

# NY\_Shooting Data Analysis

11/8/2021

## R Markdown

This is an R Markdown document regarding New york shooting Incident.I have set my directory where .csv file is located and the i am reading csv file into NYPD\_shooting\_data. This data shows every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year. Each record in csv file represents a shooting incident in NYC and includes every information about that particular event for example its location, time of occurrence etc. In addition, information related to suspect and victim demographics is also included. This data is very helpful and can be used by the anyone to explore the nature of shooting/criminal activity and to study about these events.

Embed R code chunk is shown below:

Loading all the packages needed for this rmd

Importing Data Importing my .csv data file from the local directory path “C:/Users/Study/R\_studio”

Data Tidying

Summary of NY Shooting Data

```
## INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
## Min.   : 9953245    Length:23566    Length:23566    Length:23566
## 1st Qu.: 55322799   Class :character Class :character Class :character
## Median : 83365370   Mode  :character Mode  :character Mode  :character
## Mean   :102217631
## 3rd Qu.:150772442
## Max.   :222473262
## PRECINCT      LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
## Min.   : 1.00    Length:23566    Length:23566    Length:23566
## 1st Qu.: 44.00   Class :character Class :character Class :character
## Median : 69.00   Mode  :character Mode  :character Mode  :character
## Mean   : 66.21
## 3rd Qu.: 81.00
## Max.   :123.00
## PERP_SEX      PERP_RACE      VIC_AGE_GROUP      VIC_SEX
## Length:23566   Length:23566    Length:23566    Length:23566
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
## VIC_RACE
## Length:23566
## Class :character
## Mode  :character
##
##
##
```

To change column from character type to factor type.

## DATA ANALYSIS

Number of shootings group by neighborhood so we can see how many shootings happened depending on Boro or neighborhood.

```
NYPD_shooting_data_neighborhood <- NYPD_shooting_data %>%
  group_by(Neighborhood) %>%
  summarize(number_of_shootings=n()) %>%
  arrange(desc(number_of_shootings))
NYPD_shooting_data_neighborhood
```

```
## # A tibble: 5 x 2
##   Neighborhood number_of_shootings
##   <fct>          <int>
## 1 BROOKLYN      9722
## 2 BRONX        6700
## 3 QUEENS       3526
## 4 MANHATTAN    2920
## 5 STATEN ISLAND 698
```

Number of victims group by sex to check the rate of victims depending on sex.

```
NYPD_shooting_data_sex <- NYPD_shooting_data %>%
  group_by(VIC_SEX2) %>%
  summarize(number_of_shootings=n()) %>%
  arrange(desc(number_of_shootings))
NYPD_shooting_data_sex
```

```
## # A tibble: 3 x 2
##   VIC_SEX2 number_of_shootings
##   <fct>          <int>
## 1 M          21351
## 2 F          2195
## 3 U           20
```

Number of shootings group by Race to check victims depending on race.

```
NYPD_shooting_data_race <- NYPD_shooting_data %>%
  group_by(PERP_RACE2) %>%
  summarize(number_of_shootings=n()) %>%
  arrange(desc(number_of_shootings))
NYPD_shooting_data_race
```

```
## # A tibble: 8 x 2
##   PERP_RACE2          number_of_shootings
##   <fct>          <int>
## 1 "BLACK"      9854
## 2 ""          8425
## 3 "WHITE HISPANIC" 1960
## 4 "UNKNOWN"    1869
## 5 "BLACK HISPANIC" 1081
```

```
## 6 "WHITE" 255
## 7 "ASIAN / PACIFIC ISLANDER" 120
## 8 "AMERICAN INDIAN/ALASKAN NATIVE" 2
```

