Apoorv Sharma

HCDS Course Project: A7 - Project Report

Introduction

The United States, alone, suffered several waves of COVID-19 with a total of 791,021 lives lost since early 2020. At the peak of the U.S. epidemic in January 2021, over 3,300 people died on average each day. A total of 49,161,522 cases have been reported, according to a Reuters analysis of state and county data. The widespread availability of vaccines in the United States helped to reduce cases in the spring and early summer. But the Delta variant has led to a renewed spike in cases and hospitalizations nationwide.

The virus has not only killed millions of people worldwide, but repercussions of the pandemic have also had a significant impact on everyone's lives. "The pandemic has resulted in a perfect storm of events that have each contributed to the increase in overdose death rates," says William Soares, MD. Stay-at-home orders led to profound social isolation, high unemployment, and increased stress at home — all factors that contributed to relapse for many people in recovery, Dr. Soares says. The pandemic has also limited access to addiction treatment for people trying to stop using drugs or maintain abstinence. According to a report by the CDC, Between April 2020 and April 2021, overdose deaths in the U.S. increased by 28.5 percent over the previous year, confirming many experts' worst fears about dueling crises in public health.

With the new variant of COVID-19, Omicron, already creating chaos worldwide, we are likely to see another wave and lockdown enforcements in the near future. As a result, it is extremely important to understand how various counties in the United States are reacting to the policy changes so that the new variant of COVID-19 can be addressed better and the side effects can be minimized.

In this report we analyze data from the start of the pandemic, from Philadelphia, Pennsylvania, related to COVID-19 cases and how the mask mandate changed the trajectory of cases. In addition, we also analyze one of the side effects of COVID-19: drug overdose. We analyze how the pandemic affected drug overdose cases.

We hope to discover interesting insights so that we can better our policies for the new potential lockdowns that might be seen in the near future.

Background/Related Work

As drug overdose is a common problem, there have been several papers and articles published about increase in use of opioids during the pandemic. One of the papers did a similar analysis, as my proposed analysis, in Rhode Island. The objective of this paper was to “to evaluate the rates and characteristics of deaths from drug overdose before vs during the COVID-19 pandemic”. The key finding of the paper was that “the rate of deaths from overdose in Rhode Island increased 28.1%” during the pandemic, with a p-value of 0.009. This was a very significant result. A similar research was conducted by the CDC for Cook County in Illinois. The study found that “the mean number of deaths increased to 35.1 per week (95% CI = 32.2–37.8), followed by a more pronounced increase during the 11-week stay-at-home order, with a mean of 43.4 weekly deaths (95% CI = 38.8–48.0). In the 18 weeks after the stay-at-home order was lifted, mean weekly deaths declined to 31.2 (95% CI = 28.6–33.9).”

In addition to formal papers, there have been several blogs and articles about how the “Opioid Crisis Worsened with COVID-19”. These papers and articles have strongly motivated me to understand the opioid crisis and push me to understand the effects of it in Philadelphia, Pennsylvania. Upon reading the research from several different sources, I came up with the following hypotheses:

We aim to analyze the following research questions in this paper:

1. How did masking policies change the progression of confirmed COVID-19 cases during the defined period ?
2. The number of drug overdoses cases during the COVID-19 pandemic is significantly higher than before the start of the pandemic
3. The age does not affect the dosage of Naloxone administered
4. The drug type does not affect the dosage of Naloxone administered

To analyze these questions, we make the following 2 simplifying assumptions:

1. The COVID-19 period will be defined from: February 1, 2020 through October 15, 2021
2. We only analyze data from Philadelphia, Pennsylvania

Methodology

The methodology that will be used to analyze each research question/hypothesis is stated below.

How did masking policies change the progression of confirmed COVID-19 cases during the defined period ?

For this research question, we will employ an inspection by visual analysis method. We will create 2 different graphs that show the cases and rate of cases during the masking mandate. Moreover, we will also plot +/- 3 standard deviations to see whether any increase or decrease in the rate of change of cases was significant. This methodology will aid us to understand the importance of the masking mandate.

The number of drug overdoses cases during the COVID-19 pandemic is significantly higher than before the start of the pandemic

There will be two main methods used to analyze this hypothesis. Firstly, two different graphs will be presented to visually analyze this hypothesis. This main reason the graphs are presented is because it allows the audience to grasp the information, visually draw out patterns and make their own inferences before analyzing the results from a statistical test. Moreover, data visualization sometimes reveals trends and patterns that otherwise would not have been found by statistical analysis. Next, a t-test will be conducted to analyze whether the difference between the means is significant or not. The significance level used will be 0.05, a standard value used in industry.

