Analysis of the 2016 Connecticut Senate and Presidential Elections

Lisa Jurca and Blayde Omura
Code Link: `https://github.com/blaydeomura/AI_for_redistricting_final_project.git`

April 2024

1 Introduction

Redistricting is the redrawing of electoral district boundaries and constituencies and is a fundamental process in the realm of electoral governance. The goal of redistricting is to ensure a semblance of equitable representation by striving for roughly equal populations within each district. Census data updates typically prompt revisions to accommodate shifts in population distribution. Redistricting aims to uphold the principles of democratic governance by adjusting electoral boundaries to reflect demographic changes accurately.

In this paper, we will be analyzing the 2016 Connecticut Senate and Presidential elections to assess the fairness of electoral district maps and evaluate potential instances of gerrymandering. Using the "ReCombinator", or ReCom redistricting algorithm, we were able to quickly identify and disregard unbalanced partitions by achieving roughly equal populations in each electoral district. We performed random walks by simulating different configurations of districting plans through a Markov Chain Monte Carlo method. This allowed us to conclude our outlier analysis through the creation of histograms and graphs, which we will explain in depth in this paper.

2 Legal, Partisan, and Racial Landscape in Connecticut

In Connecticut, the legal landscape of electoral districting is characterized by constitutional provisions governing the redistricting process. Connecticut law requires that districts be drawn to ensure about equal population while respecting communities of interest and minimizing the division of political subdivisions. State law requires that Congressional and legislative district lines be redrawn every 10 years. Additionally, the state constitution mandates that districts be contiguous and compact, with boundaries drawn without regard to political affiliations or incumbency. A bipartisan reapportionment committee is responsible for drafting the redistricting plan, and the plan must be approved by the Connecticut General Assembly.

As for the partisan and racial landscape, Connecticut has historically been a Democratic-leaning state, as Democratic candidates prevailed in both state and federal elections. Connecticut is a blue state government trifecta, meaning there is a Democratic majority in the House of Representatives, Senate, and Governorship. However, like many states, there have been instances of racial, partisan, and prison gerrymandering, leading to debates over the fairness and equity of electoral district boundaries. In 2011, the NAACP, in collaboration with Connecticut NAACP members, took a solemn stand against prison gerrymandering, mounting a federal lawsuit against the state's legislative maps. Their argument resonated with the principle of "one person, one vote". They contended that every citizen, regardless of their incarceration status, deserved equitable representation and political agency. Despite initial challenges, the case persevered, advancing to a three-judge panel for thorough scrutiny. Yet, both parties ultimately sought a mutual dismissal, a decision met with judicial concurrence. Though this legal chapter has drawn to a close, its implications remain profound. It underscores the enduring struggle for integrity and fairness within the electoral process.

3 Creation of Electoral District Map for Connecticut

The information used in this paper was collected from a shapefile, which includes census and election data cleaned and reformatted. Data cleaning and reformatting were performed using the Python libraries MAUP, Pandas, and Geopandas. Census data was retrieved from the 2020 PL 94-171 Data for Connecticut, sourced from the Decennial Census at the Block level on 2020 Census Redistricting Data (P.L. 94-171) shapefiles. While this data provides a comprehensive overview, it has not been validated against the official data utilized by the state's redistricting body. Variations most likely exist. Some states handle certain demographics differently, such as incarcerated people. Additionally, they may exclude non-permanent residents from the PL 94-171 data file for redistricting, among other potential modifications. The data retrieval dates for both the PL 94-171 Legacy Format Data and the Census Redistricting Data Shapefiles were recorded as August 12, 2021, and May 24, 2021, respectively.

The 2022 Connecticut Congressional Districts Plan dataset was retrieved from Nate Persily with a retrieval date of February 14, 2022. No modifications were made to the data by RDH during processing. In regards to the 2016 Connecticut precinct and election results shapefile, election results were sourced from the Connecticut Secretary of State, while the precinct shapefile was obtained from the U.S. Census Bureau's 2020 Redistricting Data Program. The RDH retrieval date for this dataset was August 17, 2021. This approach to sourcing and processing data ensures transparency and accuracy in election analysis.

