Northwestern University

MSDS Assignment 3:

Integer Programming --- New Jersey Redistricting

Assignment Submitted to

Professor Thomas W. Miller MSDS 460 Fall 2023

Master of Science in Data Science

School of Professional Studies

by

Charles Lamb, Connor Cassedy, Heidi Huckabay, Susan Alrifai

Chicago, Illinois October 27, 2023

Abstract

This assignment's aim is to explore the task of redistricting based on constraints such as population balance and adjacency while considering compactness. Using integer programming as a methodology, redistricting the state of New Jersey provided results. The results are compared to other redistricting plans and their constraint considerations as well as an evaluation of each plan's alignment with the concept of "one-person-one-vote" (Miller, 2023). The limitations of integer programming are mentioned: total population is the only constraint that the model is able to explore. After comparing the new mapping to the current one, the group observes the new creation is "shockingly close" to the present state, bringing into question whether it is worthwhile to submit a new plan to the state legislature. The only external data sources used in the model creation is the 2020 U.S Census county population totals, which the group assumes to be accurate enough for relevant purposes.

Contents

Methodology

Using census data for the population of New Jersey, a data frame of county by population was created. A secondary data frame for counties geographically adjacent from one another in New Jersey was created. Both data sets were gathered from the U.S. census government website. Since this data is gathered through a survey and its response is mandated by law, this source is a more reliable place to gather information on United States demographics (census.gov, 2023).

Due to the time it would take to run a python integer program code for the entire state of New Jersey the state was split up into two different code sections, mainly separated as north and south.[1] The North region contains 52.5% of the population, while the south region contains 47.5% of the population. The model was initially run with a population variance of +/-10%, this variance did not allow for a feasible solution, the model for the northern portion was then run with a population variance of +/- 25%. The southern portion of the state resulted in a solution at a variance percentage of +/- 195% .[2]

A python integer program code available on github, https://github.com/cglamb/MSDS_460/tree/main/Hw3, establishes decision variables representing the assignment of counties to legislative districts and establishes these constraints:

Each county is only able to be assigned to one district
A maximum population constraint for each district
A minimum population constraint for each district
A cut edge for each pair of adjacent pair
An adjacency mandate

The objective function is constructed using lpSum, which calculates the sum of the adjacent pair variables. Each adjacent pair variable represents the binary decision of whether two counties are adjacent or not. Therefore, the objective function minimizes the lack of adjacency between

[1] See Appendix figure A.1
[2] See Appendix figure B.1

counties in the redistricting plan. In other words, it encourages the creation of districts where adjacent counties are assigned to the same district, thus promoting geographic compactness. A map was drawn using the code results of counties to districts. [3]

Noting that the southern portion of the state had a large population variance, some southern districts notably Cape and Cumberland counties had a population of 249,414 people compared to 1,099,092 for the largest (Burlington and Ocean). A tweak was made to accommodate the population imbalance in addressing the "one-person-one-vote", for a final scenario a refinement of the South Jersey model. This model shrinks the number of unique districts down to 4 and offers the largest two population districts a double vote (i.e. two representatives).

Results

The redistricting plan proposed for the state of New Jersey considers a minimum and maximum population constraint for each district as well as an adjacency mandate with cut edges for each adjacency pair. Due to the inability of running an integer programming code for the entire state, New Jersey was split into two sections (North and South). Each of the two regions attempted to find six compact legislative districts.

Within New Jersey there is a population imbalance resulting in 50% of the population existing within three adjacent counties in the Northern area of the region. Within southern Jersey, the population constraints of the largest districts (Burlington and Ocean) compared to the smallest (Cape and Cumberland) differed by roughly 750,000. This was taken into account when drawing the new congressional districts, causing the Salem and Gloucester and Atlantic districts to merge with Cape and Cumberland, causing five counties to make one district, as seen in figure B.2, that only accounts for 20% of the population. The final figure proposes 10 congressional

[3] See Appendix Figure C

districts as opposed to the current 12 districts New Jersey operates under.[4] To address the population imbalance, the district consisting of Mercer and Middlesex got two votes as well as the district of Monmouth and Ocean. Mercer has a small population country while being surrounded by larger populated countries. To offset this, merging with Middlesex and giving them two votes averages down the population of a large county like Middlesex, which on its own has more people than the average population per representative. When combining Monmouth and Ocean, the merging of counties does not make a drastic difference as there is an average of 640,000 people per representative. Leaving them as two stand-alone districts would not make much of a difference as they would stand at 643,000 and 637,000 individuals per representative. Combining districts averages this number out to bring up the latter district, but otherwise does not make much of a difference.

