Analysis on the Age of Shooting Victims in NYC

Martin

2023-07-27

For this project, we will be using the Historic NYPD Shoooting incident data to see if we can answer the question:

"Are you less likely to be the victim of a shooting as you age in NYC?"

Preparing data for use

Link to the source data: https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD

Our first steps are to take a look at the data, find the fields which will be most useful for analysis, and clean up any blank records by using the exisitng "UNKNOWN" and "U" values.

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr   1.1.2     v readr   2.1.4
## v forcats 1.0.0     v stringr 1.5.0
## v ggplot2 3.4.2     v tibble  3.2.1
## v purrr   1.0.1     v tidyr   1.3.0

## -- 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
```

```r
library(dplyr)

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

NYPD_Data <- read.csv(url)

summary(NYPD_Data)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:27312       Length:27312       Length:27312
##  1st Qu.: 63860880   Class :character   Class :character   Class :character
##  Median : 90372218   Mode  :character   Mode  :character   Mode  :character
##  Mean   :120860536
##  3rd Qu.:188810230
##  Max.   :261190187
##
##  LOC_OF_OCCUR_DESC     PRECINCT      JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:27312       Min.   :  1.00   Min.   :0.0000    Length:27312
##  Class :character   1st Qu.: 44.00   1st Qu.:0.0000    Class :character
##  Mode  :character   Median : 68.00   Median :0.0000    Mode  :character
##                     Mean   : 65.64   Mean   :0.3269
##                     3rd Qu.: 81.00   3rd Qu.:0.0000
##                     Max.   :123.00   Max.   :2.0000
##                                      NA's   :2
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP
##  Length:27312       Length:27312            Length:27312
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##
##
##
##
##    PERP_SEX          PERP_RACE         VIC_AGE_GROUP        VIC_SEX
##  Length:27312       Length:27312       Length:27312       Length:27312
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    VIC_RACE           X_COORD_CD        Y_COORD_CD        Latitude
##  Length:27312       Min.   : 914928   Min.   :125757   Min.   :40.51
##  Class :character   1st Qu.:1000029   1st Qu.:182834   1st Qu.:40.67
##  Mode  :character   Median :1007731   Median :194487   Median :40.70
##                     Mean   :1009449   Mean   :208127   Mean   :40.74
##                     3rd Qu.:1016838   3rd Qu.:239518   3rd Qu.:40.82
##                     Max.   :1066815   Max.   :271128   Max.   :40.91
##                                                        NA's   :10
##    Longitude         Lon_Lat
##  Min.   :-74.25   Length:27312
##  1st Qu.:-73.94   Class :character
##  Median :-73.92   Mode  :character
##  Mean   :-73.91
##  3rd Qu.:-73.88
##  Max.   :-73.70
##  NA's   :10
```

```r
NYPD_Data <- NYPD_Data %>%
  select(-c(Latitude, Longitude, Y_COORD_CD, X_COORD_CD, Lon_Lat, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC

NYPD_Data <- NYPD_Data %>%
  mutate_at(c("PERP_RACE"), ~na_if(., ''))
```

```r
NYPD_Data[is.na(NYPD_Data)] <- "UNKNOWN"

NYPD_Data <- NYPD_Data %>%
  mutate_at(c("PERP_AGE_GROUP"), ~na_if(., ''))

NYPD_Data[is.na(NYPD_Data)] <- "UNKNOWN"

NYPD_Data <- NYPD_Data %>%
  mutate_at(c("PERP_SEX"), ~na_if(., ''))

NYPD_Data[is.na(NYPD_Data)] <- "U"

NYPD_Data <- NYPD_Data %>%
  mutate_at(c("PERP_RACE"), ~na_if(., '(null)'))

NYPD_Data[is.na(NYPD_Data)] <- "UNKNOWN"

NYPD_Data <- NYPD_Data %>%
  mutate_at(c("PERP_AGE_GROUP"), ~na_if(., '(null)'))

NYPD_Data[is.na(NYPD_Data)] <- "UNKNOWN"

NYPD_Data <- NYPD_Data %>%
  mutate_at(c("PERP_SEX"), ~na_if(., '(null)'))

NYPD_Data[is.na(NYPD_Data)] <- "U"

summary(NYPD_Data)
```

```
##   OCCUR_DATE          OCCUR_TIME             BORO               PRECINCT
##  Length:27312       Length:27312       Length:27312       Min.   :  1.00
##  Class :character   Class :character   Class :character   1st Qu.: 44.00
##  Mode  :character   Mode  :character   Mode  :character   Median : 68.00
##                                                           Mean   : 65.64
##                                                           3rd Qu.: 81.00
##                                                           Max.   :123.00
##  JURISDICTION_CODE  LOCATION_DESC      STATISTICAL_MURDER_FLAG
##  Length:27312       Length:27312       Length:27312
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##  PERP_AGE_GROUP       PERP_SEX          PERP_RACE          VIC_AGE_GROUP
##  Length:27312       Length:27312       Length:27312       Length:27312
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    VIC_SEX            VIC_RACE
##  Length:27312       Length:27312
##  Class :character   Class :character
```

```
##   Mode  :character   Mode  :character
##
##
##
```

Now that we have cleaned up some of the fields we can start to create smaller data frames for each borough and try to find other subsets of data which might interest us.

Aggregating Data

Within this sections, we'll try to break up the data across the boroughs to be able to better view everything in smaller chunks and maybe help form some data across the city that we can look more into.

We'll start with aggregating data across the city as a whole and move across boroughs alphabetically.

