

An Analysis of Covid-19 and Its Impact on Crime in New York City (Final Project Step 1)

John Herrick

2023-07-28

Introduction

Covid-19: as overblown as it may have been, or as terrible as it may have been, my work is not to assess that. Instead I aim to examine whether there was a quantifiable impact on an already unpleasant aspect of life: crime. Crime exists in every city, just as Covid-19 infected every city, so if one were to examine the quantitative impacts of Covid-19 on crime, what would one need? Well, ideally one would have a large sample size. Also, it would be helpful to have a substantial amount of both crime and Covid-19 within that sample so that if the effect size were minor it could still be detected. New York City has a massive population, and as sad as it is to say it also has large amounts of crime and was hit particularly hard by Covid-19.

Research Questions

I plan, with my assessment, to examine the following:

- To what degree did Covid-19 impact New York City (by cases and deaths)?
- To what degree did crime impact New York City before Covid (both violent and nonviolent)?
- To what degree did rates of crime change between the onset of Covid-19 and the limit of the data used into 2022?
- Did any noticeable or significant changes in crime trends coincide with surges of Covid-19 infections or deaths?
- Could the impacts of crime and/or Covid-19 be traced to an appreciable change in the city's population size, and could that change in population size perhaps have an impact on future crime or Covid trends?
- Has the dampened relationship between Covid cases and Covid deaths due to natural immunity and vaccination produced any noticeable effects on crime?

Approach and Data Resources

But why would this subject be interesting? Well, for all the factors one considers as a source of crime (i.e., poverty, fatherlessness in youth, substance abuse), plague is not commonly considered, though on its face it seems like it could be a plausible correlative factor. Going beyond just guessing, as is commonly done when opining about the causes of civil maladies like crime, I will employ data science to find if there are any statistically significant correlations between the presence of Covid-19 in a major metropolitan area, and that city's crime rates. The implications of successfully predicting crime rates with any correlative factor is desirable from the perspective of minimizing crime.

To aid me in this I have tracked down several data sets. From the CDC I have collected data on Covid-19 cases and Covid deaths (two separate data sets) in New York City, hyperlinked here:

[https://covid.cdc.gov/covid-data-tracker/#trends_weeklydeaths_select_57]

[https://covid.cdc.gov/covid-data-tracker/#trends_weeklycases_select_57]

I also collected complementary data sets for the same information, but nationwide, so that the degree of the impacts of Covid on New York City can be assessed. All data I've collected from the CDC is up to date on a weekly level, though for the most recent data there could foreseeably be some changes due to new data arriving. The hyperlink is here:

[https://covid.cdc.gov/covid-data-tracker/#trends_weeklydeaths_select_00]

[https://covid.cdc.gov/covid-data-tracker/#trends_weeklycases_select_00]

I then will compare those numbers to varying crime statistics in New York City. I have Index crime statistics by county in New York (which can easily be focused on the boroughs of New York City) in annualized data sets from 2017 through 2021 as collected by the New York State Division of Criminal Justice Services. The Division of Criminal Justice Services receives their data from Police departments, sheriffs' offices, probation departments, and the state Office of Court Administration, and in addition to their other duties maintains these records. The hyperlink to the download page is provided here:

[<https://www.criminaljustice.ny.gov/crimnet/ojsa/countycrimestats.htm>]

I also have annual reports of the rates of different categories of Index crimes in New York City ranging from 2012 to 2021, provided by the New York State Division of Criminal Justice Services, hyperlinked here:

[<https://www.criminaljustice.ny.gov/crimnet/ojsa/stats.htm>]

I, further, have crime statistics by month pertaining to Index crime rates for New York City from 2019 through September of 2022, along with crime complaints by borough. This is also from the New York State Division of Criminal Justice Services, hyperlinked here:

[<https://criminaljustice.cityofnewyork.us/system-data/>]

And, finally, I have annualized Index crime rates by crime category for New York City from 2012 to 2021, also from the New York State Division of Criminal Justice Services, as hyperlinked below:

[<https://www.criminaljustice.ny.gov/crimnet/ojsa/stats.htm>]

While this question as I've posed it, that of the impacts of Covid-19 on crime rates in New York City, is likely a subject too complex to be fully addressed by what I've put forward, it is not out of the question that I will at least find some degree of meaningful correlation here. Obviously, there are many underlying factors that come with crime rates, not the least of which would be the rise of BLM and the accompanying riots that occurred in many parts of the U.S. during the time period I am examining (though it is not clear that the riots were entirely about the activism; it could be that people were having trouble dealing with the psychological impacts of the state-mandated lockdowns which would of course be related to Covid). And beyond transient socio-political factors, there may be other issues not captured by my scope. However, performing a quantitative analysis of the possible correlations between Covid prevalence in an area and its crime rate might yield at least a partial success. My data seems well-suited to this analysis, as does my plan of inquiry, but any successes or failures can only be evident after the effort has been made.

Returning to my data, as a matter of clarification, Index crimes are the seven crime categories used by New York State and the FBI as indicators of overall crime trends (this categorization was to make crime tracking standardized across the United States). These categories are the violent crimes of murder, rape, robbery and aggravated assault; as well as the property crimes of burglary, larceny and motor vehicle theft. The variables in these data sets are substantially different from one another: in the data set assessing crime complaints by borough, the different columns are the New York City Boroughs as well as the Grand Total, and the rows are the months (this is broken down into yearly tables); in the Index crime rates by year we have the columns of County, the Population of the county (as ascertained by the US Census Bureau), and the count and rates per 100,000 persons of the major categories of Index Crime, Violent Crime, Property Crime, and Violent Crime with Firearm; and for the monthly crime analyses we have the columns of Index Crime, Violent Index Crime, and Shooting Incidents, with rows being the months (this is also broken down into annual tables).

The data from the CDC is compiled from reports given by states and territorial jurisdictions pertaining to

their numbers on local cases of Covid infections and deaths. In particular, CDC data for New York State and New York City are tallied separately, and New York City data is not counted in the New York State data. This works well for my purposes as I am only interested in New York City. Counts from many jurisdictions include counts for both confirmed and *probable* Covid-19 cases and deaths. Of potential concern for my purposes is the fact that case and death data as submitted by New York City may lag somewhat in the CDC records based on how promptly they have recorded their data, though given the length of time since the data I am examining occurred, I am unlikely to run into a substantial issue of data being updated due to recent information received from New York City. There is some variation in the recording, however; New York City reports a combination of either the event dates of cases and deaths as they occurred, or the date that the case or death in question was reported, so some degree of resolution is likely to be lost on weekly trends though hopefully this averages out on the monthly level. The data sets are simple; in this case, there are only 3 variables of interest: the city designation, the date, and the total number of cases or deaths for the week. There is an additional historical deaths column that is almost always zero on the national trends, but since I am comparing weekly data for New York City to weekly data for the nation I will not be concerning myself with that column.

Anticipated Packages Required

I anticipate that I'll need some data manipulation functions in R to carry this out, and so I need dplyr. I also anticipate that I'll be performing some linear regression and some regression analysis so I will need car and QuantPsyc, and then I'll be plotting my work so I anticipate ggplot2 will come in handy. Of course, I will need a package to read these tables which will come in the form of readxl.

Visualization Strategies

As for visualizations I think bar plots are really going to be the work horse here, given that we are dealing with weekly progressions of counts. They will essentially look like histograms, of course, but will be made in a different way. Overlapping different bar plots of interest using alpha settings and different fill colors could be a useful way to show any effects of Covid on crime rates, also creating line graphs using these counts could help show relationships as well. I also plan on zooming into periods of time where Covid is surging in either cases or deaths and overlaying that with graphs of salient crime data, either in bar plot or line graph form. Time permitting, it would be interesting to perform linear regressions on not just the effects of Covid on crime rates overall or annually, but to also zoom in on surges and see if a different model for that atypical period of time would be more accurate and significant.

Anticipated Skills and Knowledge to Develop

In order to carry this out, the only thing that I need to learn more fully is how to overlap different data graphs within a single graph while maintaining readability. This will take some time studying ggplot2 to get the best and most illustrative graphs that I can.

Step 2 (Begin)

Data Importation and Examination

The first step is to import my various tables of data. Given that so many of them are so different in the information they hold as well as their formatting, were I to combine datasets it likely would not be with a simple join operation, but rather I would likely be creating completely new dataframes formatted and sized to fit my needs and zoom in on the data in particular I wanted to analyze. First, I will import the CDC data as this will be the easiest to accomplish:

```

NYWeeklyCase <-
  read.csv("Covid and crime in NYC data/data_table_for_weekly_case_trends__new_york_city.csv",
    skip = 2)
NYWeeklyDeath <-
  read.csv("Covid and crime in NYC data/data_table_for_weekly_death_trends__new_york_city.csv",
    skip = 2)
USWeeklyCase <-
  read.csv("Covid and crime in NYC data/data_table_for_weekly_case_trends__the_united_states.csv",
    skip = 2)
USWeeklyDeath <-
  read.csv("Covid and crime in NYC data/data_table_for_weekly_death_trends__the_united_states.csv",
    skip = 2)

```

I will now present the datasets I've just imported, but I limit their substantial output using the head() function.

```
head(NYWeeklyCase)
```

```

##           State      Date Weekly.Cases
## 1 New York City Oct 19 2022      14424
## 2 New York City Oct 12 2022      12587
## 3 New York City Oct  5 2022      13841
## 4 New York City Sep 28 2022      14684
## 5 New York City Sep 21 2022      14301
## 6 New York City Sep 14 2022      12696

```

```
head(NYWeeklyDeath)
```

```

##           State      Date Weekly.Deaths
## 1 New York City Oct 19 2022          122
## 2 New York City Oct 12 2022           91
## 3 New York City Oct  5 2022           89
## 4 New York City Sep 28 2022           75
## 5 New York City Sep 21 2022           77
## 6 New York City Sep 14 2022          107

```

```
head(USWeeklyCase)
```

```

##           State      Date Weekly.Cases Historic.Cases
## 1 United States Oct 19 2022      260808             0
## 2 United States Oct 12 2022      265175             0
## 3 United States Oct  5 2022      309887             0
## 4 United States Sep 28 2022      342162             0
## 5 United States Sep 21 2022      395071             0
## 6 United States Sep 14 2022      471611             0

```

```
head(USWeeklyDeath)
```

```

##           State      Date Weekly.Deaths Historic.Deaths
## 1 United States Oct 19 2022          2566             0

```

```
## 2 United States Oct 12 2022      2582      0
## 3 United States Oct  5 2022      2994      0
## 4 United States Sep 28 2022      2974     469
## 5 United States Sep 21 2022      3158      0
## 6 United States Sep 14 2022      3230      0
```

As seen above, the dataframes themselves are fairly simple. The US dataframes have the fourth column of `Historic.Deaths`, which I shall presently remove as I don't actually need them. Furthermore, in order to make subsequent manipulations and visualizations of the data easier, I'll use the "lubridate" package to convert the dates values into actual date types.

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.2.3
```

```
##
```

```
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      date, intersect, setdiff, union
```

```
USWeeklyCase <- USWeeklyCase[,c("State", "Date", "Weekly.Cases")]
USWeeklyDeath <- USWeeklyDeath[, c("State", "Date", "Weekly.Deaths")]
USWeeklyCase$Date <- mdy(USWeeklyCase$Date)
USWeeklyDeath$Date <- mdy(USWeeklyDeath$Date)
NYWeeklyCase$Date <- mdy(NYWeeklyCase$Date)
NYWeeklyDeath$Date <- mdy(NYWeeklyDeath$Date)
```

I shall now inspect the CDC data for any outliers or excessive missing data. The easiest way to inspect this data will be by bar graph using the "ggplot2" and "scales" packages to allow me to see if the data has any implausible gaps or spikes in what I anticipate will be reasonable (if somewhat frightening) data trends.

```
library(ggplot2)
```

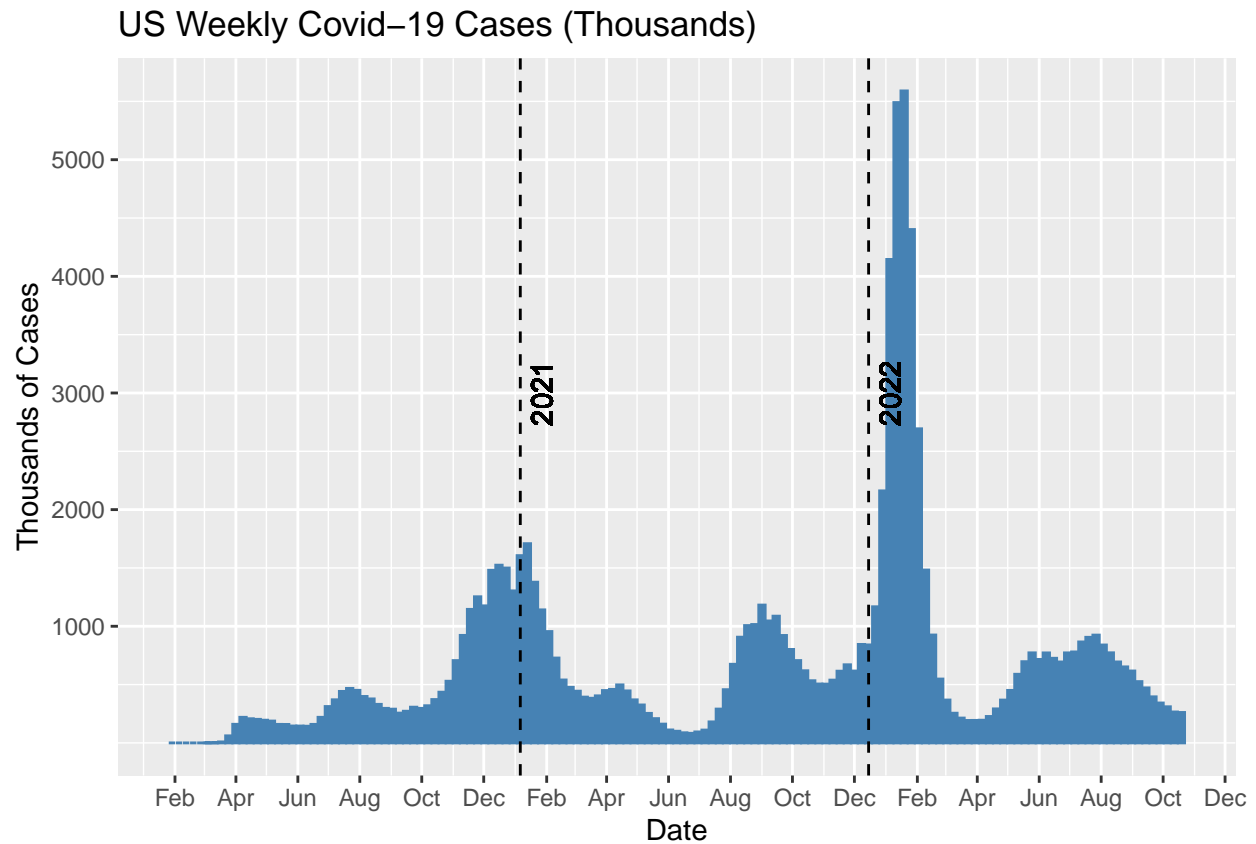
```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(scales)
USWeeklyCaseGraph <- ggplot(USWeeklyCase, aes(x = Date, y = Weekly.Cases/1000)) +
  geom_bar(stat = "identity", fill = "steel blue", color = "steel blue") +
  geom_vline(xintercept = USWeeklyCase$Date[94], linetype = "dashed") +
  geom_vline(xintercept = USWeeklyCase$Date[45], linetype = "dashed") +
  geom_text(aes(x=USWeeklyCase$Date[91], label= "2021", y=3000), colour="black",
    angle=90) + geom_text(aes(x=USWeeklyCase$Date[42], label= "2022", y=3000),
    colour="black", angle=90) + scale_x_date(date_breaks =
    "2 months", date_minor_breaks = "1 month", date_labels= c("%b")) +
  scale_y_continuous(breaks = c(1000, 2000, 3000, 4000, 5000)) + labs(y=
    "Thousands of Cases", title = "US Weekly Covid-19 Cases (Thousands)")
USWeeklyCaseGraph
```

```
## Warning: Use of 'USWeeklyCase$Date' is discouraged.
```

```
## i Use 'Date' instead.
```

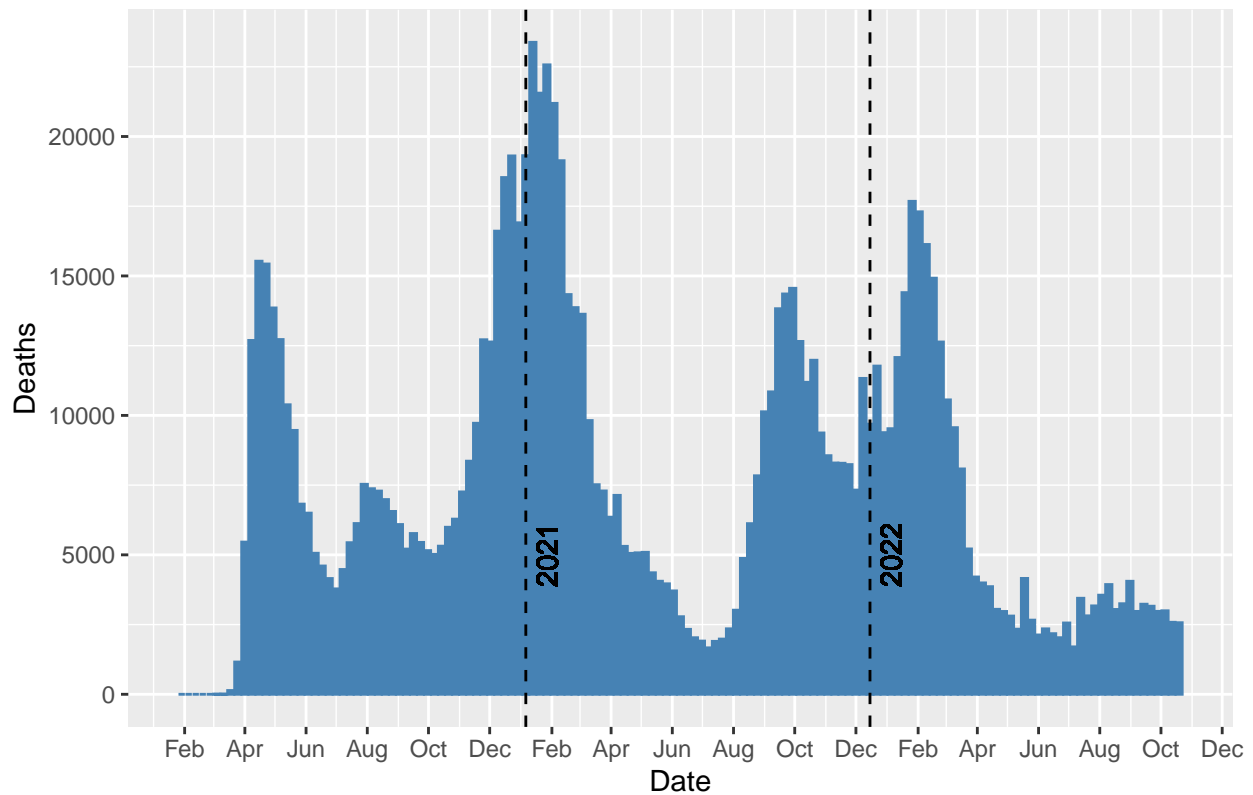
```
## Warning: Use of 'USWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```



```
USWeeklyDeathGraph <- ggplot(USWeeklyDeath, aes(x = Date, y = Weekly.Deaths)) +
  geom_bar(stat = "identity", fill = "steel blue", color = "steel blue") +
  geom_vline(xintercept = USWeeklyDeath$Date[94], linetype = "dashed") +
  geom_vline(xintercept = USWeeklyDeath$Date[45], linetype = "dashed") +
  geom_text(aes(x=USWeeklyDeath$Date[91], label= "2021", y=5000), colour="black",
    angle=90) + geom_text(aes(x=USWeeklyDeath$Date[42], label= "2022", y=5000),
    colour="black", angle=90) + scale_x_date(date_breaks =
    "2 months", date_minor_breaks = "1 month", date_labels= c("%b")) +
  labs(y= "Deaths", title = "US Weekly Covid-19 Deaths")
USWeeklyDeathGraph
```

```
## Warning: Use of 'USWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'USWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```

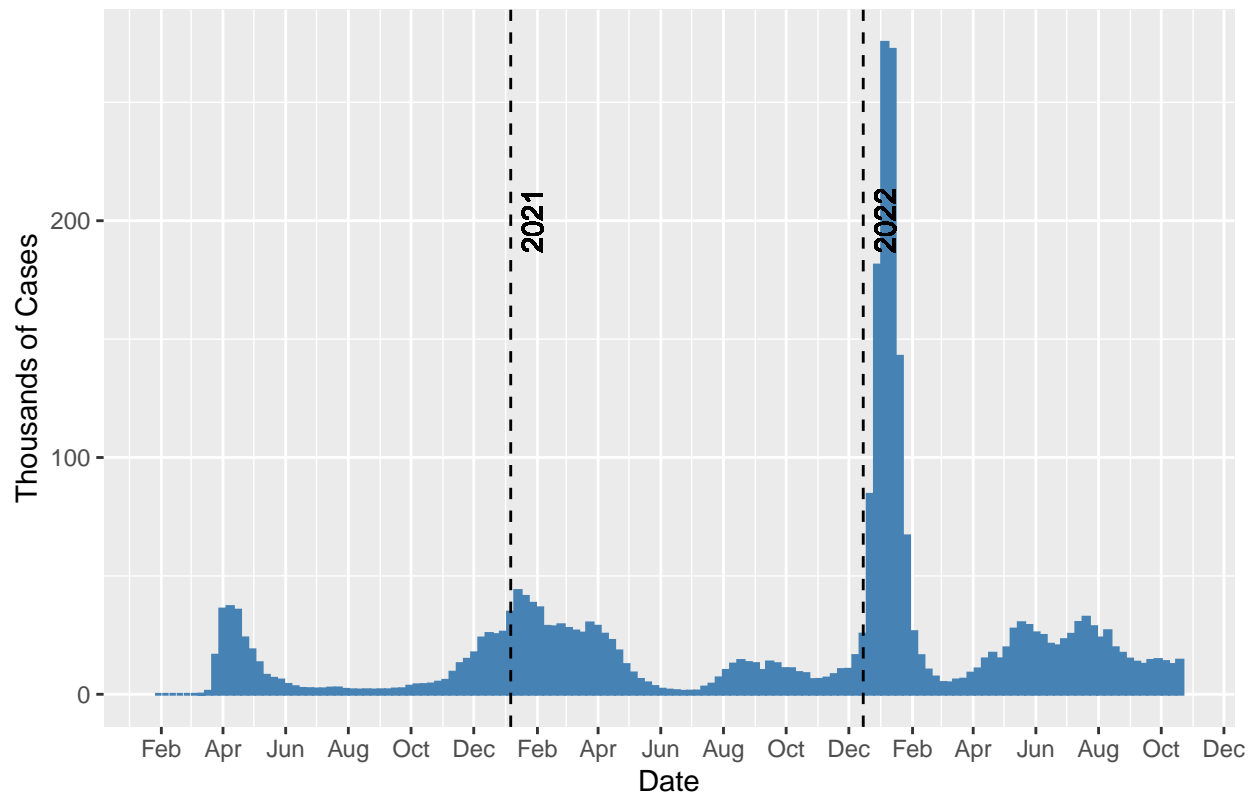
US Weekly Covid-19 Deaths



```
NYWeeklyCaseGraph <- ggplot(NYWeeklyCase, aes(x = Date, y = Weekly.Cases/1000)) +
  geom_bar(stat = "identity", fill = "steel blue", color = "steel blue") +
  geom_vline(xintercept = NYWeeklyCase$Date[94], linetype = "dashed") +
  geom_vline(xintercept = NYWeeklyCase$Date[45], linetype = "dashed") +
  geom_text(aes(x=NYWeeklyCase$Date[91], label= "2021", y=200), colour="black",
    angle=90) + geom_text(aes(x=NYWeeklyCase$Date[42], label= "2022", y=200),
    colour="black", angle=90) + scale_x_date(date_breaks =
    "2 months", date_minor_breaks = "1 month", date_labels= c("%b")) + labs(y=
    "Thousands of Cases", title = "New York City Weekly Covid-19 Cases (Thousands)")
NYWeeklyCaseGraph
```

```
## Warning: Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```

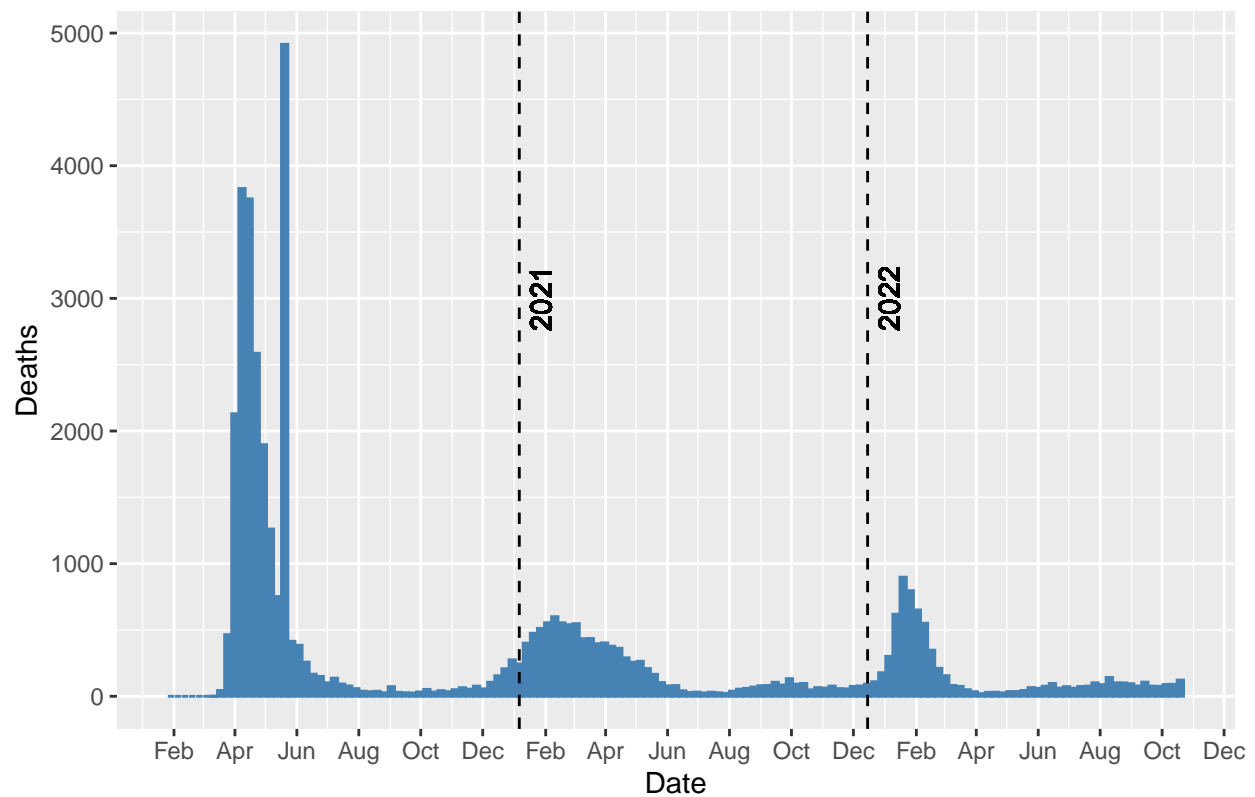
New York City Weekly Covid-19 Cases (Thousands)



```
NYWeeklyDeathGraph <- ggplot(NYWeeklyDeath, aes(x = Date, y = Weekly.Deaths)) +
  geom_bar(stat = "identity", fill = "steel blue", color = "steel blue") +
  geom_vline(xintercept = NYWeeklyDeath$Date[94], linetype = "dashed") +
  geom_vline(xintercept = NYWeeklyDeath$Date[45], linetype = "dashed") +
  geom_text(aes(x=NYWeeklyDeath$Date[91], label= "2021", y=3000), colour="black",
    angle=90) + geom_text(aes(x=NYWeeklyDeath$Date[42], label= "2022", y=3000),
    colour="black", angle=90) + scale_x_date(date_breaks =
    "2 months", date_minor_breaks = "1 month", date_labels= c("%b")) +
  labs(y= "Deaths", title = "New York City Weekly Covid-19 Deaths")
NYWeeklyDeathGraph
```

```
## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```

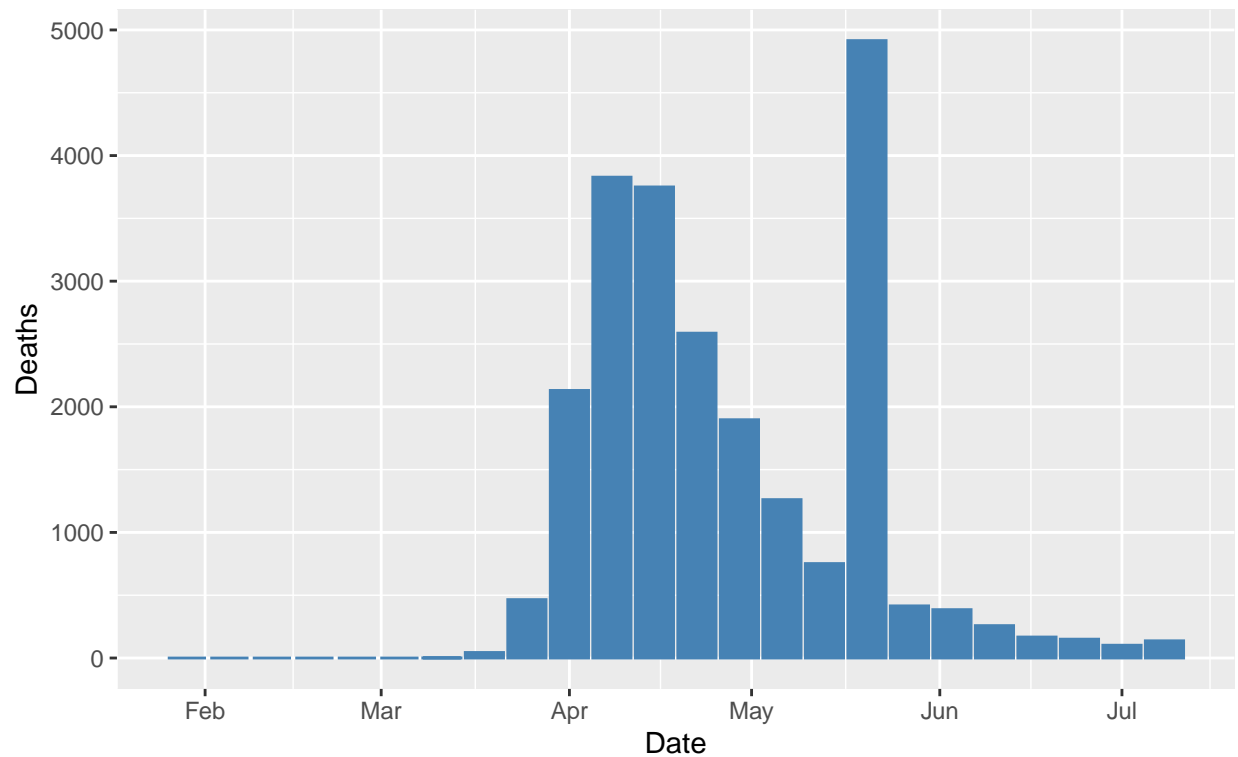

New York City Weekly Covid-19 Deaths



All of our data graphs seem reasonable so far, except for the final one. The last graph (New York City Weekly Covid-19 Deaths) shows a bizarre sudden spike between May and June of 2020. Let's zoom in on that to see what we get.