4 Markov Chains

There may be instances when we want to see a map's level of typicality. We would assess this by using a technique that compares a random sample to other maps. This technique is using a Markov Chain Monte Carlo (MCMC) process, which allows us to create an ensemble of maps. A Markov Chain is a stochastic model devoid of memory, used to represent transitions between different "states" within a system. One step involved in the process is the "recombination step". This involves the selection of two adjacent on the map, and merging their underlying graphs. Afterwards, a spanning tree is chosen for the resulting district which is double the size. Afterward, it is divided into two equal-sized halves, which make minor alterations and create a larger change in the map's configuration. This results in creating more compact districts. More compact districts aide in geographically cohesive electoral boundaries.

We performed the analysis on data from the Redistricting Data Hub, as described in section 3. This was performed with the assistance of Python libraries such as Gerrychain, Pandas, and Geopandas. Gerrychain is a python library designed for the analysis of political redistricting plans. Its abilities include generating the ensembles, as well and computing metrics to evaluate those plans. It is particularly useful when implementing algorithms such as the MCMC process. Pandas is a data manipulation and analysis library that allows for manipulation of structured data. It allows us to clean, aggregate, and analysis data. Geopandas is an extension of pandas that adds support for geospatial data. Libraries such as Matplotlib and NumPy were also used, both assigned with visualizations and numerical computing in Python.

5 Ensemble Analysis Process

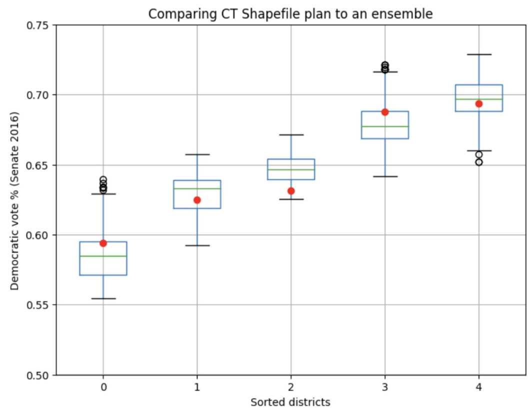
The goal of this ensemble analysis is to analyze the collective properties of an ensemble of maps. Properties may include population equality, partisan fairness, or compactness. We first needed a working CT.shp file by cleaning up the shapefiles retrieved from the Redistricting Data Hub. We then had a shapefile with working data which includes: census, precinct, congressional, and election data. Before achieving success with Connecticut, we attempted to analyze maps from Maryland, Massachusetts, Texas to no avail.

We began the ensemble analysis process by creating the initial partition. To avoid unbalanced predictions, we defined an ideal population for the districts. The ReCom proposal requires knowledge of the target population for each district. This enables us to expedite the process by discarding unbalanced partitions. This was done by dividing the number of districts by the total population. Afterward, we created the proposal. A proposal in this context refers to a mechanism for generating a new configuration in the Markov Chain. The goal is to efficiently explore the possible redistricting plans. Afterwards, the Markov Chain is created. We defined the total number of steps in the random walk as 1500. We were able to generate marginal box plots to see the results in our ensemble.

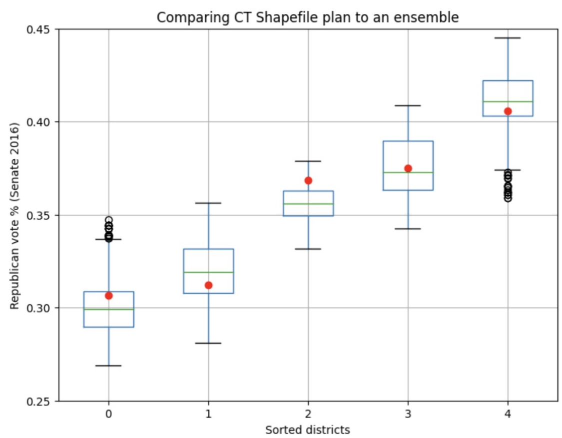
The districts in the graphs are not defined by the number of the actual districts. They are ordered by the vote percentage during a step in the walk, from least to greatest. For instance, district '0' in both graphs have the least Democratic and Republican vote percentages, while '4' have the highest vote percentages. The box plots represent the distributions, which allow us to compare this to the actual vote percentage. This can tell us if the distribution is close to the median or an outlier. We can check the dots to check for possible packing or cracking.

