NY Shootings Analysis Post Great Recession - Covid

Thales

2023-05-10

In this document, I will delve into the shootings that occurred in New York City from the post-Great Recession era (starting in 2010) up until the first two years of the Covid-19 Pandemic (2020-2022). Our goal is to analyze governmental datasets and identify potential factors that may have influenced these incidents. By doing so, we hope to gain a better understanding of the root causes of these shootings.

Fetch Reports

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- 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
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
```

```
## The following object is masked from 'package:purrr':
##
##     compact
```

```
nypd_report_url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
ny_pop_report_url <- "https://data.ny.gov/api/views/krt9-ym2k/rows.csv?accessType=DOWNLOAD&sorting=true
raw_nypd_report = read_csv(nypd_report_url)
```

```
## Rows: 27312 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
raw_ny_pop_report <- read_csv(ny_pop_report_url)
```

```
## Rows: 3717 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Geography, Program Type
## dbl (3): FIPS Code, Year, Population
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Apply necessary tranformations for further Analysis

```
report <- raw_nypd_report %>%
          select(-X_COORD_CD, -Y_COORD_CD, -Latitude, -Longitude, -INCIDENT_KEY, - Lon_Lat, -PRECINCT,


report$OCCUR_DATE <- as.Date(report$OCCUR_DATE, format = "%m/%d/%Y")

report <- report %>%
            mutate(Year = lubridate::year(OCCUR_DATE)) %>%
            filter(Year >= 2010)

nyc_counties_list <- c("Bronx County", "Kings County", "New York County", "Queens County", "Richmond Cou

ny_pop_report <- raw_ny_pop_report %>%
                    filter(Geography %in% nyc_counties_list & Year >= 2010) %>%
                    mutate(
                      BORO = case_when(
                        Geography == "Bronx County" ~ "BRONX",
                        Geography == "Kings County" ~ "BROOKLYN",
                        Geography == "New York County" ~ "MANHATTAN",
```

```
                        Geography == "Queens County" ~ "QUEENS",
                        Geography == "Richmond County" ~ "STATEN ISLAND",
                        )
                    ) %>%
                    select(-"FIPS Code", -"Program Type", -"Geography") %>%
                    arrange(Year)
```

Summarize Data for periods of interest

```
incidents_10_19 <- report %>%
    filter(Year >= 2010 & Year < 2020) %>%
    group_by(Year) %>%
    dplyr::summarize(total_cases = n())

incidents_10_22 <- report %>%
    filter(Year >= 2010 & Year < 2023) %>%
    group_by(Year) %>%
    dplyr::summarize(total_cases = n())


incidents_per_boro_10_19 <- report %>%
    filter(lubridate::year(OCCUR_DATE) < 2020) %>%
    group_by(BORO, Year) %>%
    dplyr::summarize(`Total Cases` = n())
```

```
## `summarise()` has grouped output by 'BORO'. You can override using the
## `.groups` argument.
```

```
incidents_per_boro_10_22 <- report %>%
    filter(lubridate::year(OCCUR_DATE) >= 2010) %>%
    group_by(BORO, Year) %>%
    dplyr::summarize(`Total Cases` = n())
```

```
## `summarise()` has grouped output by 'BORO'. You can override using the
## `.groups` argument.
```

```
pop_report_10_22 <- merge(
                      incidents_per_boro_10_22,
                      ny_pop_report,
                      by=c("BORO","Year"),
                      all.x=TRUE
                      )

pop_report_10_19 <- merge(
                      incidents_per_boro_10_19,
                      ny_pop_report,
                      by=c("BORO","Year"),
                      all.x=TRUE
                      )
```

Merge summarised data with dataset containg population and adding calculated columns

```
deaths_per_boro_10_19 <- report %>%
                  filter(Year < 2020 & STATISTICAL_MURDER_FLAG == TRUE) %>%
                  group_by(BORO, Year) %>%
                  dplyr::summarize(Deaths = n()) %>%
                  arrange(BORO, Year)
```

```
## `summarise()` has grouped output by 'BORO'. You can override using the
## `.groups` argument.
```

```
deaths_per_boro_10_22 <- report %>%
                  filter(Year >= 2010 & STATISTICAL_MURDER_FLAG == TRUE) %>%
                  group_by(BORO, Year) %>%
                  dplyr::summarize(Deaths = n()) %>%
                  arrange(BORO, Year)
```

```
## `summarise()` has grouped output by 'BORO'. You can override using the
## `.groups` argument.
```

