How Covid Impacted NYPD Shooting Incident Data

Brennan Han

2025-02-25

Explaination

We are going to have Pre-Covid, Covid, and Post-Covid data
and see the the amount of incidents happened during each time period
then we are going to compare buroughs to burough and from time-range
to time-range

Significant dates:
March 22, 2020 NYS on Pause Program begins, all non-essential workers must stay home
June 15, 2021 Governor Cuomo reopened the entirety of New York State

Because the Covid lock in dates were 450 days, we will use 450 days prior
as our Pre-Covid data, and 450 days after the mandate as our Post-Covid
Data, and the lock-in dates as the Covid data

R Markdown

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v lubridate 1.9.4      v tibble    3.2.1
## v purrr     1.0.4      v tidyr     1.3.1
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE, fig.show = "hold")
```

```
## Import Data
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"

NYPD_Shooting_Incident_Data <- read_csv(url)

## convert Date to Date format
```

```r
NYPD_Shooting_Incident_Data$OCCUR_DATE <- as.Date(
  NYPD_Shooting_Incident_Data$OCCUR_DATE, format = "%m/%d/%Y")
## 450 days of covid
Pre_Pre_Covid_Date <- as.Date("2018-12-28")
Pre_Covid_Date <- as.Date("2020-03-22")
Post_Covid_Date <-as.Date("2021-06-15")
Post_Post_Covid_Date <- as.Date("2022-08-08")

## Filter the data by preCovid, Covid, and PostCovid
## and put into their own tables

Pre_Covid_Data <- NYPD_Shooting_Incident_Data %>% filter(OCCUR_DATE >
Pre_Pre_Covid_Date & OCCUR_DATE < Pre_Covid_Date)

Covid_Data <- NYPD_Shooting_Incident_Data %>% filter(Pre_Covid_Date <
OCCUR_DATE & OCCUR_DATE< Post_Covid_Date)

Post_Covid_Data <- NYPD_Shooting_Incident_Data %>% filter(OCCUR_DATE
> Post_Covid_Date & OCCUR_DATE < Post_Post_Covid_Date)

## Add the counts of incidents and label them as well

Pre_Covid_Boro_Count <- Pre_Covid_Data %>% count(BORO)
Pre_Covid_Boro_Count$Date_Range <- "Pre-Covid"

Covid_Boro_Count <- Covid_Data %>% count(BORO)
Covid_Boro_Count$Date_Range <- "Covid"

Post_Covid_Boro_Count <- Post_Covid_Data %>% count(BORO)
Post_Covid_Boro_Count$Date_Range <- "Post-Covid"

## Put all the data together in one table

combined_data <- bind_rows(Pre_Covid_Boro_Count, Covid_Boro_Count,
Post_Covid_Boro_Count) %>% mutate(BORO = factor(BORO), Date_Range
=(Date_Range)) %>% droplevels()
```
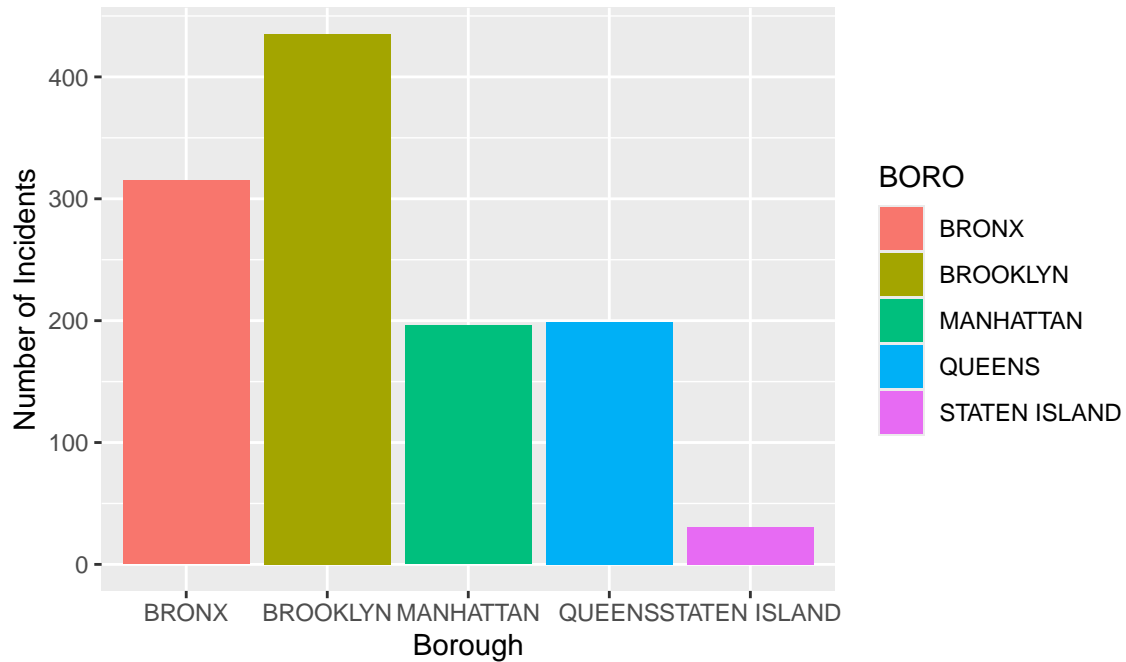
Number of Incidents by Borough Before Covid

