NYPD

2024-04-30

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

shooting_data <- read_csv(url_in)
```

```
## Rows: 28562 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.
```

```
## Cleaning Data
shooting_data <- shooting_data %>%
  rename_all(tolower) %>%
  rename(date_of_incident = occur_date, time_of_incident = occur_time) %>%
  mutate(date_of_incident = mdy(date_of_incident),
         time_of_incident = hms(time_of_incident),
         latitude = as.numeric(latitude),
         longitude = as.numeric(longitude)) %>%
  filter(!is.na(date_of_incident))
glimpse(shooting_data)
```

```
## Rows: 28,562
## Columns: 21
## $ incident_key            <dbl> 244608249, 247542571, 84967535, 202853370, 270~
## $ date_of_incident        <date> 2022-05-05, 2022-07-04, 2012-05-27, 2019-09-2~
## $ time_of_incident        <Period> 10M 0S, 22H 20M 0S, 19H 35M 0S, 21H 0M 0S, ~
## $ boro                    <chr> "MANHATTAN", "BRONX", "QUEENS", "BRONX", "BROO~
## $ loc_of_occur_desc       <chr> "INSIDE", "OUTSIDE", NA, NA, NA, NA, NA, NA, N~
## $ precinct                <dbl> 14, 48, 103, 42, 83, 23, 113, 77, 48, 49, 73, ~
## $ jurisdiction_code       <dbl> 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ loc_classfctn_desc      <chr> "COMMERCIAL", "STREET", NA, NA, NA, NA, NA, NA~
## $ location_desc           <chr> "VIDEO STORE", "(null)", NA, NA, NA, "MULTI DW~
## $ statistical_murder_flag <lgl> TRUE, TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, ~
## $ perp_age_group          <chr> "25-44", "(null)", NA, "25-44", "25-44", NA, N~
## $ perp_sex                <chr> "M", "(null)", NA, "M", "M", NA, NA, NA, NA, "~
## $ perp_race               <chr> "BLACK", "(null)", NA, "UNKNOWN", "BLACK", NA,~
## $ vic_age_group           <chr> "25-44", "18-24", "18-24", "25-44", "25-44", "~
## $ vic_sex                 <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "~
```