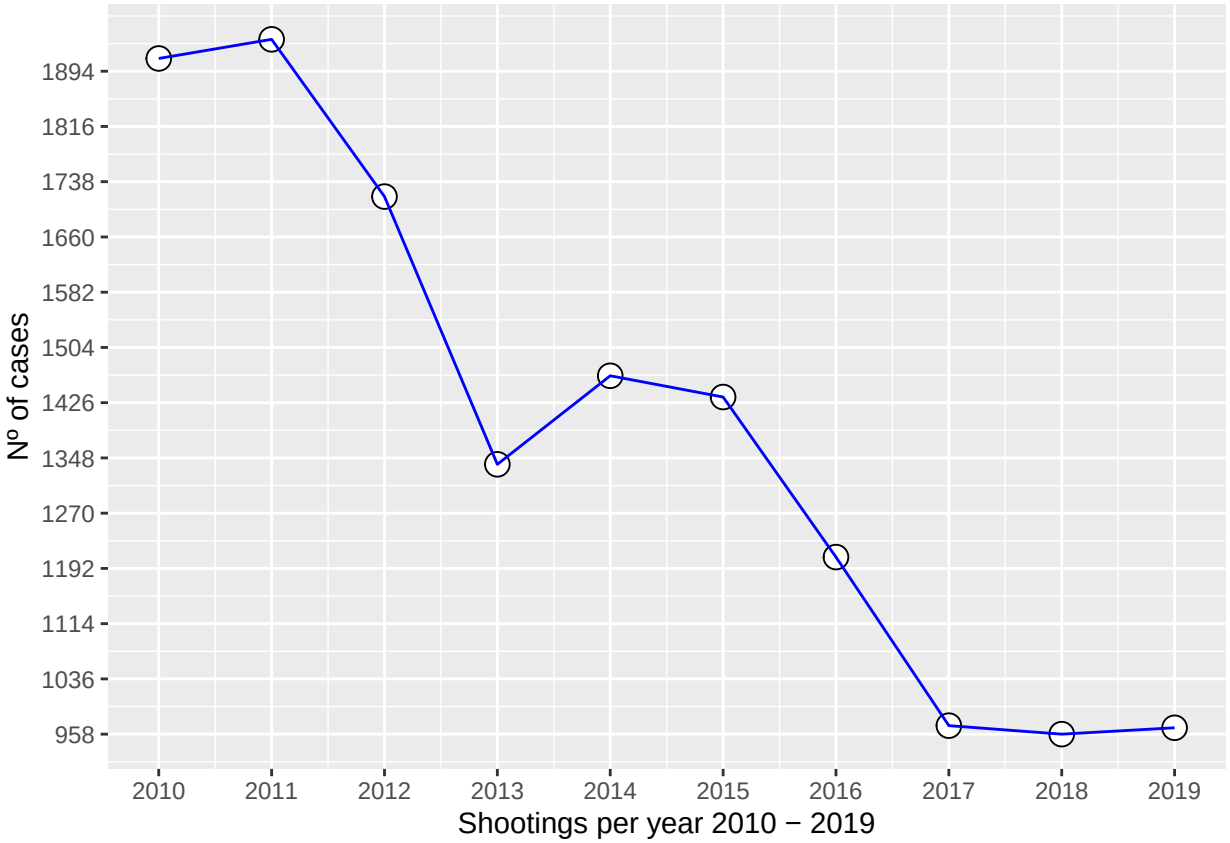
```
full_trend_report_10_19 <- merge(
                            pop_report_10_19,
                            deaths_per_boro_10_19,
                            by=c("BORO","Year"),
                            all.x=TRUE
                            )


full_trend_report_10_22 <- merge(
                            pop_report_10_22,
                            deaths_per_boro_10_22,
                            by=c("BORO","Year"),
                            all.x=TRUE
                            )

pop_report_10_19 <- full_trend_report_10_19 %>%
    mutate("Incidents per 100k residents " = (`Total Cases`/Population)*10.0**4) %>%
    mutate("Deaths per 100k residents " = (Deaths/Population)*10.0**4)


pop_report_10_19 <- pop_report_10_19 %>%
  mutate(`Incidents per 100k residents ` = round(`Incidents per 100k residents `, digits = 2)) %>%
  mutate(`Deaths per 100k residents ` = round(`Deaths per 100k residents `, digits = 2))


pop_report_10_22 <- full_trend_report_10_22 %>%
    mutate("Incidents per 100k residents " = (`Total Cases`/Population)*10.0**4) %>%
    mutate("Deaths per 100k residents " = (Deaths/Population)*10.0**4)


pop_report_10_22 <- pop_report_10_22 %>%
```

```
  mutate(`Incidents per 100k residents ` = round(`Incidents per 100k residents `, digits = 2)) %>%
  mutate(`Deaths per 100k residents ` = round(`Deaths per 100k residents `, digits = 2))
```

