Spring 4-27-2022

An Exploratory Data Analysis on COVID-19 and its Effects on Crime in New York City

Lanlie Nguyen
lnguye@bgsu.edu

An Exploratory Data Analysis on COVID-19 and its Effects on Crime in New York City

Lanlie Nguyen

Honors Project

Submitted to the Honors College at Bowling Green State University in partial fulfillment of the

requirements for graduation with

UNIVERSITY HONORS

April 27, 2022

Dr. Umar Islambekov, Department of Mathematics and Statistics, Advisor

Dr. Philip Stinson, Department of Human Services, Advisor

Table of Contents

Executive Summary

The purpose of this study was to investigate the influence that the COVID-19 pandemic had on the crime rates present in New York City, specifically comparing the years 2019 and 2020. The COVID-19 pandemic is a recent event in which the first case reported in New York City occurred on March 1, 2020. There has been focus in the media on statistics regarding only the pandemic, and not how the pandemic has affected other aspects of the world, especially crime rates.

In this study, there will be data analysis applied to understand whether the pandemic had impacted the crime rates at all. The analysis of victim demographics showed what groups of people were more likely to be victimized. Looking at the frequency of crimes over the 24-month period showed whether crime rates have decreased at all in the presence of a global pandemic. It also showed whether specific types of crime had occurred more often. Locations of the crimes were found to determine where crime would happen more frequently as well. All this information can be utilized to inform the police department of where to distribute their resources more efficiently when another pandemic like COVID-19 arises.

Introduction

For the past two years, the COVID-19 virus has been the talk all throughout the world. Everybody can look at the news and find a new headline about how many cases there are, or what new protocols are in place to keep everyone safe. The pandemic has affected everyone in the country from the forced quarantine, businesses shutting down, or the vaccine mandate, this is an event that is impossible to ignore. Its effects are extremely widespread and have changed every aspect of the country. One specific aspect of the country that COVID-19 has changed is crime rates. With fewer people going outside, many places shutting down or closing temporarily, as well as many people being unhappy or dissatisfied with the changes and enforcement that have been taking place because of the pandemic, it is easy to see why crime rates would be significantly impacted by COVID-19. Since New York City has a lot of crime due to it being a large city with many people, seeing and analyzing the effects that the pandemic has had on it should be clear.

The purpose of this study is to explore how the COVID-19 pandemic has affected the crime rates in New York City. This study will be analyzing how much each type of crime occurred before and during the pandemic. By looking at how crime has been affected during a pandemic, we can use that information for future reference in the case that we end up in another similar situation, whether that be if we get into another pandemic with a different virus, or if another variant of the COVID-19 virus appears and causes another large outbreak. Some of the questions that this research study aims to answer are:

- What kinds of trends did the crime in New York present before covid?

- What kinds of trends did the crime in New York present during covid?

- How has the pandemic affected the crime rate, if at all?

- Have certain crimes been affected more than others?

- Has the statewide stay-at-home order affected crime rates?

- Are there certain demographics that are more likely to be the victim?

- Are there certain demographics that are more likely to be the suspect?

- Where is there likely to be crime in 2019 vs 2020?

- By analyzing this data, how can this prove to be useful in the future?

- What are some steps we can take to lower crime rates and protect people?

By analyzing how a pandemic can influence crime rates within a largely populated city, there is potential that this can assist police departments to prepare for when another similar crisis appears. The results can provide insight into where their equipment and resources can be allocated better during a pandemic. Providing this information can also be used as predictive evidence. Crime analysis can show where most of the crimes are being committed, who is most likely to be targeted, and how these trends might continue into the next few months. Having knowledge of all these factors can contribute immensely to the efforts of reducing crime rates overall, whether that be in New York City or in other extremely populated cities.

Review of Literature

Criminology and criminal statistics are not new to the world. Every day on the news, there is always a new statistic being brought to light about something related to crime. The release of policing data is an idea that is becoming quite popular, as it increases transparency and allows the public to identify policing practices that may be problematic (Sibley et al.). Transparency with the community is not the only factor that makes crime statistics useful. Crime data can also be utilized to analyze areas that are targeted more and assist police departments

with anticipating where crime might happen next (Walden University). Being able to investigate data on crime rates can contribute to keeping society safe, while also keeping people informed on what is happening around them. This is especially true during a pandemic where everyone's priority becomes safety. The COVID-19 pandemic was a sudden event, where most of the statistics were focused on the virus and how many people were being affected. While everyone was talking about how long they must stay indoors or how many new cases showed up at hospitals that day, there was not enough discussion on how much crime was going on and how to address it. One study has shown that there had been an increase in homicides and gun violence during the pandemic, as the authors urge for policing reforms as well as continuing strategies that have been effective in crime reduction ("NCCCJ - Impact Report: Covid-19 and Crime"). Another study found that grand larceny of automobiles as well as burglary had increased, and that most crimes had a significant decrease around March 2020 ("An Analysis of NYC Crime Trends from January-June 2020."). Another interesting event to look at is the statewide stay-at-home order. New York City officially shut down all public spaces and urged everyone to stay at home, unless you are an essential worker, on March 22, 2020, and continued until June 8, 2020 (Kerr). Many of the research articles that investigated COVID-19 did not mention much of the stay-at-home order and whether that influenced crime rates at all. Responding to and reporting illegal activity is important regardless of the situations the public might be in; however, how they should respond to illegal activities might change.

Exploratory data analysis can assist police and inform them on how they may more efficiently allocate their resources to respond quickly to criminal complaints and overall reduce crime rates over time. Currently, there is little criminal research that has been done that connects to pandemics and other natural disasters (Esposito et al.). This project should be an addition to

current research being done about the crime rates during the pandemic using methods discussed in textbooks that were used in my undergraduate classes.

Methodology

<u>Accessing Datasets</u>

To begin this project, I needed to find an open-source dataset that contained publicly recorded criminal events in New York City. I also needed to have access to a dataset that contained information on COVID-19 cases specifically in New York City. Data on the number of cases in New York City was not hard to find, considering it is a recent event and in these times, statistics on the pandemic are crucial to determining how to combat the virus. To gain access to data on the virus, the New York Times had released many files containing COVID-19 cases on GitHub. This dataset contained numbers of cases at the county level within the United States for the year 2020, which is the year that the first case was reported. This dataset contained 888,556 observations with 10 variables. New York City also has an open-source data website at data.cityofnewyork.us that contained datasets on criminal activities reported to the NYPD. The dataset I chose to use for this project contained complaints made to the NYPD for the years of 2019 and 2020, including all valid felonies, misdemeanors, and violation crimes. This dataset contained 862,435 observations with 35 variables.

