NYC Shooting Incidents (2006-2020)

Analyzing NYPD Shooting Incident Data

This report uses public data from the New York City Police Department, explores that data, introduces additional data from the City of New York, and explores trends over time and relationships between factors. Of particular interest:

- Are there significant trends on a weekly, monthly, or yearly basis?
- Are there any notable time frames that are outliers?
- What factors may be causing certain trends or outliers?

NYC Shooting Incident Data Source

This data is provided by the City Of New York, and reported by the New York Police Department. Each row represents a single shooting incident with victim descriptors, perpetrator descriptors, location, and time.

```
# Load our data from the source, cityofnewyork.us
nypd_data <- read.csv('https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD')
```

```
shootings_by_boro <- nypd_data_clean %>%
  group_by(Borough) %>%
  tally()
summary(shootings_by_boro)
```

```
##    Borough                n
##  Length:5           Min.   : 696
##  Class :character   1st Qu.:2922
##  Mode  :character   Median :3532
##                     Mean   :4717
##                     3rd Qu.:6701
##                     Max.   :9734
```

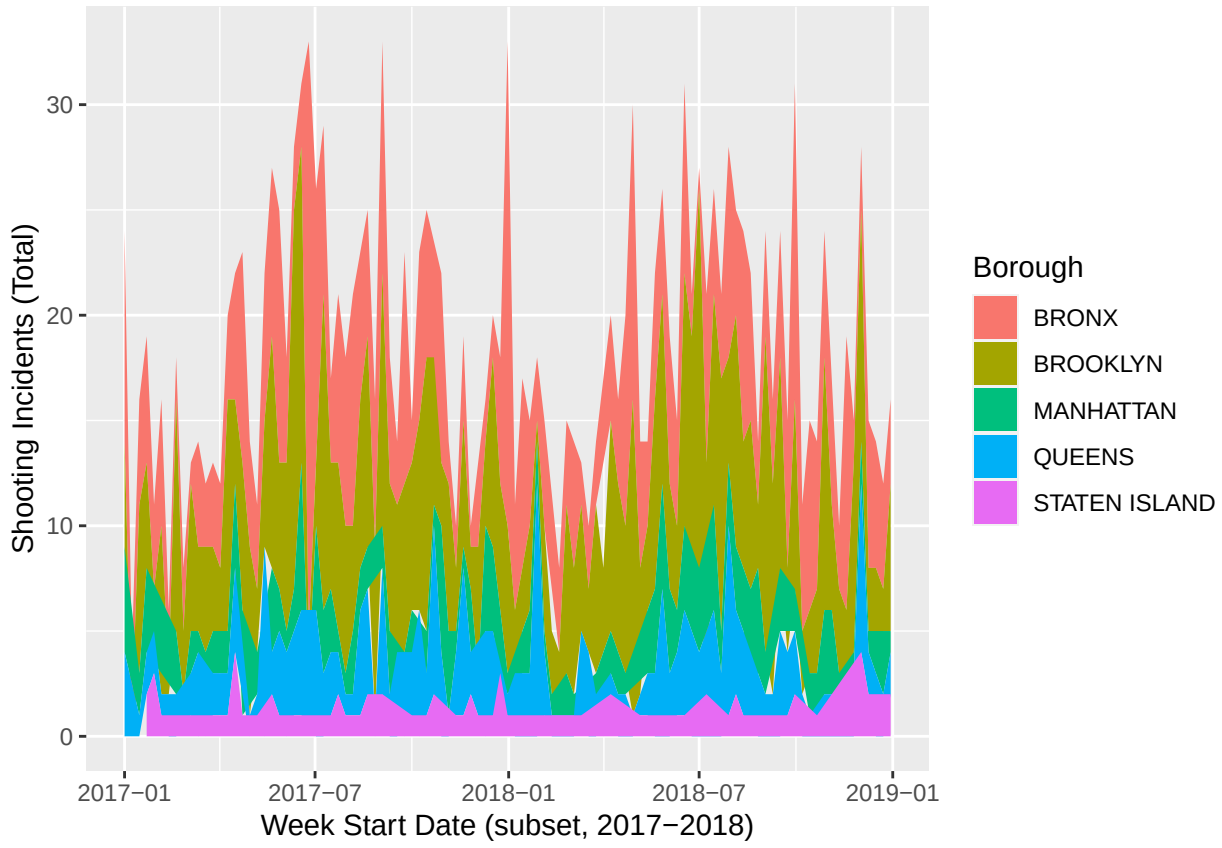
```
shootings_by_boro_by_day <- nypd_data_clean %>%
  group_by(Borough, Date) %>%
  tally()
# sample_n(shootings_by_boro_by_day, 10)

shootings_by_boro_by_day$week <- floor_date(shootings_by_boro_by_day$Date, "week")
by_week <- shootings_by_boro_by_day %>%
  group_by(Borough, week) %>%
  summarize(sum = sum(n))
```

```
## `summarise()` has grouped output by 'Borough'. You can override using the `.groups` argument.
```

