NYPD Shooting Incident Data Report

Xiao

2023-01-23

Description of the data - NYPD Shooting Incident Data (Historic):

This is a breakdown of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. Each record represents a shooting incident in NYC and includes information about the event, the location and time of occurrence. In addition, information related to suspect and victim demographics is also included. This data can be used by the public to explore the nature of shooting/criminal activity. Please refer to the attached data footnotes for additional information about this dataset.

Questions of interest

This project will analyze NYPD shooting incidence data to assess the safety of New York City. The NYPD shooting incidence dataset that we use is accessible on data.gov. This dataset is a breakdown of every shooting incident that occurred in NYC from 2006 to 2022. My research questions are: Are all areas of Brooklyn dangerous? Is Manhattan a safer area? What time should you not be out in New York City?

Step 1. Start an Rmd Document

a. loading package

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

b. reading the data

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
NYPD <- read.csv(url_in)
```

Step 2. Tidy and Transform Data

a. check for missing value

```
sapply(NYPD,function(x) table(as.character(x) =="")["TRUE"])
```

```
##           INCIDENT_KEY.NA                OCCUR_DATE.NA
##                        NA                           NA
##           OCCUR_TIME.NA                      BORO.NA
##                        NA                           NA
##              PRECINCT.NA         JURISDICTION_CODE.NA
##                        NA                           NA
##        LOCATION_DESC.TRUE STATISTICAL_MURDER_FLAG.NA
##                     14977                           NA
##        PERP_AGE_GROUP.TRUE              PERP_SEX.TRUE
##                      9344                         9310
##           PERP_RACE.TRUE             VIC_AGE_GROUP.NA
##                      9310                           NA
##              VIC_SEX.NA                   VIC_RACE.NA
##                        NA                           NA
##           X_COORD_CD.NA                 Y_COORD_CD.NA
##                        NA                           NA
##             Latitude.NA                 Longitude.NA
##                        NA                           NA
##             Lon_Lat.NA
##                        NA
```

b.tidy dataset

Due to an excessive amount of missing values, the columns perp age group, perp sex, perp race, and location desc will not be used. I will be using the columns vic_age_group, vic_sex, vic_race, boro, and precinct columns. As they have high quality data. Columns jurisdiction code and the columns from x_coord_cd to lon_lat will not be used in my analysis.

```
NYPD <- subset(NYPD, select = -c(JURISDICTION_CODE,LOCATION_DESC,
PERP_AGE_GROUP:PERP_RACE,X_COORD_CD:Lon_Lat))
```

c. summary of the dataset

```
summary(NYPD)
```

```
##   INCIDENT_KEY         OCCUR_DATE          OCCUR_TIME              BORO
##  Min.   : 9953245   Length:25596      Length:25596       Length:25596
```

```

```
1st Qu.: 61593633   Class :character   Class :character   Class :character
Median : 86437258   Mode :character   Mode :character   Mode :character
Mean   :112382648
3rd Qu.:166660833
Max.   :238490103
PRECINCT   STATISTICAL_MURDER_FLAG VIC_AGE_GROUP   VIC_SEX
Min.   :   1.00   Length:25596   Length:25596   Length:25596
1st Qu.: 44.00   Class :character   Class :character   Class :character
Median : 69.00   Mode :character   Mode :character   Mode :character
Mean   : 65.87
3rd Qu.: 81.00
Max.   : 123.00
VIC_RACE
Length:25596
Class :character
Mode :character
##
##
##
```