Shootings trend 2010 - 2019

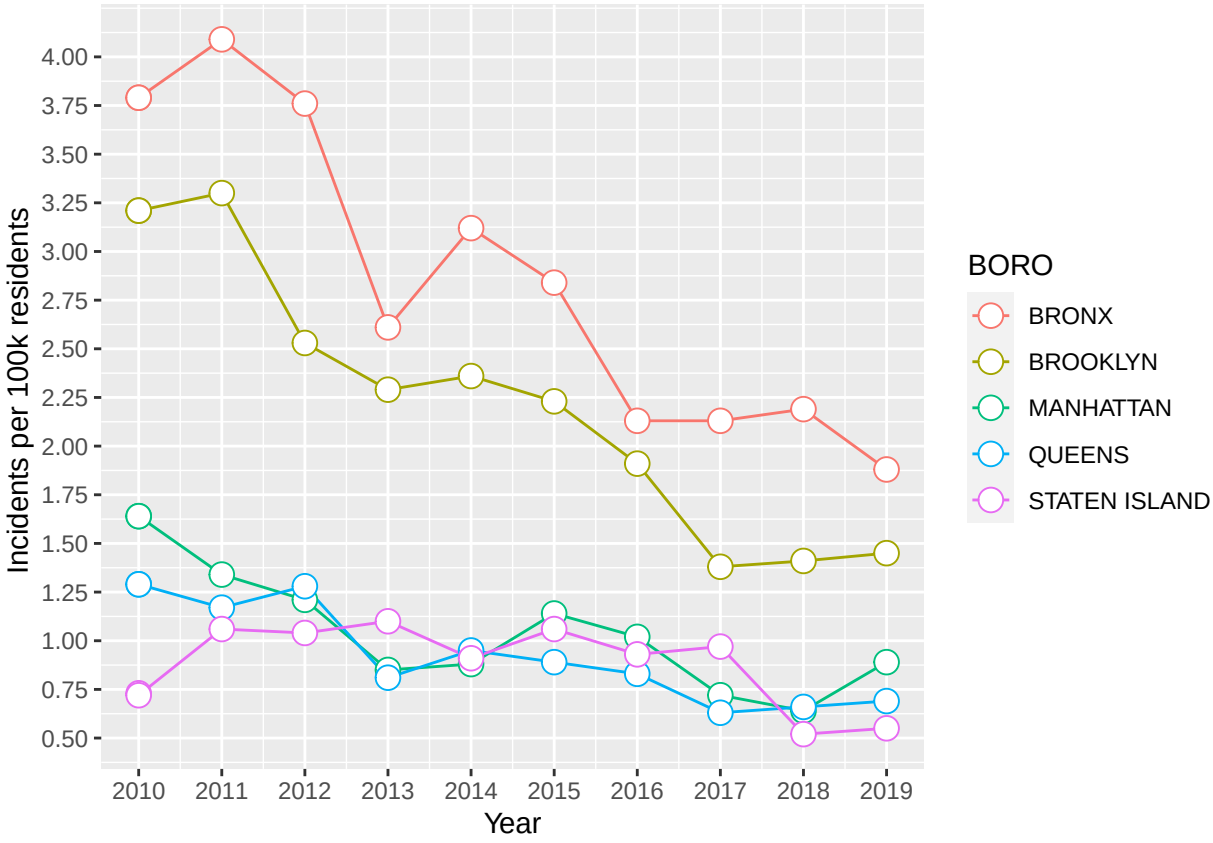
We can observe a downward trend in the number of shootings in the period after The Great Recession Until 2019 (before the Covid Pandemic Crisis)

```
ggplot(data = incidents_10_19, aes(x = Year, y = total_cases)) +
  geom_point( size=4, shape=21, fill="white") + geom_line(color = "blue") +
  xlab("Year") +
  ylab("Nº of cases") +
  ggtitle("Shootings per year 2010 - 2019") %>%
  scale_x_continuous(
    breaks=seq(
      min(incidents_10_19$Year),
      max(incidents_10_19$Year),
      by=1
      )
    ) +
  scale_y_continuous(
    breaks=seq(
      min(incidents_10_19$total_cases)
      ,max(incidents_10_19$total_cases)
      ,
      by=78
      )
    )
```



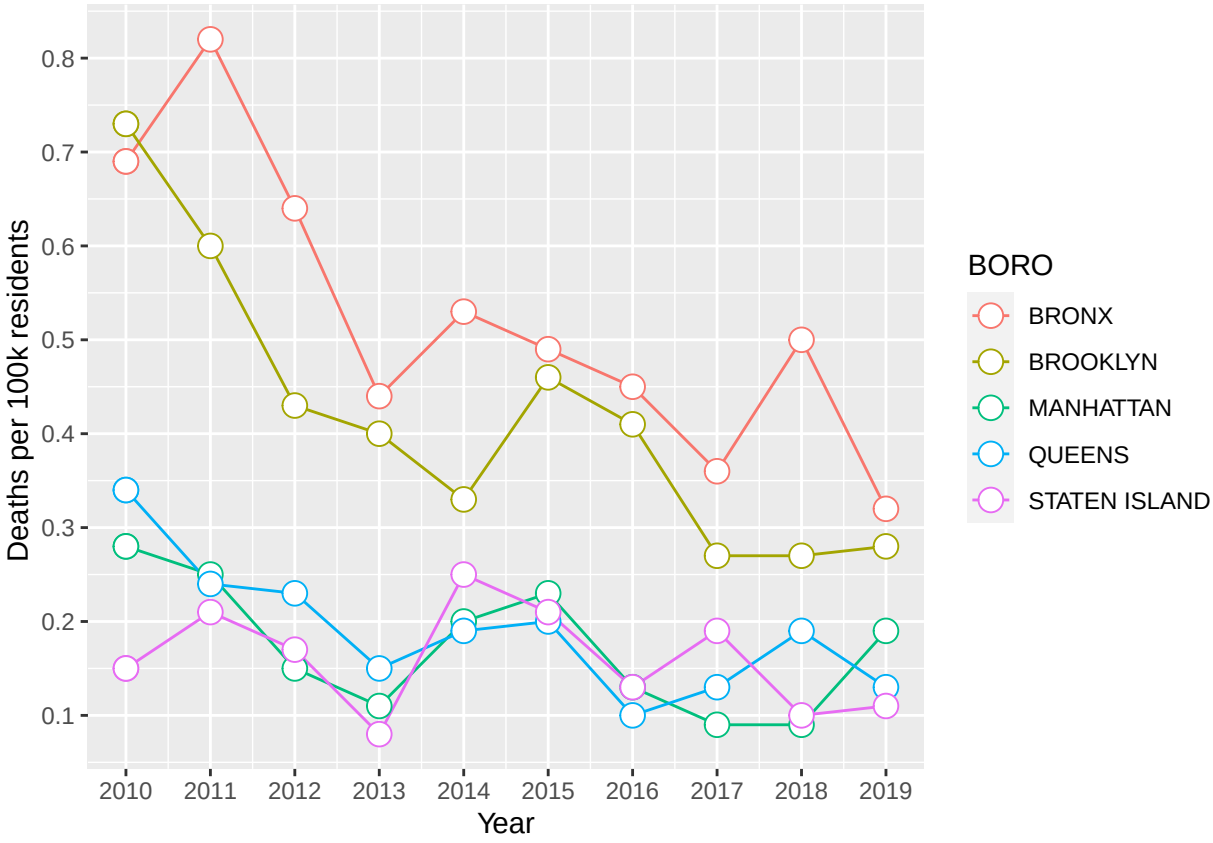
The trend can also be observed when drilling down the data. In every Borough, the amount of incidents per 100k residents decreased

```
ggplot(data=pop_report_10_19,aes(x=Year,y=`Incidents per 100k residents `,group=BORO,colour=BORO)) +
    geom_line() +
    geom_point( size=4, shape=21, fill="white") +
    scale_x_continuous(
      breaks=seq(
        min(pop_report_10_19$Year),
        max(pop_report_10_19$Year),
        by=1
        )
      ) +
    scale_y_continuous(
      breaks=seq(
        0,
        max(pop_report_10_19$`Incidents per 100k residents `),
        by=0.25
        )
      )
```