```
NY_TOTALS <- NYPD_Data %>%
  select(-c(OCCUR_DATE, OCCUR_TIME, JURISDICTION_CODE, LOCATION_DESC)) %>%
  group_by(BORO) %>%
  mutate(Incidents_by_Boro = n()) %>%
  ungroup() %>%
  group_by(VIC_RACE) %>%
  mutate(Victims_by_Race = n()) %>%
  ungroup() %>%
  select(c(BORO, everything())) %>%
  group_by(VIC_SEX) %>%
  mutate(Victims_by_Sex = n()) %>%
  ungroup() %>%
  group_by(VIC_AGE_GROUP) %>%
  mutate(Victims_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_AGE_GROUP) %>%
  mutate(Perp_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_SEX) %>%
  mutate(Perp_by_Sex = n()) %>%
  ungroup() %>%
  group_by(PERP_RACE) %>%
  mutate(Perp_by_Race = n()) %>%
  ungroup() %>%
  group_by(PRECINCT) %>%
  mutate(Incidents_by_Prec = n()) %>%
  ungroup() %>%
  group_by(STATISTICAL_MURDER_FLAG) %>%
  mutate(Incidents_by_Stat = n())

NY_TOT_Agg <- NY_TOTALS %>%
  select(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
         Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
         VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
         PERP_SEX, Perp_by_Sex) %>%
  group_by(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
           Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
           VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
```

```
           PERP_SEX, Perp_by_Sex) %>%
  arrange(BORO, PRECINCT, VIC_RACE, VIC_AGE_GROUP,
          VIC_SEX, PERP_RACE, PERP_AGE_GROUP, PERP_SEX, STATISTICAL_MURDER_FLAG)

Bronx <- NYPD_Data %>%
  filter(BORO == "BRONX")

Brooklyn <- NYPD_Data %>%
  filter(BORO == "BROOKLYN")

Manhattan <- NYPD_Data %>%
  filter(BORO == "MANHATTAN")

Queens <- NYPD_Data %>%
  filter(BORO == "QUEENS")

Staten_Island <- NYPD_Data %>%
  filter(BORO == "STATEN ISLAND")

Bronx_Totals <- Bronx %>%
  select(-c(OCCUR_DATE, OCCUR_TIME, JURISDICTION_CODE, LOCATION_DESC)) %>%
  group_by(BORO) %>%
  mutate(Incidents_by_Boro = n()) %>%
  ungroup() %>%
  group_by(VIC_RACE) %>%
  mutate(Victims_by_Race = n()) %>%
  ungroup() %>%
  select(c(BORO, everything())) %>%
  group_by(VIC_SEX) %>%
  mutate(Victims_by_Sex = n()) %>%
  ungroup() %>%
  group_by(VIC_AGE_GROUP) %>%
  mutate(Victims_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_AGE_GROUP) %>%
  mutate(Perp_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_SEX) %>%
  mutate(Perp_by_Sex = n()) %>%
  ungroup() %>%
  group_by(PERP_RACE) %>%
  mutate(Perp_by_Race = n()) %>%
  ungroup() %>%
  group_by(PRECINCT) %>%
  mutate(Incidents_by_Prec = n()) %>%
  ungroup() %>%
  group_by(STATISTICAL_MURDER_FLAG) %>%
  mutate(Incidents_by_Stat = n())

Bronx_AGG <- Bronx_Totals %>%
  select(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
         Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
         VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
```

```
         PERP_SEX, Perp_by_Sex) %>%
  group_by(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
           Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
           VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
           PERP_SEX, Perp_by_Sex) %>%
  arrange(BORO, PRECINCT, VIC_RACE, VIC_AGE_GROUP,
          VIC_SEX, PERP_RACE, PERP_AGE_GROUP, PERP_SEX, STATISTICAL_MURDER_FLAG)

Brooklyn_Totals <- Brooklyn %>%
  select(-c(OCCUR_DATE, OCCUR_TIME, JURISDICTION_CODE, LOCATION_DESC)) %>%
  group_by(BORO) %>%
  mutate(Incidents_by_Boro = n()) %>%
  ungroup() %>%
  group_by(VIC_RACE) %>%
  mutate(Victims_by_Race = n()) %>%
  ungroup() %>%
  select(c(BORO, everything())) %>%
  group_by(VIC_SEX) %>%
  mutate(Victims_by_Sex = n()) %>%
  ungroup() %>%
  group_by(VIC_AGE_GROUP) %>%
  mutate(Victims_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_AGE_GROUP) %>%
  mutate(Perp_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_SEX) %>%
  mutate(Perp_by_Sex = n()) %>%
  ungroup() %>%
  group_by(PERP_RACE) %>%
  mutate(Perp_by_Race = n()) %>%
  ungroup() %>%
  group_by(PRECINCT) %>%
  mutate(Incidents_by_Prec = n()) %>%
  ungroup() %>%
  group_by(STATISTICAL_MURDER_FLAG) %>%
  mutate(Incidents_by_Stat = n())

