Project1: Police Shooting Data

Jared Tesar

2022-09-02

```
knitr::opts_chunk$set(echo = TRUE)
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(ggplot2)
initData <- read.csv("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD")
```

NYPD Shooting Data

This data includes metrics concerning every recording shooting incident involving the New York Police Department dating back to 2006 through to the end of the previous calendar year. Each record represents a shooting incident, and each column refers to a datapoint associated with that incident.

This project will attempt to visualize the difference in age group victimhood, considered between and across sexes and neighborhoods.

Tidying Data

```
interimData <- initData %>%
  select(c(BORO, VIC_AGE_GROUP, VIC_SEX)) %>%
  filter(VIC_AGE_GROUP != "UNKNOWN")  %>%
  filter(VIC_SEX != "U")
interimData$factoredSex <- factor(interimData$VIC_SEX)
interimData$factoredAge <- factor(interimData$VIC_AGE_GROUP)
interimData$factoredBoro <- factor(interimData$BORO)
```

```
summary(interimData)
```

```
##      BORO           VIC_AGE_GROUP        VIC_SEX          factoredSex
##  Length:25530       Length:25530       Length:25530       F: 2398
##  Class :character   Class :character   Class :character   M:23132
##  Mode  :character   Mode  :character   Mode  :character
##
##
##  factoredAge          factoredBoro
##  <18  : 2681   BRONX        : 7383
##  18-24: 9600   BROOKLYN     :10337
##  25-44:11384   MANHATTAN    : 3258
##  45-64: 1698   QUEENS       : 3817
##  65+  :  167   STATEN ISLAND:  735
```

Grouping Data

```
ageByBoro <- interimData %>% group_by(factoredBoro)%>% count(factoredAge)
ageByBoro
```

```
## # A tibble: 25 x 3
## # Groups:   factoredBoro [5]
##    factoredBoro factoredAge     n
##    <fct>        <fct>       <int>
##  1 BRONX        <18           869
##  2 BRONX        18-24        2887
##  3 BRONX        25-44        3138
##  4 BRONX        45-64         445
##  5 BRONX        65+            44
##  6 BROOKLYN     <18          1060
##  7 BROOKLYN     18-24        3825
##  8 BROOKLYN     25-44        4657
##  9 BROOKLYN     45-64         727
## 10 BROOKLYN     65+            68
## # ... with 15 more rows
```

