NYPD_data_Rmd

Ron M

2025-06-01

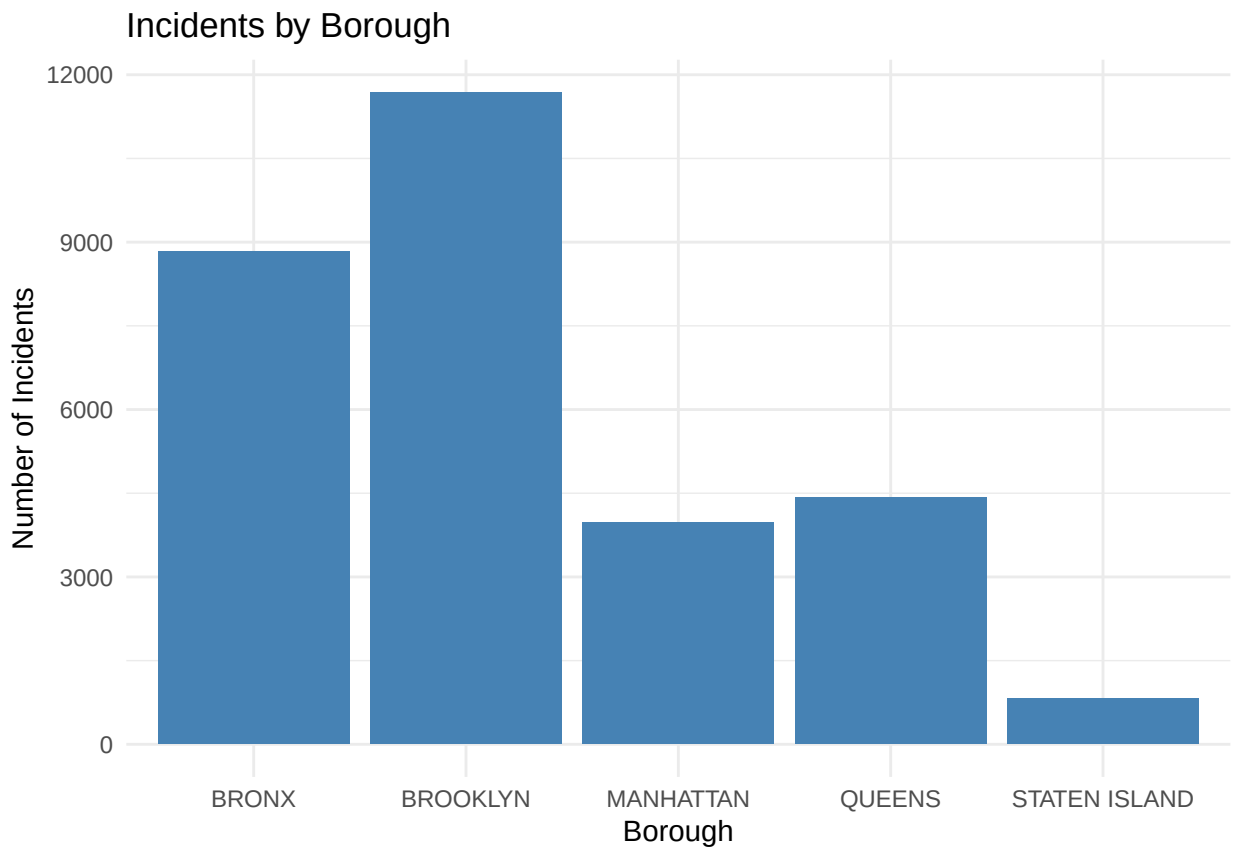
```
library(readr)
library(dplyr)
library(lubridate)
library(ggplot2)

options(width = 120)


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


nypd_data <- nypd_data %>%
  mutate(
    OCCUR_DATE = as.Date(OCCUR_DATE, format = "%m/%d/%Y"),
    Year = year(OCCUR_DATE)
  ) %>%
  filter(!is.na(OCCUR_DATE))
```