Brooklyn_AGG <- Brooklyn_Totals %>%
  select(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
         Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
         VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
         PERP_SEX, Perp_by_Sex) %>%
  group_by(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
           Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
           VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
           PERP_SEX, Perp_by_Sex) %>%
  arrange(BORO, PRECINCT, VIC_RACE, VIC_AGE_GROUP,
          VIC_SEX, PERP_RACE, PERP_AGE_GROUP, PERP_SEX, STATISTICAL_MURDER_FLAG)

Manhattan_Totals <- Manhattan %>%
  select(-c(OCCUR_DATE, OCCUR_TIME, JURISDICTION_CODE, LOCATION_DESC)) %>%
  group_by(BORO) %>%
```

```r
  mutate(Incidents_by_Boro = n()) %>%
  ungroup() %>%
  group_by(VIC_RACE) %>%
  mutate(Victims_by_Race = n()) %>%
  ungroup() %>%
  select(c(BORO, everything())) %>%
  group_by(VIC_SEX) %>%
  mutate(Victims_by_Sex = n()) %>%
  ungroup() %>%
  group_by(VIC_AGE_GROUP) %>%
  mutate(Victims_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_AGE_GROUP) %>%
  mutate(Perp_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_SEX) %>%
  mutate(Perp_by_Sex = n()) %>%
  ungroup() %>%
  group_by(PERP_RACE) %>%
  mutate(Perp_by_Race = n()) %>%
  ungroup() %>%
  group_by(PRECINCT) %>%
  mutate(Incidents_by_Prec = n()) %>%
  ungroup() %>%
  group_by(STATISTICAL_MURDER_FLAG) %>%
  mutate(Incidents_by_Stat = n())

Manhattan_AGG <- Manhattan_Totals %>%
  select(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
         Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
         VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
         PERP_SEX, Perp_by_Sex) %>%
  group_by(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
           Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
           VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
           PERP_SEX, Perp_by_Sex) %>%
  arrange(BORO, PRECINCT, VIC_RACE, VIC_AGE_GROUP,
          VIC_SEX, PERP_RACE, PERP_AGE_GROUP, PERP_SEX, STATISTICAL_MURDER_FLAG)

Queens_Totals <- Queens %>%
  select(-c(OCCUR_DATE, OCCUR_TIME, JURISDICTION_CODE, LOCATION_DESC)) %>%
  group_by(BORO) %>%
  mutate(Incidents_by_Boro = n()) %>%
  ungroup() %>%
  group_by(VIC_RACE) %>%
  mutate(Victims_by_Race = n()) %>%
  ungroup() %>%
  select(c(BORO, everything())) %>%
  group_by(VIC_SEX) %>%
  mutate(Victims_by_Sex = n()) %>%
  ungroup() %>%
  group_by(VIC_AGE_GROUP) %>%
  mutate(Victims_by_Age = n()) %>%
```

```
  ungroup() %>%
  group_by(PERP_AGE_GROUP) %>%
  mutate(Perp_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_SEX) %>%
  mutate(Perp_by_Sex = n()) %>%
  ungroup() %>%
  group_by(PERP_RACE) %>%
  mutate(Perp_by_Race = n()) %>%
  ungroup() %>%
  group_by(PRECINCT) %>%
  mutate(Incidents_by_Prec = n()) %>%
  ungroup() %>%
  group_by(STATISTICAL_MURDER_FLAG) %>%
  mutate(Incidents_by_Stat = n())

Queens_AGG <- Queens_Totals %>%
  select(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
         Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
         VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
         PERP_SEX, Perp_by_Sex) %>%
  group_by(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
           Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
           VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
           PERP_SEX, Perp_by_Sex) %>%
  arrange(BORO, PRECINCT, VIC_RACE, VIC_AGE_GROUP,
          VIC_SEX, PERP_RACE, PERP_AGE_GROUP, PERP_SEX, STATISTICAL_MURDER_FLAG)

Staten_Island_Totals <- Staten_Island %>%
  select(-c(OCCUR_DATE, OCCUR_TIME, JURISDICTION_CODE, LOCATION_DESC)) %>%
  group_by(BORO) %>%
  mutate(Incidents_by_Boro = n()) %>%
  ungroup() %>%
  group_by(VIC_RACE) %>%
  mutate(Victims_by_Race = n()) %>%
  ungroup() %>%
  select(c(BORO, everything())) %>%
  group_by(VIC_SEX) %>%
  mutate(Victims_by_Sex = n()) %>%
  ungroup() %>%
  group_by(VIC_AGE_GROUP) %>%
  mutate(Victims_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_AGE_GROUP) %>%
  mutate(Perp_by_Age = n()) %>%
  ungroup() %>%
  group_by(PERP_SEX) %>%
  mutate(Perp_by_Sex = n()) %>%
  ungroup() %>%
  group_by(PERP_RACE) %>%
  mutate(Perp_by_Race = n()) %>%
  ungroup() %>%
  group_by(PRECINCT) %>%
```

