# Analysis of Drug Poisoning Mortality in the U.S.

## Introduction and Data:

According to the CDC, "nearly 841,000 people have died since 1999 from a drug overdose" and approximately 247,000 of those have been from prescription opioids. The opioid epidemic has been at the forefront of news in the last few years, but the death toll continues to rise. With social isolation caused by the pandemic and increased illegal trafficking of prescription drugs containing lethal doses of Fentanyl, the U.S. has reached its peak historical death rate from drug overdoses according to NPR. From April 2020 to April 2021 over 100,000 people passed away from drug overdoses representing a 29% increase from the same period a year earlier. This analysis aims to identify key demographics that show higher risks for drug related deaths using data collected by the National Center for Health Statistics. This dataset spans from 1999-2018 and shows number of deaths nationally and per state based on age groups and race. The dataset also includes crude death rates and age-adjusted death rates per 100,000 people. This dataset was chosen because it provides a long time period where we more clearly see how each demographic's death rates have changed over the 20 years included. These trends were visualized using the Seaborn package in Python. By graphing the death rates of each demographic, we could clearly identify important differences and verify their significance using an ANOVA test. While there are many studies on drug overdoses by different demographics, this analysis is unique since it examines mean number of deaths by age group and race together.

An example of the dataset used can be seen below:

| State | Year | Sex | Age Group | Race and Hispanic Origin | Deaths | Population | Crude Death Rate | Standard Error for Crude Rate | Low Confidence Limit for Crude Rate | Upper Confidence Limit for Crude Rate | Age-adjusted Rate | Standard Error Age-adjusted Rate | Lower Confidence Limit for Age-adjusted Rate | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| United States | 2003 | Both Sexes | 15-24 | Hispanic | 226 | 7255772 | 3.1148 | 0.20719 | 2.7087 | 3.5209 | 3.1148 | 0.20719 | 2.7087 | |
| Alabama | 1999 | Both Sexes | All Ages | All Races-All Origins | 169 | 4430143 | 3.8148 | 0.29344 | 3.2396 | 4.3899 | 3.8521 | 0.29657 | 3.2708 | |
| Alabama | 2000 | Both Sexes | All Ages | All Races-All Origins | 197 | 4447100 | 4.4299 | 0.31561 | 3.8112 | 5.0485 | 4.4857 | 0.31985 | 3.8588 | |
| Alabama | 2001 | Both Sexes | All Ages | All Races-All Origins | 216 | 4467634 | 4.8348 | 0.32896 | 4.1900 | 5.4795 | 4.8915 | 0.33329 | 4.2382 | |
| Alabama | 2002 | Both Sexes | All Ages | All Races-All Origins | 211 | 4480089 | 4.7097 | 0.32423 | 4.0742 | 5.3452 | 4.7619 | 0.32868 | 4.1177 | |

# Analysis:

Visualizations:

The following graphs show the number of deaths for the entire U.S., for each state, each age group, and each race from 1999 to 2018. Each graph has a description below that clearly explains the trends found in the dataset.