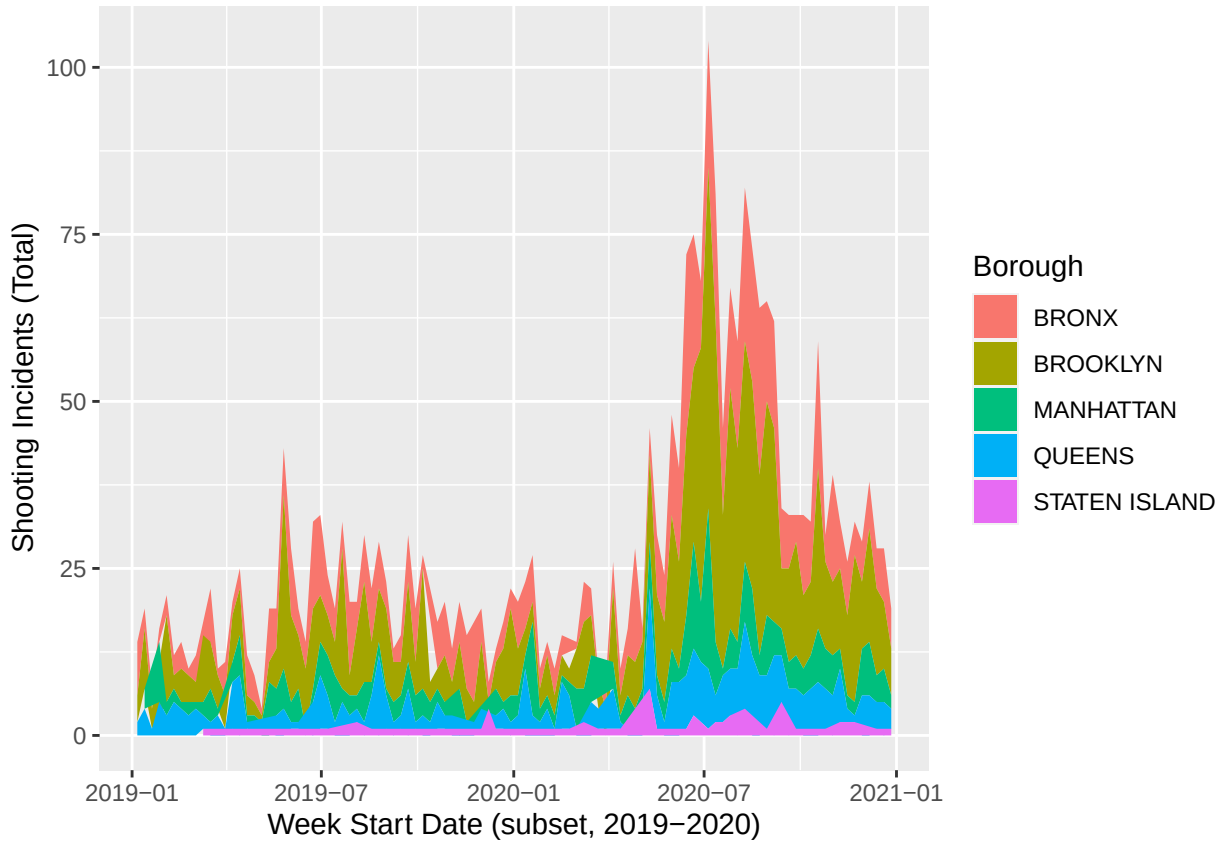
```
sample_n(by_week, 10)
```

```
## # A tibble: 50 x 3
## # Groups:   Borough [5]
##    Borough week         sum
##    <chr>   <date>     <int>
##  1 BRONX   2009-09-06     5
##  2 BRONX   2008-05-11    11
##  3 BRONX   2014-12-21     9
##  4 BRONX   2015-11-15     7
##  5 BRONX   2016-04-03     6
##  6 BRONX   2018-07-29    10
##  7 BRONX   2016-12-04     3
##  8 BRONX   2012-01-15    10
##  9 BRONX   2015-08-02    11
## 10 BRONX   2009-02-08    13
## # ... with 40 more rows
```

```
# Reduce range for effective visualization
by_week_subset = by_week[by_week$week >= "2017-01-01" & by_week$week <= "2019-01-01", ]
ggplot(by_week_subset, aes(x=week, y=sum, fill=Borough)) +
  geom_area() +
  xlab("Week Start Date (subset, 2017-2018)") +
  ylab("Shooting Incidents (Total)")
```



```
by_week_subset = by_week[by_week$week >= "2019-01-01" & by_week$week <= "2021-01-01", ]
ggplot(by_week_subset, aes(x=week, y=sum, fill=Borough)) +
  geom_area() +
  xlab("Week Start Date (subset, 2019-2020)") +
  ylab("Shooting Incidents (Total)")
```