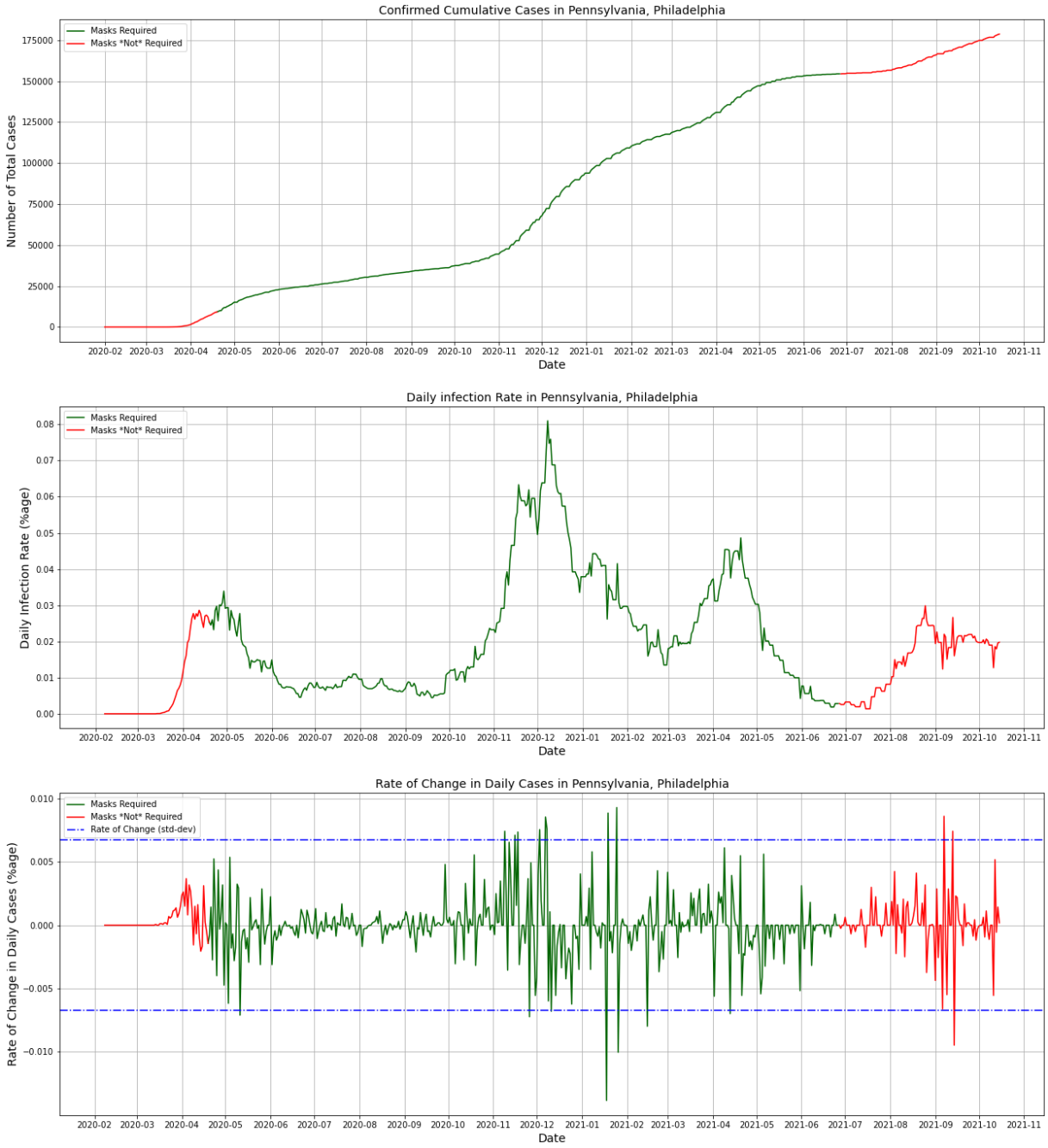
The age does not affect the dosage of Naloxone administered

There will be 2 main methods used to analyze this hypothesis as well. Firstly, a box plot will be visualized to show the audience the mean and distribution of the dosage for each age category. Secondly, an ANOVA test will be conducted to determine the p-value and the F-statistic. Allowing the audience to visualize the spread of the data before presenting the p-value is more human centered. The main reason is that it makes the analysis more ethical. The visualization aspect allows the user to internalize the data and make their own inferences before viewing the results of the ANOVA.

