Boston Crime Analysis

IST 687 Final Project

September 16, 2019

Contents

Summary

Between June 2015 and October 2018, the city of Boston experienced 327,820 incidences of crime. During that time, as few as 140 crimes or as many as 379 crimes would occur during any given day (*Table 4*). The Boston Police force is comprised of 2,015 officers and 808 civilian personnel, reference [1]. Since the city of Boston requires 24-hour coverage and the Police Department utilizes a 3 shift daily schedule, that means that during low crime rate hours, Boston will have .14 (140/941) crimes occurrences per individual police officers whereas during high crime rate hours, Boston will have .40 (379/941) crime occurrences per police officer. The variance of crime by hour of day, day of week, and month of year results in a complex logistical problem for the city to solve. How many police officers do they need to hire in fiscal year (FY) 21? How many police officers need to be schedule per month, day, and hour to ensure that the ratio of crimes to police officers remain as low as possible? Questions like these served as the foundation to the development of our final project.

Data

For this project, the website www.kaggle.com[1] was used to acquire a dataset. The data set chosen was "Boston Crime Data from 2015-2018" consisting of 327,820 observations of 18 variables (Table 1).

Table 1. Crime Data Set Description

Name	Type	Description
OCCURRED_ON_DATE	POSIXct	YYYY-MM-DD noting when crime occurred
INCIDENT_NUMBER	Chr	Accounting code for tracking
OFFENSE_CODE	Num	Categorization code for offense
OFFENSE_CODE_GROUP	Chr	Description of Offense Code
OFFENSE_DESCRIPTION	Chr	Alternate Description of Offense Code
DISTRICT	Chr	Location of crime based on chr num num (B14)
REPORTING_AREA	Num	Numerical categorization of location (XXX)
SHOOTING	Chr	NA or Y
YEAR	Num	YYYY
MONTH	Num	MM
DAY_OF_WEEK	Chr	Monday, Tuesday, etc
HOUR	Num	Military time of day
UCR_PART	Chr	Uniform Crime Report (UCR) Categorization
STREET	Chr	Street location of offense
LAT	Num	Latitude location of offense
LONG	Num	Longitude location of offense
LOCATION	Chr	Combined latitude and longitude location

[1] https://www.kaggle.com/ankkur13/boston-crime-data

To help understand what contributes to crime rate, a second data set was acquired from the NOAA website[2]. This data set, entitled "Weather" in the code, consisted 1,462 observations of seven variables (Table 2).

Table 2. Weather Data Set Description

Name	Type	Description
Name	Chr	Name of station reporting "BOSTON, MA US"
Date	Date	YYYY-MM-DD
Prcp	Num	X.XX Amount of precipitation on that date
Snow		X.XX Amount of snow on that date
AvgTemp		XX Average temperature recorded that day
MaxTemp		XX Maximum temperature recorded that day
MinTemp		XX Minimum temperature recorded that day.

Preparing the Data

To prepare the data, the team began by importing the data into R via the 'readxl' library. This allowed the team to quickly import both data sets via two lines of code.

```
Weather <- read_excel("C:/Users/jonat/Desktop/Boston Crime/Weather.xlsx")
Crime <- read_excel("C:/Users/jonat/Desktop/Boston Crime/Crime.xlsx")
```

The first issue that arose occurred as a consequence of the data fields the team intended to combine the sets by. Both data sets had a 'date' column that was determined to be the ideal column to combine by, but upon executing the *merge.data.frame()* command, an error occurred conveying that the 'Weather$Date' column was structured in a 'Date' format and the 'Crime$OCCURRED_ON_DATE' column was structured in the 'POSIXct' format, meaning that the team could not merge the two data frames without formatting.

To remedy, the following code was executed.

```
Weather$Date <- as.Date(as.POSIXct(Weather$Date,format='%Y-/%m-/%d %H:%M:%S'))
Crime$OCCURRED_ON_DATE_NOTIME <-
as.Date(as.POSIXct(Crime$OCCURRED_ON_DATE,format='%m/%d/%Y %H:%M:%S %p'))
Final <- merge.data.frame(Crime, Weather, by.x="OCCURRED_ON_DATE_NOTIME", by.y="Date",
all.x = TRUE)
```

This code did the following:

1) Changed the 'Weather$Date' to a POSIXct type to match the 'Crime$OCCURRED_ON_DATE' and then back to a 'Date' type to match the intended formatting for analysis.
2) Created a new column in the 'Crime' data frame and formatted the 'Crime$OCCURRED_ON_DATE' to a format that matched the format that the 'Weather$Date'.

[2] https://www.ncdc.noaa.gov/cdo-web/

3) Created a new data frame 'Final' that merged 'Weather' and 'Crime' via the new columns formatted in the 'Date' format.

