NYPD Shooting Project

2023-11-21

Introduction

Welcome to the Rmd documentation for the **NYPD Shooting** project.

In this project, I will be working with data NYPD Shootings (Historical) data-set in order to uncover which boroughs have had the most shootings from (2006 - 2023), and which victim race is the most common for all five boroughs. Our journey will involve data cleaning, data visualization, and modeling equipping us with a better understanding of shooting incidents that take place in the Big Apple.

Objective:

1. What is the trend of shooting incidents in NYC over time
2. Identify which borough has the most and least amount of shooting incidents
3. Identify the most frequent victim race

Data Source:

NYPD Shooting Incident Data (Historic) from 'https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD'

Data Analysis Techniques:

1. Data cleaning: Removing invalid or incomplete records, converting variables to appropriate data types
2. Data visualization: Creating charts and graphs to visualize findings

Expected Outcomes:

1. A comprehensive understanding of NYC shooting incidents
2. Identification of the boroughs and neighborhoods with the most shootings

Lets begin by importing the tidyverse

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.4      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(viridis)
```

```
## Loading required package: viridisLite
```

Importing the data using the link address and assiging it the variable 'data'

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

```
## Rows: 27312 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Data Preparation:

1. Summarizing our data
2. Identifying Missing Values and removing invalid or incomplete records
3. Confirming that each variable is the accurate datatype

Checking for missing values

```
print(colSums(is.na(data)))
```

```
##            INCIDENT_KEY             OCCUR_DATE             OCCUR_TIME
##                       0                      0                      0
##                    BORO      LOC_OF_OCCUR_DESC               PRECINCT
##                       0                  25596                      0
##       JURISDICTION_CODE     LOC_CLASSFCTN_DESC          LOCATION_DESC
##                       2                  25596                  14977
## STATISTICAL_MURDER_FLAG         PERP_AGE_GROUP               PERP_SEX
##                       0                   9344                   9310
##               PERP_RACE          VIC_AGE_GROUP               VIC_SEX
##                    9310                      0                      0
##                VIC_RACE             X_COORD_CD             Y_COORD_CD
##                       0                      0                      0
##                Latitude              Longitude                Lon_Lat
##                      10                     10                     10
```

Summerizing the data

```
summary(data)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:27312       Length:27312       Length:27312
##  1st Qu.: 63860880   Class :character   Class1:hms         Class :character
```

```
##  Median : 90372218   Mode  :character   Class2:difftime   Mode  :character
##  Mean   :120860536                      Mode  :numeric
##  3rd Qu.:188810230
##  Max.   :261190187
##
##  LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:27312       Min.   :  1.00   Min.   :0.0000    Length:27312
##  Class :character   1st Qu.: 44.00   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 68.00   Median :0.0000    Mode  :character
##                     Mean   : 65.64   Mean   :0.3269
##                     3rd Qu.: 81.00   3rd Qu.:0.0000
##                     Max.   :123.00   Max.   :2.0000
##                                      NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:27312       Mode :logical           Length:27312
##  Class :character   FALSE:22046             Class :character
##  Mode  :character   TRUE :5266              Mode  :character
##
##
##
##
##    PERP_SEX           PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:27312       Length:27312       Length:27312       Length:27312
##  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:27312       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000028   1st Qu.:182834   1st Qu.:40.67
##  Mode  :character   Median :1007731   Median :194487   Median :40.70
##                     Mean   :1009449   Mean   :208127   Mean   :40.74
##                     3rd Qu.:1016838   3rd Qu.:239518   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :10
##    Longitude        Lon_Lat
##  Min.   :-74.25   Length:27312
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :10
```

Geeting rid of the columns:

X_COORD_CD

Y_COORD_CD

Latitude

Lon_Lat

LOC_OF_OCCUR_DESC

LOC_CLASSFCTN_DESC

Longitude

```
columns_removed <- c("X_COORD_CD", "Y_COORD_CD", "Latitude", "Lon_Lat", "LOC_OF_OCCUR_DESC", "LOC_CLASSI
nypd_data <- data[, !names(data) %in% columns_removed]
```

Confirming that the columns were removed

```
column_names <- names(nypd_data)
print(column_names)
```

```
##  [1] "INCIDENT_KEY"            "OCCUR_DATE"
##  [3] "OCCUR_TIME"              "BORO"
##  [5] "PRECINCT"                "JURISDICTION_CODE"
##  [7] "LOCATION_DESC"           "STATISTICAL_MURDER_FLAG"
##  [9] "PERP_AGE_GROUP"          "PERP_SEX"
## [11] "PERP_RACE"               "VIC_AGE_GROUP"
## [13] "VIC_SEX"                 "VIC_RACE"
```

Changing "OCCUR_Date" into (datetime) datatype

```
nypd_data$OCCUR_DATE <- as.Date(nypd_data$OCCUR_DATE, format = "%m/%d/%Y")
```

