NYPD shooting data report

2022-05-03

Description

New York Police Department provided the shooting incident data from 2006 to 2021, which happened in main five boroughs in New York city. The both perpetrators and victims' sex, age group and race are provided.

The details from NYPD Shooting Incident Data (Historic) is:

https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD

In this project, we will analysis:

- Yearly shooting incidents
- Shooting incidents by time of the day
- Total shooting incident by boroughs in New York city

Settig up my environmnet

Notes : setting up my R environment by loading the 'tidyverse' and 'lubridata' packages.

```
library(tidyverse)
library(lubridate)
```

Importing Data

Note: read csv file

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

```
## # A tibble: 6 x 19
##   INCIDE~1 OCCUR~2 OCCUR~3 BORO  PRECI~4 JURIS~5 LOCAT~6 STATI~7 PERP_~8 PERP_~9
##      <dbl> <chr>   <time>  <chr>   <dbl>   <dbl> <chr>   <lgl>   <chr>   <chr>
## 1   2.36e8 11/11/~ 15:04   BROO~      79       0 <NA>    FALSE   <NA>    <NA>
## 2   2.31e8 07/16/~ 22:05   BROO~      72       0 <NA>    FALSE   45-64   M
## 3   2.31e8 07/11/~ 01:09   BROO~      79       0 <NA>    FALSE   <18     M
## 4   2.38e8 12/11/~ 13:42   BROO~      81       0 <NA>    FALSE   <NA>    <NA>
## 5   2.24e8 02/16/~ 20:00   QUEE~     113       0 <NA>    FALSE   <NA>    <NA>
## 6   2.28e8 05/15/~ 04:13   QUEE~     113       0 <NA>    TRUE    <NA>    <NA>
## # ... with 9 more variables: 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>, and abbreviated variable
## #   names 1: INCIDENT_KEY, 2: OCCUR_DATE, 3: OCCUR_TIME, 4: PRECINCT,
## #   5: JURISDICTION_CODE, 6: LOCATION_DESC, 7: STATISTICAL_MURDER_FLAG,
## #   8: PERP_AGE_GROUP, 9: PERP_SEX
```

```
dim(df)
```

```
## [1] 25596    19
```

