Project 1: Exploring Drug Overdose Deaths in Connecticut

Nikki La

3/31/2022

Introduction and the Problem

Residents of Connecticut are more likely to die from an unintentional drug overdose than from a car accident ("Opioids and Prescription Drug Overdose Prevention", n.d.). Overdoses of prescribed opiod painkillers and illegal drugs are responsible for a bulk of these deaths. Overdose deaths in Connecticut increased by 221 percent from 9.9 per 100,000 individuals in 2012 to 28.5 per 100,000 residents in 2018 ("Drug overdose deaths", n.d.). Furthermore, according to the Centers for Disease Control and Prevention (CDC), the age-adjusted drug-induced death rate in Connecticut in 2016 was 25.1 per 100,000 people, compared to 17.1 nationally ("Opioids and Prescription Drug Overdose Prevention", n.d.).

My project's goal is to look into drug overdose deaths in Connecticut, USA. People, families, and communities bear a significant weight as a result of severe drug use and overdoses. I'd like to investigate who is affected, their demographics, and the various medications that were abused more frequently utilizing the dataset I'm using.

The Dataset

The Kaggle website provided data on drug overdose deaths in Connecticut from 2012 to 2018. The download link for this dataset may be found at the end of the project. This data is broken down into 49 columns that anyone with no prior understanding of the subject may understand. There were additionally 5105 observations/rows in the data. As a result, this is a sizable dataset. This dataset achieved a readability score of 10.0 on the Kaggle website's readability scale (which is the highest and most easily readable type of dataset).

As shown above the rates in Connecticut are astounding. The data included describes drug overdose deaths that occurred from 2012 to 2018. To do this project I will explore the correlations between age, sex, race, location, year, and cause of death. Drugs correlating to death in this dataset include: heroin, cocaine, fentanyl, fentanyl analogue, oxycodone, oxymorphone, ethanol, hydrocodone, benzodiazepine, methadone, amphet, tramad, morphine, hydromorphone, opioids, and others.

Data Cleaning

I decided that I wanted one column instead of 16 columns that listed 0's and 1's showing that a specific drug was used (1 = yes, 0 = no). I simply used gather and subset to create one column that excluded the 0 values. I had many difficulties with this because of the "Other" column which listed various drugs within one column and it made it difficult to sort 1's on a single column. When I sorted the data it listed the various drugs along with all the 1's. After some time, it was realized that I could manually set the column to 1 since the dataset had already been sorted to include drugs that caused the death of the individual. I also used the drug_data dataset to make a new dataset that changed the blank cells to NA so that they would be easily changeable later. Additionally, I changed the date format and create new columns with the year so I could use it for future visualizations.

```
## code for sorting the drug columns into one column
## shows only drugs that have 1
drug_data1 = drug_data
drug_data1 = gather(drug_data, key = "drug", value = "yes", 22:38)
drug_data1 = subset(drug_data1, yes!= " " & yes != "0")
drug_data1 = drug_data1[!(is.na(drug_data1$yes) | drug_data1$yes == ""), ]
drug_data1$yes = 1

## code for converting the blank cells to NA
drug_data2 = drug_data
drug_data2 = drug_data2 %>% mutate_all(na_if, "")

## code for changing the date format
drug_data$Date = as.Date(drug_data$Date, format = "%m/%d/%Y")
drug_data$Year = format(drug_data$Date, format = "%Y")
```

A small basic summary of the drug overdose deaths in Connecticut are included below to give you a small look into the dataset I have choosen:

```
The sum of males that have overdosed on drugs:  3773
```

```
The sum of females that have overdosed on drugs:  1325
```

```
The mean for the ages in this dataset is:  41.96492
```

