

NYC Impact of Covid on Gun Violence

Ben Lewis

2025-01-10

This report is an analysis of the impact of Covid-19 on gun violence in NYC.

Step 1: Start an Rmd Document

Start an Rmd document that describes and imports the shooting project dataset in a reproducible manner.

This project uses the tidyverse library.

```
library(tidyverse)
```

It imports the publicly available data set that captures every shooting incident that occurred in NYC from 2006 to 2023.

```
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
df <- read_csv(url)
```

Step 2: Tidy and Transform Your Data

Add to your Rmd document a summary of the data and clean up your dataset by changing appropriate variables to factor and date types and getting rid of any columns not needed. Show the summary of your data to be sure there is no missing data. If there is missing data, describe how you plan to handle it.

```
# View summary of data displaying all columns with examples
glimpse(df)
```

```
## Rows: 28,562
## Columns: 21
## $ INCIDENT_KEY      <dbl> 244608249, 247542571, 84967535, 202853370, 270~
## $ OCCUR_DATE        <chr> "05/05/2022", "07/04/2022", "05/27/2012", "09/~
## $ OCCUR_TIME        <time> 00:10:00, 22:20:00, 19:35:00, 21:00:00, 21:00~
## $ BORO              <chr> "MANHATTAN", "BRONX", "QUEENS", "BRONX", "BROO~
## $ LOC_OF_OCCUR_DESC  <chr> "INSIDE", "OUTSIDE", NA, NA, NA, NA, NA, NA, N~
## $ PRECINCT          <dbl> 14, 48, 103, 42, 83, 23, 113, 77, 48, 49, 73, ~
## $ JURISDICTION_CODE  <dbl> 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ LOC_CLASSFCTN_DESC <chr> "COMMERCIAL", "STREET", NA, NA, NA, NA, NA, NA~
## $ LOCATION_DESC      <chr> "VIDEO STORE", "(null)", NA, NA, NA, "MULTI DW~
## $ STATISTICAL_MURDER_FLAG <lgl> TRUE, TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, ~
## $ PERP_AGE_GROUP     <chr> "25-44", "(null)", NA, "25-44", "25-44", NA, N~
## $ PERP_SEX           <chr> "M", "(null)", NA, "M", "M", NA, NA, NA, NA, "~
```

```
## $ PERP_RACE          <chr> "BLACK", "(null)", NA, "UNKNOWN", "BLACK", NA, ~
## $ VIC_AGE_GROUP      <chr> "25-44", "18-24", "18-24", "25-44", "25-44", "~
## $ VIC_SEX            <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "~
## $ VIC_RACE           <chr> "BLACK", "BLACK", "BLACK", "BLACK", "BLACK", "~
## $ X_COORD_CD         <dbl> 986050, 1016802, 1048632, 1014493, 1009149, 99~
## $ Y_COORD_CD         <dbl> 214231.0, 250581.0, 198262.0, 242565.0, 190104~
## $ Latitude           <dbl> 40.75469, 40.85440, 40.71063, 40.83242, 40.688~
## $ Longitude          <dbl> -73.99350, -73.88233, -73.76777, -73.89071, -7~
## $ Lon_Lat            <chr> "POINT (-73.9935 40.754692)", "POINT (-73.8823~
```

