NYC_Data

Margaret Battersby

2024-09-06

```
options(repos = c(CRAN = "https://cloud.r-project.org/"))

# read in packages

install.packages("readr")
```

```
## Installing package into 'C:/Users/Maggi/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)

## package 'readr' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'readr'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Maggi\AppData\Local\R\win-library\4.4\00LOCK\readr\libs\x64\readr.dll
## to C:\Users\Maggi\AppData\Local\R\win-library\4.4\readr\libs\x64\readr.dll:
## Permission denied

## Warning: restored 'readr'

##
## The downloaded binary packages are in
##  C:\Users\Maggi\AppData\Local\Temp\RtmpA3X9TB\downloaded_packages
```

```
install.packages("dplyr")
```

```
## Installing package into 'C:/Users/Maggi/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)

## package 'dplyr' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'dplyr'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Maggi\AppData\Local\R\win-library\4.4\00LOCK\dplyr\libs\x64\dplyr.dll
## to C:\Users\Maggi\AppData\Local\R\win-library\4.4\dplyr\libs\x64\dplyr.dll:
## Permission denied
```

```
## Warning: restored 'dplyr'

##
## The downloaded binary packages are in
##  C:\Users\Maggi\AppData\Local\Temp\RtmpA3X9TB\downloaded_packages
```

```
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/Maggi/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)

## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Maggi\AppData\Local\Temp\RtmpA3X9TB\downloaded_packages
```

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

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

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

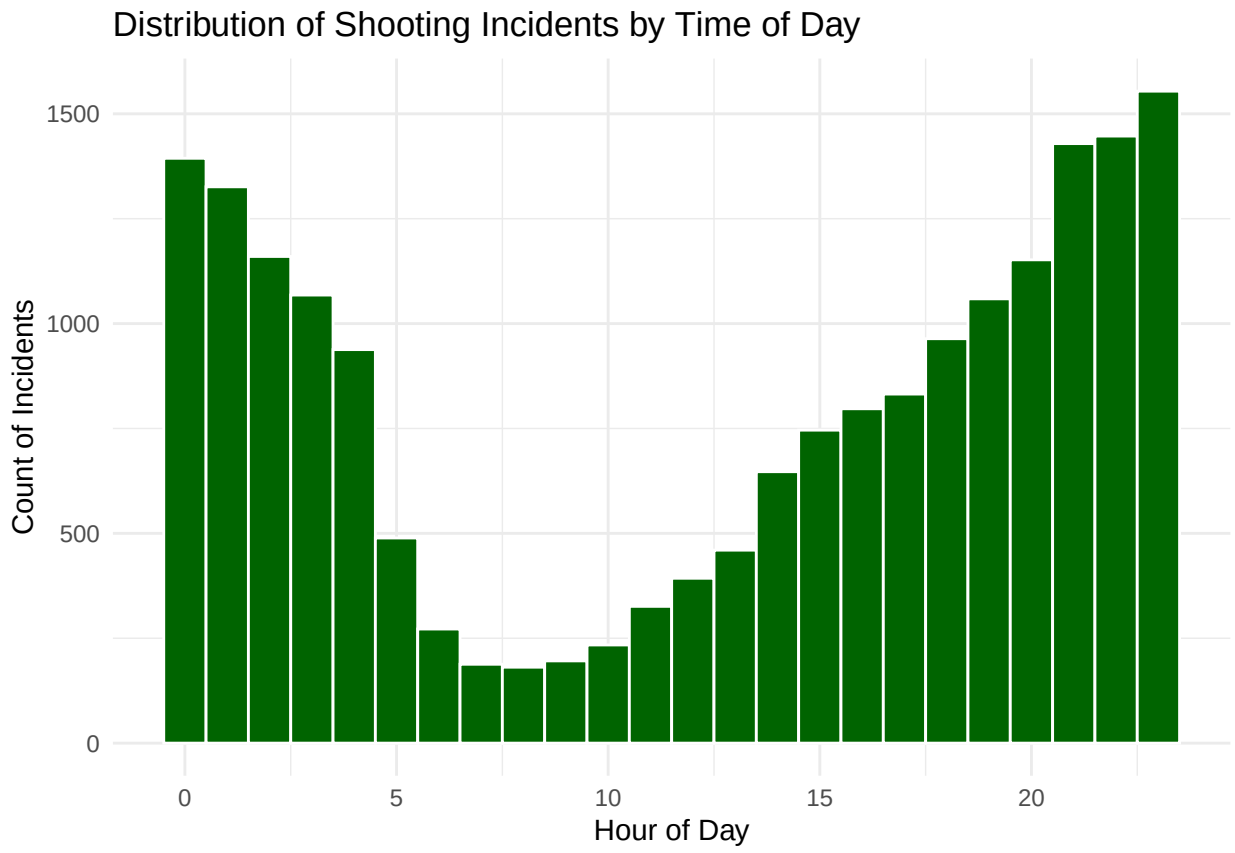
```
library(lubridate)

# read in data

## Get current Data in the file
urls <- c("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD")

nypd_data <- read_csv(urls[1])
```

```
## 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.
```