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

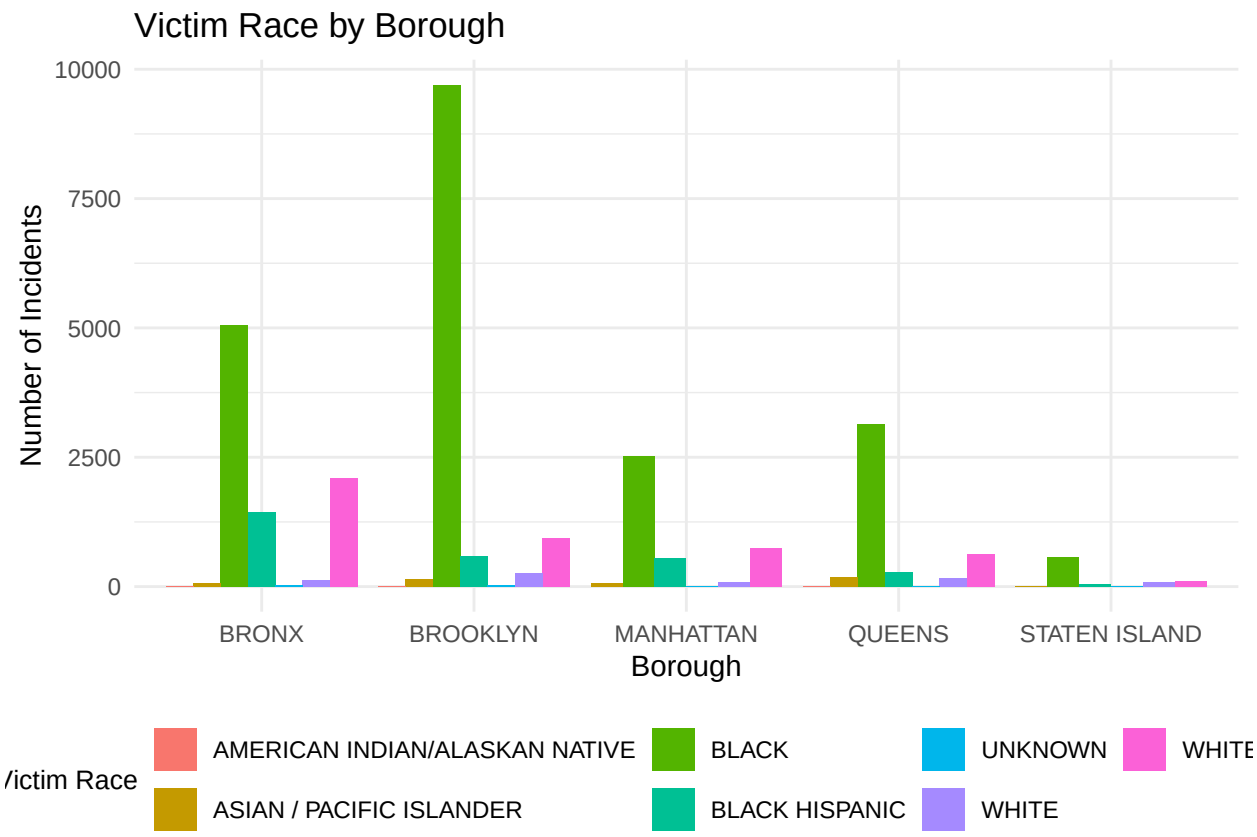
```
##   INCIDENT_KEY         OCCUR_DATE           OCCUR_TIME                    BORO
##  Min.   :  9953245   Min.   :2006-01-01   Min.   :00:00:00.000000   Length:29744
##  1st Qu.: 67321140   1st Qu.:2009-10-29   1st Qu.:03:30:45.000000   Class :character
##  Median :109291972   Median :2014-03-25   Median :15:15:00.000000   Mode  :character
##  Mean   :133850951   Mean   :2014-10-31   Mean   :12:46:10.874798
##  3rd Qu.:214741917   3rd Qu.:2020-06-29   3rd Qu.:20:44:00.000000
##  Max.   :299462478   Max.   :2024-12-31   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               Year
##  Min.   :40.51   Min.   :-74.25   Length:29744       Min.   :2006
##  1st Qu.:40.67   1st Qu.:-73.94   Class :character   1st Qu.:2009
##  Median :40.70   Median :-73.91   Mode  :character   Median :2014
##  Mean   :40.74   Mean   :-73.91                      Mean   :2014
##  3rd Qu.:40.83   3rd Qu.:-73.88                      3rd Qu.:2020
##  Max.   :40.91   Max.   :-73.70                      Max.   :2024
##  NA's   :97      NA's   :97
```