<u>Cleaning Datasets</u>

The next step required before doing any sort of analysis is cleaning up the datasets. I will be keeping these datasets separate, so I will not be joining them together into a singular dataset. The program I will be using to both clean up and analyze the data for this project is RStudio, and

I will utilize a variety of R code. To start cleaning up the first dataset, the COVID-19 case reports, I will first label it as "covid" and filter out only the cases that are in New York City. After that, I need to identify which variables are necessary and which ones are irrelevant to this study. The variables I kept were date, county, cases, cases_avg, and cases_avg_per_100k. Next, I must filter out only data that is for New York City. Now, I am left with a dataset that has 5 variables and 306 observations, with each observation representing the number of cases within a day in New York City. Note that this observation only has 306 observations, and not 365, because this dataset contains data for every day after March 1, 2020. This was the very first day that COVID-19 had been reported in New York City.

Moving onto the other dataset, which I will be labelling as "police," the first step to cleaning this is to select the necessary variables that are relevant to this project. The variables selected include RPT_DT (the date of the report), CRM_ATPT_CPTD_CD (whether the crime was completed or attempted), ADDR_PCT_CD (the precinct where the incident occurred), BORO_NM (the borough where the incident occurred), OFNS_DESC (description of the offense), PD_DESC (description of offense corresponding with police department code), LAW_CAT_CD (level of offense), PREM_TYP_DESC (description of premises), SUSP_AGE_GROUP (suspect's age), SUSP_RACE (suspect's race), SUSP_SEX (suspect's sex), VIC_AGE_GROUP (victim's age), VIC_RACE (victim's race), and VIC_SEX (victim's sex). The RPT_DT variable type needed to be changed into date type, so I mutated the variable so RStudio would recognize the variable as dates rather than characters. Then, I filtered out only the crimes that were completed rather than attempted. Since this project will be based on the idea that this will contribute to future endeavors regarding where to allocate the police department's resources, we only need to look at completed offenses in which the police were unable to

interfere with the crime before it was completed. Then, we can take out the

CRM_ATPT_CPTD_CD variable, since it is unnecessary now that we know all the observations

within the dataset are completed. We are left with a dataset containing 848,478 observations with

13 variables.

Data Analysis

I continued to use R and RStudio for the data analysis portion of this project. The library

ggplot2 was useful for creating bar graphs and line graphs for data. Bar charts were incredibly

useful for visualizing interactions between the frequency of cases against many other variables,

such as those relating to demographics, as well as locations and the types of crimes committed.

Line graphs were helpful in investigating the frequency of observations over time. The ggplot2

library was also utilized to create charts that were visually appealing and easier to understand, by

adding color to separate data by year.

Results

Victim Demographics

Out of the 848,478 total criminal complaints made to the New York City police

department, 25.96% (n = 220,228) of the identified victims are Black, 17.81% (n = 151,092) are

White Hispanic, 15.75% (n = 133,656) are White, 7.65% (n = 64,917) are Asian or Pacific

Islander, 4.22% (n = 35,802) are Black Hispanic, and 0.46% (n = 3,902) are American Indian or

Alaskan Native, with 28.15% (n = 238,881) of the cases not identifying the victim's race. All

groups except Black Hispanics saw a slight decrease in cases, while Black Hispanics saw about
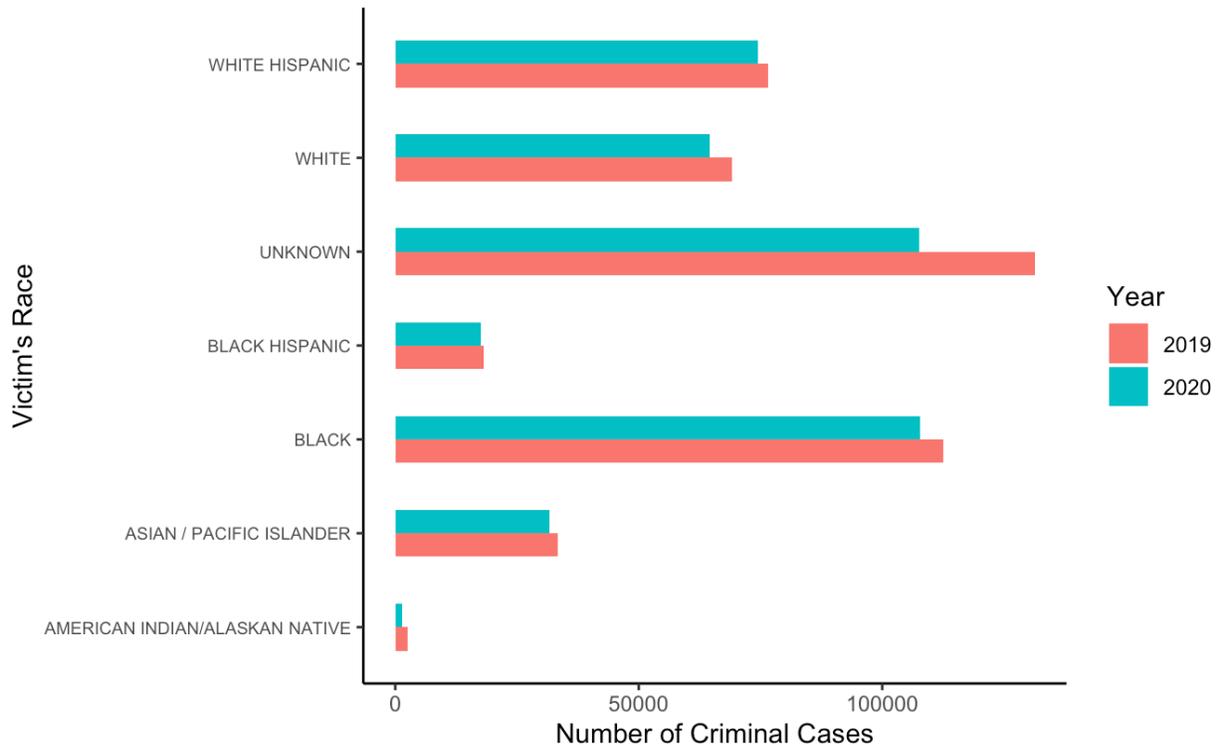
the same number of cases in 2020.

Figure 1: This bar chart displays the number of criminal cases (x-axis) and the victim's race of those cases (y-axis).

Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

As far as the genders of the victim's go, 35.85% (n = 304,146) are identified as male (M) and 39.77% (n = 337,475) are female (F). This dataset also identifies whether the victims were People of the State of New York, or if it is an organization or business under the gender variable separate from female and male. In this case, 14.66% (n = 125,306) were crimes committed against an organization or business (E), while 9.61% (n = 81,547) were committed against People of the State of New York (D), with only 4 cases that were unable to identify the victim's gender. All groups except People of the State of New York saw a slight decrease, while People of the State of New York remained about the same in 2020.
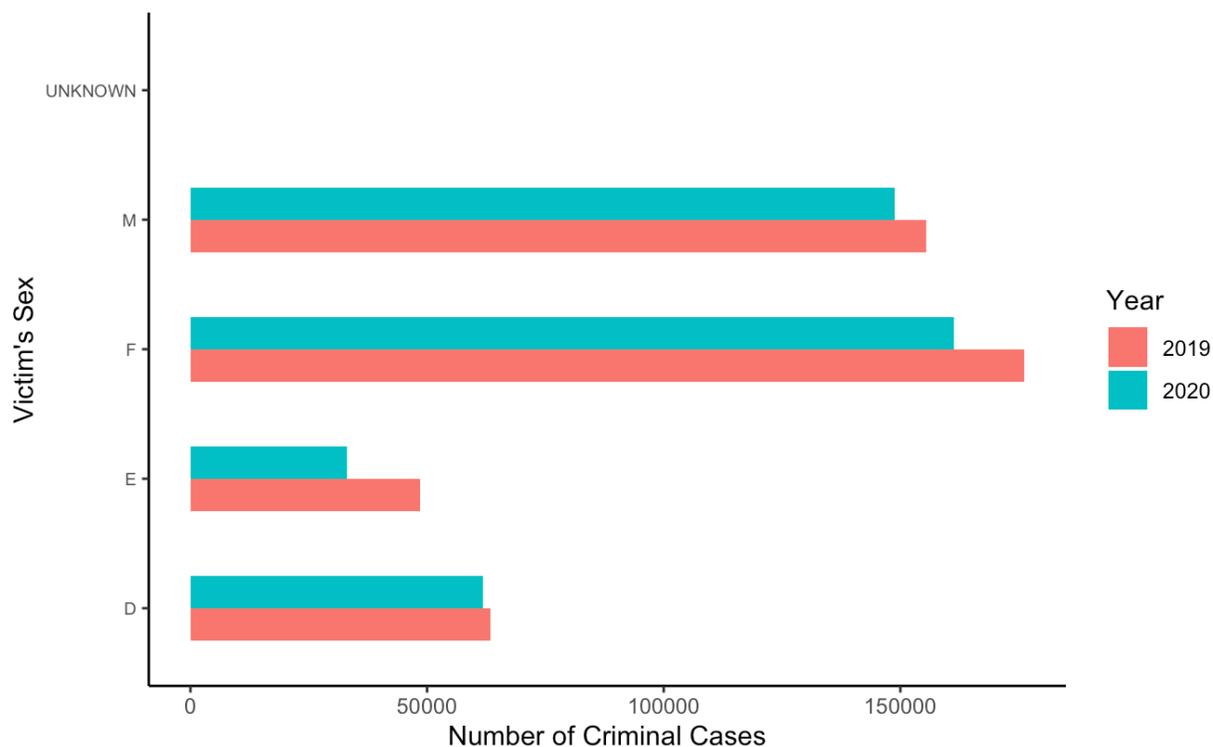
Figure 2: This bar chart displays the number of criminal cases (x-axis) and the victim's sex of those cases (y-axis).

Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

In terms of the victim's age group, 37.84% (n = 314,657) of the cases were against people between the ages of 25 and 44, 19.52% (n = 165,318) against the age group of 45 to 64, 9.36% (n = 79,544) against the age group of 18 to 24, 4.39% (n = 37,244) against victim's over 65 years old, and 24.88% (n = 221,098) cases of the victim's age group were unidentified. Those under 18, and between 18 and 24 saw a decrease in cases in 2020. Victims between ages 25 and 44, as well as above 65, had about the same number of cases. Those between the ages of 45 and 64 had more cases in 2020 than they did in 2019; however, this was a very slight increase of 0.03% (n = 289).
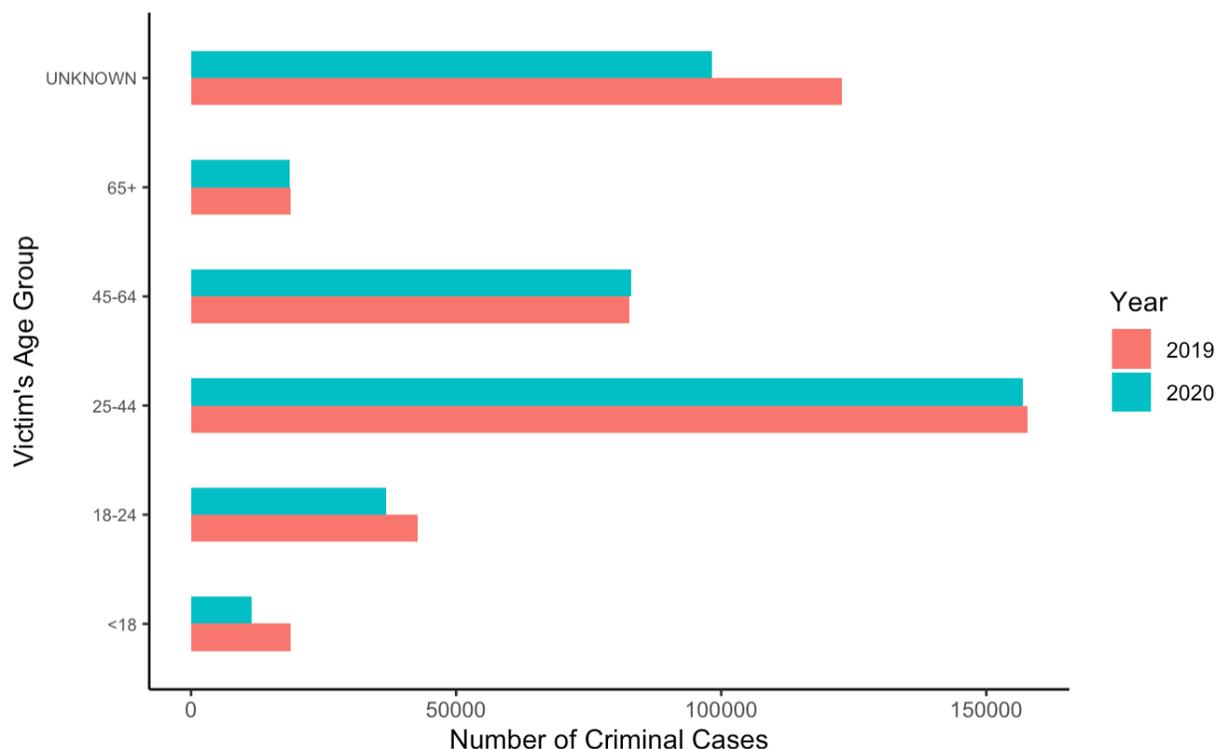
Figure 3: This bar chart displays the number of criminal cases (x-axis) and the victim's age group of those cases (y-axis). Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

