Jonathan Wise
Math 195
Final Project
Due: 12/16/22

# CONSIDERING COMPETITIVENESS
A Fresh Look at Political Redistricting

## 1. PURPOSE

---

*This study develops a model to analyze and compare the competitiveness of districts in a redistricting ensemble - identifying potential for greater competitiveness through redistricting. We run election simulations using historical voting datasets to study trade-offs, risks and opportunities of prioritizing competitiveness, and look for evidence of gerrymandering in current political redistricting maps that intentionally squelches competitiveness.*

*This study also places our three metrics of competitiveness, our use of ensembles, and our selection of data in the context of the other metrics and methods studied in Math 195 this semester.*

*The approach of this study could be of interest to other students and researchers who are interested in alternative ways to evaluate redistricting plans.*

*The results of this study could be of interest to the general public who share concern about the declining number of competitive districts, and those who are interested in new ways to think about the redistricting plans.*

---

*Political redistricting* is the process of drawing new electoral district boundaries that account for population shifts in each decennial census. Citizens generally expect redistricting to be conducted fairly, so that elections appropriately reflect the population. But, in addition to the many definitions of “fair” districting, there are many groups seeking advantages, many legal challenges, and many competing mathematical approaches and metrics.

*Gerrymandering* is the manipulation of political boundaries to favor a particular group. Such manipulation is typically considered unfair, and increasingly detected using mathematical methods and metrics.

*Competitiveness* is one (of many) dimensions of fairness. In a two party system, a *competitive* district is one that has balanced numbers of Democrat and Republican voters, it could be won by either party, and it typically has been won by each party. There are pros and cons of competitiveness: advocates argue that fewer "safe" districts will encourage stronger and more balanced candidates; opponents argue that more "competitive" districts may result in more significant swings of proportionality from election to election.

Political analysts find that the number of competitive districts is decreasing. It could be one explanation for the polarization, extremism and lack of compromise seen in congress, as safer, less-competitive districts do not incent candidates to reach across the aisle in compromise.

A deeper dive into current competitiveness by district is included in section 7 below, with the following conclusions:

- Competitive districts reduced from 46 (10.6%) to 40 (9.2%) of 435 total seats during the 2022 redistricting process
- Across the US, each state currently has between 0 and 4 competitive districts which account for 0% to 100% of their total districts
- 25 of the 50 states have zero competitive districts
- Among the states with at least one competitive district, their average number of competitive districts is 1.6, and their median number of competitive districts is 1.

## 2. SUMMARY OF FINDINGS & RECOMMENDATIONS

---

> *This study provides a framework to evaluate the degree to which political redistricting can influence competitiveness, by evaluating distribution of competitiveness in state-by-state redistricting ensembles.*
>
> *It results in a model that can be efficiently applied to every US state, allowing its competitiveness to be compared to other states and its own ensemble. This is accomplished through three Competitiveness Metrics for each state:*

(1) *Absolute Competitiveness Score on a scale of 0-1*
   - *This is the percentage of competitive districts in each state*
   - *A number close to zero indicates a small fraction of competitive districts*

(2) *Competitive-Completeness index on a scale of 0-1*
   - *This measures the fraction of Absolute Competitiveness scores in the redistricting ensemble that are at or below the state's current level.*
   - *A number close to zero indicates that a state is an outlier compared to its ensemble in terms of too-few competitive districts.*

(3) *Competitive-Potential index is on a scale of 0-1*
   - *This measures the fraction of "potential" competitive districts remaining to be added, according to its ensemble.*
   - *A number close to one indicates significant room for improvement.*

---

Applying the model to Virginia as an example:

(1) Virginia's Absolute Competitiveness Score is .091 on a scale of 0-1
- This compares to a nationwide Competitiveness Score of .092

(2) Virginia's Competitive Completeness index is 0.8 on a scale of 0-1
- This means Virginia's Competitiveness Score level is at or above 80% of the Competitiveness Scores of the plans in its redistricting ensemble.
- This indicates that Virginia is not an outlier compared to its ensemble in terms of too-few competitive districts.

(3) Virginia's Competitive Potential index is 0.67 on a scale of 0-1
- This means that, according to its ensemble, it may reasonably achieve a level of 3 competitive districts by adding 2 competitive districts.
- This indicates that Virginia has reasonable room for improvement.

Importantly, we developed this model in a way that enables it to be applied to any state. The code is available for general, public use via https://github.com/JonnyWise25/Math195Final , and the folder states/VA/Results, contains results from the study,

Key recommendation is to apply the model to all states to better understand their relative competitiveness and opportunities for change.

## 3. LEGAL STANDARDS, GERRYMANDERING AND COMPETITIVENESS

---

*Competitiveness is not a legal standard for political redistricting. But as citizens and courts consider factors such as compliance with the Voting Rights Act, gerrymandering and extreme partisanship, competitiveness could become a more interesting fairness indicator. Advocates for competitiveness hope to prevent too many safe districts with one-party dominance that lack inter party competition.*

---

***The Constitution*** requires that district maps be reviewed every ten years (following the decennial census), guaranteeing equal representation with equal population within each district. Federal law also requires compliance with the Voting Rights Act which protects the voting power of minority groups. And citizens generally expect political districting maps to be logical, explainable and fair.

Other guidelines within each state typically include:

- Contiguity (districts cannot have disconnected parts)
- Compactness (favoring plump versus snake-like districts)
- Preservation of communities of interest, cities, counties, and/or prior districts where possible and appropriate

Some states also have guidelines such as ignoring or considering incumbents, ignoring or considering partisan political data and proportionality, ignoring or considering inter-party competitiveness, ignoring or considering citizen input, allowing or disallowing multi-member districts.

***Gerrymandering*** involves manipulation of political boundaries to favor or suppress a particular group, such Democrats or Republicans, or a minority group.

***The Voting Rights Act*** covers the 15th Amendment's promise to provide equal access to voting. It prohibits restricted voting access based on race or color. It also prohibits any dilution of voting strength that makes it impossible for compact, connected minority groups to achieve proportional representation. The precedent-setting Thornburg versus Gingles case in 1986 made the Voting Rights Act actionable by highlighting three special *Gingles Factors* sufficient to claim violation of the Voting Rights Act: (1) the minority group is large enough and compact enough to constitute a majority in an additional single, legally structured district; (2) the minority group is "politically cohesive" and would vote for common leaders and causes; (3) the majority group also votes as a bloc in a way that prevents the minority group from having an impact with its votes.

***A Constitutional challenge to Gingles Factors*** is now underway as part of the Merrill vs Milligan case being considered by the Supreme Court. Milligan argues that the most recent redistricting plan for Alabama violates the Voting Rights Act by diluting Black voting power. Merrill argues that Alabama's redistricting plan is race neutral, and Milligan's demand is unconstitutional. Alabama's districts have looked largely the same for decades, and since 1992 it has had only one majority Black district. The most recent redistricting plan generally follows historic district lines and respects the industries in the area. Merrill claims that the only way to draw a second majority Black district is to intentionally sort by skin color, which violates the Constitution. Milligan argues that the Voting Rights Act cannot be addressed with racial blinders. Race must be considered to remedy race discrimination.

