NYPD Shooting Incidents - Analysis

Praveen Myakala

2023-10-23

Get NYPD Shooting Incidents & Display Summary

```
# Download data
nypd_shooting_incidents <- read_csv("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv")
```

NYPD Shooting Incident Data (Historic) from the website: https://catalog.data.gov/datase

```
## Rows: 27312 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
##
## 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.
```

Displaying Summary

```
# Display summary
summary(nypd_shooting_incidents)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:27312       Length:27312       Length:27312
##  1st Qu.: 63860880   Class :character   Class1:hms         Class :character
##  Median : 90372218   Mode  :character   Class2:difftime    Mode  :character
##  Mean   :120860536                      Mode  :numeric
##  3rd Qu.:188810230
##  Max.   :261190187
##
##  LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:27312       Min.   :  1.00   Min.   :0.0000    Length:27312
##  Class :character   1st Qu.: 44.00   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 68.00   Median :0.0000    Mode  :character
##                     Mean   : 65.64   Mean   :0.3269
```

```
##                    3rd Qu.: 81.00   3rd Qu.:0.0000
##                    Max.   :123.00   Max.   :2.0000
##                                     NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:27312       Mode :logical           Length:27312
##  Class :character   FALSE:22046             Class :character
##  Mode  :character   TRUE :5266              Mode  :character
##
##
##
##
##    PERP_SEX           PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:27312       Length:27312       Length:27312       Length:27312
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_RACE           X_COORD_CD        Y_COORD_CD        Latitude
##  Length:27312       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000029   1st Qu.:182834   1st Qu.:40.67
##  Mode  :character   Median :1007731   Median :194487   Median :40.70
##                     Mean   :1009449   Mean   :208127   Mean   :40.74
##                     3rd Qu.:1016838   3rd Qu.:239518   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :10
##    Longitude        Lon_Lat
##  Min.   :-74.25   Length:27312
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :10
```

Preparation and Cleaning

```
# Distinct data with Mutate
nypd_shooting_incidents <- nypd_shooting_incidents %>%
  distinct(INCIDENT_KEY, .keep_all = TRUE) %>%
  select(INCIDENT_KEY, PRECINCT, OCCUR_DATE, OCCUR_TIME, BORO, PERP_AGE_GROUP, PERP_SEX, PERP_RACE, VIC
  mutate(OCCUR_DATE = mdy(OCCUR_DATE)) %>%
  mutate(YEAR = year(OCCUR_DATE))
```

Mainly removing duplicates by INCIDENT_KEY and selecting only required columns to analysis with mutate

Summary after cleaning

```
# Displaying summary
summary(nypd_shooting_incidents)
```

```
##   INCIDENT_KEY          PRECINCT        OCCUR_DATE          OCCUR_TIME
##  Min.   :  9953245   Min.   :  1.00   Min.   :2006-01-01   Length:21420
##  1st Qu.: 64394528   1st Qu.: 44.00   1st Qu.:2009-08-02   Class1:hms
##  Median : 91165008   Median : 69.00   Median :2013-06-14   Class2:difftime
##  Mean   :121166392   Mean   : 66.12   Mean   :2014-01-17   Mode  :numeric
##  3rd Qu.:188062788   3rd Qu.: 81.00   3rd Qu.:2018-09-25
##  Max.   :261190187   Max.   :123.00   Max.   :2022-12-31
##      BORO           PERP_AGE_GROUP      PERP_SEX          PERP_RACE
##  Length:21420       Length:21420       Length:21420       Length:21420
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  VIC_AGE_GROUP        VIC_RACE           VIC_SEX              YEAR
##  Length:21420       Length:21420       Length:21420       Min.   :2006
##  Class :character   Class :character   Class :character   1st Qu.:2009
##  Mode  :character   Mode  :character   Mode  :character   Median :2013
##                                                           Mean   :2014
##                                                           3rd Qu.:2018
##                                                           Max.   :2022
```

```
# Grouping incidents by BORO
shootings_incidents_by_boro <- nypd_shooting_incidents %>%
  group_by(BORO) %>%
  summarize(incidents = n())
```

