

Week 4 Homework

Question 1

For this question, you will investigate the relationship between education completion rates and poverty levels. The data is currently found in two separate files. You will follow the steps outlined below to prepare the data, merge the data, and then analyze the data.

1. Read in 'education_long.csv'
2. Reshape the data into 'wide format'
3. Read in sheet 2 of the file 'PovertyReport.xlsx'
4. The poverty data was collected in 2017. Select rows from the education dataset from 2017.
5. Merge these two data sets so that all matched observations are retained.
6. Graphically investigate the relationship between general poverty levels and the college completion rates for each demographic group. What do you notice?

Question 2

For this question, you will investigate the relationship of the racial demographics of a school and the critical reading score on the SAT.

1. Load the datasets profiling NYC schools from 2006-2012 and NYC schools 2010 SAT scores, respectively.
2. For this problem, we want to merge the 2010 SAT dataset with the the NYC school profiles for that year. First, subset the NYC school profile so that it only includes data from the 2009/2010 school year.
3. If you look at the dimensions of the data, it seems that there are substantially more schools included in the profile2010 data. Look through both datasets carefully to see why this might be (Hint: search for areas where there are missing values)
4. Remove the values that do not belong.
5. Merge the datasets. Which variable is the most appropriate to use as a unique identifier?
6. Now, let's explore this data a bit. Say we want to see how the racial demographics of a school impact the school's average critical reading score on the SAT. Plot these two variables for each race on four different graphs and discuss findings.