If the Supreme Court rules in favor of Merrill, disallowing the consideration of race, the standards of redistricting will change significantly, weakening today's methods of evaluating adherence to the Voting Rights Act. This kind of change could make alternative measures such as competitiveness more important in the future.

## 4. REDISTRICTING METRICS AND COMPETITIVENESS

---

*Competitiveness is one of many metrics that can be considered to evaluate political redistricting maps. A number of commonly quoted metrics are designed to detect gerrymandering. Subsets of these metrics consider: simulated impact of voters; population alone; geography alone; composite effects.*

*The summary of metrics in the section below provides perspective about the importance of also considering competitiveness.*

---

***Fairness and discrimination*** are considerations for many redistricting metrics. Some look at how new districting plans consume historical votes (vote-based), others look at standalone factors about the districts (independent of votes). Each has advantages and disadvantages. From my perspective, a fair redistricting plan considers the people, communities and geographies of a state. Importantly, a fair redistricting plan does not 'artificially' suppress the representation of any group. I believe that there may not be a "single" metric that confirms fairness, because each metric tends to focus on a subset of dimensions of a multi-dimensional issue. For example, I believe that fairness requires redistricting plans that allow for both *proportionality and competitiveness*. I would seek a solution as fair as possible (proportional, symmetric, efficient, *and competitive*) when considering the full range of options for that state via ensemble-based metrics.

***Population Balance and Contiguity*** are essential for redistricting plans. All districts must have an equal share of the population (within a pre-specified tolerance) and a single district must be geographically connected (which can become ambiguous around bodies of water or other geographic and social divides).

***Reasonable Compactness*** is a typical standard for redistricting plans, intended to avoid meandering, gerrymandered district borders. But good measures for compactness continue to evolve. For example, in a recent Supreme Court hearing on Merrill v. Milligan, Justice Brett Kavanaugh claimed: "I don't understand how to measure whether something is reasonably compact." Geographic measures of compactness consider circles to be perfectly compact, and essentially measure circle-like attributes of each district using scores such as Polsby-Popper, Schwarzberg, Reock, and Convex Hull.

***Cut-Edge Scoring for Compactness*** is built on the premise that fair redistricting plans do not artificially divide or combine communities. It measures how often neighboring communities are divided in a given redistricting plan, and how often non-neighboring communities are artificially combined (across mountains, deserts or large bodies of water.)

> Advantage: Considers the combination of community and geography and does not penalize natural boundaries
> Disadvantage: Requires a fair precinct-based building blocks that become the nodes in the cut-edge network, and an objective standard by which nodes or precincts should be considered neighbors

***Proportionality*** (vote-based)*:* If V is the vote share (proportion) for a given group in a districting plan and S is their corresponding seat share. Then V=S indicates a perfectly proportional districting plan. Measures of proportionality reflect how much the actual result deviates from V=S.

> *Advantage:* Proportional districting plans feel fair to most citizens
> *Disadvantage:* Proportionality may be completely incompatible with some geographic population distributions; Guarantees of proportionality can also be incompatible with competitiveness (which is another important characteristic of fairness).

***Partisan Symmetry*** (vote-based)*:* This test accepts that a districting plan may not have perfect proportionality (V = S), and it instead looks for symmetry as a test of fairness. For example, Partisan Symmetry uses seat-vs-vote curves to model what would happen to each party's seat share if they gain or lose a certain percentage of votes in each district of the plan. To evaluate the symmetry, the seat-vs-vote curves are compared to each other, seeking similar behavior.

> *Advantage:* Symmetry of seat-vs-vote curves are visual and easily explained
> *Disadvantage:* While multiple Partisan Symmetry scores have been considered, none of them tell a complete story of fairness.

***Efficiency Gap*** (vote-based)*:* This score compares the number of votes that are wasted by each party. For example, if one party loses a district, they waste all votes in that district, and if they win a district by a wide margin, they waste all votes in excess of 50%. A fair map would have both sides waste the same number of votes for an efficiency gap of zero.

> *Advantages:* This score effectively highlights unfair situations such as packing one party's votes into a single district or cracking one party's votes into many districts.
> *Disadvantages:* This score can swing dramatically from year to year in stations with competitive districts (where one party "just misses" a win, and therefore wastes many votes).

***Entropy and Conditional Entropy***: this metric can be used to measure many attributes of a districting plan: its degree of segregation, its degree of split-counties, its difference versus previous plans. It does this by tracking how randomized a set of data (such as a redistricting plan) is in comparison to the other factors.

> *Advantages:* Entropy is helpful in assessing whether a plan is an outlier by comparing it to the rest of the cluster, particularly considering the distance between plans. It can mathematically assess differences that are difficult to see with the naked eye.
> *Disadvantages:* being an outlier does not necessarily reflect gerrymandering because an outlier, in terms of distance between plans, simply implies that a plan is geographically different. It makes no comment about how the partisan statistics would change because it does not reflect vote count. It seems more likely that gerrymandering would occur with minor geographic changes that cause large representation swings. Those would not appear as entropy outliers.

***"Moran's I":*** this metric can detect clustering and dispersion of minorities by measuring how similar an object (like a census block) is to the objects surrounding it. To apply it to a particular attribute (like race segregation), consider the graph G=(V,E) where the vertices are census blocks connected by an edge if they are neighbors. "Moran's I" generates a score across the entire geography. If it is close to 1: there is strong segregation. If close to 0: there is no pattern of segregation. If close to -1: there is a fine-grained checkerboard pattern where many above-average vertices are neighbors of below-average vertices in terms of minority voters.

*Advantages:* Moran's I can identify the presence or absence of underlying population distributions sometimes described as unintentional gerrymandering
*Disadvantage:* the absence of unintentional gerrymandering as measured by Moran's I does not necessarily indicate intentional gerrymandering

## 5. EMERGING USE OF ENSEMBLE METRICS

---

*This study uses ensembles to analyze the reasonableness and remaining potential for a given state's absolute competitiveness score.*

*This is accomplished by measuring and comparing the competitiveness of each district in each redistricting plan, given actual historic votes. Then, by generating redistricting ensembles for each state, and applying the same measures, we can determine whether recommended redistricting maps are statistical outliers in terms of competitiveness, as well as the reasonable range of competitiveness that we could expect in a given state.*

*This approach mirrors existing, accepted ensemble-based analysis that was designed to consider representation and gerrymandering, but not competitiveness.*

*For context and comparison, a summary of existing indices appears below.*

---

***Ensembles*** are collections of thousands of redistricting alternatives. They provide a mathematical lens by allowing statistical comparison of large collections of competing redistricting plans. Ensembles do not offer an absolute measure of fairness, but rather computer-generate many, many options of district maps that allow comparison.

