NYPD Shooting Incident Report

J.R.

2023-06-21

Introduction

This report summarizes shooting incident data from 2006 to the end of the previous year. The analysis is intended to elucidate trends in the age, location, and race of shooting victims. The analysis also takes an initial look at shooting incident trends over time, and considers further work that would provide additional insight into shooting incidents in the city of New York.

Setup the analysis

```
##Import libraries
library(tidyverse)
#library(tidyr)
#library(lubridate)
#library(ggplot2)
#library(readr)
#library(dplyr)

##This is the URL of the NYPD Shooting Death Data
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
```

Read in the data

```
NYPD_inc <- read_csv(url)
```

Tidy up the data

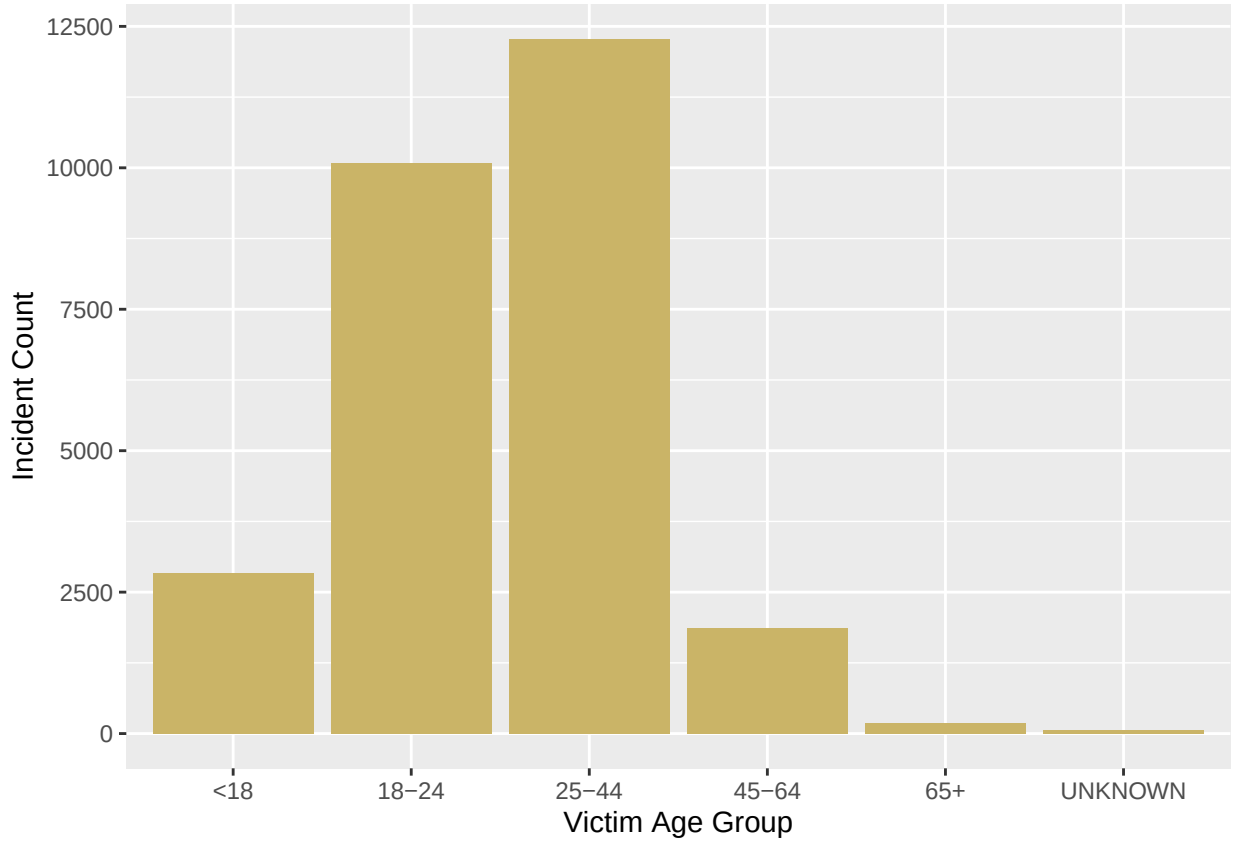
```
#Convert the date to type date instead of char
NYPD_inc$OCCUR_DATE <- lubridate::mdy(NYPD_inc$OCCUR_DATE)

#Convert other groups to categorical
NYPD_inc$VIC_AGE_GROUP <- as.factor(NYPD_inc$VIC_AGE_GROUP)
NYPD_inc$VIC_RACE <- as.factor(NYPD_inc$VIC_RACE)
NYPD_inc$VIC_SEX <- as.factor(NYPD_inc$VIC_SEX)
NYPD_inc$BORO <- as.factor(NYPD_inc$BORO)
```

```
#Get just the desired columns for this analysis
NYPD_inc <- subset(NYPD_inc, select = c(OCCUR_DATE,BORO,
                                        STATISTICAL_MURDER_FLAG,
                                        VIC_AGE_GROUP,VIC_SEX,
                                        VIC_RACE))
```