```
## $ vic_race                <chr> "BLACK", "BLACK", "BLACK", "BLACK", "BLACK", "~
## $ x_coord_cd              <dbl> 986050, 1016802, 1048632, 1014493, 1009149, 99~
## $ y_coord_cd              <dbl> 214231.0, 250581.0, 198262.0, 242565.0, 190104~
## $ latitude                <dbl> 40.75469, 40.85440, 40.71063, 40.83242, 40.688~
## $ longitude               <dbl> -73.99350, -73.88233, -73.76777, -73.89071, -7~
## $ lon_lat                 <chr> "POINT (-73.9935 40.754692)", "POINT (-73.8823~
```

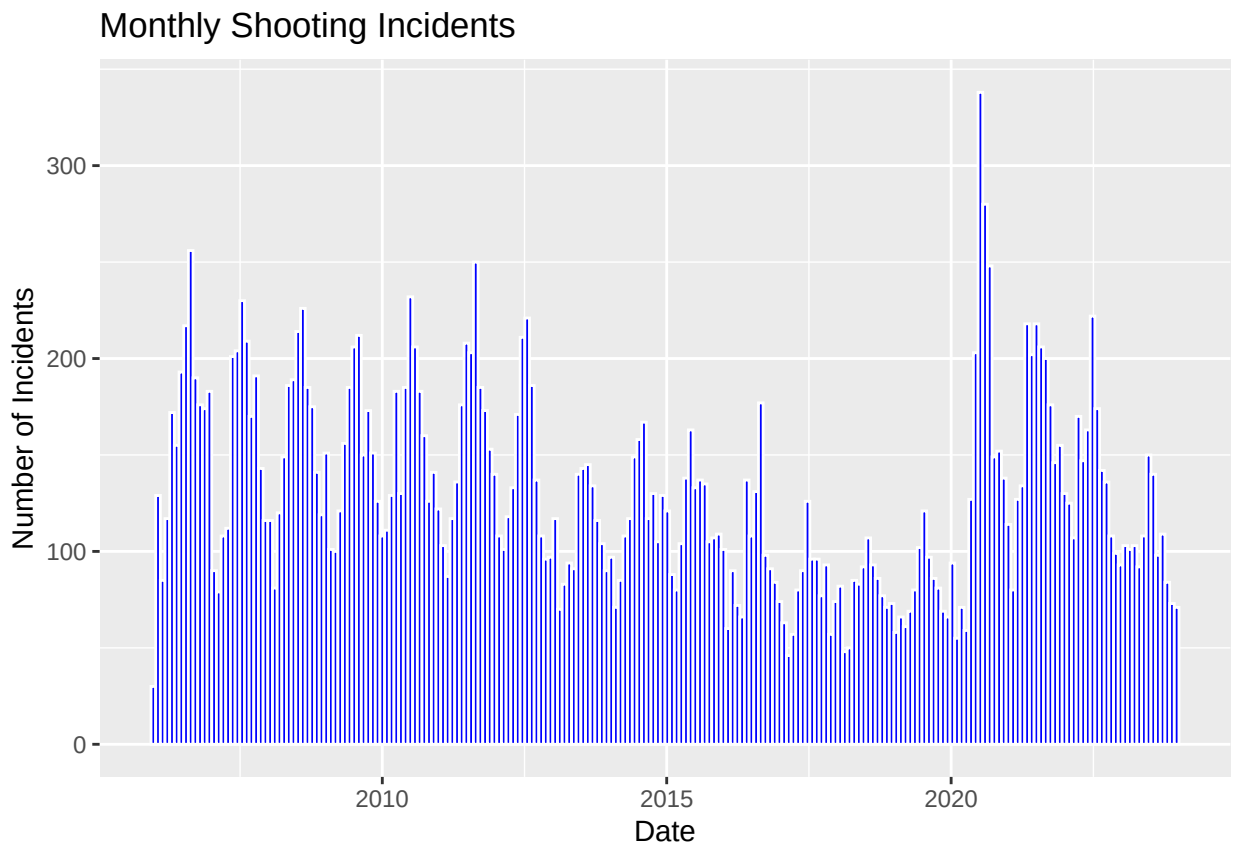
```
## Data Analysis
summary_stats <- shooting_data %>%
  group_by(boro) %>%
  summarise(total_incidents = n(),
            murders = sum(statistical_murder_flag, na.rm = TRUE))

summary_stats
```

```
## # A tibble: 5 x 3
##   boro          total_incidents murders
##   <chr>                   <int>   <int>
## 1 BRONX                    8376    1634
## 2 BROOKLYN                11346    2210
## 3 MANHATTAN                3762     672
## 4 QUEENS                   4271     840
## 5 STATEN ISLAND             807     170
```

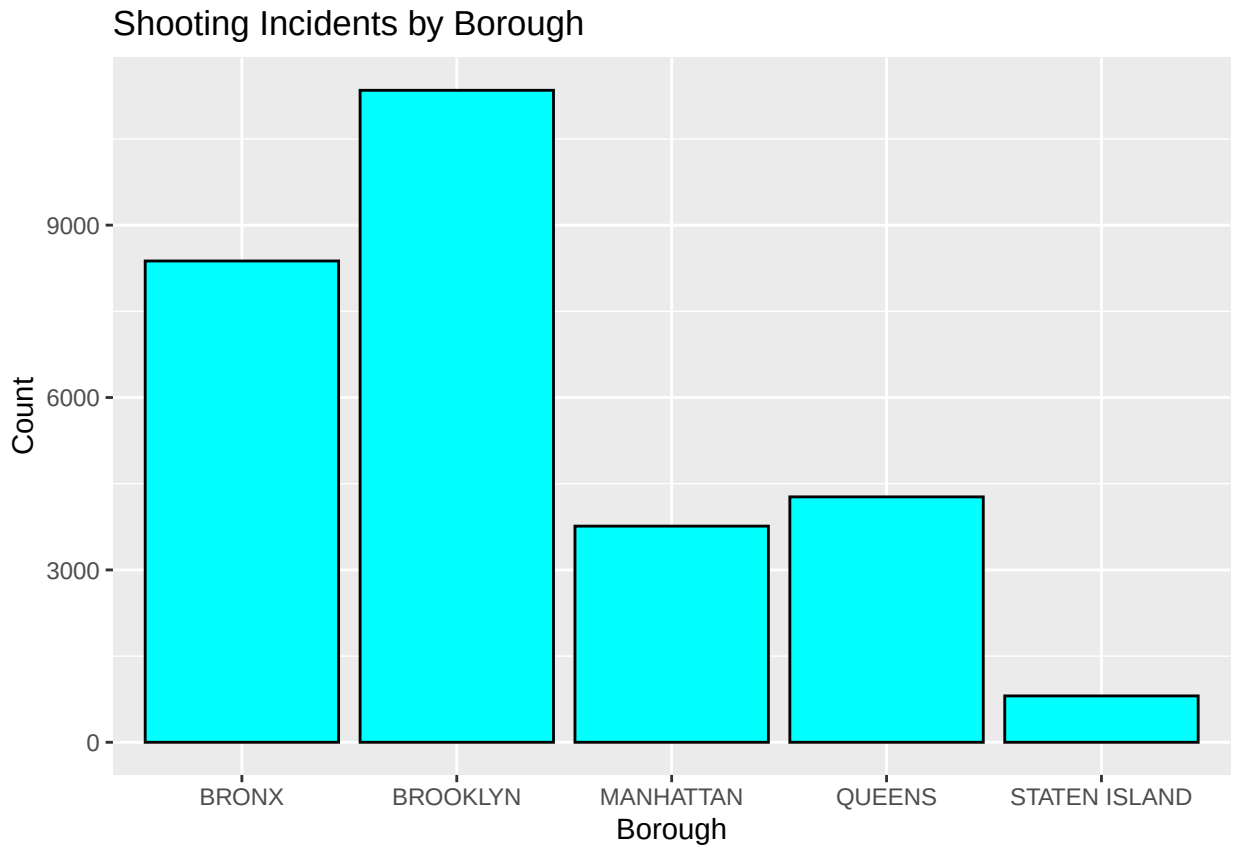
```
shooting_data_summary <- shooting_data %>%
  group_by(date_of_incident) %>%
  summarise(total_incidents = n(), .groups = 'drop')

