Is the city less safe then before?

Patrick Kendall

06 December 2023

```r
knitr::opts_chunk$set(message=FALSE)
```

```r
library(tidyverse)
```

Is the city less safe then before?

Off-topic in a status meeting, a coworker was discussing the city's safety and how it has worsened over the years. When asked if he planned on moving out, he walked back his statements slightly by stating that it was just certain regions and times of the evening. Clearly, the coworker has yet to outgrow the city. Still, it raises the question of whether it is less safe now or if city life is losing that exciting glamour? Living in the city since 2008, my coworker brings a lot of empirical analysis. I moved out of the city just after 200, so I will have to turn to historical data for analysis.

New York City publishes a breakdown of every shooting incident from 2006 through the end of the previous year. The dataset contains many attributes for location, perpetrator, and victim. The dataset is reviewed by the Office of Management Analysis and Planning. Further information about this dataset can be found at Data.Gov or NYC OpenData.

```r
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
nypd_data <- read.csv(url_in)
str(nypd_data)
```

```
## 'data.frame':    27312 obs. of  21 variables:
##  $ INCIDENT_KEY           : int  228798151 137471050 147998800 146837977 58921844 219559682 85295722
##  $ OCCUR_DATE             : chr  "05/27/2021" "06/27/2014" "11/21/2015" "10/09/2015" ...
##  $ OCCUR_TIME             : chr  "21:30:00" "17:40:00" "03:56:00" "18:30:00" ...
##  $ BORO                   : chr  "QUEENS" "BRONX" "QUEENS" "BRONX" ...
##  $ LOC_OF_OCCUR_DESC      : chr  "" "" "" "" ...
##  $ PRECINCT               : int  105 40 108 44 47 81 114 81 105 101 ...
##  $ JURISDICTION_CODE      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LOC_CLASSFCTN_DESC     : chr  "" "" "" "" ...
##  $ LOCATION_DESC          : chr  "" "" "" "" ...
##  $ STATISTICAL_MURDER_FLAG: chr  "false" "false" "true" "false" ...
##  $ PERP_AGE_GROUP         : chr  "" "" "" "" ...
##  $ PERP_SEX               : chr  "" "" "" "" ...
##  $ PERP_RACE              : chr  "" "" "" "" ...
##  $ VIC_AGE_GROUP          : chr  "18-24" "18-24" "25-44" "<18" ...
##  $ VIC_SEX                : chr  "M" "M" "M" "M" ...
##  $ VIC_RACE               : chr  "BLACK" "BLACK" "WHITE" "WHITE HISPANIC" ...
##  $ X_COORD_CD             : num  1058925 1005028 1007668 1006537 1024922 ...
##  $ Y_COORD_CD             : num  180924 234516 209837 244511 262189 ...
##  $ Latitude               : num  40.7 40.8 40.7 40.8 40.9 ...
##  $ Longitude              : num  -73.7 -73.9 -73.9 -73.9 -73.9 ...
##  $ Lon_Lat                 : chr  "POINT (-73.73083868899994 40.662964620000025)" "POINT (-73.9249423
```

