dViz FALL 2017
Carlos Sathler (cssathler@gmail.com)

INFO-I590 Data Visualization

Project title: Visualizing Gerrymandering in US Congressional Elections

Team members: Carlos Sathler

Abstract

We will perform an exploratory data analysis of potential gerrymandering [1,2] in US congressional elections. Gerrymandering "requires illicit intent" [3, p1] which is not easy to prove with readily available data [3]. For that reason, we will limit ourselves to producing visualizations that identify states where gerrymandering *could* be occurring. There are two main "red-flags" for the potential existence of gerrymandering: first, a highly irregular geographic shape of district boundaries [4]; second, a high incidence of "wasted votes" [6] across a state's congressional districts. For the mid-term project we will focus our analysis on wasted votes. Geographical analysis of gerrymandering may be explored on the final project.

Introduction

Gerrymandering is "a practice intended to establish a political advantage for a particular party or group by manipulating district boundaries" [1]. An easy way to explain gerrymandering is illustrated in Figure 1 below [7].

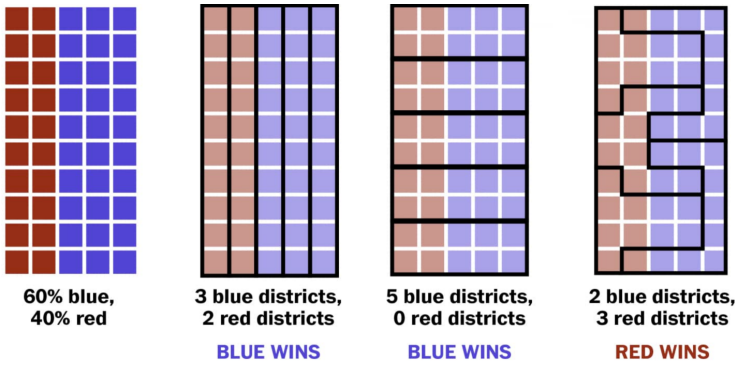

Fig 1 – Source Washington Post [7]

The leftmost picture in Figure 1 represents a population of 50 people in a hypothetical rectangular state. Imagine this state can elect 5 representatives to the US House of Representatives, each representing 10 people. Since the proportion of red and blue voters is 2:3, a perfectly fair election would elect 2 representatives of the "red" party and 3 representatives of the "blue" party to the House of Representatives.

The grouping of voters by boundaries inside a state is called "districting". Districts are redefined every 10 years based on census numbers, so that each district have the same population. Each electoral district in a congressional election will elect a single representative [8]. On Figure 1, the second picture on the left shows district boundaries achieving ideal representation; the third and fourth show examples of districting that favor one or the other party, and are both unfair. The third illustration is unfair to "red" voters and the fourth is unfair to "blue" voters.

The primary motivation for our project is the identification of states where certain parties are blatantly under/over represented due to unfair districting. We will use the concept of "efficiency gap" for our analysis. This concept was conceived for the specific purpose of assessing gerrymandering [9, 10] and is being used in the current Wisconsin gerrymandering case [11]. The efficiency gap is calculated as "the difference between the two parties' wasted votes, divided by the total number of votes" [6].

A "wasted vote" is a vote that does not contribute to win an election. So all votes for a losing candidate are "wasted" in that sense, and all votes for a winning candidate that are above of the 50% required threshold to win the election are also wasted. The efficiency gap therefore measures the percentage of votes that failed to contribute to place a party's candidate on an electoral seat.

The issue with wasted votes can be well illustrated by the 2016 congressional elections. According to the Wikipedia info box for the United States House of Representative Elections, 2016 [34] Republicans received 49.1% of the popular votes while Democrats received 48%. Yet Republicans won 241 seats while Democrats only won 194; the Republican Party was able to convert 1% more votes into 24% more seats in the congress.

Seats won	**241**	194
Seat change	▼6	▲6
Popular vote	**63,173,815**	61,776,554
Percentage	**49.1%**	48.0%

Figure 2
Section from United States House of Representative Elections, 2016 Wikipedia info box [34]

Our goal is to develop an impactful visualization to show efficiency gap numbers for all states. We plan to spend considerable time contrasting efficiency gap numbers for different states, particularly those with districts identified as highly gerrymandered districts [5]. We hope that through impactful visualization that clearly highlights unfair representation we will raise awareness to existing districting issues, so that voters possibly become motivated to join by-partisan efforts to stop gerrymandering [12,13,14,15], thus contributing to a more fair electoral system and a stronger democracy.

