

Temperature and Gun Violence in New York City

2022-08-24

Background

Temperature affects many of our day-to-day decisions. Consider the simple example of choosing an outfit. On a hot day, a person is more likely to wear shorts and less likely to wear a sweater but is complex behavior, like gun violence, also affected by temperature?

To determine if there is a relationship between temperature and gun violence, data on total shootings in New York City between 2006 - 2021 were analyzed.

The data were retrieved from from data.gov on August 24th, 2022.

```
# Loading data into R, replacing blank values with NA, nys stands for New York Shooting
nys <- read.csv('https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD', header=TRUE)

# Printing summary of data
summary(nys)
##      INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
## Min.      : 9953245  Length:25596  Length:25596  Length:25596
## 1st Qu.: 61593633  Class :character  Class :character  Class :character
## Median : 86437258  Mode  :character  Mode  :character  Mode  :character
## Mean      :112382648
## 3rd Qu.:166660833
## Max.      :238490103
##
##      PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_FLAG
## Min.      : 1.00  Min.      :0.0000  Length:25596  Length:25596
## 1st Qu.: 44.00  1st Qu.:0.0000  Class :character  Class :character
## Median : 69.00  Median :0.0000  Mode  :character  Mode  :character
## Mean      : 65.87  Mean      :0.3316
## 3rd Qu.: 81.00  3rd Qu.:0.0000
## Max.      :123.00  Max.      :2.0000
##
##      NA's      :2
##      PERP_AGE_GROUP      PERP_SEX      PERP_RACE      VIC_AGE_GROUP
## Length:25596  Length:25596  Length:25596  Length:25596
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      VIC_SEX      VIC_RACE      X_COORD_CD      Y_COORD_CD
## Length:25596  Length:25596  Min.      : 914928  Min.      :125757
## Class :character  Class :character  1st Qu.:1000011  1st Qu.:182782
## Mode  :character  Mode  :character  Median :1007715  Median :194038
## Mean      :1009455  Mean      :207894
```

```
##                               3rd Qu.:1016838  3rd Qu.:239429
##                               Max.      :1066815  Max.      :271128
##
##      Latitude      Longitude      Lon_Lat
##  Min.      :40.51   Min.      :-74.25   Length:25596
##  1st Qu.:40.67   1st Qu.: -73.94   Class :character
##  Median :40.70   Median : -73.92   Mode  :character
##  Mean    :40.74   Mean     :-73.91
##  3rd Qu.:40.82   3rd Qu.: -73.88
##  Max.    :40.91   Max.     :-73.70
##
```

Data Review and Wrangling

The output of the `summary()` function was used to identify variables to include/exclude from the analysis, and to ensure the data were in the appropriate format. After reviewing the output, 17 variables were removed from the data set. The variables were removed because they were outside the scope of the analysis (e.g., race of perpetrator, sex of victim) or had limited utility (e.g., longitude, latitude).

Date data were converted from a string variable to a date object, separated into different columns by day, month, and year, then bound to the original data frame using the `bind_rows()` function.

```
# Removing excluded categories from data frame
# nys_rmvd stands for New York Shooting Data with variables removed
nys_rmvd <- subset(nys, select = -c(INCIDENT_KEY, JURISDICTION_CODE, X_COORD_CD,
                                   Y_COORD_CD, Latitude, Longitude, Lon_Lat,
                                   PERP_RACE, PERP_SEX, PRECINCT,
                                   PERP_AGE_GROUP, LOCATION_DESC, OCCUR_TIME,
                                   STATISTICAL_MURDER_FLAG, VIC_AGE_GROUP,
                                   VIC_SEX, VIC_RACE))

# Changing format of date
nys_rmvd$OCCUR_DATE <- as.POSIXct(nys_rmvd$OCCUR_DATE, format = "%m/%d/%Y")

# Creating a new data frame with seperated date data
df <- nys_rmvd %>%
  dplyr::mutate(YEAR = lubridate::year(OCCUR_DATE),
               MONTH = lubridate::month(OCCUR_DATE),
               DAY = lubridate::day(OCCUR_DATE))

# Binding two data frames
nys_mod <- bind_rows(df, nys_rmvd)

# Removing old date data
nys_mod <- subset(nys_mod, select = -c(OCCUR_DATE))
```

The data were then piped into the `count()` function to determine the total shootings per month. Data scored as NA (i.e., not associated with a month) were not included in the analysis and were removed from the data frame.