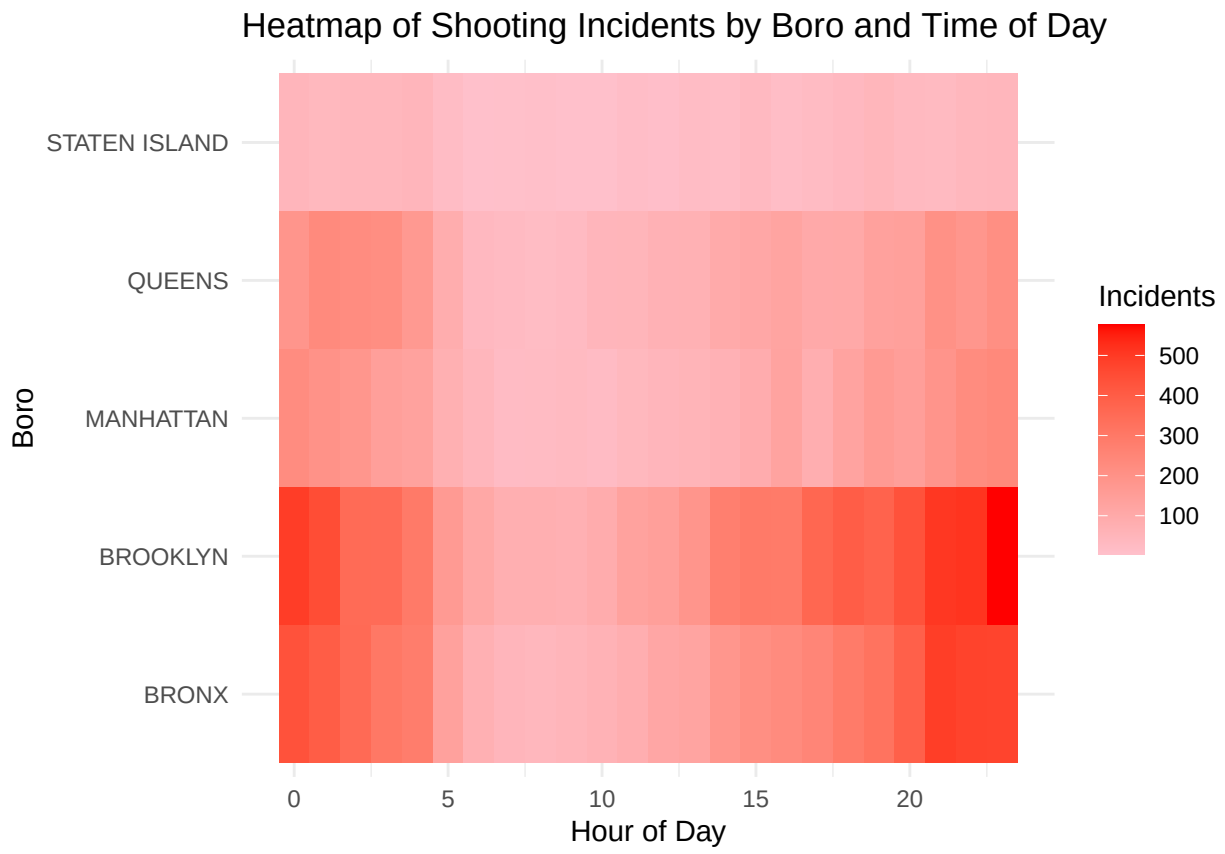
```
# clean up data

nypd_data_clean <- nypd_data %>%
  drop_na(OCCUR_DATE,OCCUR_TIME,BORO)

nypd_data_clean <- nypd_data %>%
  mutate(LOC_OF_OCCUR_DESC = replace_na(LOC_OF_OCCUR_DESC, "UNKNOWN"))

nypd_data_clean <- nypd_data %>%
    mutate(OCCUR_DATE = mdy(OCCUR_DATE),
           OCCUR_TIME = hms(OCCUR_TIME))

nypd_data_clean <- nypd_data %>%
    drop_na(VIC_SEX, PERP_SEX)

nypd_data_clean <- nypd_data_clean %>%
    mutate(OCCUR_DATE = mdy(OCCUR_DATE),
           OCCUR_TIME = hms(OCCUR_TIME))

nypd_data_clean <- nypd_data_clean %>%
    mutate(YEAR = year(OCCUR_DATE),
           MONTH = month(OCCUR_DATE, label = TRUE),
           HOUR = hour(OCCUR_TIME))

nypd_data_clean <- nypd_data_clean %>%
    mutate(YEAR = year(OCCUR_DATE),
           MONTH = month(OCCUR_DATE, label = TRUE),
           HOUR = hour(OCCUR_TIME))

nypd_data_clean <- nypd_data_clean %>%
    mutate(BORO = as.factor(BORO),
           PERP_SEX = as.factor(PERP_SEX),
           VIC_SEX = as.factor(VIC_SEX))

nypd_boro_hour <- nypd_data_clean %>%
  group_by(BORO,HOUR) %>%
  summarise(incidents = n())
```

```
## `summarise()` has grouped output by 'BORO'. You can override using the
## `.groups` argument.
```

