NYPD Shooting Incident

Source of data

NYPD Shooting Incident Data (Historic)

See https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic.

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

```
## Rows: 23568 Columns: 19

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (11): OCCUR_DATE, OCCUR_TIME, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_...
## dbl  (5): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, Latitude, Longitude
## lgl  (1): STATISTICAL_MURDER_FLAG

##
## 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.
```

```
shooting_data
```

```
## # A tibble: 23,568 x 19
##    INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO          PRECINCT JURISDICTION_CODE
##           <dbl> <chr>      <chr>      <chr>            <dbl>              <dbl>
##  1    201575314 08/23/2019 22:10:00   QUEENS             103                  0
##  2    205748546 11/27/2019 15:54:00   BRONX               40                  0
##  3    193118596 02/02/2019 19:40:00   MANHATTAN           23                  0
##  4    204192600 10/24/2019 00:52:00   STATEN ISLAND      121                  0
##  5    201483468 08/22/2019 18:03:00   BRONX               46                  0
##  6    198255460 06/07/2019 17:50:00   BROOKLYN            73                  0
##  7    194570529 03/11/2019 16:30:00   BROOKLYN            81                  0
##  8    203211777 10/03/2019 01:45:00   BROOKLYN            67                  0
##  9    193694863 02/17/2019 03:00:00   QUEENS             114                  2
## 10    199582060 07/10/2019 02:56:00   BROOKLYN            69                  0
## # ... with 23,558 more rows, and 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>
```

Transform Data

```
shooting_data <- shooting_data %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE)) %>%
  mutate(OCCUR_DAY_WDAY = wday(OCCUR_DATE, label = TRUE)) %>%
  select(-c(INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD,Y_COORD_CD,Latitude,Longitude,Lon_Lat)
  mutate(OCCUR_YEAR = str_sub(OCCUR_DATE, end = -7))
shooting_data
```