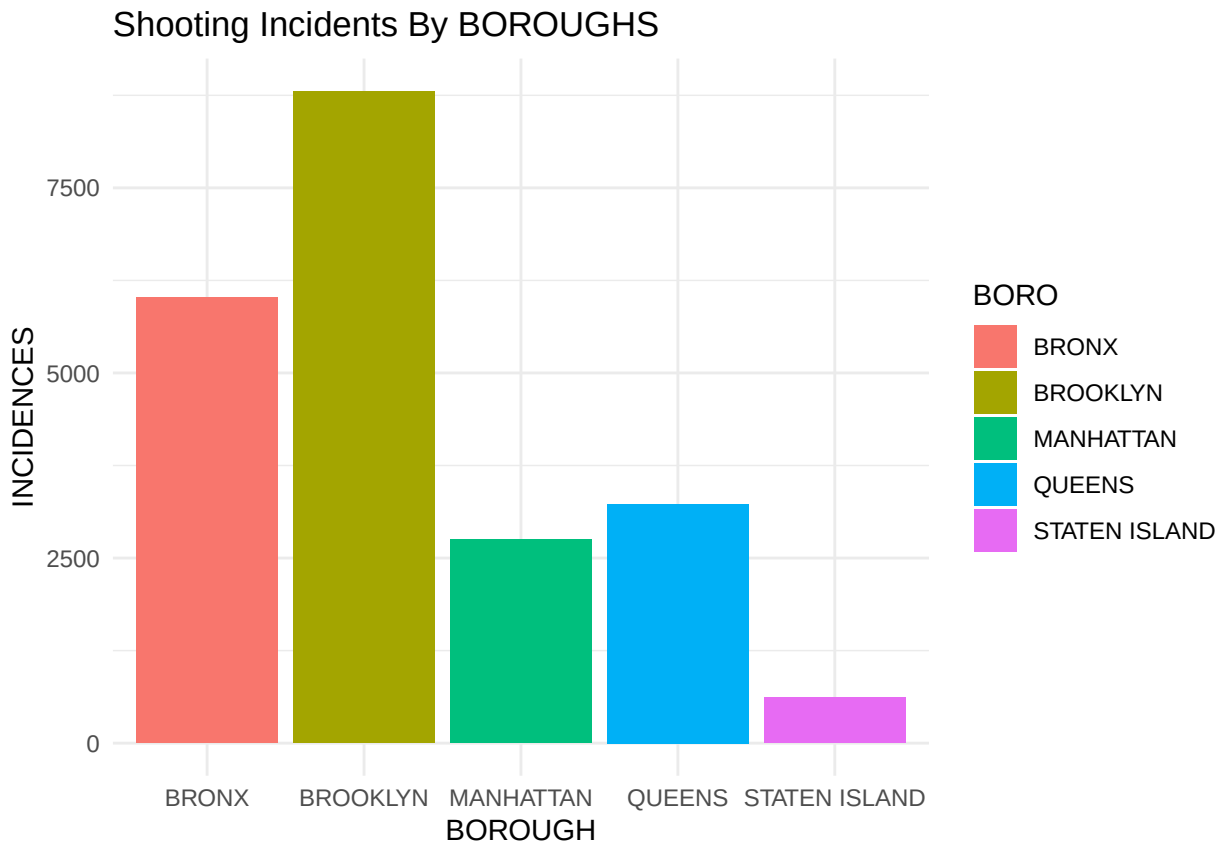
Group by BORO and calculate the sum of incidents

Displaying Summary after grouping incidents by BORO

```
# Summary
summary(shootings_incidents_by_boro)
```