***Existing Ensemble Metric: Representative Index*** : This score contextualizes whether the number of elected representatives likely from a given plan is "typical" when compared to the ensemble. Rather than looking only at the final discrete number of representatives produced by each proposed plan (which could be constructed in a way that leaves some seats vulnerable and some seats secure), this index also captures the degree of safety in the seat victories by looking at the margin of victory for democrats and republicans in the closest race that they won (how close those seats were to flipping), and extrapolates between them.

*Advantages:* it provides a continuous view of "seats-won" that considers likely stability of the expected representation from the election

*Disadvantage:* this metric is non-intuitive and it can be difficult to explain. It also may underestimate the bias because it only looks at the next most likely seat to flip (rather than the next several seats); it does not reflect all aspects of competitiveness

***Existing Ensemble Metric: Gerrymandering Index*** : This index contextualizes the level of packing and cracking in a districting plan by comparing the actual partisan votes to the ensemble. For each plan in the ensemble, it sorts the districts from least to most democrat votes, and determines the average percentage of democrat votes per district. It then compares actual to average for every district in every plan.

Advantages: This index clearly represents packing and cracking, and its graph is easier to explain.

Disadvantages: It does not consider the direction of the deviation, and further analysis is required to determine whether that deviation may be in a direction that the state desires (for example responding to legislation such as the Voting Rights Act); it does not reflect all aspects of competitiveness

## 6. PROPOSED NEW ENSEMBLE METRIC: COMPETITIVENESS INDEX

---

*This section describes how we extend ensemble concepts to focus on competitiveness.*

*Because existing ensemble indices do not reflect competitiveness, we establish a new metric that can be easily calculated for all districts in a plan, aggregated for each plan in an ensemble, and clearly viewed as part of a histogram*

*Absolute Competitiveness Score on a scale of 0-1 for each plan*

- *This is the percentage of competitive districts in each state, calculated using historical votes*
- *A number close to zero indicates a small fraction of competitive districts*

*Then comparison of any given plan to the rest of the ensemble results in two new indices:*

*Competitive-Completeness index on a scale of 0-1*

- *This measures the fraction of Absolute Competitiveness scores in the redistricting ensemble that are at or below the state's current level.*
- *A number close to zero indicates that a state is an outlier compared to its ensemble in terms of too-few competitive districts.*

*Competitive Potential index is on a scale of 0-1*

- *This measures the fraction of "potential" competitive districts remaining to be added, according to its ensemble.*
- *A number close to one indicates significant room for improvement.*

*The Competitiveness indices offer an objective mathematical lens to determine whether today's redistricting plans are statistical outliers in terms of competitiveness, and whether there exist competitive alternatives to the existing redistricting plans.*

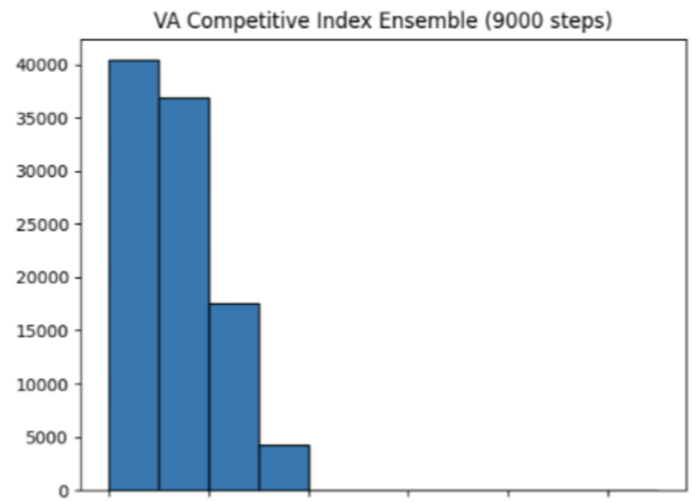
---

For the purpose of this study, we define a district as ***competitive*** if the winner of a recent two-party election (like the presidential election) won the two-party vote share by five points or less:

District_Gap = | dem votes - rep votes | / (dem votes + rep votes)

- For each district, if District_Gap < 0.05 (or five percent), that district is Competitive.
- For each districting plan, calculate the percentage of competitive districts is its ***Absolute Competitiveness Score***

For each ensemble, we can then create a histogram that shows incidence and likelihood of each discrete Absolute Competitiveness score for every plan in the ensemble:




Example of Histogram of Absolute Competitiveness Score for each of the ensemble

We can then compare the competitiveness of any given districting plan to its statewide ensemble to determine both its reasonableness (is it an outlier) and its improvement potential.

*Competitive-Completeness index on a scale of 0-1*

- This measures the fraction of Absolute Competitiveness scores in the redistricting ensemble that are at or below the given plan
- In the histogram above, the score would be approximately
    - 0.4 for a plan with zero competitive districts
    - 0.8 for a plan with one competitive district
    - 0.9 for a plan with two competitive districts
    - 1.0 for a plan with three competitive districts (since the ensemble indicates that it is unlikely to find a plan with more than 3 competitive districts
- A number close to zero could indicate that the given plan is a competitiveness outlier

*Competitive-Potential index is on a scale of 0-1*

- This measures the fraction of "potential" competitive districts remaining to be added, according to its ensemble.
- In the histogram above, the score would be approximately
    - 1.0 for a plan with zero competitive districts (it can still add 3/3)
    - 0.67 for a plan with one competitive district (it can still add ⅔)
    - 0.33 for a plan with two competitive districts (it can still add ⅓)
    - 0 for a plan with three competitive district (since the ensemble indicates that it is unlikely to find a plan with more than 3 competitive districts)
- A number close to one indicates significant room for improvement.

## 7. DEEPER DIVE INTO COMPETITIVENESS

---

*Before generating state-by-state ensembles, we establish the baseline of competitiveness in today's political redistricting plans. We draw the following conclusions below:*

- *Competitive districts in 2022 account for 40 of 435 total seats or 9.2%*
- *This is a reduction from 2020 when competitive districts accounted for 46 or 10.6% of the total seats*
- *Each US state currently has between 0 and 4 competitive districts, which is between 0% and 100% of their total districts*
- *Among the states with at least one competitive district, their average number of competitive districts is 1.6, and their median number of competitive districts is 1.*

*In the later phases of this study, we generate ensembles for each state to see how the competitiveness of the current plan compares to the range of competitiveness in the ensemble.*

---

Political scientists commonly define a district as competitive if the winner of a recent presidential election won the two-party vote share by five points or less, though some increase that span to up to eight points.

***Calculation of current competitive districts:*** As of 2022, it is widely observed that proposed redistricting plans reduce the number of competitive house seats among the 435 House districts. For this section of background perspective, this study used calculations provided by the FiveThirtyEight project, which applied 2020 presidential election results to each proposed new district, and identified the point spread between Trump and Biden. We identify a district as competitive if that point spread is less than 5 points.