We will focus on the number of incidents as they relate to the onset of Covid-19. We will examine annual trends and investigate how the shootings in 2020 correlated with key events that year. We will only need the date column and can remove the other columns. We will add columns for year and month.

```
# Remove all other columns besides incident ID and date
df <- df[, c("INCIDENT_KEY", "OCCUR_DATE")]

# Make sure there are no null values
if (sum(is.na(df$OCCUR_DATE)) > 0) {
  stop("Dates column has a null value")
}

# Transform date from character sequence to Date
df$OCCUR_DATE = mdy(df$OCCUR_DATE)

# Add year and month columns
df$OCCUR_YEAR = year(df$OCCUR_DATE)
df$OCCUR_MONTH = month(df$OCCUR_DATE)

# View transformed df
glimpse(df)
```

```
## Rows: 28,562
## Columns: 4
## $ INCIDENT_KEY <dbl> 244608249, 247542571, 84967535, 202853370, 27078636, 2303~
## $ OCCUR_DATE   <date> 2022-05-05, 2022-07-04, 2012-05-27, 2019-09-24, 2007-02-~
## $ OCCUR_YEAR   <dbl> 2022, 2022, 2012, 2019, 2007, 2021, 2021, 2021, 2021, 202~
## $ OCCUR_MONTH  <dbl> 5, 7, 5, 9, 2, 7, 6, 7, 5, 12, 9, 12, 4, 11, 7, 5, 8, 6, ~
```

```
summary(select(df, -INCIDENT_KEY))
```

```
##      OCCUR_DATE      OCCUR_YEAR  OCCUR_MONTH
##  Min.   :2006-01-01  Min.   :2006    Min.    : 1.000
## 1st Qu.:2009-09-04  1st Qu.:2009    1st Qu.: 5.000
## Median :2013-09-20  Median :2013    Median : 7.000
## Mean   :2014-06-07  Mean   :2014    Mean   : 6.805
## 3rd Qu.:2019-09-29  3rd Qu.:2019    3rd Qu.: 9.000
## Max.   :2023-12-29  Max.   :2023    Max.   :12.000
```