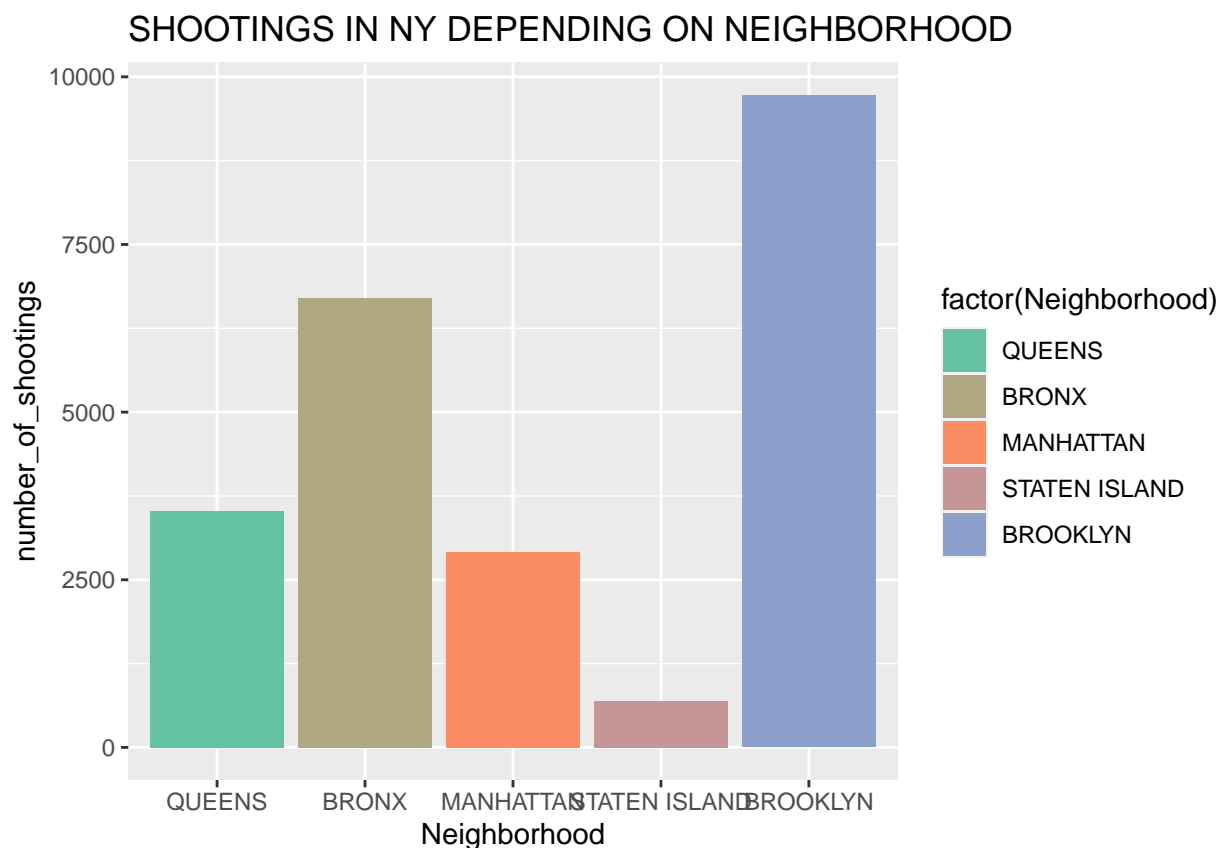
## DATA VISUALIZATON

```
#lets plot this to visualize it better. a barchart might be useful.
#need to add more colors to color palette so we can see all year totals in each bar

#define the number of colors you want
nbcols <- 15
mycolors <- colorRampPalette(brewer.pal(8, "Set2"))(nbcols)
```

