Week 3 Project

Aja Starkey

2024-09-17

Importing Data / Editing dataset

```
data <- read.csv("https://data.cityofnewyork.us/resource/833y-fsy8.csv")
# Review data set with head(data)
colnames(data)
```

```
##  [1] "incident_key"          "occur_date"
##  [3] "occur_time"            "boro"
##  [5] "loc_of_occur_desc"     "precinct"
##  [7] "jurisdiction_code"     "loc_classfctn_desc"
##  [9] "location_desc"         "statistical_murder_flag"
## [11] "perp_age_group"        "perp_sex"
## [13] "perp_race"             "vic_age_group"
## [15] "vic_sex"               "vic_race"
## [17] "x_coord_cd"            "y_coord_cd"
## [19] "latitude"              "longitude"
## [21] "geocoded_column"
```

```
df <-data %>%
  select("occur_date", "occur_time", "boro", "loc_of_occur_desc", "precinct", "loc_classfctn_desc", "lo
```

```
df <- data %>%
  select("occur_date", "occur_time", "boro", "loc_of_occur_desc", "precinct",
         "loc_classfctn_desc", "location_desc", "perp_age_group",
         "perp_sex", "perp_race", "vic_age_group", "vic_sex", "vic_race") %>%
  mutate(
    boro = as.factor(boro),
    loc_of_occur_desc = as.factor(loc_of_occur_desc),
    precinct = as.factor(precinct),
    loc_classfctn_desc = as.factor(loc_classfctn_desc),
    location_desc = as.factor(location_desc),
    perp_age_group = as.factor(perp_age_group),
    perp_sex = as.factor(perp_sex),
    perp_race = as.factor(perp_race),
    vic_age_group = as.factor(vic_age_group),
    vic_sex = as.factor(vic_sex),
    vic_race = as.factor(vic_race)
  )
```

