

NYC Shooting Data Analysis

Daniel South

2025-03-04

Violent crime is prevalent in American cities. Violence impacts young people most profoundly, especially young men, who are eight times more likely to be shot than women. New Yorkers aged 18-44 experience the highest incidents of shootings for both men and women.

New York is the largest city in the United States. It's one of the safest American cities in terms of violent crime per capita, but with a population of more than 8 million, violent crime occurs frequently throughout the city's five boroughs.

The Bronx (Bronx County)

Brooklyn (Kings County)

Manhattan (New York County)

Queens (Queens County)

Staten Island (Richmond County)

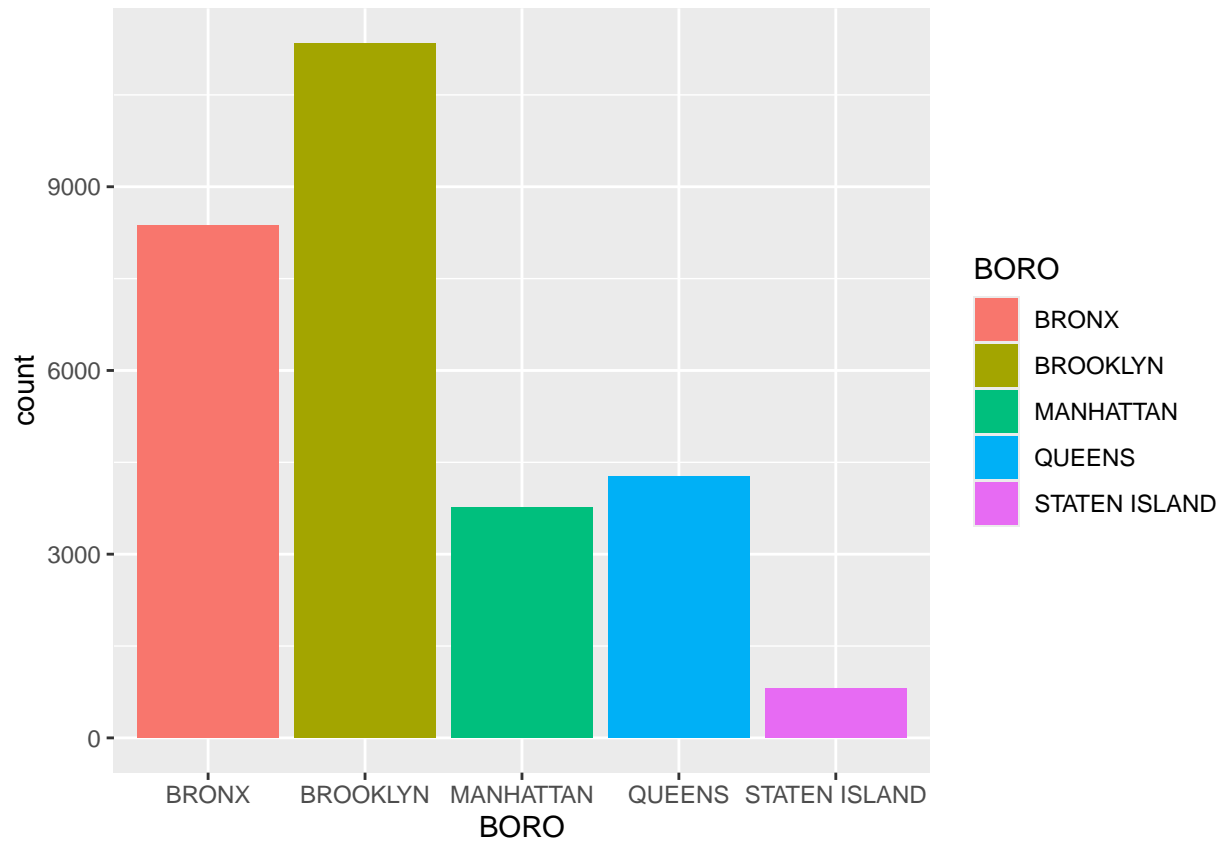
This report is based on the *NYPD_Shooting_Incident_Data_Historic_.csv* data set.

NOTE: This report is not meant to be representative of crime rates in all American cities.