```r
  mutate(Incidents_by_Prec = n()) %>%
  ungroup() %>%
  group_by(STATISTICAL_MURDER_FLAG) %>%
  mutate(Incidents_by_Stat = n())

Staten_Island_AGG <- Staten_Island_Totals %>%
  select(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
         Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
         VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
         PERP_SEX, Perp_by_Sex) %>%
  group_by(BORO, Incidents_by_Boro, PRECINCT, Incidents_by_Prec, STATISTICAL_MURDER_FLAG,
           Incidents_by_Stat,  VIC_RACE, Victims_by_Race, VIC_AGE_GROUP, Victims_by_Age,
           VIC_SEX,  Victims_by_Sex, PERP_RACE, Perp_by_Race, PERP_AGE_GROUP, Perp_by_Age,
           PERP_SEX, Perp_by_Sex) %>%
  arrange(BORO, PRECINCT, VIC_RACE, VIC_AGE_GROUP,
          VIC_SEX, PERP_RACE, PERP_AGE_GROUP, PERP_SEX, STATISTICAL_MURDER_FLAG)
```

Analyzing Across the Boroughs

Here we will take a look at some of the trends we might be able to find in each borough and then end it with a look at the city as a whole.

```r
summary(NY_TOT_Agg)
```

```
##      BORO            Incidents_by_Boro    PRECINCT      Incidents_by_Prec
##  Length:27312       Min.   :  776     Min.   :  1.00   Min.   :   1.0
##  Class :character   1st Qu.: 4094     1st Qu.: 44.00   1st Qu.: 459.0
##  Mode  :character   Median : 7937     Median : 68.00   Median : 758.0
##                     Mean   : 7786     Mean   : 65.64   Mean   : 713.3
##                     3rd Qu.:10933     3rd Qu.: 81.00   3rd Qu.: 953.0
##                     Max.   :10933     Max.   :123.00   Max.   :1557.0
##  STATISTICAL_MURDER_FLAG Incidents_by_Stat   VIC_RACE         Victims_by_Race
##  Length:27312            Min.   : 5266     Length:27312       Min.   :   10
##  Class :character        1st Qu.:22046     Class :character   1st Qu.: 4049
##  Mode  :character        Median :22046     Mode  :character   Median :19439
##                          Mean   :18811                        Mean   :14716
##                          3rd Qu.:22046                        3rd Qu.:19439
##                          Max.   :22046                        Max.   :19439
##  VIC_AGE_GROUP      Victims_by_Age    VIC_SEX          Victims_by_Sex
##  Length:27312       Min.   :    1   Length:27312       Min.   :   11
##  Class :character   1st Qu.:10086   Class :character   1st Qu.:24686
##  Mode  :character   Median :10086   Mode  :character   Median :24686
##                     Mean   : 9670                      Mean   :22563
##                     3rd Qu.:12281                      3rd Qu.:24686
##                     Max.   :12281                      Max.   :24686
##   PERP_RACE          Perp_by_Race   PERP_AGE_GROUP      Perp_by_Age
##  Length:27312       Min.   :    2   Length:27312       Min.   :    1
##  Class :character   1st Qu.:11432   Class :character   1st Qu.: 5687
##  Mode  :character   Median :11432   Mode  :character   Median : 6222
##                     Mean   :10139                      Mean   : 9022
##                     3rd Qu.:11786                      3rd Qu.:13132
##                     Max.   :11786                      Max.   :13132
```

```
##    PERP_SEX          Perp_by_Sex
##  Length:27312       Min.   :  424
##  Class :character   1st Qu.:11449
##  Mode  :character   Median :15439
##                     Mean   :13533
##                     3rd Qu.:15439
##                     Max.   :15439
```

```
max_incidents <-NY_TOT_Agg %>%
  pull(Incidents_by_Boro) %>%
  max()

min_incidents <- NY_TOT_Agg %>%
  pull(Incidents_by_Boro) %>%
  min()

max_incidents_boro <- NY_TOT_Agg %>%
  filter(Incidents_by_Boro == max_incidents) %>%
  slice_head(n = 1) %>%
  ungroup() %>%
  select(BORO) %>%
  distinct(BORO, .keep_all = FALSE)



min_incidents_boro <- NY_TOT_Agg %>%
  filter(Incidents_by_Boro == min_incidents) %>%
  slice_head(n = 1) %>%
  ungroup() %>%
  select(BORO) %>%
  distinct(BORO, .keep_all = FALSE)