As expected, the same happened in the amount of deaths per 100k residents

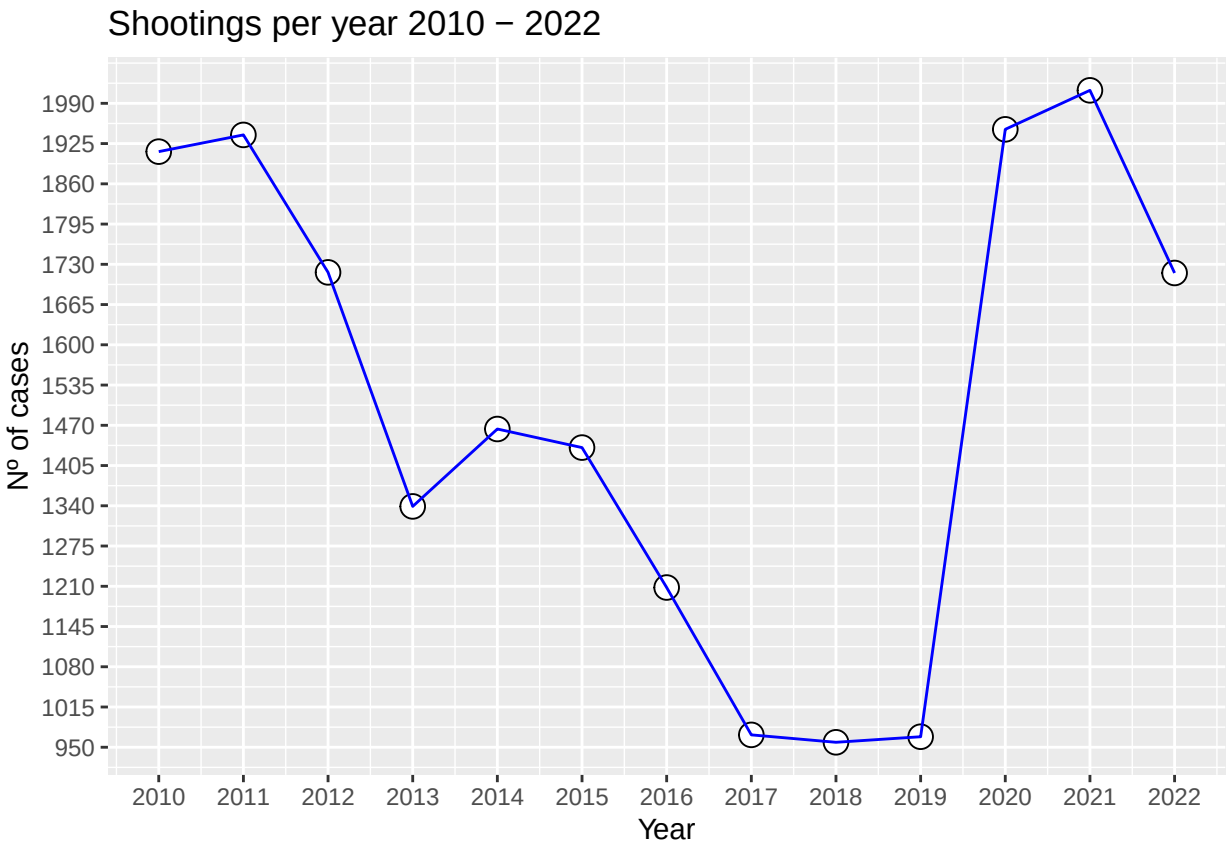
```
ggplot(data=pop_report_10_19, aes(x=Year,y=`Deaths per 100k residents `,group=BORO,colour=BORO)) +
    geom_line() +
    geom_point( size=4, shape=21, fill="white") +
    scale_x_continuous(
      breaks=seq(
        min(pop_report_10_19$Year),
        max(pop_report_10_19$Year),
        by=1
        )
      ) +
    scale_y_continuous(
      breaks=seq(
        0,
        max(pop_report_10_19$`Deaths per 100k residents `),
        by = 0.1
        )
      )
```



