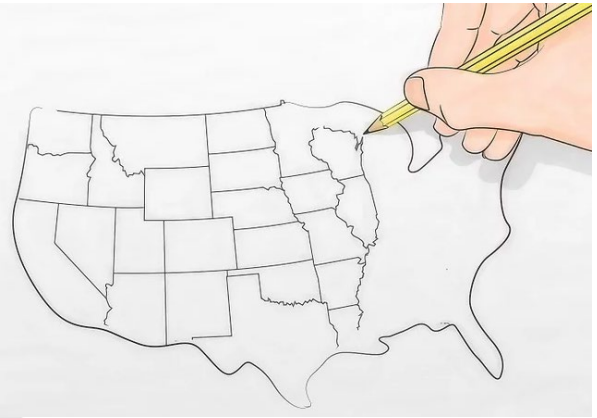


Machine Wins.

autoredistrict



Man + Machine

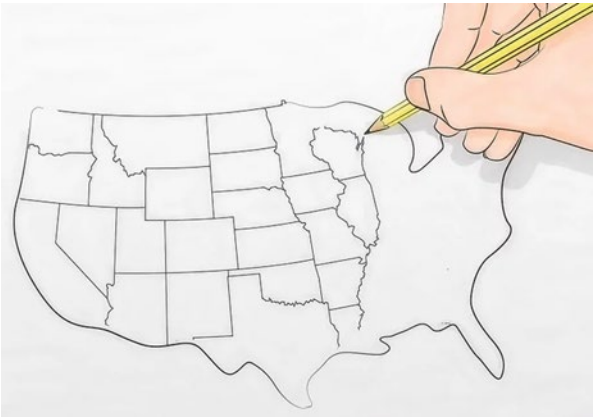


+

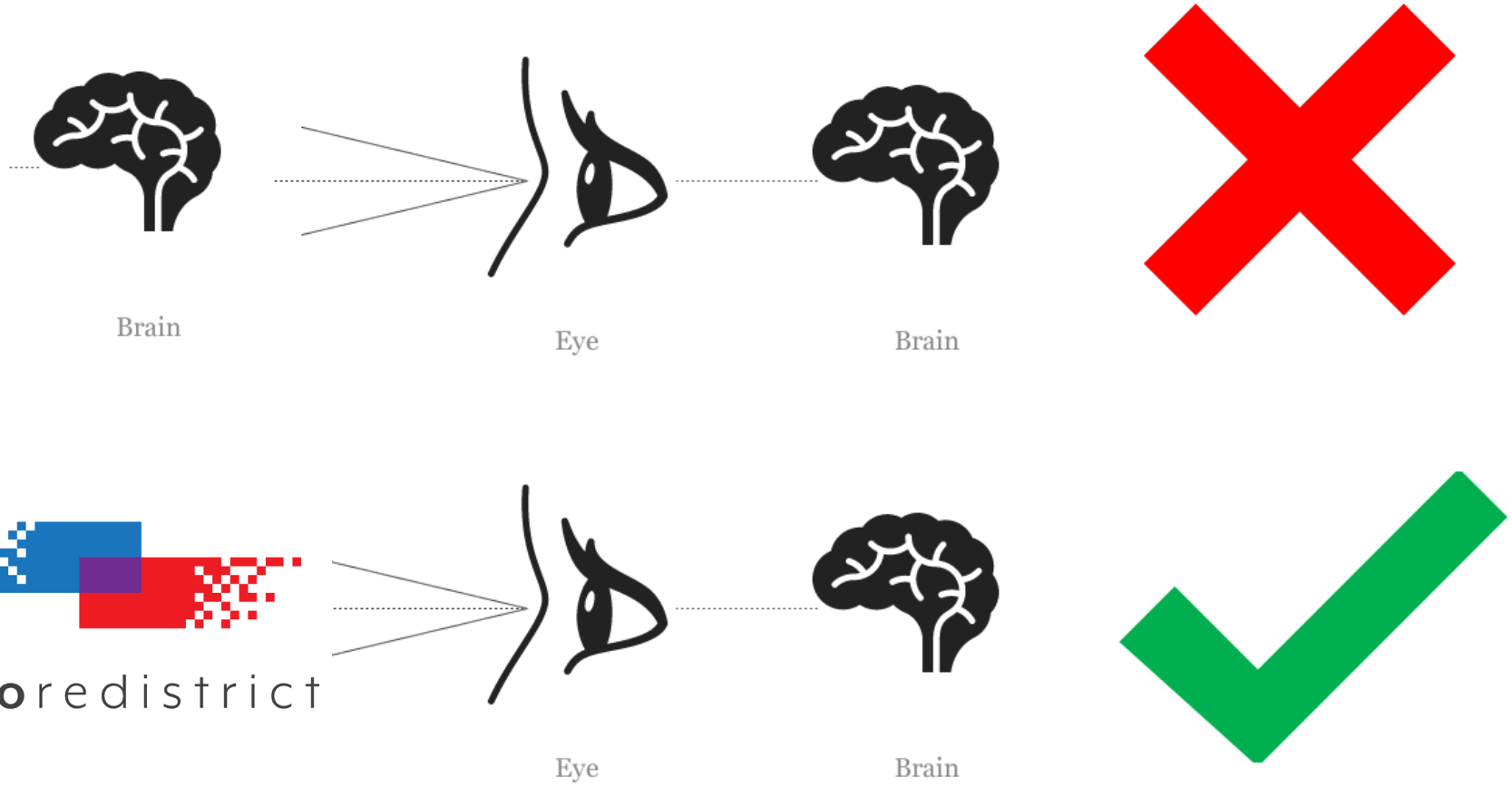
autoredistrict

=

Better map

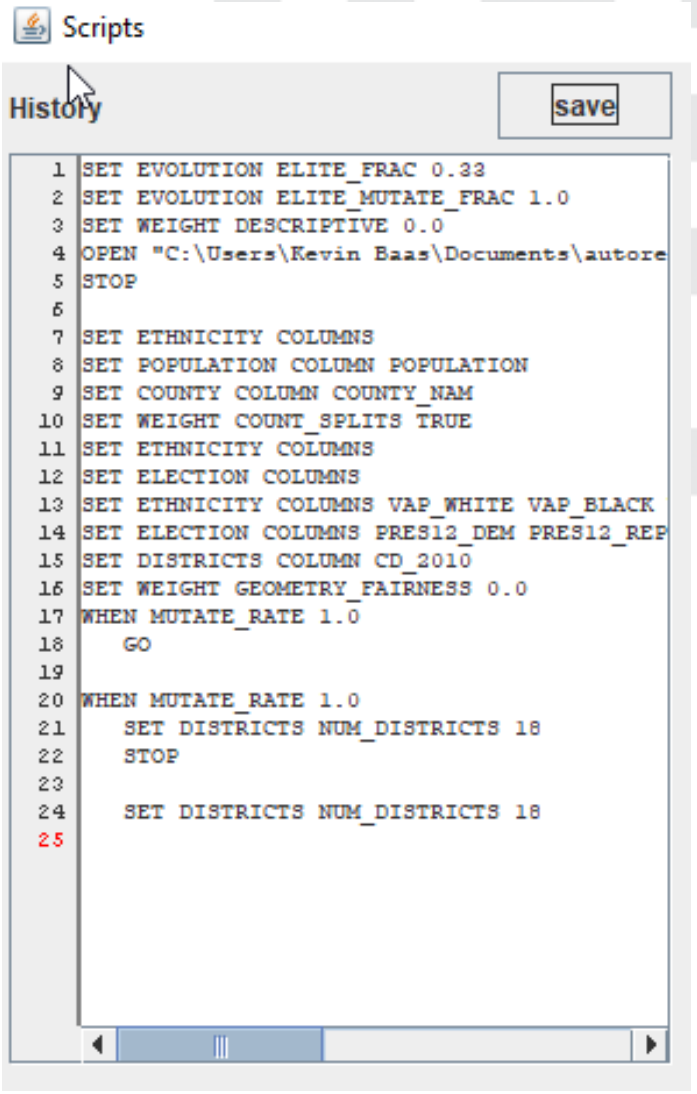


Automation **adds** Transparency



Automating AutoRedistrict

- AutoRedistrict records all user actions in a script
- Which can be played back
- Increases automation
- Increases transparency



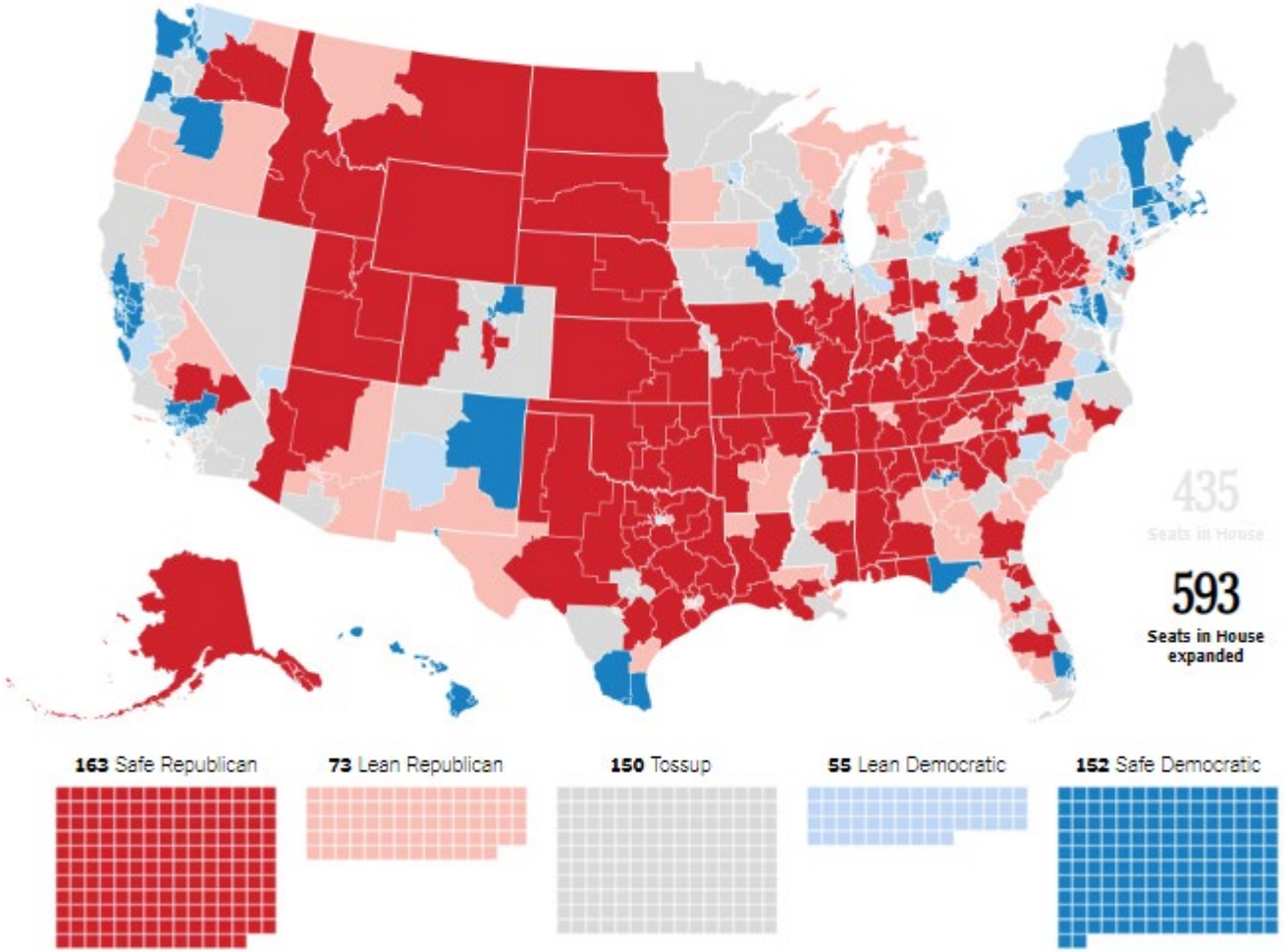
Automating Automating AutoRedistrict

- An AutoRedistrict script can be launched from the command line – without a user interface (“headless”)
- So in turn you can write a shell script to script the running of scripts
- For example…

```
xvfb-run -a -e xvfb.log  java -jar -Xmx4096M -Xms512M autoredistrict.jar run subscript1
xvfb-run -a -e xvfb.log  java -jar -Xmx4096M -Xms512M autoredistrict.jar clean 1
xvfb-run -a -e xvfb.log  java -jar -Xmx4096M -Xms512M autoredistrict.jar run subscript2
xvfb-run -a -e xvfb.log  java -jar -Xmx4096M -Xms512M autoredistrict.jar clean 2
```

The New York Times

America Needs a Bigger House

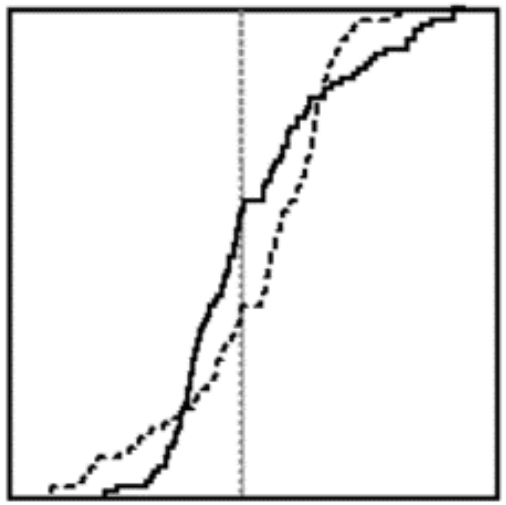


POWERED BY

autoredistrict

(Show software)

(Show website map)



The Solution
Part 2: A sound legal standard

How to win a gerrymandering lawsuit

Based on my reading of judges' opinions and defendants' filings in Supreme Court cases, in order to prevail in court you need to establish that:

- the districts are gerrymandered
- the gerrymandering is extreme
- and will continue to be

I'm going to show you how to do that.

The Supreme Court asked for a sound way to measure Gerrymandering

- That avoids counterfactuals

  adjective relating to or expressing what has not happened or is not the case.

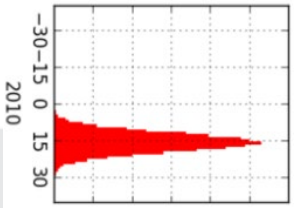
  noun a counterfactual conditional statement

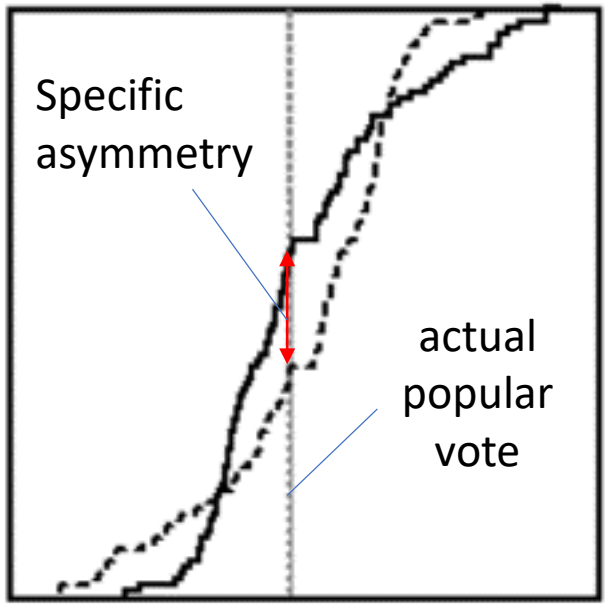
  example *If kangaroos had no tails, they would topple over.*

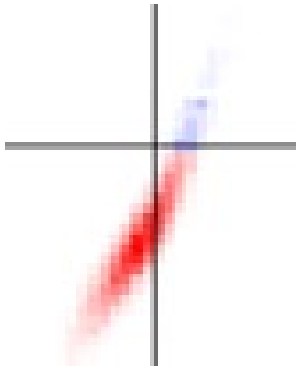
- And assesses durability
  - Was the partisan bias by chance, or will it continue to occur?

