NYPD_Shooting_Incident_Data_Report

Israel Johnson

11/21/2021

Objective

This incident data report will provide specific analysis and results, based upon the following dataset: 'NYPD Shooting Incident Data (Historic)'. To get started, we will need to install the following R libraries for further use:

1.) tidyverse 2.) sessioninfo

Once installed, the following code will load key packages from the required library that will be used for this analysis.

```
library(tidyr)
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)
```

Importing Data (Reproducibility)

We will need to import the required dataset from the following source: https://catalog.data.gov/dataset. Once you are there, proceed with the following steps:

1.) Search for the dataset titled 'NYPD Shooting Incident Data (Historic)' 2.) Right click on the CSV button next to the title, and copy the link address into your respective RMD document, as shown in the code below.

The following code completes this process and reads in the data in a CSV format.

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

Tidying and Transforming the Data

After reading in the required NYPD Shooting Incident Data, tidy and transform the data into a desired format for the analysis. For the first aspect of this analysis, the main focus will be on answering the following question:

"How many shootings occurred within each borough and how many of those shootings resulted in murder?"

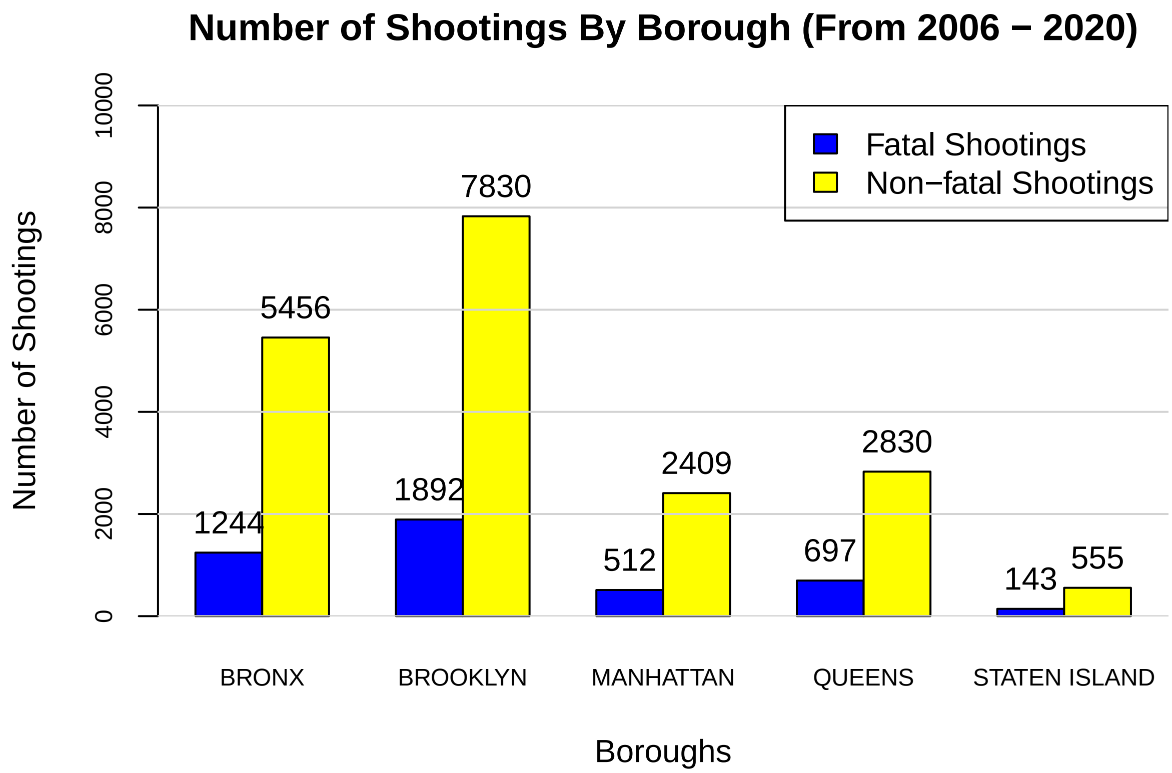
The first data visualization will focus on the answer to this question. The following code handles the initial tidying and transformation of the data.

```
incident_keys <- ny_data[1]
occurrence_dates <- ny_data[2]
occurrence_times <- ny_data[3]
boroughs <- ny_data[4]
precincts <- ny_data[5]
juris_codes <- ny_data[6]
loc_descript <- ny_data[7]
is_murder <- ny_data[8]
perp_age <- ny_data[9]
perp_sex <- ny_data[10]
perp_race <- ny_data[11]
victim_age_group <- ny_data[12]
victim_sex <- ny_data[13]
victim_race <- ny_data[14]
latitude <- ny_data[17]
longitude <- ny_data[18]
coordinates <- ny_data[19]

df <- data.frame(occurrence_dates, occurrence_times, juris_codes, precincts, boroughs, loc_descript, is_

murders <- filter(df, is_murder=='true')
not_murders <- filter(df, is_murder=='false')

murders_by_borough <- table(t(murders$BORO))
not_murders_by_borough <- table(t(not_murders$BORO))
combined <- rbind(murders_by_borough, not_murders_by_borough)
```

Data Visualization 01 : Number of Shootings Per Borough (From 2006 - 2020)

The following visualization presents a bar graph and distribution of the number of shootings that occurred in each New York borough from 2006 to 2020. In addition to this data, the visualization also shows the distinction of which of those murders in each borough resulted in murder during that timeframe.

```
bp <- barplot(combined, main="Number of Shootings By Borough (From 2006 - 2020)",
              beside = T, col=c("blue", "yellow"), width = 0.2,
              ylim = c(0,10000), cex.axis = 0.75, cex.names = 0.75,
              xlab = "Boroughs", ylab = "Number of Shootings")
grid(nx = NA, ny = NULL, lwd = 1, lty = 1, col = "lightgray")
text(bp, combined + 0.5, pos = 3, labels = combined)
legend("topright", legend = c("Fatal Shootings", "Non-fatal Shootings"),
       fill = c("blue","yellow"))
```



Data Visualization 02 : Percentage of Shootings in Brooklyn By Age Group (From 2006 - 2020)

The second visualization will require further transformations of the data. Based on the results of the first visualization, the conclusion can be made that the New York borough with the highest number of shootings from 2006 to 2020 appears to be Brooklyn. In this case, two more questions come to mind: "Why does Brooklyn have the highest number of shootings within this timeframe? Can information on the different age groups involved provide more insight on the situation in Brooklyn?". These questions form the basis of the second visualization.

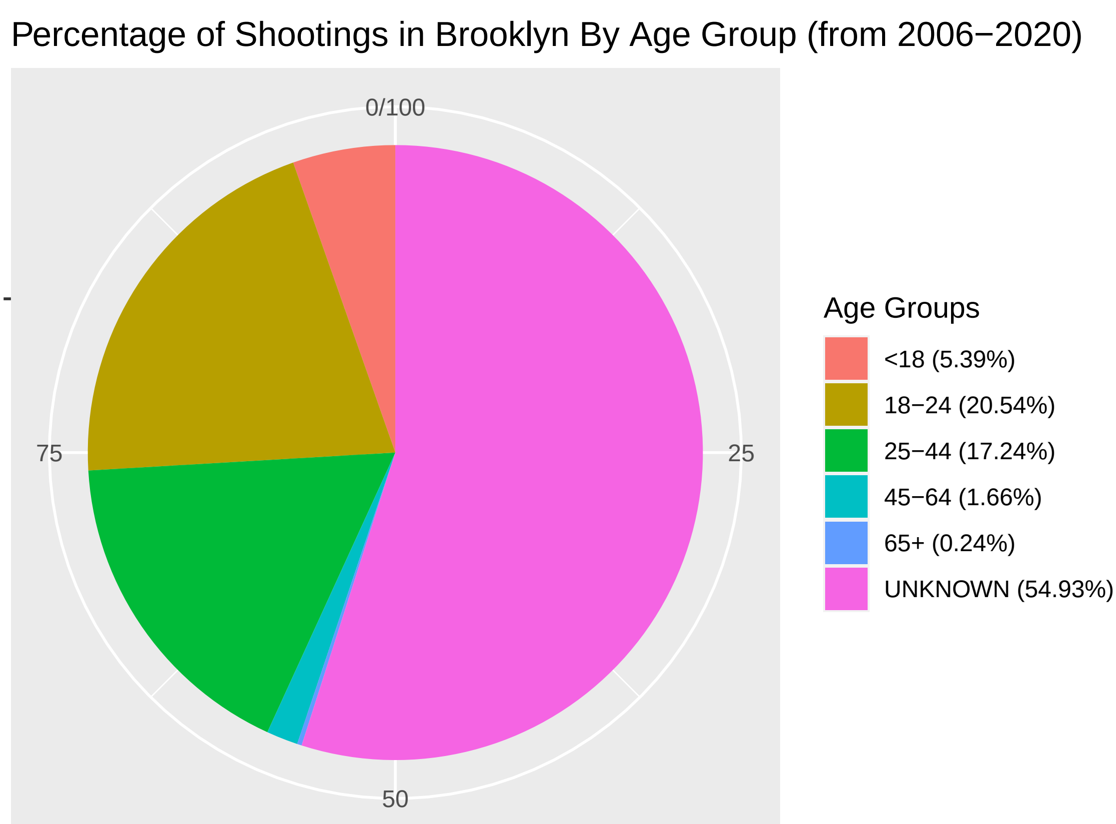
Using knowledge from the first visualization results, the next visualization focuses on how the number of shootings in Brooklyn during this time frame take place between different age groups.

One discovery that was made when further transforming the data included the fact that many shooting incidents had a missing entry or no recorded entry for the 'Age Group' attribute within the provided dataset. To handle the missing age group entries for certain observations, I assigned theses entries to the 'UNKNOWN' category, which provided more clarity on how the missing data could impact further visualization and analysis.

Further transformations are shown in the following code as well as in the second visualization, which provides further details on Brooklyn shootings among the different categories of age groups. Upon conclusion, I made another discovery that was most concerning: more than half (specifically 54.93%) of the total number of Brooklyn shooting entries were shown to have an 'UNKNOWN' age group assigned to them, indicating that more than half of the Brooklyn entries did not have a recorded age group from the original dataset. Many factors can attribute to this missing data, including: short time duration to identify suspect during incident, clothing of suspect obscuring further identifiable traits, local law enforcement not having the technology nor means to gather specific details, etc.

More data will be required to find out the reason why Brooklyn had the highest number of shootings out of the five New York boroughs. This could be a result of policy issues, lack of police presence in the borough, lack of medical technology for nearby medical facilities (to perform emergency medical procedures), or other factors.

```
new_df <- tibble(boroughs, perp_age)
new_df$PERP_AGE_GROUP <- sub("^$", "UNKNOWN", new_df$PERP_AGE_GROUP)

brooklyn_shootings <- filter(new_df, boroughs=='BROOKLYN')

m <- table(brooklyn_shootings)

shootings_per_age <- c(m[1], m[2], m[3], m[4], m[5], m[7])
age_perc <- round(shootings_per_age/sum(shootings_per_age)*100, 2)
age_groups <- c("<18 (5.39%)","18-24 (20.54%)","25-44 (17.24%)","45-64 (1.66%)",
                "65+ (0.24%)", "UNKNOWN (54.93%)")
lbls <- paste(age_perc, "%", sep="")

brooklyn_data <- data.frame(shootings_per_age, age_perc, age_groups)

ggplot(brooklyn_data, aes(x="", y=age_perc, fill=age_groups))+
  geom_bar(stat="identity", width=1)+
  theme(axis.line = element_blank())+
  labs(fill="Age Groups",x=NULL, y= NULL,
       title="Percentage of Shootings in Brooklyn By Age Group (from 2006-2020)")+
  coord_polar("y", start=0)
```



Data Model and Output

Since the data reveals that Brooklyn had the highest number of shootings during this time period, I will use this information as well as a different aspect of the Brooklyn data to form the data model. This will be a linear regression model where Brooklyn precincts will serve a function of Brooklyn dates. The independent variables will be the dates to indicate the change in time and the dependent variables will be the precincts for areas where a shooting occurred during the given period of time. Once both variables are obtained, use them as well as the NYPD Shooting Incident data for Brooklyn to generate the model.

According to the model summary output, the residual standard error was found to be fairly high, which indicates less accuracy in the model. Predicted shooting occurrences do match up with some actual occurrences from the Brooklyn area, but not with all of the actual occurrences. As a result, this model would not be the best fit when analyzing the input data. The main source of the error likely falls with the amount of data being utilized for the input data. Additional data or further transformation of the data could lead to a more accurate model. However, the initial approach and construction of the model appears to be correct, with the application of a regression model to handle the prediction of dependent variables from a set of independent variables (shootings from certain Brooklyn precincts over the selected time period).

```r
# CREATING THE MODEL
brooklyn_shootings <- filter(df, boroughs=='BROOKLYN')

brooklyn_precincts <- brooklyn_shootings$PRECINCT

brooklyn_dates <- brooklyn_shootings$OCCUR_DATE
brooklyn_dates <- format(as.Date(brooklyn_dates, "%m/%d/%Y"), "%m/%Y")


mod <- lm(brooklyn_precincts~brooklyn_dates, brooklyn_shootings)
df_w_pred <- mutate(brooklyn_shootings, pred = predict(mod))

summary(mod)
```

```
##
## Call:
## lm(formula = brooklyn_precincts ~ brooklyn_dates, data = brooklyn_shootings)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.5000  -4.5957   0.0179   4.3415  22.0357
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             75.14286    0.98369  76.388  < 2e-16 ***
## brooklyn_dates01/2007 -0.19946    1.36465  -0.146  0.88380
## brooklyn_dates01/2008  0.87987    1.43013   0.615  0.53841
## brooklyn_dates01/2009 -0.75649    1.43013  -0.529  0.59684
## brooklyn_dates01/2010 -0.06391    1.48843  -0.043  0.96575
## brooklyn_dates01/2011  1.76190    1.44796   1.217  0.22370
## brooklyn_dates01/2012 -0.30952    1.51153  -0.205  0.83775
## brooklyn_dates01/2013 -2.75155    1.41365  -1.946  0.05163 .
## brooklyn_dates01/2014 -1.38676    1.45744  -0.952  0.34137
## brooklyn_dates01/2015 -1.77076    1.43886  -1.231  0.21848
## brooklyn_dates01/2016 -1.20737    1.58024  -0.764  0.44486
## brooklyn_dates01/2017 -1.71429    1.79597  -0.955  0.33985
```

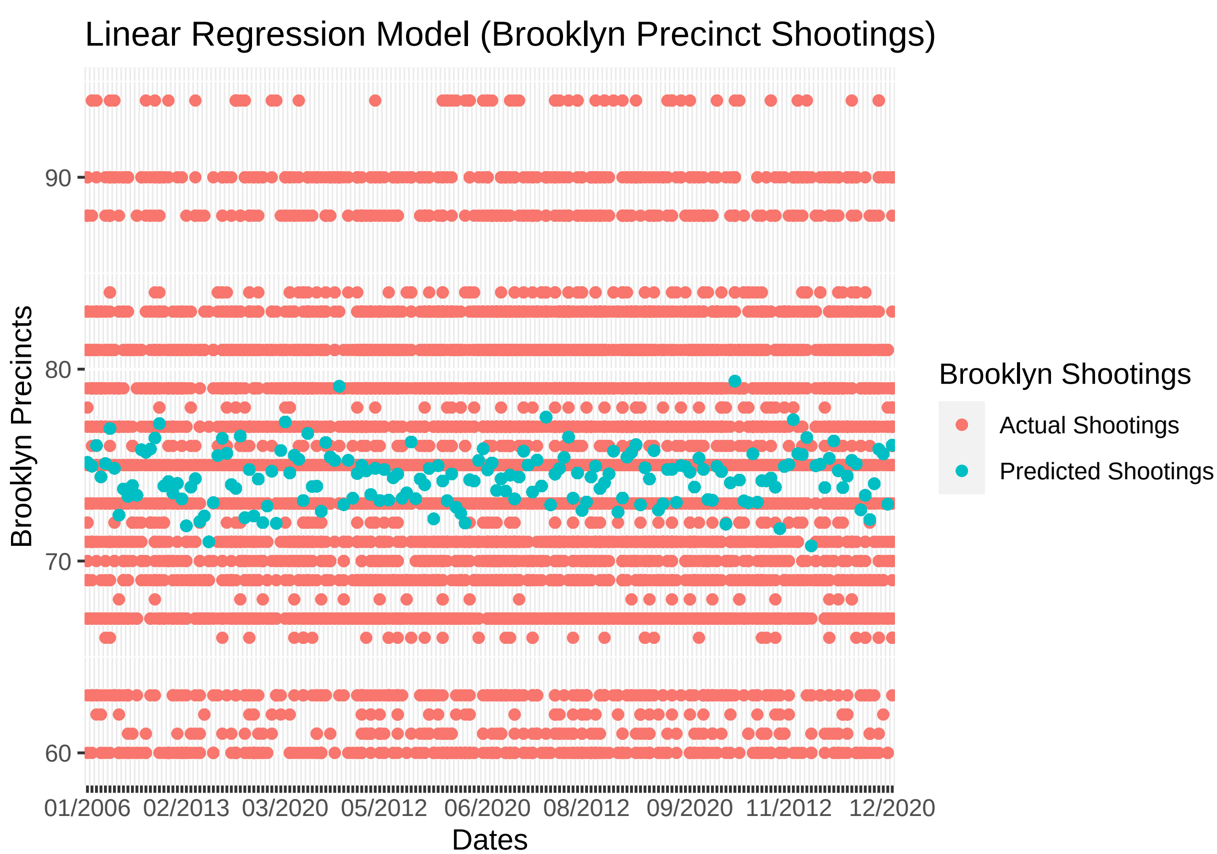
```
## brooklyn_dates01/2018  0.65714    2.03191   0.323  0.74639
## brooklyn_dates01/2019  0.52381    1.79597   0.292  0.77055
## brooklyn_dates01/2020  0.69714    1.69241   0.412  0.68041
## brooklyn_dates02/2006  1.26891    1.53695   0.826  0.40905
## brooklyn_dates02/2007  2.01993    1.43886   1.404  0.16040
## brooklyn_dates02/2008 -1.23377    1.55064  -0.796  0.42625
## brooklyn_dates02/2009 -1.00000    1.52393  -0.656  0.51171
## brooklyn_dates02/2010 -1.59740    1.55064  -1.030  0.30296
## brooklyn_dates02/2011 -1.09740    1.43013  -0.767  0.44290
## brooklyn_dates02/2012 -1.90602    1.48843  -1.281  0.20038
## brooklyn_dates02/2013 -3.30952    1.51153  -2.190  0.02858 *
## brooklyn_dates02/2014 -1.29286    1.82713  -0.708  0.47922
## brooklyn_dates02/2015 -0.85119    1.71560  -0.496  0.61980
## brooklyn_dates02/2016 -3.08225    1.55064  -1.988  0.04687 *
## brooklyn_dates02/2017 -2.79286    1.82713  -1.529  0.12641
## brooklyn_dates02/2018 -4.14286    2.08673  -1.985  0.04714 *
## brooklyn_dates02/2019 -2.09938    1.74045  -1.206  0.22776
## brooklyn_dates02/2020  0.35714    2.08673   0.171  0.86411
## brooklyn_dates03/2006  1.25298    1.39838   0.896  0.37026
## brooklyn_dates03/2007  0.46690    1.45744   0.320  0.74871
## brooklyn_dates03/2008 -1.16209    1.37094  -0.848  0.39665
## brooklyn_dates03/2009 -1.35597    1.32096  -1.027  0.30468
## brooklyn_dates03/2010  1.36825    1.42173   0.962  0.33588
## brooklyn_dates03/2011 -2.88704    1.43886  -2.006  0.04483 *
## brooklyn_dates03/2012 -0.37815    1.37745  -0.275  0.78368
## brooklyn_dates03/2013 -2.79402    1.43886  -1.942  0.05219 .
## brooklyn_dates03/2014 -0.87202    1.39838  -0.624  0.53291
## brooklyn_dates03/2015 -3.14286    1.65038  -1.904  0.05690 .
## brooklyn_dates03/2016 -2.27922    1.76717  -1.290  0.19717
## brooklyn_dates03/2017 -0.46104    1.76717  -0.261  0.79418
## brooklyn_dates03/2018 -3.17857    1.63127  -1.949  0.05138 .
## brooklyn_dates03/2019  0.61576    1.61327   0.382  0.70270
## brooklyn_dates03/2020  2.10714    1.63127   1.292  0.19649
## brooklyn_dates04/2006 -0.55026    1.35857  -0.405  0.68546
## brooklyn_dates04/2007  0.36778    1.40587   0.262  0.79363
## brooklyn_dates04/2008  0.15903    1.36465   0.117  0.90723
## brooklyn_dates04/2009 -1.98901    1.37094  -1.451  0.14686
## brooklyn_dates04/2010  1.50831    1.23247   1.224  0.22106
## brooklyn_dates04/2011 -1.26286    1.26485  -0.998  0.31810
## brooklyn_dates04/2012 -1.24490    1.39115  -0.895  0.37088
## brooklyn_dates04/2013 -2.55102    1.39115  -1.834  0.06672 .
## brooklyn_dates04/2014  1.03663    1.47764   0.702  0.48298
## brooklyn_dates04/2015  0.27714    1.38418   0.200  0.84131
## brooklyn_dates04/2016  0.09524    1.79597   0.053  0.95771
## brooklyn_dates04/2017  3.96429    1.63127   2.430  0.01511 *
## brooklyn_dates04/2018 -2.20737    1.58024  -1.397  0.16249
## brooklyn_dates04/2019  0.10714    1.82713   0.059  0.95324
## brooklyn_dates04/2020 -1.86700    1.61327  -1.157  0.24719
## brooklyn_dates05/2006 -0.58036    1.24914  -0.465  0.64222
## brooklyn_dates05/2007 -0.15385    1.22012  -0.126  0.89966
## brooklyn_dates05/2008 -0.44423    1.27168  -0.349  0.72685
## brooklyn_dates05/2009 -1.67774    1.23247  -1.361  0.17346
## brooklyn_dates05/2010 -0.31235    1.33090  -0.235  0.81445
## brooklyn_dates05/2011 -1.98623    1.24053  -1.601  0.10939
```

```
## brooklyn_dates05/2012 -0.37143    1.28258  -0.290  0.77213
## brooklyn_dates05/2013 -1.96639    1.53695  -1.279  0.20078
## brooklyn_dates05/2014 -0.80415    1.31621  -0.611  0.54124
## brooklyn_dates05/2015 -0.61121    1.25214  -0.488  0.62547
## brooklyn_dates05/2016 -1.86786    1.46732  -1.273  0.20306
## brooklyn_dates05/2017 -1.59113    1.61327  -0.986  0.32402
## brooklyn_dates05/2018  1.06227    1.47764   0.719  0.47222
## brooklyn_dates05/2019 -1.89286    1.56505  -1.209  0.22652
## brooklyn_dates05/2020 -0.86626    1.40587  -0.616  0.53780
## brooklyn_dates06/2006 -1.16695    1.24053  -0.941  0.34689
## brooklyn_dates06/2007 -0.31786    1.24914  -0.254  0.79914
## brooklyn_dates06/2008 -2.94589    1.29848  -2.269  0.02331 *
## brooklyn_dates06/2009 -0.17184    1.28640  -0.134  0.89373
## brooklyn_dates06/2010 -0.96855    1.18434  -0.818  0.41349
## brooklyn_dates06/2011 -1.99770    1.31621  -1.518  0.12911
## brooklyn_dates06/2012 -0.60338    1.26156  -0.478  0.63246
## brooklyn_dates06/2013 -2.34286    1.30273  -1.798  0.07214 .
## brooklyn_dates06/2014 -2.65848    1.30710  -2.034  0.04199 *
## brooklyn_dates06/2015 -3.16558    1.43013  -2.213  0.02689 *
## brooklyn_dates06/2016 -0.91071    1.34698  -0.676  0.49898
## brooklyn_dates06/2017 -0.97619    1.39838  -0.698  0.48514
## brooklyn_dates06/2018  0.10039    1.49971   0.067  0.94663
## brooklyn_dates06/2019  0.70821    1.40587   0.504  0.61445
## brooklyn_dates06/2020 -0.39056    1.18434  -0.330  0.74158
## brooklyn_dates07/2006 -0.03061    1.20477  -0.025  0.97973
## brooklyn_dates07/2007 -1.46825    1.15929  -1.267  0.20536
## brooklyn_dates07/2008 -0.86081    1.17171  -0.735  0.46257
## brooklyn_dates07/2009 -1.49929    1.19879  -1.251  0.21109
## brooklyn_dates07/2010 -0.66713    1.19499  -0.558  0.57667
## brooklyn_dates07/2011 -1.90327    1.20895  -1.574  0.11545
## brooklyn_dates07/2012 -0.76355    1.22990  -0.621  0.53473
## brooklyn_dates07/2013  0.58128    1.33610   0.435  0.66353
## brooklyn_dates07/2014 -0.14286    1.32096  -0.108  0.91388
## brooklyn_dates07/2015 -1.54608    1.31621  -1.175  0.24017
## brooklyn_dates07/2016  0.12030    1.34145   0.090  0.92854
## brooklyn_dates07/2017 -1.23377    1.55064  -0.796  0.42625
## brooklyn_dates07/2018  2.35714    1.35857   1.735  0.08277 .
## brooklyn_dates07/2019 -2.21429    1.44796  -1.529  0.12624
## brooklyn_dates07/2020 -0.61558    1.12028  -0.549  0.58268
## brooklyn_dates08/2006 -0.32389    1.17320  -0.276  0.78250
## brooklyn_dates08/2007  0.25714    1.26485   0.203  0.83891
## brooklyn_dates08/2008  1.31497    1.24053   1.060  0.28917
## brooklyn_dates08/2009 -1.86917    1.21110  -1.543  0.12277
## brooklyn_dates08/2010 -0.55665    1.22990  -0.453  0.65085
## brooklyn_dates08/2011 -2.50456    1.21329  -2.064  0.03902 *
## brooklyn_dates08/2012 -2.06494    1.25835  -1.641  0.10083
## brooklyn_dates08/2013 -0.75930    1.27168  -0.597  0.55047
## brooklyn_dates08/2014 -0.20083    1.28640  -0.156  0.87594
## brooklyn_dates08/2015 -1.36905    1.23779  -1.106  0.26874
## brooklyn_dates08/2016 -1.06494    1.25835  -0.846  0.39741
## brooklyn_dates08/2017 -0.60000    1.52393  -0.394  0.69380
## brooklyn_dates08/2018  0.58128    1.61327   0.360  0.71862
## brooklyn_dates08/2019 -2.59286    1.46732  -1.767  0.07725 .
## brooklyn_dates08/2020 -1.86138    1.14842  -1.621  0.10509
```

```
## brooklyn_dates09/2006  0.27273    1.25835   0.217  0.82842   
## brooklyn_dates09/2007  0.50714    1.24914   0.406  0.68476   
## brooklyn_dates09/2008  0.92208    1.25835   0.733  0.46372   
## brooklyn_dates09/2009 -2.21042    1.26823  -1.743  0.08138 . 
## brooklyn_dates09/2010 -0.28571    1.23779  -0.231  0.81745   
## brooklyn_dates09/2011 -0.86667    1.19131  -0.727  0.46694   
## brooklyn_dates09/2012  0.60714    1.34698   0.451  0.65218   
## brooklyn_dates09/2013 -2.48768    1.33610  -1.862  0.06265 . 
## brooklyn_dates09/2014 -2.16071    1.34698  -1.604  0.10872   
## brooklyn_dates09/2015 -0.37013    1.43013  -0.259  0.79579   
## brooklyn_dates09/2016 -0.36508    1.31159  -0.278  0.78075   
## brooklyn_dates09/2017 -2.08730    1.51153  -1.381  0.16734   
## brooklyn_dates09/2018 -0.16988    1.49971  -0.113  0.90981   
## brooklyn_dates09/2019 -0.22857    1.52393  -0.150  0.88078   
## brooklyn_dates09/2020 -0.51190    1.23779  -0.414  0.67920   
## brooklyn_dates10/2006 -1.28802    1.31621  -0.979  0.32781   
## brooklyn_dates10/2007  0.21859    1.24053   0.176  0.86014   
## brooklyn_dates10/2008 -0.36508    1.31159  -0.278  0.78075   
## brooklyn_dates10/2009 -1.93531    1.36465  -1.418  0.15617   
## brooklyn_dates10/2010 -1.97619    1.32586  -1.490  0.13613   
## brooklyn_dates10/2011 -0.21182    1.22990  -0.172  0.86326   
## brooklyn_dates10/2012 -0.48432    1.45744  -0.332  0.73966   
## brooklyn_dates10/2013 -3.22350    1.31621  -2.449  0.01434 * 
## brooklyn_dates10/2014 -1.06739    1.36465  -0.782  0.43414   
## brooklyn_dates10/2015  4.23214    1.46732   2.884  0.00393 **
## brooklyn_dates10/2016 -0.92411    1.56505  -0.590  0.55489   
## brooklyn_dates10/2017 -2.00772    1.49971  -1.339  0.18069   
## brooklyn_dates10/2018 -2.10286    1.69241  -1.243  0.21407   
## brooklyn_dates10/2019  0.44538    1.53695   0.290  0.77199   
## brooklyn_dates10/2020 -2.07834    1.31621  -1.579  0.11436   
## brooklyn_dates11/2006 -0.95714    1.28258  -0.746  0.45553   
## brooklyn_dates11/2007 -0.96375    1.29435  -0.745  0.45654   
## brooklyn_dates11/2008 -0.81633    1.39115  -0.587  0.55735   
## brooklyn_dates11/2009 -1.29286    1.32586  -0.975  0.32953   
## brooklyn_dates11/2010 -3.45767    1.35857  -2.545  0.01094 * 
## brooklyn_dates11/2011 -0.21558    1.35268  -0.159  0.87338   
## brooklyn_dates11/2012 -0.11654    1.48843  -0.078  0.93759   
## brooklyn_dates11/2013  2.22381    1.59629   1.393  0.16362   
## brooklyn_dates11/2014  0.45289    1.40587   0.322  0.74735   
## brooklyn_dates11/2015  0.40476    1.44796   0.280  0.77984   
## brooklyn_dates11/2016  1.29464    1.56505   0.827  0.40813   
## brooklyn_dates11/2017 -4.35338    1.86096  -2.339  0.01934 * 
## brooklyn_dates11/2018 -0.17734    1.61327  -0.110  0.91247   
## brooklyn_dates11/2019 -0.09740    1.76717  -0.055  0.95605   
## brooklyn_dates11/2020 -1.31235    1.33090  -0.986  0.32413   
## brooklyn_dates12/2006  0.21157    1.25214   0.169  0.86582   
## brooklyn_dates12/2007  1.11246    1.40587   0.791  0.42879   
## brooklyn_dates12/2008 -0.43619    1.26485  -0.345  0.73021   
## brooklyn_dates12/2009 -1.31786    1.46732  -0.898  0.36913   
## brooklyn_dates12/2010 -0.71733    1.40587  -0.510  0.60990   
## brooklyn_dates12/2011  0.10243    1.36465   0.075  0.94017   
## brooklyn_dates12/2012 -0.08036    1.56505  -0.051  0.95905   
## brooklyn_dates12/2013 -2.47619    1.47764  -1.676  0.09382 . 
## brooklyn_dates12/2014 -1.71693    1.35857  -1.264  0.20634   
```

```
## brooklyn_dates12/2015 -2.98377    1.43013  -2.086  0.03697 *
## brooklyn_dates12/2016 -1.11429    1.52393  -0.731  0.46468
## brooklyn_dates12/2017  0.68473    1.61327   0.424  0.67126
## brooklyn_dates12/2018  0.44048    1.71560   0.257  0.79738
## brooklyn_dates12/2019 -2.17989    1.65038  -1.321  0.18659
## brooklyn_dates12/2020  0.89560    1.37094   0.653  0.51359
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.886 on 9542 degrees of freedom
## Multiple R-squared:  0.03217,    Adjusted R-squared:  0.01401
## F-statistic: 1.772 on 179 and 9542 DF,  p-value: 1.507e-09
```

```r
ggplot(df_w_pred)+
  geom_point(aes(x=brooklyn_dates, y=brooklyn_precincts,color="Actual Shootings"))+
  geom_point(aes(x=brooklyn_dates, y=pred,color="Predicted Shootings"))+
  scale_x_discrete(guide = guide_axis(check.overlap = TRUE))+
  guides(color=guide_legend(title = "Brooklyn Shootings"))+
  labs(title="Linear Regression Model (Brooklyn Precinct Shootings)",y="Brooklyn Precincts", x="Dates")
```



Identification of Bias and Conclusion (Communication of Results and Summary)

There is potential for different biases to occur here in the given data and analysis, but the most likely case for this process would be selection bias, especially in plotting the model and selection of input data. The given data model and related plot are based upon specific aspects of the NYPD data (namely data associated with one New York borough: Brooklyn). The inclusion of additional data from the NYPD dataset may have improved the output of the model and provided more accurate results. Another form of bias that could be found is measurement bias, which also relates back to the differences between the predicted occurrences and actual occurrences from the plot of the created data model.

Overall, the analysis has been able to reveal key aspects of information, based upon the NYPD dataset. The analysis (visualizations included) have been able to answer critical questions, including: which New York boroughs had the highest (and lowest) number of shootings within the given period of time as well as how many of those shootings resulted in murder? In the case study of Brooklyn shootings between the different age groups, I revealed not only which age groups were involved in the shootings, but also a significant amount of missing data for shooting incidents that do not have an associated age group. Further data in this regard can be beneficial for law enforcement and other users in determining other factors that are a cause to the high number of shootings in the Brooklyn borough, as well as for the shootings in other New York boroughs.

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_3.3.5 dplyr_1.0.7   tidyr_1.1.4
##
## loaded via a namespace (and not attached):
##  [1] pillar_1.6.4     compiler_4.1.2   highr_0.9        tools_4.1.2
##  [5] digest_0.6.28    evaluate_0.14    lifecycle_1.0.1  tibble_3.1.5
##  [9] gtable_0.3.0     pkgconfig_2.0.3  rlang_0.4.12     DBI_1.1.1
## [13] yaml_2.2.1       xfun_0.27        fastmap_1.1.0    withr_2.4.2
## [17] stringr_1.4.0    knitr_1.36       generics_0.1.1   vctrs_0.3.8
## [21] grid_4.1.2       tidyselect_1.1.1 glue_1.4.2       R6_2.5.1
## [25] fansi_0.5.0      rmarkdown_2.11   purrr_0.3.4      farver_2.1.0
## [29] magrittr_2.0.1   scales_1.1.1     ellipsis_0.3.2   htmltools_0.5.2
## [33] assertthat_0.2.1 colorspace_2.0-2 labeling_0.4.2   utf8_1.2.2
## [37] stringi_1.7.5    munsell_0.5.0    crayon_1.4.2
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.