Upward trend in shootings starting in 2020

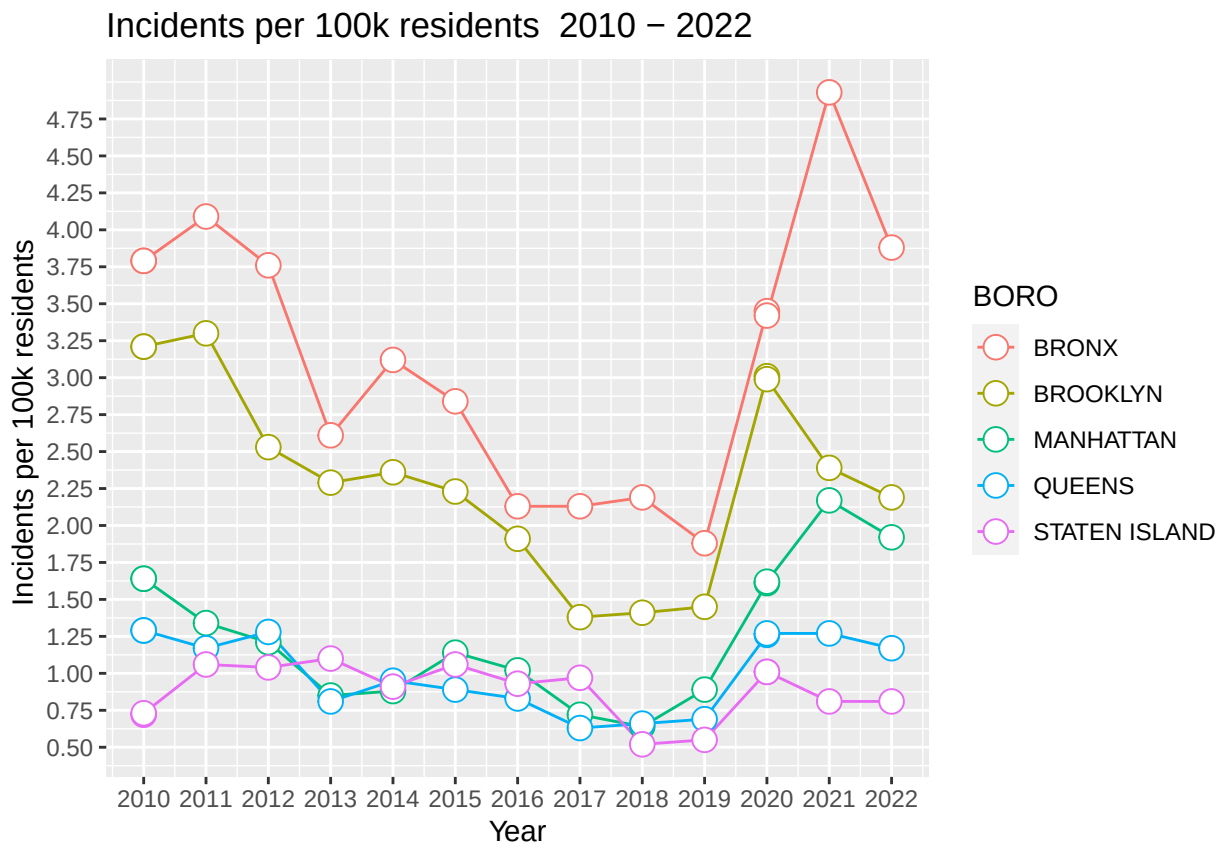
During the Pandemic, we can observe a spike in the number of cases

```
ggplot(data = incidents_10_22, aes(x = Year, y = total_cases)) +
  geom_point(size=4, shape=21, fill="white") + geom_line(color = "blue") +
  xlab("Year") +
  ylab("Nº of cases") +
  ggtitle("Shootings per year 2010 - 2022") +
  scale_x_continuous(
    breaks=seq(
      min(incidents_10_22$Year),
      max(incidents_10_22$Year),
      by=1
      )
    ) +
  scale_y_continuous(
    breaks=seq(
      950,
      max(incidents_10_22$total_cases),
      by = 65
      )
    )
```