Min_Max_boro_combined <- bind_rows(
  min_incidents_boro %>% mutate(Type = "Min_Incidents_Boro"),
  max_incidents_boro %>% mutate(Type = "Max_Incidents_Boro")
)


Min_Max_incidents_combined <- bind_rows(
  data.frame(Incidents = min_incidents, Type = "Min_Incidents"),
  data.frame(Incidents = max_incidents, Type = "Max_Incidents")
)


Min_Max_Incidents <- bind_cols(Min_Max_incidents_combined, Min_Max_boro_combined)
```

```
## New names:
## * `Type` -> `Type...2`
## * `Type` -> `Type...4`
```

```
med_incidents <-NY_TOT_Agg %>%
  pull(Incidents_by_Boro) %>%
```

```r
  median()

med_incidents_boro <- NY_TOT_Agg %>%
  filter(Incidents_by_Boro == med_incidents) %>%
  slice_head(n = 1) %>%
  ungroup() %>%
  select(BORO) %>%
  distinct(BORO, .keep_all = FALSE)

boro_combined <- bind_rows(
  min_incidents_boro %>% mutate(Type = "Min_Incidents_Boro"),
  max_incidents_boro %>% mutate(Type = "Max_Incidents_Boro"),
  med_incidents_boro %>% mutate(Type = "Med_Incidents_Boro")
)

incidents_combined <- bind_rows(
  data.frame(Incidents = min_incidents, Type = "Min_Incidents"),
  data.frame(Incidents = max_incidents, Type = "Max_Incidents"),
  data.frame(Incidents = med_incidents, Type = "Med_Incidents")
)

AGG_Incidents <- bind_cols(incidents_combined, boro_combined)
```

```
## New names:
## * `Type` -> `Type...2`
## * `Type` -> `Type...4`
```

```r
incidents_total <- nrow(NY_TOT_Agg)

Bronx_In <- data.frame(Boro = "Bronx", Incidents = nrow(Bronx))

Brooklyn_In <- data.frame(Boro = "Brooklyn", Incidents = nrow(Brooklyn))

Manhattan_In <- data.frame(Boro = "Manhattan", Incidents = nrow(Manhattan))

Queens_In <- data.frame(Boro = "Queens", Incidents = nrow(Queens))

Staten_Island_In <- data.frame(Boro = "Staten Island", Incidents = nrow(Staten_Island))

NYC_Incidents_Tot <- bind_rows(Bronx_In, Brooklyn_In, Manhattan_In, Queens_In, Staten_Island_In)

NYC_AVG_Incidents <- sum(NYC_Incidents_Tot$Incidents) / nrow(NYC_Incidents_Tot)

closest_mean <- min(abs(NY_TOT_Agg$Incidents_by_Boro - NYC_AVG_Incidents))

closest_mean_boro <- NY_TOT_Agg %>%
  filter(abs(Incidents_by_Boro - NYC_AVG_Incidents) == closest_mean) %>%
  slice_head(n = 1) %>%
  ungroup() %>%
  select(BORO) %>%
  distinct(BORO, .keep_all = FALSE)

NYC_age_VIC <- NYPD_Data %>%
```

```r
  group_by(BORO, VIC_AGE_GROUP) %>%
  count() %>%
  ungroup()

names(NYC_age_VIC)[3] <- "Count"

NYC_age_VIC_filtered <- NYC_age_VIC %>%
  filter(VIC_AGE_GROUP != 1022)

average_incidents_per_age_group <- NYC_age_VIC_filtered %>%
  group_by(VIC_AGE_GROUP) %>%
  summarise(avg_incidents = mean(Count))