Number of shootings depending on neighborhood

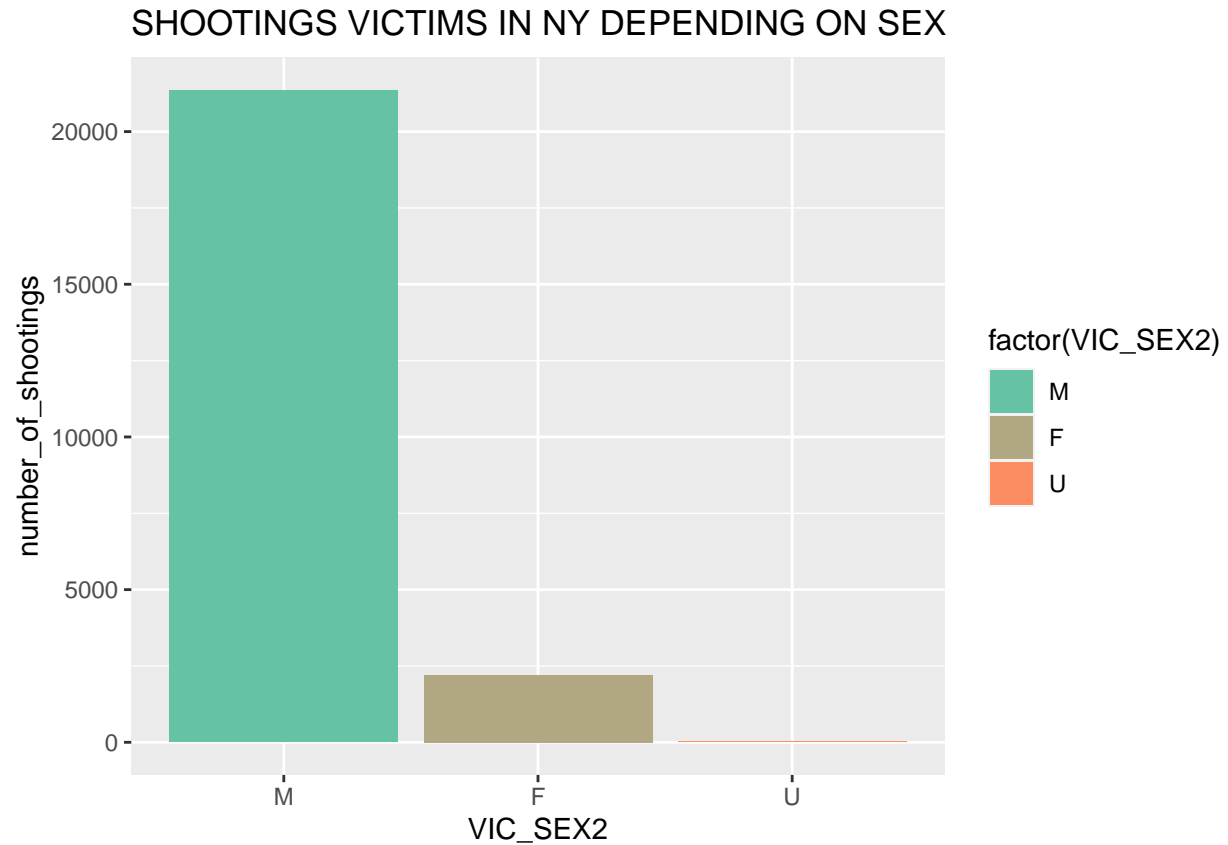
```
ggplot(NYPD_shooting_data_neighborhood, aes(x=Neighborhood, y=number_of_shootings,
fill = factor(Neighborhood))) +
geom_col() + scale_fill_manual(values = mycolors) + ggtitle("SHOOTINGS IN NY DEPENDING ON NEIGHBORHOOD")
```



Looking at this graph we can conclude that Brooklyn has the most number of shootings and Staten Island has the least.

Number of victims depending on Sex

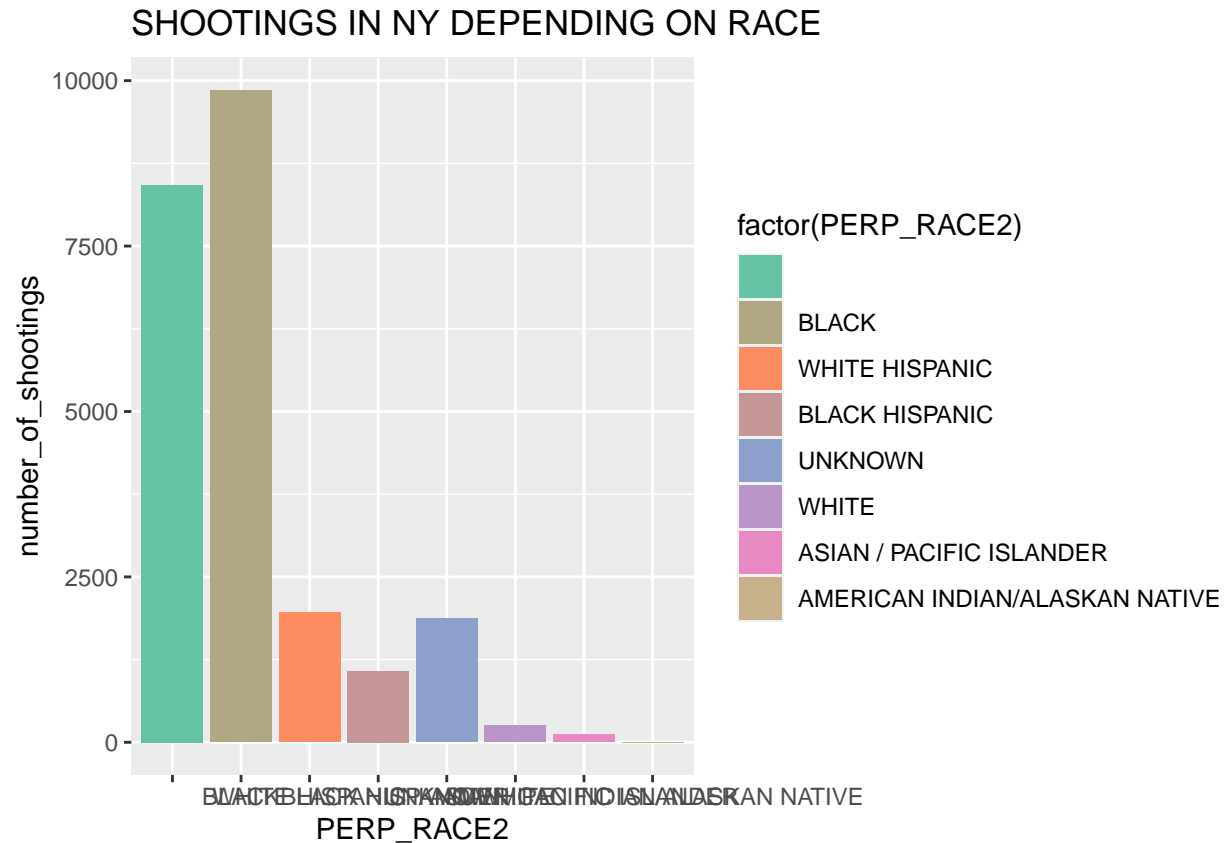
```
ggplot(NYPD_shooting_data_sex, aes(x=VIC_SEX2, y=number_of_shootings, fill = factor(VIC_SEX2))) +
geom_col() + ggtitle("SHOOTINGS VICTIMS IN NY DEPENDING ON SEX") + scale_fill_manual(values = mycolors)
```