Data on the monthly average temperature were not included in the original data set; therefore, they were retrieved from an external source. Monthly averages from 1991 to 2020 were retrieved from Current Results and manually entered into R. The data can be accessed using the hyper-linked text.

After the data were manually entered into an R vector, they were attached to a data frame containing the total number of shootings organized by month. The new data frame was used to generate scatter plots and build simple linear regression models.

```
# Data frame with total shootings organized by month
nys_count <- nys_mod %>% count(MONTH)

# Removing NAs from data frame
nys_count <- nys_count[1:12,]

# Average high temperature for NYC by month
average_high <- c(39, 42, 50, 63, 73, 80, 86, 84, 78, 66, 54, 45)
average_low <- c(27, 29, 35, 46, 56, 64, 71, 69, 63, 53, 41, 35)

# Adding average high temperature to data frame
nys_count$average_high <- average_high
nys_count$average_low <- average_low
```

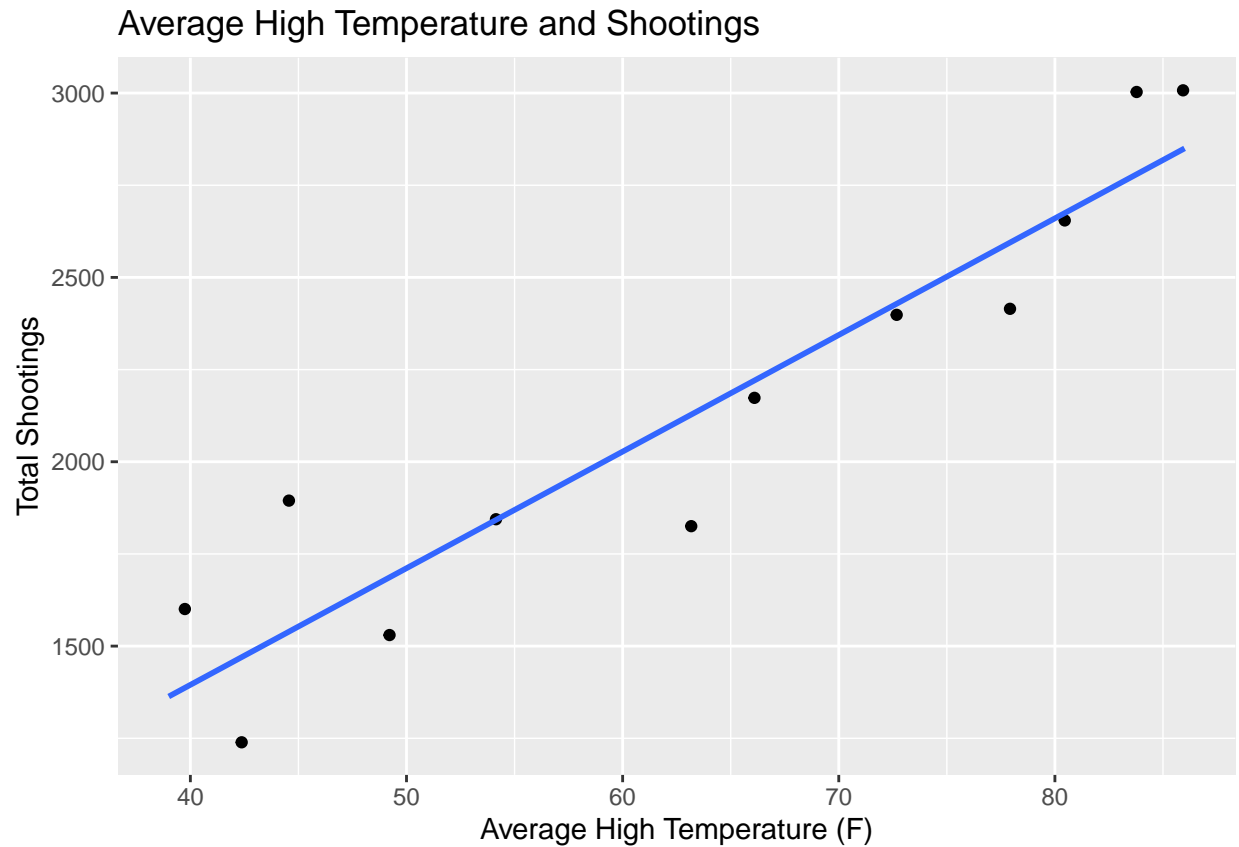
Analysis

Scatter Plots

The nys_count data frame and ggplot2 package were used to create two scatter plots. The purpose of creating the scatter plots was to determine if there was a correlation between temperature and total number of shootings.

```
# Scatter plot for average high temperature and total shootings
ggplot(data = nys_count, mapping = aes(x = average_high, y = n)) +
  geom_jitter() +
  geom_smooth(method = lm, se = F) +
  ylab("Total Shootings") +
  xlab("Average High Temperature (F)") +
  ggtitle("Average High Temperature and Shootings")
```