Plot the analyses

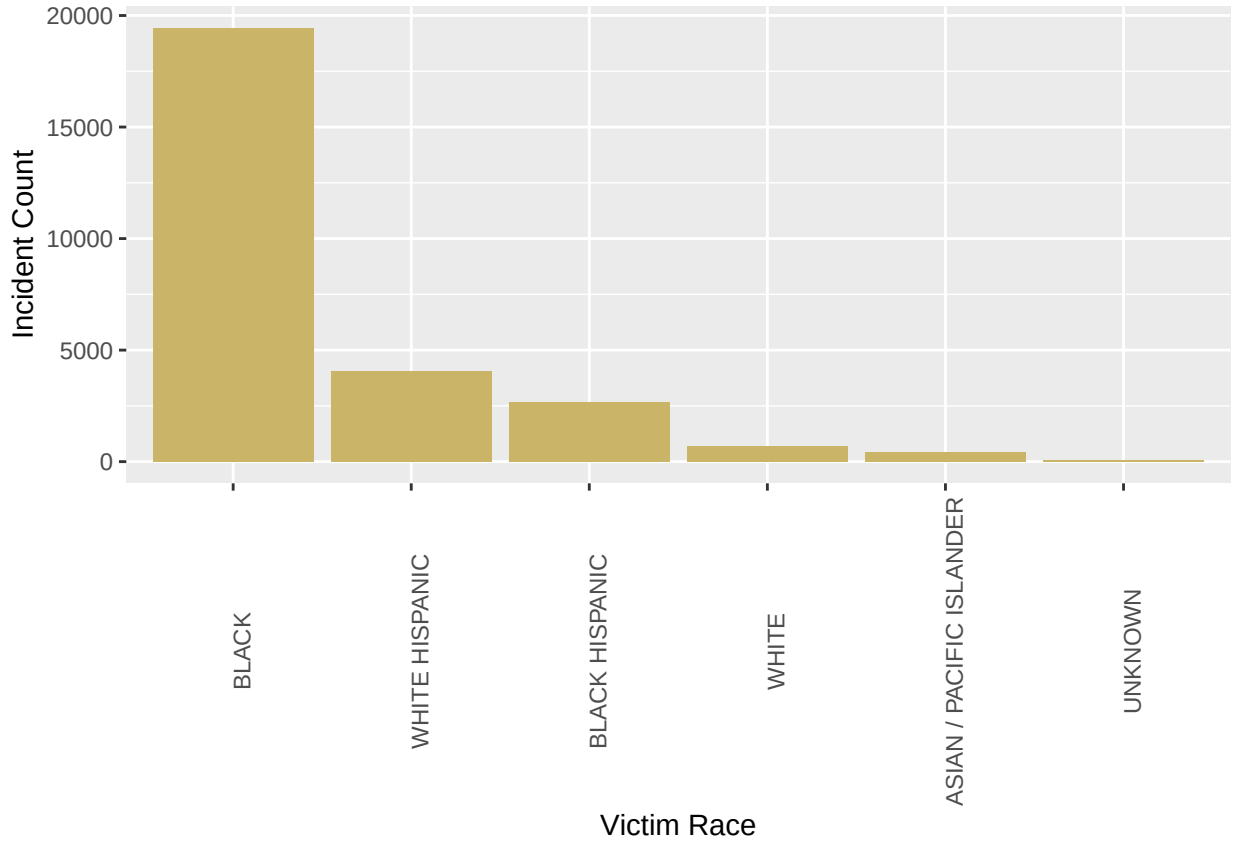
Shooting incidents by victim age group

```
NYPD_inc %>%
  group_by(VIC_AGE_GROUP) %>%
  mutate(group_num = n()) %>%
  dplyr::filter(group_num>1) %>%
  ggplot(aes(x=VIC_AGE_GROUP)) +
  geom_bar(fill = "#CAB467") +
  labs(x='Victim Age Group') +
  labs(y='Incident Count')
```



