R Homework — Due Week 3

*Creating & Manipulating Data*

Introduction:

In R programming, you often will have to modify your dataset before you can effectively parse through and run analysis on it. In this homework we will put together concepts we’ve learned in Week 2 and Week 3 in order to better examine the Open Policing Project data. We’ll be using piping, and these can get pretty in depth, so come to me (Michael) if you have any questions.

Download the Dataset:  
The Open Policing Project is run by Stanford and tracks data on police traffic stops. It is freely available for your analysis online.

1. Use the link: https://openpolicing.stanford.edu/data/
2. Navigate to the DATA tab, scrolling down to the section on North Carolina. Click on the R icon for any Raleigh, and it will download as a .csv file. Save it wherever your workspace is located! Begin by reading the .csv file into R. Save this dataframe as df\_1.

The Assignment

Warm-up

1. Subset the data so you are only looking at the data where the subject\_sex is equal to “female”.
2. Using piping, look at the subject\_age and get the average and standard deviation for this value.

Advanced Analysis

1. Clean the dataset so you are back to the original df.
2. Get the number of stops for the year 2013. Print your result to the console.
   1. Hint: here, you can use filter().
3. Get the proportion breakdown of each race in the Raleigh dataset. Print your result to the console.
   1. Hint: you can use mutate() here to create the new variable.
4. Find the “hit rate” for Raleigh, separated by subject race.
   1. Hint: Filter the dataset and group by race. Then you would need to summarize. What column should you summarize? Print your result to the console.
      1. Utilize piping to make your code more efficient.
5. Further break down the traffic stops, by “year” and by “subject\_race”
   1. Hint: we looked at in the lesson ways to break down dates so that we only have the year. As for the break down by year and race, the count() function would be helpful here.
6. Finally, let’s put some more concepts together. Group the data by “subject\_race” and get the search and risk rate breakdown for each race.
   1. Hint: group the data first, then you can run the analysis on it. What would you need to get search and frisk rate from the data?