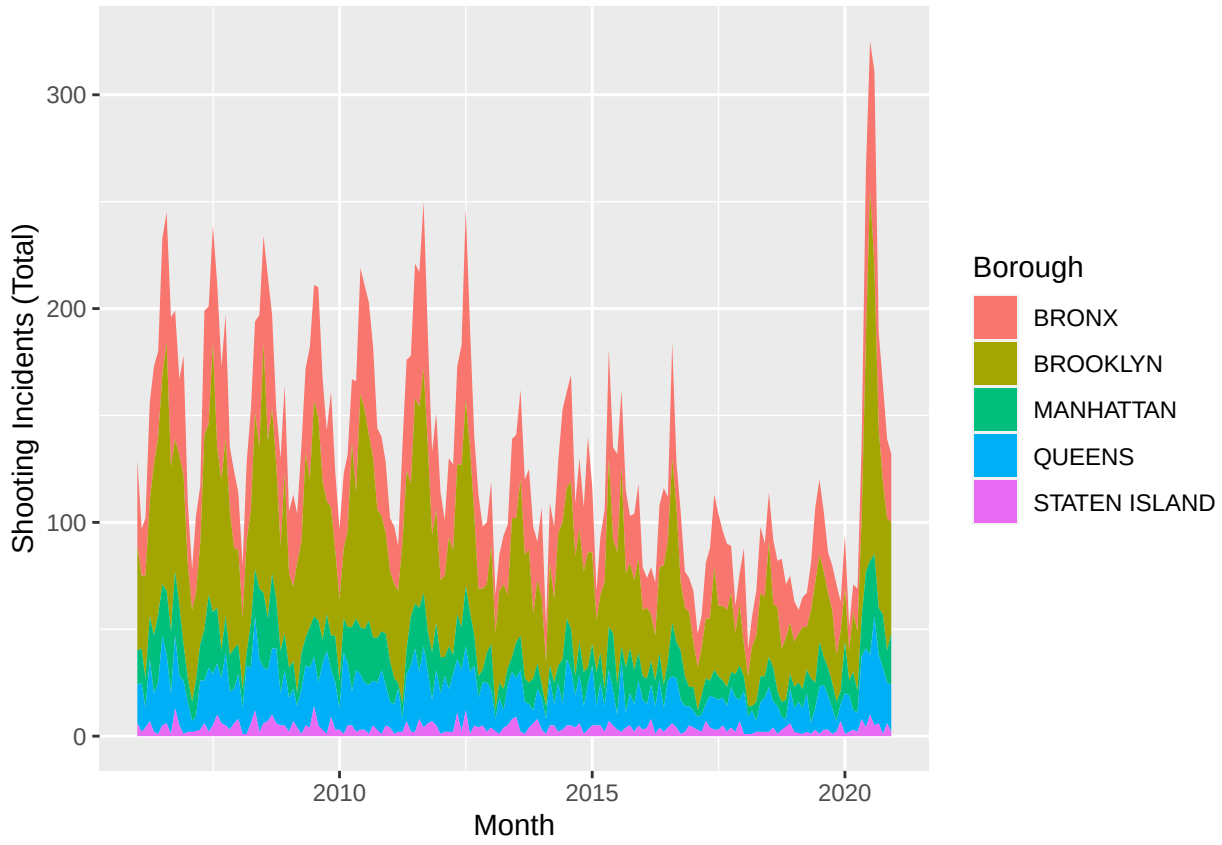
NYC Shooting Trends in 2020

After aggregating the shooting data by week and by borough, and plotting it across time, we can see a clear difference between 2017-2019 and 2020. There is a spike in incidents across most boroughs starting in the spring of 2020 (notably aligning with the beginning of the pandemic). To further confirm this trend we can zoom out a little more, summarizing monthly data for the full time span.

```
shootings_by_boro_by_day$month <- floor_date(shootings_by_boro_by_day$Date, "month")
by_month <- shootings_by_boro_by_day %>%
  group_by(Borough, month) %>%
  summarize(sum = sum(n))
```

```
## `summarise()` has grouped output by 'Borough'. You can override using the `.groups` argument.
```

```
ggplot(by_month, aes(x=month, y=sum, fill=Borough)) +
  geom_area() +
  xlab("Month") +
  ylab("Shooting Incidents (Total)")
```



Combination of Two Trends

Looking at this graph of data by month for the full time span, we can see a rhythm in incidents, with regular spikes during the warmer months. Until 2013 these spikes were quite consistent, then in the following years incidents began to drop with 2017-2019 seeing the lowest spikes in shooting incidents. This downward trend produces a stark contrast between 2019 and 2020, with shooting incidents nearly tripling for 2020. This notably the year the pandemic began. When comparing 2020 numbers to 2013 and earlier the increase is closer to a 50% jump, rather than a multiple. This does not change the unfortunate fact that violence jumped dramatically in 2020, but it does help contextualize it.

Comparing with COVID-19 Data

In order to draw conclusions around shooting incidents and the influence of the COVID-19 pandemic, we will pull in another historical dataset from the City of New York, which contains daily counts for confirmed cases, hospitalizations, and deaths.

```
nyc_covid_data <- read.csv('https://data.cityofnewyork.us/resource/rc75-m7u3.csv') %>%
  mutate(CLEAN_DATE=str_replace_all(date_of_interest,"T00:00:00.000","")) %>%
  mutate(Date = ymd(CLEAN_DATE)) %>%
  select(Date, case_count, hospitalized_count, death_count)
head(nyc_covid_data)
```

```
##         Date case_count hospitalized_count death_count
```

```
## 1 2020-02-29          1                 1           0
## 2 2020-03-01          0                 1           0
## 3 2020-03-02          0                 2           0
## 4 2020-03-03          1                 7           0
## 5 2020-03-04          5                 2           0
## 6 2020-03-05          3                14           0
```

```
tail(nyc_covid_data)
```

```
##           Date case_count hospitalized_count death_count
## 639 2021-11-28       1063                 48           7
## 640 2021-11-29       2254                 58           7
## 641 2021-11-30       1783                 62           5
## 642 2021-12-01       1909                 53           9
## 643 2021-12-02       1653                 21           5
## 644 2021-12-03       1281                  3           1
```

