ALGORITHMIC REDISTRICTING FOR MASSACHUSSETS

Josh Muckensturm, Scott Serpa, Shawn Sie, Timothy Chiu

I. Data Sources

Our team worked on the redistricting plan for the state of Massachusetts, which entailed assigning one of their nine U.S. representatives across 14 counties, while achieving racial balance and ensuring approximately the same population across the districts. For our data sources, we used the Massachusetts Census Bureau to find the total state population, each county's respective population, and to find statistics about each county's population broken down by race. The exact sources can be found in the Appendix under *Data Sources*.

Although economic census data can have sampling and non-sampling errors that may affect the accuracy of the data, we thought the census data from the government bureaus that we utilized gave us the best chance to come up with an equal and fair redistricting model. Therefore, we do not have too much concern with the sources we used. Of course, having the most up-to-date data is ideal, but given that we used the most recent data the census bureaus provided (2020), that concern was alleviated.

II. Goal Program and Algebraic Formulation

Our methodology included the use of linear goal programming to achieve the best redistricting plan for our state. With goal programming, we can suggest a target goal for our objective through the implementation of a constraint that included both a positive and negative slack variable. The use of these slack variables allows the program to go above and below our suggested target value, or right-hand-side. In the objective, however, we aim to minimize both slack variables to achieve as little deviation from the target value as possible. In our program, the target value was the amount of racial equality to attempt to achieve in all nine districts. Below is the algebraic formulation for our goal program.

Indices:
$i \in \{1,2,\ldots,9\}$ Representative/congressional district
$j \in \{1,2,\ldots,14\}$ County index number

Constant Variables:
P_j = current population percentage of state for county *j* (2020 US Census)
W_j = percentage of population identified as white for county *j*
$B_{j,d}$ = immediate adjacency matrix for county *j* and county *d*
$B'_{j,d}$ = relaxed (clustered) adjacency matrix for county *j* and county *d*

Decision Variables:
$X_{i,j} \in \{0,1\}$ where 1 if representative *i* assigned to county *j*
$A_{i,j}$ = population to allocate to district *i* from county *j*
UR_i = negative slack for racial equality for representative/district *i*
ER_i = positive slack for racial equality for representative/district *i*

Constraints:

(1) $\sum_1^j X_{i,j} \geq 1 \quad \forall i$ ensure each district has at least one county
(2) $\sum_1^i X_{i,j} \geq 1 \quad \forall j$ ensure each county assigned to at least one district
(3) $\sum_1^i A_{i,j} = P_j \quad \forall j$ ensure each county's total population is allocated
(4) $A_{i,j} \leq M * X_{i,j} \quad \forall i \forall j$ binary switch to control upper bound allocation of population
(5) $A_{i,j} \geq X_{i,j} \quad \forall i \forall j$ binary switch to control lower bound allocation of population
(6) $X_{i,j} \leq \sum_1^{d:d \neq j}(B_{j,d} * X_{i,j}) \quad \forall i \forall j$ ensure clustered counties together in district assignments
(7) $X_{i,j} + X_{i,d} \leq B'_{j,d} + 1 \quad \forall i \forall j \forall d$ restrict distant counties from getting assigned together
(8) $669{,}035 \leq \sum_1^j A_{i,j} \geq 905{,}1066 \quad \forall i$ upper and lower bounds on district population sizes
(9) $\sum_1^j (W_j * A_{i,j}) + UR_i - ER_i = 425{,}000 \quad \forall i$ goal: minimize racial inequalities among districts
(10) $X_{i,j} \in \{0,1\}$ decision variable constraints
$A_{i,j} \geq 0$
$UR_i, ER_i \geq 0$

Objective Function:
Z = MINIMIZE [$100 \sum_1^i \sum_1^j X_{i,j} + \sum_1^i (UR_i + ER_i)$]