```r
ggplot(Pre_Covid_Boro_Count, aes(x = BORO, y = n, fill = BORO)) +
geom_bar(stat = "identity") + labs( title = "Number of Incidents by
Borough The Year Before Covid", x = "Borough", y = "Number of Incidents"
)
```
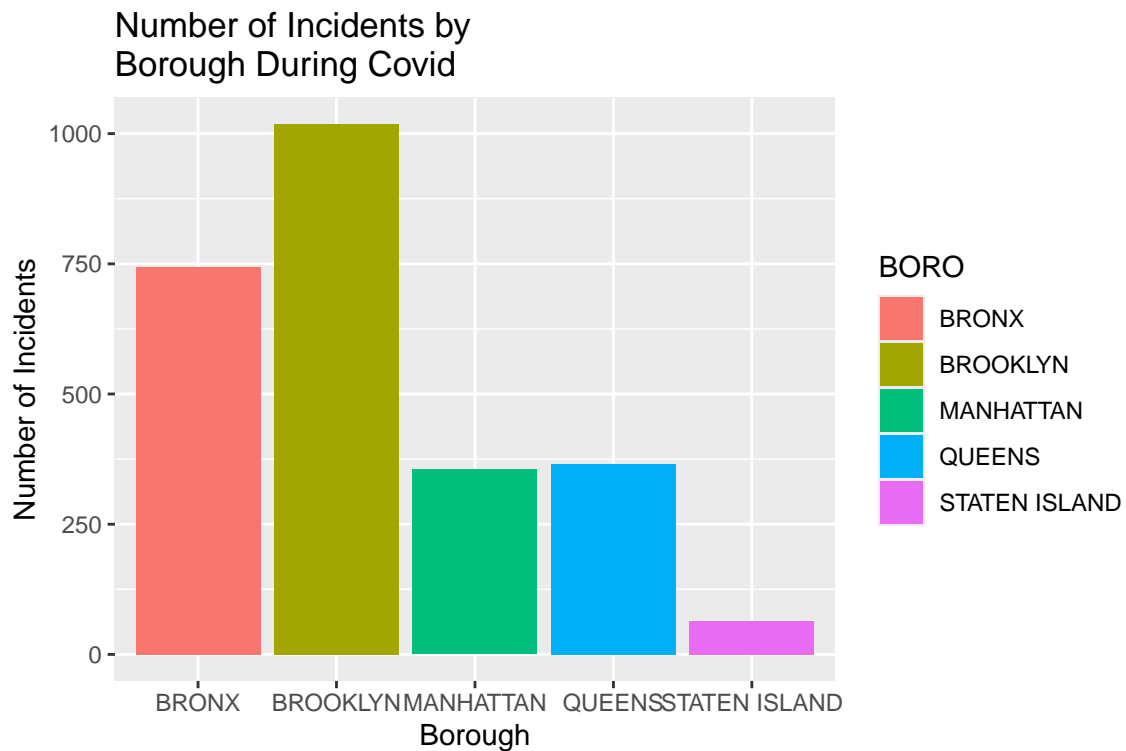
Number of Incidents by Borough The Year Before Covid
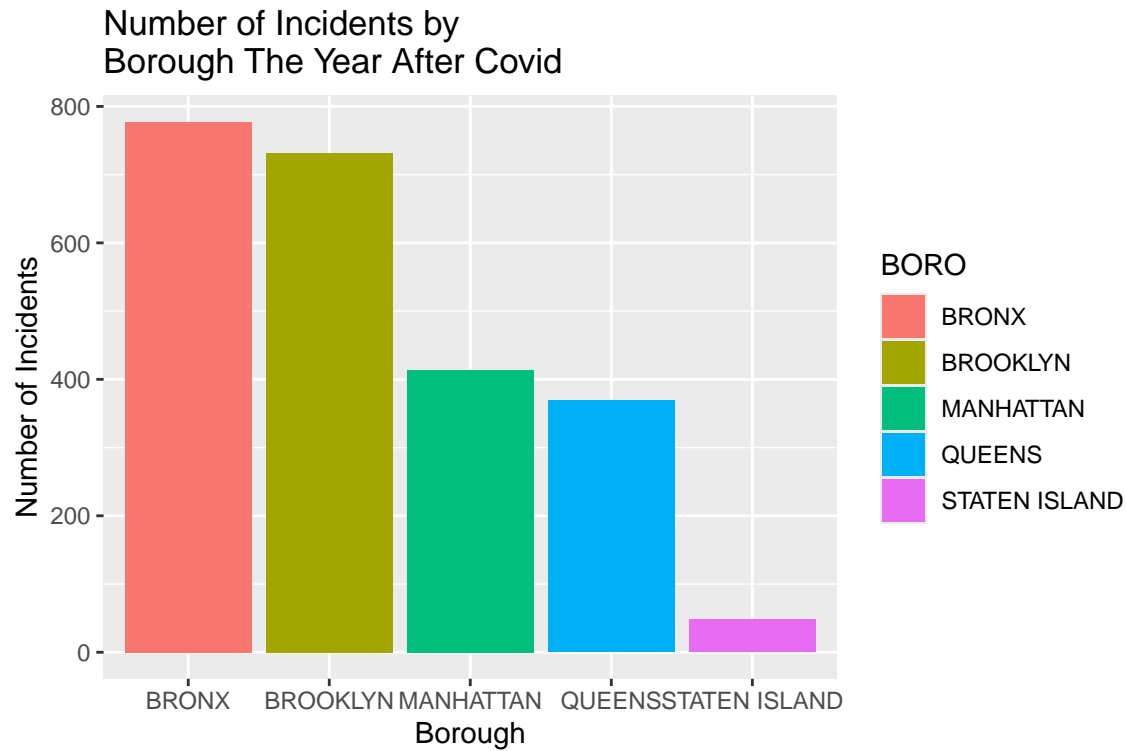
Number of Incidents by Borough During Covid

