NYPD Shooting Analysis

Anna Behlke

2026-02-08

Introduction

This project analyzes the NYPD Shooting Incident dataset to examine trends over time, differences across boroughs, victim demographics, and factors associated with fatal outcomes.

Data Preparation

Packages used: tidyverse, lubridate, hms, scales, broom Cleaned data: mutate occur_date, occur time, year, create binary for fatal, filter by is not !is.na for occur_date, BORO, VIC_AGE_GROUP, VIC_SEX

```
shootings <- read_csv("NYPD_Shooting_Incident_Data__Historic_.csv")
```

```
## Rows: 29744 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## 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.
```

```
shootings_clean <- shootings %>%
  mutate(
    occur_date = mdy(OCCUR_DATE),
    occur_time = as_hms(OCCUR_TIME),
    year = year(occur_date),
    fatal = if_else(STATISTICAL_MURDER_FLAG == TRUE, 1, 0)
  ) %>%
  filter(
    !is.na(occur_date),
    !is.na(BORO),
    !is.na(VIC_AGE_GROUP),
    !is.na(VIC_SEX)
  )

summary(shootings_clean)
```

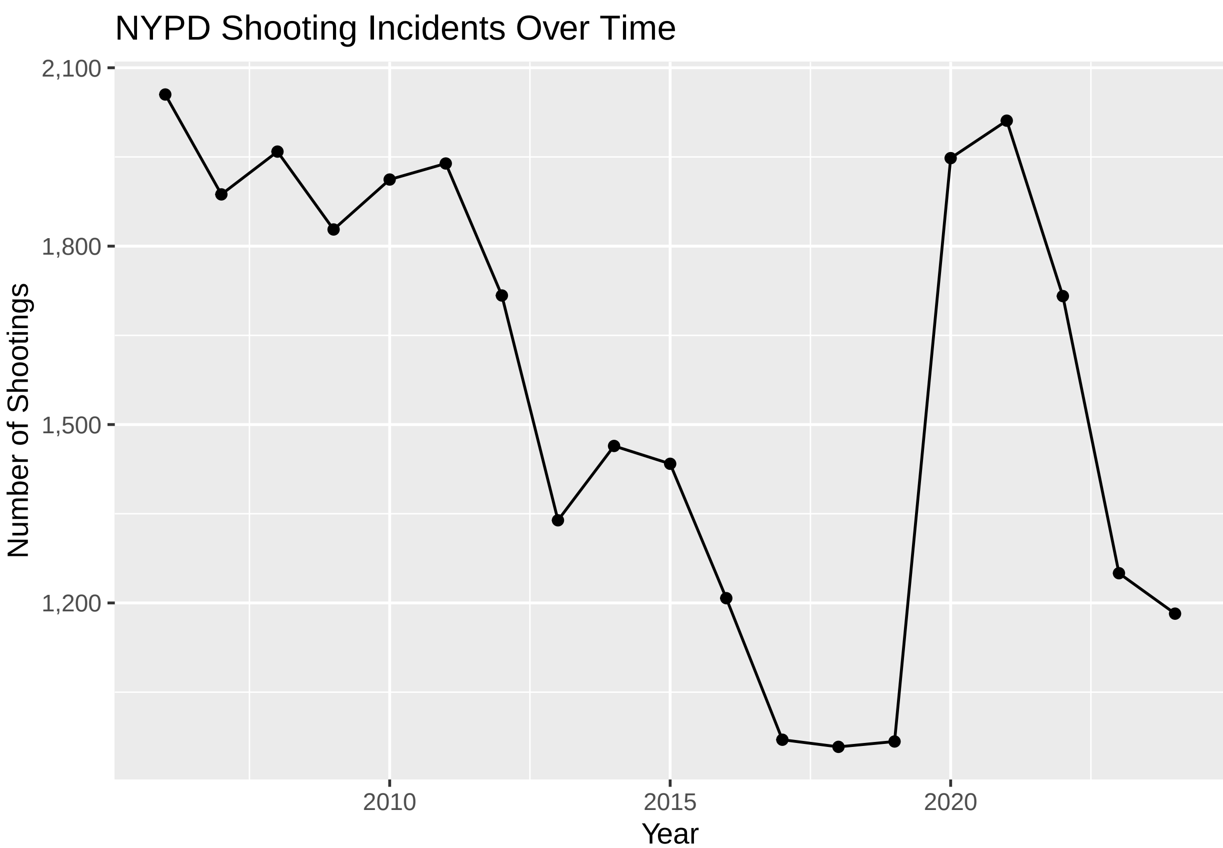
```
##   INCIDENT_KEY        OCCUR_DATE          OCCUR_TIME
##  Min.   :  9953245   Length:29744       Min.   :00:00:00.000000
##  1st Qu.: 67321140   Class :character   1st Qu.:03:30:45.000000
##  Median :109291972   Mode  :character   Median :15:15:00.000000
```

```
## Mean   :133850951                     Mean   :12:46:10.874798
## 3rd Qu.:214741917                     3rd Qu.:20:44:00.000000
## Max.   :299462478                     Max.   :23:59:00.000000
##
##     BORO           LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE
## Length:29744       Length:29744       Min.   :  1.00   Min.   :0.0000
## Class :character   Class :character   1st Qu.: 44.00   1st Qu.:0.0000
## Mode  :character   Mode  :character   Median : 67.00   Median :0.0000
##                                       Mean   : 65.23   Mean   :0.3181
##                                       3rd Qu.: 81.00   3rd Qu.:0.0000
##                                       Max.   :123.00   Max.   :2.0000
##                                                        NA's   :2
## LOC_CLASSFCTN_DESC LOCATION_DESC      STATISTICAL_MURDER_FLAG
## Length:29744       Length:29744       Mode :logical
## Class :character   Class :character   FALSE:23979
## Mode  :character   Mode  :character   TRUE :5765
##
##
##
##
## PERP_AGE_GROUP       PERP_SEX          PERP_RACE         VIC_AGE_GROUP
## Length:29744       Length:29744       Length:29744       Length:29744
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   VIC_SEX            VIC_RACE           X_COORD_CD        Y_COORD_CD
## Length:29744       Length:29744       Min.   : 914928   Min.   :125757
## Class :character   Class :character   1st Qu.:1000094   1st Qu.:183042
## 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           occur_date
## Min.   :40.51   Min.   :-74.25   Length:29744       Min.   :2006-01-01
## 1st Qu.:40.67   1st Qu.:-73.94   Class :character   1st Qu.:2009-10-29
## Median :40.70   Median :-73.91   Mode  :character   Median :2014-03-25
## Mean   :40.74   Mean   :-73.91                      Mean   :2014-10-31
## 3rd Qu.:40.83   3rd Qu.:-73.88                      3rd Qu.:2020-06-29
## Max.   :40.91   Max.   :-73.70                      Max.   :2024-12-31
## NA's   :97      NA's   :97
##   occur_time                   year          fatal
## Min.   :00:00:00.000000   Min.   :2006   Min.   :0.0000
## 1st Qu.:03:30:45.000000   1st Qu.:2009   1st Qu.:0.0000
## Median :15:15:00.000000   Median :2014   Median :0.0000
## Mean   :12:46:10.874798   Mean   :2014   Mean   :0.1938
## 3rd Qu.:20:44:00.000000   3rd Qu.:2020   3rd Qu.:0.0000
## Max.   :23:59:00.000000   Max.   :2024   Max.   :1.0000
##
```