average_incidents_Age <- data.frame(
  VIC_AGE_GROUP = average_incidents_per_age_group$VIC_AGE_GROUP,
  avg_incidents = average_incidents_per_age_group$avg_incidents)
```

At the end of the analysis, I have decided to look into the ages of the victims as well as the incidents in each borough. This is to help get an understanding of the possible trends that the ages of the victims might give us and because, I feel this is the attribute in the data that least requires any further understanding of the population and demographics of the city.

Visualizing the data

In this section, our goal is to use what we have found to create some visuals to help us look at our analysis of the shooting across the boroughs and how they stack up against the city as a whole.

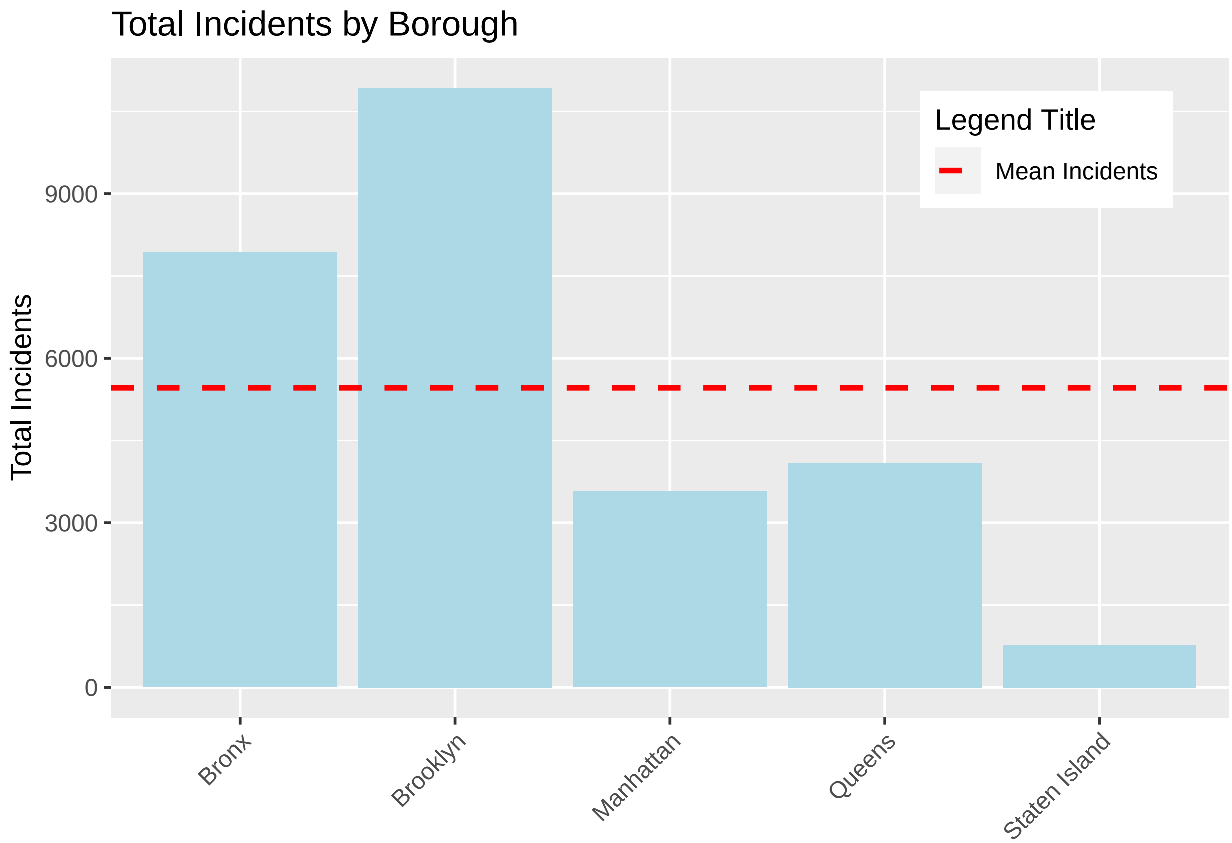
```r
library(ggplot2)

incidents_total <- nrow(NY_TOT_Agg)

