NYPD Shooting Incident Data Report

Table of Contents

Introduction

The information determines the functional execution of the shooting incidents by the NYPD in New York City. This determines the execution of the collected data which is found on the website of the NYPD. This data contains the details of the shooting from the year 2006. The R-programming concept can be implemented to determine the relational evaluation between various locations in New York City where the incidents are found.

Finding & Analysis

```
# install.packages("tidyverse")
library(tidyverse)
library(lubridate)
```

Figure 1: Libraries implementation

(Source: Provided)

The libraries of the R-programming are applicable to make a suitable test bench where the testing of the data can be implemented. This determines the involvement of two libraries such as 'tidyverse', and 'lubridate'. In this case 'tidyverse' implements 'Tudy' which is applicable to implement some necessary functions such as summation, and so on. This also determines the implementation of the pipelining operator and approach ('%>%') which is implemented to derive some functional determination. On the other hand, 'libridate' is implemented to make a date value column using a list of date columns.

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


##
## -- Column specification --------------------------------------------------------
## cols(
##   INCIDENT_KEY = col_double(),
##   OCCUR_DATE = col_character(),
##   OCCUR_TIME = col_time(format = ""),
##   BORO = col_character(),
##   PRECINCT = col_double(),
##   JURISDICTION_CODE = col_double(),
##   LOCATION_DESC = col_character(),
##   STATISTICAL_MURDER_FLAG = col_logical(),
##   PERP_AGE_GROUP = col_character(),
##   PERP_SEX = col_character(),
##   PERP_RACE = col_character(),
##   VIC_AGE_GROUP = col_character(),
##   VIC_SEX = col_character(),
```

Figure 2: Reading of collected data

(Source: Provided)

The reading approach of the collected data is determined using the 'read_csv' approach. This method is implemented to collect or download the necessary data from a data source (Silva, 2022). The data is then assigned to a data frame which is named 'df'.

```
head(df)
```

```
## # A tibble: 6 x 19
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO          PRECINCT JURISDICTION_CODE
##          <dbl> <chr>      <time>     <chr>            <dbl>              <dbl>
## 1    201575314 08/23/2019 22:10      QUEENS             103                  0
## 2    205748546 11/27/2019 15:54      BRONX               40                  0
## 3    193118596 02/02/2019 19:40      MANHATTAN           23                  0
## 4    204192600 10/24/2019 00:52      STATEN ISLAND      121                  0
## 5    201483468 08/22/2019 18:03      BRONX               46                  0
## 6    198255460 06/07/2019 17:50      BROOKLYN            73                  0
## # ... with 13 more variables: LOCATION_DESC <chr>,
## #   STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>, PERP_SEX <chr>,
## #   PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>, VIC_RACE <chr>,
## #   X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>, Longitude <dbl>,
## #   Lon_Lat <chr>
```

Figure 3: Details of the data

(Source: Provided)

The data details are derived by using the 'head' mechanism. This determines the involvement of the character, string, and integer value type data. The determination also highlights the name of the columns which is present in the collected data.

```
df_2 = df %>% select(INCIDENT_KEY,
                     OCCUR_DATE,
                     OCCUR_TIME,
                     BORO,
                     STATISTICAL_MURDER_FLAG,
                     PERP_AGE_GROUP,
                     PERP_SEX,
                     PERP_RACE,
                     VIC_AGE_GROUP,
                     VIC_SEX,
                     VIC_RACE,
                     Latitude,
                     Longitude)

# Return the column name along with the missing values
lapply(df_2, function(x) sum(is.na(x)))
```

```
## $INCIDENT_KEY
## [1] 0
##
## $OCCUR_DATE
## [1] 0
```

Figure 4: Missing value check

(Source: Provided)

The missing value checking approach is applicable to find out which column has the blank section. This is determined by using the 'select' approach with the assistance of 'tidy'. The sum of 'is.na' approach is implemented to determine the presence of a blank in the data.

```
# Tidy and transform data
df_2 = df_2 %>%
  replace_na(list(PERP_AGE_GROUP = "Unknown", PERP_SEX = "Unknown", PERP_RACE = "Unknown"))
```

Figure 5: Data replacement

(Source: Provided)

The data replacement approach is implemented to transform the data section into a specific data type. The 'replace_na' approach is applicable to fill up the blank section with 'Unknown' string value data.

```
# Remove extreme values in data
df_2 = subset(df_2, PERP_AGE_GROUP!="1020" & PERP_AGE_GROUP!="224" & PERP_AGE_GROUP!="940")

df_2$PERP_AGE_GROUP = recode(df_2$PERP_AGE_GROUP, UNKNOWN = "Unknown")
df_2$PERP_SEX = recode(df_2$PERP_SEX, U = "Unknown")
df_2$PERP_RACE = recode(df_2$PERP_RACE, UNKNOWN = "Unknown")
df_2$VIC_SEX   = recode(df_2$VIC_SEX, U = "Unknown")
df_2$VIC_RACE   = recode(df_2$VIC_RACE, UNKNOWN = "Unknown")
df_2$INCIDENT_KEY = as.character(df_2$INCIDENT_KEY)
df_2$BORO = as.factor(df_2$BORO)
df_2$PERP_AGE_GROUP = as.factor(df_2$PERP_AGE_GROUP)
df_2$PERP_SEX = as.factor(df_2$PERP_SEX)
df_2$PERP_RACE = as.factor(df_2$PERP_RACE)
df_2$VIC_AGE_GROUP = as.factor(df_2$VIC_AGE_GROUP)
df_2$VIC_SEX = as.factor(df_2$VIC_SEX)
df_2$VIC_RACE = as.factor(df_2$VIC_RACE)

# Return summary statistics
summary(df_2)
```

```
##  INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME                   BORO
##  Length:23565       Length:23565       Length:23565      BRONX        :6698
##  Class :character   Class :character   Class1:hms        BROOKLYN     :9721
##  Mode  :character   Mode  :character   Class2:difftime   MANHATTAN    :2921
##                                        Mode  :numeric    QUEENS       :3527
##                                                          STATEN ISLAND: 698
```

Figure 6: Removal of unused parts from the data

(Source: Provided)

The removal approach is the part of cleaning mechanism. This determines the summary of the statistical value of the data. The sectional determination provides the removal of the extreme values that are present in the main data section (Mezmir, 2020).