Editing and Adding Variables

```r
df_2024 <- df %>% filter(str_detect(occur_date, "^2024"))
df_2023 <- df %>% filter(str_detect(occur_date, "^2023"))

# Create new column with shooting counts
df <- df %>%
  add_count(boro, name = "boro_count") %>%
  relocate(boro) %>%
  relocate(boro_count, .after = boro)

# Edit time and date variables
df <- df %>%
  mutate(occur_date = gsub("^2023-", "", gsub("T.*", "", occur_date)))

df <- df %>%
  mutate(occur_time = gsub(":00$", "", occur_time))

  monthly_counts <- df %>%

  mutate(month = substr(occur_date, 1, 2)) %>%

  group_by(month, boro) %>%

  summarise(incident_count = n(), .groups = "drop")


december_data <- df %>%
  filter(substr(occur_date, 1, 2) == "12")
# Review dataset with head(df)
```
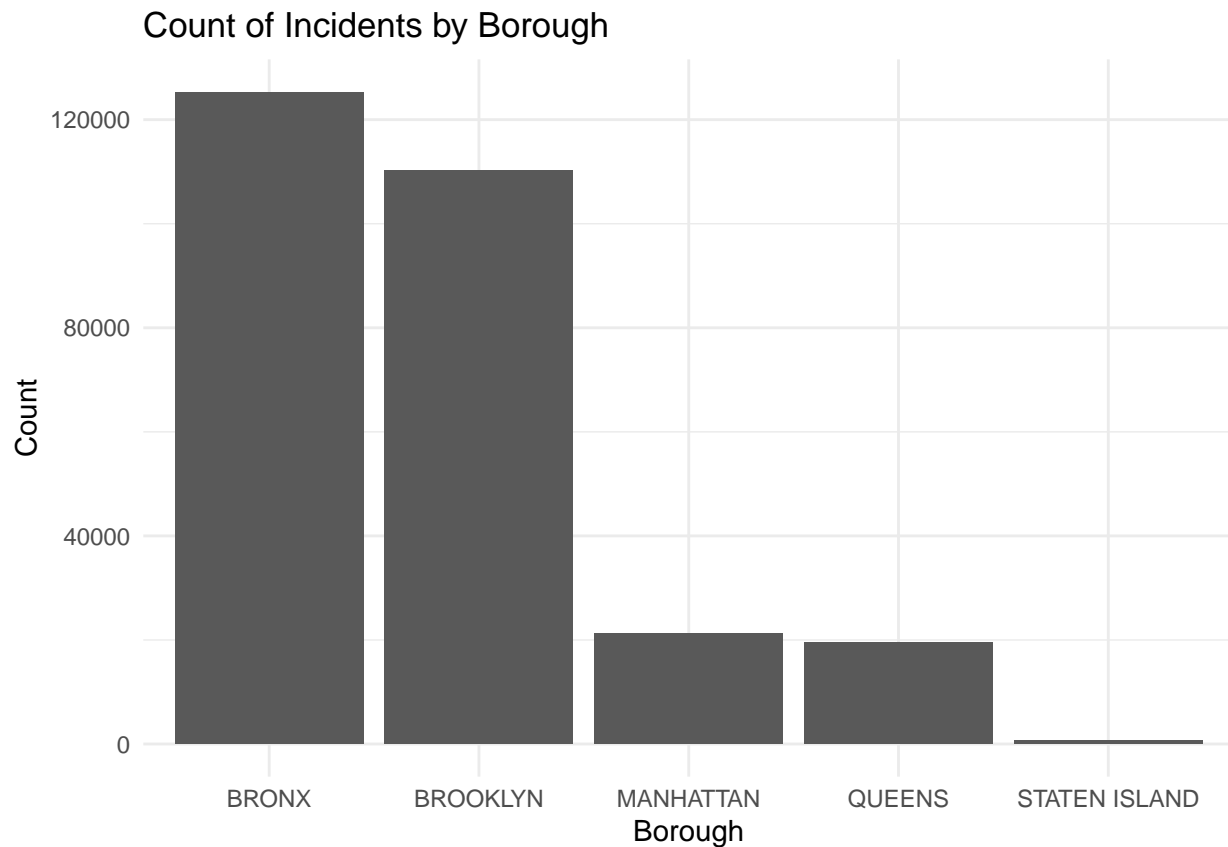
Data Visualizations

```r
ggplot(df, aes(x = boro, y = boro_count)) + geom_bar(stat = "identity") + labs(x = "Borough", y = "Coun
```

