

Impact of Community Centers on Shooting Incidents in New York City

It's often reported that community centers have an active effect in reducing crime. I was interested in performing analysis to show that connection. I have performed analysis on community centers in New York City and their relation to shooting incidents. I have also selected a specific community center that opened in November 2020 to investigate whether it's opening had an effect on shooting incidents in the surrounding area.

Import Libraries and Data

Import Libraries

Import the following libraries: *tidyverse*: Cleaning and plotting *lubridate*: Date transformation *geosphere*: Geographic distance transformation

Import and Clean Data

I have selected three primary data sources. NYPD Shooting data (provided in course materials), NYC facilities data that provides lists of public facilities in the New York City area, and New York City precinct population data.

NYPD Shooting Data: <https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic>

NYC Facilities Data: <https://catalog.data.gov/dataset/facilities-database>

NYC Precinct Population data: <https://johnkeefe.net/nyc-police-precinct-and-census-data>

```
nypd_data = read.csv('NYPD_Shooting_Incident_data__Historic_.csv')
nypd_data = nypd_data %>% mutate(OCCUR_DATE = mdy(OCCUR_DATE))
```

```
fac_data = read.csv('Facilities_Database.csv')
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## EOF within quoted string
```

```
pop_data = data.frame(c("BRONX", "BROOKLYN", "MANHATTAN", "QUEENS", "STATEN ISLAND"), c(1472654, 2736074, 1694000, 2405235, 468526))
colnames(pop_data) <- c("borough", "population")
```

```
precinct_pop_data = read.csv('nyc_precinct_2020pop.csv')
precinct_pop_data = precinct_pop_data %>% select(c("precinct", "P1_001N"))
colnames(precinct_pop_data) <- c("precinct", "population")
```

Data Exploration

Data exploration and preliminary analysis. Visualizations and processing in this section is used to understand the data and get an preliminary view of what the data looks like.

NYPD Shooting Data Exploration

Prepare shooting data for plotting:

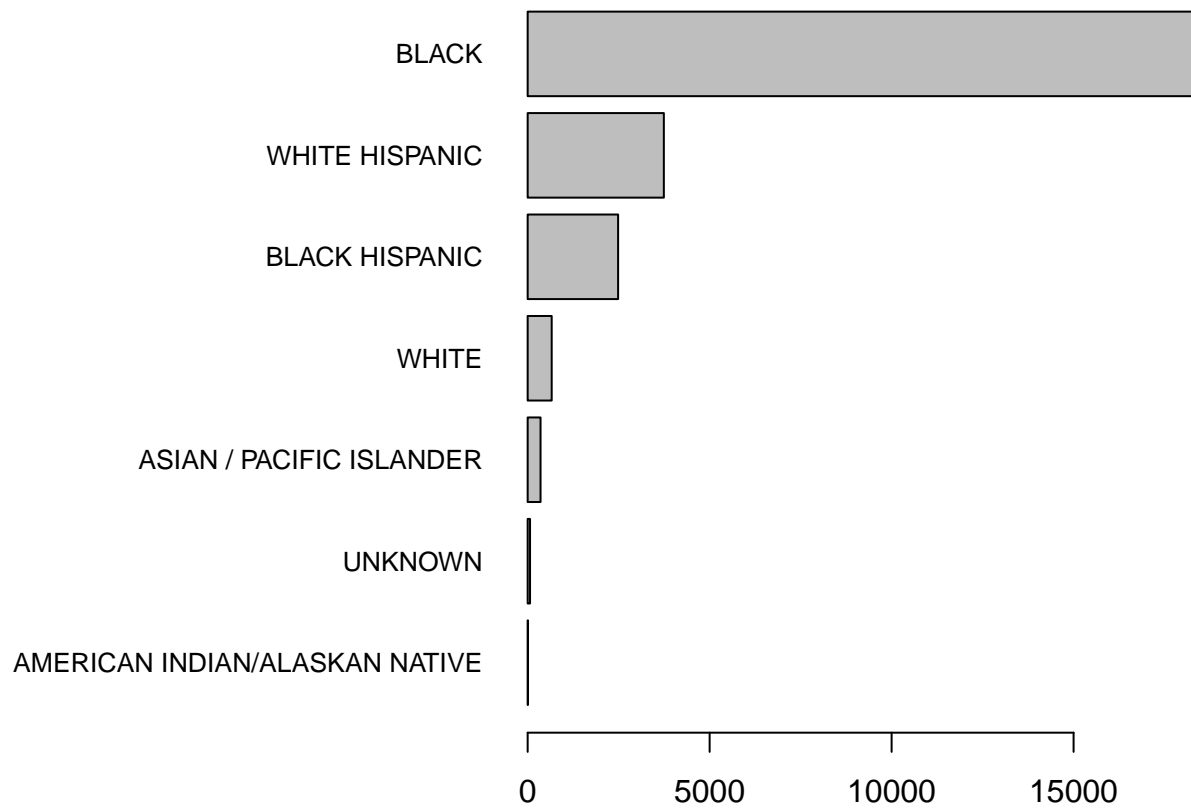
```
s_race_table = as.data.frame(table(nypd_data$VIC_RACE))
s_race_table = s_race_table[order(s_race_table$Freq),]

s_boro_table = as.data.frame(table(nypd_data$BORO))
colnames(s_boro_table) <- c("borough", "Freq")
s_boro_table = merge(s_boro_table, pop_data, by="borough")
s_boro_table = s_boro_table[order(s_boro_table$Freq),]
s_boro_table$per_thous = s_boro_table$Freq / (s_boro_table$population/1000)

s_precinct_table = as.data.frame(table(nypd_data$PRECINCT))
colnames(s_precinct_table) <- c("precinct", "Freq")
s_precinct_table = merge(s_precinct_table, precinct_pop_data, by="precinct")
s_precinct_table = s_precinct_table[order(s_precinct_table$Freq),]
s_precinct_table$per_thous = s_precinct_table$Freq / (s_precinct_table$population/1000)
```

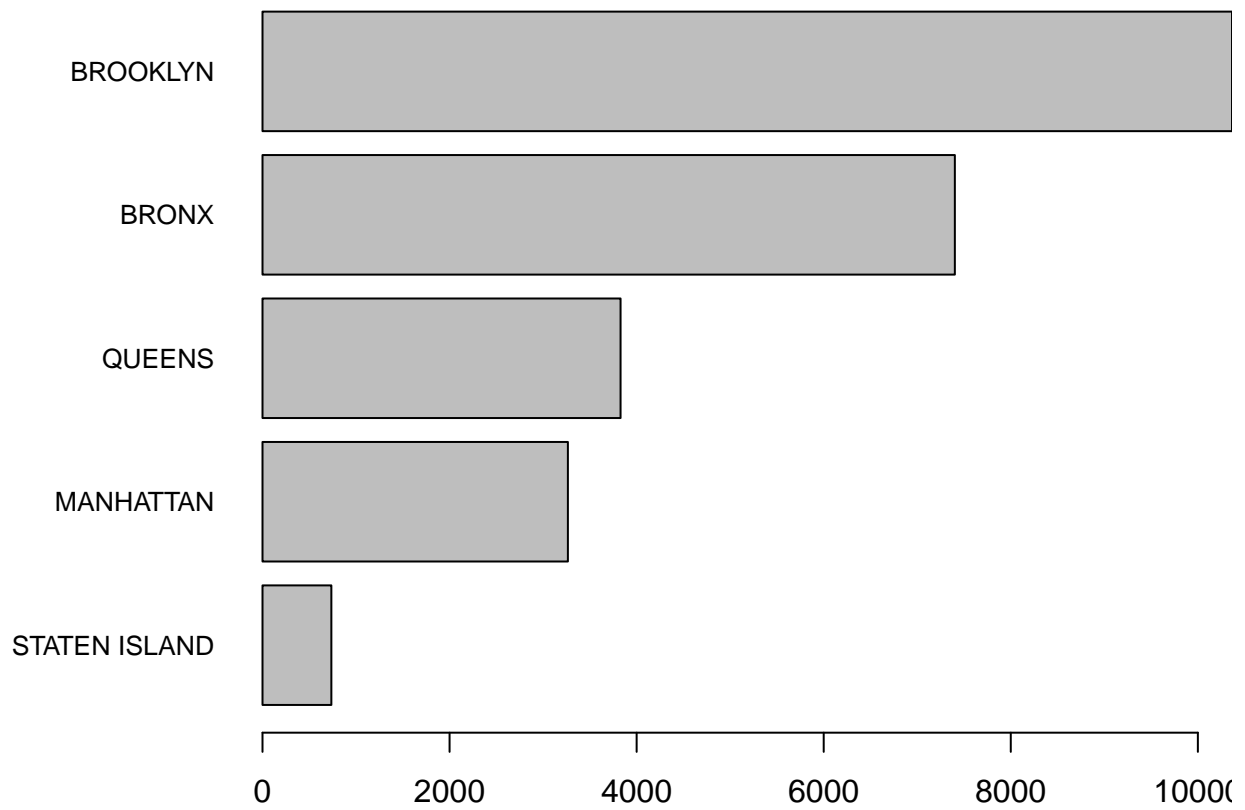
Shooting Incidents by Race

```
par(mar=c(3,15,0,0))  
barplot(names = s_race_table$Var1, height = s_race_table$Freq, horiz=TRUE, cex.names=0.8, las=1)
```



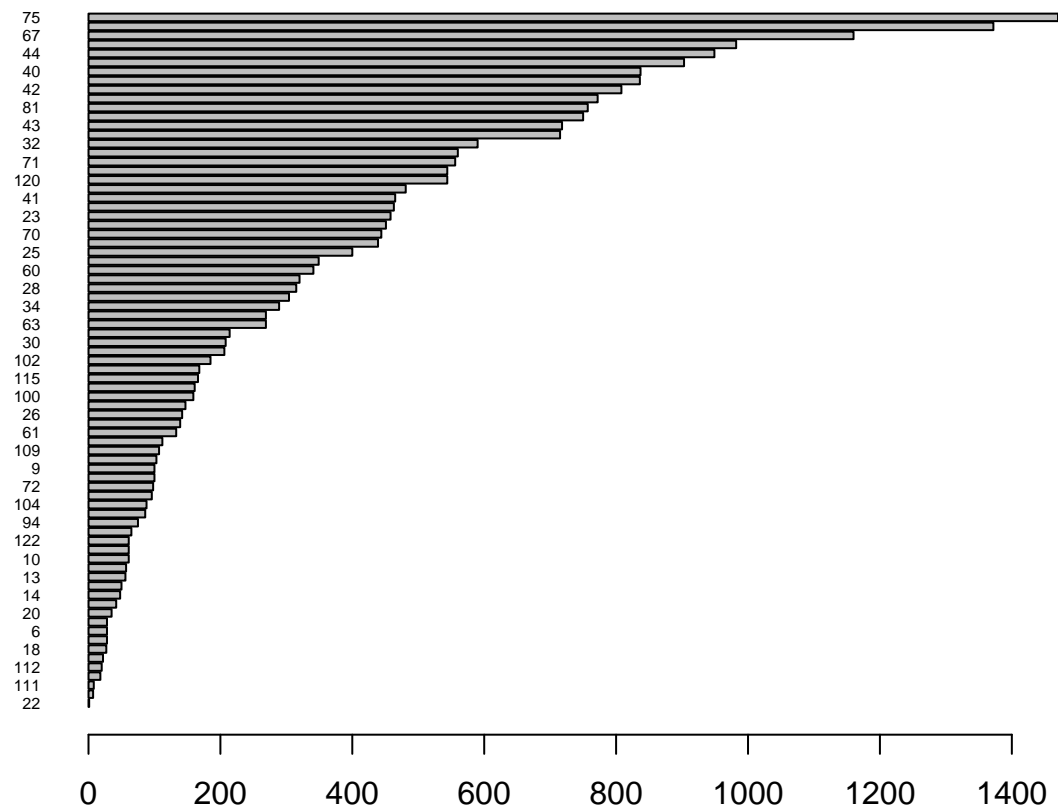
Shooting Incidents by Borough

```
par(mar=c(3,7,0,0))  
barplot(names = s_boro_table$borough, height = s_boro_table$Freq,horiz=TRUE,cex.names=0.8, las=1)
```



Shooting Incidents by Precinct

```
par(mar=c(3, 7,0,0))  
barplot(names = s_precinct_table$precinct, height = s_precinct_table$Freq,horiz=TRUE,cex.names=0.5, las=
```



NYC Facilities Data Exploration

Prepare facilities data for plotting:

```
services_groups = c("YOUTH SERVICES", "CULTURAL INSTITUTIONS", "HUMAN SERVICES", "PARKS AND PLAZAS", "LIBRARIES")
education_groups = c("SCHOOLS (K-12)", "DAY CARE AND PRE-KINDERGARTEN", "CHILD SERVICES AND WELFARE", "VOCATIONAL TRAINING")
services_education_groups = c("YOUTH SERVICES", "CULTURAL INSTITUTIONS", "HUMAN SERVICES", "PARKS AND PLAZAS", "LIBRARIES", "SCHOOLS (K-12)", "DAY CARE AND PRE-KINDERGARTEN", "CHILD SERVICES AND WELFARE", "VOCATIONAL TRAINING")

fac_data = fac_data[fac_data$facgroup %in% services_education_groups,]

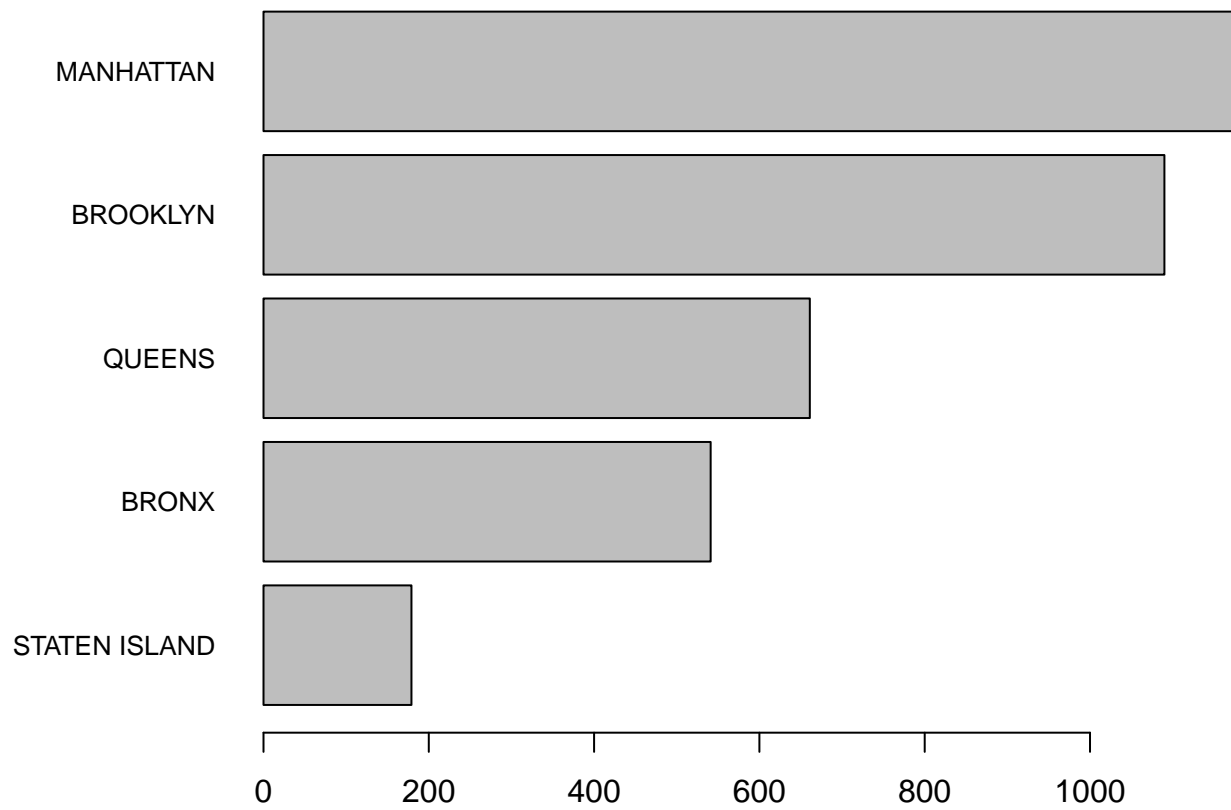
f_facgroup_freq = as.data.frame(table(fac_data$facgroup))
f_facsubgrp_freq = as.data.frame(table(fac_data$facsubgrp))
f_factype_freq = as.data.frame(table(fac_data$factype))

f_boro_table = as.data.frame(table(fac_data$borough))
colnames(f_boro_table) <- c("borough", "Freq")
f_boro_table = merge(f_boro_table, pop_data, by="borough")
f_boro_table = f_boro_table[order(f_boro_table$Freq),]
f_boro_table$per_thous = f_boro_table$Freq / (f_boro_table$population/1000)

f_precinct_table = as.data.frame(table(fac_data$policeprct))
colnames(f_precinct_table) <- c("precinct", "Freq")
f_precinct_table = merge(f_precinct_table, precinct_pop_data, by="precinct")
f_precinct_table = f_precinct_table[order(f_precinct_table$Freq),]
f_precinct_table$per_thous = f_precinct_table$Freq / (f_precinct_table$population/1000)
```

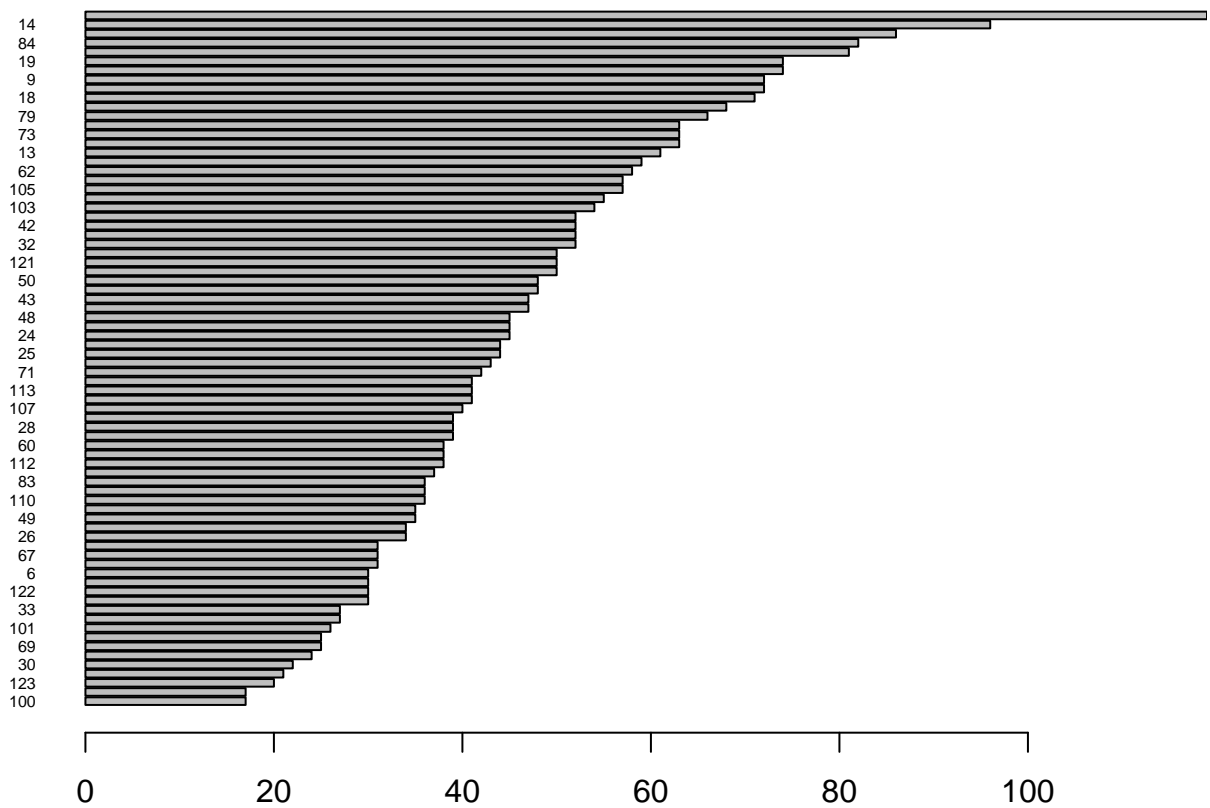
Facilities by Borough

```
par(mar=c(3,7,0,0))  
barplot(names = f_boro_table$borough, height = f_boro_table$Freq,horiz=TRUE,cex.names=0.8, las=1)
```



Facilities by Precinct

```
par(mar=c(3,3,0,0))  
barplot(names = f_precinct_table$precinct, height = f_precinct_table$Freq, horiz=TRUE, cex.names=.5, las=
```



NYC Facilities vs. Shootings Analysis and Modelling

In the following plot, for each precinct, I have shown the number of facilities available per thousand people vs. the number of shootings per thousand people. This data is plotted in blue.

The facilities used in this analysis are both services-based facilities that may offer other opportunities for residents to engage with their community and educational facilities.

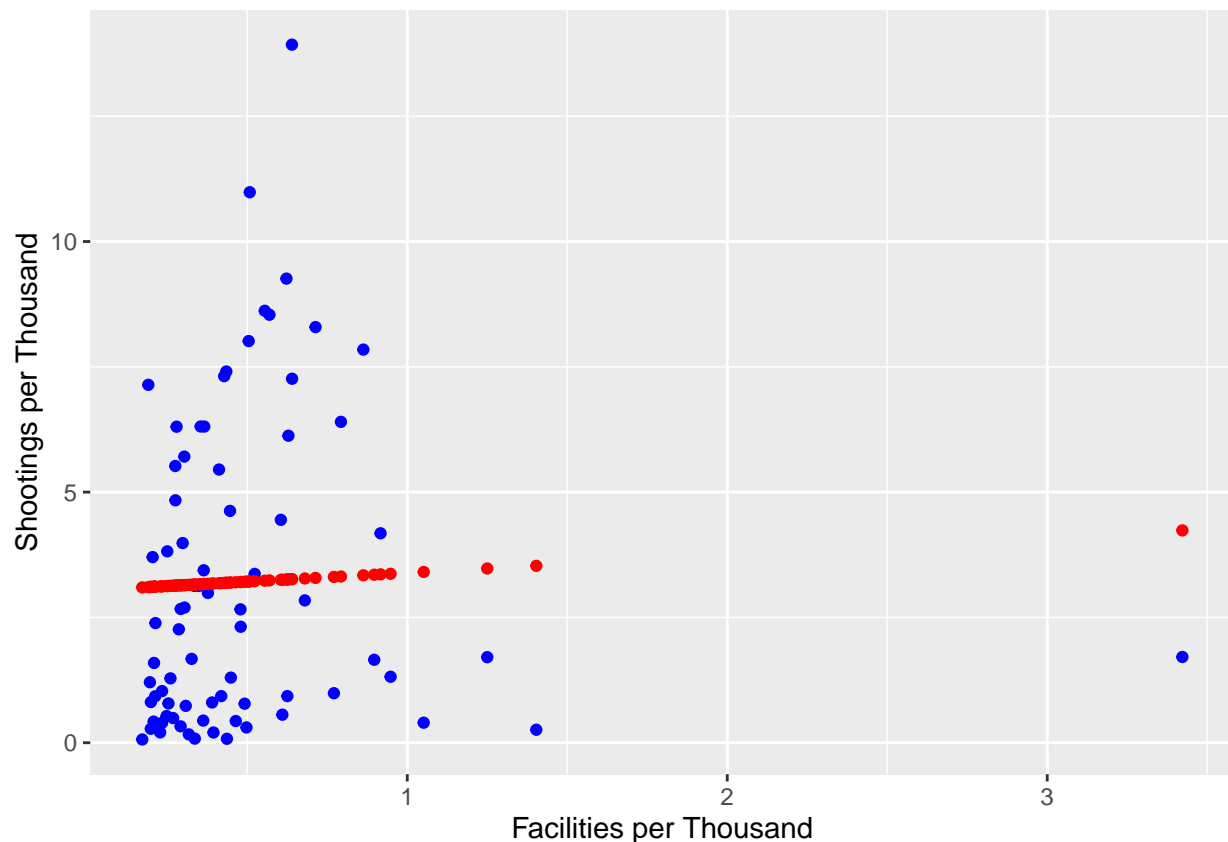
I have also plotted a linear model derived from this data. This data is plotted in red.

As is evident from the plot, this analysis appears to show a weak linear relationship between the number of facilities available in a precinct and the number of shootings that have occurred in that precinct. However, due to the spread of data, I wouldn't want to conclude that there is a clear relationship between facilities available and shootings.

```
combined = merge(s_precinct_table,f_precinct_table,by="precinct")
mod = lm(per_thous.x ~ per_thous.y, data=combined)

combined_w_pred = combined %>% mutate(pred = predict(mod))

combined_w_pred %>% ggplot() + geom_point(aes(x=per_thous.y,y=per_thous.x), color="blue") + geom_point(
```



Marcy Houses Community Center Case Study

I have selected a specific community center, the Marcy Houses Community Center, which opened in November 2020.

<https://www1.nyc.gov/office-of-the-mayor/news/766-20/mayor-de-blasio-celebrates-opening-marcy-houses-community-center#/0>

To determine if the opening of this community center had an impact on shootings, I looked the the shooting incidents within one mile of the facility. I plotted the monthly shooting incidents in this area that occurred between November 2018 through December 2021. I then reviewed the data before and after the facility opened to determine impact. Viewing this time series, it does appear that shooting incidents reduced when looking at the summer months between 2020 and 2021. However, shooting incidents in both years appear significantly higher than 2018 and 2019.

With this analysis, it does not appear that the opening of this facility had a direct impact on shooting incidents within the mile surrounding the facility.

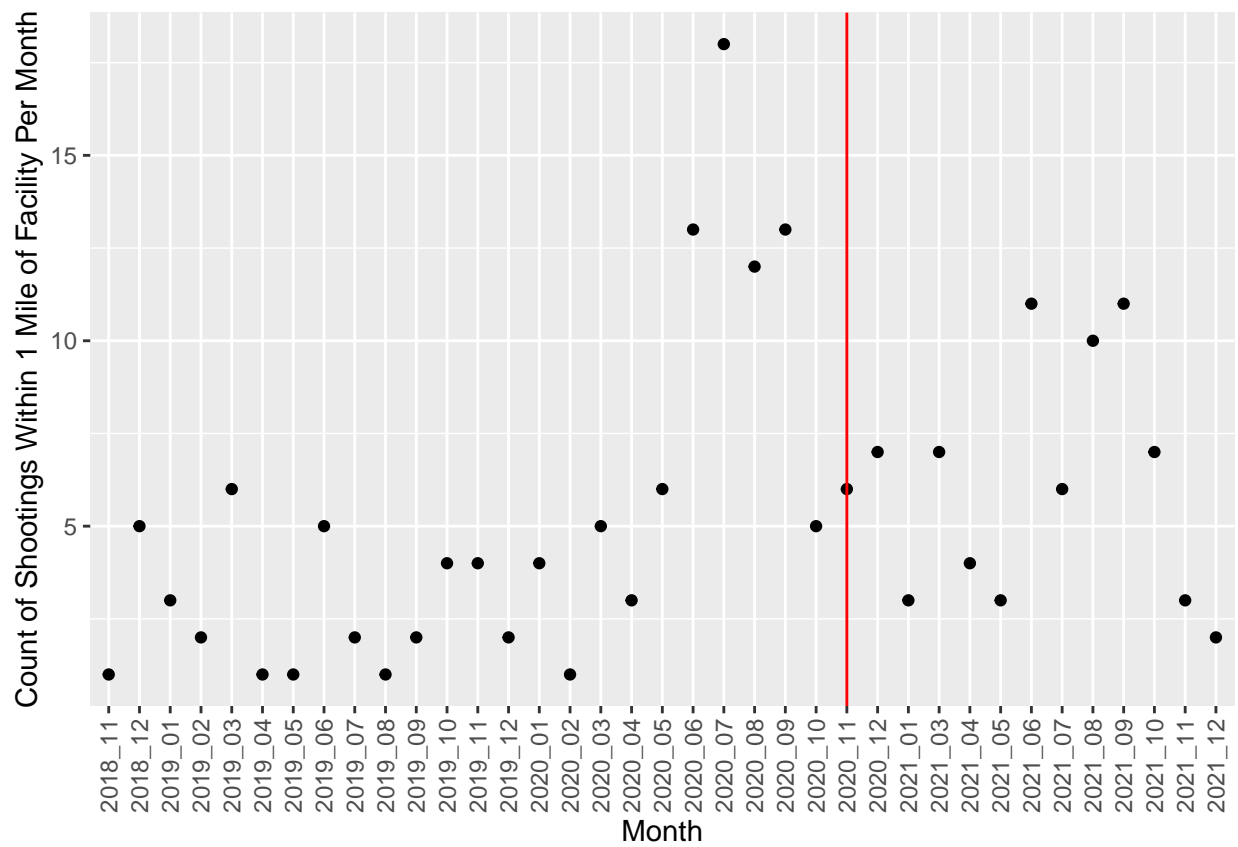
```
dist_from_comm_center = 1609.34
date_opened = date("2020-11-06")
start_obs_period = date_opened - 30*24
end_obs_period = date('2021-12-31')

dist_comm_center = function(coord_lat, coord_long) {
  comm_center_coord = c(-73.95000304, 40.69722182)
  coord = c(coord_long, coord_lat)
  distm(comm_center_coord, coord, fun=distHaversine)
}

nypd_data_dist = nypd_data
nypd_data_dist$dist = mapply(dist_comm_center, nypd_data$Latitude, nypd_data$Longitude)
nypd_data_dist = nypd_data_dist[nypd_data_dist$OCCUR_DATE > start_obs_period & nypd_data_dist$OCCUR_DATE < end_obs_period,]
nypd_data_dist$year_month = format(nypd_data_dist$OCCUR_DATE, format = "%Y_%m")
nypd_data_dist = nypd_data_dist[nypd_data_dist$dist < dist_from_comm_center,]

comm_center_table = as.data.frame(table(nypd_data_dist$year_month))
colnames(comm_center_table) <- c("year_month", "count")
comm_center_table = comm_center_table[order(comm_center_table$year_month),]

comm_center_table %>% ggplot() + geom_point(aes(x=year_month, y=count)) + theme(axis.text.x = element_t
```



Bias and Future Analysis

Bias in this analysis should be considered carefully.

I personally come into this analysis with strong feelings about gun violence. I like the idea that fostering a sense of community in an area could promote empathy, provide opportunity, and reduce violence. I was disappointed when the data didn't clearly show this relationship.

Further analysis should be done on how the shooting data was collected and if there was bias in the reported data. Has the data been reported accurately and completely by the police? Was the location data of the shooting incidents recorded accurately?

In addition, the time period reviewed in this analysis does include the summer of 2020. This includes the civil unrest that followed the murder of George Floyd, which I suspect may account for some of the increased violence in the data. Additional analysis should be done, expanded to more cities, more community centers, and a broader time period, to achieve a more representative picture.

Gun violence is an incredibly complex and substantial subject. Much more in depth analysis will certainly be needed, both on how community centers impact gun violence, and on the many other factors that contribute to it.