The program includes data for each county to include total population, racial proportions (white vs. non-white), and adjacency connections to allow us to build a network of the state. Our decision variables include the use of binary assignment variables, continuous population allocation variables, and continuous slack variables. Constraints 1 and 2 force at least one county in each district and at least one district for each county, respectively. Constraint 3 forces the entirety of a county's total population to be completely allocated among the districts it was assigned. Constraints 4 and 5 are binary switches to only allow population allocation if the county is assigned to the selected district. Constraints 6 and 7 attempt to enforce continuity of counties assigned to a district using relaxed adjacency matrices. *B* is the strict adjacency matrix to ensure adjacent counties assigned together. *B'* is a relaxed matrix where we loosen immediate adjacency requirements and instead cluster counties by region and population to allow more flexibility in assignment but prevent distant counties from being assigned together. Constraint 8 enforces minimum and maximum district population requirements. The minimum and maximum values were created based on the total state's population divided by nine districts (787,101) with a $\pm 15\%$ margin for flexibility. Constraint 9 is our goal constraint to achieve racial equality. Our target value was the weighted average across the counties (71.2% white) in each district. Constraint 10 enforces binary and non-negative constraints on our decision variables. The objective function seeks to minimize the racial inequalities while minimizing the number of split counties.

III. Optimal Redistricting Solution and Maps

We came up with two optimal redistricting solutions for the state of Massachusetts: one that redistricted by only total population and one that redistricted by total population and racial equality. The visualization comparisons among both solutions and the current actual district plan are provided in the Appendix under *Map Comparisons.*

Between our optimal solutions and the state's current map, the map we would recommend that is most "fair and equitable" is our map that redistricts based on total population and racial equality. Although our solution does not actually provide the geographical splits of each county, our optimal solution hits on our goals of being able to keep total populations adequately consistent across counties, while reaching our racial balance objective. Of course, there are always improvements that can be made especially with such an imperfect process, but our optimal solution at the very least meets the principle we sought out to solve for—citizens should have an equal representation in voting.

IV. Discussion on Potential Improvements

Our goal program was able to achieve a relatively fair and equitable plan based on only two values: population and racial equality. One could factor in more variables that might attribute to what constitutes fair and equitable like economic status. Another possible improvement would be going down to the city/town level for more fidelity of the results. A brief look into this strategy showed a lack of reliable data sources required.

A final consideration would be the continuity constraints. Our district continuity constraints do not guarantee continuity but seek to limit the discontinuity of the results. Our team did attempt other solutions before this program's relaxation. Our initial strategy was to utilize another decision variable, $Y_{i,j,d}$ which would track the use of the edges in our network. This variable was binary where it would be 1 if edge *j-d* was used in district *i*. We enforced constraints to make $Y_{i,j,d}$ equal 1 only when $X_{i,j}$ and $X_{i,d}$ were both 1 meaning both those nodes were assigned to the same district. It would take on a value of 0 otherwise. We then used a heuristic constraint where the sum of the number of unique edges in each district must be less than or equal to three times the number of counties in the district plus two. This metric seemed to work well on smaller toy problems but still did not force continuity in some cases. For reference, see the *Heuristic Examples* in the Appendix.

We ultimately decided to relax the hard continuity constraint and focus more on preventing discontinuity than ensuring it. The hard continuity constraint seems to be challenging and requires an internal searching algorithm to ensure each district's selected subgraph is connected by at least one adjacent edge. This would require going out multiple degrees which would be hard to formulate in a linear program.

APPENDIX

Data Sources:

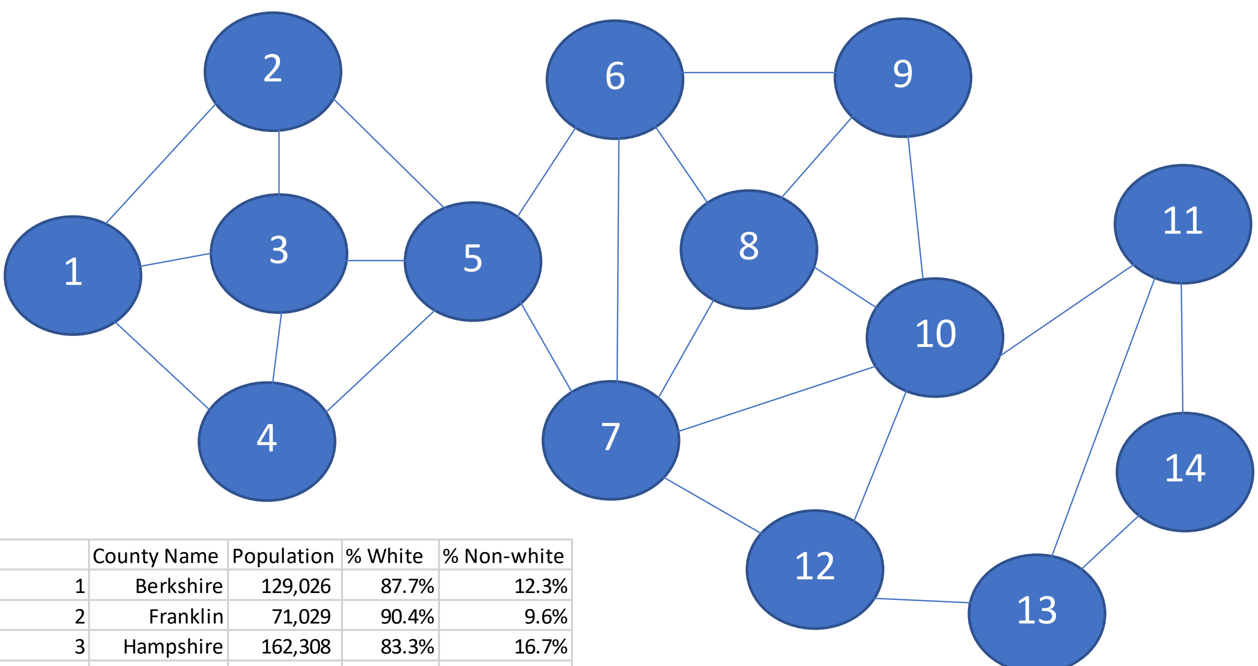
Massachusetts county population and racial splits: https://www.census.gov/quickfacts/MA

Massachusetts population by city/town:

https://malegislature.gov/Redistricting/MassachusettsCensusData/CityTown

Massachusetts current congressional district map: https://www.sec.state.ma.us/cis/cispdf/MA-Congressional-Map-2019.pdf

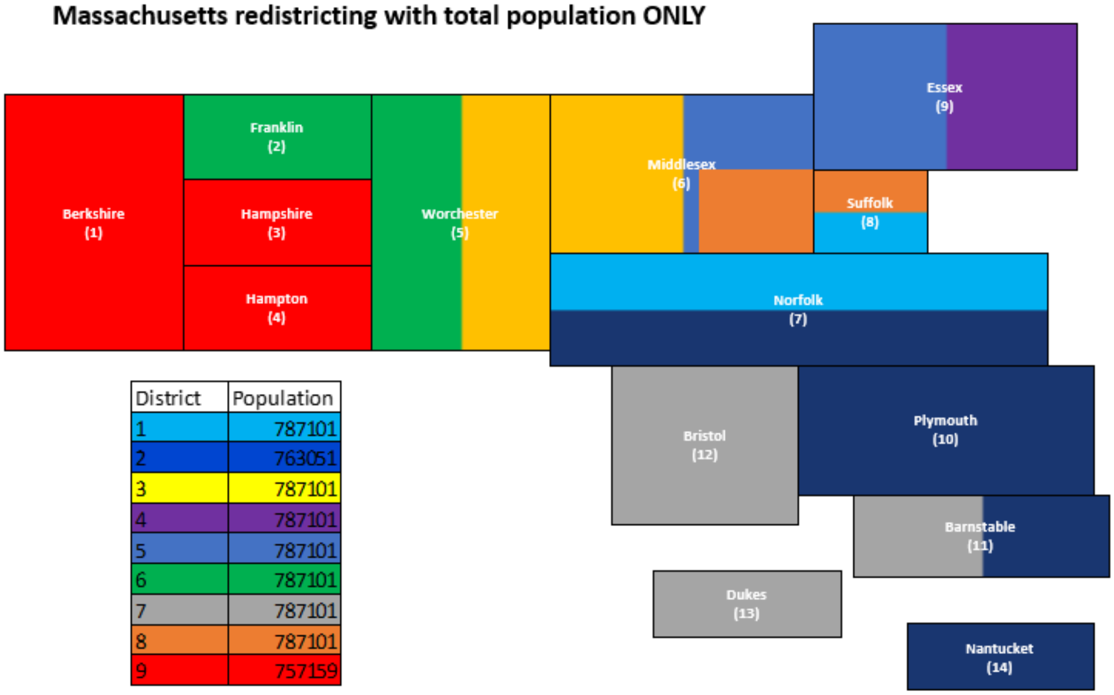
State Network Representation and Population Statistics:



	County Name	Population	% White	% Non-white
1	Berkshire	129,026	87.7%	12.3%
2	Franklin	71,029	90.4%	9.6%
3	Hampshire	162,308	83.3%	16.7%
4	Hampden	465,825	61.4%	38.6%
5	Worcester	862,111	75.7%	24.3%
6	Middlesex	1,632,002	71.2%	28.8%
7	Norfolk	725,981	73.8%	26.2%
8	Suffolk	797,936	45.2%	54.8%
9	Essex	809,829	69.0%	31.0%
10	Plymouth	530,819	81.1%	18.9%
11	Barnstable	228,996	89.5%	10.5%
12	Bristol	579,200	81.8%	18.2%
13	Dukes	20,600	86.9%	13.1%
14	Nantucket	14,255	71.1%	28.9%

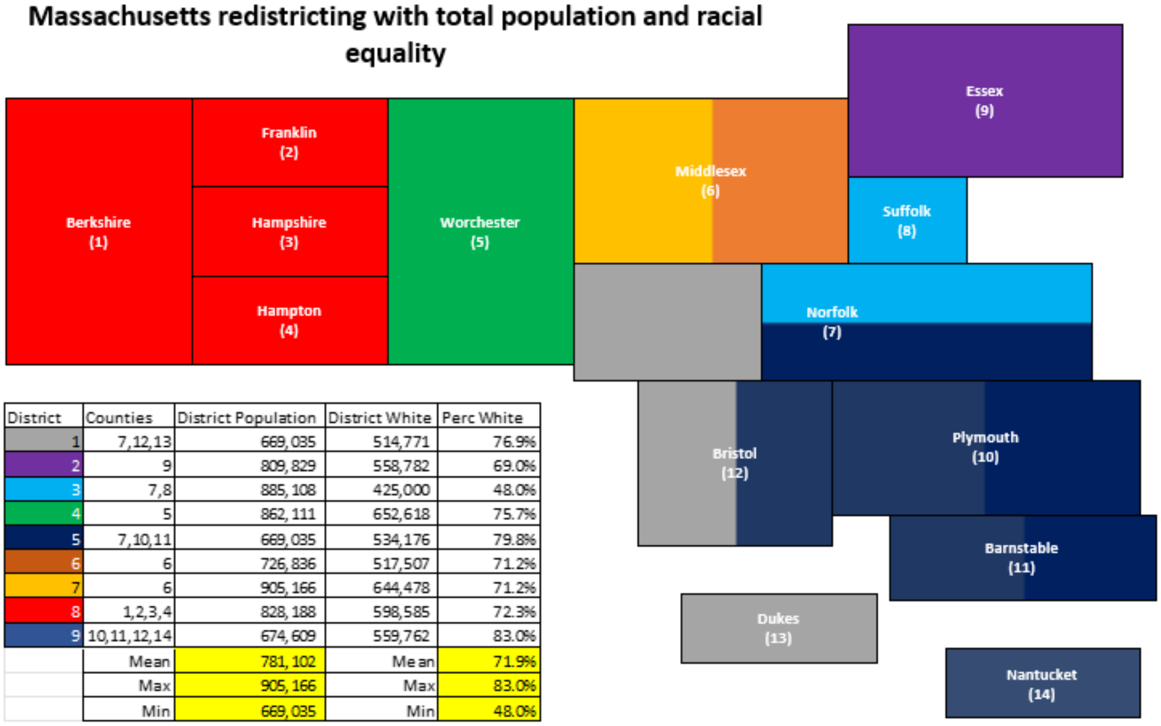
Map Comparisons:

Massachusetts redistricting with total population ONLY

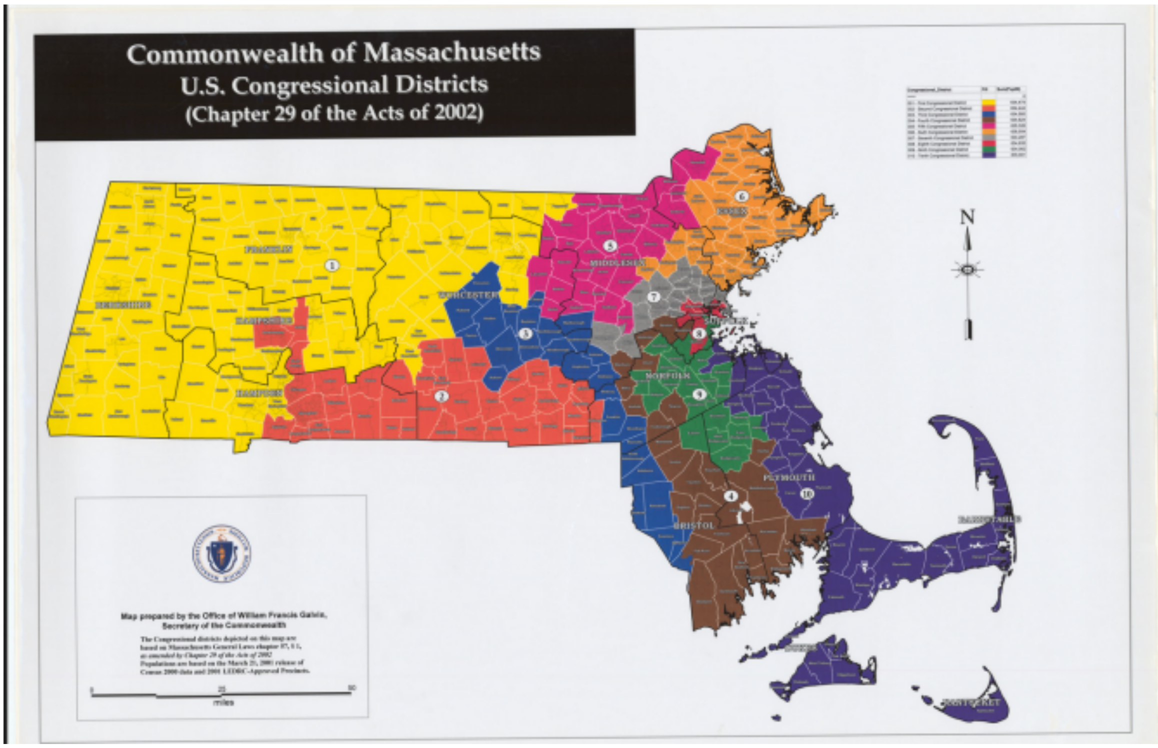


District	Population
1	787101
2	763051
3	787101
4	787101
5	787101
6	787101
7	787101
8	787101
9	757159