```
NYWeeklyDeathZoom <- NYWeeklyDeath[120:143, ]
NYWeeklyDeathZoomGraph <- ggplot(NYWeeklyDeathZoom, aes(x = Date, y =
  Weekly.Deaths)) + geom_bar(stat = "identity", fill = "steel blue", color =
  "steel blue") + labs(y = "Deaths", title = "New York City Weekly Covid-19
  Deaths, 2020")
NYWeeklyDeathZoomGraph
```

New York City Weekly Covid-19 Deaths, 2020



It looks as though the death rates suddenly jumped from under 1000 deaths a week to almost 5000 deaths a week, with no similar movement in the weekly cases for either the US or NYC datasets. If the data were to mirror the trends seen in the other 3 graphs, then it would be approximately 500, but it's just under 5000. It seems likely that a data entry accidentally has an extra digit.

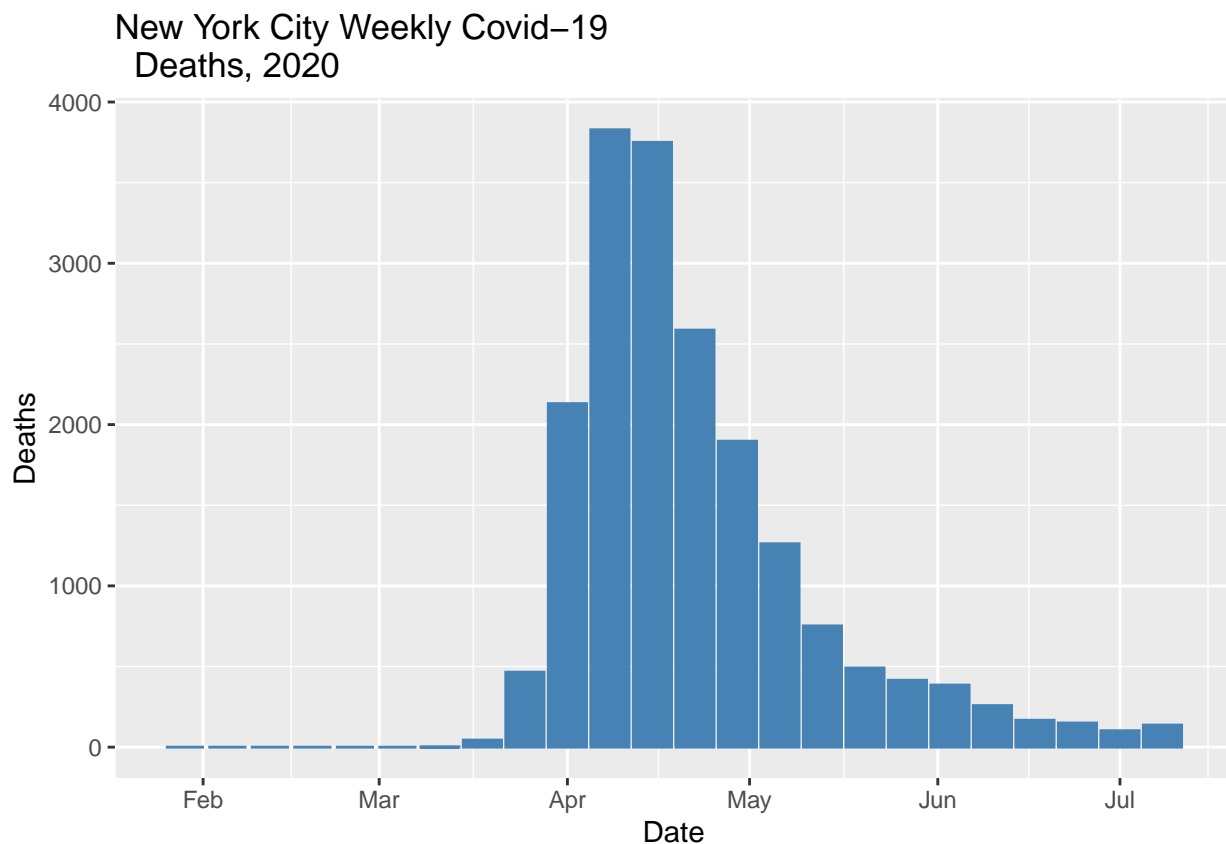
NYWeeklyDeathZoom

##	State	Date	Weekly.Deaths
## 120	New York City	2020-07-08	137
## 121	New York City	2020-07-01	102
## 122	New York City	2020-06-24	150
## 123	New York City	2020-06-17	167
## 124	New York City	2020-06-10	258
## 125	New York City	2020-06-03	385
## 126	New York City	2020-05-27	415
## 127	New York City	2020-05-20	4915
## 128	New York City	2020-05-13	752
## 129	New York City	2020-05-06	1261
## 130	New York City	2020-04-29	1897
## 131	New York City	2020-04-22	2586
## 132	New York City	2020-04-15	3750
## 133	New York City	2020-04-08	3828
## 134	New York City	2020-04-01	2130
## 135	New York City	2020-03-25	465
## 136	New York City	2020-03-18	44
## 137	New York City	2020-03-11	1

```
## 138 New York City 2020-03-04 0
## 139 New York City 2020-02-26 0
## 140 New York City 2020-02-19 0
## 141 New York City 2020-02-12 0
## 142 New York City 2020-02-05 0
## 143 New York City 2020-01-29 0
```

Given that the the data spike violates the trends present in the other graphs (with special relationship to the weekly case graphs, which show no surge in cases to explain this) and present in this graph as well, I am moving forward under the assumption that there was a data entry error, though it may be possible that a slew of deaths from earlier in the month were reported on that one date. I will remove the last digit from the entry for 2020-05-20, so the value will go from 4915 to 491. We'll check and see how this lines up with the trends in the data (I could just remove the datapoint, but I have decided to try this first and see if my suspicion at least seems reasonable). I have of course looked online for any indication of what may have happened regarding the data on that particular date, but nothing turned up.

```
NYWeeklyDeath$Weekly.Deaths[127] <- 491
NYWeeklyDeathZoom <- NYWeeklyDeath[120:143, ]
NYWeeklyDeathZoomGraph <- ggplot(NYWeeklyDeathZoom, aes(x = Date, y =
  Weekly.Deaths)) + geom_bar(stat = "identity", fill = "steel blue", color =
  "steel blue") + labs(y= "Deaths", title = "New York City Weekly Covid-19
  Deaths, 2020")
NYWeeklyDeathZoomGraph
```



The zoomed in graph looks much more reasonable. Let's examine the larger graph.

```

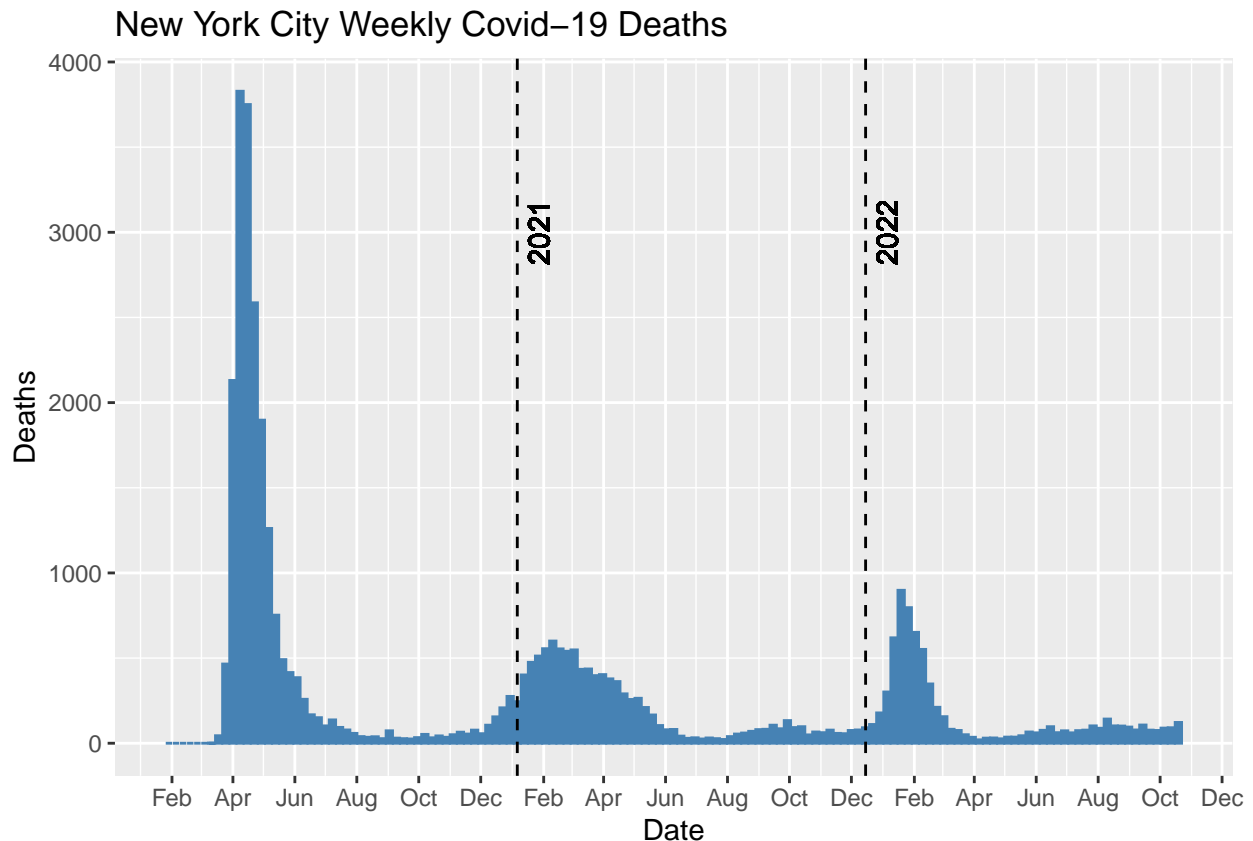
NYWeeklyDeathGraph <- ggplot(NYWeeklyDeath, aes(x = Date, y = Weekly.Deaths)) +
  geom_bar(stat = "identity", fill = "steel blue", color = "steel blue") +
  geom_vline(xintercept = NYWeeklyDeath$Date[94], linetype = "dashed") +
  geom_vline(xintercept = NYWeeklyDeath$Date[45], linetype = "dashed") +
  geom_text(aes(x=NYWeeklyDeath$Date[91], label= "2021", y=3000), colour="black",
angle=90) + geom_text(aes(x=NYWeeklyDeath$Date[42], label= "2022", y=3000),
colour="black", angle=90) + scale_x_date(date_breaks =
"2 months", date_minor_breaks = "1 month", date_labels= c("%b")) +
labs(y= "Deaths", title = "New York City Weekly Covid-19 Deaths")
NYWeeklyDeathGraph

```

```

## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.

```



Well, the larger graph no longer has its highest value come from an outlier, and the overall graph looks much more reasonable. Furthermore, no data points are missing in any of the graphs, so I will consider the CDC data ready for further use.

Moving on, I will now import and inspect my crime data. The County Index Crime Rates for 2017 through 2021 are .xls files, so I'll need the "readxl" package.

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.2.3
```

```
CIR2017 <- read_xls("Covid and crime in NYC data/2017-county-index-rates.xls", skip = 4)
```

```
## New names:
## * 'Count' -> 'Count...3'
## * 'Rate' -> 'Rate...4'
## * 'Count' -> 'Count...5'
## * 'Rate' -> 'Rate...6'
## * 'Count' -> 'Count...7'
## * 'Rate' -> 'Rate...8'
## * 'Count' -> 'Count...9'
## * 'Rate' -> 'Rate...10'
```

```
colnames(CIR2017) <- c("County", "Population", "Index_Crime_Count",
  "Index_Crime_Rate", "Violent_Crime_Count", "Violent_Crime_Rate",
  "Property_Crime_Count", "Property_Crime_Rate", "Violent_Crime_With_Firearm_Count",
  "Violent_Crime_With_Firearm_Rate")
```

Due to formatting issues with the dataset, I need to skip the first four rows, and then manually assign the categories of “Index Crime”, “Violent Crime”, “Property Crime”, and “Violent Crime with Firearm” to their respective columns pertaining to count and rate. I’ll have to repeat this for each of the datasets in this collection.

```
CIR2018 <- read_xls("Covid and crime in NYC data/2018-county-index-rates.xls", skip = 4)
```

```
## New names:
## * 'Count' -> 'Count...3'
## * 'Rate' -> 'Rate...4'
## * 'Count' -> 'Count...5'
## * 'Rate' -> 'Rate...6'
## * 'Count' -> 'Count...7'
## * 'Rate' -> 'Rate...8'
## * 'Count' -> 'Count...9'
## * 'Rate' -> 'Rate...10'
```

```
colnames(CIR2018) <- c("County", "Population", "Index_Crime_Count",
  "Index_Crime_Rate", "Violent_Crime_Count", "Violent_Crime_Rate",
  "Property_Crime_Count", "Property_Crime_Rate", "Violent_Crime_With_Firearm_Count",
  "Violent_Crime_With_Firearm_Rate")
```

```
CIR2019 <- read_xls("Covid and crime in NYC data/2019-county-index-rates.xls", skip = 4)
```

```
## New names:
## * 'Count' -> 'Count...3'
## * 'Rate' -> 'Rate...4'
## * 'Count' -> 'Count...5'
## * 'Rate' -> 'Rate...6'
## * 'Count' -> 'Count...7'
## * 'Rate' -> 'Rate...8'
## * 'Count' -> 'Count...9'
## * 'Rate' -> 'Rate...10'
```

```
colnames(CIR2019) <- c("County", "Population", "Index_Crime_Count",
  "Index_Crime_Rate", "Violent_Crime_Count", "Violent_Crime_Rate",
  "Property_Crime_Count", "Property_Crime_Rate", "Violent_Crime_With_Firearm_Count",
  "Violent_Crime_With_Firearm_Rate")
CIR2020 <- read_xls("Covid and crime in NYC data/2020-county-index-rates.xls", skip = 4)
```

```
## New names:
## * 'Count' -> 'Count...3'
## * 'Rate' -> 'Rate...4'
## * 'Count' -> 'Count...5'
## * 'Rate' -> 'Rate...6'
## * 'Count' -> 'Count...7'
## * 'Rate' -> 'Rate...8'
## * 'Count' -> 'Count...9'
## * 'Rate' -> 'Rate...10'
```

```
colnames(CIR2020) <- c("County", "Population", "Index_Crime_Count",
  "Index_Crime_Rate", "Violent_Crime_Count", "Violent_Crime_Rate",
  "Property_Crime_Count", "Property_Crime_Rate", "Violent_Crime_With_Firearm_Count",
  "Violent_Crime_With_Firearm_Rate")
CIR2021 <- read_xls("Covid and crime in NYC data/2021-county-index-rates.xls", skip = 4)
```

```
## New names:
## * 'Count' -> 'Count...3'
## * 'Rate' -> 'Rate...4'
## * 'Count' -> 'Count...5'
## * 'Rate' -> 'Rate...6'
## * 'Count' -> 'Count...7'
## * 'Rate' -> 'Rate...8'
## * 'Count' -> 'Count...9'
## * 'Rate' -> 'Rate...10'
```

```
colnames(CIR2021) <- c("County", "Population", "Index_Crime_Count",
  "Index_Crime_Rate", "Violent_Crime_Count", "Violent_Crime_Rate",
  "Property_Crime_Count", "Property_Crime_Rate", "Violent_Crime_With_Firearm_Count",
  "Violent_Crime_With_Firearm_Rate")
```

I'll present only one of the resulting dataframes as they are all formatted the same, just for different years.

```
head(CIR2017)
```

```
## # A tibble: 6 x 10
##   County      Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>         <dbl>         <dbl>         <dbl>         <dbl>
## 1 Albany      310352           8787          2831.          1162
## 2 Allegany    46878            509          1086.           80
## 3 Bronx      1470074          35507          2415.         12534
## 4 Broome      194923           5584          2865.           697
## 5 Cattaraugus 75611            1176          1555.           130
## 6 Cayuga      77712            1371          1764.           248
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

Of all of these data, I'm only interested in data for New York City. There are five counties in New York City which are co-extensive with their boroughs. The boroughs are: the Bronx (Bronx County), Brooklyn (Kings County), Manhattan (New York County), Queens (Queens County), and Staten Island (Richmond County). I will now condense this information down into only the counties which represent the five boroughs of New York City, as well as making a dataframe containing the information for the grand total of all of New York City. I'll use the "dplyr" package to make this simple.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
CIR2017 <- CIR2017%>%filter(County %in% c("Bronx", "Kings", "Queens", "New York",
"Richmond", "New York City"))
CIR2018 <- CIR2018%>%filter(County %in% c("Bronx", "Kings", "Queens", "New York",
"Richmond", "New York City"))
CIR2019 <- CIR2019%>%filter(County %in% c("Bronx", "Kings", "Queens", "New York",
"Richmond", "New York City"))
CIR2020 <- CIR2020%>%filter(County %in% c("Bronx", "Kings", "Queens", "New York",
"Richmond", "New York City"))
CIR2021 <- CIR2021%>%filter(County %in% c("Bronx", "Kings", "Queens", "New York",
"Richmond", "New York City"))
```

The transformed dataframes are now in the format presented below:

```
CIR2017
```

```
## # A tibble: 6 x 10
```

```
##   County      Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 Bronx      1470074          35507          2415.          12534
## 2 Kings      2646816          49766          1880.          14593
## 3 New York   1663498          47146          2834.           8671
## 4 Queens     2356841          32860          1394.           9184
## 5 Richmond    479104           5969          1246.           1451
## 6 New York Ci~ 8616333          171248          1987.          46433
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

CIR2018

```
## # A tibble: 6 x 10
##   County      Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 Bronx          1454179          35355          2431.          12514
## 2 Kings          2618198          50106          1914.          14384
## 3 New York      1645512          49830          3028.           8951
## 4 Queens        2331358          33135          1421.           8856
## 5 Richmond       473924           5738          1211.           1408
## 6 New York Ci~  8523171         174164          2043.          46113
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

CIR2019

```
## # A tibble: 6 x 10
##   County      Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 Bronx          1426931          35255          2471.          13157
## 2 Kings          2575649          47735          1853.          14194
## 3 New York      1638724          49858          3042.           9779
## 4 Queens        2267722          32483          1432.           9393
## 5 Richmond       470017           4789          1019.           1298
## 6 New York Ci~  8379043         170120          2030.          47821
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

CIR2020

```
## # A tibble: 6 x 10
##   County      Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 Bronx          1413534          36217          2562.          13129
## 2 Kings          2551468          50938          1996.          14590
## 3 New York      1623340          47676          2937.           9264
## 4 Queens        2246431          37446          1667.           9629
## 5 Richmond       465604           5042          1083.           1347
## 6 New York Ci~  8300377         177319          2136.          47959
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

CIR2021

```
## # A tibble: 6 x 10
##   County      Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 Bronx          1426273          38311          2686.          14819
## 2 Kings          2643508          50476          1909.          15249
```



```
## 3 New York      1578342      53543      3392.      11721
## 4 Queens        2333311      39330      1686.      10262
## 5 Richmond      493953       4845       981.       1313
## 6 New York Ci~  8475387     186505     2201.     53364
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

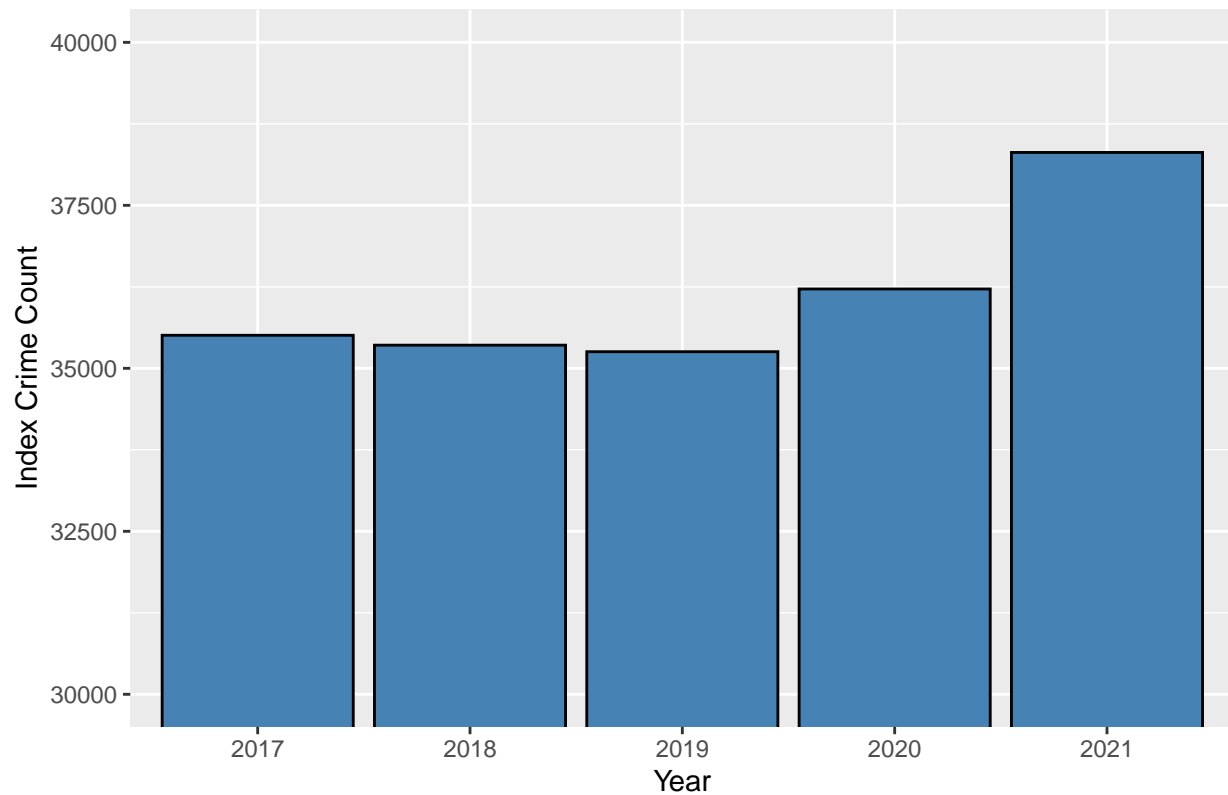
Scanning through the datasets I find nothing that stands out as an outlier. From year to year the data points all seem close enough to each other over time to be reasonable, though a more meaningful way to view this would likely be to arrange this collection of data into six different dataframes by county (and the grand total by itself) so that the changes in data points over time could be seen. I will begin by arranging a dataframe for the Bronx, and I will change the County field to Year, and then input the proper year for each row. I will also join on the individual years together for the borough such that all the data from 2017 through 2021 can be viewed in one dataframe (and ideally graphed in a single graph over that time span).

```
bronx2017 = CIR2017[CIR2017$County == "Bronx", ]
bronx2018 = CIR2018[CIR2018$County == "Bronx", ]
bronx2019 = CIR2019[CIR2019$County == "Bronx", ]
bronx2020 = CIR2020[CIR2020$County == "Bronx", ]
bronx2021 = CIR2021[CIR2021$County == "Bronx", ]
bronx <- rbind(bronx2017, bronx2018, bronx2019, bronx2020, bronx2021)
colnames(bronx)[1] = c("Year")
bronx$Year <- c("2017", "2018", "2019", "2020", "2021")
```

Trying out this plotting strategy below, I can see a clear trend of increase into 2020 and 2021 of the Index Crime Count.

```
ggplot(bronx, aes(x = Year, y = Index_Crime_Count)) + geom_bar(stat = "identity",
  fill = "steel blue", color = "black") + coord_cartesian(ylim = c(30000,40000)) +
  labs(y = "Index Crime Count", title = "Bronx Index Crime Count by Year")
```

Bronx Index Crime Count by Year



Repeating the process for the other boroughs and New York City overall, I get:

```
Kings2017 = CIR2017[CIR2017$County == "Kings", ]
Kings2018 = CIR2018[CIR2018$County == "Kings", ]
Kings2019 = CIR2019[CIR2019$County == "Kings", ]
Kings2020 = CIR2020[CIR2020$County == "Kings", ]
Kings2021 = CIR2021[CIR2021$County == "Kings", ]
kings <- rbind(Kings2017, Kings2018, Kings2019, Kings2020, Kings2021)
colnames(kings)[1] = c("Year")
kings$Year <- c("2017", "2018", "2019", "2020", "2021")
NY2017 = CIR2017[CIR2017$County == "New York", ]
NY2018 = CIR2018[CIR2018$County == "New York", ]
NY2019 = CIR2019[CIR2019$County == "New York", ]
NY2020 = CIR2020[CIR2020$County == "New York", ]
NY2021 = CIR2021[CIR2021$County == "New York", ]
ny <- rbind(NY2017, NY2018, NY2019, NY2020, NY2021)
colnames(ny)[1] = c("Year")
ny$Year <- c("2017", "2018", "2019", "2020", "2021")
Queens2017 = CIR2017[CIR2017$County == "Queens", ]
Queens2018 = CIR2018[CIR2018$County == "Queens", ]
Queens2019 = CIR2019[CIR2019$County == "Queens", ]
Queens2020 = CIR2020[CIR2020$County == "Queens", ]
Queens2021 = CIR2021[CIR2021$County == "Queens", ]
queens <- rbind(Queens2017, Queens2018, Queens2019, Queens2020, Queens2021)
colnames(queens)[1] = c("Year")
queens$Year <- c("2017", "2018", "2019", "2020", "2021")
```

```

Richmond2017 = CIR2017[CIR2017$County == "Richmond", ]
Richmond2018 = CIR2018[CIR2018$County == "Richmond", ]
Richmond2019 = CIR2019[CIR2019$County == "Richmond", ]
Richmond2020 = CIR2020[CIR2020$County == "Richmond", ]
Richmond2021 = CIR2021[CIR2021$County == "Richmond", ]
richmond <- rbind(Richmond2017, Richmond2018, Richmond2019, Richmond2020,
  Richmond2021)
colnames(richmond)[1] = c("Year")
richmond$Year <- c("2017", "2018", "2019", "2020", "2021")
NYC2017 = CIR2017[CIR2017$County == "New York City", ]
NYC2018 = CIR2018[CIR2018$County == "New York City", ]
NYC2019 = CIR2019[CIR2019$County == "New York City", ]
NYC2020 = CIR2020[CIR2020$County == "New York City", ]
NYC2021 = CIR2021[CIR2021$County == "New York City", ]
nyc <- rbind(NYC2017, NYC2018, NYC2019, NYC2020, NYC2021)
colnames(nyc)[1] = c("Year")
nyc$Year <- c("2017", "2018", "2019", "2020", "2021")

```

The newly transformed and combined dataframes for Bronx, Kings, Ny, Queens, Richmond, and NYC are as follows:

bronx

```

## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017    1470074          35507          2415.          12534
## 2 2018    1454179          35355          2431.          12514
## 3 2019    1426931          35255          2471.          13157
## 4 2020    1413534          36217          2562.          13129
## 5 2021    1426273          38311          2686.          14819
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>

```

kings

```

## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017    2646816          49766          1880.          14593
## 2 2018    2618198          50106          1914.          14384
## 3 2019    2575649          47735          1853.          14194
## 4 2020    2551468          50938          1996.          14590
## 5 2021    2643508          50476          1909.          15249
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>

```

ny

```

## # A tibble: 5 x 10

```

```
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017      1663498          47146          2834.          8671
## 2 2018      1645512          49830          3028.          8951
## 3 2019      1638724          49858          3042.          9779
## 4 2020      1623340          47676          2937.          9264
## 5 2021      1578342          53543          3392.          11721
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

queens

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017      2356841          32860          1394.          9184
## 2 2018      2331358          33135          1421.          8856
## 3 2019      2267722          32483          1432.          9393
## 4 2020      2246431          37446          1667.          9629
## 5 2021      2333311          39330          1686.          10262
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

richmond

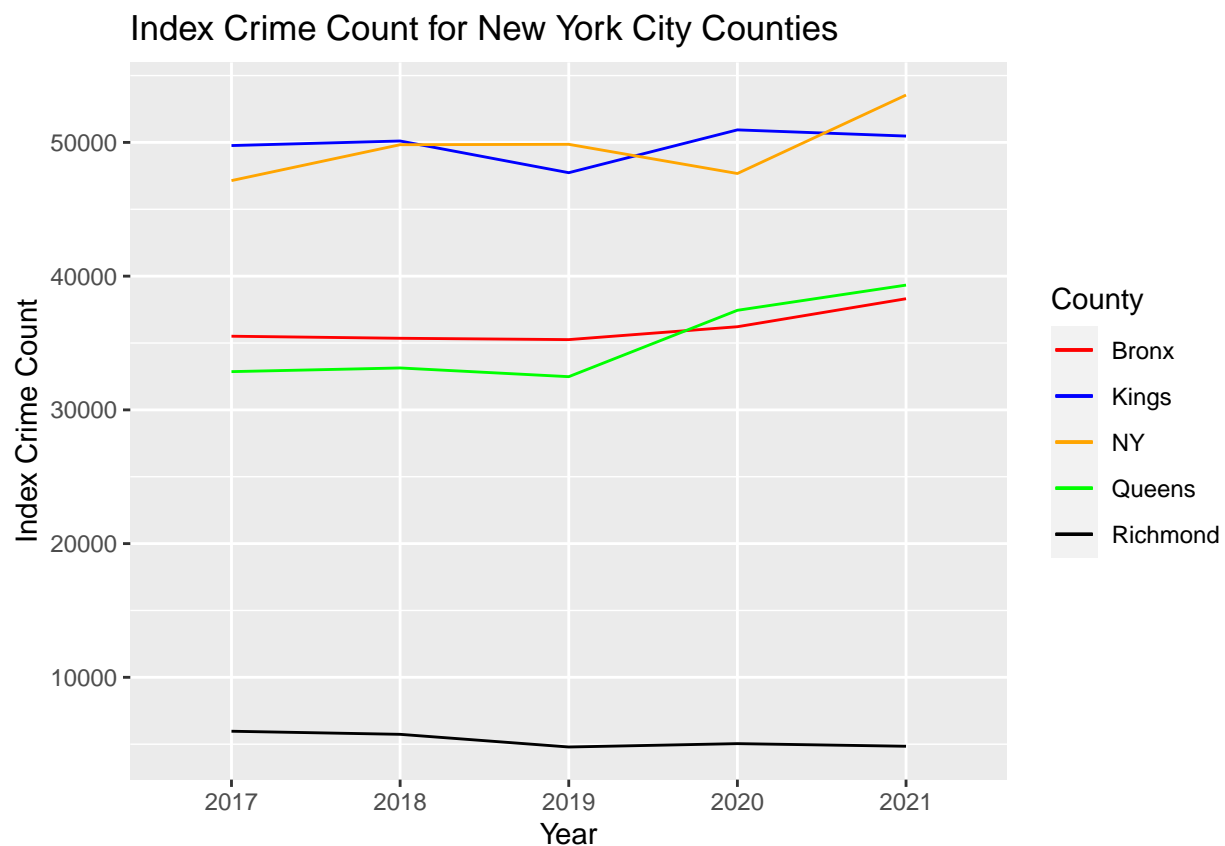
```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017      479104          5969          1246.          1451
## 2 2018      473924          5738          1211.          1408
## 3 2019      470017          4789          1019.          1298
## 4 2020      465604          5042          1083.          1347
## 5 2021      493953          4845           981.          1313
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

nyc

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017      8616333          171248          1987.          46433
## 2 2018      8523171          174164          2043.          46113
## 3 2019      8379043          170120          2030.          47821
## 4 2020      8300377          177319          2136.          47959
## 5 2021      8475387          186505          2201.          53364
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

I think it would be interesting to see the data presented in a line graph, which I've accomplished below. The five boroughs of New York City (as named by their counties) have all been graphed over 2017-2021 against the counts for the index crimes in their areas (summed annually).

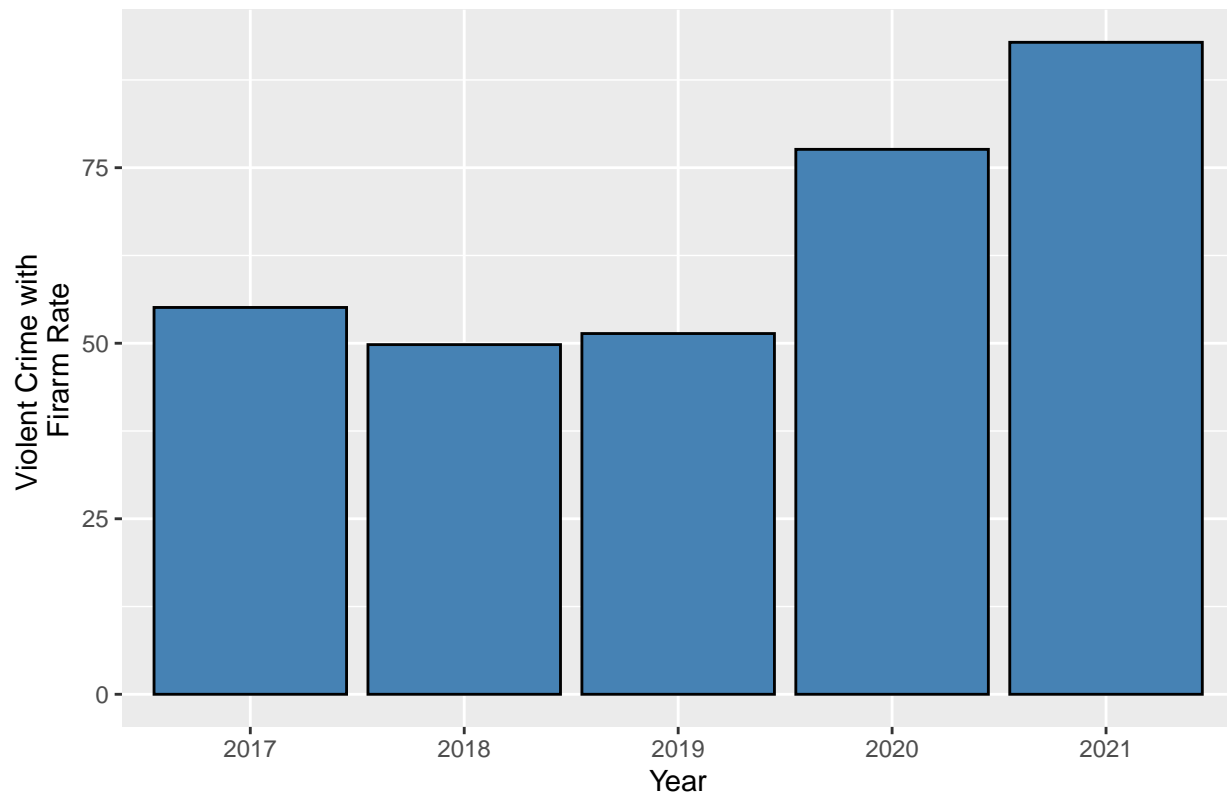
```
ggplot(bronx, aes(x=Year)) + geom_line(aes(y=Index_Crime_Count, color = "Bronx"),
  group = 1) + geom_line(data = kings, aes(x=Year, y = Index_Crime_Count,
  color = "Kings"), group =1) + geom_line(data = queens, aes(x=Year, y =
  Index_Crime_Count, color = "Queens"), group =1, ) + geom_line(data = ny,
  aes(x=Year, y = Index_Crime_Count, color = "NY"), group =1) + geom_line(data =
  richmond, aes(x=Year, y = Index_Crime_Count, color = "Richmond"), group =1) +
  scale_color_manual(name = "County", values = c("Bronx" = "red", "Kings" =
  "blue", "Queens" = "green", "NY" = "orange", "Richmond" = "black")) + labs(y=
  "Index Crime Count", title = "Index Crime Count for New York City Counties")
```



Briefly examining New York City overall for Violent Crime with a Firearm Rate, we find:

```
ggplot(nyc, aes(x=Year, y = Violent_Crime_With_Firearm_Rate)) +geom_bar(stat=
  "identity", fill = "steel blue", color= "black") + labs(y= "Violent Crime with
  Firarm Rate", title = "Violent Crime with Firearm Rate for New York City")
```

Violent Crime with Firearm Rate for New York City



As an interesting initial pass of the whole of New York City, we find that the rate of violent crimes using a firearm per 100,000 people jumped substantially after Covid-19 arrived to almost double what it had been in 2019. I will likely come back to examine this data more closely later, but first I will continue with the cleaning and preparing of my other data. Now I will import my more specific crime category data by year for New York City.

