A State, Divided - Visualizing the Effect of Computationally Generated Districting Plans on Local Communities

Rishabh Chandra
Massachusetts Institute of Technology
rchandra@mit.edu

Abstract

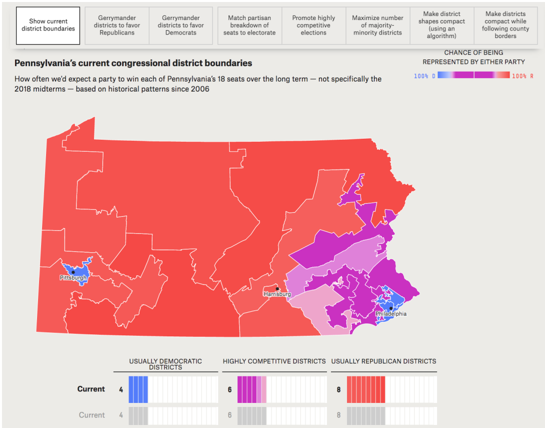
The word ”gerrymandering” means a lot to all of us in today’s raucous political climate, but we often struggle to truly understand the effect of redistricting a state on the democratic power of people in those districts. A state can be split in quintillions of ways, many of which can be considered ”fair” cuts. Researchers seek to distinguish between ”fair” and ”unfair” by sampling millions of plans and calculating summary statistics that can adjudge the existing plan to be biased. But what do those millions of plans look like? How do they affect local communities? We sought to answer that question with this visualization, and developed a library of more than 300 computationally drawn plans in Iowa, Georgia, and Pennsylvania that helps provide perspective on the scope of redistricting strategies that exist in those states.

1 Introduction

In 2018, following a legal challenge to Pennsylvania’s 2011-drawn congressional districts, the state Supreme Court struck down the old plan in favor of a different one that supposedly was less fair. In the aftermath of that discussion, many people outside of the mathematical world wondered aloud what the specific rationale was for the decision. In an aim to explain the context of gerrymandering to the public, researchers at FiveThirtyEight.com developed The Atlas of Gerrymandering, which was a publication designed to interactively demonstrate the differences between various different choices of districting plans on a state. These plans could be toggled through with tabs, and tooltips explained the likelihood of these districts to vote for a political party.

While the work done there is extremely valuable, however, it does not truly explain the mathematics behind the decision to declare the 2011 map biased. For that, we need to sample from far more than just eight plans. Broadly speaking, it takes millions of plans to generate a distribution of partisan lean, and it was determined that the existing plan was an outlier in that distribution. But this methodology is significantly harder to visualize, for a few key reasons:

Figure 1: FiveThirtyEight’s visualization for the districts of Pennsylvania, allowing for eight different state level plans.



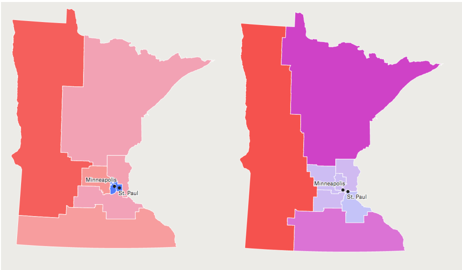
1. It is impossible to show millions of images of districting plans and have people understand the statistical significance of one outlier.

2. Not all gerrymandered or unfair plans look aesthetically troubling. Some, like the gerrymandered plans in Figure 2, look compact.

3. Each individual computational plan differs from the previous one only very slightly, for reasons that will be discussed. Therefore, from the naked eye many of these plans will appear redundant.

Therefore, our target is to find a middle ground between the aesthetic appeal of the visualizations

Figure 2: On the left, a Republican Gerrymander for Minnesota according to FiveThirtyEight; on the right, a Democratic Gerrymander. Note how both plans look compact, yet Democrats win only two of the districts on the left and six on the right, in a state that is about 50% for each party.



produced by FiveThirtyEight and the computational models produced by researchers. We seek to visualize computationally generated districts that are representative of the entire space of plans that could exist for a particular state, and to allow the user to toggle through these plans to get a sense of how a particular state can be divided.

2 Theory & Related Work

Researchers at the Metric Geometry and Gerrymandering Group (MGGG) have worked with teams from all over the United States to understand the space of all plans that can be generated for a particular state. We seek to express the problem of redistricting as the mathematical problem of graph partitioning, where we turn a state into a graph of n nodes that we wish to split into k partitions, with some constraints c on each partition (it must be contiguous, roughly equal population, compactness, etc).