Massachusetts redistricting with total population and racial equality

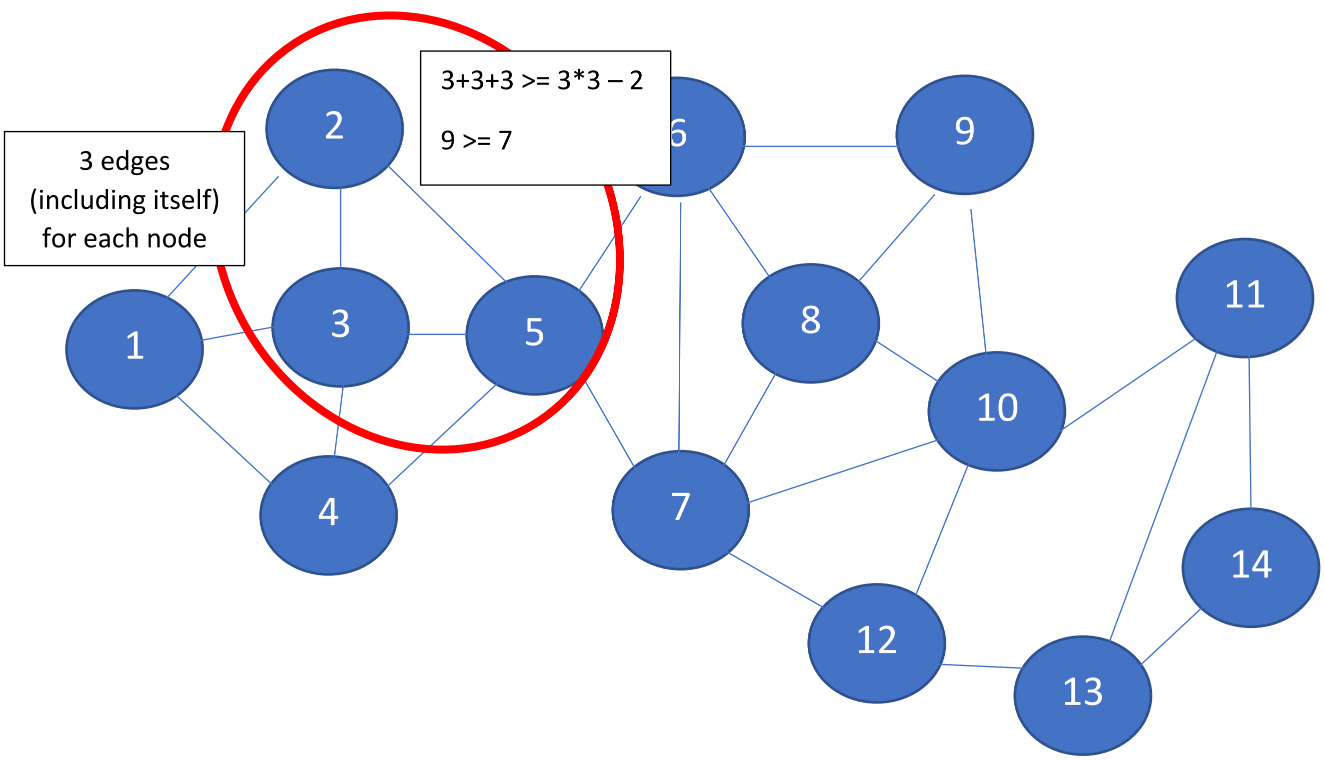


District	Counties	District Population	District White	Perc White
1	7,12,13	669,035	514,771	76.9%
2	9	809,829	558,782	69.0%
3	7,8	885,108	425,000	48.0%
4	5	862,111	652,618	75.7%
5	7,10,11	669,035	534,176	79.8%
6	6	726,836	517,507	71.2%
7	6	905,166	644,478	71.2%
8	1,2,3,4	828,188	598,585	72.3%
9	10,11,12,14	674,609	559,762	83.0%
	Mean	781,102	Mean	71.9%
	Max	905,166	Max	83.0%
	Min	669,035	Min	48.0%



Heuristic Examples:

Example 1: Heuristic continuity metric successful:



Example 2: Heuristic continuity metric **NOT** successful: