(1) Regression

1. We started by downloading data from OpenDataSoftwre titled “USA 2016 Presidential Election by County”. We entered “Pennsylvania” as the only filter to only extract data on counties from the state of Pennsylvania.
2. Once data was downloaded and uploaded onto an Excel sheet, we pulled data on racial makeup and voting results from each county and placed data onto a second excel sheet.
3. After the data was organized by 2016 candidate votes and racial makeup of each county, we clicked on “Data, Data Analysis”. A box offering different forms of data analysis should appear. We chose “Regression” as our analytical tool.
4. From there we placed “% of Republican/Democrat vote per county” in Y axis while placing all ethnic makeups in X axis. We chose the “Labels” and “Residuals” boxes so they could be included in the final Regression. After boxes are included, we pressed “Okay” and the regressions were made.
5. Repeat the process for the political party not chosen in the first Multiple Linear Regression. Finally, we analyzed results and drew conclusions on the two Multiple Linear Regressions.

(2) Clustering

1. We gathered the necessary columns for clustering: State, Fips code, County name, percentage of Democratic votes in 2016, percentage of Republican votes in 2016, and percentage of the various races we analyzed.
2. We made one sheet to cluster the Democrat votes and another to cluster the Republican vote.

***We did the following steps for each sheet:***

1. We calculated the z-scores for each of the variables for each county using the STANDARDIZE function.
2. We numbered each of the counties 1-67 by adding a column “anchor number” next to the county names.
3. We chose five random counties (1, 33, 25, 47, 60) and wrote the corresponding number down. These would be anchors 1, 2, 3, 4, and 5.
4. Using VLOOKUP, we filled out all of the z-scores for each of the variables for the five selected counties.
5. We calculated the distance squared for each of the 67 counties’ z-scores from the z-scores of anchors 1, 2, 3, 4, and 5 using the SUMXMY2 function.
6. We calculated the lowest distance squared among each of the anchors for each county and matched that minimum number with the corresponding randomly chosen anchor numbers using the MIN and MATCH functions.
7. We calculated the sum of the minimum distance squared for all of the counties and used that as the objective for the Solver tool.
8. We used the Solver tool to determine what the actual anchors are for the data – anchors that correctly represent the five distinct clusters. We set constraints for the values to be >= 1, <= 67, and an integer.
9. We used the new anchors to analyze and interpret the data.

(3) Pivot Table and Chart

1. Highlight all the MATCH results from the cluster analysis and press “Insert, PivotChart”. Choose for the PivotChart to appear on a blank sheet
2. Once the PivotChart appears, organize field names. “Anchor numbers” should be placed in rows and count of counties should be placed in values.
3. After this, data PivotChart and table should clearly show the distribution of counties into each anchor number