NYPD_data_Rmd

Ron M

2025-06-01

```
library(readr)


options(width = 120)

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


nypd_data <- read_csv(url)
```

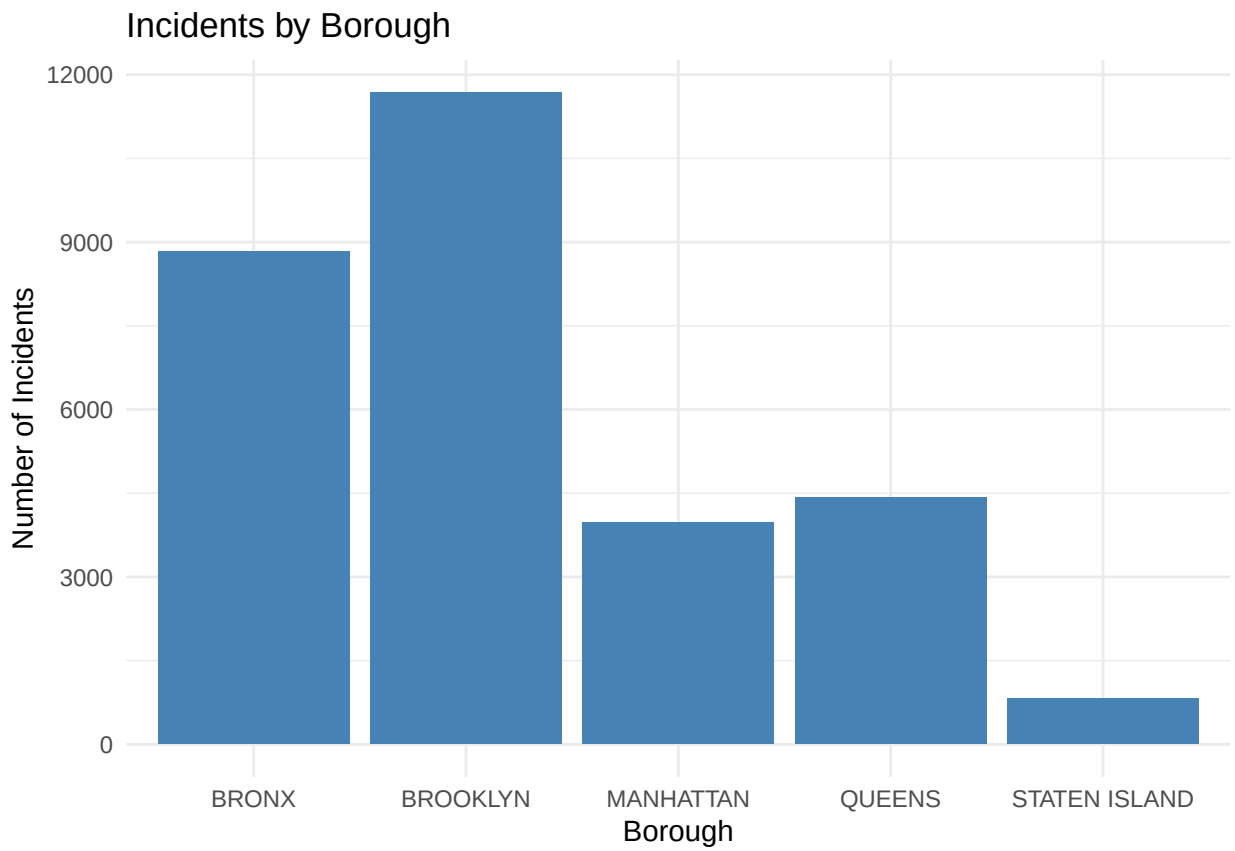
```
## Rows: 29744 Columns: 21
## -- Column specification --------------------------------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION_DESC, PERP_AGE_GROUP, PE
## dbl   (5): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, Latitude, Longitude
## num   (2): 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.
```

```
print(summary(nypd_data), width = 100)
```

```
##   INCIDENT_KEY        OCCUR_DATE          OCCUR_TIME                     BORO
##  Min.   :  9953245   Length:29744       Min.   :00:00:00.000000   Length:29744
##  1st Qu.: 67321140   Class :character   1st Qu.:03:30:45.000000   Class :character
##  Median :109291972   Mode  :character   Median :15:15:00.000000   Mode  :character
##  Mean   :133850951                      Mean   :12:46:10.874798
##  3rd Qu.:214741917                      3rd Qu.:20:44:00.000000
##  Max.   :299462478                      Max.   :23:59:00.000000
##
##  LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC LOCATION_DESC
##  Length:29744       Min.   :  1.00   Min.   :0.0000    Length:29744       Length:29744
##  Class :character   1st Qu.: 44.00   1st Qu.:0.0000    Class :character   Class :character
##  Mode  :character   Median : 67.00   Median :0.0000    Mode  :character   Mode  :character
##                     Mean   : 65.23   Mean   :0.3181
##                     3rd Qu.: 81.00   3rd Qu.:0.0000
##                     Max.   :123.00   Max.   :2.0000
##                                      NA's   :2
##  STATISTICAL_MURDER_FLAG PERP_AGE_GROUP       PERP_SEX          PERP_RACE
```

```
##  Mode :logical           Length:29744       Length:29744       Length:29744
##  FALSE:23979             Class :character   Class :character   Class :character
##  TRUE :5765              Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  VIC_AGE_GROUP        VIC_SEX            VIC_RACE           X_COORD_CD        Y_COORD_CD
##  Length:29744       Length:29744       Length:29744       Min.   : 914928   Min.   :125757
##  Class :character   Class :character   Class :character   1st Qu.:1000094   1st Qu.:183042
##  Mode  :character   Mode  :character   Mode  :character   Median :1007826   Median :195506
##                                                           Mean   :1009442   Mean   :208722
##                                                           3rd Qu.:1016739   3rd Qu.:239980
##                                                           Max.   :1066815   Max.   :271128
##
##     Latitude       Longitude        Lon_Lat
##  Min.   :40.51   Min.   :-74.25   Length:29744
##  1st Qu.:40.67   1st Qu.:-73.94   Class :character
##  Median :40.70   Median :-73.91   Mode  :character
##  Mean   :40.74   Mean   :-73.91
##  3rd Qu.:40.83   3rd Qu.:-73.88
##  Max.   :40.91   Max.   :-73.70
##  NA's   :97      NA's   :97
```

```r
library(ggplot2)
ggplot(nypd_data, aes(x= BORO))+
  geom_bar(fill = "steelblue")+
  labs(
  title = "Incidents by Borough",
  x = "Borough",
  y = "Number of Incidents"
  )+
theme_minimal()
```




```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

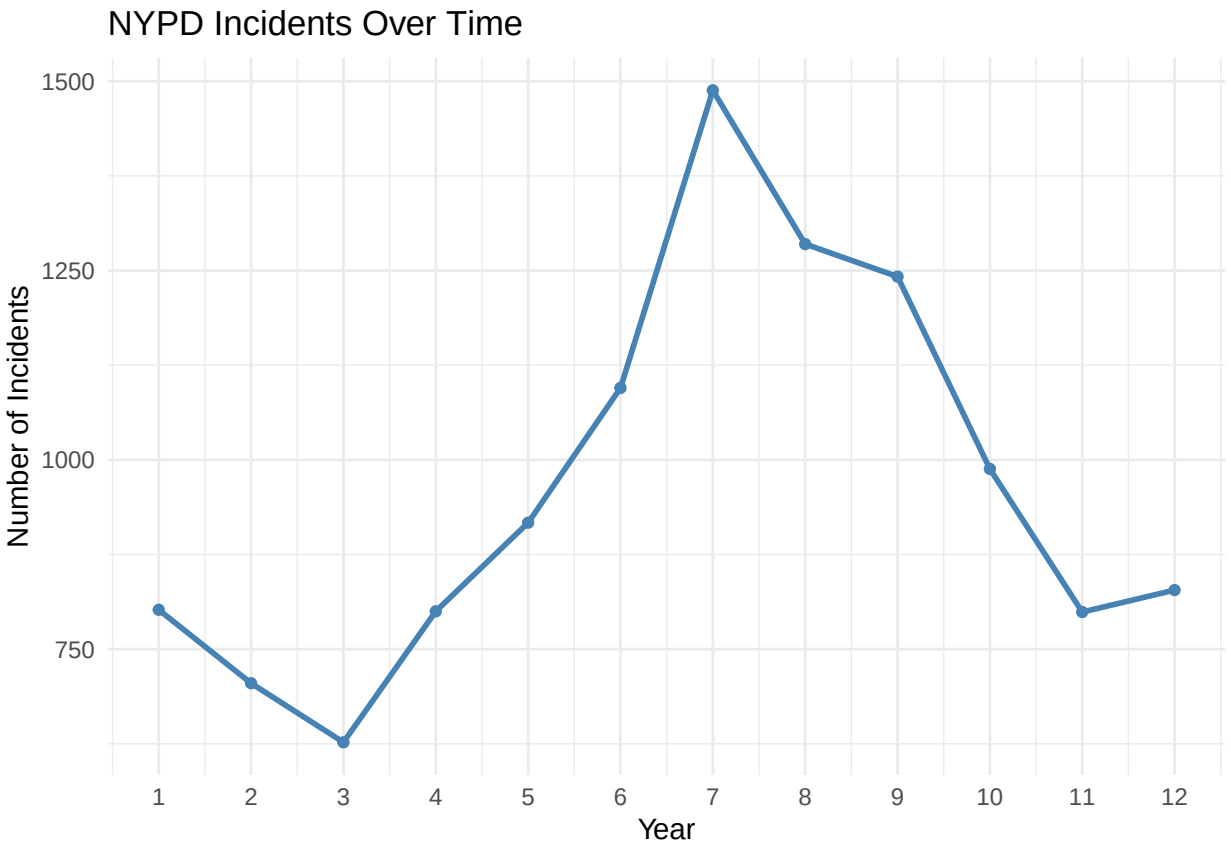
```r
library(ggplot2)

nypd_data <- nypd_data %>%
  mutate(OCCUR_DATE = as.Date(OCCUR_DATE)) %>%
  filter(!is.na(OCCUR_DATE)) %>%
  mutate(Year = year(OCCUR_DATE))


yearly_counts <- nypd_data %>%
  group_by(Year) %>%
  summarise(Incidents = n()) %>%
  filter(!is.na(Year))  # Just in case


ggplot(yearly_counts, aes(x = Year, y = Incidents)) +
  geom_line(color = "steelblue", size = 1) +
  geom_point(color = "steelblue") +
  scale_x_continuous(breaks = yearly_counts$Year) +
  labs(
    title = "NYPD Incidents Over Time",
    x = "Year",
    y = "Number of Incidents"
  ) +
  theme_minimal()
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was generated.
```




```
#Bias Identification:
 # I do not have any sort of bias towards this project or data set. This is not
#a data set I have chosen but it is the data set that was provided for this
#project. I do not feel any way regarding this data. I understand the
#importance of identifying and accounting for your personal biases during
#projects like this one but it does not pertain to this specific example.
```