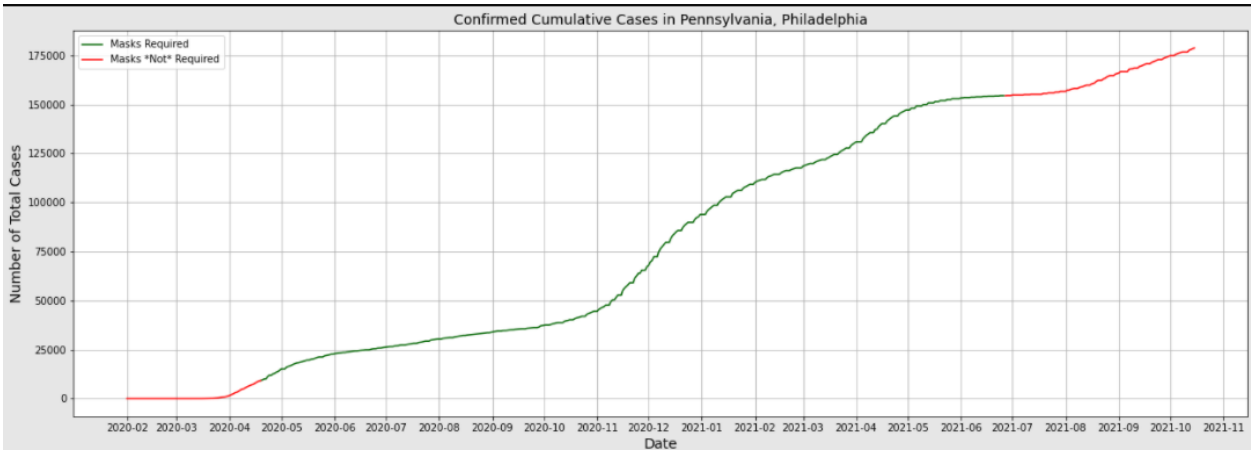
The drug type does not affect the dosage of Naloxone administered

This hypothesis will be tested in a similar fashion to the one above. However, a challenge associated with real-world data and running ANOVA tests is that the number of samples across the categories is not usually similar. As a result, if this happens to be the case, I will need to downsample or upsample some of my categories to ensure the test is valid.

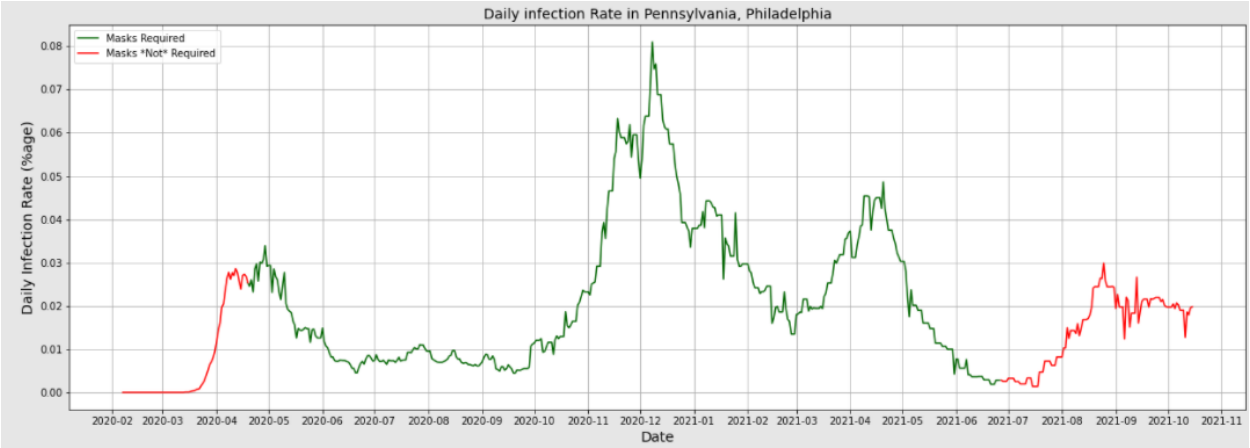
Findings



The line plots above depict my analysis of investigating how masking policies change the progression of confirmed COVID-19 cases from Feb, 1 2020 to Oct 15, 2021, in Pennsylvania, Philadelphia. Each subplot in the graph above is supposed to aid the user to gain a new perspective on how the confirmed COVID cases were impacted by masking policies. We now discuss each subplot.