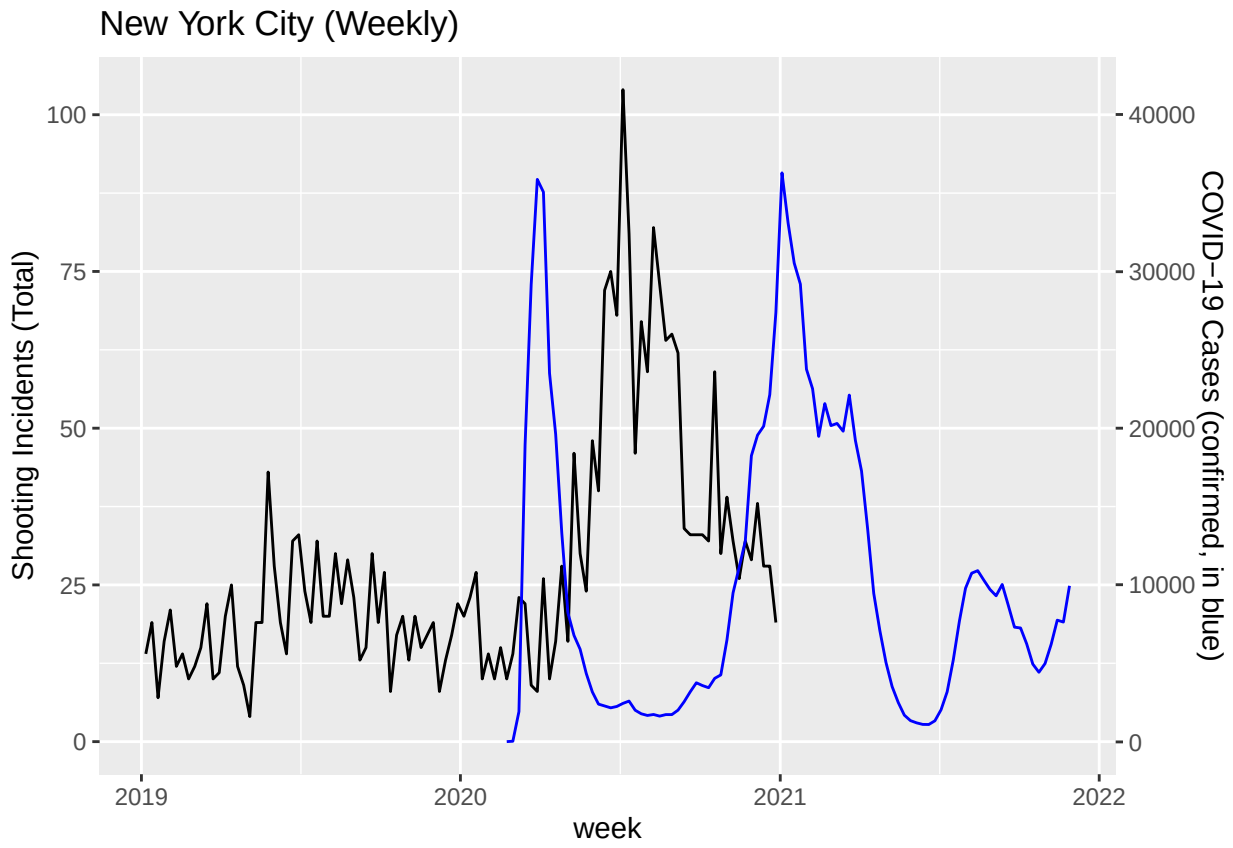
```
# Group by Week for visualization
nyc_covid_data$week <- floor_date(nyc_covid_data$Date, "week")
covid_by_week <- nyc_covid_data %>%
  group_by(week) %>%
  summarize(case_sum = sum(case_count), death_sum = sum(death_count), hospitalized_sum = sum(hospitaliz
sample_n(covid_by_week, 10)
```

```
## # A tibble: 10 x 4
##    week       case_sum death_sum hospitalized_sum
##    <date>        <int>     <int>            <int>
##  1 2021-02-14    19477       510             2082
##  2 2020-11-08     9491        61              592
##  3 2020-11-22    12793        89              962
##  4 2020-07-26     1777        40              229
##  5 2021-05-09     3514       155              437
##  6 2021-10-10     6237        58              359
##  7 2021-06-06     1200        48              190
##  8 2020-09-13     2546        29              214
##  9 2021-08-15    10910        74              783
## 10 2020-12-06    19556       178             1586
```

Comparing Shooting Incidents with COVID-19 Cases

```
total_by_week_subset <- by_week_subset %>%
  group_by(week) %>%
  summarize(total = sum(sum))
scale <- 400
ggplot() +
  ggtitle("New York City (Weekly)") +
  geom_line(aes(x=week, y=total), data=total_by_week_subset, color="black") +
  geom_line(aes(x=week, y=case_sum/scale), data=covid_by_week, color="blue") +
  scale_y_continuous(
    # First Y
```

```
    name = "Shooting Incidents (Total)",
    # Second Y
    sec.axis = sec_axis(~.*scale, name="COVID-19 Cases (confirmed, in blue)")
  )
```



