NYPD Shooting Incident Report

Krunal.Desai

2024-02-09

Step-1 Identify and import the data

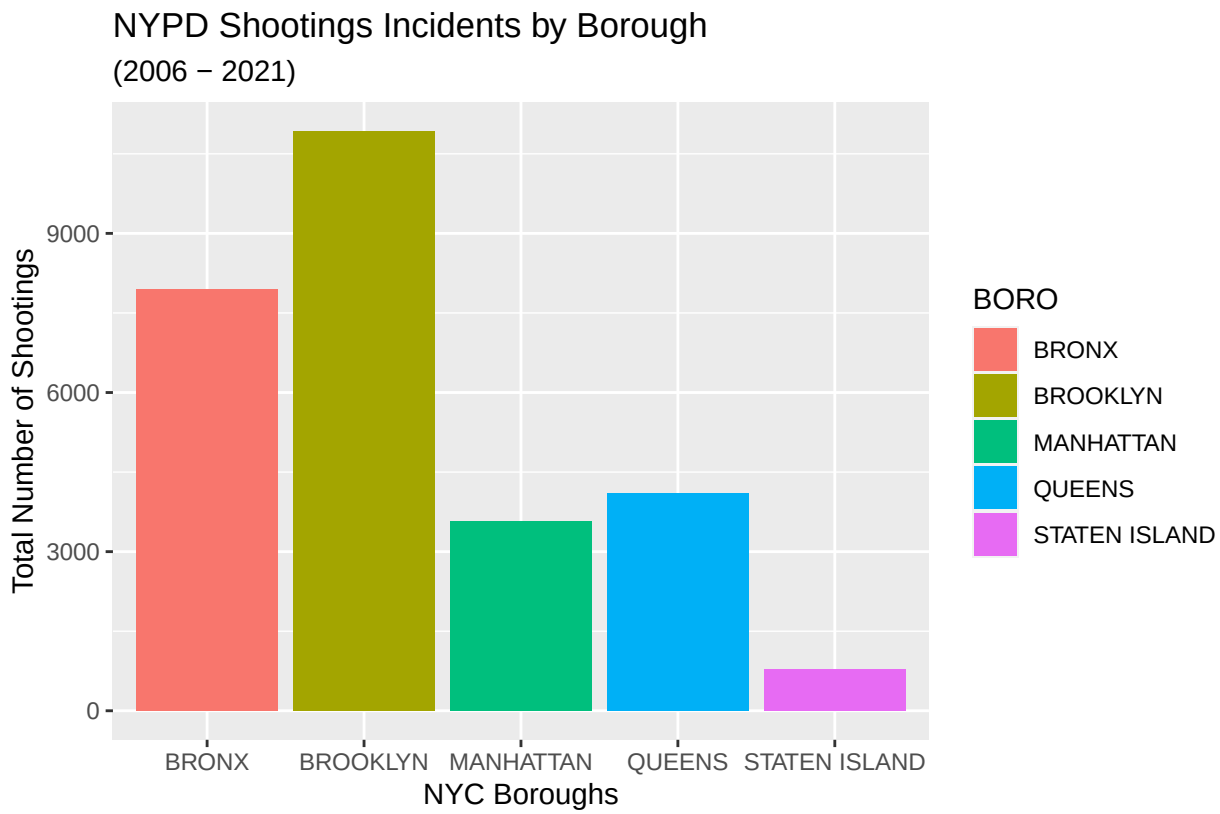
I will start by reading in the data from the main csv files.