5.1 Marginal Box Plots Senate 2016 Election

The following marginal box plots utilize data from Connecticut's 2016 Senate Election. Here we are comparing the Connecticut shapefile plan with an ensemble of plans, which illustrates the distribution of Democratic and Republican vote percentages across various districting plans.



(a) Democratic Senate 2016.



(b) Republican Senate 2016.

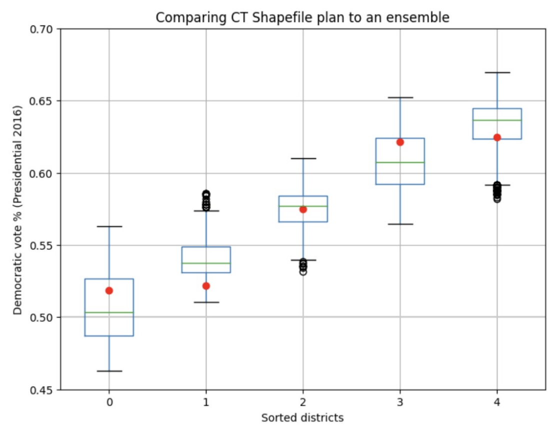
Figure 1: Marginal Box Plots Senate 2016

Figure a represents the Democratic vote percentage of the 2016 Senate election. The red dot in the marginal box plot represents the initial districting plan. We can see in this plot that the initial districting plan's Democratic vote percentage is generally near the median of the distribution. Additionally, the lowest Democratic district is still about 50 percent, which allows all districts to be democratic.

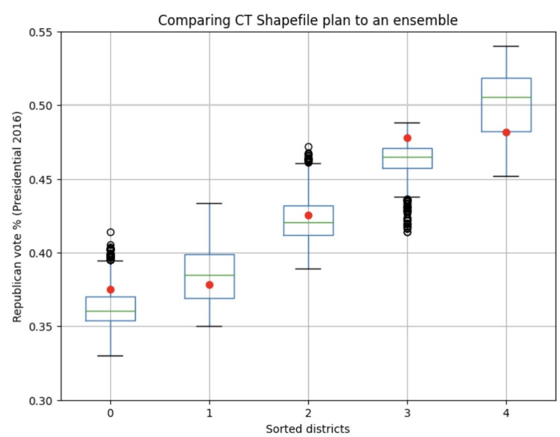
Figure b represents the Republican vote percentage of the 2016 Senate Election. In this box plot, we can observe that the district with the highest Republican vote percentage is still lower than 50 percent. This illustrates how Republicans were not able to win a district in this election.

5.2 Marginal Box Plots Presidential 2016 Election

The following marginal box plots are utilizing data from Connecticut's 2016 Presidential Election. Here we are comparing the Connecticut shapefile plan with an ensemble of plans, which illustrates the distribution of Democratic and Republican vote percentages across various districting plans.



(a) Democratic Presidential 2016.



(b) Republican Presidential 2016.

Figure 2: Marginal Box Plots Presidential 2016

In the figure a, we can see the Democratic Vote Percentages. The votes range from about 50 percent, which demonstrates a relatively high level of support for the Democratic candidate. Additionally, the interquartile ranges are fairly tight. In figure b, we can see the Republican vote percentages are consistently lower than the Democratic percentages. There are significant outliers here where the Republican vote exceeds the typical range.

When we compare both plots, we can see that Democratic votes are higher across all the districts. This is a consistent trend within the Presidential election, along with the previous Senate election. This aligns with the Democratic trends Connecticut has had in previous national elections. Democratic votes show less variability compared to Republican votes. This suggests a very stable Democratic base across districting plans. In terms of outliers, the Democratic outliers show low support. This differs from the Republican graph, where outliers are generally higher. This show strong Republican support within the outliers.