The first subplot shows the number of cumulative confirmed cases over the specified time period. This plot is supposed to help the user to get a general idea of the progression of COVD-19 in the specified state and county. Moreover, the color encodes the make mandate policy; red being mask were not required and green being masks were required in public.



The second plot depicts a more granular level of information. Specifically, the subplot shows the infection rate of COVID-19 cases in the specified county and time period. Here, the infection rate was calculated using the following formula:

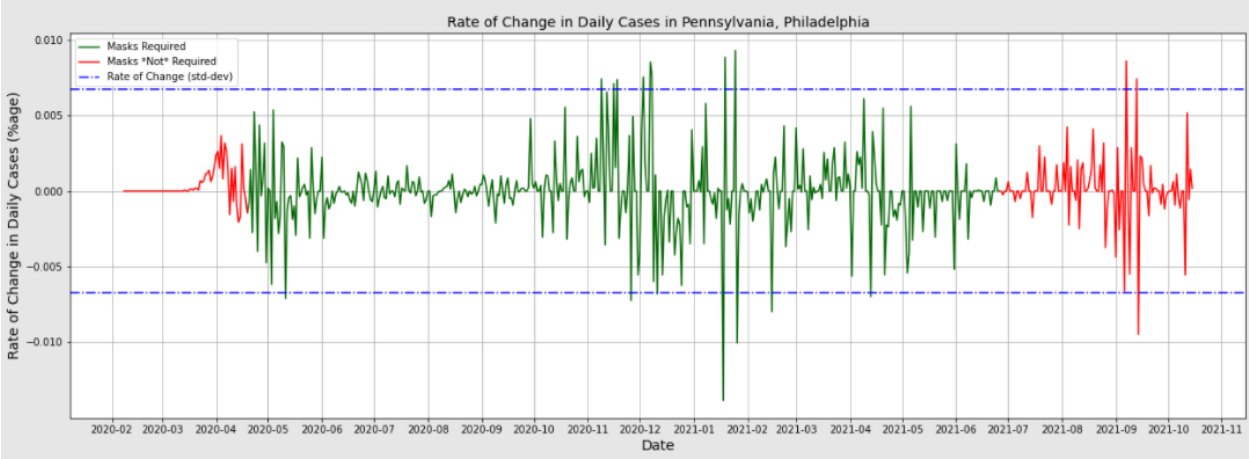
$$Infection\ Rate\ = \frac{7\ day\ moving\ average\ of\ daily\ cases}{Susceptible\ population}$$ where the 'Susceptible population' is defined as follows:

$$Susceptible\ population\ = \ Population\ (in\ 2021)\ \ - Cumulative\ Deaths\ to\ Date - \ Cumulative\ Cases\ to\ Date$$

The main objective of this graph is to allow the user to visualize how the rate was changing on a daily basis. As can be seen, when the mask mandate was enforced (April 2020), the rate of infection dropped and remained low until the end of November 2020. Moreover, once the mask mandate was lifted (July 2021), the number of cases started to rise again. Based on these two treds, one possible conclusion is that the masking mandate was able to slow down the infection