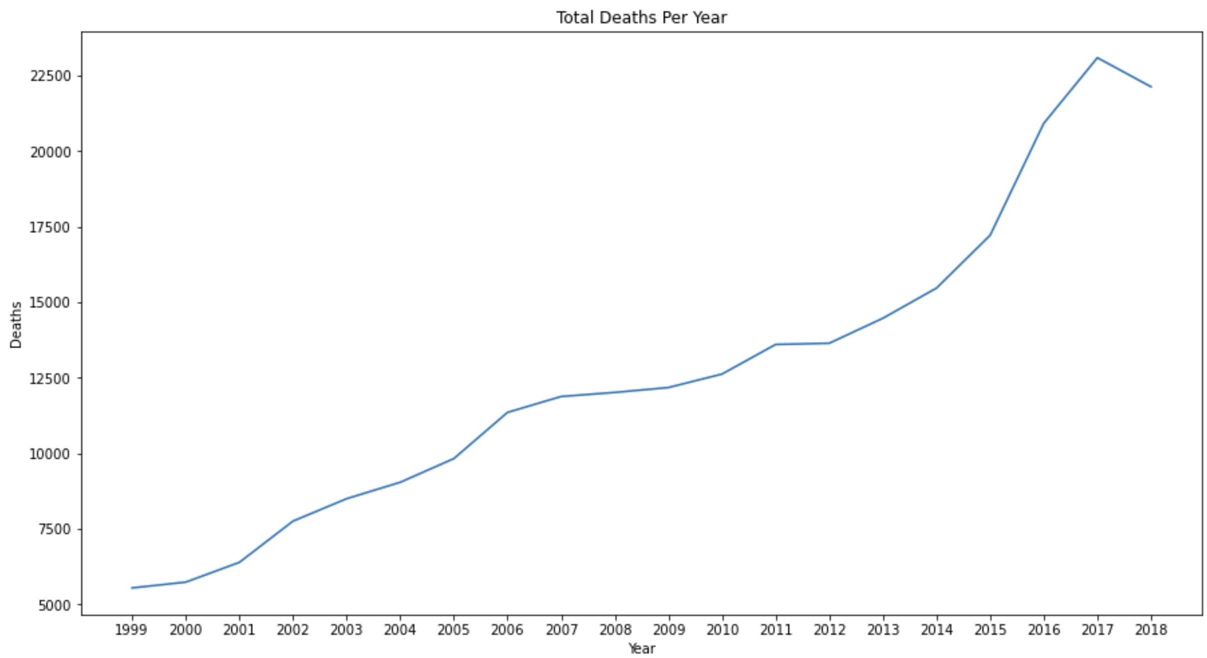



Figure 1:

Figure 1 shows the total amount of drug mortalities in the U.S. from 1999-2018. Each year the number of deaths increase by a few hundred and by 2014 there is a sharp increase that reaches its peak in 2017. This is strong evidence that drug mortalities have become more frequent each year and have reached a historical maximum in the last few years. There is a slight dip in total deaths in 2018 but based on current literature this trend does not continue as 2020-2021 had the highest record number of drug related deaths.

Figure 2: (A sample of states)