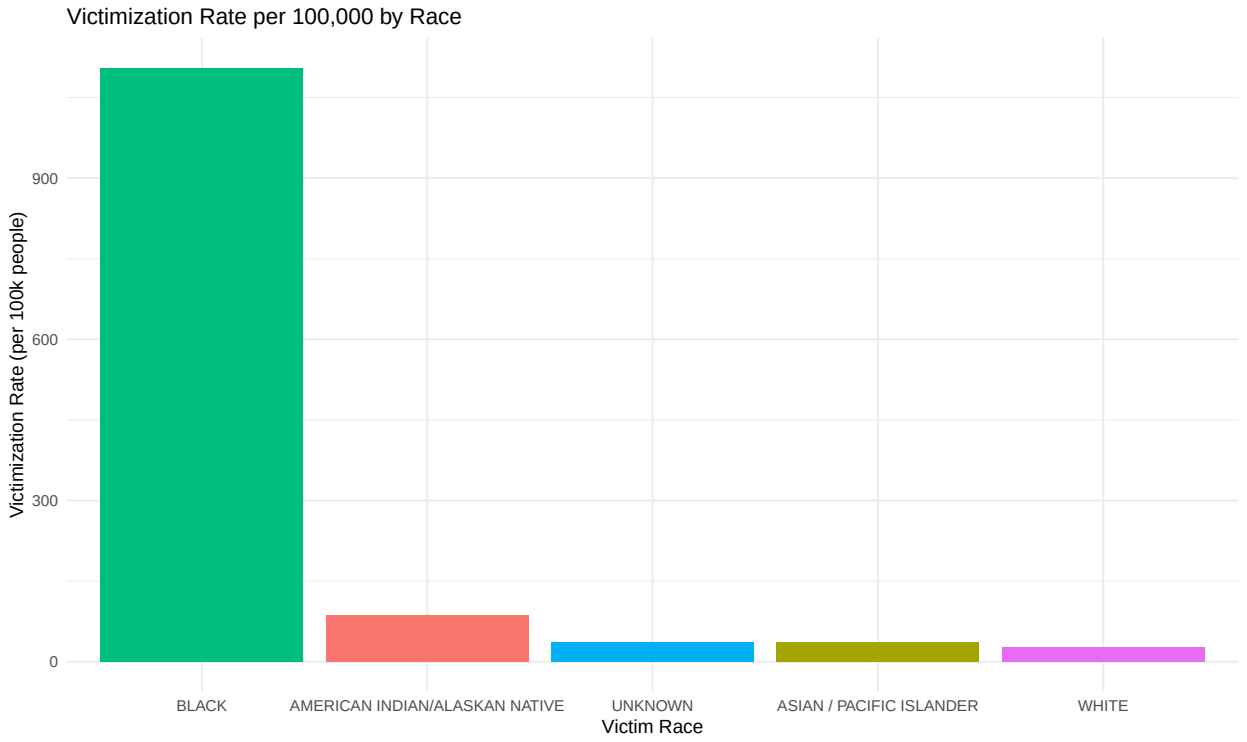
```r
ggplot(nypd_data, aes(x = BORO)) +
  geom_bar(fill = "steelblue") +
  labs(
    title = "Incidents by Borough",
    x = "Borough",
    y = "Number of Incidents"
  ) +
  theme_minimal()
```




```
race_by_boro <- nypd_data %>%
  filter(!is.na(VIC_RACE), !is.na(BORO)) %>%
  group_by(BORO, VIC_RACE) %>%
  summarise(Count = n(), .groups = "drop") %>%
  arrange(desc(Count))

ggplot(race_by_boro, aes(x = BORO, y = Count, fill = VIC_RACE)) +
  geom_col(position = "dodge") +
  labs(
    title = "Victim Race by Borough",
    x = "Borough",
    y = "Number of Incidents",
    fill = "Victim Race"
  ) +
  theme_minimal() +
  theme(
    legend.position = "bottom",
    legend.title = element_text(size = 10),
    axis.text.x = element_text(angle = 0, hjust = 0.5)
  )
```

