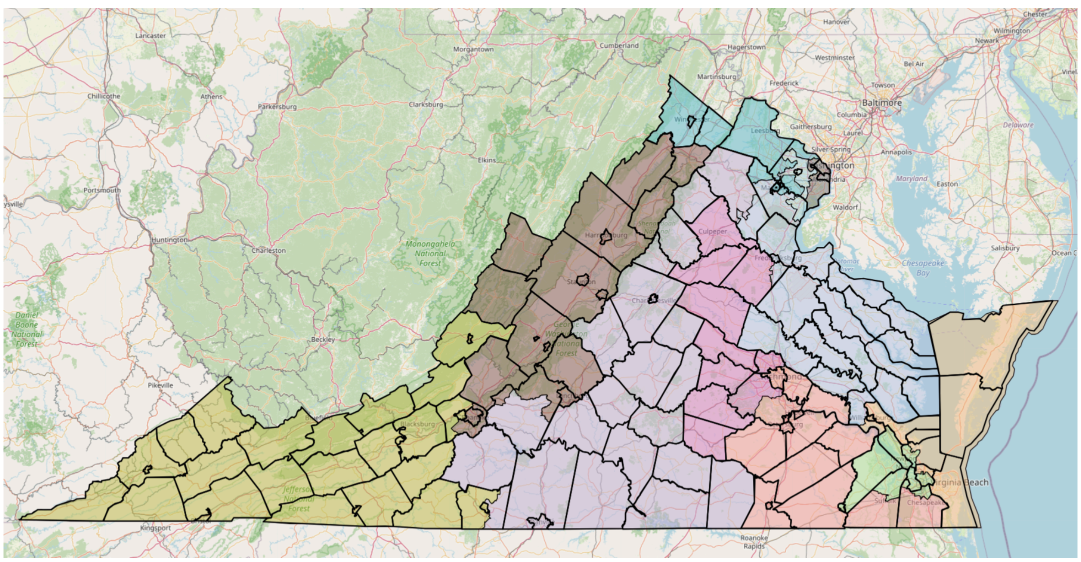
MSDS 460 Assignment 3: Algorithmic Redistricting through Integer Programming

Group Members
- Albert Lee
- Alberto Olea
- Maddie Stapulionis
- Thomas Young
- Migus Wong

Introduction

This project focused on developing an integer programming approach to redesign Virginia’s legislative districts. The goal was to craft an algorithmic solution that ensured a fair and equitable Virginia voter population distribution while maintaining geographic adjacency. By integrating demographic data with county adjacency, the model aimed to balance the “one person, one vote” principle with practical constraints. Our work demonstrates how optimizing for fairness can address complex real-world challenges with electoral representation. Below is a visual representation of how the countries/ independent cities of Virginia are split into districts. The reason our group decided to pick Virginia is because it’s a good example of a “Purple State” and is oftentimes a bellwether state for future elections. Political pollsters and the media tend to hyperfocus on local Virginia results because it may indicate how national election results may play out. Virginia is also unique in that its gubernatorial race does not allow for consecutive term limits so there’s never an incumbent governor serving two consecutive terms. In addition to the uniqueness of Virginia’s election, it’s also a Purple State. Virginia’s governor is a Republican but the state legislature is controlled by Democrats.



Methods

Methodologically we began with data collection and preprocessing. Population data for Virginia’s counties and independent cities was sourced from public census records, while adjacency information between counties was obtained from the U.S. Census Bureau. Preprocessing ensured consistency by reconciling mismatched entries between population and adjacency datasets and preparing the data for integration into the optimization framework.

The optimization model was built in Python using the PuLP library, With the primary goal of assigning counties to one of eleven legislative districts while minimizing deviations from a target population. We introduced binary decision variables to represent district assignments, whereas the object function sought to reduce absolute differences between district populations.

Visualization played an important role- utilizing GeoPandas and Folium to map district assignments and population distributions. The visuals supported the interpretation of results, thus offering a spatial perspective on the effectiveness of the proposed model and its adherence to the constraints. Integrating computational tools and visualization techniques highlighted the potential of integer programming to inform equitable redistricting.

Results

The optimization model successfully distributed Virginia's population of 8,715,698 across eleven districts, achieving a target population of 792,336 per district, though Fairfax County's population exceeded the target by over 40%. The model effectively balanced populations across the remaining districts despite this challenge. By minimizing the deviations from the target population the model upheld the 1P1V principle and ensured each of Virginia's 133 counties and independent cities was assigned to a district.