Exploring Basic Demographics

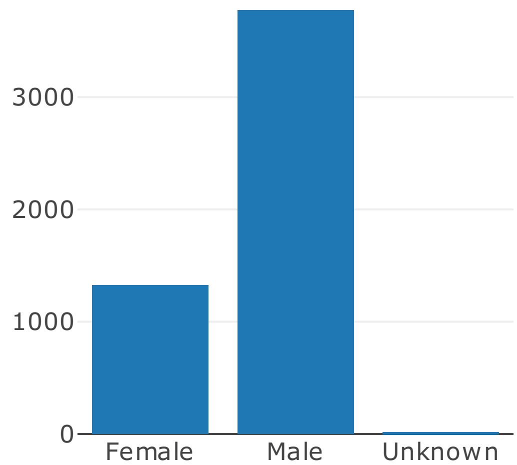
The goal of my project was to investigate the demographics of Connecticut drug overdose deaths in order to reach a conclusion. In this part, we'll look at some of important points about overdose deaths in Connecticut. This part is also really critical in allowing me to draw conclusions about drug overdose deaths in Connecticut.

When we look at the bar chart, we can observe that men had more than double the rate of overdose deaths than women (3773 men, 1325 females).

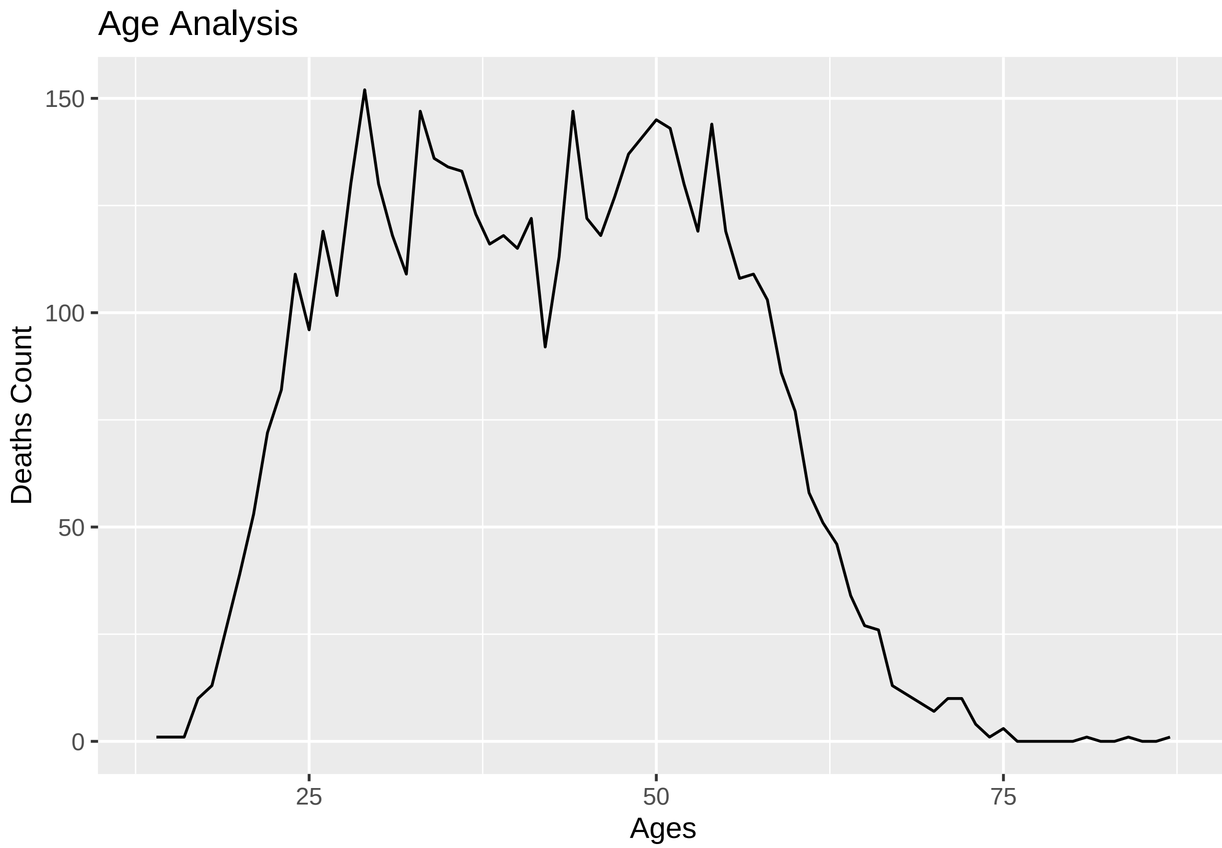
```
table(drug_data$Sex)
```

```
         Female    Male Unknown
      6    1325    3773       1
```

```
plot_ly(x = c("Unknown", "Female", "Male"), y = c(7, 1325, 3773), type = "bar", name = "Sex")
```

