NYPD Shooting Data project for R

LWarne

September 10, 2021

Library in Tidyverse and Lubridate

library(lubridate) library(tidyverse)

Data selection

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

Import Data

```
nypd <- read.csv("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD")
summary(nypd)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME            BORO
##  Min.   :  9953245   Length:23568       Length:23568       Length:23568
##  1st Qu.: 55317014   Class :character   Class :character   Class :character
##  Median : 83365370   Mode  :character   Mode  :character   Mode  :character
##  Mean   :102218616
##  3rd Qu.:150772442
##  Max.   :222473262
##
##     PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Min.   :  1.00   Min.   :0.0000    Length:23568       Length:23568
##  1st Qu.: 44.00   1st Qu.:0.0000    Class :character   Class :character
##  Median : 69.00   Median :0.0000    Mode  :character   Mode  :character
##  Mean   : 66.21   Mean   :0.3323
##  3rd Qu.: 81.00   3rd Qu.:0.0000
##  Max.   :123.00   Max.   :2.0000
##                   NA's   :2
##  PERP_AGE_GROUP       PERP_SEX          PERP_RACE         VIC_AGE_GROUP
##  Length:23568       Length:23568       Length:23568       Length:23568
##  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:23568       Length:23568       Length:23568       Length:23568
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     Latitude       Longitude        Lon_Lat
##  Min.   :40.51   Min.   :-74.25   Length:23568
##  1st Qu.:40.67   1st Qu.:-73.94   Class :character
##  Median :40.70   Median :-73.92   Mode  :character
##  Mean   :40.74   Mean   :-73.91
##  3rd Qu.:40.82   3rd Qu.:-73.88
##  Max.   :40.91   Max.   :-73.70
##
```

Show internal

```
str(nypd)
```

```
## 'data.frame':    23568 obs. of  19 variables:
##  $ INCIDENT_KEY           : int  201575314 205748546 193118596 204192600 201483468 198255460 1945705
##  $ OCCUR_DATE             : chr  "08/23/2019" "11/27/2019" "02/02/2019" "10/24/2019" ...
##  $ OCCUR_TIME             : chr  "22:10:00" "15:54:00" "19:40:00" "00:52:00" ...
##  $ BORO                   : chr  "QUEENS" "BRONX" "MANHATTAN" "STATEN ISLAND" ...
##  $ PRECINCT               : int  103 40 23 121 46 73 81 67 114 69 ...
##  $ JURISDICTION_CODE      : int  0 0 0 0 0 0 0 0 2 0 ...
##  $ LOCATION_DESC          : chr  "" "" "" "PVT HOUSE" ...
##  $ STATISTICAL_MURDER_FLAG: chr  "false" "false" "false" "true" ...
##  $ PERP_AGE_GROUP         : chr  "" "<18" "18-24" "25-44" ...
##  $ PERP_SEX               : chr  "" "M" "M" "M" ...
##  $ PERP_RACE              : chr  "" "BLACK" "WHITE HISPANIC" "BLACK" ...
##  $ VIC_AGE_GROUP          : chr  "25-44" "25-44" "18-24" "25-44" ...
##  $ VIC_SEX                : chr  "M" "F" "M" "F" ...
##  $ VIC_RACE               : chr  "BLACK" "BLACK" "BLACK HISPANIC" "BLACK" ...
##  $ X_COORD_CD             : chr  "1037451" "1006789" "999347" "938149" ...
##  $ Y_COORD_CD             : chr  "193561" "237559" "227795" "171781" ...
##  $ Latitude               : num  40.7 40.8 40.8 40.6 40.9 ...
##  $ Longitude              : num  -73.8 -73.9 -73.9 -74.2 -73.9 ...
##  $ Lon_Lat                : chr  "POINT (-73.80814071699996 40.697805308000056)" "POINT (-73.9185706
```

- Confirmed as a **data.frame** with 23,568 rows of observations and 19 columns - dates **mdy** and times *as characters* along with the majority of data of interest for analysis.

Tidy and Transform

"This *.csv* is already a data.frame with vectors which simplifies the cleaning. Copying the list of columns allows for easier writing of code with fewer errors as some are capitalized, some are not and not all contain the *underscore*."

