New York Shootings

James Barry

5/22/2021

R Markdown

The following project will go into analysis surrounding shooting incidents in New York City. The data is collected from https://catalog.data.gov/dataset.

The below code reads in the NYPD shooting data in a reproducible way. The OCCUR_DATE variable is changed to a date variable. Columns I will not be using and removed include Longitude,Latitude,Lon_Lat,Y_COORD_CD, and X_COORD_CD. Three bar graphs are created using ggplot that are helpful analyzing the categorical variables. Analysis focuses on the victims since the data is not complete for the perpetrators. The missing data is not an issue for the purpose of this analysis, but I would remove blank rows if looking at the perpetrator. It is also most likely more accurate data since the victims of shootings are more easily identifiable. The summary shows the data included.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'readr' was built under R version 3.6.3
```

```
## Warning: package 'purrr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

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

```r
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.6.3
```

```
##
## Attaching package: 'lubridate'
```

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

```r
library(dplyr)
url_in <- "https://catalog.data.gov/dataset"
file_names <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
urls <- str_c(file_names)
NYPD_Shootings <- read_csv(urls[1])
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   INCIDENT_KEY = col_double(),
##   OCCUR_DATE = col_character(),
##   OCCUR_TIME = col_time(format = ""),
##   BORO = col_character(),
##   PRECINCT = col_double(),
##   JURISDICTION_CODE = col_double(),
##   LOCATION_DESC = col_character(),
##   STATISTICAL_MURDER_FLAG = col_logical(),
##   PERP_AGE_GROUP = col_character(),
##   PERP_SEX = col_character(),
##   PERP_RACE = col_character(),
##   VIC_AGE_GROUP = col_character(),
##   VIC_SEX = col_character(),
##   VIC_RACE = col_character(),
##   X_COORD_CD = col_number(),
##   Y_COORD_CD = col_number(),
##   Latitude = col_double(),
##   Longitude = col_double(),
##   Lon_Lat = col_character()
## )
```