# Step 3. Visualization and Analysis

## a. visualizing Data

```r
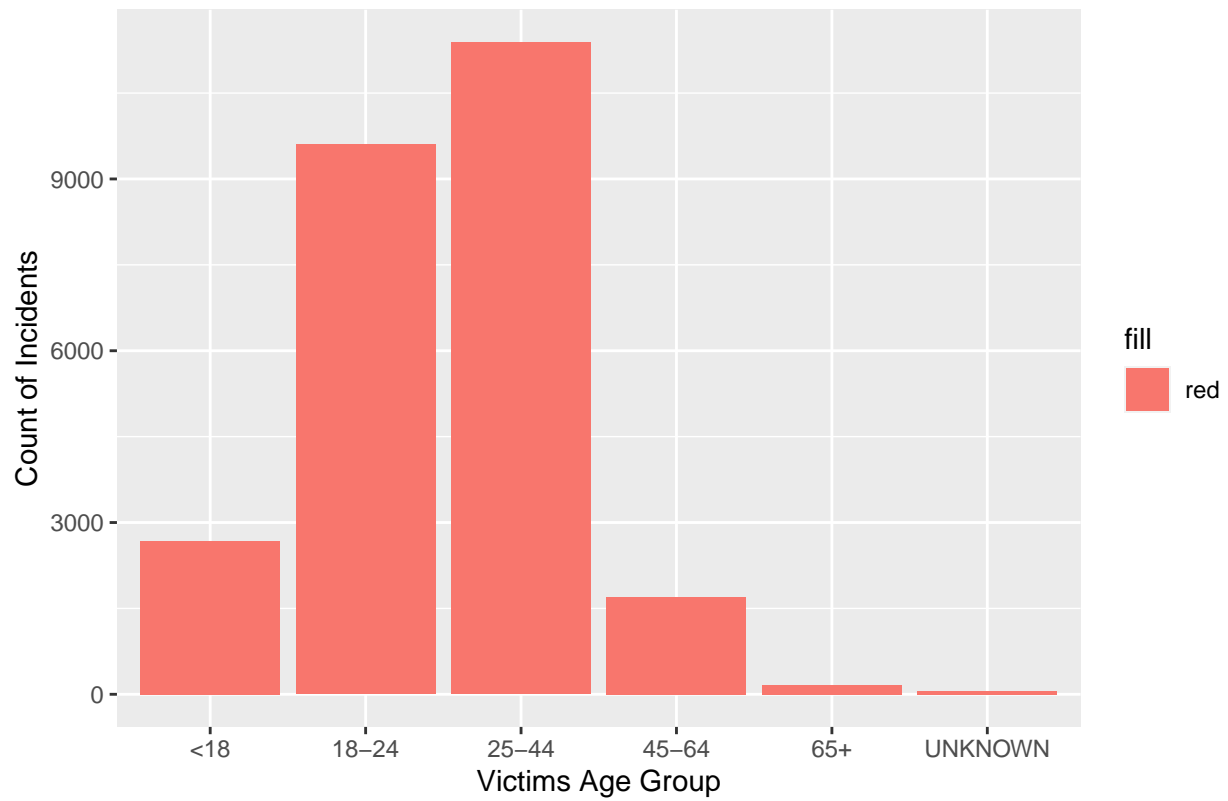ggplot(NYPD, aes(x=BORO,fill="red")) + geom_bar() +
  labs(title = "Distribution of the NYC shooting incidents",
       x="Boroughs", y="Count of Incidents")
```