“Specific Asymmetry” + Probability model

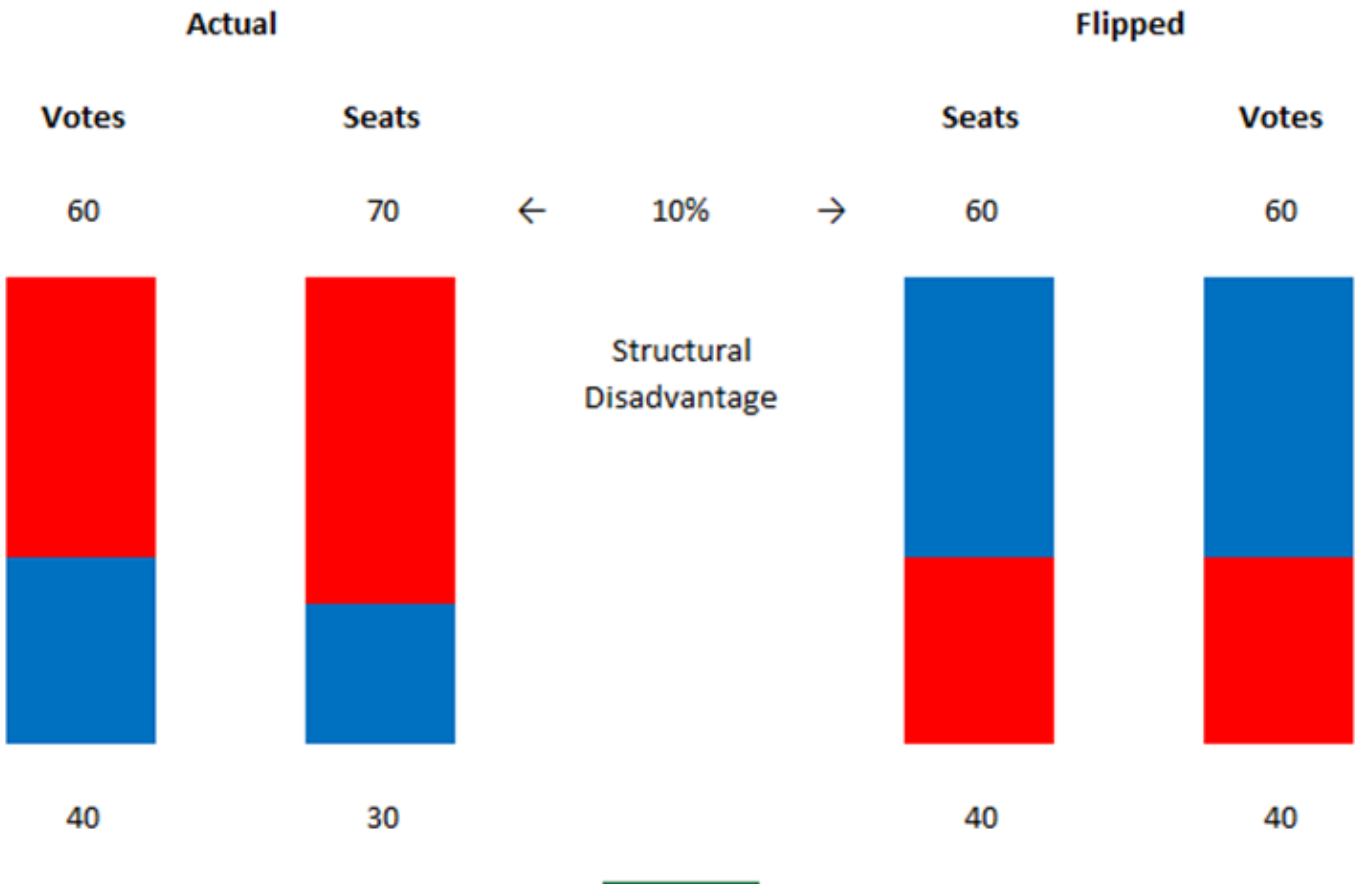
- ✔ Avoids counterfactuals
- ✔ Assess durability







Specific Asymmetry



Generating seats-votes curve from 1 election

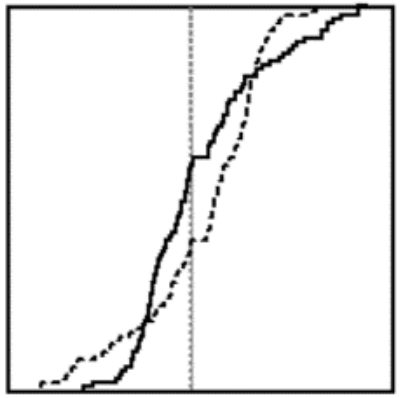
$$DemDistrictVote = \frac{OriginalDemVote}{OriginalDemTotal} \cdot PopularVotePercent$$

X (independent variable)

$$RepDistrictVote = \frac{OriginalRepVote}{OriginalRepTotal} \cdot PopularVotePercent$$

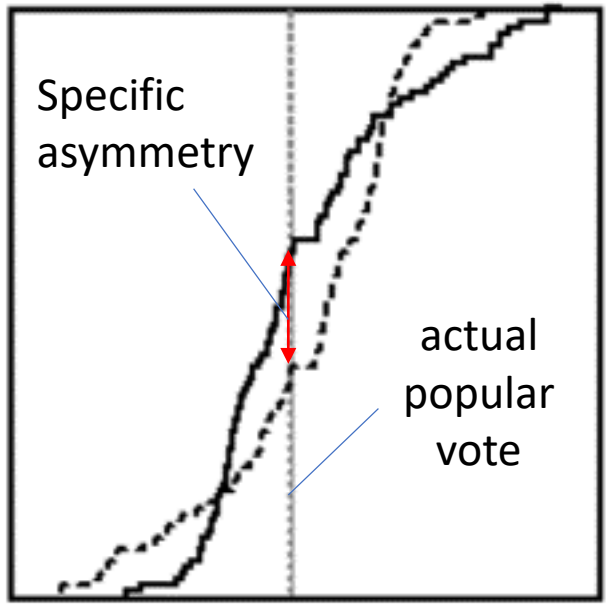
$$DemSeats = \sum_{Districts} DemDistrictVote > RepDistrictVote \begin{cases} true : 1 \\ false : 0 \end{cases}$$

Y (dependent variable)

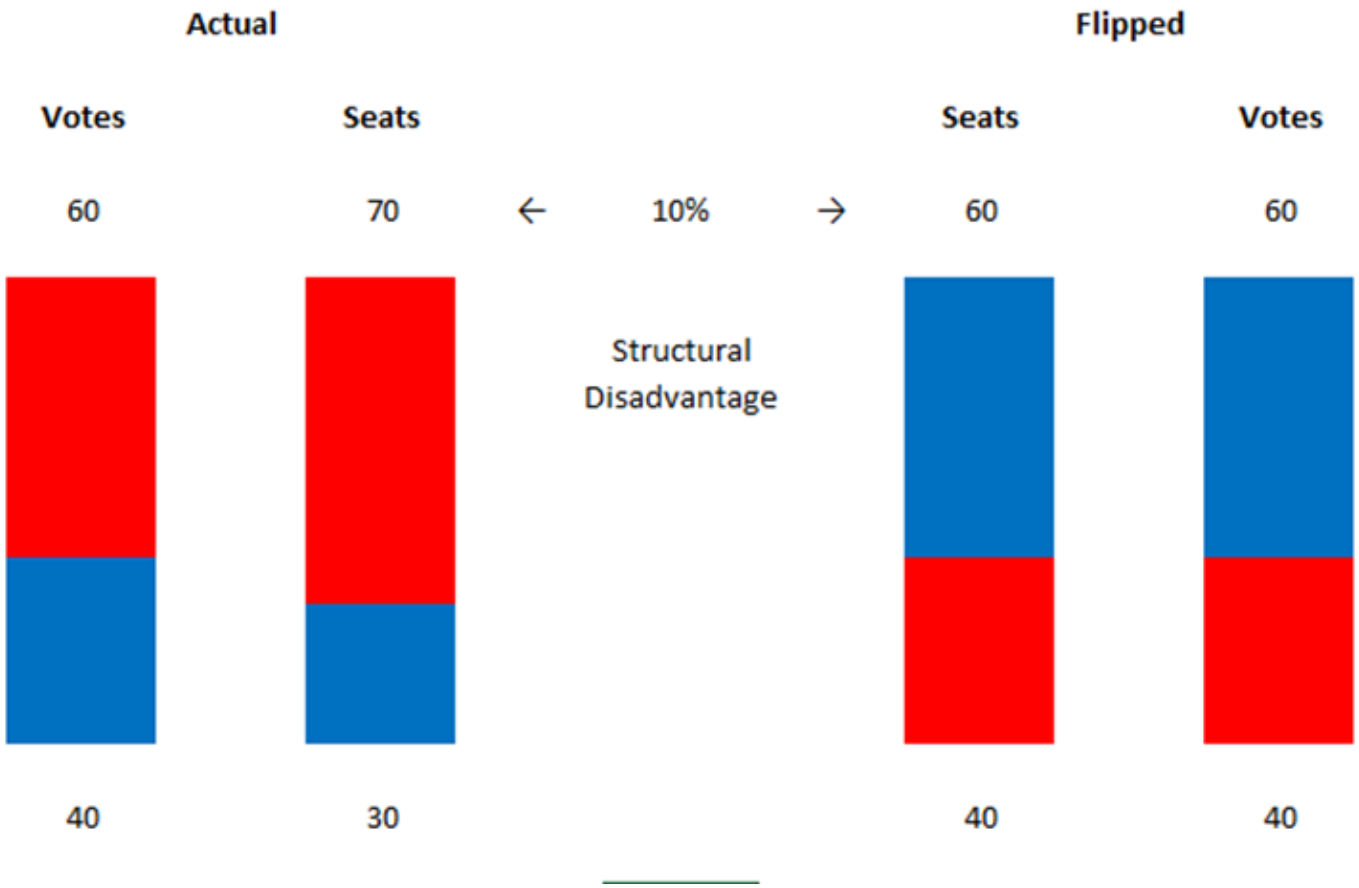


Specific asymmetry

- “Specific asymmetry” is the vertical distance (# of seats) between the seats votes curve and its reflection, measured at the actual popular vote
  - Avoids counterfactuals

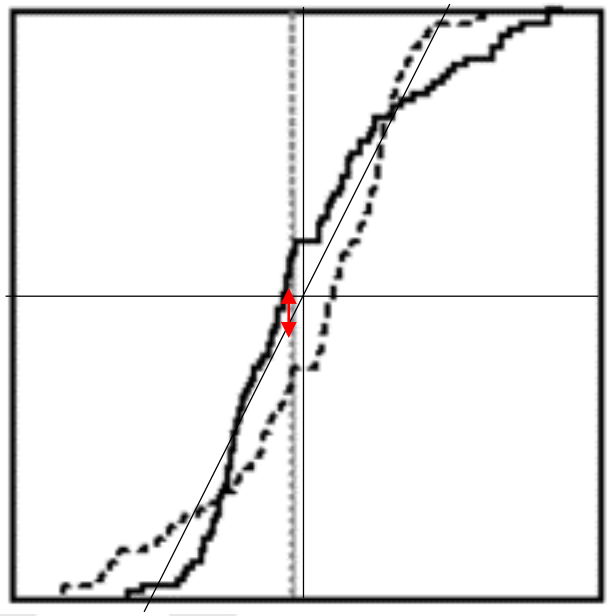


Specific Asymmetry



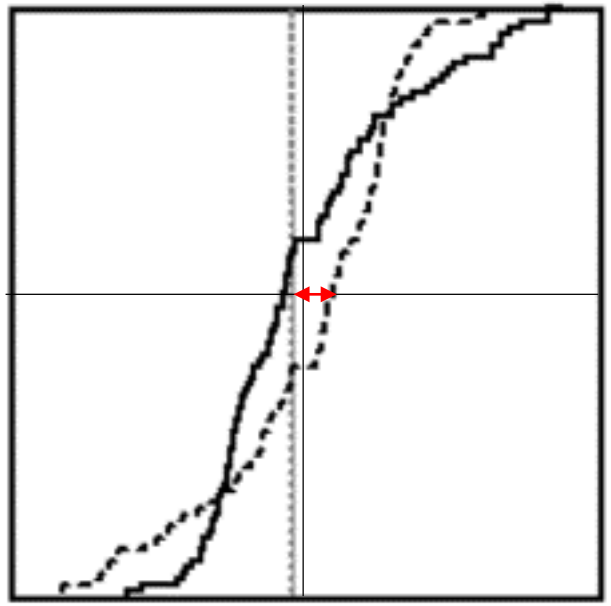
Different measures of gerrymandering

Efficiency gap
Steph. & McGhee



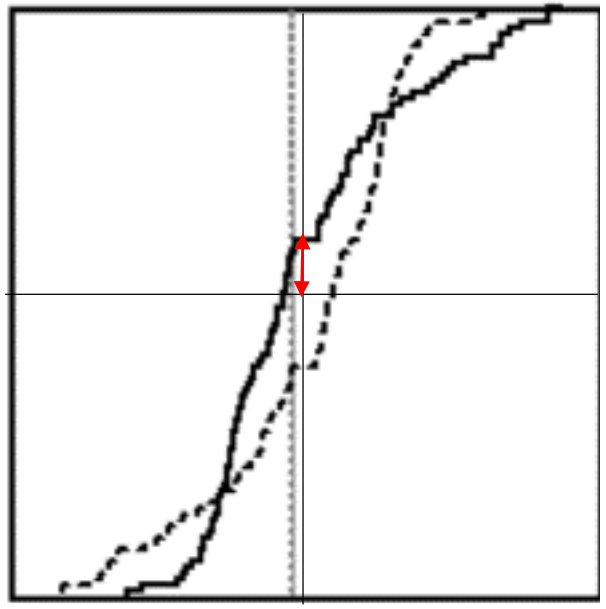
Assumes linearity

Median minus Mean
Sam Wang et. al.



