

NYPD Shooting Data

T. Pacheco

2022-07-25

R Markdown

Load Data

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --v ggplot2 3.3.6
## 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     vforcats  0.5.1

## Warning: package 'tidyverse' was built under R version 3.6.2
## Warning: package 'readr' was built under R version 3.6.2
## Warning: package 'purrr' was built under R version 3.6.2
## Warning: package 'forcats' was built under R version 3.6.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

NYPD_shooting <- read_csv(url_in)

## Rows: 25596 Columns: 19-- Column specification -----
## Delimiter: ","
## chr (10): OCCUR_DATE, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_SEX, PERP_R...
## 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.

NYPD_shooting <- subset(NYPD_shooting, select = -c(LOCATION_DESC, PERP_AGE_GROUP:PERP_RACE))

library(dplyr)
library(lubridate)
```

```

## Warning: package 'lubridate' was built under R version 3.6.2

##
## Attaching package: 'lubridate'

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

NYPD_shooting <- mutate(NYPD_shooting, OCCUR_DATE = mdy(OCCUR_DATE))

grouped_boro <- NYPD_shooting %>%
  select(OCCUR_DATE, BORO) %>%
  arrange(BORO)

```

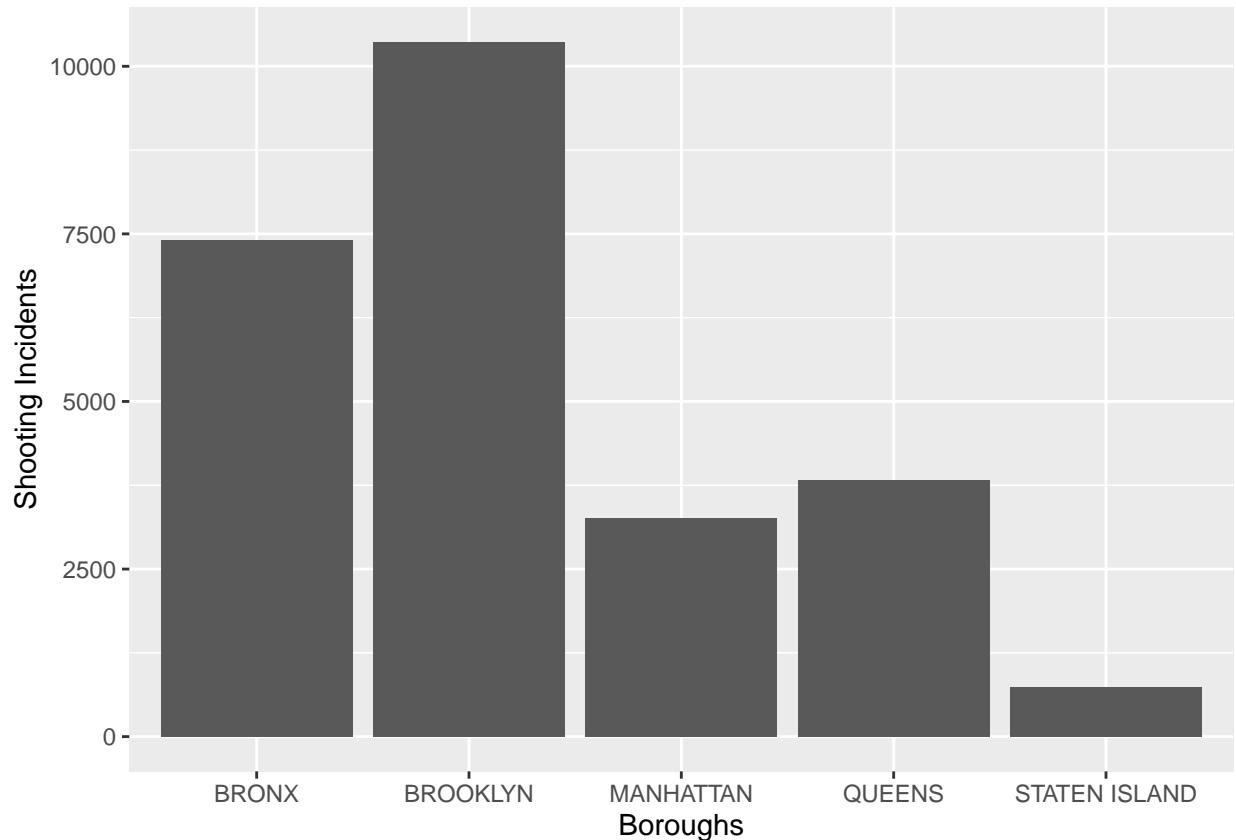
Amount of shootings by Borough and shooting victims by gender