Distribution of the NYC shooting incidents
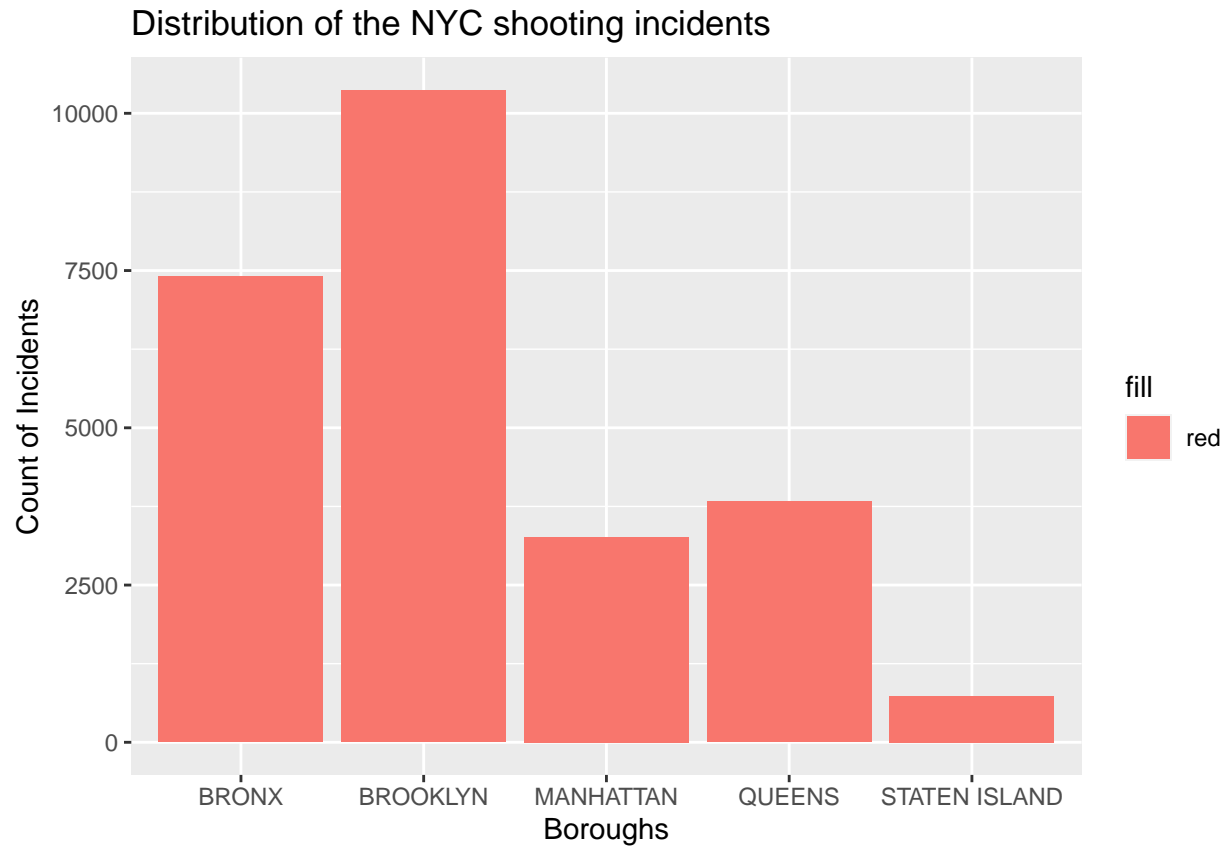

```
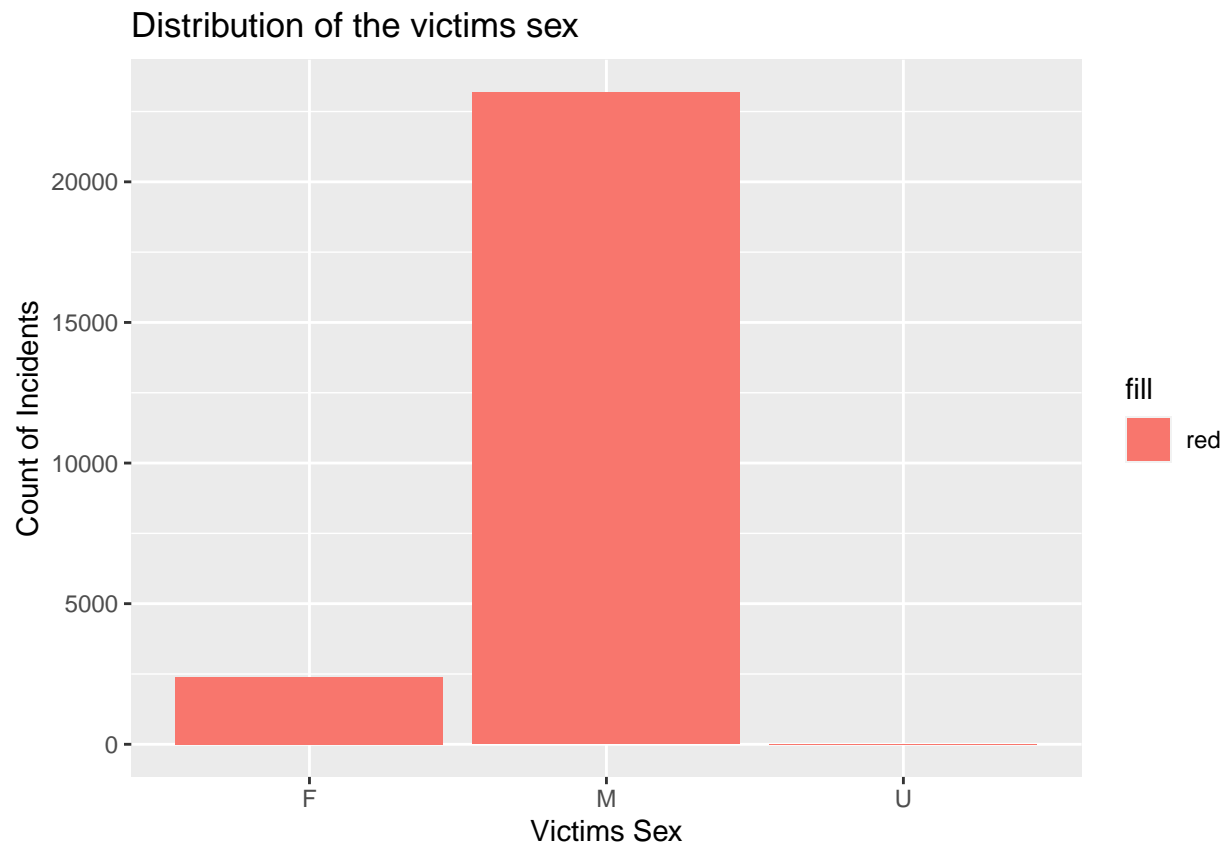ggplot(NYPD, aes(x=VIC_AGE_GROUP,fill="red")) + geom_bar() +
  labs(title = "Distribution of the victims age group",
       x="Victims Age Group",y="Count of Incidents")
```

Distribution of the victims age group


```
ggplot(NYPD, aes(x=VIC_SEX,fill="red")) + geom_bar() +
  labs(title = "Distribution of the victims sex",
       x="Victims Sex", y="Count of Incidents")
```