To ensure the 'Final' date frame was set up as intended, a 'str()' command was executed with the results shown in Figure 1.

```
## 'data.frame':    327820 obs. of  24 variables:
##  $ OCCURRED_ON_DATE_NOTIME: Date, format: "2015-06-15" "2015-06-15" ...
##  $ INCIDENT_NUMBER        : chr  "I152049780" "I152049771" "I152049775" "I152049773" ...
##  $ OFFENSE_CODE           : num  3125 3114 801 3502 3501 ...
##  $ OFFENSE_CODE_GROUP     : chr  "Warrant Arrests" "Investigate Property" "Simple Assault" "Missing Person Located" ...
##  $ OFFENSE_DESCRIPTION    : chr  "WARRANT ARREST" "INVESTIGATE PROPERTY" "ASSAULT - SIMPLE" "MISSING PERSON - LOCATED"
...
##  $ DISTRICT               : chr  "D4" "B2" "E13" "A1" ...
##  $ REPORTING_AREA         : num  171 586 304 111 111 624 794 28 741 172 ...
##  $ SHOOTING               : chr  NA NA NA NA ...
##  $ OCCURRED_ON_DATE       : POSIXct, format: "2015-06-15 23:15:00" "2015-06-15 22:19:00" ...
##  $ YEAR                   : num  2015 2015 2015 2015 2015 ...
##  $ MONTH                  : num  6 6 6 6 6 6 6 6 6 6 ...
##  $ DAY_OF_WEEK            : chr  "Monday" "Monday" "Monday" "Monday" ...
##  $ HOUR                   : num  23 22 22 22 22 22 22 22 21 21 ...
##  $ UCR_PART               : chr  "Part Three" "Part Three" "Part Two" "Part Three" ...
##  $ STREET                 : chr  "HARRISON AVE" "TERRACE ST" "BRAGDON ST" "TREMONT ST" ...
##  $ Lat                    : num  42.3 42.3 42.3 42.4 42.4 ...
##  $ Long                   : num  -71.1 -71.1 -71.1 -71.1 -71.1 ...
##  $ Location               : chr  "(42.33555954, -71.07436364)" "(42.32746198, -71.09852527)" "(42.31731834, -71.0967899
2)" "(42.35428377, -71.06380404)" ...
##  $ Name                   : chr  "BOSTON, MA US" "BOSTON, MA US" "BOSTON, MA US" "BOSTON, MA US" ...
##  $ Prcp                   : num  0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 ...
##  $ Snow                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ AvgTemp                : num  58 58 58 58 58 58 58 58 58 58 ...
##  $ MaxTemp                : num  63 63 63 63 63 63 63 63 63 63 ...
##  $ MinTemp                : num  54 54 54 54 54 54 54 54 54 54 ...
```

Figure 1. Structure of the 'Final' Data Frame

Prepping the Data Conclusion

The data initially imported in a very clean nature. Minimal to no data munging was required and the only coding required to finish prepping the data for analysis was the formatting of dates to ensure combination of data frames could occur.

Analysis

To scope the project, the team looked at 10 separate questions the team believed would be beneficial to look into. Of those 10 questions, some were combined or removed near the end of the term as the team progressed through the class. The final five questions were:

1) Has there been an increase of crime from 2015-2018?
2) Do particular crimes occur more frequently at a certain time, day, week, or month?
3) Do warrants act as precursors for crime?
4) Is there a difference between the neighborhoods?
5) Is temperature a predictor of crime?

