

Analysis NYPD Shooting Incidents

2022-09-02

Introduction

This report is made from New York City Shooting Incidents (Historical) dataset. At this time, the dataset includes data from January 1st, 2006 to December 31st, 2021. There is information about date, time, location, perpetrator and victim. There were two questions that I wanted to check: 1. What is the overall trend over time 2. What are the profiles of the victims regarding age group and sex

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

Import data from NYC shooting episodes

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
shooting_episodes <- read_csv(url_in, show_col_types = FALSE)
```

Data summary after import

```
summary(shooting_episodes)
```

```

## INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
## Min.      : 9953245      Length:25596      Length:25596      Length:25596
## 1st Qu.: 61593633      Class :character      Class1:hms      Class :character
## Median : 86437258      Mode  :character      Class2:difftime      Mode  :character
## Mean    :112382648      Mode  :numeric
## 3rd Qu.:166660833
## Max.    :238490103
##
## PRECINCT      JURISDICTION_CODE      LOCATION_DESC      STATISTICAL_MURDER_FLAG
## Min.      : 1.00      Min.      :0.0000      Length:25596      Mode :logical
## 1st Qu.: 44.00      1st Qu.:0.0000      Class :character      FALSE:20668
## Median : 69.00      Median :0.0000      Mode  :character      TRUE :4928
## Mean    : 65.87      Mean    :0.3316
## 3rd Qu.: 81.00      3rd Qu.:0.0000
## Max.    :123.00      Max.    :2.0000
## NA's    :2
## PERP_AGE_GROUP      PERP_SEX      PERP_RACE      VIC_AGE_GROUP
## Length:25596      Length:25596      Length:25596      Length:25596
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
## VIC_SEX      VIC_RACE      X_COORD_CD      Y_COORD_CD
## Length:25596      Length:25596      Min.      : 914928      Min.      :125757
## Class :character      Class :character      1st Qu.:1000011      1st Qu.:182782
## Mode  :character      Mode  :character      Median :1007715      Median :194038
## Mean    :1009455      Mean    :207894
## 3rd Qu.:1016838      3rd Qu.:239429
## Max.    :1066815      Max.    :271128
##
## Latitude      Longitude      Lon_Lat
## Min.      :40.51      Min.      : -74.25      Length:25596
## 1st Qu.:40.67      1st Qu.: -73.94      Class :character
## Median :40.70      Median : -73.92      Mode  :character
## Mean    :40.74      Mean    : -73.91
## 3rd Qu.:40.82      3rd Qu.: -73.88
## Max.    :40.91      Max.    : -73.70
##
##

```

Convert dates to Date format using Lubridate and remove some columns

```

shooting_episodes <- shooting_episodes %>% mutate(OCCUR_DATE=mdy(OCCUR_DATE))
shooting_episodes <- shooting_episodes %>% select(-c(INCIDENT_KEY,JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD))
summary(shooting_episodes)

```

```

## OCCUR_DATE      OCCUR_TIME      BORO      PRECINCT
## Min.      :2006-01-01      Length:25596      Length:25596      Min.      : 1.00
## 1st Qu.:2009-05-10      Class1:hms      Class :character      1st Qu.: 44.00
## Median :2012-08-26      Class2:difftime      Mode  :character      Median : 69.00
## Mean    :2013-06-13      Mode  :numeric      Mean    : 65.87

```

```
## 3rd Qu.:2017-07-01                      3rd Qu.: 81.00
## Max.      :2021-12-31                    Max.      :123.00
## LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
## Length:25596        Mode :logical          Length:25596
## Class :character    FALSE:20668            Class :character
## Mode :character     TRUE :4928              Mode :character
##
##
##
## PERP_SEX            PERP_RACE            VIC_AGE_GROUP      VIC_SEX
## Length:25596        Length:25596        Length:25596      Length:25596
## Class :character    Class :character    Class :character  Class :character
## Mode :character     Mode :character     Mode :character   Mode :character
##
##
##
## VIC_RACE
## Length:25596
## Class :character
## Mode :character
##
##
##
```

Visualize Shooting episodes per days and trends

```
shooting_per_date <- shooting_episodes %>%
  count(OCCUR_DATE)
most_bloody <- shooting_per_date[which.max(shooting_per_date$n),]
paste("Date with the highest number of shooting incidents: ", most_bloody$OCCUR_DATE)
```