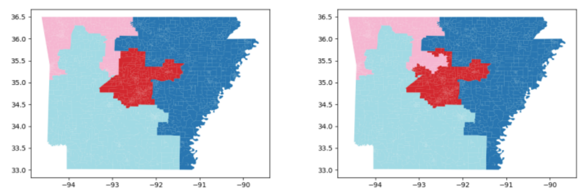
The process of sampling these partitions occurs through adaptive MCMC (Markov Chain Monte Carlo), which is used for a host of problems involving sampling from distributions without a closed-form expression. Mathematically speaking, according to Andrieu et. al, 2008, given a distribution π of districting plans, we can sample that distribution described by $\int_X f(x)\pi(x)$ for some function we wish to sample by calculating $\frac{1}{N}\Sigma_N f(X_i)$.

Intuitively, we input into the model a starting partition Q, and then propose to Q a proposed alteration Q' that differs from Q in some simple way. The algorithm then decides with some probability p whether it will accept the proposed change, and let $Q = Q'$, so as long as Q' still has the same n nodes, k partitions, and abides by the constraints c. We repeat this process many many times, each time obtaining a new partition Q' that is one of the quintillions of valid cuts of a particular state that are available. In so doing, the Markov Chain acts as a random walk through the space of all districting plans. Over many iterations, it is shown that the chain in fact begins to sample the entire space representatively.

A large part of the generative process is the way in which we generate our proposal changes Q'. The state of the art in most of the country is an algorithm called SINGLE-EDGE-FLIP, which, on each iteration i, moves one node on the border of a partition k_1 over the border to become a member of k_2. This slowly results in the desired random walk. The research conducted at MGGG, however, uses a graph recombination technique that has been shown to lead to faster space exploration (See figure 3). GRAPH-RECOM works as follows:

1. On iteration i with plan Q, select at random two bordering partitions k_1 and k_2.
2. Combine k_1 and k_2 to produce the MST for a combined graph k_c.
3. Split k_c by deleting any one edge of this MST to produce two new partitions k_1 and k_2
4. Return the modified partition Q'

Figure 3: The two images above show the result of one step of recombination for the state of Arkansas. Note how the blue districts remain unchanged, but the red and pink district exchanged some territory.



It is not guaranteed that Q' will meet the constraints c, but this is checked before Q' is accepted into the chain. This recombination technique results in larger changes on each individual step, which results in an exploration of the full space more quickly.

In research, these individual plans are never plotted, but are instead used to calculate statistics on the overall distribution of the possible plans. One such statistic is the **efficiency gap**, which is defined as $e(Q) = |\frac{\Sigma_k w_{dk} - \Sigma_k w_{rk}}{V}|$, where V is the total number of voters in a state in an election, w_{dk} is the number of wasted democratic votes in a district, and w_{rk} is the number of wasted republican votes in a district.

The number of wasted votes for a party in a district is equal to the number of votes for the party if the party lost, or the number of votes for the party minus the number needed to win if the party won. For example, if in District A there were 59 Republican and 41 Democratic votes, there would be 41 Democratic wasted votes and 59-42 = 17 Republican wasted votes.

3 Methods

The primary data source for this project were the available MGGG shapefiles for states around the country, available at https://github.com/mggg-states. These shapefiles contain geometric information about the precincts/counties of each state (depending on state law, one or both of those boundaries are used to draw congressional districts), as well as feature level information about each of these geometric objects. This allows us to draw districts as we choose, by aggregating low level features. Shapefiles from Iowa, Pennsylvania, and Georgia were used for the visualization. Many other states have shapefiles available, but were either old (Missouri), represented states that were too small to have meaningful value to explore the scope of the project (Utah, Rhode Island), or had deficiencies that made them impossible to use to render images, despite being fine for research purposes (Texas, Ohio, Wisconsin, North Carolina, Virginia, Michigan, Illinois, Massachusetts). The remaining states do not have shapefiles as of yet, or are states with only one congressional district (Alaska, Montana, Wyoming, North Dakota, South Dakota, Vermont).

For each of Iowa, Georgia, and Pennsylvania, a shapefile dissolving all boundaries into current congressional boundaries was created in QGIS 3.4. Features were extracted using MGGG code (gerrychain). In Pennsylvania, shapefiles were also generated for the old 2011 plan, as well as data on possible Democratic and Republican Gerrymanders as projected by FiveThirtyEight.

Next, each of the raw shapefiles was loaded into gerrychain and entered into the Markov Chain using the procedure described above. 1000 plans were drawn using GRAPH-RECOM for each of the three states, and every tenth plan was downloaded into a new shapefile for each state. This allowed for each plan to be noticeably different from the previous plans. The 100 downloaded plans served as the basis for the backend of the visualization.