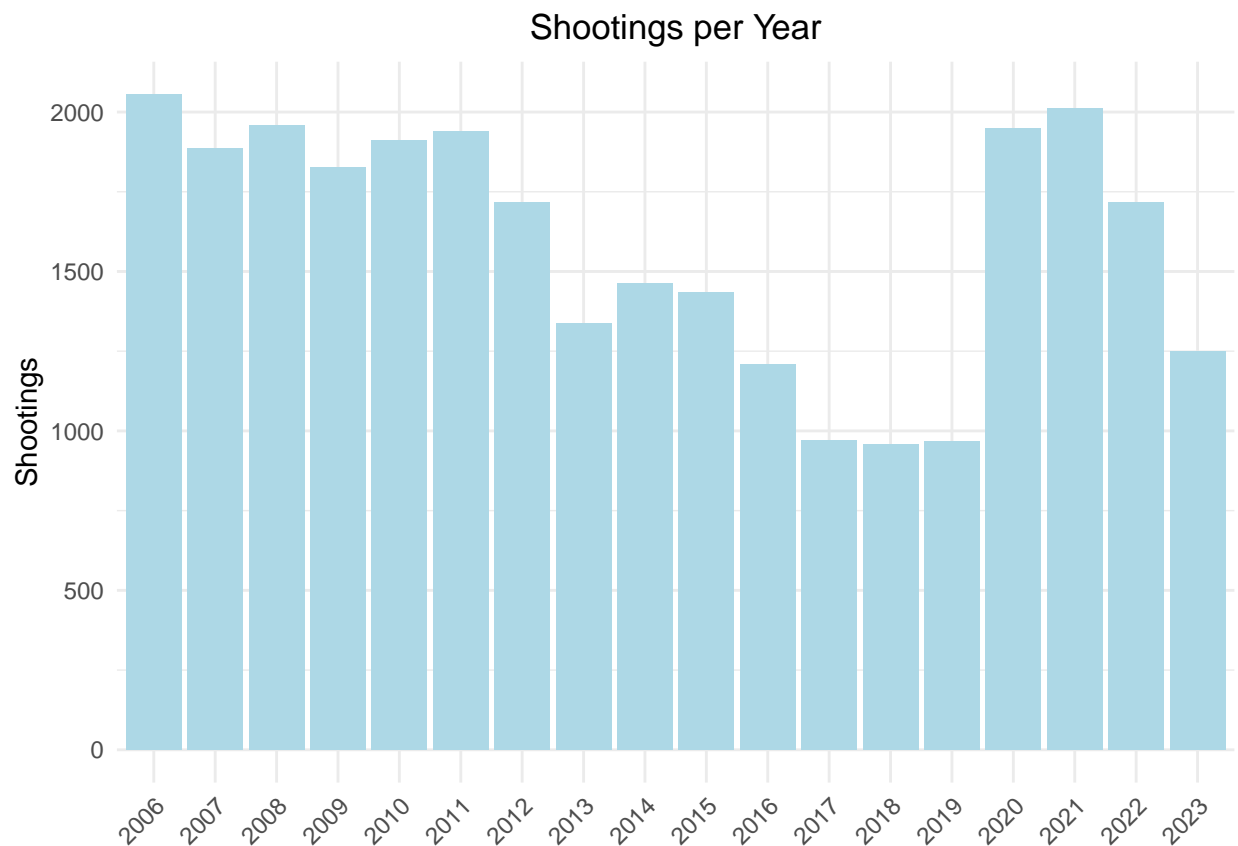
Step 3: Add Visualizations and Analysis

Add at least two different visualizations & some analysis to your Rmd. Does this raise additional questions that you should investigate?

Annual Trends

```
# Dataframe grouping by year
year_counts = as.data.frame(table(df$OCCUR_YEAR))
colnames(year_counts) <- c("Year", "Shootings")

# Bar plot
ggplot(year_counts, aes(x = Year, y = Shootings)) +
  geom_bar(stat = "identity", fill = "lightblue") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(hjust = 0.5)
  ) +
  labs(
    title = "Shootings per Year",
    x = NULL
  )
```



As you can see the number of shootings experienced a steady decline from over 2000 incidents in 2006 to under 1000 in 2017 and remained level at slightly under 1000 from 2017-2019. The first reported case of Covid in NYC occurred on March 1, 2020. That year the number of incidents doubled from the previous year. After a slight increase in 2021 there was steep decline in 2022 and 2023.

Monthly Trends

The surge in shootings in 2020 was influenced by several factors. Below are some key events:

- **Full Lockdown**, *March 22, 2020* – The State of New York implemented an executive order closing all non-essential businesses and ordering all residents to stay home except for essential needs like groceries or medical care.
- **Protests Following George Floyd's Death**, *May-June, 2020* – The killing of George Floyd by police on May 25, 2020 sparked widespread protests and heightened tensions between communities and law enforcement.
- **Police Budget Cuts**, *July, 2020* – NYC government cut approximately \$1 billion from the NYPD budget in response to public pressure.

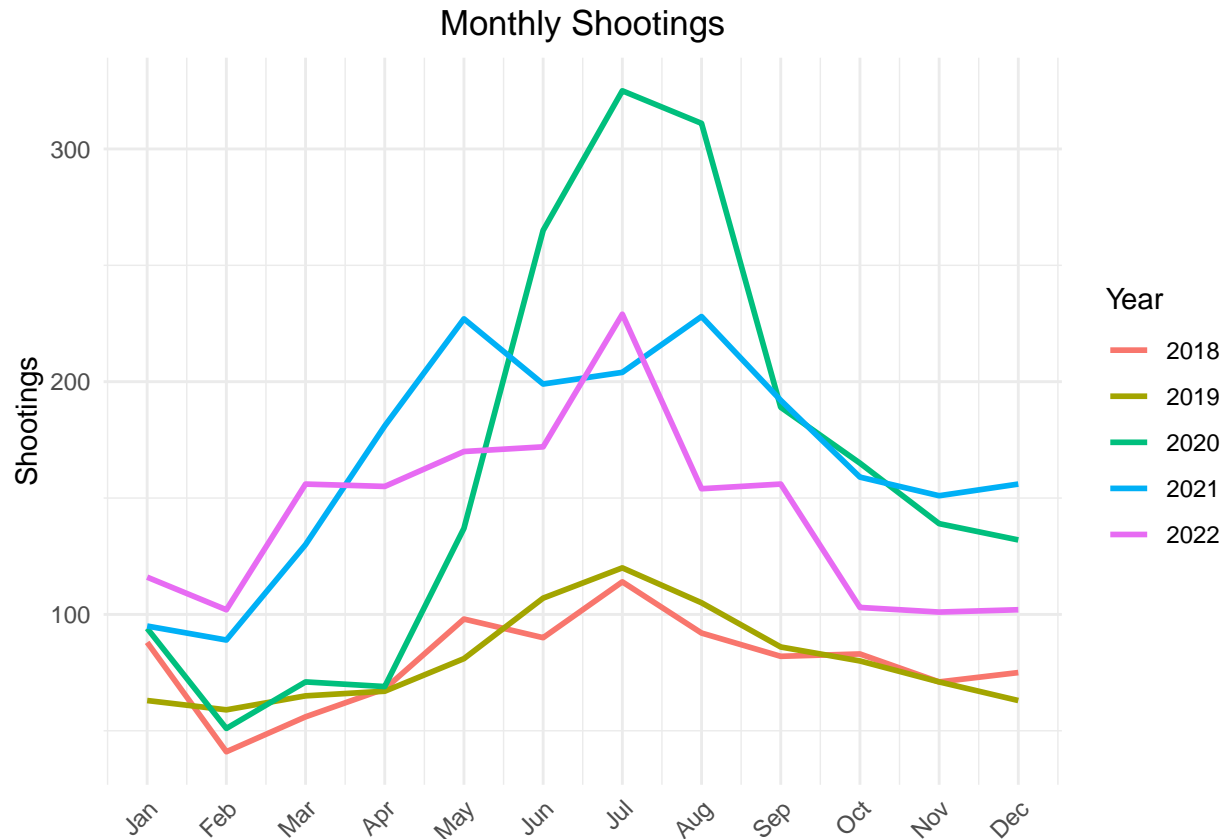
It is important to remember these are just a few of the events that may have had an impact. Some other events and factors to consider are the economic and social disruption, strain on the courts and criminal justice system, bail reform laws that took effect in January 2020, gang violence, and increase in gun sales among other factors.