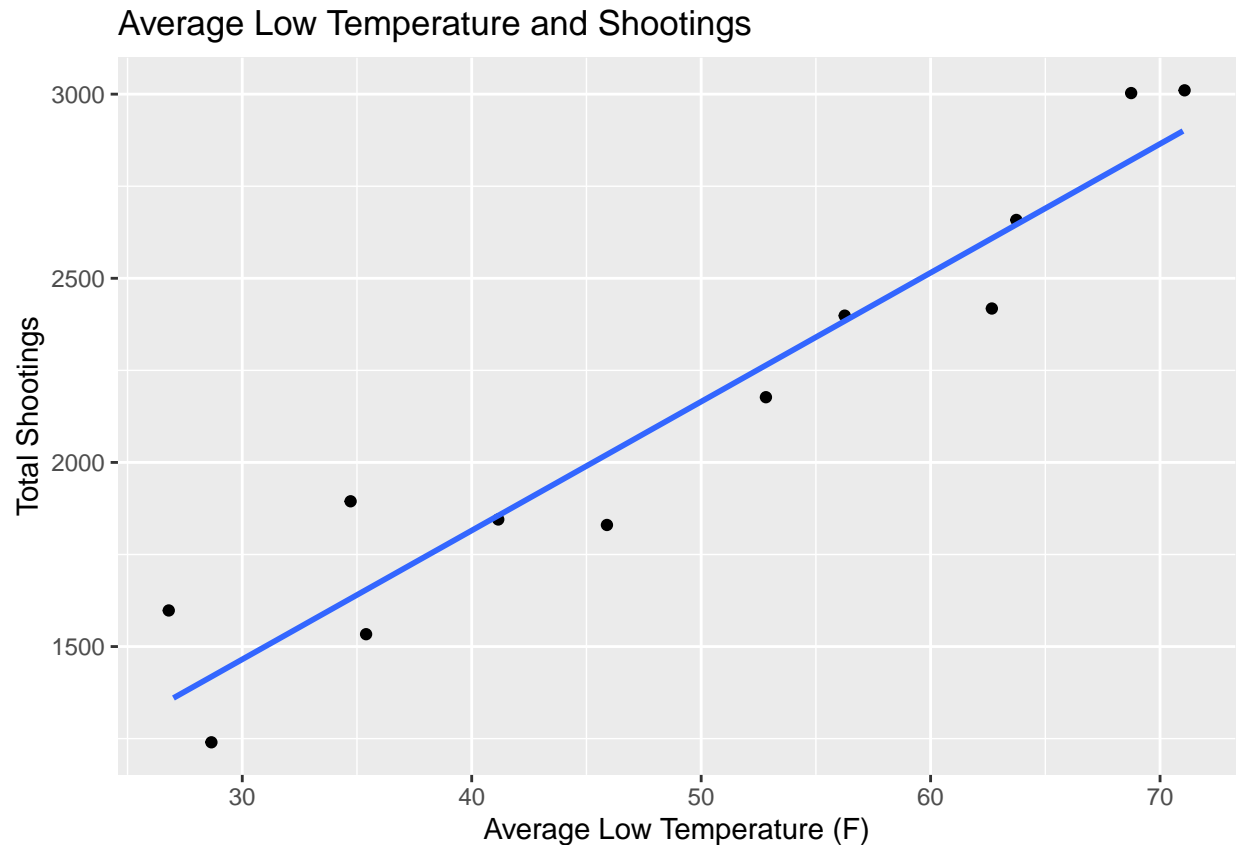
```
## 'geom_smooth()' using formula 'y ~ x'
```



The data displayed above indicate a positive correlation between average high temperature and total shootings.

```
# Scatter plot for average low temperature and total shootings
ggplot(data = nys_count, mapping = aes(x = average_low, y = n)) +
  geom_jitter() +
  geom_smooth(method = lm, se = F) +
  ylab("Total Shootings") +
  xlab("Average Low Temperature (F)") +
  ggtitle("Average Low Temperature and Shootings")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



Similar to the scatter plot examining average high temperature and total shootings, the scatter plot for average low temperature and total shootings also show a positive correlation.

Linear Models

```
# Linear model
linear_mod <- lm(average_high ~ n, data = nys_count)

# Summary of linear model
summary(linear_mod)
```

```
##
## Call:
## lm(formula = average_high ~ n, data = nys_count)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.708  -2.069   1.808   3.365   8.122
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.202113    7.235995   0.581   0.574
## n             0.027722    0.003285   8.439 7.35e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.26 on 10 degrees of freedom
## Multiple R-squared:  0.8769, Adjusted R-squared:  0.8646
## F-statistic: 71.22 on 1 and 10 DF,  p-value: 7.351e-06
```

The summary output indicates that average high temperature accounts for 86.46% of variance in the data.

```
# Linear model
linear_mod <- lm(average_low ~ n, data = nys_count)

# Summary of linear model
summary(linear_mod)
```

```
##
## Call:
## lm(formula = average_low ~ n, data = nys_count)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.1188 -1.4885  0.5415  2.9339  6.5021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.800901   5.497538  -1.237   0.244
## n             0.026200   0.002496  10.498 1.02e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.756 on 10 degrees of freedom
## Multiple R-squared:  0.9168, Adjusted R-squared:  0.9085
## F-statistic: 110.2 on 1 and 10 DF,  p-value: 1.016e-06
```

The summary output indicates that average low temperature accounts for 90.85% of variance in the data.

Discussion

Based on the results of the analysis, there appears to be a relationship between temperature and gun violence. Both scatter plots indicate a positive correlation between temperature and total shootings, and both models account for a large amount of the variance in the data.

How the Analysis Can Be Used

The information from the analysis might be useful to law enforcement agencies and other stakeholder groups in New York City. For instance, it can inform when to allocate limited resources (e.g., access to indoor activities with air conditioning).

Limitations

Although the data from this analysis could be used to address gun violence, there are at least two limitations that affect its utility. First, the analysis used monthly averages that were aggregated across several years, some of which were before 2006. Second, the analysis did not identify areas (e.g., public housing) associated with gun violence. Ideally, future iterations of this project will account for the limitations of the analysis.

Potential Biases

To minimize bias, variables involving race, age, and sex were not included in the analysis; however, it may have been affected by other sources of bias. Specifically, the researcher's previous experience may have influenced the type of models used (i.e., selection of simple linear regression).

Session Info

```
sessionInfo()
```

```
## R version 4.0.4 (2021-02-15)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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] lubridate_1.8.0 forcats_0.5.1  stringr_1.4.0  dplyr_1.0.8
## [5] purrr_0.3.4    readr_2.1.2    tidyr_1.2.0    tibble_3.1.6
## [9] ggplot2_3.3.5  tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.1.2 xfun_0.30      lattice_0.20-45 splines_4.0.4
## [5] haven_2.4.3      colorspace_2.0-3 vctrs_0.3.8    generics_0.1.2
## [9] htmltools_0.5.2  mgcv_1.8-40    yaml_2.3.5     utf8_1.2.2
## [13] rlang_1.0.2      pillar_1.7.0   glue_1.6.2     withr_2.5.0
## [17] DBI_1.1.2        dbplyr_2.1.1   modelr_0.1.8   readxl_1.4.0
## [21] lifecycle_1.0.1  munsell_0.5.0  gtable_0.3.0   cellranger_1.1.0
## [25] rvest_1.0.2      evaluate_0.15  labeling_0.4.2 knitr_1.38
## [29] tzdb_0.3.0       fastmap_1.1.0  fansi_1.0.3    highr_0.9
## [33] broom_0.8.0      backports_1.4.1 scales_1.2.0   jsonlite_1.8.0
## [37] farver_2.1.0     fs_1.5.2       hms_1.1.1      digest_0.6.29
## [41] stringi_1.7.6    grid_4.0.4     cli_3.2.0      tools_4.0.4
## [45] magrittr_2.0.3   crayon_1.5.1   pkgconfig_2.0.3 Matrix_1.4-1
## [49] ellipsis_0.3.2   xml2_1.3.3     reprex_2.0.1   assertthat_0.2.1
## [53] rmarkdown_2.13   httr_1.4.2     rstudioapi_0.13 R6_2.5.1
## [57] nlme_3.1-157     compiler_4.0.4
```