```
nycCrimeCats <- read_xls("Covid and crime in NYC data/Regions.xls", skip = 4)
```

```
## New names:
## * ' ' -> '...1'
## * ' ' -> '...2'
```

```
head(nycCrimeCats)
```

```
## # A tibble: 6 x 14
##   ...1      ...2 '2012' '2013' '2014' '2015' '2016' '2017' '2018' '2019' '2020'
##   <chr>    <chr>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 <NA>    <NA>    NA     NA     NA     NA     NA     NA     NA     NA     NA
## 2 New York~ Inde~ 452363 432695 410181 391448 378001 363177 349127 336996 343861
## 3 <NA>    Viol~ 79301 76481 72365 75157 74126 70709 68578 69898 70708
## 4 <NA>    Murd~ 688   645   616   620   631   551   568   570   843
## 5 <NA>    Rape 2831  2616  2537  6151  6247  6386  6726  6648  5650
## 6 <NA>    Robb~ 28554 27179 23994 23888 22286 20069 18194 18076 17542
## # i 3 more variables: '2021' <dbl>, '2020-21' <dbl>, '2012-21' <dbl>
```

The data is useful, but the dataframe uses the years as the column names and the categories of crimes as the rows. I want to be able to more easily compare my data, and so I will transform the dataframe to be more in line with what I had for my previous dataframe, which is to say with the years as its own column and the categories of crime as column names. I will use the “tibble” library to allow me to use the `add_column()` function to insert the year column at the front, and I will also use the transpose function (`t()`) to accomplish the bulk of my transformation work. The final product will be shown after the code. Also, at the end of the transformation, all of my columns were characters, and so I adjusted all but the Year column back to numeric (the Year column I left as characters due to the last two entries that have to do with percent change).

```
library(tibble)
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
nycCrimeCats <- nycCrimeCats[13:22, ]
nycCrimeCats <- nycCrimeCats[ , 2:14]
colnames(nycCrimeCats)[1] <- "Crime_Category"
namevect<- nycCrimeCats[["Crime_Category"]]
nycCrimeCats <- t(nycCrimeCats)
nycCrimeCats <- data.frame(nycCrimeCats)
colnames(nycCrimeCats) <- namevect
Year <- rownames(nycCrimeCats)[2:13]
nycCrimeCats <- nycCrimeCats[2:13, ]
nycCrimeCats <- add_column(nycCrimeCats, Year, .before =1)
nycCrimeCats$Year[11:12] <- c("%change_2020-21", "%change 2012-21")
rownames(nycCrimeCats) <- NULL
nycCrimeCats$`Index Crime` <- round(as.numeric(nycCrimeCats$`Index Crime`),
  digits =0)
nycCrimeCats$`Violent Crime` <- round(as.numeric(nycCrimeCats$`Violent Crime`),
  digits =0)
nycCrimeCats$Murder <- round(as.numeric(nycCrimeCats$Murder), digits =0)
nycCrimeCats$Rape <- round(as.numeric(nycCrimeCats$Rape), digits =0)
nycCrimeCats$Robbery <- round(as.numeric(nycCrimeCats$Robbery), digits =0)
nycCrimeCats$`Agg. Assault` <- round(as.numeric(nycCrimeCats$`Agg. Assault`),
  digits =0)
nycCrimeCats$`Property Crime` <- round(as.numeric(nycCrimeCats$`Property Crime`),
  digits =0)
nycCrimeCats$Burglary <- round(as.numeric(nycCrimeCats$Burglary), digits =0)
nycCrimeCats$Larceny <- round(as.numeric(nycCrimeCats$Larceny), digits =0)
nycCrimeCats$`MV Theft` <- round(as.numeric(nycCrimeCats$`MV Theft`), digits =0)
nycCrimeCats$Date <- as.numeric(nycCrimeCats$Year)
```

```
## Warning: NAs introduced by coercion
```

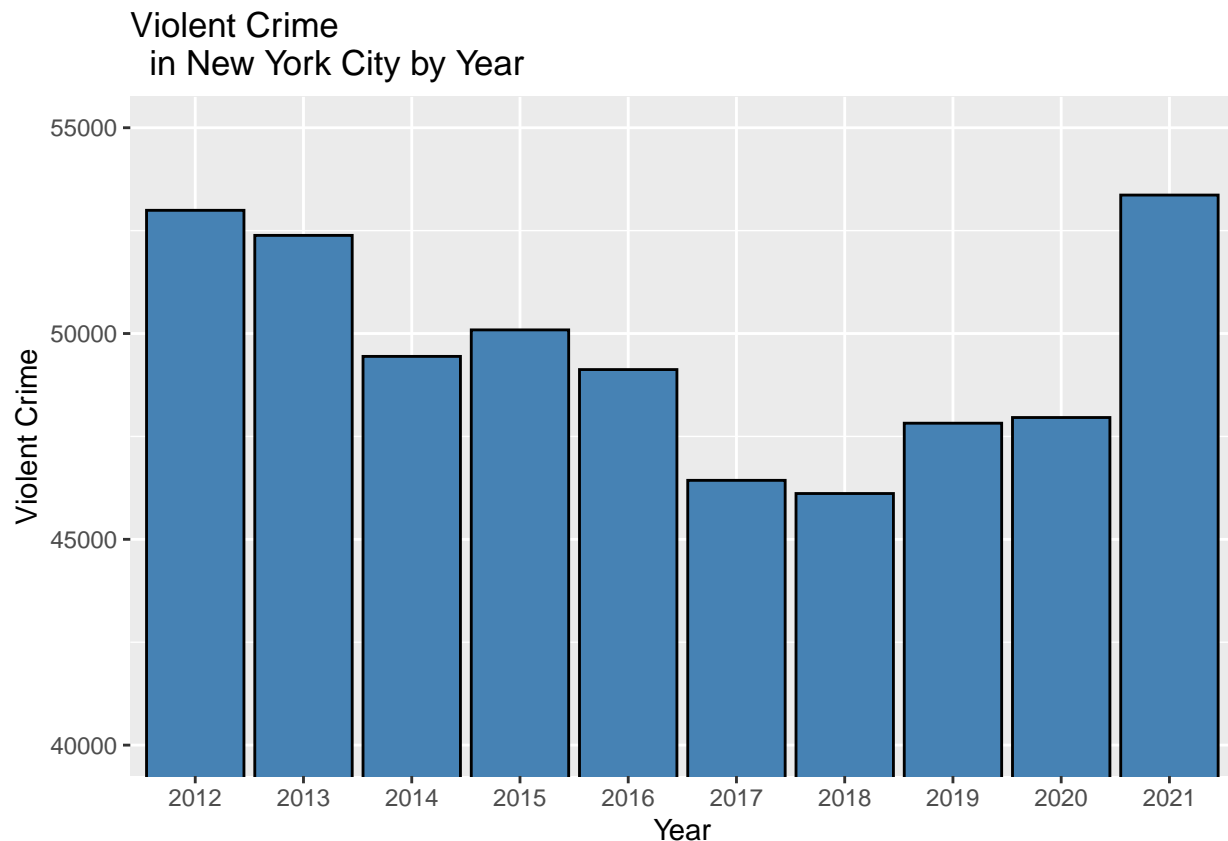
```
head(nycCrimeCats)
```

```
##   Year Index Crime Violent Crime Murder Rape Robbery Agg. Assault
## 1 2012      195753      52993    419 1162   20201      31211
## 2 2013      194355      52384    335 1112   19170      31767
## 3 2014      185191      49444    333 1070   16581      31460
## 4 2015      179948      50088    352 2244   16946      30546
## 5 2016      174402      49124    335 2372   15544      30873
```

## 6	2017	171248	46433	292 2375	13995	29771
##	Property Crime	Burglary	Larceny	MV Theft	Date	
## 1	142760	18635	115935	8190	2012	
## 2	141971	16606	117931	7434	2013	
## 3	135747	15916	112107	7724	2014	
## 4	129860	14098	108376	7386	2015	
## 5	125278	12041	106868	6369	2016	
## 6	124815	11104	107976	5735	2017	

Now, let's see what kind of graphical representation of the data I can get from my new dataframe:

```
nycCrimeCatsNoChange <- nycCrimeCats[1:10, ]
ggplot(nycCrimeCatsNoChange, aes(x=Year, y = `Violent Crime`)) + geom_bar(stat =
  "identity", fill = "steel blue", color = "black") + labs(title = "Violent Crime
  in New York City by Year") + coord_cartesian(ylim = c(40000,55000))
```



The transformation was a success, and we can see that although violent crime was trending down leading up to 2018, the trend reversed itself afterwards and spiked in 2021.

And now for my final data import, there will be several sheets. The first sheet is the crime complaints by year in New York City. For all of the sheets present from this excel file the unit of time in these tables is months, and they are presented in their character form (i.e., "January", "February", etc...); in order to make this dataframe compatible with my other dataframes I will change the information in those columns to a year-month-day format using lubridate. I do, however, end up having to factor these dates because my graphs present sideways in this case unless I make them into a factor. In addition, since the data of interest in this dataframe are crime totals tabulated for the month, I am setting my values in the date column such

that the month is identified by its final day, ensuring that comparing the data in this dataframe to others doesn't have a false displacement of one month in the data. Also, my crime data itself is presenting numbers as characters, and I want them in numeric, so I will be correcting that as well. Here is the dataframe prior to transformation:

```
crimeComplaints <- read_xlsx("Covid and crime in NYC data/Crime_thruSep2022.xlsx", sheet = 4)
```

```
## New names:
## * ' ' -> '...2'
## * ' ' -> '...3'
## * ' ' -> '...4'
## * ' ' -> '...5'
## * ' ' -> '...6'
## * ' ' -> '...7'
```

```
head(crimeComplaints)
```

```
## # A tibble: 6 x 7
##   '2020 CRIME COMPLAINTS' ...2 ...3 ...4 ...5 ...6 ...7
##   <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 <NA> BRONX BROOKLYN MANHATTAN QUEENS STATEN ISLAND Grand T~
## 2 January 8504 11111 9962 7910 1626 39113
## 3 February 7681 10304 9285 7576 1540 36386
## 4 March 7284 9667 7704 6652 1376 32683
## 5 April 5624 7191 5069 5202 1075 24161
## 6 May 7495 9372 6724 6956 1337 31884
```

Moving on now to the transformation work outlined just above, with a finished dataframe presented afterwards (it is for the year 2022, but the dataframes for the other years are formatted the same).

```
crimeComplaints <- crimeComplaints[, c(1,7) ]
crimeComplaints2020 <- crimeComplaints[c(2:13), ]
crimeComplaints2020[1] <- ymd(c("2020-01-31", "2020-02-29", "2020-03-31",
  "2020-04-30", "2020-05-31", "2020-06-30", "2020-07-31", "2020-08-31",
  "2020-09-30", "2020-10-31", "2020-11-30", "2020-12-31"))
colnames(crimeComplaints2020) <- c("Date", "Total")
crimeComplaints2020$Date <- factor(crimeComplaints2020$Date, levels =
  c("2020-01-31", "2020-02-29", "2020-03-31", "2020-04-30", "2020-05-31",
  "2020-06-30", "2020-07-31", "2020-08-31", "2020-09-30", "2020-10-31",
  "2020-11-30", "2020-12-31"))
crimeComplaints2020$Total <- as.numeric(crimeComplaints2020$Total)
crimeComplaints2021 <- crimeComplaints[c(18:29), ]
crimeComplaints2021[,1] <- ymd(c("2021-01-31", "2021-02-28", "2021-03-31",
  "2021-04-30", "2021-05-31", "2021-06-30", "2021-07-31", "2021-08-31",
  "2021-09-30", "2021-10-31", "2021-11-30", "2021-12-31"))
colnames(crimeComplaints2021) <- c("Date", "Total")
crimeComplaints2021$Date <- factor(crimeComplaints2021$Date, levels =
  c("2021-01-31", "2021-02-28", "2021-03-31", "2021-04-30", "2021-05-31",
  "2021-06-30", "2021-07-31", "2021-08-31", "2021-09-30", "2021-10-31",
  "2021-11-30", "2021-12-31"))
crimeComplaints2021$Total <- as.numeric(crimeComplaints2021$Total)
crimeComplaints2022 <- crimeComplaints[c(34:38), ]
```

```

crimeComplaints2022[,1] <- ymd(c("2022-01-31", "2022-02-28", "2022-03-31",
  "2022-04-30", "2022-05-31"))
colnames(crimeComplaints2022) <- c("Date", "Total")
crimeComplaints2022$Date <- factor(crimeComplaints2022$Date, levels =
  c("2022-01-31", "2022-02-28", "2022-03-31", "2022-04-30", "2022-05-31"))
crimeComplaints2022$Total <- as.numeric(crimeComplaints2022$Total)
head(crimeComplaints2022)

```

```

## # A tibble: 5 x 2
##   Date      Total
##   <fct>     <dbl>
## 1 2022-01-31 38433
## 2 2022-02-28 37828
## 3 2022-03-31 44375
## 4 2022-04-30 42639
## 5 2022-05-31 44613

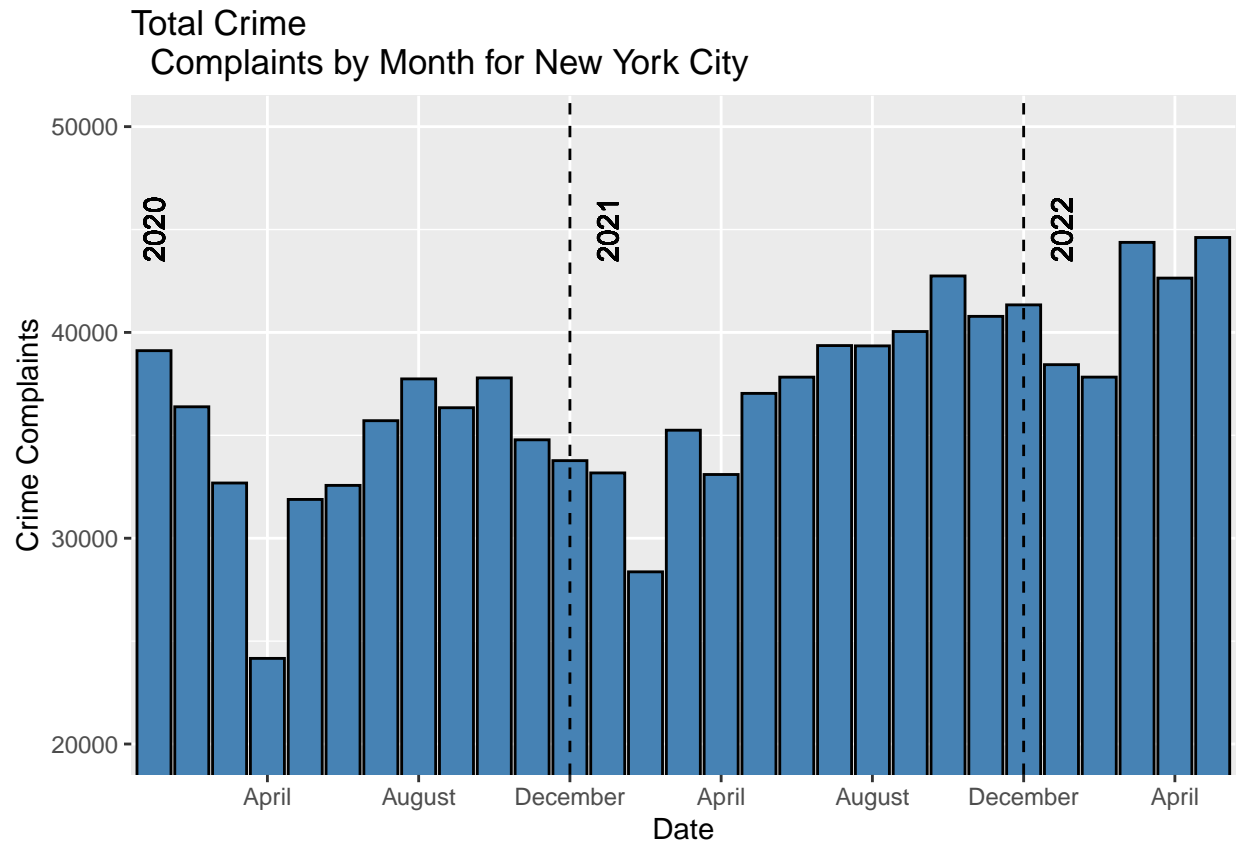
```

If I have a need of borough specific information then I can return to my earlier datasets and retrieve it, but I think I will stick with the total complaints for New York City for now. As I've been doing throughout, I will now graphically examine my new dataframes to see if any new or unexpected information reveals itself. First I will combine these dataframes into one dataframe in chronological order so that any long term trends can be discerned. What my graph ends up showing is that there is a cyclical ebb and growth of crime complaints with an overall trend for growth.

```

crimeComplaintsTot <- rbind(crimeComplaints2020[,], crimeComplaints2021[,],
  crimeComplaints2022[,])
crimeComplaintsTot <- crimeComplaintsTot[nrow(crimeComplaintsTot):1,]
ggplot(crimeComplaintsTot, aes(x = Date, y = Total)) + geom_bar(stat="identity",
  fill = "steel blue", color = "black") + coord_cartesian(ylim=c(20000,50000)) +
  geom_vline(xintercept = c("2019-12-31", "2020-12-31", "2021-12-31"), linetype =
    "dashed") + geom_text(aes(x=Date[5], y=45000, label = "2022", angle = 90)) +
  geom_text(aes(x=Date[17], y=45000, label = "2021", angle = 90)) +
  geom_text(aes(x=Date[29], y=45000, label = "2020", angle = 90)) +
  scale_x_discrete(breaks = (c("2020-04-30", "2020-08-31", "2020-12-31",
    "2021-04-30", "2021-08-31", "2021-12-31", "2022-04-30", "2022-08-31",
    "2022-12-31")), labels = c("April", "August", "December", "April", "August",
    "December", "April", "August", "December")) + labs(title = "Total Crime
    Complaints by Month for New York City", y = "Crime Complaints")

```



Now I shall import the remaining sheets. There are four years and three sheets in question, but there is enough overlap of data between the sheets that I can get the necessary data from only two imports. Given that they are all different tables containing the same rows and columns but for different years, I shall combine these separate tables into a single dataframe such that all of their information spanning 2019-2022 is included. The earlier transformations regarding dates and conversions of number information into numeric is carried out below as well. An uncombined and untransformed dataframe is shown below, followed by the finished dataframe. Again, we see an increasing trend of growth for crime, but we also see cyclical ebbs and surges, with the surges occurring late in the year.

```
crimesByMonth2019 <- read_xlsx("Covid and crime in NYC data/Crime_thruSep2022.xlsx", sheet = 3)
```

```
## New names:
## * ' ' -> '...3'
## * ' ' -> '...4'
```

```
crimesByMonth2021 <- read_xlsx("Covid and crime in NYC data/Crime_thruSep2022.xlsx", sheet = 1)
```

```
## New names:
## * ' ' -> '...3'
## * ' ' -> '...4'
```

```
crimesByMonth2022 <- crimesByMonth2021[2:10, ]
crimesByMonth2021 <- crimesByMonth2021[20:31, ]
crimesByMonth2020 <- crimesByMonth2019[2:13, ]
crimesByMonth2019 <- crimesByMonth2019[18:29, ]
head(crimesByMonth2019)
```

```
## # A tibble: 6 x 4
##   '2020'   Crime ...3   ...4
##   <chr>    <chr> <chr> <chr>
## 1 January  7215  2598  52
## 2 February 6228  2343  42
## 3 March    6966  2731  56
## 4 April    7162  2733  62
## 5 May      8075  3209  61
## 6 June     8297  3274  89
```

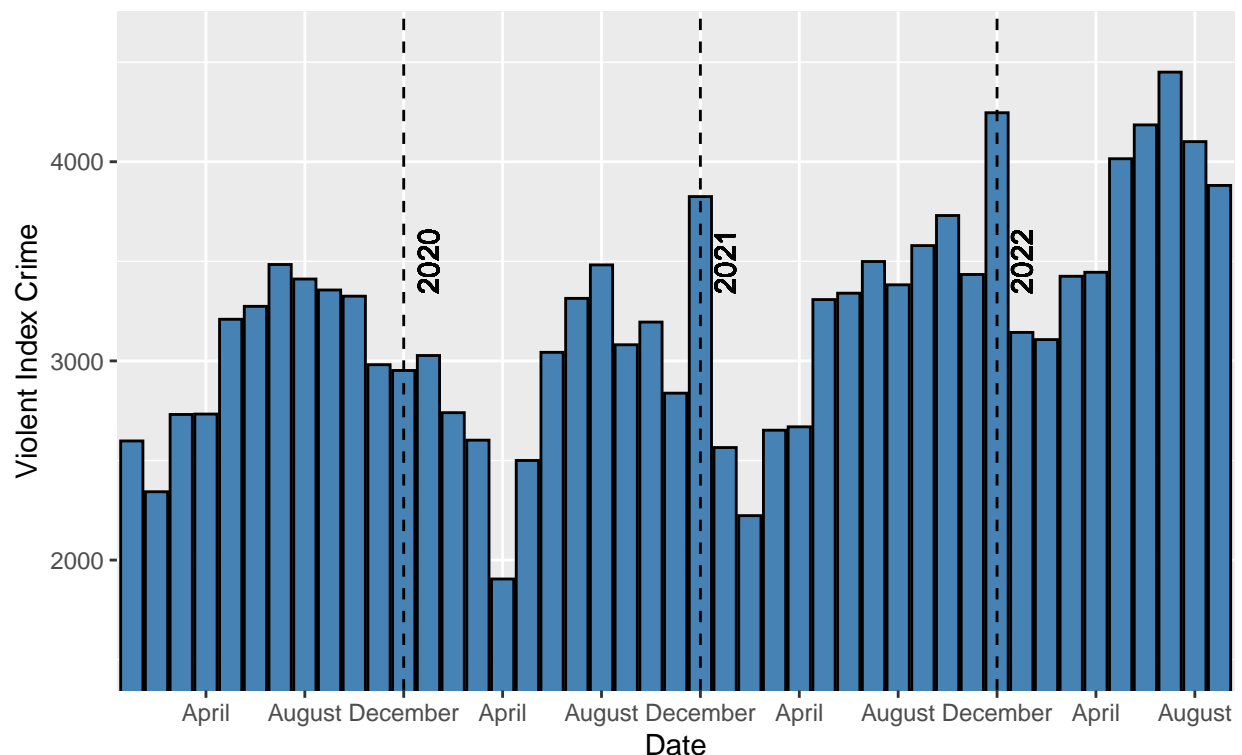
```
colnames(crimesByMonth2019) <- c("Date", "Index_Crime", "Violent_Index_Crime",
  "Shooting_Incidents")
colnames(crimesByMonth2020) <- c("Date", "Index_Crime", "Violent_Index_Crime",
  "Shooting_Incidents")
colnames(crimesByMonth2021) <- c("Date", "Index_Crime", "Violent_Index_Crime",
  "Shooting_Incidents")
colnames(crimesByMonth2022) <- c("Date", "Index_Crime", "Violent_Index_Crime",
  "Shooting_Incidents")
crimesByMonth2019$Date <- ymd(c("2019-01-31", "2019-02-28", "2019-03-31",
  "2019-04-30", "2019-05-31", "2019-06-30", "2019-07-31", "2019-08-31",
  "2019-09-30", "2019-10-31", "2019-11-30", "2019-12-31"))
crimesByMonth2019$Date <- factor(crimesByMonth2019$Date, levels = c("2019-01-31",
  "2019-02-28", "2019-03-31", "2019-04-30", "2019-05-31", "2019-06-30",
  "2019-07-31", "2019-08-31", "2019-09-30", "2019-10-31", "2019-11-30",
  "2019-12-31"))
crimesByMonth2020$Date <- ymd(c("2020-01-31", "2020-02-29", "2020-03-31",
  "2020-04-30", "2020-05-31", "2020-06-30", "2020-07-31", "2020-08-31",
  "2020-09-30", "2020-10-31", "2020-11-30", "2020-12-31"))
crimesByMonth2020$Date <- factor(crimesByMonth2020$Date, levels = c("2020-01-31",
  "2020-02-29", "2020-03-31", "2020-04-30", "2020-05-31", "2020-06-30", "2020-07-31",
  "2020-08-31", "2020-09-30", "2020-10-31", "2020-11-30", "2020-12-31"))
crimesByMonth2021$Date <- ymd(c("2021-01-31", "2021-02-28", "2021-03-31",
  "2021-04-30", "2021-05-31", "2021-06-30", "2021-07-31", "2021-08-31",
  "2021-09-30", "2021-10-31", "2021-11-30", "2021-12-31"))
crimesByMonth2021$Date <- factor(crimesByMonth2021$Date, levels = c("2021-01-31",
  "2021-02-28", "2021-03-31", "2021-04-30", "2021-05-31", "2021-06-30", "2021-07-31",
  "2021-08-31", "2021-09-30", "2021-10-31", "2021-11-30", "2021-12-31"))
crimesByMonth2022$Date <- ymd(c("2022-01-31", "2022-02-28", "2022-03-31",
  "2022-04-30", "2022-05-31", "2022-06-30", "2022-07-31", "2022-08-31",
  "2022-09-30"))
crimesByMonth2022$Date <- factor(crimesByMonth2022$Date, levels = c("2022-01-31",
  "2022-02-28", "2022-03-31", "2022-04-30", "2022-05-31", "2022-06-30", "2022-07-31",
  "2022-08-31", "2022-09-30"))
crimesByMonth2019_22 <- rbind(crimesByMonth2019, crimesByMonth2020,
  crimesByMonth2021, crimesByMonth2022)
crimesByMonth2019_22$Violent_Index_Crime <-
  as.numeric(crimesByMonth2019_22$Violent_Index_Crime)
crimesByMonth2019_22$Index_Crime <- as.numeric(crimesByMonth2019_22$Index_Crime)
crimesByMonth2019_22$Shooting_Incidents <-
  as.numeric(crimesByMonth2019_22$Shooting_Incidents)
head(crimesByMonth2019_22)
```

```
## # A tibble: 6 x 4
##   Date      Index_Crime Violent_Index_Crime Shooting_Incidents
```

##	<fct>	<dbl>	<dbl>	<dbl>
## 1	2019-01-31	7215	2598	52
## 2	2019-02-28	6228	2343	42
## 3	2019-03-31	6966	2731	56
## 4	2019-04-30	7162	2733	62
## 5	2019-05-31	8075	3209	61
## 6	2019-06-30	8297	3274	89

```
ggplot(crimesByMonth2019_22, aes(x= Date, y = Violent_Index_Crime)) +
  geom_bar(stat = "identity", fill="steel blue", color = "black") +
  coord_cartesian(ylim = c(1500,4600)) + scale_x_discrete(breaks = c("2019-04-30",
"2019-08-31", "2019-12-31", "2020-04-30", "2020-08-31", "2020-12-31", "2021-04-30",
"2021-08-31", "2021-12-31", "2022-04-30", "2022-08-31", "2022-12-31"), labels =
c("April", "August", "December", "April", "August", "December", "April", "August",
"December", "April", "August", "December"))+ geom_vline(xintercept = c("2019-12-31",
"2020-12-31", "2021-12-31"), linetype = "dashed") + geom_text(aes(x=Date[37],
y=3500, label = "2022", angle = 90)) + geom_text(aes(x=Date[25], y=3500, label =
"2021", angle = 90)) + geom_text(aes(x=Date[13], y=3500, label = "2020", angle =
90)) + labs(y = "Violent Index Crime", title = "Violent Index Crime in New York
City by Month (2019-2022)")
```

Violent Index Crime in New York
City by Month (2019–2022)



Data Analysis

Although some data analysis was occurring throughout the data importation and cleaning steps, partially to guide my efforts in arranging my data, I can now focus fully on data analysis.

Having imported, cleaned, transformed, and combined all of my data into useful dataframes, I will now briefly present the ones that I anticipate will be useful for my analysis:

```
head(USWeeklyCase)
```

```
##           State      Date Weekly.Cases
## 1 United States 2022-10-19      260808
## 2 United States 2022-10-12      265175
## 3 United States 2022-10-05      309887
## 4 United States 2022-09-28      342162
## 5 United States 2022-09-21      395071
## 6 United States 2022-09-14      471611
```

```
head(USWeeklyDeath)
```

```
##           State      Date Weekly.Deaths
## 1 United States 2022-10-19         2566
## 2 United States 2022-10-12         2582
## 3 United States 2022-10-05         2994
## 4 United States 2022-09-28         2974
## 5 United States 2022-09-21         3158
## 6 United States 2022-09-14         3230
```

```
head(NYWeeklyCase)
```

```
##           State      Date Weekly.Cases
## 1 New York City 2022-10-19      14424
## 2 New York City 2022-10-12      12587
## 3 New York City 2022-10-05      13841
## 4 New York City 2022-09-28      14684
## 5 New York City 2022-09-21      14301
## 6 New York City 2022-09-14      12696
```

```
head(NYWeeklyDeath)
```

```
##           State      Date Weekly.Deaths
## 1 New York City 2022-10-19         122
## 2 New York City 2022-10-12          91
## 3 New York City 2022-10-05          89
## 4 New York City 2022-09-28          75
## 5 New York City 2022-09-21          77
## 6 New York City 2022-09-14         107
```

```
head(bronx)
```

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017      1470074          35507          2415.          12534
## 2 2018      1454179          35355          2431.          12514
```

```
## 3 2019      1426931      35255      2471.      13157
## 4 2020      1413534      36217      2562.      13129
## 5 2021      1426273      38311      2686.      14819
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

```
head(kings)
```

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 2017      2646816      49766      1880.      14593
## 2 2018      2618198      50106      1914.      14384
## 3 2019      2575649      47735      1853.      14194
## 4 2020      2551468      50938      1996.      14590
## 5 2021      2643508      50476      1909.      15249
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

```
head(queens)
```

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 2017      2356841      32860      1394.      9184
## 2 2018      2331358      33135      1421.      8856
## 3 2019      2267722      32483      1432.      9393
## 4 2020      2246431      37446      1667.      9629
## 5 2021      2333311      39330      1686.      10262
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