Shootings by Borough (Categorical)

The graph below shows the number of shootings in each of New York City's five boroughs.

```
ggplot(df, aes(x = BORO, fill = BORO)) + geom_bar()
```



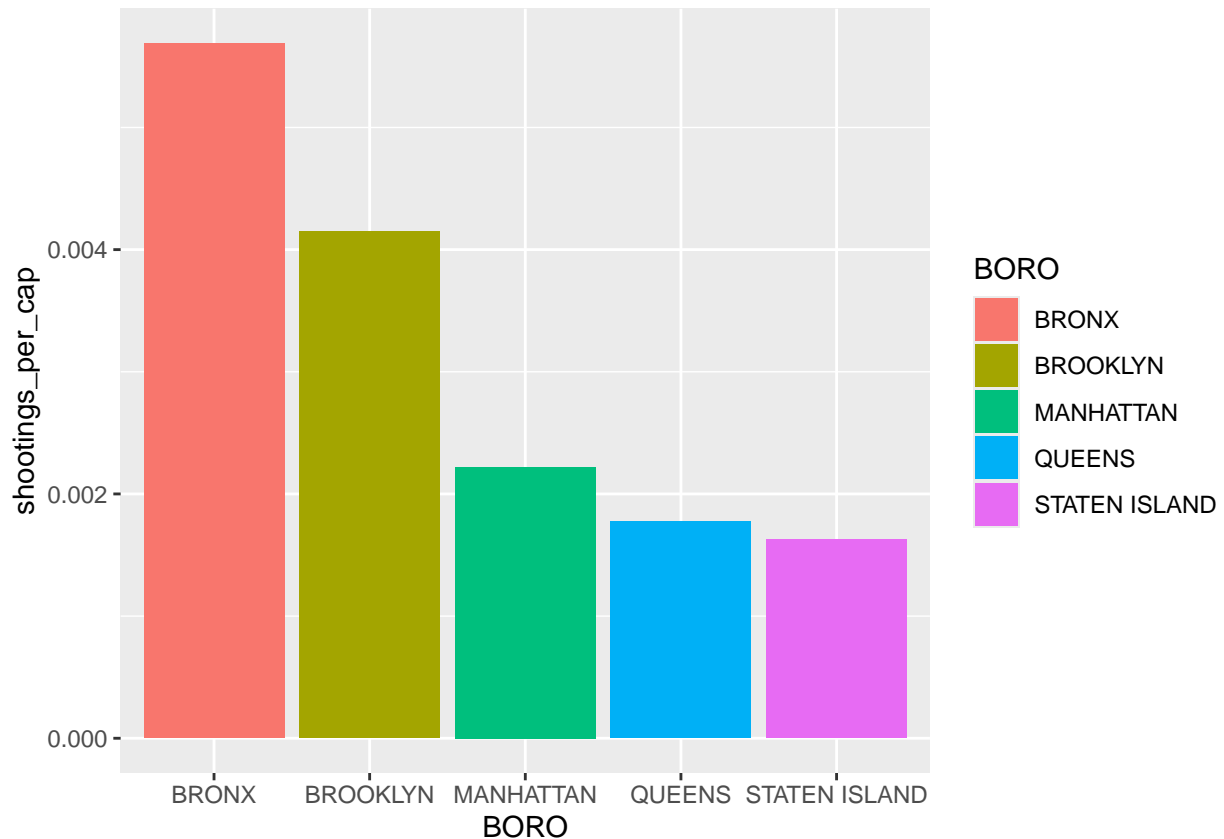
Shootings Per Capita By Borough

Brooklyn has the highest number of shootings, but it also has the largest population of any of the boroughs. The Bronx has nearly as many shootings incidents, but the population is about half that of Brooklyn, so the Bronx stands out as the borough with the highest number of shootings per capita.

