NYC Shooting Project

student

7/12/2021

NYC Shooting:

This is a breakdown of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year. (source: https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic)

Questions:

1) What is the trend with regard to years?

2) What is the lowest/highest number of incidents per victim age group?

Bias:

Coming from New York, by bias was based on pride. Finding out what borough had the highest number was my focus and the age groups within that area. I think I wanted to prove the area that I grew up in had a lower incident rate than others. In regards to selecting the data, well, that was the assignment, but I asked the question if I would do this on my own, and yes, I would – so that needs to be accounted for as well. Even though this RMD file does not have further studies, I reviewed later, other factors, including race, age, and murder count. These to have a question of bias attached to them.

Note: Other bias reasons: A byproduct of limited knowledge of the field of analytics & knowledge of R, which is the reason for taking said class.

Loading R Packages:

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

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

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.0.4
```

```
## Warning: package 'tidyr' was built under R version 4.0.4
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
## Warning: package 'purrr' was built under R version 4.0.3

## Warning: package 'dplyr' was built under R version 4.0.4

## Warning: package 'stringr' was built under R version 4.0.3

## Warning: package 'forcats' was built under R version 4.0.5

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

## Warning: package 'lubridate' was built under R version 4.0.5

##
## Attaching package: 'lubridate'

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

Reading in the Data:

```
#url_in <-"https://data.cityofnewyork.us/resource/833y-fsy8.csv" - appears to be an issue, seek other p
url_in <-"https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
dataShooting <- read.csv(url_in)
```

Summary of Data:

```
summary(dataShooting)
```

```
##   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
##
```

```
head(dataShooting)
```

```
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME          BORO PRECINCT JURISDICTION_CODE
## 1    201575314 08/23/2019   22:10:00        QUEENS      103                 0
## 2    205748546 11/27/2019   15:54:00         BRONX       40                 0
## 3    193118596 02/02/2019   19:40:00     MANHATTAN       23                 0
## 4    204192600 10/24/2019   00:52:00 STATEN ISLAND      121                 0
## 5    201483468 08/22/2019   18:03:00         BRONX       46                 0
## 6    198255460 06/07/2019   17:50:00      BROOKLYN       73                 0
##   LOCATION_DESC STATISTICAL_MURDER_FLAG PERP_AGE_GROUP PERP_SEX      PERP_RACE
## 1                                false
## 2                                false            <18        M          BLACK
## 3                                false          18-24        M WHITE HISPANIC
## 4     PVT HOUSE                   true          25-44        M          BLACK
## 5                                false          25-44        M BLACK HISPANIC
## 6                                false          45-64        M WHITE HISPANIC
##   VIC_AGE_GROUP VIC_SEX       VIC_RACE X_COORD_CD Y_COORD_CD Latitude Longitude
## 1         25-44       M          BLACK    1037451     193561 40.69781 -73.80814
## 2         25-44       F          BLACK    1006789     237559 40.81870 -73.91857
## 3         18-24       M BLACK HISPANIC     999347     227795 40.79192 -73.94548
## 4         25-44       F          BLACK     938149     171781 40.63806 -74.16611
## 5         18-24       M          BLACK    1008224     250621 40.85455 -73.91334
## 6         25-44       M          BLACK    1009650     186966 40.67983 -73.90843
##                                      Lon_Lat
## 1 POINT (-73.80814071699996 40.697805308000056)
## 2  POINT (-73.91857061799993 40.81869973000005)
## 3 POINT (-73.94547965999999 40.791916091000076)
## 4  POINT (-74.16610830199996 40.63806398200006)
## 5  POINT (-73.91333944399999 40.85454734900003)
## 6  POINT (-73.90842523899994 40.67982701600005)
```

```
# DETAILS OF SUMMARY:
  # INCIDENT_KEY: provides a unique key/id to every row
  # OCCUR_DATE: provides date of incident
  # OCCUR_TIME: provides time of incident
  # BORO: provides the location per boro (five boros to NY City)
  # PRECINCT: provides the police station reasonable
  # JURISDICTION_CODE: provides police station location and j code
  # LOCATION_DESC: provides location of incident
  # STATISTICAL_MURDER_FLAG: provides details of shooting (fatal)
  # PERP_AGE_GROUP, PERP_SEX, PERP_RACE: provides details of perpetrator
  # VIC_AGE_GROUP, VIC_SEX, VIC_RACE: provides details of victim
  # X_COORD_CD, Y_COORD_CD, LATITUDE, LONGITUDE, GEOCODED_COLUMN: provides location information of inci
```