```
head(richmond)
```

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 2017      479104      5969      1246.      1451
## 2 2018      473924      5738      1211.      1408
## 3 2019      470017      4789      1019.      1298
## 4 2020      465604      5042      1083.      1347
## 5 2021      493953      4845      981.      1313
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

```
head(ny)
```

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017      1663498          47146          2834.          8671
## 2 2018      1645512          49830          3028.          8951
## 3 2019      1638724          49858          3042.          9779
## 4 2020      1623340          47676          2937.          9264
## 5 2021      1578342          53543          3392.          11721
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

```
head(nyc)
```

```
## # A tibble: 5 x 10
##   Year Population Index_Crime_Count Index_Crime_Rate Violent_Crime_Count
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 2017      8616333          171248          1987.          46433
## 2 2018      8523171          174164          2043.          46113
## 3 2019      8379043          170120          2030.          47821
## 4 2020      8300377          177319          2136.          47959
## 5 2021      8475387          186505          2201.          53364
## # i 5 more variables: Violent_Crime_Rate <dbl>, Property_Crime_Count <dbl>,
## #   Property_Crime_Rate <dbl>, Violent_Crime_With_Firearm_Count <dbl>,
## #   Violent_Crime_With_Firearm_Rate <dbl>
```

```
head(nycCrimeCats)
```

```
##   Year Index Crime Violent Crime Murder Rape Robbery Agg. Assault
## 1 2012      195753      52993    419 1162  20201      31211
## 2 2013      194355      52384    335 1112  19170      31767
## 3 2014      185191      49444    333 1070  16581      31460
## 4 2015      179948      50088    352 2244  16946      30546
## 5 2016      174402      49124    335 2372  15544      30873
## 6 2017      171248      46433    292 2375  13995      29771
##   Property Crime Burglary Larceny MV Theft Date
## 1      142760      18635  115935    8190 2012
## 2      141971      16606  117931    7434 2013
## 3      135747      15916  112107    7724 2014
## 4      129860      14098  108376    7386 2015
## 5      125278      12041  106868    6369 2016
## 6      124815      11104  107976    5735 2017
```

```
head(crimeComplaintsTot)
```

```
## # A tibble: 6 x 2
##   Date      Total
##   <fct>    <dbl>
## 1 2022-05-31 44613
## 2 2022-04-30 42639
## 3 2022-03-31 44375
## 4 2022-02-28 37828
```



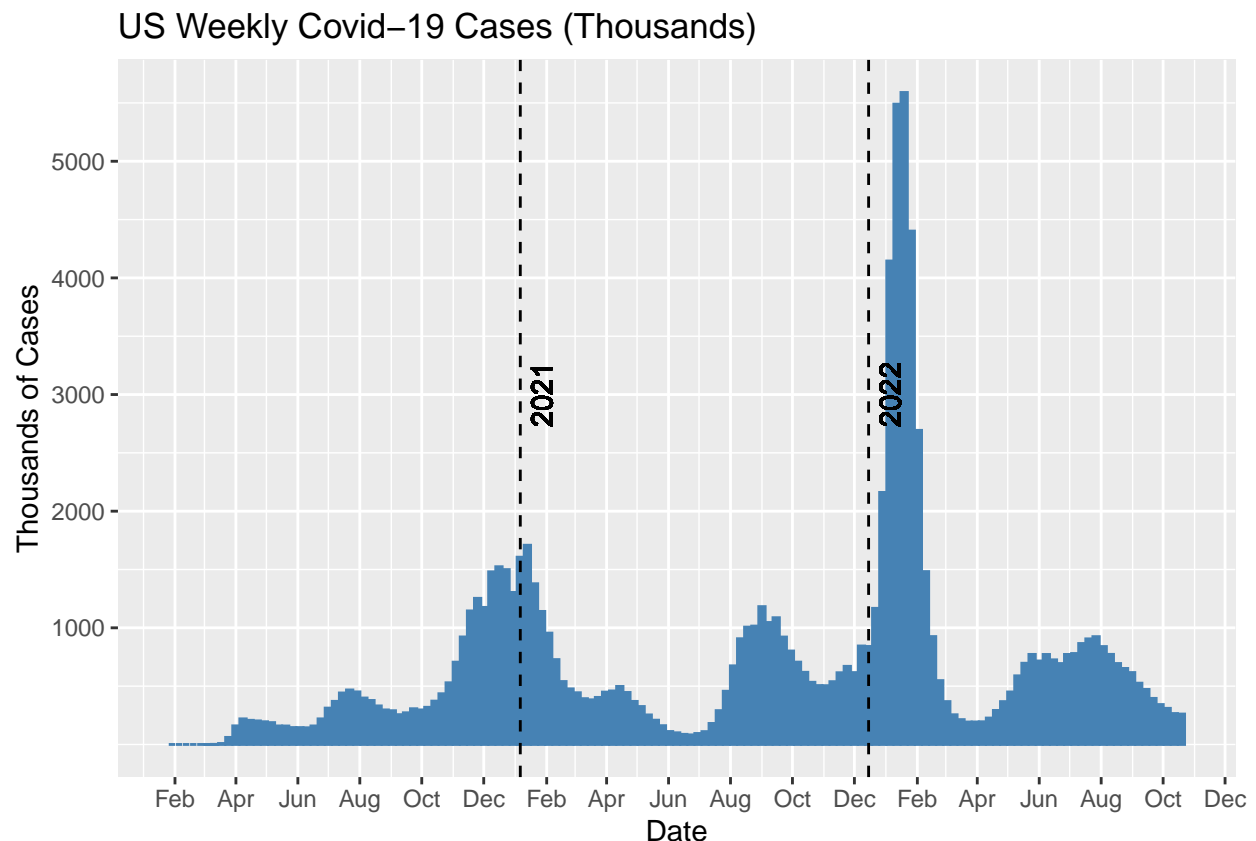
```
## 5 2022-01-31 38433
## 6 2021-12-31 41338
```

My goal in creating these dataframes was to have the various datapoints for Covid infections and deaths available to peruse over time, as well as to have the same for crime by category and/or location. It is not immediately clear whether Covid-19 cases or deaths are correlated to criminal activity in New York City, and the best way to find information that isn't necessarily clear in tabular form is to examine things graphically. Also, getting some summary statistics (especially for my crime data) would possibly yield some interesting fruits.

First, just for the sake of observing overall trends, I would like to examine the graphs pertaining to Covid-19 infections and deaths for New York City as compared to the US overall to see if New York City was in some way disproportionately affected.

USWeeklyCaseGraph

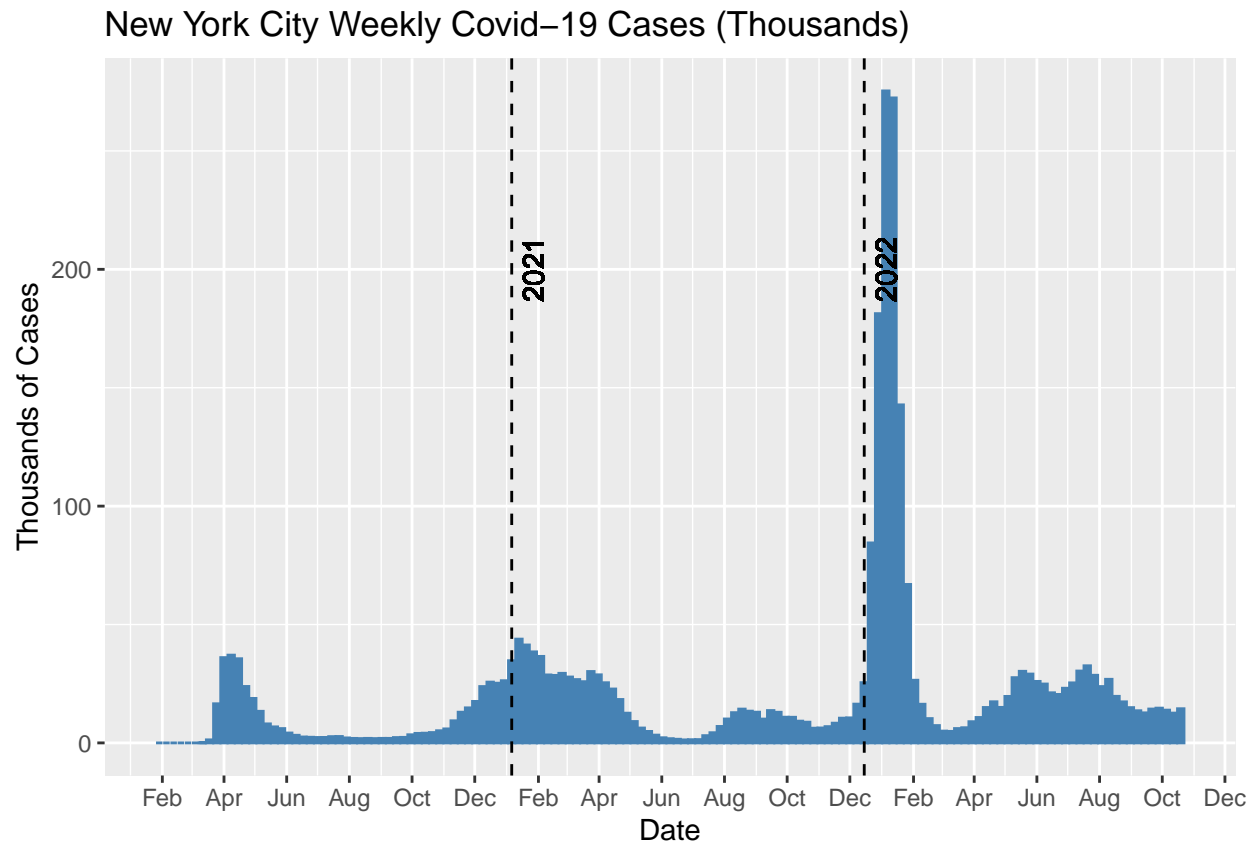
```
## Warning: Use of 'USWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'USWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```



NYWeeklyCaseGraph

```
## Warning: Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```

```
## Warning: Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```

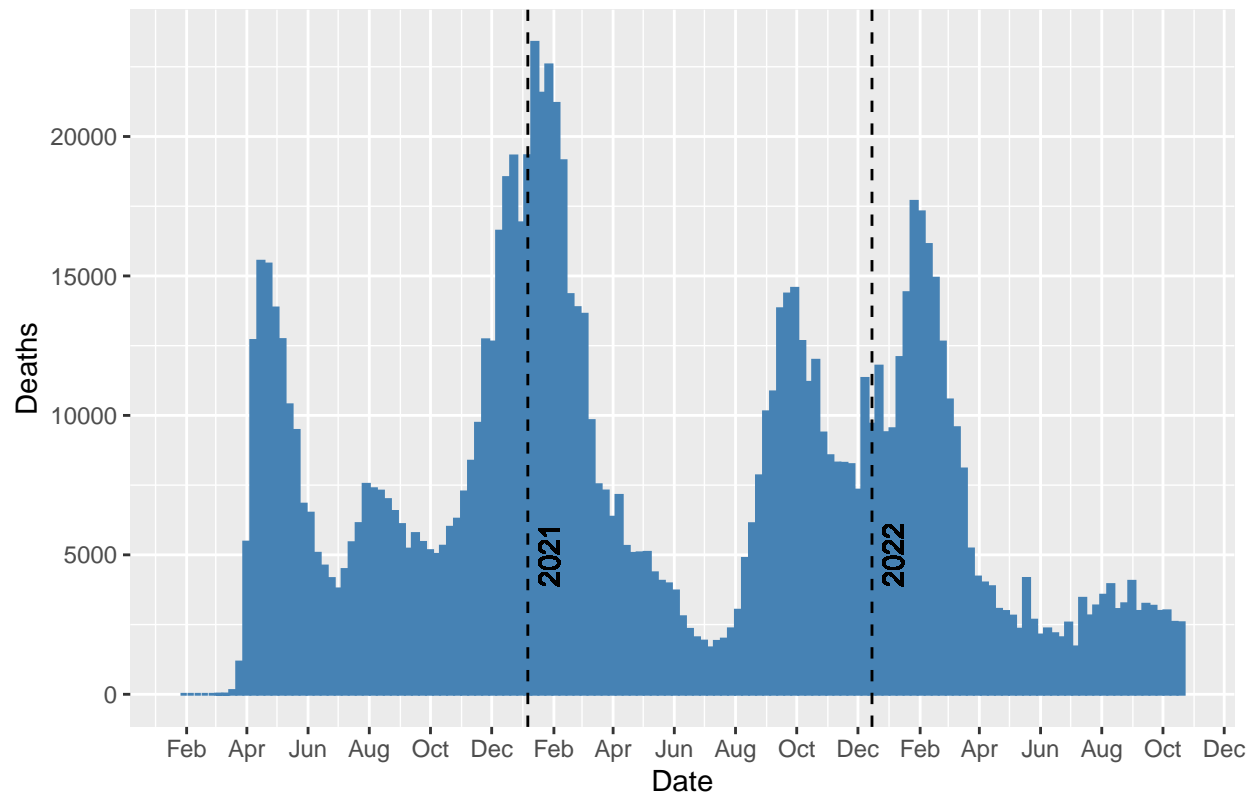


USWeeklyDeathGraph

```
## Warning: Use of 'USWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```

```
## Warning: Use of 'USWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```

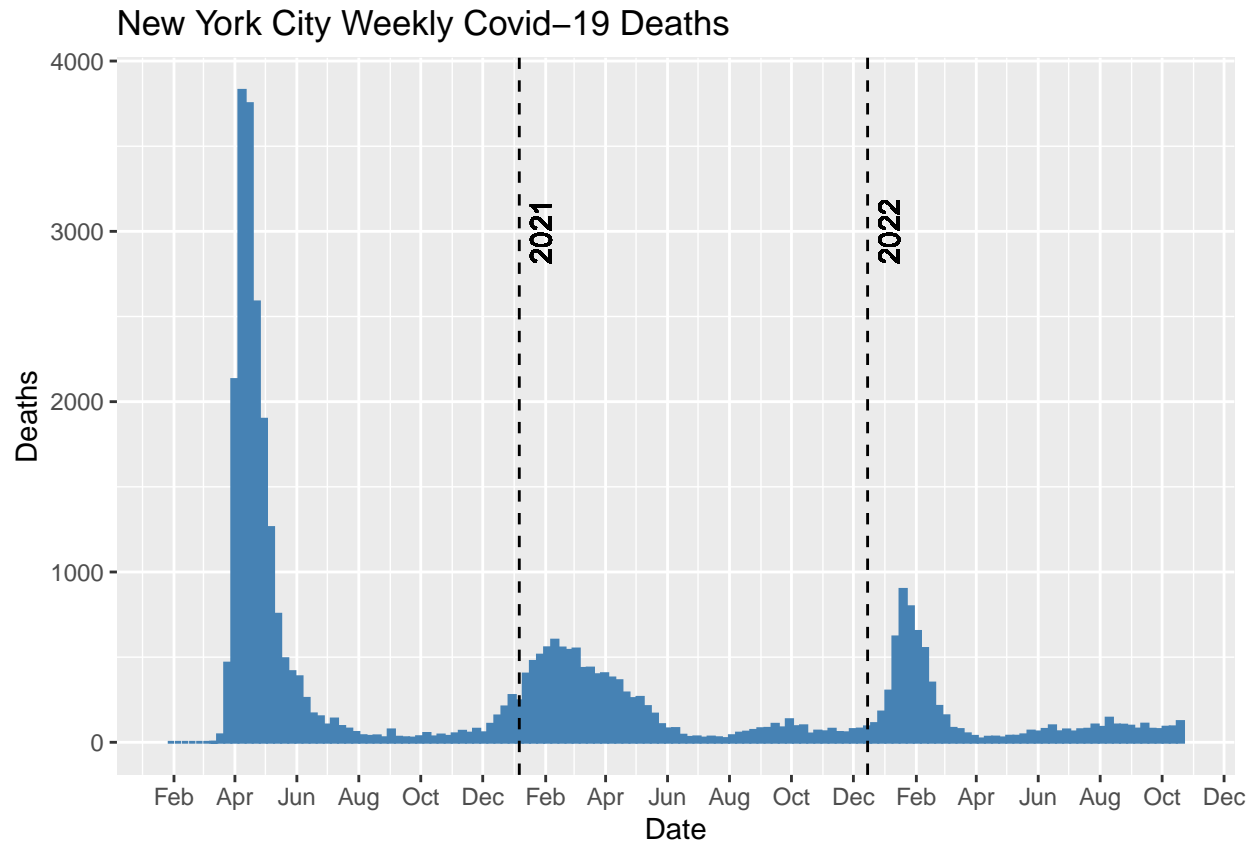
US Weekly Covid-19 Deaths



NYWeeklyDeathGraph

```
## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```

```
## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```



For the weekly cases of Covid-19, the graphs are extremely similar, with the only real difference being the magnitudes between the 2022 spike and the other lesser peaks in Covid cases throughout the rest of the timeline. For the US cases the major spike was of a lesser magnitude when compared to the other three most prominent hills in the data, whereas for New York City the spike in 2022 was of a substantially greater magnitude relative to the other large crests in the data trend. Also, New York City did have a small spike in April of 2020 that does not correspond to a similar jump in cases nationally. For my assessment of crime relationships between Covid cases, the hills in that graph would be useful places to examine, with the most interesting place possibly coinciding with the early 2022 spike.

As pertains to death from Covid-19 we see that nationally there are strong spikes that begin when COVID data was first being tracked around April of 2020 and then reoccur on a seasonal winter/fall basis, dropping off as spring arrives. The greatest spike in national deaths occurs in the fall/winter period of late 2020 and early 2021. For New York City, however, the greatest spike by far occurs in the early 2020 period (beginning near April and lasting through May), and is in fact approximately 4 times greater than the next greatest spike of data in that trend (which occurs around February of 2022). Since I am interested in how crime is impacted by Covid cases and deaths, this initial spike would be very interesting to compare to crime data for that period of time, though the other two spikes could be useful too.

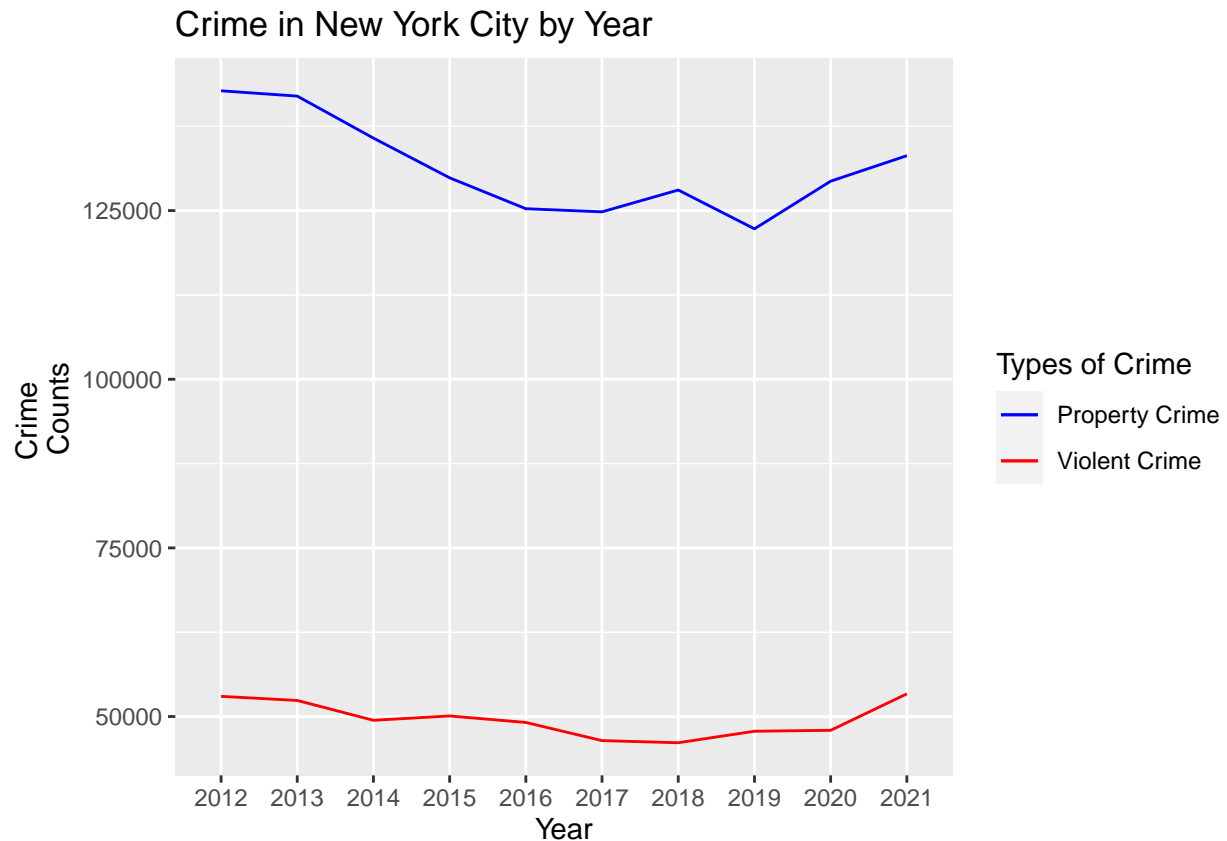
Moving on to assessing trends in crime with an eye for those interesting periods, we will first look at New York City crime over time as categorized into violent and nonviolent crime. We find that for categories of crime other than larceny, there is actually a fairly large spread between the means and their corresponding max's and mins.

```
nycCrimeCatsNoChangeGraph <- ggplot(nycCrimeCatsNoChange, aes(x = Year)) +
  geom_line(aes(y = nycCrimeCatsNoChange$`Violent Crime`, color = "Violent Crime"),
    group = 1) + geom_line(aes(y = nycCrimeCatsNoChange$`Property Crime`, color =
    "Property Crime"), group = 1) + scale_color_manual(name = "Types of Crime", values =
```

```
(c("Violent Crime" = "red", "Property Crime" = "blue"))) + labs(y = "Crime
Counts", title = "Crime in New York City by Year")
nycCrimeCatsNoChangeGraph
```

```
## Warning: Use of ‘‘ nycCrimeCatsNoChange$‘Violent Crime‘ ‘‘ is discouraged.
## i Use ‘Violent Crime’ instead.
```

```
## Warning: Use of ‘‘ nycCrimeCatsNoChange$‘Property Crime‘ ‘‘ is discouraged.
## i Use ‘Property Crime’ instead.
```



```
summary(nycCrimeCatsNoChange)
```

```
##      Year      Index Crime      Violent Crime      Murder
## Length:10      Min.   :170120      Min.   :46113      Min.   :292.0
## Class :character 1st Qu.:174224      1st Qu.:47856      1st Qu.:322.5
## Mode  :character Median :178634      Median :49284      Median :335.0
##                      Mean  :180901      Mean  :49572      Mean  :363.6
##                      3rd Qu.:186177      3rd Qu.:51810      3rd Qu.:402.2
##                      Max.   :195753      Max.   :53364      Max.   :488.0
##      Rape      Robbery      Agg. Assault      Property Crime
## Min.   :1070      Min.   :12962      Min.   :29771      Min.   :122299
## 1st Qu.:1432      1st Qu.:13524      1st Qu.:30628      1st Qu.:125971
## Median :2312      Median :14770      Median :31274      Median :129610
## Mean   :2063      Mean   :15589      Mean   :31557      Mean   :131328
```

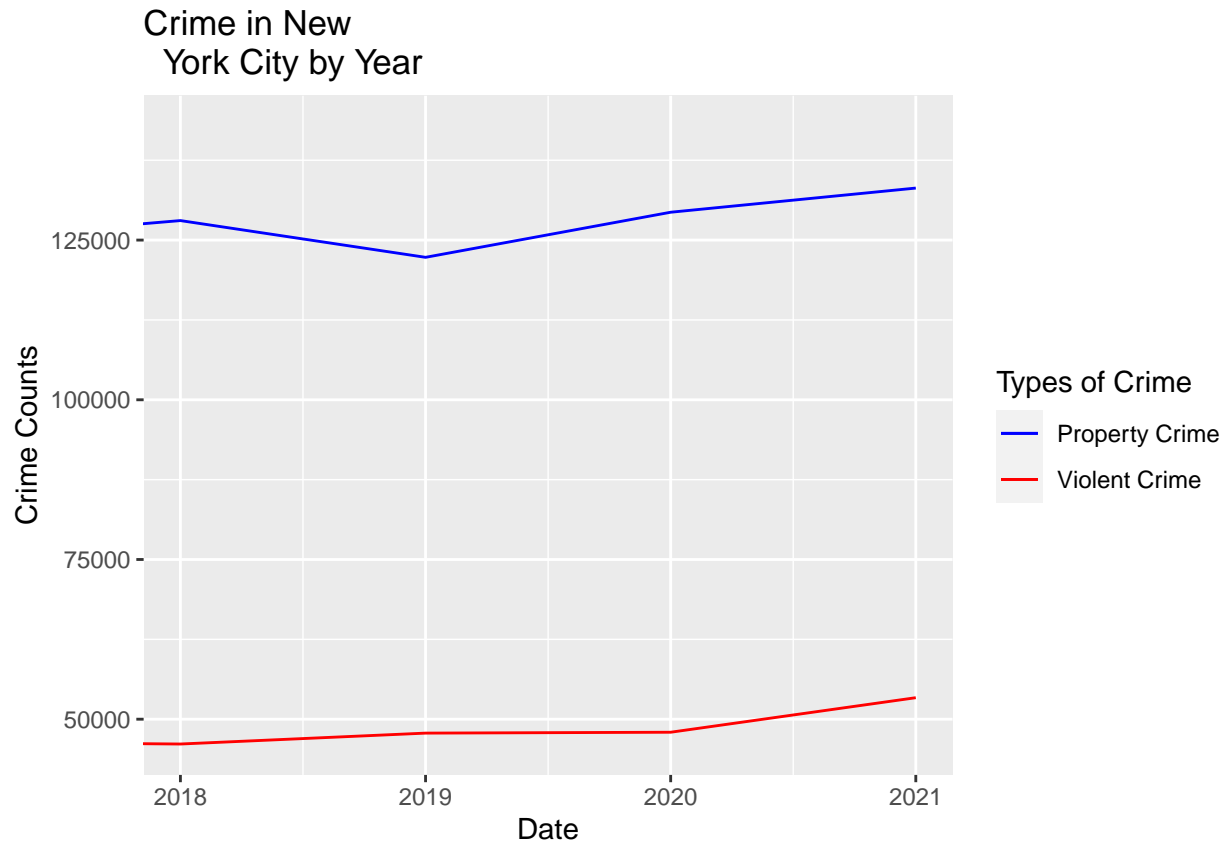
##	3rd Qu.:2439	3rd Qu.:16855	3rd Qu.:31690	3rd Qu.:135096
##	Max. :2814	Max. :20201	Max. :36507	Max. :142760
##	Burglary	Larceny	MV Theft	Date
##	Min. : 9846	Min. :106196	Min. : 5522	Min. :2012
##	1st Qu.:11210	1st Qu.:107192	1st Qu.: 5894	1st Qu.:2014
##	Median :12974	Median :109663	Median : 7410	Median :2016
##	Mean :13452	Mean :110495	Mean : 7382	Mean :2016
##	3rd Qu.:15462	3rd Qu.:112000	3rd Qu.: 8074	3rd Qu.:2019
##	Max. :18635	Max. :117931	Max. :10664	Max. :2021

Although this timeframe is substantially greater than Covid on the front-end, I felt it would be useful to see if any pre-existing trends would carry over into the Covid-19 timeframe and perhaps mislead our inquiry. I see no pre-existing strong trends, but I do see an interesting reversal in property crime. The overall trend of property crime was a downward trajectory with 2018 being a slight uptick, but it was otherwise trending down, even after 2018 into 2019. However, the trend in property crime after 2019 has been to sharply increase. Something similar was occurring with violent crime as well, though 2019-2020 was largely stagnant. Perhaps zooming in would be more useful.

```
nycCrimeCatsNoChangeGraph <- ggplot(nycCrimeCatsNoChange, aes(x = Date)) +
  geom_line(aes(y = nycCrimeCatsNoChange$`Violent Crime`, color = "Violent Crime"),
    group =1) + geom_line(aes(y = nycCrimeCatsNoChange$`Property Crime`, color =
      "Property Crime"), group =1) + coord_cartesian(xlim = c(2018,2021))+
  scale_color_manual(name = "Types of Crime", values = (c("Violent Crime" = "red",
    "Property Crime" = "blue")))) + labs(y = "Crime Counts", title = "Crime in New
    York City by Year")
nycCrimeCatsNoChangeGraph
```

```
## Warning: Use of ‘nycCrimeCatsNoChange$`Violent Crime`’ is discouraged.
## i Use ‘Violent Crime’ instead.
```

```
## Warning: Use of ‘nycCrimeCatsNoChange$`Property Crime`’ is discouraged.
## i Use ‘Property Crime’ instead.
```



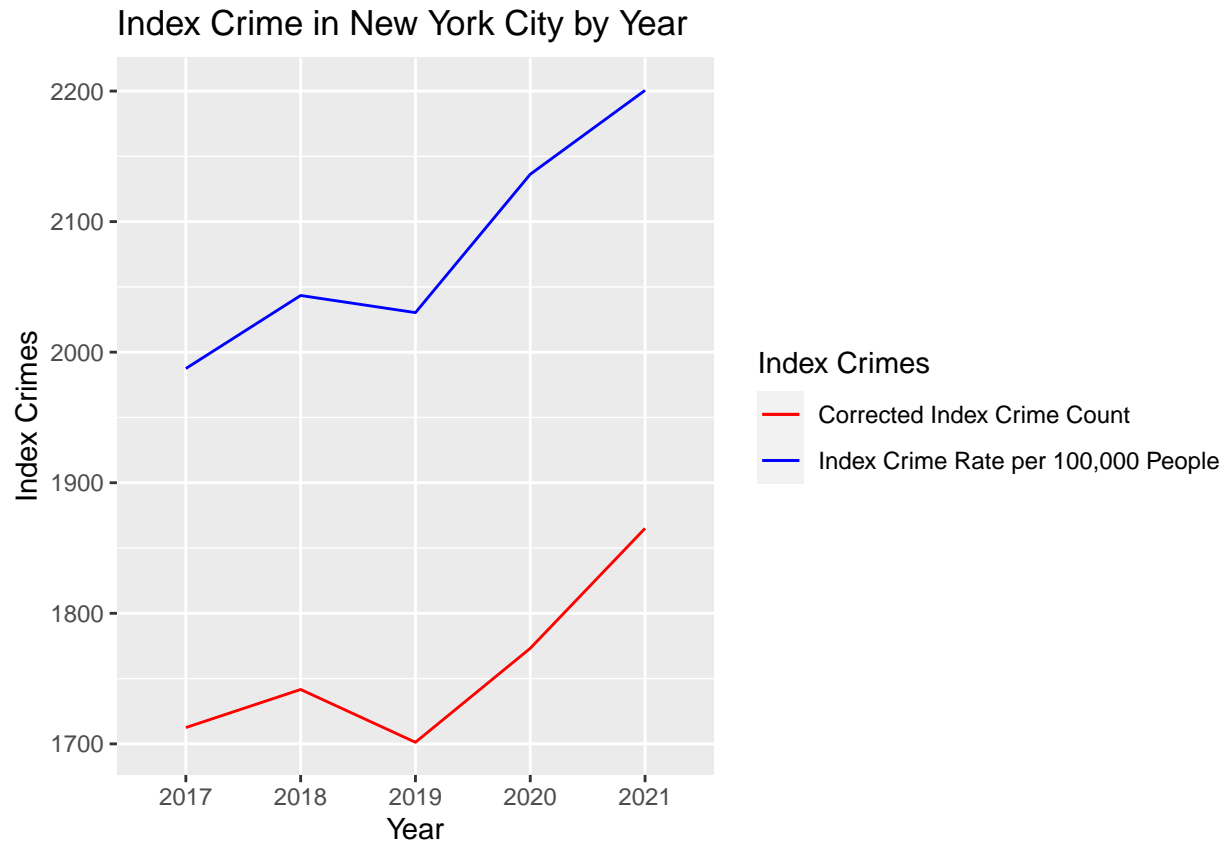
Zooming in doesn't really add much to our analysis. However, we have seen that from 2019 to 2021 there was an increase in both violent crime and property crime, which may or may not be related to Covid-19.