To perform basic statistical analysis of the data frame, two methods were utilized. First, using the *summary()* function of R, the following information was processed.

```
##  OCCURRED_ON_DATE_NOTIME INCIDENT_NUMBER     OFFENSE_CODE
##  Min.   :2015-06-15      Length:327820      Min.   : 111
##  1st Qu.:2016-04-20      Class :character   1st Qu.:1001
##  Median :2017-02-14      Mode  :character   Median :2907
##  Mean   :2017-02-09                         Mean   :2318
##  3rd Qu.:2017-11-30                         3rd Qu.:3201
##  Max.   :2018-10-03                         Max.   :3831
##
##  OFFENSE_CODE_GROUP OFFENSE_DESCRIPTION   DISTRICT         REPORTING_AREA
##  Length:327820      Length:327820       Length:327820      Min.   :  0.0
##  Class :character   Class :character    Class :character   1st Qu.:177.0
##  Mode  :character   Mode  :character    Mode  :character   Median :343.0
##                                                           Mean   :383.2
##                                                           3rd Qu.:544.0
##                                                           Max.   :962.0
##                                                           NA's   :20920
##    SHOOTING         OCCURRED_ON_DATE                 YEAR
##  Length:327820      Min.   :2015-06-15 00:00:00   Min.   :2015
##  Class :character   1st Qu.:2016-04-20 09:43:45   1st Qu.:2016
##  Mode  :character   Median :2017-02-14 15:49:00   Median :2017
##                     Mean   :2017-02-10 07:26:53   Mean   :2017
##                     3rd Qu.:2017-11-30 18:23:45   3rd Qu.:2017
##                     Max.   :2018-10-03 20:49:00   Max.   :2018
##
##      MONTH        DAY_OF_WEEK            HOUR         UCR_PART
##  Min.   : 1.000   Length:327820      Min.   : 0.00   Length:327820
##  1st Qu.: 4.000   Class :character   1st Qu.: 9.00   Class :character
##  Median : 7.000   Mode  :character   Median :14.00   Mode  :character
##  Mean   : 6.672                      Mean   :13.11
##  3rd Qu.: 9.000                      3rd Qu.:18.00
##  Max.   :12.000                      Max.   :23.00
##
##     STREET               Lat             Long          Location
##  Length:327820      Min.   :-1.00   Min.   :-71.18   Length:327820
##  Class :character   1st Qu.:42.30   1st Qu.:-71.10   Class :character
##  Mode  :character   Median :42.33   Median :-71.08   Mode  :character
##                     Mean   :42.21   Mean   :-70.91
##                     3rd Qu.:42.35   3rd Qu.:-71.06
##                     Max.   :42.40   Max.   : -1.00
##                     NA's   :20632   NA's   :20632
##      Name                Prcp             Snow            AvgTemp
##  Length:327820      Min.   :0.0000   Min.   : 0.0000   Min.   : 0.00
##  Class :character   1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.:42.00
##  Mode  :character   Median :0.0000   Median : 0.0000   Median :57.00
##                     Mean   :0.1073   Mean   : 0.1009   Mean   :55.63
##                     3rd Qu.:0.0500   3rd Qu.: 0.0000   3rd Qu.:71.00
##                     Max.   :2.6800   Max.   :14.5000   Max.   :89.00
##
##     MaxTemp         MinTemp        MonthName
##  Min.   :12.00   Min.   :-9.00   Aug    : 35137
##  1st Qu.:49.00   1st Qu.:36.00   Jul    : 34640
##  Median :65.00   Median :50.00   Sep    : 34023
##  Mean   :63.55   Mean   :48.41   Jun    : 30622
##  3rd Qu.:79.00   3rd Qu.:63.00   Oct    : 26437
##  Max.   :98.00   Max.   :81.00   May    : 26242
##                                  (Other):140719
```

Figure 2. Summary(Final) Output in R

As shown in Figure 2, minus the weather parameters, the remaining calculations do not provide much insight into the data. To gain insight into basic descriptive statistics of different parameters (e.g. average crime rate per hour of the day), a function 'DescStat' was written to output the desired analysis.

```
DescStat <- function(inputVector)
{
 cat("mean:",mean(inputVector))
 cat("\n")
 cat("median:",median(inputVector))
 cat("\n")
 cat("min:",min(inputVector))
 cat("\n")
 cat("max:",max(inputVector))
 cat("\n")
 cat("Standard Deviation:",sd(inputVector))
 cat("\n")
 cat("Quantile (0.05 - 0.95):", quantile(inputVector, probs = c(0.05,0.95)))
 cat("\n")
 cat("Skewness:", skewness(inputVector, na.rm = TRUE))
}
```

The following analysis was completed to determine the frequency of crime given the hour of day, day of week, month of the year, and the year for the combined data set.

```
## mean: 13659.17
## median: 15345.5
## min: 3409
## max: 21350
## Standard Deviation: 5597.182
## Quantile (0.05 - 0.95): 3685.75 20827.75
```

Figure 3.Basic Stats on Hourly (Total) Crime Rate

```
## mean: 46831.43
## median: 47726
## min: 41374
## max: 49758
## Standard Deviation: 2668.821
## Quantile (0.05 - 0.95): 42752.5 49275.9
```

Figure 4. Basic Stats on Daily (Total) Crime Rate

```
## mean: 27318.33
## median: 25199
## min: 21661
## max: 35137
## Standard Deviation: 4921.163
## Quantile (0.05 - 0.95): 22663.65 34863.65
```

Figure 5. Basic Stats on Monthly (Total) Crime Rate

```
## mean: 81955
## median: 86745
## min: 53392
## max: 100938
## Standard Deviation: 22576.26
## Quantile (0.05 - 0.95): 56536.6 100667.4
```

Figure 6. Basic Stats on Yearly (Total) Crime Rate

Given the analysis now completed, the team was able to start drawing conclusions on aspects of the entire dataset from 2015-2018.

1) Crime ranged from 3,409 occurrences to 21,350 given any hour of the day with an average of 13,659 occurrences (*Figure 3*).
2) Any given day of the week had 41,374 to 49,758 occurrences of crime with a standard deviation of only 2,668 occurrences from the mean of 46,381 occurrences (*Figure 4*).
3) Per month, 27,318 occurrences of crime on average would be recorded with high months reaching 35,137 occurrences and low months only seeing 21,661 occurrences (*Figure 5*).
4) A yearly average of 81,955 occurrences was recorded (*Figure 6*).

Has Crime increased from 2015-2018?

