

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 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 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 .csv data file from the <https://data.cityofnewyork.us/api/views/833y-fsy8/rows>.
csv path

```
##
## -- 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()
## )
```

Data Tidying

Summary of NY Shooting Data

##	INCIDENT_KEY	OCCUR_DATE	OCCUR_TIME	BORO
----	--------------	------------	------------	------

```
## Min. : 9953245 Length:6843 Length:6843 Length:6843
## 1st Qu.: 37294982 Class :character Class1:hms Class :character
## Median : 71859341 Mode :character Class2:difftime Mode :character
## Mean : 81934179 Mode :numeric
## 3rd Qu.: 92982910
## Max. :222299954
## PRECINCT LOCATION_DESC STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
## Min. : 1.00 Length:6843 Mode :logical Length:6843
## 1st Qu.: 43.00 Class :character FALSE:5344 Class :character
## Median : 69.00 Mode :character TRUE :1499 Mode :character
## Mean : 65.56
## 3rd Qu.: 81.00
## Max. :123.00
## PERP_SEX PERP_RACE VIC_AGE_GROUP VIC_SEX
## Length:6843 Length:6843 Length:6843 Length:6843
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## VIC_RACE
## Length:6843
## 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 2709
## 2 BRONX 1903
## 3 QUEENS 992
## 4 MANHATTAN 983
## 5 STATEN ISLAND 256
```

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             5979
## 2 F             861
## 3 U              3
```

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: 7 x 2
##   PERP_RACE2          number_of_shootings
##   <fct>                <int>
## 1 BLACK                4633
## 2 WHITE HISPANIC       802
## 3 UNKNOWN              760
## 4 BLACK HISPANIC       449
## 5 WHITE                140
## 6 ASIAN / PACIFIC ISLANDER  58
## 7 AMERICAN INDIAN/ALASKAN NATIVE  1
```

DATA VISUALIZATON

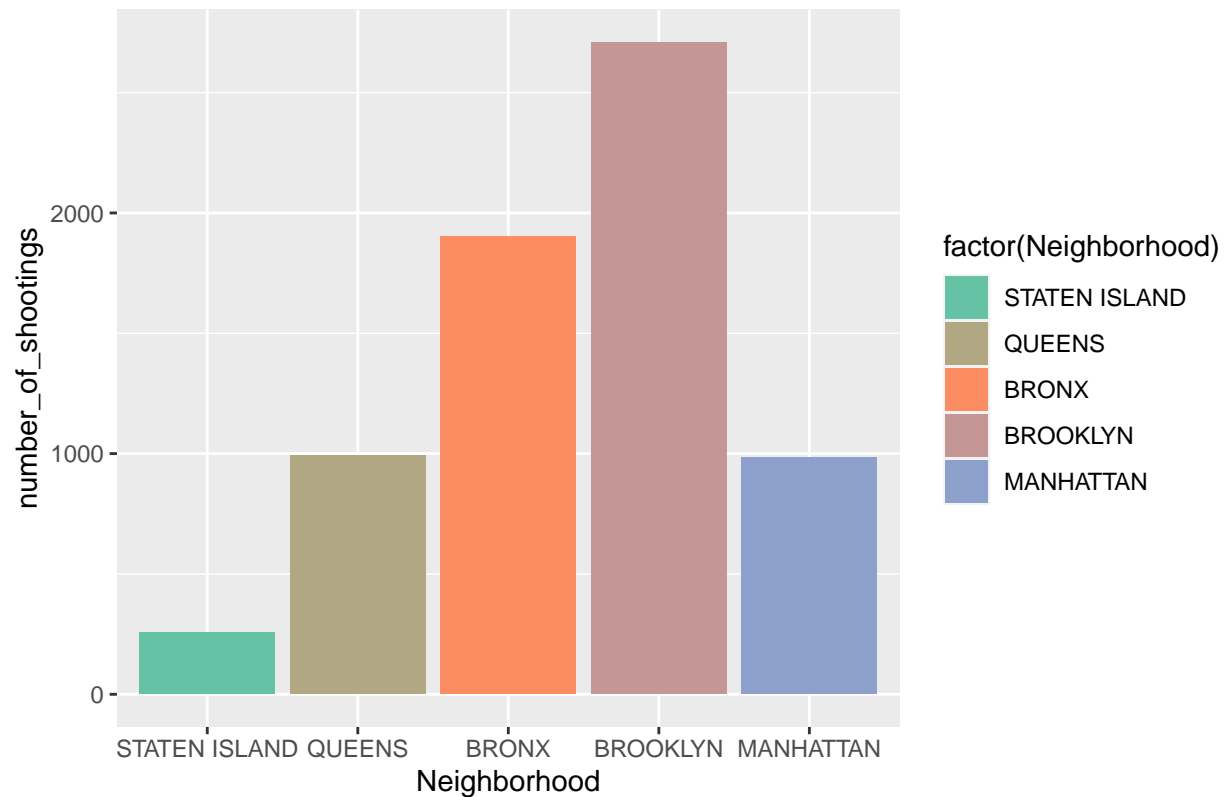
```
#lets plot this to visualize it better. a barchart might be useful.
#need to add more colors to color palette so we can see all year totals in each bar

#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")
```