Examining this slightly differently, we can assess index crime versus index crime rates (which controls for population changes). Examining the graph below, we do see marked increases in both categories after 2019, though for the sake of this graph I have corrected the crime counts by dividing its values by 100 so the graph scales well with the lower magnitude crime rates per 100,000 persons.

```
nycGraph <- ggplot(nyc, aes(x= Year)) + geom_line(aes(y = nyc$Index_Crime_Count/100,
  color = "Corrected Index Crime Count"), group =1) + geom_line(aes(y =
  nyc$Index_Crime_Rate, color = "Index Crime Rate per 100,000 People"), group =1) +
  scale_color_manual(name = "Index Crimes", values = (c("Corrected Index Crime Count" =
  "red", "Index Crime Rate per 100,000 People" = "blue"))) + labs(y =
  "Index Crimes", title = "Index Crime in New York City by Year")
nycGraph
```

```
## Warning: Use of 'nyc$Index_Crime_Count' is discouraged.
## i Use 'Index_Crime_Count' instead.
```

```
## Warning: Use of 'nyc$Index_Crime_Rate' is discouraged.
## i Use 'Index_Crime_Rate' instead.
```



Moving on to a monthly analysis of crime in New York City below we see some interesting trends. First, to be clear, I did have to correct the Shooting Incidents as their numbers were otherwise too small to graph alongside the other values, so the Shooting Incidents were multiplied by 100.

What we see initially is that shooting incidents do not seem to be directly related to Covid-19 cases or deaths in terms of their timings, unless there is an inverse relationship between the two, which is possible. The reason I state that Covid cases and deaths are not driving shootings is that the shootings seem to happen in the summer and drop off in those times when Covid 19 cases and deaths are surging, but once again there could possibly be a negative correlation between the two. We do also see the categories of Index Crimes and Violent index Crimes being high during the winter and fall periods, though this is particularly noticeable with their spikes that occur in the winter of 2021 and 2022.

Following up with an examination of the summary statistics pertaining to the crime in New York City by month, we also find the means for our variables of interest. I'm less interested in the other numbers, but the mean is useful to assess when data points are above and below the mean, we just need to remember to apply the same correction factor to the summary statistics for Shooting Incidents as was applied in the graph.

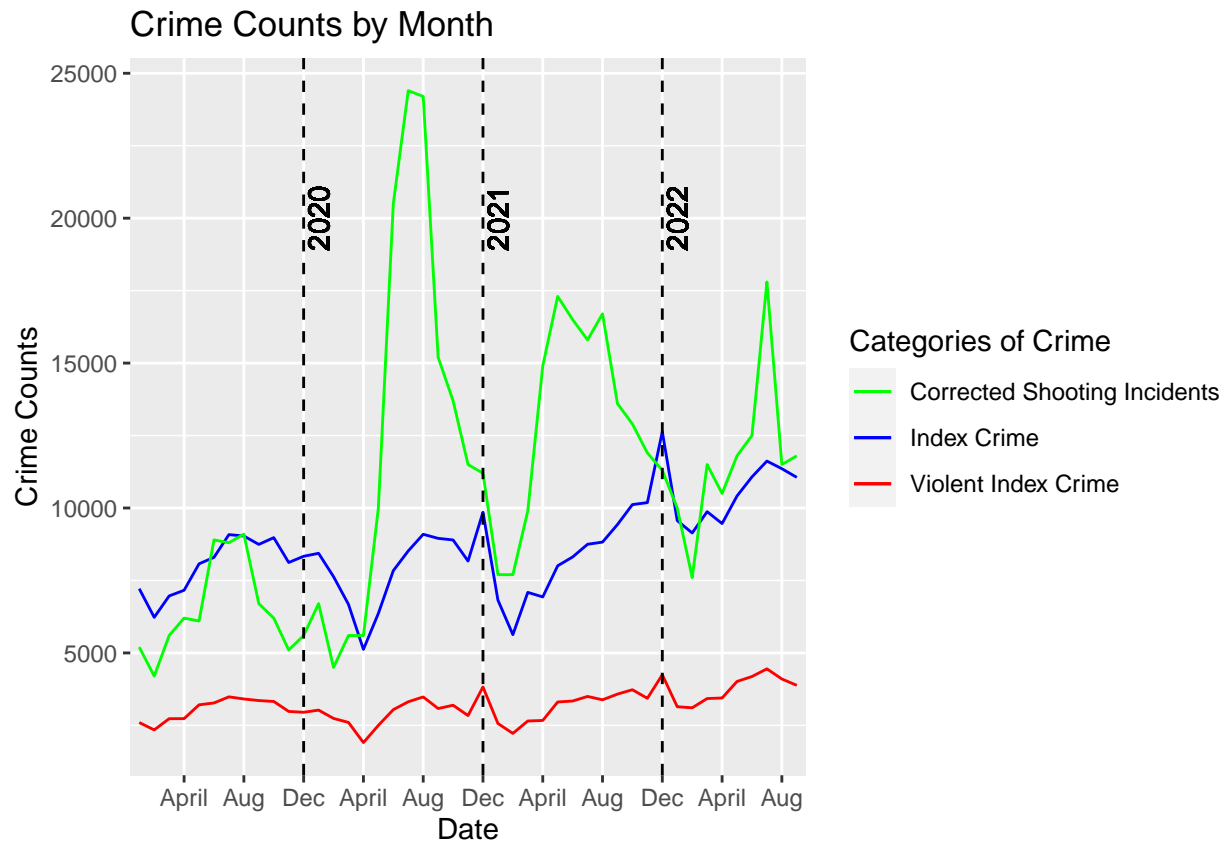
```
cbmgraph19_22 <- ggplot(crimesByMonth2019_22, aes(x=Date)) + geom_line(aes(y =
  Index_Crime, color = "Index Crime"), group = 1) + geom_line(aes(y =
  Violent_Index_Crime, color = "Violent Index Crime"), group = 1) +
  geom_line(aes(y = Shooting_Incidents*100, color = "Corrected Shooting Incidents"),
  group = 1) + scale_x_discrete(breaks = c("2019-04-30", "2019-08-31",
  "2019-12-31", "2020-04-30", "2020-08-31", "2020-12-31", "2021-04-30", "2021-08-31",
  "2021-12-31", "2022-04-30", "2022-08-31", "2022-12-31"), labels = c("April",
  "Aug", "Dec", "April", "Aug", "Dec", "April", "Aug", "Dec", "April",
  "Aug", "Dec")) + geom_vline(xintercept = c("2019-12-31", "2020-12-31",
  "2021-12-31"), linetype = "dashed") + geom_text(aes(x=Date[37], y=20000, label =
```



```

"2022", angle = 90)) + geom_text(aes(x=Date[25], y=20000, label = "2021",
angle = 90)) + geom_text(aes(x=Date[13], y=20000, label = "2020", angle = 90)) +
labs(y = "Crime Counts", title= "Crime Counts by Month") +
scale_color_manual(name = "Categories of Crime",
values = c("Violent Index Crime" = "red", "Index Crime" = "blue",
"Corrected Shooting Incidents" = "green"))
cbmgraph19_22

```



```
summary(crimesByMonth2019_22)
```

```
##          Date      Index_Crime  Violent_Index_Crime Shooting_Incidents
## 2019-01-31: 1   Min.   : 5121    Min.   :1905         Min.   : 42.0
## 2019-02-28: 1   1st Qu.: 7632    1st Qu.:2740         1st Qu.: 67.0
## 2019-03-31: 1   Median : 8744    Median :3274         Median :105.0
## 2019-04-30: 1   Mean    : 8624    Mean    :3207         Mean    :109.3
## 2019-05-31: 1   3rd Qu.: 9463    3rd Qu.:3482         3rd Qu.:136.0
## 2019-06-30: 1   Max.    :12632    Max.    :4450         Max.    :244.0
## (Other)      :39
```

To see how this lines up with Covid cases in New York City, I will overlay their graphs. I will have to perform some corrections to the data so that they will all scale properly and still be readable. In this instance, the Shooting Instances will be uncorrected, but the categories of Violent Index Crime and Index Crime will need to be divided by 100, and the Crime Complaints will need to be divided by 250.

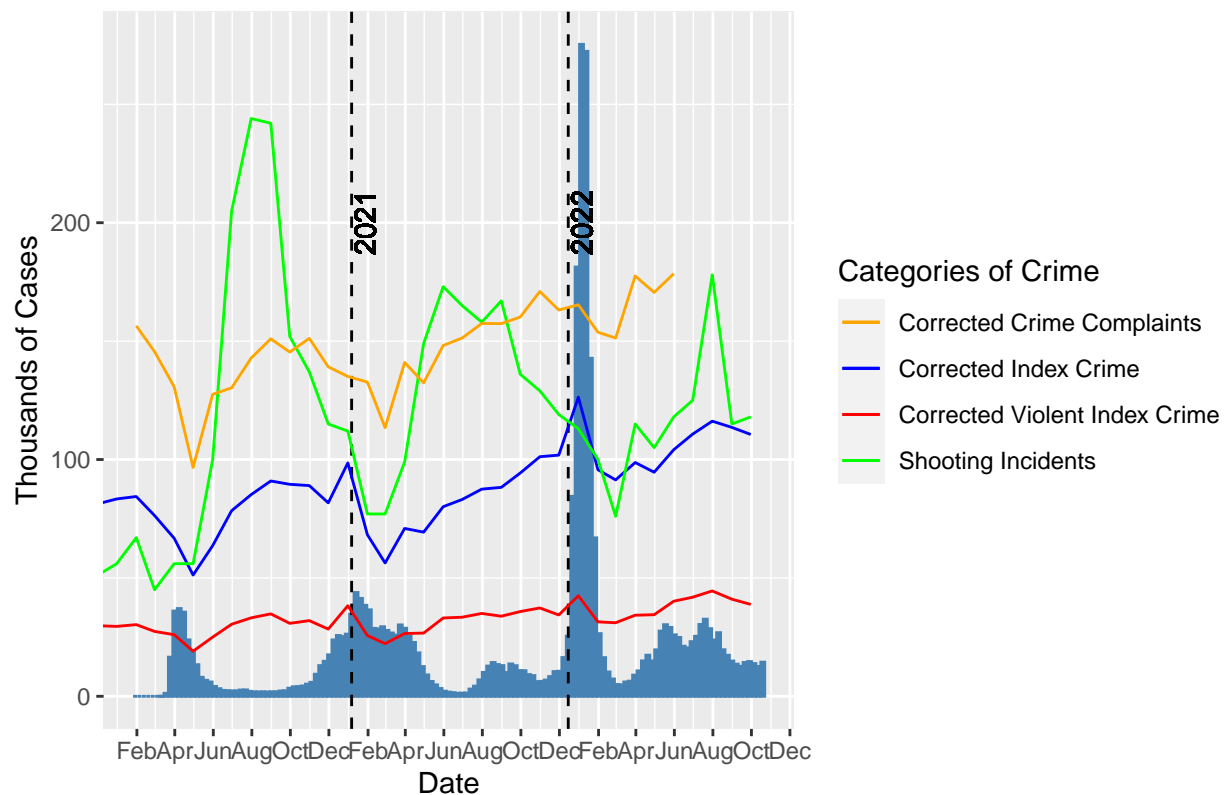
It is evident how clearly the peaks for the categories of Violent Index Crime and Index Crime line up with the 2021 and 2022 spikes for Covid cases, though these trends are the opposite for what occurred during the

initial spike in Covid cases in early 2020. Also, it seems that the category of Shooting Instances might have a negative correlation with Covid cases, but it is not a clean one if it exists at all, due partly to its poor mirroring of the 2022 spike in cases. The category of Crime Complaints mirrors Index Crime so closely that it looks to be a redundant variable, except for the fact that Crime Complaints do not spike in the winters in the same way that the Index Crimes do.

```
correctedCrimesByMonth2019_22 <- crimesByMonth2019_22
correctedCrimesByMonth2019_22$Date <- ymd(crimesByMonth2019_22$Date)
crimeComplaintsTotcorrected <- crimeComplaintsTot
crimeComplaintsTotcorrected$Date <- ymd(crimeComplaintsTotcorrected$Date)
NYWeeklyCaseGraph + geom_line(data = correctedCrimesByMonth2019_22, aes(y =
  Index_Crime/100, color = "Corrected Index Crime"), group =1) + geom_line(data =
  correctedCrimesByMonth2019_22, aes(y = Violent_Index_Crime/100, color =
  "Corrected Violent Index Crime"), group = 1) + geom_line(data =
  correctedCrimesByMonth2019_22, aes(y = Shooting_Incidents, color =
  "Shooting Incidents"), group = 1) + geom_line(data= crimeComplaintsTotcorrected,
  aes(y = Total/250, color = "Corrected Crime Complaints"), group = 1)+
  scale_color_manual(name = "Categories of Crime", values =
  c("Corrected Violent Index Crime" = "red", "Corrected Index Crime" = "blue",
  "Shooting Incidents" = "green", "Corrected Crime Complaints" = "orange")) +
  coord_cartesian(xlim = as.Date(c("2020-01-29", "2022-10-19"))) + labs(title =
  "New York City Weekly Covid-19 Cases in Relation to Crime")
```

```
## Warning: Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```

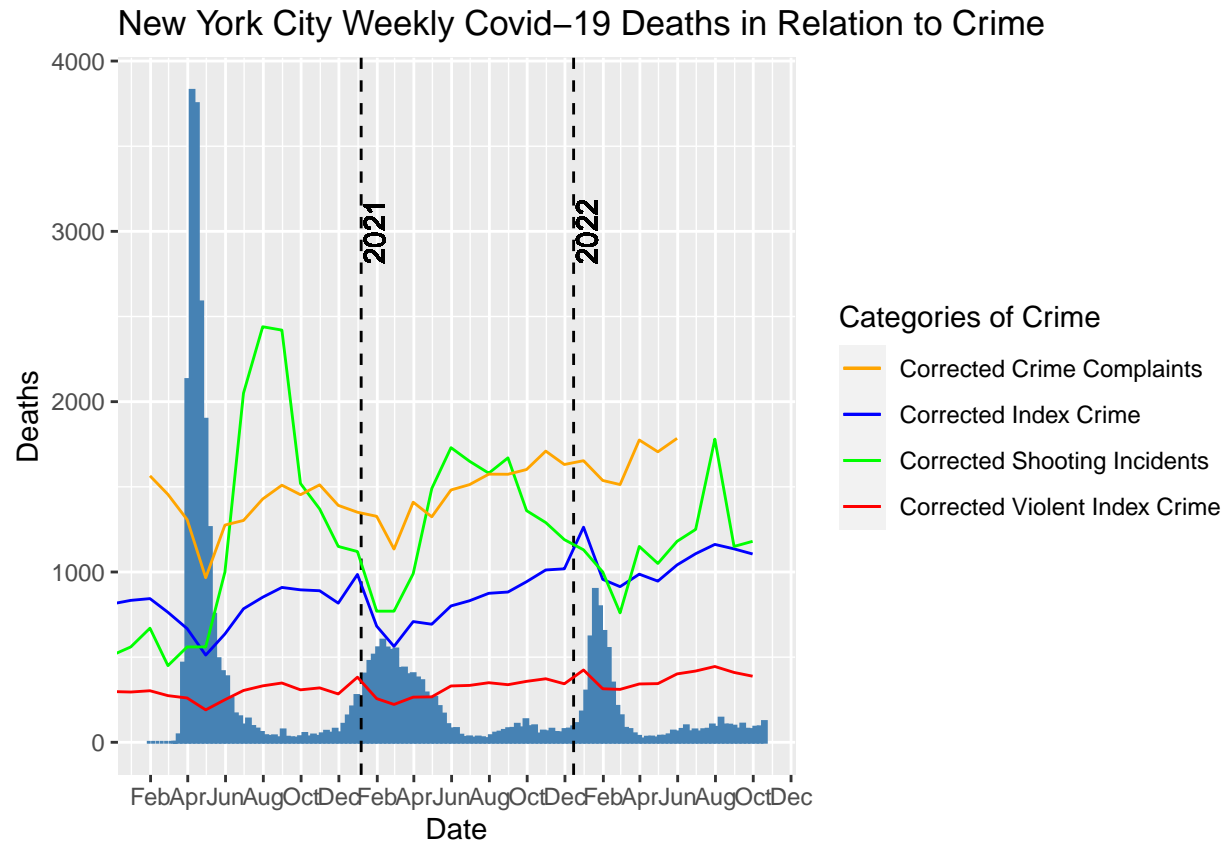
New York City Weekly Covid-19 Cases in Relation to Crime



To perform the same with weekly Covid deaths in New York City, I needed to correct the Index Crime and Violent Index Crime categories by dividing by 10, Crime Complaints I divided by 25, and then I needed to adjust the Shooting Incidents category by multiplying by 10. The relationships are much the same as they were before except for the early 2020 spike, which completely overshadows the other spikes due to lack of natural immunity and vaccines during that wave of infections.

```
NYWeeklyDeathGraph + geom_line(data = correctedCrimesByMonth2019_22,
  aes(y = Index_Crime/10, color = "Corrected Index Crime"), group = 1) +
  geom_line(data = correctedCrimesByMonth2019_22, aes(y = Violent_Index_Crime/10,
  color = "Corrected Violent Index Crime"), group = 1) + geom_line(data =
  correctedCrimesByMonth2019_22, aes(y = Shooting_Incidents*10, color =
  "Corrected Shooting Incidents"), group = 1) + geom_line(data=
  crimeComplaintsTotcorrected, aes(y = Total/25, color =
  "Corrected Crime Complaints")) + scale_color_manual(name = "Categories of Crime",
  values = c("Corrected Violent Index Crime" = "red", "Corrected Index Crime" =
  "blue", "Corrected Shooting Incidents" = "green", "Corrected Crime Complaints" =
  "orange")) + coord_cartesian(xlim = as.Date(c("2020-01-29", "2022-10-19"))) +
  labs(title = "New York City Weekly Covid-19 Deaths in Relation to Crime")
```

```
## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```



Given that I am not yet knowledgeable in time series analysis techniques and this data seems to be a time series, I am not able to go as far as I would like in this analysis. The data distributions, given that they occur over time and are in some ways recurring over time, are not normal. However, if I were to isolate individual cycles the covid data becomes much more normal in its distributions, if not perfectly normal. Creating two different dataframes for the periods of time where the greatest spikes of Covid deaths and cases were occurring would let me at least perform some correlation analyses.

Since the Covid data is weekly, but the crime data is monthly, I will make a dataframe wherein the value for the crime data is repeated every week within a given month until the month changes (essentially like checking the correlation between a line graph and a histogram). Since the dates within the dataframes in question do not line up perfectly this will take some manual tuning.

Due to the small number of cases in this test, and also the repetition of the crime data to account for the lower resolution of the crime data relative to the Covid data, I am using Kendall's Tau to assess the possible correlations between the weekly Covid numbers and the variables of interest. The first analysis will be during the greatest spike of Covid deaths, which occurred in early 2020. In order to accomplish this I have to truncate my different dataframes such that we are only examining the period of time from March to June of 2020, and since the dates are staggered differently across the dataframes this will require some manual tuning. As I've already mentioned above, I'll need to compensate for the lower resolution of crime data by ensuring that each of its monthly data points exists as the same value for each week in a given month of Covid data.

```
NYWeeklyDeathFirstChunk <- NYWeeklyDeath[NYWeeklyDeath$Date > "2020-03-04", ]
NYWeeklyDeathFirstChunk <- NYWeeklyDeathFirstChunk[NYWeeklyDeathFirstChunk$Date <
  "2020-06-01", ]
correctedCrimesByMonth2019_22altered <- correctedCrimesByMonth2019_22[5,2:4]
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
```

```

correctedCrimesByMonth2019_22[5,2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[5,2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[5,2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[4, 2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[4, 2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[4, 2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[4, 2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[4, 2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[3,2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[3,2:4])
correctedCrimesByMonth2019_22altered <- rbind(correctedCrimesByMonth2019_22altered,
correctedCrimesByMonth2019_22[3,2:4])
crimeComplaintsTotcorrectedalt <-
data.frame(c(rep(crimeComplaintsTotcorrected$Total[1], 4),
rep(crimeComplaintsTotcorrected$Total[2], 5),
rep(crimeComplaintsTotcorrected$Total[3], 3)))
NYWeeklyDeathFirstChunk <- cbind(NYWeeklyDeathFirstChunk,
correctedCrimesByMonth2019_22altered, crimeComplaintsTotcorrectedalt[, 1])
colnames(NYWeeklyDeathFirstChunk)[7] <- "Total_Complaints"
cor.test(NYWeeklyDeathFirstChunk$Weekly.Deaths, NYWeeklyDeathFirstChunk$Index_Crime,
method = "kendall")

```

```

## Warning in cor.test.default(NYWeeklyDeathFirstChunk$Weekly.Deaths,
## NYWeeklyDeathFirstChunk$Index_Crime, : Cannot compute exact p-value with ties

```

```

##
## Kendall's rank correlation tau
##
## data: NYWeeklyDeathFirstChunk$Weekly.Deaths and NYWeeklyDeathFirstChunk$Index_Crime
## z = 0.36894, p-value = 0.7122
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
## 0.0897737

```

```

cor.test(NYWeeklyDeathFirstChunk$Weekly.Deaths,
NYWeeklyDeathFirstChunk$Violent_Index_Crime, method = "kendall")

```

```

## Warning in cor.test.default(NYWeeklyDeathFirstChunk$Weekly.Deaths,
## NYWeeklyDeathFirstChunk$Violent_Index_Crime, : Cannot compute exact p-value with
## ties

```

```

##

```

```
## Kendall's rank correlation tau
##
## data: NYWeeklyDeathFirstChunk$Weekly.Deaths and NYWeeklyDeathFirstChunk$Violent_Index_Crime
## z = 0.36894, p-value = 0.7122
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.0897737
```

```
cor.test(NYWeeklyDeathFirstChunk$Weekly.Deaths,
  NYWeeklyDeathFirstChunk$Shooting_Incidents, method = "kendall")
```

```
## Warning in cor.test.default(NYWeeklyDeathFirstChunk$Weekly.Deaths,
## NYWeeklyDeathFirstChunk$Shooting_Incidents, : Cannot compute exact p-value with
## ties
```

```
##
## Kendall's rank correlation tau
##
## data: NYWeeklyDeathFirstChunk$Weekly.Deaths and NYWeeklyDeathFirstChunk$Shooting_Incidents
## z = 3.3205, p-value = 0.0008987
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.8079633
```

```
cor.test(NYWeeklyDeathFirstChunk$Weekly.Deaths,
  NYWeeklyDeathFirstChunk$Total_Complaints, method = "kendall")
```

```
## Warning in cor.test.default(NYWeeklyDeathFirstChunk$Weekly.Deaths,
## NYWeeklyDeathFirstChunk$Total_Complaints, : Cannot compute exact p-value with
## ties
```

```
##
## Kendall's rank correlation tau
##
## data: NYWeeklyDeathFirstChunk$Weekly.Deaths and NYWeeklyDeathFirstChunk$Total_Complaints
## z = -1.8447, p-value = 0.06508
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.4488685
```

Examining our correlations for the different variables with Covid deaths we find only one that is statistically significant, which is Shooting Incidents with a strong correlation coefficient of 0.81 and a p-value of 0.0009. The variable of Total Complaints (criminal) is borderline with a p-value of 0.065, and its coefficient is a considerable -0.49. In this instance, one variable (Shooting Incidents) has a positive correlation with Covid Deaths (i.e., when its values are increasing, Covid Deaths are increasing), whereas the other (Total Complaints) is negative (i.e., while its values are increasing, Covid deaths are decreasing).

Looking now at the strongest spike of Covid cases, I will once again have to truncate the dataframes to zoom in on a specific period of time. In this case it will be from November of 2021 through the end of March, 2022. Dataframe tuning similar to what was done previously will have to be repeated for this timeframe.

```

NYWeeklyCaseChunk <- NYWeeklyCase[(NYWeeklyCase$Date > "2021-11-03") &
  (NYWeeklyCase$Date < "2022-03-30"), ]
correctedCrimesByMonth2019_22case <-
  correctedCrimesByMonth2019_22[(correctedCrimesByMonth2019_22$Date >=
    "2021-11-01") & (correctedCrimesByMonth2019_22$Date <= "2022-03-31"), ]
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22case[5, ],
  correctedCrimesByMonth2019_22case[5, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[5, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[5, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[4, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[4, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[4, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[4, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[3, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[3, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[3, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[3, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[2, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[2, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[2, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[2, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[1, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[1, ])
correctedCrimesByMonth2019_22casealt <- rbind(correctedCrimesByMonth2019_22casealt,
  correctedCrimesByMonth2019_22case[1, ])
correctedCrimesByMonth2019_22casealt <- correctedCrimesByMonth2019_22casealt[, 2:4]
NYWeeklyCaseChunk <- cbind(NYWeeklyCaseChunk, correctedCrimesByMonth2019_22casealt)
crimeComplaintsTotcorrectedaltddeath <-
  data.frame(c(rep(crimeComplaintsTotcorrected$Total[15], 4),
    rep(crimeComplaintsTotcorrected$Total[16], 4),
    rep(crimeComplaintsTotcorrected$Total[17], 4),
    rep(crimeComplaintsTotcorrected$Total[18], 5),
    rep(crimeComplaintsTotcorrected$Total[19], 3)))
NYWeeklyCaseChunk <- cbind(NYWeeklyCaseChunk, crimeComplaintsTotcorrectedaltddeath)
colnames(NYWeeklyCaseChunk)[7] <- "Total_Complaints"

```

```
cor.test(NYWeeklyCaseChunk$Weekly.Cases, NYWeeklyCaseChunk$Index_Crime,  
         method = "kendall")
```

```
## Warning in cor.test.default(NYWeeklyCaseChunk$Weekly.Cases,  
## NYWeeklyCaseChunk$Index_Crime, : Cannot compute exact p-value with ties
```

```
##  
## Kendall's rank correlation tau  
##  
## data: NYWeeklyCaseChunk$Weekly.Cases and NYWeeklyCaseChunk$Index_Crime  
## z = 0.36592, p-value = 0.7144  
## alternative hypothesis: true tau is not equal to 0  
## sample estimates:  
## tau  
## 0.06328741
```

```
cor.test(NYWeeklyCaseChunk$Weekly.Cases, NYWeeklyCaseChunk$Violent_Index_Crime,  
         method = "kendall")
```

```
## Warning in cor.test.default(NYWeeklyCaseChunk$Weekly.Cases,  
## NYWeeklyCaseChunk$Violent_Index_Crime, : Cannot compute exact p-value with ties
```

```
##  
## Kendall's rank correlation tau  
##  
## data: NYWeeklyCaseChunk$Weekly.Cases and NYWeeklyCaseChunk$Violent_Index_Crime  
## z = 0.36592, p-value = 0.7144  
## alternative hypothesis: true tau is not equal to 0  
## sample estimates:  
## tau  
## 0.06328741
```

```
cor.test(NYWeeklyCaseChunk$Weekly.Cases, NYWeeklyCaseChunk$Shooting_Incidents,  
         method = "kendall")
```

```
## Warning in cor.test.default(NYWeeklyCaseChunk$Weekly.Cases,  
## NYWeeklyCaseChunk$Shooting_Incidents, : Cannot compute exact p-value with ties
```

```
##  
## Kendall's rank correlation tau  
##  
## data: NYWeeklyCaseChunk$Weekly.Cases and NYWeeklyCaseChunk$Shooting_Incidents  
## z = -1.9627, p-value = 0.04968  
## alternative hypothesis: true tau is not equal to 0  
## sample estimates:  
## tau  
## -0.3394507
```

```
cor.test(NYWeeklyCaseChunk$Weekly.Cases, NYWeeklyCaseChunk$Total_Complaints,  
         method = "kendall")
```



```
## Warning in cor.test.default(NYWeeklyCaseChunk$Weekly.Cases,
## NYWeeklyCaseChunk$Total_Complaints, : Cannot compute exact p-value with ties

##
## Kendall's rank correlation tau
##
## data: NYWeeklyCaseChunk$Weekly.Cases and NYWeeklyCaseChunk$Total_Complaints
## z = -2.761, p-value = 0.005762
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.4775323
```

This time there were two significant correlations: Shooting Incidents (but this time with a negative coefficient of -0.34 and a p-value of 0.0497), and Total Complaints (with a coefficient of -0.48 and a p-value of 0.0058). Total Complaints is much more significant this time than the last correlational analysis, but the magnitude and direction of its coefficient are roughly equivalent. Shooting Incidents is now in the opposite direction and less than half the magnitude than before, and this may be because the initial correlational analysis looked at the first major spike of Covid in 2020 (the biggest of the death spikes); this spike was counter to many of the trends later seen in the winter of 2020/2021 and 2021/2022. The correlational analysis in early 2020 took place in a timeframe before natural immunity and vaccines had become factors, and so was meaningfully different from this second analysis which concerns itself with a timeframe when natural immunity and immunizations were factors. This could help explain why the coefficients for Shooting Incidents were in opposite directions. Total Complaints however has been constant, so it seems that crime generally has a negative correlation with Covid cases and deaths, whereas Shooting Incidents have been mixed in the directions of their correlations.

As an additional note, though I've only done two analyses (one for each of the largest spikes in both Covid cases and deaths), since those two variables mimic each other in their general behaviours my tests essentially perform double duty. That is to say, even though the greatest spikes for Covid cases and Covid deaths occurred in different years, both variables were surging at the same times and so would have similar correlational results with the crime variables in question. That means that although only two analyses were done, we essentially have the results of four.

This concludes the data analysis portion of my final project, step 2. In step 3 I will bring together my insights for a final summary. Given that I have not yet learned how to implement machine learning, I cannot say if I will find a use for it in my final step.

Step 3 (Begin)

Having conducted much of the data analysis already, there is little to do in the way of deeper data diving, however I would like to finish up with a time series analysis on the data. This could not be done earlier as we had not covered that subject, but having recently done so I believe that my time series data would be well served by such an examination.