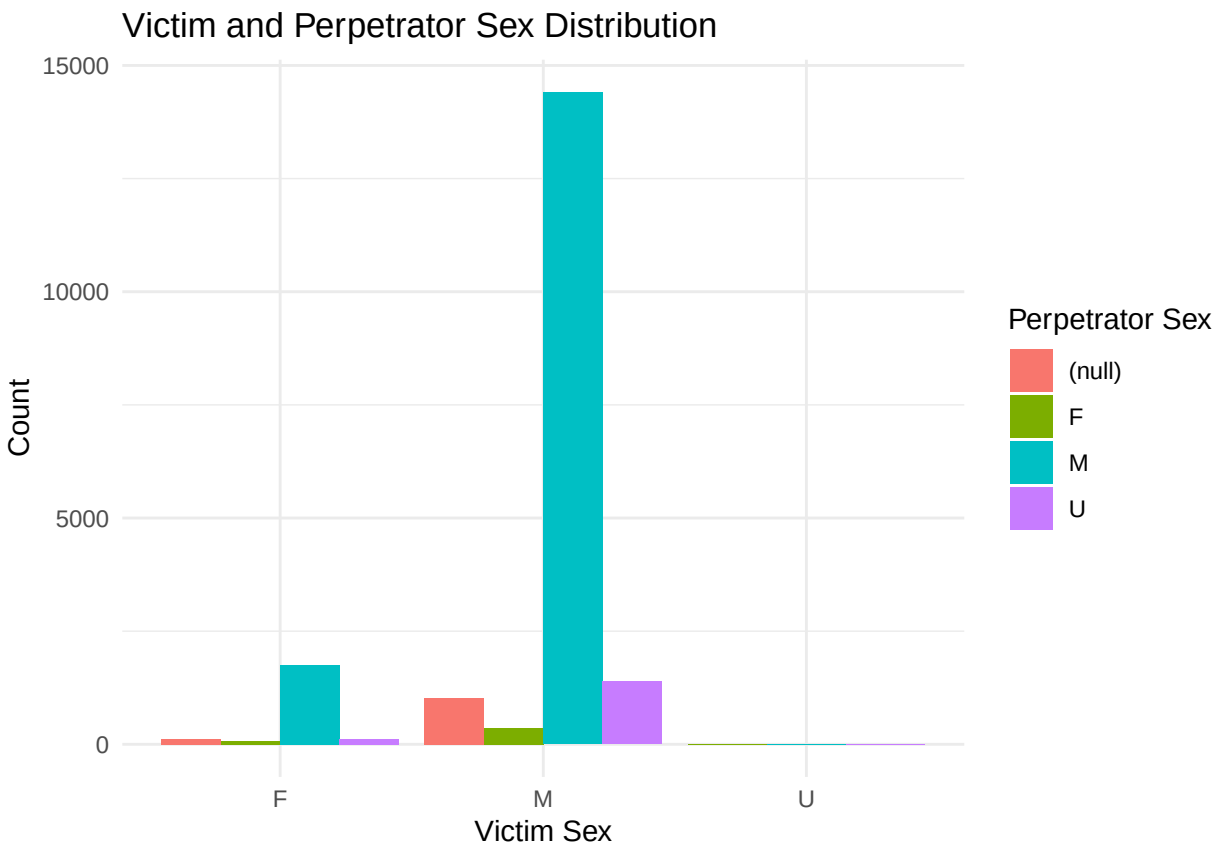
```r
# visualizations

ggplot(nypd_data_clean, aes(x = HOUR)) +
    geom_histogram(binwidth = 1, fill = "darkgreen", color = "white") +
    labs(title = "Distribution of Shooting Incidents by Time of Day",
         x = "Hour of Day", y = "Count of Incidents") +
    theme_minimal()
```




```r
ggplot(nypd_boro_hour, aes(x = HOUR, y = BORO, fill = incidents)) +
    geom_tile() +
    scale_fill_gradient(low = "pink", high = "red") +
    labs(title = "Heatmap of Shooting Incidents by Boro and Time of Day",
         x = "Hour of Day", y = "Boro", fill = "Incidents") +
    theme_minimal()
```




```
ggplot(nypd_data_clean, aes(x = VIC_SEX, fill = PERP_SEX)) +
    geom_bar(position = "dodge") +
    labs(title = "Victim and Perpetrator Sex Distribution",
         x = "Victim Sex", y = "Count", fill = "Perpetrator Sex") +
    theme_minimal()
```


Victim and Perpetrator Sex Distribution
15000
10000
5000
0
Count
F
M
U
Victim Sex
Perpetrator Sex
(null)
F
M
U