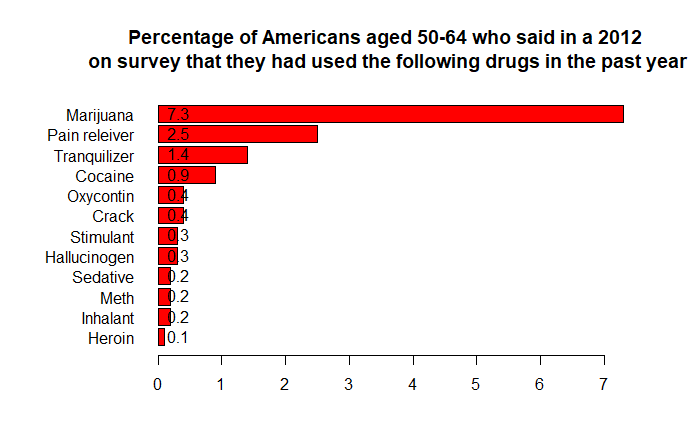
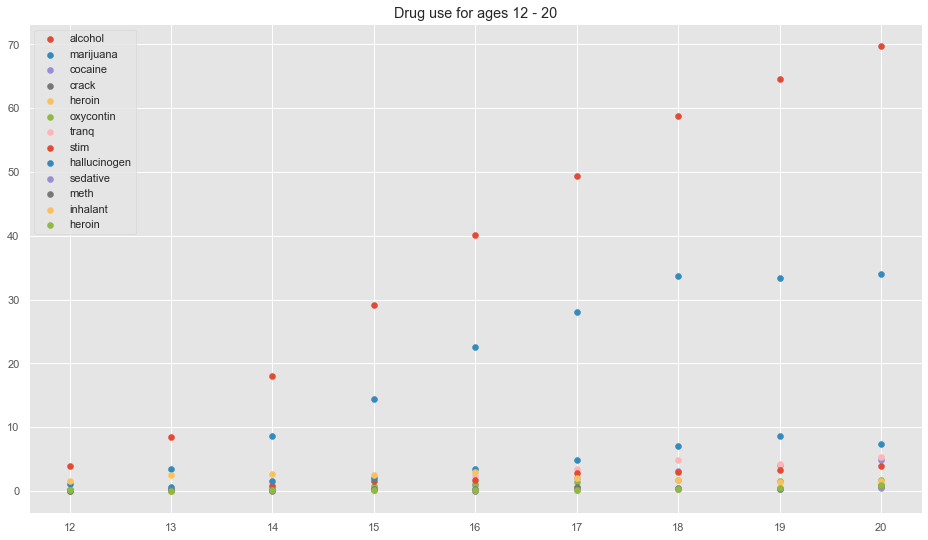
Final Report

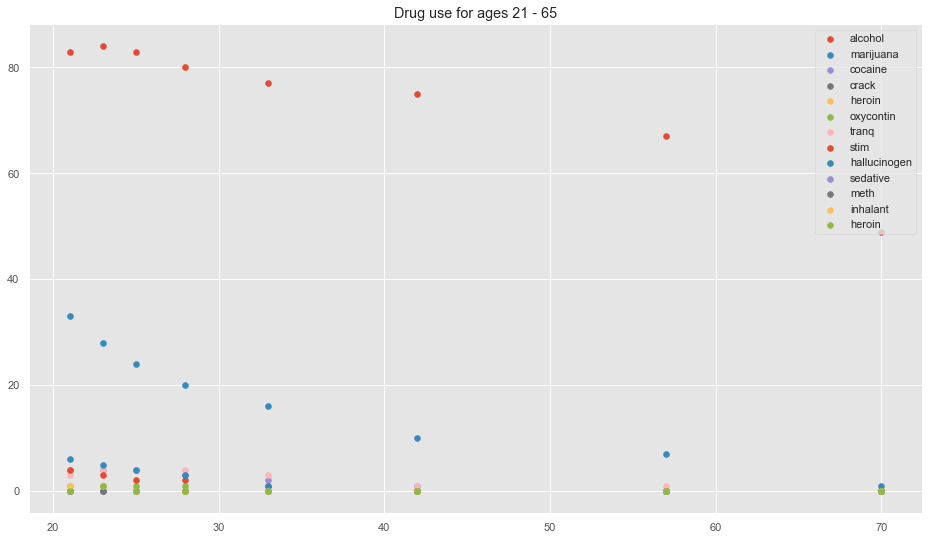
Over the past decade we have seen a rise in the use of drugs in our daily lives. This may have been helped through the legalization of some of these drugs, such as marijuana. An article done by FiveThirtyEight reports on how the baby boomer generation has seen an increase in drug related hospital admissions and overdoses. Their studies show that 7.29% of Americans, aged 50-64, who participated in a 2012 survey, has used marijuana in the previous year. There were some things that the article does not mention, such as, how likely is it that these drugs are correlated with one another? What ages do people usually spike their interest in taking drugs? And how closely related is one instance of drug consumption to another? This is what this project will be trying to find out. This will be done from correlation, plotting, and using kmeans.

