

DATA 512: Final Report

Trevor Nims

December 14, 2021

Introduction

As the Covid-19 pandemic has raged on for far longer than what many of us had originally foreseen, it has caused both irreparable damage as well as phenomenal growth. This growth, is especially pertinent in analytical data, and the stoking of public interest in this type of data/analysis. This sort of public-facing analysis is what I sought to do in this project.

My analysis will investigate the relationship between Covid-19 and reported crime rates in Prince George's County MD. Although it would not be fair to claim that this analysis is an example of entirely original research, it would be fair to claim that it is in fact novel, and that it tackles an interesting issue using statistically-sound methodology.

Background

When the pandemic began - or rather, when organized government reaction to the pandemic began - many people confined themselves to their homes. As intuition would dictate, less people out-and-about would result in lower crime rates. This intuition was put to the test by USA Today's analytics team, who in April of 2020 published an article noting that crime rates had dropped dramatically throughout the month of March in 2020 [1]. This preliminary analysis interested me, and caused me to formulate the following hypothesis:

Are crime rates in Prince George's County MD lower during the pandemic than prior to it?

While the question was relatively open-ended, I hoped to use all tools at my disposal to answer it.

Methodology

After some initial data exploration, my plan of attack had taken form. First, I would split the reported crimes dataset into two timeframes, one with data

prior to Covid-19 and one with the data afterwards. To choose the date when Covid "began," I went with the first date a Covid-19 case was recorded in the John Hopkins confirmed case count dataset.

Now that I had a temporal partition of the dataset, I looked into increasing the granularity of my analysis. Each reported crime in the dataset had an associated label with it classifying its crime sub-type (examples include Accident, Theft, Homicide, etc.). Wanting to incorporate these labels into my analysis, I grouped the data by crime sub-type, creating temporally correlated pairs of crime sub-types I could use for comparison. However, simply the raw count data wouldn't work for the t-tests I planned on using, so I further aggregated the data by week, obtaining counts of each sub-type of crime per week for each of the temporal divisions.

Finally, I possessed data I was able to test; first, however, I needed to check my assumptions. Out of the gate, I made the assumption that the data in the two timeframes were independent of one another. Next, after examining the variances of my data and realizing that they were far from equal for each of the sub-type temporal pairs, I decided to change the test to a Welch's t-test that does not require the assumption of equal variance. Finally, I ran a normality assessment (using a Kolmogorov Smirnov test) on the distributions of all sub-type temporal pairs, removing from my analysis any pairs which contained at least one member that failed the normality assessment at a p-value of .05.

Once this process had been completed, I was able to run a one-sided Welch's t-test on each of the sub-type temporal pairs, comparing the means of reported crimes per week with the following hypotheses:

H_0 : Mean number of crimes reported by week of each sub-type were the same during and not during covid

H_A : Mean number of crimes reported by week of each sub-type was less during covid than not during covid

