**FINAL PROJECT**

I started this project wanting to find some correlation between factors such as age, sex, or race and the probability of death due to accidental drug overdose. Because of the dataset, I changed my question many times, as the data was in a format that was not acceptable for what I wanted to do. I removed some extraneous demographics columns, as well as certain variables which were small or irrelevant. For the date, I removed all except the year, as I wanted to see if certain drugs were more popular in certain years. My hypothesis was that certain drugs would be more popular to certain age groups, or even between genders. I feel the analysis told a pretty good story of the correlation between the variables. Alas, I did not see any real connection between certain drugs and age groups. A worrying statistic I did notice, was that deaths from overdoses were increasing as the years passed. And, a couple outliers were the 14-year-old female and 15-year-old male, along with an 84-year-old male, which is extremely saddening.

After completing this analysis, I would have liked a few more data points to determine correlation and possibly causation of drug use to begin with. I would have liked to know the subjects’ family information, such as if they were married, or if they had kids. Maybe if they were prescribed medication, this could have been a factor as well, or even if they suffered from any ailments such as depression, anxiety, etc. I believe some of these factors could help paint a better picture of reasons for excessive drug use. I did face some challenges as well along the way. The data had to be manipulated is a way that I am familiar with using it. Also, some tools might be available to better use the data that I have not even seen yet. Hopefully I can use a similar dataset in a future course to produce some better and more accurate results.