```

library(ggplot2)

ggplot(NYPD_shooting, aes(x = BORO)) + geom_bar() + xlab("Boroughs") + ylab("Shooting Incidents")

```

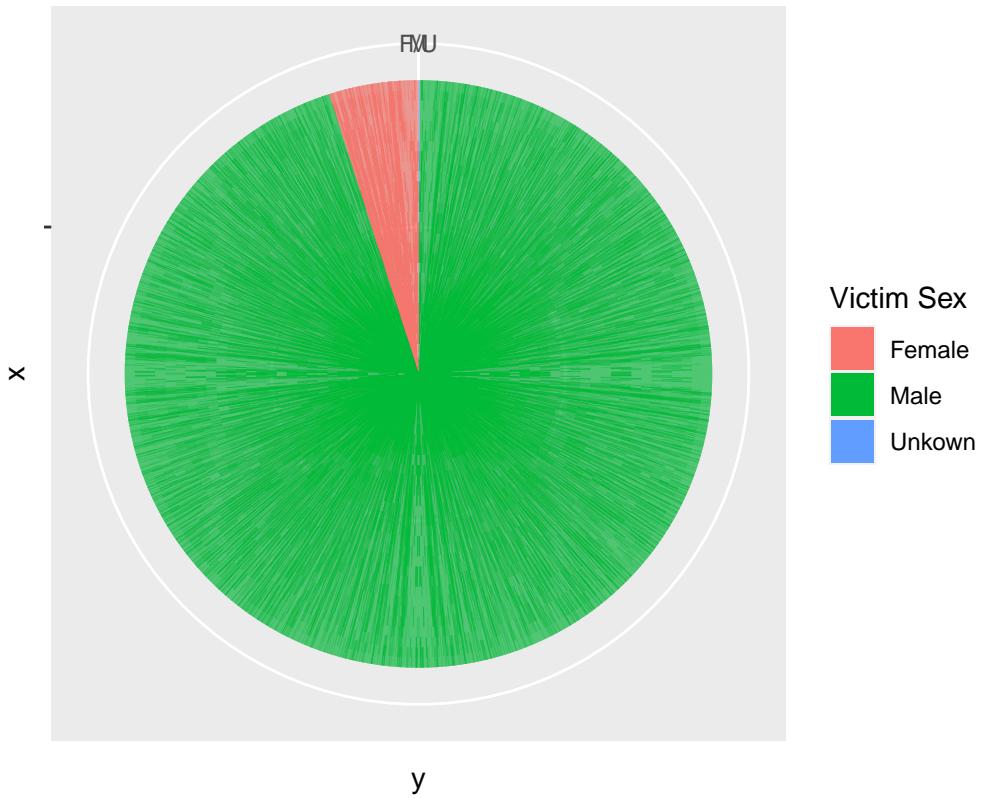


```

ggplot(NYPD_shooting, aes(x = "", y = VIC_SEX, fill = VIC_SEX)) + geom_col() + ylab("y") + coord_polar()

```

Incidents by Victim Sex



Amount of shootings per day reported in New York

The amount and percentage of shootings by borough

```
library(scales)
```

```
##  
## Attaching package: 'scales'  
  
## The following object is masked from 'package:purrr':  
##  
##     discard  
  
## The following object is masked from 'package:readr':  
##  
##     col_factor  
  
shootings_per_day = (length(NYPD_shooting$INCIDENT_KEY))/(length(unique(NYPD_shooting$OCCUR_DATE)))  
  
cat("Shootings per day =",shootings_per_day)  
  
## Shootings per day = 4.732113
```

```

total_shootings = length(NYPD_shooting$BORO)

incidents_by_boro <- grouped_boro %>%
  count(BORO) %>%
  rename(Shootings = n)

shooting_rate = (incidents_by_boro$Shootings/total_shootings)

percent_boro <- data.frame(incidents_by_boro, percent(shooting_rate))