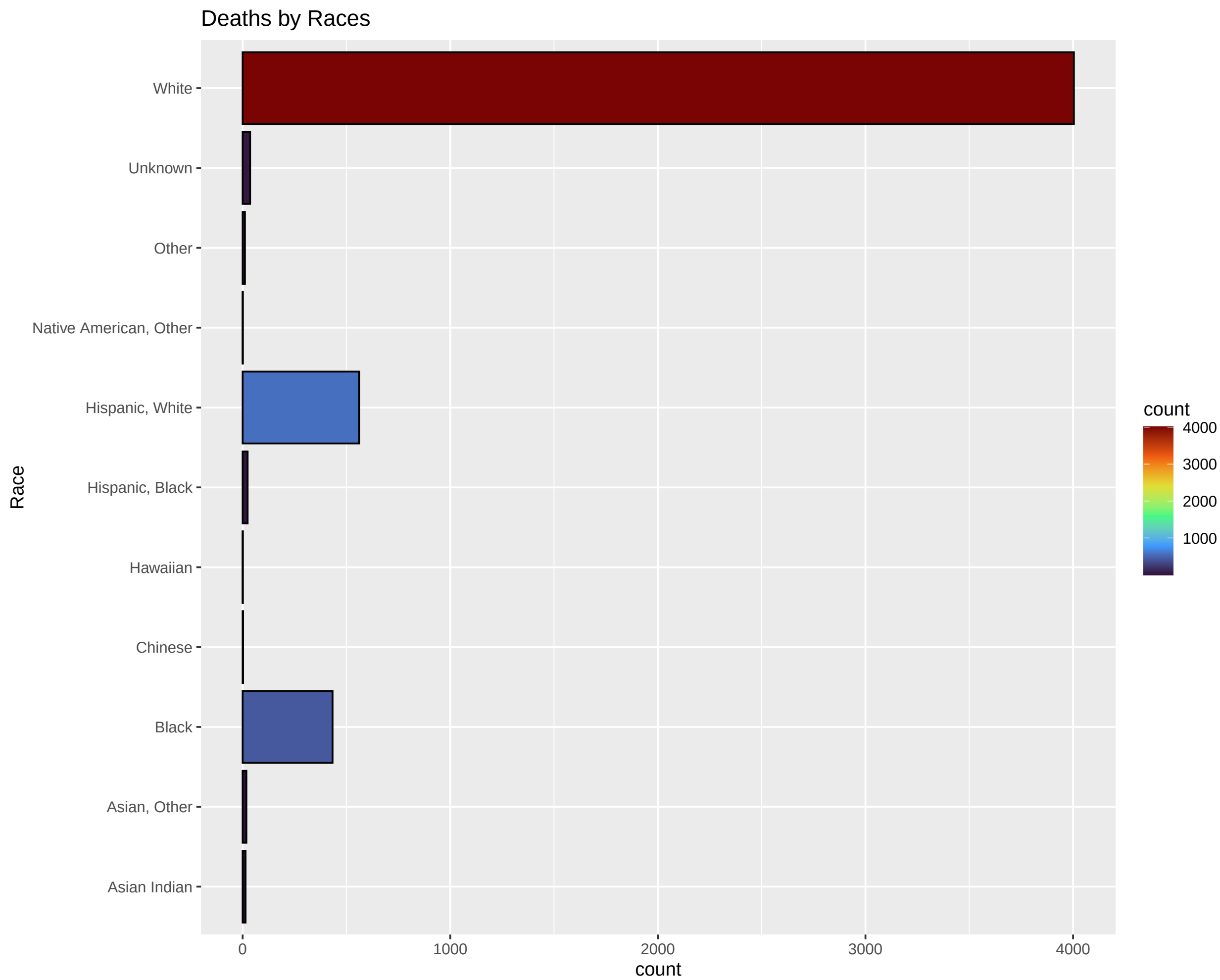



```
age_analysis = ggplot(drug_data, aes(Age)) + geom_line(aes(fill = ..count..),
stat = "bin", binwidth = 1)
age_analysis + labs(title = "Age Analysis", y = "Deaths Count", x = "Ages")
```