Confirming that the OCCUR_Date is a datetime datatype

```
str(nypd_data)
```

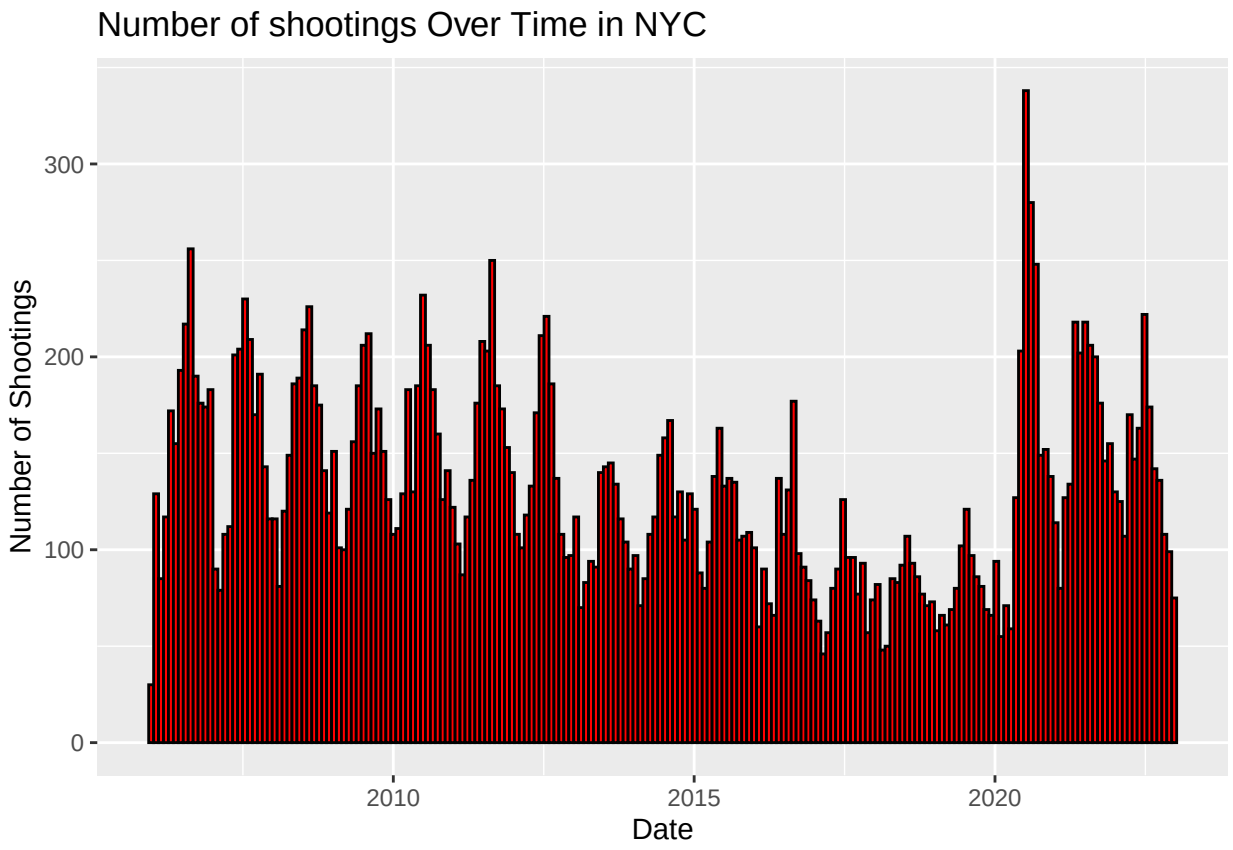
```
## tibble [27,312 x 14] (S3: tbl_df/tbl/data.frame)
##  $ INCIDENT_KEY           : num [1:27312] 2.29e+08 1.37e+08 1.48e+08 1.47e+08 5.89e+07 ...
##  $ OCCUR_DATE             : Date[1:27312], format: "2021-05-27" "2014-06-27" ...
##  $ OCCUR_TIME             : 'hms' num [1:27312] 21:30:00 17:40:00 03:56:00 18:30:00 ...
##   ..- attr(*, "units")= chr "secs"
##  $ BORO                   : chr [1:27312] "QUEENS" "BRONX" "QUEENS" "BRONX" ...
##  $ PRECINCT               : num [1:27312] 105 40 108 44 47 81 114 81 105 101 ...
##  $ JURISDICTION_CODE      : num [1:27312] 0 0 0 0 0 0 0 0 0 0 ...
##  $ LOCATION_DESC          : chr [1:27312] NA NA NA NA ...
##  $ STATISTICAL_MURDER_FLAG: logi [1:27312] FALSE FALSE TRUE FALSE TRUE TRUE ...
##  $ PERP_AGE_GROUP         : chr [1:27312] NA NA NA NA ...
##  $ PERP_SEX               : chr [1:27312] NA NA NA NA ...
##  $ PERP_RACE              : chr [1:27312] NA NA NA NA ...
##  $ VIC_AGE_GROUP          : chr [1:27312] "18-24" "18-24" "25-44" "<18" ...
##  $ VIC_SEX                : chr [1:27312] "M" "M" "M" "M" ...
##  $ VIC_RACE               : chr [1:27312] "BLACK" "BLACK" "WHITE" "WHITE HISPANIC" ...
```