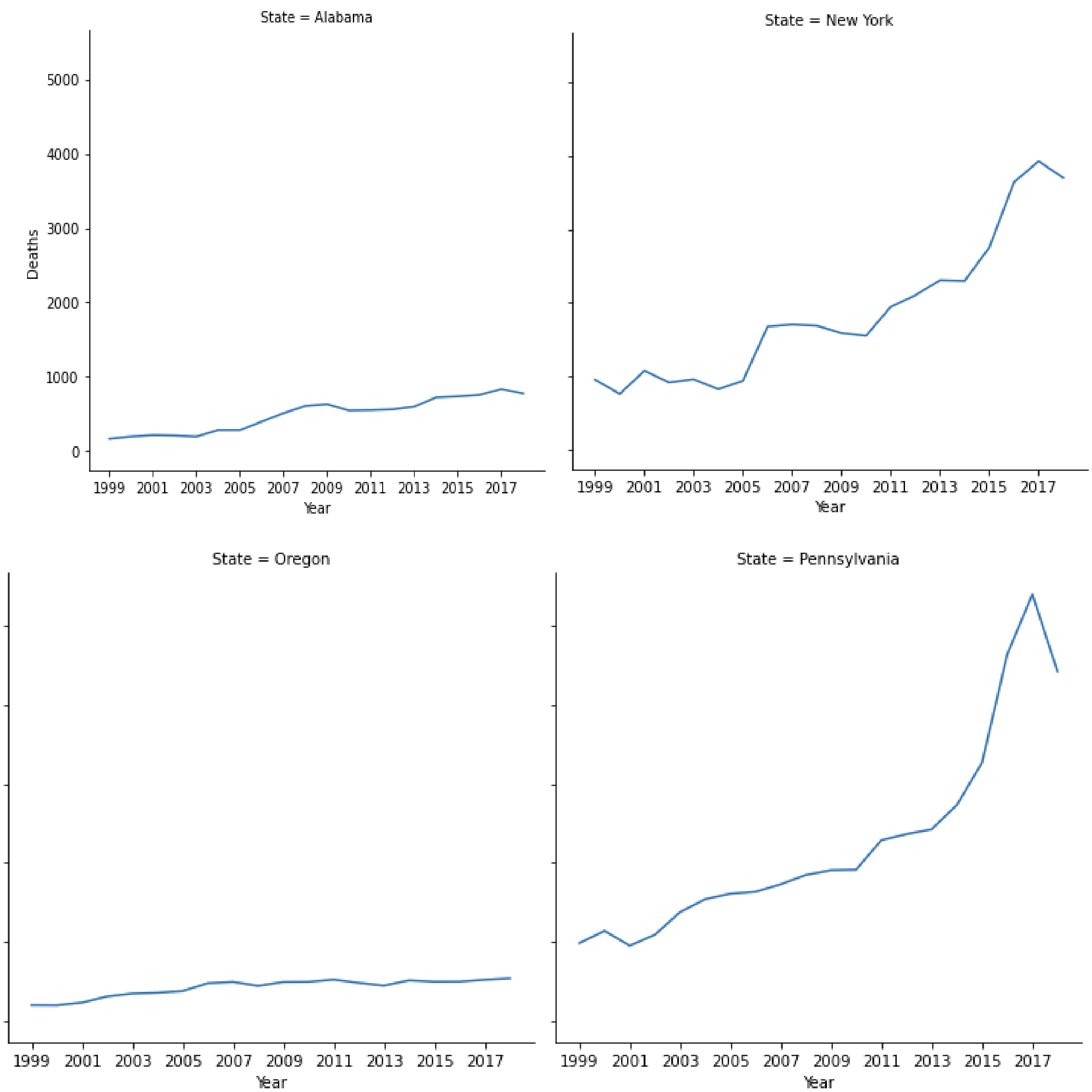
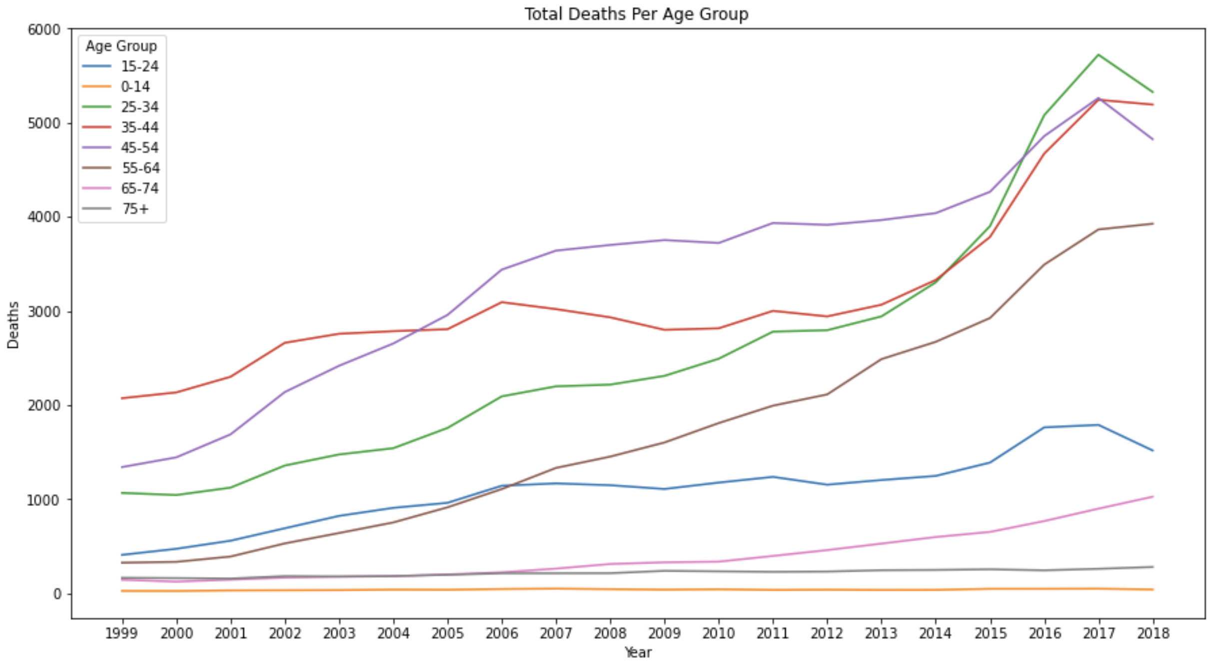



Figure 2 shows the frequency of deaths for a sample of states to show the variation. It is clear that death rates vary greatly between states with large populations (New York) versus states with smaller populations (Oregon). The majority of states have a flat or slightly increasing trend. The states with the highest frequencies and greatest increase in deaths are: California, Florida, New York, Ohio, and Pennsylvania. All of these states are among the top ten most populated states in the U.S. which likely contributes to this trend. We can clearly tell that drug mortalities in these 4 states are at an alarming rate when compared to other states and should be targeted by health organizations.

