NYPD Shooting Incident Data

Loading Libraries

Importing Data I will start by reading in the data from the CSV file by using the link from https://catalog.data.gov/dataset

```
NYPD_shooting_data <- read.csv(url("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessTy
```

Reading the dataset to see what we got.

Will be using the head function to see the first 5 rows of the dataset.

```
head(NYPD_shooting_data,5)
```

```
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME          BORO PRECINCT JURISDICTION_CODE
## 1    201575314 08/23/2019   22:10:00        QUEENS      103                 0
## 2    205748546 11/27/2019   15:54:00         BRONX       40                 0
## 3    193118596 02/02/2019   19:40:00     MANHATTAN       23                 0
## 4    204192600 10/24/2019   00:52:00 STATEN ISLAND      121                 0
## 5    201483468 08/22/2019   18:03:00         BRONX       46                 0
##   LOCATION_DESC STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX      PERP_RACE
## 1                                 false
## 2                                 false            <18        M          BLACK
## 3                                 false          18-24        M WHITE HISPANIC
## 4     PVT HOUSE                    true          25-44        M          BLACK
## 5                                 false          25-44        M BLACK HISPANIC
##   VIC_AGE_GROUP VIC_SEX       VIC_RACE X_COORD_CD Y_COORD_CD Latitude Longitude
## 1         25-44       M          BLACK    1037451     193561 40.69781 -73.80814
## 2         25-44       F          BLACK    1006789     237559 40.81870 -73.91857
## 3         18-24       M BLACK HISPANIC     999347     227795 40.79192 -73.94548
## 4         25-44       F          BLACK     938149     171781 40.63806 -74.16611
## 5         18-24       M          BLACK    1008224     250621 40.85455 -73.91334
##                                    Lon_Lat
## 1 POINT (-73.80814071699996 40.697805308000056)
## 2  POINT (-73.91857061799993 40.81869973000005)
## 3 POINT (-73.94547965999999 40.791916091000076)
## 4  POINT (-74.16610830199996 40.63806398200006)
## 5  POINT (-73.91333944399999 40.85454734900003)
```

##Changing the Data Frame.

After looking through the data, I decided to keep: Date, Boro, Perp_Age_group, Perp_sex, Perp_Race, Vic_Age_group, Vic_sex,and Vic_RAce

```r
NYPD_New <- NYPD_shooting_data[c("OCCUR_DATE", "OCCUR_TIME", "BORO", "PERP_AGE_GROUP", "PERP_SEX", "PERP
```