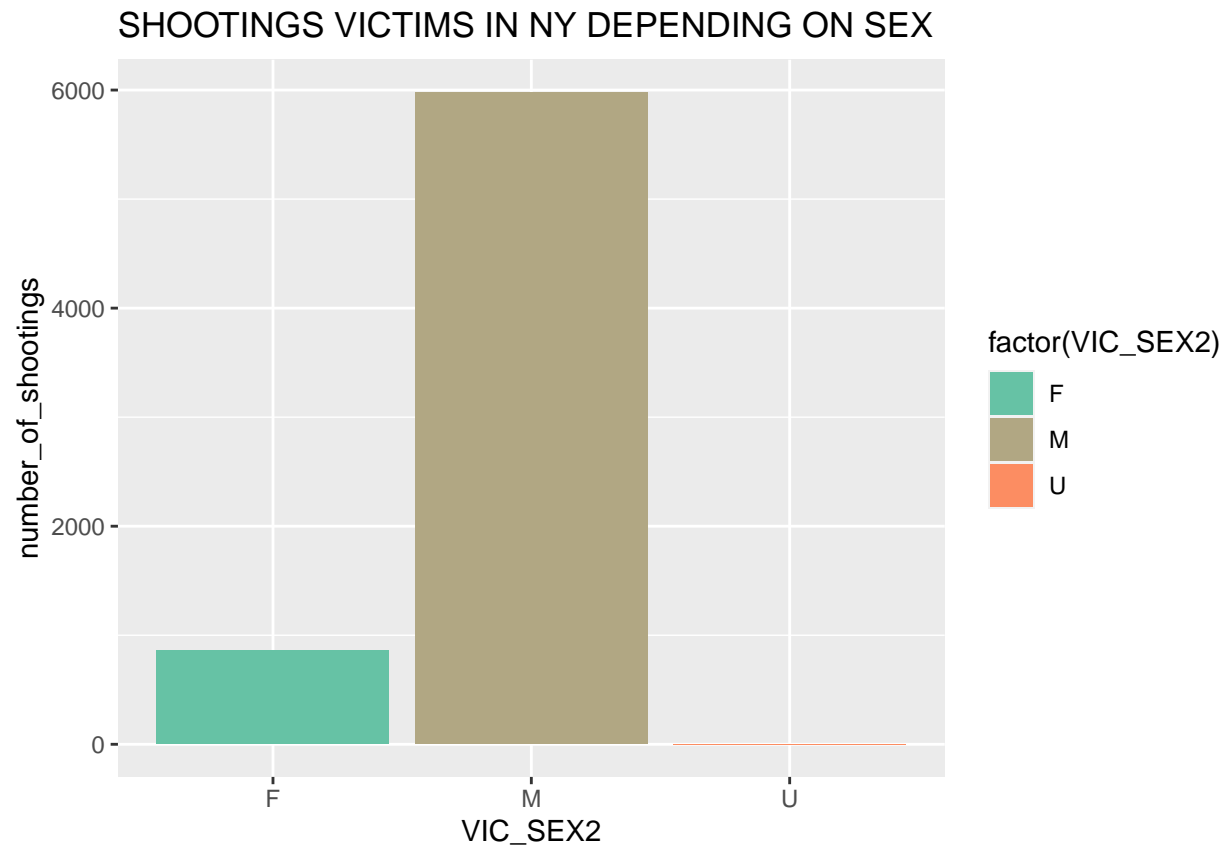
SHOOTINGS IN NY DEPENDING ON NEIGHBORHOOD