The dataset that was used was provided by the FiveThirtyEight website, which was taken from a 2012 survey. The dataset describes the age of the people taking the survey, and the drugs that were in question. These drugs include marijuana, cocaine, meth, oxycontin, and a few others.

 FiveThirtyEight’s approach to the problem was by creating a bar plot that shows what percentage was the highest in terms of drug use for a certain age group. An example of the bar plot can be seen below.

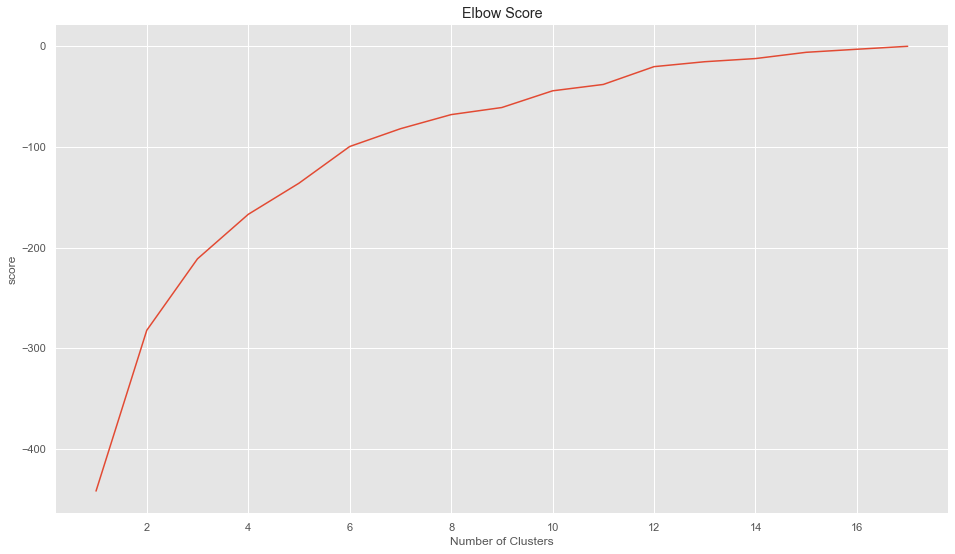
As can be seen from the plot above, marijuana is the most common drug taken between the ages 50-64 (when alcohol is not considered), with 7.29% of the survey participants saying that they have taken the drug in the year before. While heroin is the least common drug with only 0.1% having taken it.

 How does this data hold up against other age groups? Below shows a plot of 2 different age groups, one is 12-20, and the other is 20 and up to 65+.

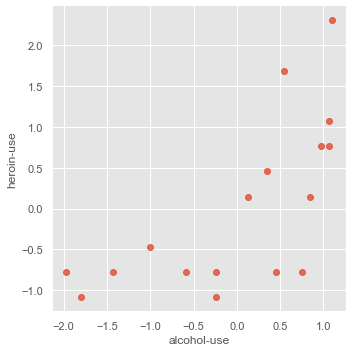


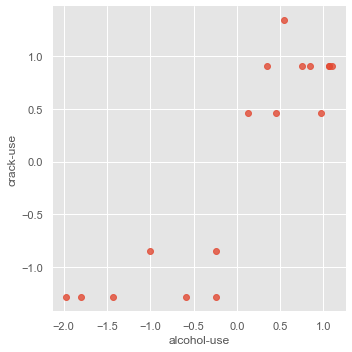
From the 2 plots above it can be seen that most of the drugs hit a peak around the ages of 20-21, except for alcohol, which has its peak around the age of 24. Interestingly as the participants reach the legal age of drinking alcohol, the use of alcohol decrease. However, this trend can be seen from all the other drugs included in the data.

How closely do these drugs relate to one another? Does the use of one drug influence the use of another? This dataset cannot predict this with 100% certainty, as there could be more attributes that contribute to the correlation of one drug to another. The plot on the back of this paper shows how the use of one drug may correlate to another. One thing that can be taken away from the plot below is that use of marijuana is very closely related to the use of many other drugs. This can be seen by how it is very closely related to drugs such as pain relievers, oxycontin, tranquilizers, and stimulants. While inhalant drugs has a very low correlation to any other drugs listed in this dataset, even reaching negative correlation values.

 To take a close look at how related the use of these drugs are to one another, kmeans was used to look at the clustering of the usage of these drugs. The number of clusters used to figure out the clustering of the data was figured out by using the elbow score method. The elbow score plot can be seen below.

As can be seen from the plot above, the graph spikes up until it reaches around 6 in the x-axis. Therefore 6 was used as the elbow score.