To determine if crime has increased from 2015-2018, the following calculates were performed using the *DescStat()* function. Since the only complete years in our dataset was 2016 and 2017, we choose to utilize the average occurrences per day versus the frequency of crime per year.

Table 3. Average Occurrences of Crime Per Day

Year	Frequency of Crime	Number of Days	Average Occurrences Per Day
2015	53,392	199	268.3
2016	99,134	365	271.6
2017	100,938	365	276.5
2018	74,356	275	270.4

Table 4. Average Occurrence Per Day Analysis

Parameter	Value
Mean	271.6
Median	273.0
Minimum	140.0
Maximum	379.0
Standard Deviation	33.0
Quantile (0.05-0.95)	216.3 | 321.0

Analysis Conclusion

The range of crime given the hour of day, day of week, month of the year, and given year represents a monumental logistical problem. According to the City of Boston's website[3], the expenditures of the Police Department have increased from FY17-FY18 and the budget projection shows an increase for FY19-FY20 (Figure 7).

[3] https://www.boston.gov/departments/budget/fy20-operating-budget

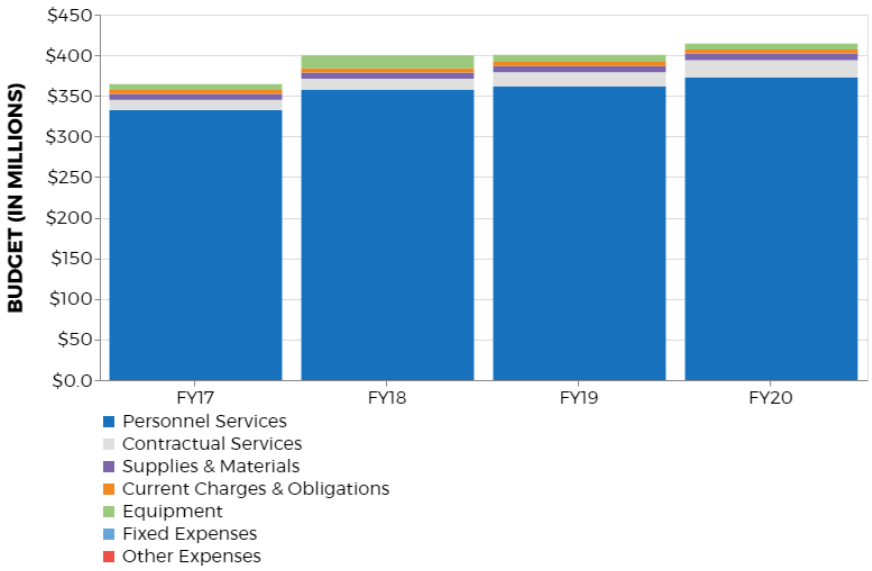

Figure 7. Boston FY17-FY18 Expenditures and FY19-FY20 Budgets

This increase in budget shows that the city of Boston understands they have a crime rate problem and now have the ability to use increase funding to not only increase the force, but to increase the level of Data Analytics they use to predict crime given historical averages.

Visualizations

The intent of this section is to show a summary of the different style graphs we used to visualize our initial questions. Not all graphs were used to answer the final five questions, but rather were used to understand the data. Additional graphs are included in the Rmarkdown file.