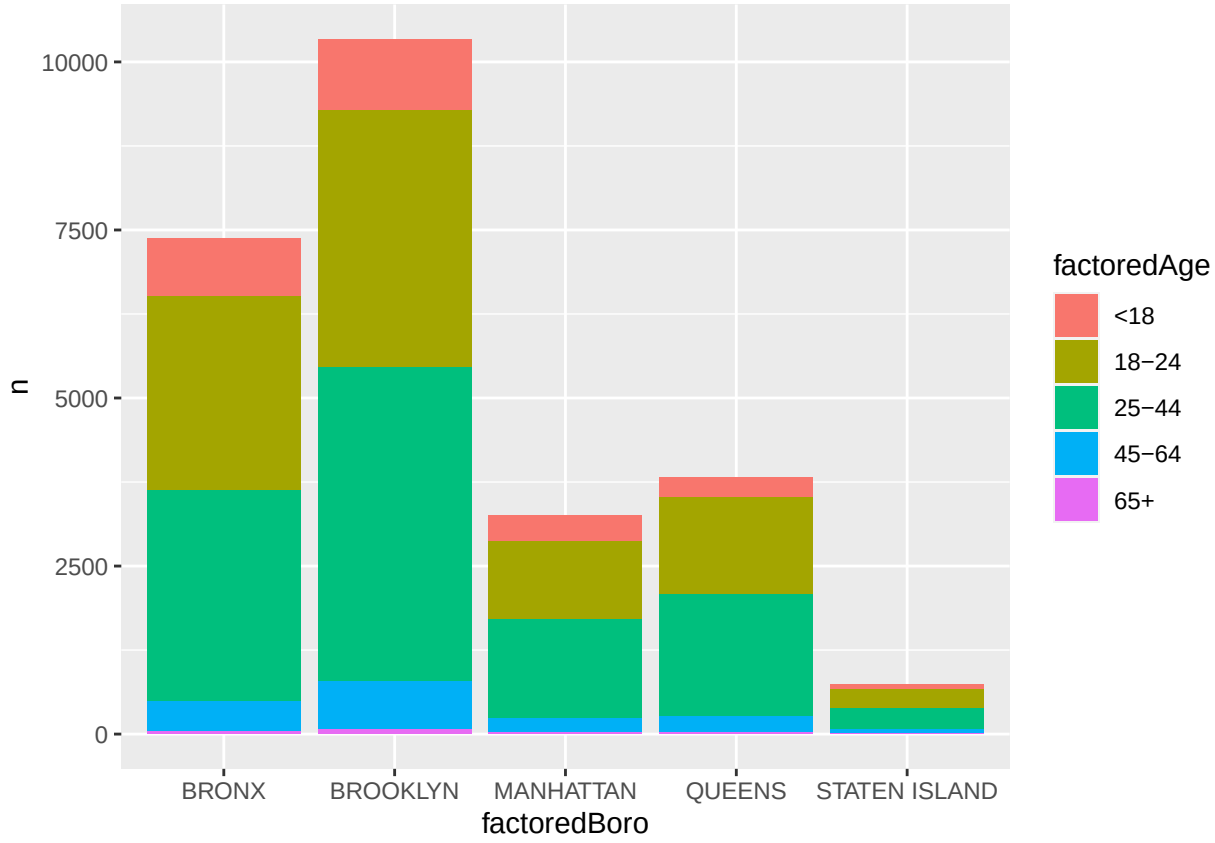
```
ageBySex <- interimData %>% group_by(factoredSex) %>% count(factoredAge)
ageBySex
```

```
## # A tibble: 10 x 3
## # Groups:   factoredSex [2]
##    factoredSex factoredAge     n
##    <fct>       <fct>       <int>
##  1 F           <18           376
##  2 F           18-24         732
##  3 F           25-44         914
##  4 F           45-64         322
##  5 F           65+            54
##  6 M           <18          2305
##  7 M           18-24        8868
```

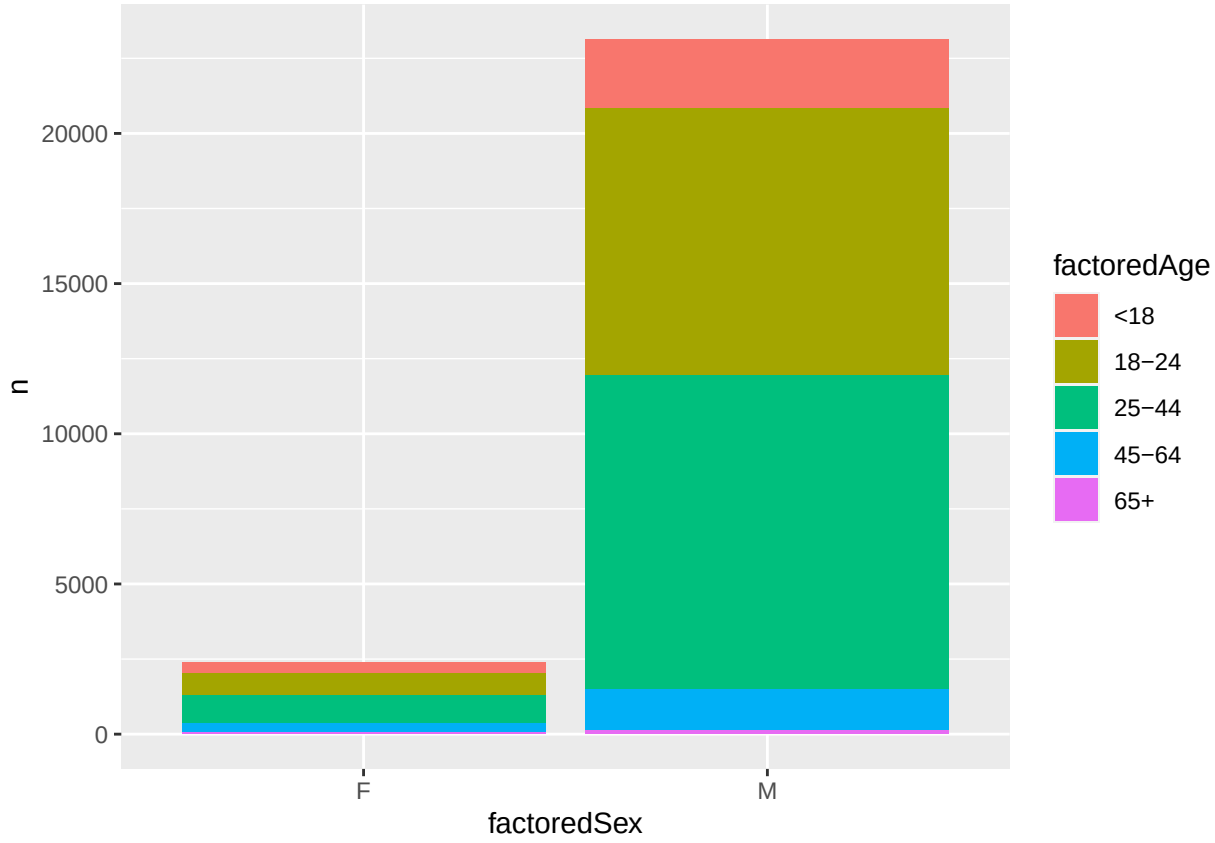
```
##  8 M           25-44       10470
##  9 M           45-64        1376
## 10 M           65+           113
```