The line graph above shows a breakdown of the ages of people who have overdosed on drugs. Between the ages of 26 and 54, we notice a significant increase. The largest death rate occurs around the age of 29, when the death count exceeds 150. We can observe that drug overdoses are a problem within this large age group of 26 to 54.

```
drug_data2$Race[is.na(drug_data2$Race)] = "Unknown"
ggplot(drug_data2, aes(y =Race)) + geom_bar(aes(fill = ..count..),
stat = "count", color = "black") +
scale_fill_viridis_c(option = "H") + ggtitle("Deaths by Races")
```

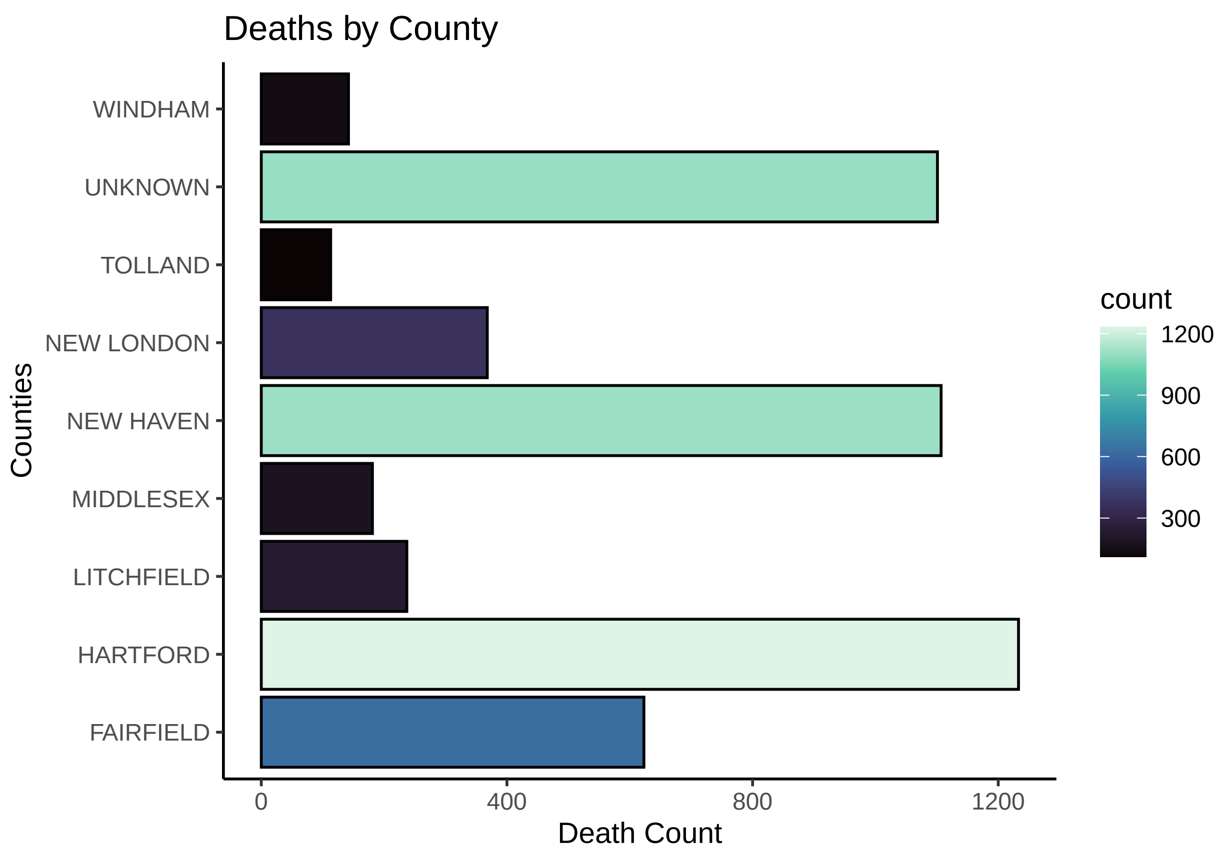



```
table(drug_data2$Race)
```

```
        Asian Indian           Asian, Other                  Black
                  14                     18                    433
             Chinese               Hawaiian        Hispanic, Black
                   2                      1                     24
     Hispanic, White Native American, Other                  Other
                 561                      1                     11
             Unknown                  White
                  36                   4004
```