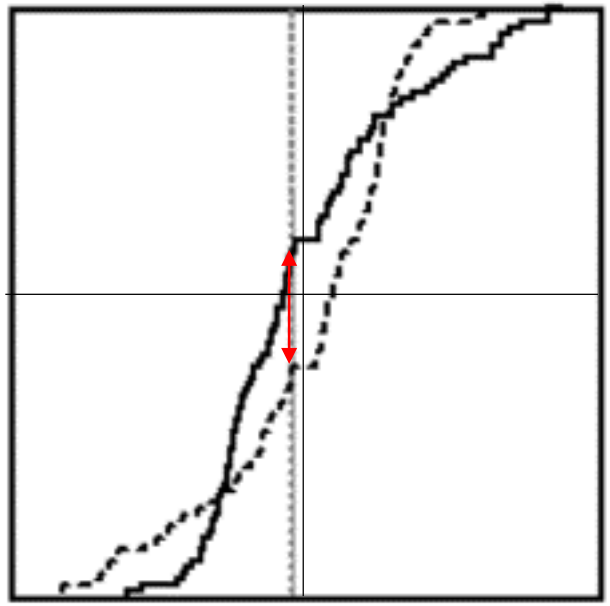
Measures at a hypothetical seat count (implicitly)

Partisan Symmetry
Grofman & King



Measures at a hypothetical popular vote (50:50)

Specific Asymmetry
Baas & McAuliffe



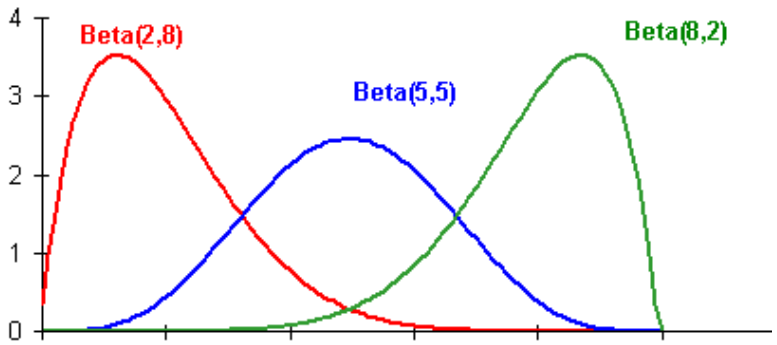
No counterfactuals

Voter sentiment = weighted coin

- When a random voter shows up at the polls, **which way they vote can be modeled by a flip of a weighted coin.**

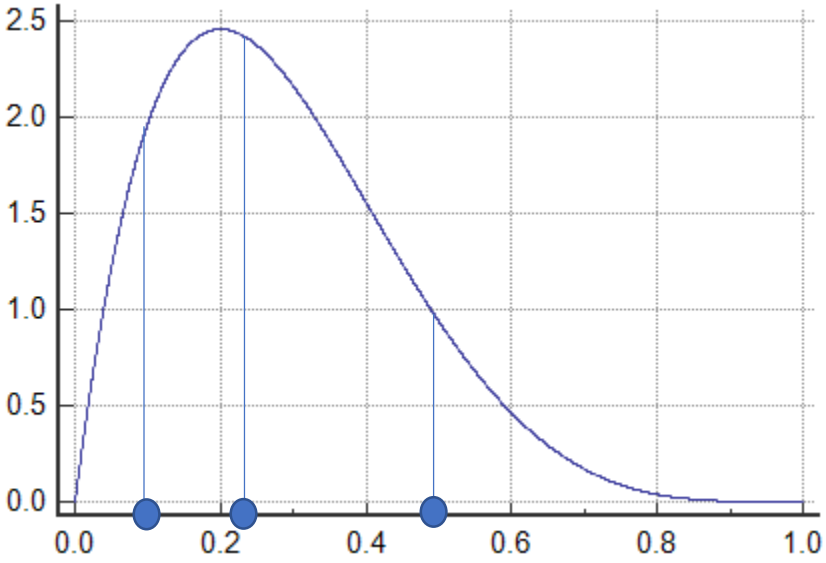
- The probability that a coin has any given weight is modelled by the "**Beta distribution**", pictured below.

- So **we use a Beta distribution to model voter sentiment.**

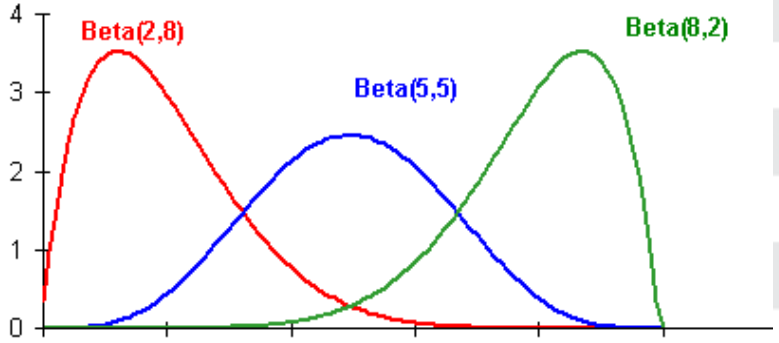


Maximum Likelihood Estimation

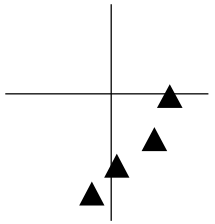
- Maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model so the observed data is most probable.

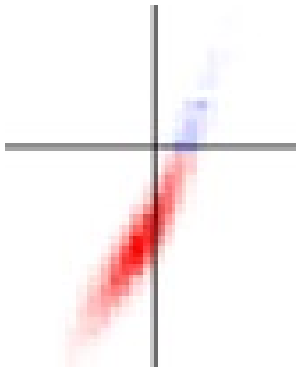


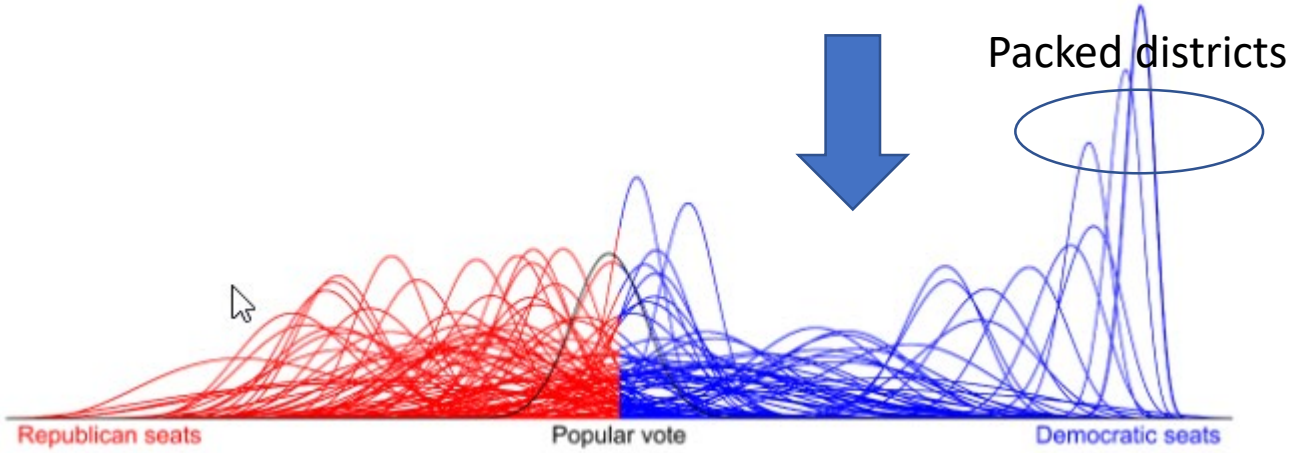
Probability model: 2-level Beta



- 1st level: The popular vote Beta distribution models the shared co-variance among the districts
- 2nd level: The district Beta distributions then take individual district deviations from that
- An unbiased estimator is used to avoid overfitting
- Then just pull random samples

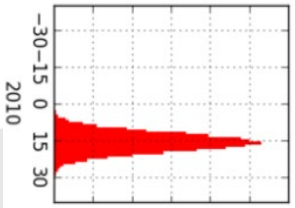


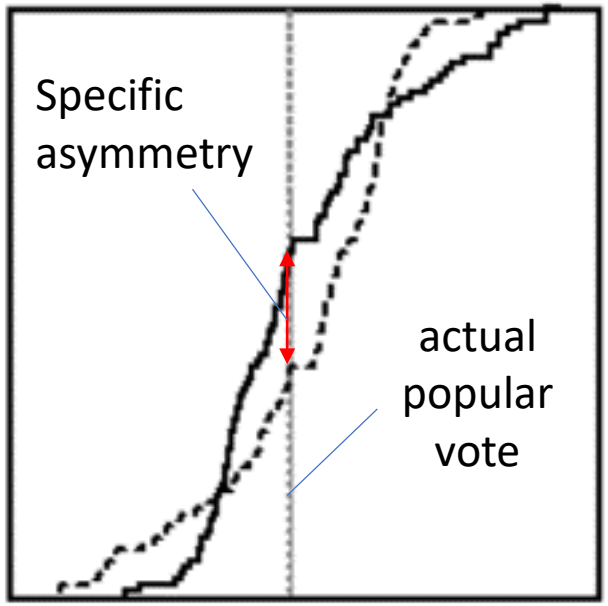


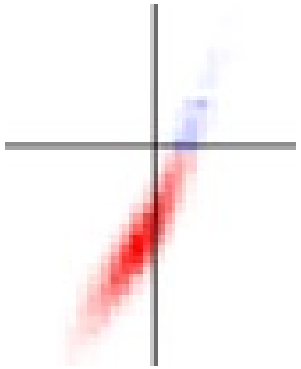


“Specific Asymmetry” + Probability model

- ✔ Avoids counterfactuals
- ✔ Assess durability







Recap

Criteria in → Map out

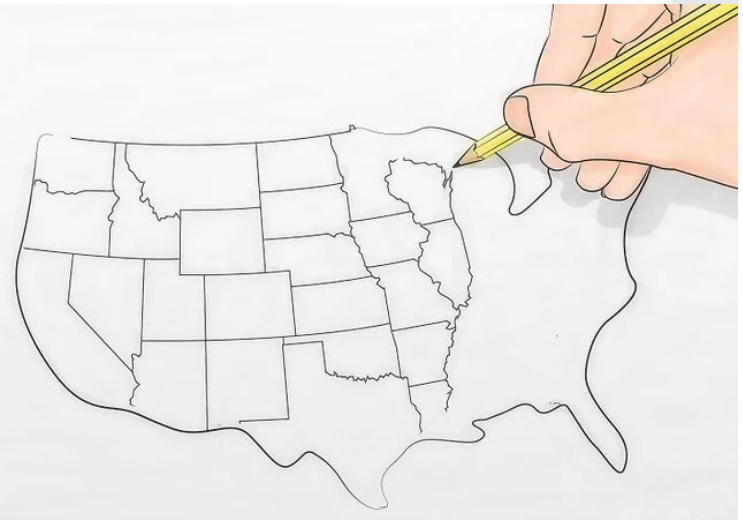
Custom criteria

Custom criteria

Custom criteria

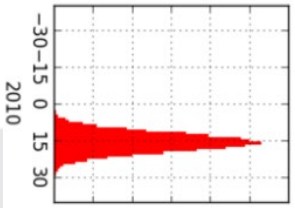
Custom criteria

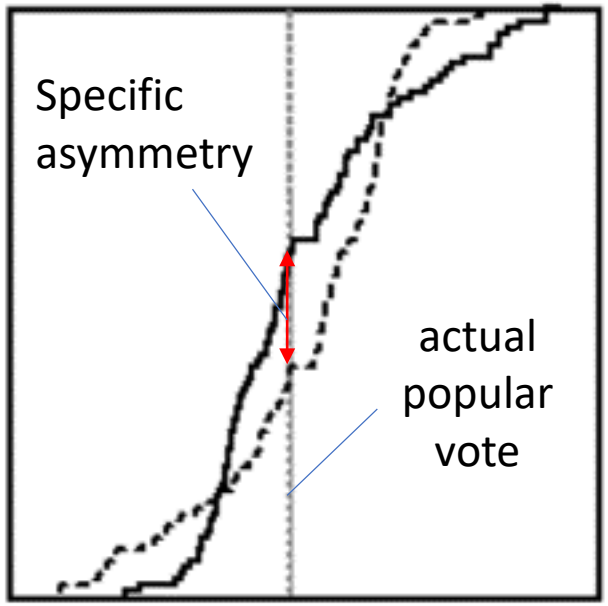
autoredistrict

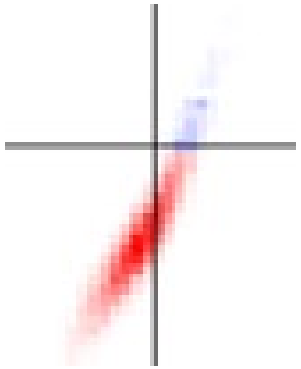


“Specific Asymmetry” + Probability model

- ✔ Avoids counterfactuals
- ✔ Assess durability







The Right to Vote protects all other rights

“The right of voting for representatives is the primary right by which other rights are protected. To take away this right is to reduce a man to slavery, for slavery consists in being subject to the will of another, and he that has not a vote in the election of representatives is in this case.”

- Thomas Paine, Dissertation on First Principles of Government

I want to get this out there. Contact me.

- Give demos of the software
- Explain the partisan gerrymandering metric
- Answer questions
- Generate maps for you
- Add new criteria into the software

Website: autoredistrict.org
Email: kbaas@autoredistrict.org
Facebook group: **AutoRedistrict**
Name: **Kevin Baas**

Thank you.
Come visit my booth,
I'd love to explain more.

Questions? Comments?

Website: autoredistrict.org
Email: kbaas@autoredistrict.org
Facebook group: **AutoRedistrict**
Name: **Kevin Baas**

autoredistrict

Extra slides

Potential improvements to AutoRedistrict

- Major refactoring
- Make criteria more modular and extensible
- KML export / google maps integration
- Shared public repo for source data and result data
- Plugin for ArcGIS (or QGIS)

Countering common legal arguments

Common legal arguments:

- Outcome is due to changes in voter sentiment
- Outcome is a natural consequence of geography
- Etc.

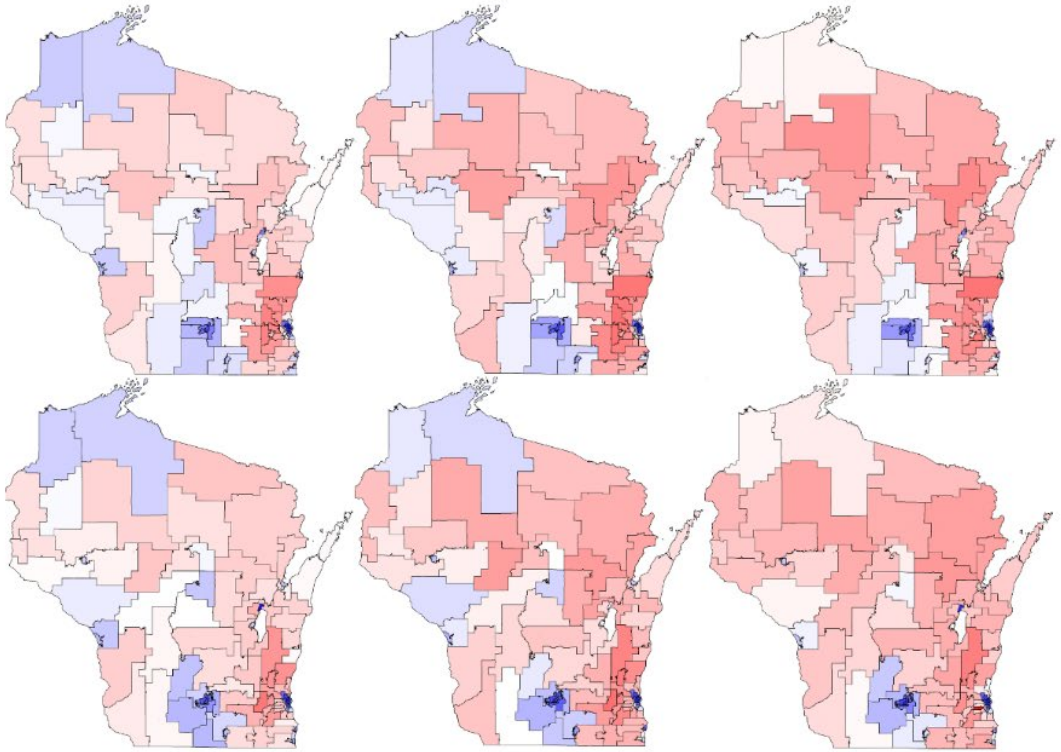
Solution: Make everything the same except district shapes.
Since everything else is held constant, all differences in election outcomes must be due to district shapes alone.