Suspect Demographics

For racial demographics of the suspects in these cases, 28.45% (n = 241,396) were identified to be Black, 12.76% (n = 108,302) are White Hispanic, 7.33% (n = 62,223) are White, 4.05% (n = 34,399) are Black Hispanic, 2.79% (n = 23,634) are Asian or Pacific Islander, 0.20% (n = 1,684) are American Indian or Alaskan Native, and 44.14% (n = 376,840) of the cases are left unknown. Those who are Black seem to be both the most common victim and suspect group in terms of race within this dataset, but this may not be indicative of the population proportion of New York City. All groups saw a slight decrease in cases in 2020. There are also much more unknown cases of suspects than there are victims.
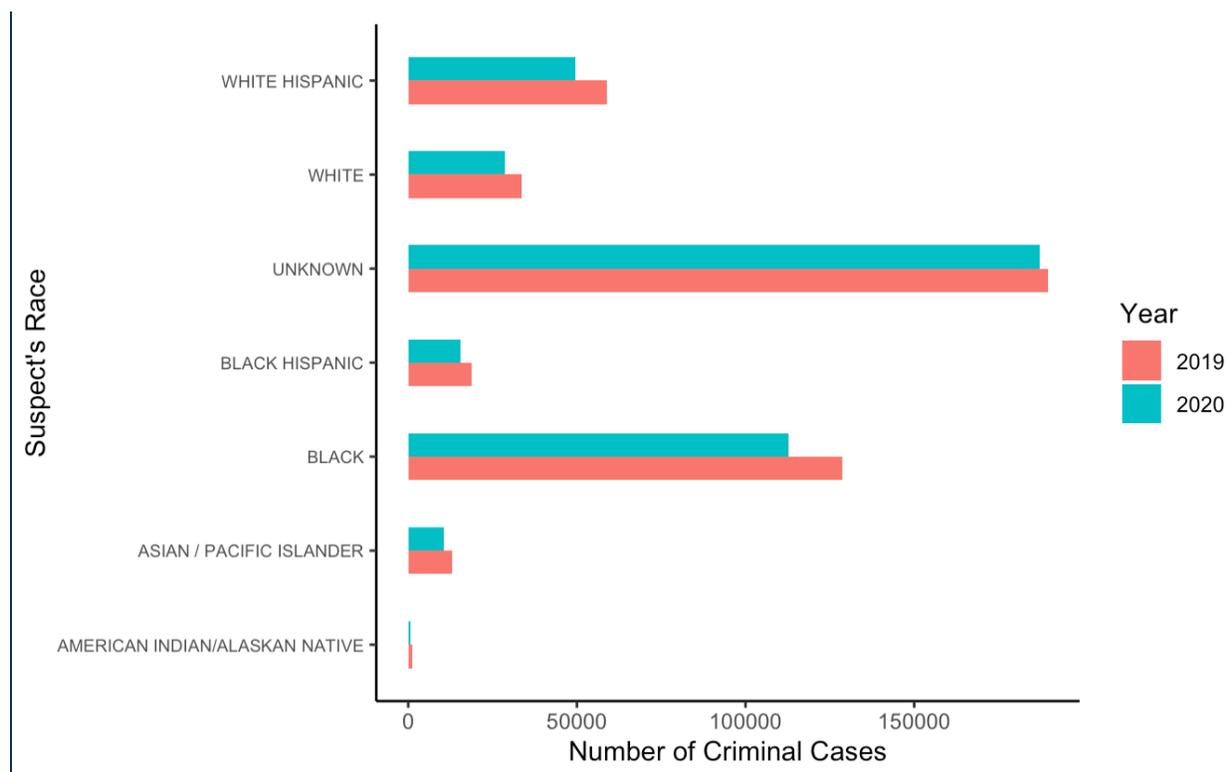
Figure 4: This bar chart displays the number of criminal cases (x-axis) and the suspect's race of those cases (y-axis). Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