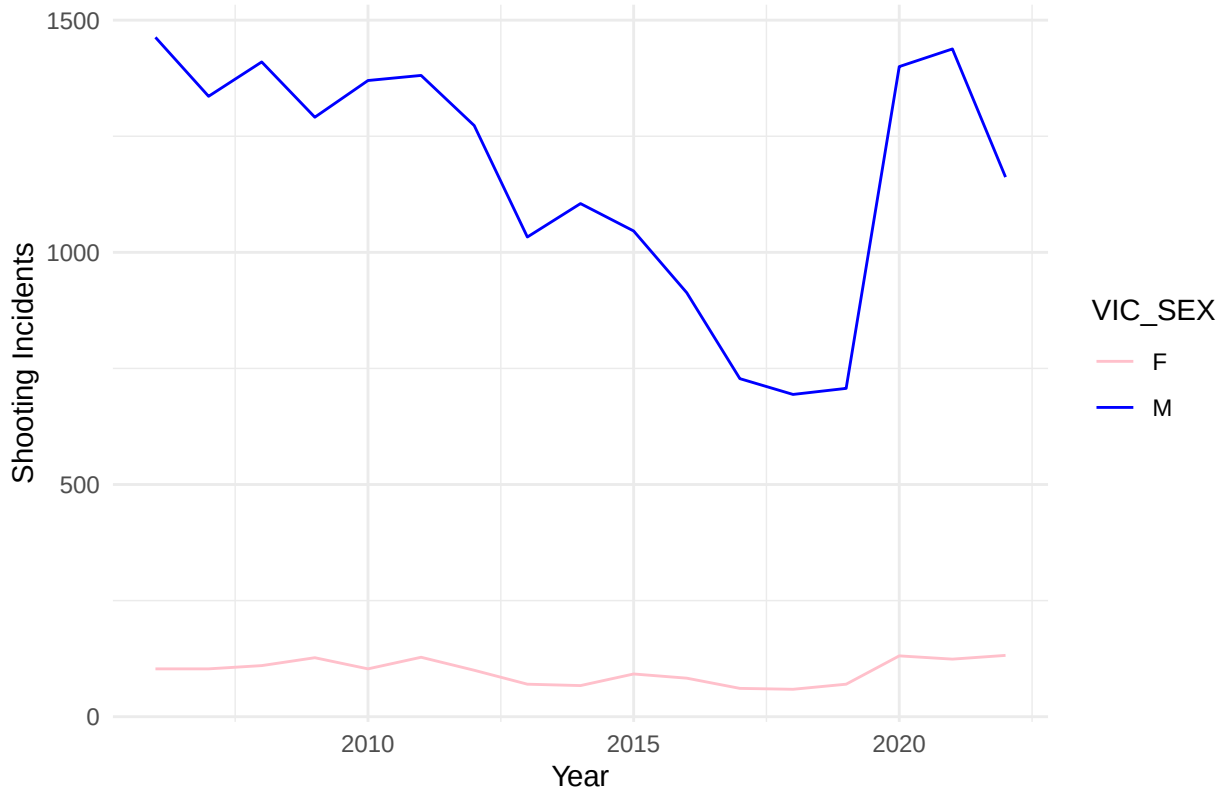
```
##      BORO             incidents
##  Length:5           Min.   : 619
##  Class :character   1st Qu.:2747
##  Mode  :character   Median :3229
##                     Mean   :4284
##                     3rd Qu.:6019
##                     Max.   :8806
```

```
# ggplot code for bar graph

ggplot(shootings_incidents_by_boro, aes(x=BORO, y=incidents, fill=BORO)) +
  geom_bar(stat="identity") +
  xlab("BOROUGH") + ylab("INCIDENCES") +
  ggtitle("Shooting Incidents By BOROUGHS") +
  theme_minimal()
```



