Final Report for Syslab

A Novel Approach to Algorithmic Redistricting: Combating Gerrymandering with Artificial Intelligence Tools

Deven Hagen

5/27/25

Yilmaz - Period 1

Table of Contents

Abstract

Gerrymandering is the process by which district borders are drawn to benefit a particular group, most often a political party. Gerrymandering has been around for over 200 years in the United States, and both parties have used it to their advantage. The issue is most prominent regarding congressional district maps, which determine the boundaries for the 435 districts that each elect one member to the House of Representatives. Typically, in each state, the state legislature is tasked with drawing the lines every ten years, so the party in control can create a map that will enable it to keep power for many electoral cycles. In recent years, some states, including Arizona and California, have established Independent Redistricting Commissions (IRCs) that aim to remove bias in their congressional maps by tasking a group of citizens, balanced by party identification, with drawing them. Thus, the development of unbiased algorithms can aid IRCs in making fairer maps. I used Markov Chain Monte Carlo methods (MCMC) through the GerryChain Python library to create an ensemble of possible congressional district maps for different states. I then developed a novel, broad-based heuristic to grade maps and select for those with a high score. My approach improved upon many current maps in compactness (measured with Reock/Polsby-Popper score), competitiveness, and partisan fairness (measured with efficiency gap). Hopefully, IRCs can use my algorithm in concert with others when creating maps, helping to remove biases from an important process and ultimately strengthening our democracy.

I. Introduction

The United States House of Representatives has 435 members. Each state receives a number of representatives to the House proportional to its population in the decennial Census. For example, California, with its population of almost 40 million, has 52 representatives, while Wyoming and its 600,000 citizens have just one. The representatives each have their own district within their state; the population of each district is roughly the same, and each district must be contiguous, as stipulated by federal law. The Census is highly detailed–it provides a population breakdown down to the block level. Thus, after the Census is tabulated, each state legislature creates its own district map, using Census data to verify that the districts meet the necessary criteria. Unfortunately, given the two-party system in our country, this has led to the proliferation of "gerrymandering," where political parties draw the district lines to give themselves a majority of seats, even if they do not receive a majority of votes.

A prominent example of this took place after the 2010 elections with Project REDMAP. Republicans swept to power in many state legislatures that year because of backlash against President Obama and Democrats, owing to the slow economic recovery after the Great Recession. This gave them the power to draw the congressional district maps in key swing states such as Wisconsin and North Carolina, which they took advantage of. After the Republicans redrew the lines, for example, they achieved a 13-5 majority in the Pennsylvania delegation to the House of Representatives after the subsequent 2012 election, even though Democratic candidates received 50.3% of the votes cast. Essentially, Republicans packed Democratic voters into a few districts, where Democratic candidates won by large majorities, and spread out their own voters over a number of districts, enabling them to win more districts by slimmer margins. I use this example of Republican gerrymandering not to single out the party, as this is certainly a bipartisan issue. Democrats, for example, have gerrymandered states such as Maryland and Illinois for decades. In addition, another consideration is that, because of the high correlation between race and political party, many cases of gerrymandering run afoul of the 1965 Voting Rights Act, which stipulates that minority voters must have “opportunity districts”—districts where they make up more than half of the voting-age population—to ensure that they have representation in Congress. Thus, combating gerrymandering can also prevent the instances of disenfranchisement that often come with unfairly drawn maps.

In recent years, a few states, such as Virginia, have created independent redistricting commissions (IRCs) to draw the lines, in an attempt to prevent political parties from gaming the process. These commissions have drawn markedly more fair maps than partisan state legislatures, but unfortunately, both parties seem reluctant to adopt this approach for fear of losing power to the people. Thus, an algorithmic approach to gerrymandering, which would be nonpartisan and without racial bias, could serve both as a tool for IRCs and as a general starting point for each state’s congressional map. Using artificial intelligence to replace human involvement in drawing the maps completely is not a viable solution in the short term—the redistricting process is quite complex, and requires human oversight to ensure all considerations are fulfilled—but it can surely be of great use to those designing the maps, and it can be a means to ascertain whether existing maps are biased.