> *Advantage*: This data is readily available to be compiled for analysis. It is a credible source and even if slightly skewed, it effectively represents observations that have been widely described
> *Disadvantage:* Conclusions on the precise number of remaining competitive districts vary slightly between different reporting sources. These differences could be related to timing, assumptions about unfinalized redistricting plans or differences in definition of competitive districts. In any case, it would therefore be ideal to independently verify these calculations as the underlying data is readily available.

Using data from the FiveThirtyEight project, and applying it district-by-district and state-by-state, we make the following observations:

- Competitive districts reduced from 46 (10.6%) to 40 (9.2%) of 435 total seats during the 2022 redistricting process
- Across the US, each state currently has between 0 and 4 competitive districts which account for 0% to 100% of their total districts
- 25 of the 50 states have zero competitive districts
- Among the states with at least one competitive district, their average number of competitive districts is 1.6, and their median number of competitive districts is 1.

The complete list is below (based upon new districting plans, ordered from greatest to least percentage of competitive districts:

STATES WITH COMPETITIVE DISTRICTS

| STATE | COMPETITIVE | TOTAL DISTRICTS | PERCENTAGE |
|---|---|---|---|
| NH | 2 | 2 | 100% |
| NM | 2 | 3 | 67% |
| IA | 2 | 4 | 50% |
| NV | 2 | 4 | 50% |
| CT | 2 | 5 | 40% |
| NE | 1 | 3 | 33% |
| KS | 1 | 4 | 25% |
| MI | 3 | 13 | 23% |
| PA | 3 | 17 | 18% |
| OR | 1 | 6 | 17% |
| NY | 4 | 26 | 15% |
| OH | 2 | 15 | 13% |
| CO | 1 | 8 | 13% |
| MD | 1 | 8 | 13% |
| MN | 1 | 8 | 13% |
| AZ | 1 | 9 | 11% |
| WA | 1 | 10 | 10% |
| VA | 1 | 11 | 9% |
| NJ | 1 | 12 | 8% |
| FL | 2 | 28 | 7% |
| GA | 1 | 14 | 7% |
| NC | 1 | 14 | 7% |
| IL | 1 | 17 | 6% |
| CA | 2 | 52 | 4% |
| TX | 1 | 38 | 3% |

STATES WITHOUT COMPETITIVE DISTRICTS

| STATE | COMPETITIVE | TOTAL DISTRICTS | PERCENTAGE |
|---|---|---|---|
| AK | 0 | 4 | 0% |
| AL | 0 | 7 | 0% |
| AS | 0 | 1 | 0% |
| DL | 0 | 1 | 0% |
| HI | 0 | 2 | 0% |
| ID | 0 | 2 | 0% |
| IN | 0 | 9 | 0% |
| KY | 0 | 6 | 0% |
| LA | 0 | 6 | 0% |
| ME | 0 | 2 | 0% |
| MA | 0 | 9 | 0% |
| MO | 0 | 8 | 0% |
| MS | 0 | 4 | 0% |
| MT | 0 | 2 | 0% |
| ND | 0 | 1 | 0% |
| OK | 0 | 5 | 0% |
| RI | 0 | 2 | 0% |
| SC | 0 | 7 | 0% |
| SD | 0 | 1 | 0% |
| TN | 0 | 9 | 0% |
| UT | 0 | 4 | 0% |
| VT | 0 | 1 | 0% |
| WI | 0 | 8 | 0% |
| WV | 0 | 2 | 0% |
| WY | 0 | 1 | 0% |

## 8. GENERATING REDISTRICTING PLAN ENSEMBLES

---

*To apply the competitiveness index, we generate **redistricting ensembles** using the Markov Chain logic in Gerrychain and Networkx, using code that we customized to apply to any and every state.*

*We ensure that each plan meets legal requirements and guidelines (such as contiguity and compactness). We measure compactness via cut-edge for each ensemble, and we use this metric to indicate convergence of the Markov chain. We then apply the competitiveness index to each ensemble for analysis.*

*This section provides a summary of why we use Gerrychain and Networkx, why we trust Recom-like Markov chains, and how we think about convergence. The next section will describe relevant data sources, and the sections after that will describe relevant data and our simplifying assumption about single member districts.*

*After that context, we will review sample results from the Competitiveness study.*

---

***Best, worst or most extreme*** districting plans along any dimension (including competitiveness) are mathematically complex to identify because they depend on competing metrics, and the space of all possible districting plans is too large to catalog. This is why researchers increasingly seek large ensembles of redistricting plans that statistically represent the space of all plans.

***GerryChain*** software was developed as an open-source toolkit to enable anyone to generate and study large ensembles of districting plans. It is a modular version of Markov chain sampling and Markov Chain Monte Carlo that randomly walks from one districing plan to the next by repeatedly repartitioning the graphical representation of the state G=(V,E).

***Networkx*** is a Python software package that simplifies the creation, manipulation and study of graphs and networks.

***Dual graph G=(V,E)*** aligns with the premise that any given districting plan can be represented by a partition of a graph that geographically represents voters. Typically in the graph G=(V,E) each vertex is a VTD voting tabulation district and each edge indicates that those VTDs are geographic neighbors. Therefore the graph captures the interconnectedness of the populations within the state. Districting plans partition the graph, assigning each VTD node to one district, respecting rules such as connectedness, compactness, population balance, and more. Non-compliant partitions, which reflect illegal districting plans, are disallowed.

***Recombination Algorithm:*** starts by representing a state as a dual graph G = (V,E) and a redistricting plan P that partitions V into a legal set of districts. Recom steps *randomly* from districting plan P to the next districting plan Q by *randomly* selecting a boundary edge (u,v) where u and v are neighboring VTDs or census blocks who are in different districts W1 and W2. Recom then merges districts W1 and W2 to make a single mega district, which Recom will subsequently divide into two new districts.

***Confirming Recom has converged:*** Since the space of all possible districting plans is too large to catalog, we rely on a large set of districting plans that statistically represents the space of all districting plans. Therefore we need to confirm that Recombination Markov chain has converged to a stationary distribution. Since we cannot easily prove it, we rely on heuristics to explain why we think it has converged. For example, we look for evidence that even if we start from different initial positions, our distribution for those relevant metrics is generally converging towards the same end-point distribution. We also confirm that the Markov chain appears to be mixing and sampling from many different possible plans. In addition, we continue to run the Markov chain for many steps beyond the supposed stationary distribution, and we show that the degree of change from step to step becomes minimal after it approaches the supposed stationary distribution.