Of course, there are different ways to approach redistricting beyond an integer programming problem. The site https://districtr.org allows the user to customize districts for all 50 states in a way that is much more hands-on and interactive. The user can color districts as desired and observe many important parameters involving population and race. However, the group found this software difficult to work with since there is little guidance in construction, and the race and population factors are only able to be observed after districts are colored. If the website allowed users to upload their own maps for each state and users could interact with these creations, then perhaps the group would have found more utility from the program.

Perhaps the most difficult task for election planners is to create voting districts that reflect the common interests and values of particular regions. Embedded in this is a desire for fairness and equitability that leads towards equal representation in voting. But this is not always the case: government officials who construct districts almost always shun data-driven approaches towards

[4] See Appendix Figure C.1

redistricting in a way that enhances their own agendas. Throughout the 2010s and 2020s, these issues have reached the Supreme Court, and many constitutional factors have arisen in arguments from lawyers and political scholars (balllotpedia.org, 2023).

Evenwel v. Abbot (2016) ruled unanimously that a state can use total population as a factor for districting, but there is a continuous discussion as to how illegal immigrants, prisoners, and children should interact with redistribution. Overall, there is a hope that new computational techniques will allow a greater variety of demographic constraints to be considered in new districting plans. This is the obvious next step in creating plans that allow for equal representation when Election Day comes along.

At its current stage, there appears to be no benefit in submitting a new map to the New Jersey State Legislature since there is little obvious difference between what the group created and the map New Jersey currently uses. This will hold to be the case until more advanced techniques can incorporate other Census data like race, citizenship status, age, income level, prisoner status, etc… in developing a map. A singular population constraint provided the group a computationally intensive process, and it is almost certain that adding more constraints in an integer programming problem would not offer adequate results. New approaches beyond integer programming may allow for certain groups to be represented more properly, although factors like adjacency and compactness among districts will still need to be considered.

Bibliography

Miller, Thomas. “Homework Assignment 3: Integer Programming “Example---Algorithmic R…” 2023. https://canvas.northwestern.edu/courses/200718/assignments/1276268?module_item_id=276170

U.S. Census, “County population totals.” 2020. https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html

Ballot Pedia, “Timeline of redistricting cases heard by the Supreme Court of the United States.”2023 **https://ballotpedia.org/Timeline_of_redistricting_cases_heard_by_the_Supreme_Court_of_the_United_States**

Appendix A

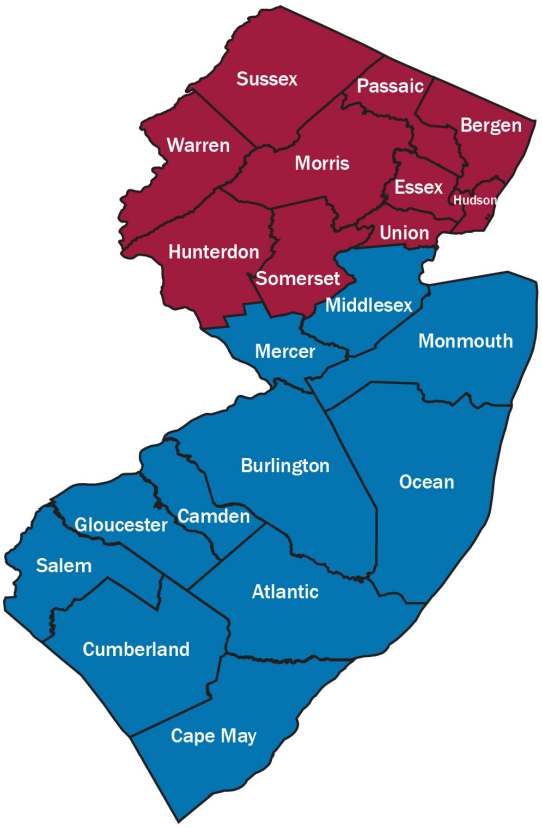

Figure A.1 NJ divided for code segments
the northern region contains 52.5% of the population
the southern region contains 47.5% of the population