Incidents_Boro <- ggplot(NYC_Incidents_Tot, aes(x = Boro, y = Incidents)) +
  geom_bar(stat = "identity", fill = "lightblue") +
  geom_hline(aes(yintercept = NYC_AVG_Incidents,linetype = "Mean Incidents"),
             color = "red", size = 1) +
  labs(title = "Total Incidents by Borough",
       x = "Boro",
       y = "Total Incidents",
       linetype = "Mean Incidents") +
       scale_linetype_manual(name = "Legend Title", values = "dashed", labels = "Mean Incidents") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        axis.title.x = element_blank(),
        legend.position = c(0.95, 0.95),
        legend.justification = c(1, 1))
```

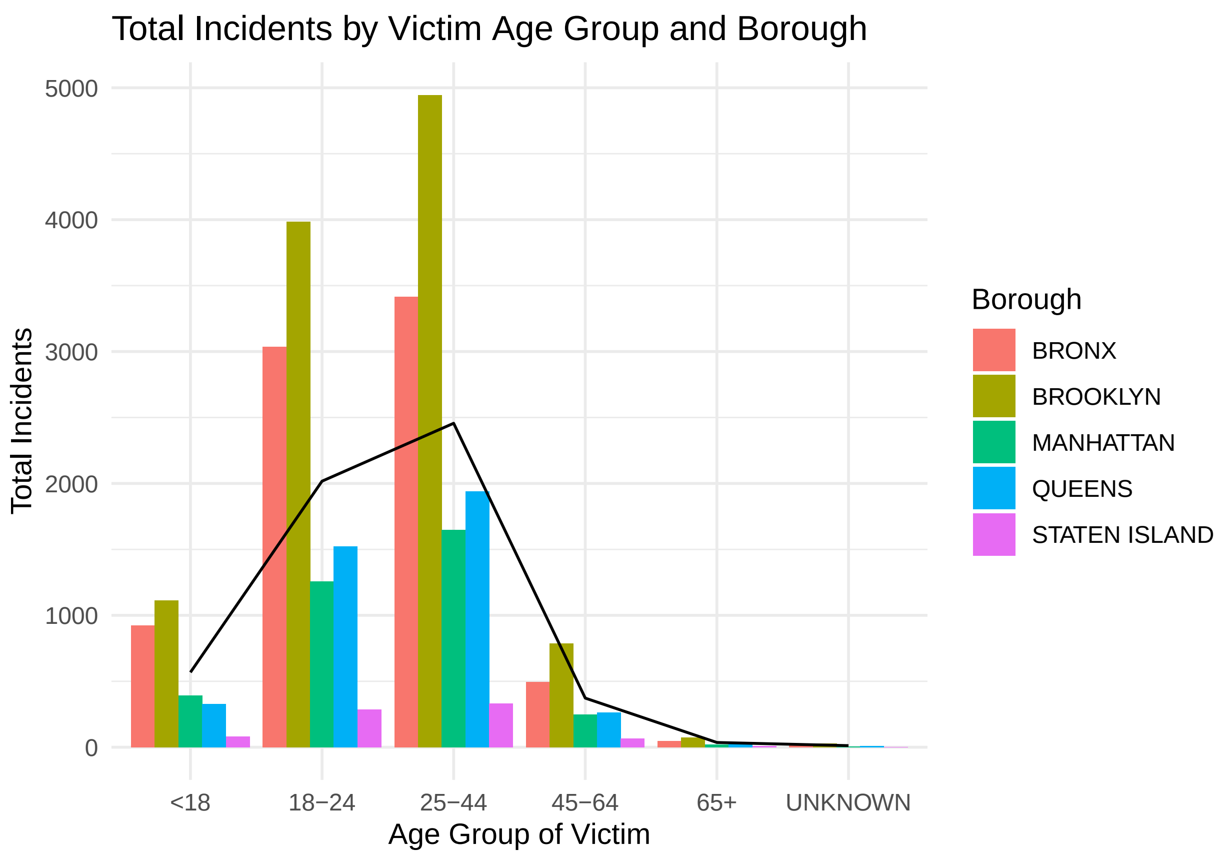
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
Incidents_Boro
```




```
Incidents_Age_Boro <- ggplot() +
  geom_bar(data = NYC_age_VIC_filtered, aes(x = VIC_AGE_GROUP, y = Count, fill = BORO),
           stat = "identity", position = "dodge") +
  geom_line(data = average_incidents_per_age_group, aes(x = VIC_AGE_GROUP, y = avg_incidents, group = 1)
            color = "black", size = 0.5) +
  labs(title = "Total Incidents by Victim Age Group and Borough",
       x = "Age Group of Victim",
       y = "Total Incidents",
       fill = "Borough") +
  theme_minimal()

Incidents_Age_Boro
```



With this data we can see that there is a trend across the boroughs showing that shootings tend to occur the most with younger victims, building up to the 25-44 age range with Brooklyn and the Bronx having much higher numbers of incidents than the other boroughs.

We can use this to create a model and see what relationships might exist between the age of the victim and where they live.

Modeling the data

For the model for our data, I decided to use a "Mixed Effect" model as there are repeated incidents centering on the 25-44 age range.

I wanted to both show the total incidents we are seeing while showing the general trend which exists for each borough:

That the older one is, the less likely they are to be a victim of a shooting in New York City.

```
library(lme4)
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
mixed_model <- lmer(Count ~ VIC_AGE_GROUP + (1 | BORO), data = NYC_age_VIC_filtered)


summary(mixed_model)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Count ~ VIC_AGE_GROUP + (1 | BORO)
##    Data: NYC_age_VIC_filtered
##
## REML criterion at convergence: 402.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.97639 -0.41185  0.06858  0.30752  2.32179
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  BORO     (Intercept) 342664   585.4
##  Residual             588321   767.0
## Number of obs: 30, groups:  BORO, 5
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)             567.8      431.5   1.316
## VIC_AGE_GROUP18-24     1449.4      485.1   2.988
## VIC_AGE_GROUP25-44     1888.4      485.1   3.893
## VIC_AGE_GROUP45-64     -195.2      485.1  -0.402
## VIC_AGE_GROUP65+       -531.6      485.1  -1.096
## VIC_AGE_GROUPUNKNOWN   -555.6      485.1  -1.145
##
## Correlation of Fixed Effects:
##               (Intr) VIC_AGE_GROUP1 VIC_AGE_GROUP2 VIC_AGE_GROUP4
## VIC_AGE_GROUP1 -0.562
## VIC_AGE_GROUP2 -0.562  0.500
## VIC_AGE_GROUP4 -0.562  0.500          0.500
## VIC_AGE_GROUP6 -0.562  0.500          0.500          0.500
## VIC_AGE_GROUPU -0.562  0.500          0.500          0.500
##               VIC_AGE_GROUP6
## VIC_AGE_GROUP1
## VIC_AGE_GROUP2
## VIC_AGE_GROUP4
## VIC_AGE_GROUP6
## VIC_AGE_GROUPU  0.500
```