## 9. RELEVANT DATA FOR ANALYSIS OF COMPETITIVENESS

---

*We run election simulations using historical voting datasets, such as the Virginia 2018 senate race to study trade-offs, risks and opportunities of prioritizing competitiveness, and look for evidence of gerrymandering in current political redistricting maps that squelches competitiveness.*

*In addition, when GerryChain is run to generate ensembles for each state that can be shown to be compact, contiguous and converging, GerryChain uses a dual graph created from the  aforementioned*
*The survey below describes pros and cons of the array of data used to analyze political redistricting.*

---

***Electoral Data:*** Each election, each year in each state produces results that should be able to be broken down to vote count within each precinct or voting tabulation district. For example, we could look at a 2012 senate race in Virginia to see the number of votes for the Democrat and Republican in each VTD, or we could look at the 2020 presidential race in Minnesota and similarly see the number of votes for Biden and Trump in each precinct. This data allows us to simulate the district-by-district results of the same election under a variety of redistricting plans. This is critical to our ensemble analysis

*Advantages:* this data is publicly available and generally comprehensive. It allows comparison across time and across a variety of races involving different candidates.
*Disadvantages:* because of the array of choices of electoral data, each of which has a unique set of candidates and environmental conditions, it is possible that two single sets of data applied to the same model could drive very different results. Analysts must proceed with caution when drawing generalized conclusions

***Decennial Census Data*** is collected during the decennial attempt to count and collect data about every resident of the United States, tied to exactly where they live. The Decennial Census asks questions about race, ethnicity, and citizenship. It also asks about heads-of-household, number of children, relationships, gender, age, commuting patterns, schools, housing, economic data and more. This data is then rolled-up from the individual to their census block - which is the smallest building block for many other geographies (some of which are nested, and some of which are not).

*Advantages:* it is a stunning volume of comprehensive, publicly available data

*Disadvantages:* (1) it is known to undercount and overcount: some people are not counted at all, and others (like college students, children with divorced parents, military) are counted twice; (2) the race and ethnicity categories change over time, and do not necessarily match the ways that people think of themselves; (3) the question of citizenship (which could be critical to drawing boundaries for voting), has become highly politicized and therefore less accurate; (4) for the purpose of redistricting, it does not perfectly align with voting data.

***Geographic data (particularly in a GIS)*** looks at many different layers of information about each position on the earth's surface: streets, rivers, buildings, vegetation, businesses, people and more. By dividing the map into fine-grained census blocks, we can determine where (on the map) each person, each voting-age person, and each voting-age citizen (with all of their personal attributes) resides, and what conditions they face in terms of surrounding streets, water, buildings, vegetation, etc.

*Advantages:* we can use this data to analyze concentration of races, ethnicities and languages on the map, and within precincts and congressional districts.

*Disadvantages:* (1) the translation from a globe to a computer screen causes a degree of unavoidable distortion; (2) change of both resolution and coordinate systems will impact most geographic measures of compactness, which are considered important for re-districting; (3) it can be difficult to join geographic data with other data sets that are not collected or maintained at the same level of detail (such as electoral data); (4) because districts and boundaries change over time, the *Vintage* or exact age of geographic data is important.

# 10.COMPETITIVENESS IN SINGLE-MEMBER DISTRICTS

---

*For initial simplicity, this study assumes a plurality-based social choice function and single-member districts. We recognize that competitiveness could be explored with different metrics, methods, outcomes and importance in multimember districts, with or without ranked choice. This may be a topic for future study.*

---

***Winner-takes-all Plurality Voting***: each voter selects one candidate, and the candidate who earns the most votes wins

***Ranked Choice Voting***: each voter considers a slate of candidates and (rather than voting for a single candidate) ranks multiple candidates as their first choice, second choice, third choice and so forth. A behind-the-scenes algorithm then tabulates votes to find the winner, often instituting an "instant run-off" if the leader achieves plurality rather than majority.

***Single member districts:*** one seat is elected in each district.

***Multi-member districts:*** more than one seat is elected per district. In these elections, different social choice voting algorithms can be applied: from winner-takes-all to ranked-choice. Elections in multi-member districts tend to be egregiously gerrymandered when using a social choice function like winner takes all, and they can be more fair and proportional when using a ranked-choice voting algorithm.

> *Advantages:* (1) multi-member districts with ranked choice can limit the effect of bad gerrymandering (making it such that even the most Republican or most Democratic gerrymander is nearly proportional); (2) it makes it possible for the most fair map to become proportional in ways that may not be possible for a Single Member District.
>
> *Disadvantages*: Multi-member districts with winner-takes-all can negatively impact proportionality.

## 11. COMPETITIVENESS STUDY RESULTS

---

*The model developed through this Competitiveness Study allows state by state analysis of competitiveness, as compared to other states and as compared to its own ensemble.*

*Considering Virginia as an example:*

*Virginia's Absolute Competitiveness Score is 9.1% on a scale of 0-100%*
- *This compares to a nationwide Competitiveness Score of 9.2%*

*Virginia's Competitive-Completeness index is 0.8 on a scale of 0-1*
- *This means Virginia's Competitiveness Score level is at or above 80% of the Competitiveness Scores of the plans in its redistricting ensemble.*
- *This indicates that Virginia is not an outlier compared to its ensemble in terms of too-few competitive districts.*

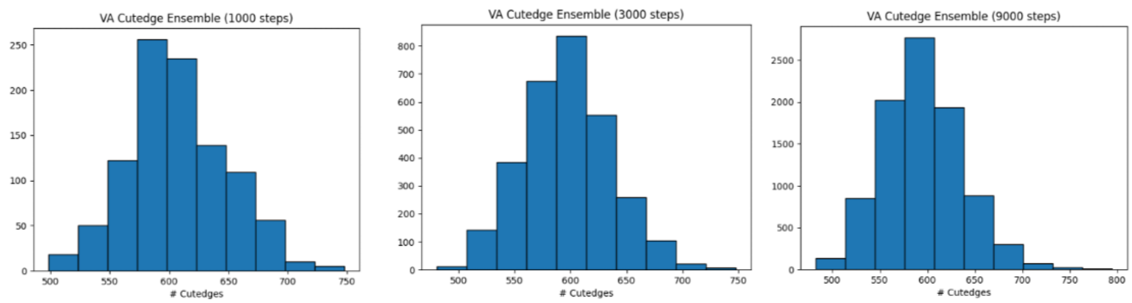
*Virginia's Competitive-Potential index is 0.67 on a scale of 0-1*
- *This means that according to its ensemble, it may reasonably achieve a level of 3 competitive districts by adding 2 more competitive districts.*
- *This indicates that Virginia still has reasonable room for improvement.*

---

***Test runs:*** we performed multiple test runs to refine our code to produce reasonable results (error-free, effective graphical output, reasonable run-times). We also generalized the code to enable output for all fifty states. An illustrative subset appears below for Virginia. All runs will be cataloged and available at https://github.com/JonnyWise25/Math195Final

---

***Stationary point / Compactness Distribution:*** we ran our code to generate completely separate redistricting ensembles for each state under consideration. In the case of Virginia, we ran it for 1,000 steps, 3,000 steps and then for 9,000 steps. This allowed us to consider how close we may be to a stationary point, and a representative sample of the universe of all possible redistricting plans for Virgina. After each run, we evaluated the ensemble for the distribution of cut-edges (as a proxy for compactness). For 1,000 steps, 3,000 steps and 9,000 steps the distribution of the compactness of the resulting districting plans become similar, with a relatively normal distribution around just under 600 cut-edges.