Exploratory Data Analysis

Shootings Over Time

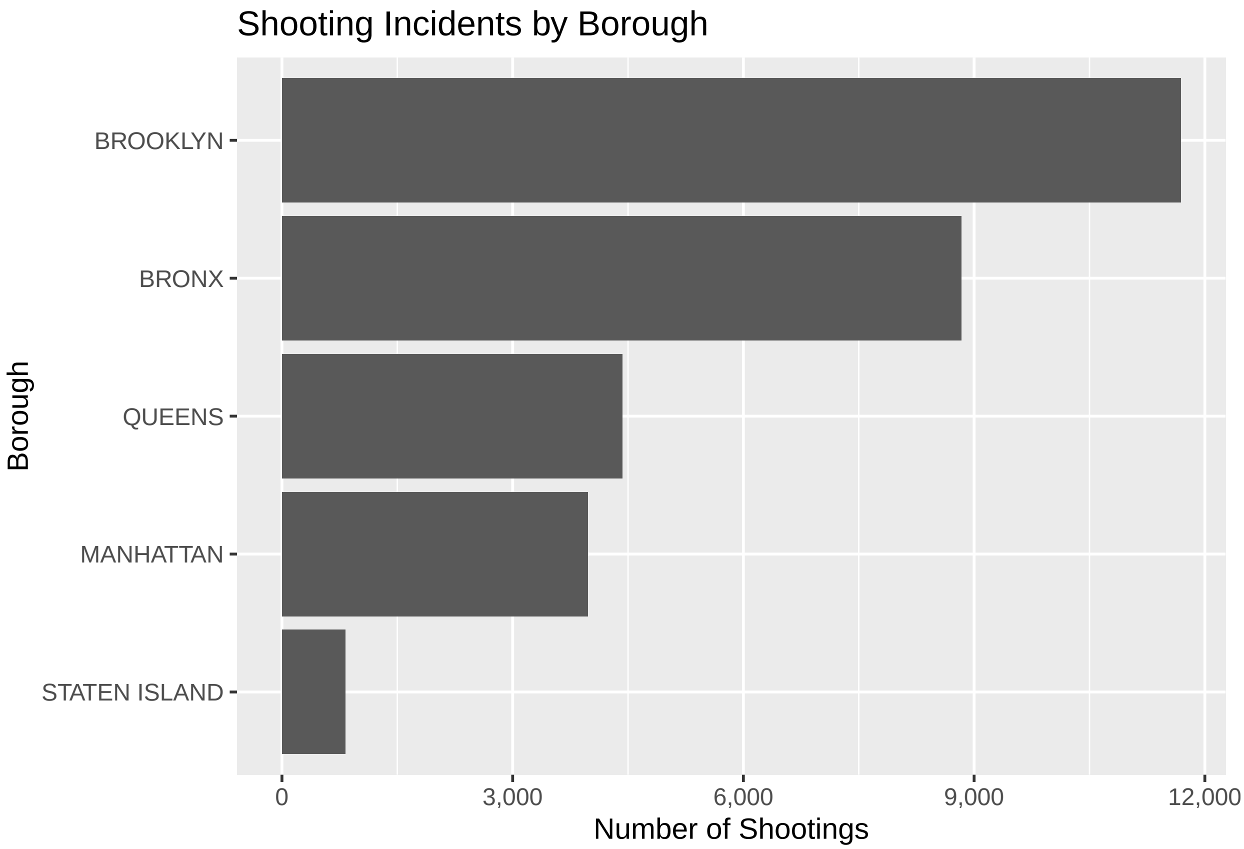
```
shootings_clean %>%
  count(year) %>%
  ggplot(aes(year, n)) +
  geom_line() +
  geom_point() +
  scale_y_continuous(labels = comma) +
  labs(
    title = "NYPD Shooting Incidents Over Time",
    x = "Year",
    y = "Number of Shootings"
  )
```



