NYPD Shooting Incidents

DATA SETS IMPORTING, TIDY & TRANSFORMING, AND VISUALIZING DATA

I am proud to present my first Data Science project to my colleagues in Data Science as a Field Course.

- STUDENT IN MSDS PROGRAM AT CU-BOULDER.

Step 1 : Importing Data

```
library(tidyverse)

library(tinytex)

 library(lubridate)

url <- “https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD”

nypd <- read_csv(url)
```

Step 2: Tidy & Transform Data

summary(nypd)

Removing Unwanted Columns from the Data Set

nypd = nypd %>% select(-c(INCIDENT_KEY, OCCUR_TIME, PRECINCT, JURISDICTION_CODE, STATISTICAL_MURDER_FLAG, PERP_RACE, VIC_RACE, X_COORD_CD, Y_COORD_CD, Latitude, Longitude, Lon_Lat))

Changing the "OCCUR_DATE" column data type into date

class(nypd$OCCUR_DATE)

```
[1] "character"
```

nypd <- nypd %>% mutate (OCCUR_DATE = mdy(OCCUR_DATE))

class(nypd$OCCUR_DATE)

```
[1] "Date"
```

Step 2: Tidy & Transform Data

changing names for some columns

colnames(nypd)[1] <- “Date”

colnames(nypd)[2] <- “Borough”

colnames(nypd)[3] <- “place”

colnames(nypd)[4] <- “Perp_Age”

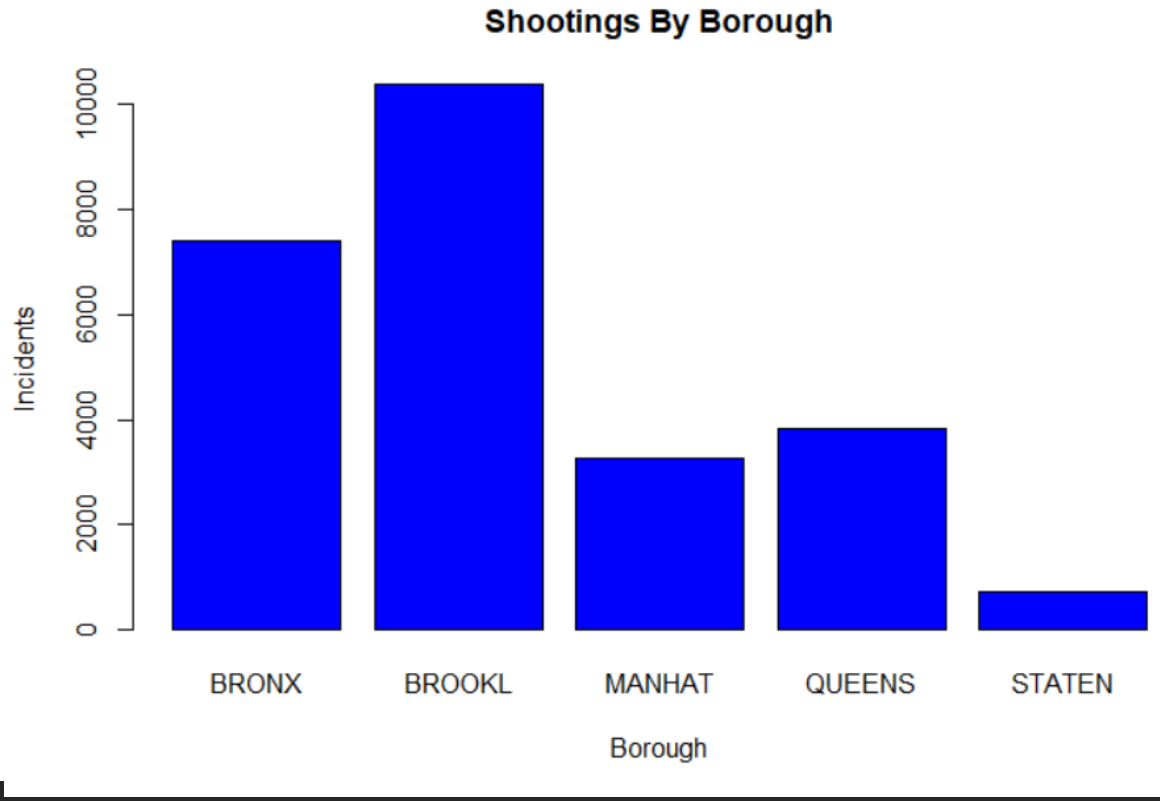
colnames(nypd)[5] <- “Perp_Gender”

colnames(nypd)[6] <- “Vic_Age”

colnames(nypd)[7] <- “Vic_Gender”

Step 3: Visualizing & Analyzing

barplot(a$n, names = a$Borough, xlab="Borough", ylab="Incidents", main="Shootings By Borough", col="blue")



Step 3: Visualizing & Analyzing

barplot(b*n*,*names*=*b*Vic_Age, xlab="Victom age", ylab="Shootings", main="Incidents By Victom Age", col="green")

Step 4: Bias Identification

To avoid bias I did not like to analyze the data in terms of race.