```
g <- ggplot(df_2, aes(x = BORO)) +
  geom_bar() +
  labs(title = "Boroughs of New York City",
       x = "Boroughs of New York City",
       y = "Count of Incidents") +
  theme_minimal()
g
```

Figure 7: Boroughs of New York City Plotting using ggplot

(Source: Provided)

The evaluated approach determines the Boroughs of New York City by using the bar plot concept. This introduces the section of x, and y where x determines the 'Boroughs of New York City', and y determines the 'Count of Incidents'.

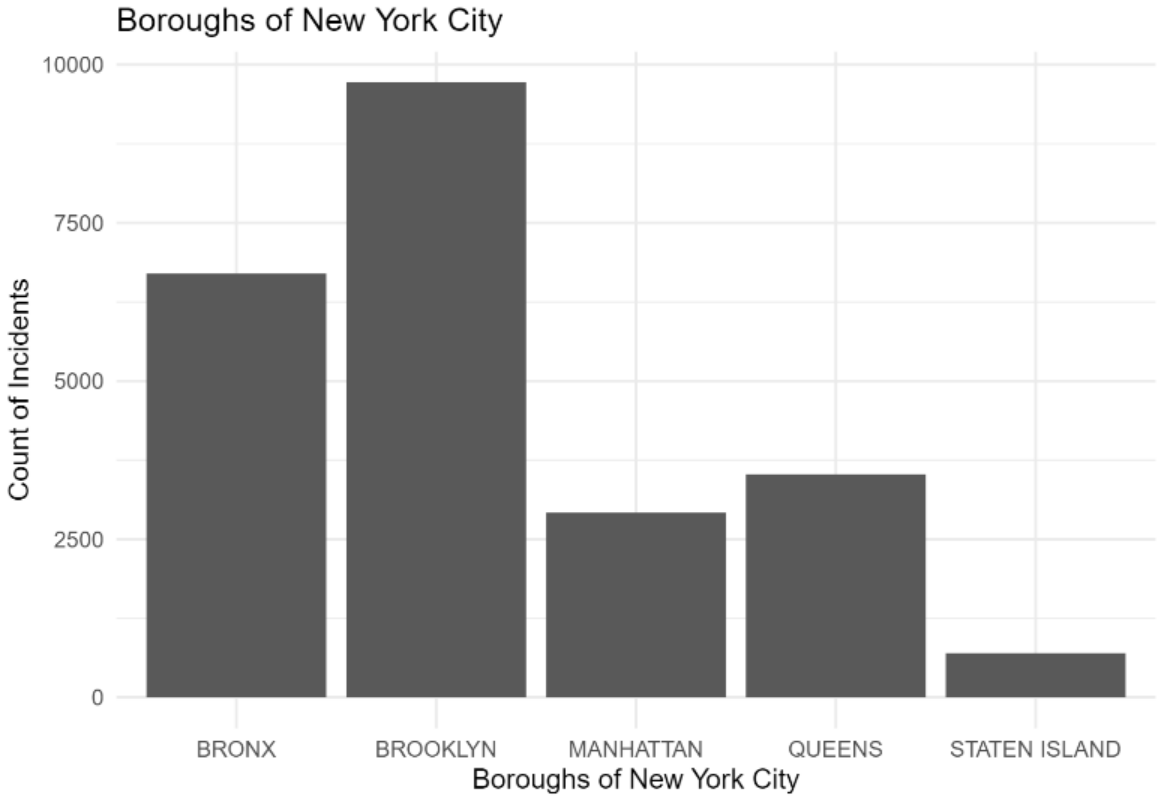

Figure 8: Outcome of the plot

(Source: Provided)

The outcome of the plotted section determines the construction of the bar where 'Boroughs of New York City', and 'Count of Incidents' are found. According to this plot determination, the maximum incidents are found in Brooklyn. Whereas, the minimum incidents are found in Staten Island.

```
table(df_2$BORO, df_2$STATISTICAL_MURDER_FLAG)
```

```
##
##                FALSE TRUE
##   BRONX         5454 1244
##   BROOKLYN      7829 1892
```

Figure 9: Table with shooting data values

(Source: Provided)

The table determines the false, and true numbers of shooting incidents. According to the table data, Bronx has 5454 false, and 1244 true shooting cases. Whereas, Brooklyn has 7829 false, and 1892 true cases found.

Conclusion

The execution determination demonstrates the implementation of the R-programming to determine the main part of the evaluation. The determining data values support the context of the shooting incident finding and analysis concept. This concept is usable to find those locations of the New York City where the maximum shooting cases are found. As per the determination, maximum shooting cases are found in Brooklyn.

References

Silva, J.R., 2022. A Crime Script Analysis of Fatal Police Shootings in New York. Criminology, Crim. Just. L & Soc'y, 23, p.1.

Mezmir, E.A., 2020. Qualitative data analysis: An overview of data reduction, data display, and interpretation. Research on humanities and social sciences, 10(21), pp.15-27.