 # TIDY DATA (a wee bit):

```
#dataShooting <- separate(dataShooting,occur_date, into = c("date", "timeRemove"), sep = "T") - basic, 
#dataShooting$BORO <- as.factor(dataShooting$BORO) - factor works, but issues with plots, dropped!

dataShooting <- mutate(dataShooting, OCCUR_DATE = mdy(OCCUR_DATE), dsSHOOTINGS = 1,
                       YEARS = format_ISO8601(OCCUR_DATE, precision = "y"),
                       YEAR_MONTHS = format_ISO8601(OCCUR_DATE, precision = "ym"))
```

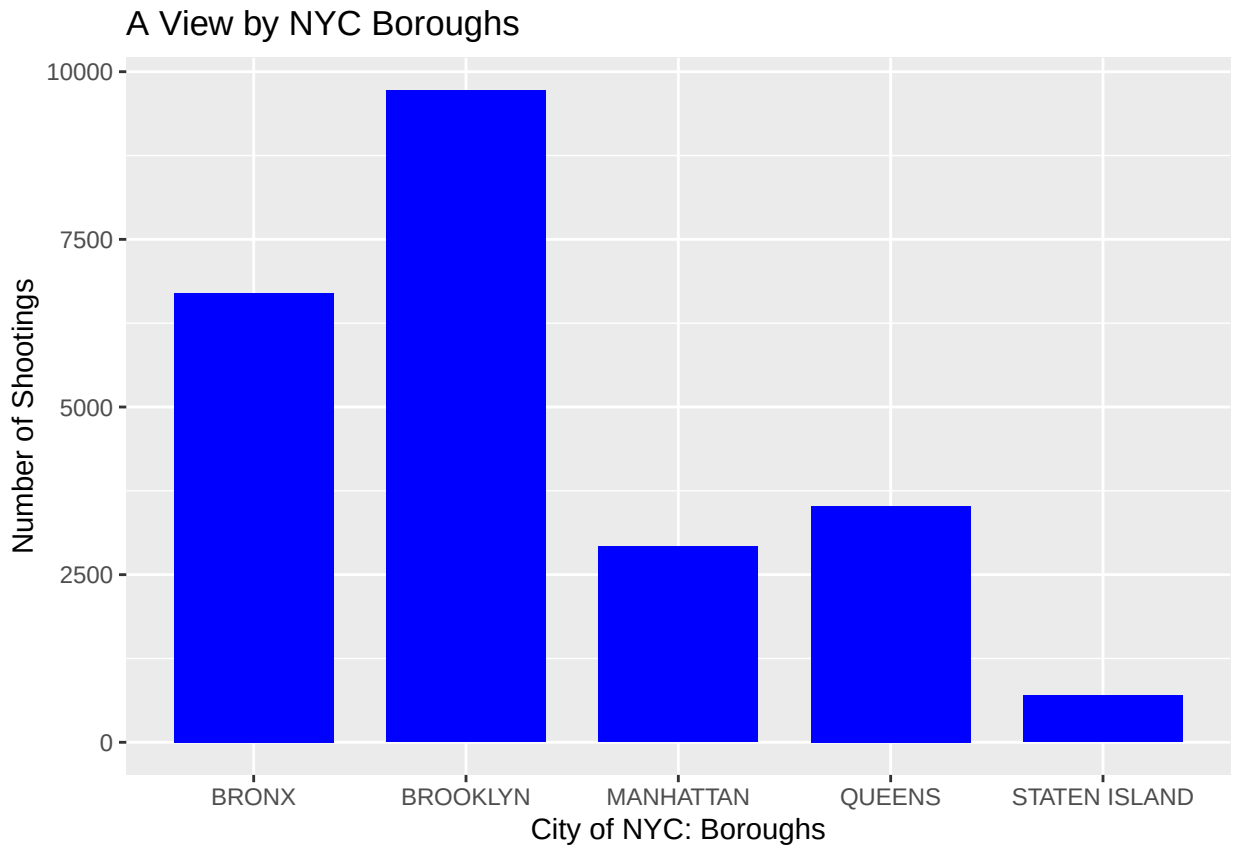
SELECTING COLUMNS:

```
dsPlotReview <- select(dataShooting, OCCUR_DATE, VIC_AGE_GROUP, YEAR_MONTHS, YEARS, OCCUR_TIME, BORO, ds
  filter(dsSHOOTINGS == "1")
```

Three Basic Visualizations:

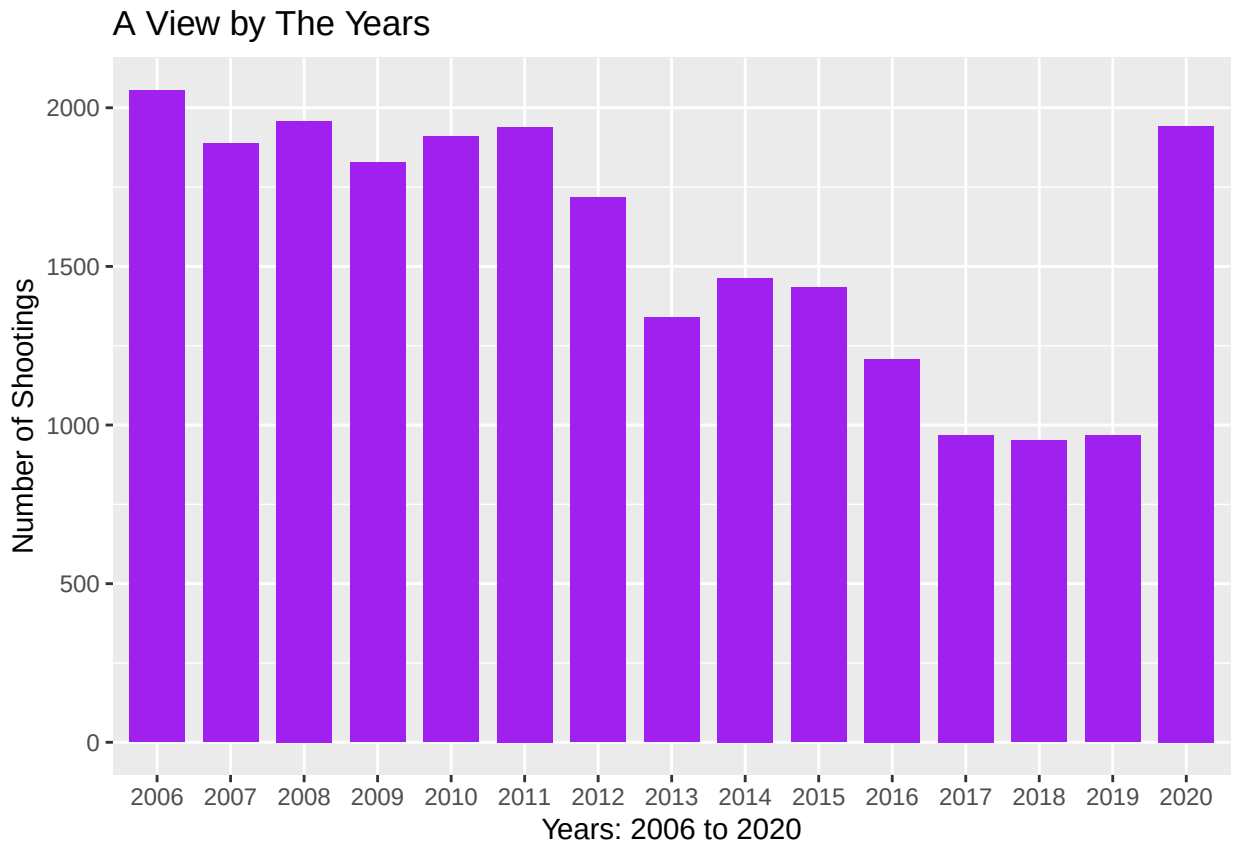
```
dsPlotOne <- dsPlotReview %>%
  group_by(BORO) %>%
  count()

dsPlotViewOne <-ggplot(dsPlotOne, aes(x = BORO, weight =n)) +
  geom_bar(width = .75, fill = "blue") +
  labs(x = "City of NYC: Boroughs", y = "Number of Shootings",
       title = "A View by NYC Boroughs")
plot(dsPlotViewOne)
```