Time	same
Space	same
Elections	same
District shapes	different

Wisconsin Assembly before and after Gerrymandering using cross-aggregated vote counts

2000 districts

2010 districts



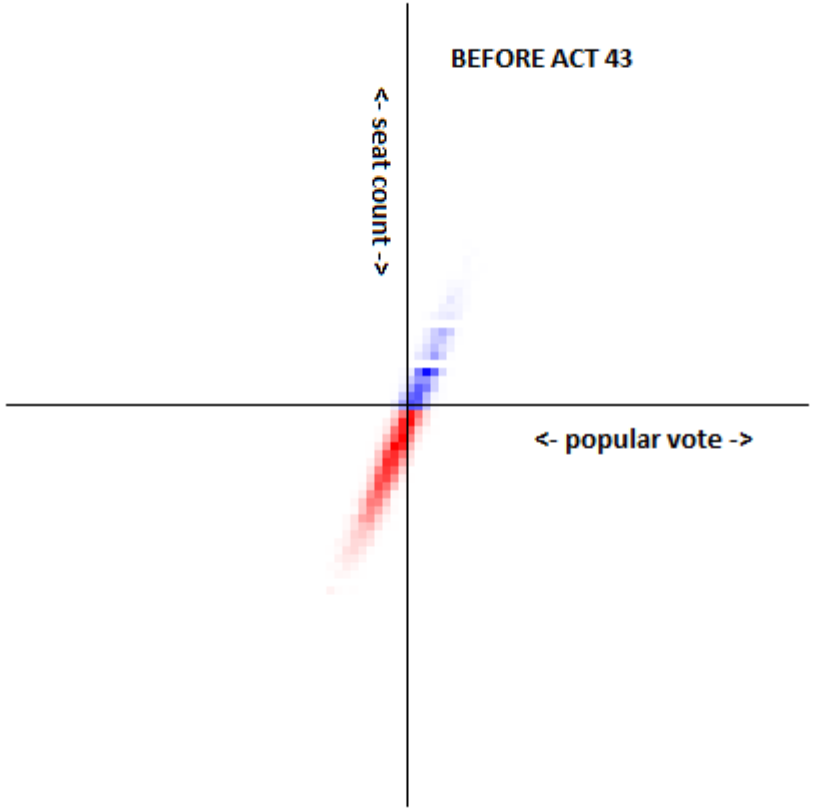
re-aggregate to voting ward resolution

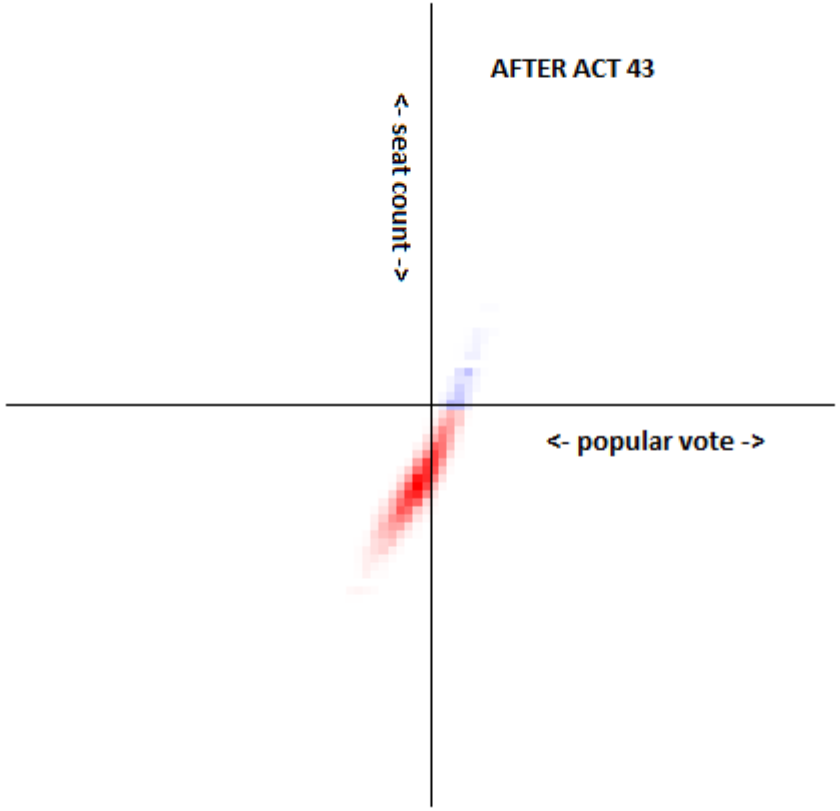
de-aggregate to block resolution

voting ward resolution

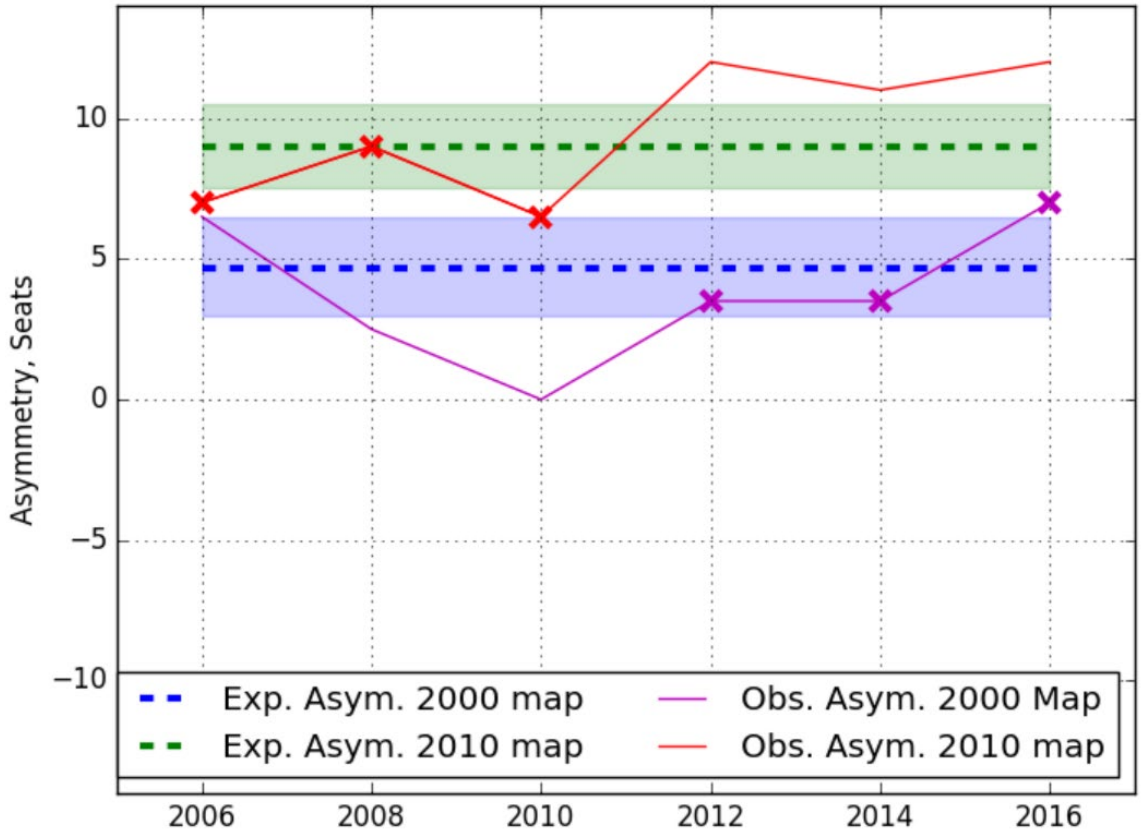
Figure 16: Partisan vote packing in Wisconsin Assembly elections. Top: 2000 cycle districts, Bottom: 2010 cycle districts, Left to right: 2012, 2014, and 2016 elections.

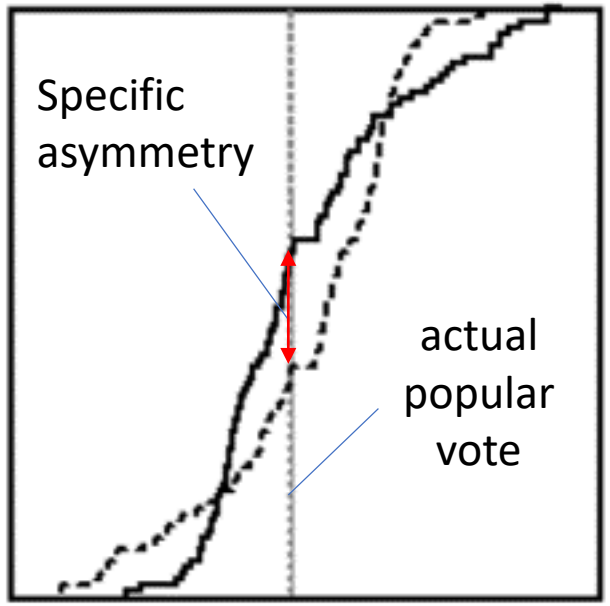
Wisconsin Assembly, before and after Gerrymandering Seats-votes pictures





Wisconsin Assembly before and after Gerrymandering Specific Asymmetry, Expected and Actual





Myths about automated redistricting

- Automated redistricting can be used for evil (not a myth)
  - Yeah, and it **is** being used for evil, and that evil will flourish if we don't fight back by using it for good
  - You don't need automated redistricting to gerrymander, self-sorting makes it trivial: just draw a circle around the cities. Done.
- A computer can't decide the criteria for creating a district (not a myth)
  - …but it can create districts based on these criteria better than any human could do.
- Automated redistricting removes human input
  - The results can be pre-processed and post-processed
  - Different criteria and priorities can be chosen in advance and adjusted in real-time
  - People can choose between a number of proposed/generated solutions
- Automated redistricting removes transparency
  - Automated redistricting **adds** transparency
  - You can read the source code – you can't read a person's mind
    - An open-source license such as GNU-GPL protects against malicious code tampering
  - You can record every action – you can't read a person's mind
  - It's repeatable / reproducible

Myths about automated redistricting (cont'd)

- Constitutional criteria must come first – automated redistricting can't do that
  - In mathematics and computer science we call this a "constraint". Computers are **really** good at them.
  - It's trivial to feed a computer user-supplied constraints
  - A genetic algorithm can churn through user-supplied constraints like butter
  - This includes "communities of interest"
- Compactness and other such measurements are subjective and a computer can't measure them
  - These are trivial to compute, not even hard… not even average.
  - If you can't put it in a formula, you are being biased and that's bad.
- Automated redistricting is deterministic – produces only one solution
  - Heuristic optimization algorithms produce a different solution each time
- The computing power needed is unmanageable
  - With today's computing power, it can be done on a typical desktop computer
  - The solution can start at a low resolution and go to progressively higher resolutions

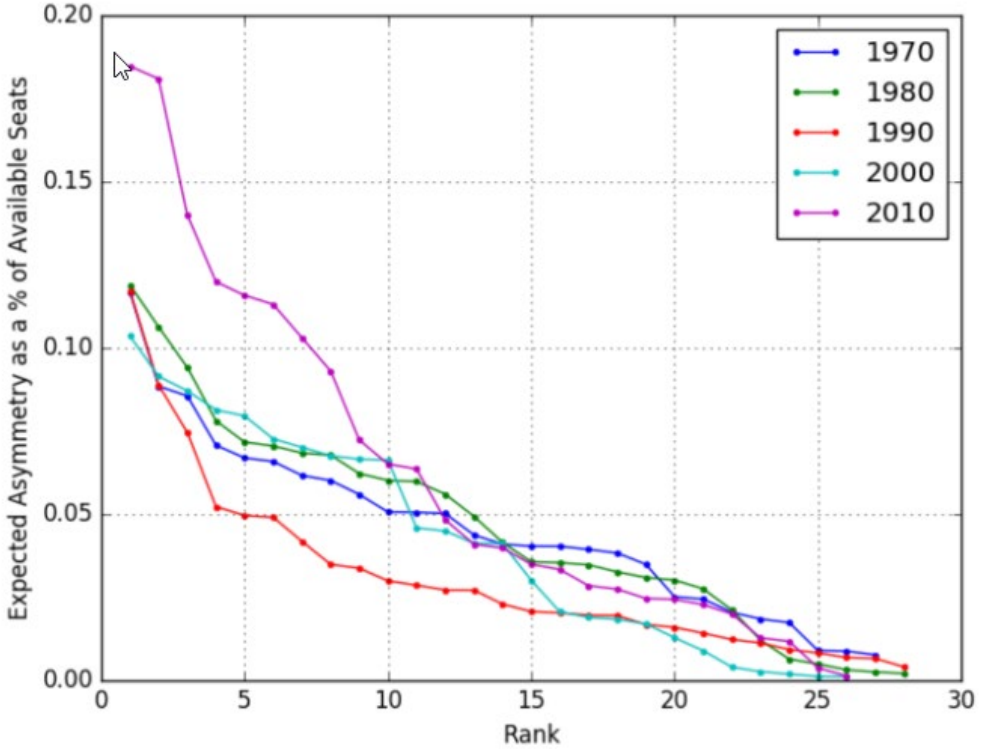
Gerrymandering is getting *more* extreme

In 2010, Republicans gerrymandered the entire country, openly, and bragged about it

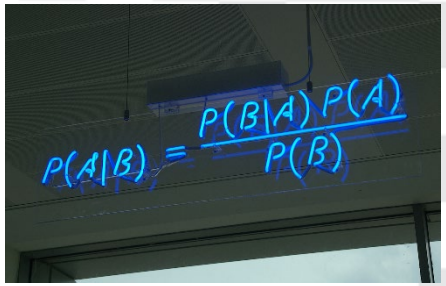
"Republicans have an opportunity to create 20-25 new Republican Congressional Districts through the redistricting process over the next five election cycles, solidifying a Republican House majority."

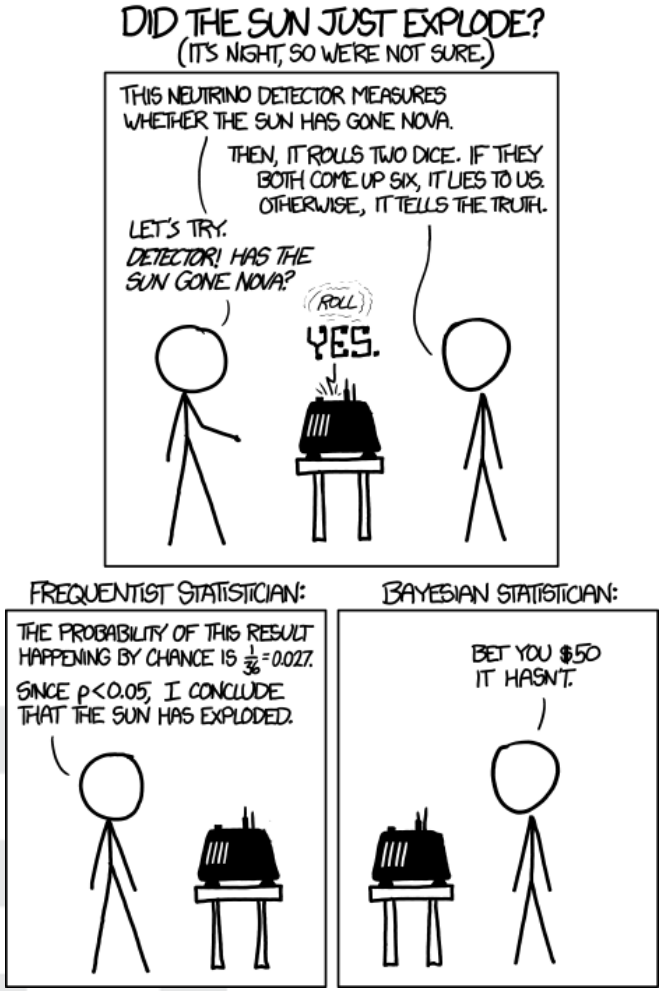
– *redistrictingmajorityproject.com*

2010 gerrymanders were significantly more extreme than all previous decades



Bayesian probability





The Frequentist is using a simple significance test:

"P value" = p(sun **not** exploded|yes) = (almost 1/36).

that's less than 0.05, so the sun has exploded.