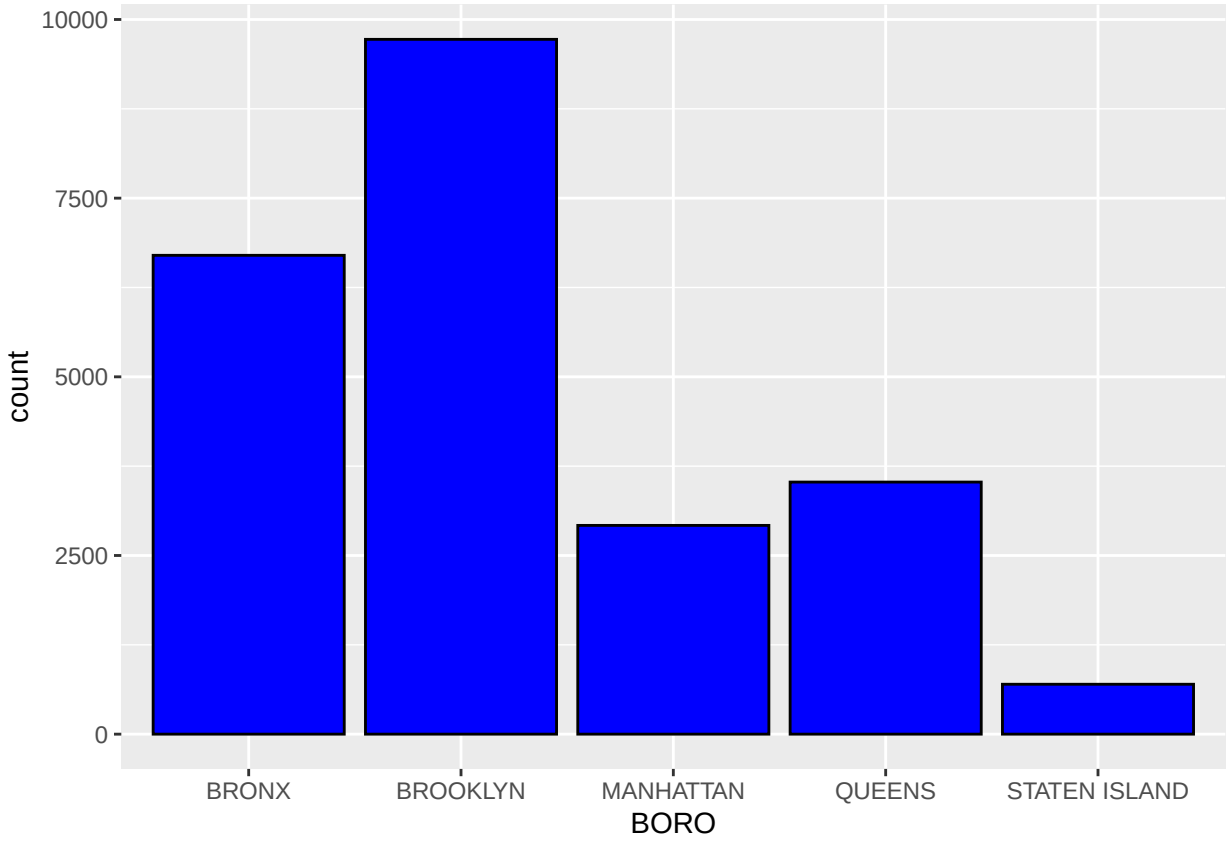
```r
NYPD_Shootings <- NYPD_Shootings %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE)) %>%
  select (-c(Longitude,Latitude,Lon_Lat,Y_COORD_CD,X_COORD_CD,))

summary(NYPD_Shootings)
```

```
##   INCIDENT_KEY         OCCUR_DATE          OCCUR_TIME           BORO
##  Min.   :  9953245   Min.   :2006-01-01   Length:23568      Length:23568
##  1st Qu.: 55317014   1st Qu.:2008-12-30   Class1:hms        Class :character
##  Median : 83365370   Median :2012-02-26   Class2:difftime   Mode  :character
```

```
##  Mean   :102218616   Mean   :2012-10-03   Mode  :numeric
##  3rd Qu.:150772442   3rd Qu.:2016-02-28
##  Max.   :222473262   Max.   :2020-12-31
##
##     PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   :  1.00   Min.   :0.0000    Length:23568       Mode :logical
##  1st Qu.: 44.00   1st Qu.:0.0000    Class :character   FALSE:19080
##  Median : 69.00   Median :0.0000    Mode  :character   TRUE :4488
##  Mean   : 66.21   Mean   :0.3323
##  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:23568       Length:23568       Length:23568       Length:23568
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_SEX            VIC_RACE
##  Length:23568       Length:23568
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

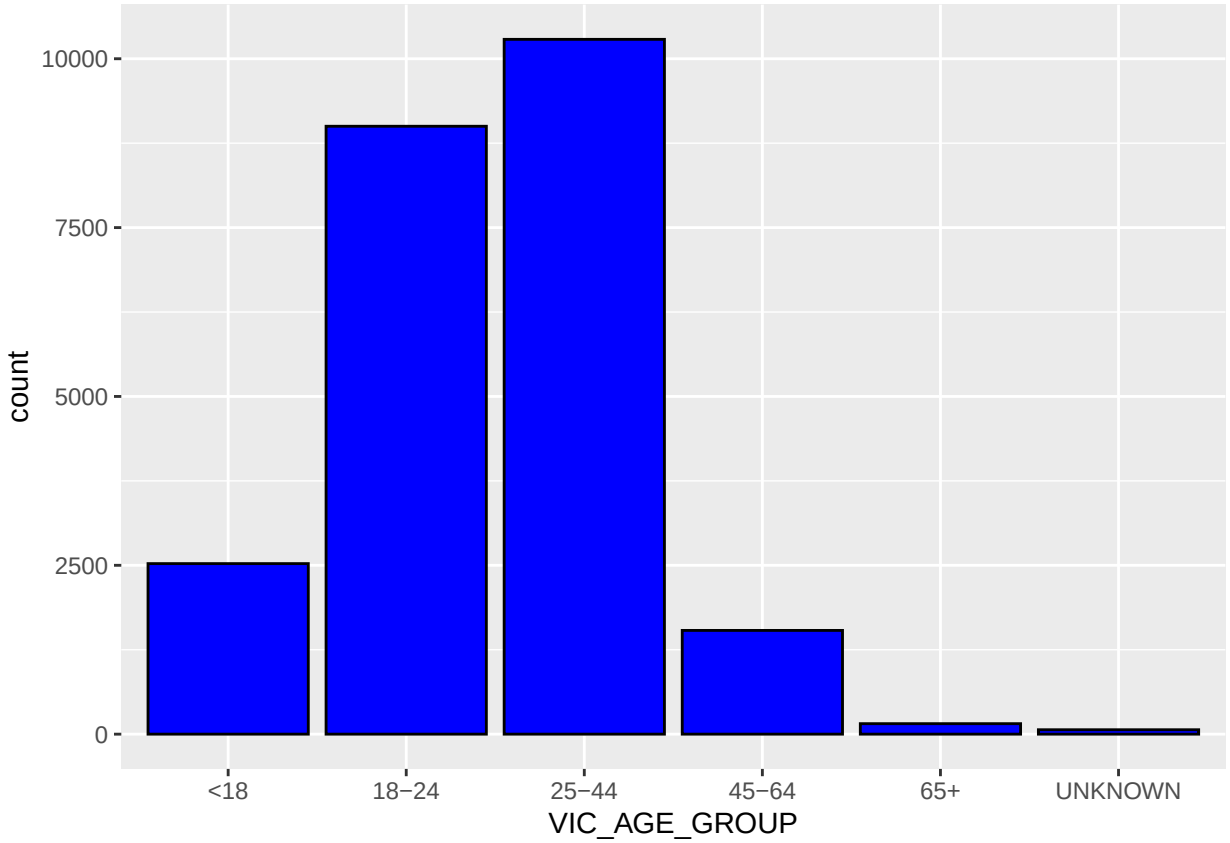
Graph 1 - Shootings By Boro