```
## [1] "Date with the highest number of shooting incidents: 2020-07-05"
```

```
paste("Maximun number shooting per day: ", max(shooting_per_date$n))
```

```
## [1] "Maximun number shooting per day: 47"
```

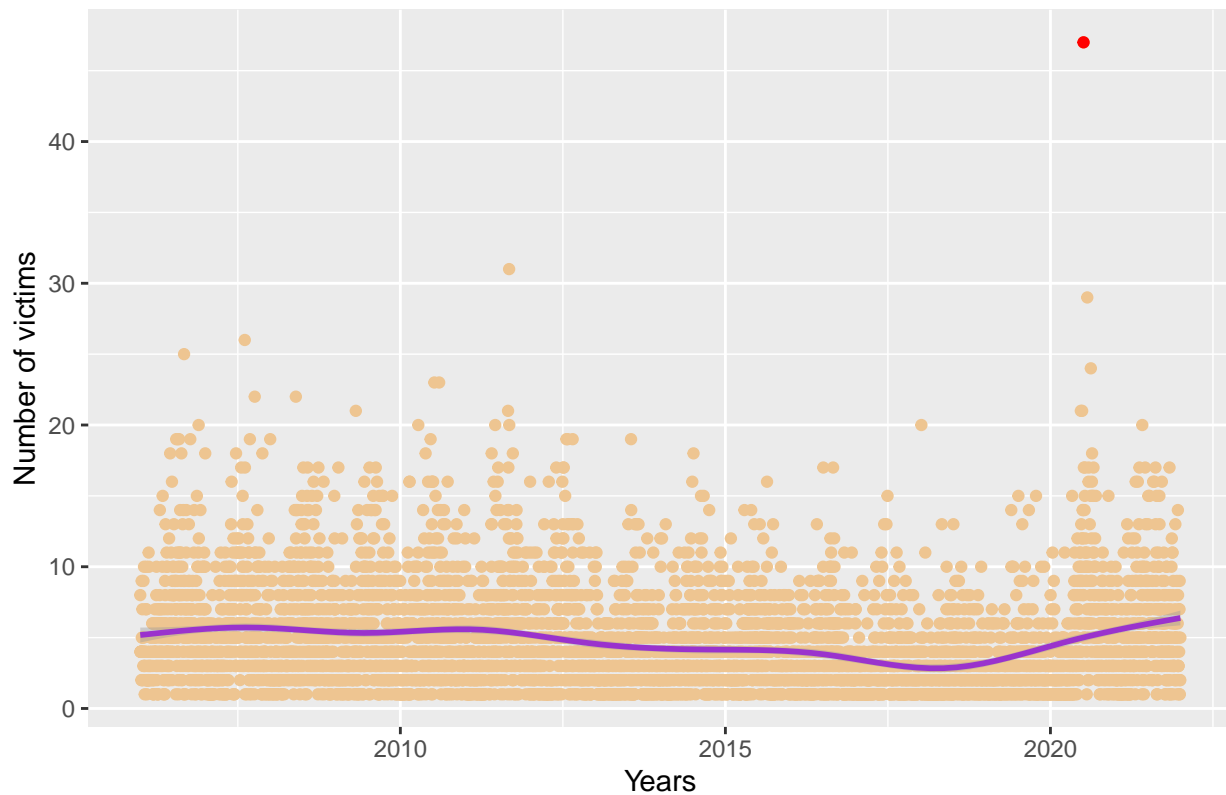
```
paste("Average number of shootings per day: ", mean(shooting_per_date$n))
```

```
## [1] "Average number of shootings per day: 4.73211314475874"
```

```
plot1 <- ggplot(shooting_per_date, mapping = aes(x = OCCUR_DATE)) +
  geom_point(mapping = aes(y = n), color = "burlywood2") +
  geom_point(most_bloody, mapping = aes(y = n), color = "red") +
  geom_smooth(mapping = aes(y = n), color = "darkorchid3") +
  labs(title = "Number of Shooting Victims per day", x = "Years", y = "Number of victims")
plot1
```

```
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

Number of Shooting Victims per day



Some questions can be made:

- From the trend-line: why are shooting episodes increasing in number the last years?
- What were the reasons from the decline in the the 2010's
- July 5th, 2020 was a day with many shooting episodes. What happened that date?

Some possible explanations could be:

- Consequences of sanitary measures against COVID 19 pandemic and acceptance in different sectors in society
- Changes in policies towards crime (policy, local anti-crime organizations, funding)
- As the pandemic receded one could see a surge in crime in general after months of restrictions and lockdowns
- Some extra info after the weekend of July 4th, 2021: <https://www.bbc.com/news/world-us-canada-57743694>