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

NYPD <- read.csv(url_NYPD)
```

Step-2 Cleaning the data

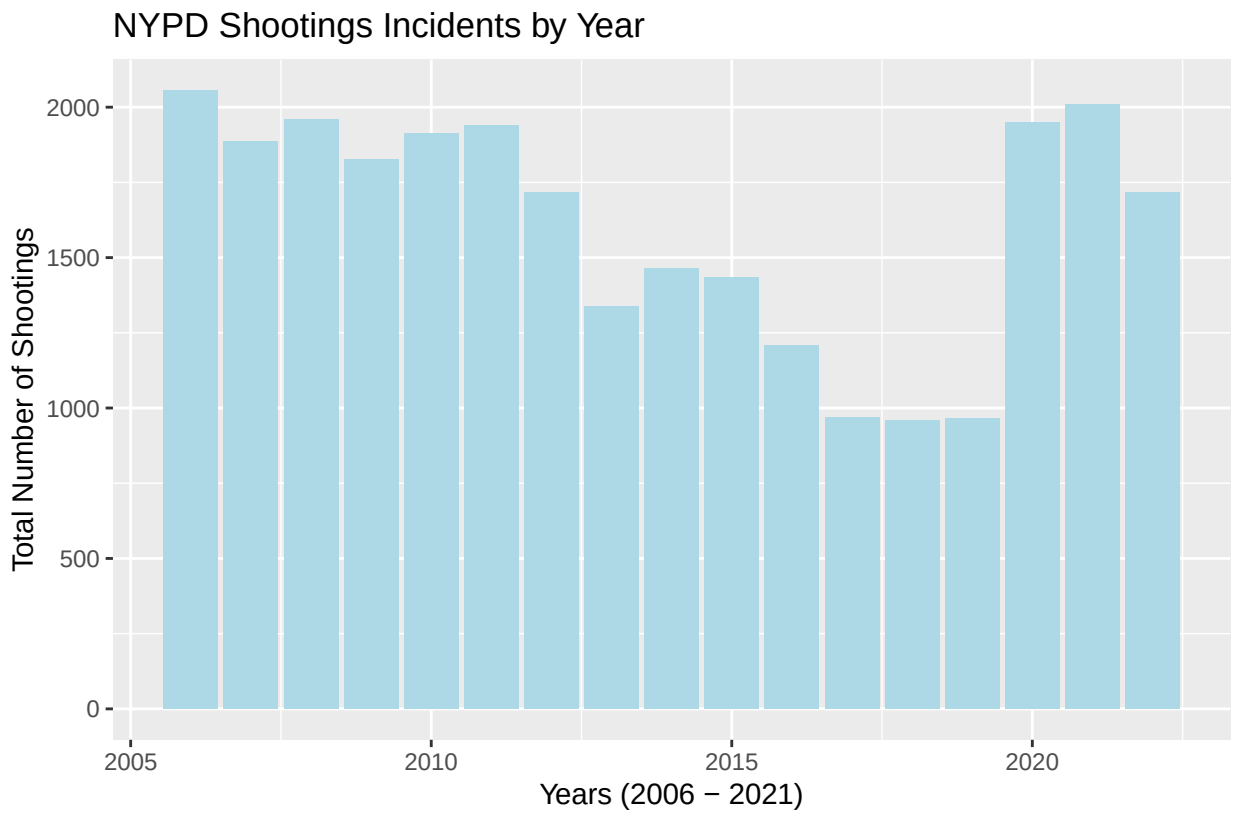
Now, I will cleaning gthe data and visualize to see

```
NYPD_clean <- NYPD %>%
  select(c("OCCUR_DATE","OCCUR_TIME","BORO","PRECINCT",
           "STATISTICAL_MURDER_FLAG","VIC_AGE_GROUP","VIC_SEX","VIC_RACE")) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE),
         OCCUR_TIME = hms(OCCUR_TIME),
         STATISTICAL_MURDER_FLAG = as.logical(STATISTICAL_MURDER_FLAG),
         Shootings = 1,
         Year = year(OCCUR_DATE))


NYPD_clean %>%
  ggplot(aes(x = BORO, fill = BORO)) +
  geom_bar() +
  labs(title = "NYPD Shootings Incidents by Borough",
       subtitle = "(2006 - 2021)",
       x = "NYC Boroughs",
       y = "Total Number of Shootings",
       caption = "(Figure - 1)")
```



(Figure − 1)