Population figures courtesy of Wikipedia per the 2020 US Census.

```
df_pop = data.frame(BORO=c('BRONX', 'BROOKLYN', 'MANHATTAN', 'QUEENS', 'STATEN ISLAND'), shootings=c(8376, 10000, 2000, 1500, 1200))
df_pop$shootings_per_cap = df_pop$shootings / df_pop$pop

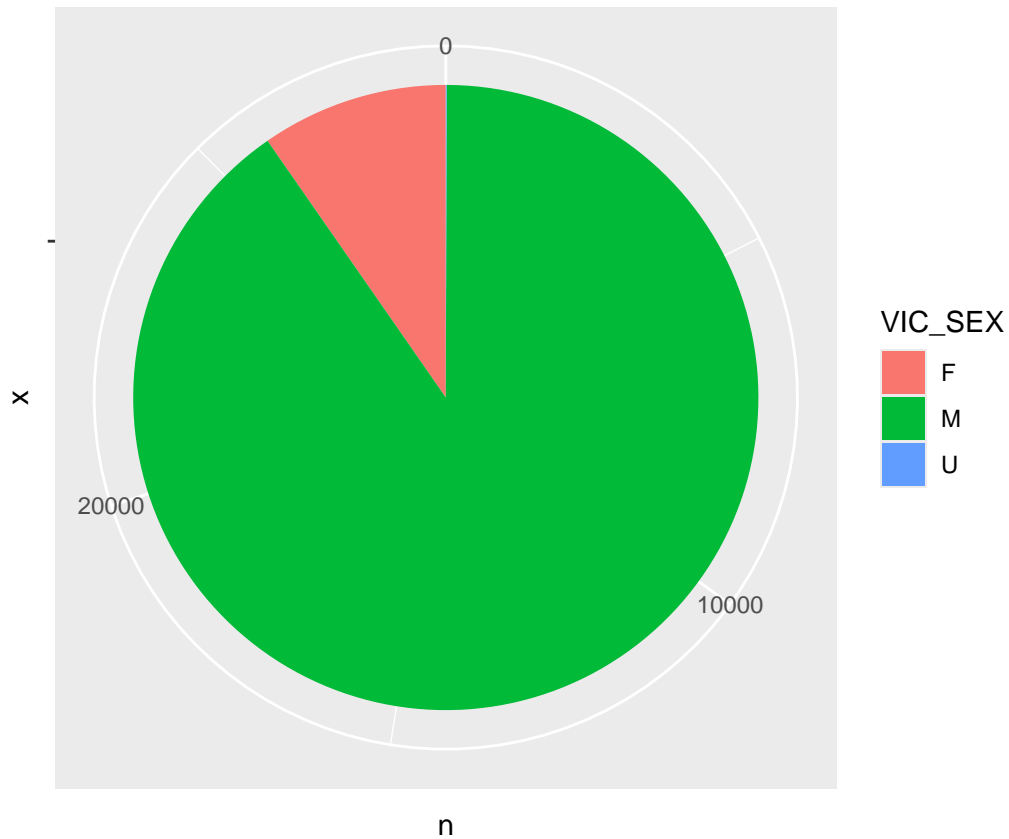
ggplot(df_pop, aes(x=BORO, y=shootings_per_cap, fill = BORO)) + geom_bar(stat="identity")
```



Shooting Victims by Sex

The vast majority of shooting victims in New York City are male.

