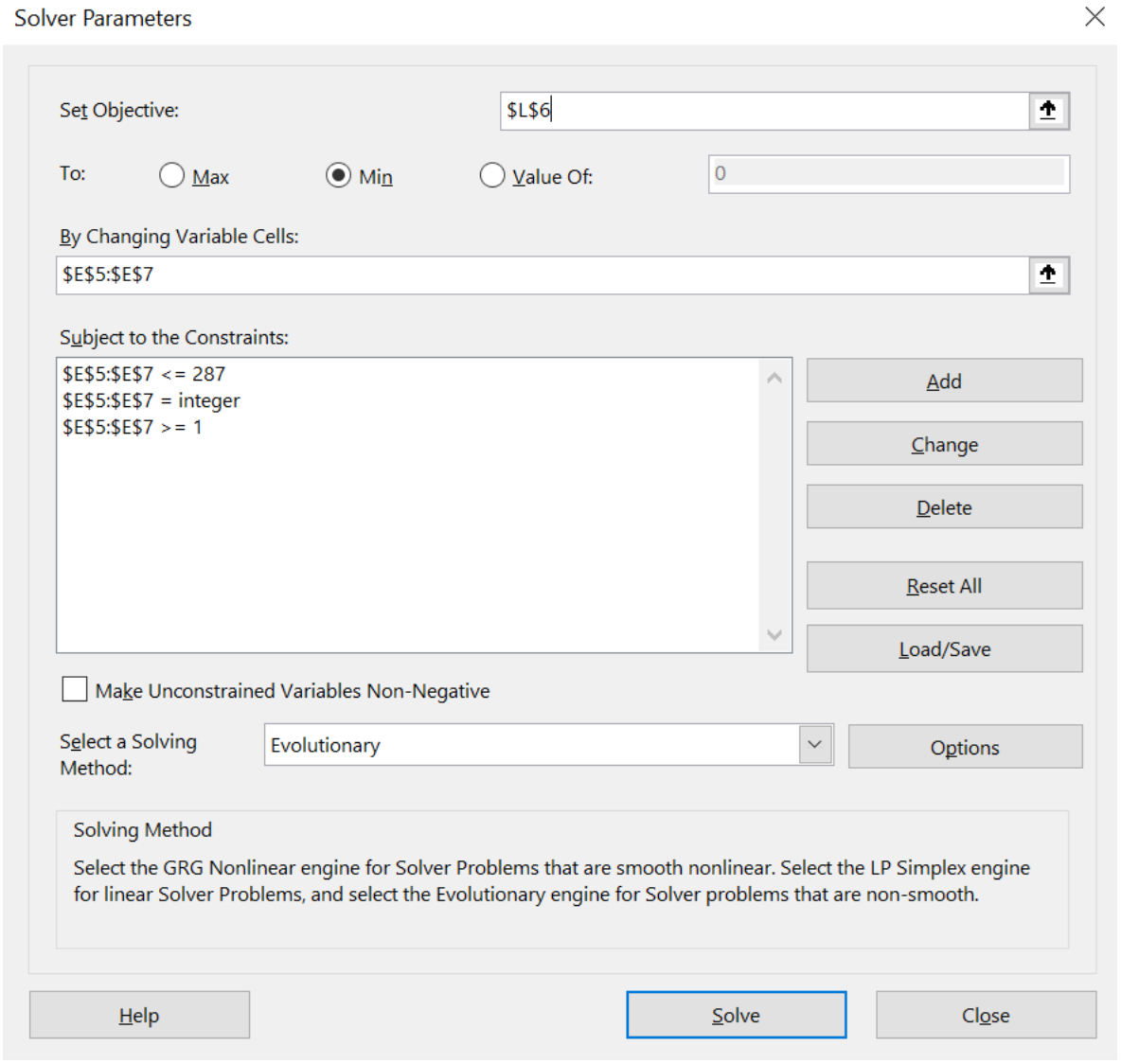
Instructions for Replicating Analysis

1. Download the data titled “Statewide by Party and Precinct (CSV)” from https://elections.maryland.gov/elections/2016/index.html. Save it as an Excel workbook (.xlsx).
2. Apply several filters to the data by going to data and then filter. Filter the column titled LBE such that it only includes “Baltimore City.” Filter by “PRECICNT” to skip any records where the precinct is listed as “unable to be determined.” By filtering on “ELIGIBILE VOTERS,” remove any records for which there are less than 10 eligible voters. Copy all data that is still showing after these filters have been applied into a new worksheet titled “Filtered Data.”
3. In “Filtered Data,” add a column titled “Precinct-Party Key,” which will serve as a lookup key indicating a precinct and the party affiliation of a given group of voters. Do this by concatenating the value of the “PRECINCT” column with an underscore and the value of the party column, with one caveat: if the content of the “Party” column is not “DEMOCRAT” or “REPUBLICAN” concatenate the underscore with “OTHER”. The formula should look something like the following: “=CONCATENATE(D2, "_",IF(OR(E2="DEMOCRAT", E2="REPUBLICAN"), E2, "OTHER"))”
4. Create a PIVOT table out of the Filtered Data workbook (which now includes the Precinct-Party Key column). Set Precinct-Party key to be a row and sum of “ELIGIBLE_VOTERS” and sum of “Count of Voters” to be the columns in the PIVOT table.
5. Create a new worksheet called “Lookup.” Copy and paste the entire precinct column from the initial dataset download (the worksheet titled “Official by Party and Precinct”). Highlight what you just pasted and navigate to data and then the remove duplicates button which looks like three cells in a column where the top and bottom cells are blue and there is a red x in the right corner. Remove duplicate values.
6. In the subsequent columns of “Lookup” to the right of the “Precinct” column, use either VLOOKUP or a combination of MATCH and INDEX to lookup the following quantities from the Pivot table you created in step 4 for each precinct: number of eligible Democrat voters, number of Democrat voters, number of eligible Republican voters, number of Republican voters, number of eligible Other voters, and number of Other voters. Concatenate the value of the precinct column with an underscore and either “DEMOCRAT,” “REPUBLICAN,” or “OTHER” depending on the column. This formula should look something like the following: “=INDEX('Voters by Party and Precinct'!$A$3:$C$877, MATCH(CONCATENATE($A2, "_DEMOCRAT"), 'Voters by Party and Precinct'!$A$3:$A$877, 0), 2)”
7. Create two columns titled “Overall number of eligible voters” and “Overall number of voters” by summing the appropriate column for each of the party categorizations (DEMOCRAT, REPUBLICAN, and OTHER). Create a new column called Overall Voter Turnout.
8. Copy the Precinct column, number of eligible Democrats column, number of eligible Republicans column, and Overall Voter Turnout column into a clean worksheet and prepare to perform a cluster analysis.
9. Insert two rows at the top of the pile that contain the mean and sample standard deviation of the number of eligible Democrats, number of eligible Republicans, and Overall Voter Turnout rate.
10. Insert a column to the left of the precinct code that will serve as the cluster number. Using the STANDARDIZE function, compute z-scores for the number of eligible Democrats, number of eligible Republicans, and Overall Voter Turnout rate.