```
NYPD_clean %>%
  ggplot(aes(x = Year)) +
  geom_bar(fill = "lightblue",show.legend = FALSE) +
  labs(title = "NYPD Shootings Incidents by Year",
       x = "Years (2006 - 2021)",
       y = "Total Number of Shootings",
       caption = "(Figure - 2)")
```

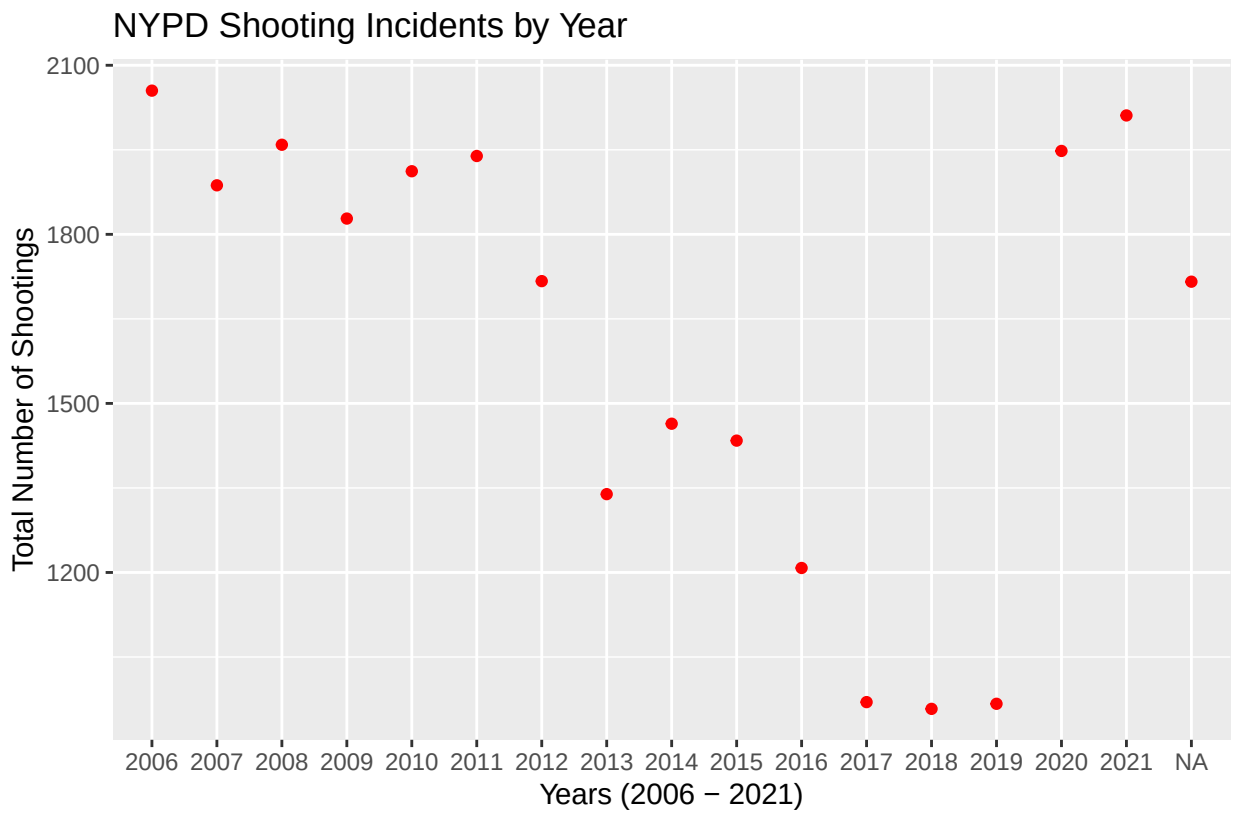


(Figure − 2)

Step-3 Analyze and Visulize data

```
NYPD_year <- NYPD_clean %>%
  group_by(Year) %>%
  summarize(Shootings = sum(Shootings))

NYPD_year %>%
  ggplot(aes(x = as.factor(Year), y = Shootings)) +
  geom_line() +
  geom_point(color = "red") +
  scale_x_discrete(labels = as.character(2006:2021)) +
  labs(
    title = "NYPD Shooting Incidents by Year",
    x = "Years (2006 - 2021)",
    y = "Total Number of Shootings",
    caption = "(Figure - 3)"
  )
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```



(Figure − 3)