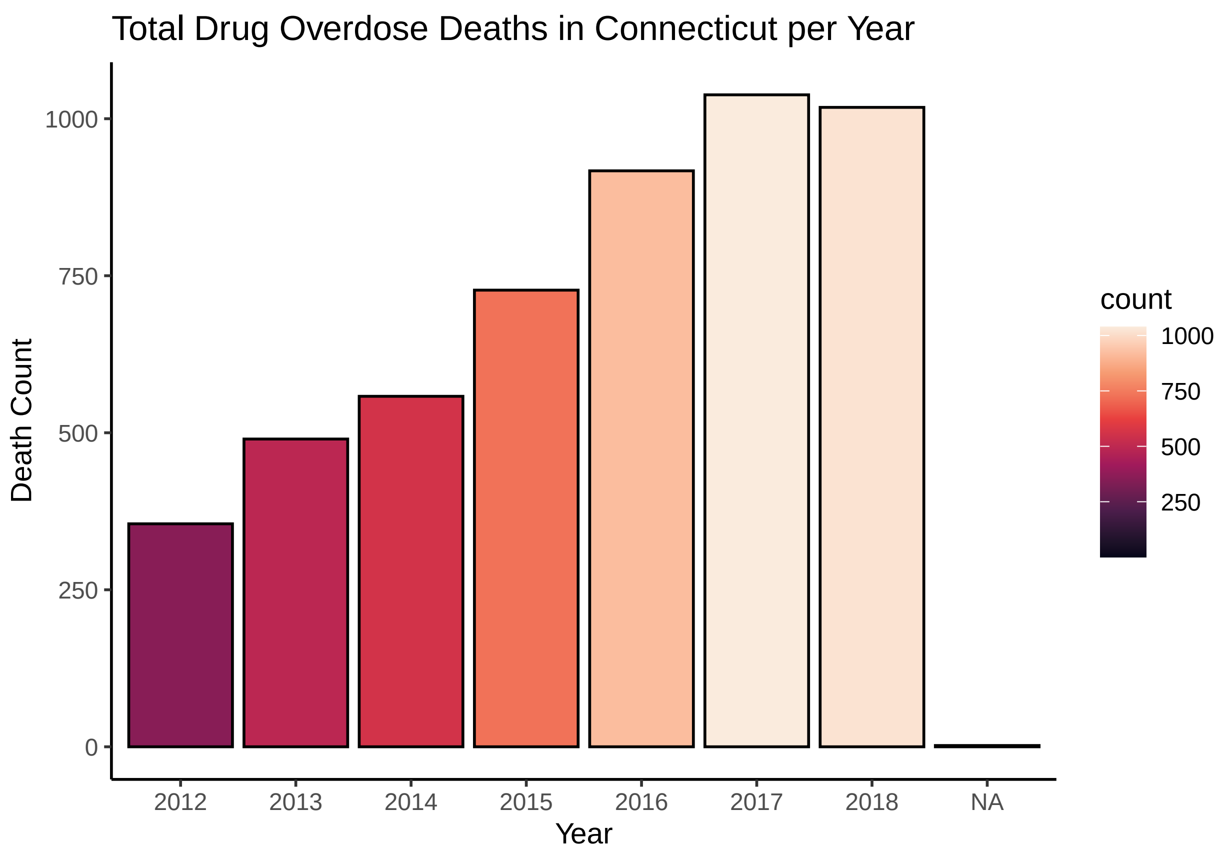
Additionally, we also notice a significant disparity in drug overdoses between races. In comparison to Hispanics, Whites, Hawaiians, Blacks, and others, Whites had 4004 deaths in the second bar plot.

```
drug_data2$DeathCounty[is.na(drug_data2$DeathCounty)] = "UNKNOWN"
drug_data2$DeathCounty = gsub("USA", "UNKNOWN", drug_data2$DeathCounty)
pia = ggplot(drug_data2, aes(y =DeathCounty)) + geom_bar(aes(fill = ..count..),
stat = "count", color = "black") +
scale_fill_viridis_c(option = "G") + theme_classic()
pia + labs(title = "Deaths by County", y = "Counties", x = "Death Count")
```