Count of Incidents by Borough


```r
bronx_data <- df %>%
  mutate(month = substr(occur_date, 1, 2)) %>%
  filter(boro == "BRONX") %>%
  group_by(month) %>%
  summarise(incident_count = n())

ggplot(bronx_data, aes(x = month, y = incident_count)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Number of Incidents in Bronx per Month", x = "Month", y = "Number of Incidents") +
  theme_minimal()
```
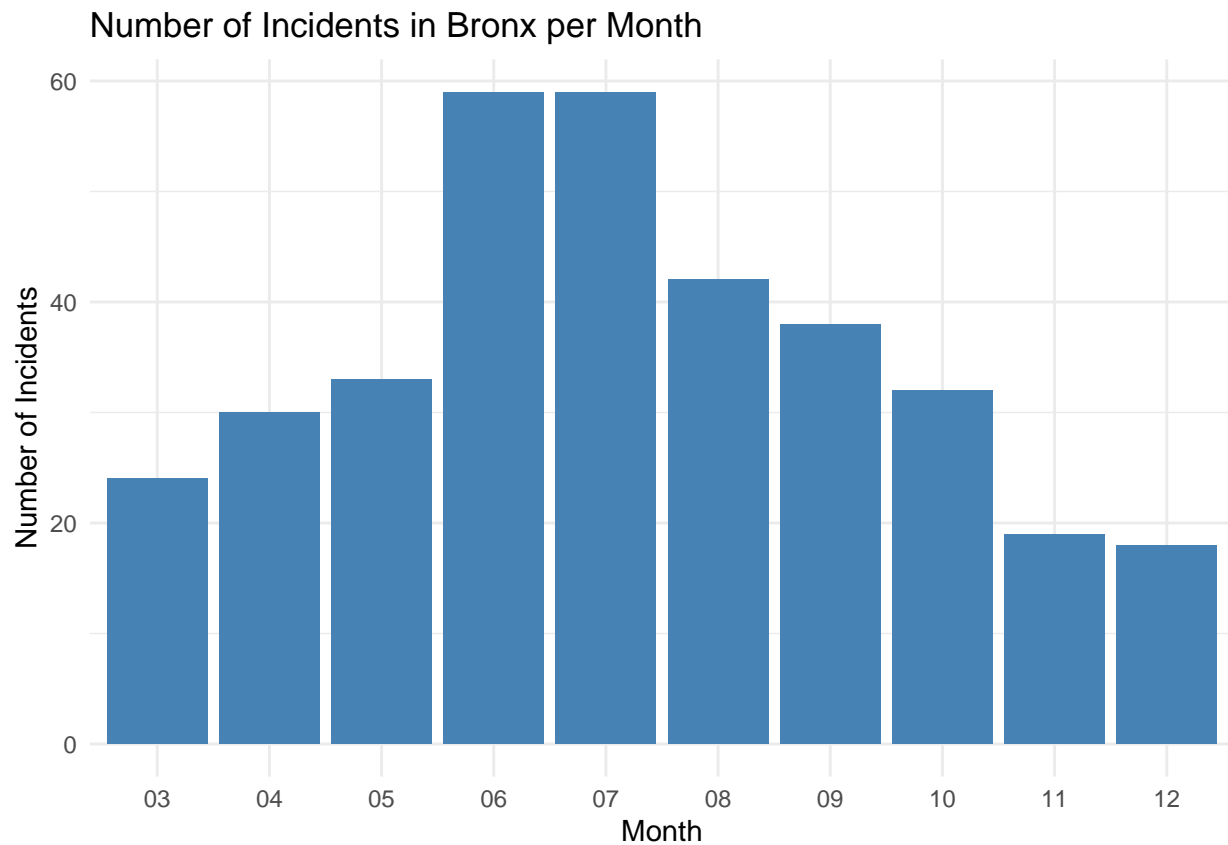
Number of Incidents in Bronx per Month


```
monthly_counts <- df %>%
  mutate(month = substr(occur_date, 1, 2)) %>%
  group_by(month, boro) %>%
  summarise(incident_count = n(), .groups = 'drop')

ggplot(monthly_counts, aes(x = month, y = incident_count, fill = boro)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Number of Incidents per Month by Borough", x = "Month", y = "Number of Incidents") +
  scale_fill_manual(values = c("BRONX" = "steelblue", "MANHATTAN" = "grey",
                               "BROOKLYN" = "pink", "QUEENS" = "purple",
                               "STATEN ISLAND" = "cyan")) +
  theme_minimal()
```