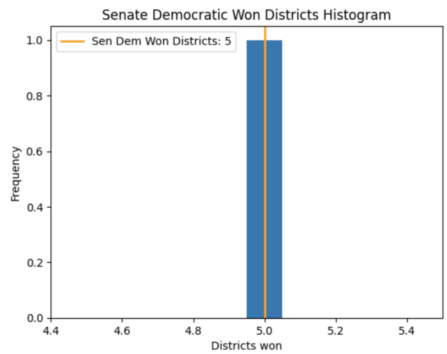
6 Gerrymandering Metrics

Gerrymandering is a term entrenched in political discourse. Gerrymandering represents a practice where electoral district boundaries are deliberately manipulated for political advantage. Gerrymandering is typically seen in two forms: "packing" and "cracking". "Packing" is seen when voters of a particular political affiliation are concentrated into a few electoral districts. This strategy aims to minimize their overall impact on election outcomes by ensuring that they win those specific districts overwhelmingly. Conversely, "cracking" is where voters of a particular political affiliation are dispersed across multiple electoral districts. This strategy aims to dilute their voting power, preventing them from forming a majority in any single district and reducing their overall influence on election outcomes.

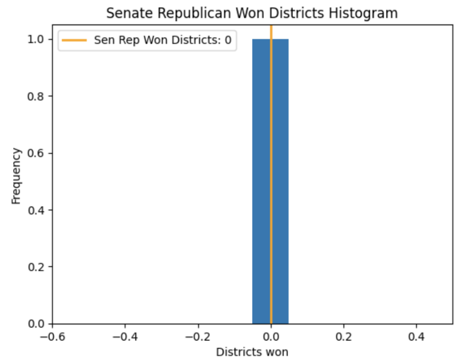
In the following sections, we will explain the results of our findings during our outlier analysis. To generate the histograms, we ran our random walk. We extracted the results from the Senate 2016 Election from an initial partition, as well as determined the number of districts won by each party. Afterwards, we iterate over a Markov Chain, using the MCMC method.

6.1 Democratic vs Republican Won Districts

Here we will discuss the amount of districts that were won by each party. More specifically, Democratic and Republican when discussing larger scale elections. This is an important metric because it is crucial to know this ratio when discussing gerrymandering as it helps us recognize disproportionate representation as a whole. However, the use of other metrics is crucial for determining possible bias that leads to these results.



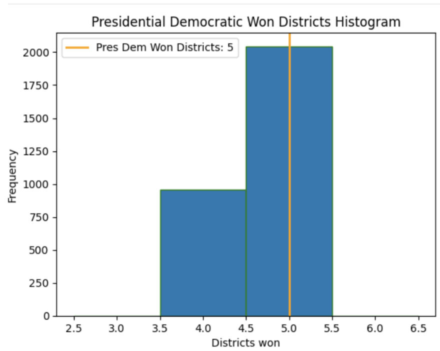
(a) Democratic Districts.



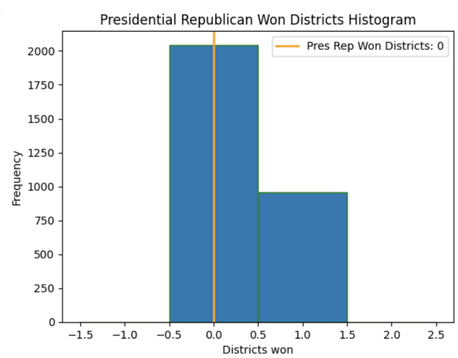
(b) Republican Districts.

Figure 3: Senate Won Districts Histograms

The above figures show the amount of districts won by the Democrats and Republicans during the 2016 Senate election. In this election, the Democrats won 5 districts. The Republicans won 0 districts.



(a) Democratic Districts.



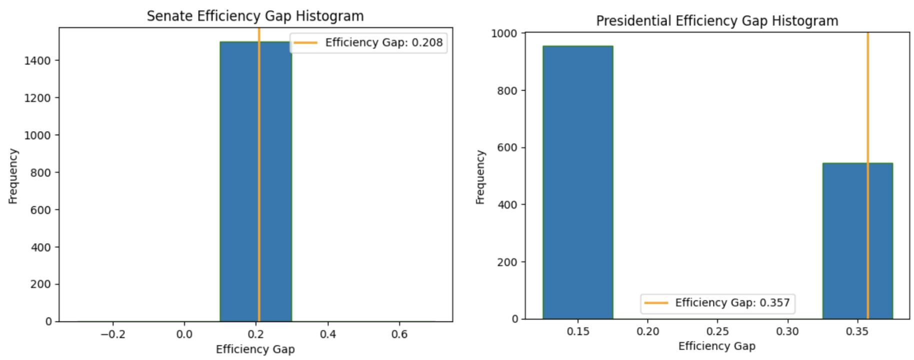
(b) Republican Districts.