The differences in demographics of victims and suspects becomes more apparent when comparing the genders. For suspects, 45.58% (n = 386,764) were identified as male, while only 13.17% (n = 111,785) of suspects were female, leaving 41.24% (n = 349,929) of cases unknown. While females saw cases of being victimized slightly more than males, males are overwhelmingly more common of a suspect. Both groups saw a slight decrease in cases in 2020.
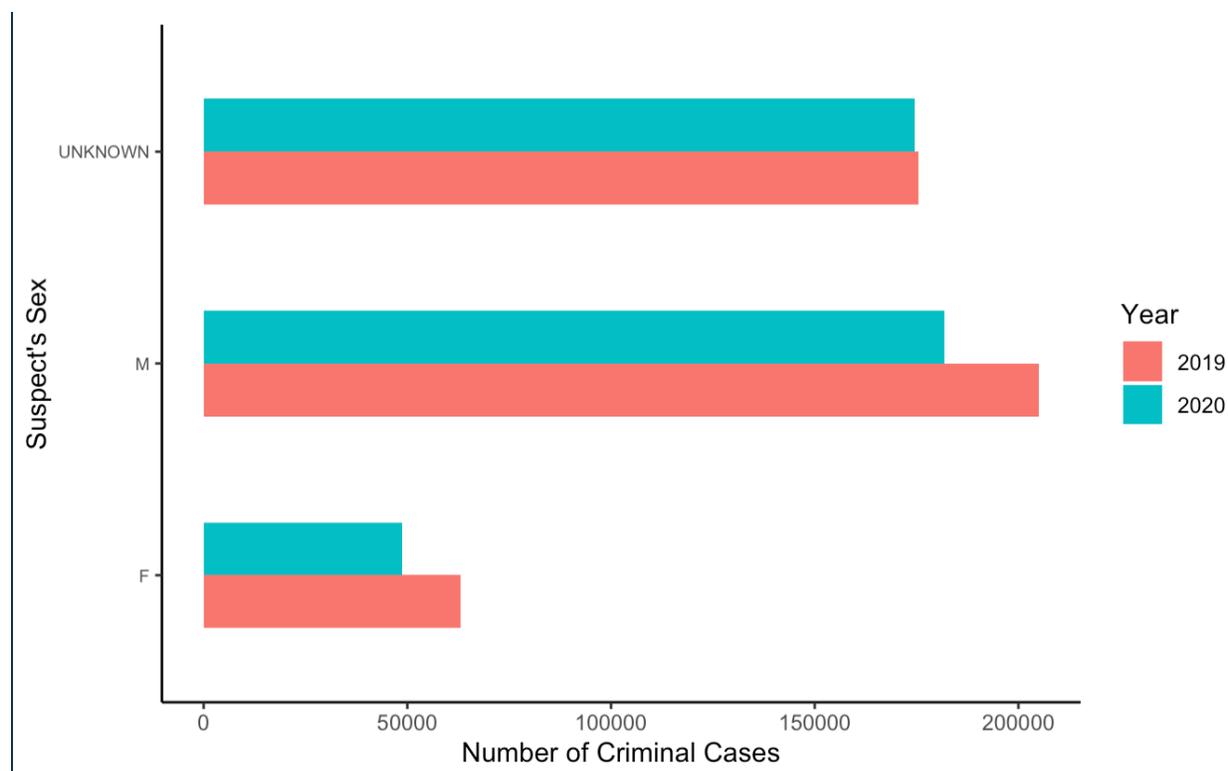
Figure 5: This bar chart displays the number of criminal cases (x-axis) and the suspect's sex of those cases (y-axis).

Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

Similar to the racial demographic comparison, the results of the age groups of suspects were not much different than the victims. Suspects within the age group of 25-44 took up 25.02% (n = 212,248) of cases, 45-64 took up 8.82% (n = 74,837) of cases, 18-24 took up 7.85% (n = 66,632) of cases, less than 18 took up 2.11% (n = 17,921) of cases, above 65 took up 0.84% (n = 7,132) of cases, and 55.35% (n = 469,663) are left unknown. All age groups saw a slight decrease in cases in 2020. Like victims, the most common age group to be a suspect is those between 25 and 44. The biggest difference is a significant increase in unknown cases for suspects, having 30.47% (n = 248,565) more unknowns than the victim's age demographics.
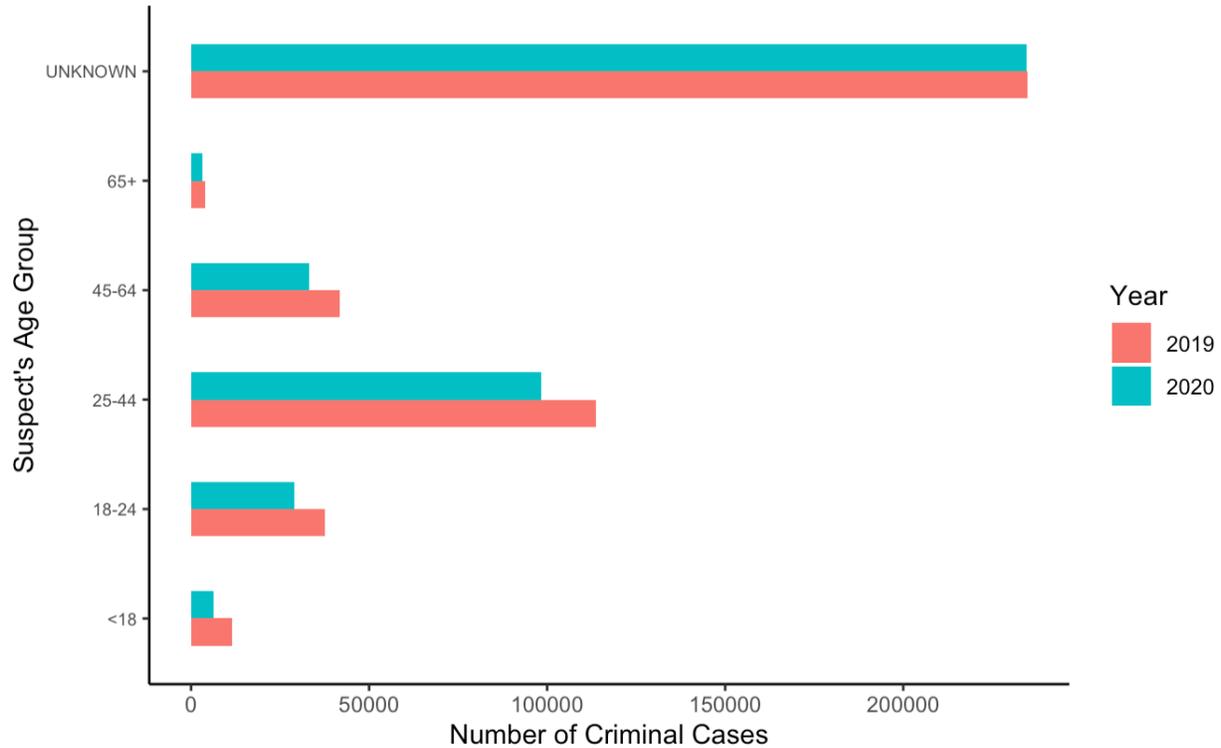
Figure 6: This bar chart displays the number of criminal cases (x-axis) and the suspect's age group of those cases