Number of Incidents per Month by Borough
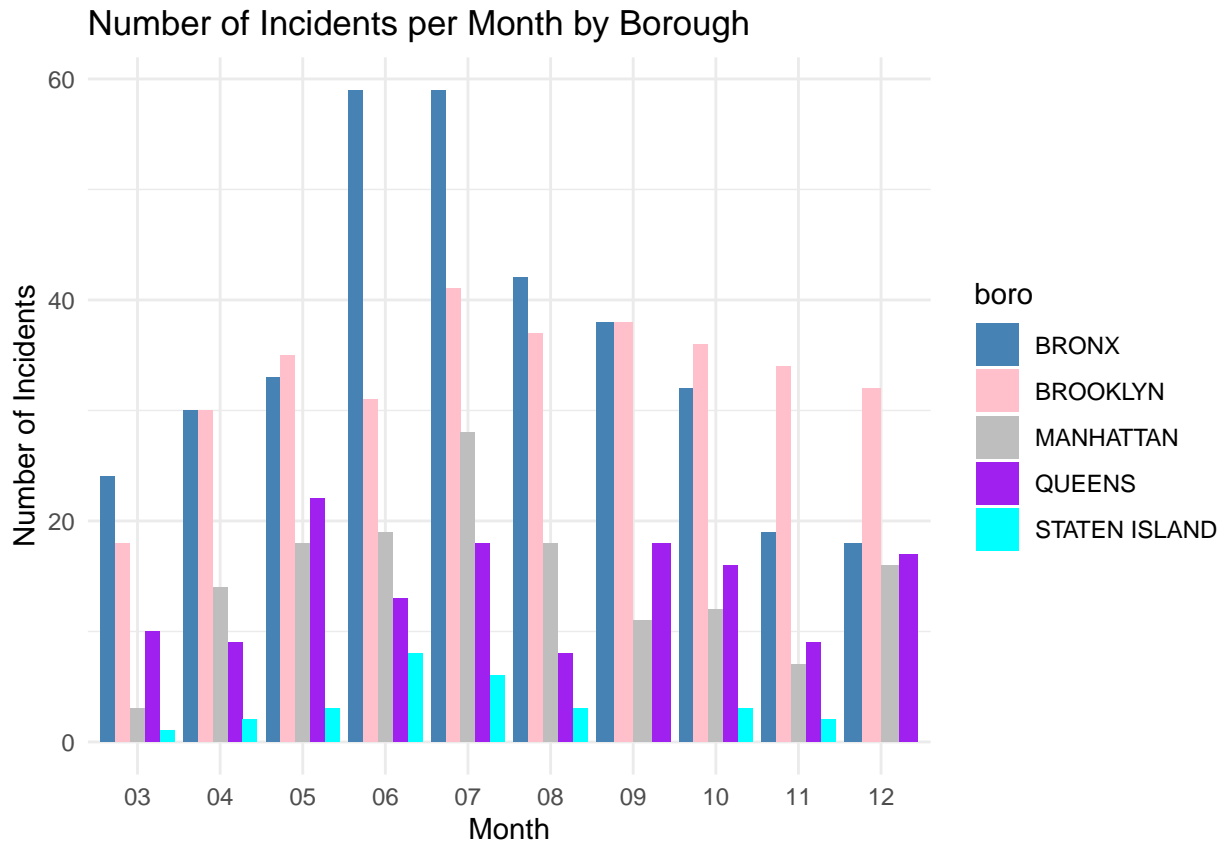
Analysis / Discussion:

The three graphs above raise several questions I'd like to explore further. In the first graph, the Bronx had the highest number of shooting incidents in 2023, while Staten Island had the fewest. This leads me to ask: what is the ratio of shootings to population in each borough?

The second graph highlights the monthly shooting trends in the Bronx, showing an increase in incidents during the summer months (June and July) and a decrease in the winter months (November and December). This prompts me to wonder if the same pattern holds true across all New York City boroughs.

The third graph illustrates shooting incidents per month for all boroughs. Based on this visualization, it may be useful to conduct a statistical analysis to determine if there is a significant difference in shooting incidents between July and December in each borough.

Conclusion:

In this report I wrangled the data by first editing the columns. I took out a few of the spacial variables such as longitude and latitude, so that the focus of the analysis was on each borough as a whole. I then edited the time and date variables, so that they would be easier to interpret, and made a new column to represent the total amount of shooting incidents by borough.

Although I did not have a specific bias while completing this data analysis project, If I was living in New York city and had more context to each neighborhood, then I may have a specific bias against certain neighborhoods if I was aware that they were unsafe.