Reviewing specific district outcomes, District 2, containing Fairfax County, had the highest population at 1,141,878 residents. Other districts demonstrated more balanced populations: District 1 with 784,629 residents, district 3 with 777.058 residents, and District 4 with 789,654 residents. District 8 ..

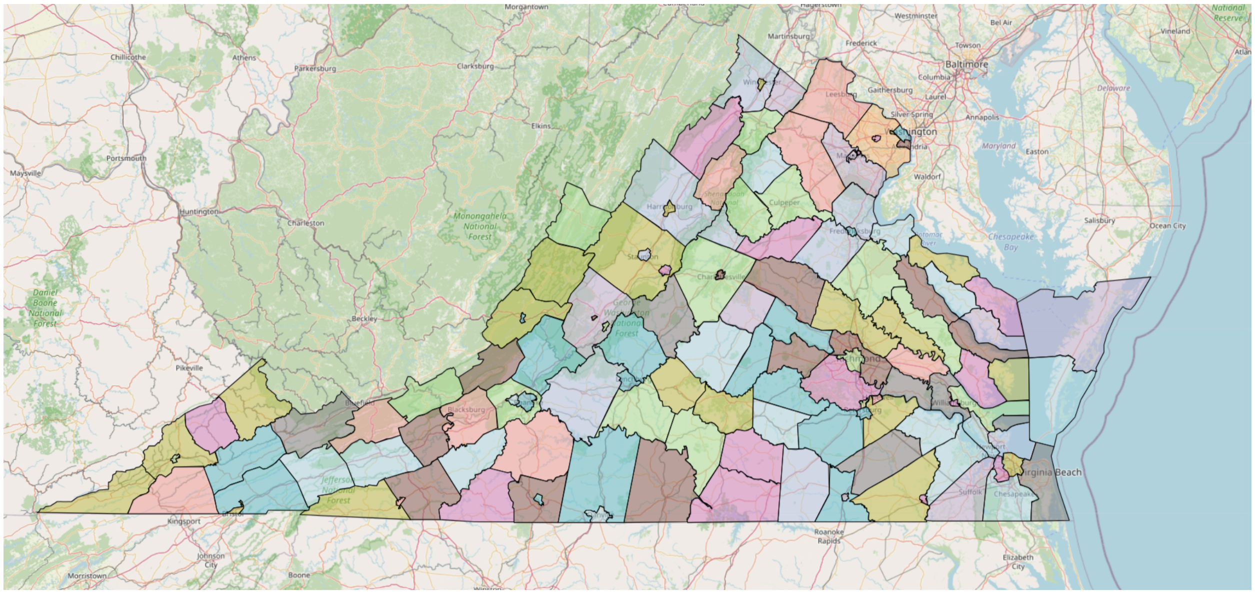
Assignments adhered to geographic closeness by incorporating adjacent constraints. Accomack County was placed in District 5, which achieved a population of 743,379 residents, maintaining natural groupings. These assignments were guided by the network flow model that ensured district proximity. Each district formed a connected component where every county could reach every other county within the same district through an adjacent neighbor.

Optimization achieved an optimal status, minimizing population imbalances while respecting contiguity constraints. Visualizations created via GeoPandas and Folium provided clarity into the district boundaries and population distributions. An example would be Fairfax County's large population required careful balancing; other districts combined urban and rural counties to achieve equity. Such visuals offered a spatial perspective, thus supporting the validation of the model's results and confirming the geographic coherence of district assignments.

The redistricting plan model could benefit from adding additional variables and constraints. Constraints like number of available polling places, yearly road closures, and There are gaps in county level data that are not easy to address with current online resources.

Base Model Setup (Optimizing for population Variance)