The Bayesian is using Bayes' rule:

p(a|b) = p(b|a)*p(a)/p(b)

a = sun exploded, b = machine says yes

p(sun exploded|yes)

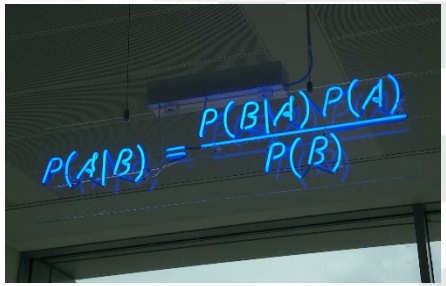
= p(yes|sun exploded) * p(sun exploded) / p(yes)

= (35/36) * (almost zero) / (almost 1/36)

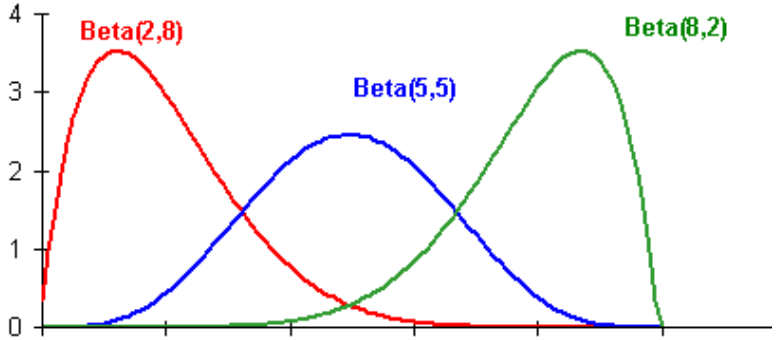
= 35 * almost zero

(35 times more likely than it was before the machine said yes)

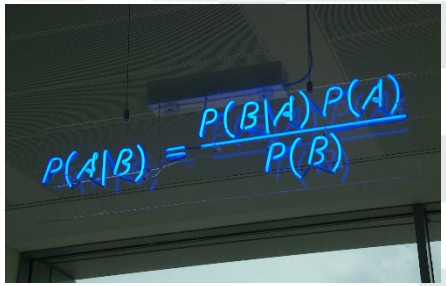
The Beta Distribution



- After 5 flips of a coin, you get 4 heads and 1 tail. What's the likelihood that it's a fair coin?
- Bayes' Rule: p(fair|4h,1t) ∝ p(4h,1t|fair)
- More generally: p(w=x|outcomes) ∝ p(outcomes|w=x)
- Draw out the full curve for every value of w
  - That's called the "Beta Distribution"



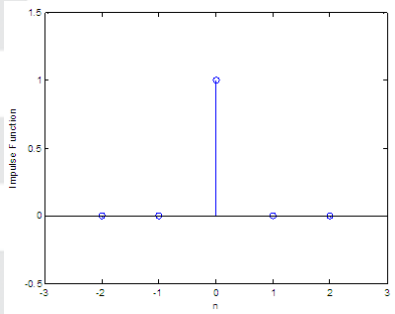
Bayesian probability

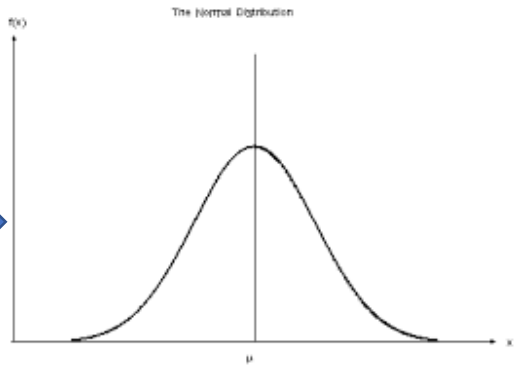


Frequentist

We don't know the exact outcome each time, but **we know the probability distribution of the outcomes**. (e.g. 50% heads, 50% tails)

Uses observations to **estimate the single most likely** parameters of a probability distribution. (e.g. mean and variance)

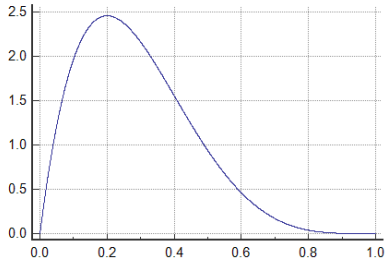


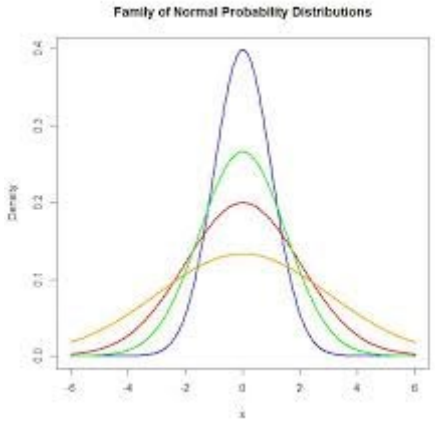


Bayesian

No, actually we don't know that either. We can only infer from the data **that some distributions are more likely than others.** (e.g. we give different likelihoods to each possible weight of a coin.)

Uses observations to **estimate a likelihood for ALL possible** parameters of a distribution.





Towards adoption - outreach

- Education / spreading awareness
- Political Action Committees
- Contacting representatives
- Academic publications
- News / Opinion (such as the NYT)

Towards adoption - action

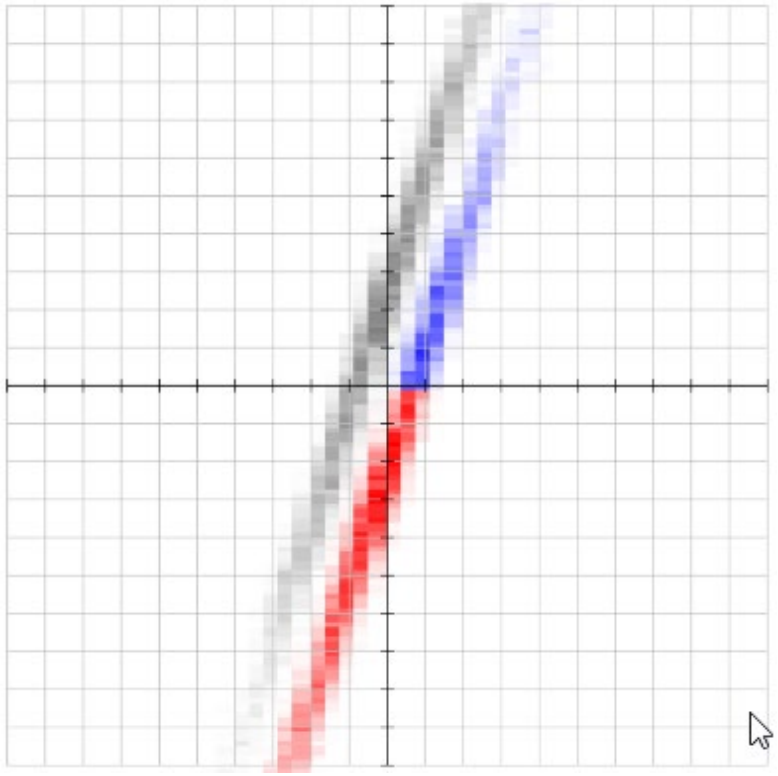
- Lawsuits based on sound mathematics
- Ballot initiatives
- Individual municipalities reaching out

Partisan impacts of redistricting methods

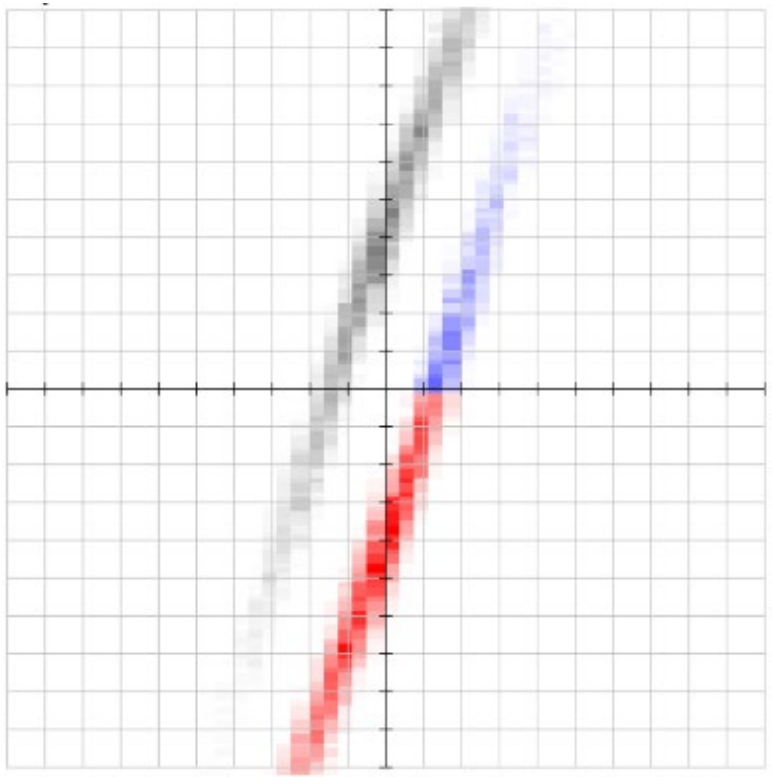
- Used presidential vote counts from 6 elections
- Used the probability model
- Used all 50 states, so the results aren’t idiosyncratic to any state
- Used 4 different redistricting methods:
  - Actual 2000 districts
  - Actual 2010 districts
  - Compactness optimized districts
  - Multi-member districts
- Generated seats-votes picture for each

All congressional districts
Seats-votes likelihood pictures

Actual 2000 districts

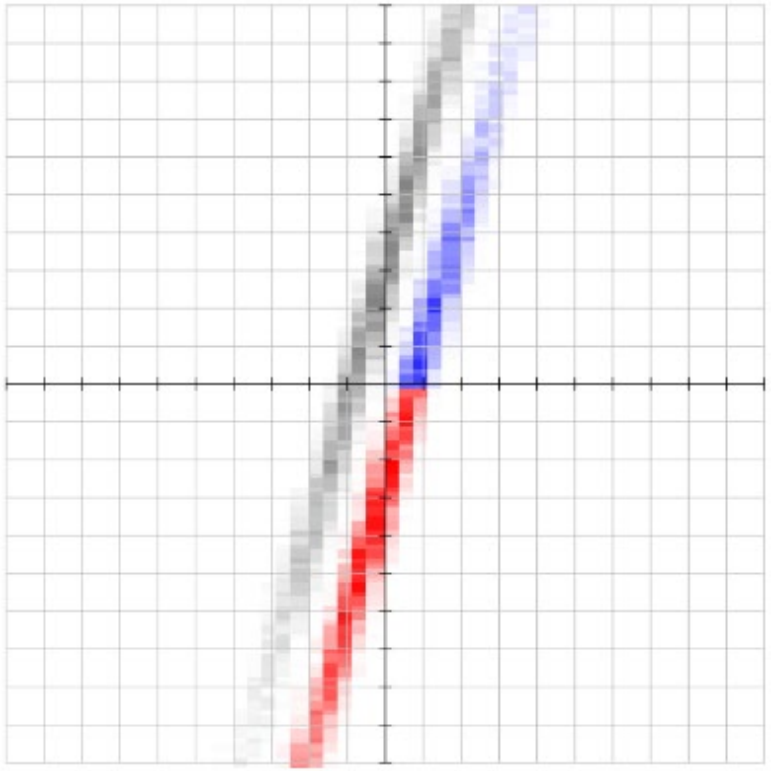


Actual 2010 districts

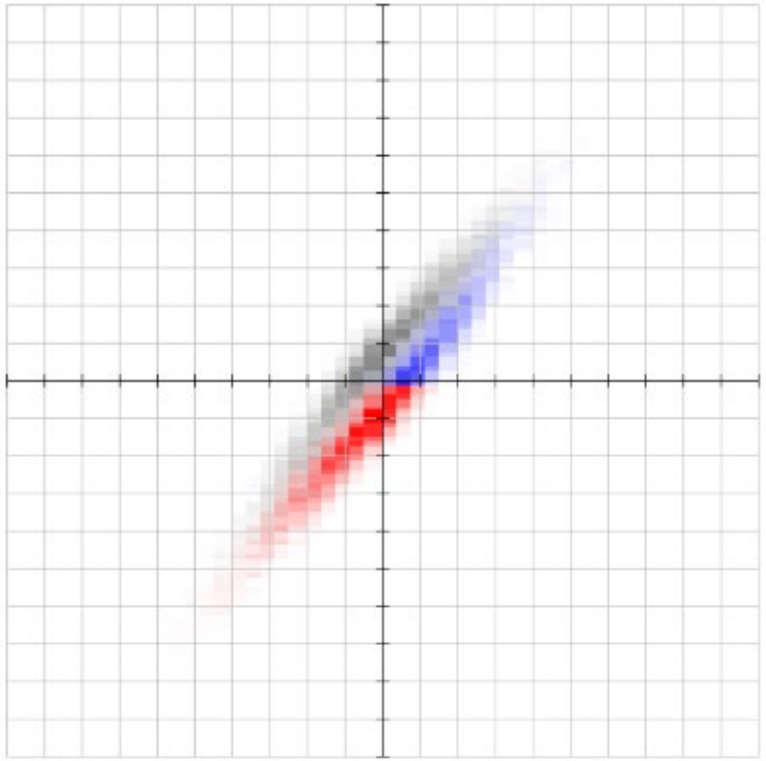


All congressional districts
Seats-votes likelihood pictures

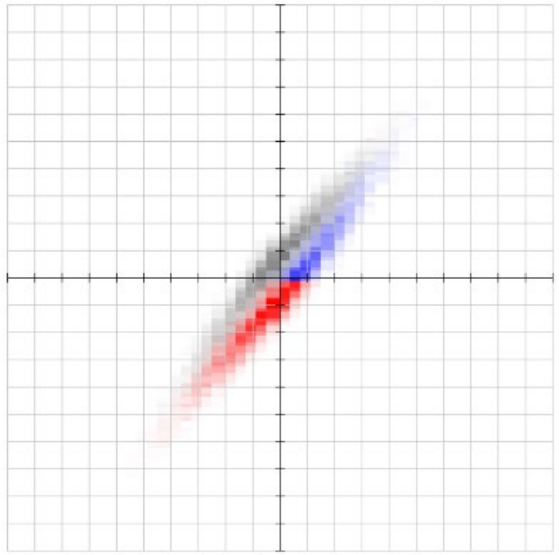
Compactness optimized (Bdistricting)



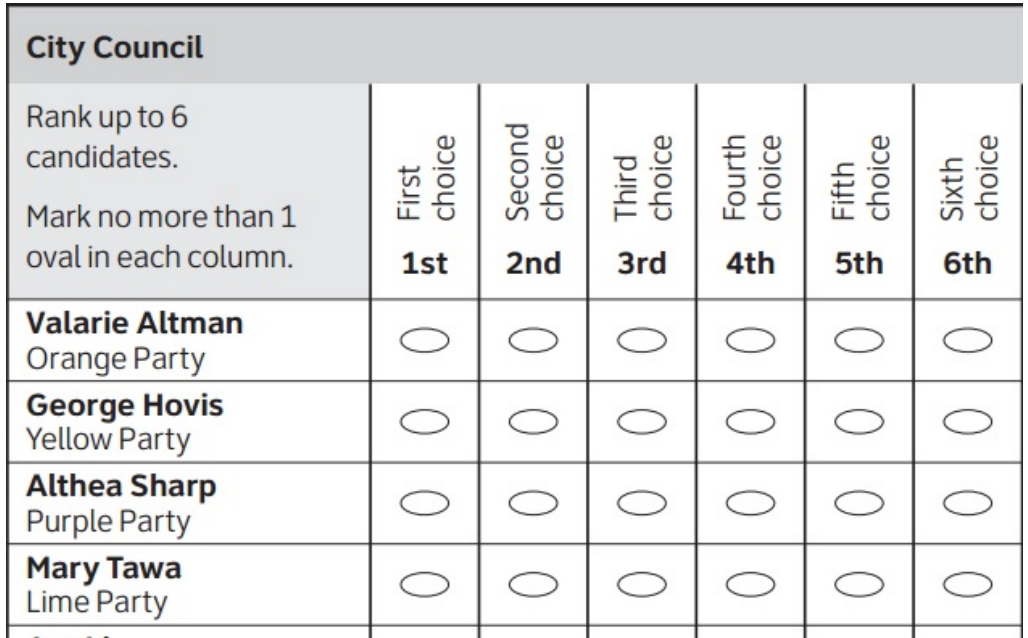
Multi-member districts



Multi-member proportional districts, such as Ranked Choice, are the solution



- Gives voters more expression
- Eliminates the need for party primaries
- Gives third parties a chance
- More proportionally represents minorities
- Produces a diagonal seats-votes curve (instead of sigmoidal)
- **Eliminates gerrymandering**



City Council						
Rank up to 6 candidates. Mark no more than 1 oval in each column.	First choice **1st**	Second choice **2nd**	Third choice **3rd**	Fourth choice **4th**	Fifth choice **5th**	Sixth choice **6th**
Valarie Altman Orange Party	⬭	⬭	⬭	⬭	⬭	⬭
George Hovis Yellow Party	⬭	⬭	⬭	⬭	⬭	⬭
Althea Sharp Purple Party	⬭	⬭	⬭	⬭	⬭	⬭
Mary Tawa Lime Party	⬭	⬭	⬭	⬭	⬭	⬭

Visualizing the Genetic Algorithm

- **The Genetic Algorithm is a probabilistic binary search through a very high dimensional space, operating on each dimension in parallel**
  - Think about the negative space – the candidates that *aren't* selected for recombination. These are areas of the solution search space that are being eliminated.
  - Each "gene" is a dimension. So this elimination is happening in parallel on all dimensions ("genes" and even gene combinations) at once
- Since it's a binary search, and thus eliminates areas of the search space exponentially, **the mutation rate should drop exponentially** over time ("annealing") to confine the search space
- When the mutation rate is very small, it's "complete".