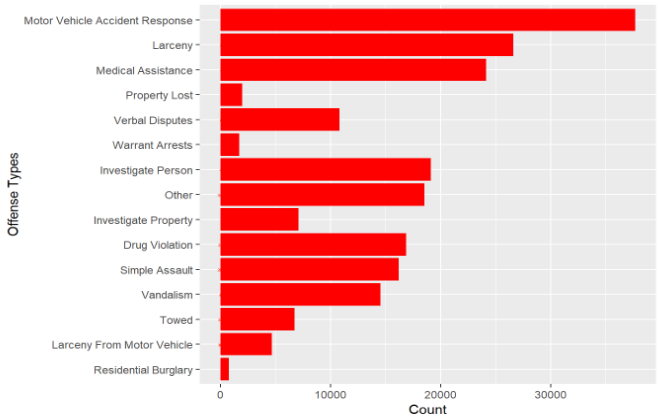

Figure 8. Summary Total of Offenses

Figure 9. Offense Type Summarized by District

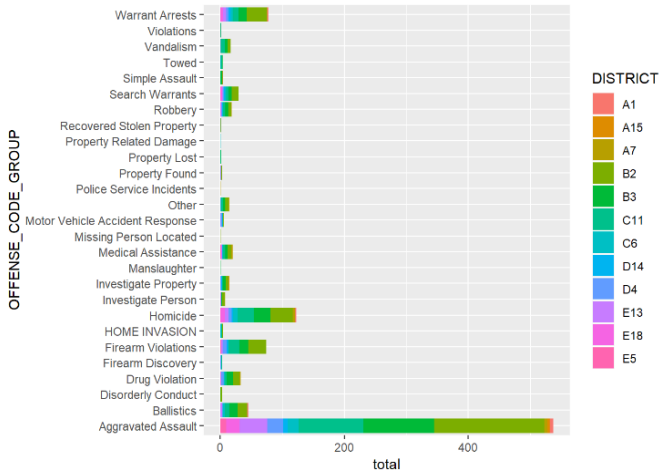

Figure 10. Offense Type Summarized by District Where a Shooting Occurred

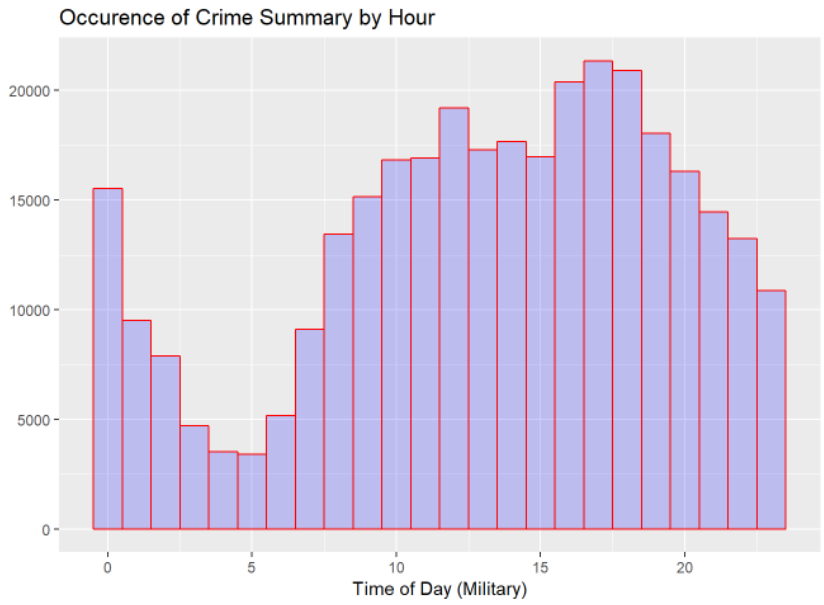