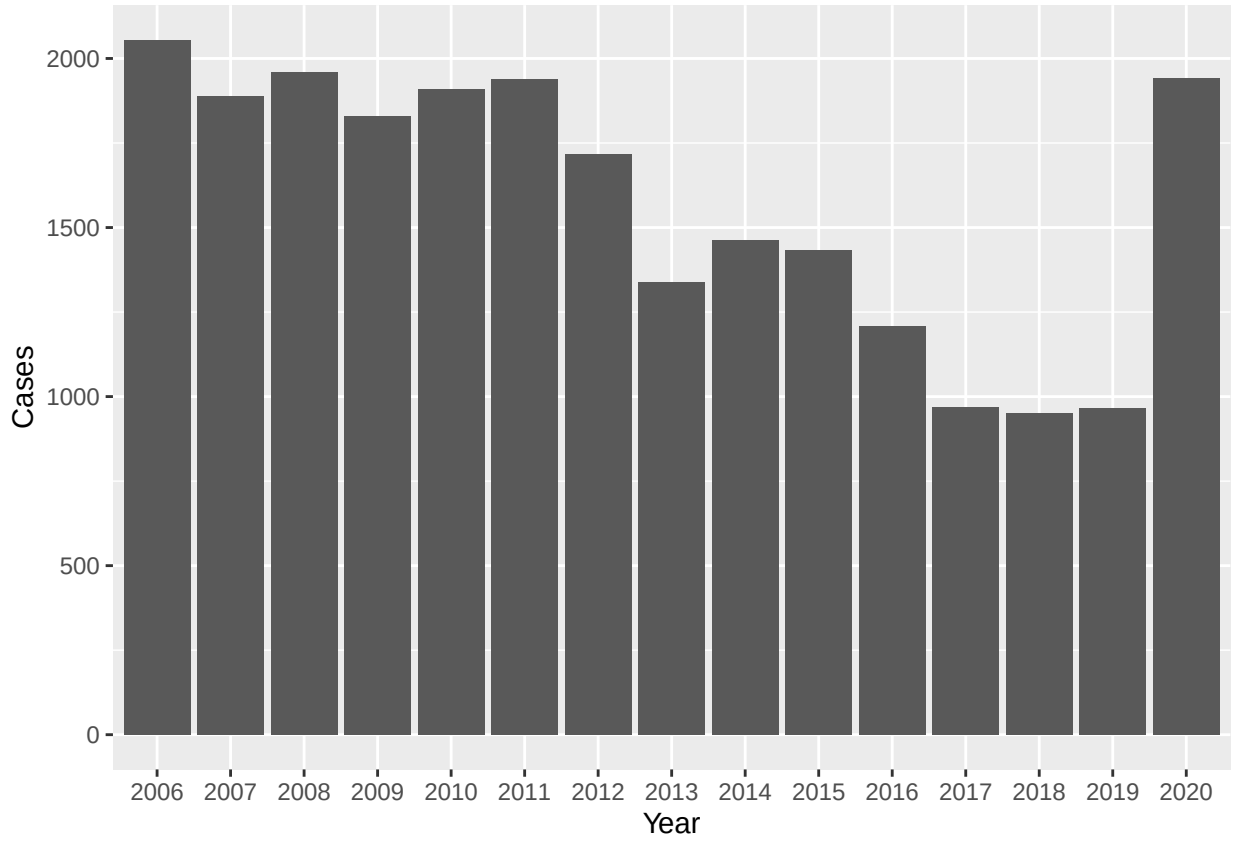
```
## # A tibble: 23,568 x 13
##    OCCUR_DATE OCCUR_TIME BORO          LOCATION_DESC STATISTICAL_MUR~ PERP_AGE_GROUP
##    <date>     <chr>      <chr>         <chr>         <lgl>            <chr>
##  1 2019-08-23 22:10:00   QUEENS        <NA>          FALSE            <NA>
##  2 2019-11-27 15:54:00   BRONX         <NA>          FALSE            <18
##  3 2019-02-02 19:40:00   MANHATTAN     <NA>          FALSE            18-24
##  4 2019-10-24 00:52:00   STATEN ISLAND PVT HOUSE     TRUE             25-44
##  5 2019-08-22 18:03:00   BRONX         <NA>          FALSE            25-44
##  6 2019-06-07 17:50:00   BROOKLYN      <NA>          FALSE            45-64
##  7 2019-03-11 16:30:00   BROOKLYN      <NA>          FALSE            18-24
##  8 2019-10-03 01:45:00   BROOKLYN      MULTI DWELL ~ TRUE             <NA>
##  9 2019-02-17 03:00:00   QUEENS        MULTI DWELL ~ FALSE            18-24
## 10 2019-07-10 02:56:00   BROOKLYN      <NA>          FALSE            25-44
## # ... with 23,558 more rows, and 7 more variables: PERP_SEX <chr>,
## #   PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>, VIC_RACE <chr>,
## #   OCCUR_DAY_WDAY <ord>, OCCUR_YEAR <chr>
```

Shootings: Year-Over-Year Growth

```
shootings_by_year <- shooting_data %>%
  count(OCCUR_YEAR)
shootings_by_year
```

```
## # A tibble: 15 x 2
##    OCCUR_YEAR     n
##    <chr>      <int>
##  1 2006        2055
##  2 2007        1887
##  3 2008        1958
##  4 2009        1828
##  5 2010        1910
##  6 2011        1939
##  7 2012        1717
##  8 2013        1339
##  9 2014        1464
## 10 2015        1434
## 11 2016        1208
## 12 2017         969
## 13 2018         951
## 14 2019         967
## 15 2020        1942
```

```
ggplot(data = shootings_by_year, mapping = aes(x = OCCUR_YEAR, y= n)) +
  geom_bar(stat='identity') +
  labs(y = "Cases", x = "Year")
```

