

New York City Boroughs Shooting Incidents Report

Phanitta Chomsinsap

2023-10-15

This report evaluates the **New York Police Department (NYPD) Shooting Incident (Historic)** dataset to better understand the number of shooting incidents in each borough over time from 2006 to 2022. This report will also discuss steps to import, tidy, transform, and analyze data as well as address any possible bias on the analysis.

Libraries

Two libraries used in this report are:

```
library(tidyverse)
library(lubridate)
```

Data Import

Data is loaded in as a csv format. The dataset contains many columns of information such as the date and time of the incident, location of occurrence, information about the perpetrator and the victim, and whether the incident led to a murder.

```
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv"
nypd <- read_csv(url)
nypd
```

```
## # A tibble: 27,312 x 21
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO      LOC_OF_OCCUR_DESC PRECINCT
##   <dbl> <chr>      <time> <chr>      <chr>      <dbl>
## 1 228798151 05/27/2021 21:30  QUEENS    <NA>      105
## 2 137471050 06/27/2014 17:40  BRONX     <NA>      40
## 3 147998800 11/21/2015 03:56  QUEENS    <NA>      108
## 4 146837977 10/09/2015 18:30  BRONX     <NA>      44
## 5 58921844 02/19/2009 22:58  BRONX     <NA>      47
## 6 219559682 10/21/2020 21:36  BROOKLYN  <NA>      81
## 7 85295722 06/17/2012 22:47  QUEENS    <NA>      114
## 8 71662474 03/08/2010 19:41  BROOKLYN  <NA>      81
## 9 83002139 02/05/2012 05:45  QUEENS    <NA>      105
## 10 86437261 08/26/2012 01:10  QUEENS    <NA>      101
## # i 27,302 more rows
## # i 15 more variables: JURISDICTION_CODE <dbl>, LOC_CLASSFCTN_DESC <chr>,
## # LOCATION_DESC <chr>, STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>,
## # PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>,
## # VIC_RACE <chr>, X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>,
## # Longitude <dbl>, Lon_Lat <chr>
```

Data Preprocessing

Since this report focuses on the trend of shooting incident over time within each borough, some columns that are not relevant to the analysis are removed. Columns that are removed include LOC_OF_OCCUR_DESC, PRECINCT, JURISDICTION_CODE, LOC_CLASSFCTN_DESC, LOCATION_DESC, X_COORD_CD, Y_COORD_CD, PERP_AGE_GROUP, PERP_SEX, PERP_RACE, VIC_AGE_GROUP, VIC_SEX, VIC_RACE, Latitude, Longitude, and Lon_Lat.

```
df <- subset(nypd, select = -c(LOC_OF_OCCUR_DESC, PRECINCT, JURISDICTION_CODE,
                               LOC_CLASSFCTN_DESC, LOCATION_DESC, X_COORD_CD,
                               Y_COORD_CD, PERP_AGE_GROUP, PERP_SEX, PERP_RACE,
                               VIC_AGE_GROUP, VIC_SEX, VIC_RACE,
                               Latitude, Longitude, Lon_Lat))
```

The column OCCUR_DATE is converted from chr to date type.

```
df$OCCUR_DATE <- str_replace_all(df$OCCUR_DATE, "/", "-")
df$OCCUR_DATE <- mdy(df$OCCUR_DATE)
```

A new column called OCC_COUNT or occurrence count is created. The purpose of this column is to use it to sum the number of total incidents in a month.

```
df <- df %>%
  mutate(OCC_COUNT = 1)
```

Tidy data is shown below:

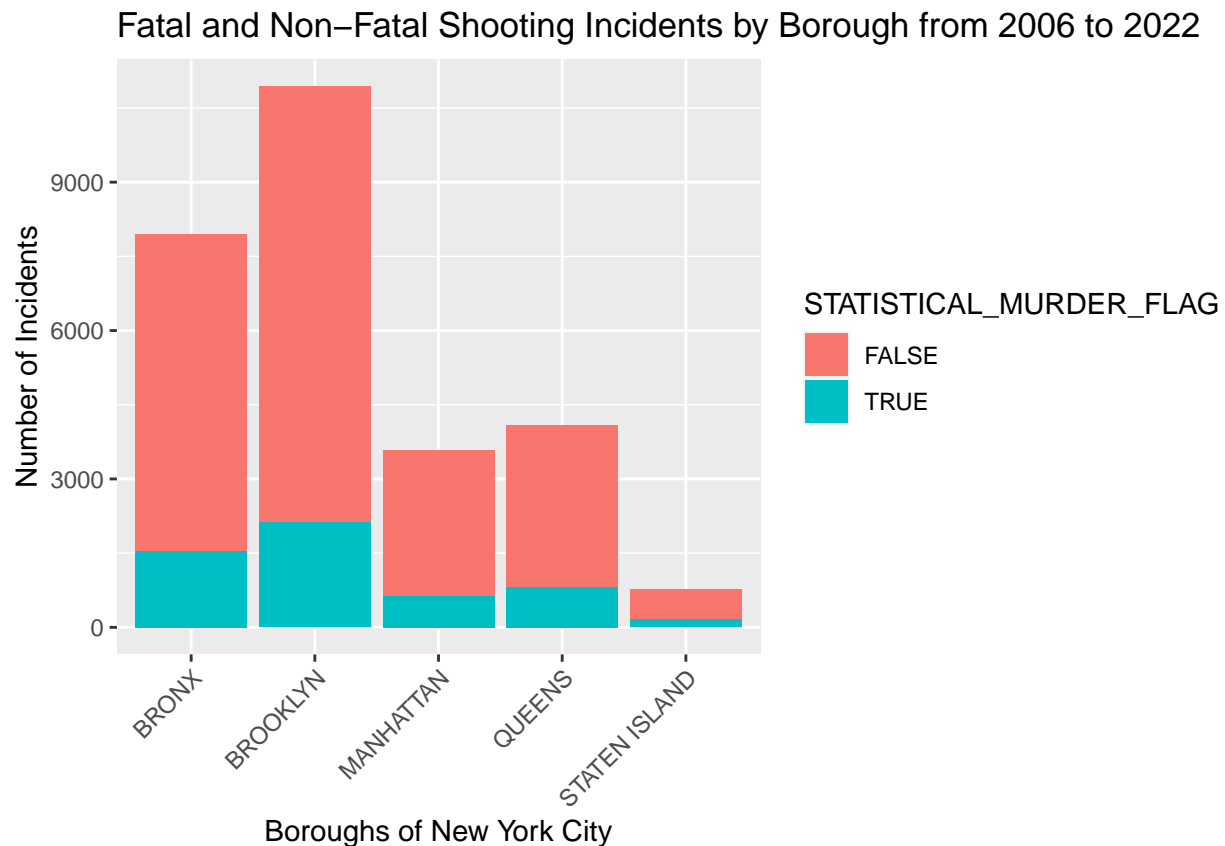
```
df

## # A tibble: 27,312 x 6
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO      STATISTICAL_MURDER_FLAG OCC_COUNT
##   <dbl> <date>      <time> <chr>      <lg1>
## 1  228798151 2021-05-27 21:30  QUEENS    FALSE      1
## 2  137471050 2014-06-27 17:40  BRONX     FALSE      1
## 3  147998800 2015-11-21 03:56  QUEENS    TRUE       1
## 4  146837977 2015-10-09 18:30  BRONX     FALSE      1
## 5   58921844 2009-02-19 22:58  BRONX     TRUE       1
## 6  219559682 2020-10-21 21:36  BROOKLYN  TRUE       1
## 7   85295722 2012-06-17 22:47  QUEENS    FALSE      1
## 8   71662474 2010-03-08 19:41  BROOKLYN  TRUE       1
## 9   83002139 2012-02-05 05:45  QUEENS    FALSE      1
## 10  86437261 2012-08-26 01:10  QUEENS    FALSE      1
## # i 27,302 more rows
```

Analysis & Bias Identification

The first observation is the cumulative number of incidents by each borough from 2006 to 2022. The following bar chart is color coded based on the STATISTICAL_MURDER_FLAG. A value of TRUE means that a shooting incident led to murder.

```
ggplot(df, aes(x = BORO, fill = STATISTICAL_MURDER_FLAG)) +
  geom_bar() +
  scale_color_brewer(palette = "Dark2") +
  labs(x = "NYPD Shooting Incidents by Borough from 2006 to 2022") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  xlab("Boroughs of New York City") +
  ylab("Number of Incidents") +
  ggtitle("Fatal and Non-Fatal Shooting Incidents by Borough from 2006 to 2022")
```