Distribution of the victims sex


```
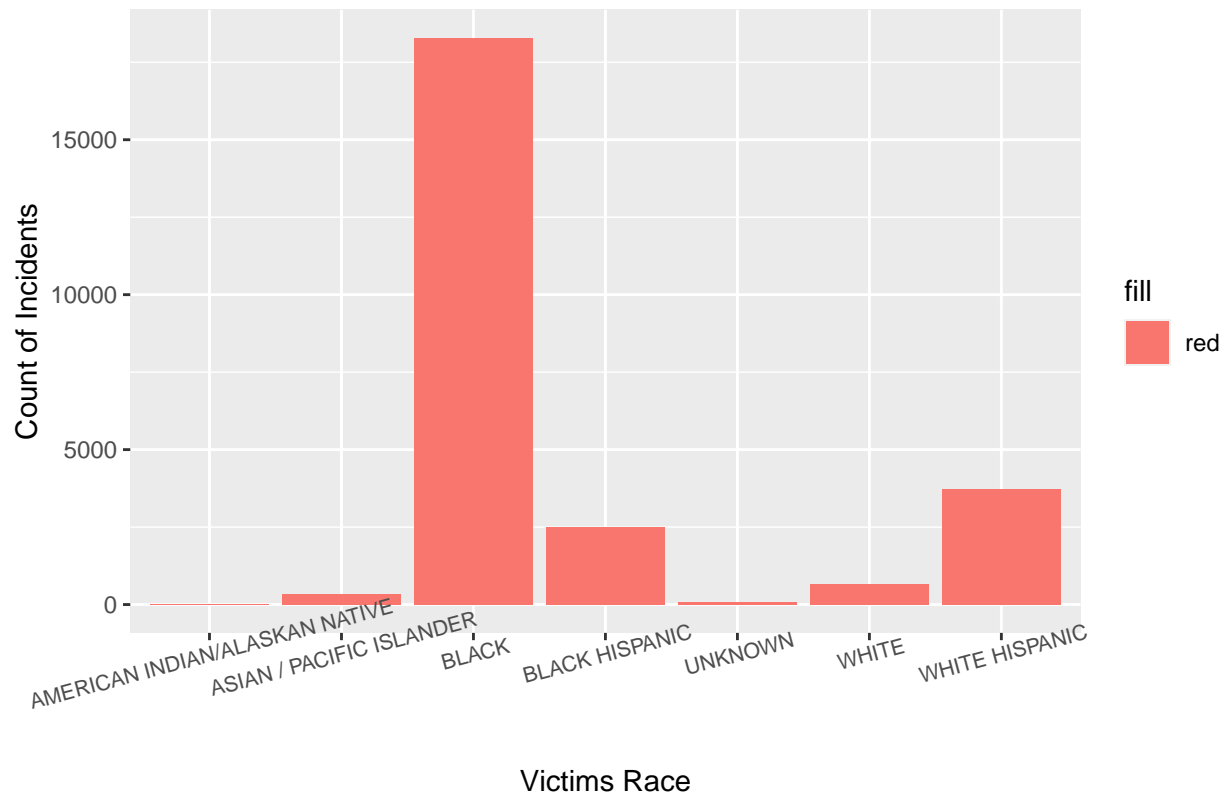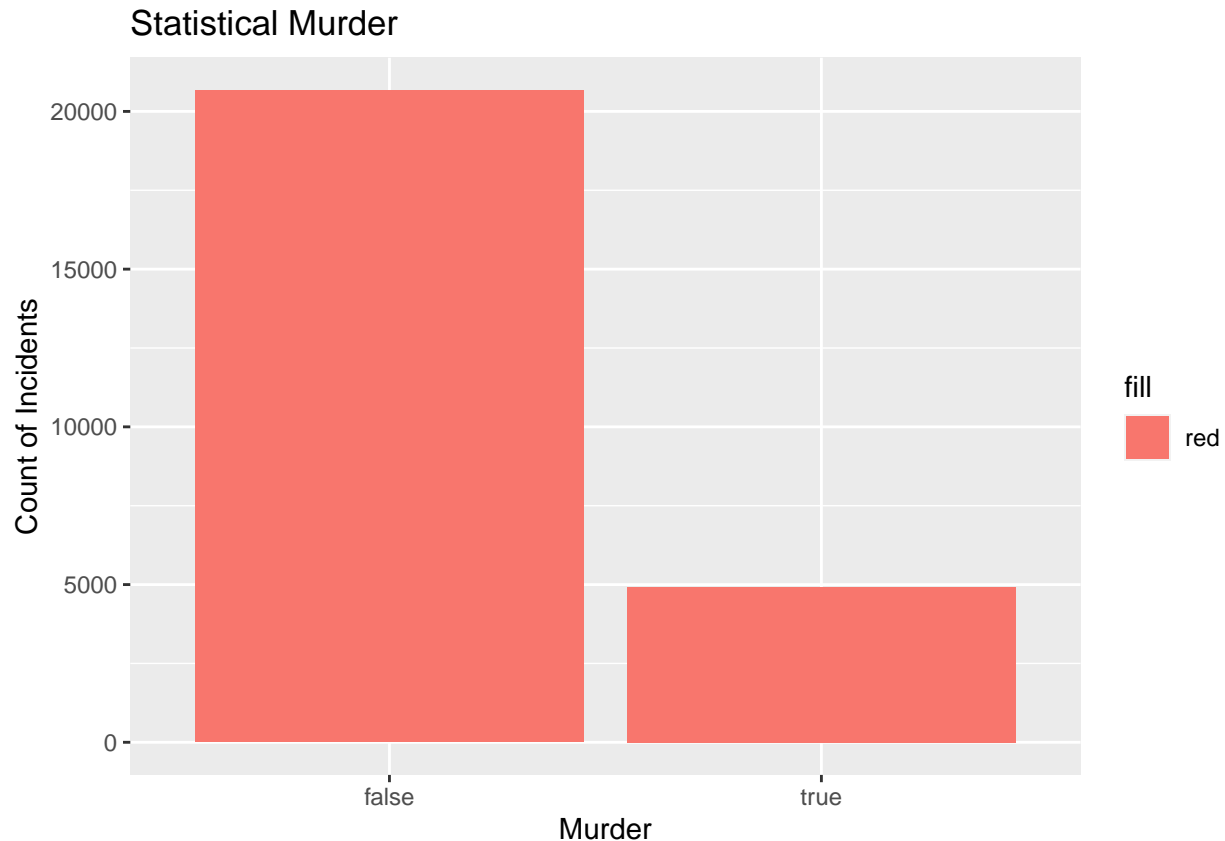ggplot(NYPD, aes(x=VIC_RACE,fill="red")) + geom_bar() +
  labs(title = "Distribution of the victims race",
       x="Victims Race", y="Count of Incidents") +
  theme(axis.text.x= element_text(size=8, angle=15))
```