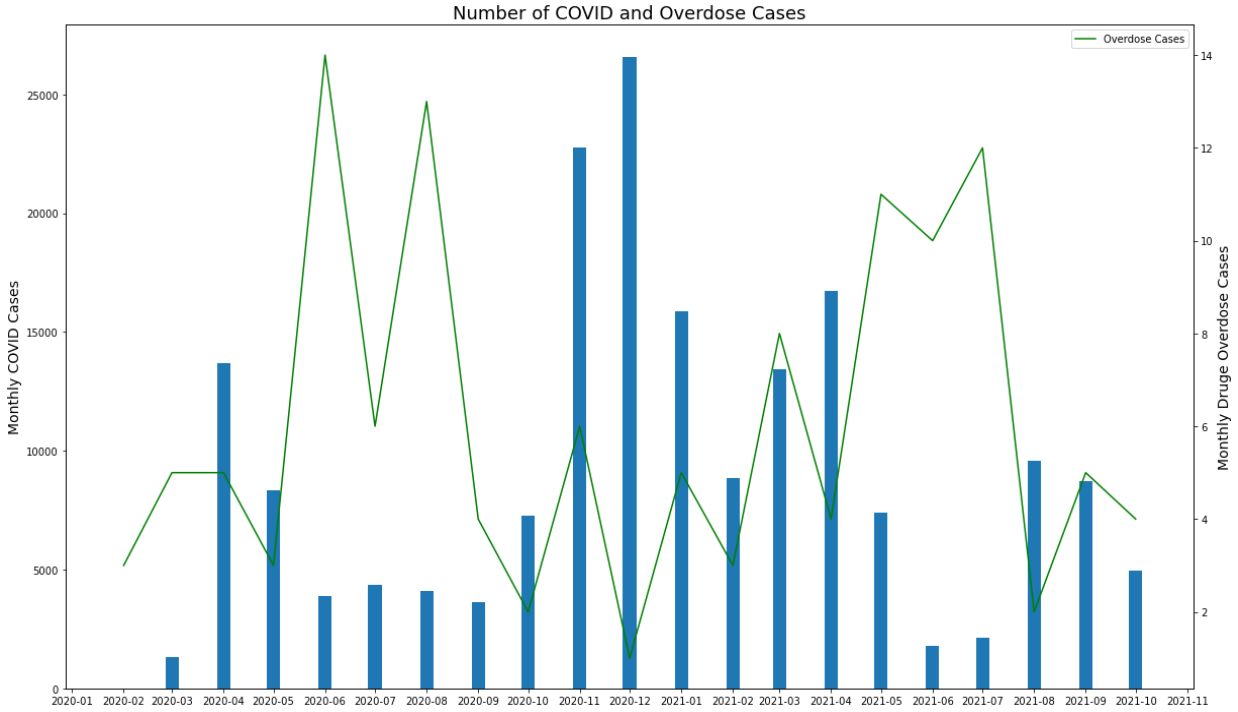
rate. There are, however, other factors that also helped slow down the rate. Potential factors include social distancing rules and lockdowns.

However, based on this graph, there is an anomaly. Even though the masking was required, between November 2020 to mid December, 2020, there was a rapid increase in the number of COVID-19 cases. This could potentially be explained by the increase in travel due to the holiday season and more people meeting each other.

