

NYPD Shooting Incident (Historical)

Matthew Young

2022-04-29

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr    0.3.4
## v tibble   3.1.6     v dplyr    1.0.8
## v tidyr    1.2.0     v stringr  1.4.0
## v readr    2.1.1     vforcats  0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

library(tibble)
library(ggplot2)
library(dplyr)
library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
## 
##     date, intersect, setdiff, union

nypd_shootings <- read_csv("NYPD_Shooting_Incident_Data__Historic_.csv")

## # Rows: 23585 Columns: 19
## -- Column specification -----
## Delimiter: ","
## chr (10): OCCUR_DATE, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_SEX, PERP_R...
## 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.

tibble_nypd <- as_tibble(nypd_shootings)
nypd <- tibble_nypd
nypd <- select(nypd, -INCIDENT_KEY)
nypd <- select(nypd, -Lon_Lat)
nypd$PERP_AGE_GROUP[nypd$PERP_AGE_GROUP == ""] <- "UNKNOWN"
nypd$PERP_AGE_GROUP[nypd$PERP_AGE_GROUP == 224] <- "UNKNOWN"
nypd$PERP_AGE_GROUP[nypd$PERP_AGE_GROUP == 940] <- "UNKNOWN"
nypd$PERP_AGE_GROUP[nypd$PERP_AGE_GROUP == 1020] <- "UNKNOWN"

nypd$LOCATION_DESC[nypd$LOCATION_DESC == "UNKNOWN"] <- "NA"

nypd$STATISTICAL_MURDER_FLAG[nypd$STATISTICAL_MURDER_FLAG == "false" ] <- 0
nypd$STATISTICAL_MURDER_FLAG[nypd$STATISTICAL_MURDER_FLAG == "true" ] <- 1

nypd$PERP_SEX[nypd$PERP_SEX == "U"] <- "UNDISCLOSED"

nypd$PERP_RACE[nypd$PERP_RACE == ""] <- "UNKNOWN"

nypd$OCCUR_DATE <- mdy(nypd$OCCUR_DATE)
nypd$OCCUR_TIME <- hms(nypd$OCCUR_TIME)

nypd <- arrange(nypd, nypd$OCCUR_DATE, nypd$OCCUR_TIME)

```

Plot Summaries

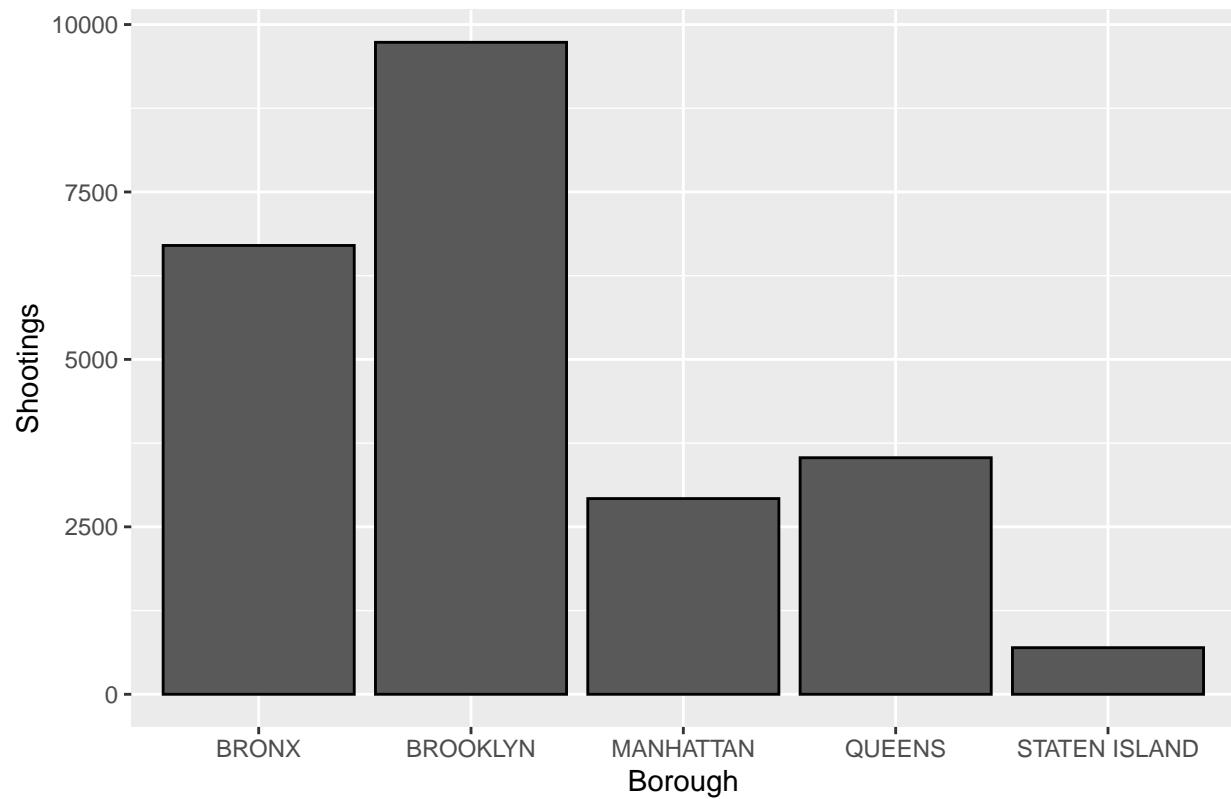
Shooting Incidents and Shootings Resulting in Death by Borough