Figure 3:



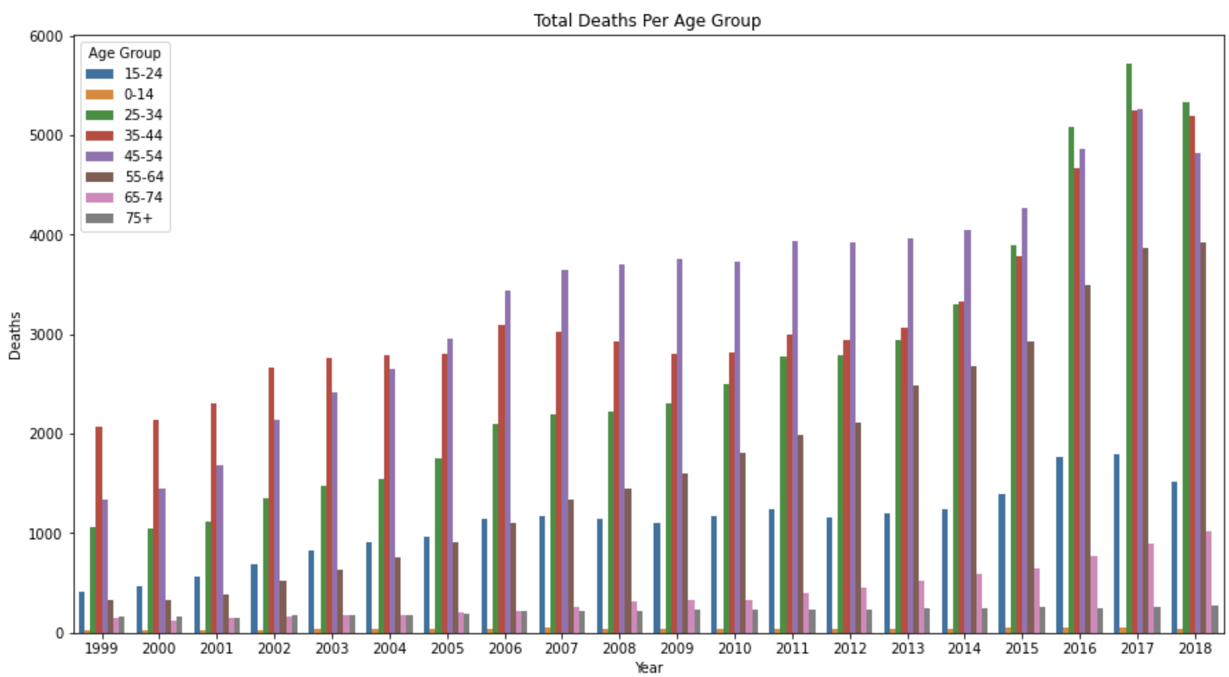
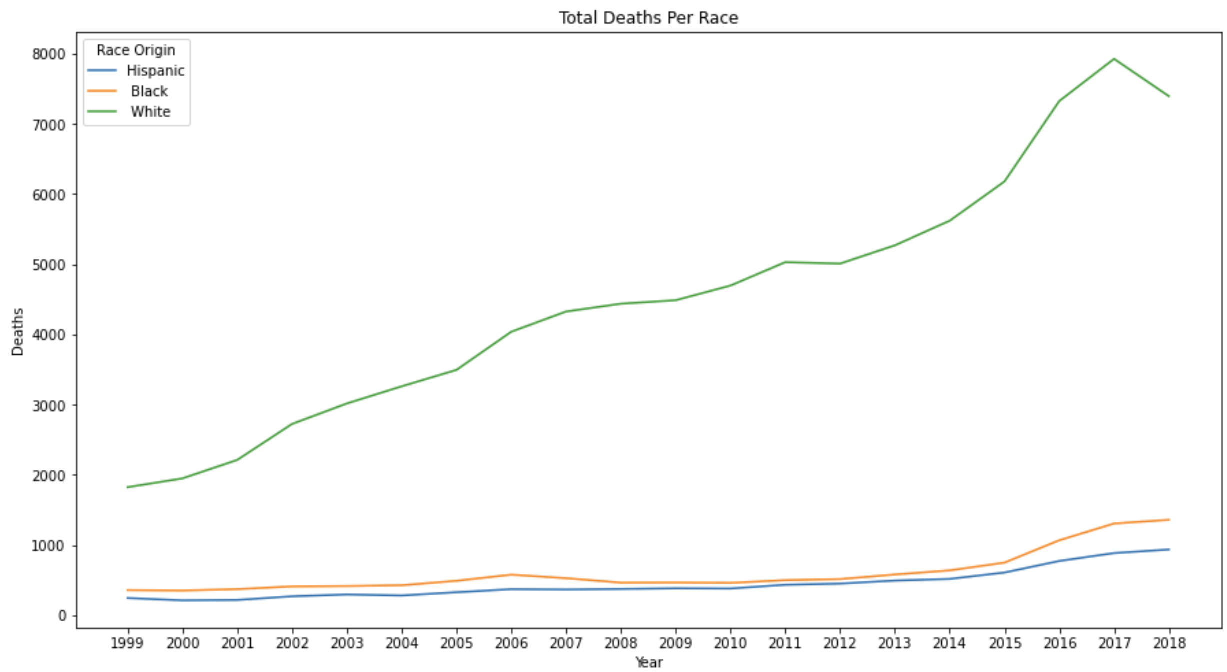