Cut Edge distributions for Virginia Ensembles after 1000, 3000 and 9000 steps

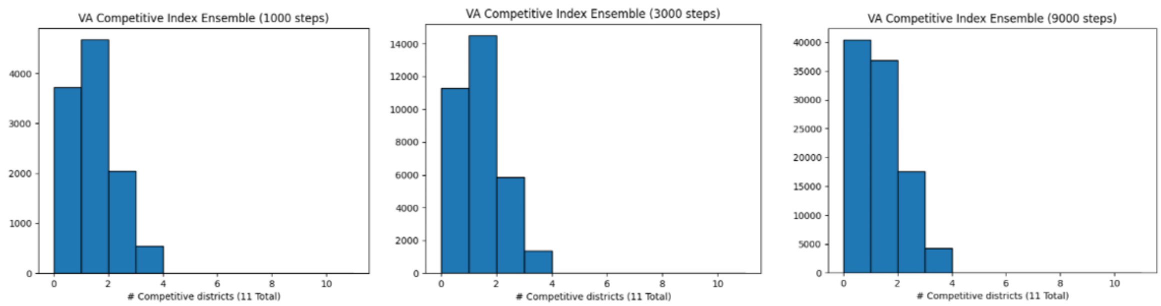
---

***Competitiveness Distribution:*** Next, we look at the distribution of the competitiveness index that results from applying 2018 election results to each plan in the ensemble. For the purpose of this study, we define a district as competitive if the winner achieved victory by a share of five points or less:

$$District\_Gap = | dem\ votes - rep\ votes | / (dem\ votes + rep\ votes)$$

- For each district, if District_Gap < 0.05 (or five percent), that district is Competitive.
- For each districting plan, identify the number (and percentage) of competitive districts
- For each ensemble, create a histogram that shows incidence and likelihood of of each number (or percentage) competitive districts

In the case of Virginia, after 1,000 steps, 3,000 steps and 9,000 steps the distribution of the competitiveness index is shown below:




Competitiveness distributions for Virginia Ensembles after 1000, 3000 and 9000 steps

---

***Analysis of  Ensemble:*** Approximately 80% of the plans in Virginia's redistricting ensemble have 0 or 1 competitive districts (comprising 0% - 9% of its eleven districts). And approximately 20% have 2 or 3 competitive districts (comprising 18% - 27% of its eleven districts). Today, Virginia has one competitive district,  9.1%, nearly at the national average of 9.2%. And compared to its ensemble, this falls at or above 80% of the likely set of outcomes and appears not to have an outlier plan. However, if competitiveness was a priority, Virginia's political map makers have a robust set of alternatives to consider, as over 20% of the districting plans in the ensemble offer more competitive districts that could reasonably bring Virginia's score up to 18-27% percent from its current level of 9%. More specifically, the ensemble indicates that Virginia could reasonably add 2 competitive districts to get to a total of 3.

As such,

Virginia's Absolute Competitiveness Score is 9.1% on a scale of 0-100%

- This compares to a nationwide Competitiveness Score of 9.2%

Virginia's Competitive-Completeness index is 0.8 on a scale of 0-1

- This means Virginia's Competitiveness Score level is at or above 80% of the Competitiveness Scores of the plans in its redistricting ensemble.
- This indicates that Virginia is not an outlier compared to its ensemble in terms of too-few competitive districts.

Virginia's Competitive-Potential index is 0.67 on a scale of 0-1

- This means that according to its ensemble, it may reasonably achieve a level of 3 competitive districts by adding 2 competitive districts.
- This indicates that Virginia still has reasonable room for improvement.

## 12. RECOMMENDATIONS AND WAYS TO IMPROVE THE STUDY

---

*Key recommendation is to apply the model to all states to better understand their relative competitiveness and opportunities for change.*

*Other recommendations include:*

- *Independently verify the data used from the FiveThirtyEight project, which was used to compute the Absolute Competitiveness Score of the current Districting plans for each state*

- *recalculate each state's Absolute Competitiveness Scores, Competitive-Completeness index and Competitive-Potential index using multiple alternative election results applied to the ensembles developed for each state. Identify and explain differences*

- *Test alternative starting points for ensemble generation and alternative metrics to track convergence*

---

## 13. COPY OF CODE AND SPECIFICATION OF DATA USED

---

*With the artifacts and instructions below, readers should have everything that they need to replicate the results of the Competitiveness Study, and apply to additional states and election data sets.*
*In addition to the image below, a copy of the code is available on Github* *https://github.com/JonnyWise25/Math195Final*

---

The most important file in the repo is the ensemble.py file, responsible for creating and running the ensemble

Below are the necessary imports to run the ensemble…
It is necessary to install networkx, matplotlib, numpy, gerrychain, and csv before running the import code

```
# Import necessary packages
import networkx as nx
import matplotlib.pyplot as plt
from numpy import random
import gerrychain
from gerrychain import Graph, Partition, proposals, updaters, constraints, accept, MarkovChain
from gerrychain.updaters import cut_edges, Tally
from gerrychain.tree import recursive_tree_part
from gerrychain.proposals import recom
from gerrychain.accept import always_accept
from functools import partial
import csv
```

The next chunk of code defines the 'runEnsemble' method, and loads in the dual graph

```
def runEnsemble(stateCode, filePath, num_dist, total_steps = 9000, pop_tolerance = .02, district_gap_threshold= .05):
    #stateCode = States Code ex: VA MN WI
    #total_steps =  Number of steps in random walk ensemble  Default:(1000, 3000, 9000)
    #num_dist =  Number of Congressional Districts in State
    #pop_tolerance = population variation tolerance allowed in redistricting plan (.02/2% default)
    #district_gap_threshold = threshold for a district to be competitive (.05/5% default)
    #filePath = path to dual graph data ex: "VA/VA_precincts.json"

    # import dual graph
    data = Graph.from_json(filePath)
```

After that, the total and ideal population are calculated. Then the initial partial and random walk proposal are instantiated and the population constraint is created

```python
# Get total population
tot_pop = 0
for v in data.nodes():
    tot_pop = tot_pop + data.nodes()[v]['TOTPOP20']


## Make an initial districting plan using recursive_tree_part (from last time)
ideal_pop = tot_pop/num_dist
initial_plan = recursive_tree_part(data, range(num_dist), ideal_pop, 'TOTPOP20', pop_tolerance, 10)

# initial partition
initial_partition = Partition(
    data, # dual graph
    assignment = initial_plan, #initial districting plan
    updaters={
    "our cut edges": cut_edges,
    "district population": Tally("TOTPOP20", alias = "district population"), # across districts, add total population
    "district democrat votes": Tally("G18DSEN", alias = "district democrat votes"),
    "district republican votes": Tally("G18RSEN", alias = "district republican votes"),
    "district independent votes": Tally("G18OSEN", alias = "district independent votes"),
})

# random walk proposal
rw_proposal = partial(recom, ## how you choose a next districting plan
                      pop_col = "TOTPOP20", ## What data describes population?
                      pop_target = ideal_pop, ## What the target/ideal population is for each district
                                              ## (we calculated ideal pop above)
                      epsilon = pop_tolerance,  ## how far from ideal population you can deviate
                                              ## (we set pop_tolerance above)
                      node_repeats = 1 ## How many times you attempt before resrating
                      )

## Contraint on population: stay within pop_tolerance of ideal
population_constraint = constraints.within_percent_of_ideal_population(
    initial_partition,
    pop_tolerance,
    pop_key="district population")
```

