

NYPD Shooting Data

2024-04-30

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- 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(lubridate)
```

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

```
shootings <- read_csv(NYPD_data)
```

```
## Rows: 28562 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.
```

```
shootings$OCCUR_DATE <- mdy(shootings$OCCUR_DATE)
```

```
shootings <- shootings %>%
  mutate(across(c(BORO, PERP_AGE_GROUP, PERP_SEX, PERP_RACE, VIC_AGE_GROUP, VIC_SEX, VIC_RACE, STATISTICAL_MURDER_FLAG), as.factor))
```

```
plot_victim_age <- shootings %>%
  filter(STATISTICAL_MURDER_FLAG == FALSE) %>%
  ggplot(aes(x = VIC_AGE_GROUP)) +
  geom_bar(fill = "red") +
```

```

theme_bw() +
labs(x = "Victim Age Group",
     y = "Non-Fatal Shooting Incidents",
     title = "Non-Fatal Shooting Incidents by Victim Age Group")

plot_borough <- shootings %>%
  ggplot(aes(x = BORO, fill = STATISTICAL_MURDER_FLAG)) +
  geom_bar() +
  theme_bw() +
  labs(x = "Borough",
       y = "Number of Shootings",
       fill = "Statistical Murder Flag",
       title = "Shootings by Borough")

plot_perpetrator_age <- shootings %>%
  ggplot(aes(x = PERP_AGE_GROUP)) +
  geom_bar(fill = "blue") +
  theme_bw() +
  labs(x = "Perpetrator Age Group",
       y = "Number of Shootings",
       title = "Shootings by Perpetrator Age Group")

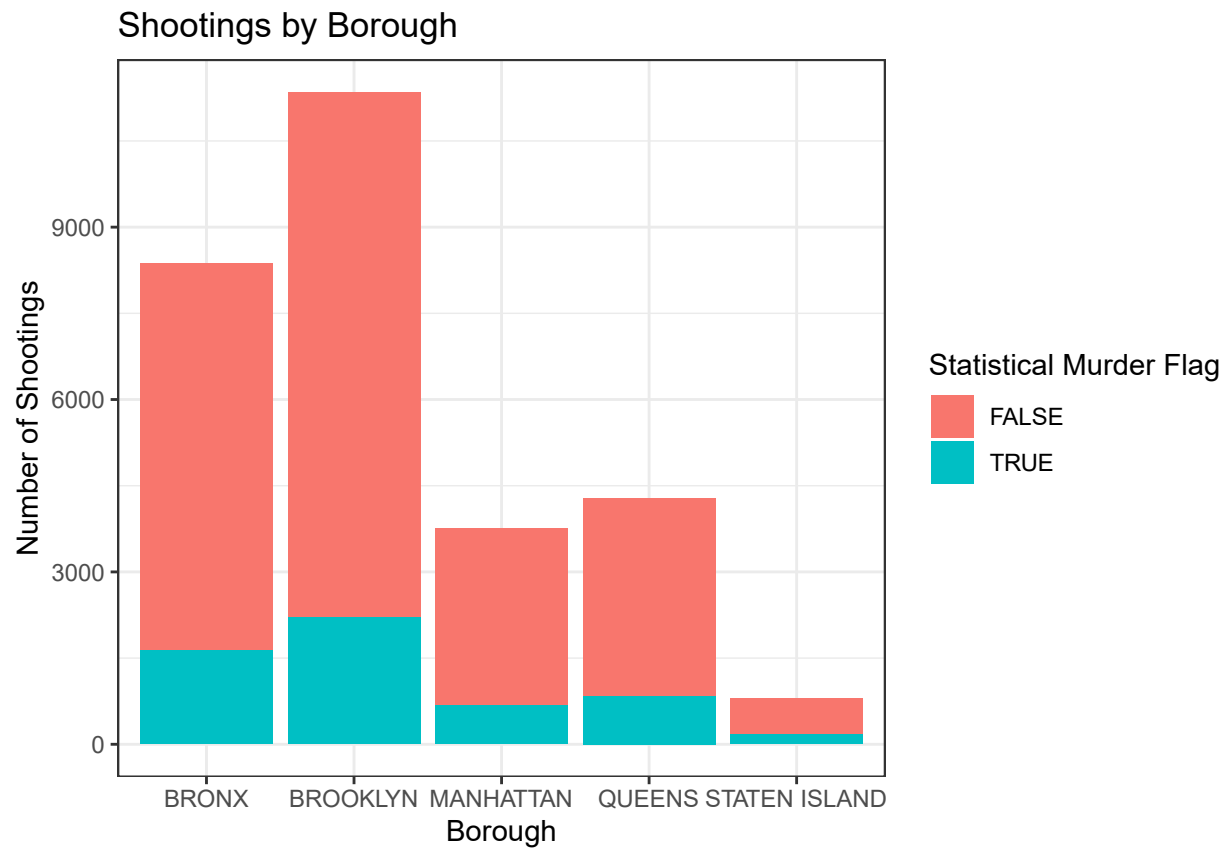
plot_victim_race <- shootings %>%
  ggplot(aes(x = VIC_RACE)) +
  geom_bar(fill = "green") +
  theme_bw() +
  labs(x = "Victim Race",
       y = "Number of Shootings",
       title = "Shootings by Victim Race")