(y-axis). Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

Analysis of Covid and Crime Cases

The main topic of interest within this study was determining whether COVID-19 had an impact at all on the number and types of criminal cases committed in New York City. To begin this analysis, I looked at the general number of cases for both the police and covid datasets. In 2019, there were 443,493 criminal complaints (52.27% of total complaints) and 0 cases of COVID-19. In 2020, there were 404,985 criminal complaints (47.73% of total complaints) and 430,863 cases of COVID-19. This results in an 8.68% decrease in total criminal complaints in the year 2020.
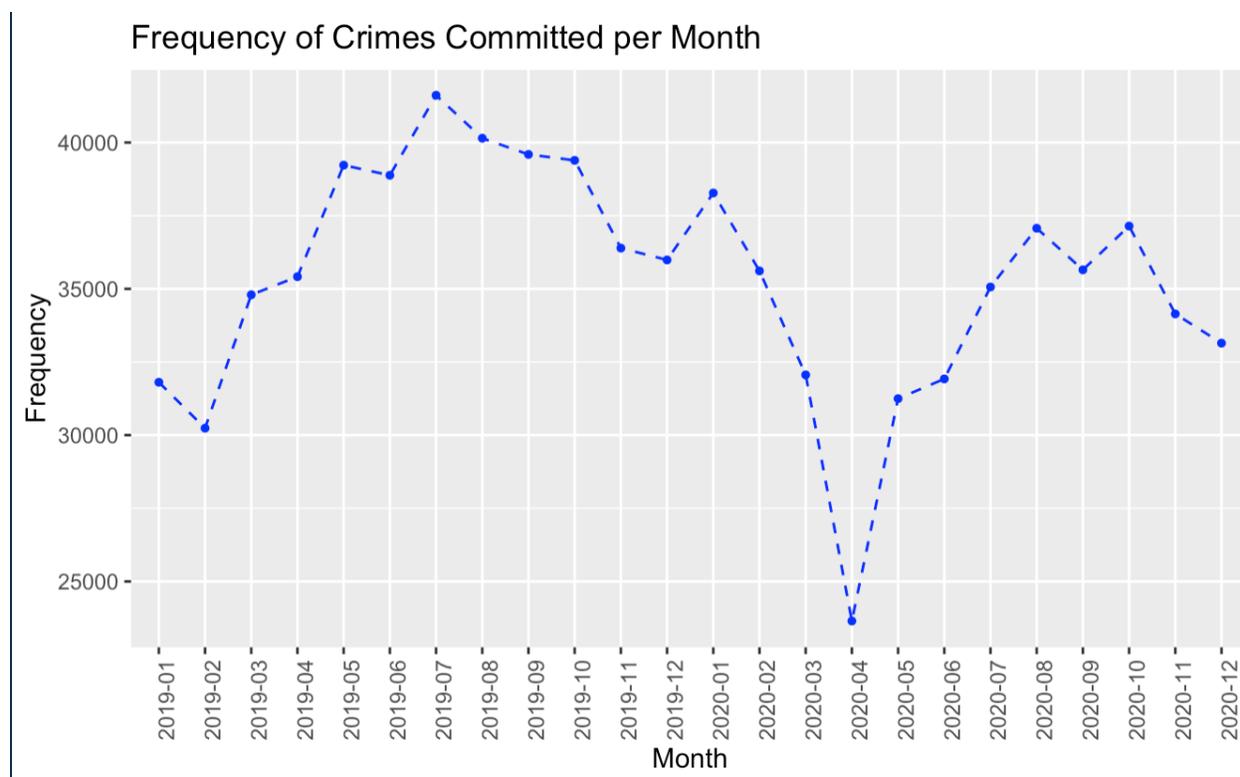
Figure 7: This bar chart displays the month (x-axis) and the number of cases (y-axis) that have occurred for that month.

Over the full 24 months, April 2020 had the least number of criminal cases, but also had by far the most covid cases. The month of April in 2020 had 2.79% (n = 23,649) of total criminal complaints, with 30.00% (n = 129,266) of the total number of covid cases. This was the only momentous change that had occurred within the police dataset when comparing it to the covid dataset when it comes to the number of cases. COVID-19 cases had a dramatic fall in cases in the month of May 2020 and continues to decrease for the next couple of months, which could have been due to the statewide stay-at-home order. The statewide stay-at-home order, however, did not stop the increase of criminal cases after April, which saw a rise in cases for each month between May and August. There is a trend in which the rise of COVID-19 cases also relates to the drop in criminal cases, and vice versa, during the months of the statewide stay-at-home order.

It seems the rate of crime in 2019 increased around summer and fall, while decreased when it

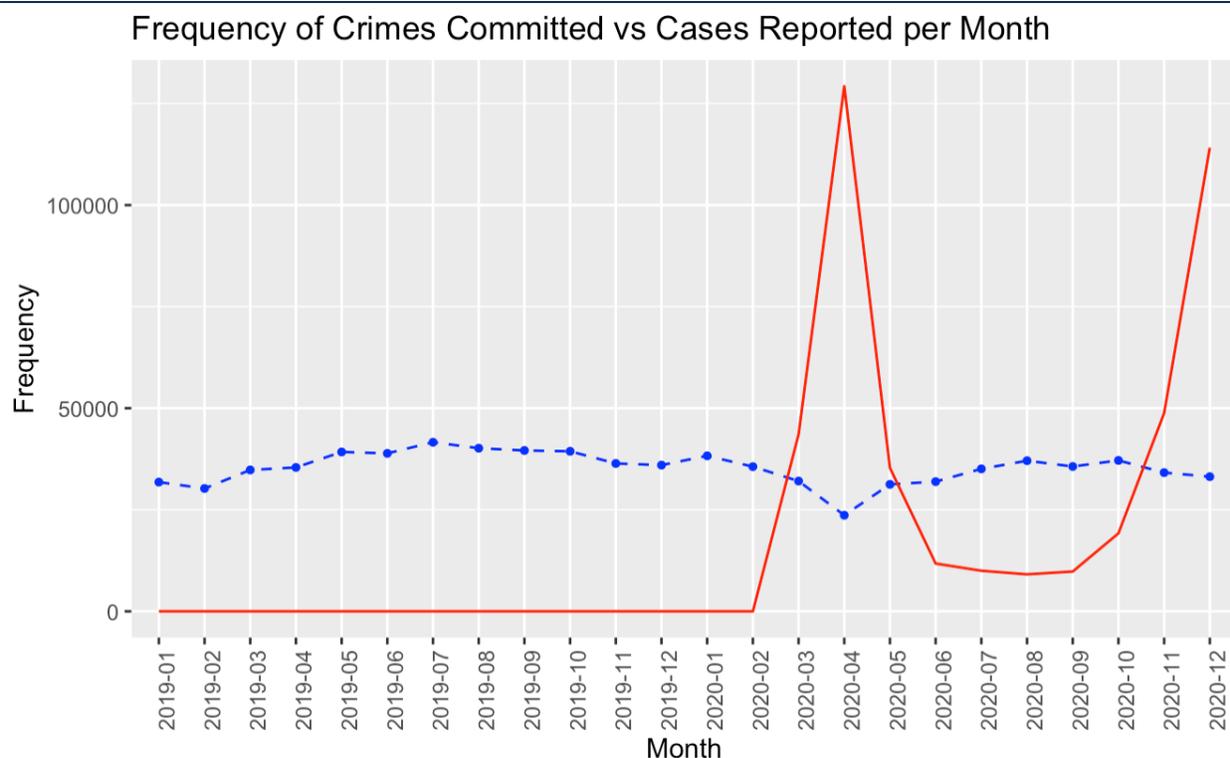became winter. This trend holds true into 2020, with the exception of a dramatic drop in April.