Figure 4: Presidential Won Districts Histograms

The above histograms depict the amount of districts won in the 2016 Presidential election. Similar to the previous histograms, the Democrats won 5 districts while Republicans won 0 districts. This aligns with Connecticut's tendency to be a Democratic-leaning state.

6.2 Efficiency Gap

The Efficiency Gap is a metric that measures the distribution of wasted votes between political parties. Wasted votes are those cast for losing candidates, or surplus votes beyond what is needed for a candidate to win. A significant gap between the wasted votes of two parties suggests possible gerrymandering. This is particularly plausible if one party consistently has more wasted votes than the other. In terms of gerrymandering, the Efficiency Gap is a fine metric, but it has its flaws. For example, it assumes the distribution of votes should proportionally reflect the distribution of seats. Additionally, it does not account for natural geographic sorting, or instances were those who live in a certain area are like-minded.



(a) Senate Election EG.

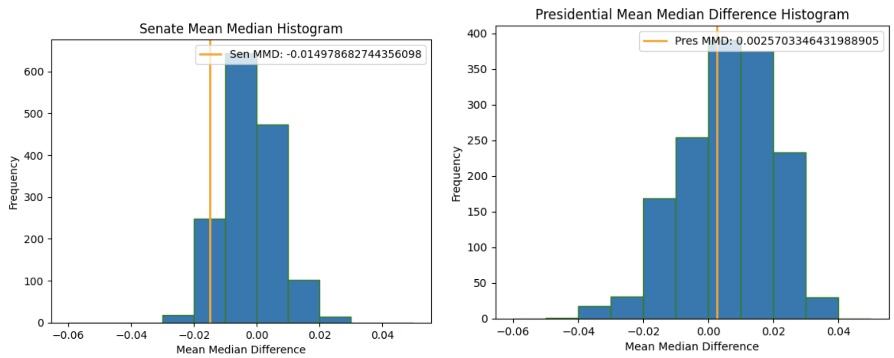
(b) Presidential Election EG.

Figure 5: Efficiency Gap Histograms

In figures a and b, we see histograms that represents the Efficiency Gap calculations for the 2016 Senate and Presidential Elections. We calculated the Efficiency Gap using Gerrychain's built-in methods. Both histograms, a and b, show proof that Democrats did better in both elections via the positive efficiency gap. This means that republicans have more wasted votes in both histograms. You might imagine these efficiency gap metrics prove gerrymandering, but it may be more indicative that Democrats simply won every district. This claim is further backed up by the marginal box plots previously mentioned. Republicans wasted their votes for each district because they lost every district.

6.3 Mean-Median Difference

The Mean-Median Difference examines the disparity between average and median voting preferences within districts. A larger difference implies political manipulation. It suggests certain districts may have an unusually high or low concentration of voters with specific preferences. The mean-median difference metric has been under scrutiny as one of the primary metrics of gerrymandering. This is due to the mean-median difference reflecting the distribution of numerical data, while not accounting for geographical or political boundaries of electoral districts. However, mean-median difference is a fine metric for calculating political skew of the data. We used the built-in method for mean median difference by Gerrychain for our calculations.



(a) Senate Election MM.

(b) Presidential Election MM.

Figure 6: Mean-Median Difference Histograms

Figure a, or the Senate 2016 histogram, illustrates the mean-median difference histogram. As labeled by the orange line, the mean-median difference is approximately -0.015. Firstly, we can notice that the average percentage of votes received by one party is lower than the median percentage of votes received by this same party. This suggests that there is a skewed distribution of votes toward lower percentages in some districts, although very slight. This indicates there may be some districts where a party receives a low percentage of votes, in comparison to the overall average. This opens up the possibility for disparities in representation or various other factors.

Figure b is a histogram that represents the mean-median difference of the 2016 Presidential election. As labeled by the orange line, the mean-median difference is approximately 0.0026. This signifies an extremely minimal skew in the distribution of votes across the districts. The mean-median difference is close to 0 which suggests that the distribution of votes is relatively balanced. There is no distinct skew towards any party. Despite the slight skew, we may conclude that there are no substantial disparities in this specific election across districts.