Shootings by Borough

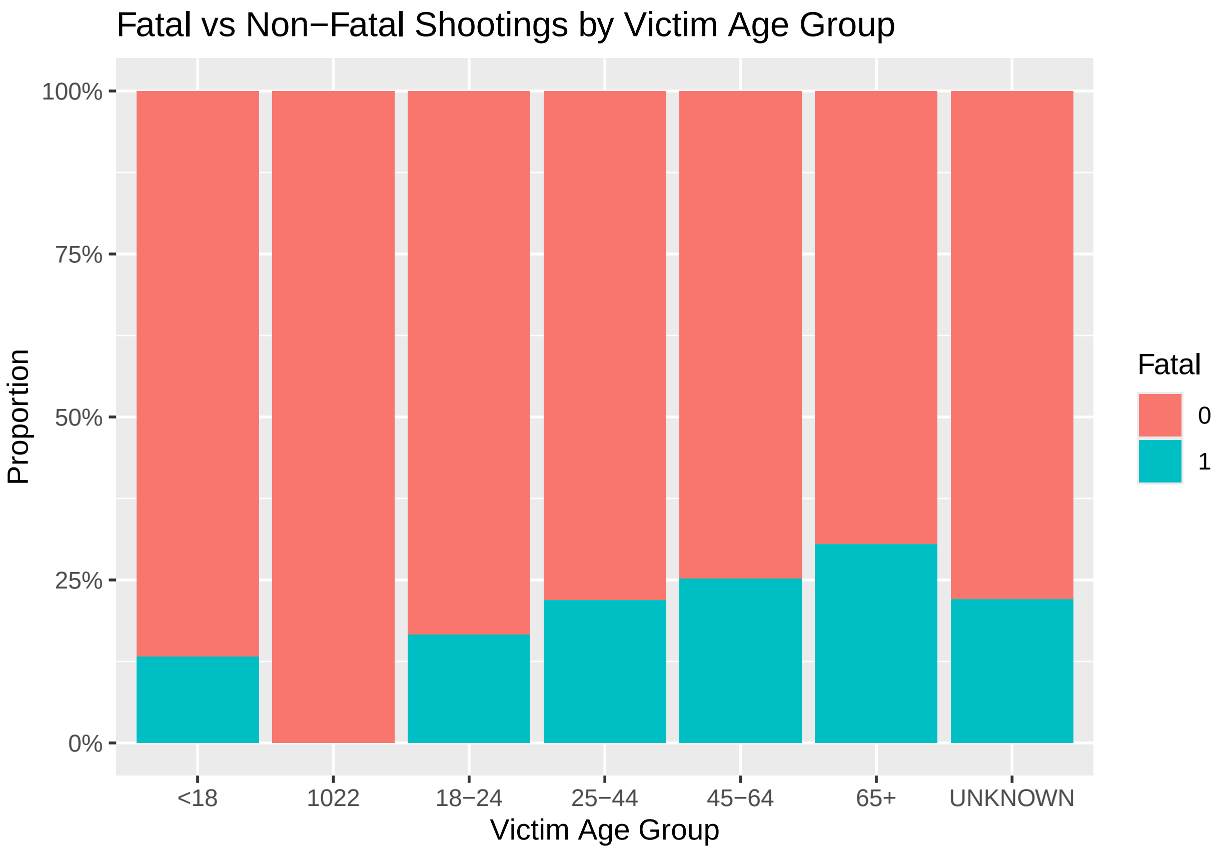
```
shootings_clean %>%
  count(BORO) %>%
  ggplot(aes(reorder(BORO, n), n)) +
  geom_col() +
  coord_flip() +
  scale_y_continuous(labels = comma) +
  labs(
    title = "Shooting Incidents by Borough",
    x = "Borough",
```

```
    y = "Number of Shootings"
  )
```



Fatal vs Non-Fatal by Victim Age Group

```
shootings_clean %>%
  ggplot(aes(VIC_AGE_GROUP, fill = factor(fatal))) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = percent) +
  labs(
    title = "Fatal vs Non-Fatal Shootings by Victim Age Group",
    x = "Victim Age Group",
    y = "Proportion",
    fill = "Fatal"
  )
```