2. Data visualization: Creating charts and graphs to visualize findings

Ploting shootings over time

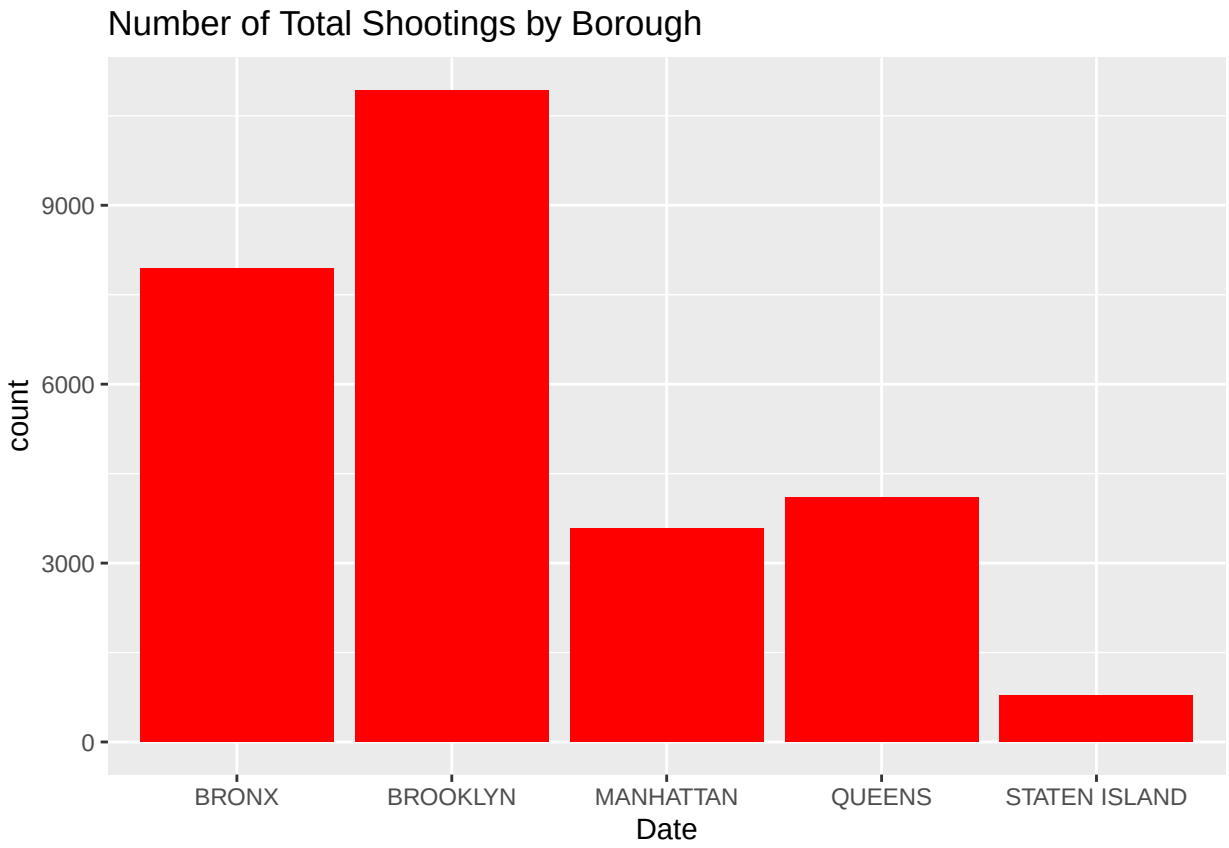
```
nypd_data$OCCUR_DATE <- as.Date(nypd_data$OCCUR_DATE)
```

```r
nypd_data %>%
  ggplot(aes(x = OCCUR_DATE)) +
  geom_histogram(binwidth = 30, fill = "red", color = "black") +
  labs(title = "Number of shootings Over Time in NYC",
       x = "Date",
       y = "Number of Shootings")
```