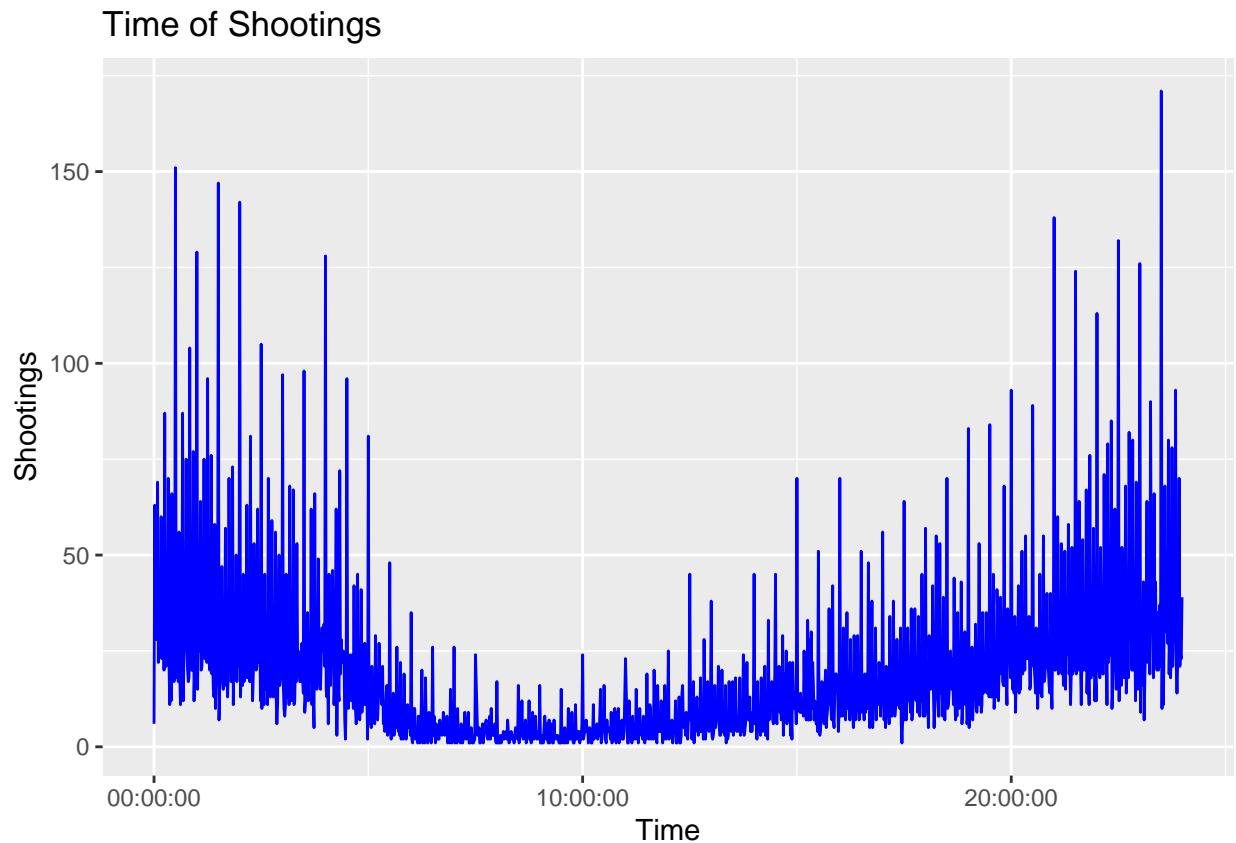
percent_boro

##           BORO Shootings percent.shooting_rate.
## 1      BRONX     7402        28.9%
## 2    BROOKLYN    10365        40.5%
## 3  MANHATTAN     3265        12.8%
## 4    QUEENS     3828        15.0%
## 5  STATEN ISLAND     736         2.9%

incident_time = NYPD_shooting %>%
  group_by(OCCUR_TIME) %>%
  count()

ggplot(incident_time, aes(x = OCCUR_TIME, y = n)) + geom_line(color = "blue") + labs(title = "Time of Shootings")

```



As shown in the “Time of Shootings” Plot, the bulk of the shootings throughout New York take place when the sun is down.

Based on the percentage of shootings the safest borough by gun violence is Staten Island.

Possible bias may come from the reporting of the shootings, the human element in data entry.