Tacit assumptions of the Genetic Algorithm

- **Randomly combining two maps with good scores is more likely than not to be produce a map with a better score than randomly combining two maps with bad scores.**
  - This is the underlying assumption of a genetic algorithm. If this is false, the algorithm can't work.
- **High score (or conversely low score) represents good fulfillment of the objective.**
- **A score can take on many different values.**
- **Time to calculate the score does not grow too fast with the size of the solution (in bits).**
  - In computer science, computation time is measured in what's called "Big-O notation". In Big -O notation, we are concerned only with how the number of computations scale with the number of data points. "N" signifies the number of data points. For instance if we are sorting a 52-card deck, N=52. If to sort them, we have to compare every card to every other card, then the number of computations is proportional to N*N. (aka N^2) This would take too long to calculate for a genetic algorithm to be practical.
  - However, genetic algorithms only need an *approximately* correct scoring system to work, so they can use shortcuts or heuristics instead of exact scores. And almost all problems admit linear-time (O(N)) heuristics.

Reading and writing shapefiles in Java

- Shapefiles are open data formats, thoroughly spec'd by ESRI
- Consequently, many open-source libraries exist for reading and writing them in various programming languages, including libraries published by ESRI
- To keep the code easy to maintain, I selected a very small and simple library for reading the shapefile polygons
- Separately I found a very small and simple library for reading and writing the .dbf (dbase) file that stores the tabular data.

Accidental Gerrymandering and Self-Sorting

- Geography is **not** a neutral criteria
  - Democrats tend to concentrate in urban areas
  - They are unintentionally “packing” their own votes, reducing the number of congressional seats they can get

- Consequently, party-blind redistricting leads to systemic partisan bias
  - “Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures” - *Jowei Chen and Jonathan Rodden*
  - Driving blind is no way to avoid hitting pedestrians

- Without proper analytic tools and training, an independent commission can still gerrymander – accidentally.
  - A carpenter is only as good as their ruler
  - Independent commissions are necessary **but not sufficient.**

The $$ cost of Gerrymandering

Redistricting legal fight on track to cost Wisconsin taxpayers $3.5 million

Patrick Marley, Milwaukee Journal Sentinel Published 12:19 p.m. CT Jan. 22, 2019 | Updated 5:29 p.m. CT Jan. 22, 2019

North Carolina's redistricting cases cost taxpayers $5.6 million and counting

By Taft Wireback taft.wireback@greensboro.com Jan 6, 2018

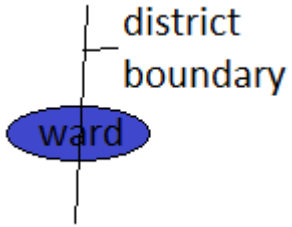
Public Could Pay $10 Million for Redistricting Lawsuits

Virginia taxpayers could be on the hook for more than $10 million in legal fees related to three redistricting lawsuits.

July 17, 2018, at 11:52 a.m.

Complications – Getting the data!

- There is no central national repository – or standardized format – for election vote counts integrated with geospatial data. Each state publishes their own data separately.
  - Florida mails you a CD because apparently they don't have … email?!?
  - Some open source efforts exist, but the naming and formatting is not consistent enough for automation
- In many states, district boundaries don't follow voting ward boundaries, splitting voting wards in half
  - Just make equal population voting wards first,
    and then make the electoral districts out of the voting wards.
    (Wisconsin does this and it works great!)



- Many states will change their voting wards mid-decade
  - Why?!
  - Makes data analysis difficult
    - You have to de-aggregate to census block level, then re-aggregate to the new (or old) districts

Probability of Map vs. Probability of Outcome

Probability of a **map** being equally gerrymandered or more given **one** election's vote counts

- The random map generating algorithm implicitly pulls from a probability distribution of maps
  - Which is not explicitly stated
  - And there is no empirical evidence for
- Can't extract probability densities about the election outcomes for a given map
  - Doesn't demonstrate durability of the gerrymander over multiple elections

Probability of an **outcome** given the map and **multiple** election's vote counts

- Doesn't assume a probability distribution of various maps
  - Fit a Bayesian prior distribution from actual vote counts using the Empirical Bayes Method
- Can extract various probability density functions for the map
  - Including durability over multiple elections

Aggregating the needed data

- This was way too much work! This should be compiled by the federal government in a simple and consistent format!
- Initial shapefiles of voting tabulation districts from census.gov
  - (4 states were not available, so had to use census tracks instead)
- Population and demographics from census.gov (block files .csv)
- (estimated) Ward-resolution presidential vote counts from Stephen Wolf of DailyKo's google drive (Thanks to FairVote.org for referring me to this.)

History & Motivation

- A friend of mine become my state assembly person
- He posted something on social media about using independent redistricting committees to end gerrymandering
- I thought to myself: that’s not a solution, you need to know how to optimize it for fairness, and for that what you need is an algorithm and a fast computer
- The algorithm was an interesting intellectual problem: optimizing multiple conflicting criteria at once, related to regions defined on a space, so I pondered it (I enjoy reasoning spatially)
- After I thought I had all the pieces solved, I wanted to see if it worked, so I built what was in my head
- And because the software was unique and had a noble purpose, my motivation kept up to see it through to all the way
- (And yes it took a long time to write!)

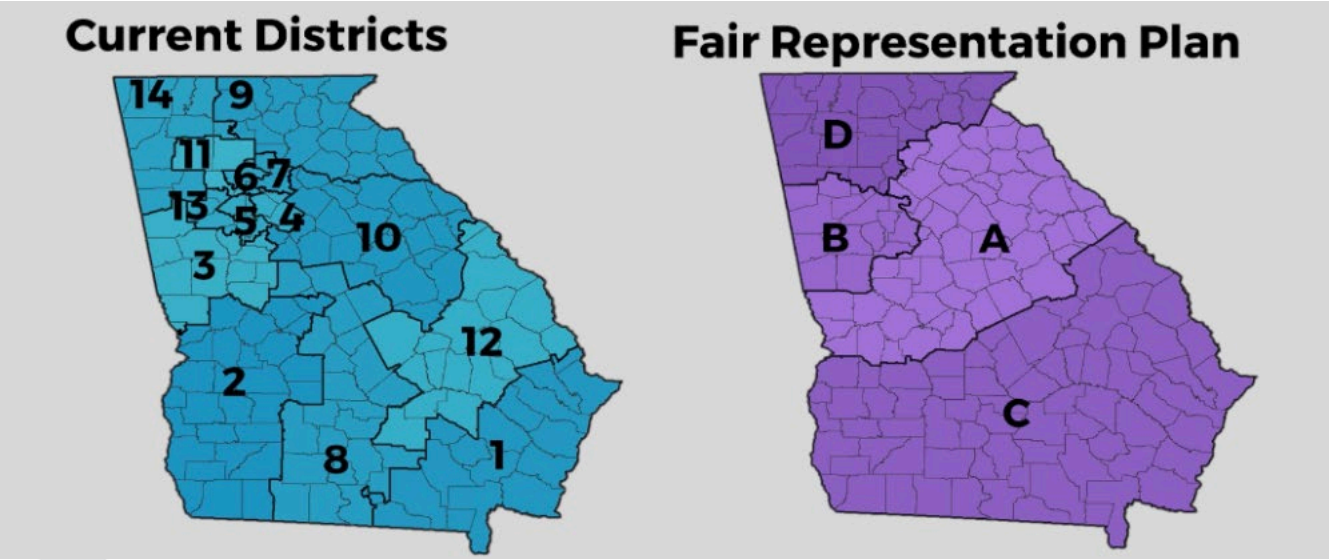
Multi-objective heuristic optimization

The Genetic Algorithm is part of a larger class of algorithms called "Multi-objective heuristic optimization". *These algorithms are not deterministic.* They are random.

This class of algorithms is used to find good solutions to multiple goals at once, when deterministic methods are not feasible.

- Other algorithms in this class include:
  - Particle swarm optimization
  - Ant colony optimization
  - Learning classifier systems
  - Simulated annealing
  - Stochastic gradient descent

Multi-Member districts / Ranked Choice

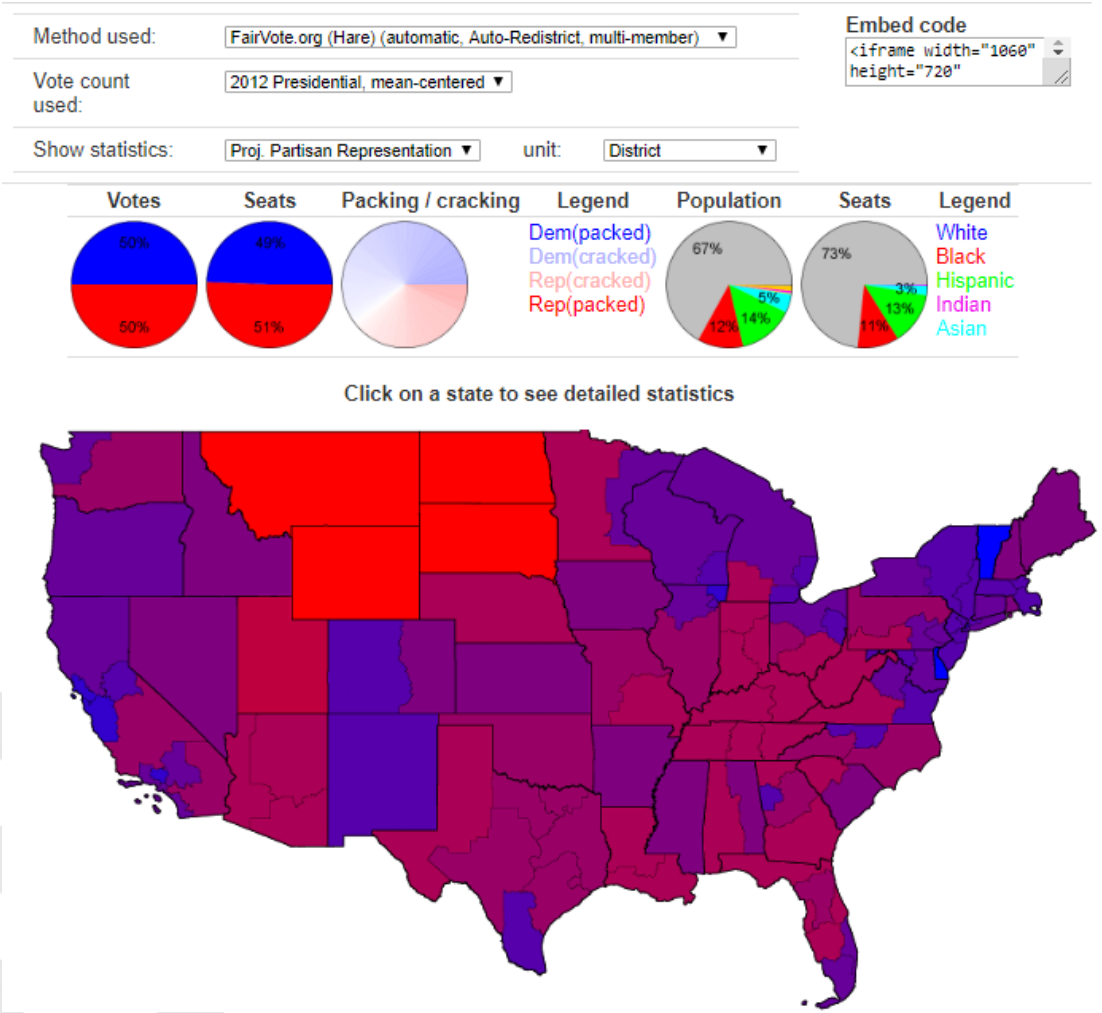


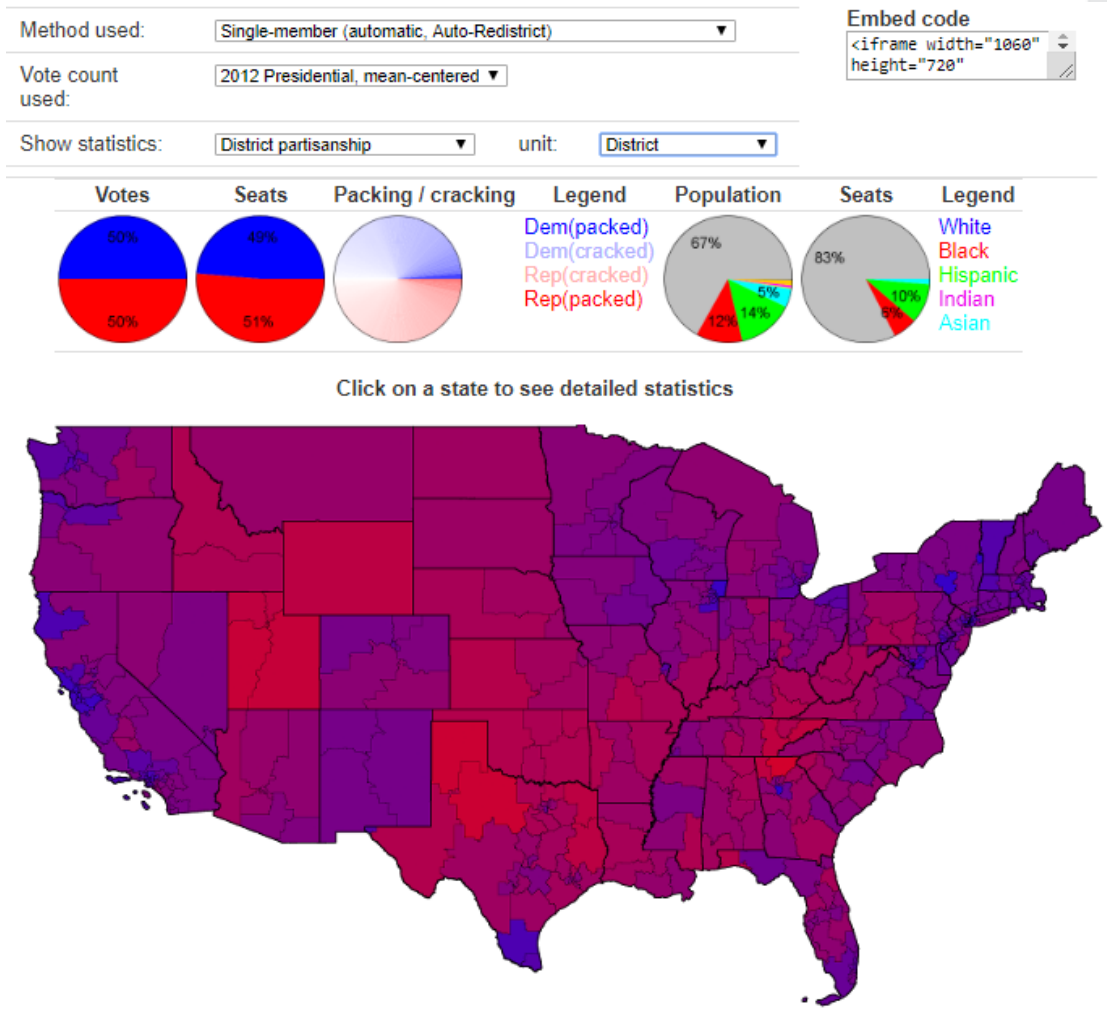
City Council						
Rank up to 6 candidates. Mark no more than 1 oval in each column.	First choice **1st**	Second choice **2nd**	Third choice **3rd**	Fourth choice **4th**	Fifth choice **5th**	Sixth choice **6th**
Valarie Altman Orange Party	⬭	⬭	⬭	⬭	⬭	⬭
George Hovis Yellow Party	⬭	⬭	⬭	⬭	⬭	⬭
Althea Sharp Purple Party	⬭	⬭	⬭	⬭	⬭	⬭
Mary Tawa Lime Party	⬭	⬭	⬭	⬭	⬭	⬭
Joe Li Tan Party	⬭	⬭	⬭	⬭	⬭	⬭
Phil Wilkie Independent	⬭	⬭	⬭	⬭	⬭	⬭

