NYPD Shooting Incident Project

Allison Flors

2023-11-02

Introduction

This is an exploratory analysis of the NYPD Shooting Incident Data from 2006-2022. This data is provided by the city of New York (data.cityofnewyork.us).

Here is a summary of the data taken directly from the source:

"This is a breakdown of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. Each record represents a shooting incident in NYC and includes information about the event, the location and time of occurrence. In addition, information related to suspect and victim demographics is also included."

Visualizations and models will be used to begin exploring trends and questions that come from this data.

Import Data

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

library(readr)  # Load the readr package
nypd_shooting <- read_csv(url_in)
```

```
## 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.
```

Incidents Per Year (Graph & Model)

```
library(tidyverse)
library(lubridate)

nypd_shooting$OCCUR_DATE <- dmy(nypd_shooting$OCCUR_DATE)
```

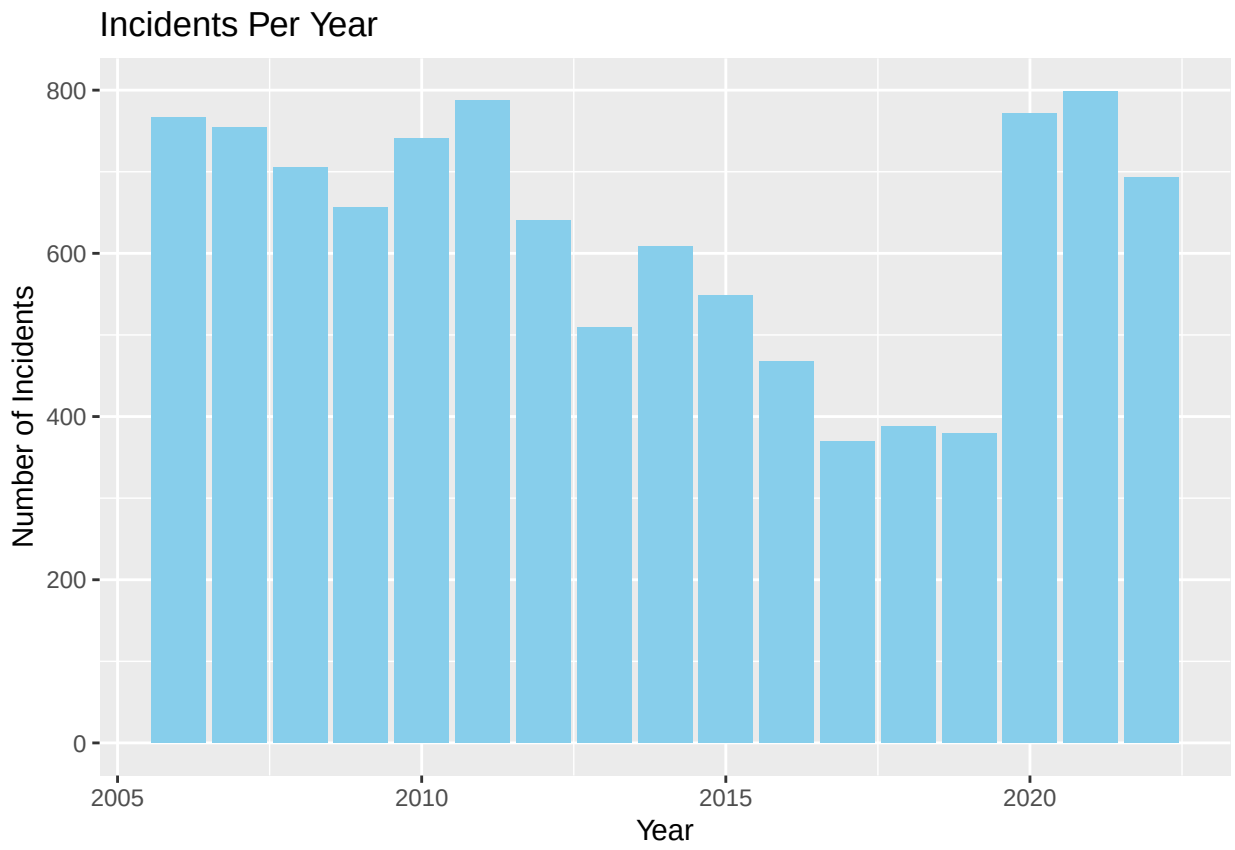
```
## Warning: 16728 failed to parse.
```