Comparing COVID-19 Cases and Shootings

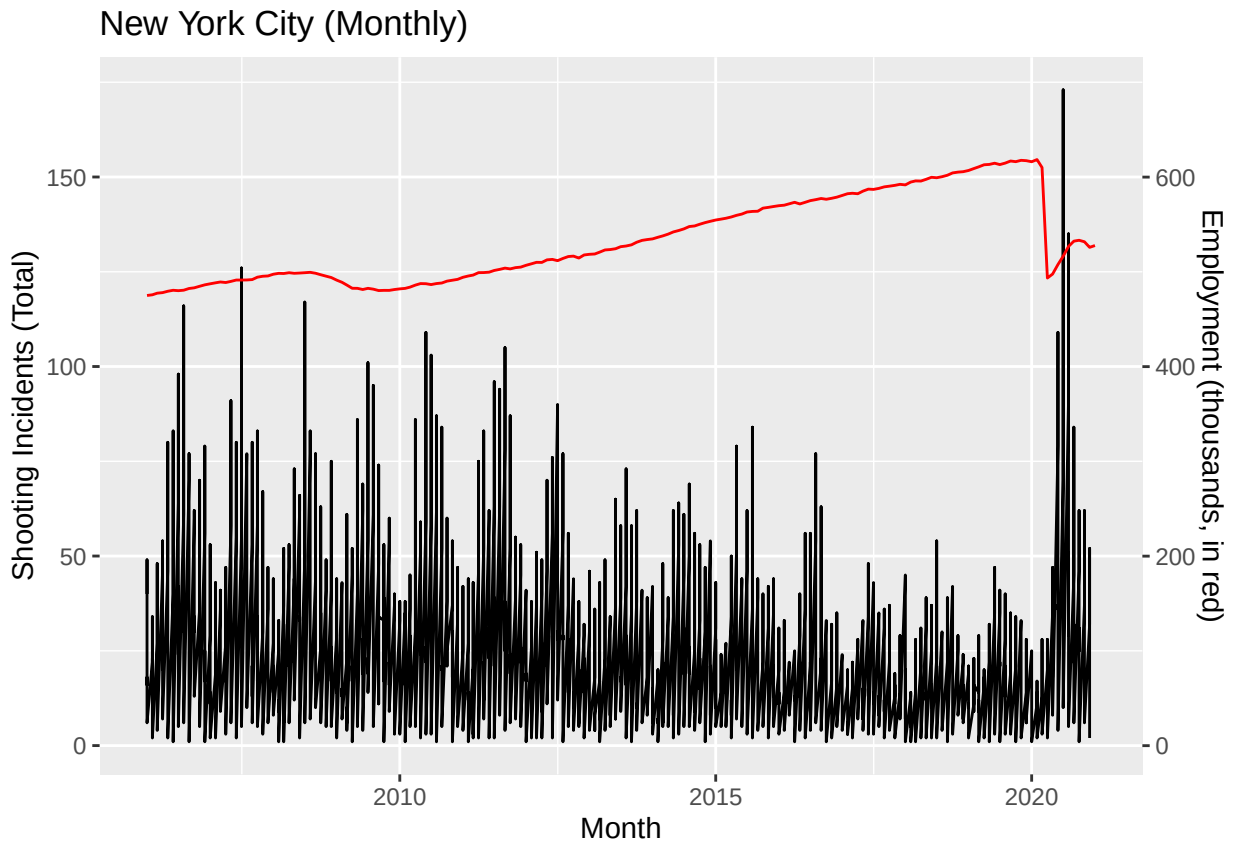
Comparing COVID-19 cases with shooting incidents presents an interesting relationship which could be misconstrued as inverse. More recent data would complete the picture, but other factors influenced by the pandemic are likely the culprit here. That is, violence was not escalating when more people had COVID-19 (as in the disease did not directly cause people to shoot people), but rather secondary factors like lockdown measures and job losses increased shooting incidents, violence. We can investigate this hypothesis further by pulling in employment data and comparing it with shooting incidents.