Generating National Maps for FairVote

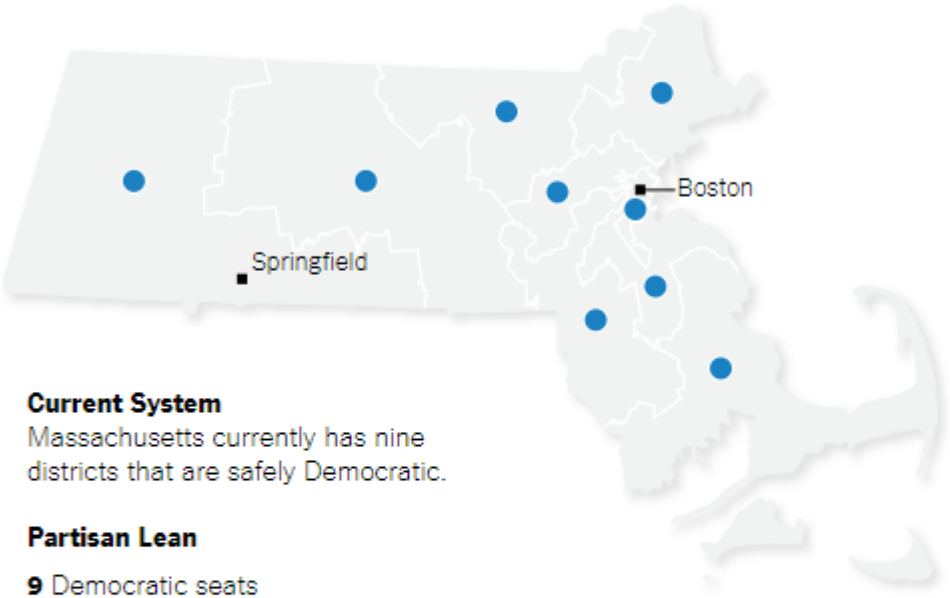
- Added option for multi-member with “Hare quota” to AutoRedistrict
- Added FairVote.org’s rules for seats per district (3/4/5)
- Recorded a script using AutoRedistrict’s instruction window
- Wrote a linux shell script to copy that script 50 times, changing the state and seat counts
- Wrote a small script to run those 50 scripts
- Wrote php and javascript for the interactive map
- Wrote php scripts to aggregate the statistics of all 50 maps, and sent links to FairVote.org for analysis.

Interactive maps compiled from AutoRedistrict exported renderings and stats





NYT Maps – expanded house, multi-member



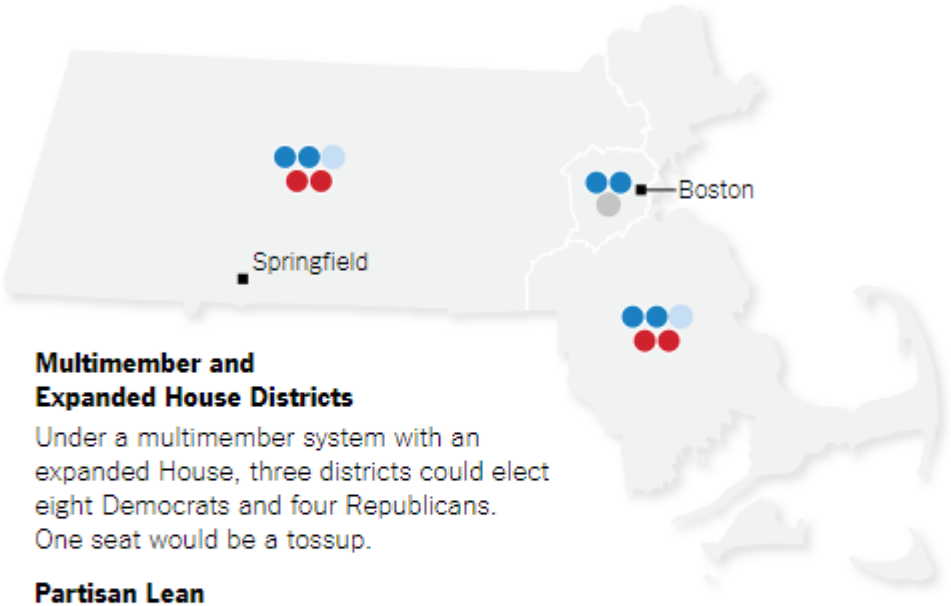
Current System
Massachusetts currently has nine districts that are safely Democratic.

Partisan Lean

9 Democratic seats

0 Republican seats

0 Toss-ups



Multimember and Expanded House Districts
Under a multimember system with an expanded House, three districts could elect eight Democrats and four Republicans. One seat would be a tossup.

Partisan Lean

8 Democratic seats

4 Republican seats

1 Toss-up

Transparency

- Due to the political nature of redistricting, I decided that I had to make AutoRedistrict Free and Open Source.
- The code (and executable) is hosted on github.
- It’s licensed GNU-GPL 3. (“copy-lefted”)
- The shapefiles for the interactive map are published on my ftp site.
- The program records all actions in a script and you can play it back or use the script on a different state. So even the settings and process is transparent. (and reusable!)

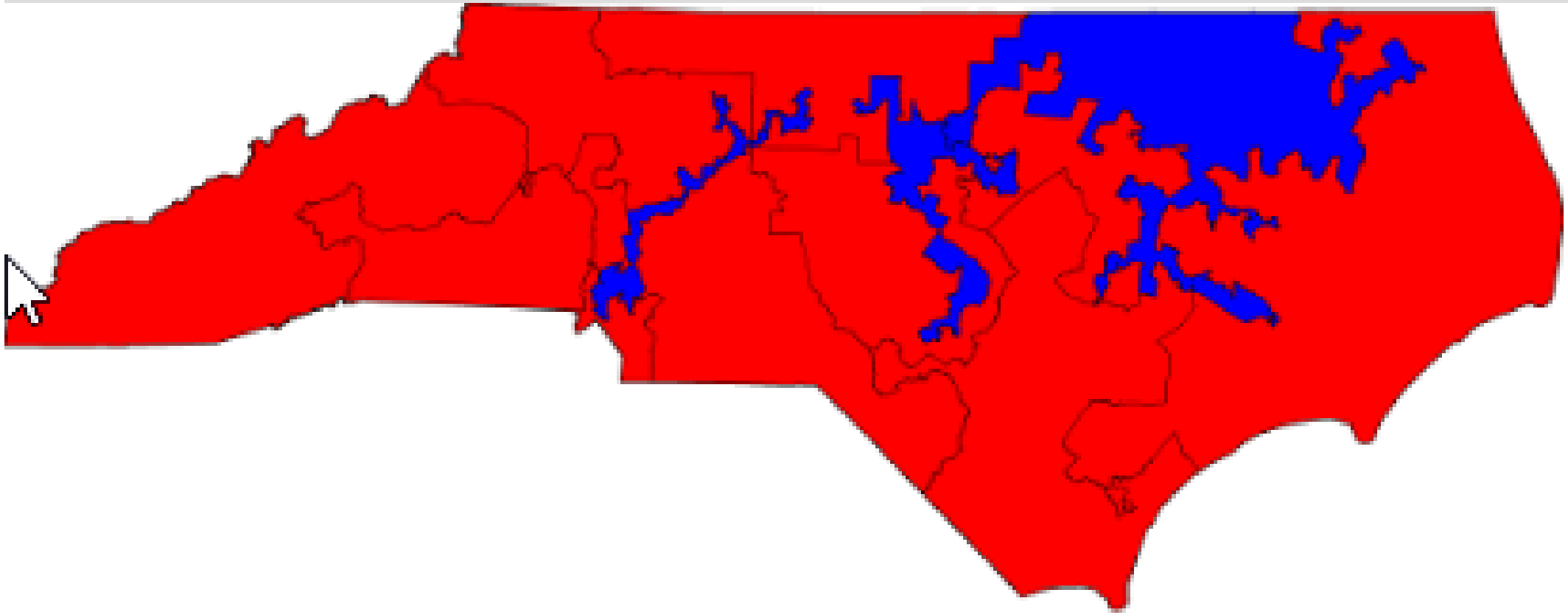
About myself

- Fully self-taught software developer (no college)
- I've been programming since I was a little kid
- Interested in artificial intelligence, simulation, and modeling
- Senior Systems Analyst in the Government sector
- I wrote AutoRedistrict in my spare time
  - Eliminate gerrymandering
  - Makes redistricting cheaper

Wisconsin Assembly, before and after Gerrymandering, Methodology Highlights

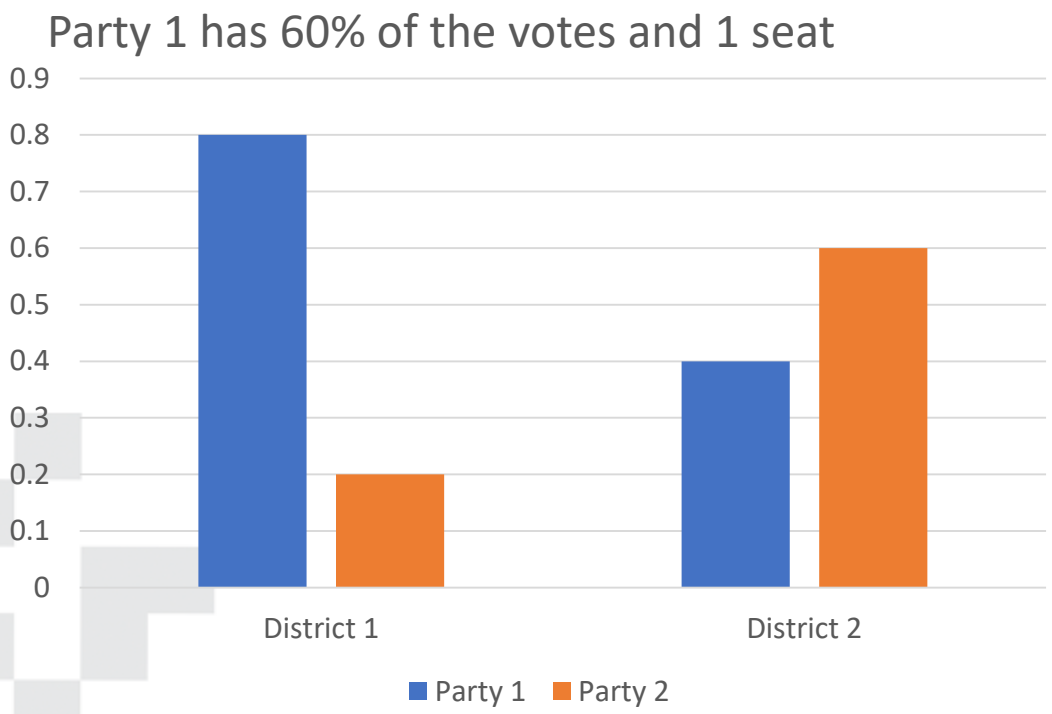
- Highest resolution available
- Equal number of elections aggregated forward and back (3 & 3)
- Same exact elections are used in both before and after picture
- Can’t argue that it’s caused by changes in voter sentiment over time, because they both cover the same time period
- Can’t argue that the gerrymandering is a natural consequence of geography, if it’s absent in the before picture

Gerrymandering is *extreme*

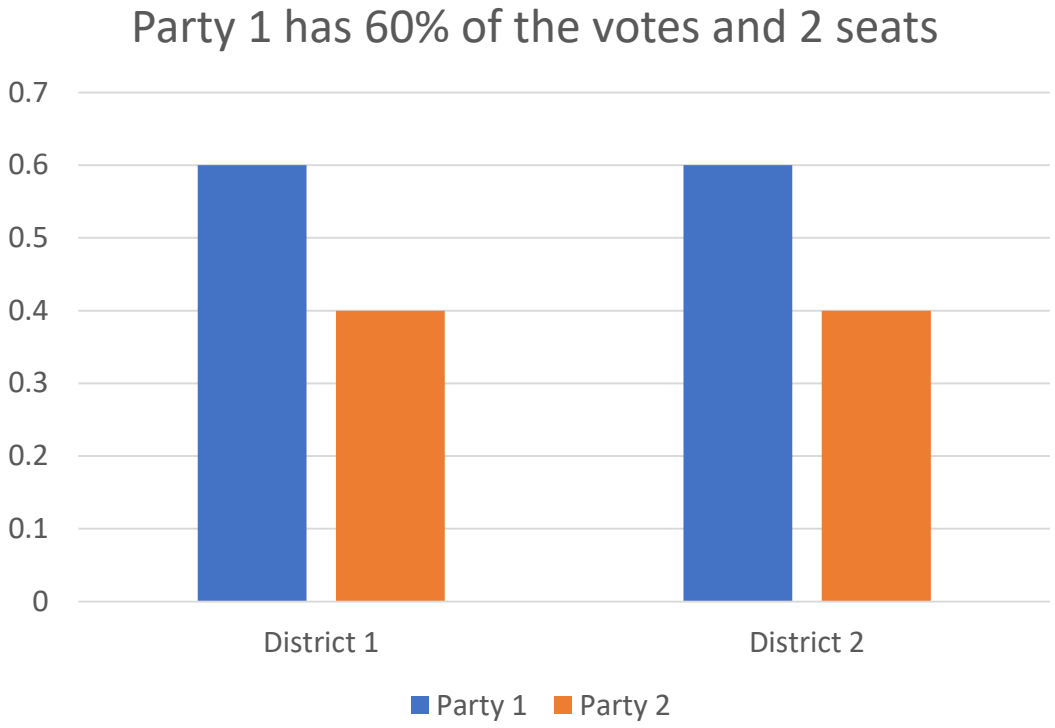


How gerrymandering works

Packed



Cracked



Justice Delayed is Justice Denied - indefinitely

- ~~2012 election~~ irreparable harm
- ~~2014 election~~ irreparable harm
- ~~2016 election~~ irreparable harm
- ~~2018 election~~ irreparable harm
- 2020 election only one election left,
  - then the same people who gerrymandered last time will be able to gerrymander again,
    - securing stolen seats for another 10 years,
      - a new lawsuit is filed,
        - and the cycle repeats.