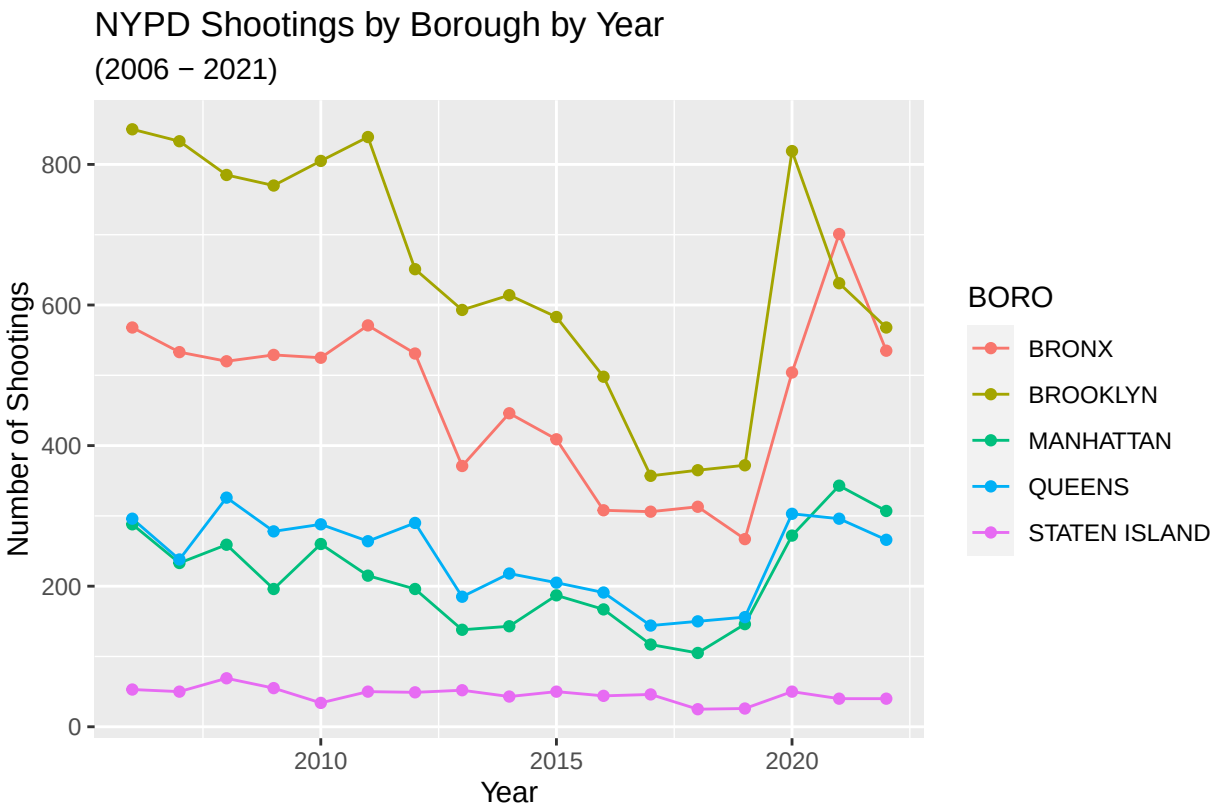
```
NYPD_boro <- NYPD_clean %>%
  group_by(BORO, OCCUR_DATE,Shootings) %>%
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG),
            .groups = 'drop') %>%
  select(BORO,OCCUR_DATE,Shootings,STATISTICAL_MURDER_FLAG) %>%
  ungroup()

NYPD_boro_year <- NYPD_clean %>%
  mutate(Year = year(OCCUR_DATE)) %>%
  group_by(BORO, Year, Shootings) %>%
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG),
            .groups = 'drop') %>%
  select(BORO, Year,Shootings,STATISTICAL_MURDER_FLAG) %>%
  ungroup()

NYPD_boro_total <- NYPD_boro_year %>%
  group_by(BORO) %>%
  summarize(Shootings = sum(Shootings))
(7402 + 10365)/ sum(NYPD_boro_total$Shootings)
```

```
## [1] 0.6505199
```

```
736/ sum(NYPD_boro_total$Shootings)
```

```
## [1] 0.02694786
```