II. Background

Much research has already been done on this topic, likely owing to its importance as an issue. I shall provide a non-exhaustive summary of some of the most promising projects in the past decade. Brian Olson gained prominence in the early 2010s for developing software to create districts that were maximally compact. Chen and Rodden took this a step further, developing an algorithm that would ensure minority-majority districts while satisfying the population and contiguity constraints (2015). Liu et al. approached the redistricting problem from an optimization standpoint, rating maps based on a carefully crafted heuristic and finding the best ones using a genetic algorithm (2016). In 2020, Haas et al. used a method called Seed-Fill-Shift-Repair (SFSR) to create a huge number of maps, allowing for selection of maps that meet specified criteria and testing to see how biased a particular map is compared to a general sample. The most extensive software, however, was created by Kevin Baas at AutoRedistrict in 2018. His evolutionary algorithm approach can optimize maps based on pre-selected metrics, and it is also the first and possibly only that can also create multi-member district maps. Multi-member district maps involve the creation of districts that elect multiple representatives, and are mathematically more likely to allow for true proportional outcomes in legislative allocation (Volić, 2024). However, Baas’s heuristic uses relatively few metrics; my project’s novel heuristic is more detailed and complex, helping it to create better maps. I have experimented with a wide range of algorithms, and the final product does an excellent job of creating maps that score highly across the board in key metrics.

III. Applications:

As I have stated above, my algorithm can be used by IRCs (in conjunction with other algorithms) to create fairer maps. As this process occurs every 10 years, the next time this could be put into use would be the 2030 redistricting cycle.

IV. Methods:

1. Obtain relevant Census and electoral data, and use necessary Python libraries to form my own dataset
2. Preprocess the data, removing columns or instances with missing values
3. Experiment with a variety of optimization strategies (short bursts, simulated annealing, etc.) for map-making
4. Create and expand upon a novel heuristic to use in optimization
5. Compare generated maps to current ones; run them through testing to ensure that they lack both racial and partisan bias

Much of the data I need to complete this project can be located for free online. The Census Bureau (www.census.gov) has all of its data online in parsable formats. I used the Python library Pandas to read in Census numerical data and the related library GeoPandas to handle spatial data in shapefile format. As for the election data, each state's Department of Elections posts detailed election results at the precinct level online; both the Harvard Dataverse and sites such as Redistricting Data Hub (here is Virginia's database: https://redistrictingdatahub.org/state/virginia/) have this readily available online. I merged these datasets into a singular GeoPandas GeoDataFrame with all the relevant information, and I used the Python library GerryChain to transform this data into a single graph. From that point, I used GerryChain's *SingleMetricOptimizer* sublibrary to optimize for my heuristic (which I will discuss more in the next paragraph), and I used the Python csv library to output my data as a .csv file with precinct to district mappings.

Some common considerations in the redistricting process include:

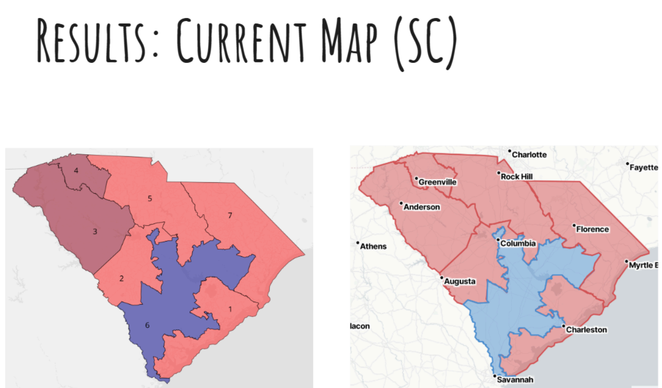
Population balance: The districts should have roughly equal populations
Contiguity: Each district should be one piece, with no holes or tears
Compactness: The districts should have a large area to perimeter ratio
Partisan bias: The map should not benefit one party over the other
Racial bias: The map should not disenfranchise any racial group