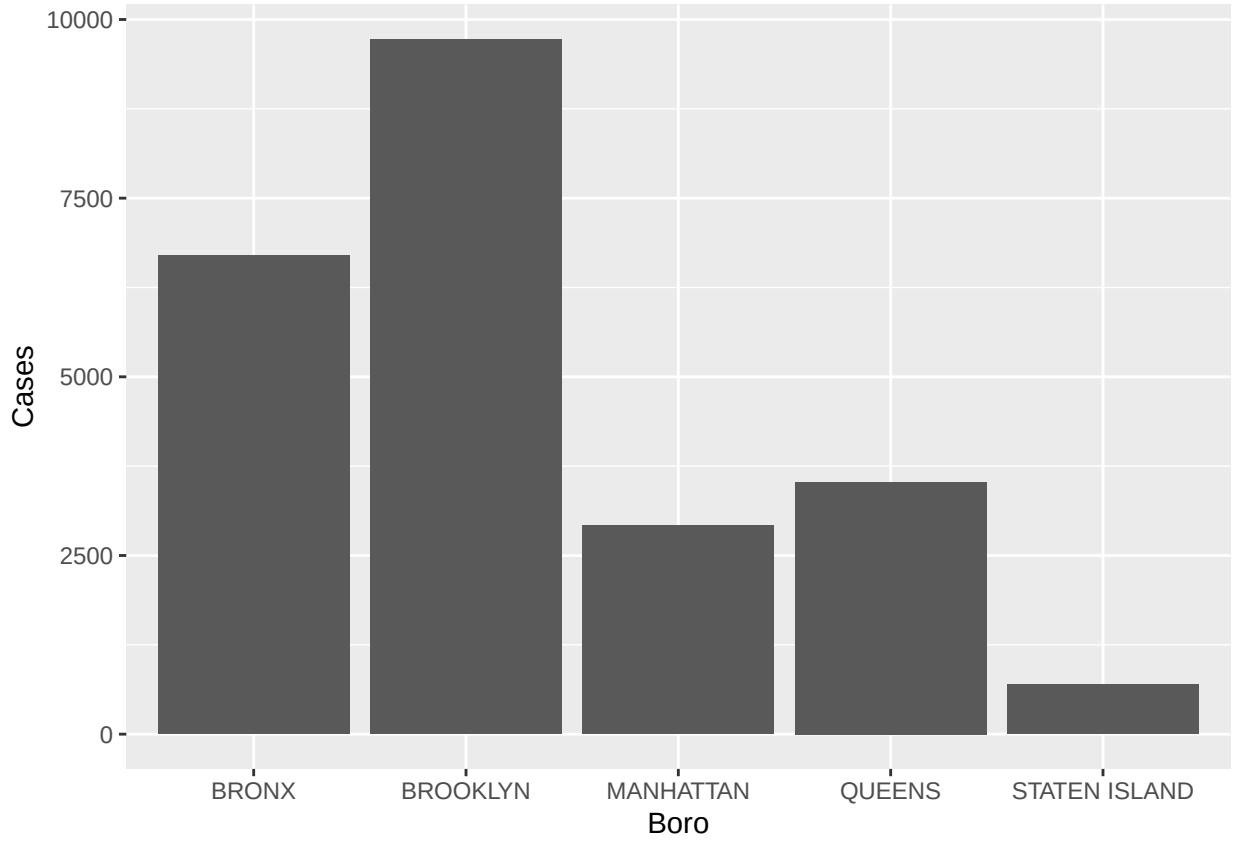


Number of Shooting by Boro

```
shootings_by_boro <- shooting_data %>%
  count(BORO)
shootings_by_boro
```

```
## # A tibble: 5 x 2
##   BORO              n
##   <chr>         <int>
## 1 BRONX          6700
## 2 BROOKLYN       9722
## 3 MANHATTAN      2921
## 4 QUEENS         3527
## 5 STATEN ISLAND   698
```

```
ggplot(data = shootings_by_boro, mapping = aes(x = BORO, y= n)) +
  geom_bar(stat='identity') +
  labs(y = "Cases", x = "Boro")
```