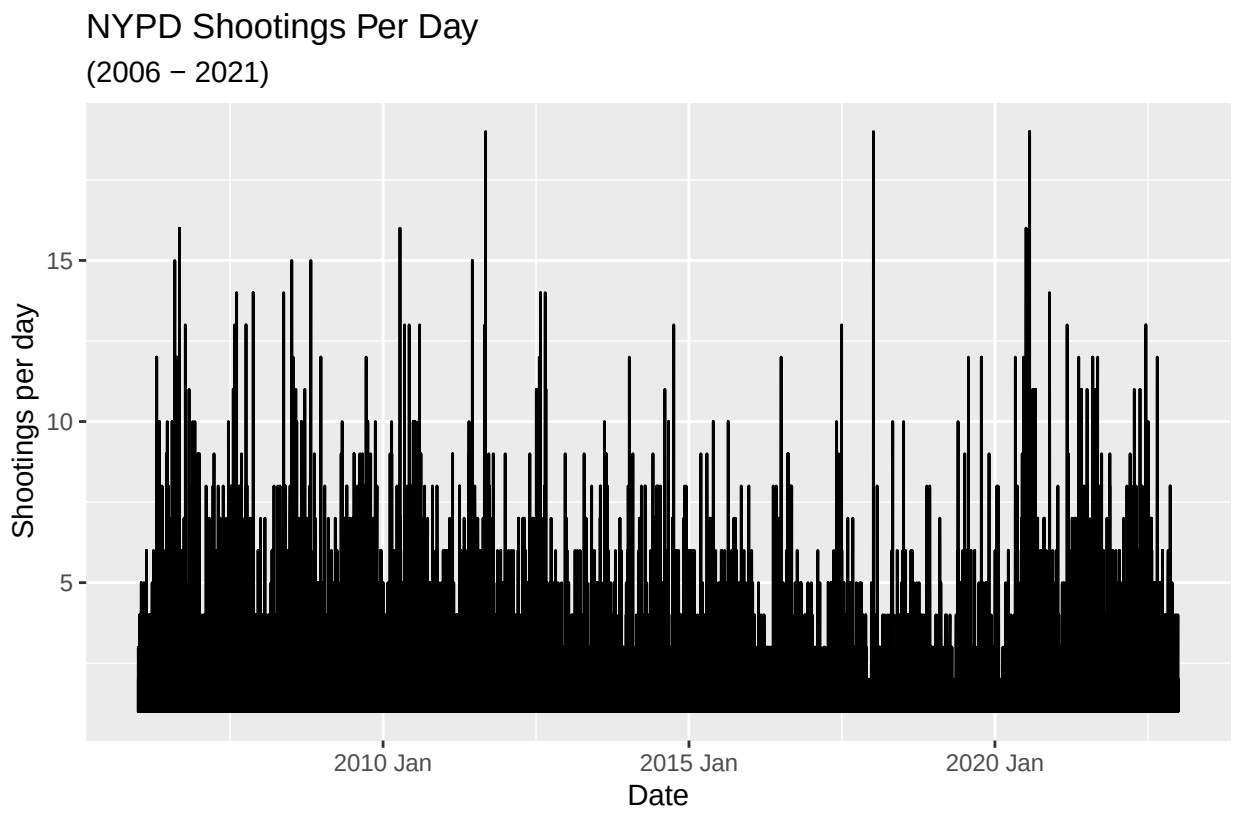
```
NYPD_boro_year %>%
  ggplot(aes(x = Year, y = Shootings,color = BORO)) +
  geom_line() +
  geom_point() +
  labs(title = "NYPD Shootings by Borough by Year",
       subtitle = "(2006 - 2021)",
       x = "Year",
       y = "Number of Shootings",
       caption = "(Figure - 4)")
```



(Figure − 4)

```
NYPD_boro <- NYPD_clean %>%
  group_by(BORO, OCCUR_DATE, Shootings) %>%
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG),
            .groups = 'drop') %>%
  select(BORO, OCCUR_DATE, Shootings, STATISTICAL_MURDER_FLAG) %>%
  ungroup()

NYPD_boro %>%
  ggplot(aes(x = OCCUR_DATE, y = Shootings)) +
```

```
  geom_line() +
  scale_x_date(date_labels = "%Y %b") +
  labs(title = "NYPD Shootings Per Day",
       subtitle = "(2006 - 2021)",
       x = "Date",
       y = "Shootings per day",
       caption = "(Figure - 5)")
```