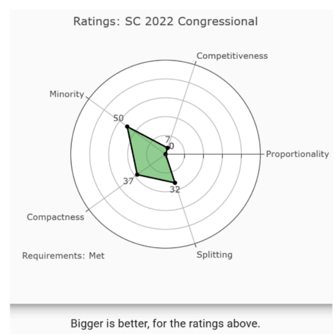
I set a population balance of 0.75% and a contiguous set of districts as my map constraints. To measure compactness, I used the average Polsby-Popper score for a district, equal to 4*pi*area over perimeter squared; a circle has a value of 1, the maximal value. To measure partisan bias, I used a compound statistic formed from the average of three metrics: partisan bias, efficiency gap, and mean-median difference. Partisan bias measures the difference in the statewide vote that a party gets and the percentage of seats it receives. Efficiency gap measures the ratio of wasted votes for each party; a wasted vote is any vote in a district for the winning party past 50%, or any vote in a district for the losing party. Mean-median difference subtracts the result in the median district by partisanship from the mean vote statewide. Because African-Americans tend to vote strongly for Democrats, partisan and racial bias go hand-in-hand, so I expected my heuristic to combat racial bias as well. I actually used 1 minus the average district Polsby-Popper score, as well as the absolute value of the average of the three partisan bias metrics, so that I could minimize the overall heuristic as opposed to maximizing it. I used a scaling factor on the compound partisan bias metric so that both terms would have influence on the algorithm. Thus, my heuristic was equal to:

20 * the average of the partisanship metrics + 1 - the average district Polsby-Popper score

V. **Results**:

I used the map grading features on both Dave's Redistricting (https://davesredistricting.org/) and PlanScore (https://planscore.org/) to assess the merits of the map that my algorithm created. The different facets I chose to grade include, but are not limited to, population balance, contiguity, compactness, county splitting, partisan bias, racial bias, and competitiveness. Below is a summary of the results for the current map in South Carolina and a map of South Carolina that my algorithm created:



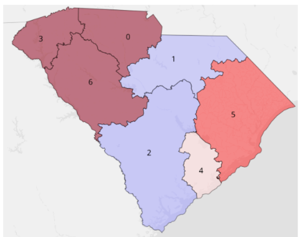


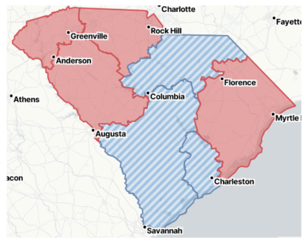
Metric	
• Reock	0.3611
• Polsby-Popper	0.2147

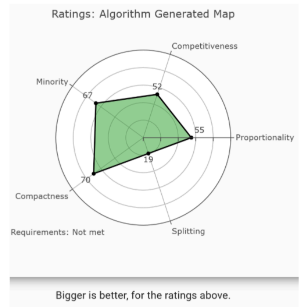
Metric	Value	Favors Democrats in this % of Scenarios*	More Skewed than this % of Historical Plans‡	More Pro-Democratic than this % of Historical Plans‡
Efficiency Gap	19.3% Pro-Republican	<1%	99%	1%
Declination	0.51 Pro-Republican	<1%	91%	5%

Figure 1: A summary of results for South Carolina's current map

Results: My Map (SC)







Metric	
• Reock	0.4816
• Polsby-Popper	0.2891

Metric	Value	Favors Democrats in this % of Scenarios*	More Skewed than this % of Historical Plans‡	More Pro-Democratic than this % of Historical Plans‡
Efficiency Gap	5.0% Pro-Republican	34%	51%	22%
Declination	0.02 Pro-Democratic	68%	6%	54%

Figure 2: A summary of results for a map of South Carolina created by my algorithm

As you can see from the above, my algorithm outperformed the current map in almost every facet, resulting in a much larger area of the green polygon above. The lone exception was county splitting; however, this makes sense, because my algorithm looks at states on a precinct level and thus does not see counties. The most important thing to note is that my map essentially had no partisan skew, allowing for a more fair electoral playing field.

