# Data Storytelling

So after completing the wrangling portion of the project, I wondered if there was any additional prep that I needed to do before starting the comparisons and I didn’t see anything too problematic. However, what I decided to do was one last wrangling type of analysis before doing comparison analysis. I wanted to see what sizes the towns/cities we were going to deal with in case it affected analysis later on. I also wanted to make sure there were reasonable group sizes so I have split up the towns on population. Those with less than 25,000 residents are in one group, then 25,000-50,000, then 50,000-100,000, 100-500,000, and anything larger were all grouped separately. Later on during analysis, we can see if this is significant or not, but I wanted to have them grouped out in case.

The first thing I wanted to see was the percentage of homeless people and those in shelters. This statistic was not included in the original data but I was able to calculate it relatively easily. Those who are able to stay shelters are more hidden from view of others and may not be subject as much to both violent and non-violent types of crime. They most likely do not have as big of an impact as those out on the street. The first plot is of the percentage of those homeless and those in shelters. Immediately when looking at the graph we can see that the larges # of those living in shelters seems to be around 2% which is good since it means there isn’t a high homeless pattern in any of the towns, leaving one less variable to worry about, since a significantly large homeless population when compared to other cities, may affect crime rates more than expected. I expected there to possibly include more higher amounts of homeless to create a better training set. However, even in New York City, the percentage seems to be extremely low. Generally, the percentage of those on the streets are lower than those in shelters, meaning the city has room to accommodate most of the people in the shelters, which could mean that since the city has enough money to combat the homeless situation, it would typically have a higher presence of officers present, leading to less crime, or a higher socio-economic level in the community, also possibly leading to lower crime rates.

The next plot is them plotted against each other. Specifically, I wanted to do this to find outliers in the ratio of those living in the two separate places. These two types of cities would be particularly interesting to look at during the analysis: those with a higher percentage of those in shelters and extremely low on the streets, as well as an extremely low percentage in shelters and a higher percentage on the street. The former would usually mean a higher QOL and lower crime rates while the latter could mean higher crime rates.

In the following plot, I wanted to see the distribution of cities in the data, and it seems like there is a rather unequal amount of urban to rural communities. There was not a baseline given for what was considered urban or not, but perhaps the best data set would have had a relatively equal number of urban and rural communities and a small but substantial number of communities within the two indicating varying levels of suburbia. I do believe that there is a connection between crime rate and cities, with the hypothesis, yet to be proven by this data, that urban areas have much more crime than rural ones.

I then plotted urbanism to poverty, as it would seem to logically make sense that urban areas would have higher percentages of people living in poverty, yet urbanism does not seem to be directly correlated as based on the plot output there seems to be a high poverty level in rural areas as well. It is intriguing that the rates of poverty are among the lowest in semi urban then start climbing rapidly at 60% urban, yet completely rural neighborhoods show extremely high amounts of poverty.

Unemployment and poverty were also noted as it is important to note that in a lot of these area, tons of people are living in poverty and yet are gainfully employed, but do not make enough to support themselves and/or family leading them to possibly commit crime out of necessity. These would be more likely to be possible non-violent crimes over violent ones.

Visualizing the median income for races over the cities show the discrepancy that different races face and may be a contributing factor towards crime rates that seem to point toward certain races.