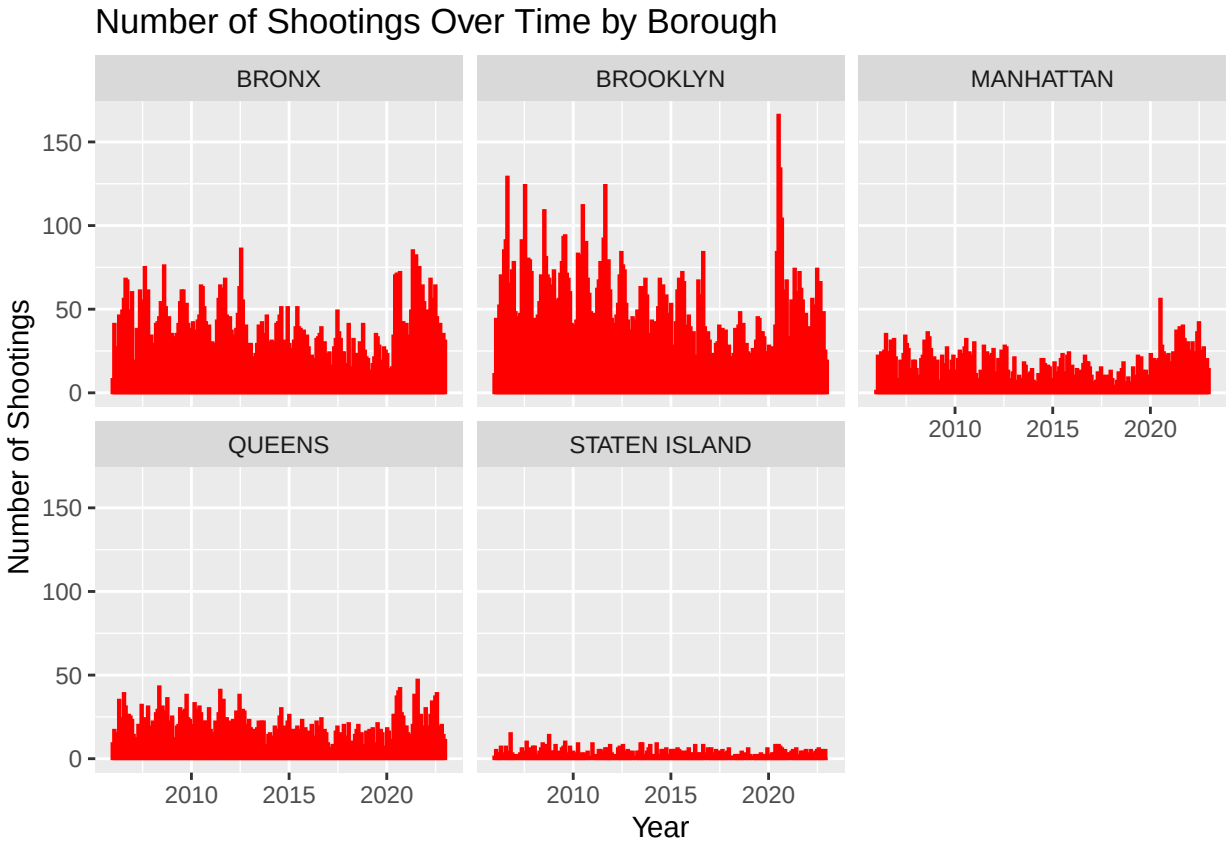
Total Shootings by Borough

```r
nypd_data %>%
  ggplot(aes(x = BORO)) +
  geom_bar(fill = "red") +
  labs(title = "Number of Total Shootings by Borough ",
       x = "Date",
       Y = "Total Shootings")
```



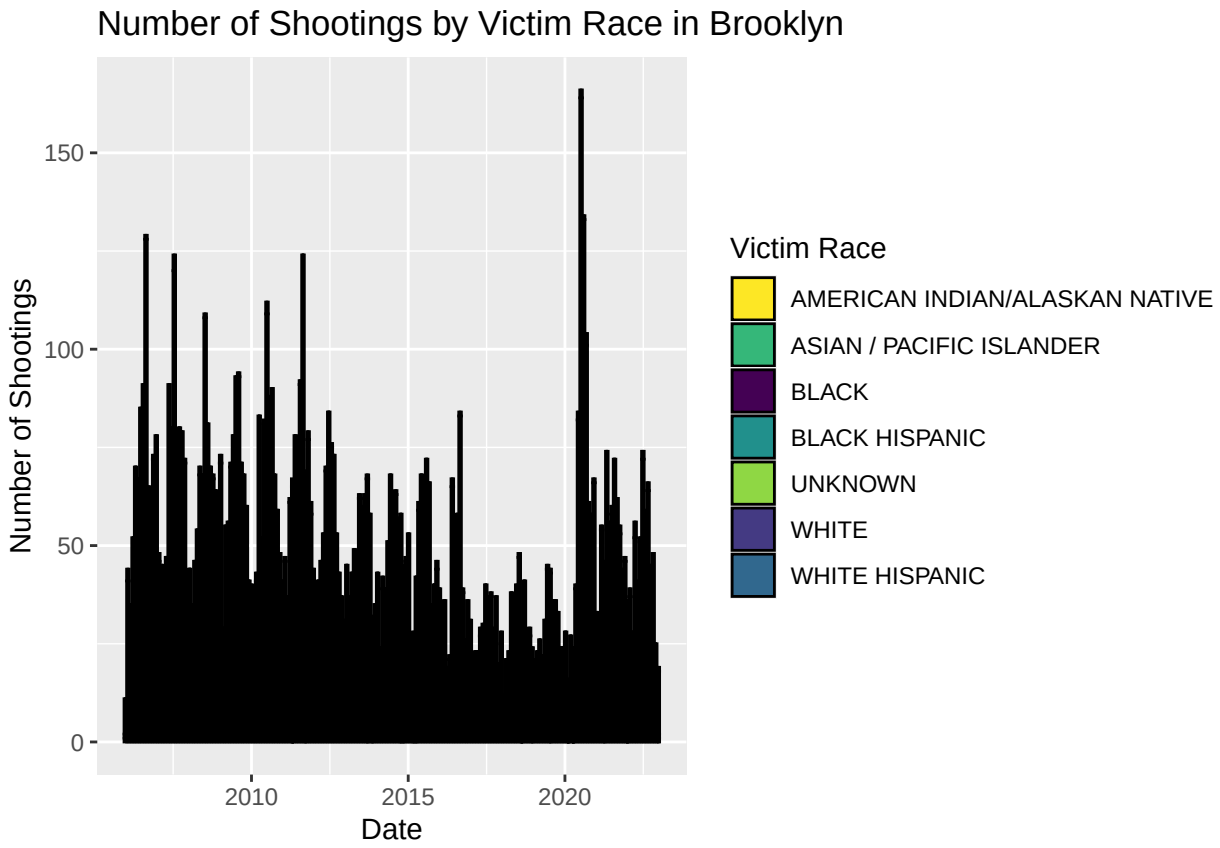
Shootings Over Time by Borough

```
nypd_data %>%
  ggplot(aes(x = OCCUR_DATE)) +
  geom_histogram(binwidth = 30, fill = "black", color = "red") +
  facet_wrap(~ BORO) +
  labs(title = "Number of Shootings Over Time by Borough ",
       x = "Year",
       y = "Number of Shootings")
```