Background and related work

We identified many visualizations that explain gerrymandering with play data [7, 16], including the one included herein as Figure 1. There are many visualizations on the web showing maps of districts with highly irregular boundaries that clearly indicate gerrymandering, but do not include any data analysis [15, 16, 17]. The online game "The Redistricting Game" [18] is worth mentioning here because it offers an excellent opportunity for players to understand how gerrymandering really works. In the game players engage in districting a fictitious state to achieve political advantage.

In addition to the issue of wasted votes and unfair representation that we will explore on our project, there are at least two other issues that motivated our interest in the topic of gerrymandering. One of them is well explained in two videos featuring governor Arnold Schwarzenegger [19, 20]. The videos explain how politicians engage in gerrymandering to perpetuate themselves in the US House of Representatives, regardless of their performance.

We found visualization that corroborates Mr. Schwarzenegger assertion. According to OpenSecrets.org the US House of Representatives re-election rate in the period of 1964-2016 was seldom less than 85% [21]. Mr. Schwarzenegger created a "Crowdpac" to combat gerrymandering [22].

The second issue that motivated our project is related to the possible impact of gerrymandering in the perceived growing polarization of our political system [22, 23, 24]. One possible way to approach the topic is the Cook Partisan Voting Index [25, 26] which measures "how strongly a United States congressional district or state leans toward the Democratic or Republican Party, compared to the nation as a whole" [25]. We found one visualization that analyzes this data: The State of Gerrymandering, created by an organization called Silicon Valley Data Science [27]. While we will not analyze polarization for this project we will likely use datasets used by the Silicon Valley Data Science in their study. This data is available in github [28].

In addition to the above mentioned study, three other works on gerrymandering offer great explanations of gerrymandering, and include visualizations with real data:

1. gerrymander.princeton.edu [29] has many useful links, proposes a test to determine if a state has gerrymandering, and allows users to explore how gerrymandered their state is; it includes visualizations but data is not available.
2. Quantifying Gerrymandering [30] has many useful links, shows interactive visualization, and list data sources.
3. The Most Gerrymandered States Ranked by Efficiency Gap and Seat Advantage [31] has many useful links and analyzes efficiency gap; data is not available.

Datasets

We will create the dataset for the project by parsing the New York Times webpage "House Election Results: G.O.P. Keeps Control" [32] which contains information on election results for House of Representatives for all districts in the union. For districts where candidates won "uncontested" however, the site does not offer vote count, so we will rely on Wikipedia to obtain vote count for districts in which candidates had an "uncontested" victory. Vote count is required information for calculation of "efficiency gap".

Figure 3 below shows a screenshot of the New York Times webpage "House Election Results: G.O.P. Keeps Control" [32].