```
ggplot(Covid_Boro_Count, aes(x = BORO, y = n, fill = BORO) ) +
geom_bar(stat = "Identity") + labs( title = "Number of Incidents by
Borough During Covid", x = "Borough", y = "Number of Incidents" )
```


Number of Incidents by Borough During Covid

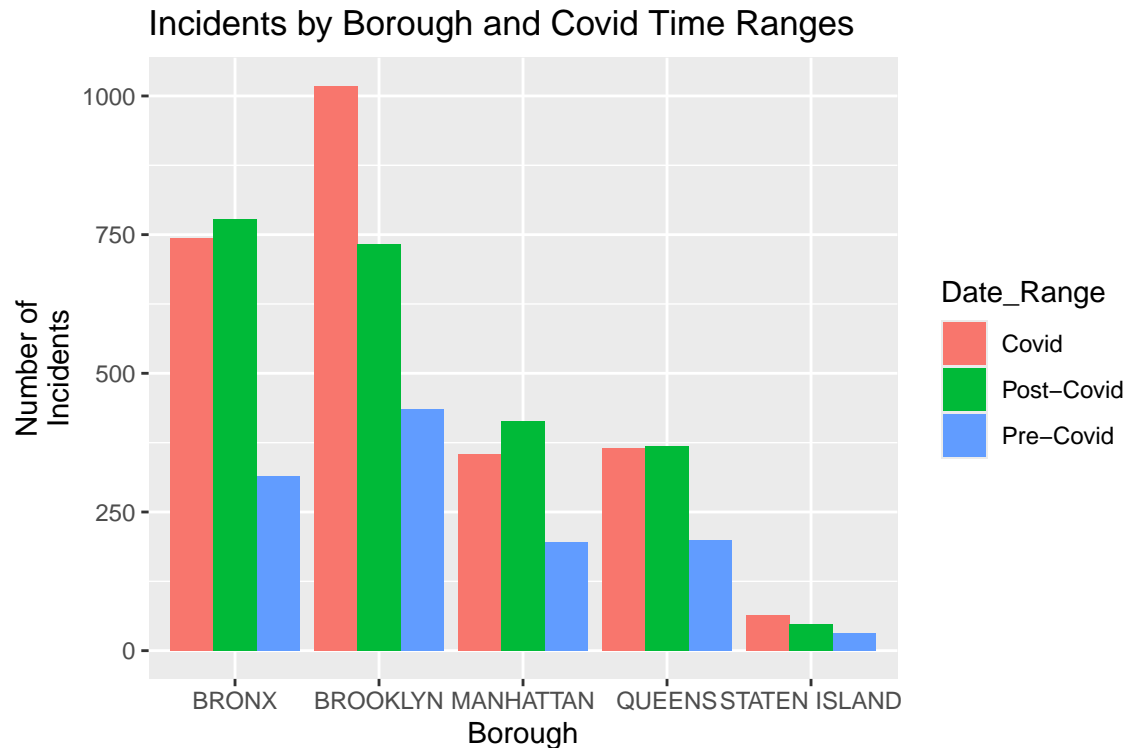Number of Incidents by Borough The Year After Covid