Distribution of the victims race


```
ggplot(NYPD, aes(x=STATISTICAL_MURDER_FLAG,fill="red")) + geom_bar() +
  labs(title ="Statistical Murder", x="Murder", y="Count of Incidents")
```

Statistical Murder

According to the above bar charts, Brooklyn has the most shooting incidents. The majority of the victims are male, between the ages of 18 and 44, and are of the black race. 80% of the incidents are not murder. In light of this, I now have a research question. Are all areas of Brooklyn dangerous? Is Manhattan a safer area? What time should you not be out in New York City?

b. analysis and research questions

Analysis for research question for 1 and 2

1st to 34th precinct represented in Manhattan. 40th to 52nd precinct represented in Bronx. 60th to 94th precinct represented in Brooklyn. 100th to 115th precinct represented in Queens. 120th to 123rd represented in Staten Island.

```
ggplot(NYPD, aes(x=PRECINCT,fill=BORO)) + geom_bar() +
  labs(title = "Distribution of the incidents by precinct",
       x="Precinct", y="Count of Incidents")
```

Distribution of the incidents by precinct
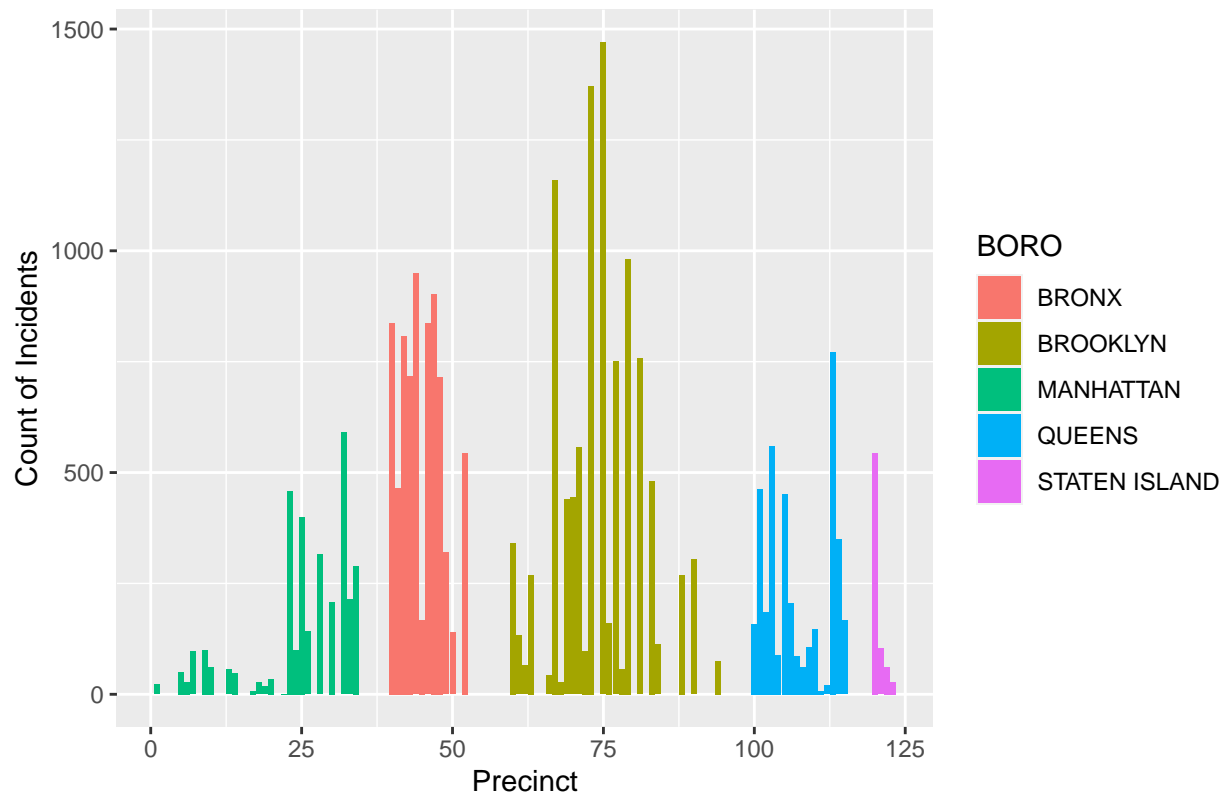
The first column of the table is the number of precincts.

```
table(NYPD$PRECINCT,NYPD$BORO)
```

```
##
##        BRONX BROOKLYN MANHATTAN QUEENS STATEN ISLAND
##   1        0        0        22      0             0
##   5        0        0        50      0             0
##   6        0        0        28      0             0
##   7        0        0        96      0             0
##   9        0        0       100      0             0
##   10       0        0        61      0             0
##   13       0        0        56      0             0
##   14       0        0        48      0             0
##   17       0        0         7      0             0
##   18       0        0        27      0             0
##   19       0        0        18      0             0
##   20       0        0        35      0             0
##   22       0        0         1      0             0
##   23       0        0       458      0             0
##   24       0        0       100      0             0
##   25       0        0       400      0             0
##   26       0        0       142      0             0
```

```
## 28      0      0    315      0      0
## 30      0      0    208      0      0
## 32      0      0    590      0      0
## 33      0      0    214      0      0
## 34      0      0    289      0      0
## 40    837      0      0      0      0
## 41    465      0      0      0      0
## 42    808      0      0      0      0
## 43    718      0      0      0      0
## 44    949      0      0      0      0
## 45    168      0      0      0      0
## 46    836      0      0      0      0
## 47    903      0      0      0      0
## 48    715      0      0      0      0
## 49    320      0      0      0      0
## 50    139      0      0      0      0
## 52    544      0      0      0      0
## 60      0    341      0      0      0
## 61      0    133      0      0      0
## 62      0     65      0      0      0
## 63      0    269      0      0      0
## 66      0     42      0      0      0
## 67      0   1160      0      0      0
## 68      0     28      0      0      0
## 69      0    439      0      0      0
## 70      0    444      0      0      0
## 71      0    556      0      0      0
## 72      0     98      0      0      0
## 73      0   1372      0      0      0
## 75      0   1470      0      0      0
## 76      0    161      0      0      0
## 77      0    750      0      0      0
## 78      0     57      0      0      0
## 79      0    982      0      0      0
## 81      0    757      0      0      0
## 83      0    481      0      0      0
## 84      0    112      0      0      0
## 88      0    269      0      0      0
## 90      0    304      0      0      0
## 94      0     75      0      0      0
## 100     0      0      0    159      0
## 101     0      0      0    463      0
## 102     0      0      0    185      0
## 103     0      0      0    560      0
## 104     0      0      0     88      0
## 105     0      0      0    451      0
## 106     0      0      0    206      0
## 107     0      0      0     86      0
## 108     0      0      0     61      0
## 109     0      0      0    107      0
## 110     0      0      0    147      0
## 111     0      0      0      8      0
## 112     0      0      0     20      0
## 113     0      0      0    772      0
```

```
## 114      0         0         0     349            0
## 115      0         0         0     166            0
## 120      0         0         0       0          544
## 121      0         0         0       0          103
## 122      0         0         0       0           61
## 123      0         0         0       0           28
```

research question 1. Are all areas of Brooklyn dangerous?