Then we create our MarkovChain (this step doesn't actually run the chain however)

```python
## Creating the chain
# Set sup the chain, but doesn't run it!
our_random_walk = MarkovChain(
    proposal = rw_proposal,
    constraints = [population_constraint],
    accept = always_accept, # Accept every proposed plan that meets the population constraints
    initial_state = initial_partition,
    total_steps = total_steps)
```

Then we loop through the MarkovChain, actually running the random walk. As we do this, we calculate the district gap and determine if the district is competitive. We then add the number of competitive districts in that plan to the ensemble list.

```python
    # What ensembles we want to build
    ensemble = {}
    ensemble['cutedge'] = []
    ensemble['competitiveness_index'] = []
    # threshold for the competitiveness index set to 5%
    ensemble['distGap_threshold'] = district_gap_threshold
    counter = 1


    # This actually runs the random walk!
    for part in our_random_walk:
        if counter % 100 == 0:
            print("Running Dist #:")
            print(counter)
        ensemble['cutedge'].append(len(part["our cut edges"]))

        num_competitive = 0
        for i in range(num_dist):
            # calculate district gap for each district in ensemble
            # total votes = democrat votes + republican votes
            dist_gap = abs(part["district democrat votes"][i] - part["district republican votes"][i]) / (part["district democrat votes"][i] + part["district republican votes"][i])

            # if the district gap is less than the threshold (.05 default) then the district is competitive
            if dist_gap < ensemble['distGap_threshold']:
                num_competitive = num_competitive + 1

            # add the # of competitive districts in the plan to the ensemble
            ensemble['competitiveness_index'].append(num_competitive)
```

Finally, we create and save histograms for the cutedge score and competitiveness index. Following that, we also save the results from the ensemble as a csv file.

```python
    plt.figure()
    plt.hist(ensemble['cutedge'], edgecolor = "black")
    plt.title("{} Cutedge Ensemble ({} steps)".format(stateCode,total_steps))
    plt.xlabel("# Cutedges")
    plt.savefig("States/"+stateCode+"/Results/"+str(total_steps)+"Cutedge.png")
    #plt.show()


    print(ensemble['competitiveness_index'])
    #bins = range(0,num_dist+1)
    bins = [0,1,2,3,4,5,6,7,8,9,10,11]
    plt.figure()
    plt.hist(ensemble['competitiveness_index'], bins = bins, edgecolor = "black")
    plt.title("{} Competitive Index Ensemble ({} steps)".format(stateCode,total_steps))
    plt.xlabel("# Competitive districts ({} Total)".format(num_dist))
    plt.savefig("States/"+stateCode+"/Results/"+str(total_steps)+"CompetitivenessIndex.png")
    #plt.show()

    newFilePath = ""+stateCode+"/Results/"+str(total_steps)+"Ensemble.csv"
    with open(newFilePath, 'w') as myFile:
        writer = csv.writer(myFile)
        writer.writerow(['cutedge', 'competitiveness_index'])
        for i in range(0, len(ensemble['cutedge'])):
            writer.writerow([ensemble['cutedge'][i],ensemble['competitiveness_index'][i]])
```

The ensemble program can be imported and called in the main.py file as shown below

```python
import ensemble
# Run ensemble (1k,3k,9k steps) for VA
ensemble.runEnsemble(stateCode = "VA", filePath = "States/VA/Data/VA_precincts.json", num_dist=11, total_steps = 1000, pop_tolerance = .02, district_gap_threshold= .05)
ensemble.runEnsemble(stateCode = "VA", filePath = "States/VA/Data/VA_precincts.json", num_dist=11, total_steps = 3000, pop_tolerance = .02, district_gap_threshold= .05)
ensemble.runEnsemble(stateCode = "VA", filePath = "States/VA/Data/VA_precincts.json", num_dist=11, total_steps = 9000, pop_tolerance = .02, district_gap_threshold= .05)
```

14. COPY OF PRESENTATION MATERIALS

*The slides below support a ten-minute presentation of the Competitiveness Project and its results*

## Why Competitiveness

Numbers of competitive districts are decreasing

Possible contributor to polarization & extremism

Competitiveness ***dropped*** from 46 to 40 (of 435  districts)

25 states have ***zero*** competitive districts

25 states range from 1 to 4 competitive districts

- Their average is 1.6

—

# Political Districting Impacts Competitiveness

To what degree is this true?

How can **Competitiveness Index** guide us to opportunities?

## 3 Competitiveness Metrics

- **Absolute Competitiveness**
- **Competitive Completeness** — Versus Ensemble
- **Competitive Potential** — Versus Ensemble

## Absolute Competitiveness Score.

**scale of 0 - 1**

**Percentage of competitive districts in any given plan**

—

## Competitive-Completeness Index.

## scale of 0 - 1

**Compare to ensemble:** Is this plan strongly competitive (1) or weakly competitive (0)

**Fraction of ensemble plans with same or fewer competitive districts**

—

## Competitive-Potential Index.

## scale of 0 - 1

**Compare to ensemble:** Is the improvement potential strong (1) or weak (0)?

**Fraction of potential competitive districts yet to be added**

# 3 Competitiveness Metrics

**Absolute Competitiveness**

% of districts that are competitive

**Competitive Completeness**

**Versus Ensemble**

0 =weakly competitve

1 =strongly competitive

**Competitive Potential**

**Versus Ensemble**

0 = no more potential

1 = much more potential

LET'S LOOK AT VIRGINIA:

1-OF-11 Competitive Districts

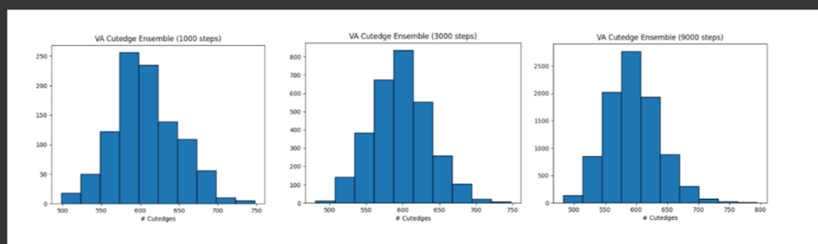
## ABSOLUTE COMPETITIVENESS

SCORE OF .091

—

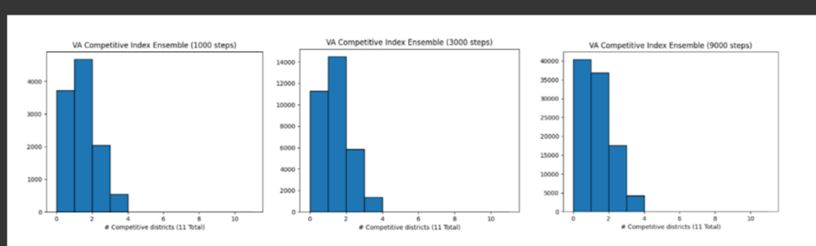
# Use GerryChain to create Virginia's Ensemble 9000 steps