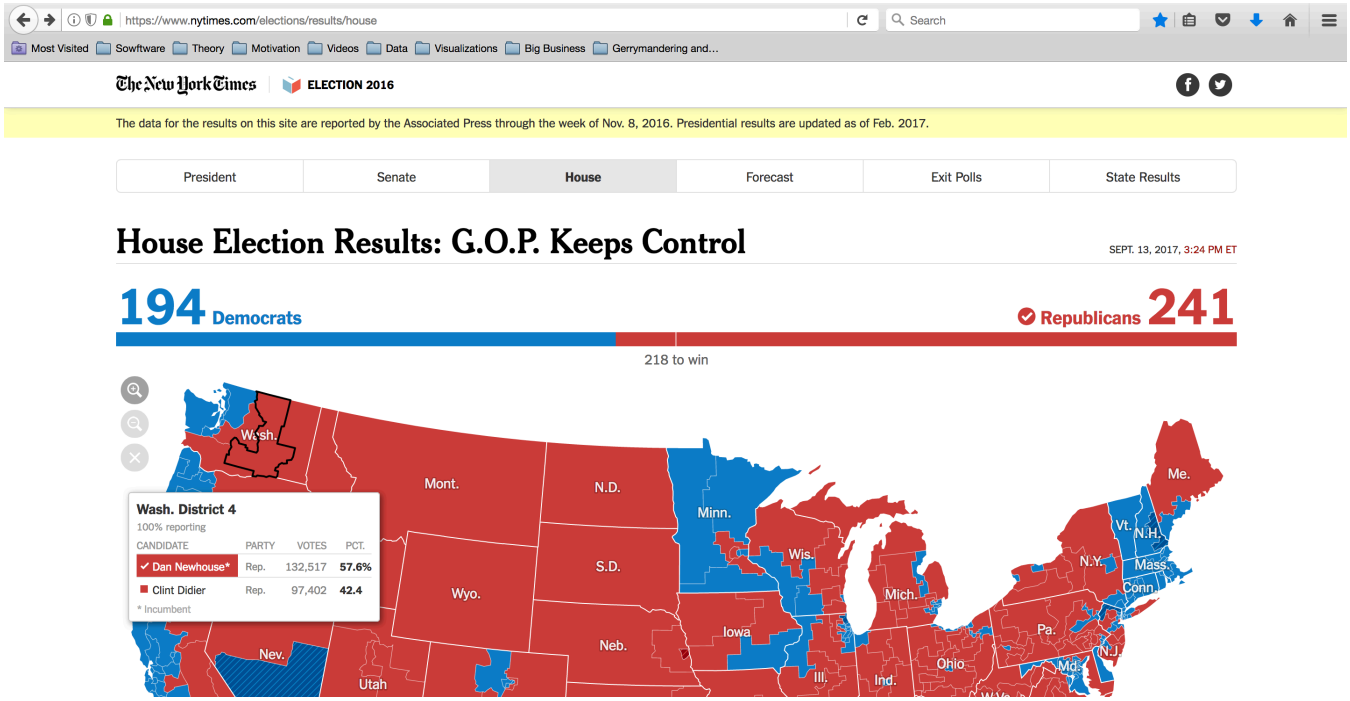

Figure 3
New York Times webpage "House Election Results: G.O.P. Keeps Control" [32]

Preliminary analysis of the html file for this page shows that all election results data is stored in a single variable in the html code, in JSON format, so it will be relatively easy to parse the file to obtain the data, and create the dataset for the project. Vote count for uncontested district wins will be compiled manually, as stated above, from Wikipedia pages.

Visualization Ideas

To illustrate initial visualization ideas for the project, we explore 3 examples of efficiency gap, using Figure 1 from the introduction. The calculations presented below are based on Eric Petry's explanation of the efficiency gap in the article How the Efficiency Gap Works [33].

Example 1: No gerrymandering - Figure 1, 2nd scenario.
Assume districts are numbered 1 to 5 from left to right.

	Red Votes	Blue Votes	Wasted Red	Wasted Blue	Net Wasted
District 1	10	0	4	0	4 Red
District 2	10	0	4	0	4 Red
District 3	0	10	0	4	4 Blue
District 4	0	10	0	4	4 Blue
District 5	0	10	0	4	4 Blue
Totals	**20**	**30**	**8**	**12**	**4 Blue**

Proportion of Blue votes = 60%
Proportion of Blue wins = 60%
Efficiency Gap = 4/50 = 8%

Example 2: Heavy gerrymandering, no irregular district boundaries - Figure 1, 2nd scenario.
Assume districts are numbered top to bottom.

	Red Votes	Blue Votes	Wasted Red	Wasted Blue	Net Wasted
District 1	4	6	4	0	4 Red
District 2	4	6	4	0	4 Red
District 3	4	6	4	0	4 Red
District 4	4	6	4	0	4 Red
District 5	4	6	4	0	4 Red
Totals	**20**	**30**	**12**	**0**	**20 Red**

Proportion of Blue votes = 60%
Proportion of Blue wins = 100%
Efficiency Gap = 20/50 = 40%

Example 3: Heavy gerrymandering, high irregular district boundaries - Figure 1, 3nd scenario.
Assume districts are numbered top to bottom, left to right.

