NYPD Shooting Incident Data Analysis

Ryan Phillips

2023-06-14

Necessary packages for this analysis: tidyverse core packages.

Step 0: Import Library

```r
##install.packages("tidyverse") to install
library(tidyverse)
```

Step 1: Read in Data

Begin by reading in data from csv file from the City of New York.

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

Step 2: Clean and Transform Data

Select only columns relevant to our investigation of this data

```r
NYPD_data <- NYPD_data %>% mutate(OCCUR_DATE= mdy(OCCUR_DATE)) %>% select(-c(JURISDICTION_CODE,INCIDENT_
```

Rename Date and time

```r
NYPD_data <- NYPD_data %>% rename(Date = OCCUR_DATE, Time = OCCUR_TIME)
```

Filter out NA and UNKNOWN variables

```r
NYPD_data_tidy <- NYPD_data %>% drop_na()
# filter out Unknown vic age group variables
NYPD_data_tidy <- NYPD_data_tidy %>% filter(!grepl('UNKNOWN', VIC_AGE_GROUP))
NYPD_data_tidy <- NYPD_data_tidy %>% filter(!grepl('1022', VIC_AGE_GROUP))
```

Summary of our cleaned data

```r
summary(NYPD_data_tidy)
```

```
##       Date              Time                BORO          PERP_AGE_GROUP
##  Min.   :2006-01-01   Length:17911       Length:17911       Length:17911
##  1st Qu.:2008-08-06   Class1:hms         Class :character   Class :character
##  Median :2011-11-17   Class2:difftime    Mode  :character   Mode  :character
##  Mean   :2013-05-12   Mode  :numeric
##  3rd Qu.:2018-04-28
##  Max.   :2022-12-31
##     PERP_SEX          VIC_AGE_GROUP        VIC_SEX
##  Length:17911       Length:17911       Length:17911
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

Step 3: Add Analysis and Visualization

Lets sort victim age groups in descending order by total number, to see which age group has the most
instances.

```
NYPD_data_tidy %>% group_by(VIC_AGE_GROUP) %>% summarise(Total=n()) %>% arrange(desc(Total))
```

```
## # A tibble: 5 x 2
##   VIC_AGE_GROUP Total
##   <chr>         <int>
## 1 25-44          7939
## 2 18-24          6518
## 3 <18            2027
## 4 45-64          1290
## 5 65+             137
```

Create a bar chart to visualize instances per age group.

```
vic_age_data <- NYPD_data_tidy %>% group_by(VIC_AGE_GROUP) %>% summarize(incidents = n())