shooting_data_summary
```

```
## # A tibble: 6,095 x 2
##    date_of_incident total_incidents
##    <date>                    <int>
##  1 2006-01-01                    8
##  2 2006-01-02                    4
##  3 2006-01-03                    4
##  4 2006-01-04                    4
##  5 2006-01-05                    4
##  6 2006-01-06                    4
##  7 2006-01-07                    2
##  8 2006-01-08                    4
##  9 2006-01-09                    9
## 10 2006-01-10                    5
## # i 6,085 more rows
```

```
## Time series of incidents
shooting_data %>%
  ggplot(aes(x = date_of_incident)) +
  geom_histogram(binwidth = 30, fill = "blue", color = "white") +
  labs(title = "Monthly Shooting Incidents", x = "Date", y = "Number of Incidents")
```




```
## Incidents by borough
shooting_data %>%
  ggplot(aes(x = factor(boro))) +
  geom_bar(fill = "cyan", color = "black") +
  labs(title = "Shooting Incidents by Borough", x = "Borough", y = "Count")
```




```
## Linear model predicting incidents based on time
model <- lm(total_incidents ~ month(date_of_incident) + year(date_of_incident), data = shooting_data_su
summary(model)
```

```
##
## Call:
## lm(formula = total_incidents ~ month(date_of_incident) + year(date_of_incident),
##     data = shooting_data_summary)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.909 -2.473 -0.883  1.514 42.751
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              175.869995  17.320042  10.154  < 2e-16 ***
## month(date_of_incident)    0.093292   0.013200   7.068 1.75e-12 ***
## year(date_of_incident)    -0.085284   0.008598  -9.919  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.512 on 6092 degrees of freedom
## Multiple R-squared:  0.02375,    Adjusted R-squared:  0.02343
## F-statistic: 74.11 on 2 and 6092 DF,  p-value: < 2.2e-16
```

```
## Bias Discussion

# - Reporting Bias: Differences in reporting due to location, time, and victim/witness reluctance.
# - Selection Bias: Over / under representation of certain incident types.
# - Geographical Bias: Variance in reporting intensity across different areas.
# - Outcome Bias: More complete data for cases with severe outcomes.
# - Perpetrator and Victim Bias: Categorization biases based on race, age, or gender.
# - Data Entry and Classification Bias: Errors in documentation affecting data accuracy.
```