6.4 Cut Edges

Generally speaking, cut edges refer to the edges of a graph that connect to adjacent points. In the context of redistricting, cut edges refer to the boundaries of a graph that connect two districts. They play a crucial role in defining the boundaries of districts, as the configuration determines the composition of each district. In turn, this can influence the representation of different communities and populations. One of the objectives during the redistricting process is to create districts in a way that makes them contiguous or ensures they are a single connected region. In addition to this, they attempt to hit criteria such as population equality and compactness per the Voting Rights Act. Checking the number of cut edges in redistricting plans provides insight into the fragmentation of districts. A higher amount of cut edges leads to higher concerns of representation, while few suggest more compact districts. More compact districts are believed to demonstrate more equitable representation.

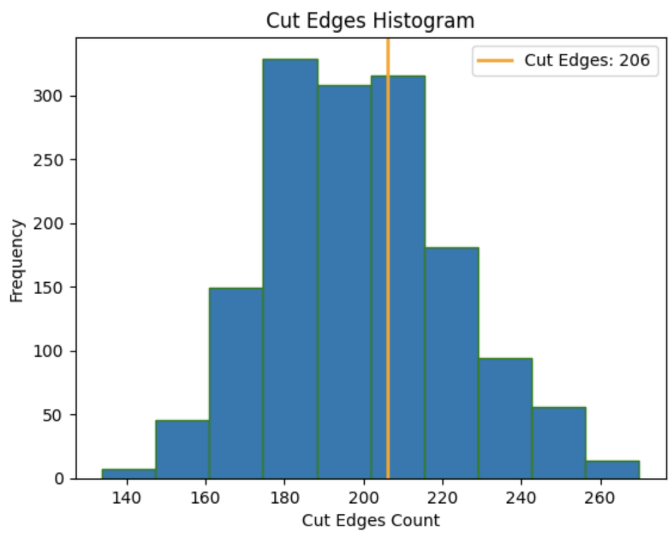


Figure 7: Cut Edges Histograms

The effects of the cut edges is dependent on various factors, such as population and the size of the state. Connecticut is a relatively small state with a high population density. The 2020 census lists Connecticut's population as 3.6 million people. In addition, Connecticut has diverse geographic features, such as urban and rural areas. Distinct communities and geographic boundaries could contribute to the cut edges in the redistricting process. Considering these factors, 206 cut edges may be considered relatively high for Connecticut. However, this is still not substantial evidence for apparent gerrymandering. When we take a look at the histogram, it falls under the normal distribution.

7 Results

The results of our analysis of the 2016 Connecticut Senate and Presidential elections have given us valuable insights into the integrity of electoral districting maps. This allowed us to shed light on potential instances of gerrymandering. By utilizing the Recom districting algorithm, we ensured equitable population distributions across the electoral districts. This enabled us to have a comprehensive examination of the plans. In addition to the Recom districting algorithm, we utilized Markov Chain Monte Carlo simulations to conduct outlier analysis. This allows us to generate histograms to assess the electoral landscape.

By utilizing Democratic and Republican vote percentages in both elections, we found consistent trends within Connecticut's districts. There was consistent favoritism towards Democratic candidates. Using marginal box plots, we were able to illustrate the distribution of those percentages and highlight stable Democratic support, along with noting Republican outliers. This emphasizes Connecticut's Democratic-leaning political landscape and allows us to ensure fairness in representation.

Metrics such as the Efficiency Gap, Mean-Median Difference, and Cut Edges allowed us to gain additional insights into potential gerrymandering. The efficiency gap indicated a relatively fair distribution of votes to seats. However, the mean-median differences suggest slight disparities in vote distribution. The number of Cut edges highlighted concerns regarding district fragmentation, but there were generally no significant signs of gerrymandering. These analyses underscore the complications of electoral districting, as well as the importance of ensuring equity in electoral processes.

References

[1] Proto, Jennifer. "Legal Challenges to Connecticut Redistricting Plans." Legal Challenges to Connecticut Redistricting Plans, 1 Dec. 2012,

www.cga.ct.gov/2021/rpt/pdf/2021-R-0012.pdf.

[2]Staff, A. C. (2023, July 17). Connecticut's population inched up 0.9 Percent last decade. Census.gov.

https://www.census.gov/library/stories/state-by-state/ connecticut-population-change-between-census-decade.html