Figure 11. Histogram of Crime Rates by Time of Day

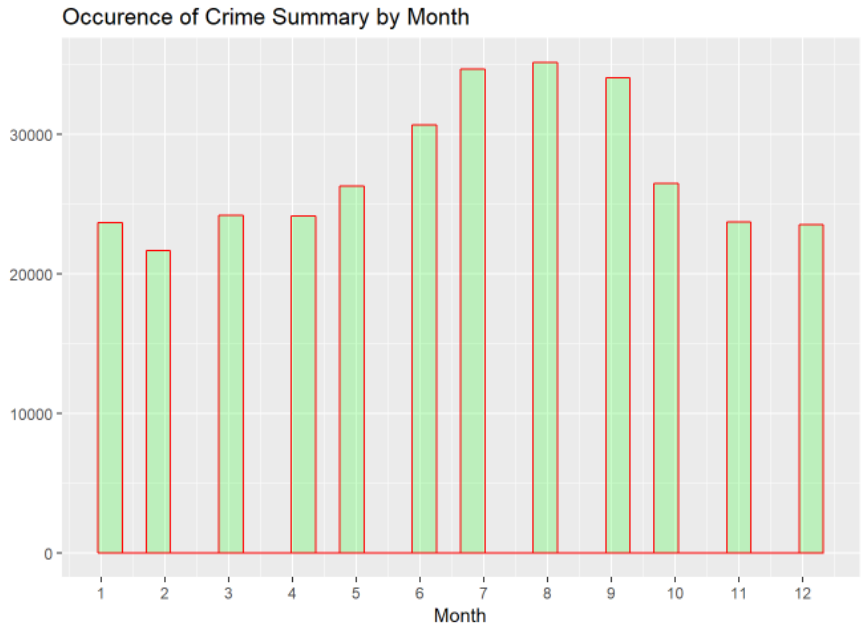

Figure 12. Histogram of Crime Rates by Month

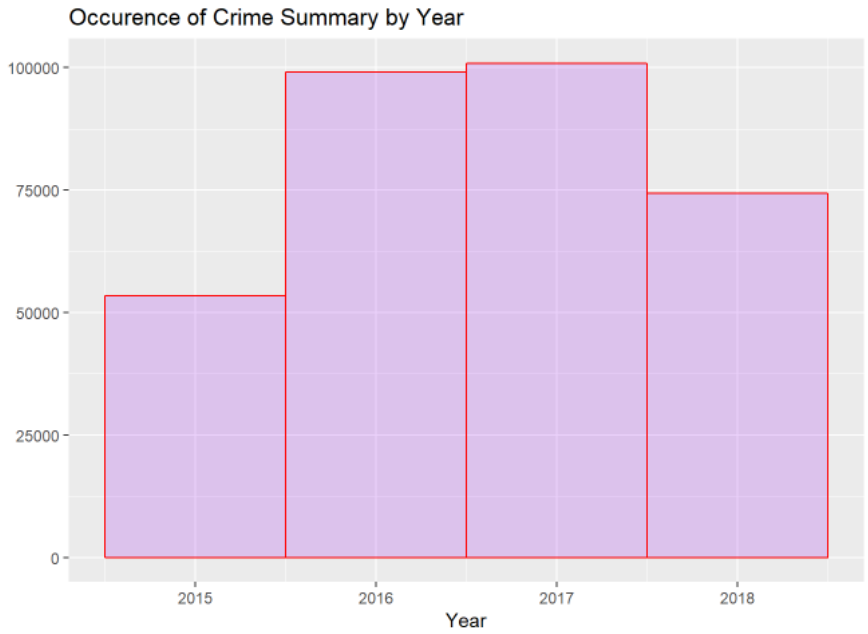

Figure 13. Histogram of Crime Rates by Year

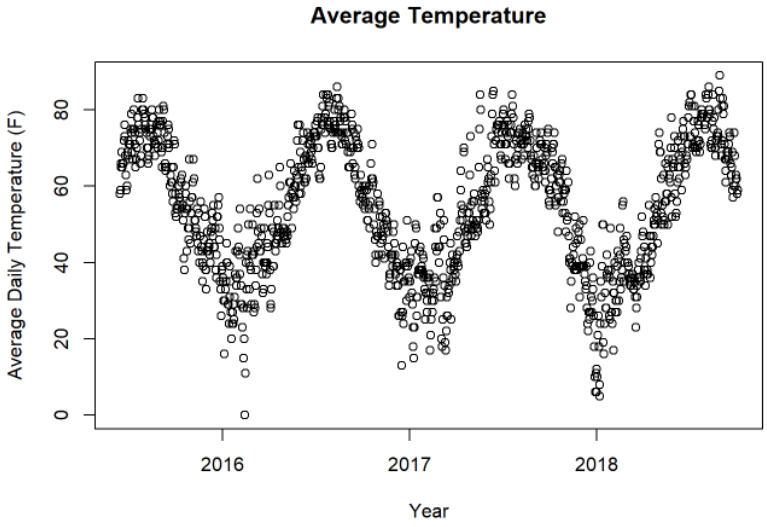