The dataset is just shooting incidents, so it isn't all crimes that would impact public safety or even all crimes where a gun was involved but never fired. The dataset, however, is complete, verified, publicly available, and spans the desired range we are looking for. Unfortunately the dataset has too many columns and the date columns are of type chr.

```r
nypd_data <- subset(nypd_data,select = c(INCIDENT_KEY,OCCUR_DATE,OCCUR_TIME,BORO,PRECINCT))
nypd_data$dateof <- lubridate::mdy(stringr::str_c(nypd_data$OCCUR_DATE))
nypd_data$year <- lubridate::year(nypd_data$dateof)
nypd_data <- nypd_data %>% filter(year < 2023)

nypd_data <- subset(nypd_data,select = -c(OCCUR_DATE,OCCUR_TIME))
nypd_data <- dplyr::rename(nypd_data,key = INCIDENT_KEY,boro = BORO,precinct = PRECINCT)

by_year <- nypd_data %>% group_by(year) %>% count()
by_year <- dplyr::rename(by_year,incidents = n)
stkd_cntlst <- as.list(by_year$incidents)
stkd_mean <- round(mean(unlist(stkd_cntlst)),digits = 1)
stkd_sd <- round(sd(unlist(stkd_cntlst)),digits = 1)
stkd_2008 <- round((by_year %>% filter(year == 2008) %>%
                    select(incidents))$incidents,digits = 0)
stkd_2022 <- round((by_year %>%
                    filter(year == 2022) %>%