Visualizing Data

```
boroPlot <- ggplot(aes(x = factoredBoro, y = n, fill=factoredAge), data = ageByBoro) +
  geom_bar(stat = "identity")
boroPlot
```




```
sexPlot <- ggplot(aes(x = factoredSex, y = n, fill=factoredAge), data = ageBySex) +
  geom_bar(stat = "identity")
sexPlot
```



Analysis

These visualizations show stark differences in total shooting incidents, but the ratio between age groups appears to be mostly consistent. Males are much more commonly involved in shootings, as are individuals between 18 and 44 years of age. Different neighborhoods have different rates of shootings. It would be interesting to further investigate differences like wealth among these neighborhoods to possibly explain the disparities.

Modelling Data

```
summary(lm(n ~ factoredBoro, data = ageByBoro))
```

```
##
## Call:
## lm(formula = n ~ factoredBoro, data = ageByBoro)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1999.4  -630.6  -138.0   679.6  2589.6
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1476.6      539.4   2.738   0.0127 *
```

```
## factoredBoroBROOKLYN         590.8      762.8   0.775   0.4477
## factoredBoroMANHATTAN       -825.0      762.8  -1.082   0.2923
## factoredBoroQUEENS          -713.2      762.8  -0.935   0.3609
## factoredBoroSTATEN ISLAND  -1329.6      762.8  -1.743   0.0967 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1206 on 20 degrees of freedom
## Multiple R-squared:  0.2806, Adjusted R-squared:  0.1367
## F-statistic:  1.95 on 4 and 20 DF,  p-value: 0.1414
```

```
summary(lm(n ~ factoredSex, data = ageBySex))
```

```
##
## Call:
## lm(formula = n ~ factoredSex, data = ageBySex)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4513.4 -1847.5  -130.6   388.9  5843.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     479.6     1491.1   0.322   0.7560
## factoredSexM   4146.8     2108.7   1.967   0.0848 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3334 on 8 degrees of freedom
## Multiple R-squared:  0.3259, Adjusted R-squared:  0.2416
## F-statistic: 3.867 on 1 and 8 DF,  p-value: 0.0848
```

Biases

This analysis does not take into account the relative populations of men, women, or each age group considered. A larger population would result in a higher likelihood for random chance to cause the discrepancy portrayed here. Nor does this analysis take into account groups affiliated with like gangs or general criminal behavior rates between sexes and age groups. As for personal bias, I have a hard time noticing any. Perhaps I just need more practice in this skill. I would say I am biased to expect a difference in data between age groups and sexes, and I mitigated this by accepting the age groups as defined by the NYPD data and did not otherwise attempt to guarantee a difference among the data groups.