Figure 8: This bar chart displays the month (x-axis) and the number of cases (y-axis) that have occurred for that

month, both criminal and covid. The blue dashed line represents the criminal cases, and the red solid line represents

the covid cases.

While the number of cases may not have overall changed significantly, this does not

address the other topic of interest, which is whether specific types of crime have been affected by

the pandemic. Due to there being a substantial number of several types of crime occurring in

New York City, I separated and analyzed the top 20 most common crimes committed. For most

of these, there was a slight decrease in the number of offenses in 2020, which follows the earlier

observation of there being overall less offenses in 2020. Some others, however, showed an

increased number of offenses when compared to 2019. These include grand larceny of motor

vehicles, criminal mischief and related offenses, and burglary. Compared to 2019, grand larceny of motor vehicle offenses increased by 70.04% (n = 3,663), criminal mischief and related offenses increased by 1.89% (n = 868), and burglary offenses increased by 48.15% (n = 4,724).
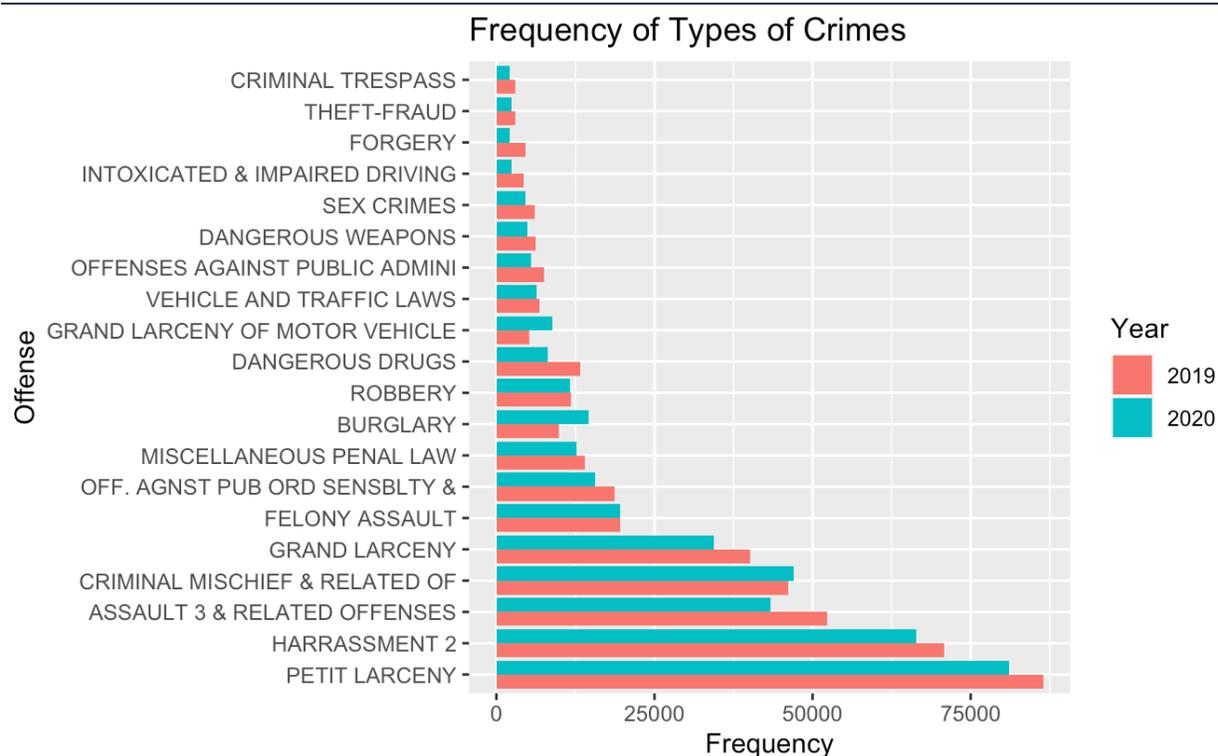

Figure 9: This bar chart displays the number of cases (x-axis) and the type of criminal offense (y-axis). Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

The last topic of interest is seeing where crime has occurred the most, and whether that has changed in the presence of a pandemic. In all boroughs of New York City, the frequency of crimes in 2020 has decreased. Brooklyn continued to have the greatest number of offenses in both years, with a total of 28.77% (n = 244,087) of offenses. Queens and Staten Island showed minor change in numbers.
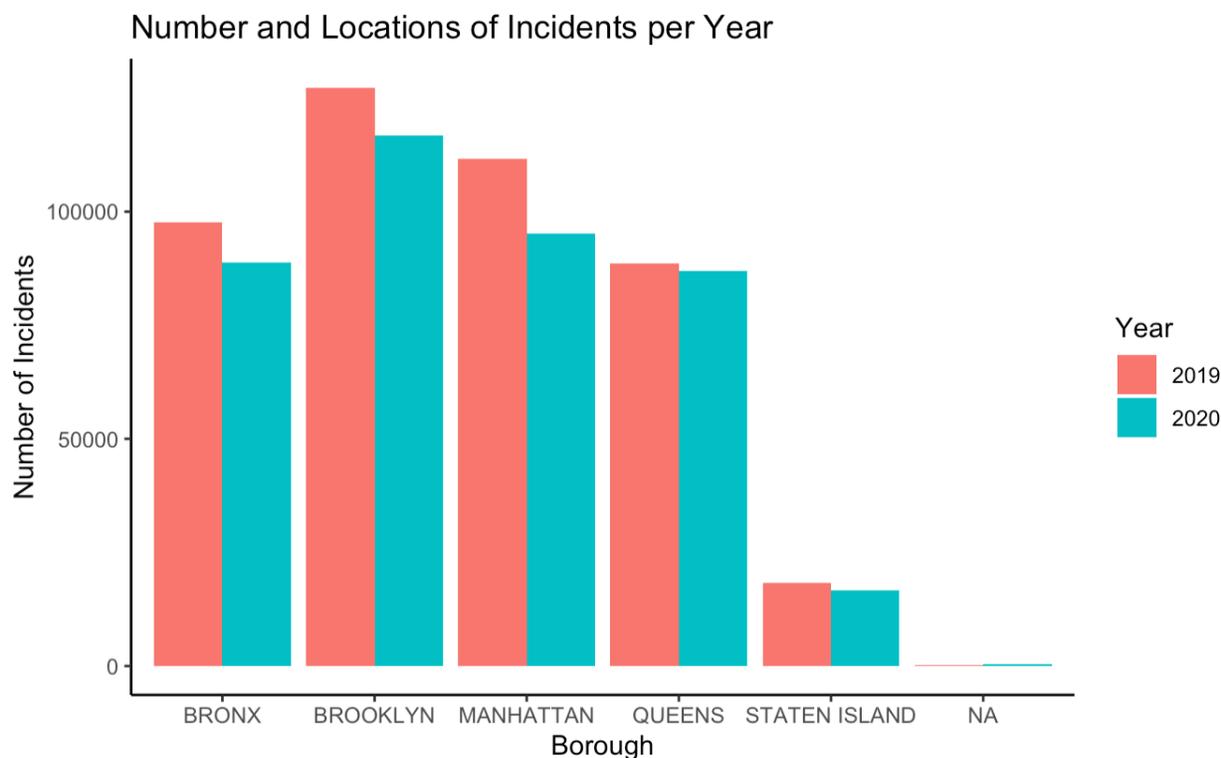
Number and Locations of Incidents per Year