Figure 3 shows that the youngest and oldest age groups (0-14, 15-24, 65-74, and 75+) have the lowest rates of drug mortalities whereas the middle age groups (25-34, 35-44, 45-54, and 55-64) have the highest rates of drug mortalities, all hitting their peaks in 2017. In the bar plot we can more clearly see that the 35-44 and 45-54 age groups, historically had the highest drug mortalities but in just a few years (2015-2018) the 25-34 age group out grows the others and has the highest all time drug mortalities of any age group in 2017. These trends are not surprising as age groups in the middle make up a larger majority of the population and are much more likely to be users of drugs that can be more lethal. What is surprising is the sharp increase in drug mortalities in the 25-34 age group. I expected this younger age group to always have been the leading age group in mortalities.

Figure 4:



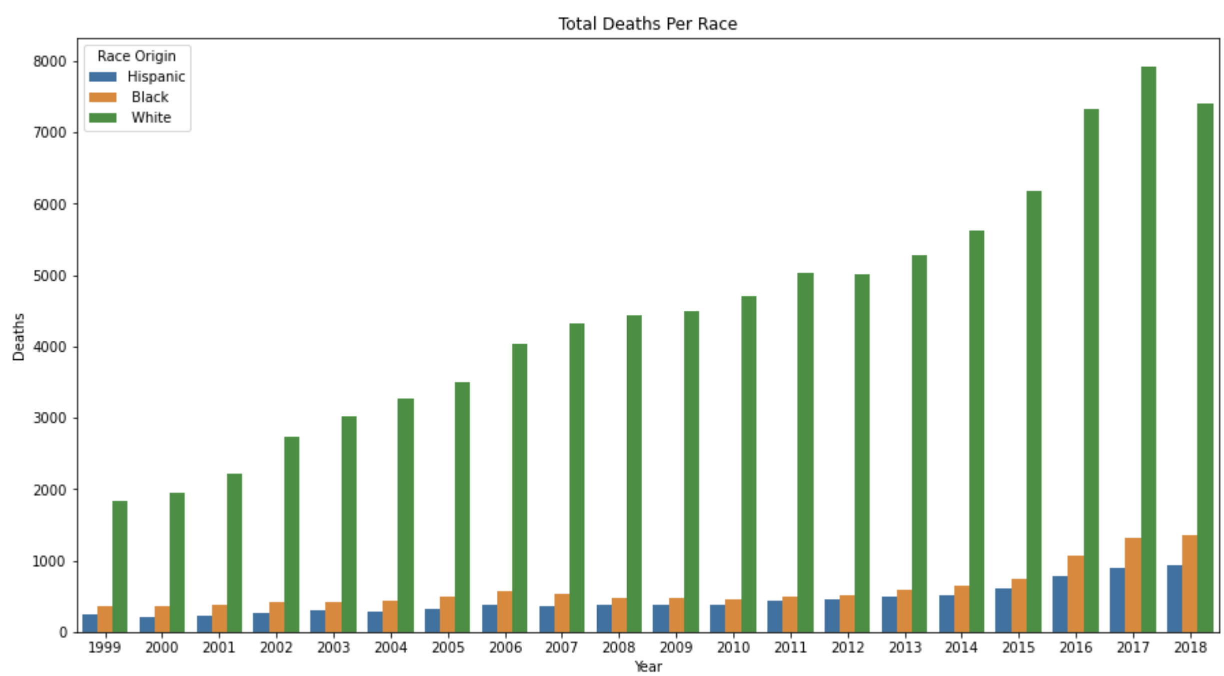



Figure 4 shows that people of White origin make up for almost all of the drug mortalities from 1999-2018 in the United States. Drug mortalities for Black and Hispanic origins remain very low and have flat trends with slight increases in 2016-2018 which is consistent with our other findings when looking at drug mortalities for the entire U.S. This huge disparity may seem surprising but when looking at opioid related deaths in the U.S. by race this trend is also echoed in those statistics because opioid use in white communities are statistically much higher than in other race communities. This trend strongly supports the public health warnings on the opioid crisis.

Figure 5:

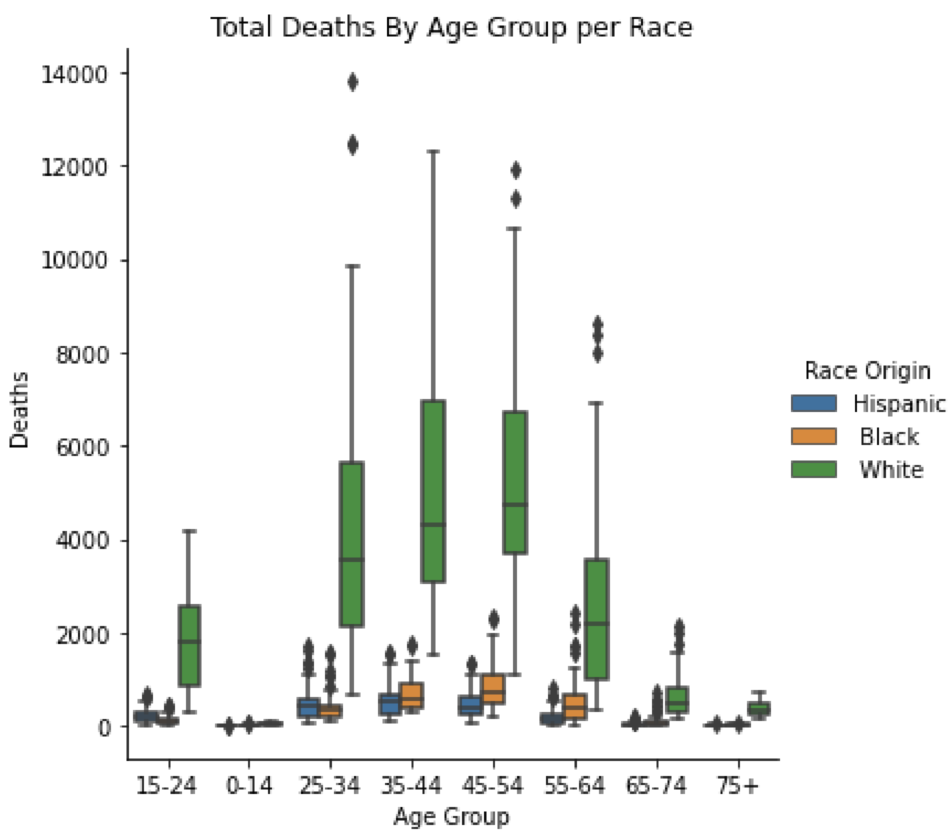
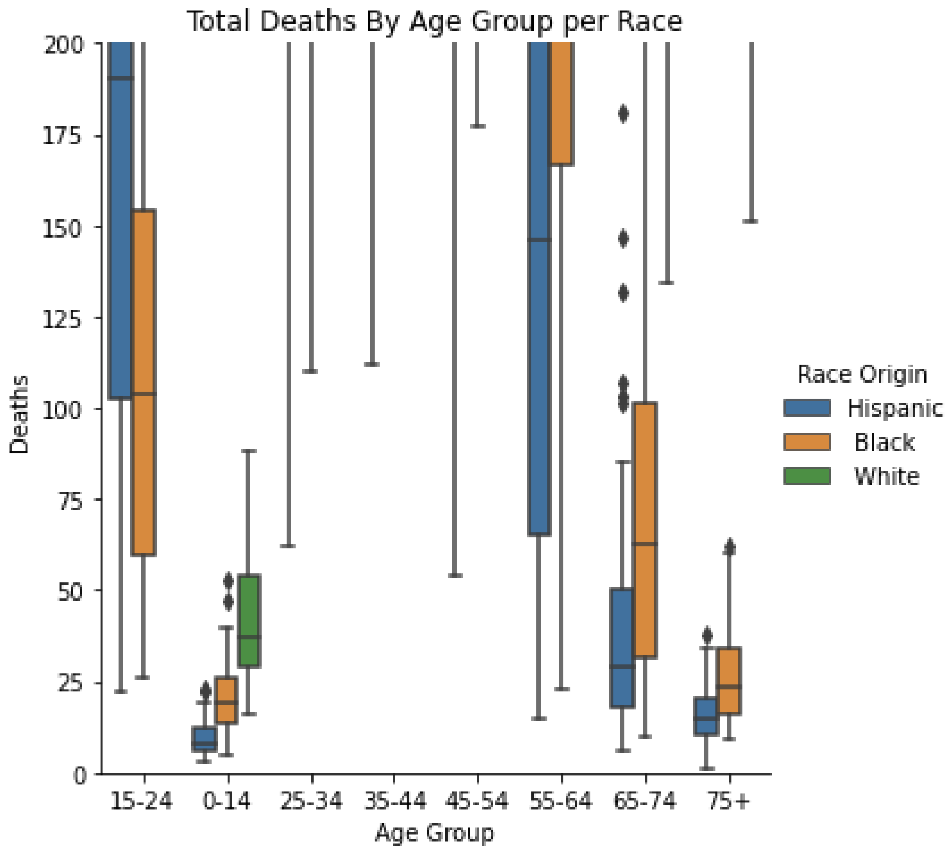