```
nypd_shooting <- nypd_shooting[!is.na(nypd_shooting$OCCUR_DATE), ]

incident_counts <- nypd_shooting %>%
  group_by(Year = year(OCCUR_DATE)) %>%
  summarize(IncidentCount = n())

incident_counts <- na.omit(incident_counts)

ggplot(incident_counts, aes(x=Year, y=IncidentCount)) +
  geom_bar(stat="identity", fill = "skyblue") +
  labs(title = "Incidents Per Year", x = "Year", y = "Number of Incidents")
```




```
model <- lm(IncidentCount ~ Year, data = incident_counts)
summary(model)
```

```
##
## Call:
## lm(formula = IncidentCount ~ Year, data = incident_counts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -221.529 -123.941   -3.294   77.000  248.882
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21473.412  14631.579   1.468    0.163
## Year          -10.353      7.265  -1.425    0.175
##
## Residual standard error: 146.7 on 15 degrees of freedom
## Multiple R-squared:  0.1192, Adjusted R-squared:  0.06053
## F-statistic: 2.031 on 1 and 15 DF,  p-value: 0.1746
```

```
new_data <- data.frame(Year = 2023)
predicted_incidents_2023 <- predict(model, newdata = new_data)

predicted_incidents_2023
```

```
##        1
## 529.4118
```

Graph & Model Analysis

The most notable data point that sticks out in the "Incidents Per Year" graph is the large spike in incidents between 2019 and 2020. This is the largest year over year change between two years. 2020 was the beginning of the COVID-19 global pandemic and it was a time when racial justice issues were a hot topic in the United States. Based on this chart, it seems like there is a need to investigate those factors or others that could have led to this dramatic increase in incidents.

A model was run to predict the number of incidents in 2023. It came back with 529.41. This would be a drop from 2022 following the pattern of a dip in incidents following a spike year. However, the p-value for this model was .17 and thus not fully reliable though it does seem likely to see a drop of some sort (though maybe not as drastic) for 2023.

Victim Race Analysis

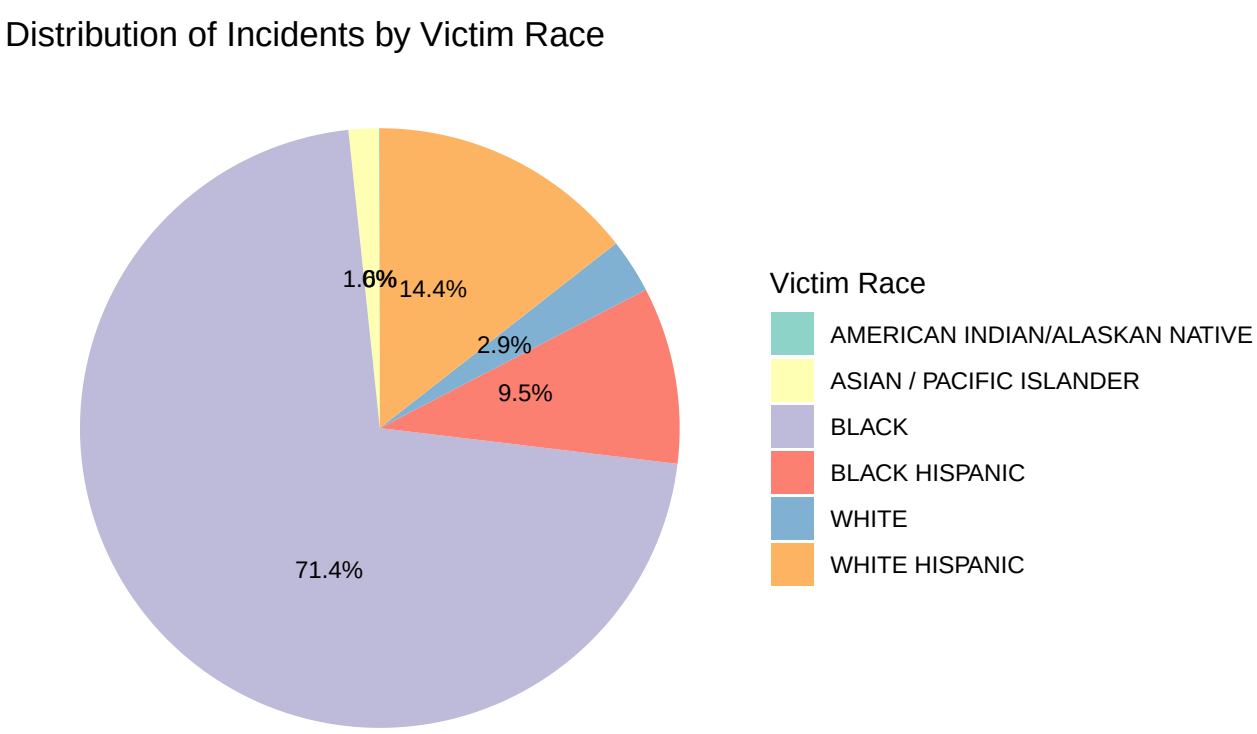
```
library(tidyverse)