print(plot_victim_age)

```

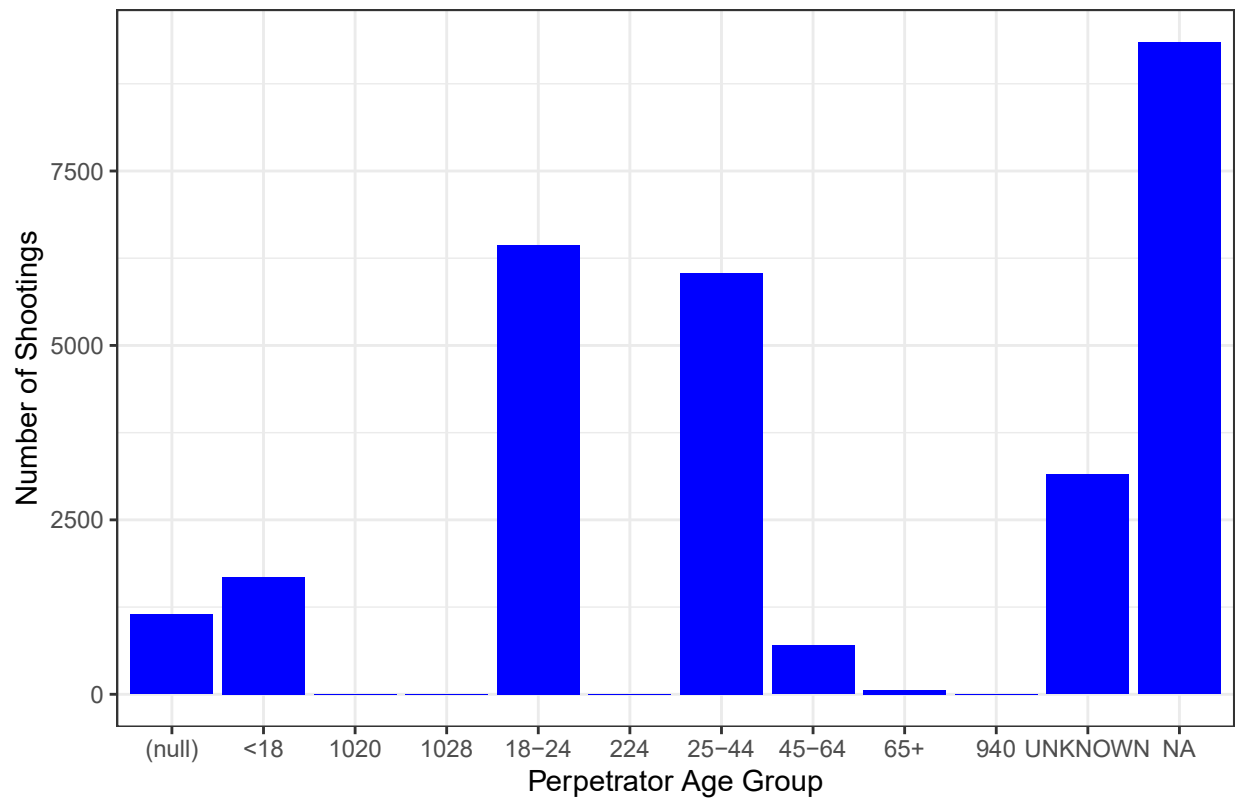


```
print(plot_borough)
```

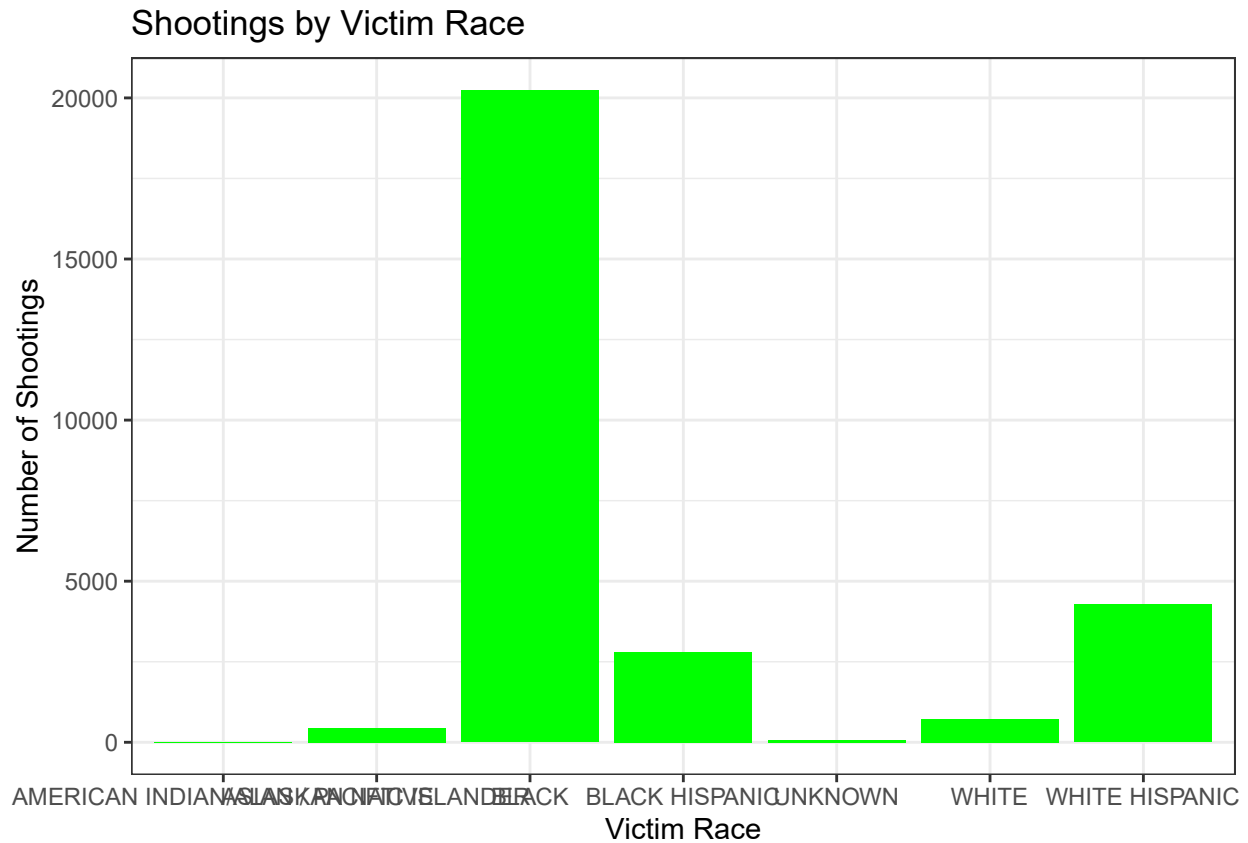


```
print(plot_perpetrator_age)
```

Shootings by Perpetrator Age Group



```
print(plot_victim_race)
```



In Conclusion we can see that certain age groups make up a large portion of the victims. In addition Brooklyn had the highest fatalaties. I think my biases are in regards to age, race and gender, which is why I chose to investigate those categories to see if there might be any links.