Figure 5:




To examine deaths for each age group by race, a box and whisker plot was utilized. We can see that people of White origin by far have the highest mean number of deaths in every single age group. The 45-54 age group had the overall highest mean number of deaths for people of White and Black origin. The 35-44 age group had the overall highest mean number of deaths for people of Hispanic origin. In the 15-24 and 25-34 age groups, those of Hispanic origin have a slightly higher mean number of deaths than those of Black origin. In the 0-14, 35,44, 45-54, 55-64, and 75+ those of Black origin have slightly higher mean number of deaths than those of Hispanic origin.

The diamonds in the plot represent outliers. In some cases, it would be best to remove these points from the data to normalize the distributions but in this case, it does not make sense to do so. The outliers represent the last few years in the dataset where deaths have reached their historical peak. Compared to the historical means of each boxplot these death tolls are technically outliers but, they have become the new normal, so it is important to leave them in the data to show how much higher the death tolls are in the last few years.

ANOVA Tests:

To measure if the differences between group means are statistically significant a One-Way ANOVA test was utilized. Age group and race origins are our categorical variables of interest and we are looking at the differences in mean deaths between each group. If these tests return a significant p-value (less than .05) than we can conclude that the differences in mean number of deaths between each demographic is statistically significant and should be noted as a key indicator when predicting drug related deaths in different areas of the country.