List of columns

INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO PRECINCT JURISDICTION_CODE LOCATION_DESC STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX PERP_RACE VIC_AGE_GROUP VIC_SEX VIC_RACE X_COORD_CD Y_COORD_CD Latitude Longitude Lon_Lat

Remove columns unused for Analysis

```
nypd$JURISDICTION_CODE <-NULL
nypd$PRECINCT <-NULL
nypd$X_COORD_CD <-NULL
nypd$INCIDENT_KEY <-NULL
nypd$Y_COORD_CD <-NULL
nypd$Lon_Lat <-NULL
nypd$Latitude <-NULL
nypd$Longitude <-NULL
```

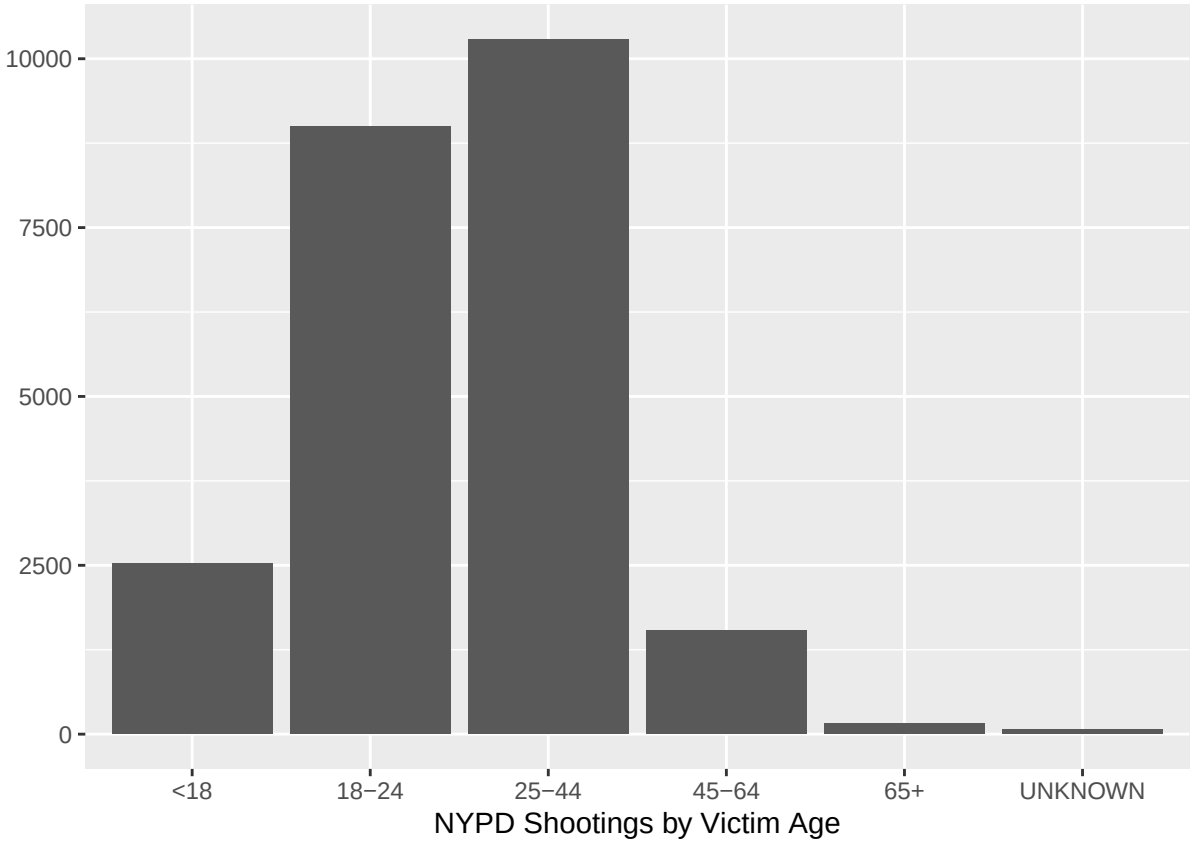
Missing data

The NYPD Historic Shooting Data runs for fourteen years 2006-2020, for many of the rows (observation points) data is missing in relation to perpetrators. This includes 8,425 cases where Perp Age Group, Sex and Race are all missing. The PERP_AGE_GROUP contains three notably inaccurate age groups - 224,940 and 1020 - that were most likely entry errors. In addition, sex in both perpetrator and victim columns is listed as U for unknown in some cases instead of being blank. Data is complete for date, time, and BORO which provides insight into timing and location of occurrences. For this purpose, the data is the data and will allow the blank counts to appear as blanks.