We can see that Hartford is the most populous county, with New Haven coming in second. As a result of the demographic distribution of the counties, we may conclude that the higher deaths in Hartford and New Haven counties are likely due to increased population size rather than increased drug use and death. However, there was a significant amount of null data in this section. In terms of deaths by county, unknown data ranks third, making it difficult to appropriately calculate deaths by county.

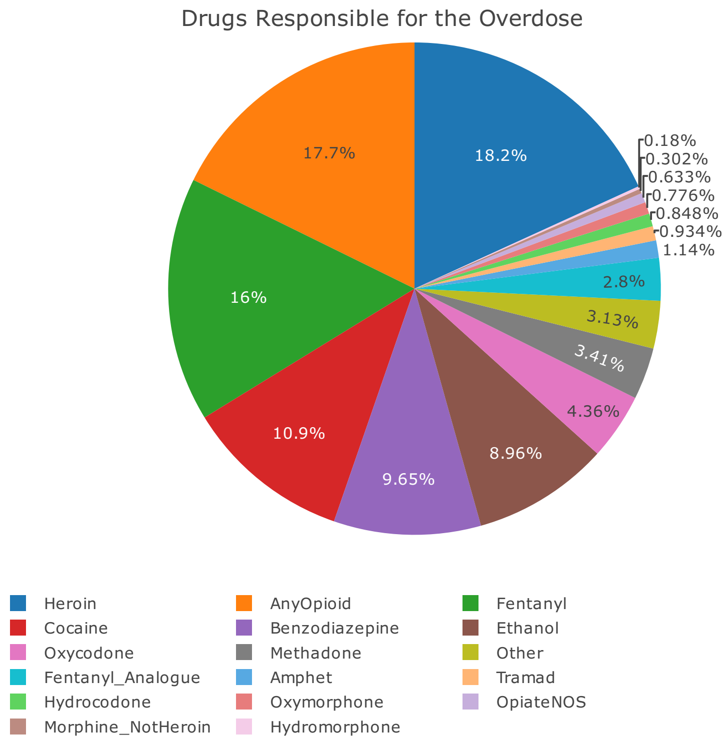
```
fig = ggplot(data = drug_data, aes(Year)) + geom_bar(aes(fill = ..count..),
stat = "count", color = "black") +
scale_fill_viridis_c(option = "F") + theme_classic()
fig + labs(title = "Total Drug Overdose Deaths in Connecticut per Year",
y = "Death Count")
```