Renaming the columns so that they are more easily readable. Transforming all the blank columns into NA.

```r
colnames(NYPD_New) <- c("Date", "Time", "Region", "Attacker_Age_Group", "Attacker_Gender", "Attacker_Rac
NYPD_New[NYPD_New==""]<-NA

NYPD_New<-NYPD_New %>%
  mutate(Date = mdy(Date))

head(NYPD_New,5)
```

```
##          Date     Time         Region Attacker_Age_Group Attacker_Gender
## 1 2019-08-23 22:10:00         QUEENS               <NA>            <NA>
## 2 2019-11-27 15:54:00          BRONX                <18               M
## 3 2019-02-02 19:40:00      MANHATTAN              18-24               M
## 4 2019-10-24 00:52:00 STATEN ISLAND              25-44               M
## 5 2019-08-22 18:03:00          BRONX              25-44               M
##    Attacker_Race Victim_Age Victim_Gender     Victim_Race
## 1           <NA>      25-44             M           BLACK
## 2          BLACK      25-44             F           BLACK
## 3 WHITE HISPANIC      18-24             M BLACK HISPANIC
## 4          BLACK      25-44             F           BLACK
## 5 BLACK HISPANIC      18-24             M           BLACK
```

```r
#replacing Attacker Gender
NYPD_New$Attacker_Gender[NYPD_New$Attacker_Gender =="M"] <- 1
NYPD_New$Attacker_Gender[NYPD_New$Attacker_Gender =="F"] <- 0
NYPD_New$Attacker_Gender[NYPD_New$Attacker_Gender =="U"] <- NA

#replacing Victim Gender
NYPD_New$Victim_Gender[NYPD_New$Victim_Gender =="M"]<-1
NYPD_New$Victim_Gender[NYPD_New$Victim_Gender =="F"]<-0

#Replacing Attacker Age Group
NYPD_New$Attacker_Age_Group[NYPD_New$Attacker_Age_Group=="<18"] <- 0
NYPD_New$Attacker_Age_Group[NYPD_New$Attacker_Age_Group=="18-24"] <- 1
NYPD_New$Attacker_Age_Group[NYPD_New$Attacker_Age_Group=="25-44"] <- 2
NYPD_New$Attacker_Age_Group[NYPD_New$Attacker_Age_Group=="45-64"] <- 3
NYPD_New$Attacker_Age_Group[NYPD_New$Attacker_Age_Group=="65+"] <- 4
NYPD_New$Attacker_Age_Group[NYPD_New$Attacker_Age_Group=="UNKNOWN"] <- NA