Cleaning and Transformating Data

Cleaning data, select variables, rename columns, and convert data type

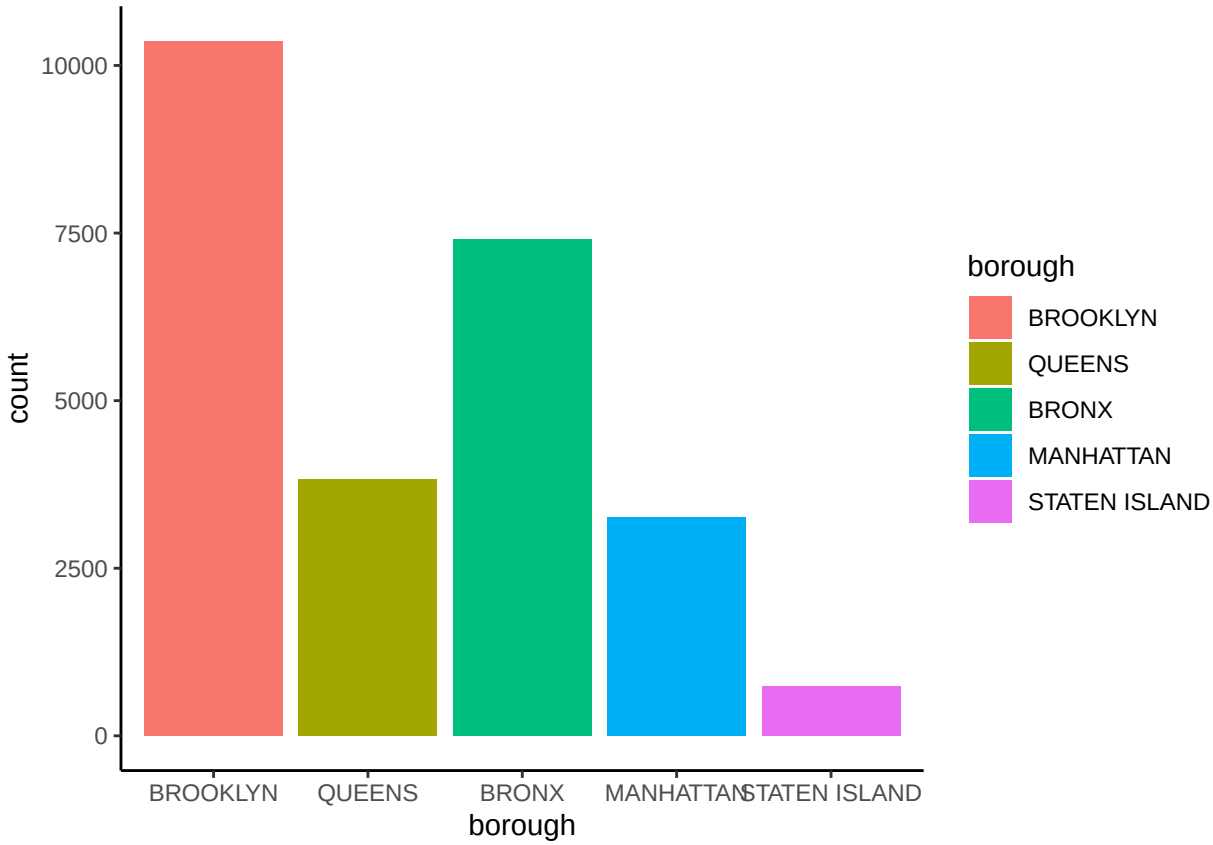
```
df <- df %>%
  rename(date = OCCUR_DATE, time =OCCUR_TIME,  borough = BORO, murder_flag = STATISTICAL_MURDER_FLAG, vi
  mutate(date = mdy(date), borough = as_factor(borough), murder_flag = as.factor(murder_flag), vic_age =
         vic_sex = as_factor(vic_sex), vic_race = as_factor(vic_race)) %>%

  select(date,time, borough, murder_flag, vic_age, vic_sex, vic_race)

head(df)
```

```
## # A tibble: 6 x 7
##   date       time   borough  murder_flag vic_age vic_sex vic_race
##   <date>     <time> <fct>    <fct>       <fct>   <fct>   <fct>
## 1 2021-11-11 15:04  BROOKLYN FALSE       18-24   M       BLACK
## 2 2021-07-16 22:05  BROOKLYN FALSE       25-44   M       ASIAN / PACIFIC ISLAND~
## 3 2021-07-11 01:09  BROOKLYN FALSE       25-44   M       BLACK
## 4 2021-12-11 13:42  BROOKLYN FALSE       25-44   M       BLACK
## 5 2021-02-16 20:00  QUEENS   FALSE       25-44   M       BLACK
## 6 2021-05-15 04:13  QUEENS   TRUE        25-44   M       BLACK
```

Visualizing and Analyzing Data

```
summary(df)
```

```
##       date                time                      borough      murder_flag
##  Min.   :2006-01-01   Length:25596      BROOKLYN     :10365   FALSE:20668
##  1st Qu.:2009-05-10   Class1:hms        QUEENS       : 3828   TRUE : 4928
##  Median :2012-08-26   Class2:difftime   BRONX        : 7402
##  Mean   :2013-06-13   Mode  :numeric    MANHATTAN    : 3265
##  3rd Qu.:2017-07-01                     STATEN ISLAND:  736
##  Max.   :2021-12-31
##
##     vic_age      vic_sex                              vic_race
##  18-24  : 9604   M:23182   BLACK                          :18281
##  25-44  :11386   F: 2403   ASIAN / PACIFIC ISLANDER       :  354
##  <18    : 2681   U:   11   BLACK HISPANIC                 : 2485
##  45-64  : 1698             WHITE HISPANIC                 : 3742
##  65+    :  167             WHITE                          :  660
##  UNKNOWN:   60             AMERICAN INDIAN/ALASKAN NATIVE:    9
##                            UNKNOWN                        :   65
```