Figure 6:

```
Age Group
0-14     [74, 74, 91, 97, 100, 117, 112, 137, 151, 131,...
15-24    [226, 148, 127, 1240, 1435, 1700, 2095, 2491, ...
25-34    [3236, 3169, 3410, 4125, 4488, 4680, 5340, 634...
35-44    [6295, 6469, 6968, 8064, 8358, 8439, 8506, 937...
45-54    [4067, 4389, 5115, 6466, 7325, 8040, 8968, 104...
55-64    [991, 1013, 1185, 1601, 1943, 2283, 2761, 3355...
65-74    [434, 371, 440, 508, 534, 549, 607, 677, 791, ...
75+      [496, 483, 470, 552, 536, 555, 596, 644, 644, ...
Name: Deaths, dtype: object
P-Value for Anova is:  6.239572982291277e-107
```

Figure 6 shows the number of deaths per age group by year and gives you better sense of the groups we are comparing in this test. The ANOVA returned an extremely small p-value close to 0 which is less than our level of significance of .05. This means that there is very strong evidence that the differences between group means are statistically significant for age groups. This supports what we saw in the line and bar plot as certain age groups had extremely high rates of deaths compared to others.

Figure 7:

```
Race Origin
 Black      [22, 14, 24, 28, 30, 32, 31, 37, 40, 35, 24, 2...
 White      [37, 46, 55, 53, 54, 72, 63, 75, 80, 79, 71, 8...
Hispanic    [226, 148, 127, 146, 199, 14, 7, 9, 11, 15, 8,...
Name: Deaths, dtype: object
P-Value for Anova is:  1.3458071332241042e-63
```

Figure 7 shows the number of deaths per race by year and gives you better sense of the groups we are comparing in this test. The ANOVA returned an extremely small p-value close to 0 which is less than our level of significance of .05. This means that there is very strong evidence that the differences between group means are statistically significant. This supports what we saw in the line and bar plot as the white race origin had extremely high rates of deaths compared to Black and Hispanic race origins.

## Conclusion and Literature Review

The results from analysis support the facts and narrative that drug overdoses have steadily increased over the last twenty years and within the last couple of years have reached alarming numbers that do not seem to be decreasing anytime soon. Our analysis revealed that that the demographic with by far the highest mean number of deaths are those of White origin in the age ranges of 25-54. For people of Hispanic and Black origin, their deaths tolls were much lower and remained well below 1000 each year until the last few years of the dataset. Between these two groups we found that overall, people of Black origin had slightly higher mean number of deaths in every age group except for 15-24 and 25-34. While other research has looked at age group and race individually, this analysis uncovered differences when looking at these categorical variables together. There is strong evidence that drug overdoses are epidemic in predominantly White communities of all age groups examined. These findings are key indicators for future work to examine why these large differences exist. If we were to further this research, we would target the differences in these groups and combined with other factors such as indicators of wealth, we could uncover specific risk factors that make people in certain types of communities more at risk for drug related deaths. Using this data, a predictive model could be created to help public health workers identify specific communities that may be at high risk.

This type of prediction model has already been employed by researchers at the University of California San Diego School of Medicine, San Diego State University, and other international collaborators. These researchers studied yearly patterns of drug overdoses in every county in the United States and created a predictive model to identify future counties that may be at risk for high numbers of death rates. The first author Charlie Marks, states that their model "brought substantial improvement to predicting counties with high fatal overdose rates compared to a simple benchmark that relied on past year rates alone. We also found that increased overdoses in a neighboring county are very predictive of future overdoses in a given county, indicating that the overdose epidemic spreads geographically." This study is extremely important as it will give public health officials forewarning on which counties may be at risk and allow for proper action before it is too late. However, Marks warns that "our model will only be useful in predicting and preventing deaths if there is no lag in getting data from local and national agencies." This seems to be a common trend with government agencies and their ability to collect data so moving forward a policy change to enable the swift exchange of data may be the best hope for having a useful predictive model.

Without systemic policy changes drug overdoses will likely continue to increase in the United States. Public health workers can do nothing to prevent these overdoses and are simply trying to save as many lives as they can. Policy reform on drugs is a polarizing topic but many experts point to European countries great success with dedicated therapy centers for addiction and regulated drugs which helps to cut out illegal trafficking. Some U.S. states have begun to move in that direction such as Oregon with their Drug Addiction Treatment and Recovery Act in

November of 2020. With a more liberal population in Oregon, this act was widely supported and passing a similar act is far less likely in East Coast states such as New York where their drug related policies tend to be more conservative. As a data scientist, I believe that the continued work of researchers such as Charlie Marks are the only way to change this trend. Data speaks for itself and if more research and impactful data visualizations were presented to policy makers, I believe they would be more inclined to react to the out-of-control situation.