Visualizing Gerrymandering

Chuyan Huang, Wenwei Liu, Kim-Anh-Nhi Nguyen
Massachusetts Institute of Technology, 6.894 Interactive Data Visualization—Spring 2019
77 Massachusetts Ave, Cambridge, MA 02139, USA
{chuang2, vickyliu, anhnhi}@mit.edu

Abstract

Gerrymandering is a practice intended to establish a political advantage for a particular party of group by manipulating district boundaries. Historically, different measurements of gerrymandering have been presented to courts to support rationales to claim illegal gerrymandering. Currently, there is not yet a universally agreed-upon metric for evaluating splitting of municipal units with districting plans. Commonly used metrics usually focus on either geographical compactness (number of cuts and splits) or partisan symmetry and vote efficiency (efficiency gap, mean-median, number of seats won by a certain party).

However, there is intricate interplay between the legal constraints and the measurements of interest, and among the measurements themselves. Researchers have struggled to understand the trade-offs between these intertwined metrics.

Visualizing Gerrymandering is a visual analytics system designed to enable the interactive exploration and analysis of different gerrymandering-related metrics' distributions across the United States. The toolkit utilizes 100,000 simulated redistricting plans with data from the Metric Geometry and Gerrymandering Group (MGGG). Using the state of Virginia and Pennsylvania as case studies, the authors develop a methodology of exploring joint probability distributions of five different metrics and demonstrate how such distribution data can be used to help policy makers and the general public understand the interactions between different metrics and demystify some common false beliefs.

Keywords: gerrymandering, interactive visualization.

1 Introduction

The MGGG has previously generated large ensembles of districting plans for different states in the United States. They used a Markov Chain Monte Carlo simulation and recombination steps to generate data. The data used to build the visualization took 100k steps to ensemble data. At each step of the chain, the statistics are measured, generated for every congressional district in any given state, and then aggregated on a collection of measurements (population balance, expected voting behaviour, geographic compactness, etc).[1]

The metrics generated for each observation includes: number of cuts, number of splits, efficiency gap of the plan, mean-median of the plan, and number of seats won by the Democratic party. Using the underlying data, the authors of this paper then generated histogram distributions of five metrics used to measure gerrymandering to further illustrate the underlying interactions among metrics.[2]

Other than work previously done by the MGGG, there have been other groups that visualize gerrymandering to better explain the concept. For example, Sathler visualizes the 2016 US House of Representatives election results and analyzes the imbalance at national or state levels using with efficiency gap analysis, where she plots histograms based on geo-political distributions in each state.[3] Princeton University also has a Gerrymandering Group that maintains an open-source database housing election precinct geographies for all 50 states.[4] However, these tools and systems mostly focus on one particular metric or districting plan. Instead, the authors of this paper propose a novel approach to inspect gerrymandering through distribution data, which will be the educational focus of *Visualizing Gerrymandering* as an interactive visualization system for policymakers, activists and the general public.

2 Methods

Visualizing Gerrymandering is designed with three different tools to allow users to interact on the webpage.

The first tool allows users to view some common metrics used to measure gerrymandering. The interactive visualization shows distribution histograms of the five metrics, while dynamically querying filtered data using the sliding selection bars as the range for the particular metric being filtered on and applying the filter on the other four distributions. Through these interactive techniques, this tool demonstrates the connections among the five metrics and how a given metric influences the other four.

The interactive functionalities of this tool further break down into three different components:

1. Drop down function that allows users to see definitions of any give metric and short descriptions of how to interpret these definitions
2. Drop down option for users to choose a state of interest and see how different states have different distributions based on the histograms of the above five metrics;
3. Radio buttons that allow users to choose their metrics of interest and filter data to see how the other metrics respond to the filter criterion.

[1] MGGG developed a Python library named GerryChain to build ensembles of districting plans. Documentation and more details available at: https://github.com/mggg/GerryChain.