According to the bar chart and the following summary table, Brooklyn has the highest number of shooting incidents in both fatal (led to murder) and non-fatal category of 2122 and 8811 incidents respectively.

```
table(df$BORO, df$STATISTICAL_MURDER_FLAG)
```

```
##
##           FALSE TRUE
##  BRONX         6395 1542
##  BROOKLYN      8811 2122
##  MANHATTAN     2942  630
##  QUEENS        3284  810
##  STATEN ISLAND  614  162
```

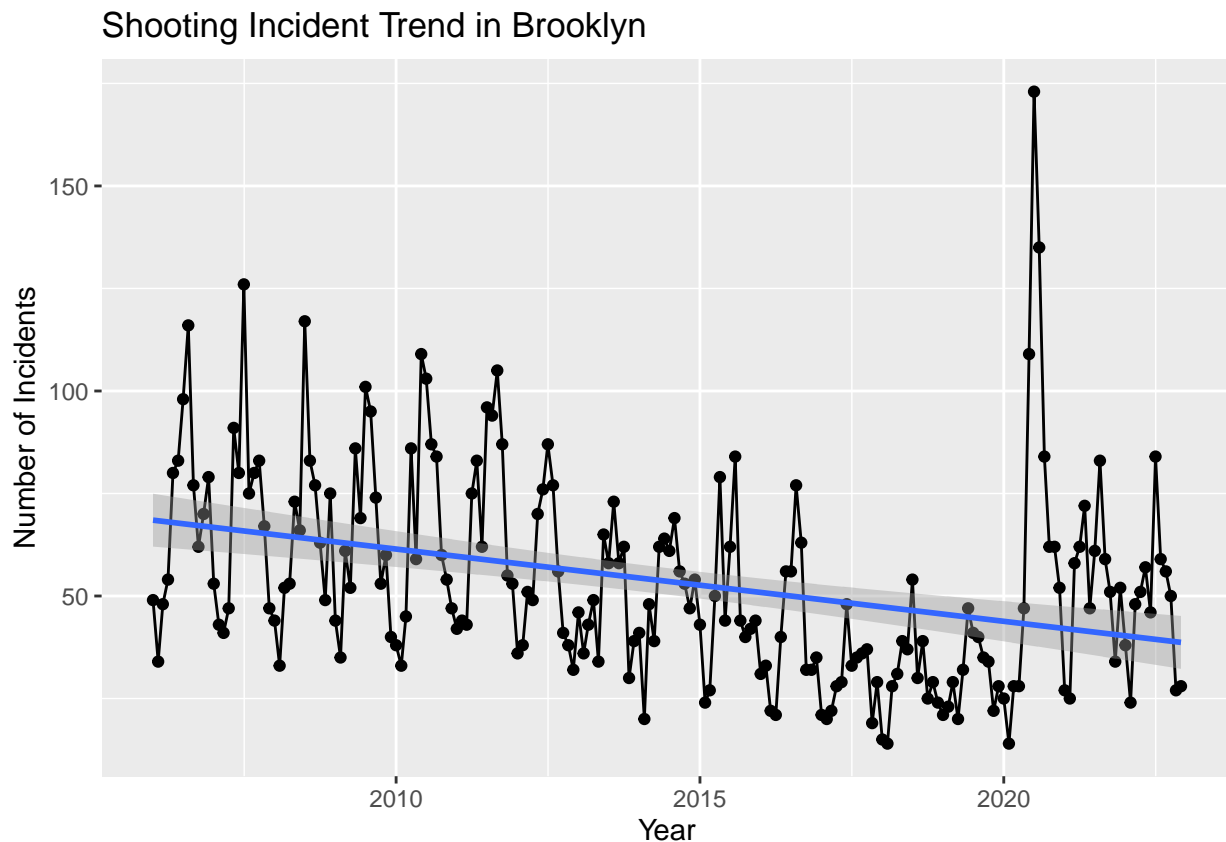
In order to further analyze the trend of the shooting incidents in Brooklyn, the time series chart is generated. Incident data are aggregated by month for the ease of analysis. From 2006 to 2020, the overall number of shooting incidents in Brooklyn is trending down despite the ups and downs. In late 2020, the number of

shooting incident reaches its peak. This is the same year that COVID-19 has a major outbreak. It is possible that during the time of hardship with many shutdowns of businesses, job losses, and sickness, more crime and violence tend to happen. From 2021 onward, the number of incidents has dropped significantly. This aligns with the fact that New York City was fully reopen on July 1, 2021 where everything slowly returned to a new normal.