The bar chart above shows the total number of drug-related deaths per year. As a result, we can see that the number of drug overdose deaths is rising. We can observe that the death rate was approximately 350 in 2012, but it has risen to nearly 1000 in 2018. We can also see that 2017 had the greatest death rate on the list, with a difference of around 20 deaths between 2017 and 2018.

Drug Analysis

```
fig1 = plot_ly(data = drug_data1, labels = ~drug, values = ~yes, type = 'pie')
fig1 = fig1 %>%
layout(title = "Drugs Responsible for the Overdose", xaxis = list(showgrid = FALSE,
zeroline = FALSE, showticklabels = FALSE),
yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),
legend = list(orientation = 'h'))
fig1
```



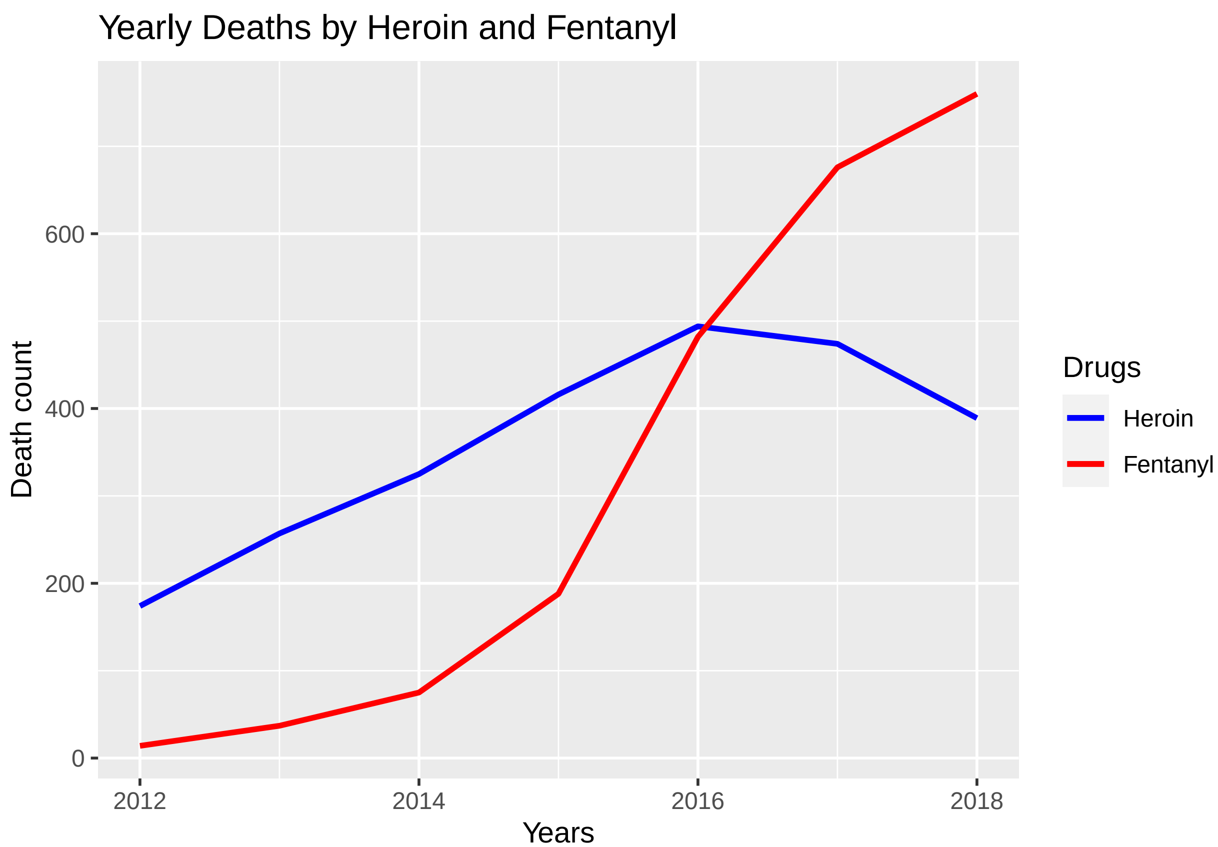
I decided to make a pie chart to look at the various drugs that people in Connecticut abused. There are 16 medications on the list, plus "others," which covers a smaller number of other drugs. We can see that three medications are more regularly used and are more frequently linked to drug overdose deaths. Fentanyl has a 16 percent prevalence, opioids have a 17.7 percent prevalence, and heroin has an 18.2 percent prevalence.

```
drug_data3 = drug_data1
drug_data3 <- drug_data3[grep("Heroin",drug_data3$drug),]
drug_data3 <- subset(drug_data3, drug_data3$drug != "Morphine_NotHeroin")
heroin_count = data.frame(year = c(2012:2018), heroin = c(174, 257,
325, 416, 494, 474, 389))


drug_data4 = drug_data1
drug_data4 <- drug_data4[grep("Fentanyl",drug_data4$drug),]
drug_data4 <- subset(drug_data4, drug_data4$drug != "Fentanyl_Analogue")
fentanyl_count = data.frame(year = c(2012:2018), fentanyl = c(14, 37, 75,
188, 482, 676, 760))
drug_table1 = left_join(heroin_count, fentanyl_count, by = "year")


colors = c("Heroin" = "blue", "Fentanyl" = "red")
```

```
drug_plot = ggplot(drug_table1, aes(x = year)) +
geom_line(aes(y = heroin, color = "Heroin"),
size = 1) +
geom_line(aes(y = fentanyl, color = "Fentanyl"), size = 1) +
labs(title = "Yearly Deaths by Heroin and Fentanyl", x = "Years", y = "Count")
drug_plot + scale_color_manual(values = colors) +
labs(y= "Death count", color = "Drugs")
```