We can see from the bar chart and table that there are three precincts in Brooklyn—the 73rd, 75th, and 67th—where there have been more than 1000 shooting events. However, some regions, like the 62nd, 63rd, 68th, and 78th precincts, have fewer than 100 shooting incidents.

research question 2. Is Manhattan a safer area?

The table shows that the 1st through the 22nd and 24th precincts are fairly safe, with the number of gunshot incidents being equal to or below 100. The average number of shooting incidents in the 23rd, 25th to 34th precincts is similar to the other boroughs.

Analysis for research question 3

```r
hour <- as.integer(substr(NYPD$OCCUR_TIME,1,2))

NYPD <- data.frame(INCIDENT_KEY=NYPD$INCIDENT_KEY,OCCUR_DATE=NYPD$OCCUR_DATE, OCCUR_TIME=NYPD$OCCUR_TIME,
                   STATISTICAL_MURDER_FLAG=NYPD$STATISTICAL_MURDER_FLAG,
                   VIC_AGE_GROUP=NYPD$VIC_AGE_GROUP,VIC_SEX=NYPD$VIC_SEX,
                   VIC_RACE =NYPD$VIC_RACE)

ggplot(NYPD, aes(x=hour)) + geom_bar(fill="coral2") +
  labs(title = "Distribution of the incidents by Hours",
       x="Hour", y="Count of Incidents")
```

Distribution of the incidents by Hours
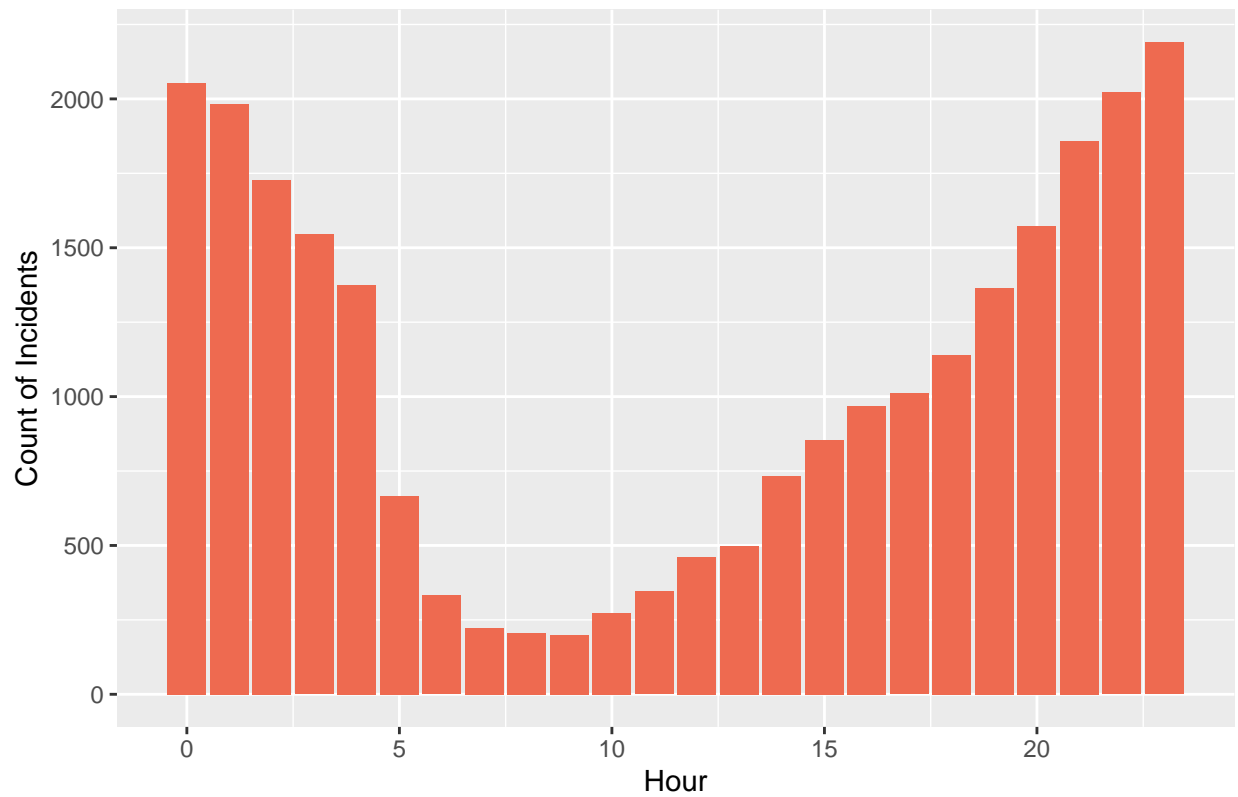

```
ggplot(NYPD, aes(x=hour,y=BORO)) + geom_jitter(color="coral2") +
    labs(title = "Distribution of the incidents by Hours for each Boroughs",
        x="Hour", y="Boroughs")
```