VI. **Limitations**:

First of all, I used precincts as the building blocks for districts in my algorithm. As a result, I cannot make districts of the ideal size, because my data is not granular enough to make those micro-adjustments. In addition, to decrease runtime, I did not examine every combination of district sizes, instead choosing to focus on a few, more important cases. For instance, it is quite unlikely that a state with 12 members would create one district of 7 members and one of 5; I only focused on creating maps with districts of equal population. Beyond that, it may simply be impossible to find a "perfect heuristic." After all, this process is heavily nuanced, and any optimization will necessarily neglect certain factors.

VII. **Conclusion**:

I was able to successfully create a novel approach for automated redistricting that combats gerrymandering. My algorithm can be used in concert with others to increase fairness in the redistricting process, hopefully in conjunction with IRCs. Eventually, it is possible that algorithms could take over the process entirely, although we are not at that stage yet.

VIII. Future Work:

I'd like to continue to make my heuristic more complex, taking into account other considerations such as county splitting and competitiveness. I want to explore the use of Simulated Annealing and Tilted Runs to a much greater extent; I focused primarily on Short Bursts, but I believe that these methods could yield excellent maps as well. I'd also like to grade current congressional district maps with my heuristic, similar to what Sam Wang at the Princeton Gerrymandering Project did a couple of years ago. This approach doesn't just have to work for federal redistricting: I want to examine the merits of this approach with redistricting for state legislatures and even research whether it can be scaled to other countries, potentially with multiparty systems. As I said earlier, the next redistricting cycle isn't until 2030, but I hope to contact existing independent redistricting commissions and provide them with my algorithm.

IX. Materials:

A link to my code and my datasets can be found here: https://drive.google.com/drive/folders/1Y062shl1EHVQcmqkgdO7LVpIyt-Z0RUz?usp=drive_link.

References:

Altman, M., & McDonald, M. P. (2013). Paradoxes of political reform: Congressional redistricting in florida. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2421908

Chen, J., & Rodden, J. (2015). Cutting through the thicket: Redistricting simulations and the detection of partisan gerrymanders. *Election Law Journal*, *14*(4).

Cirincione, C., Darling, T. A., & O'Rourke, T. G. (2000). Assessing south carolina's 1990s congressional districting. *Political Geography*, *19*(2), 189-211. https://doi.org/10.1016/s0962-6298(99)00047-5

Duignan, B. (2024, September 21). Gerrymandering. Encyclopedia Britannica. https://www.britannica.com/topic/gerrymandering

Fifield, B., Higgins, Michael, Imai, K., & Tarr, A. (2020). Automated redistricting simulation using markov chain monte carlo. *Journal of Computational and Graphical Statistics*, *29*(4), 715-728. https://doi.org/10.1080/10618600.2020.1739532

Haas, C., Hachadoorian, L., Kimbrough, S. O., Miller, P., & Murphy, F. (2020). Seed-Fill-Shift-Repair: A redistricting heuristic for civic deliberation. *PloS one*, *15*(9), e0237935. https://doi.org/10.1371/journal.pone.0237935

Hagen, D. (2025, March 24). *Sys-Lab-Project*. GitHub. https://github.com/DASH758/Sys-Lab-Project/

Kevin Baas. (2018, May 11). *About*. AutoRedistrict. http://autoredistrict.org/

Liu, Y. Y., Cho, W. K. T., & Wang, S. (2016). PEAR: A massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation*, *30*, 78-92. https://doi.org/10.1016/j.swevo.2016.04.004

Metric Geometry and Gerrymandering Group. (2022). *GerryChain documentation*. GerryChain. https://gerrychain.readthedocs.io/en/latest/

Olson, B. (2020, October 17). *2010 redistricting results*. BDistricting. https://bdistricting.com/2010/

Virginia data. (2024). Redistricting Data Hub. https://redistrictingdatahub.org/state/virginia

Volić, I. (2024). *Making democracy count: How mathematics improves voting, electoral maps, and representation*. Princeton University Press.

APPENDIX:

I. **CODE:**

   Code and Data