```
ggplot(Post_Covid_Boro_Count, aes(x = BORO, y = n, fill = BORO) ) +
geom_bar(stat = "Identity") + labs( title = "Number of Incidents by
Borough The Year After Covid", x = "Borough", y = "Number of Incidents"
)
```


##

Number of Incidents by Borough and Covid Time Ranges

```
ggplot(combined_data, aes(x = BORO, y = n, fill = Date_Range) ) +
geom_bar(stat = "identity", position = "dodge") + labs(title =
"Incidents by Borough and Covid Time Ranges", x = "Borough", y = "Number of
Incidents")
```

Incidents by Borough and Covid Time Ranges

Put together all the combined Covid Time ranges data so we can run a model on them

```
combined_data$Date_Range <- factor(combined_data$Date_Range, levels =
                                   c("Pre-Covid", "Covid", "Post-Covid"))
```

Run a linear model on them

```
lm_model <- lm(n ~ BORO + Date_Range, data = combined_data)
summary(lm_model)
```

```
##
## Call:
## lm(formula = n ~ BORO + Date_Range, data = combined_data)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -128.067  -67.633   -5.867   50.100  184.600
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            443.07      86.86    5.101 0.000929 ***
## BOROBROOKLYN           116.33     103.82    1.121 0.294987
## BOROMANHATTAN         -290.33     103.82   -2.797 0.023321 *
## BOROQUEENS            -301.00     103.82   -2.899 0.019913 *
## BOROSTATEN ISLAND     -564.33     103.82   -5.436 0.000619 ***
## Date_RangeCovid        274.00      80.42    3.407 0.009261 **
## Date_RangePost-Covid   232.80      80.42    2.895 0.020048 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 127.2 on 8 degrees of freedom
## Multiple R-squared:  0.894,  Adjusted R-squared:  0.8145
## F-statistic: 11.25 on 6 and 8 DF,  p-value: 0.001587
```

Some notes on potential Bias

```
I am aware that there is bias in that I did not add data about
whether or not there was a death in the incidents so an incident may have
a varying amount of range in its definition. Also, each borough is different
in area and population which could make it seem like boroughs have a higher
or lower incidents because of this.
```