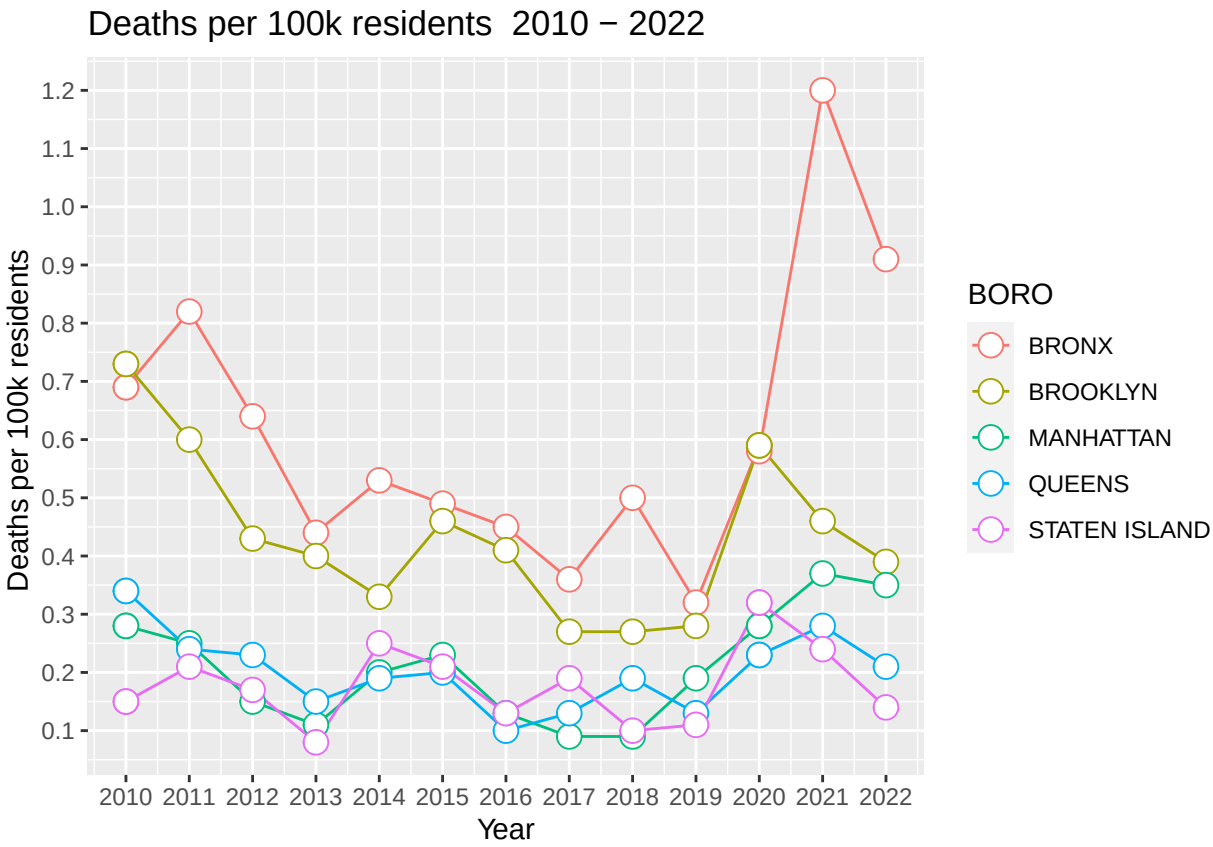


Also observed in every Borough in NYC

```
ggplot(data=pop_report_10_22,aes(x=Year,y=`Incidents per 100k residents `,group=BORO,colour=BORO)) +
    geom_line() +
    geom_point( size=4, shape=21, fill="white") +
    ggtitle("Incidents per 100k residents  2010 - 2022") +
    scale_x_continuous(
      breaks=seq(
        min(pop_report_10_22$Year),
        max(pop_report_10_22$Year),
        by=1
        )
      ) +
    scale_y_continuous(
      breaks=seq(
        0,
        max(pop_report_10_22$`Incidents per 100k residents `),
        by=0.25
        )
      )
```




```
ggplot(data=pop_report_10_22, aes(x=Year, y=`Deaths per 100k residents `, group = BORO, colour = BORO))
    geom_line() +
    geom_point( size=4, shape=21, fill="white") +
    ggtitle("Deaths per 100k residents  2010 - 2022") +
    scale_x_continuous(breaks = seq(min(pop_report_10_22$Year), max(pop_report_10_22$Year), by = 1)) +
    scale_y_continuous(breaks = seq(0, max(pop_report_10_22$`Deaths per 100k residents `), by = 0.1))
```



The unemployment rate could be a factor that contributes to the increase in gun violence. We'll import a New York State Dataset and analyze if there is a relationship for further discussion

Importing NY unemployment dataset

```
unemployment_url  <- "https://fred.stlouisfed.org/graph/fredgraph.csv?bgcolor=%23e1e9f0&chart_type=line

unemployment_report = read_csv(unemployment_url)
```

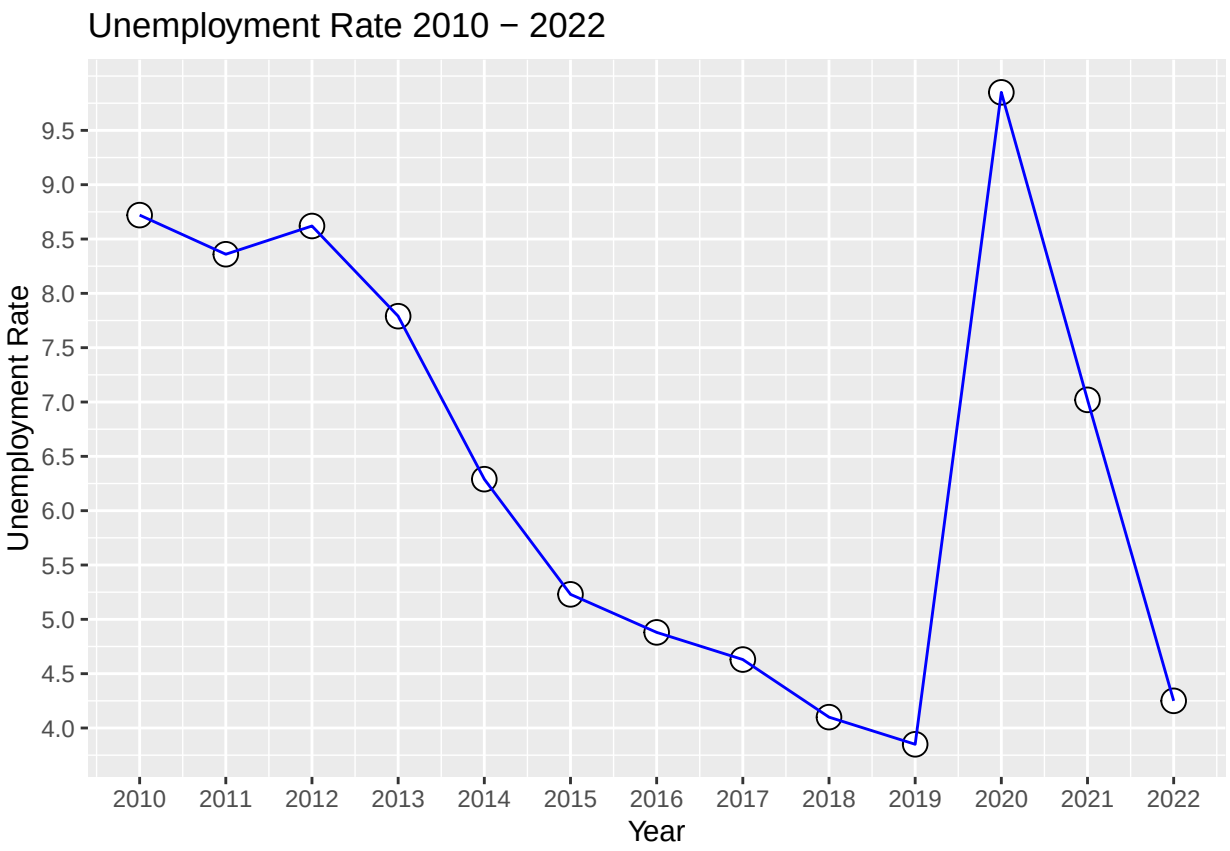
```
## Rows: 567 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (1): NYUR
## date (1): DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
unempleyment_yearly_summary <- unemployment_report %>%
  mutate(Year = lubridate::year(DATE)) %>%
  filter(Year > 2005 & Year < 2023) %>%
  group_by(Year) %>%
```

```
  summarise_at(vars(NYUR), list(`Unemployment Rate` = mean)) %>%
  mutate(`Unemployment Rate` = round(`Unemployment Rate`, digits = 2))
```