```
#Visuals
p<-ggplot(data=NYPD_Shootings, aes(x=BORO)) + geom_bar(fill ="blue", color ="black")
p
```



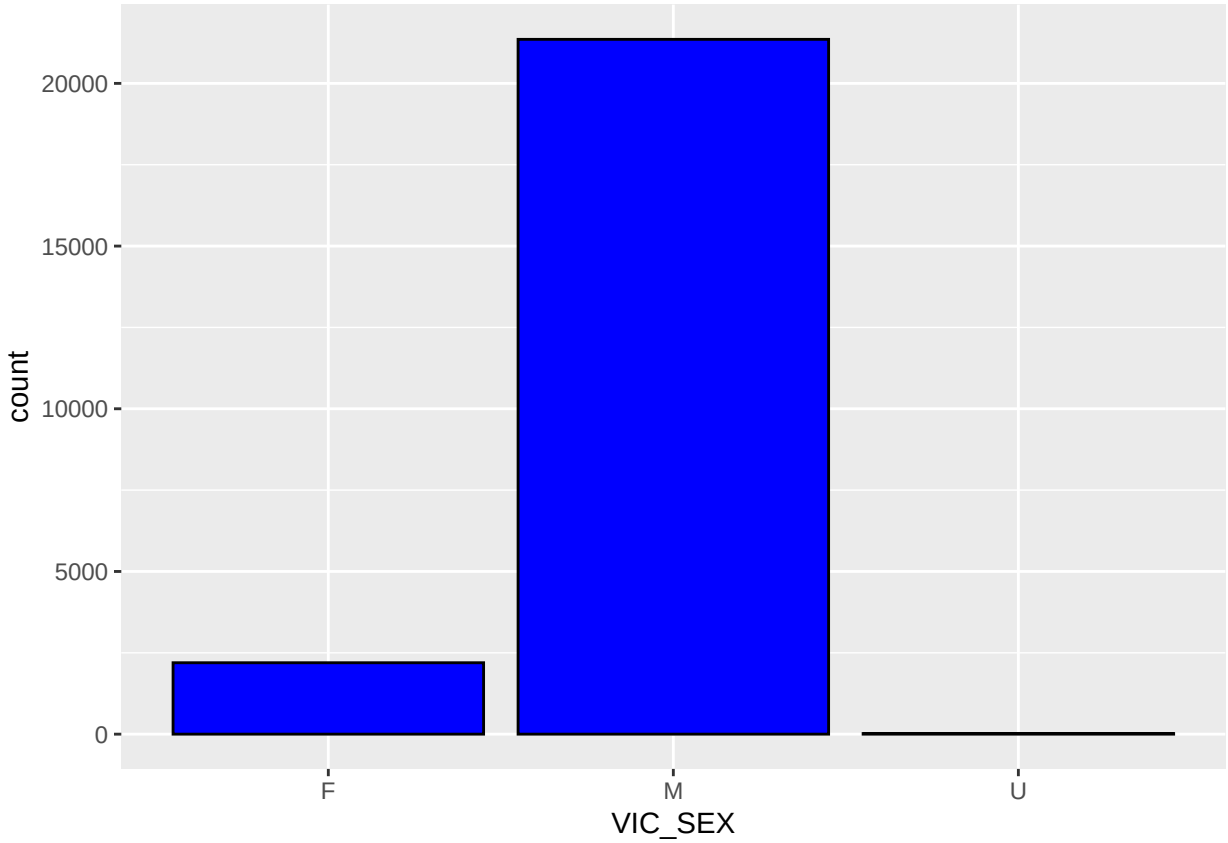
Graph 2 - Shooting Victims by Age

```
p<-ggplot(data=NYPD_Shootings, aes(x=VIC_AGE_GROUP)) + geom_bar(fill ="blue", color ="black")
p
```



Graph 3 - Shootings by Sex

```
p<-ggplot(data=NYPD_Shootings, aes(x=VIC_SEX)) + geom_bar(fill ="blue", color ="black")
p
```



Analysis Graph 1 shows the Brooklyn has the highest total number of shooting incidents. These numbers are *especially* interesting when looking at the population. Brooklyn has around 2.6 million people compared to Queens which only has a population of around 2.3 million. This shows that Brooklyn has signigantly higher shooting rate with more than double. Graph 2 shows the age group that is most likely to be a victim of a shooting is ages 18-44. 18-24 is especially high.Graph 3 shows that males are the victims of shootings far more than females. Since I only looked at the victim profile, it would be interesting to see if the perpertrator profile was shared any similarities. If it lined up with the victim profile or was different. Also looking further into the Brooklyn Boro would be interesting because the incident rate is so high.

Conclusion The overall takeaway of the above analysis is the likelihood of being a victim of a shooting in New York City. The most common is in Brooklyn for males between the age of 18-24. To avoid bias I graphed variables and then looked at the data. A source of bias could include only focusing on the victims. Assuming the shooter has the same profile is not necesarily the case and should be looked into further.