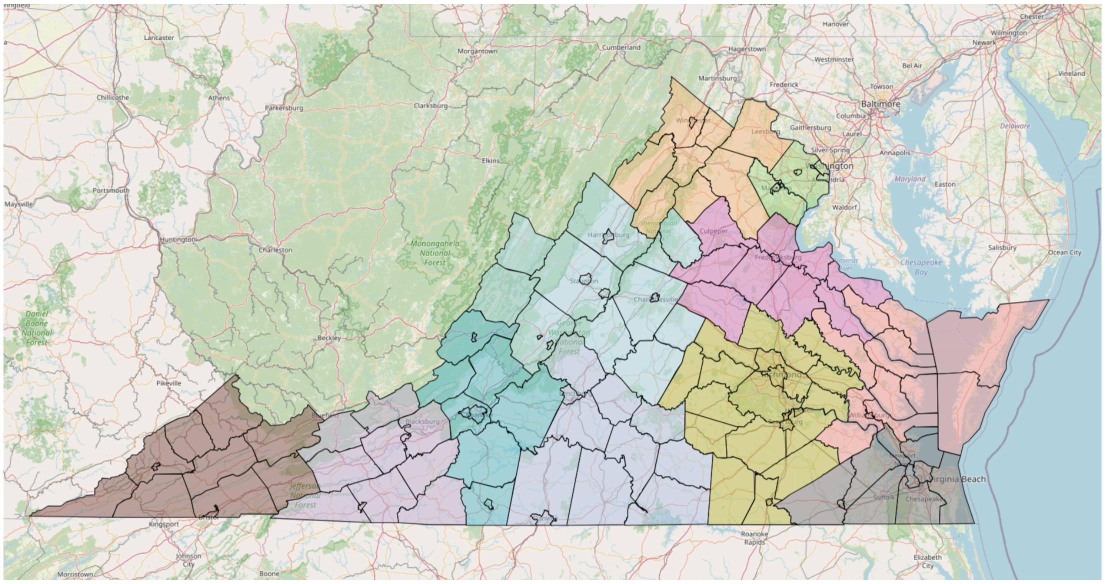
The first model developed for these assignments was an integer linear programming model optimizing for population variance with the goal being to make district populations as equal as possible. The resulting map shows the distribution of districts across Virginia. Variables were mainly set up to represent a specific county/ independent city being assigned to a specific district with a total 11 possible districts.



While this map would ensure that votes-per-district were as close as possible, there are several issues with this proposal, the main one being the lack of adjacency between districts. This was addressed in the next developed model which added constraints for contiguity via the concept of flow.

Adding Contiguity and Moment-of-Inertia

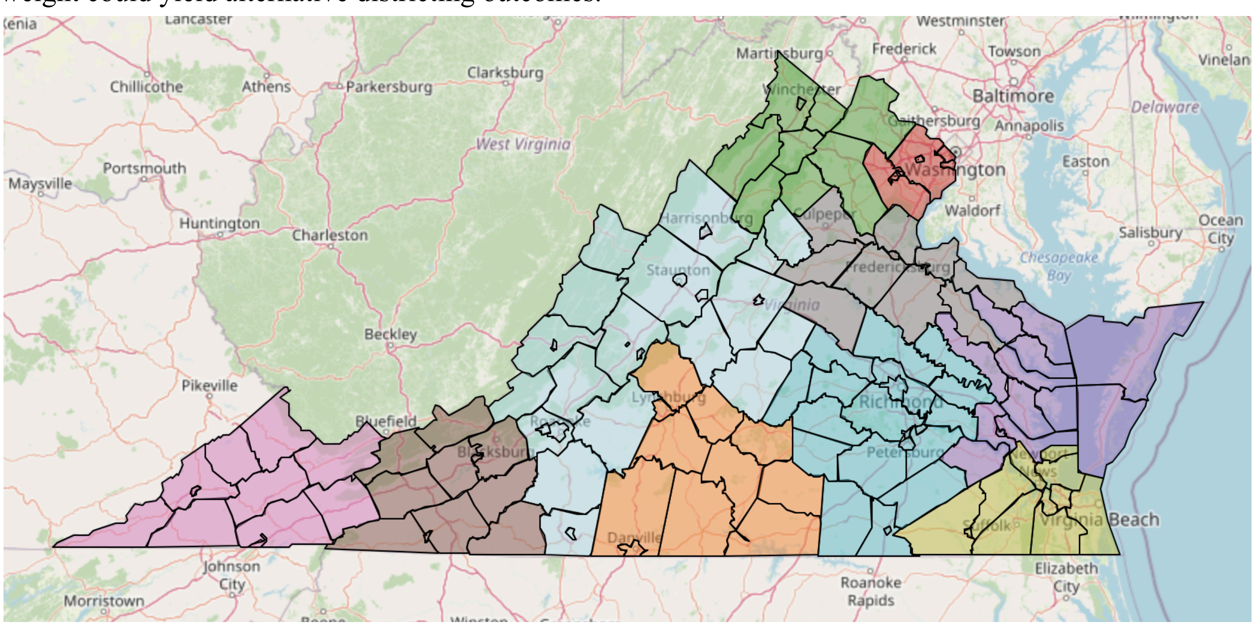
In order to achieve contiguity, a new objective was selected along with additional constraints. This new objective was based off Hess' Integer Programming solution [Validi, 2022] and proposes that the new objective equation should be the minimization of the "moment of inertia" which, in this context, is the sum of distances between the a county and its "district county anchor" squared times the population of the county. Distance calculation was conducted by determining the centroid of each county. Variables were also set up slightly differently in this model compared to the "base" model; instead of variables representing a county in a district, a different approach was employed where each county would be assigned to an "anchor county" which represents the best county to minimize a district's distance objective. Additionally, more constraints were added to the problem which define the concept of "flow" helps set up the integer program as a network model. Although contiguity is not necessarily directly enforced by the objective or constraints, it appears that this MOI based approach is enough to produce a set of contiguous districts across Virginia.