	Red Votes	Blue Votes	Wasted Red	Wasted Blue	Net Wasted
District 1	6	4	0	4	4 Blue
District 2	6	4	0	4	4 Blue
District 3	6	4	0	4	4 Blue
District 4	1	9	1	3	2 Blue
District 5	1	9	1	3	2 Blue
Totals	**20**	**30**	**2**	**18**	**16 Blue**

Proportion of Blue votes = 60%
Proportion of Blue wins = 40%
Efficiency Gap = 16/50 = 32%

Plot ideas:

1. Efficiency gap by state (bar chart or heat map with tabular data)
2. Districts with highest relative number of wasted votes (same as above)
3. Seats lost due to districting, by state (percentage of votes vs. percentage of seats)
4. Seats gained due to districting, by state (same as above)

Ideas for iterative capabilities:

1. Allow user to generate plot(s) by first selecting a state
2. Allow user to generate plot(s) by first select top/bottom states on a particular metric
3. Allow user to generate plot(s) that display total for entire country
4. Add links to the plots so user can jump to Congressperson's webpage
5. Add links to organizations that are fighting gerrymandering
6. Add links to pages that educate on the subject

Research questions and working hypotheses

Are wasted votes and particularly the efficiency gap good indicators of gerrymandering? We expect that yes; they will be good indicators of gerrymandering. We will confirm our findings by analyzing (1) the current shape of the districts in the states where the efficiency gap suggests gerrymandering is occurring and (2) by analyzing the historical changes to district boundaries in those states. Since this first phase of the project does not include map displays, the verification will be done manually.

References

[1] https://en.wikipedia.org/wiki/Gerrymandering
[2] http://reclaimtheamericandream.org/brief-gerrymander/
[3] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2583528
[4] https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see/?utm_term=.0c2983149766
[5] https://en.wikipedia.org/wiki/Wasted_vote
[6] https://www.washingtonpost.com/news/wonk/wp/2014/05/15/americas-most-gerrymandered-congressional-districts/?utm_term=.6acac63037ea
[7] https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see/?utm_term=.ce5c534365c6
[8] http://archive.fairvote.org/?page=526
[9] http://www.brennancenter.org/sites/default/files/legal-work/Stephanopoulos_McGhee_Efficiency_Gap_2015.pdf
[10] https://newrepublic.com/article/118534/gerrymandering-efficiency-gap-better-way-measure-gerrymandering
[11] http://www.npr.org/sections/thetwo-way/2017/06/19/533519165/supreme-court-agrees-to-hear-wisconsin-gerrymandering-case
[12] http://www.fairvote.org/gerrymandering#gerrymandering_key_facts
[13] http://www.commoncause.org/issues/voting-and-elections/redistricting/
[14] https://www.crowdpac.com/communities/arnold
[15] https://rantt.com/the-top-10-most-gerrymandered-states-in-america-bd962843ba1f
[16] https://www.lobbyists4good.org/single-post/2017/01/23/Is-this-the-root-cause-of-our-political-dysfunction
[17] https://www.wired.com/2016/01/gerrymandering-is-even-more-infuriating-when-you-can-actually-see-it/
[18] http://www.redistrictinggame.org/index.php
[19] https://www.youtube.com/watch?v=Qfc3N0ZngXs
[20] https://www.youtube.com/watch?v=2w-pgvRsIuE
[21] https://www.opensecrets.org/overview/reelect.php
[22] https://www.brookings.edu/blog/brookings-now/2017/07/06/a-primer-on-gerrymandering-and-political-polarization/
[23] https://www.jstor.org/stable/25548144?seq=1#page_scan_tab_contents
[24] http://endgerrymanderingnow.org/2014/07/17/does-gerrymandering-cause-polarization/
[25] https://en.wikipedia.org/wiki/Cook_Partisan_Voting_Index
[26] https://www.dailykos.com/stories/2013/12/10/1261624/-Partisan-Voting-Index-Calculator
[27] http://svds.com/gerrymandering/
[28] https://github.com/silicon-valley-data-science/gerrymandering-congress
[29] https://gerrymander.princeton.edu
[30] https://services.math.duke.edu/projects/gerrymandering/
[31] https://www.azavea.com/blog/2017/07/19/gerrymandered-states-ranked-efficiency-gap-seat-advantage/
[32] https://www.nytimes.com/elections/results/house
[33] https://www.brennancenter.org/sites/default/files/legal-work/How_the_Efficiency_Gap_Standard_Works.pdf
[34] https://en.wikipedia.org/wiki/United_States_House_of_Representatives_elections,_2016