**Instructions**

1. Download data from Maryland Open Data Portal, the Maryland Food System Map, and the JHU CSSE daily reports of COVID-19 data.
2. For socioeconomic/demographic data:
   1. Create a master workbook with all necessary variables, pulling data from the Open Data Portal and the Food System Map files.
   2. Create rough scatter plots comparing two variables to see how many rough clusters there might be (if applicable).
   3. In a separate worksheet: calculate standard deviation and mean for each variable.
   4. Calculate the z values for each of the variables using the STANDARDIZE function.
   5. Create another table with the columns “anchor,” “county #,” “county name,” and the z values for each of the variables. Pull the county name and z variables using a VLOOKUP function and the “statewide” table. Use three anchors.
   6. Add three columns to the original “statewide” table. Name these “dist 2\_1,” “dist 2\_2,” and “dist 2\_3.”
   7. Use the SUMXMY2 function to pull values for each of the “dist 2\_x” columns.
   8. Add another column and call it “min dist^2” (minimum distance squared). Input numbers in this column by using the MIN function, with the values coming from one row of the “dist 2\_x” columns.
   9. Create a new column at the end called “anchor” and use the MATCH function to see which anchor it corresponds with.
   10. Calculate the sum of the minimum distance squared by using the SUM function for all the values in the “min dist^2” column.
   11. Run a cluster using the Solver add-in.
   12. Make sure the solving method is “Evolutionary.”
   13. On the same worksheet, make a list of which counties are in which anchor cluster.
3. For health data:
   1. Create a master workbook with all necessary variables, pulling data from different datasheets from the Food System Map.
   2. Calculate the mean and standard deviation for each variable column.
   3. Then, determine the z-scores for each variable by using the function STANDARDIZE.
   4. Create a separate table above the data with the chosen anchor points with their corresponding z-scores by using the function VLOOKUP.
   5. Continuing to the right of the data table, calculate the distance squared to each anchor point by using the function SUMXMY2.
   6. Calculate the minimum distance squared out of the four distance squared in the same row (same tract/town).
   7. By using the function MATCH, determine the anchor number of the minimum distance squared of the row.
   8. Calculate the sum of the minimum distance squared.
   9. Use the Solver analysis tool to conduct the cluster analysis calculations.
   10. Add constraints in the Solver window.
   11. Use the function COUNTIF to count the tracts in each cluster.
   12. Create a bar graph that visualizes the number of tracts in each group.
   13. Add chart title and axes titles to the chart.
4. For COVID-19 data:
   1. Download data from the JHU CSSE GitHub. Use data from 10/9/2020 to 10/23/2020 and compile it in one master document.
   2. Calculate an average value for each variable by county.
   3. Remove the “Recovered” column as there were no available data.
   4. Calculate the mean and standard deviation for each variable column.
   5. Then, determine the z-scores for each variable by using the function STANDARDIZE.
   6. Create a separate table above the data with the chosen anchor points with their corresponding z-scores by using the function VLOOKUP.
   7. Continuing to the right of the data table, calculate the distance squared to each anchor point by using the function SUMXMY2.
   8. Calculate the minimum distance squared out of the four distance squared in the same row (same tract/town).
   9. By using the function MATCH, determine the anchor number of the minimum distance squared of the row.
   10. Calculate the sum of the minimum distance squared.
   11. Use the Solver analysis tool to conduct the cluster analysis calculations.
   12. Add constraints in the Solver window.
   13. Use the function COUNTIF to count the tracts in each cluster.
   14. Create a bar graph that visualizes the number of tracts in each group.
   15. Add chart title and axes titles to the chart.
5. Creating cluster maps:
   1. Download an outline of Maryland state and counties.
   2. Use Photoshop and the “fill bucket” to color-code the anchors to create a cluster map, one for each group of data (socioeconomic/demographic, health, and COVID-19).
6. For correlation:
   1. Compile all three groups of data into a master document.
   2. Select all the data and use the Solver add-in to run a correlation.
   3. Add a conditional formatting rule to color-code on a color scale the correlation coefficients (-1 is red, 0 is yellow, 1 is green).
   4. Bold any interesting correlation coefficients (beyond what might be expected, like a higher population having a higher correlation with the number of COVID-19 deaths).
7. For cluster interpretation:
   1. Make a worksheet containing the cluster anchors for all three groups of data and interpret the results for each anchor/cluster. Name it the “anchors + clusters” worksheet.
   2. Compile a list of clusters with their constituent counties. Add in the anchor number and the corresponding interpretation for each cluster. Name this worksheet “interpretation.”
   3. Make a master table to visualize which county belongs to which cluster for each group of data. Color-code it by anchor number (1, 2, or 3). Create a “Master Cluster” column that combines the three anchor numbers for each colony, and use Conditional Formatting and a color scale to visualize which counties share similar clustering tendencies across the three data groups.
   4. Add in the corresponding interpretations from the “interpretation” worksheet.
   5. Consolidate the interpretations into an “Overall Master Interpretation” column for easy viewing.