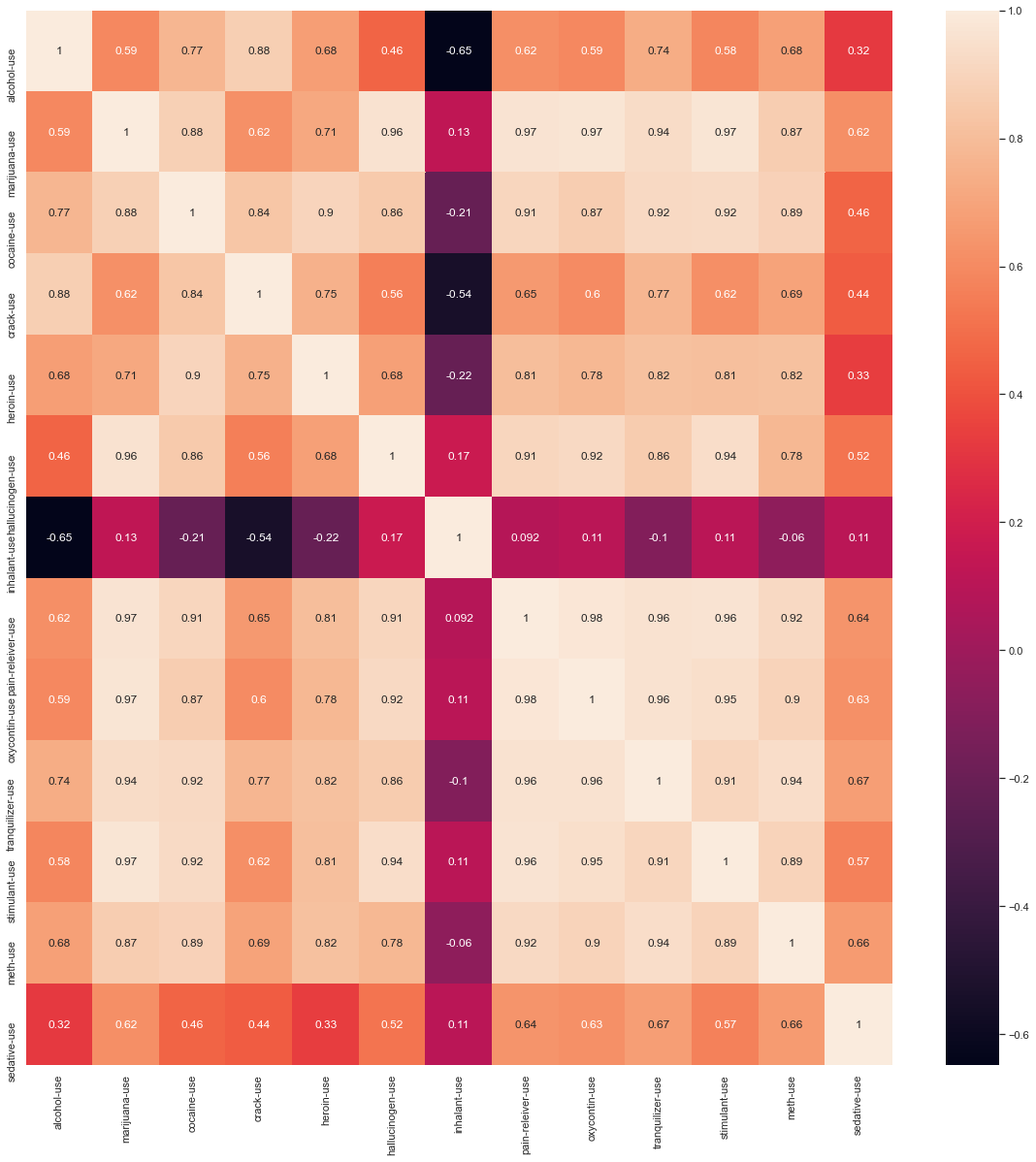
 As for the clustering in itself, it showed that some of the drugs were somewhat related, and some were not. The use of alcohol and heroin were pretty close to one another as can be seen from the plot below.

While the use of crack and alcohol were not, as can be seen from the plot below.

Alcohol was chosen as a baseline to compare to the other drugs as it was the most used across all age groups when compared to the others.

In conclusion, yes there are some correlations between the use of some drugs to another. However, the data from this dataset alone cannot conclude theses correlations with absolute certainty, as there can be many other attributes that contribute to drug usage. What is absolute is that marijuana is the most common drug taken in all age groups when alcohol is not considered in the data. Also, the age for highest drug consumption is around the early 20s.

Future implementations of this project are to get more data on the people being surveyed, to actually study more about the correlation between the use of one drug to another. Another would be to use a better model, while clustering did show some correlations between the use of one drug to another, there could be a better way to do it .



References

Flowers, Anna Maria Barry-Jester and Andrew. “How Baby Boomers Get High | FiveThirtyEight.” *FiveThirtyEight*, FiveThirtyEight, 23 Apr. 2015, <https://fivethirtyeight.com/features/how-baby-boomers-get-high/>.

James, Gareth, et al. *An Introduction to Statistical Learning*. Springer Science & Business Media, 2013.