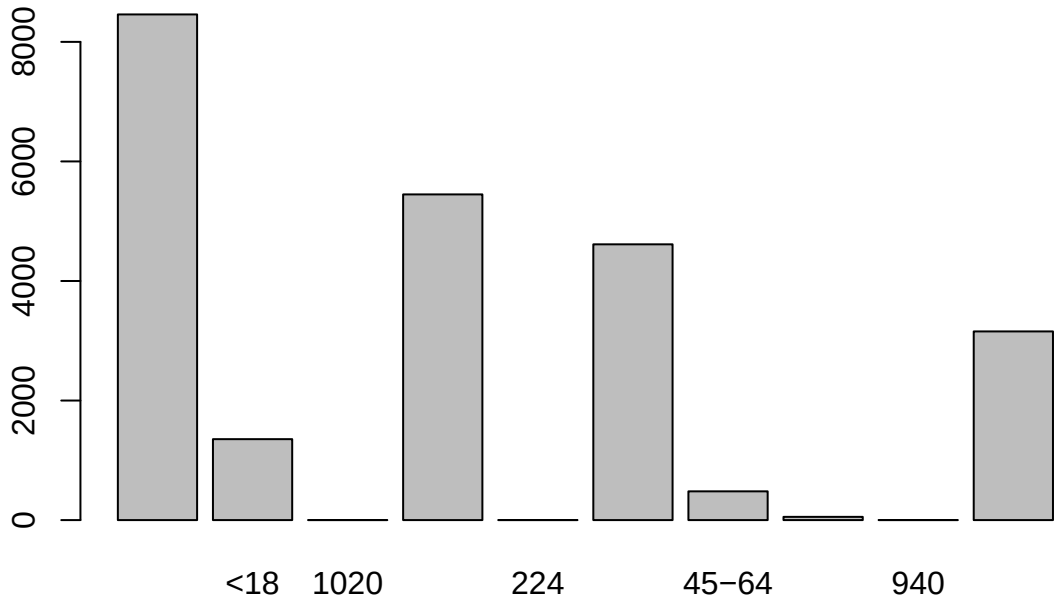
Visualize data points

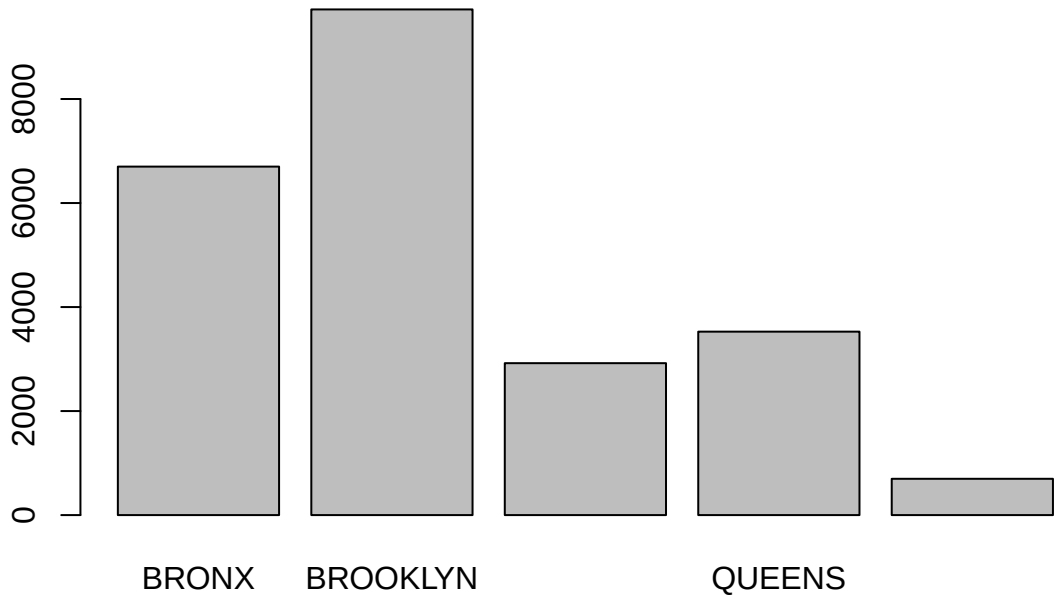
Raw data Plot of Occurences by Victim Age Groups

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

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



Comparison with Barplot of Occurrences by Pepetrator Age Groups


0
2000
4000
6000
8000
<18
1020
224
45−64
940



Manhattan (2921) and **Staten Island** (698) shootings respectively are not labeled

Conclusion and Bias

Serious bias - this is my first time creating an R Markdown document and I kept it simple. I enjoyed seeing the data become real, but it was a challenge to write the code and then make the visuals appear.

In thinking of my own Human Resources perspective, the only conclusion I would draw from this data, would be to evaluate whether enough effective mindfulness or mental health supports are available for officers. For usefulness, in addition to shooting data, BORO population numbers would be useful along with size of police forces in each BORO. I would also consider the possibility of increasing community information regarding gun safety.

Before providing this information for legislative or business decision making, I feel that additional data is required for accurate analysis and comparison. This data might incorprate data from similar cities or additional internal data that accurately documents police shooting data to determine if any of the blanks are NYPD department related.