nypd_shooting <- nypd_shooting %>%
  filter(!is.na(VIC_RACE) & VIC_RACE != "UNKNOWN")

incident_counts_by_race <- nypd_shooting %>%
  group_by(VIC_RACE) %>%
  summarize(IncidentCount = n())

ggplot(incident_counts_by_race, aes(x = "", y = IncidentCount, fill = VIC_RACE)) +
  geom_bar(stat="identity", width=1) +
  coord_polar(theta = "y") +
  labs (
    title = "Distribution of Incidents by Victim Race",
    fill = "Victim Race",
    x= NULL,
    y=NULL
  ) +
```

```
  theme_minimal() +
  scale_fill_brewer(type = "qual", palette = "Set3") +
  geom_text(aes(label = paste0(round(IncidentCount / sum(IncidentCount)*100,1),"%")),
            position = position_stack(vjust = 0.4),
            size=3) +
  theme_void()
```



Graph Analysis

The "Distribution of Incidents by Victim Race" shows an overwhelming majority of the incident victims have been Black (over 80% include Black Hispanic victims). This is concerning because the population of New York has only fluctuated around 25-20% in the years this data spans (according to the US census online quick facts).

Potential Bias

An important note here - there were some records were the victims race was "UNKNOWN" or missing in the original data set. These records were excluded from the analysis. Also, it is not clear if victims are self reporting their race in some manner or if someone completing the report is choosing a race for the victim. This could lead to some biased or incorrect records.

Session Inforamtion