11. Arbitrarily choose 3 precinct numbers that will serve as our three initial cluster anchors. List their cluster numbers above the data and one column to the left of the first column containing z-scores. Two cells above the topmost cluster number, type the label “Column.” For the three cells to the right of this cell, put the integer number corresponding to the index number of that column. For example, since column F is the sixth column in the workbook, the cell in column F should be 6 (the cell in column G would be 7 and so on).
12. Define the data from the Cluster Number column to the rightmost z-score column to be the named range named “Precinct_Cluster_stats.”
13. Use VLOOKUP in the cells immediately to the left of the cluster numbers discussed in step 11 to lookup the precinct code associated with the given cluster number. Your formula should look something like: “=VLOOKUP(E5, Precinct_Cluster_Stats, 2)”
14. Use VLOOKUP in the set of 3x3 cells immediately to the right of the cluster numbers discussed in step 11 to lookup the z-scores for the clusters that have been specified as anchors. Your formula should look something like the following (note the dollar signs): “=VLOOKUP($E5, Precinct_Cluster_Stats, F$3)”
15. In the normal data set below the data for the anchor clusters, compute the square distance of each precinct in the data from each of the anchor clusters. Use the SUMXMY2 function and the cells containing the z-scores for the three anchor clusters you produced in the previous step. For the leftmost column of square distances, the formula should look something like this: “=SUMXMY2($F9:$H9, $F$5:$H$5).” Every time you move a column to the right, increase the number indicating the row in the second parameter by 1.
16. Insert a column to the right of the rightmost distance column titled “Minimum Square Distance.” Utilize the MIN function to take the minimum of the square distances from the anchor clusters that you calculated in the prior three cells immediately to the left.
17. Above the Minimum Distance column header, compute the sum of the minimum square distances. The formula should look something like the following: “=MIN(I9:K9)”
18. Create a column to the right of the “Minimum Square Distance” column and title it “Assigned to”. The formula should mirror the following construction: =MATCH(L9, I9:K9, 0). This formula basically identifies which anchor cluster the current cluster was closest to in terms of squared distance.
19. Go to the data tab and select solver (assuming it is already installed). Configure you solver’s settings as shown in the following screenshot. Also, select “Options,” “Evolutionary,” and set the mutation rate to 0.5. The key is that the cell currently represented as “$L$6” in the screenshot is the sum of minimum square distances and the cells currently represented as “$E$5:$E$7” in the screenshot correspond to the cluster numbers created in step 11. Note that we restrict the value of these cells to being an integer between 1 (our lowest cluster number) and 287 (our highest cluster number).


Solver Parameters
Set Objective:
$L$6
To:
Max
Min
Value Of:
0
By Changing Variable Cells:
$E$5:$E$7
Subject to the Constraints:
$E$5:$E$7 <= 287
$E$5:$E$7 = integer
$E$5:$E$7 >= 1
Add
Change
Delete
Reset All
Load/Save
Make Unconstrained Variables Non-Negative
Select a Solving Method:
Evolutionary
Options
Solving Method
Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simplex engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.
Help
Solve
Close