NYPD_Shooting_Incident_Data_Report

2024-02-24

Introduction

This report is to provide a data analysis of NYPD Shooting Incidents. The data was collected by the city of New York.

Data Import and Cleaning

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

```
## Rows: 27312 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## dbl   (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl   (1): STATISTICAL_MURDER_FLAG
## time  (1): OCCUR_TIME
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(nypd_data)
```

```
## # A tibble: 6 x 21
##   INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO     LOC_OF_OCCUR_DESC PRECINCT
##          <dbl> <chr>      <time>     <chr>    <chr>                <dbl>
## 1    228798151 05/27/2021 21:30      QUEENS   <NA>                   105
## 2    137471050 06/27/2014 17:40      BRONX    <NA>                    40
## 3    147998800 11/21/2015 03:56      QUEENS   <NA>                   108
## 4    146837977 10/09/2015 18:30      BRONX    <NA>                    44
## 5     58921844 02/19/2009 22:58      BRONX    <NA>                    47
## 6    219559682 10/21/2020 21:36      BROOKLYN <NA>                    81
## # i 15 more variables: JURISDICTION_CODE <dbl>, LOC_CLASSFCTN_DESC <chr>,
## #   LOCATION_DESC <chr>, STATISTICAL_MURDER_FLAG <lgl>, PERP_AGE_GROUP <chr>,
## #   PERP_SEX <chr>, PERP_RACE <chr>, VIC_AGE_GROUP <chr>, VIC_SEX <chr>,
## #   VIC_RACE <chr>, X_COORD_CD <dbl>, Y_COORD_CD <dbl>, Latitude <dbl>,
## #   Longitude <dbl>, Lon_Lat <chr>
```

```
summary(nypd_data)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME           BORO
##  Min.   :  9953245   Length:27312       Length:27312      Length:27312
##  1st Qu.: 63860880   Class :character   Class1:hms        Class :character
##  Median : 90372218   Mode  :character   Class2:difftime   Mode  :character
##  Mean   :120860536                      Mode  :numeric
##  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       Mode :logical           Length:27312
##  Class :character   FALSE:22046             Class :character
##  Mode  :character   TRUE :5266              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.:1000028   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
```

Cleaning Data to remove any unneeded columns

```
# Removing Columns that are not needed for this analysis
ny_shootings <- select(nypd_data,
                       -INCIDENT_KEY,
                       -PRECINCT,
                       -LOCATION_DESC,
                       -LOC_OF_OCCUR_DESC,
                       -JURISDICTION_CODE,
                       -LOC_CLASSFCTN_DESC,
                       -X_COORD_CD,
                       -Y_COORD_CD,
                       -Latitude,
                       -Longitude,
                       -Lon_Lat,
                       )
```

I removed most of the data that was not necessary like the coordinates, Lat, and Long also other codes/numbers that were not needed.

Cleaning Age Group Data

Lots of unknown numbers and mixture of null and unknown. Moved them all into the unknown column.

```
ny_shootings <- ny_shootings %>%
    mutate(PERP_AGE_GROUP = ifelse(PERP_AGE_GROUP %in% c("UNKNOWN","unknown", NA, NULL, "(null)", "940"

ny_shootings <- ny_shootings %>%
    mutate(VIC_AGE_GROUP = ifelse(VIC_AGE_GROUP %in% c("UNKNOWN","unknown", "1022"), "unknown", VIC_AGE_
```

Data Analysis

Summary Statistics

```
summary(ny_shootings)
```

```
##   OCCUR_DATE          OCCUR_TIME           BORO
##  Length:27312       Length:27312      Length:27312
##  Class :character   Class1:hms        Class :character
##  Mode  :character   Class2:difftime   Mode  :character
##                     Mode  :numeric
##  STATISTICAL_MURDER_FLAG PERP_AGE_GROUP       PERP_SEX
##  Mode :logical           Length:27312       Length:27312
##  FALSE:22046             Class :character   Class :character
##  TRUE :5266              Mode  :character   Mode  :character
##
##   PERP_RACE          VIC_AGE_GROUP        VIC_SEX            VIC_RACE
##  Length:27312       Length:27312       Length:27312       Length:27312
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
```