#Replacing Victim Age Group
NYPD_New$Victim_Age[NYPD_New$Victim_Age=="<18"] <-0
NYPD_New$Victim_Age[NYPD_New$Victim_Age=="18-24"] <-1
NYPD_New$Victim_Age[NYPD_New$Victim_Age=="25-44"] <-2
NYPD_New$Victim_Age[NYPD_New$Victim_Age=="45-64"] <-3
NYPD_New$Victim_Age[NYPD_New$Victim_Age=="65+"] <-4
NYPD_New$Victim_Age[NYPD_New$Victim_Age=="UNKNOWN"] <-NA
```
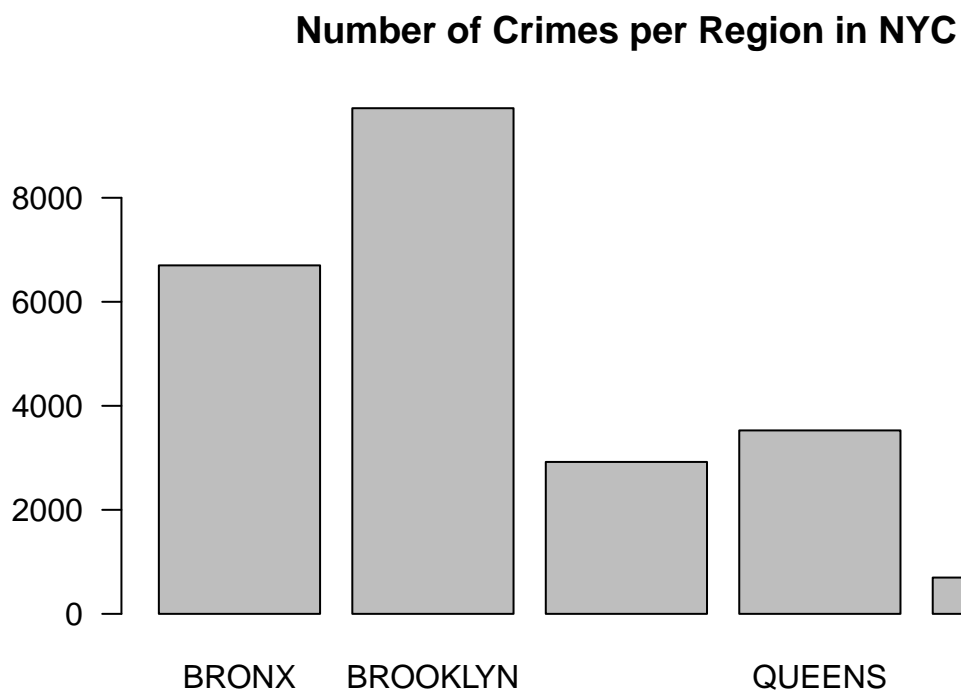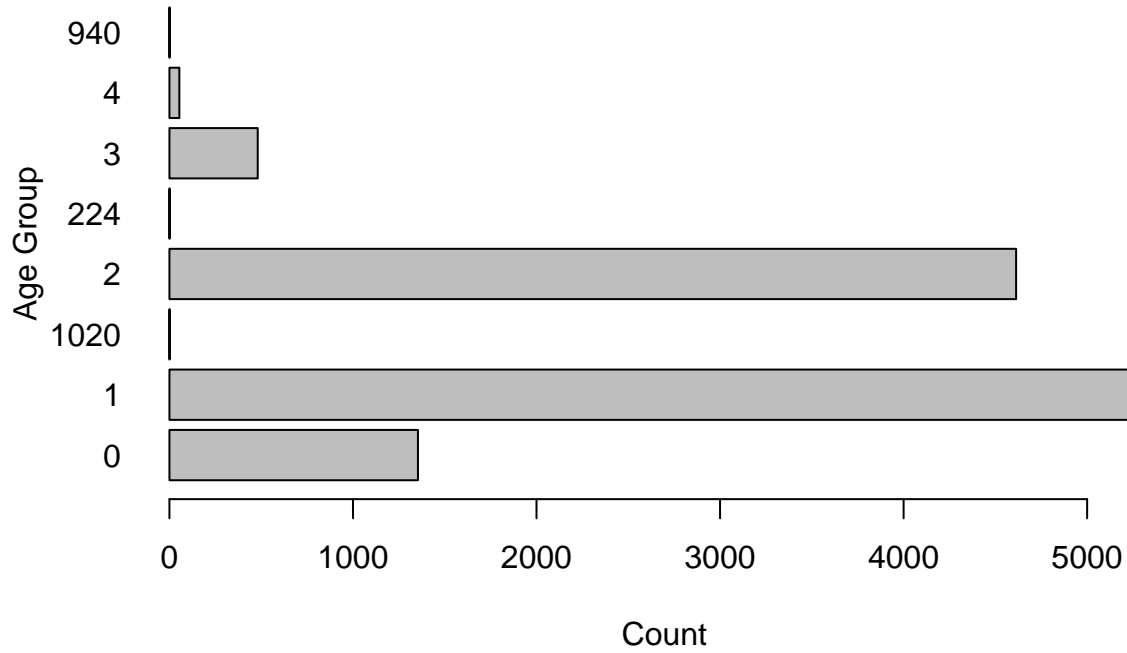
Creating a bar plot to see where the most shooting incidents occur. By this plot, we can easily see Brooklyn has the most shooting incidents.

```r
barplot(table(NYPD_New$Region), main = "Number of Crimes per Region in NYC", las = 1)
```
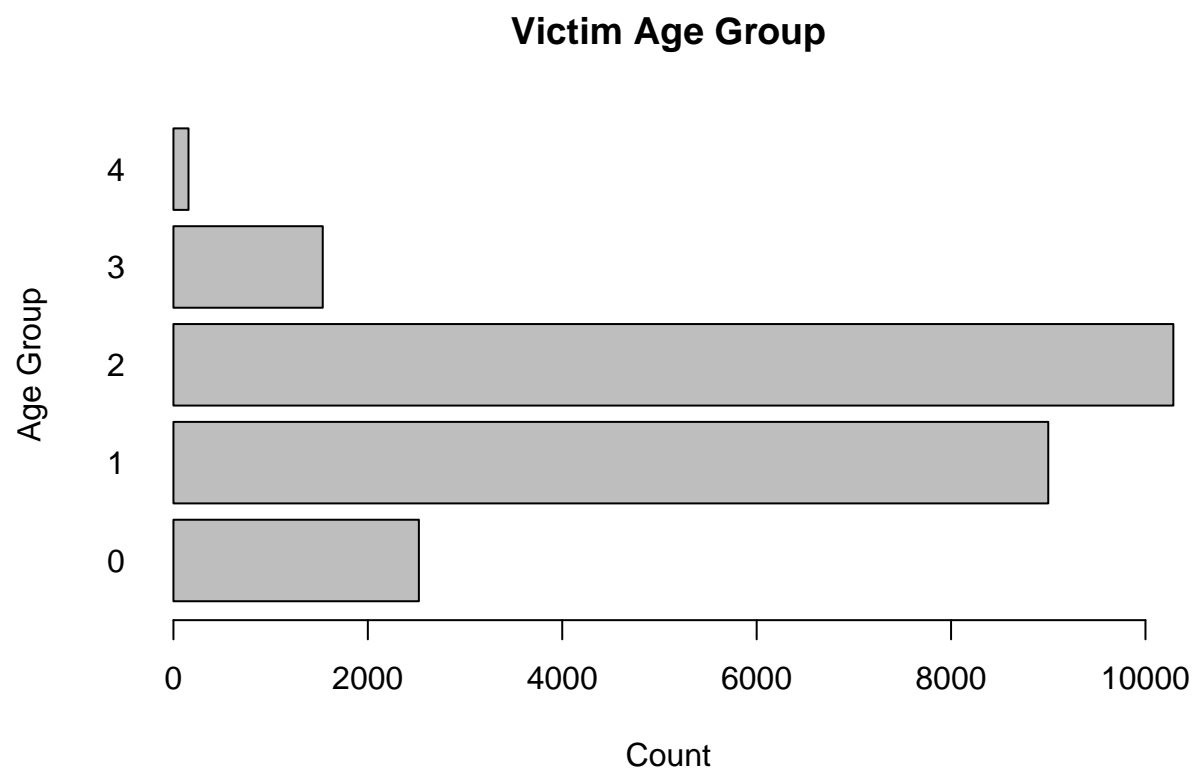
Number of Crimes per Region in NYC


```r
barplot(table(NYPD_New$Attacker_Age_Group), horiz = T, xlab = "Count", ylab = "Age Group", las =1, main
```
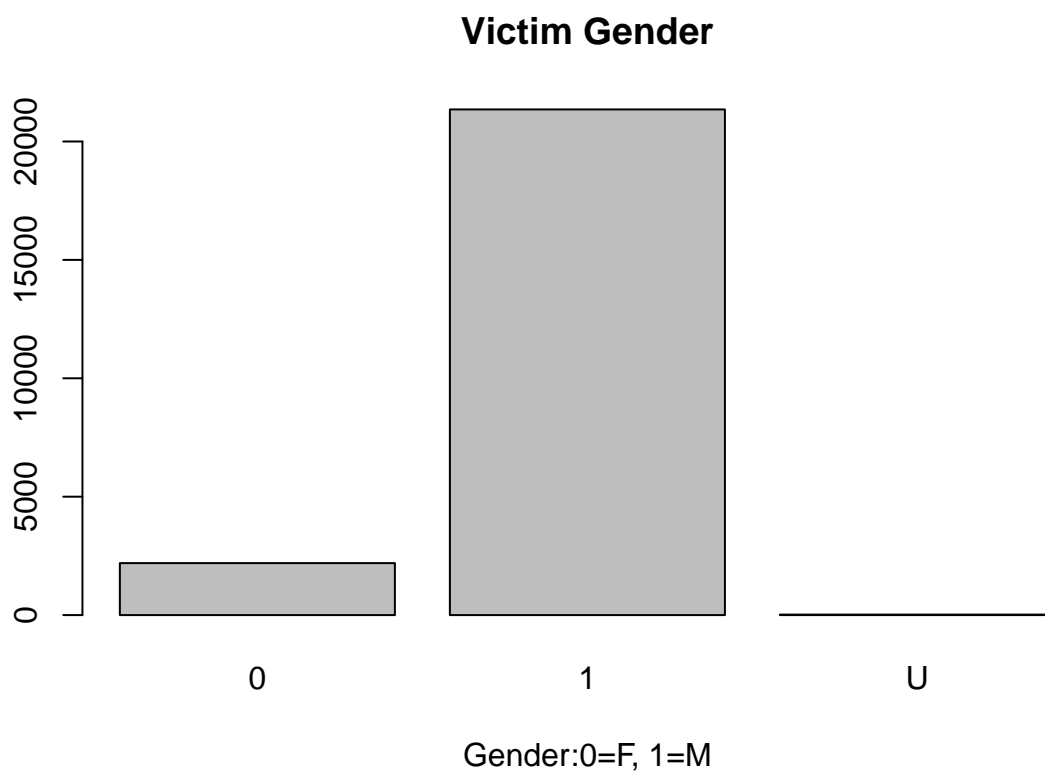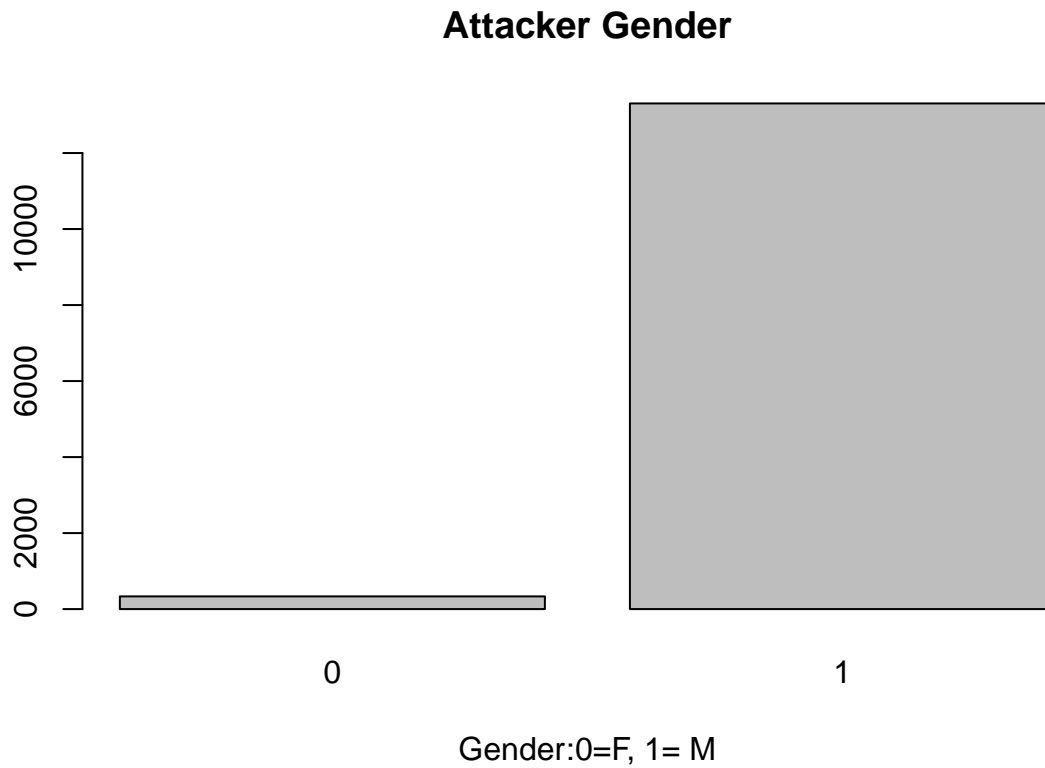
Attacker Age Group


```
barplot(table(NYPD_New$Victim_Age), horiz = T, xlab = "Count", ylab = "Age Group", las =1, main = "Victi
```

Victim Age Group


```r
barplot(table(NYPD_New$Victim_Gender), xlab = "Gender:0=F, 1=M", main = "Victim Gender")
```

Victim Gender

Gender:0=F, 1=M

```r
barplot(table(NYPD_New$Attacker_Gender), xlab = "Gender:0=F, 1= M", main = "Attacker Gender")
```

Attacker Gender

Gender:0=F, 1= M

```
#using only columns that i want to use
for_model <- NYPD_New[c("Attacker_Gender", "Victim_Gender", "Attacker_Age_Group", "Victim_Age")]

mod1 <- lm(Attacker_Age_Group ~ Victim_Gender + Victim_Age, data = for_model)
```

##Possible Bias Sources: Step 4 Conclusion to my project and include any possible sources of bias. Be sure to identify what your personal bias might be and how you have mitigated that. I view data for NYC as a whole rather than per borough. I disregarded race of the victim, as that may have something to do with the attacker.