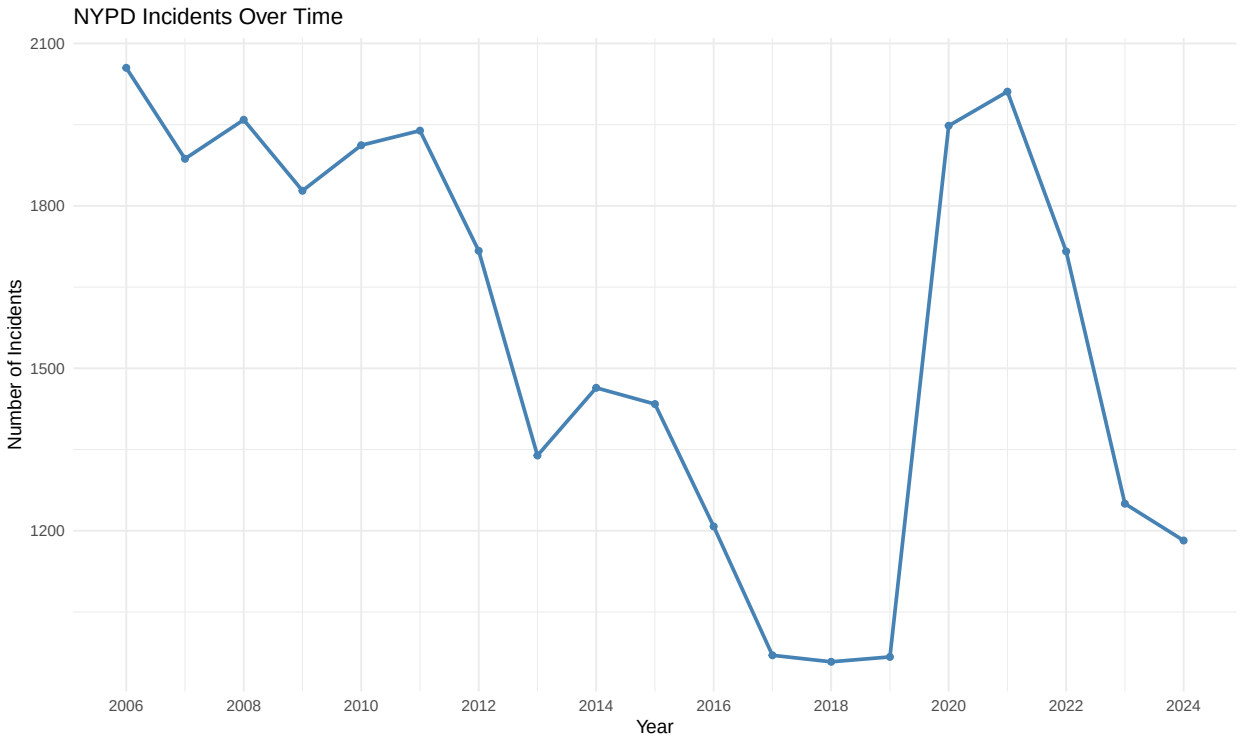



```r
# NYC Race population estimates
race_population <- tibble(
  VIC_RACE = c("BLACK", "WHITE", "ASIAN / PACIFIC ISLANDER", "HISPANIC", "AMERICAN INDIAN/ALASKAN NATIVE
  Population = c(1900000, 2700000, 1350000, 2400000, 15000, 200000)
)


victim_counts <- nypd_data %>%
  filter(!is.na(VIC_RACE)) %>%
  group_by(VIC_RACE) %>%
  summarise(Count = n(), .groups = "drop")


victim_rates <- victim_counts %>%
  inner_join(race_population, by = "VIC_RACE") %>%
  mutate(Rate_per_100k = (Count / Population) * 100000)


ggplot(victim_rates, aes(x = reorder(VIC_RACE, -Rate_per_100k), y = Rate_per_100k, fill = VIC_RACE)) +
  geom_col(show.legend = FALSE) +
  labs(
    title = "Victimization Rate per 100,000 by Race",
    x = "Victim Race",
    y = "Victimization Rate (per 100k people)"
  ) +
  theme_minimal()
```