When you're asked to name two common drugs, what comes to mind? I think of opiates like heroin and fentanyl. I included these two well-known drugs in the line plot above. Around 2015 to 2016, we notice a rise in the use of fentanyl. This indicates that the drug's popularity may be increasing. We also see that around the same time that fentanyl becomes popular, heroin starts to decline. This shows a negative correlation between these two common drugs.

Conclusion

The purpose of this project was to explore drug overdose deaths in Connecticut using visualizations and exploratory analysis to come to a proper conclusion. To accomplish so, I looked into the fundamental demographics of those in Connecticut as well as some drug analysis. With this information and numbers, I found that the drug overdose epidemic will continue to rise in huge yearly waves. I also observed that persons who overdosed on drugs were more often males, whites, and between the ages of 26 and 54. I didn't want to stop there; I wanted to learn more about the drugs that were commonly overused. I decided to build a pie chart using the information from the dataset and discovered that heroin, opioids, and fentanyl were the most often abused substances and were the leading causes of drug deaths. In conclusion, this analysis emphasizes the need of finding solutions to the drug overdose crisis. Drug overdose deaths can be avoided if the right precautions and measures are taken such as increased public awareness, treatment, and recovery support.

Links

Drug overdose deaths data: https://www.kaggle.com/datasets/ruchi798/drug-overdose-deaths/download
Connecticut drug overdose information: https://portal.ct.gov/dph/Health-Education-Management--Surveillance/The-Office-of-Injury-Prevention/Opioids-and-Prescription-Drug-Overdose-Prevention-Program