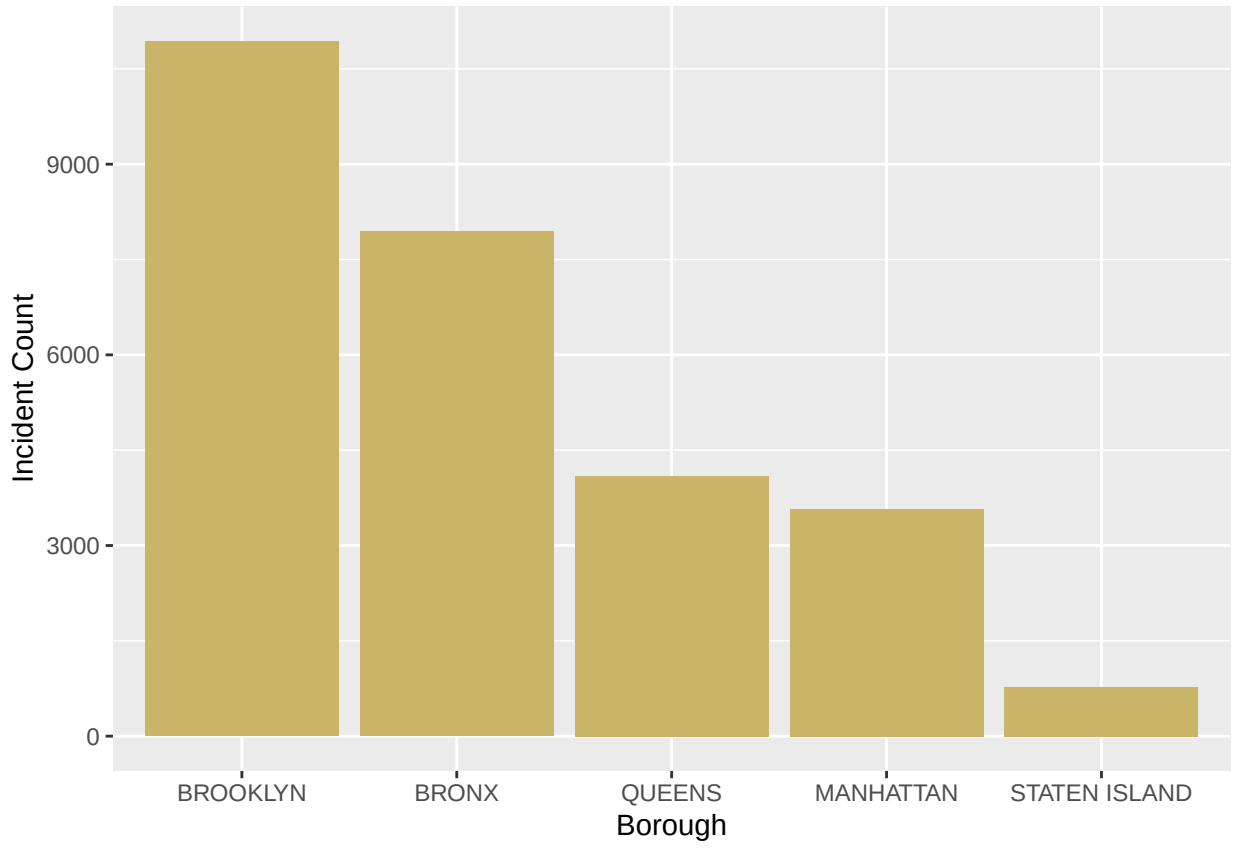
Shooting incidents by victim race

```
NYPD_inc %>%
  group_by(VIC_RACE) %>%
  mutate(group_num = n()) %>%
  dplyr::filter(group_num>50) %>%
  ggplot(aes(x=reorder(VIC_RACE,VIC_RACE,function(x)-length(x)))) +
  geom_bar(fill = "#CAB467") +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(x='Victim Race') +
  labs(y='Incident Count')
```