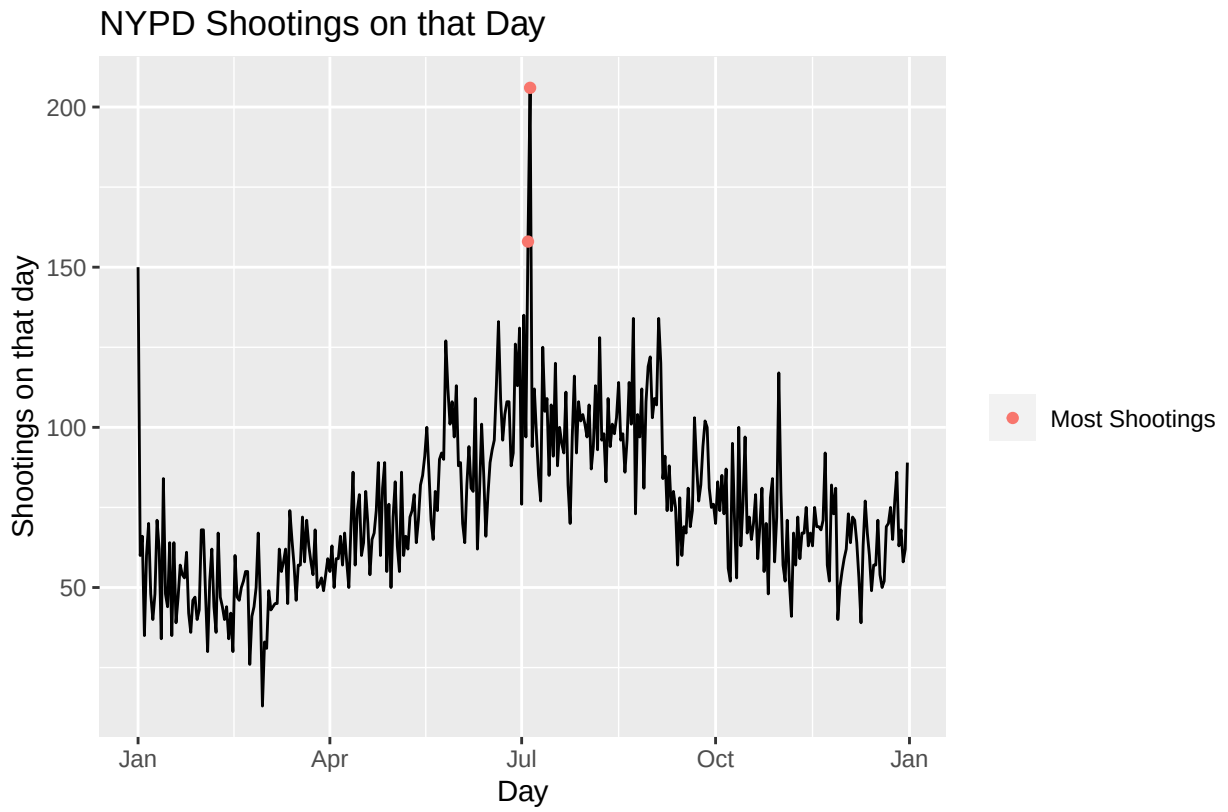


(Figure − 5)

```
NYPD_time_year <- NYPD_clean %>%
  mutate(Time_year = format(as.Date(OCCUR_DATE), "%m/%d")) %>%
  mutate(Time_year = as.Date(Time_year,"%m/%d")) %>%
  group_by(Time_year,Shootings) %>%
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG),
            .groups = 'drop') %>%
  select(Time_year,Shootings,STATISTICAL_MURDER_FLAG) %>%
  ungroup()

NYPD_time_year %>% slice_max(Shootings, n = 2)
```

```
## # A tibble: 2 x 3
##   Time_year  Shootings STATISTICAL_MURDER_FLAG
##   <date>         <dbl>                  <int>
## 1 2024-07-05       206                     33
## 2 2024-07-04       158                     26
```