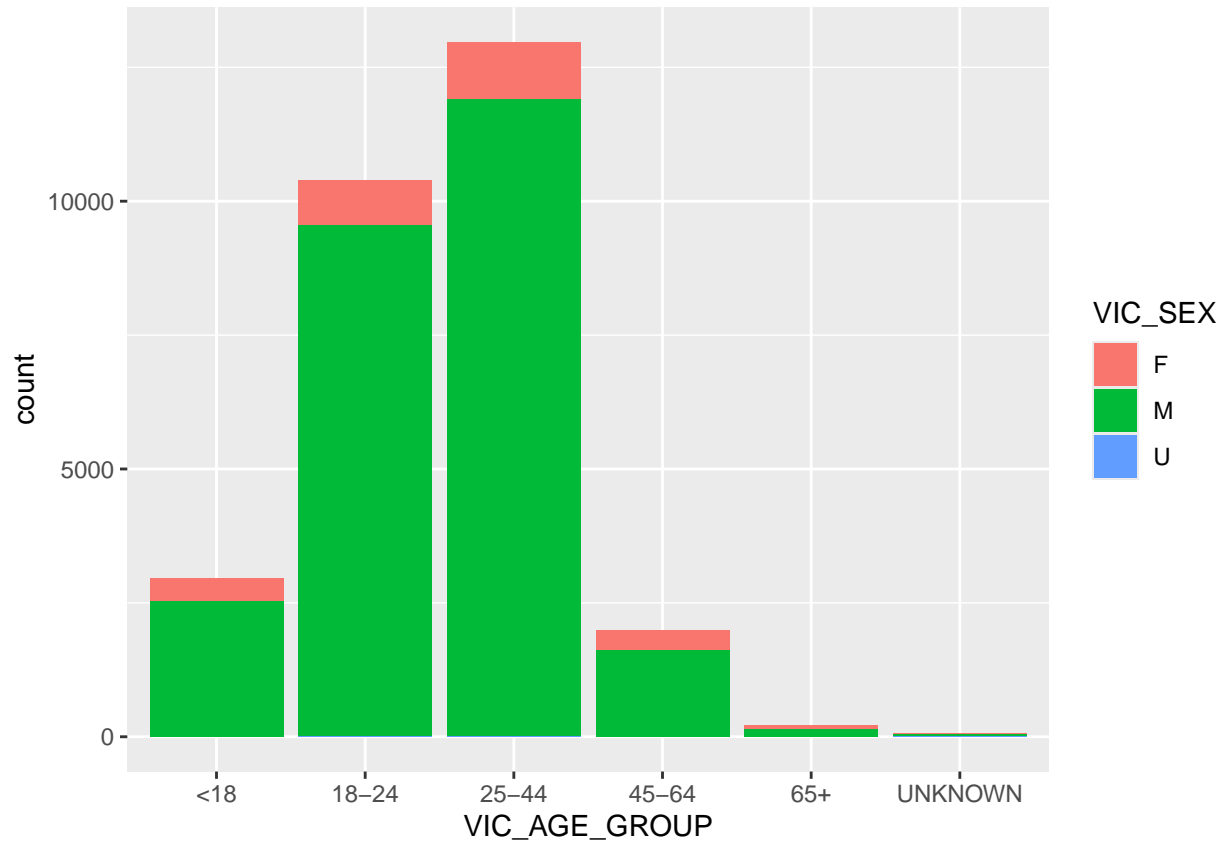
```
df2 = df %>%  
  group_by(VIC_SEX) %>%  
  count() %>%  
  ungroup()  
ggplot(df2, aes(x = "", y = n, fill = VIC_SEX)) + geom_col() + coord_polar(theta="y")
```



Victim by Age Group and Sex

Victim by age bracket further broken down by sex.

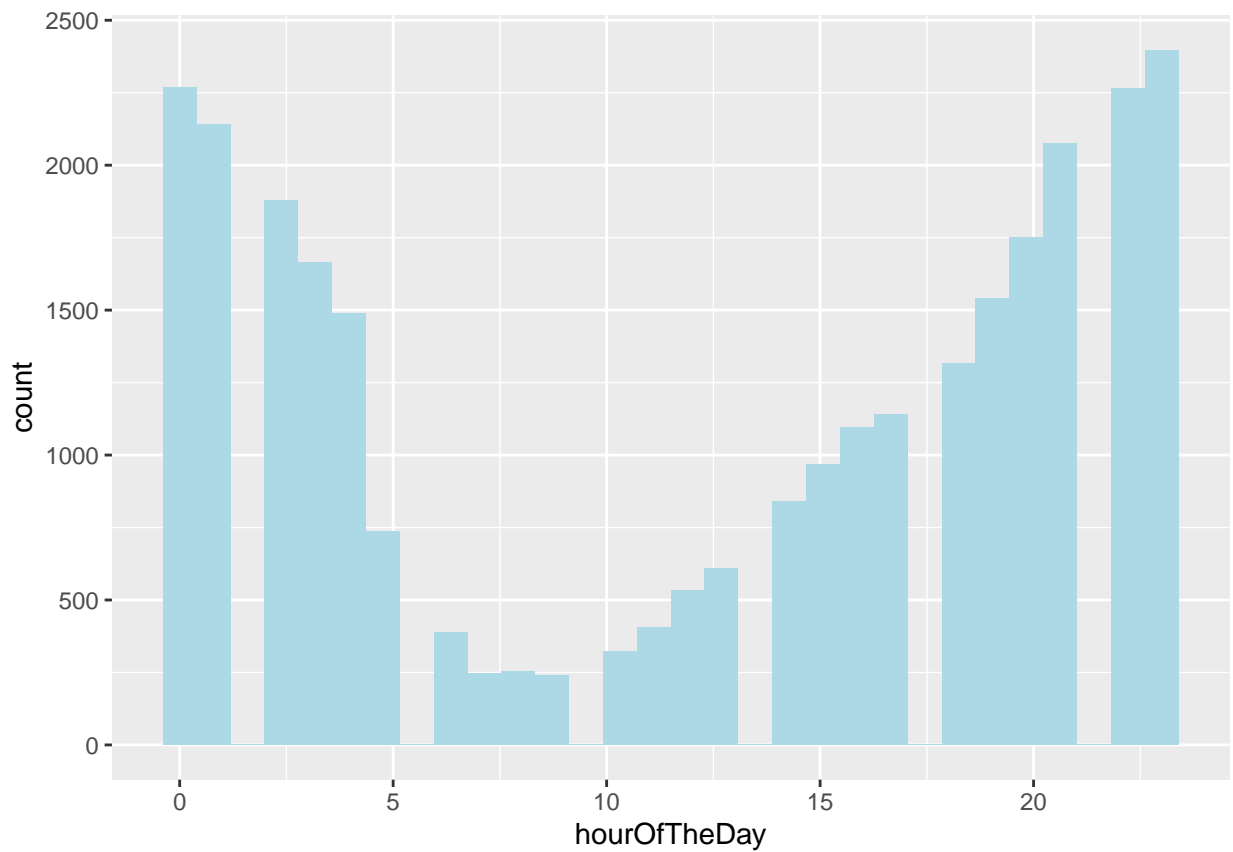
```
df4 = filter(df, VIC_AGE_GROUP!=1022)  
ggplot(df4, aes(x=VIC_AGE_GROUP, fill=VIC_SEX)) + geom_bar()
```