Looking at this graph we can conclude that most of the victims were male.

Number of Shootings depending on Race

```
ggplot(NYPD_shooting_data_race, aes(x=PERP_RACE2, y=number_of_shootings, fill = factor(PERP_RACE2))) +
  geom_col() + ggtitle("SHOOTINGS IN NY DEPENDING ON RACE")+ scale_fill_manual(values = mycolors)
```



looking at this bar graph we can say that most victims were black.

## DATA MODELLING

Creating a Model where the number of shooting incidents which involved race and how many were turned into murders in neighborhood(Boro).

```
shootings2 <- NYPD_shooting_data %>% group_by(Neighborhood,
PERP_RACE2, STATISTICAL_MURDER_FLAG == "true") %>%
summarise(tot = n())
```

## 'summarise()' has grouped output by 'Neighborhood', 'PERP\_RACE2'. You can override using the '.group

```
glimpse(shootings2)
```

```
## Rows: 72
## Columns: 4
## Groups: Neighborhood, PERP_RACE2 [37]
## $ Neighborhood      <fct> QUEENS, QUEENS, QUEENS, QUEENS, QU~
## $ PERP_RACE2         <fct> , , BLACK, BLACK, WHITE HISPANIC, ~
## $ 'STATISTICAL_MURDER_FLAG == "true"' <lgl> FALSE, TRUE, FALSE, TRUE, FALSE, T~
## $ tot                <int> 967, 243, 1215, 296, 230, 73, 76, ~
```

```
lm_mod <- lm(tot ~ PERP_RACE2, data = shootings2)
```

```
# Summary of the model
summary(lm_mod)
```

```
##
## Call:
## lm(formula = tot ~ PERP_RACE2, data = shootings2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -906.40 -139.65   -8.25   21.62 2473.60
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      842.5      173.1   4.867 7.74e-06
## PERP_RACE2BLACK      142.9      244.8   0.584 0.56145
## PERP_RACE2WHITE HISPANIC    -646.5      244.8  -2.641 0.01038
## PERP_RACE2BLACK HISPANIC    -734.4      244.8  -3.000 0.00384
## PERP_RACE2UNKNOWN    -655.6      244.8  -2.678 0.00940
## PERP_RACE2WHITE     -817.0      244.8  -3.337 0.00141
## PERP_RACE2ASIAN / PACIFIC ISLANDER    -830.5      244.8  -3.392 0.00119
## PERP_RACE2AMERICAN INDIAN/ALASKAN NATIVE  -841.5      424.0  -1.985 0.05148
##
## (Intercept)          ***
## PERP_RACE2BLACK
## PERP_RACE2WHITE HISPANIC          *
## PERP_RACE2BLACK HISPANIC          **
## PERP_RACE2UNKNOWN          **
## PERP_RACE2WHITE          **
## PERP_RACE2ASIAN / PACIFIC ISLANDER  **
## PERP_RACE2AMERICAN INDIAN/ALASKAN NATIVE .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 547.4 on 64 degrees of freedom
## Multiple R-squared:  0.3417, Adjusted R-squared:  0.2697
## F-statistic: 4.745 on 7 and 64 DF,  p-value: 0.0002438
```

Possible sources of Bias. I visualize data by checking victims depending on sex. According to my analysis I saw that most of the victims are male. i might be bias on my analysis as I checked on sex. I wanted to check shootings depending upon race. Others can analyze depending on date or based on any other field.