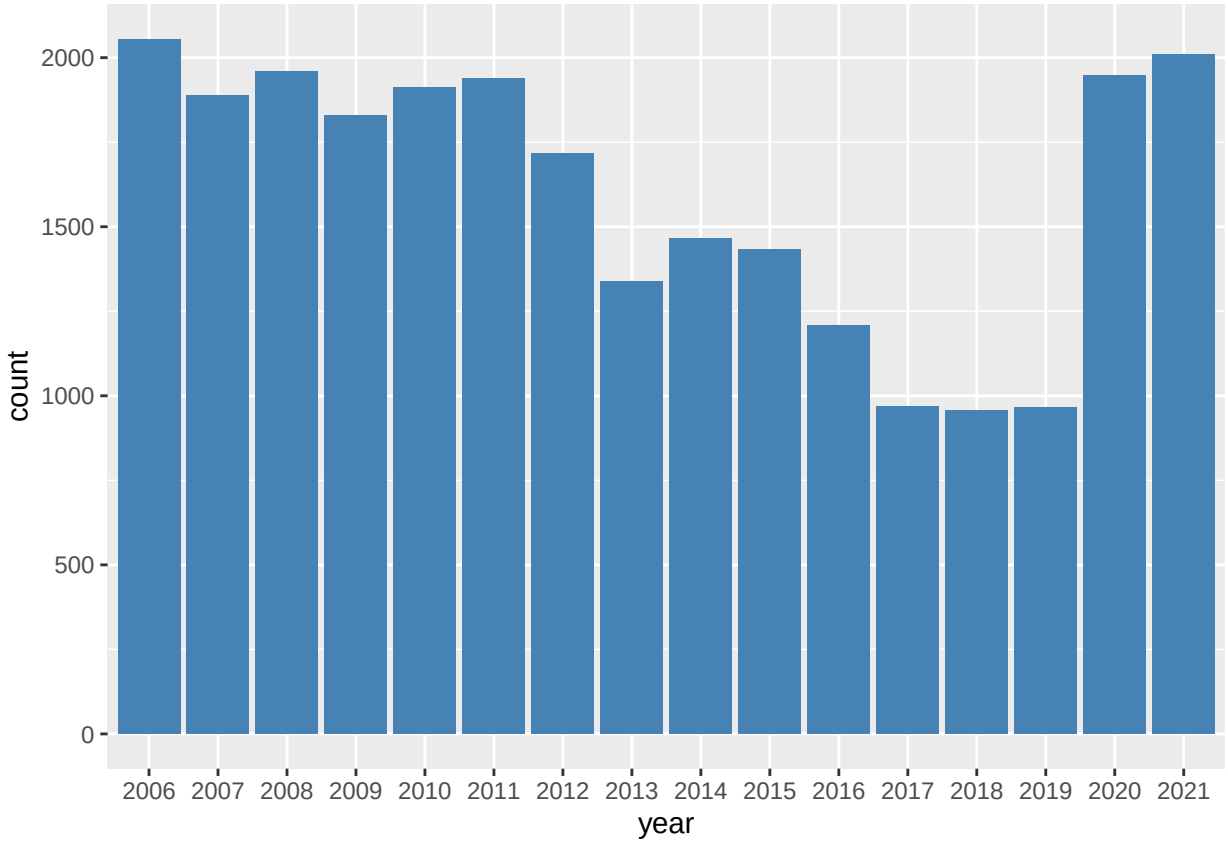


The total shooting incident by boroughs in new york city. Brookly has the most shooting incidents in total, Bronx comes the second. Staten Island has the lowest shooting incidents among the 5 boroughs.

```
## # A tibble: 16 x 2
##    year  count
##    <chr> <int>
##  1 2006   2055
##  2 2007   1887
##  3 2008   1959
##  4 2009   1828
##  5 2010   1912
##  6 2011   1939
##  7 2012   1717
##  8 2013   1339
##  9 2014   1464
## 10 2015   1434
## 11 2016   1208
## 12 2017    970
## 13 2018    958
## 14 2019    967
## 15 2020   1948
## 16 2021   2011
```