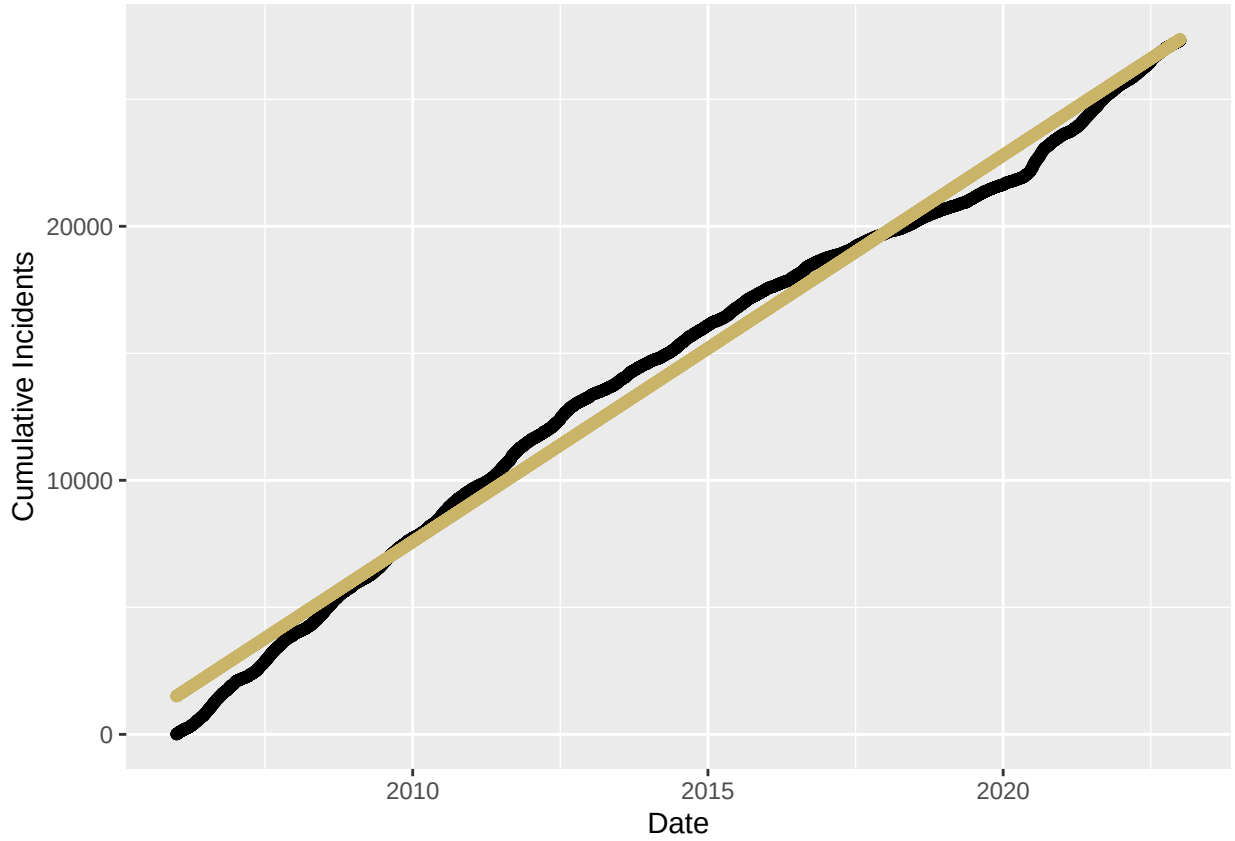
Shooting incidents by borough

```
NYPD_inc %>%
  group_by(BORO) %>%
  mutate(group_num = n()) %>%
  dplyr::filter(group_num>5) %>%
  ggplot(aes(x=reorder(BORO,BORO,function(x)-length(x))), fill = x) +
  geom_bar(fill = "#CAB467") +
  labs(x='Borough') +
  labs(y='Incident Count')
```



Cumulative shooting incidents by date

```
NYPD_inc <- arrange(NYPD_inc, OCCUR_DATE)
cum_inc <- NYPD_inc %>%
  group_by(OCCUR_DATE) %>%
  summarize(count = n()) %>%
  mutate(count = count, cum_count = cumsum(count))

#make a linear model
mod <- lm(cum_count ~ OCCUR_DATE, data = cum_inc)

cum_inc_w_pred <- cum_inc %>% mutate(pred = predict(mod, newdata=cum_inc))

cum_inc_w_pred %>% ggplot() +
  geom_point(aes(x = OCCUR_DATE, y = cum_count), color = "black") +
  geom_point(aes(x = OCCUR_DATE, y = pred), color = "#CAB467") +
  labs(x='Date') +
  labs(y='Cumulative Incidents')
```



Conclusion

The majority of shooting incidents reported by the NYPD since 2006 have been in the 25-44 year old age group. Black victims represent the largest share of incidents, and the borough of Brooklyn had the most reported shooting incidents. The cumulative incidents have increased in approximately linear fashion over this period of time. However, leading up to 2020 there was a decrease in the shooting incident rate relative to the long term average, and there was a rapid increase in the rate just after 2020. It would be interesting to look into the cause of the slope discontinuity around this time and see if it was related to the onset of COVID-19.It would also be interesting to plot the local derivative of the cumulative incidents vs. time in order to observe incident rate versus time and season of year.

There are possible sources of bias in the data. The data include certain demographic information such as race, that is a police officer's interpretation of a person's skin color. In particular, the perpetrator's race may be subject to bias in how the data were collected or measured. Since the perpetrator may never have been caught, or the wrong perpetrator could have been caught, this data may be biased against certain races based on the collection method. For this reason, this analysis does not consider the perpetrator data. The time reports for the shooting are described as "the exact time of the shooting incident", but the exact time would almost be impossible to know. In certain boroughs, police response time may be faster or slower, leading to biasing of time data.

I personally chose to analyze the incident count by race, borough, and age because I suspected there would be interesting observations that identify racial inequities. To mitigate this potential bias, I could plot the data in a variety of ways beyond what is presented here. For example, rather than plotting absolute incident count, it would be interesting to look at incident rate normalized by population, or by population of a certain race within a borough. This would help mitigate potential bias and highlight whether and to what extent there is a disproportionate number of victims of any race.