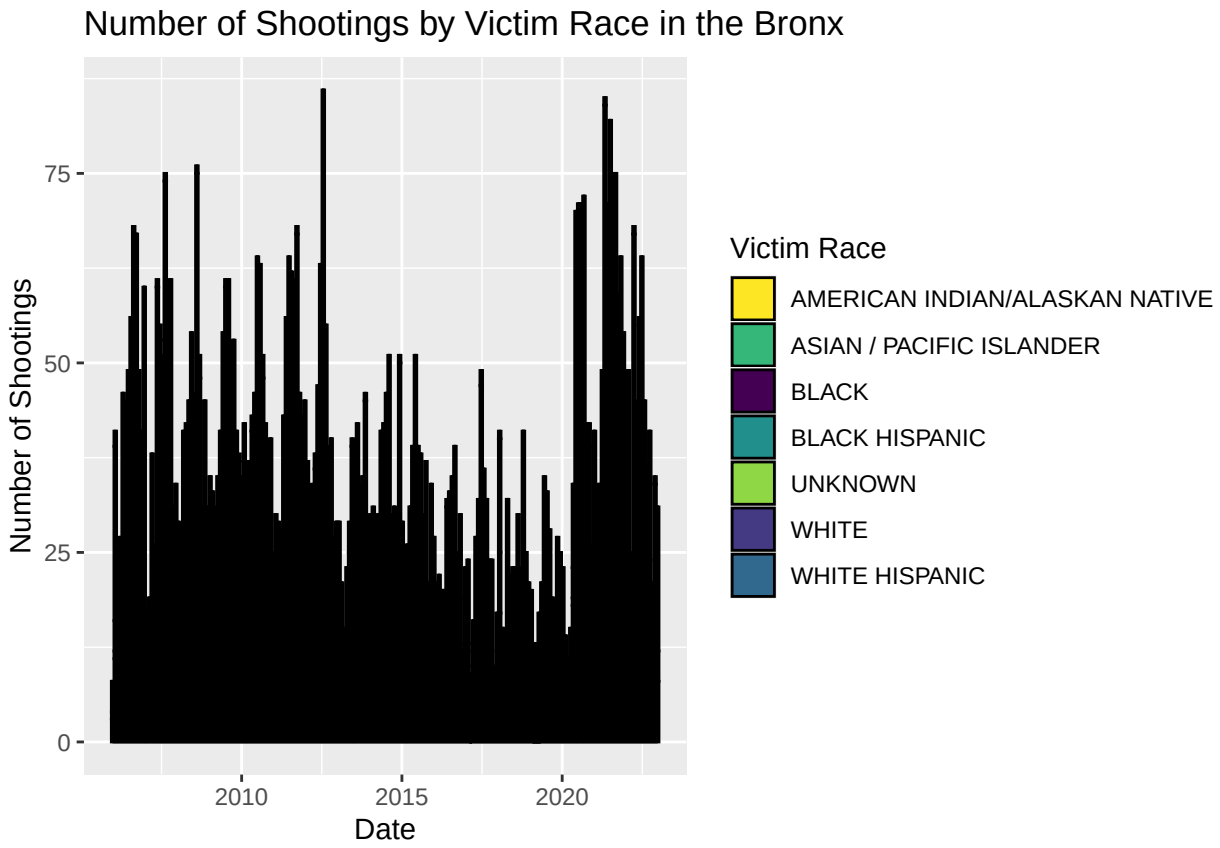
Number of Shootings Over Time by Victim Race In Brooklyn

```
unique_vic_race <- unique(nypd_data$VIC_RACE)
colors <- viridis_pal()(length(unique_vic_race))
brooklyn_data <- nypd_data %>% filter(BORO == "BROOKLYN")
ggplot(brooklyn_data, aes(x = OCCUR_DATE, fill = VIC_RACE)) +
  geom_histogram(binwidth = 30, color = "black", position = "stack") +
  labs(title = "Number of Shootings by Victim Race in Brooklyn",
       x = "Date",
       y = "Number of Shootings",
       fill = "Victim Race") +
  scale_fill_manual(values = setNames(colors, unique_vic_race))
```