```

ggplot(nypd, aes(x = `BORO`, fill=`STATISTICAL_MURDER_FLAG`)) +
  geom_bar(color="black") +
  labs(x = "Borough", y = "Shootings", title="Shooting Incidents and Shootings Resulting in Death (2006-2016)", subtitle="by Borough") +
  scale_fill_discrete(name = "Shooting Type", labels = c("Non-Fatal Shootings", "Fatal Shootings"))

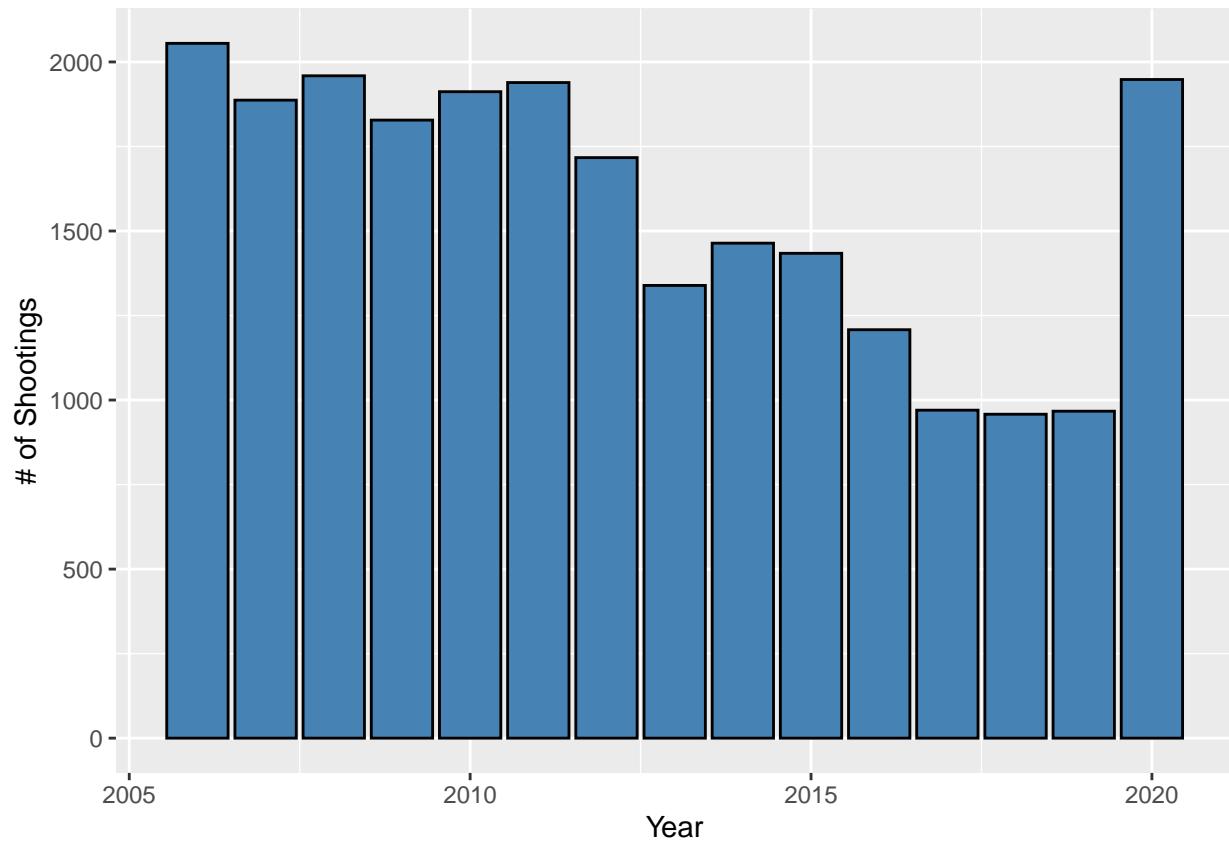
```

Shooting Incidents and Shootings Resulting in Death (2006–2020)



Total Shootings by Year