Employment Data

```
# Aggregate data across industries
nyc_employ_monthly <- nyc_employ_data %>%
  mutate(month = ym(year_month)) %>%
  group_by(month) %>%
  summarize(employ_avg = mean(EMPLOYMENT))
nyc_employ_monthly_subset <- nyc_employ_monthly[nyc_employ_monthly$month >= "2006-01-01" & nyc_employ_m
head(nyc_employ_monthly_subset)
```

```
## # A tibble: 6 x 2
##   month      employ_avg
##   <date>          <dbl>
## 1 2006-01-01       475.
## 2 2006-02-01       476.
## 3 2006-03-01       477.
## 4 2006-04-01       478.
## 5 2006-05-01       479.
## 6 2006-06-01       480.
```

```r
# Compare shootings with employment
scale <- 4
ggplot() +
  ggtitle("New York City (Monthly)") +
  xlab("Month") +
  geom_line(aes(x=month, y=sum), data=by_month, color="black") +
  geom_line(aes(x=month, y=employ_avg/scale), data=nyc_employ_monthly_subset, color="red") +
  #geom_line(aes(x=week, y=death_sum/scale), data=covid_by_week, color="purple") +
  scale_y_continuous(
    # First Y
    name = "Shooting Incidents (Total)",
    # Second Y
    sec.axis = sec_axis(~.*scale, name="Employment (thousands, in red)")
  )
```



Initial Analysis: Employment and Shooting Incidents

Comparing employment numbers with shooting incidents, we can see there does seem to be some correlation between rising employment and falling shooting incidents. This is most clearly seen in early 2020, when there is a sharp drop in employment due to the pandemic, and a corollary spike in shooting incidents.

Modeling Employment and Shooting Incidents

Seen below, a linear model with shooting incidents as a function of employment numbers does indeed describe a significant relationship. The model summary implies it is not the strongest relationship, but it is enough to help clarify the relationship between COVID-19 and violence.