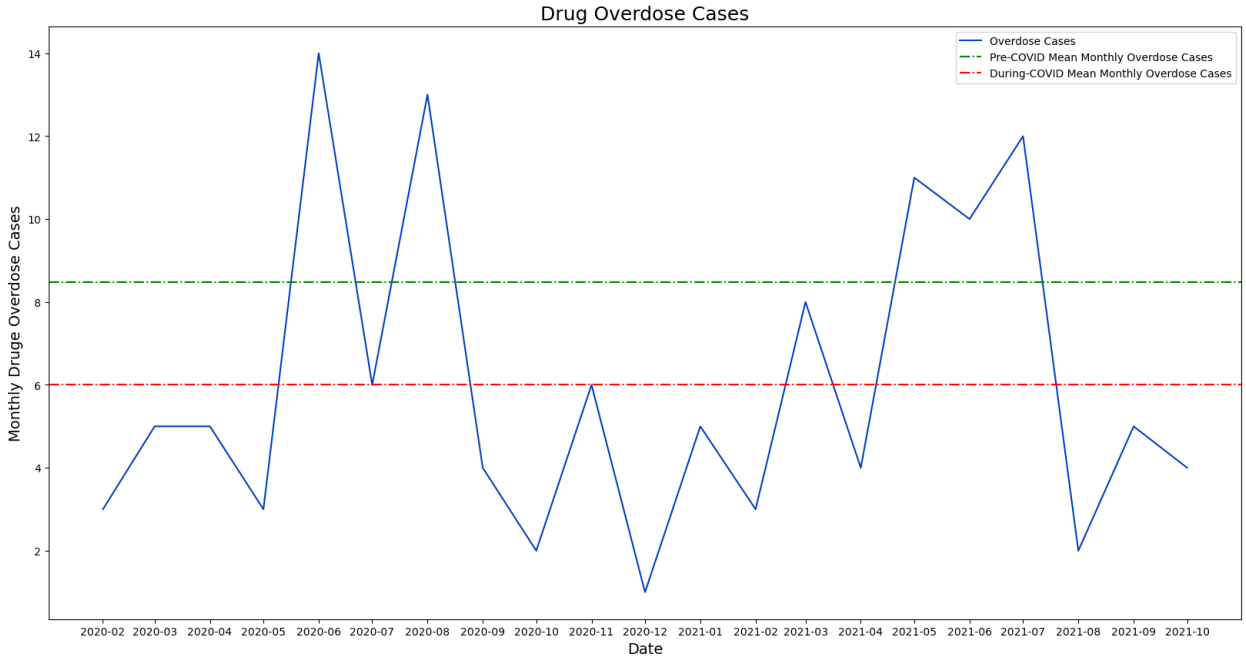


The last subplot is the second derivative of the number of cases. It shows the rate of change of new cases being added on a daily basis. This allows the user to dive deeper into the analysis and understand how ‘fast’ the new confirmed cases are being added. The graph also shows a 95% confidence interval (3 standard deviations), in the blue line. Using this information, one way to see what days had a significant number of cases (rise or fall) is by looking at the days where the rate of change of cases is outside the 95% confidence interval.

In regards to the second hypothesis, we analyze the following visualizations:



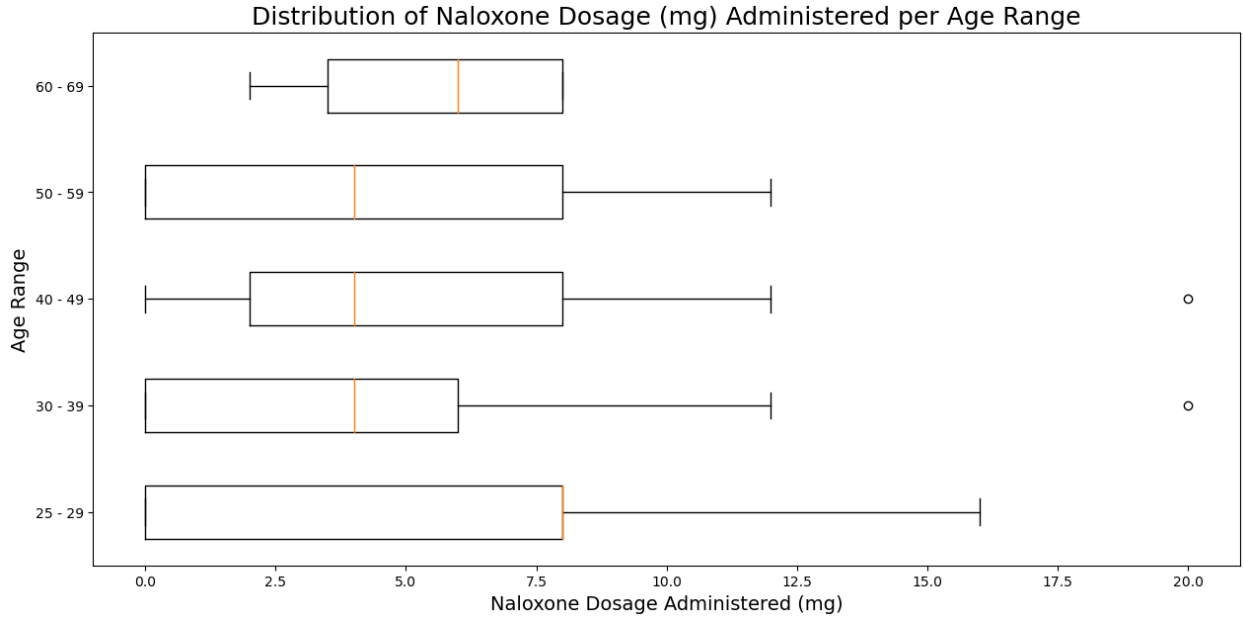
The figure above shows the number of monthly COVID-19 (left y-axis), represented by bars and number of monthly overdoses cases (right y-axis), represented by the line chart. Visually, it is evident that there is a delay in the number of COVID-19 cases peaking and the number of overdose cases following the same trajectory and peaking. The same can be said about the troughs of the chart. This was an extremely interesting find, one that might not have been revealed by a statistical test. One potential reason for this delay could be the enforcement and relaxation of stay-at-home orders. After an increase in COVID cases, the government most likely introduced lockdowns, thus increasing the loneliness factor. This may have led recovering addicts into relapse. Moreover, as the number of cases rose, the economy also suffered and people could have lost their jobs, thus leading to financial insecurity and depression.