ggplot(vic_age_data, aes(x=VIC_AGE_GROUP, y=incidents, fill=VIC_AGE_GROUP)) +
  geom_bar(stat = "identity") +
  xlab("AGE") + ylab("INCIDENTS") +
  ggtitle("SHOOTING INCIDENTS BY AGE OF VICTIMS")+
  theme_light()
```

SHOOTING INCIDENTS BY AGE OF VICTIMS
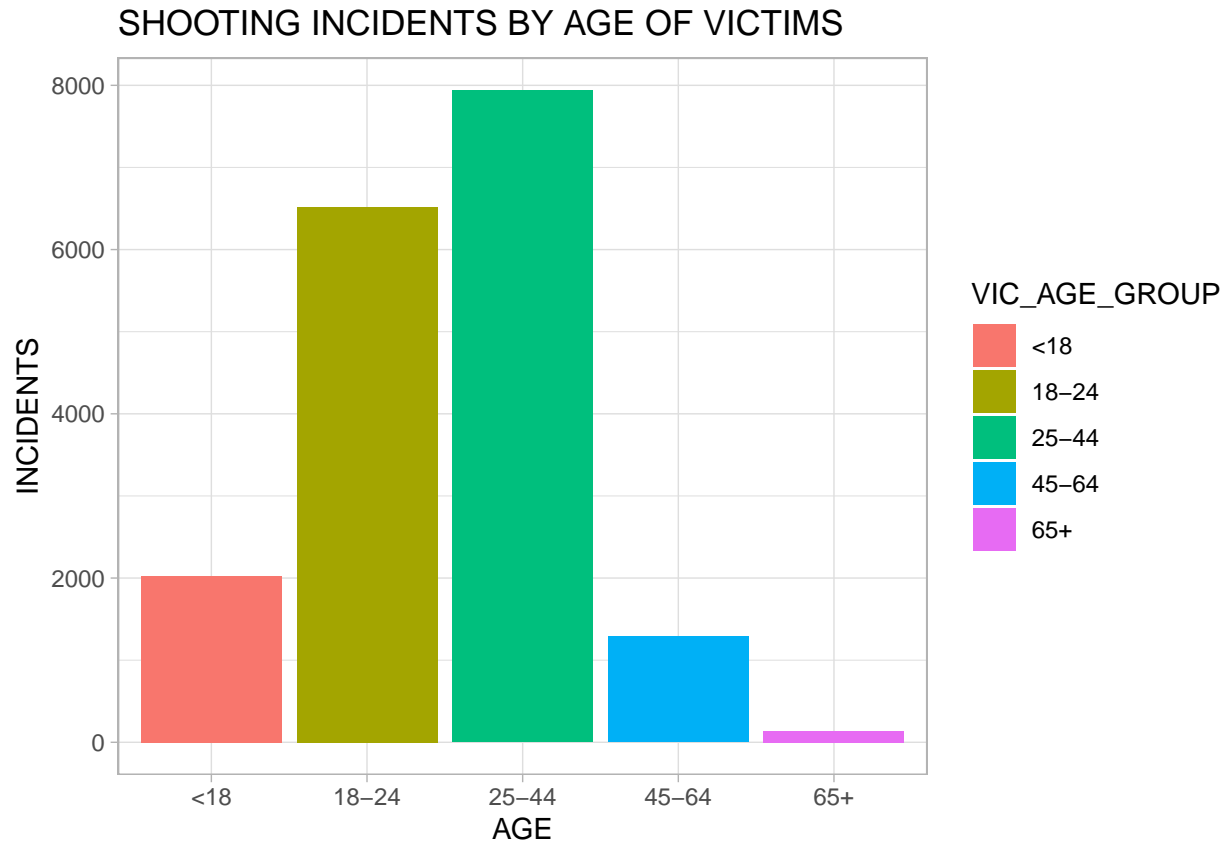
I want to see which age group has the highest chance of becoming a victim of a shooting. I begin by creating a population column for each age range.

```
NYPD_data_tidy <- NYPD_data_tidy %>%
  mutate(Population = case_when(
    VIC_AGE_GROUP == "<18" ~ 1798842,
    VIC_AGE_GROUP == "18-24" ~ 773258,
    VIC_AGE_GROUP == "25-44" ~ 2708853,
    VIC_AGE_GROUP == "45-64" ~ 2101599,
    VIC_AGE_GROUP == "65+" ~ 1155075,
    TRUE ~ NA_real_
  ))
```

Next I calculate and plot the shooting incident rate per 100,000 people for each age group.

```
vic_age_data_rate <- NYPD_data_tidy %>%
  group_by(VIC_AGE_GROUP) %>%
  summarise(total_shootings = n(),
            population = unique(Population),
            shooting_rate = total_shootings / (population / 100000)) %>%
  arrange(desc(shooting_rate))

ggplot(vic_age_data_rate, aes(x = VIC_AGE_GROUP, y = shooting_rate)) +
  geom_bar(stat = "identity", fill = "turquoise", color = "black") +
  ggtitle("Chance of Getting Shot by Age") +
  xlab("Age") +
```

3