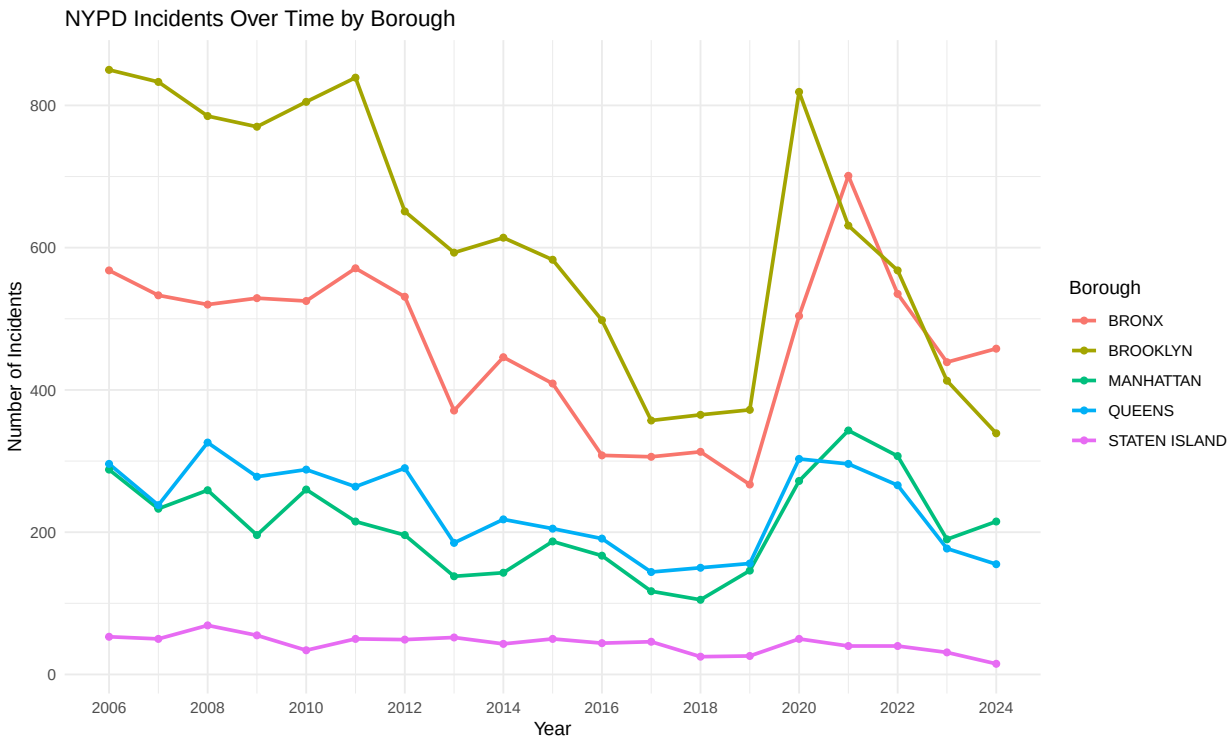



```
yearly_counts <- nypd_data %>%
  group_by(Year) %>%
  summarise(Incidents = n()) %>%
  arrange(Year)



ggplot(yearly_counts, aes(x = Year, y = Incidents)) +
  geom_line(color = "steelblue", size = 1) +
  geom_point(color = "steelblue") +
  scale_x_continuous(breaks = seq(min(yearly_counts$Year), max(yearly_counts$Year), 2)) +
  labs(
    title = "NYPD Incidents Over Time",
    x = "Year",
    y = "Number of Incidents"
  ) +
  theme_minimal()
```