While results are now contiguous, the population variance across the districts is not nearly as fair meaning some votes in some districts may be theoretically weaker than others.

Adding Election Data

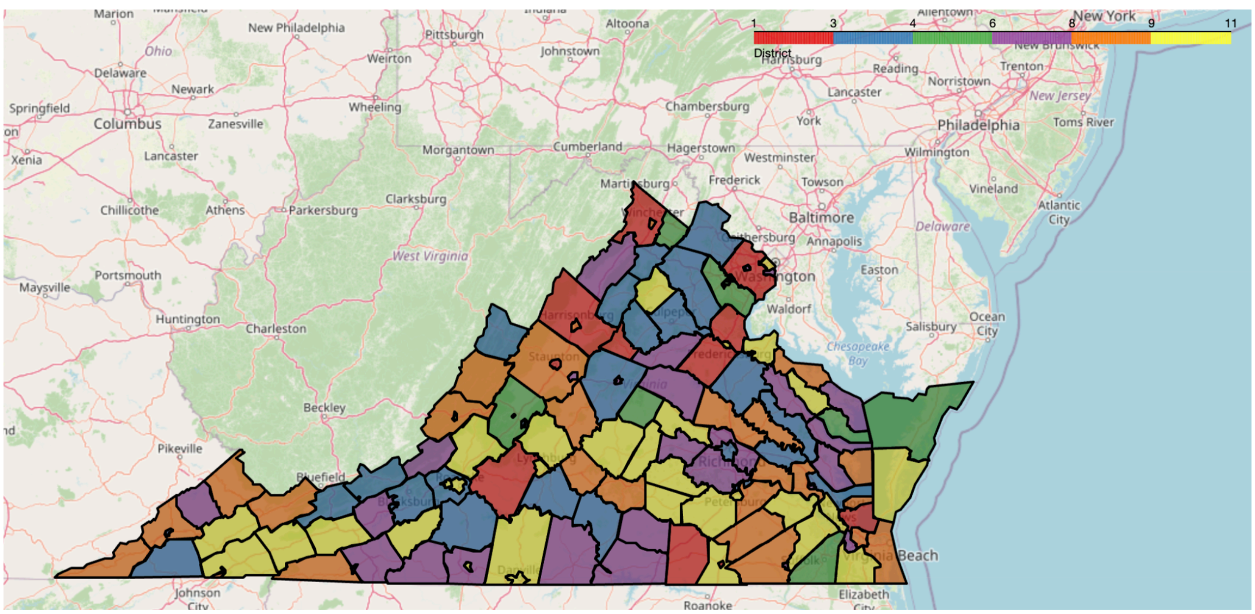
To enhance fairness in congressional district assignments, a model was developed that incorporates vote share data from the 2020 general presidential election. Political proportions were used to balance the distribution of major parties within districts, ensuring that no single district had an extreme partisan skew. The objective function minimizes partisan imbalance by reducing the difference between Democratic and Republican vote shares while also considering geographic compactness. Additionally, a political fairness weight was introduced to control the trade-off between political balance and district adjacency. While the final district assignments remained unchanged from the congruity model, adjusting the political fairness weight could yield alternative districting outcomes.