Distribution of the incidents by Hours for each Boroughs
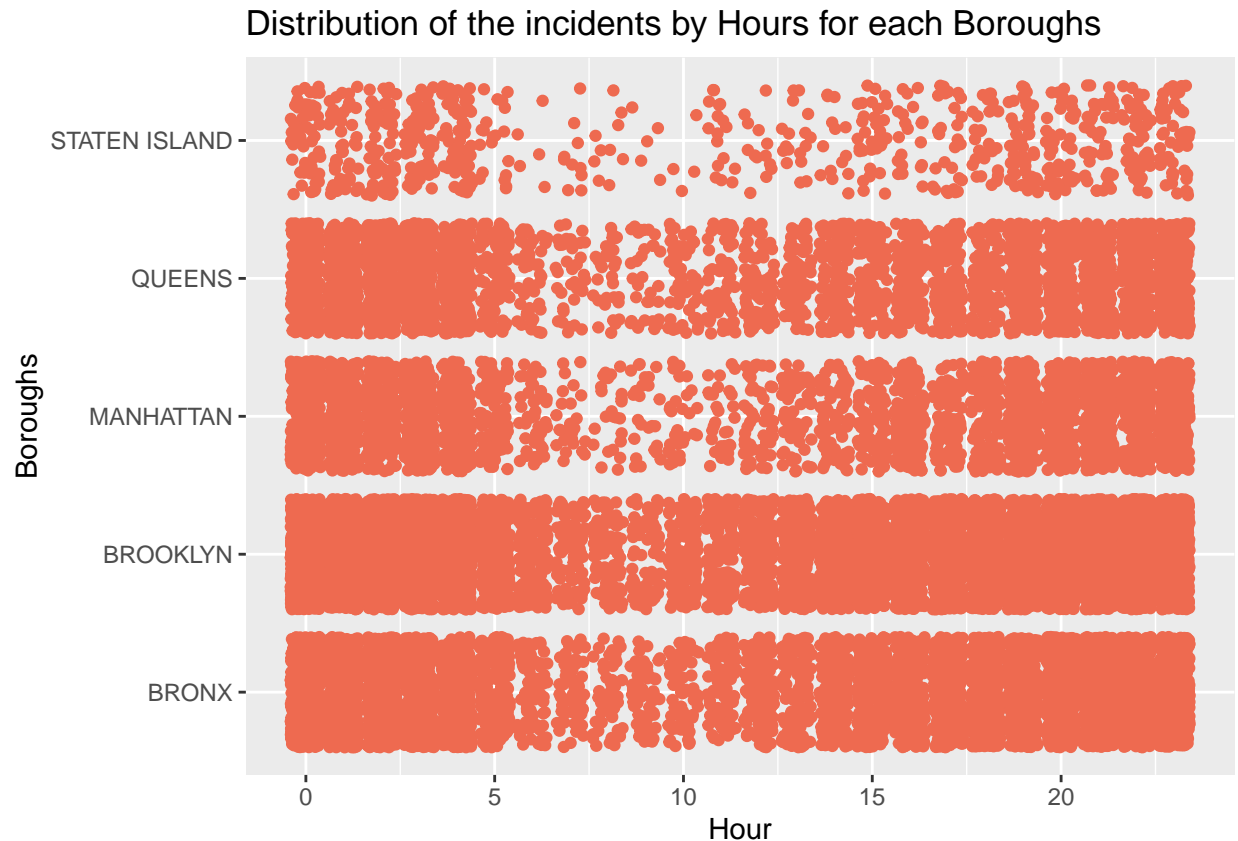
research question 3. What time should you not be out in New york City?

We can observe from the bar chart and jitter graph that the safest time is from 5:30 am to 12:30 pm. The dangerous period is from 9:30 pm to 2:30 am. Midnight is the most dangerous time to go out.

Step 4: Conclusion and Bias Identification

Conclusion:

Even though Brooklyn has the most shootings, certain areas are relatively safe with fewer than 100 shootings from the year of 2006 to 2022. Lower Manhattan and Downtown are relatively safe, where there have been less than 100 shootings from the year of 2006 to 2022. Midnight is a dangerous time go out. I have attached a map image of new york city by the number of precincts.

Bias:

My bias for this project is overusing bar charts for analysis and visualization. I mitigated that by using table and jitter graph for visualization and analysis rather than more bar chart.

A image for NYPD Police precincts

https://www.nyc.gov/site/nypd/bureaus/patrol/precincts-landing.page

Figure 1: NYPD Police Precincts

Source: New York City Department of Planning

14