A linear model, as shown in the code `geom_smooth(method = "lm")`, is used to predict number of incidents in the future based on the current trend. Overall, the number of shooting incidents is trending downward and therefore based on the linear regression (blue line), it is likely that 2023 and onward this trend might continue, unless there is a disrupting event such as pandemics or great depression.

```
incd_by_month <- df %>%
  filter (BORO=="BROOKLYN") %>%
  group_by(month = lubridate::floor_date(OCCUR_DATE, 'month')) %>%
  summarize(sum_occ_count = sum(OCC_COUNT))

ggplot(incd_by_month, aes(x = month, y = sum_occ_count)) +
  geom_point() +
  geom_line() +
  geom_smooth(method = "lm") +
  xlab("Year") +
  ylab("Number of Incidents") +
  ggtitle("Shooting Incident Trend in Brooklyn")
```



To ensure that my observation of more violence during COVID-19 time is not just an opinion or a bias, a similar trend should also be seen across all boroughs in the New York City. The next four plots show the shooting trends in Bronx, Queens, Manhattan, and Staten Island.

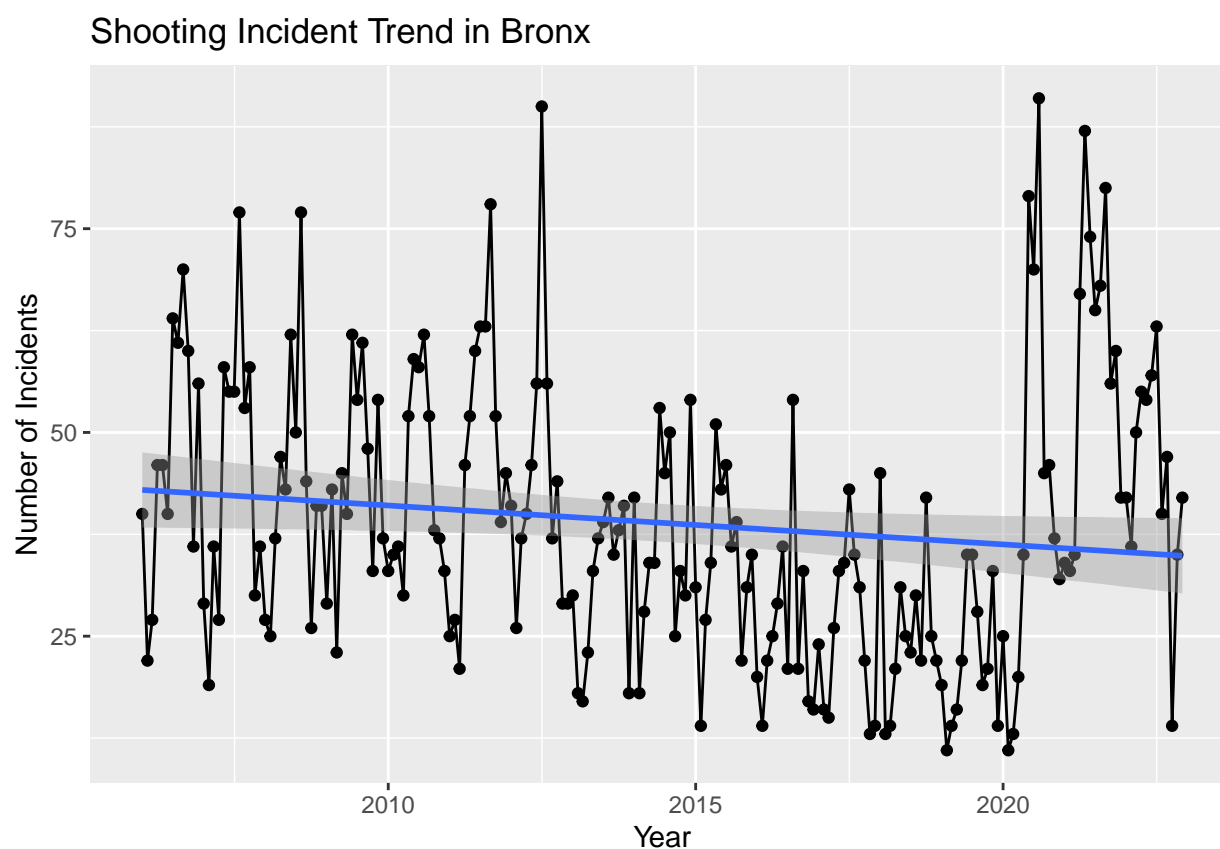
The following four plots contain a similar trend as that of Brooklyn where there is a ramp up of the number of shooting incidents from 2019 to 2020, where in 2020 COVID-19 exploded full force in the U.S. Manhattan does not have a downward trend in the cases of shooting incidents like other boroughs. This may be due to other circumstances unique to Manhattan that was not present in the dataset.

```

incd_by_month <- df %>%
  filter (BORO=="BRONX") %>%
  group_by(month = lubridate::floor_date(OCCUR_DATE, 'month')) %>%
  summarize(sum_occ_count = sum(OCC_COUNT))

ggplot(incd_by_month, aes(x = month, y = sum_occ_count)) +
  geom_point() +
  geom_line() +
  geom_smooth(method = "lm") +
  xlab("Year") +
  ylab("Number of Incidents") +
  ggtitle("Shooting Incident Trend in Bronx")