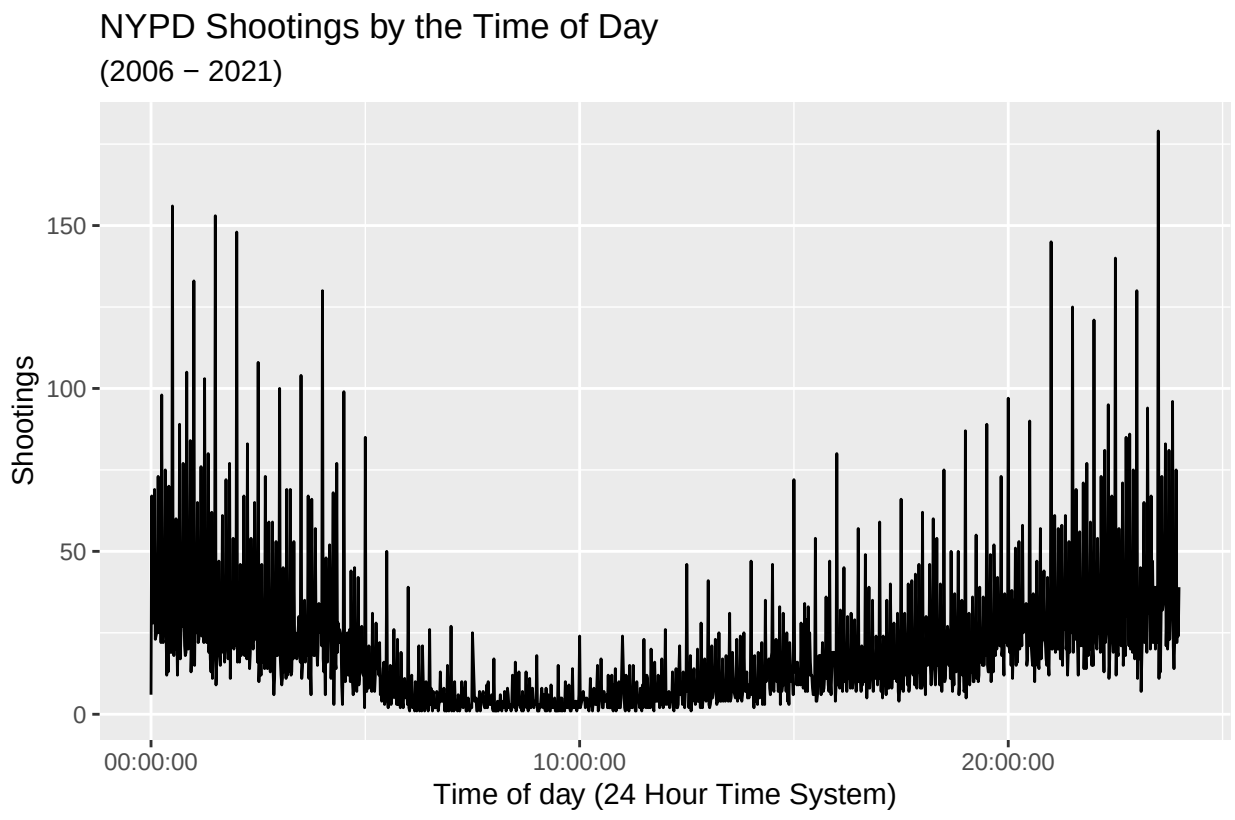
```
NYPD_July_5 <- NYPD_clean %>%
  mutate(Time_year = format(as.Date(OCCUR_DATE), "%m/%d"),
         Hour = hour(OCCUR_TIME)) %>%
  mutate(Time_year = as.Date(Time_year,"%m/%d")) %>%
  filter(Time_year == "2022-07-05") %>%
  group_by(Hour,Shootings) %>%
  summarize(Shootings = sum(Shootings),
            .groups = 'drop')


NYPD_time_year %>%
  ggplot(aes(x = Time_year, y = Shootings)) +
  geom_line() +
  geom_point(data = NYPD_time_year %>% slice_max(Shootings, n = 2),
             aes(color="Most Shootings")) +
  scale_x_date(date_labels = "%b") +
  labs(title = "NYPD Shootings on that Day",
       subtile = "(2006 - 2021)",
       colour = "",
       x = "Day",
       y = "Shootings on that day",
       caption = "(Figure - 6)")
```



(Figure − 6)

```
NYPD_time_day <- NYPD_clean %>%
  group_by(OCCUR_TIME,Shootings) %>%
```

```
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG),
            .groups = 'drop') %>%
  select(OCCUR_TIME,Shootings,STATISTICAL_MURDER_FLAG)

NYPD_time_day %>%
  ggplot(aes(x = OCCUR_TIME, y = Shootings)) +
  geom_line() +
  scale_x_time() +
  labs(title = "NYPD Shootings by the Time of Day",
       subtitle = "(2006 - 2021)",
       x = "Time of day (24 Hour Time System)",
       y = "Shootings",
       caption = "(Figure - 7)")
```



(Figure − 7)