Appendix B

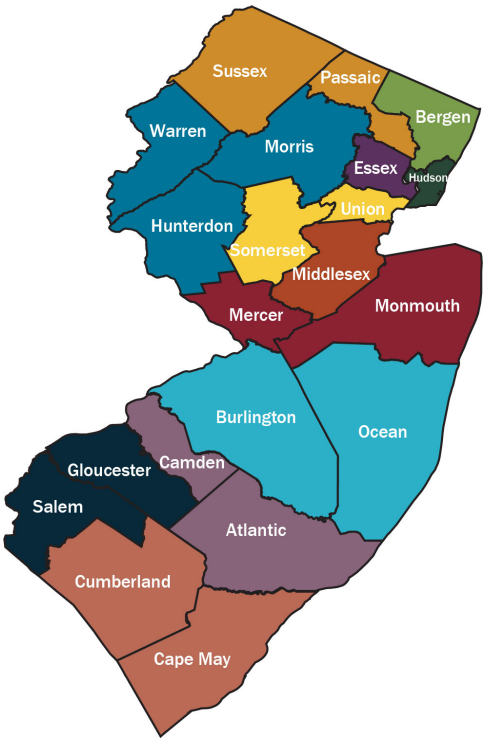

Figure B.1. Each color group has 1 representative, total reps 12

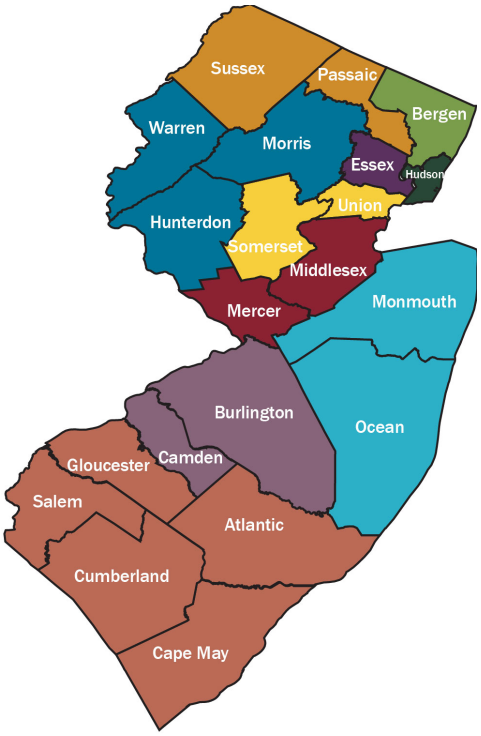

Figure B.2. Counties in groups of 10, Mercer and Middlesex, Monmouth and Ocean each have 1 representative, total reps 12

Appendix C

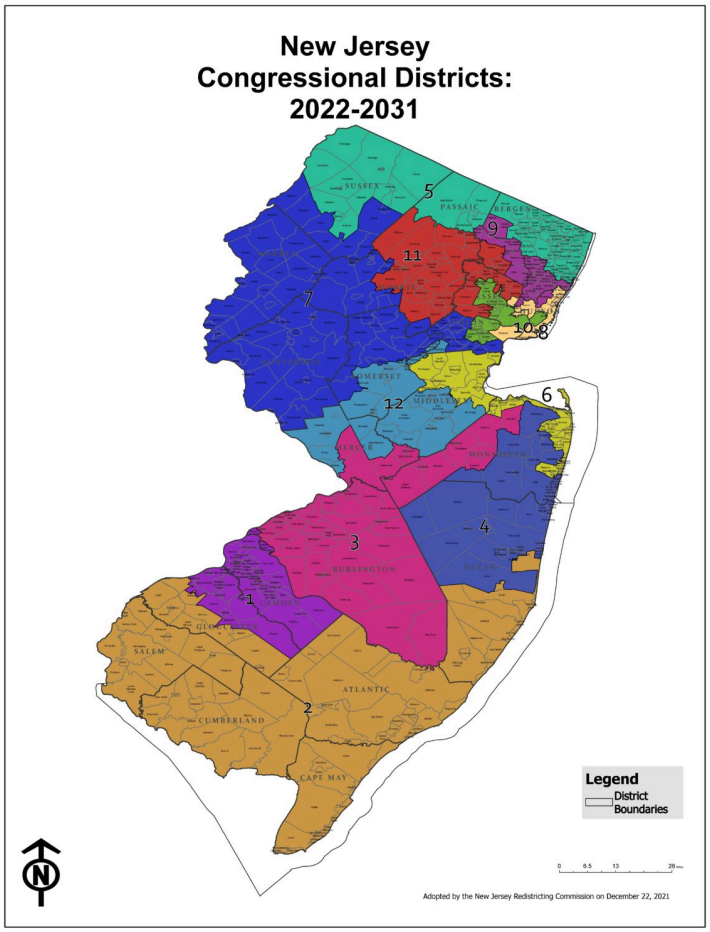

Figure C.1 Current New Jersey Congressional Districts