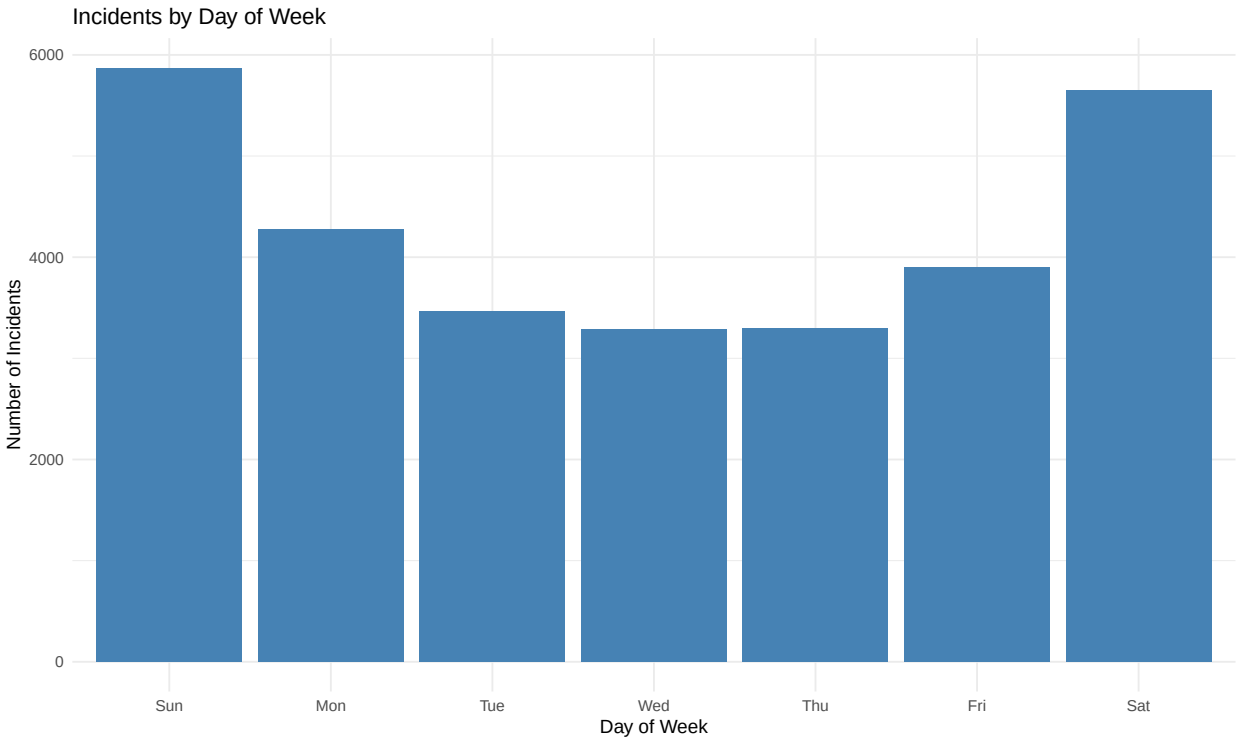
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was generated.
```




```r
borough_trends <- nypd_data %>%
  group_by(Year, BORO) %>%
  summarise(Incidents = n(), .groups = "drop")

# Plot
ggplot(borough_trends, aes(x = Year, y = Incidents, color = BORO)) +
  geom_line(size = 1) +
  geom_point() +
  scale_x_continuous(breaks = seq(min(borough_trends$Year), max(borough_trends$Year), 2)) +
  labs(
    title = "NYPD Incidents Over Time by Borough",
    x = "Year",
    y = "Number of Incidents",
    color = "Borough"
  ) +
  theme_minimal()
```




```r
nypd_data <- nypd_data %>%
  mutate(Weekday = wday(OCCUR_DATE, label = TRUE))

weekday_summary <- nypd_data %>%
  group_by(Weekday) %>%
  summarise(Incidents = n())

ggplot(weekday_summary, aes(x = Weekday, y = Incidents)) +
  geom_col(fill = "steelblue") +
  labs(
    title = "Incidents by Day of Week",
    x = "Day of Week",
    y = "Number of Incidents"
  ) +
  theme_minimal()
```




```
#Bias Identification:
 # Racial biased is the most prevalent bias within this study. The data shows
#blacks are disproportionately more affected by gun violence in New YorK City than
# any other race. In order to account for this bias, I included general census
#data for New York City from 2020.
```