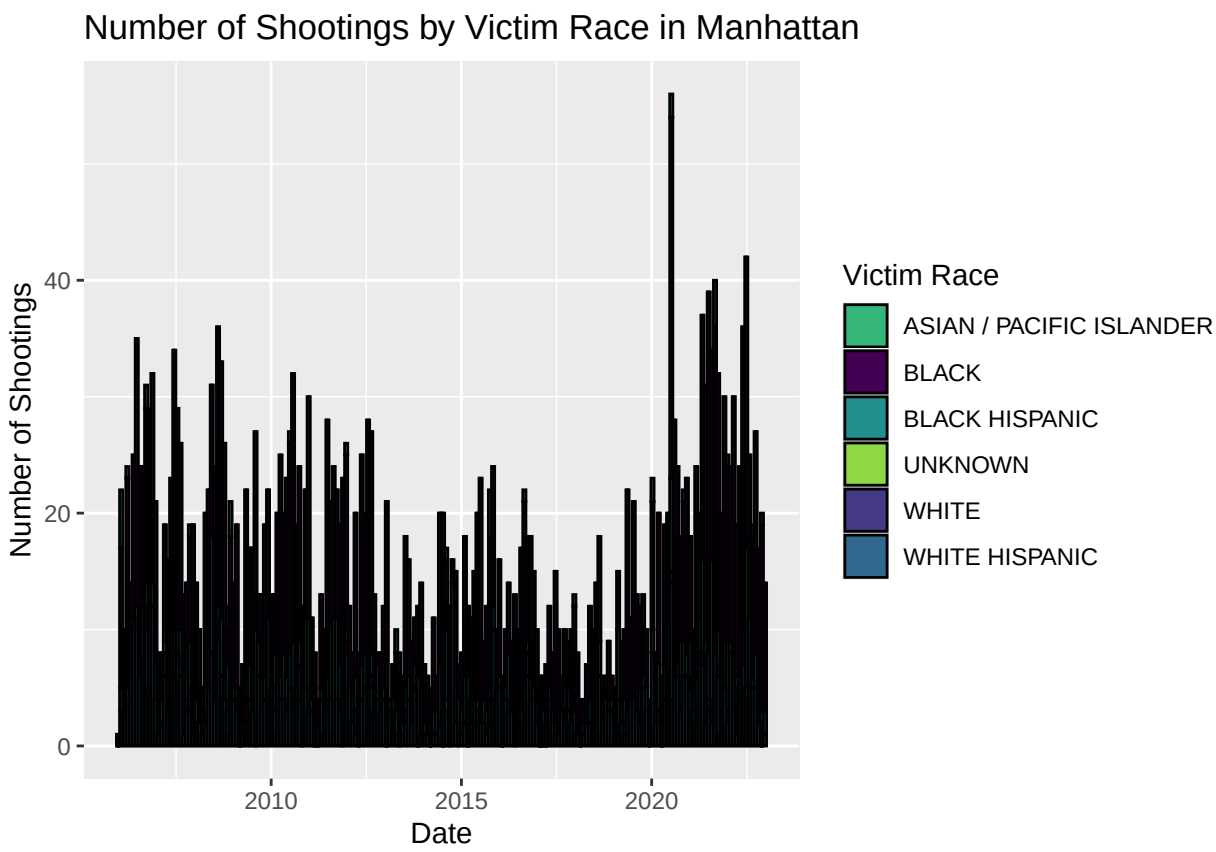
Number of Shootings Over Time by Victim Race In the Bronx

```
unique_vic_race <- unique(nypd_data$VIC_RACE)
colors <- viridis_pal()(length(unique_vic_race))
brooklyn_data <- nypd_data %>% filter(BORO == "BRONX")
ggplot(brooklyn_data, aes(x = OCCUR_DATE, fill = VIC_RACE)) +
  geom_histogram(binwidth = 30, color = "black", position = "stack") +
  labs(title = "Number of Shootings by Victim Race in the Bronx",
       x = "Date",
       y = "Number of Shootings",
       fill = "Victim Race") +
  scale_fill_manual(values = setNames(colors, unique_vic_race))
```