                    select(incidents))$incidents,digits = 0)
rm(by_year)
rm(stkd_cntlst)

nypd_boroyear_aggr <- nypd_data %>% group_by(boro,year) %>% count()
nypd_boroyear_aggr <- dplyr::rename(nypd_boroyear_aggr,incidents = n)

str(nypd_data)
```

```
## 'data.frame':    27312 obs. of  5 variables:
##  $ key     : int  228798151 137471050 147998800 146837977 58921844 219559682 85295722 71662474 83002
##  $ boro    : chr  "QUEENS" "BRONX" "QUEENS" "BRONX" ...
##  $ precinct: int  105 40 108 44 47 81 114 81 105 101 ...
##  $ dateof  : Date, format: "2021-05-27" "2014-06-27" ...
##  $ year    : num  2021 2014 2015 2015 2009 ...
```

```r
summary(nypd_boroyear_aggr)
```

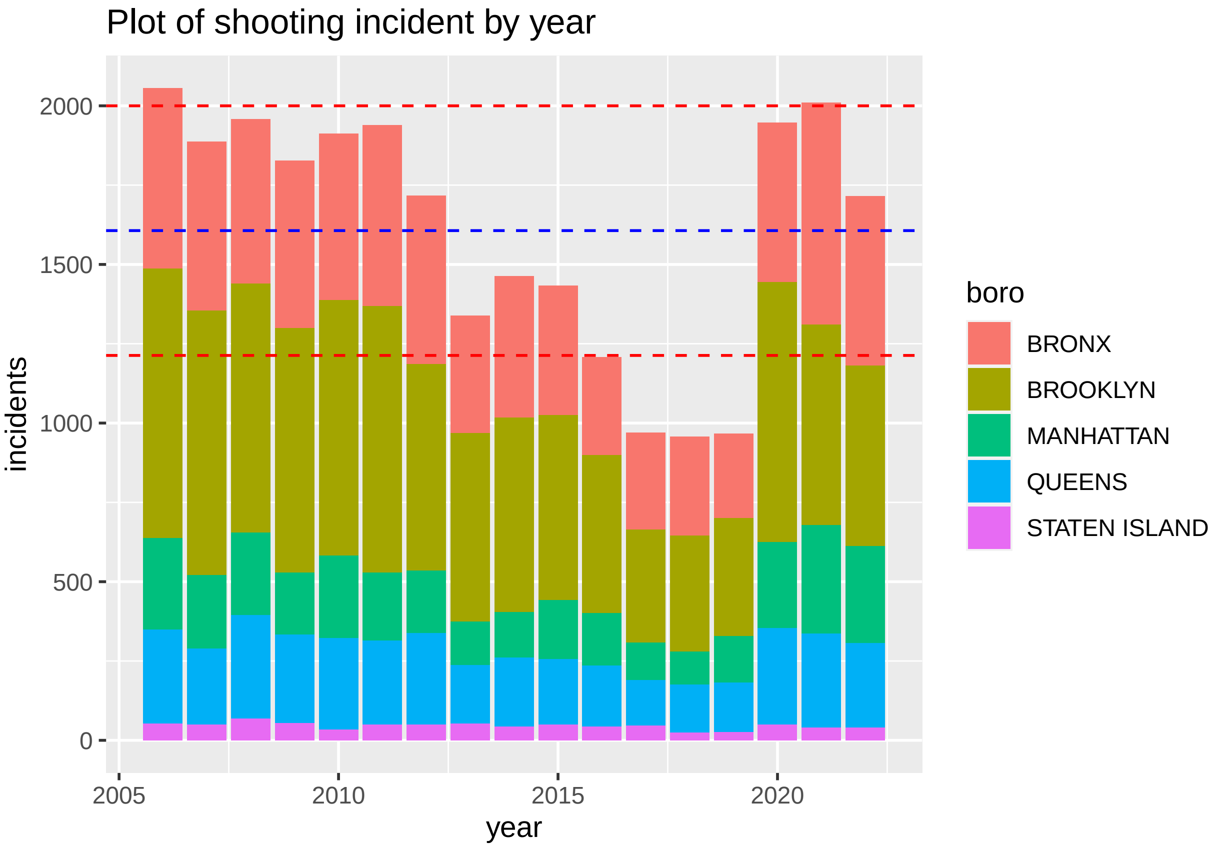
```
##      boro                year        incidents
##  Length:85          Min.   :2006   Min.   : 25.0
##  Class :character   1st Qu.:2010   1st Qu.:144.0
##  Mode  :character   Median :2014   Median :278.0
##                     Mean   :2014   Mean   :321.3
##                     3rd Qu.:2018   3rd Qu.:520.0
##                     Max.   :2022   Max.   :850.0
```

```r
sbs_cntlst <- as.list(nypd_boroyear_aggr$incidents)
sbs_mean <- round(mean(unlist(sbs_cntlst)),digits = 1)
sbs_sd <- round(sd(unlist(sbs_cntlst)),digits = 1)
rm(sbs_cntlst)
```