Modeling

Create logistic regression model using the cleaned data to determine if fatality is realated to victim age, sex, and borough.

```
model_data <- shootings_clean %>%
  select(fatal, VIC_AGE_GROUP, VIC_SEX, BORO) %>%
  drop_na()

fatal_model <- glm(
  fatal ~ VIC_AGE_GROUP + VIC_SEX + BORO,
  data = model_data,
  family = binomial
)

summary(fatal_model)
```

```
##
## Call:
## glm(formula = fatal ~ VIC_AGE_GROUP + VIC_SEX + BORO, family = binomial,
##     data = model_data)
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -1.82137    0.07139 -25.511  < 2e-16 ***
## VIC_AGE_GROUP1022      -7.58894   72.46288  -0.105   0.9166
## VIC_AGE_GROUP18-24      0.26494    0.05930   4.468 7.90e-06 ***
```

```
## VIC_AGE_GROUP25-44    0.61013    0.05719  10.668  < 2e-16 ***
## VIC_AGE_GROUP45-64    0.79153    0.07306  10.834  < 2e-16 ***
## VIC_AGE_GROUP65+      1.04987    0.15127   6.941 3.91e-12 ***
## VIC_AGE_GROUPUNKNOWN  0.68748    0.30233   2.274   0.0230 *
## VIC_SEXM             -0.04024    0.04953  -0.812   0.4165
## VIC_SEXU             -1.08936    1.06043  -1.027   0.3043
## BOROBROOKLYN         -0.02016    0.03574  -0.564   0.5727
## BOROMANHATTAN        -0.11540    0.04939  -2.336   0.0195 *
## BOROQUEENS           -0.01714    0.04658  -0.368   0.7130
## BOROSTATEN ISLAND     0.05413    0.09065   0.597   0.5504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 29251  on 29743  degrees of freedom
## Residual deviance: 28992  on 29731  degrees of freedom
## AIC: 29018
##
## Number of Fisher Scoring iterations: 8
```

```
tidy(fatal_model, exponentiate = TRUE, conf.int = TRUE)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## # A tibble: 13 x 7
##    term                 estimate std.error statistic   p.value conf.low conf.high
##    <chr>                   <dbl>     <dbl>     <dbl>     <dbl>    <dbl>     <dbl>
##  1 (Intercept)          0.162       0.0714   -25.5   1.48e-143   0.141      0.186
##  2 VIC_AGE_GROUP1022    0.000506   72.5       -0.105 9.17e-  1  NA       5063.
##  3 VIC_AGE_GROUP18-24   1.30        0.0593     4.47  7.90e-  6   1.16       1.47
##  4 VIC_AGE_GROUP25-44   1.84        0.0572    10.7   1.43e- 26   1.65       2.06
##  5 VIC_AGE_GROUP45-64   2.21        0.0731    10.8   2.38e- 27   1.91       2.55
##  6 VIC_AGE_GROUP65+     2.86        0.151      6.94  3.91e- 12   2.11       3.83
##  7 VIC_AGE_GROUPUNKNO~  1.99        0.302      2.27  2.30e-  2   1.06       3.51
##  8 VIC_SEXM             0.961       0.0495    -0.812 4.17e-  1   0.872      1.06
##  9 VIC_SEXU             0.336       1.06      -1.03  3.04e-  1   0.0181     1.81
## 10 BOROBROOKLYN         0.980       0.0357    -0.564 5.73e-  1   0.914      1.05
## 11 BOROMANHATTAN        0.891       0.0494    -2.34  1.95e-  2   0.808      0.981
## 12 BOROQUEENS           0.983       0.0466    -0.368 7.13e-  1   0.897      1.08
## 13 BOROSTATEN ISLAND    1.06        0.0906     0.597 5.50e-  1   0.881      1.26
```

Bias and Limitations

This dataset includes only NYPD-reported incidents and may reflect reporting bias, missing demographic data, and changes in policing practices over time. The model is exploratory and does not imply causation.

Conclusion

The analysis shows variation in shooting incidents over time and across boroughs, with differences in fatal outcomes across victim demographics. Further analysis could include spatial or policy-related factors.