We'll begin with time series analyses examining Covid cases and deaths in New York City. Below, I check the autocorrelation, partial autocorrelation, and I use the `auto.arima` function to create the best model from my data using autoregression and moving averages as well as diffing, and then I finish up with a 10 week projection plot.

```
library(xts)
```

```
## Loading required package: zoo
```

```
##  
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':  
##  
##   as.Date, as.Date.numeric
```

```
##  
## Attaching package: 'xts'
```

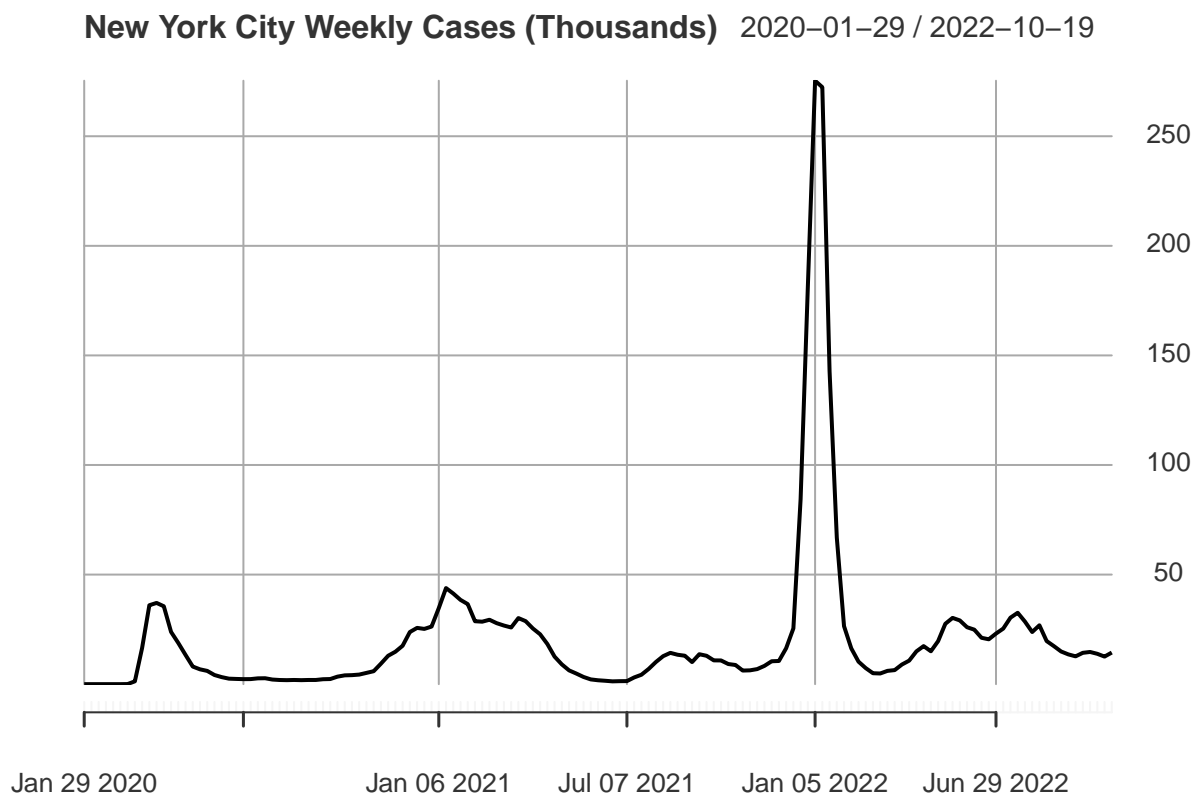
```
## The following objects are masked from 'package:dplyr':  
##  
##   first, last
```

```
library(forecast)
```

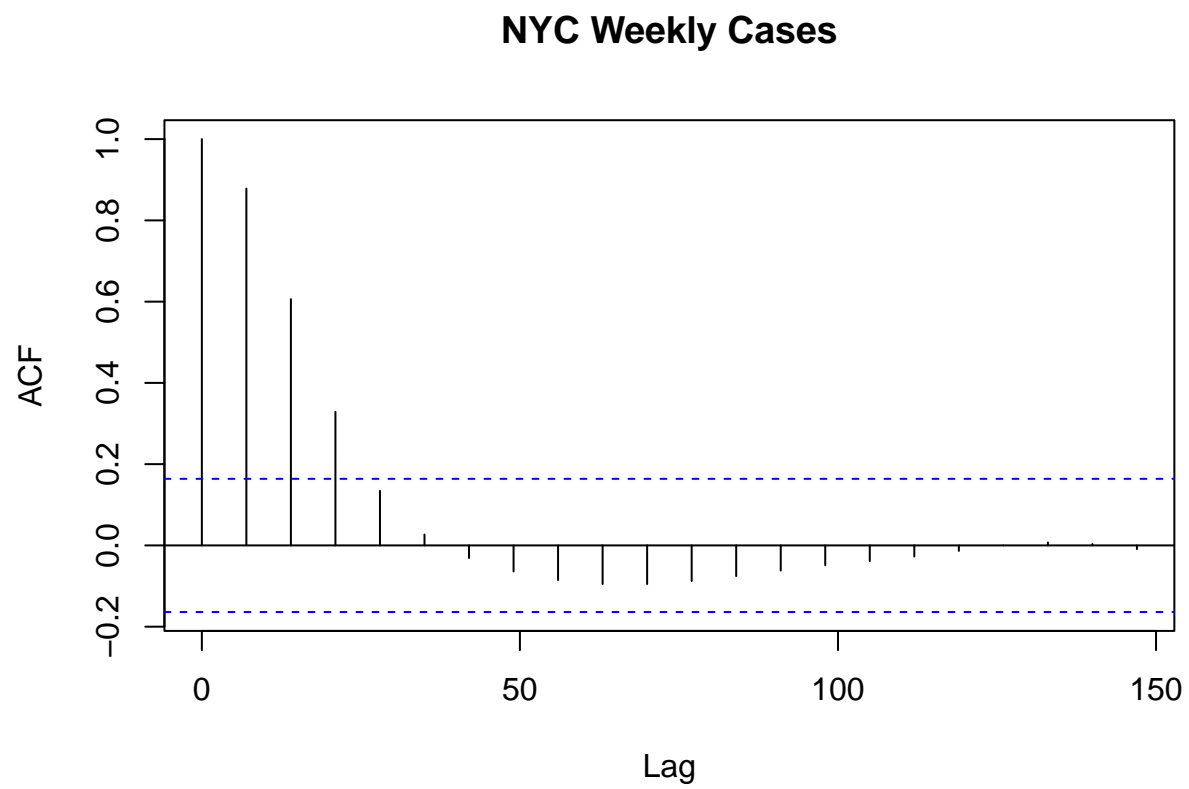
```
## Warning: package 'forecast' was built under R version 4.2.2
```

```
## Registered S3 method overwritten by 'quantmod':  
##   method      from  
##   as.zoo.data.frame zoo
```

```
NY_TimeSeries_Weekly <- select(NYWeeklyCase, -State)  
NY_TimeSeries_Weekly$Weekly.Cases <- NY_TimeSeries_Weekly$Weekly.Cases/1000  
NY_TimeSeries_Weekly <- xts(NY_TimeSeries_Weekly$Weekly.Cases, NY_TimeSeries_Weekly$Date)  
plot(NY_TimeSeries_Weekly, main = "New York City Weekly Cases (Thousands)", yaxis.left = FALSE)
```

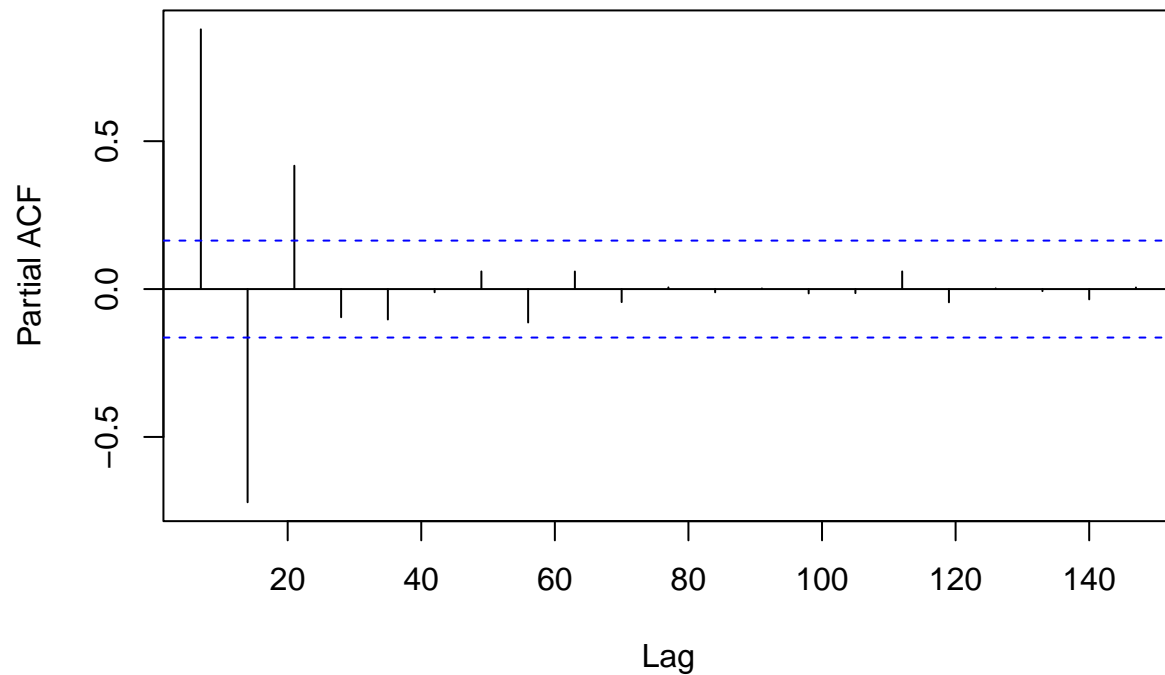


```
acf(NY_TimeSeries_Weekly, main = "NYC Weekly Cases")
```



```
pacf(NY_TimeSeries_Weekly, main = "NYC Weekly Cases")
```

NYC Weekly Cases



```
NYBest <- auto.arima(NY_TimeSeries_Weekly)
NYBest
```

```
## Series: NY_TimeSeries_Weekly
## ARIMA(1,0,3) with non-zero mean
##
## Coefficients:
##          ar1      ma1      ma2      ma3      mean
##          0.5287  1.3554  1.0262  0.4349  20.0645
## s.e.    0.0965  0.1027  0.1406  0.0919   7.1528
##
## sigma^2 = 119.4: log likelihood = -544.24
## AIC=1100.47   AICc=1101.09   BIC=1118.25
```

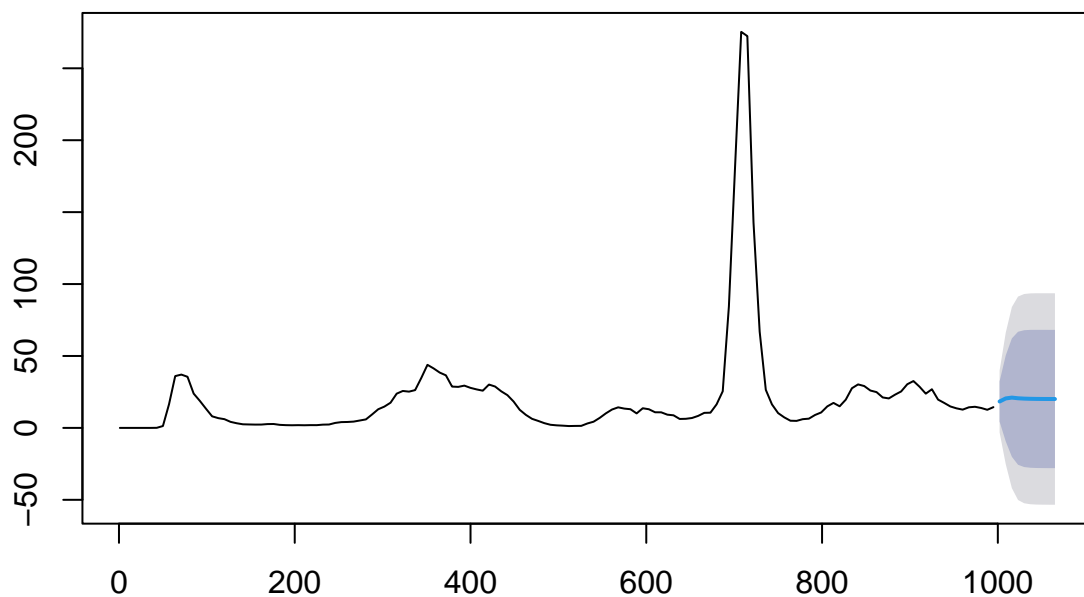
```
predict(NYBest, n.ahead=10, se.fit = TRUE)
```

```
## $pred
## Time Series:
## Start = 1002
## End = 1065
## Frequency = 0.142857142857143
## [1] 18.34465 20.47858 21.05934 20.59050 20.34261 20.21154 20.14224 20.10559
## [9] 20.08622 20.07598
##
## $se
```

```
## Time Series:
## Start = 1002
## End = 1065
## Frequency = 0.142857142857143
## [1] 10.92700 23.30787 32.11890 36.08023 37.11211 37.39549 37.47433 37.49634
## [9] 37.50249 37.50421
```

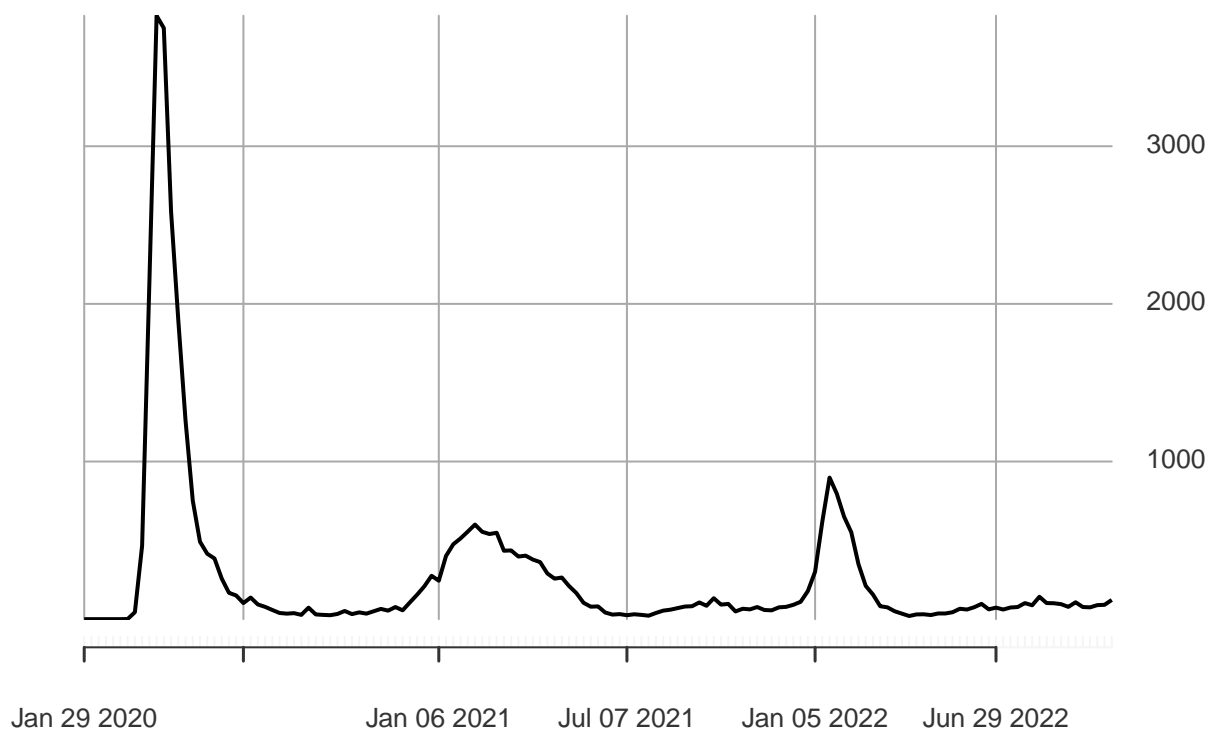
```
NYCaseForecast <- forecast(NYBest, h = 10)
plot(NYCaseForecast)
```

Forecasts from ARIMA(1,0,3) with non-zero mean



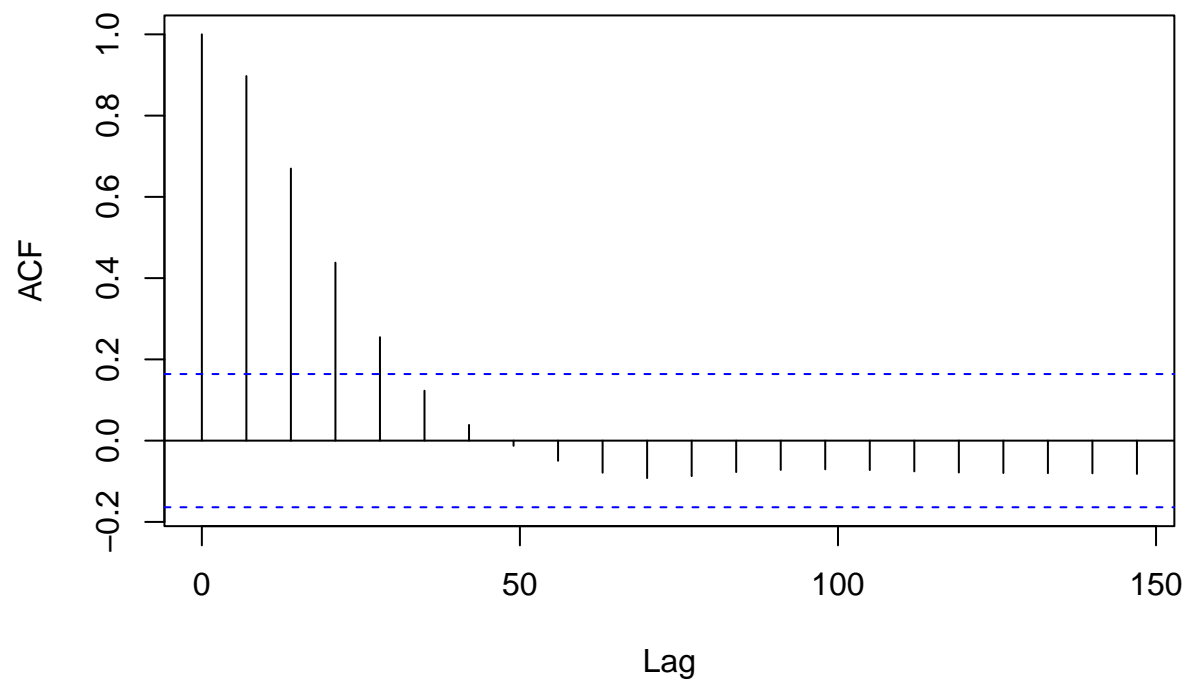
```
library(xts)
library(forecast)
NY_TimeSeries_WeeklyD <- select(NYWeeklyDeath, -State)
NY_TimeSeries_WeeklyD$Weekly.Deaths <- NY_TimeSeries_WeeklyD$Weekly.Deaths
NY_TimeSeries_WeeklyD <- xts(NY_TimeSeries_WeeklyD$Weekly.Deaths, NY_TimeSeries_WeeklyD$Date)
plot(NY_TimeSeries_WeeklyD, main = "New York City Weekly Deaths Time Series", yaxis.left = FALSE)
```

New York City Weekly Deaths Time Series 2020-01-29 / 2022-10-19



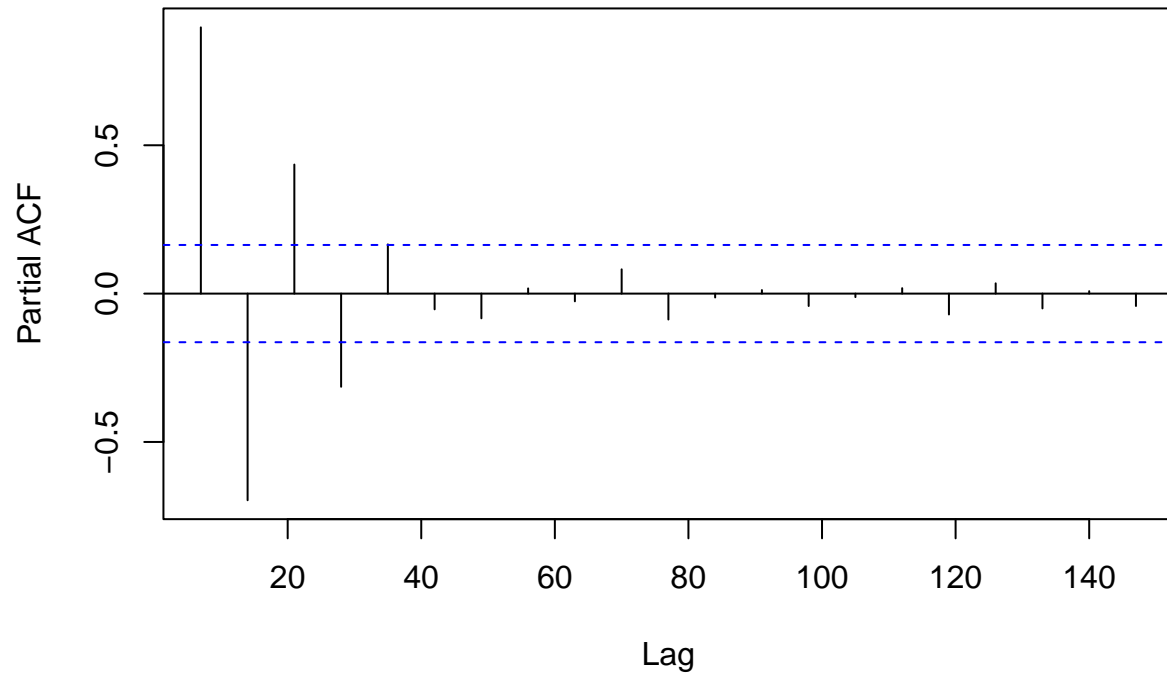
```
acf(NY_TimeSeries_WeeklyD, main = "NYC Weekly Deaths")
```

NYC Weekly Deaths



```
pacf(NY_TimeSeries_WeeklyD, main = "NYC Weekly Deaths")
```

NYC Weekly Deaths



```
NYBestD <- auto.arima(NY_TimeSeries_WeeklyD)
NYBestD
```

```
## Series: NY_TimeSeries_WeeklyD
## ARIMA(0,1,2)
##
## Coefficients:
##      ma1      ma2
##      1.2556  0.5292
## s.e.  0.0718  0.0760
##
## sigma^2 = 23192: log likelihood = -915.03
## AIC=1836.07  AICc=1836.24  BIC=1844.93
```

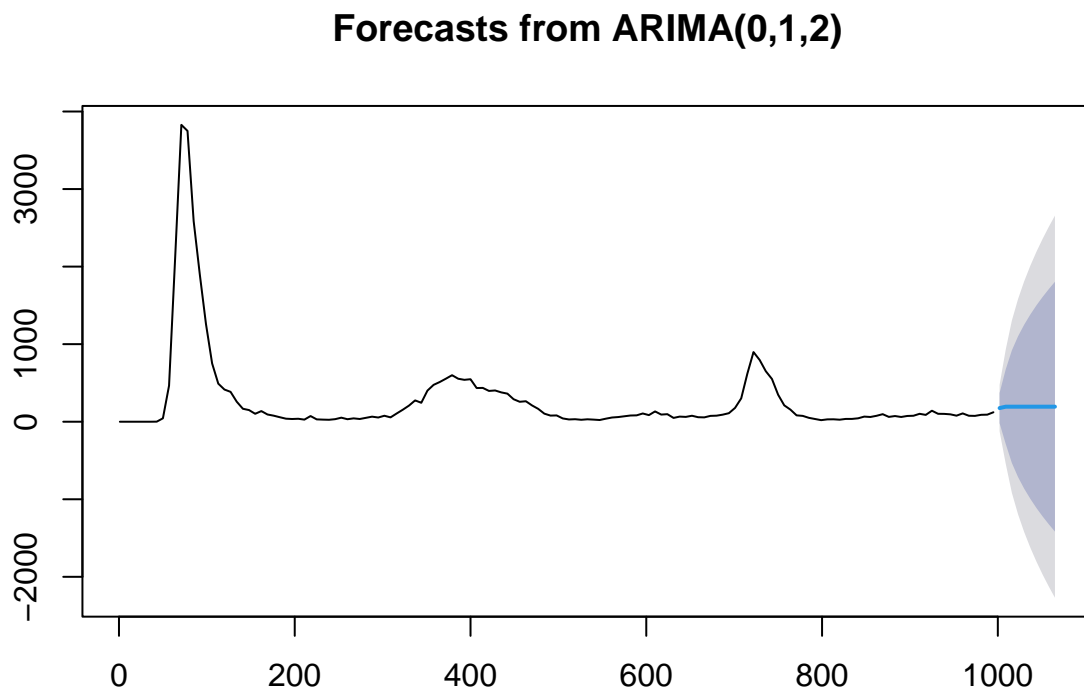
```
predict(NYBestD, n.ahead=10, se.fit = TRUE)
```

```
## $pred
## Time Series:
## Start = 1002
## End = 1065
## Frequency = 0.142857142857143
## [1] 175.7221 193.7641 193.7641 193.7641 193.7641 193.7641 193.7641 193.7641
## [9] 193.7641 193.7641
##
## $se
```



```
## Time Series:
## Start = 1002
## End = 1065
## Frequency = 0.142857142857143
## [1] 152.2894 375.7483 566.6102 707.7482 825.0866 927.7005 1020.0435
## [8] 1104.6941 1183.3044 1257.0083
```

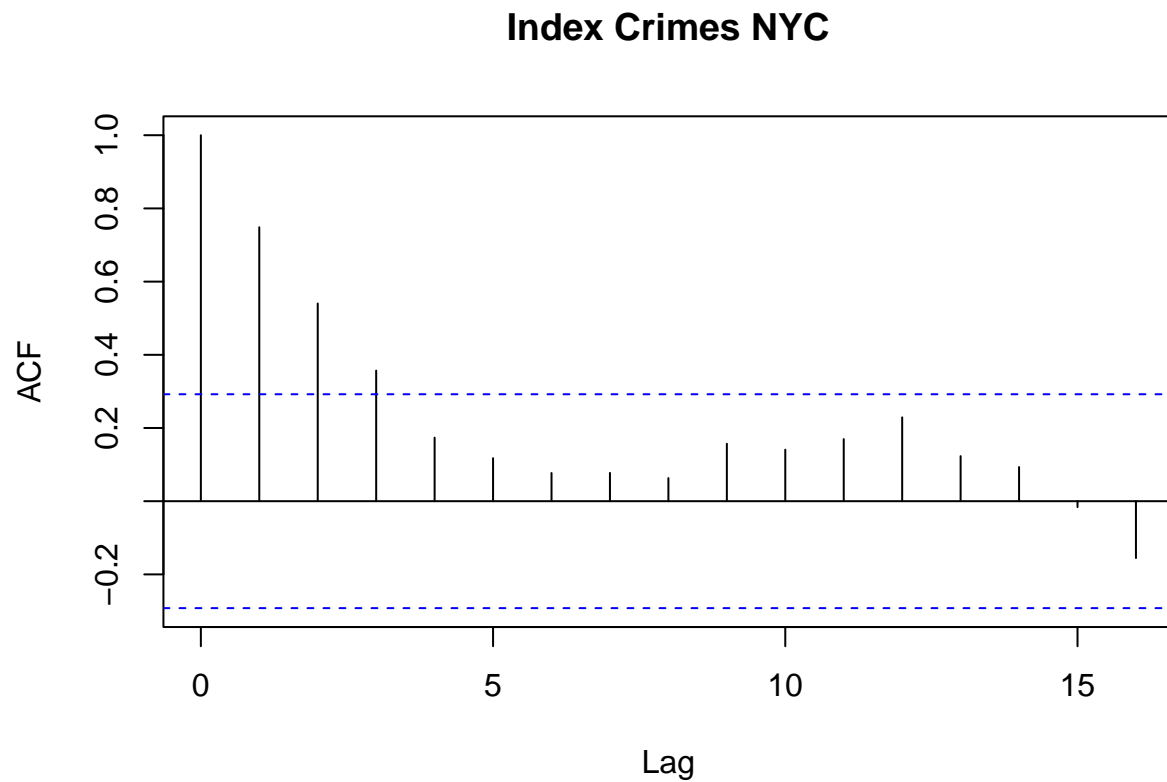
```
NYCaseForecastD <- forecast(NYBestD, h = 10)
plot(NYCaseForecastD)
```



The autocorrelations show significant relationships amongst the time series data going back several lags, whereas the partial autocorrelation shows that these connections are not as strong as initially supposed, and even see-saw between positive and negative correlations. For both weekly cases and weekly deaths, the resulting projections both have fairly wide confidence intervals, even just 10 weeks out. While the `auto.arima` function does create models that project into the negative territory, it is obviously not the case that we could have negative numbers in this data set.

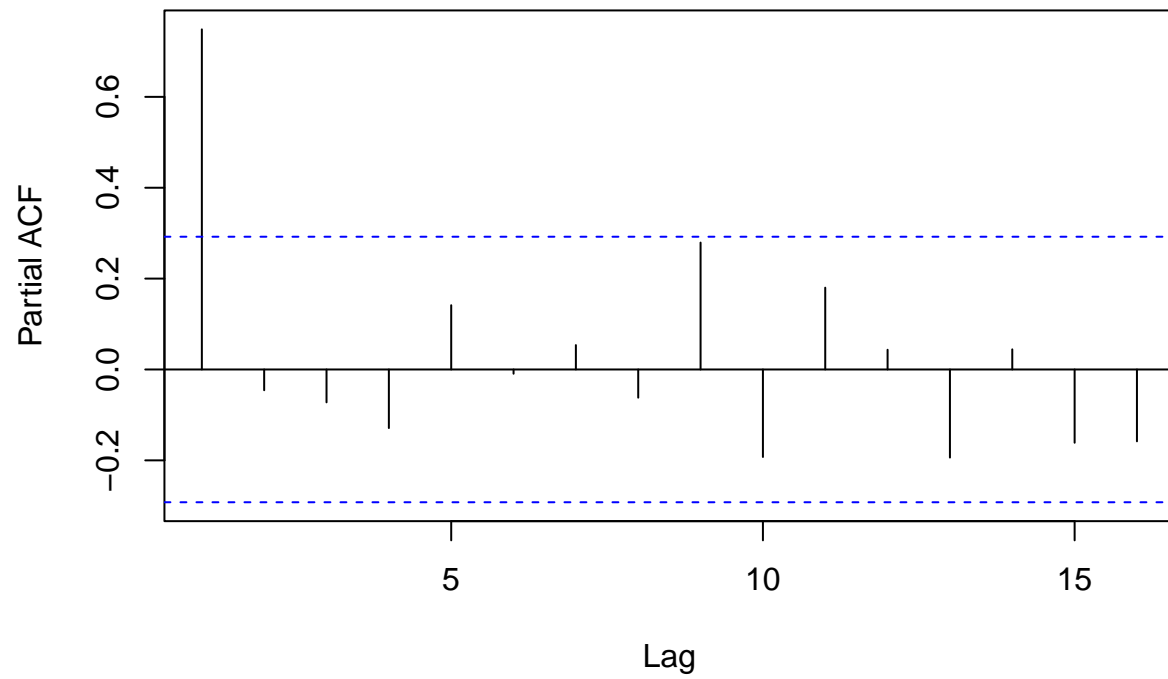
I then proceed to assessing the crime data (Index Crime, Violent Index Crime, and Shooting Incidents) as time series'. With all three of these categories, however, we find the `auto.arima` output to contain ARIMA (0,1,0), which is to say a “random walk”. In “random walks” it is the case that the data exhibits autocorrelation (as is shown by the acf), but is essentially as random as a person aimlessly strolling. This is evidenced by applying the diff of 1 (which is the “1” in the ARIMA output of (0,1,0)) and plotting that to find essentially noise. That the pacf shows few to no significant correlations (depending on which category of crime we are inspecting) demonstrates that the acf outputs are essentially self-reinforcing, and not indicative of a pattern.

```
crimesByMonth19_22_Date<- crimesByMonth2019_22
crimesByMonth19_22_Date$Date<- ymd(crimesByMonth2019_22$Date)
crimesByMonth19_22_Indexts <- xts(crimesByMonth19_22_Date$Index_Crime, crimesByMonth19_22_Date$Date)
acf(crimesByMonth19_22_Indexts, main = "Index Crimes NYC")
```



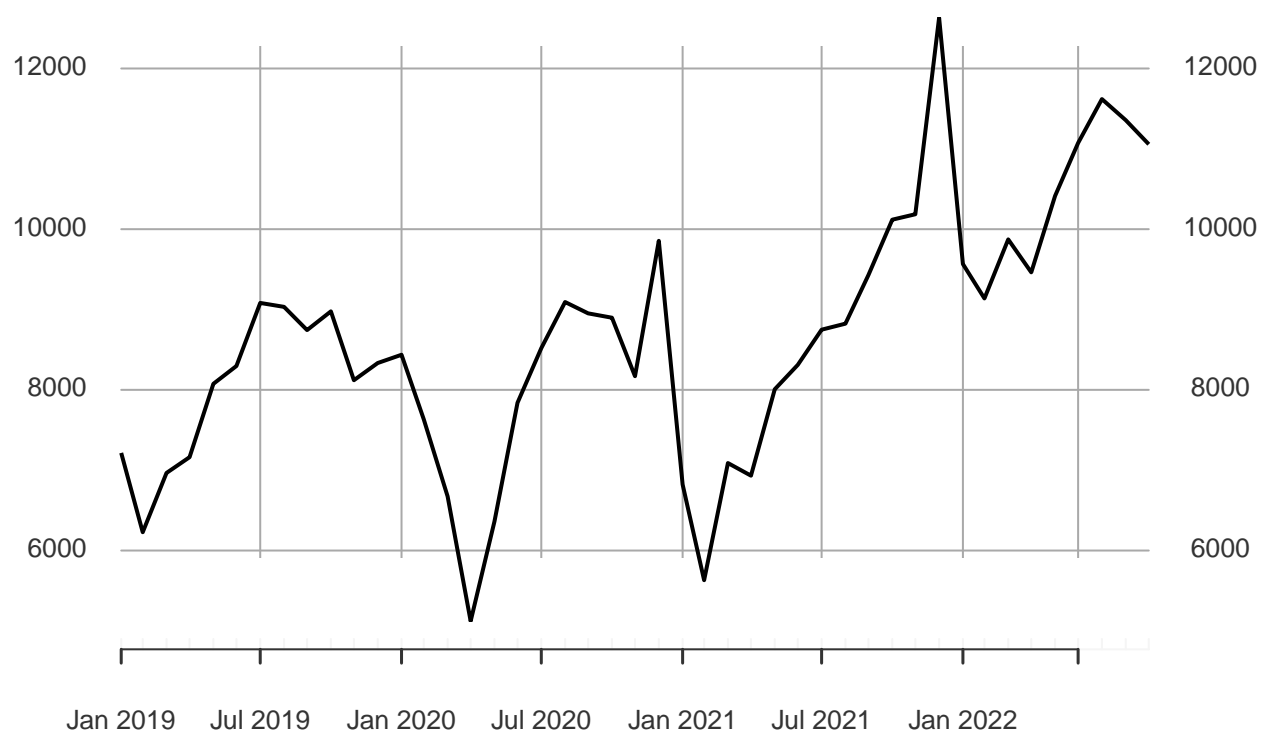
```
pacf(crimesByMonth19_22_Indexts, main = "Index Crimes NYC")
```

Index Crimes NYC



```
plot(crimesByMonth19_22_Indexts, main = "Index Crimes In New York City Time Series")
```