2 Detailed definitions of metrics available at: https://github.mit.edu/pages/6894-sp19/Visualizing_Gerrymandering
3 Sathler, 2017, Visualizing Gerrymandering: Efficiency Gap Analysis of the 2016 House of Representatives Election Results. https://www.researchgate.net/publication/324091753_Visualizing_Gerrymandering_Efficiency_Gap_Analysis_of_the_2016_House_of_Representatives_Election_Results.
4 http://gerrymander.princeton.edu/.

2.1 Interactive Histograms

Histogram is used to demonstrate the distribution of metric values over all possible redistricting plans and showcases the notion that the shape of other metrics' distributions does not change monotonically with one metric skewed to one side. All five histogram charts have two layers: a background drop of a lighter color with lower opacity to remind users what the original distribution of such metric looks like, and a histogram of brighter color(s) and higher opacity in the front to represent the updated distributions based on the filter applied through the sliding selection bars.

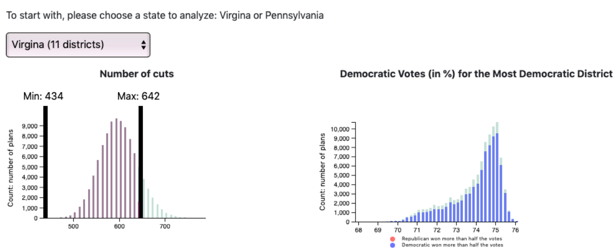

Figure 1: On the left filter on number of cuts, on the right how Democratic vote percentage change based on filtering

The color encoding on each metric also follows the political representation as red for Republican and blue for Democratic. Thus, the user can have a visual understanding of how each state's political composition is significantly different, and how the numerical value of the metrics affects the state's election results visually.

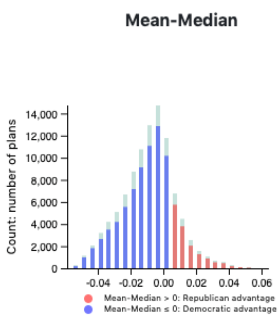

Figure 2: The mean-median metric was broken down by Republican advantage and Democratic advantage based on party color encodings

With the above tool, users would have a much better understanding of the interactions between metrics and how one particular metric impacts the others.

2.2 Interactive Map

The second tool is an interactive design that serves as quiz for users to play with. Users are first asked to make a guess, based on the maps shown, which one looks as there exists gerrymandering based on the color for each district (color gradient correlated with political representation – red for Republican, blue for Democratic). The takeaway designed here is that the judge or decision maker should not decide if there exists gerrymandering or not simply based on the map or how each district behaves. Rather, to have a more comprehensive understanding of the overall well-being of the state, a judge should look at where the districting plan stands in the overall distribution.

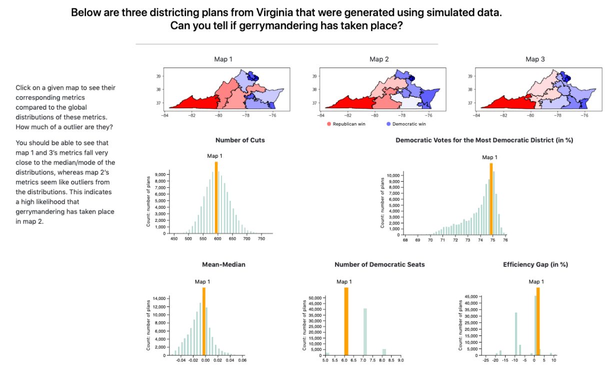

Figure 3: Based on map selection, highlight corresponding metrics on distributions

2.3 Dropdown Comparison

With the third tool, users are asked to observe how Virginia and Pennsylvania differ in all five previously mentioned metrics. The drop-down option allows users to choose a metric of interest and compare Virginia and Pennsylvania in their distribution based on different histograms. This is to further reinforce the notion that gerrymandering metrics should not be compared across states, because states differ dramatically both geographically and demographically.