Shootings by Hour of the Day

Most shootings occur between 6 pm and 4 am (18:00 - 04:00).

```
df$hourOfTheDay = as.numeric(substr(df$OCCUR_TIME, 1, 2))  
ggplot(data = df, aes(x = hourOfTheDay)) + geom_histogram(fill = "lightblue")
```



Summary of the Data Set

```
summary(df)
```

```
## INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
## Min. : 9953245    Length:28562    Length:28562    Length:28562
## 1st Qu.: 65439914  Class :character  Class :character  Class :character
## Median : 92711254  Mode  :character  Mode  :character  Mode  :character
## Mean :127405824
## 3rd Qu.:203131993
## Max. :279758069
##
## LOC_OF_OCCUR_DESC  PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
## Length:28562      Min. : 1.0      Min. :0.0000      Length:28562
## Class :character  1st Qu.: 44.0   1st Qu.:0.0000      Class :character
## Mode :character   Median : 67.0   Median :0.0000      Mode :character
##                      Mean : 65.5      Mean :0.3219
##                      3rd Qu.: 81.0   3rd Qu.:0.0000
##                      Max. :123.0   Max. :2.0000
##                      NA's :2
## LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
## Length:28562      Length:28562      Length:28562
## Class :character   Class :character   Class :character
## Mode :character    Mode :character    Mode :character
##
##
##
## PERP_SEX           PERP_RACE           VIC_AGE_GROUP      VIC_SEX
## Length:28562      Length:28562      Length:28562      Length:28562
## Class :character   Class :character   Class :character   Class :character
## Mode :character    Mode :character    Mode :character    Mode :character
##
##
##
## VIC_RACE           X_COORD_CD           Y_COORD_CD          Latitude
## Length:28562      Min. : 914928      Min. :125757      Min. :40.51
## Class :character   1st Qu.:1000068    1st Qu.:182912    1st Qu.:40.67
## Mode :character    Median :1007772    Median :194901    Median :40.70
##                      Mean :1009424      Mean :208380      Mean :40.74
##                      3rd Qu.:1016807  3rd Qu.:239814    3rd Qu.:40.82
##                      Max. :1066815      Max. :271128      Max. :40.91
##                      NA's :59
##
## Longitude          Lon_Lat              hourOfTheDay
## Min. : -74.25      Length:28562      Min. : 0.00
## 1st Qu.: -73.94    Class :character   1st Qu.: 3.00
## Median : -73.92    Mode :character    Median :15.00
## Mean : -73.91      Mean :12.27
## 3rd Qu.: -73.88    3rd Qu.:20.00
## Max. : -73.70      Max. :23.00
## NA's :59
```