Let's examine the trend for the shootings that occurred within 2020 by comparing the monthly incidents with the 2 years before and after 2020.

```
# Filter years 2018 - 2022
df_2018_to_2022 <- df %>%
  filter(OCCUR_YEAR >= 2018 & OCCUR_YEAR <= 2022)

# Group each year by month
monthly_shootings <- df_2018_to_2022 %>%
  group_by(OCCUR_YEAR, OCCUR_MONTH) %>%
  summarise(incident_count = n(), .groups = "drop")

# Line chart for each year
ggplot(monthly_shootings, aes(x = OCCUR_MONTH, y = incident_count, color = as.factor(OCCUR_YEAR), group
  geom_line(linewidth = 1) +
  scale_x_continuous(breaks = 1:12, labels = month.abb) + # Show months as abbreviations
  theme_minimal() +
  labs(
    title = "Monthly Shootings",
    x = NULL,
    y = "Shootings",
    color = "Year"
  ) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(hjust = 0.5)
  )
)
```



One of the first things to notice is that all years follow a seasonal trend with incidents peaking in the summer and at their lowest in the winter. In 2020 shootings were similar to the 2 years prior until May when they started skyrocketing. The summer months of 2020 were significantly higher than the those same months in the 2 years before and after 2020.

Some things that warrant further investigation would be to include events in the 2 years prior and after 2020 that may have influenced gun violence in those years. Another thing to consider is how the population may have changed in those years. The population in NYC declined after 2020 so it would be interesting to compare the shootings per capita.

Step 4: Add Bias Identification

Write the conclusion to your project report and include any possible sources of bias. Be sure to identify what your personal bias might be and how you have mitigated that.

Conclusion

The analysis of gun violence in NYC during the COVID-19 pandemic reveals several important trends. There was a significant increase in shootings in 2020, coinciding with major events such as the onset of the pandemic, social unrest, and law enforcement policy changes. While the data shows correlation, causation cannot be assumed. Additional factors such as economic distress and social isolation may have contributed.

Possible Sources of Bias For full disclosure I have lived in Chicago since 2013, and there were multiple shootings in very close proximity to where I lived in 2020. I came in with the expectation of a surge in shootings in NYC in the summer of 2020. I did my best to mitigate my bias by focusing on the data rather than trying to slice the data in a way that fit my preconceived notions.