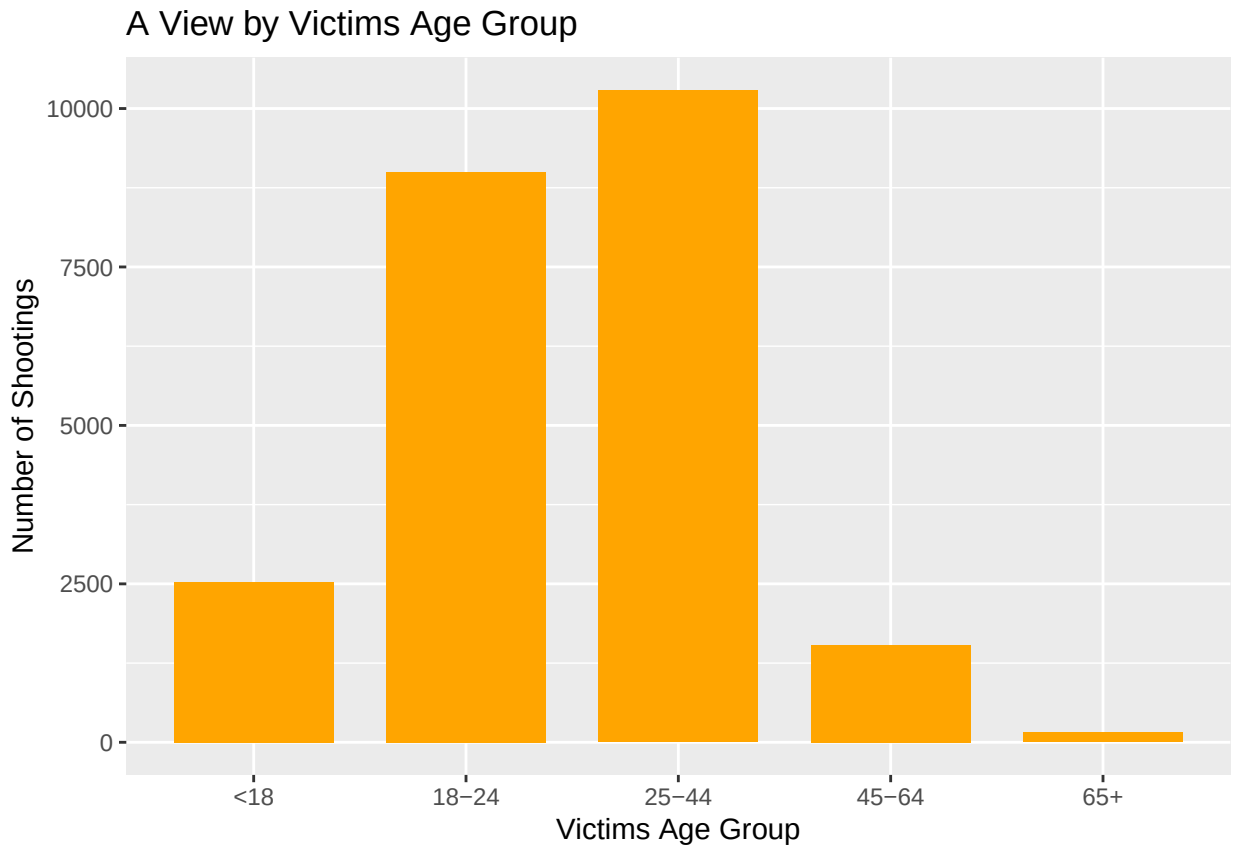

```
#-------------------------------------------------------------
dsPlotTwo <- dsPlotReview %>%
  group_by(YEARS) %>%
  count()

dsPlotViewTwo <-ggplot(dsPlotTwo, aes(x = YEARS, weight =n)) +
  geom_bar(width = .75, fill = "purple") +
  labs(x = "Years: 2006 to 2020", y = "Number of Shootings",
       title = "A View by The Years")
plot(dsPlotViewTwo)
```




```
#-------------------------------------------------------------
dsPlotThree <- dsPlotReview %>%
  group_by(VIC_AGE_GROUP) %>%
  filter(VIC_AGE_GROUP != "UNKNOWN") %>%
  count()

dsPlotViewThree <-ggplot(dsPlotThree, aes(x = VIC_AGE_GROUP, weight =n)) +
  geom_bar(width = .75, fill = "orange") +
  labs(x = "Victims Age Group", y = "Number of Shootings",
       title = "A View by Victims Age Group")
plot(dsPlotViewThree)
```



Conclusion:

Review of Bias:

Coming from New York, by bias was based on pride. Finding out what borough had the highest number was my focus and the age groups within that area. I think I wanted to prove the area that I grew up in had a lower incident rate than others. In regards to selecting the data, well, that was the assignment, but I asked the question if I would do this on my own, and yes, I would – so that needs to be accounted for as well. Even though this RMD file does not have further studies, I reviewed later, other factors, including race, age, and murder count. These to have a question of bias attached to them.

Note: Other bias reasons: A byproduct of limited knowledge of the field of analytics & knowledge of R, which is the reason for taking said class.

Question & Answers:

1) What is the trend with regard to years? - According to the data, incidents were going down, but made a sharp turn up during the past year (2020).

2) What is the lowest/highest number of incidents per victim age group? - According to the data, the age group of "25-44" has the greatest number incidents.