```
nypd %>%
  group_by(year(`OCCUR_DATE`), `BORO`) %>%
  ggplot(aes(year(`OCCUR_DATE`))) +
  geom_bar(color="black", fill="steelblue") +
  labs(x = "Year", y = "# of Shootings")
```

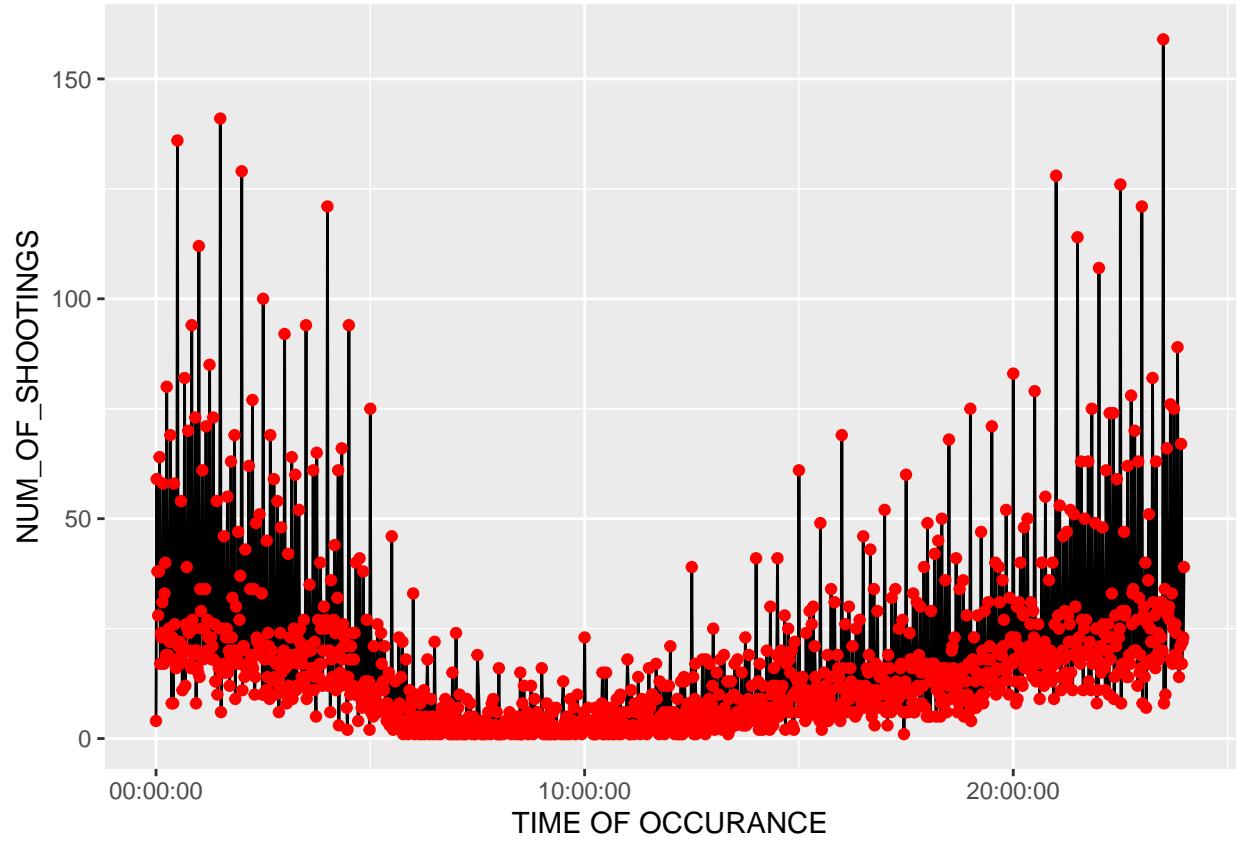


```
scale_x_continuous(breaks = 2006:2020)
```

```
## <ScaleContinuousPosition>
##   Range:
##   Limits:    0 --     1
```

Total Shootings by Hour

```
nypd %>%
  arrange(`OCCUR_TIME`) %>%
  group_by(`OCCUR_TIME`) %>%
  summarise(NUM_OF_SHOOTINGS = n(), groups = 'drop') %>%
  ggplot(aes(`OCCUR_TIME` , `NUM_OF_SHOOTINGS`)) +
  geom_line(color = "black") +
  geom_point(color = "red") +
  scale_x_time() +
  labs(x = "TIME OF OCCURANCE")
```



Conclusion

Based on the data, Brooklyn has the most shooting incidents with Staten Island having the least. While shooting incidents have been falling since 2006, they have returned to 2006 levels starting with 2020. Based on the number of shootings that happen in total within a hour the majority of shootings happen between 20:00 and 02:00