Analyzing unemployment rate over the same period

```
ggplot(data = unempleyment_yearly_summary %>% filter(Year >= 2010),aes(x=Year,y =`Unemployment Rate`)) +
  geom_point( size=4, shape=21, fill="white") + geom_line(color = "blue") +
  xlab("Year") +
  ylab("Unemployment Rate") +
  ggtitle("Unemployment Rate 2010 - 2022") +
  scale_x_continuous(
    breaks=seq(
      min(unempleyment_yearly_summary$Year),
      max(unempleyment_yearly_summary$Year),
      by=1
      )
    ) +
  scale_y_continuous(
    breaks=seq(
      0,
      max(unempleyment_yearly_summary$`Unemployment Rate`),
      by=0.5
      )
    )
```



Merging Unenmployment and Shooting Incidents Datasets

```
shootings_unemployment_set <- merge(
                                incidents_10_22,
                                unempleyment_yearly_summary %>%
                                  filter(Year >= 2010),
                                by=c("Year"),
                                all.x=TRUE
                                )
```

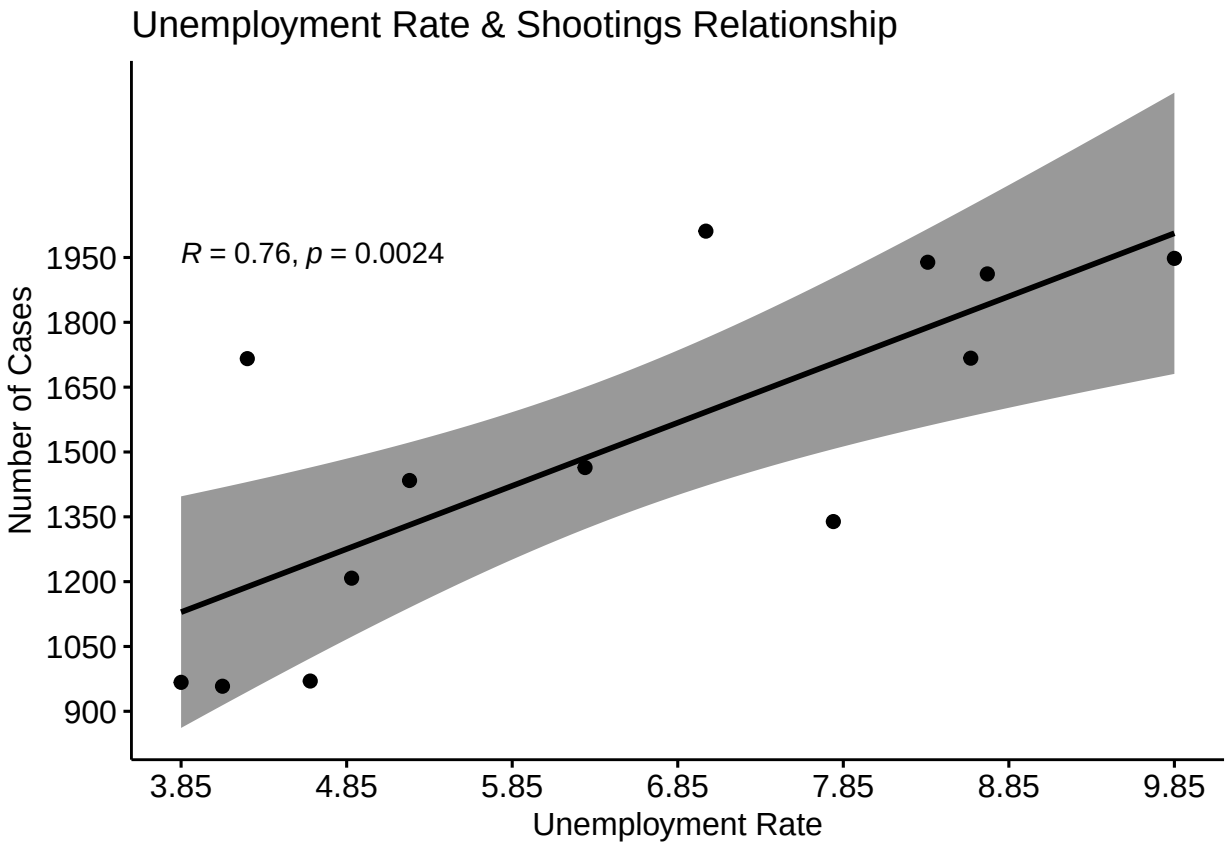
Pearson correlation Analysis

```
library("ggpubr")
```

```
##
## Attaching package: 'ggpubr'
```

```
## The following object is masked from 'package:plyr':
##
##     mutate
```

```
ggscatter(shootings_unemployment_set, x = "Unemployment Rate", y = "total_cases",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Unemployment Rate", ylab = "Number of Cases") +
          ggtitle("Unemployment Rate & Shootings Relationship") +
          scale_x_continuous(
            breaks=seq(
              min(shootings_unemployment_set$`Unemployment Rate`),
              max(shootings_unemployment_set$`Unemployment Rate`),
              by=1
              )
            ) +
          scale_y_continuous(
            breaks=seq(
              0,
              max(shootings_unemployment_set$total_cases),
              by=150
              )
            )
```