Figure 14. Summary of Average Temperature by Year

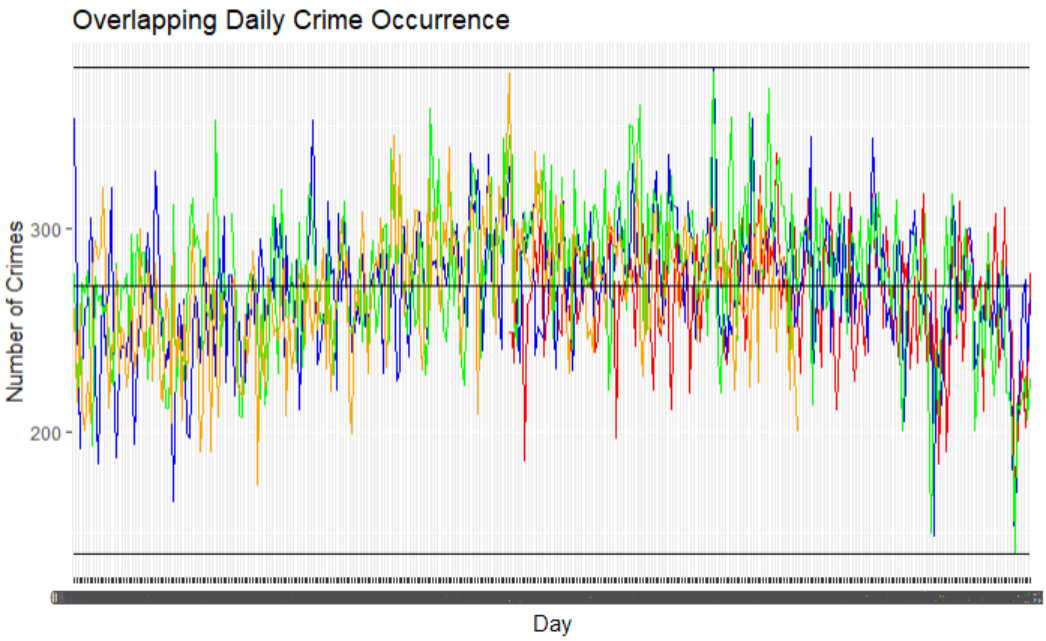

Figure 15. Daily Crime Occurrence Plot

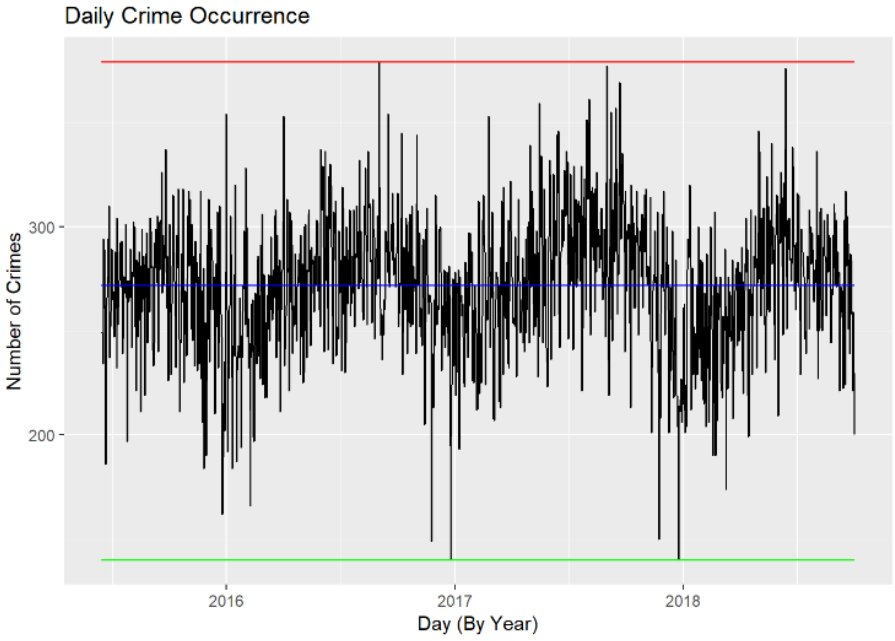

Figure 16. Daily Crime Occurrence Grouping by Year

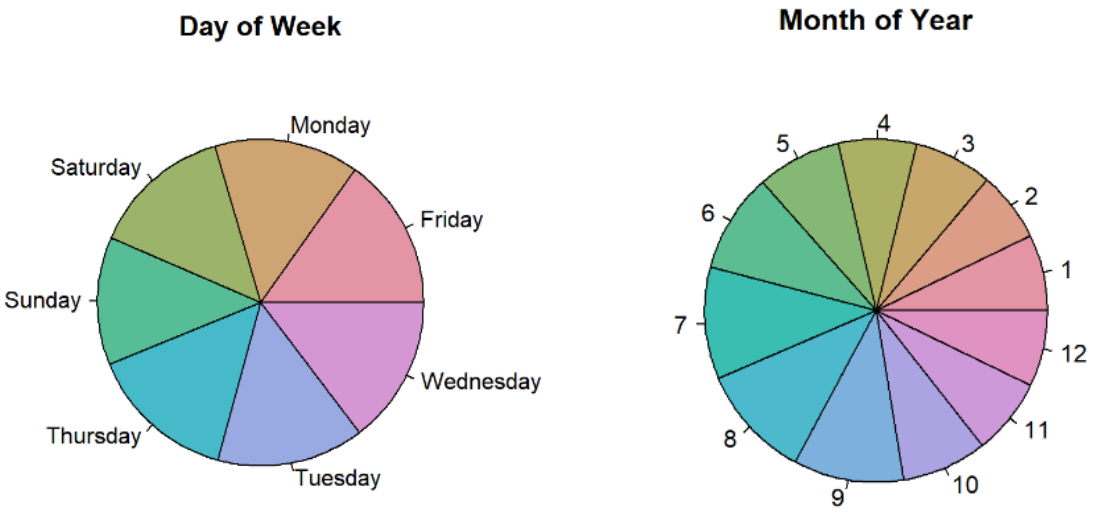

Figure 17. Pie Charts of Crime Breakdown by Day and Month