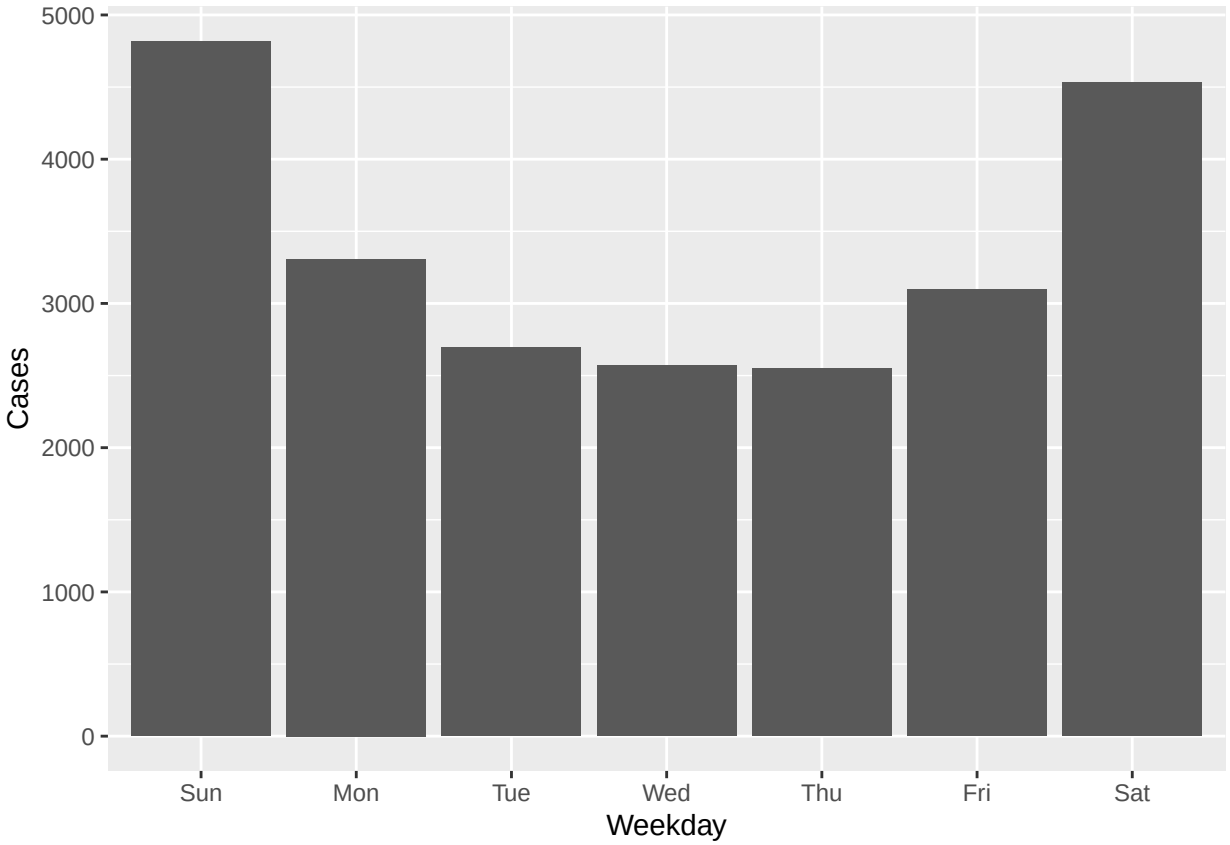


Number of Shooting by Week

```
shootings_by_week <- shooting_data %>%
  count(OCCUR_DAY_WDAY)
shootings_by_week
```

```
## # A tibble: 7 x 2
##   OCCUR_DAY_WDAY     n
##   <ord>          <int>
## 1 Sun             4817
## 2 Mon             3309
## 3 Tue             2694
## 4 Wed             2572
## 5 Thu             2549
## 6 Fri             3095
## 7 Sat             4532
```

```
ggplot(data = shootings_by_week, mapping = aes(x = OCCUR_DAY_WDAY, y= n)) +
  geom_bar(stat='identity') +
  labs(y = "Cases", x = "Weekday")
```


5000
4000
3000
2000
1000
0
Cases
Sun
Mon
Tue
Wed
Thu
Fri
Sat
Weekday