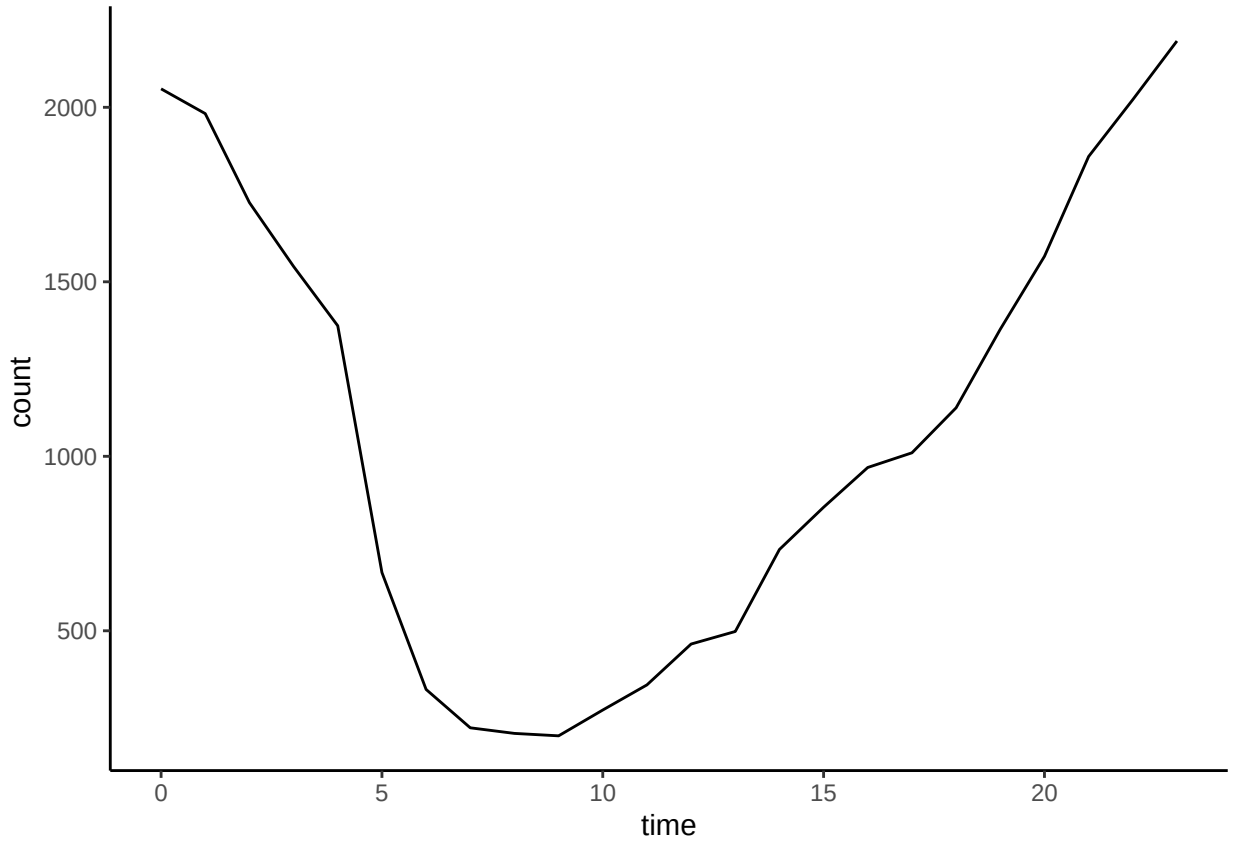
```
##      year               count
##  Length:16          Min.   : 958
##  Class :character   1st Qu.:1306
##  Mode  :character   Median :1772
##                     Mean   :1600
```

```
##                   3rd Qu.:1941
##                   Max.   :2055
```



This is the yearly shooting incidents chart. In 2006, it has the most shooting incidents. From 2017 to 2019, shooting incidents reach the bottom. From 2020 to 2021 it suddenly shoots up, almost doubled.

```
## # A tibble: 6 x 2
## # Groups:   hour [6]
##    hour     n
##   <dbl> <int>
## 1     0  2053
## 2     1  1982
## 3     2  1727
## 4     3  1544
## 5     4  1374
## 6     5   667
```



This is the shooting incidents by time of the day chart. Early morning it has the lowest shooting incidents. Incidents number peaks around midnight. During the day time the incident climbs steadily from the bottom in early morning, all the way to midnight.

Conclusion

In conclusion, from 2006 shooting incidents drops steadily until 2020, which shoots up sharply back to the 2006 level. In early morning New York City has the lowest shooting incidents. By boroughs, Staten Island has the total lowest shooting incidents.