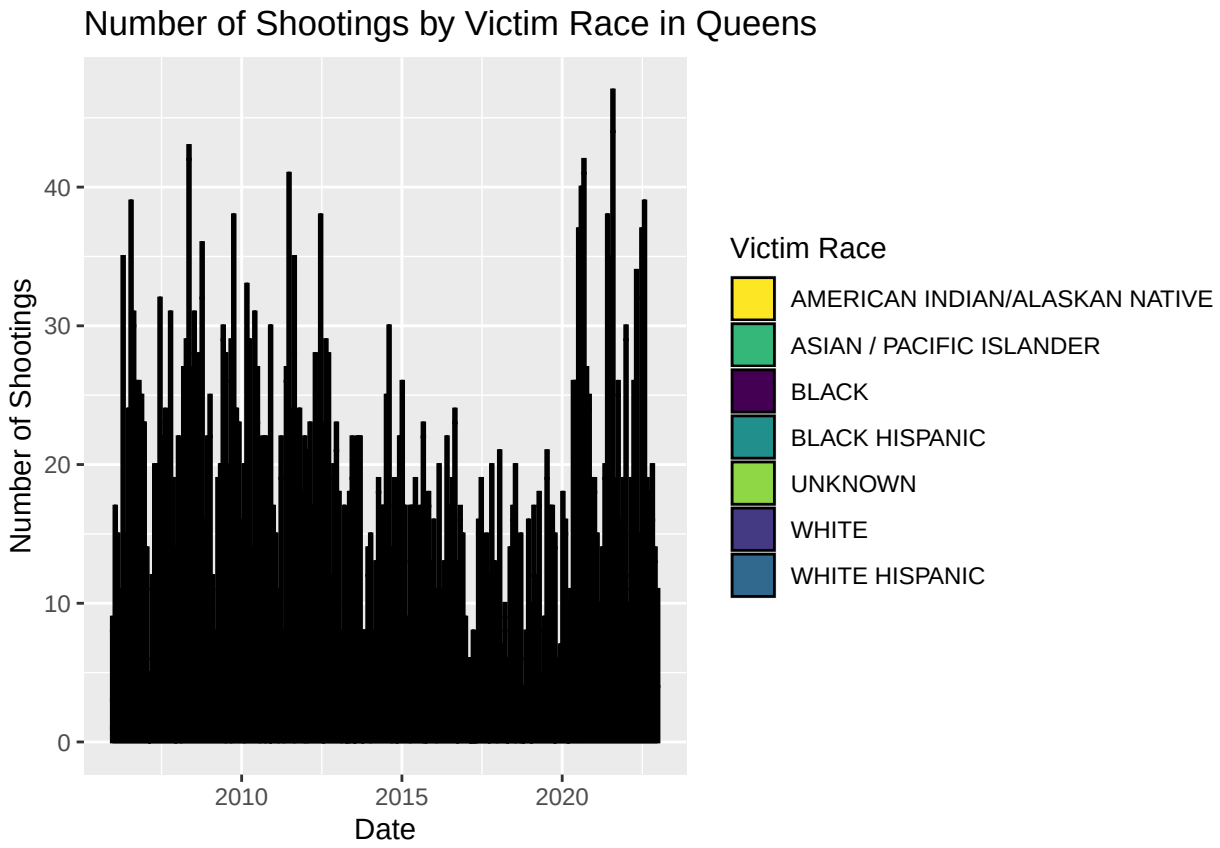
Number of Shootings Over Time by Victim Race In Manhattan

```
unique_vic_race <- unique(nypd_data$VIC_RACE)
colors <- viridis_pal()(length(unique_vic_race))
brooklyn_data <- nypd_data %>% filter(BORO == "MANHATTAN")
ggplot(brooklyn_data, aes(x = OCCUR_DATE, fill = VIC_RACE)) +
  geom_histogram(binwidth = 30, color = "black", position = "stack") +
  labs(title = "Number of Shootings by Victim Race in Manhattan",
       x = "Date",
       y = "Number of Shootings",
       fill = "Victim Race") +
  scale_fill_manual(values = setNames(colors, unique_vic_race))
```