```
ylab("Shooting Rate per 100,000 People") +
theme_minimal()
```

Chance of Getting Shot by Age
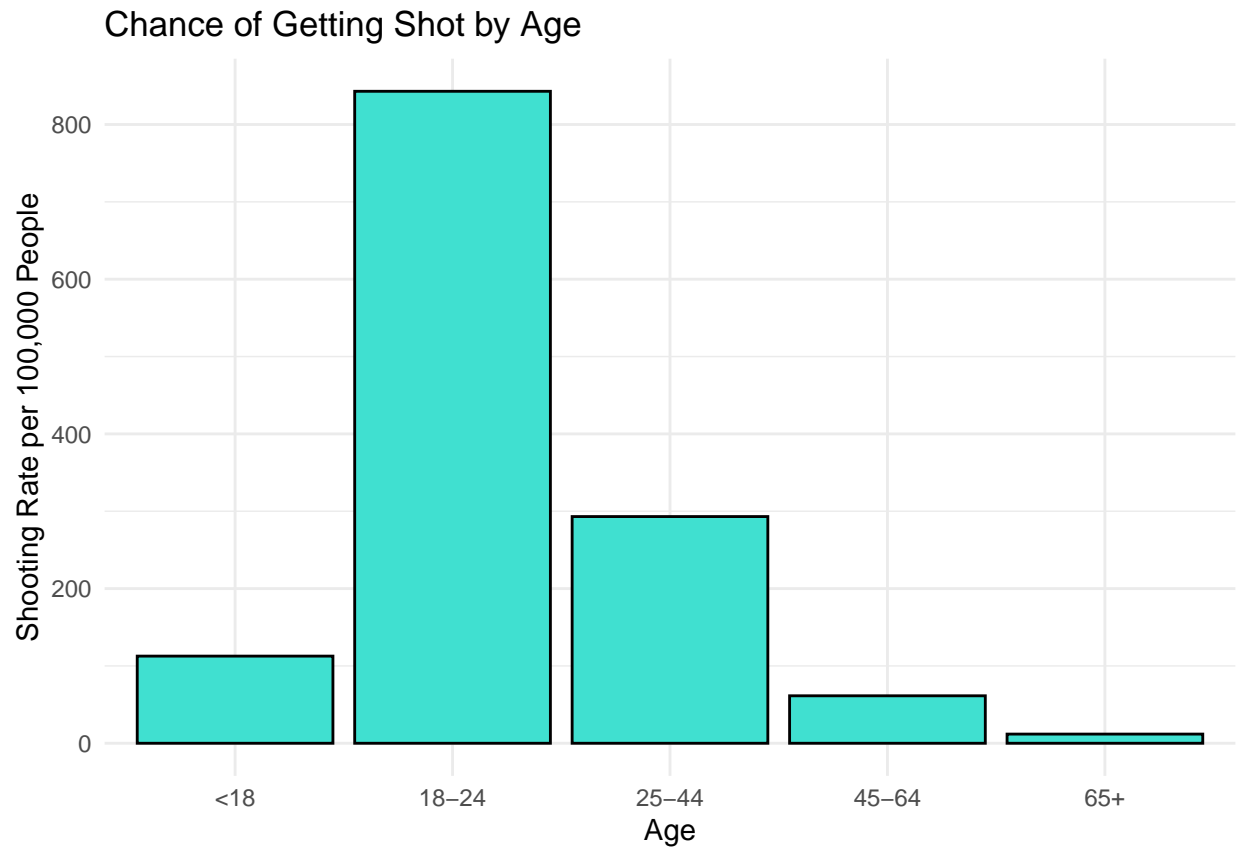
Now I would like to see what time of day that shootings are most likely to occur.

```
NYPD_data_tidy$HOUR = hour(hms(as.character(NYPD_data_tidy$Time)))

occur_hour = NYPD_data_tidy %>% group_by(HOUR) %>% count()

ggplot(occur_hour, aes(x=HOUR, y=n)) +
  geom_line() +
  labs(title = "What time should people be most cautious of shootings in NYC",
       x = "Hour Occured",
       y= "Number of Incidents") +
  theme_minimal()
```

What time should people be most cautious of shootings in NYC