Visualizing the Model

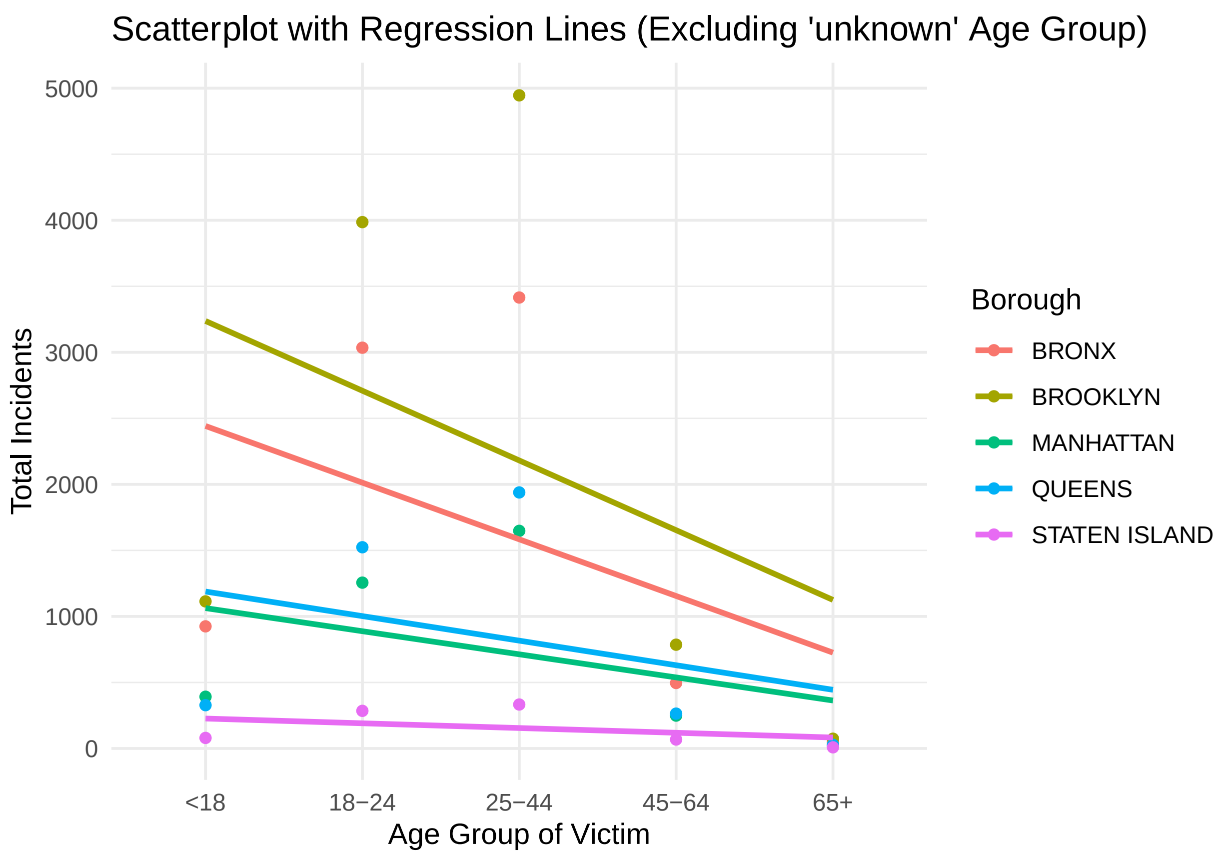
The chosen visualization is a scatterplot with regression lines. The plots themselves give us the ability to see the total number of victims per borough.

The regression lines show us the trend of shooting victims being more likely to be of a younger age group.

```r
NYC_age_VIC_filtered <- NYC_age_VIC_filtered %>%
  filter(VIC_AGE_GROUP != "UNKNOWN")

ggplot(NYC_age_VIC_filtered, aes(x = VIC_AGE_GROUP, y = Count, color = BORO)) +
  geom_point() +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE, aes(group = BORO, color = BORO)) +
  labs(title = "Scatterplot with Regression Lines (Excluding 'unknown' Age Group)",
       x = "Age Group of Victim",
       y = "Total Incidents",
       color = "Borough") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```



Closing notes

A few pieces of consideration come to mind which may help with future analysis. A breakdown of the population of the boroughs might help with verifying if the trend found is proportional to the percentage of the population across the city and in each borough.

An example of how this data additional data might change how we can interpret this trend, would be if we found that the percentage of the population which is 65+ is small enough that the few shooting victims that are of that age group represents a much larger proportion compared to other age groups. Then, we would see that the likelihood of being shot as you age would not actually decrease significantly and might actually increase.

It is this lack of population data which made me initially hesitant to find trends by race. Without knowing the demographics of the city or the individual boroughs, any data that I would find could not be truly justified as. For example, if I had solid demographic data which shows that say, Brooklyn is a neighborhood of mostly individuals who identify as White, and the highest number of shooting victims were White, this would make sense as it would be proportionate to the population of the borough.

Another piece of the data I struggled to discount bias in, is that the age groups are not equal in size. the 18-24 range is the smallest age range while the 25-44 range is one of the two likely to be the largest. If the ranges were closer to ten year intervals starting with 18-27 the graph may show a less steep trend.