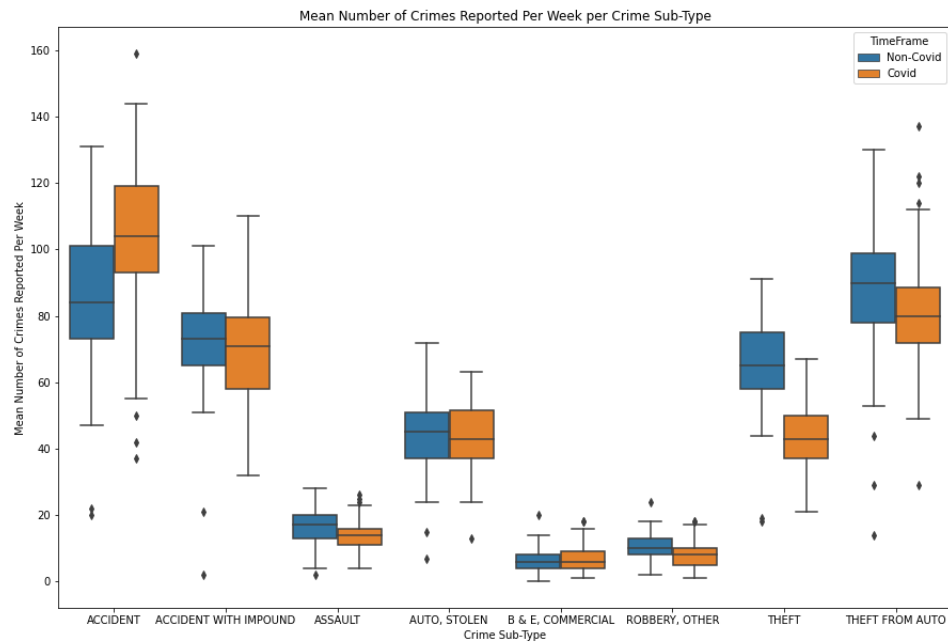
Findings

Ultimately, eight sub-type temporal pairs were tested, with the following results:

1) With a p-value below .05, the data suggests that the number of Accidents reported by week was greater during Covid-19 than prior.

2) With p-values greater than .95, the data suggests that the number of Accidents with Impounds, Assaults, Robberies (Other), Thefts, and Thefts from Auto were greater prior to Covid-19 than during. The distributions of each temporal sub-type can be further analyzed with a visual display, shown below

in the form of a boxplot:



As can be seen above, distribution pairs with significant p-values are often visually quite different as well.

Discussion

While this analysis does not do justice to the complex relationship between crime rates and Covid-19, it does shed some light on which kinds of crime in Prince George's county have increased or decreased during the pandemic. Future researchers could further granularize the research, perhaps focusing on one sub-type of crime and assessing both the large scale data at hand as well as conducting more personal research in the form of interviews with criminals, law enforcement, and (possibly) victims. This combination of big data and thick data would aid in creating a more holistic view of the issue, providing context for a more motivating story than the one told in this analysis.

Limitations

While the limitations of this analysis are myriad, there are a few that especially come to mind:

1. The choice of timeframes was (relatively) arbitrary. While using the Johns Hopkins first recorded Covid-19 case was an easy heuristic approach, a better analysis would have taken local government stay-at-home orders (among other factors) into account.
2. A number of crime sub-types were excluded from the analysis. While the initial dataset possessed over thirty different sub-types, the number that were eliminated through either the temporal partitioning (some crimes only occurred either prior to or during covid) as well as the normality testing suggest that a number of crime sub-types necessitate a different kind of analysis in concert with the one presented in this paper.
3. The changing distribution of crime reporting over time. While the analysis made the assumption that crime reporting did not differ between the two timeframes, this is most likely not the case.

Conclusion

While this analysis suggested the impact of Covid-19 on the rates of a number of sub-types of crime in Prince George's county MD, it by no means was a comprehensive study. Further research, performed in a more granular manner, would greatly intensify the meaning of the study and lend additional insight into the relationship between crime and Covid-19.

References

[1] Jacoby, K., Stucka, M., & Phillips, K. (2020, April 16). Crime rates plummet amid the coronavirus pandemic, but not everyone is safer in their home. USA Today. Retrieved December 15, 2021, from <https://www.usatoday.com/story/news/investigations/2020/04/04/coronavirus-crime-rates-drop-and-domestic-violence-spikes/2939120001/>

Data Sources

John's Hopkins Confirmed Covid-19 Case Count Data:

url: <https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university>

New York Times Masking Usage Survey Responses:

url: <https://github.com/nytimes/covid-19-data/tree/master/mask-use>

CDC Mask Mandate Data:

url: <https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i>

Prince George's County MD Crime Reporting Data:

url: <https://data.princegeorgescountymd.gov/Public-Safety/Crime-Incidents-February-2017-to-Present/wb4e-w4nf>