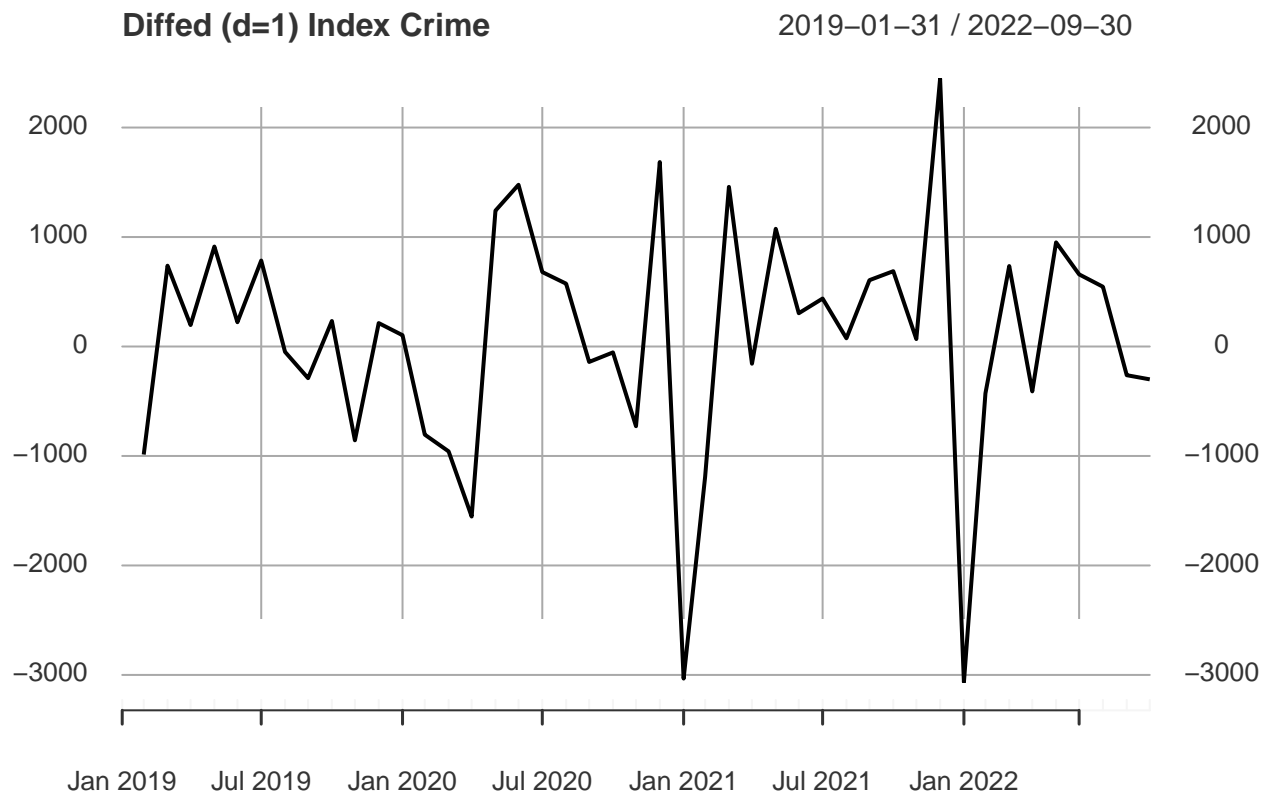
Index Crimes In New York City Time Series 2019-01-31 / 2022-09-30



```
bestIndexC <- auto.arima(crimesByMonth19_22_Indexts)
bestIndexC
```

```
## Series: crimesByMonth19_22_Indexts
## ARIMA(0,1,0)
##
## sigma^2 = 1113624: log likelihood = -368.74
## AIC=739.48 AICc=739.58 BIC=741.27
```

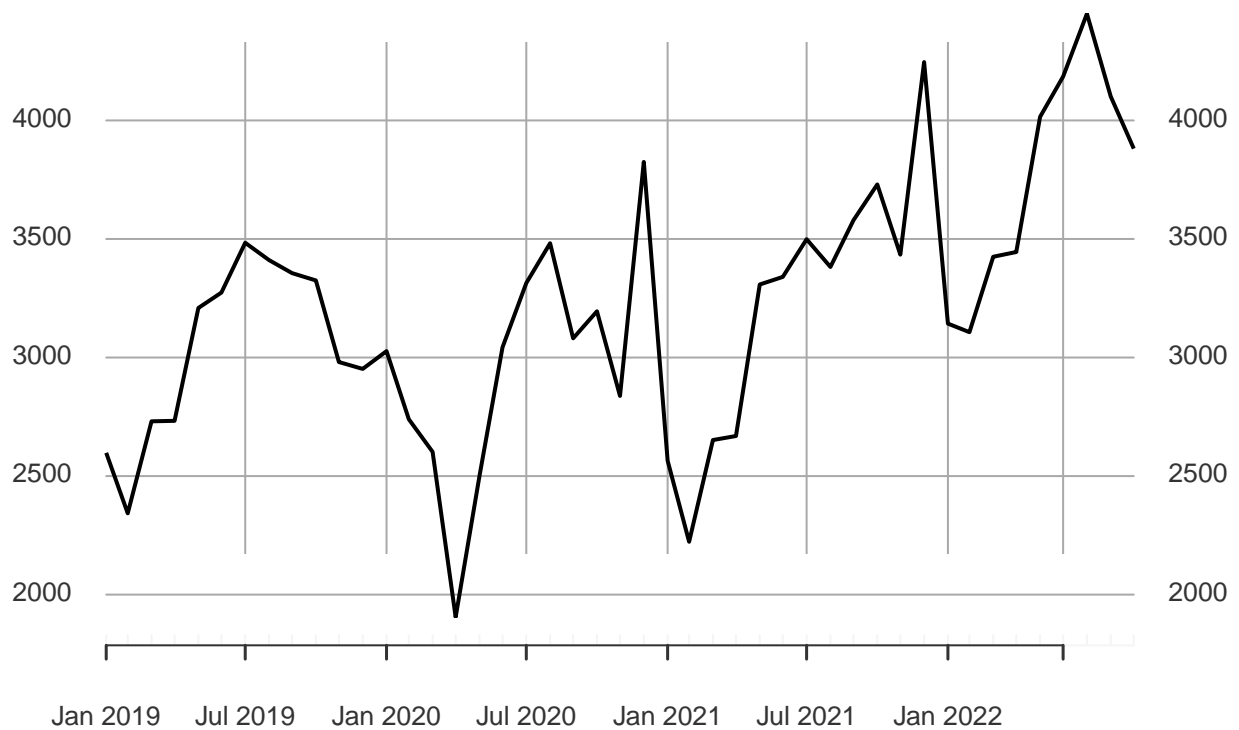
```
cbmdiffed <- diff(crimesByMonth19_22_Indexts, 1)
plot(cbmdiffed, main = "Diffed (d=1) Index Crime")
```



```
crimesByMonth19_22_Violent <- xts(crimesByMonth19_22_Date$Violent_Index_Crime, crimesByMonth19_22_Date$Date)
plot(crimesByMonth19_22_Violent, main = "NYC Violent Crime")
```

NYC Violent Crime

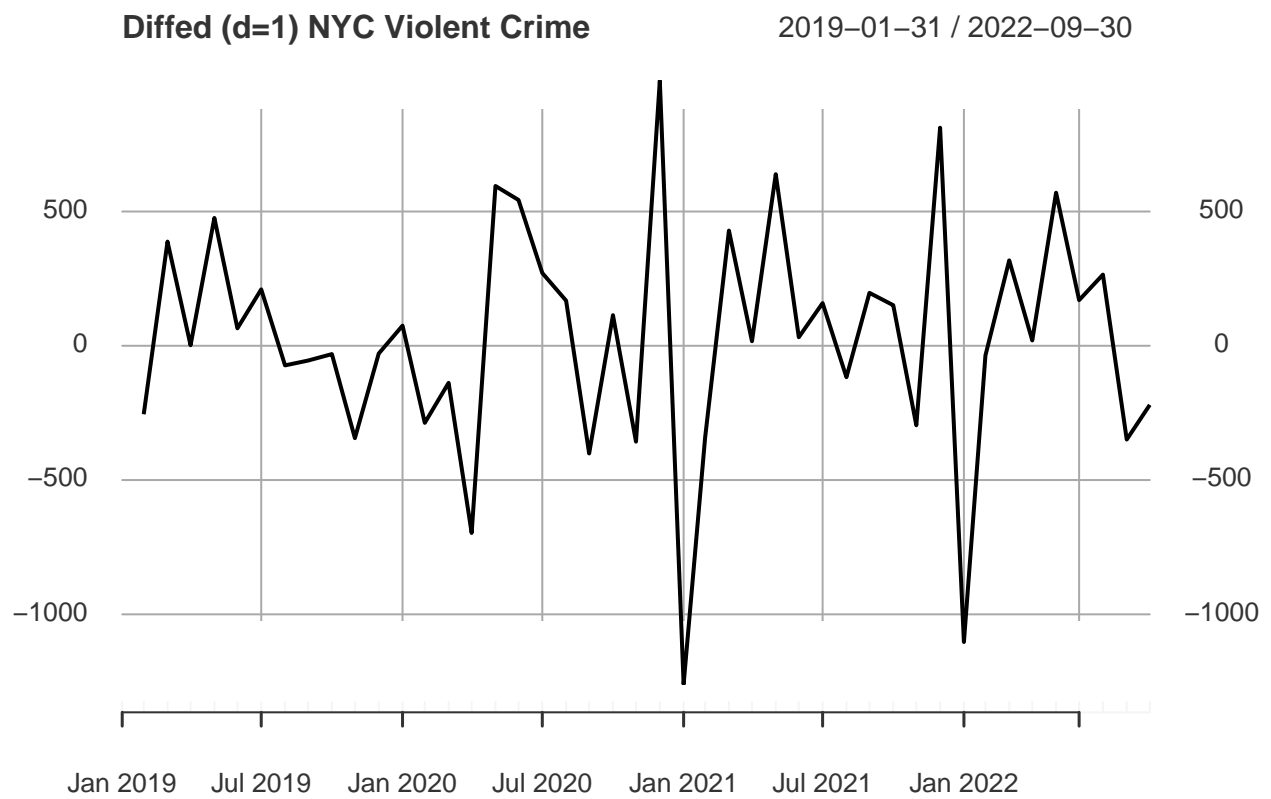
2019-01-31 / 2022-09-30



```
bestIndexV <- auto.arima(crimesByMonth19_22_Violent)
bestIndexV
```

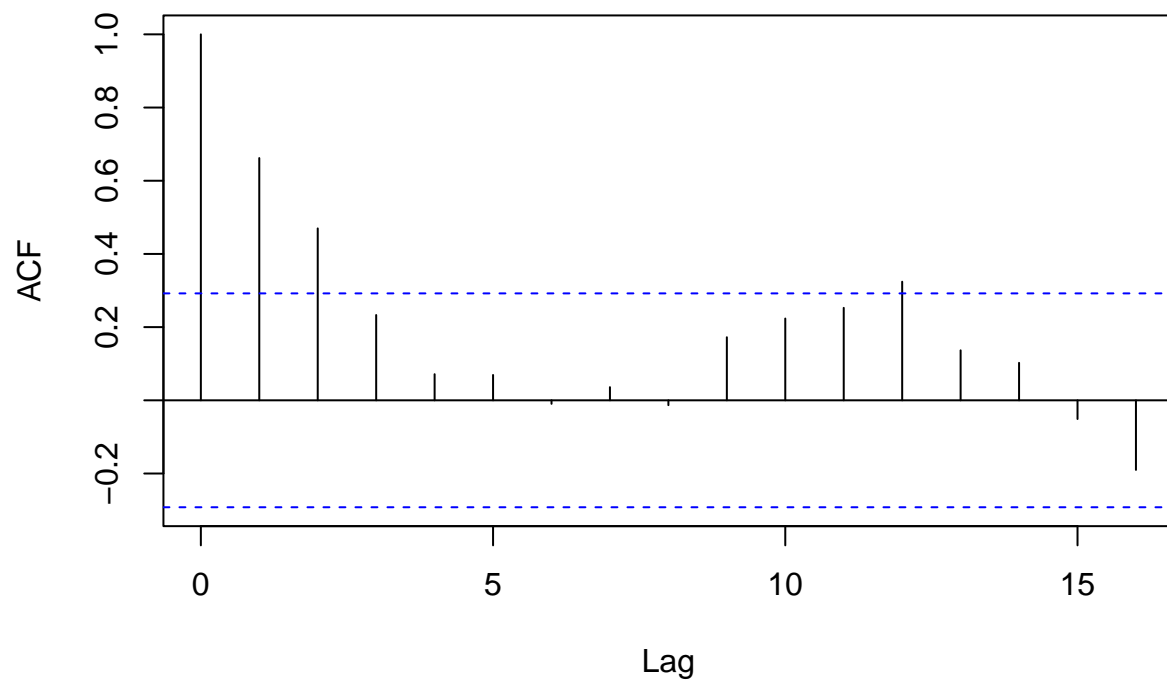
```
## Series: crimesByMonth19_22_Violent
## ARIMA(0,1,0)
##
## sigma^2 = 188573: log likelihood = -329.67
## AIC=661.35 AICc=661.44 BIC=663.13
```

```
cbmvdifffed <- diff(crimesByMonth19_22_Violent, 1)
plot(cbmvdifffed, main = "Difffed (d=1) NYC Violent Crime")
```



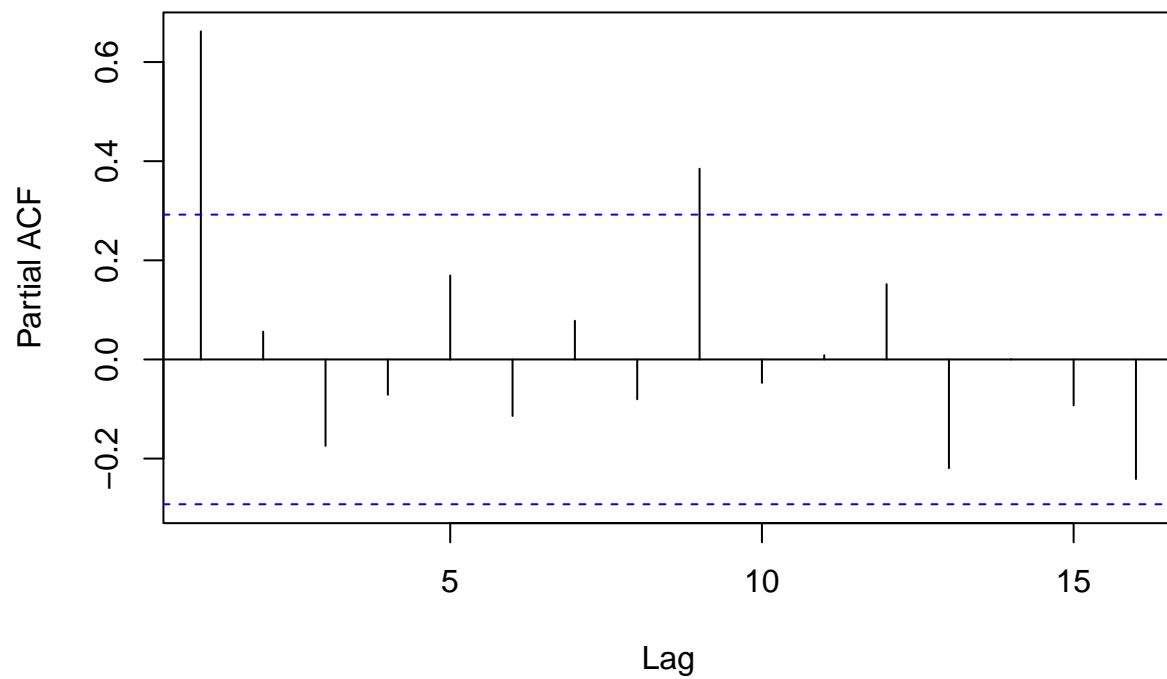
```
acf(crimesByMonth19_22_Violent, main = "Violent Crime")
```

Violent Crime



```
pacf(crimesByMonth19_22_Violent, main = "Violent Crime")
```

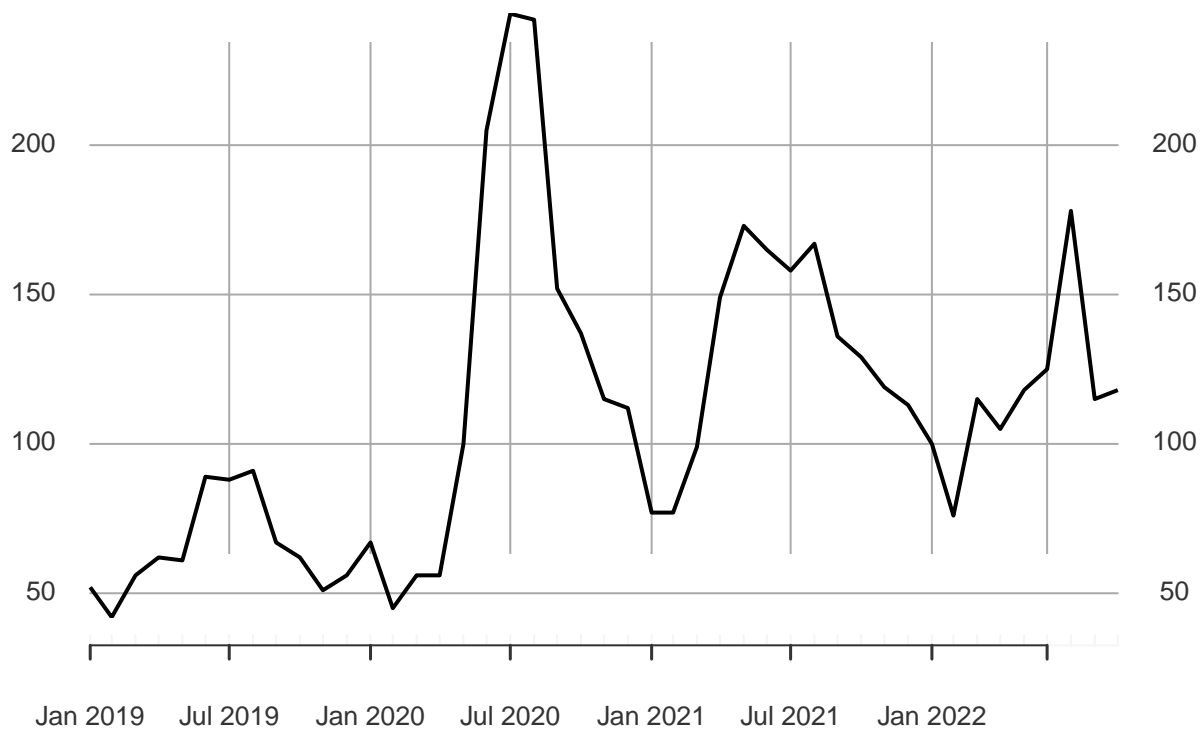

Violent Crime



```
crimesByMonth19_22_Shoot <- select(crimesByMonth19_22_Date, c(Date, Shooting_Incidents))
crimesByMonth19_22_Shoots <- xts(crimesByMonth19_22_Shoot$Shooting_Incidents, crimesByMonth19_22_Shoot$Date)
plot(crimesByMonth19_22_Shoots, main = "Shooting Incidents NYC Time Series")
```

Shooting Incidents NYC Time Series

2019-01-31 / 2022-09-30

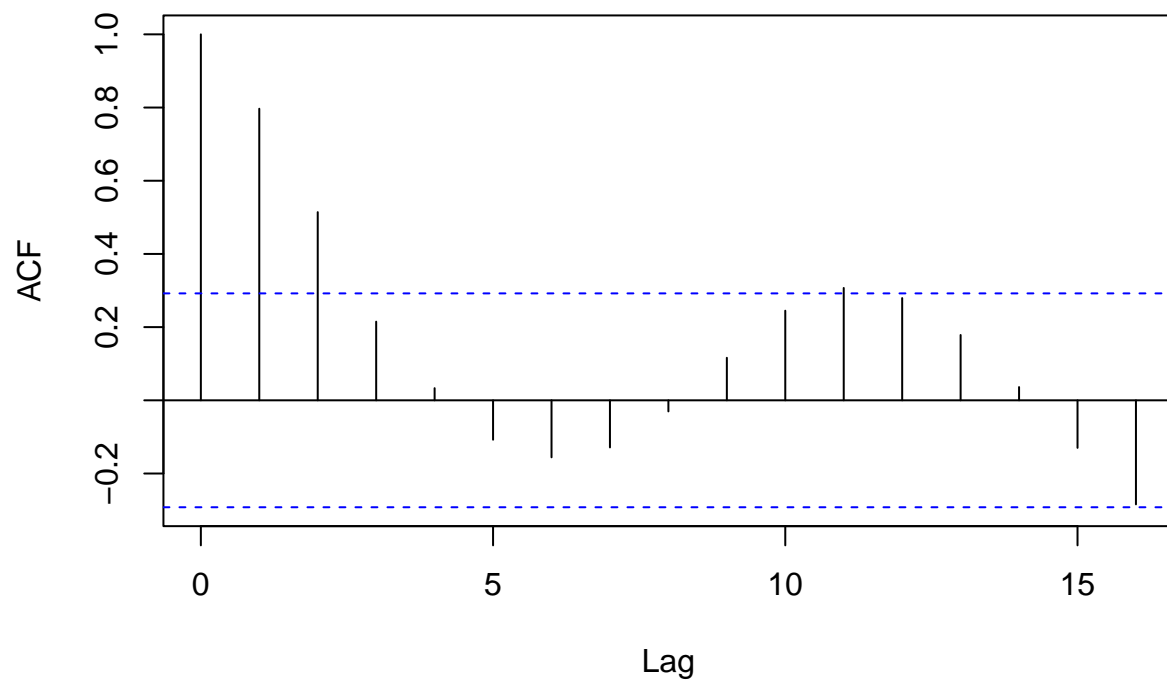


```
bestShoot <- auto.arima(crimesByMonth19_22_Shoots)
bestShoot
```

```
## Series: crimesByMonth19_22_Shoots
## ARIMA(0,1,0)
##
## sigma^2 = 940.9: log likelihood = -213.06
## AIC=428.13 AICc=428.22 BIC=429.91
```

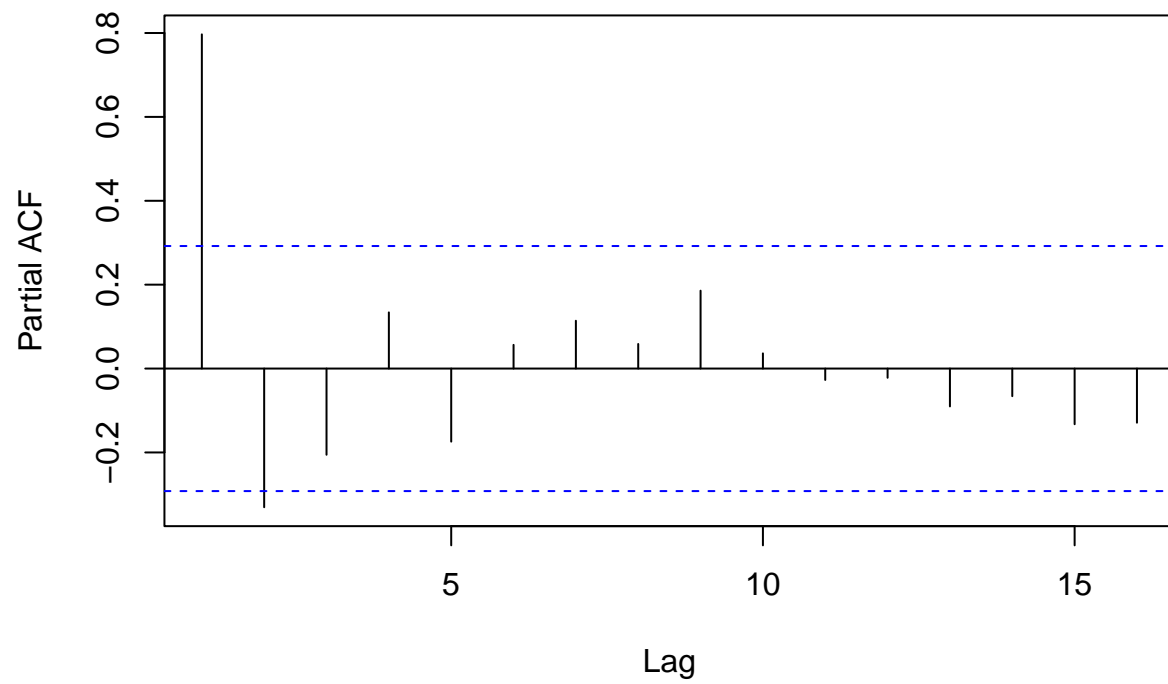
```
acf(crimesByMonth19_22_Shoots, main = "Shooting Incidents")
```

Shooting Incidents

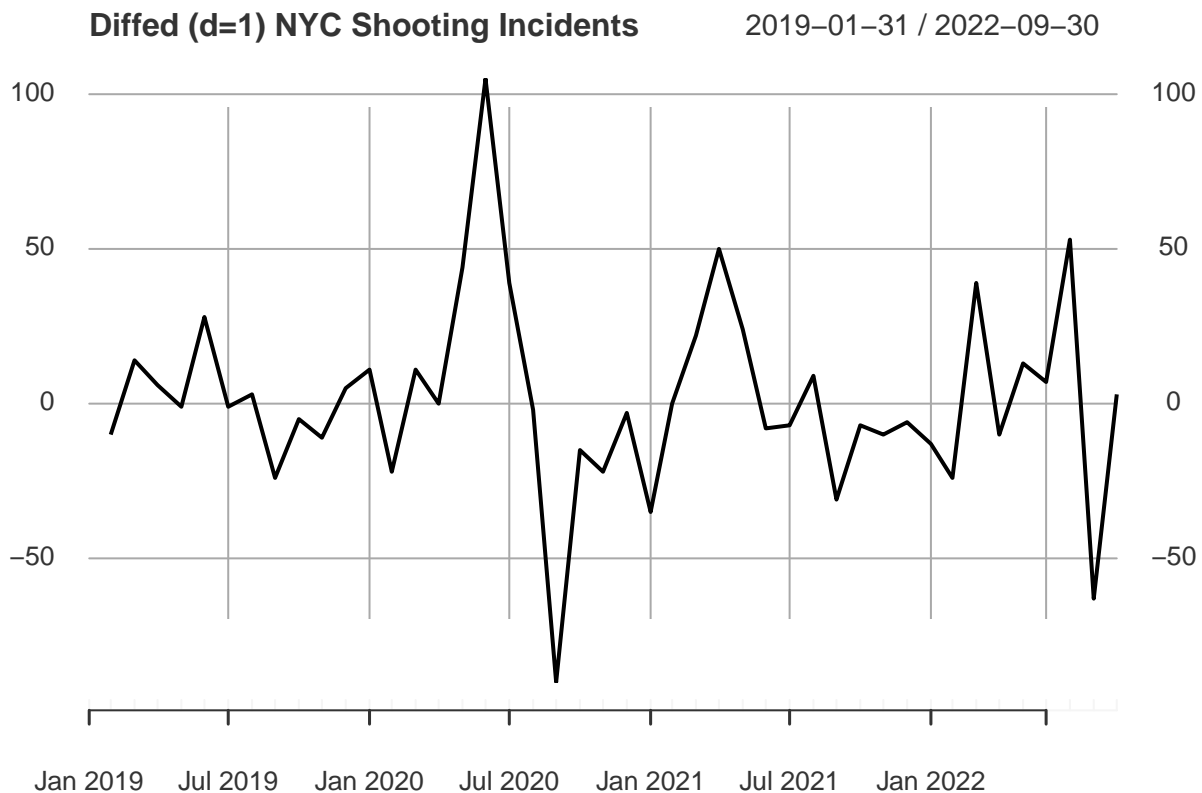


```
pacf(crimesByMonth19_22_Shoots, main = "Shooting Incidents")
```

Shooting Incidents



```
cbmSdiffed <- diff(crimesByMonth19_22_Shoots, 1)
plot(cbmSdiffed, main = "Differenced (d=1) NYC Shooting Incidents")
```



Having now conducted my time series analyses, I am free to move on to the presentation of my data and findings.

My Data Journey

The relationships between crime and many other factors has been oft-cited but seldom proved. It is assumed, for instance, that poverty leads to crime, and not vice-versa, once a correlation between the two has been shown to exist. Many such correlations can and have been collected and produced throughout the centuries, sometimes out of genuine concern for human well-being, and other times for the more base end of aiding ambition. I asked the question “Is there a quantitative impact of Covid 19 on crime in New York City?” This question allowed for positive and negative correlations, weak or strong correlations; any correlations at all, really. And having searched for weeks through the data I collected I am finally ready to give my answer. But first, a recap on the data used.

For the Covid-19 data I was able to acquire weekly Covid case numbers as well as weekly death numbers for New York City; this was collected from the CDC. Only minor data cleaning was needed for one extreme outlier. As for the crime data, I was able to collect crime statistics from the New York State Division of Criminal Justice Services. This data ranged from annualized crime statistics on various categories of crime, to monthly statistics on the same, as well as data on criminal complaints. Not all of the data I collected was used or found to be useful. For instance, given the time available and the scope of my inquiry, I did not spend much time analyzing the crime statistics as broken down by borough, and rather focused on the overall city itself. Some transformation of the data was found to be useful, such as joining similar annualized datasets into larger composite ones so that progression over a period of time greater than one year could be assessed. I also at times had to correct various crime-statistics when plotting them together so that their

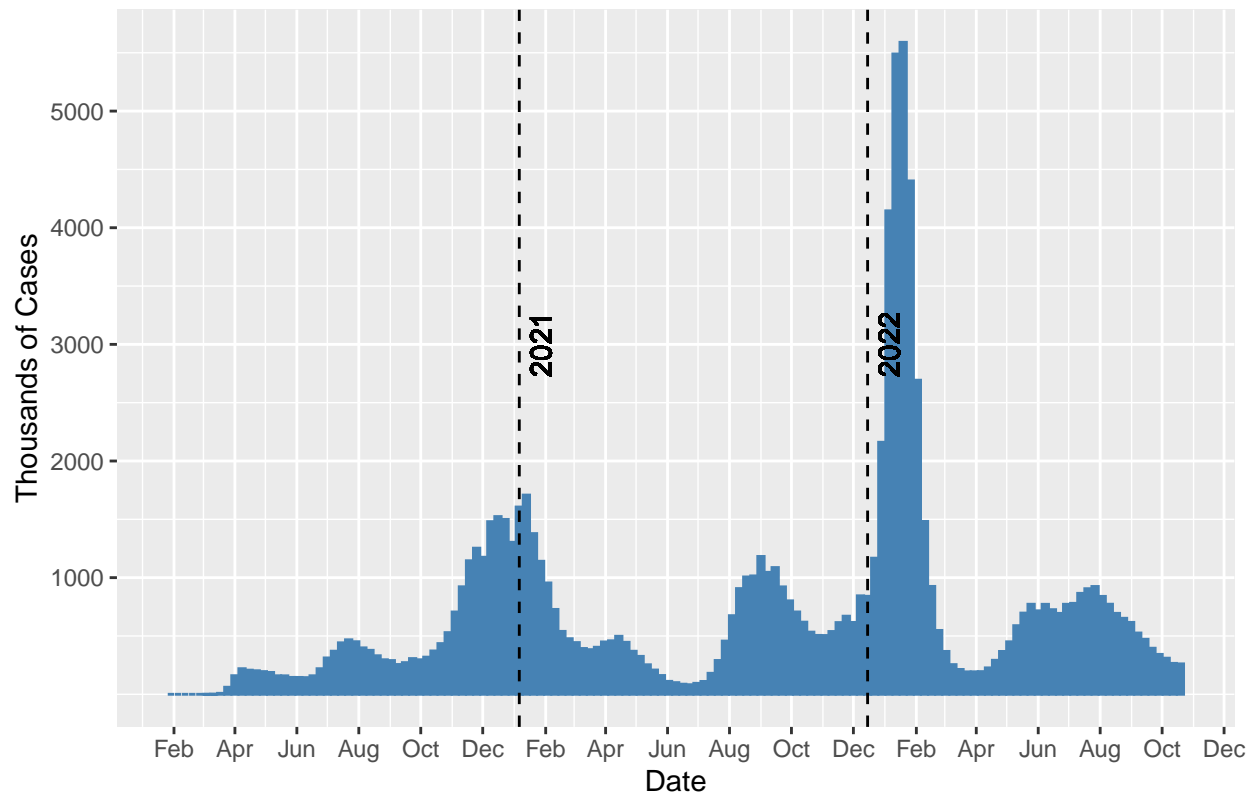
resulting plotted graphs would scale properly.