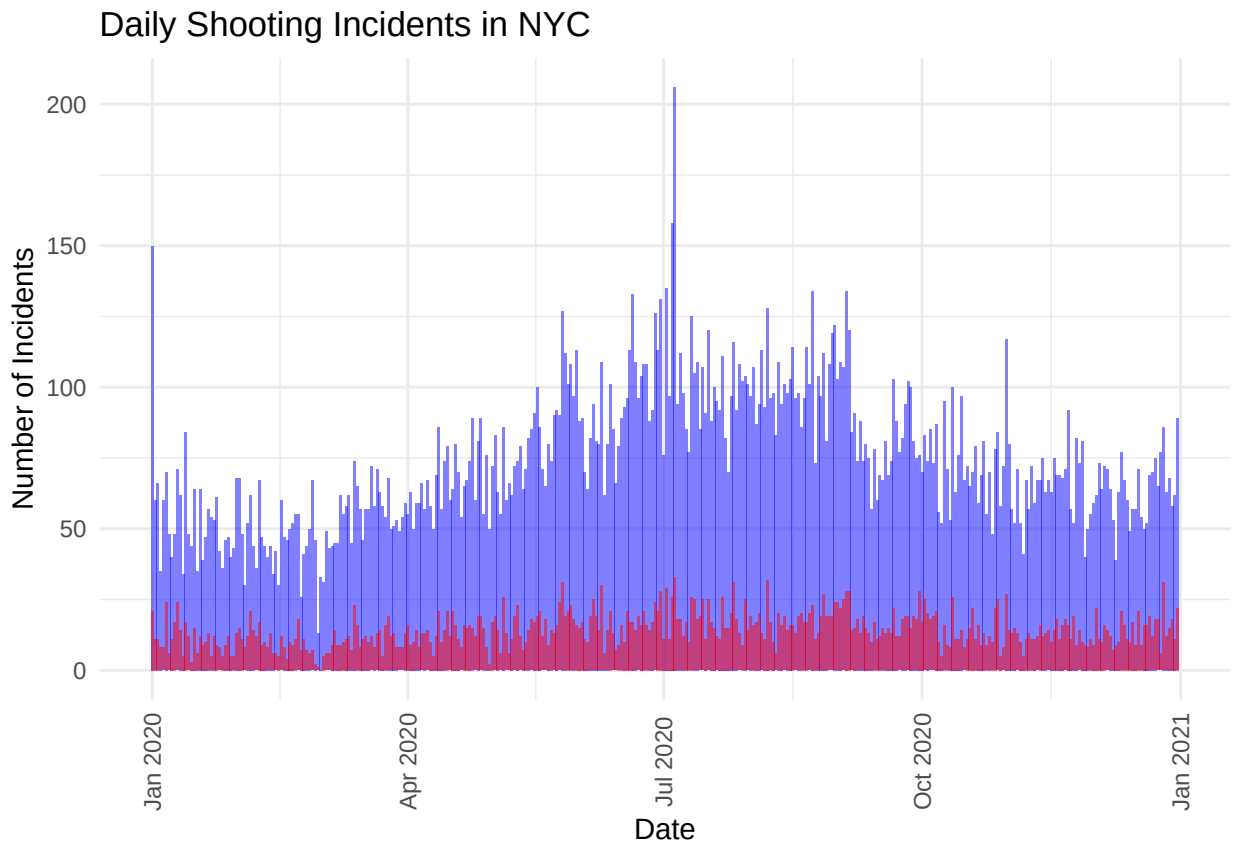
Data Visualization

Now creating a simple visualization to see the distribution of data.

```
# Ensure 'OCCUR_DATE' is in Date format
Dates <- as.Date(ny_shootings$OCCUR_DATE, format = "%m/%d/%y")
ny_shootings$OCCUR_DATE <- Dates

# Summarize the data to get the count of events for each date
daily_counts <- ny_shootings %>%
  group_by(OCCUR_DATE) %>%
  summarise(count = n(),
            deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE", na.rm = TRUE))
```

Layer Bar Graph

```
ggplot(daily_counts, aes(x = OCCUR_DATE)) +
  geom_col(aes(y = count), fill = "blue", alpha = 0.5) +  # Layer for total count
  geom_col(aes(y = deaths), fill = "red", alpha = 0.5) +  # Layer for deaths
  theme_minimal() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90, vjust = 0.5)) +
  labs(title = "Daily Shooting Incidents in NYC", x = "Date", y = "Number of Incidents")
```

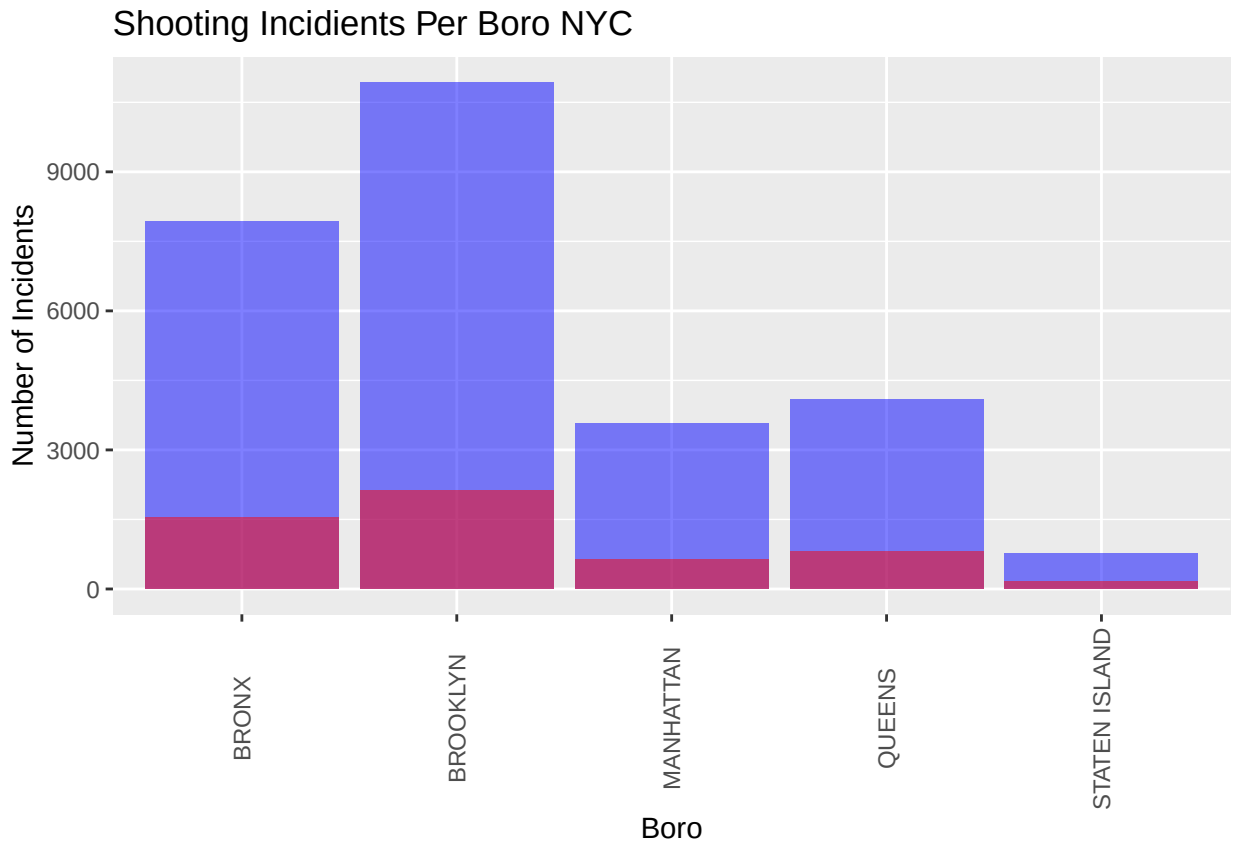


Finding Demographic

Location of the incidents along with the proper age range. I already cleaned the age range there was lots of unknown.

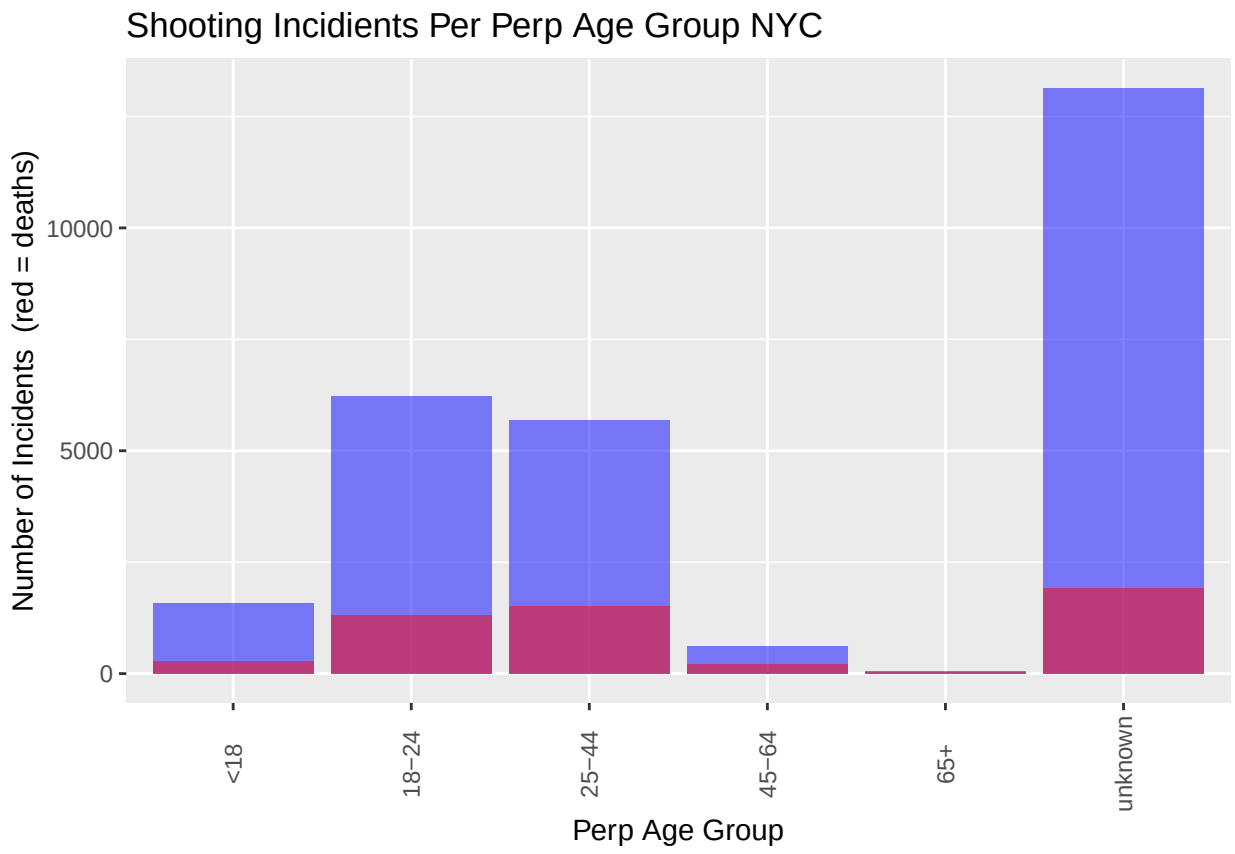
```
# Summarize the data to get the count of events for each Boro
boro_counts <- ny_shootings %>%
  group_by(BORO) %>%
  summarise(count = n(),
            deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE", na.rm = TRUE))
```

```
ggplot(boro_counts, aes(x = BORO, y = count)) +
    geom_col(aes(y = count), fill = "blue", alpha = 0.5) +
    geom_col(aes(y = deaths), fill = "red", alpha = 0.5) +
    theme(legend.position = "bottom",
          axis.text.x = element_text(angle = 90, vjust = 0.5))+
    labs(title = "Shooting Incidients Per Boro NYC", x = "Boro", y = "Number of Incidents")
```



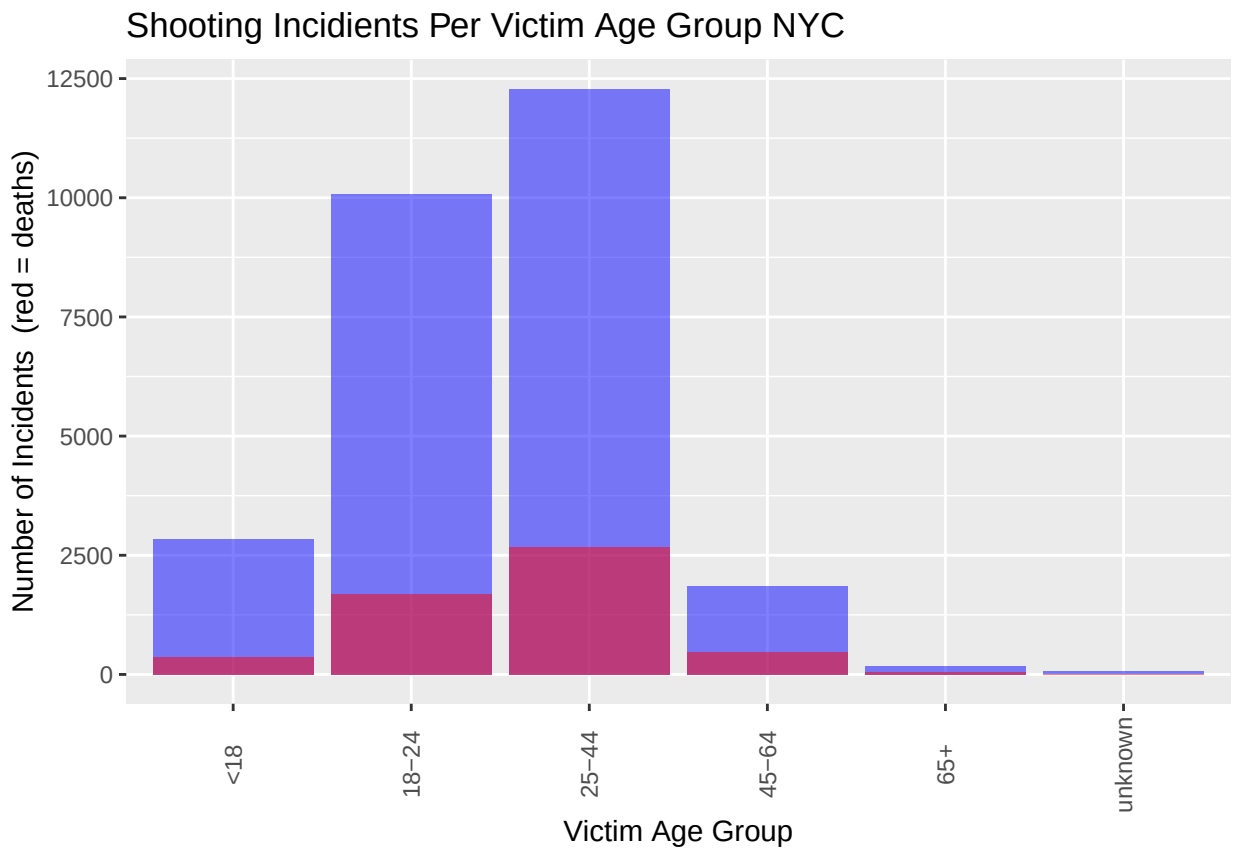

```
# Summarize the data to get the count of events for each Age Group
perp_age_group_counts <- ny_shootings %>%
  group_by(PERP_AGE_GROUP) %>%
  summarise(count = n(),
            deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE", na.rm = TRUE))
```

```
ggplot(perp_age_group_counts, aes(x = PERP_AGE_GROUP, y = count)) +
    geom_col(aes(y = count), fill = "blue", alpha = 0.5) +
    geom_col(aes(y = deaths), fill = "red", alpha = 0.5) +
    theme(legend.position = "bottom",
          axis.text.x = element_text(angle = 90, vjust = 0.5))+
    labs(title = "Shooting Incidients Per Perp Age Group NYC", x = "Perp Age Group", y = "Number of Inci
```




```
# Summarize the data to get the count of events for each Age Group
vic_age_group_counts <- ny_shootings %>%
  group_by(VIC_AGE_GROUP) %>%
  summarise(count = n(),
            deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE", na.rm = FALSE))
```

```
ggplot(vic_age_group_counts, aes(x = VIC_AGE_GROUP, y = count)) +
    geom_col(aes(y = count), fill = "blue", alpha = 0.5) +
    geom_col(aes(y = deaths), fill = "red", alpha = 0.5) +
    theme(legend.position = "bottom",
          axis.text.x = element_text(angle = 90, vjust = 0.5))+
    labs(title = "Shooting Incidients Per Victim Age Group NYC", x = "Victim Age Group", y = "Number of
```



Model

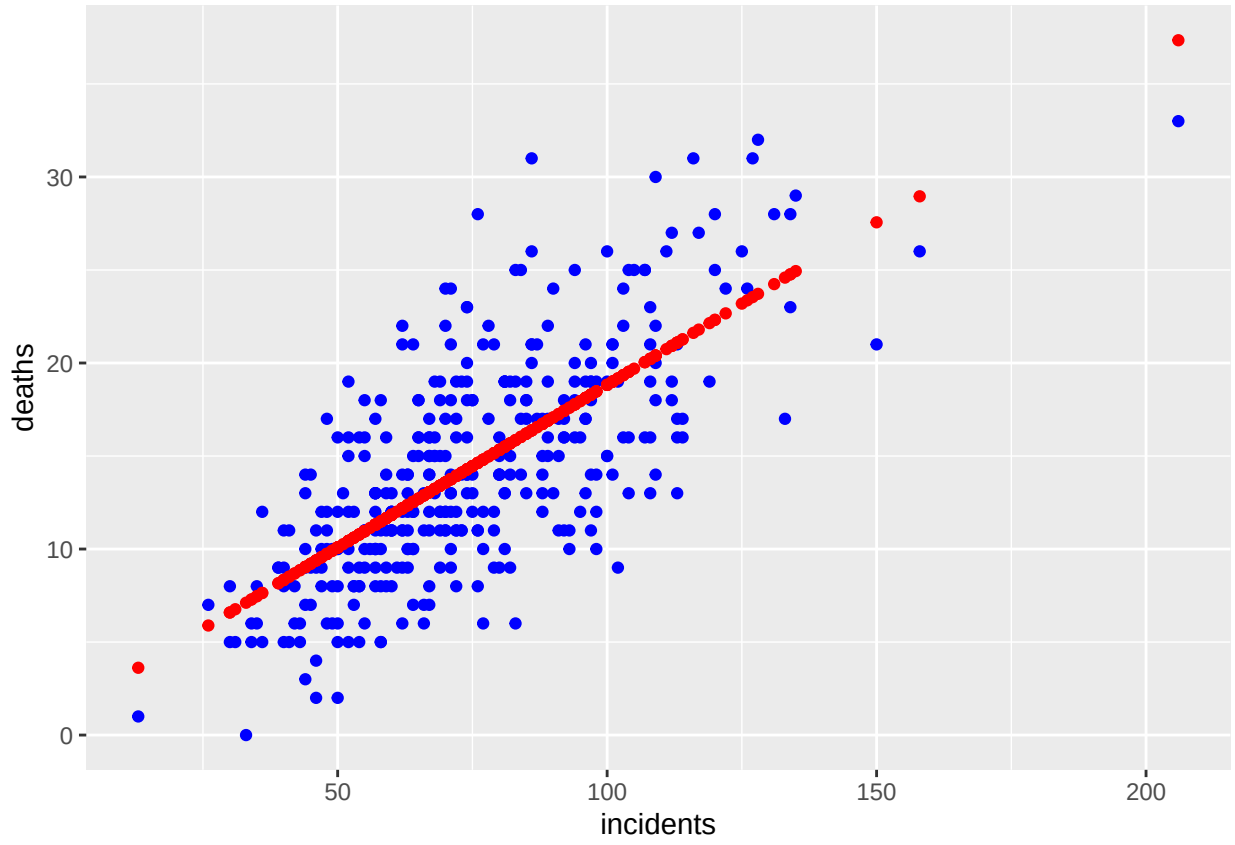
Creating the Model using date of incident, incident resulting in deaths, and incidents.

```
# Aggregate data by date to get daily counts of incidents and deaths
daily_summary <- ny_shootings %>%
  group_by(OCCUR_DATE) %>%
  summarise(incidents = n(),
            deaths = sum(STATISTICAL_MURDER_FLAG == "TRUE", na.rm = TRUE))

# Fit a GLM (e.g., Poisson regression for count data)
mod <- glm(deaths ~ incidents, data = daily_summary)
summary(mod)
```

```
##
## Call:
## glm(formula = deaths ~ incidents, data = daily_summary)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.34610    0.70431   1.911   0.0568 .
## incidents    0.17477    0.00897  19.484   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for gaussian family taken to be 17.5667)
##
##     Null deviance: 13062.9  on 365  degrees of freedom
## Residual deviance:  6394.3  on 364  degrees of freedom
## AIC: 2091.6
##
## Number of Fisher Scoring iterations: 2
```

```
daily_pred <- daily_summary %>% mutate(pred = predict(mod))

daily_pred %>% ggplot() +
    geom_point(aes(x = incidents, y = deaths), color = "blue")+
    geom_point(aes(x = incidents, y = pred), color = "red")
```



Analysis

Bias:

can come from the reports and also people not reporting any incidents. I do not find myself having much of a bias other than i do not like gang violence.

Daily Counts of incidents

This shows a very high correlation with more incidents in the warmer seasons of the year and a drop off in the colder seasons of the year. This is to expected with more activity outside and mixing with other people during the warm seasons and less activity in the colder seasons.

Boros

Based on the information provided by the State of New York we can see some distinct trends. Two of the five Boros make up the majority of the shooting incidents (Bronx and Brooklyn) with the next Boro Queens coming in third with half the amount of shootings as Bronx and nearly a 1/3 of the incidents of Brooklyn. It's hard to justify exactly why that is base on the limited information and only having the shooting reports. Also to keep in mind these are only the reported incidents and there should be some acknowledgment for those incidents that did not go reported. From this information alone one could guess this would have to do with either gang activity or lack of police in these areas compared to other Boros of NYC.

Age Range

Taking a look at the Perpetrator and Victim age groupings is very interesting these can tell a good amount of stories on their own. From what you can see i combined null NA and unknown into one unknown column. Having such a significant Perpetrator column as "Unknown" most likely means they were never caught. These Perpetrators not being caught could also support the suspicion that there is lower policing in these areas leading to more Perpetrators going uncaught. This could also be related to gang activity as well if they go uncaught and continue to commit shooting incidents which would lead to more incidents in specific areas compared to others. The age groups also so a very large number of people in the military fighting age which is to be expected with gang activity mixed with the facts of having a firearm in NYC being illegal for most also would correlate these with gang activity as well.