R Notebook

Code

Hide

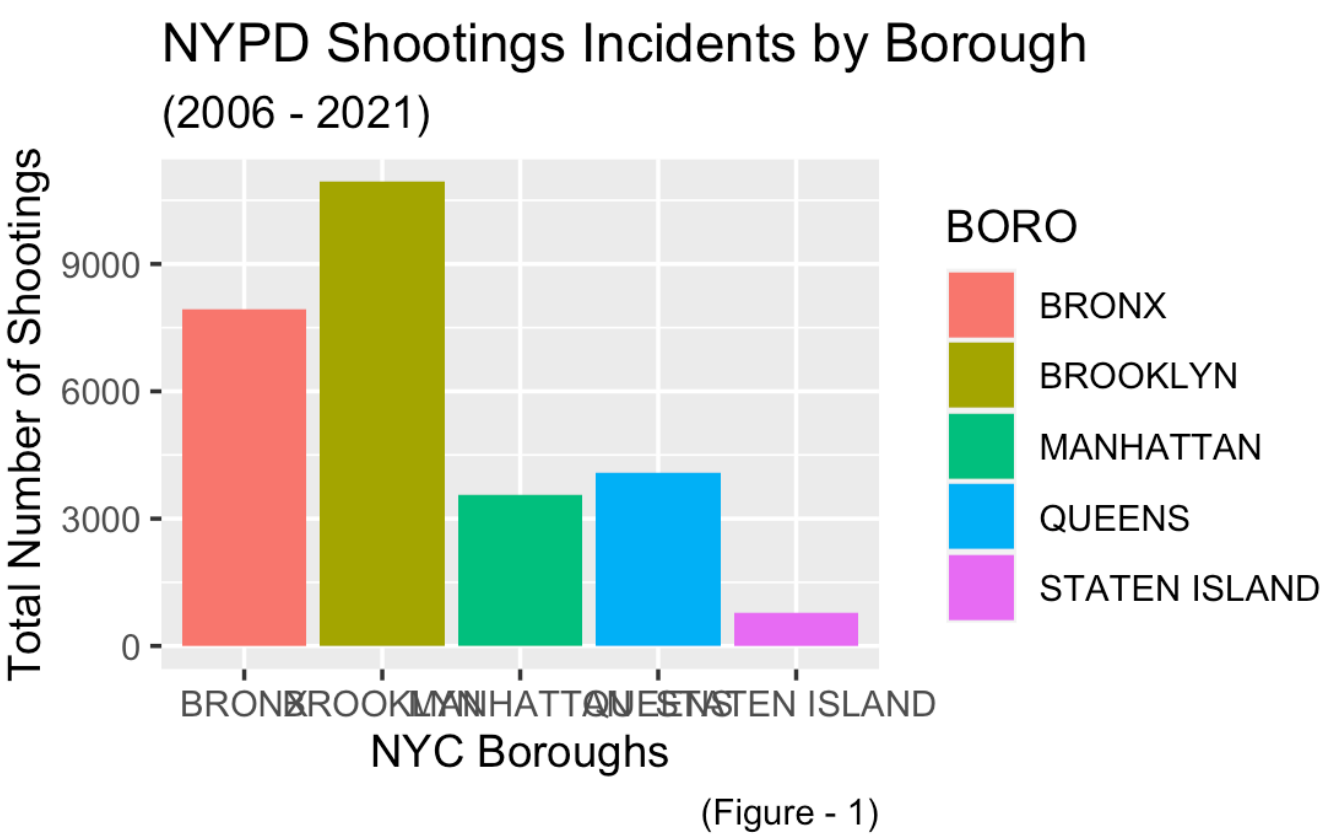
```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(lubridate)
library(ggplot2)
```

Hide

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

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

Hide

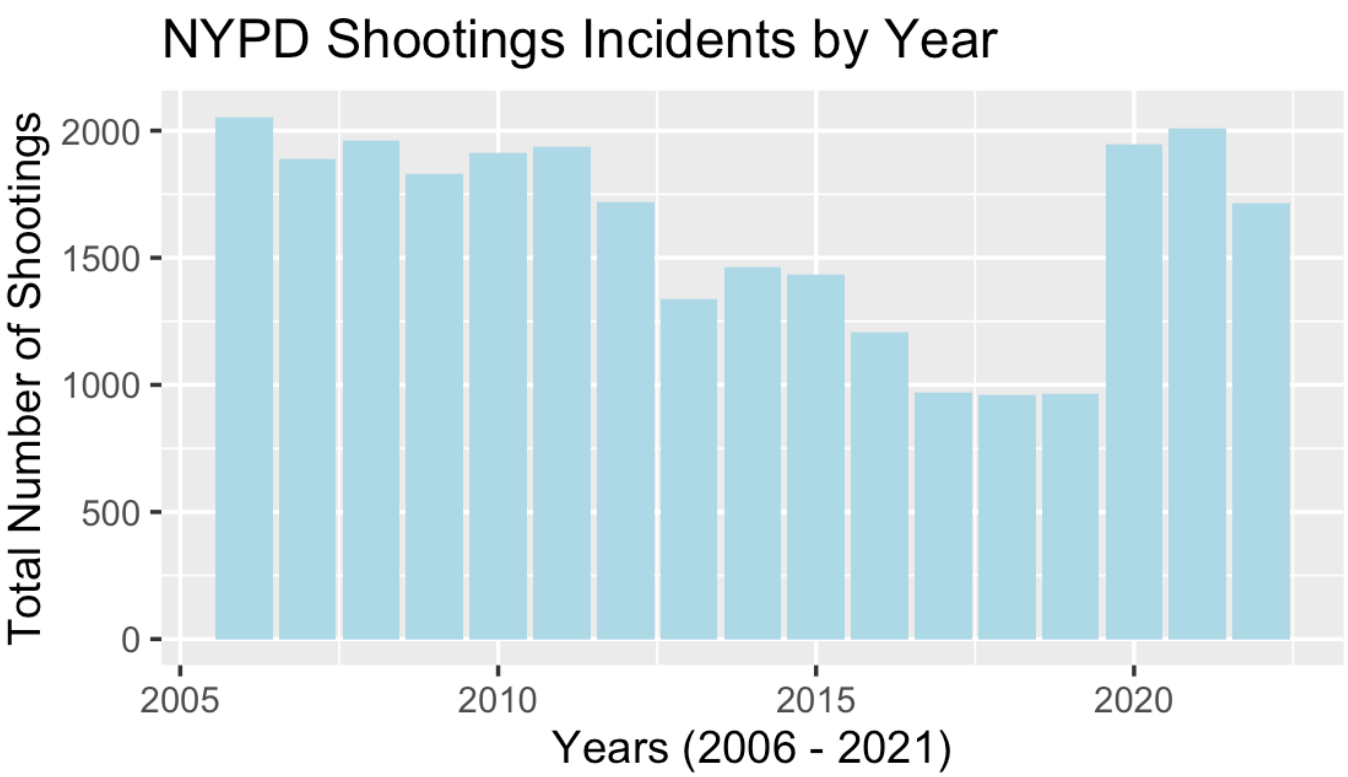
```
NYPD_clean <- NYPD %>%
  select(c("OCCUR_DATE","OCCUR_TIME","BORO","PRECINCT",
           "STATISTICAL_MURDER_FLAG","VIC_AGE_GROUP","VIC_SEX","VIC_RACE")) %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE),
         OCCUR_TIME = hms(OCCUR_TIME),
         STATISTICAL_MURDER_FLAG = as.logical(STATISTICAL_MURDER_FLAG),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

Hide

```
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)

Hide

```
NYPD_year <- NYPD_clean %>%
  group_by(Year,Shootings) %>%
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG)) %>%
  select(Year,Shootings,STATISTICAL_MURDER_FLAG) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'Year'. You can override using the `.groups` argu
ment.
```

Hide

```
NYPD_year %>% slice_max(Shootings, n = 4)
```

Year	Shootings	STATISTICAL_MURDER_FLAG
<dbl>	<dbl>	<int>
2006	2055	445
2021	2011	428
2008	1959	362
2020	1948	366

4 rows

Hide

```
NYPD_year %>% slice_min(Shootings, n = 4)
```

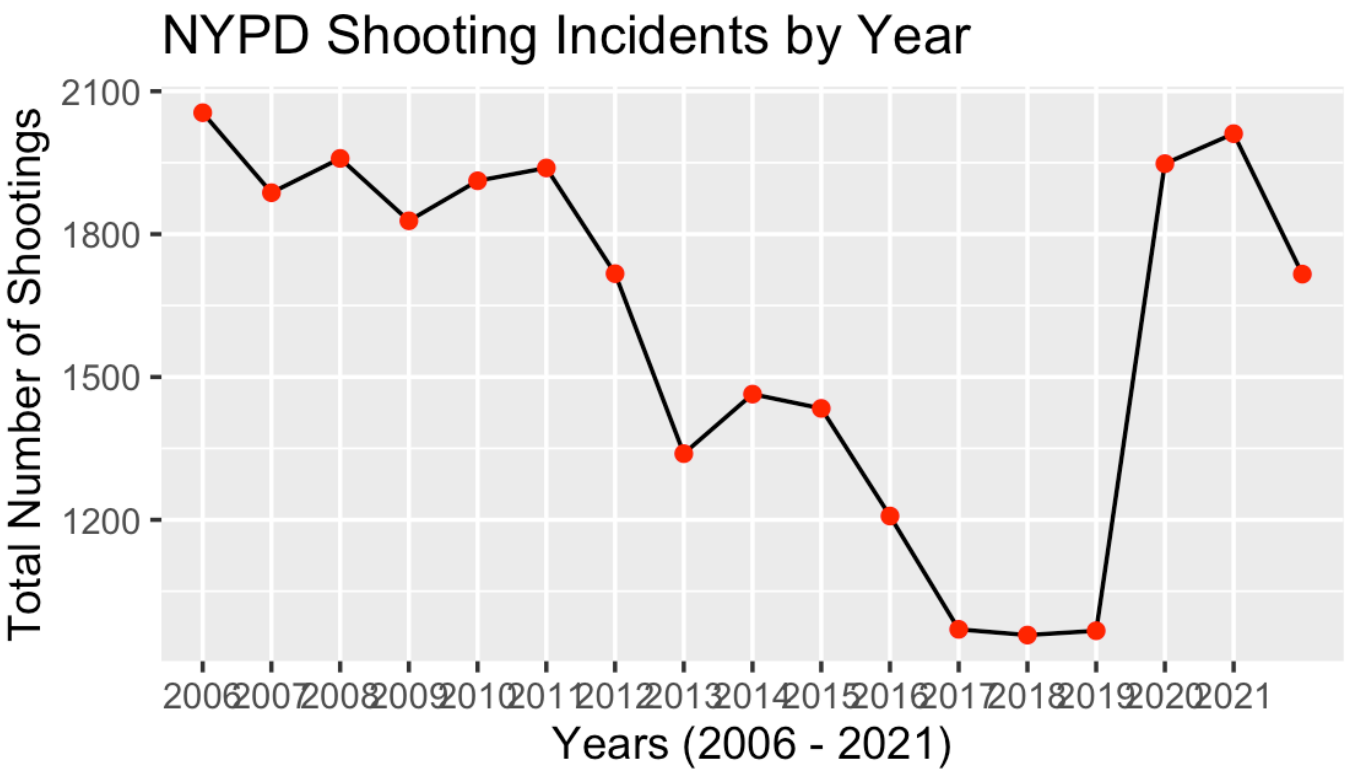
Year <dbl>	Shootings <dbl>	STATISTICAL_MURDER_FLAG <int>
2018	958	204
2019	967	184
2017	970	174
2016	1208	223

4 rows

Hide

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

```
Warning: Continuous limits supplied to discrete scale.
ℹ Did you mean `limits = factor(...)` or
  `scale_*_continuous()`?
```



(Figure - 3)

Hide

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

```
`summarise()` has grouped output by 'BORO', 'OCCUR_DATE'. You can override using the
`.groups` argument.
```

Hide

```
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)) %>%
  select(BORO, Year,Shootings,STATISTICAL_MURDER_FLAG) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'BORO', 'Year'. You can override using the `.grou
ps` argument.
```

Hide

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

```
[1] 0.6505199
```

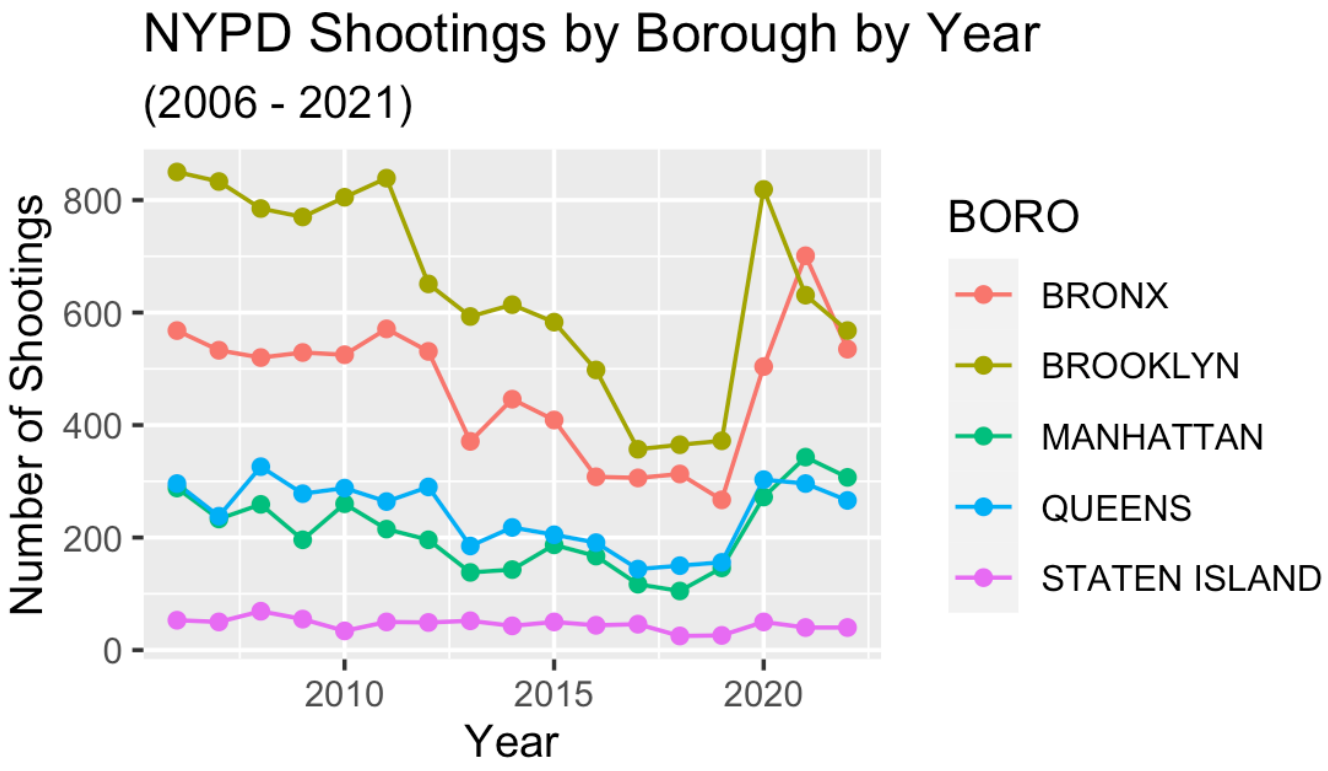
Hide

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

```
[1] 0.02694786
```

Hide

```
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)")
```



Hide

```
NYPD_date <- NYPD_clean %>%
  group_by(OCCUR_DATE,Shootings,STATISTICAL_MURDER_FLAG) %>%
  summarize(Shootings = sum(Shootings),

            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG)) %>%
  select(OCCUR_DATE,Shootings,STATISTICAL_MURDER_FLAG) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'OCCUR_DATE', 'Shootings'. You can override using
the `.groups` argument.
```

Hide

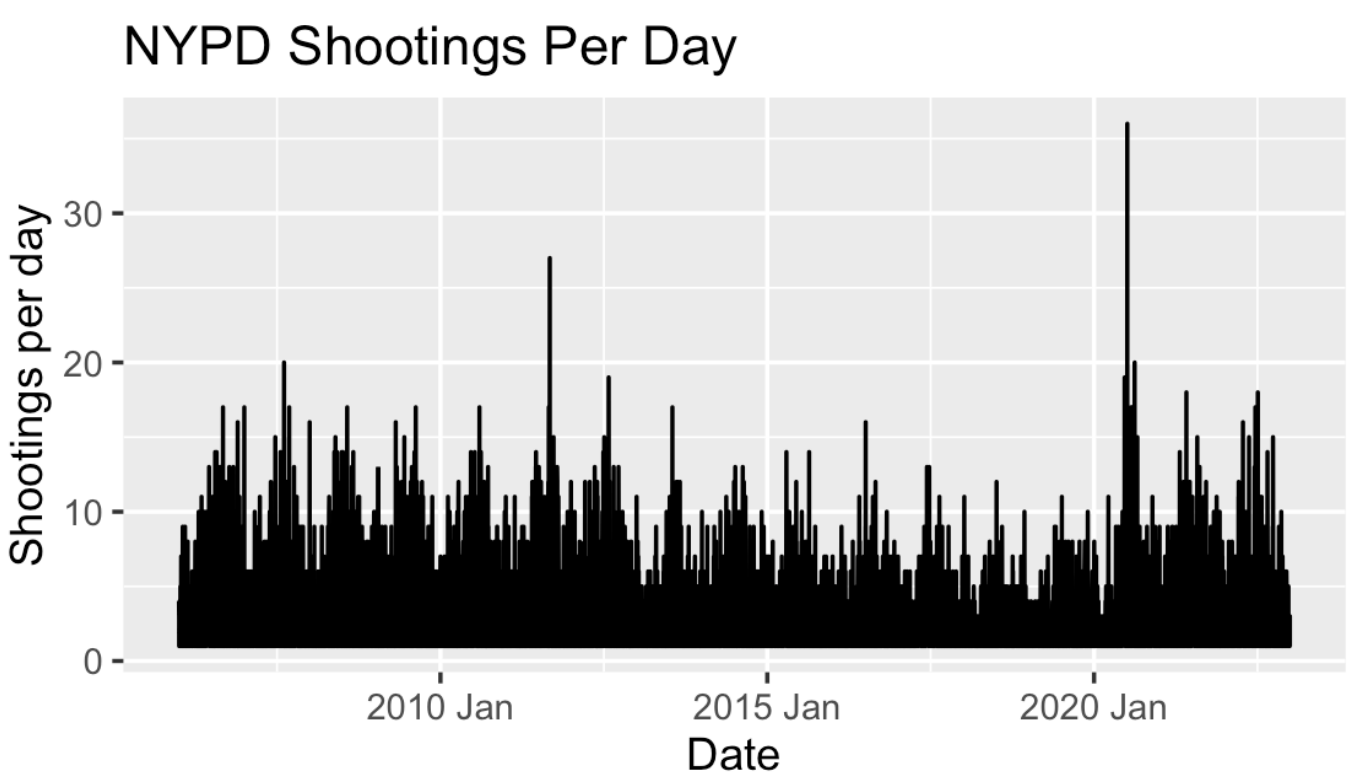
```
NYPD_date %>% slice_max(Shootings, n=2)
```

OCCUR_DATE <date>	Shootings <dbl>	STATISTICAL_MURDER_FLAG <int>
2020-07-05	36	0
2011-09-04	27	0

2 rows

Hide

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



(Figure - 5)

Hide

```
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)) %>%
  select(Time_year,Shootings,STATISTICAL_MURDER_FLAG) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'Time_year'. You can override using the `.groups`
argument.
```

Hide

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

Time_year <date>	Shootings <dbl>	STATISTICAL_MURDER_FLAG <int>
2023-07-05	206	33
2023-07-04	158	26

2 rows

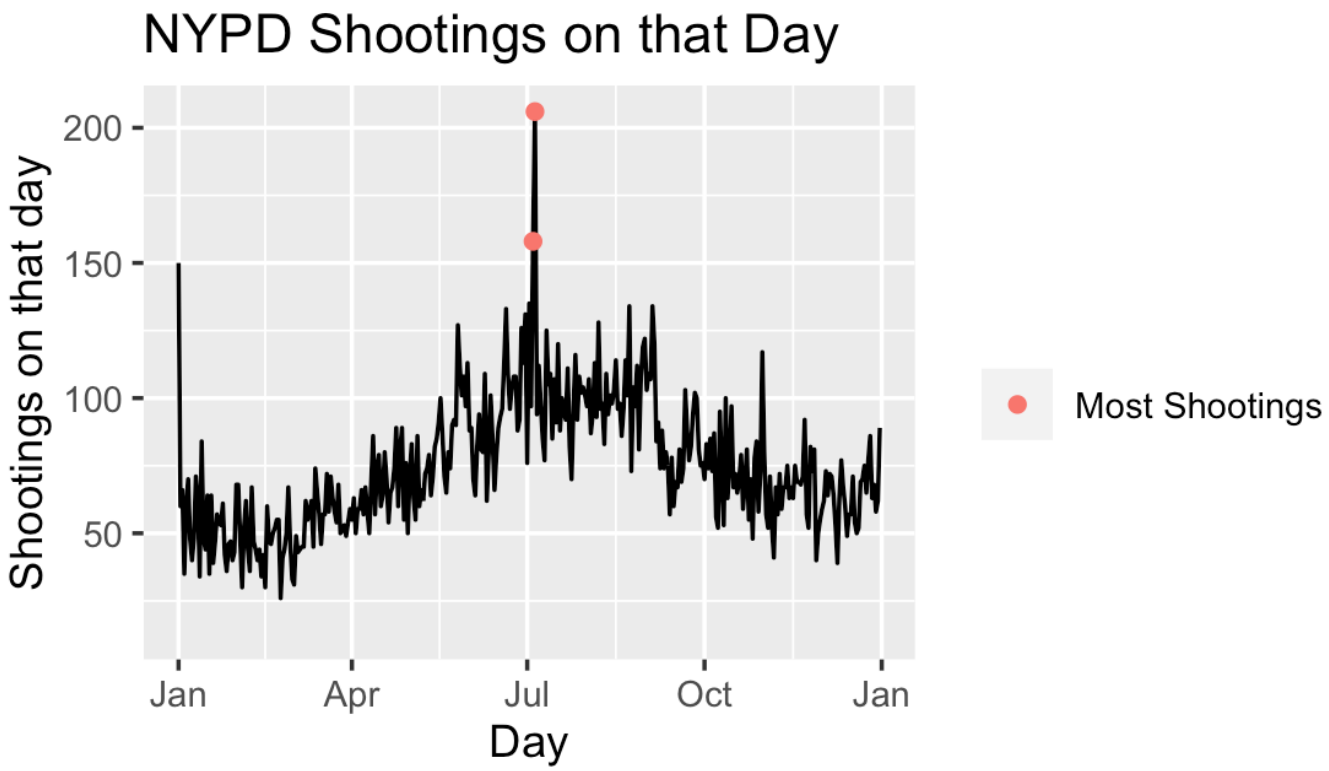
Hide

```
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))
```

```
`summarise()` has grouped output by 'Hour'. You can override using the `.groups` argu
ment.
```

Hide

```
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)

Hide

```
summary(clean)
```

```
  OCCUR_DATE                   BORO       LOC_OF_OCCUR_DESC
 Min.   :2006-01-01   BRONX        : 7937   Length:27312
 1st Qu.:2009-07-18   BROOKLYN     :10933   Class :character
 Median :2013-04-29   MANHATTAN    : 3572   Mode  :character
 Mean   :2014-01-06   QUEENS       : 4094
 3rd Qu.:2018-10-15   STATEN ISLAND:  776
 Max.   :2022-12-31

    PRECINCT     LOC_CLASSFCTN_DESC    Latitude       Longitude
 75     : 1557   Length:27312       Min.   :40.51   Min.   :-74.25
 73     : 1452   Class :character   1st Qu.:40.67   1st Qu.:-73.94
 67     : 1216   Mode  :character   Median :40.70   Median :-73.92
 44     : 1020                      Mean   :40.74   Mean   :-73.91
 79     : 1012                      3rd Qu.:40.82   3rd Qu.:-73.88
 47     :  953                      Max.   :40.91   Max.   :-73.70
 (Other):20102                      NA's   :10      NA's   :10
```

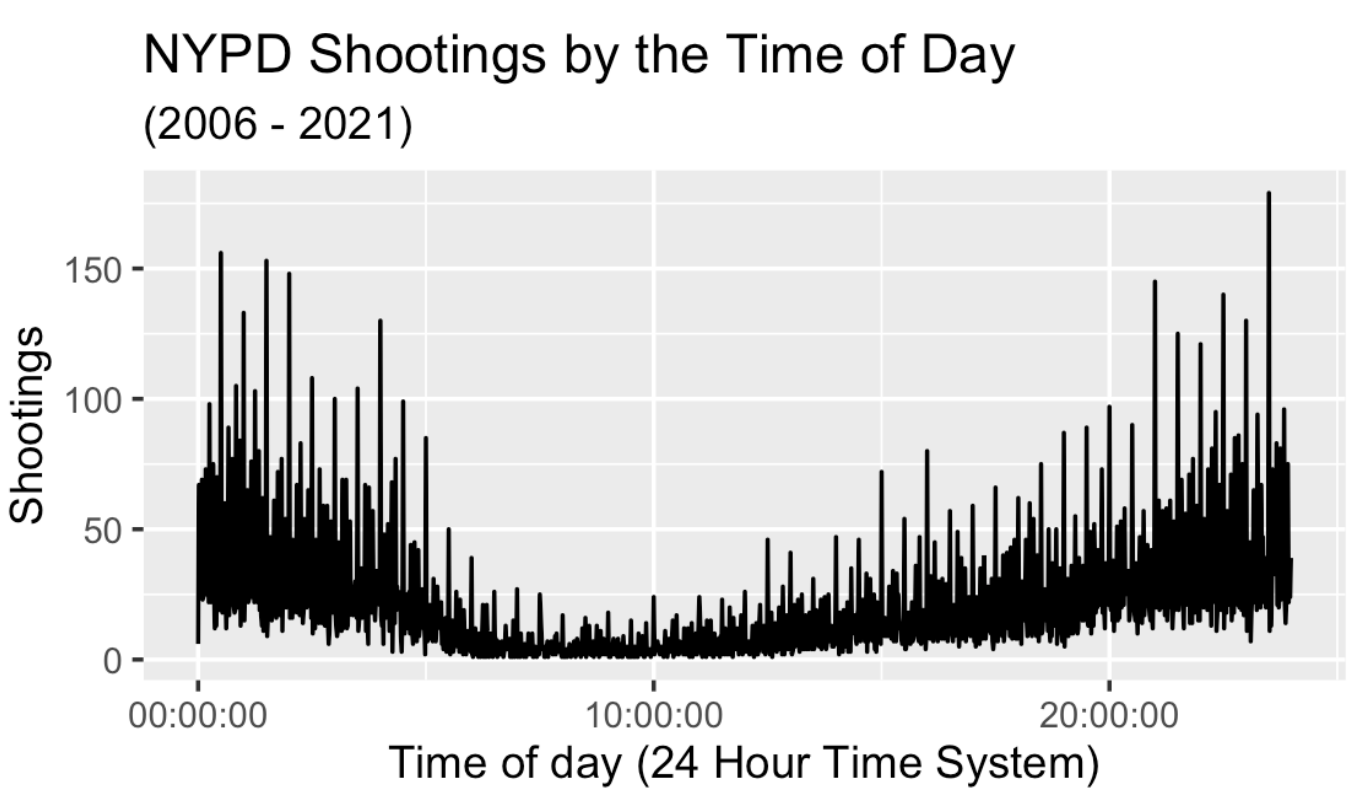
Hide

```
NYPD_time_day <- NYPD_clean %>%
  group_by(OCCUR_TIME,Shootings) %>%
  summarize(Shootings = sum(Shootings),
            STATISTICAL_MURDER_FLAG = sum(STATISTICAL_MURDER_FLAG)) %>%
  select(OCCUR_TIME,Shootings,STATISTICAL_MURDER_FLAG)
```

```
`summarise()` has grouped output by 'OCCUR_TIME'. You can override using the `.groups
` argument.
```

Hide

```
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)

Hide

```
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)) %>%
  mutate(Hour2 = Hour^2) %>%
  select(Hour,Shootings,STATISTICAL_MURDER_FLAG, Hour2)
```

```
`summarise()` has grouped output by 'Hour'. You can override using the `.groups` argu
ment.
```

Hide

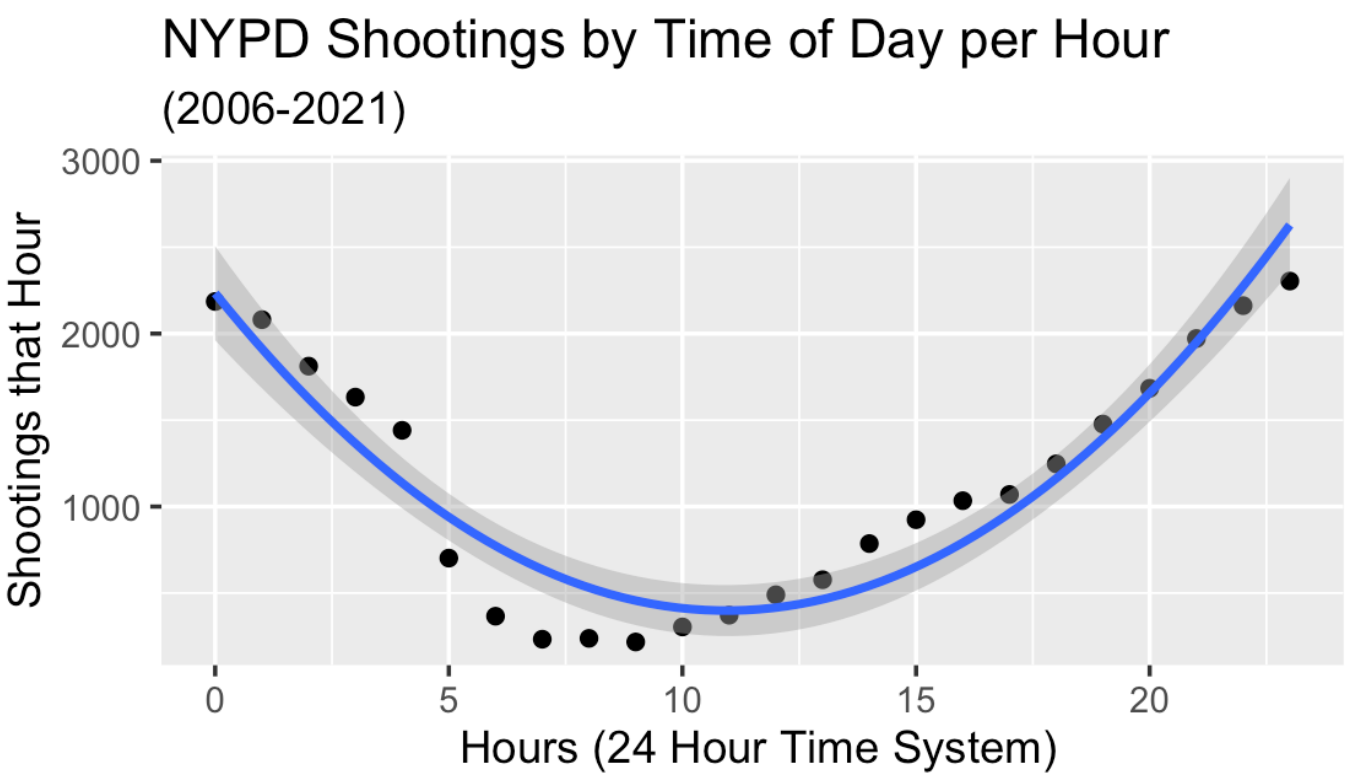
```
NYPD_time_hour_model <- lm(data = NYPD_time_hour, Shootings ~ Hour + Hour2)
summary(NYPD_time_hour_model)
```

```
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
```

Hide

```
NYPD_time_hour %>%
  ggplot(aes(x = Hour, y = Shootings)) +
  geom_point() +
  stat_smooth(method = "lm", formula = y ~ x + I(x^2), size = 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)")
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
Please use `linewidth` instead.
```


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)