```
## - Session info ---------------------------------------------------------------
##  setting  value
##  version  R version 4.3.1 (2023-06-16 ucrt)
##  os       Windows 11 x64 (build 22621)
##  system   x86_64, mingw32
##  ui       RTerm
##  language (EN)
##  collate  English_United States.utf8
##  ctype    English_United States.utf8
##  tz       America/New_York
##  date     2023-11-02
##  pandoc   3.1.1 @ C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/ (via rmarkdown)
##
## - Packages -------------------------------------------------------------------
##  package      * version date (UTC) lib source
##  bit            4.0.5   2022-11-15 [1] CRAN (R 4.3.1)
##  bit64          4.0.5   2020-08-30 [1] CRAN (R 4.3.1)
##  cli            3.6.1   2023-03-23 [1] CRAN (R 4.3.1)
##  colorspace     2.1-0   2023-01-23 [1] CRAN (R 4.3.1)
##  crayon         1.5.2   2022-09-29 [1] CRAN (R 4.3.1)
##  curl           5.1.0   2023-10-02 [1] CRAN (R 4.3.1)
##  digest         0.6.33  2023-07-07 [1] CRAN (R 4.3.1)
##  dplyr        * 1.1.3   2023-09-03 [1] CRAN (R 4.3.1)
##  evaluate       0.22    2023-09-29 [1] CRAN (R 4.3.1)
##  fansi          1.0.5   2023-10-08 [1] CRAN (R 4.3.1)
##  farver         2.1.1   2022-07-06 [1] CRAN (R 4.3.1)
##  fastmap        1.1.1   2023-02-24 [1] CRAN (R 4.3.1)
##  forcats      * 1.0.0   2023-01-29 [1] CRAN (R 4.3.1)
##  generics       0.1.3   2022-07-05 [1] CRAN (R 4.3.1)
##  ggplot2      * 3.4.4   2023-10-12 [1] CRAN (R 4.3.1)
##  glue           1.6.2   2022-02-24 [1] CRAN (R 4.3.1)
##  gtable         0.3.4   2023-08-21 [1] CRAN (R 4.3.1)
##  hms            1.1.3   2023-03-21 [1] CRAN (R 4.3.1)
##  htmltools      0.5.6.1 2023-10-06 [1] CRAN (R 4.3.1)
##  knitr          1.44    2023-09-11 [1] CRAN (R 4.3.1)
##  labeling       0.4.3   2023-08-29 [1] CRAN (R 4.3.1)
##  lifecycle      1.0.3   2022-10-07 [1] CRAN (R 4.3.1)
##  lubridate    * 1.9.3   2023-09-27 [1] CRAN (R 4.3.1)
##  magrittr       2.0.3   2022-03-30 [1] CRAN (R 4.3.1)
##  munsell        0.5.0   2018-06-12 [1] CRAN (R 4.3.1)
##  pillar         1.9.0   2023-03-22 [1] CRAN (R 4.3.1)
##  pkgconfig      2.0.3   2019-09-22 [1] CRAN (R 4.3.1)
##  purrr        * 1.0.2   2023-08-10 [1] CRAN (R 4.3.1)
##  R6             2.5.1   2021-08-19 [1] CRAN (R 4.3.1)
##  RColorBrewer   1.1-3   2022-04-03 [1] CRAN (R 4.3.1)
##  readr        * 2.1.4   2023-02-10 [1] CRAN (R 4.3.1)
##  rlang          1.1.1   2023-04-28 [1] CRAN (R 4.3.1)
##  rmarkdown      2.25    2023-09-18 [1] CRAN (R 4.3.1)
##  rstudioapi     0.15.0  2023-07-07 [1] CRAN (R 4.3.1)
##  scales         1.2.1   2022-08-20 [1] CRAN (R 4.3.1)
##  sessioninfo    1.2.2   2021-12-06 [1] CRAN (R 4.3.1)
##  stringi        1.7.12  2023-01-11 [1] CRAN (R 4.3.1)
```

```
##  stringr      * 1.5.0   2022-12-02 [1] CRAN (R 4.3.1)
##  tibble       * 3.2.1   2023-03-20 [1] CRAN (R 4.3.1)
##  tidyr        * 1.3.0   2023-01-24 [1] CRAN (R 4.3.1)
##  tidyselect     1.2.0   2022-10-10 [1] CRAN (R 4.3.1)
##  tidyverse    * 2.0.0   2023-02-22 [1] CRAN (R 4.3.1)
##  timechange     0.2.0   2023-01-11 [1] CRAN (R 4.3.1)
##  tzdb           0.4.0   2023-05-12 [1] CRAN (R 4.3.1)
##  utf8           1.2.4   2023-10-22 [1] CRAN (R 4.3.1)
##  vctrs          0.6.4   2023-10-12 [1] CRAN (R 4.3.1)
##  vroom          1.6.4   2023-10-02 [1] CRAN (R 4.3.1)
##  withr          2.5.1   2023-09-26 [1] CRAN (R 4.3.1)
##  xfun           0.40    2023-08-09 [1] CRAN (R 4.3.1)
##  yaml           2.3.7   2023-01-23 [1] CRAN (R 4.3.1)
##
##  [1] C:/Users/allis/AppData/Local/R/win-library/4.3
##  [2] C:/Program Files/R/R-4.3.1/library
##
## ------------------------------------------------------------------------------
```