It was initially our target to run Markov Chain simulations live - that is, to allow the user to select a state, ask for a recombination, and have the code execute at the time (so as to produce a new random plan every time, as opposed to the pseudorandomness of the 100 available plans). Unfortunately, existing technology does not allow this procedure to cohere with the nature of visualization, as it would take upwards of 90 seconds for the image to re-render every time.

In developing the front end, we sought a few key goals: (1) ease of use for the user - the user should be able to see a map of the United States, and on clicking a state, be easily able to understand how to navigate so as to generate new districting plans; (2) Mathematical understanding without lecture - it should be fairly intuitive that new computational plans are being generated, so that the user is able to see the art of these plans, while at the same time learning something about how these lines affect the drawing of districts; (3) minimalist design - there is plenty of documentation regarding gerrymandering both in this paper and in thousands of other sources - this web application itself should focus on the state map, without giving more details than necessary.

During the design process, there were many ideas that had to be cut for the sake of time or available resources. There was an idea to show multiple elections in the database, but it was determined that each shapefile had inconsistent data on the elections available. Therefore, the 2016 Presidential election was used across the board. We also considered showing a bar plot of each of the congressional districts in a given plan, but instead went ahead with a text-based design that also shows the efficiency gap.

4 Results

Figures 4, 5, 6, and 7 depict scenes from the application. In the end, the process generated a fully navigable visualization capable of showing interesting and new districting plans for each of the three studied states. In addition, the use of race as a feature allows for the understanding of majority-minority districts and looks for ways in which they can be made.

Figure 4: The main page of the visualization has a map of the United States, clearly depicting the states that are available for exploration and the states for which shapefiles currently exist in the MGGG database.

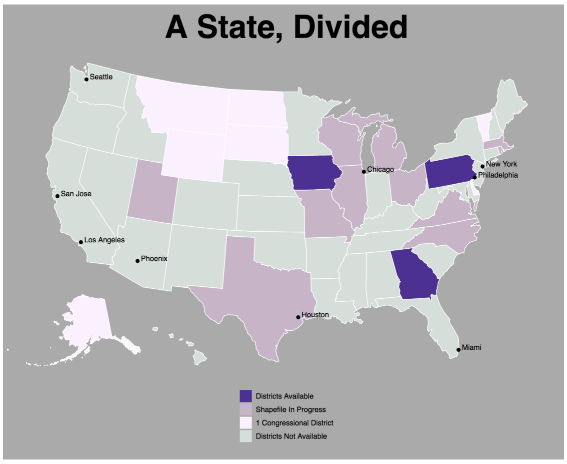
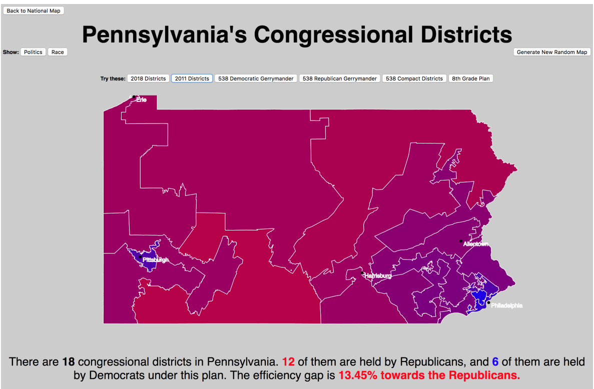


Figure 5: These are the political partisanships of each Pennsylvania district under the 2011 plan. Notice how the efficiency gap is 13.45% towards the Republicans - it was on this basis that the plan was found to be an outlier (analysis showed that the mean effifiency gap of the distribution is closer to 2.2% towards Republicans)



5 Discussion

These maps proved to be highly interesting for the user in understanding how gerrymandering actually works. On the request of a new map from the sample, the user can analyze how certain districts can be engineered to favor Republicans or Democrats despite having been generated at random using a Markov Chain. For example, even though Iowa was fairly Republican in

Figure 6: This is the racial makeup of each Georgia congressional district. In a state that is only 55% white, it makes sense that five of the fourteen districts in Georgia are majority-minority.