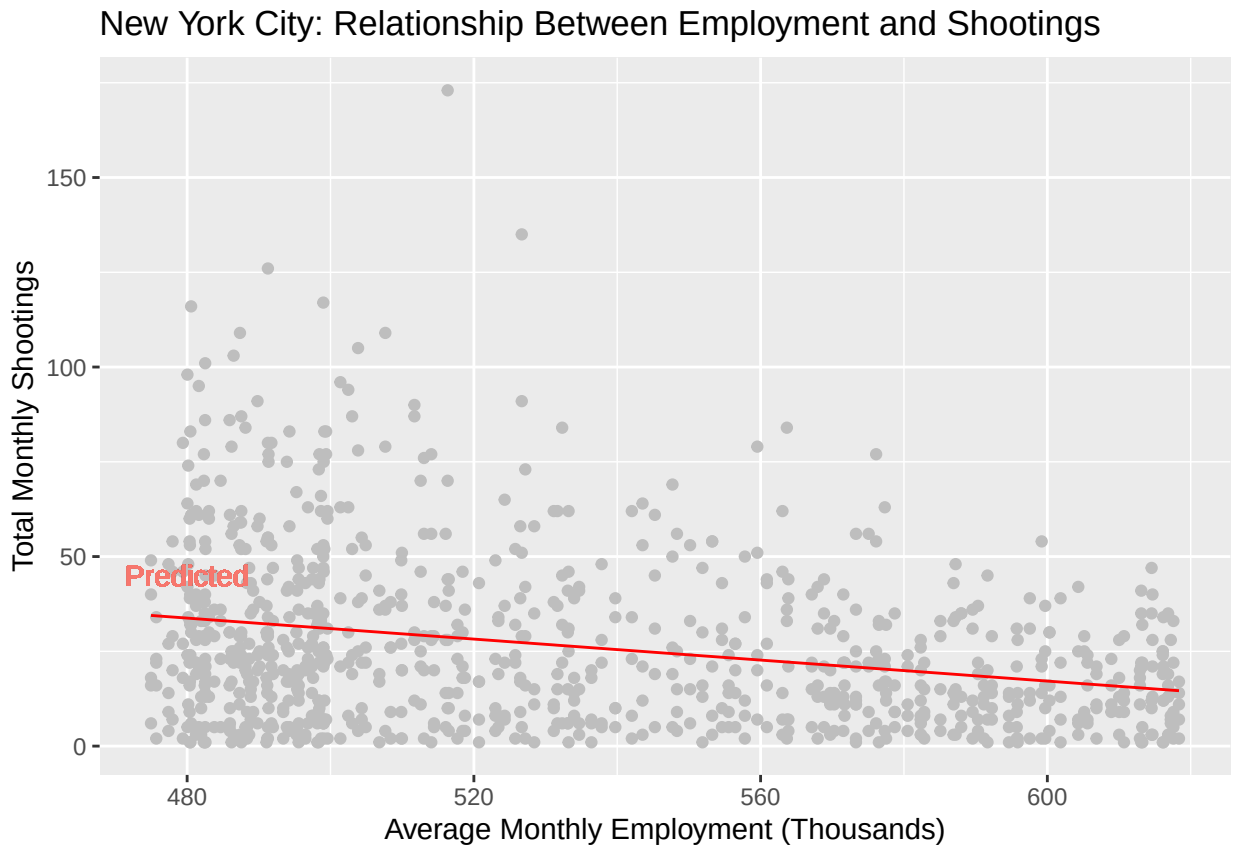
```
# Combine shooting and employment data on month value
merged <- merge(by_month, nyc_employ_monthly_subset, by = c("month", "month"), all.x=TRUE)
employ_monthly_avgs <- nyc_employ_monthly_subset %>%
  select(employ_avg)
# Build and Evaluate Model
linear_model <- lm(sum ~ employ_avg, data=merged)
summary(linear_model)
```

```
##
## Call:
## lm(formula = sum ~ employ_avg, data = merged)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.715 -15.189  -4.526  10.807 144.265
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 100.23910    8.74206  11.466   <2e-16 ***
## employ_avg   -0.13848    0.01633  -8.479   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.99 on 892 degrees of freedom
## Multiple R-squared:  0.07458,    Adjusted R-squared:  0.07354
## F-statistic: 71.89 on 1 and 892 DF,  p-value: < 2.2e-16
```

```
predictions <- predict(linear_model)
modtest <- bind_cols(merged, predictions)
```

```
## New names:
## * NA -> ...5
```

```
ggplot(modtest, aes(x=employ_avg, y=sum)) +
  ggtitle("New York City: Relationship Between Employment and Shootings") +
  theme(legend.position = "none") +
  xlab("Average Monthly Employment (Thousands)") +
  ylab("Total Monthly Shootings") +
  geom_point(color="gray") +
  geom_text(aes(x=480, y = 45, label = "Predicted", color = "red")) +
  geom_line(aes(x=employ_avg, y=predictions), color="red")
```



Possible Bias

The source data is provided by the Office of Management Analysis and Planning (OMAP), the "evaluation and assessment arm" of the New York City Police Department. The office focuses on optimizing the organization of the NYPD, which implies that data reported will impact officer assignments and precinct funding. Since this data is specifically shootings reported to the police, it is also possible that police response time, neighborhood affiliation, and historical proceedings would lead to under-reporting in some neighborhoods.

New York City's official reporting for COVID-19 data was debated early on during the pandemic, since case numbers directly influenced federal aid and resources available to the city. It is not clear whether the dataset presented here would have been influenced by this supposed bias.

New York City's Seasonally Adjusted Employment numbers are provided by the Mayor's Office of Management & Budgets (OMB), which may have political incentives which would introduce bias.

Additionally, this report's author previously lived in New York City, primarily in Manhattan, which may also have introduced some bias into this report. However, in providing an assessment, I have decided not to focus on any one borough, and I have no particular leanings regarding the NYPD.

Conclusion

New York City saw a large jump in shooting incidents during the spring and summer of 2020. When compared to previous years, 2013 and earlier, this increase is not as dramatic but it is still much higher than the worst monthly spike since 2006. This is related to this time frame aligning with the beginning of the COVID-19 pandemic, but it was the secondary impacts of COVID-19 which were the culprit. Comparing COVID-19 cases directly with shooting incidents presented an unclear, almost inverse relationship which could be

attributed to the delay in secondary impacts of the pandemic. One of these impacts was confirmed by comparing employment data with shooting incidents, presenting a significant relationship between decreased employment and increased shooting incidents.