Cut-edge compactness Converges




—

# Use ElectionData to compare Absolute Competitiveness in Virginia Ensemble



# 3 Competitiveness Metrics

**Absolute Competitiveness**

% of districts that are competitive

**Competitive Completeness**

**Versus Ensemble**

0 =weakly competitve

1 =strongly competitive

**Competitive Potential**

**Versus Ensemble**

0 = no more potential

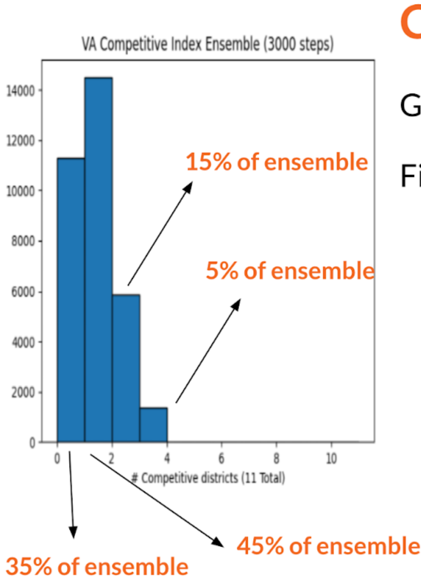
1 = much more potential

For any DISTRICTING PLAN

fraction of Competitive Districts

# ABSOLUTE COMPETITIVENESS

## SCORE 0 to 1




## Competitive-Completeness Index

Given Absolute Competitiveness Score for a current plan

Find **FRACTION** of scores in ensemble at/below your score

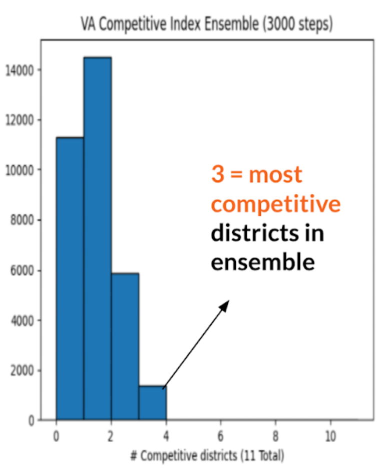
Your Comp Dists =**0**… Competitive-Completeness =**0.35**

Your Comp Dists =**1**… Competitive-Completeness =**0.80**

Your Comp Dists =**2** … Competitive-Completeness =**0.95**

Your Comp Dists =**3+**… Competitive-Completeness =**1.0**

**0 means outlier: weakly competitive districts**




## Competitive-Potential Index

Given Absolute Competitiveness Score for a current plan

Compare to the **most competitive** districts in ensemble

Your Comp Dists =**0**… Competitive-Potential = 3/3 = 1.0

Your Comp Dists =**1**… Competitive-Potential = 2/3 = .67

Your Comp Dists =**2** … Competitive-Potential = 1/3 = .33

Your Comp Dists =**3**… Competitive-Potential = 0/3 = 0.0

**1 means most potential for improvement**

# 3 Competitiveness Metrics

**Absolute Competitiveness**

% of districts that are competitive

**Competitive Completeness**

Versus Ensemble

0 =weakly competitve

1 =strongly competitive

**Competitive Potential**

Versus Ensemble

0 = no more potential

1 = much more potential

Back to **VIRGINIA:**

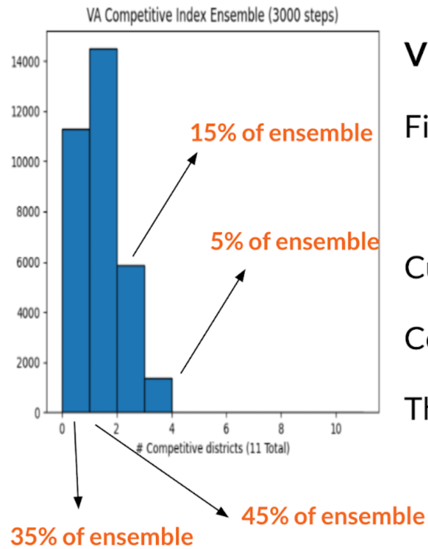
**1-OF-11** Competitive Districts

## ABSOLUTE COMPETITIVENESS

SCORE OF 1/11 =

.091

## Competitive-Completeness Index




**Virginia = 0.091 … 1 out of 11 competitive districts**

Find **FRACTION** of scores in ensemble at/below your score

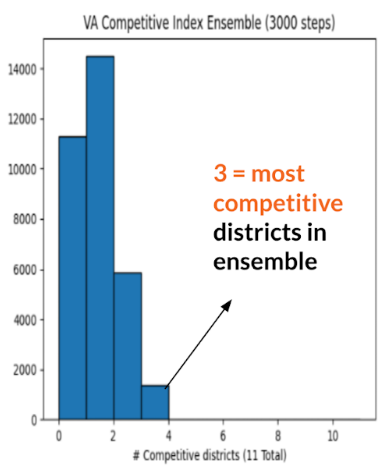
Current Plan's Competive Dists =**1**…

Competitive-Completeness =**0.80 (=.35+.45)**

The Current Plan is relatively complete and competitve

---

## Competitive-Potential Index




**Virginia = 0.091 … 1 out of 11 competitive districts**

Compare to the **most competitive** districts in ensemble

Current Plan's Competive Dists =**1**…

Competitive-Potential = 2/3 = .67

There is still relatively strong room for improvement

## Conclusions for VA's Current Redistricting Plan

At 0.091, VA slightly less competitive vs nationwide avg
**At 0.8 Competitive Completeness, VA is not an outlier vs its ensemble**
At 0.67 competitive potential, VA has headroom and options to become more competitive

**Absolute Competitiveness = 0.091**

**Competitive Completeness = 0.8**

**Competitive Potential = 0.67**

—

# Model highlights opportunities for Political Districting to Impact Competitiveness

- Code available on GitHub
- Data available for multiple states, completed
- Next: complete for all states
- Next: vary election data inputs

## 15. WORKS CITED

Most sources used for background reference and knowledge came directly from Math 195

Additional sources not used in class

Best, R., Bycoffe, A., & Rakich, N. (2021, August 9). *What Redistricting looks like in every state*. FiveThirtyEight. https://projects.fivethirtyeight.com/redistricting-2022-maps/

---