```



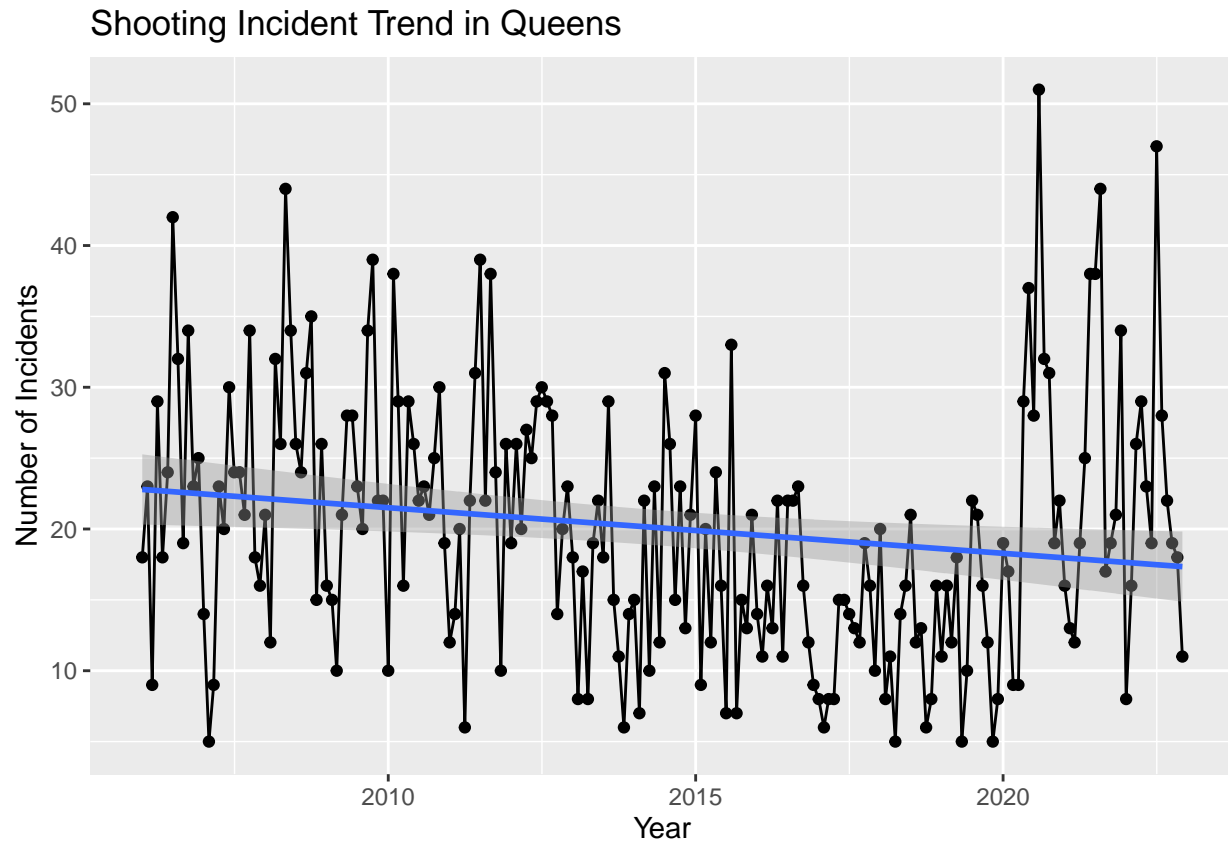
```

incd_by_month <- df %>%
  filter (BORO=="QUEENS") %>%
  group_by(month = lubridate::floor_date(OCCUR_DATE, 'month')) %>%
  summarize(sum_occ_count = sum(OCC_COUNT))

ggplot(incd_by_month, aes(x = month, y = sum_occ_count)) +
  geom_point() +