Adding Home Ownership Percentages by County

To enhance fairness in congressional district assignments, a model was developed that incorporates Virginia home ownership rates by county. These rates of homeownership have been shown to have an impact on voter behavior and could help reduce partisan skew in districts. The objective function addresses partisan imbalance by minimizing the variance between districts and the state average ensuring that districts home ownership rates reflect the state as much as possible. This model's contiguity was not completed in time for submission but the model is optimal and included for the sake of future contiguity completion. Contiguity will be important in reducing the difference between Democratic and Republican vote shares at district borders while also considering geographic compactness.

```
Assigned counties per district:
District 0: []
District 1: ['Albemarle County', 'Amelia County', 'Augusta County', 'Bath County', 'Bland County', 'Botetourt County', 'Buchanan County', 'Carroll County', 'Cha
District 2: ['Lancaster County', 'Mecklenburg County', 'Nottoway County', 'Powhatan County', 'York County', 'Galax City']
District 3: ['Amherst County', 'Arlington County', 'Bedford County', 'Greene County', 'Chesapeake City', 'Fairfax City']
District 4: ['Accomack County', 'Isle of Wight County', 'Poquoson City', 'Williamsburg City']
District 5: ['Alleghany County', 'Appomattox County', 'Buckingham County', 'Campbell County', 'Caroline County', 'Craig County', 'Culpeper County', 'Essex Count
District 6: ['Goochland County', 'Montgomery County', 'Danville City']
District 7: ['Halifax County', 'Sussex County']
District 8: ['Hanover County', 'Prince Edward County']
District 9: ['Greensville County', 'Northampton County', 'Orange County', 'Prince George County']
District 10: ['Brunswick County', 'Shenandoah County']
District 0 - Average Homeownership Rate: 0.00%
District 1 - Average Homeownership Rate: 78.75%
District 2 - Average Homeownership Rate: 77.55%
District 3 - Average Homeownership Rate: 74.30%
District 4 - Average Homeownership Rate: 76.95%
District 5 - Average Homeownership Rate: 76.77%
District 6 - Average Homeownership Rate: 73.68%
District 7 - Average Homeownership Rate: 69.93%
District 8 - Average Homeownership Rate: 74.26%
District 9 - Average Homeownership Rate: 69.60%
District 10 - Average Homeownership Rate: 71.31%
```



Adding Gender and Age Demographics Enhancements to Base Model

A consideration for encouraging fair and equal representation across the districts, we decided to add Gender and Age demographics data from the US Census bureau. The demographics were broken up into a variety of factors ranging from total female, total male, and various age groups by gender along with totals. For simplicity, we went with male, female, total population under 5, and total population over 85.

What also proved to be difficult in this dataset is the number of years. Again, for simplicity, we just summed all the years' population as a totality rather than picking a single year to try and perform the analysis on.

The first step in the Gender and Age demographic enhancement was to merge the two dataframes together (VirginiaCounty) and demographic. Because of potential variations in the County Names, we decided to use the GEOID instead. In the initial model, our constraints were too tight that the linear problem solution was infeasible so we went ahead and loosened the constraints. The primary change was by updating the allowed deviation from 10% to 15% which differs from our base model. We're also adding additional constraints around age and gender demographics so we were not too concerned around changing the deviation. After performing data checks, the model was ran and we were able to generate an optimal linear programming model.

Solution

While it would be nice for all of us to get recognition beyond this class for submitting an optimal redistricting solution for the state of Virginia, I think we would need more time and to tweak the model more. Where we ran into issues was performing contiguity for the various additional models that we were tweaking to the base model. If contiguity wasn't a major factor, I believe we would have a better starting point. As mail-in ballots become more popular ever since COVID, do districts really need to be adjacent? If we had more time, that would be another area our team would look into. Another factor that Virginia closely identifies is the amount of transient military members. Depending on how long they're stationed at the bases in Virginia, they may decide to become residents vs remain a non-resident. Those factors have a sway in how the districts are formed since Virginia is home to some of the most well known military installations - Pentagon, Fort Belvoir, Norfolk Naval Base, and Quantico.

References

Validi,Hamidreza, Austin Buchanan, and Eugene Lykhovyd. 2022, March–April. "Imposing Contiguity Constraints in Political Districting Models." Operations Research, 70(2): 867-892. GitHub C++ code repository at https://github.com/zhelih/districtingLinks to an external site. . Online abstract and link to paper at http://www.optimization-online.org/DB_HTML/2020/01/7582.html

Pew Research Center. "Partisanship by Family Income, Home Ownership, Union Membership, and Veteran Status." *Pew Research Center*, April 9, 2024. https://www.pewresearch.org/politics/2024/04/09/partisanship-by-family-income-home-ownership-union-membership-and-veteran-status/.

Federal Reserve Bank of St. Louis. "Release Tables: 172256." *Federal Reserve Economic Data (FRED)*. Last modified February 2025. https://fred.stlouisfed.org/release/tables?eid=172256&rid=406.