MSDS 460-57
Thomas Miller
May 6, 2022
Group 4

Homework Assignment 3:
Algorithmic Redistricting

Reed Ballesteros
Chris Lee
Lena Lu
Brett Mele
Zach Watson

Workgroup Four's goal was to create a redistricting plan for Tennessee that would encourage equal representation across races in congressional districts. We pulled our data from the 2020 Census and from the Federal Information Processing System (FIPS). Two sets of data: counties and adjacency data came from the https://www2.census.gov/geo/docs/reference/ website, and racial data by county was obtained from the data.census.gov website. We also obtained the FIPS codes to obtain the latitude and longitude of the counties (tn_fips.csv).

Evidence suggests that minorities continue to be underrepresented in the census, as outlined in a report from the U.S. Census Bureau. At a national level, latinos had a net undercount rate of 4.99% and blacks had a net undercount rate of 3.30%. Due to the use of census data for our congressional maps, our data also likely underrepresents minority populations. However, due to variability in our racial breakdown constraints, these undercounts are not likely to significantly impact our model.

Tennessee is allocated nine members in the House of Representatives for 2022. The total population of Tennessee is 6,910,840 which means an average population per district of 767,871 over nine districts. 71% of the population identifies as white-only, whereas 29% of the population identifies as non-white. We came up with three linear programming models to try to best solve the problem. One model prioritized adjacency, one prioritized racial balance, but our models for both failed in regards to the adjacency requirement. The third model tries to combine models 1 & 2. Our three models are explained below.

Model 1 minimizes the difference in the average population per district from the lower bound of the ideal district while also attempting to create compact districts within the population constraints. The initial plan was to iterate through and increase the number of adjacencies required for each county, but that became too much for PuLP to handle beyond if it increased

the requirements beyond three. In addition, we attempted to include constraints for the maximum allowable distance between two counties in a district (540km) but found it to be not much of a factor. Once the process gets below 70% of the maximum distance between counties in the state, the solver stops working. We defined our objective function for Model 1 with the following equation: $\sum_{j=1}^{7}\sum_{i=1}^{95} county\_pop_{ij} - lower\_lim$

The objective function includes only seven districts and acknowledges that Shelby (population 929744) and Davidson (population 715884) counties represent their own respective districts due to their large populations meeting or exceeding the population criteria below.

Model 1 is bound by the following criteria:

- Create a balanced distribution of populations by requiring each district to be within 10% of the ideal average population per district (between 691,000 and 844,000)
- A county may only be in one district
- Counties must be adjacent (starting with at least one adjacency)
- Maximum allowable distance between counties: 540km

Model 2 aims to distribute racial balance across districts by minimizing the difference between the proportion of white people per district and the proportion of white people at state-level. We defined our objective function with the following equation:

$$\sum_{i=1}^{7} \left|\frac{white\_pop_i}{total\_pop_i} - white\_pop\_prop\right|$$

Since PuLP does not intake division, the objective function is rewritten to:

$$\sum_{i=1}^{7} |white\_pop_i - total\_pop_i \times white\_pop\_prop|$$

Model 2 is bound by the following criteria:

- Create a balanced distribution of populations by requiring each district to be within 20% of the average population per district (614,000-921,000)
- Require adjacent counties to be in the same district; however, this resulted in pockets of adjacent counties rather than a true district adjacency
- A county may only be in one district
- Davidson County (Population: 929,744; 34% white) and Shelby County (Population 715,884; 54% white) have a much lower white proportion than the state level. They also have enough population to be their own districts. Consequently, they are districted into "double-districts" where the district will receive two representatives, but also has double the population of other districts. This way, one representative per almost 768,000 population average still balances out

Model 3 uses the same objective and constraints as Model 2. Model 3 uses a different adjacency constraint algorithm to try to reduce the number of "pockets" and make districts more centralized. This model ran for 13 hours and was still trying to find a solution. .

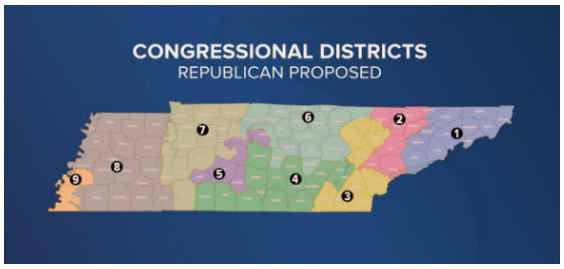
The optimal redistricting solution is Model 2. Although the districts are more disjointed, non-white people in Tennessee would receive more representation through the double-district models. The majority non-white population in Davidson and Shelby would be represented by four representatives compared to the two they had before. Shelby and Davidson county's non-white population makes up 46% of Tennessee's non-white population. By minimizing the districts' difference in white proportion (and subsequently non-white proportion) to the state's, we create a situation where each representative represents the state's demographic. Since the state is 70% white, the demographic of each representative's constituents are also about 70%.