```

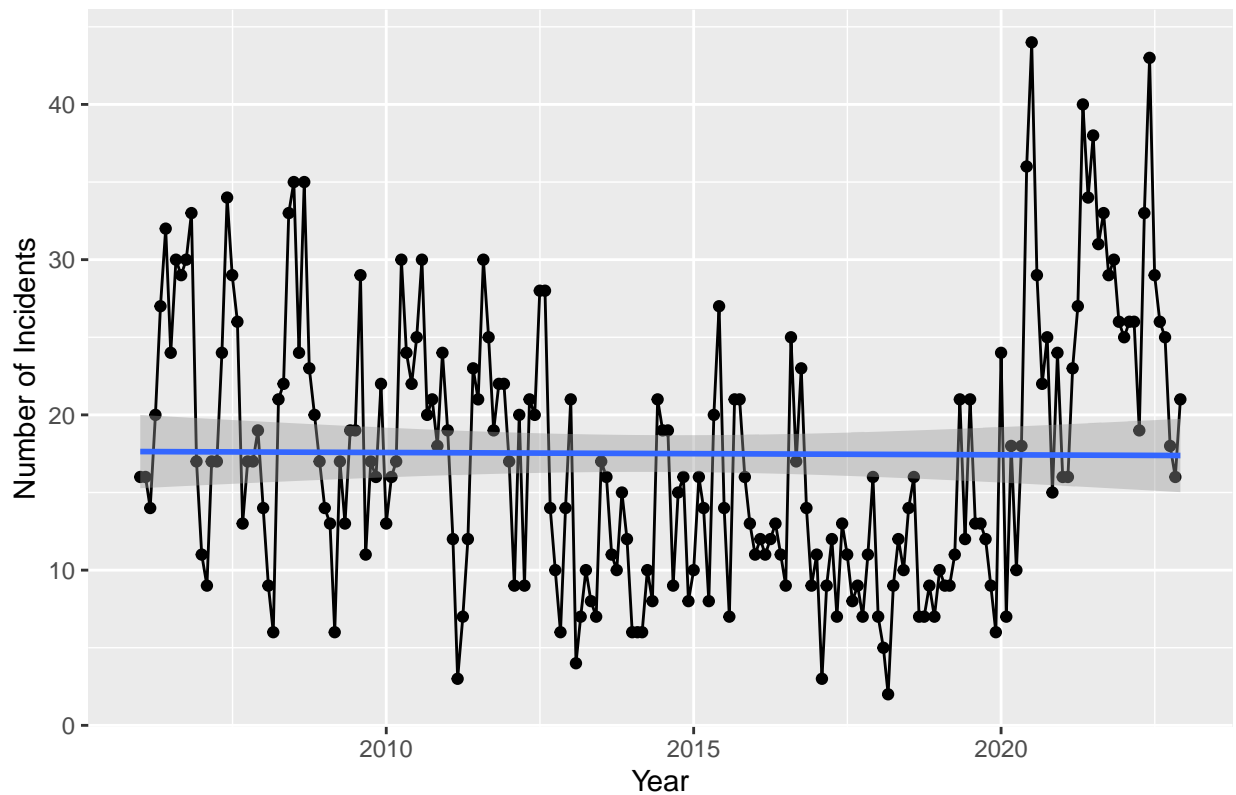
```
geom_line() +
geom_smooth(method = "lm") +
xlab("Year") +
ylab("Number of Incidents") +
ggtitle("Shooting Incident Trend in Queens")
```



```
incd_by_month <- df %>%
  filter (BORO=="MANHATTAN") %>%
  group_by(month = lubridate::floor_date(OCCUR_DATE, 'month')) %>%
  summarize(sum_occ_count = sum(OCC_COUNT))

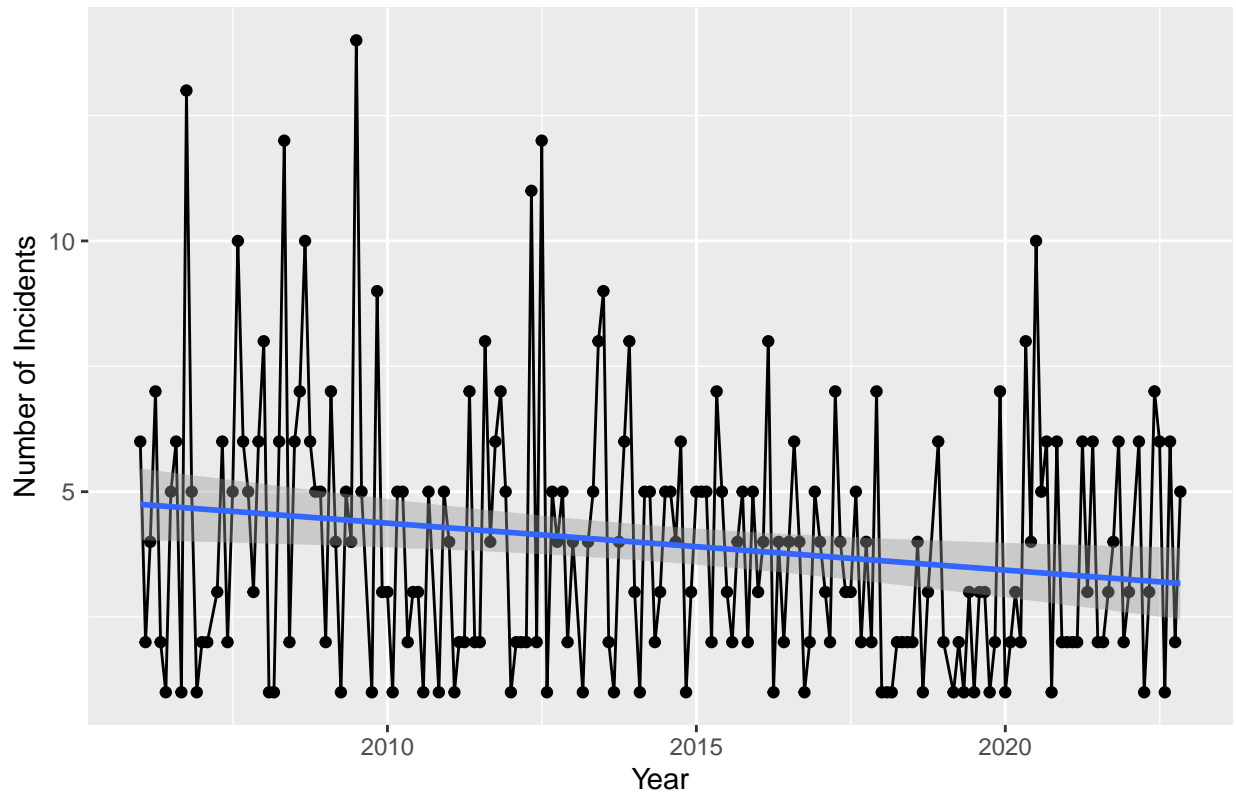
ggplot(incd_by_month, aes(x = month, y = sum_occ_count)) +
  geom_point() +
  geom_line() +
  geom_smooth(method = "lm") +
  xlab("Year") +
  ylab("Number of Incidents") +
  ggtitle("Shooting Incident Trend in Manhattan")
```

Shooting Incident Trend in Manhattan



```
incd_by_month <- df %>%  
  filter (BORO=="STATEN ISLAND") %>%  
  group_by(month = lubridate::floor_date(OCCUR_DATE, 'month')) %>%  
  summarize(sum_occ_count = sum(OCC_COUNT))  
  
ggplot(incd_by_month, aes(x = month, y = sum_occ_count)) +  
  geom_point() +  
  geom_line() +  
  geom_smooth(method = "lm") +  
  xlab("Year") +  
  ylab("Number of Incidents") +  
  ggtitle("Shooting Incident Trend in Staten Island")
```

Shooting Incident Trend in Staten Island



Conclusion

In conclusion, the number of shooting incidents in Brooklyn, Bronx, Queens, and Staten Island seem to be trending down over time. Disrupting events such as the COVID-19 pandemics may have caused the violence to increase, resulting in higher number of shooting occurrences during that time. As the time becomes more stable, the shooting incidents may decline over time. Since the dataset is only for New York City region, it is not possible to generalize the same result to the rest of the U.S.