Figure 10: This bar chart displays the borough (x-axis) where an offense has occurred and the frequency of offenses (y-axis). Bar chart colors are colored by the year, with red representing 2019 and blue representing 2020.

Limitations

One limitation of this study is that the datasets used are specific to New York City. The results of this research may not be the same in other parts of the world. Considering every city within the United States would become a significantly much larger task, let alone the entirety of the world. The police data set already contained over 800,000 cases just for New York City, so including more large cities would be difficult for my current technology to handle. There could also be several factors within New York City that could have contributed to the results found that may not be indicative of all cities. Each city has a different culture, groups of people, and many more factors that are not included in these data sets, therefore I would not be able to fully

analyze why some cities might be affected more than others, or what causes why some crimes occur more often. There also may be factors about the pandemic itself that may contribute to the trends presented in this study. Every pandemic may have its own unique factors, and this study does not go in-depth about what could be unique to the COVID-19 pandemic that does not apply to other pandemics.

Another limitation is that this dataset only looks at the years 2019 to 2020, the start of the COVID-19 pandemic. This does not include more recent activities of the pandemic, including the appearance of new variants such as Omicron. If I were to continue this study to include up to the most recent dates of 2022, this project would be nonstop. Since this study does not include the entire pandemic, any insights regarding the periods where a different variant had occurred, and may have caused an increase in cases, are not considered in this study.

Regarding the demographics of the victims in this study, I was only able to look at the demographics of the victims in the police dataset. This study was not able to compare it to the overall population proportions within New York City, and only observes what groups of people have been more often a victim in these cases. There were also many instances where the data sets were incomplete. A large number of demographics of the victims were listed as "UNKNOWN," which could have affected the data analysis greatly if these had been identified.

Conclusion

All the data used within this study was found open-sourced, online by the City of New York. These data sets have been cleaned up for the purpose of analyzing whether COVID-19 had an impact on any scale on the criminal offenses occurring in New York City. The findings

suggest that overall, crime rates have reduced slightly in the presence of the pandemic but follow many of the same trends that had occurred before the pandemic.

Demographic analyzation of victims shows that Black people, females, and those within the ages of 25 and 44 are the groups most likely to be victimized. There were significantly more cases with victims that were Black and those aged 25 to 44 compared to the other groups within their categories. With this information in mind, the police department may be able to find areas in which these groups of people are more densely populated and provide more helpful resources to them.

The frequency of the crimes committed increased during summer and fall months and fell when it became closer to the winter months in 2019. For 2020, this trend continued with the exception of April, where COVID-19 cases had a major spike. The decrease in crime rate, however, did not follow a similar pattern in December 2020 despite having another major COVID-19 spike. Instead, it had followed a similar trend presented in 2019 as if there were no COVID-19 cases at all. This may indicate that by December of the pandemic, people understood what they were dealing with and felt no need to change ways in terms of criminal activity. Despite the statewide stay-at-home order taking place between late March into early June, criminal activity began to rise again between May and August. There was a relationship between COVID-19 and crime rates during the stay-at-home order, where COVID-19 increased and crime rates decreased for that month, and vice versa. There is also overall about 8% less crime being committed during the pandemic. In terms of frequency of specific crimes, the only changes overall were that most crimes had a slight decrease, with the exception of burglary, criminal mischief and related offenses, and grand larceny of motor vehicles which had all increased. Other than Staten Island and Queens, the rest of the boroughs in New York City saw a slight

decrease in the frequency of crimes. Staten Island and Queens had remained about the same number of crimes despite the pandemic. This information could help police determine and prepare for how much crime is committed, what crimes could see an increase, and where they would be committed specifically during a pandemic, and how people react to a situation like this as the pandemic rolls on.

References

"An Analysis of NYC Crime Trends from January-June 2020." *CAP Index*, CAP Index, Inc., 8

Sept. 2020, capindex.com/covid-19-crime-new-york-city-crime-study-for-2020/.

Esposito, Michelle Marie, and Anna King. "New York City: Covid-19 Quarantine and Crime."

Emerald Insight, Emerald Publishing Limited, 18 May 2021,

www.emerald.com/insight/content/doi/10.1108/JCP-10-2020-0046/full/html#sec013.

Kerr, Alexandra. "Timeline: How Covid-19 Took over NYC." Investopedia, Dotdash Meredith,

5 Apr. 2021, www.investopedia.com/historical-timeline-of-covid-19-in-new-york-city-

5071986.

"NCCCJ - Impact Report: Covid-19 and Crime" *National Commission on COVID-19 and

Criminal Justice*, Council on Criminal Justice, 21 May 2021,

covid19.counciloncj.org/2021/05/21/impact-report-covid-19-and-crime-4/.

Sibley, Ryan, et al. "The Benefits of Data in Criminal Justice: Improving Police-Community

Relations." Sunlight Foundation, 30 Apr. 2015, sunlightfoundation.com/2015/04/30/the-

benefits-of-data-in-criminal-justice-improving-police-community-relations/.

Walden University. "Why National Crime Statistics Are Important." *Walden University*, Walden

University LLC, 25 Mar. 2021, www.waldenu.edu/online-bachelors-programs/bs-in-

criminal-justice/resource/why-national-crime-statistics-are-important.