The methodology for the employment of this corpus of data was to graphically examine these various phenomena of crime and sickness over time for any patterns or trends, as well as examining crime statistics and Covid together graphically for any type of correlations (positive or negative) both overall and seasonally. After the graphical examinations I also turned to Kendall's Tau to look for any quantitative correlations between Covid and crime during the largest spikes of both weekly Covid cases and weekly Covid deaths. Finally, I performed some time series analyses on the Covid case and death data, as well as my three strongest performing crime statistics. What I found is presented below.

First, I asked myself if there was anything unique about the Covid experience in New York City, specifically as could be gleaned from weekly case and death data. What I found is that although the national and the city-wide trends for those parameters were very similar, New York City experienced both stronger death and case spikes as relative to the other surges in its own trends when compared to the much more mild and even trends of the country overall. What this meant, was that the initial surge of deaths in New York City was much stronger relative to any other surges of death in in that place than could be said of the surges of death statistics for the country overall. Similarly, in early 2022, New York City experienced a tremendous surge in Covid cases relative to any other surges seen before in that city, as opposed to the much more even ebbs and surges of the national trend. I decided that although there were cyclical trends in both the surges and ebbs of Covid cases and deaths in New York City, I would be well served to pay special attention to these spikes as any relationships or effects that would be otherwise too small to notice might be evident during these times. There was, of course, the obvious visual correlation between cases and deaths, though what was truly odd was that the two greatest spikes (i.e., the one for weekly cases, and the one for weekly deaths) did not coincide with each other. I reasoned that this was likely because initially there was no natural immunity or vaccines available, and so the disease was at its peak lethality, and that later when Covid cases strongly surged, perhaps due to the more transmissible variants that later emerged, not only was natural immunity present but so were vaccines and so the death tolls were more subdued. For easy reference, I have included the relevant graphs below.

```
## Warning: Use of 'USWeeklyCase$Date' is discouraged.  
## i Use 'Date' instead.  
## Use of 'USWeeklyCase$Date' is discouraged.  
## i Use 'Date' instead.
```

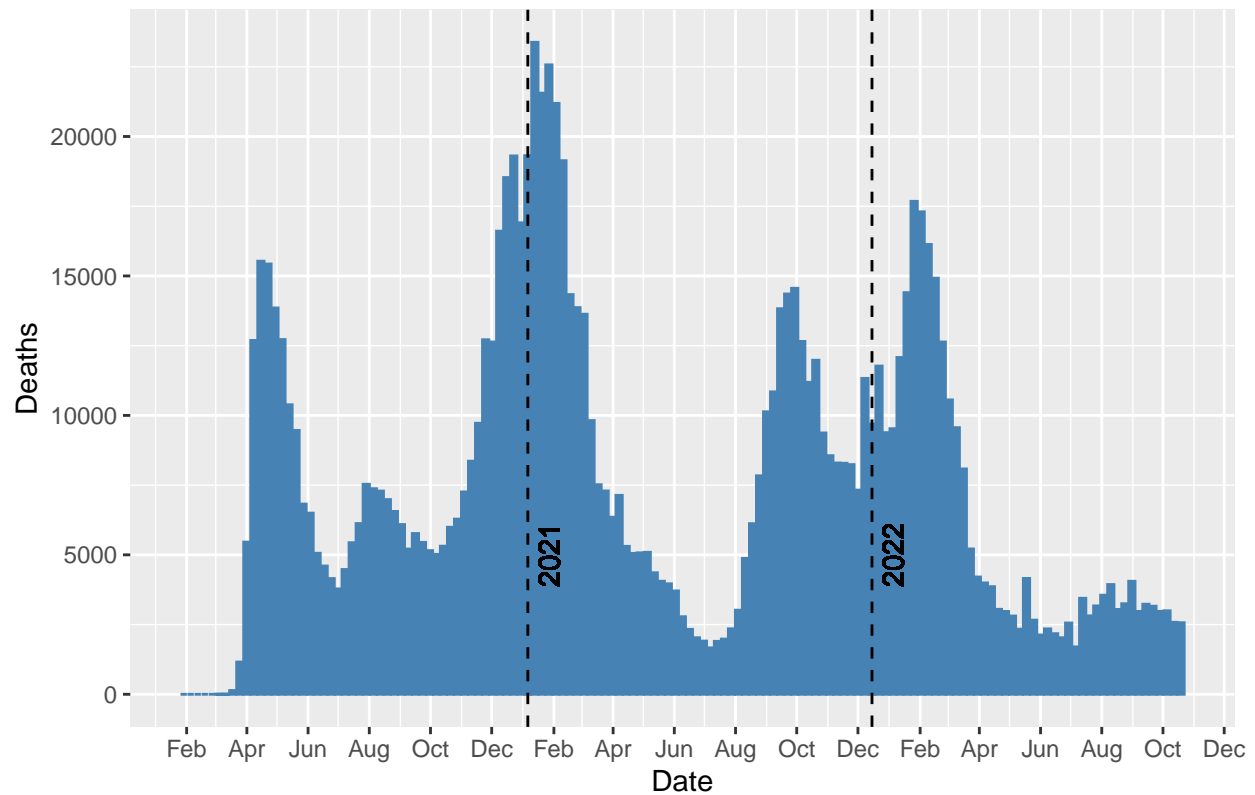
US Weekly Covid-19 Cases (Thousands)



```
## Warning: Use of 'USWeeklyDeath$Date' is discouraged.  
## i Use 'Date' instead.
```

```
## Warning: Use of 'USWeeklyDeath$Date' is discouraged.  
## i Use 'Date' instead.
```

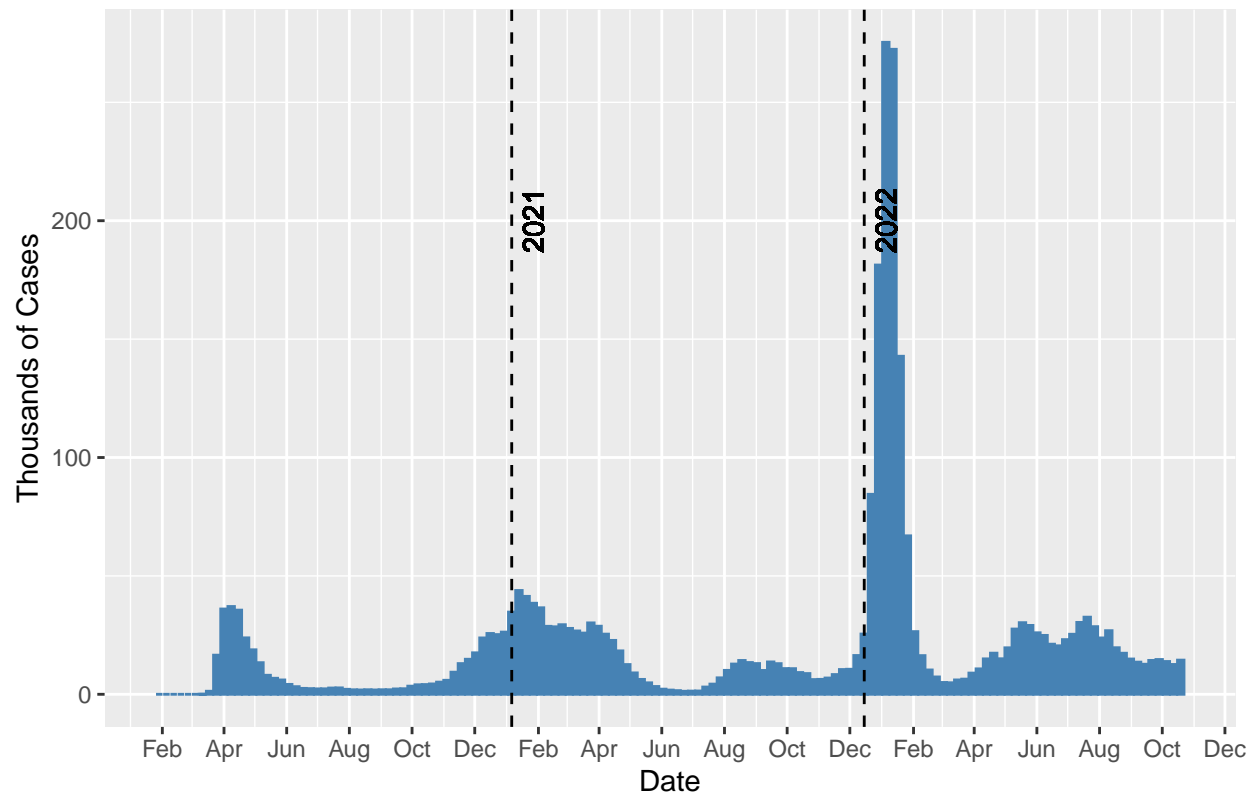
US Weekly Covid-19 Deaths



```
## Warning: Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```

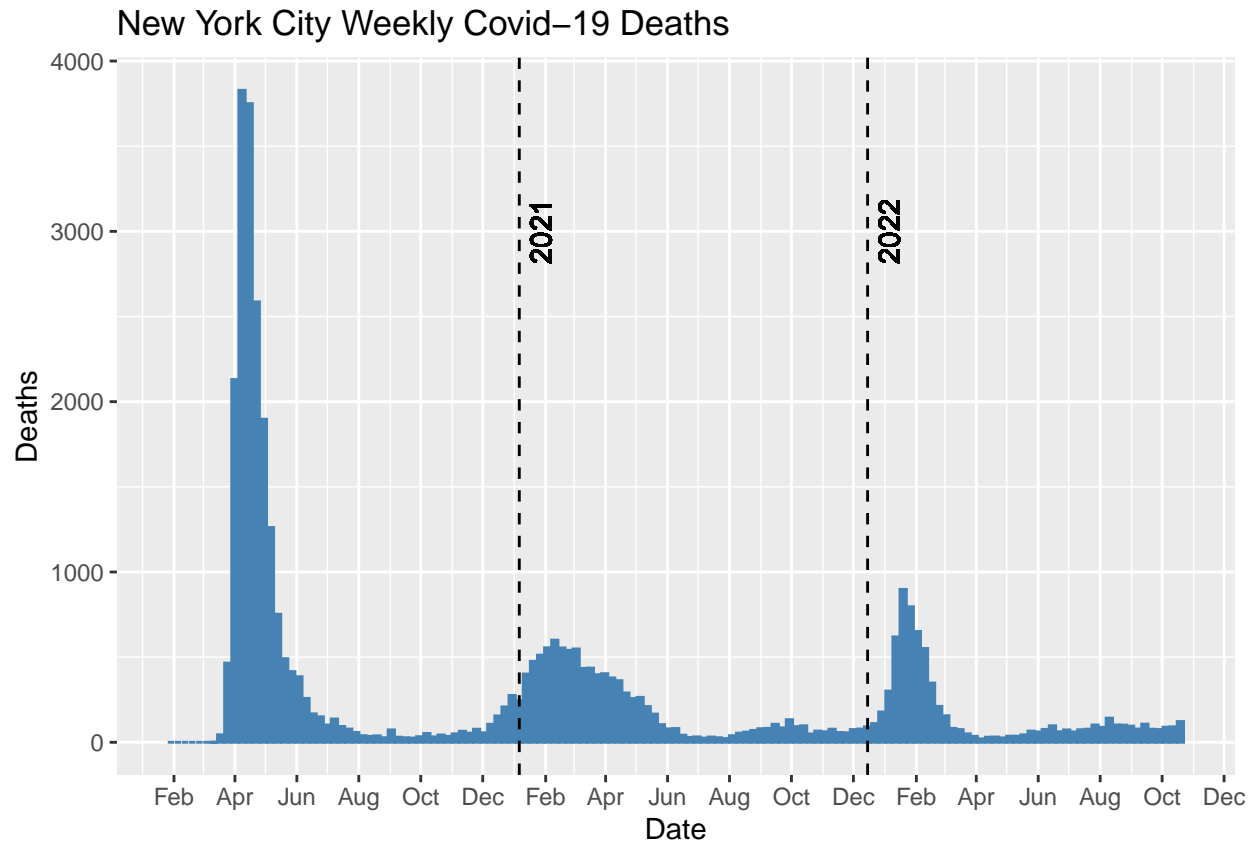
```
## Warning: Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```


New York City Weekly Covid-19 Cases (Thousands)



```
## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```

```
## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```



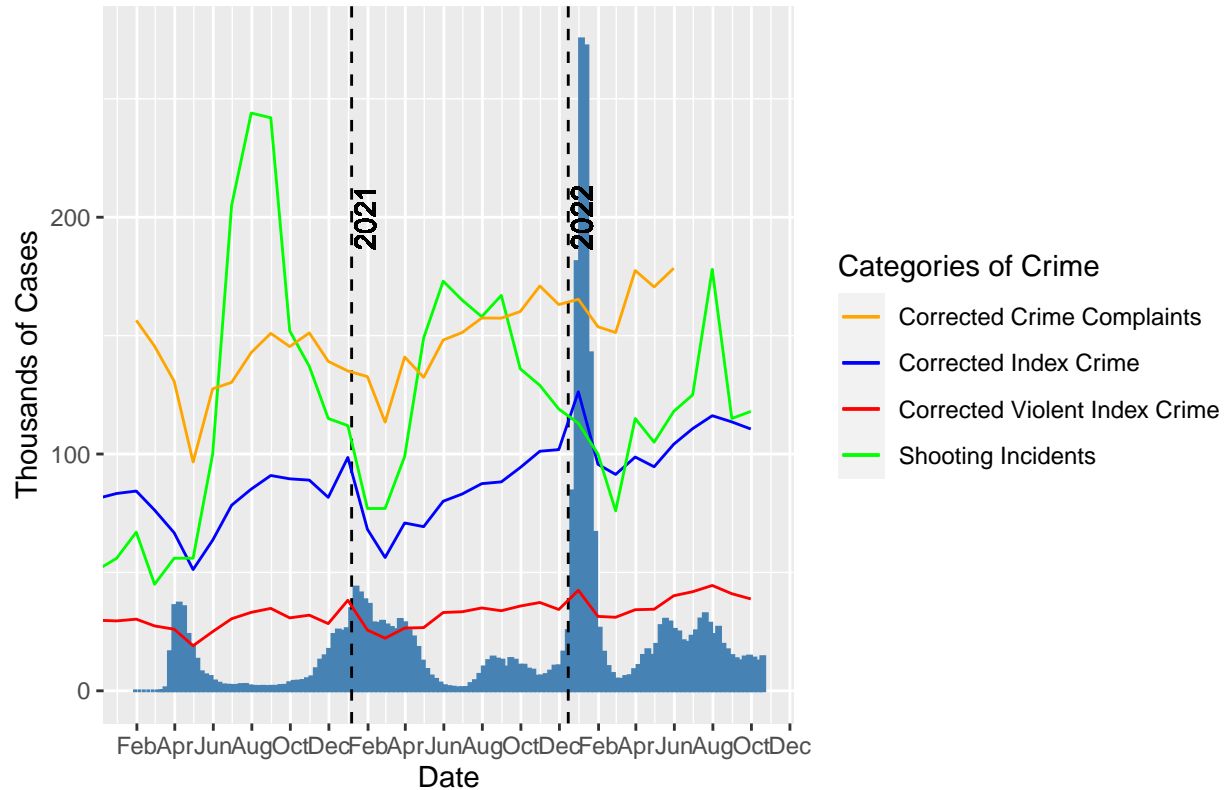
Examining crime statistics for Property Crime, Violent Crime, and Violent Crimes Committed With Guns leading up to the 2019 - 2022 period I found those categories of crime were generally on a downward trend just prior to the arrival of Covid, but then turned around very sharply for a strongly increasing trend going forward. While examining monthly criminal complaints in New York City for the period of time from 2019 to 2022, I found a cyclical ebb and flow in those complaints as well as an overall tendency to increase. Of particular note was the late winter/early spring down-tick of Criminal Complaints that led to yearly lows, only to be followed by a new all-period-high crest in Criminal Complaints later in the year. Given the seasonality of Covid 19 cases and deaths, specifically that both seemed to be low in the warmer months and high in the colder, this was an interesting inverse relationship. This trend of low Covid numbers in the warmer months was broken in 2022, but may have been caused by either new and more transmissible variants such as omicron, or perhaps just greater availability of Covid testing.

I then examined what ended up being my three major categories of crime: Index Crime, Violent Index Crime, and Shooting Incidents, alongside Criminal Complaints. Examining them together graphically, some interesting things became apparent. Violent Index Crime and Index Crime largely seemed to mirror each other, with a lesser relationship between Index Crime and Criminal Complaints. All three had an ebb and a flow such that they tended to be highest in warm weather with a drop-off that occurred early in the year, bottoming out near the tail-end of winter. Shooting Incidents seemed to follow this overall pattern as well, but its surges were much tighter, and followed the odd pattern of having its peak spread in 2020 and then tightening up as the years went on. When compared to Covid, it seemed almost like there might be a negative correlation between Covid cases and deaths and crime overall, except that the overlaps for the winter surges of Covid did not properly line up with the timing of the decline in crime. Also, oddly, Index Crime and Violent Index Crime, in the years 2021 and 2022 showed spikes in December, which seemed out of line compared to the general trends of crime and definitely out of step with a negative correlation with Covid. Furthermore, the initial part of the graph showed that when Covid began with its initial spike, all

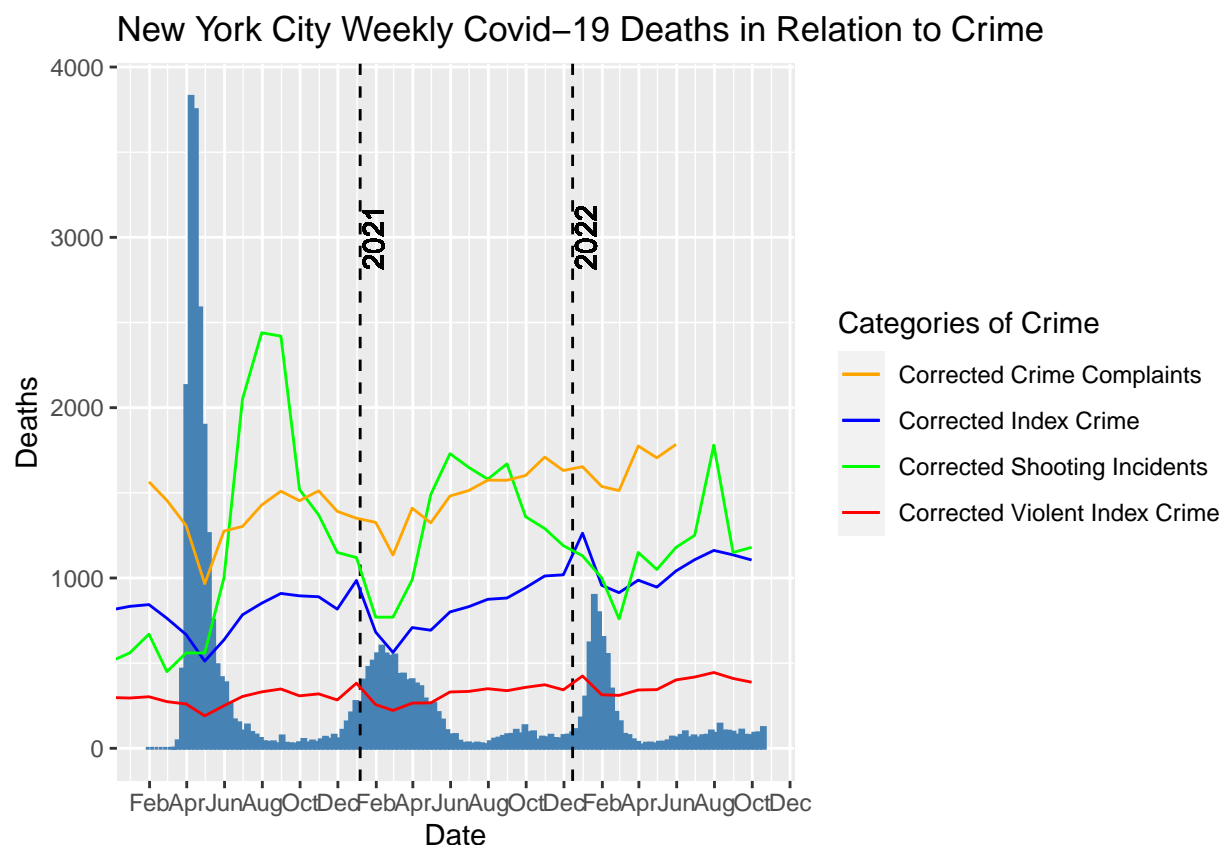
categories of crime dropped fairly strongly. except for Shooting Incidents which remained just somewhat low until after the initial spike in Covid. This initial trend may be because of the government imposed lock-downs. For easy reference, I am inserting the relevant graphs below (as seen earlier in my project).

```
## Warning: Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'NYWeeklyCase$Date' is discouraged.
## i Use 'Date' instead.
```

New York City Weekly Covid-19 Cases in Relation to Crime



```
## Warning: Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
## Use of 'NYWeeklyDeath$Date' is discouraged.
## i Use 'Date' instead.
```



Wanting to assess quantitatively just what correlations might emerge between these different data, I had to deal with a problem: my datasets were time series'. This meant that my usual preferences for correlational analyses likely would not work as my data would have autocorrelation and would not be normally distributed. At this point in my studies I had not yet learned how to deal with time series, so I opted to instead isolate these data trends to the periods of time during which the greatest interesting movement occurred. Specifically, I isolated the window of time in early 2020 when Covid deaths had their highest magnitude peak and I isolated the period of time in early 2022 when Covid cases were experiencing the same. The idea was that although there would be some autocorrelation, if I isolated my data sufficiently I wouldn't have to worry about seasonal effects and I might even have a normal distribution of data points. However, since my datasets would naturally be fairly small in terms of numbers of occurrences, I would need to rely on Kendall's Tau. This was also a necessity because my crime data was monthly but my Covid data was weekly, which meant that if I wanted these data to line up I need to include repetitions of the crime data such that every month's worth of Covid data would have the same corresponding crime data. Due to this repetition of data, Kendall's Tau was also desirabel. After a substantial amount of data massaging to get all of my required data points to line up, I then assessed the various aforementioned categories of crime for their correlation with Covid. What I found was somewhat disappointing.

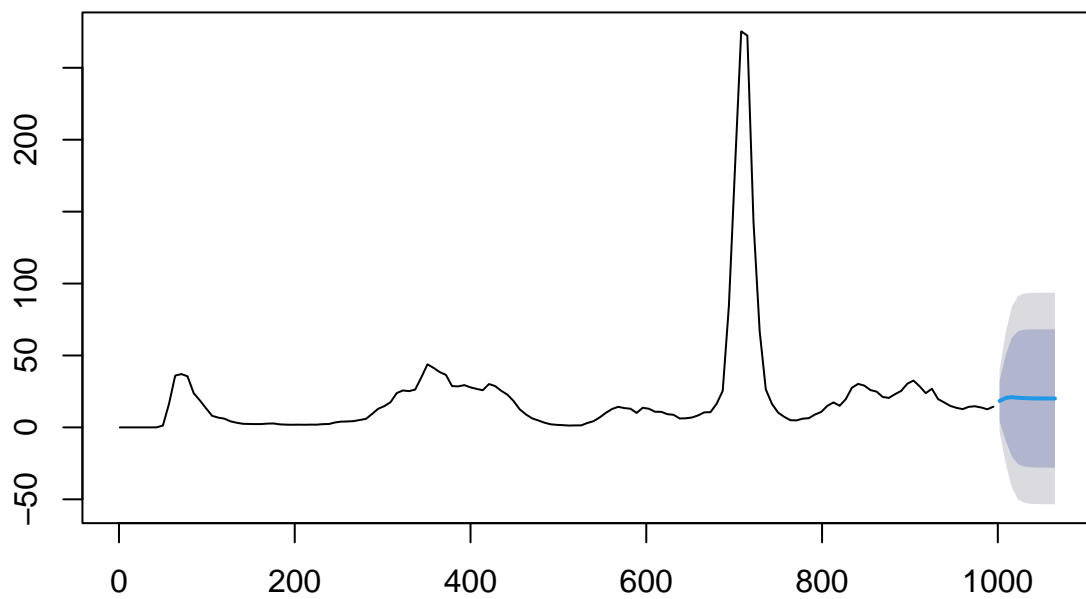
For Covid deaths, there was only one clearly significant correlation, Shooting Incidents, and one borderline correlation, Criminal Complaints. Shooting Incidents came in with a strong correlation coefficient of 0.81 and a p-value of 0.0009. Criminal Complaints had a p-value of 0.065, and its coefficient was a considerable -0.49. In this instance, one variable (Shooting Incidents) has a positive correlation with Covid Deaths (i.e., when it's values were increasing, Covid Deaths were also increasing), whereas the other (Criminal Complaints) was negative (i.e., while its values were increasing, Covid deaths were decreasing). Neither Index Crime nor Violent Index Crime were significant at all, with p-values above 0.7.

Given that the other peak I was assessing was approximately two years later, I wondered if I would find any better results there. Assessing the Covid case spike in early 2022 for Kendall's Tau yielded two significant correlations: the return of Shooting incidents and Criminal Complaints. Except this time around, Shooting Incidents had a negative coefficient of -0.34 and a p-value of 0.0497, while Criminal Complaints maintained the direction of its coefficient at -0.48 and a p-value of 0.0058. Whereas Shooting Incidents had a substantial change in both the size and direction of its correlation, Criminal Complaints was fairly comparable but much more significant this time around. The different correlation coefficient magnitude and direction for Shooting Incidents may be because the correlational analysis of Covid deaths looked at the first major spike of Covid in 2020 which behaved counter to many of the trends later seen in the winters of 2020/2021 and 2021/2022. That correlational analysis in early 2020 was meaningfully different in that it took place in a time when there was no natural immunity or vaccine, whereas the analysis in early 2022 concerns itself with a time frame when natural immunity and immunizations were factors.

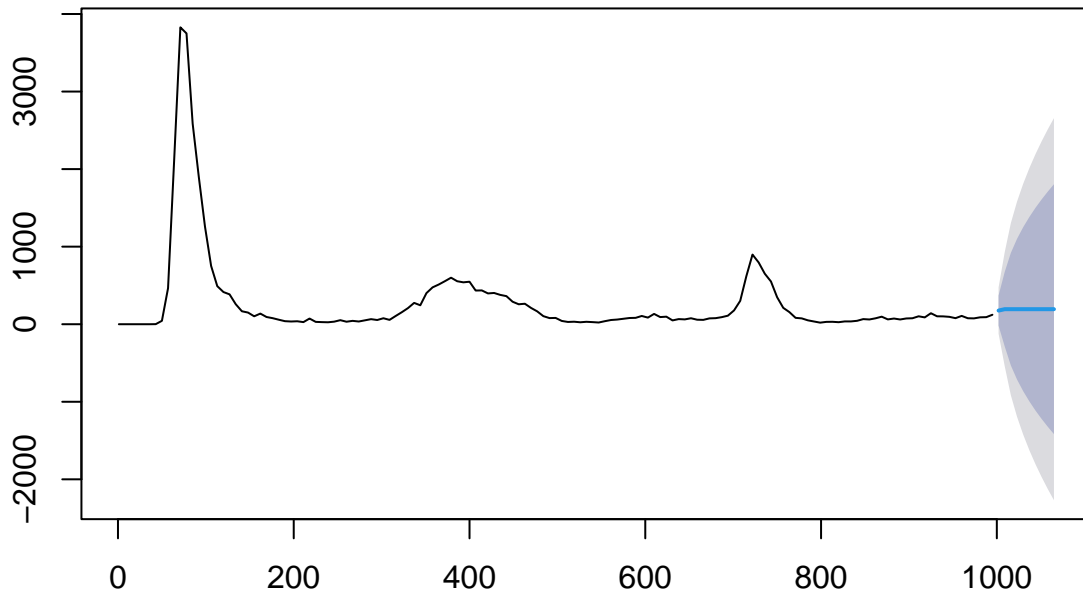
What was especially pleasing about these analyses was that they essentially performed double duty. Since Covid cases and deaths largely mirrored each other (although their chief spikes occurred in different years), even though we only directly examined Covid deaths in 2020 and cases in 2022, their strong correlations with each other means that these results are at least somewhat transferable.

Moving on to time series analyses, I assessed Covid cases and deaths as well as Index Crime, Violent Index Crime, and Shooting incidents. For Covid cases and deaths I found that the autocorrelations showed significant relationships amongst the time series data going back several lags, but the partial autocorrelations revealed that these connections were not actually as strong as the autocorrelations might suggest (they even see-sawed between positive and negative correlations). Using `auto.arima` to examine the data for auto regression, moving averages, and differencing I was able to create a model that projected cases and deaths out for 10 weeks. As seems to frequently be the case for this type of model, the spread for confidence intervals was fairly wide. Also, the model projected that negative Covid cases and deaths were possible, which is obviously not actually the case. Assessing the crime data (Index Crime, Violent Index Crime, and Shooting Incidents) resulted in `auto.arima` outputs of ARIMA (0,1,0), which is to say "random walks". In these "random walks" the data exhibited autocorrelation (as was borne out by the `acf`'s I ran on my data), but it was essentially as random as a person wandering about. This was further evidenced by applying the diff of 1 (which refers to the "1" in the ARIMA output of (0,1,0)) and plotting that to find essentially noise. The `pacf`'s also showed few to no significant correlations (depending on which category of crime I was inspecting) and that demonstrated that the `acf` outputs were essentially self-reinforcing, and not indicative of a pattern. For reference, I've included the projection plots below with the first being for Covid cases and the second being for Covid deaths.

Forecasts from ARIMA(1,0,3) with non-zero mean



Forecasts from ARIMA(0,1,2)



So, what did I find? Was crime really impacted by Covid-19, either cases or deaths? Well, if there is such a relationship between Covid and crime my data doesn't conclusively show it. Graphically, the data is very mixed, as evidenced by the low Kendall's Tau correlational coefficients I found for many of my variables, and any relationships that were present seemed to be inversely related. This would mean that crime was generally more of an issue when Covid was not, and vice-versa. Now, had I found many significant negative correlations, then that would have been something interesting, but I only was able to track down two variables that had statistically significant correlations with Covid-19: Shooting Incidents and Criminal Complaints. One might anticipate that Index Crime or Violent Index Crime would be significant if Criminal Complaints was, but that wasn't the case. To make matters worse, Shooting Incidents changed the direction and magnitude of its correlation between the two quantitative tests that I ran, which makes any information gleaned from it unreliable. The only consistent information I got was for Criminal Complaints, but that was only clearly significant in the second test, although in both tests the magnitude and direction of the coefficients were the same. In this instance, it suggested that there was a moderate inverse relationship between Criminal Complaints and Covid-19.

While it is nice to not walk away completely empty-handed, it is also the case that there are many confounding factors. Attempting to isolate the cause of crime in any city as massive as New York City would be rife with complications under even the best conditions, and this particular analysis was no different. For instance, New York City gets very cold in the winters, and also snowy. This would coincide with the Covid surges, which makes sense as everybody is clustering together in warm buildings which is the ideal place for Covid to spread, and it would also coincide with nature impeding the commission of various crimes. So, in that way, my analysis may be noticing nothing more than the effects of the weather. Furthermore, there was great political and social strife during the time assessed, with one substantial example of that being the BLM riots which would strongly impact the criminal statistics without being directly related to Covid.

Future analyses would be well-served to isolate the timings of government imposed lock-downs (which obviously would impact crime without necessarily being disease driven; both when the lock-downs occurred and during the aftermath of the lock-downs), isolating the times and locations of the BLM riots or other similar occurrences, and assessing these effects by borough as that would possibly be able to show trends that might be otherwise missed when examining New York City as a whole. Furthermore, approaching this study as a time series analysis from the beginning would be beneficial, especially for finding meaningful variables of interest. Once that is completed, I would suggest a multivariate time series model be constructed using the Covid data (which I've already shown to produce meaningful results with my own efforts), as well as with criminal statistics that show strong characteristics of overlap with Covid without being merely "random walk" data. Given that I wanted to find correlational relationships between Covid and crime, and both of those are time series data, a multivariate time series analysis would be the only sensible way to proceed, and might actually produce meaningful results.

As for the average person reading this, the implications of my research are only that if you want to avoid Covid-19 in New York City you should generally not visit in the winter, and if you want to avoid crime, you should generally avoid visiting every other season of the year. It would seem that when it comes to these two unpleasant facts about New York City, you can only pick your poison, not avoid both.

References

Centers for Disease Control and Prevention. COVID Data Tracker. Atlanta, GA: U.S. Department of Health and Human Services, CDC; 2022, October 24. <https://covid.cdc.gov/covid-data-tracker>.

New York State Division of Criminal Justice Services. <https://www.criminaljustice.ny.gov/>.