Visualize Shooting victims sex and age group

```
shooting_victim <- shooting_episodes %>%
  count(VIC_AGE_GROUP, VIC_SEX)
colnames(shooting_victim) <- c("Age", "Sex", "Episodes")
shooting_victim$Sex[shooting_victim$Sex == "F"] <- "Female"
shooting_victim$Sex[shooting_victim$Sex == "M"] <- "Male"
```

```

shooting_victim$Sex[shooting_victim$Sex == "U"] <- "Unknown"

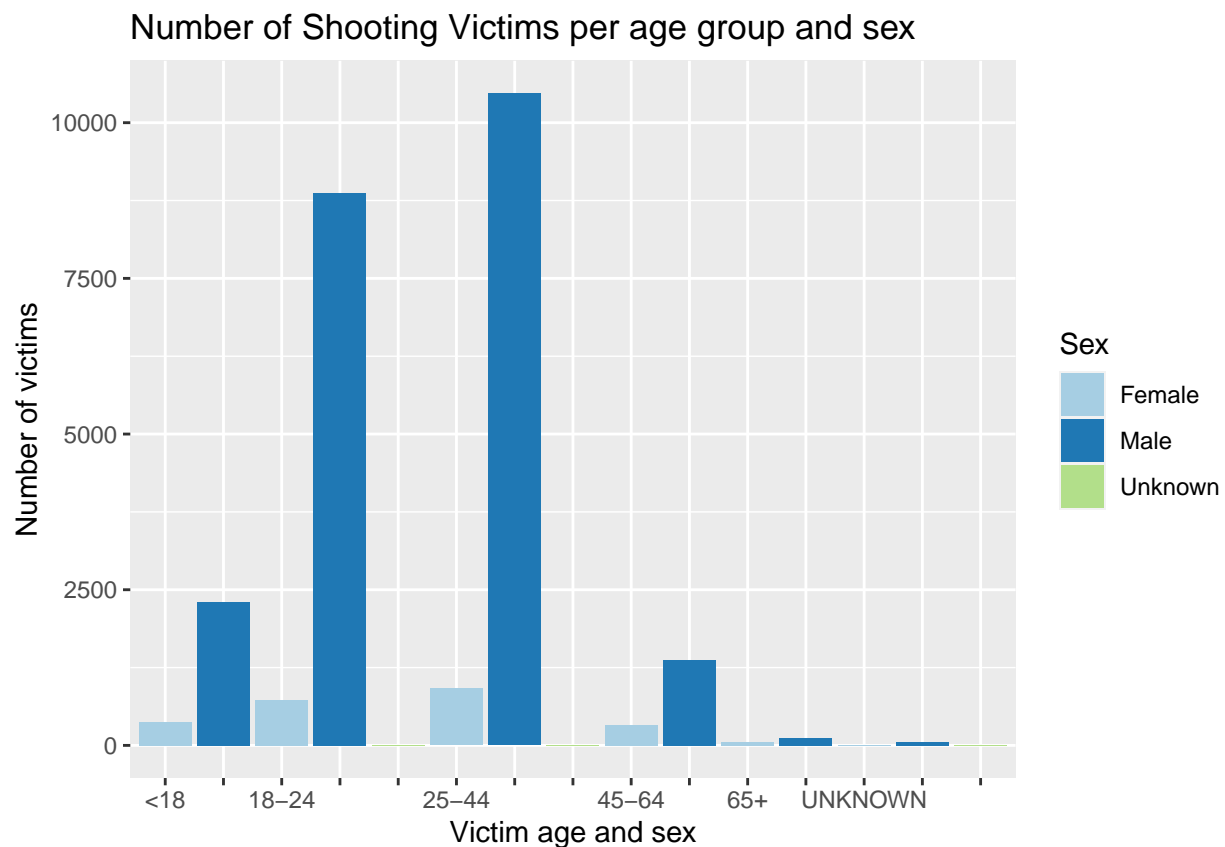
shooting_victim$Labels <- " "

shooting_victim$Labels <- ifelse(!(duplicated(shooting_victim$Age)), shooting_victim$Age, shooting_victim$Labels)

xlabels <- shooting_victim$Labels

plot2 <- ggplot(shooting_victim, mapping = aes(x = interaction(Sex, Age), y = Episodes, fill = Sex)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  scale_x_discrete("Victim age and sex", labels = xlabels) +
  scale_fill_brewer(palette = "Paired") +
  labs(title = "Number of Shooting Victims per age group and sex", y = "Number of victims", col = "Victim sex") +
  scale_color_manual(labels = c("Female", "Male", "Unknown"))
plot2

```



Bias and Conclusion

Bias

There are several possible bias that one could list:

1. On the data itself: when I analyzed the data regarding victim sex, it is remarkable to find “U” only in certain age groups. What does that mean and what does U represent? I called it “Unknown” in my graph labels

2. Omission: is there other variables that has been omitted and could have been more helpful to understand this problem
3. Personal: I was expecting that male victims were over represented in the statistics. Both because I have the impression that men have bigger chances to get into violent episodes, been outside more, etc.

Conclusion

Trend: After a decline from the first half of the 2010's, a significant increase of shooting episodes can be seen the last couple of years

Victims: Men are definitely overrepresented in the victim statistics and the most common victim is between 25-44 years old