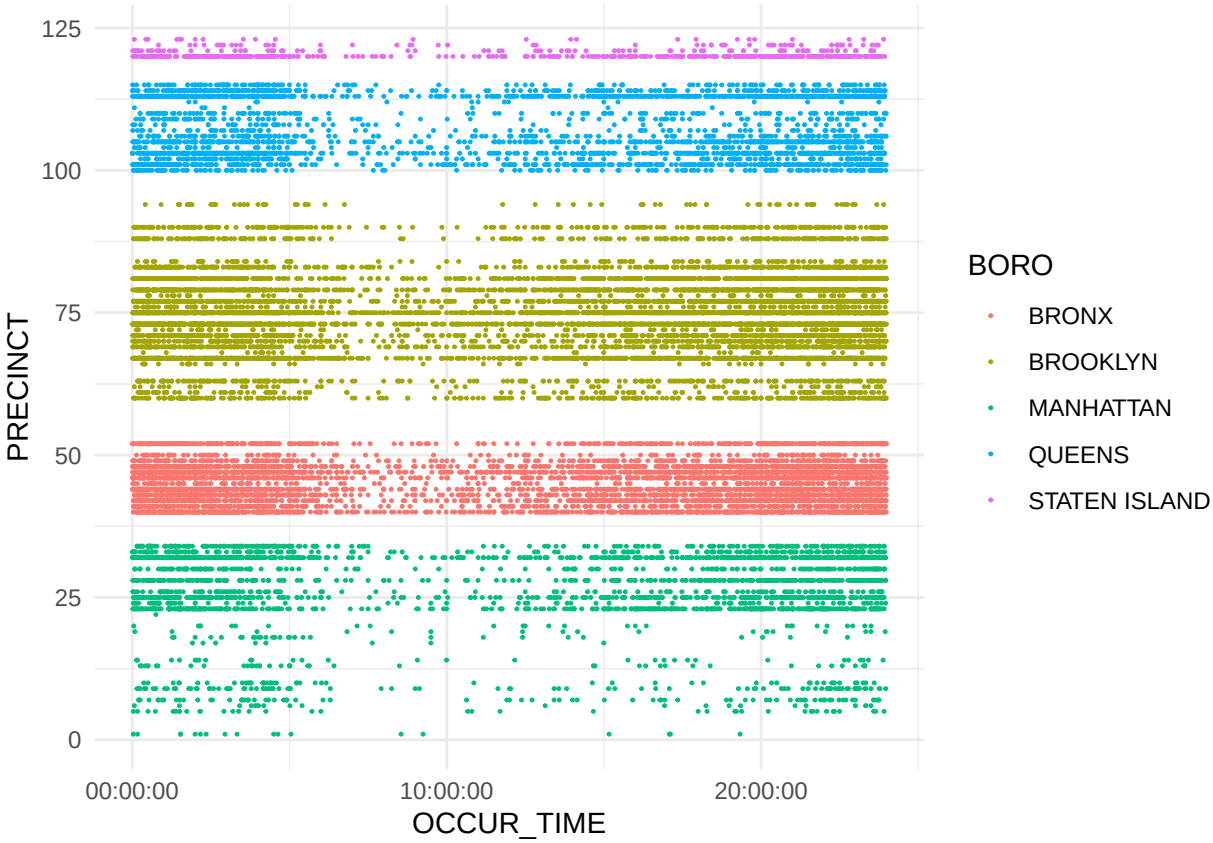
NYPD Shooting incidents by victim gender

```
# Graph with victim gender

nypd_shooting_incidents %>%
  filter(VIC_SEX %in% c("F", "M")) %>%
  count(YEAR, VIC_SEX) %>%
  ggplot(aes(x = YEAR, y = n, color = VIC_SEX)) +
  geom_line() +
  labs(x = "Year", y = "Shooting Incidents", title = "") +
  theme_minimal() +
  scale_color_manual(values = c("M" = "blue", "F" = "pink")) +
  theme_minimal()
```



NYPD Shooting incidents over the course of the day by precint

```
# Graph incidents over time period

ggplot(nypd_shooting_incidents, aes(x = OCCUR_TIME, y = PRECINCT, color = BORO)) +
  geom_point(size = 0.2) +
  theme_minimal()
```


125
100
75
50
25
0
PRECINCT
00:00:00
10:00:00
20:00:00
OCCUR_TIME
BORO
BRONX
BROOKLYN
MANHATTAN
QUEENS
STATEN ISLAND