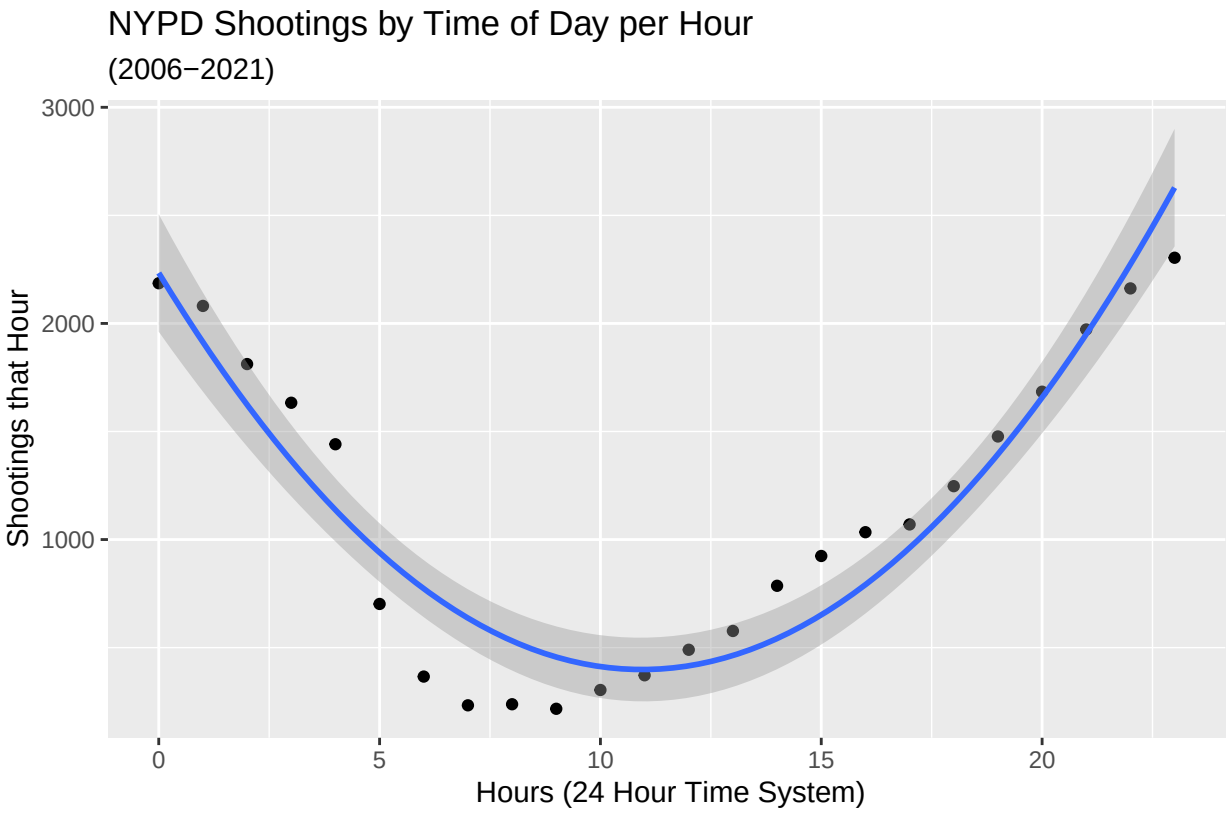
```
NYPD_time_hour <- NYPD_clean %>%
  mutate(Hour = hour(OCCUR_TIME)) %>%
  group_by(Hour,Shootings) %>%
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG),
            .groups = 'drop') %>%
  mutate(Hour2 = Hour^2) %>%
  select(Hour,Shootings,STATISTICAL_MURDER_FLAG, Hour2)
```

Step-4 Choose model

```
NYPD_time_hour_model <- lm(data = NYPD_time_hour, Shootings ~ Hour + Hour2)
summary(NYPD_time_hour_model)
```

```
##
## Call:
## lm(formula = Shootings ~ Hour + Hour2, data = NYPD_time_hour)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -406.73 -143.32   50.61  172.71  303.99
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2233.526    130.753   17.08 8.56e-14 ***
## Hour        -335.455     26.333  -12.74 2.40e-11 ***
## Hour2         15.331      1.106   13.87 4.86e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 231.6 on 21 degrees of freedom
## Multiple R-squared:  0.9044, Adjusted R-squared:  0.8952
## F-statistic: 99.28 on 2 and 21 DF,  p-value: 1.981e-11
```

```
NYPD_time_hour %>%
  ggplot(aes(x = Hour, y = Shootings)) +
  geom_point() +
  stat_smooth(method = "lm", formula = y ~ x + I(x^2), linewidth = 1) +
  labs(title = "NYPD Shootings by Time of Day per Hour",
       subtitle = "(2006-2021)",
       x = " Hours (24 Hour Time System)",
       y = "Shootings that Hour",
       caption = "(Figure - 8)")
```


NYPD Shootings by Time of Day per Hour
(2006−2021)
Shootings that Hour
3000
2000
1000
0
5
10
15
20
Hours (24 Hour Time System)

(Figure − 8)