80% of the damage is already done

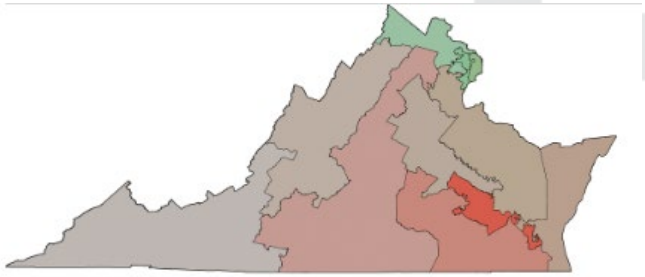
…and when it hits 100%,
it repeats.

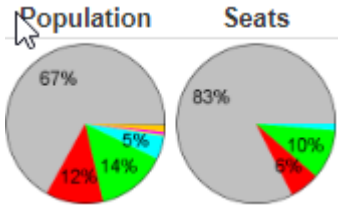
Means of prevention

- Early identification (identify and contest gerrymandering *before* maps are approved)
  - Requires transparency
  - And automated tools
- Clear (and mathematically sound) legal standards
  - Automated redistricting tools to enforce them
- Good actors
  - Independent commissions
  - Oversight

Integrated analytics

- Maps
  - Population density, partisan swing, demographics, etc.
  - Per capita, per precinct, per district
- Charts
  - Pie charts, seats-votes pictures, probability densities
- Tables
  - By district, by party, by demographic, global statistics
- All exportable





District	Population	Winner	PVI	Vote gap	Wasted votes	FV Safe D	FV Lean D	FV Tossup	FV Lean R	FV Safe R	Pop per seats	Compactness
1	725,957	0,1,	R+5	32958	152397	0.0	0.0	0.0	1.0	0.0	725,957	0.19488820216547642
2	708,231	0,1,	R+1	7646	141899	0.0	0.0	1.0	0.0	0.0	708,231	0.2585711092796171
3	747,534	1,0,	D+29	168902	146745	1.0	0.0	0.0	0.0	0.0	747,534	0.08282255183975036
4	726,613	0,1,	R+4	26480	159305	0.0	0.0	0.0	1.0	0.0	726,613	0.22932503100987145
5	727,430	0,1,	R+5	30983	159783	0.0	0.0	0.0	1.0	0.0	727,430	0.16240362377197312
6	727,447	0,1,	R+10	59144	147375	0.0	0.0	0.0	0.0	1.0	727,447	0.1891544244160684

*Prevention is much better than correction

Prevents a lot more harm

- Prevents irreparable harm to voters and the country each election by keeping elections truly democratic
- Prevents irreparable harm to the world via policy (e.g. climate change)

Correction often simply **isn't an option**

- Lawsuit is the only means, and defendants will use delay tactics and the courts move very slowly and are hesitant to rule
  - Many years and millions of dollars later, nothing

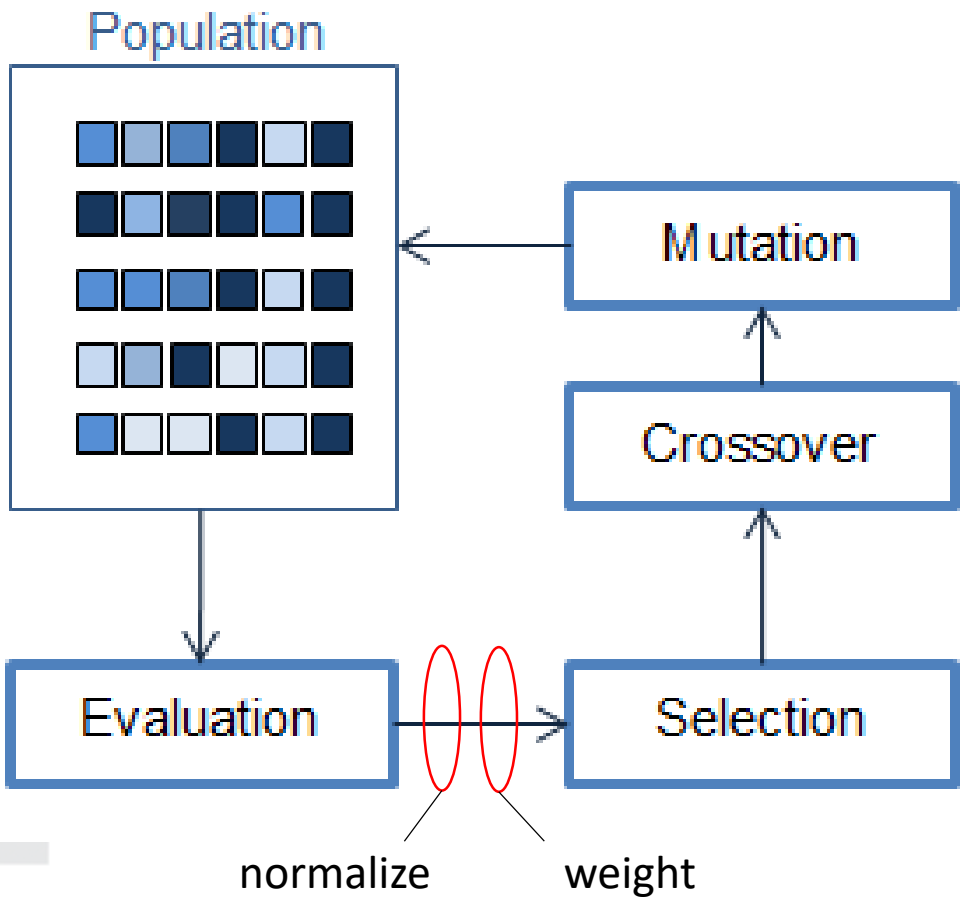
Gerrymandering determines the policies

- Takes the power to control the composition of congress away from the citizens - turns democracy into autocracy
- Determines who has control of congress
  - Which determines the laws and policies, which affect the country
- You can’t impact what policies get implemented if you can’t impact the composition of congress

Means of prevention

- Early identification (identify and contest gerrymandering *before* maps are approved)
  - Requires transparency
  - And automated tools
- Clear (and mathematically sound) legal standards
  - Automated redistricting tools to enforce them
- Good actors
  - Independent commissions
  - Oversight

The Genetic Algorithm: Steps



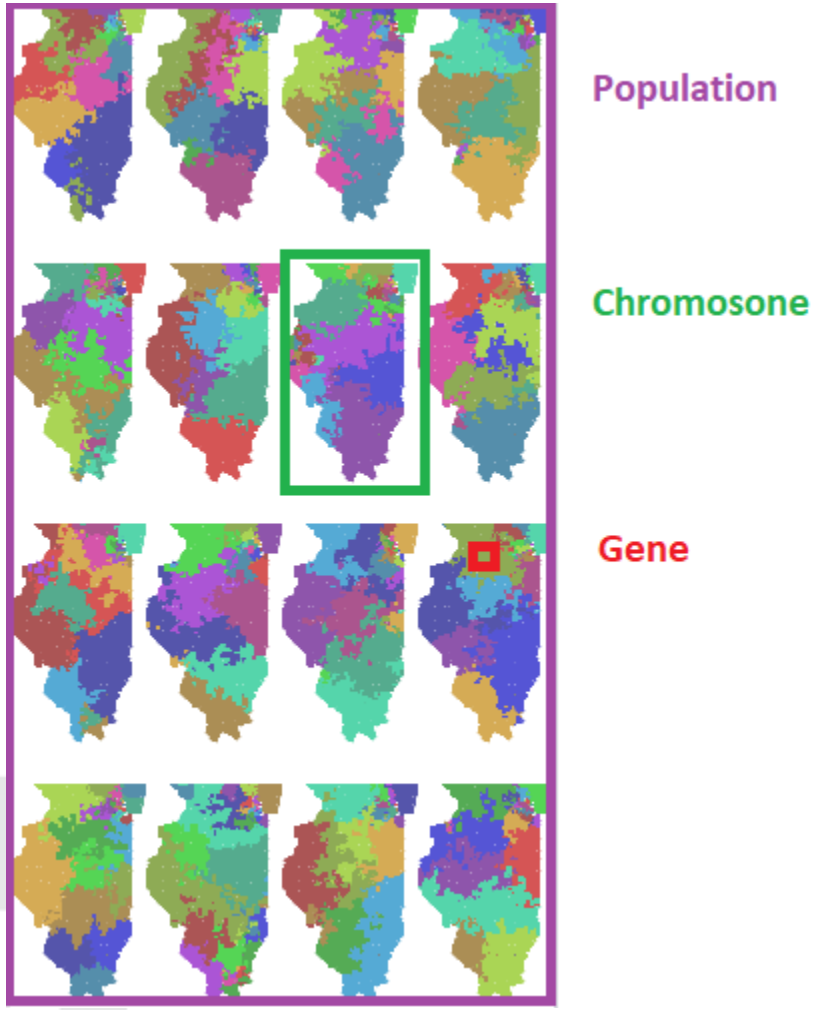
1) Evaluate – score the fitness

1.2) Normalize - map all scores into a fixed and smooth range

1.3) Weight – Adjust importance of sub-scores

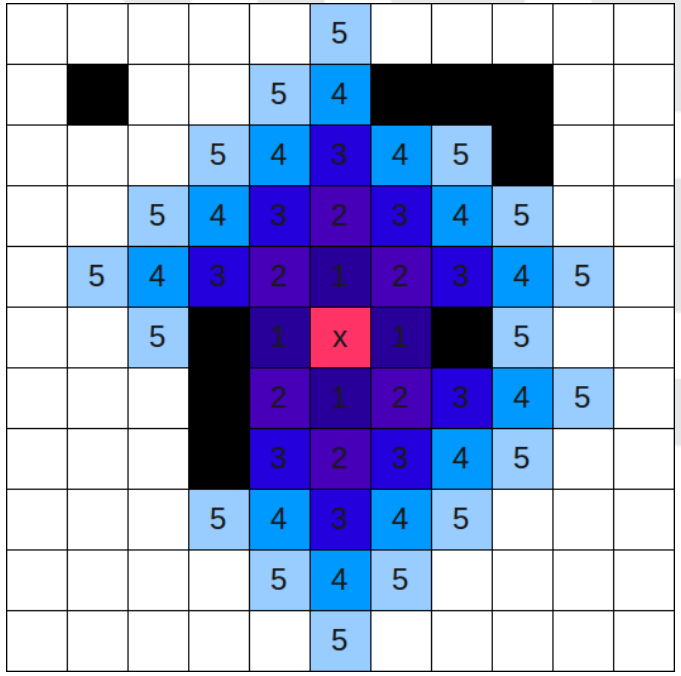
2) Select – A few different methods, truncation is the simplest

The Genetic Algorithm: Definitions



- The "**Population**" is the set of all maps currently being evaluated
- Each "**Chromosome**" is a map
- The "**Genes**" are a list of what district each voting ward is assigned to (in the order that the voting wards occur in the tabular data)
  - Ex. 1,5,3,1,2,6….

Step 0: Initialization

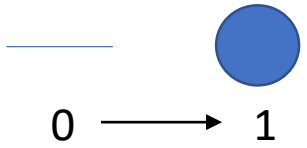


- A random “seed” ward is chosen for each district
- Remaining districts are added via a
  randomized breadth-first flood fill
- All districts are flood-filled at the same time,
  with the lowest population district always taking the next fill iteration
- This results in
  - Roughly equal population districts
  - That are fairly compact
  - And contiguous
  - And yet are random
- Time to complete is proportional to number of voting wards (O(N))

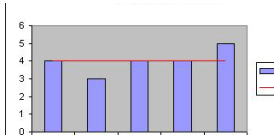
Step 1: Computing the scores – The scores

- Dis-contiguity is a district's total population, minus the population that's not in the most highly populated region

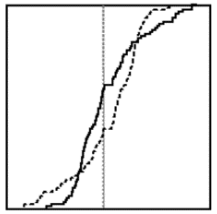
- Compactness is a district's area divided by it's perimeter squared ("isoperimetric quotient") $\frac{4\pi A}{p^2}$



- Population inequality is total squared deviation from perfect equality

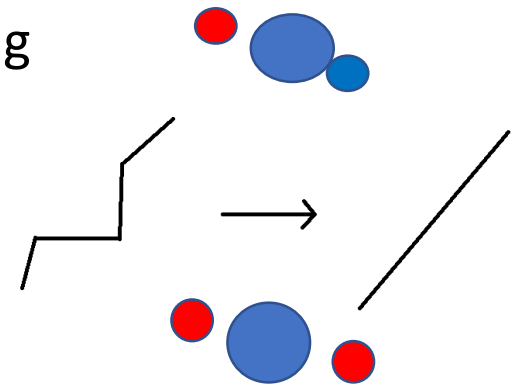


- Partisan gerrymandering is computed from the seats-votes curve and will be explained later in this slideshow

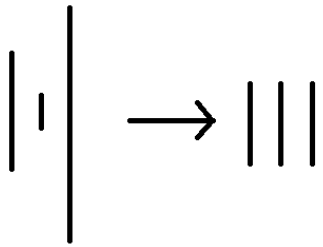


Normalization - Why

- Smooths out sudden jumps in scores
  - Prevents a single optimization from dominating
  - Gives tough-to-optimize areas a boost

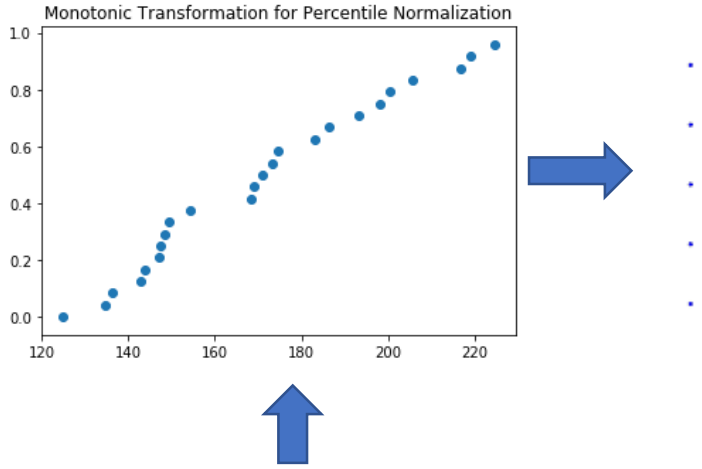


- Puts all criteria on the same scale
  - e.g. population is in thousands,
    compactness is between 0 and 1
  - Equalizes their impact



Normalization - How

- “Rank normalization” is used
  - For each criteria, order all maps from best to worst
  - Replace their “score” with their place in that order
  - Produces evenly spaced scores



evenly spaced scores

unevenly spaced scores

Step 2: Selection

- “Truncation selection” is used
  - Just select the top 50% or so (adjustable)
  - This is the simplest selection method – there are others

- An adjustable amount of “Elitism” is included
  - “Elitism” is where some of the “parents” survive for multiple generations
  - This allows them to pass on more information to future generations
  - Keeps the gene pool closer to recently discovered optima
  - Has been shown to improve convergence

Step 3: Recombination

- Two parents are selected at random from the survivors
- “Uniform recombination” is done
  - As opposed to “single point crossover”
  - For each voting ward, one of the two parents is selected at random
- This can add discontinuities to the maps, but those are slowly weeded out by selection pressure