With a correlation of 0.76 and a P value of 0.0024, the data suggests that the unemployment could be related to the Number of Shootings in NYC.

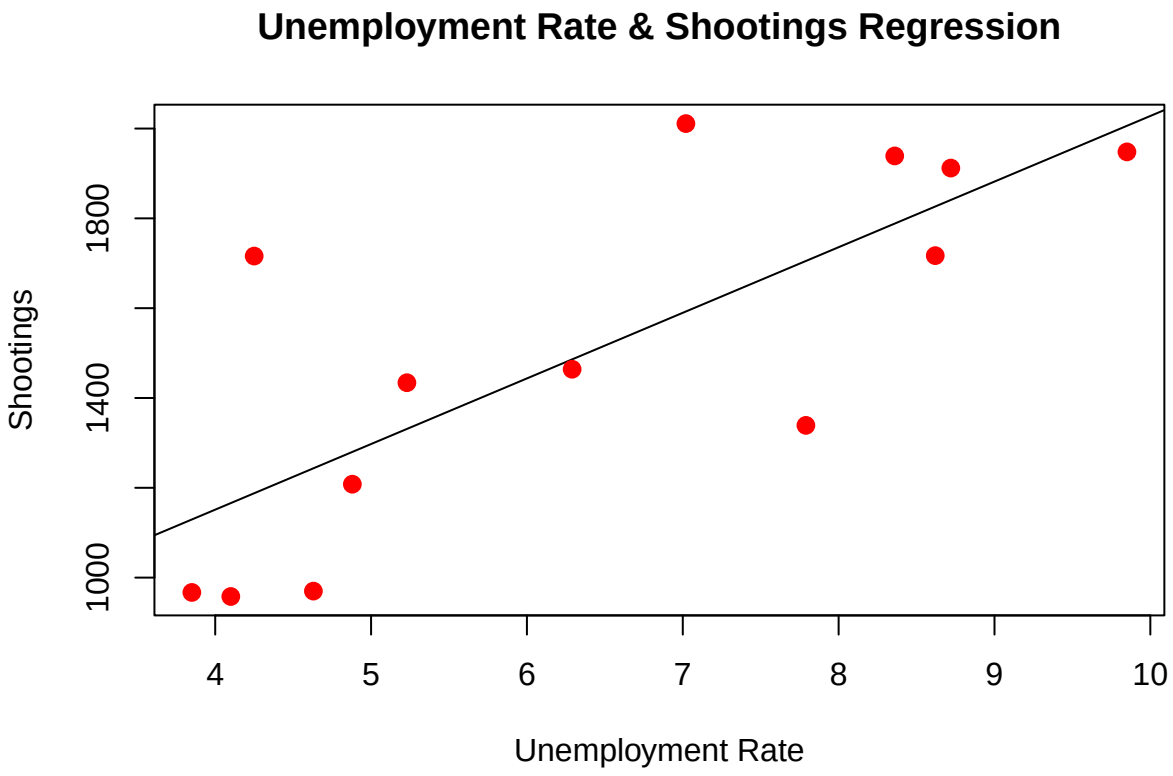
Linear regression Analysis

```
relation <- lm(shootings_unemployment_set$`Unemployment Rate`~shootings_unemployment_set$total_cases)
relation
```

```
##
## Call:
## lm(formula = shootings_unemployment_set$`Unemployment Rate` ~
##     shootings_unemployment_set$total_cases)
##
## Coefficients:
##                               (Intercept)  shootings_unemployment_set$total_cases
##                                  0.431318                                0.003982
```

```
# Plot the chart.
plot(
  shootings_unemployment_set$`Unemployment Rate`,
  shootings_unemployment_set$total_cases,
  col = "red",
  main = "Unemployment Rate & Shootings Regression",
  abline(
```

```
    lm(
      shootings_unemployment_set$total_cases~shootings_unemployment_set$`Unemployment Rate`
      ))
 ,cex = 1.3,
 pch = 16,
 xlab = "Unemployment Rate",
 ylab = "Shootings"
 )
```



The linear regression plot also suggests that variations in the Unemployment Rate could be a factor that contributes to variations in shooting incidents in NYC, as observed in the spike during 2020-2022.

Possible Bias Sources

- Unemployment rates were based on average value for New York State over 2010-2022, therefore including rates for all Boroughs/Counties in the state
- Boroughs and Counties were assumed to be interchangeable (eg: "Kings County" ~ "Brooklyn") for population mapping
- A more in depth analysis would be necessary, including other factors like income distribution, poverty levels, and correlation with other criminal activities in order to better understand the observed trend