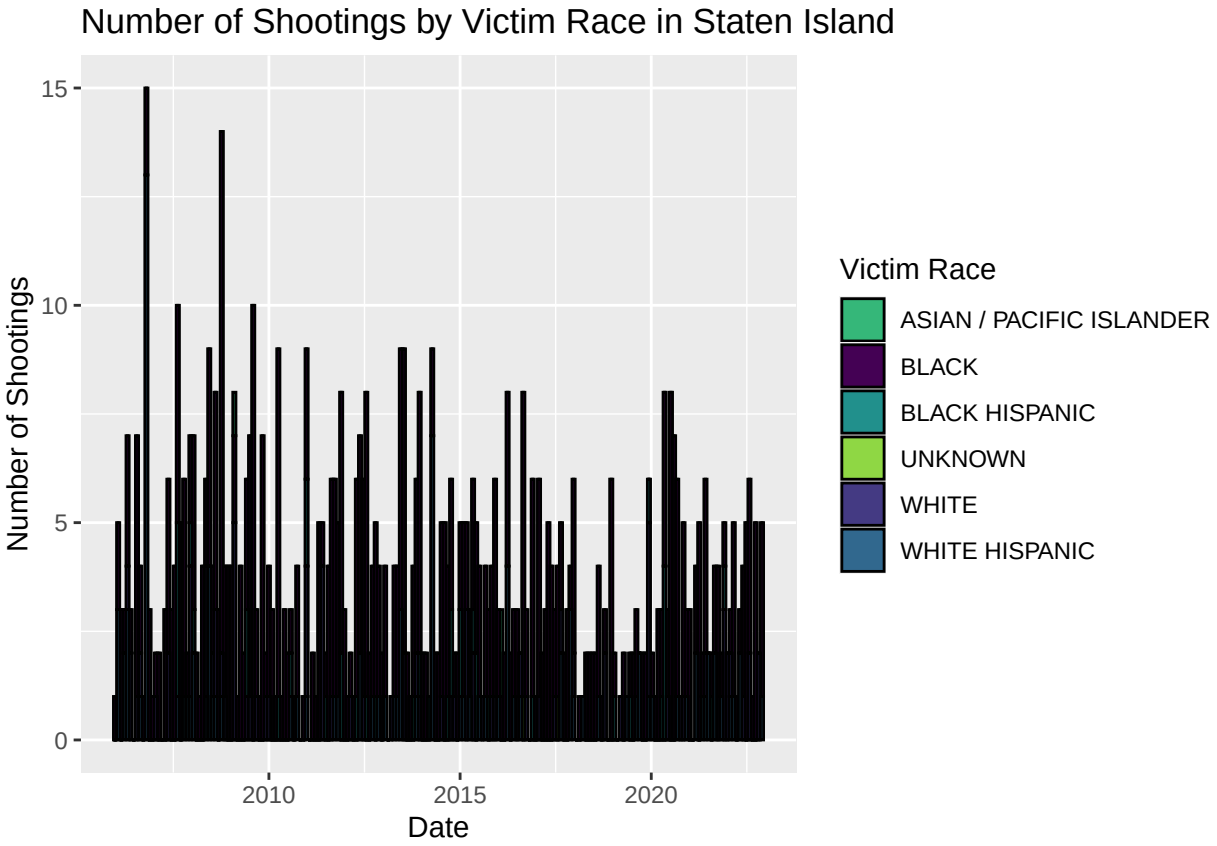
Number of Shootings Over Time by Victim Race In Queens

```
unique_vic_race <- unique(nypd_data$VIC_RACE)
colors <- viridis_pal()(length(unique_vic_race))
brooklyn_data <- nypd_data %>% filter(BORO == "QUEENS")
ggplot(brooklyn_data, aes(x = OCCUR_DATE, fill = VIC_RACE)) +
  geom_histogram(binwidth = 30, color = "black", position = "stack") +
  labs(title = "Number of Shootings by Victim Race in Queens",
       x = "Date",
       y = "Number of Shootings",
       fill = "Victim Race") +
  scale_fill_manual(values = setNames(colors, unique_vic_race))
```



Number of Shootings Over Time by Victim Race In Staten Island

```
unique_vic_race <- unique(nypd_data$VIC_RACE)
colors <- viridis_pal()(length(unique_vic_race))
brooklyn_data <- nypd_data %>% filter(BORO == "STATEN ISLAND")
ggplot(brooklyn_data, aes(x = OCCUR_DATE, fill = VIC_RACE)) +
  geom_histogram(binwidth = 30, color = "black", position = "stack") +
  labs(title = "Number of Shootings by Victim Race in Staten Island",
       x = "Date",
       y = "Number of Shootings",
       fill = "Victim Race") +
  scale_fill_manual(values = setNames(colors, unique_vic_race))
```



#Summary: This R markdown presents my findings of my analysis of the NYPD Shooting Incident data-set. The goal of my analysis was to uncover the trends and patterns of shootings in New York City (NYC) from 2006 - 2023, with a specific focus on all five boroughs and victim race demographics.

Biases:

Because I live near NYC and wanted to avoid any favoritism, I focused on all five boroughs rather than analyzing a single one.

A total of eight graphs were created but only six will be used to visualize my findings.

Number of Shootings Over Time in NYC: This graph visualizes the number of shootings from (2006 - 2023) for NYC as a whole.

Number of Total Shootings by Borough: This graph ranks the five boroughs by the total number of shooting incidents from (2006 - 2023).

Number of Shootings by Victims Race: This graph shows the number of shootings for each victim race category in Queens, Staten Island, the Bronx, and Manhattan from (2006 - 2023).

Findings:

The Number of Shootings Over Time in NYC graph shows that the number of shootings increased significantly during the middle of the COVID-19 pandemic (2020-2021) compared to the previous years (2006 - 2019).

The Number of Total Shootings by Borough graph reveals that Brooklyn has the highest number of shooting incidents, followed by the Bronx, Queens, Manhattan, and Staten Island in decreasing order.

The Number of Shootings by Victims Race graphs for Queens, Staten Island, the Bronx, and Manhattan all show similar patterns, with "Black" victims being the most frequent targets and "American Indian/Alaskan Native" victims being the least frequent.

My findings reveal that the number of shootings increased during the COVID-19 pandemic and that Brooklyn has the highest number of shooting incidents out of the five boroughs. Additionally, Black individuals are disproportionately affected by shootings. Further analysis would be needed inorder to truly understand what the true causes of these incidents are and why Black individuals are the most affected by these shooting incidents.

To offer a deeper understanding of my findings, I have prepared accompanying PowerPoints showcasing the previously presented visualizations. These slides offer a better view of my findings. In addition you'll be able to fully view the stacked bar charts for the number of shootings by victim race for each borough .

Thank you.

Link to Graphs PDF

file:///Users/fritzathis/Desktop/NYPD-%20Shooting%20-%20Presentation.pdf