This next figure, above, shows the monthly number of cases seen by the line chart. In addition, it also shows the mean number of monthly overdose cases before and during the pandemic, depicted by the green and red lines respectively. Visually it is evident that the number of cases did not increase during the pandemic; Rather, there was a decline in the number of cases. However, was the difference significant ?

We conduct a T-Test to statistically check the result and obtain a p-value of 0.06112. Comparing this to a significance level of 0.05, we can fail to reject the null hypothesis that there is no difference between the 2 means. However, this p-value is very close to 0.05 and more analysis can be done to confirm/deny the test results.

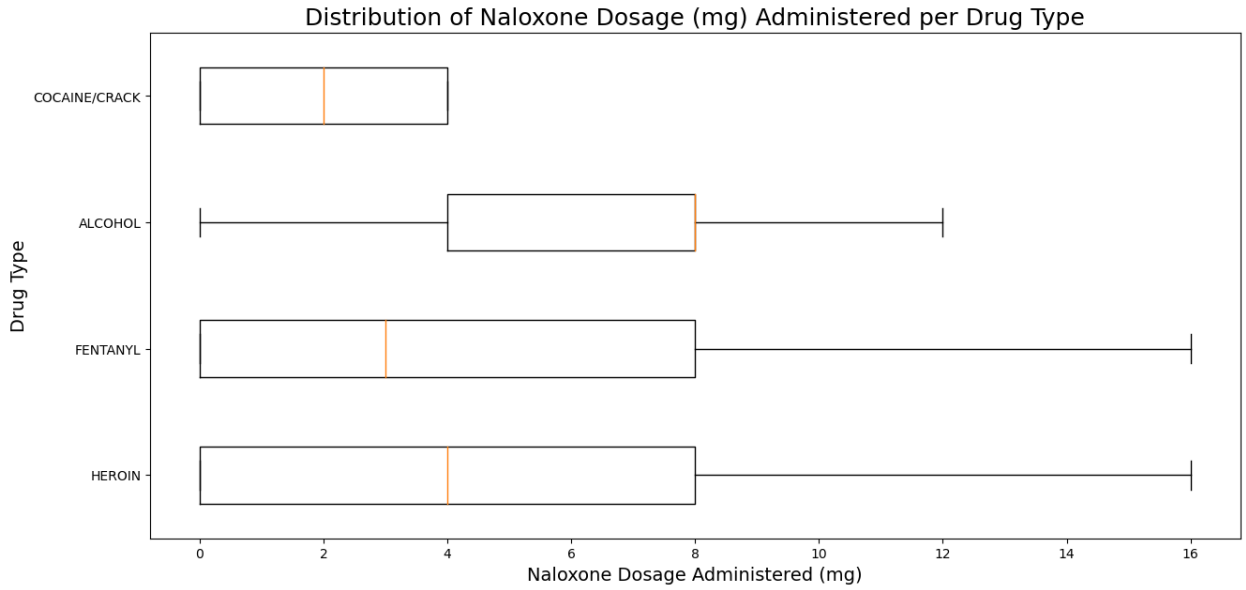
In regards to the third hypothesis, we analyze the following visualization:



Based on visual analysis, it can be observed that the means are relatively equal for all age ranges other than 25-29 and 60-69. The mean Naloxone dosage administered is slightly higher. Moreover, for 40-49 and 30-39, there seem to be large outliers where the dosage administered is extremely high. Additionally, we can also observe that the general trend of the dosage is downward as the person gets older.

We can further analyze this data by running an ANOVA. Conducting this test, we obtain a p-value of 0.6871, thus we fail to reject the null hypothesis.

For the last hypothesis we analyze the following visualization:



Based on visual analysis it can be seen that the Alcohol has a slightly higher dosage of Naloxone administered compared to the other drug types. However, Heroin and Fentanyl have outliers for the dosage administered. Based on this we can also argue that Heroin and Fentanyl are more lethal drugs as they need a larger dosage of Naloxone to reverse the opioid effect.

We can further analyze this data by running an ANOVA test. Conducting this test, we obtain a p-value of 0.49, thus we fail to reject the null hypothesis.