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)
```

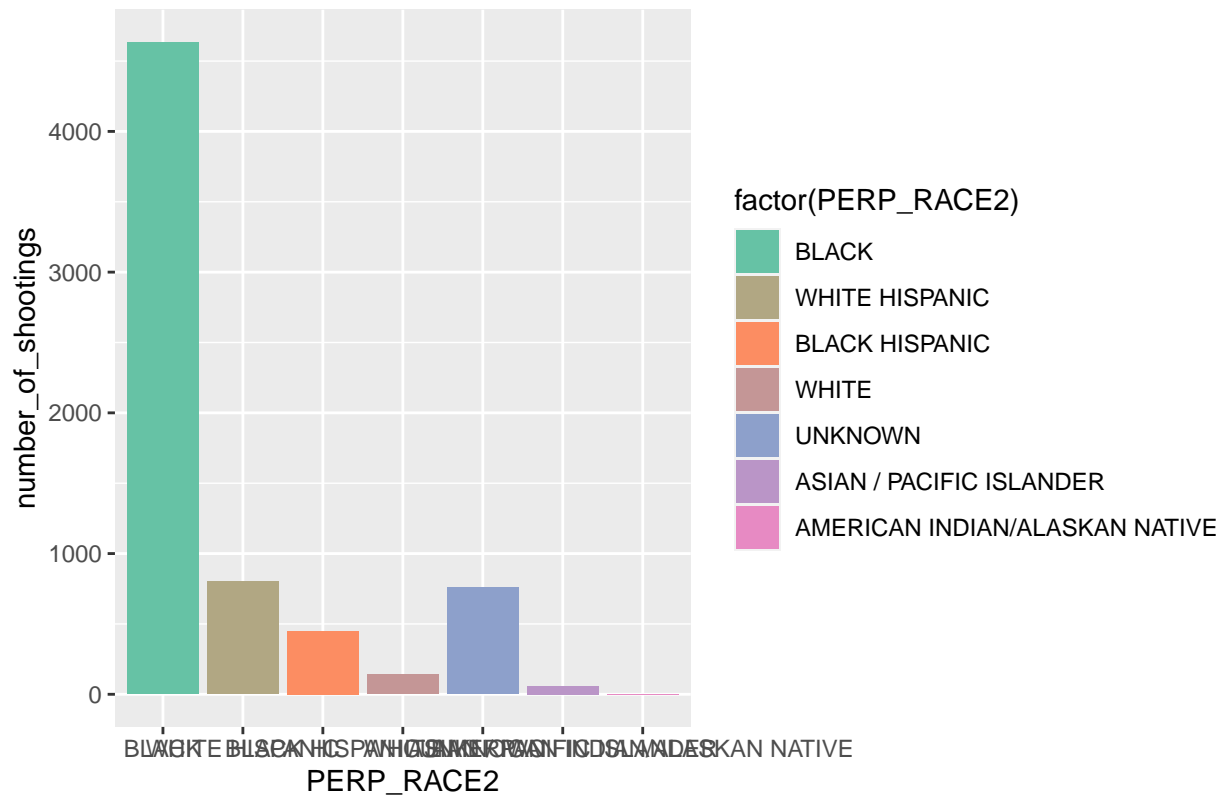


Looking at this graph we can conclude that most of the victims were male.

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)
```

SHOOTINGS IN NY DEPENDING ON RACE



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: 31
## Columns: 4
## Groups: Neighborhood, PERP_RACE2 [31]
## $ Neighborhood      <fct> STATEN ISLAND, STATEN ISLAND, STAT~
## $ PERP_RACE2         <fct> BLACK, WHITE HISPANIC, BLACK HISPANIC, WHITE, UNKNOWN, ASIAN / PACIFIC ISLANDER, AMERICAN INDIAN/ALASKAN NATIVE
## $ 'STATISTICAL_MURDER_FLAG == "true"' <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE
## $ tot                <int> 171, 35, 7, 24, 15, 4, 657, 112, 3~
```

```
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
## -755.6   -50.1    -4.6    12.3  1156.4
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   926.6      136.9   6.770 5.29e-07
## PERP_RACE2WHITE HISPANIC      -766.2      193.5  -3.959 0.000585
## PERP_RACE2BLACK HISPANIC      -836.8      193.5  -4.323 0.000232
## PERP_RACE2WHITE                -898.6      193.5  -4.643 0.000103
## PERP_RACE2UNKNOWN              -774.6      193.5  -4.002 0.000524
## PERP_RACE2ASIAN / PACIFIC ISLANDER -915.0      193.5  -4.728 8.30e-05
## PERP_RACE2AMERICAN INDIAN/ALASKAN NATIVE -925.6      335.2  -2.761 0.010867
##
## (Intercept)                  ***
## PERP_RACE2WHITE HISPANIC      ***
## PERP_RACE2BLACK HISPANIC      ***
## PERP_RACE2WHITE                ***
## PERP_RACE2UNKNOWN              ***
## PERP_RACE2ASIAN / PACIFIC ISLANDER ***
## PERP_RACE2AMERICAN INDIAN/ALASKAN NATIVE *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 306 on 24 degrees of freedom
## Multiple R-squared:  0.5774, Adjusted R-squared:  0.4718
## F-statistic: 5.466 on 6 and 24 DF,  p-value: 0.0011
```

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.