Is the city safer now than it was in the past?

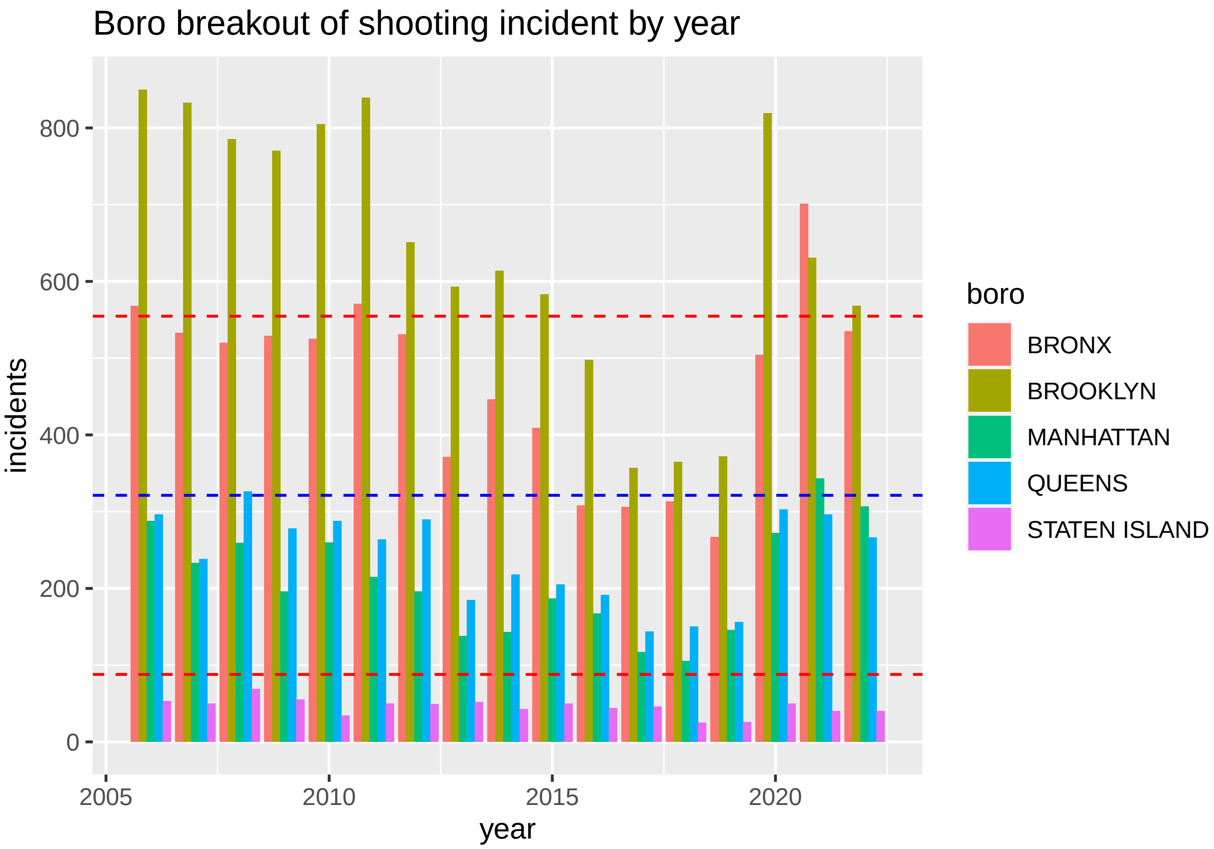
By aggregating all data from each Boro into a year, we see that the mean is 1606.6 incidents, and the standard deviation is 393.4. The stacked bar chart shows that the most recent year is slightly above average, yet only 7 years out of the last 17 have lower incidences. The years 2013 through 2019 have a low number of incidences compared to the rest of the dataset. So, if someone moved into the city during these years, it is safe to say that the city is less safe now. The most recent year's incidences of 1716 is less than when my coworker moved to the city, which was 1959 back in 2008.

```
ggplot(nypd_boroyear_aggr,aes(x = year,y = incidents,fill = boro)) +
  geom_col() + geom_hline(yintercept=stkd_mean +stkd_sd, linetype="dashed", color = "red") +
  geom_hline(yintercept=stkd_mean - stkd_sd, linetype="dashed", color = "red") +
  geom_hline(yintercept=stkd_mean, linetype="dashed", color = "blue") +
  ggtitle("Plot of shooting incident by year")
```



By unstacking the Boros, the mean has moved to 321.3 incidences with a standard deviation of 233.4 incidences. The chart indicates that 2 of the 5 Boros are consistently above the mean and add to most incidences. One interesting thing is that the trend lines of the stacked chart appear to exist in the unstacked chart for each Boro.

```
ggplot(nypd_boroyear_aggr,aes(x = year,y = incidents,fill = boro)) +
  geom_bar(position = "dodge",stat = "identity") +
  geom_hline(yintercept=sbs_mean + sbs_sd, linetype="dashed", color = "red") +
  geom_hline(yintercept=sbs_mean - sbs_sd, linetype="dashed", color = "red") +
  geom_hline(yintercept=sbs_mean, linetype="dashed", color = "blue") +
  ggtitle("Boro breakout of shooting incident by year")
```



Was my coworker correct in saying the city is less safe than in prior years? Since this is only shooting incidents, I can't say definitely, but outside of 2013 through 2019, I am leaning towards no. I moved out of the city well before this period. To further remove any bias, I didn't state what number of incidents would even be considered safe. New York City isn't even a city where my coworker or I have lived.

Additional Questions

- Another question that is still outstanding and the dataset can help answer is the aspect of time concerning safety.
- A question that we don't have enough data in the dataset is why there was a jump back up after the start of COVID.
- If the future data stays consistent with the prior 3 years and similar to the first 7 years, then what caused the low incidents in the city from 2013 through 2019.

```
rm(stkd_mean)
rm(stkd_sd)
rm(sbs_mean)
rm(sbs_sd)
rm(nypd_boroyear_aggr)
rm(nypd_data)
rm(url_in)
```

Built with 4.3.2