Discussion/Implications

Out of the 3 hypothesis tests that were conducted, one was marginally significant, whereas the other two were insignificant. As aforementioned, the following hypothesis had significantly marginal results: The number of drug overdoses cases during the COVID-19 pandemic is significantly higher than before the start of the pandemic. Based on the graph above, it was observed that drug overdose cases were **lower** during the pandemic than after the pandemic. This is completely against what other research has found and was extremely surprising. One avenue for further research could be to investigate why exactly drug overdose cases were lower during the pandemic than before the pandemic, where in the rest of the country, drug overdose cases seemed to rise.

The area of research that this project involved is linked to many principles in human centered data science. Firstly, the analysis of drug overdose is a huge concern to data privacy. Drug consumption is a very controversial and sensitive topic. Consequently, any data that is published about drug consumption is extremely sensitive. In the dataset, even though the latitude and longitude coordinates of the incident were reported, I ensure not to create any visualizations about this.

Another driving factor for this research project was the societal impact of government policies. One of the main reasons I decided to conduct this analysis and look into drug overdoses cases was to see how lockdowns, due to COVID-19, would impact society. My prediction here was that drug overdoses would increase due to lockdowns as people would become more lonely and depressed. The fact that the hypothesis proved otherwise was a surprise.

Limitations

There are several different limitations in some of the hypothesis tests that were conducted. Firstly, the number of drug overdose cases reported in Philadelphia was extremely small. Having extremely small samples may have caused invalid results. A potential improvement can be to increase the focus area to all of Pennsylvania, and thus increase the number of samples.

Another limitation, connected to the small sample size, is that some of the t-test and ANOVA assumptions were not completely met. For example, a main assumption of the ANOVA is that the samples must follow a normal distribution. With real world data and small sample sizes, this was not the case. As a result, this could be another potential flaw in running the test.

Conclusion

Throughout the paper we analyzed 4 different research questions/hypotheses:

1. How did masking policies change the progression of confirmed COVID-19 cases during the defined period ?
2. The number of drug overdoses cases during the COVID-19 pandemic is significantly higher than before the start of the pandemic
3. The age does not affect the dosage of Naloxone administered
4. The drug type does not affect the dosage of Naloxone administered

We were able to see that the result of the second hypothesis was marginally significant as the p-value for this was ~0.06. However, the remainder of the questions/hypotheses turned out to be insignificant.

Although the opioid overdose issue is not apparent in Philadelphia, Pennsylvania, it is not to say that opioid overdose is not a problem nationwide. Based on the various research papers and articles analyzed, this is clearly a problem in several different parts of the country.

I hope that this paper has highlighted and improved the understanding of human centered data science. Firstly, it is important to know that the policies that are implemented such as the mask mandate and lockdowns have side-effects that affect humans. This was seen in one of the figures, where there was a delay in the COVID cases and overdose cases. Moreover, this paper also highlights the various ethical and privacy concerns there are when talking about any data that is related to humans, especially when there is a sensitive topic involved.

References

1. Macmadu A, Batthala S, Correia Gabel AM, et al. Comparison of Characteristics of Deaths From Drug Overdose Before vs During the COVID-19 Pandemic in Rhode Island. JAMA Netw Open. 2021;4(9):e2125538. doi:10.1001/jamanetworkopen.2021.25538
2. Ama-assn.org. 2021. [online] Available at: <https://www.ama-assn.org/system/files/issue-brief-increases-in-opioid-related-overdose.pdf> [Accessed 10 December 2021].
3. Commonwealthfund.org. 2021. “It’s Really, Truly Everywhere”: How the Opioid Crisis Worsened with COVID-19. [online] Available at: <https://www.commonwealthfund.org/publications/podcast/2021/jun/its-really-truly-everywhere-how-opioid-crisis-worsened-covid-19> [Accessed 10 December 2021].
4. KFF. 2021. Drug Overdose Deaths Rose During the COVID-19 Pandemic, Particularly Among Black and American Indian/Alaska Native People. [online] Available at: <https://www.kff.org/coronavirus-covid-19/press-release/drug-overdose-deaths-rose-during-the-covid-19-pandemic-particularly-among-black-and-american-indian-alaska-native-people/> [Accessed 10 December 2021].

Data Sources

1. Overdose Information Network Data CY January 2018 - Current Monthly County State Police
2. COVID-19 data from John Hopkins University