Model 2: Breakdown by District

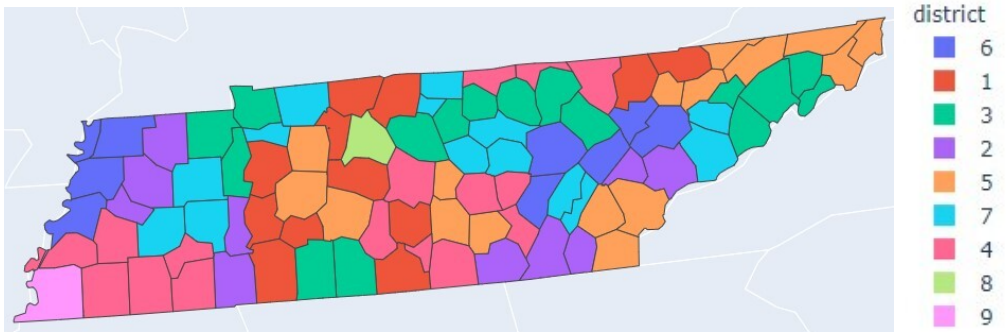
District	1 (Two Reps)	2 (Two Reps)	3	4	5	6	7

Total Pop	1,801,426	1,842,454	698,001	657,342	663,862	625,076	622,679
White Alone	1,278,148	1,145,122	548,090	533,271	506,100	444,476	445,039
White %	71	62	79	81	76	71	71

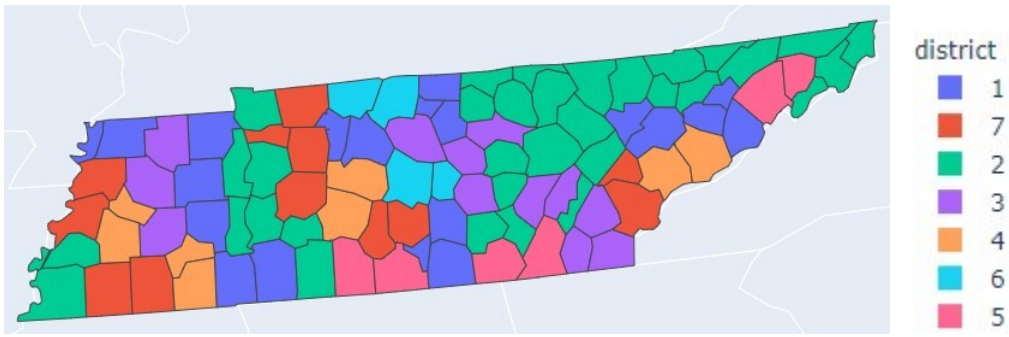
Government Proposed Map:



Model 1 Map:



Model 2 Map:



Model 3 Map: None (no solution)

We would recommend our map for model 2 as the most fair and equitable of the maps above. Although it does not meet the boundary constraints, our double-districting model allows for the most equal representation by race. In comparison, the racial makeup of the currently proposed Tennessee map leans more heavily white in all but the 9th district (southwest corner), which is majority black. In order to meet the boundary criteria in a future model, the adjacency constraints should be written in a way that two counties belonging to the same district must have a path that connects them without leaving the district boundaries. This would eliminate the pockets of counties that were dispersed throughout the map. Unfortunately that idea is beyond the scope of our abilities as it stands.

References:

National Census Coverage Estimates for People in the United States by ... https://www2.census.gov/programs-surveys/decennial/coverage-measurement/pes/national-census-coverage-estimates-by-demographic-characteristics.pdf.

Bureau, US Census. “Census.gov Homepage.” *Census.gov*, 6 May 2022, https://www.census.gov/en.html.

Horvat, Sabi. “How to Draw Congressional Districts in Python with Linear Programming.” *Medium*, Towards Data Science, 8 Jan. 2022, https://towardsdatascience.com/how-to-draw-congressional-districts-in-python-with-linear-programming-b1e33c80bc52.

Ryanabest. “What Redistricting Looks like in Every State - Tennessee.” *FiveThirtyEight*, 8 May 2022, https://projects.fivethirtyeight.com/redistricting-2022-maps/tennessee/.