## Introduction

This analysis uses Indiana's 2020 US Census data in an integer programming problem to identify an optimal solution for the distribution of counties into representative districts for the state. The optimal solution is the result of the following constraints: districts should have approximately the same population, districts should be made up of counties which are adjacent to one another, the population of the districts should be representative of the distribution of minorities throughout the state, and finally that the entirety of each county is associated with exactly one district.

## Data sources & Outliers

This project sourced data from the US Census Bureau's Census Data API ("*Available APIs*", 2024). Since this project is focused on the function of redistricting by the state governments, the data that those governments use is the same data used here. While there may be nominal inaccuracies in how the census describes the target populations, any inaccuracies that exist, the state would also be subject to them.

One significant conflating factor discovered early was the presence of Marion County within the dataset. This single county contains a population close to a million, significantly greater than the goal population of any single district. Not only that, but it also contains a disproportionate distribution of the minority population of Indiana, enough so that, without specific consideration of how to treat the data with and without the outlier county, other districts would likely be less representative of the state demographics than would be appropriate for the goal of this redistricting exercise.

## Specification

Four specifications were made to identify the optimal solution. The objective function chosen was to minimize the number of cut edges between districts, which prioritizes the geographic compactness provided the remaining constraints are also satisfied. This approach is drawn from Becker and Solomon (2020) and their study of redistricting algorithms, specifically drawing from their methods of integer programming. While their methods use census units, this paper uses a more simplistic approach with counties.

During testing of our redistricting code we had run into difficulties with the code completing within a reasonable time, so the PuLP solver was limited to a 10-minute run. The first pass lowered the number of districts that had to be modeled from 9 to 4, along with a larger tolerance in deviation from the desired average population of each district (from a 5% tolerance to a 50%), and finally we raised the gap tolerance of the solver itself to .02.

Table 1: Four District Result Population Distribution

| District | Population |
|---|---|
| 1 | 2432765 |
| 2 | 971737 |
| 3 | 2518863 |
| 4 | 861038 |

Each county was to be assigned to exactly one district; this specification ensured both that each county is included in a district and that no county was included multiple times in the solution. The next specification is that the resultant populations of each district must be within 10% of each other with the exception of the Marion County district.

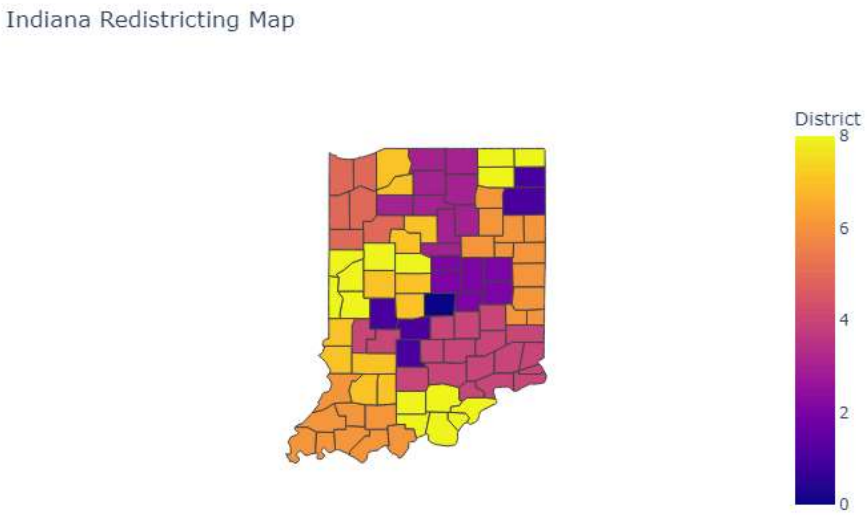
The final specification is to restrain the percentage of white population in each district to be no more than 70%. This provides a roughly 10% margin from the demographic makeup of the state as a whole, which is about 61% white (America Counts Staff 2021). The demographics of each of a state's voting districts are expected to be approximately similar to the demographics of the overall state. This specification is difficult to define mathematically for two reasons; first, there are no definitive legal definitions for how much deviation from the state's demographics is acceptable for a given set of districts. Secondly, because Marion County makes up slightly more than 1/7th of the population of the state and close to half of the minority demographic population, finding a solution in which the remaining districts were approximately representative of the broader state's demographics is more of a judicial calculation than a mathematical one.

## Discussion

The optimal solution struggles to completely adhere to the given constraints. The requirement that prevents districts from cutting into counties is a clear limitation on the ability to have clean borders between them. However, the approach to set Marion County aside and run the integer programming on the rest of Indiana does draw some contiguous counties within the constraints. The first six districts have a neat grouping of counties, but the remaining three are spread out across the state. It makes sense that the algorithm is seeking balance between the adjacency constraints and the population constraints.

This solution failed to find districts with a white population 70% or less in all but one district, and the population metrics show a range of roughly 300,000 between the most populous and least populous districts. This solution could be improved by using a more granular geographic boundary, such as census tracts, which would necessitate a much longer run time to find the optimal solution.

Figure 1: Algorithmic Redistricting Solution for Indiana




Algorithmic Redistricting Solution for Indiana: Produced using data from Plotly. Districts are zero indexed, where Marion County represents district 0.

Table 2: Nine District Solution with population and race constraints

| District | Population | Pct. White |
|---|---|---|
| 0 | 977,203 | 52.5 |
| 1 | 677532 | 81.0 |
| 2 | 736639 | 86.1 |
| 3 | 789233 | 81.5 |
| 4 | 690495 | 91.3 |
| 5 | 751250 | 68.5 |
| 6 | 788854 | 90.0 |
| 7 | 679878 | 86.3 |
| 8 | 698785 | 87.2 |

When it comes to actually submitting a redistricting plan, it's important to consider how litigious a process redistricting frequently is. Whether or not a plan is approved seems to be based more on the political leanings of the legislative bodies confirming the plan rather than any claims to objective fairness. The fact that this plan is based on theoretically objective mathematics rather than deliberate gerrymandering may not only not be persuasive, but in fact actually be a mark against the viability of this plan when it comes to whether or not it could be approved. Before this redistricting plan could be submitted, a significant political campaign would need to be implemented familiarizing the plan with the various decision makers of the process to get their own input and buy-in for the plan, and each of those decision makers could request changes to the specifics of the methodologies.

# References

America Counts Staff. “Indiana’s Population Rose 4.7% Since 2010.” U.S. Census Bureau, August 25, 2021. https://www.census.gov/library/stories/state-by-state/indiana-population-change-between-census-decade.html.

“Available APIs”, U.S. Census Bureau. Last Revised September 18, 2024. https://www.census.gov/data/developers/data-sets.html.

Becker, Amariah, and Justin Solomon. “Redistricting Algorithms. ” Arxiv, November 18, 2020. https://arxiv.org/abs/2011.09504.

United States Census Bureau, County Adjacency Data. June 6, 2015 . https://www2.census.gov/geo/docs/reference/county_adjacency.txt