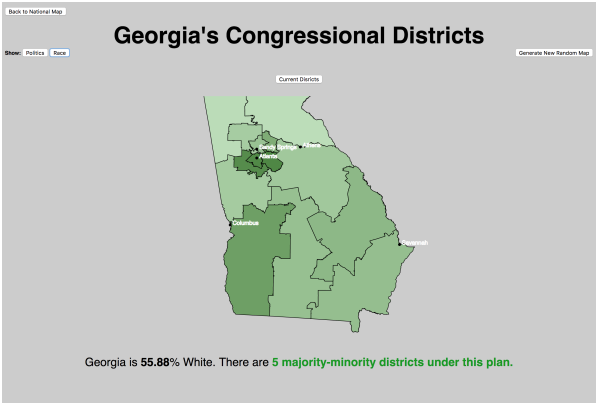
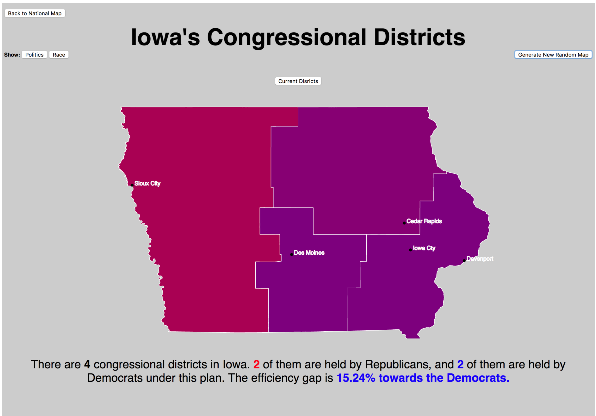


Figure 7: This randomized map of Iowa shows one way to make two Democratic Districts in a state that heavily preferred Donald Trump over Hillary Clinton. The plan makes one district around the left-leaning college towns in the center of the state, and another in liberal eastern Iowa.



2016, by concentrating one district around the college towns in Des Moines and Ames and another in the liberal eastern part of the state, Iowa could have had split representation in Congress in 2016. Similarly, in Pennsylvania the metropolitan area of Pittsburgh can be split in two to create two Democratic districts, or can be kept together to create only one. Most plans we saw that had one Pittsburgh district had an efficiency gap that favored Republicans.

On a broader point, while this visualization makes it easy to see how gerrymandered plans get engineered, it is quite difficult to actually gerrymander using this data, since no one plan has been specifically constructed to benefit one party. In other words, none of the computationally generated shapefiles have a specific prejudice to gerrymander, even though they demonstrate how it could be done in a different environment.

6 Future Work

This is a project that is tied to the author's research at MGGG, where research is being conducted on more improvements to the proposal process beyond GRAPH-RECOM. These proposal techniques are useful for getting even further understanding of what the space looks like, and can be used to develop even more maps just like this one. MGGG seeks to procure and tabulate more shapefiles in the future, and in doing so have the opportunity to populate this visualization with more states and more plans.

It is a goal of the author to one day have the ability to generate plans live and at random, which is currently computationally impossible since it takes too long to develop a new plan. A tool should exist which would allow the user to recombine two districts at a time, so as to structurally engineer a plan of their choosing (in essence, remove some of the randomization of the Markov Chain and allow the user to select a specific k_1, k_2 that they would like to redraw). It is also a goal to be able to compare time series across multiple elections.

Finally, it is a goal of the author to be able to communicate better with the public the mathematics behind gerrymandering. This visualization does do that by introducing random computational plans into a model largely resembling FiveThirtyEight, and therefore is novel in finding a central point between the public and research understanding of gerrymandering. However, a lot more can be done simply by removing some of the common misconceptions, such as assuming that biased plans need to look "ugly". We as a research community can also do better to explain to the public how these plans affect them directly, and visualizations like these can help to do exactly that.

Acknowledgments

Special thanks should go to Arvind Satyanarayan, professor of 6.894 for this term, for advice on how to proceed given the extremely tight schedule and the nature of a single-person team. Also thanks to Soya Park, the teaching assistant for this term, for her guidance. Finally thanks should be given to Daryl Deford, Postdoctoral Associate at MGGG and the supervisor to the author, and Justin Solomon, professor and PI for MGGG.

References

Christophe Andrieu and Johannes Thomas. 2008. A tutorial on adaptive mcmc. Stat Comput, (18):343-373.

Yves Atchade and Jeffery Rosenthal. 2005. On adaptive Markov chain Monte Carlo algorithms. Bernoulli, 5(11):815-828.

Gareth Roberts and Jeffery Rosenthal. 2008. Examples of Adaptive MCMC. probability.ca.

Allan Bycoffe et. al. The Atlas of Redistricting. ESPN, 2018. https://projects.fivethirtyeight.com/redistricting-maps/