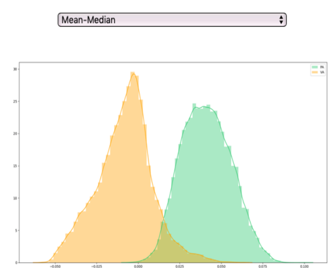

Figure 4: Based on the dropdown selection, display of metric comparison for Virginia and Pennsylvania

3 RESULTS

Visualizing Gerrymandering comes in as a tool which can serve to demonstrate the hidden correlations of multiple distributions and effectively represent the distribution values within high dimensional data. In addition, the MGGG is looking to have a tool that not only can explain the metrics and plans underlying generated data, but a tool that can easily demonstrate how current districting plans and current measurements of districting plans are not the best study to measure Gerrymandering by the court system or other users.

The address these pain points, *Visualizing Gerrymandering* aims to achieve three main objectives:

1. Show interactions between the different metrics by making users filter on the metric they want, and show how the other metrics change accordingly.
2. Make the user guess from the map and show how the underlying metrics are significantly different from what users might have previously assumed by looking at plan maps, to further reinforce of notion that distributions, instead of a single plan or metric, should be used to evaluate gerrymandering.
3. Demonstrate how each state have significantly different metrics, and thus it is not meaningful to compare among states or create 1 metric to measure the overall Gerrymandering behavior across the U.S.

4 Discussion

From observing users interact with *Visualizing Gerrymandering*, it seems that the system is effective to communicate and to inform users on the following aspects:

1. Help user better understand the interactions between different metrics.
2. Use the tool to observe the impact on change in certain redistricting criteria on partisan and population balance.
3. Looking at any particular metric or districting plan by itself is meaningless - this is why distributions of metrics are used across large ensembles of districting plans. What could be meaningful is how much of an outlier a plan is compared to the distribution of plans.
4. Metrics should not be compared across different states because states are geographically and demographically different by nature.

5 Future work

With students, faculty and guests who were interested in gerrymandering stopping by our table during the poster session, we have found that there is potential for future applications beyond the original scope of the project and where we could improve for next steps.

First, there could be more informative descriptions and pictures displayed to better demonstrate the meaning of each defined metric. The webpage was designed to assist MGGG when they educate political decision makers and activist groups on their understandings of different metrics, with the underlying assumption that they all understand what mean-median, number of cuts, efficiency gap, etc. mean. However, with more people stopping by our table during the day of presentation, we came to realize that there is a large group of people who are educated with the idea of gerrymandering but are not familiar with specific metrics used to measure gerrymandering.

We have made a conscious decision to include a dropdown bar which explains different metrics in text. However, it would be hard for people who have no previous understanding of such metrics to read heavy text. For next step, to educate people and entertain them at the same time, we would like to use graphical representations of different metrics. For example, in our video poster, we have presented a way of interpreting number of cuts using a graphical example. We would like to use similar graphical representations to illustrate other metrics that used in our analysis.

Second, we could allow users to have access to different redistricting map plans, including the current and historical plans for different states. Virginia and Pennsylvania were used as case studies to illustrate some of the educational lessons we would like users to walk away with. However, it seems like people would be interested in using this tool to inspect other states as well.

Third, we would like to build more functionalities to enable the interactive visualization system to go beyond just educational purposes and support actual decision making. After understanding and agreeing with some of the findings from the case studies, people seemed interested in what purpose this tool could serve in the court system and how it could help policymakers make better decisions. Of course, further user research would need to be conducted before building such functionalities.

Acknowledgements

We thank Arvind Satyanarayan, Assistant Professor at MIT EECS/CSAIL, and Soya Park, Ph.D. student at MIT CSAIL, who designed and intructed the class. We thank Justin Solomon, Principal Investigator, and Daryl DeFord, Postdoctoral Associate, at Metric Geometry and Gerrymandering Group, MIT CSAIL, who made the districting plans data available to us and gave valuable feedback on the visualization system.

References

[1] Open-sourced Python library developed and maintained by the Metric Geometry and Gerrymandering Group, https://github.com/mggg/GerryChain.

[2] Sathler, 2017, Visualizing Gerrymandering: Efficiency Gap Analysis of the 2016 House of Representatives Election Results. https://www.researchgate.net/publication/324091753_Visualizing_Gerrymandering_Efficiency_Gap_Analysis_of_the_2016_House_of_Representatives_Election_Results.

[3] Princeton Gerrymandering Project website, http://gerrymander.princeton.edu/.