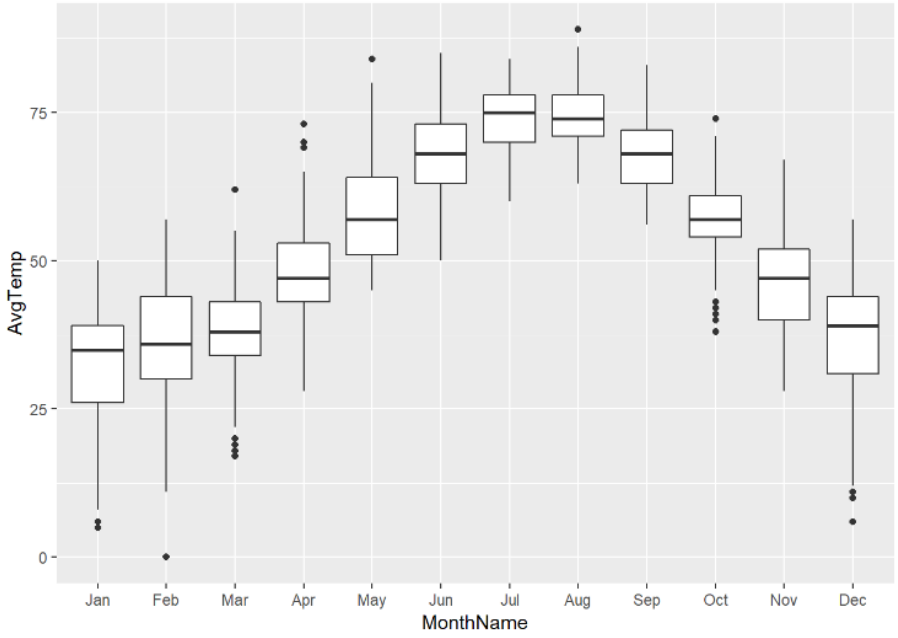

Figure 18. Boxplot of Month vs Average Temperature

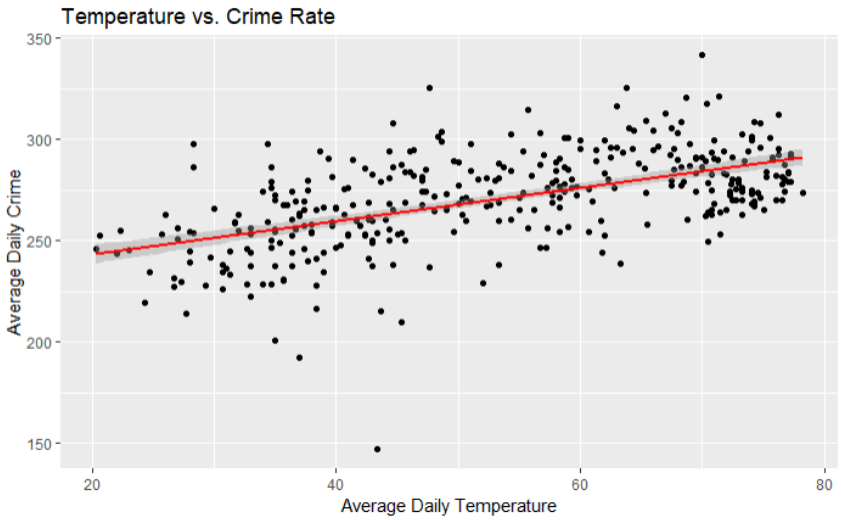

Figure 19. Temperature vs. Crime Rate

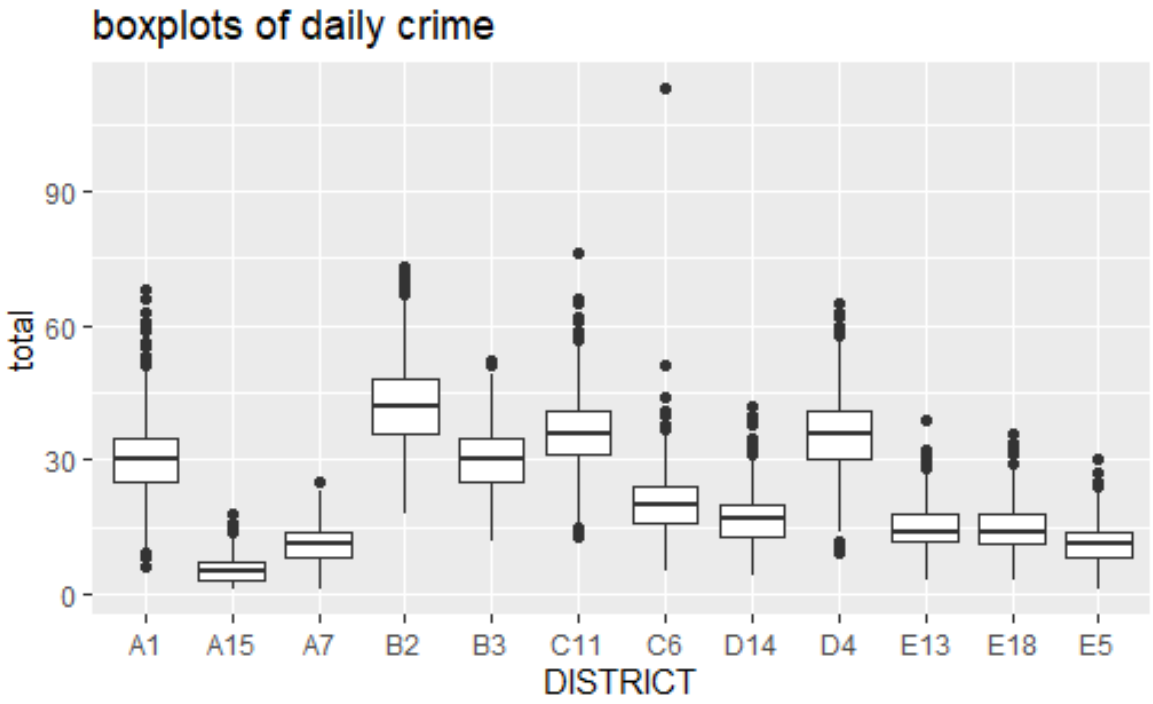

Figure 20. Boxplots of daily crime by district

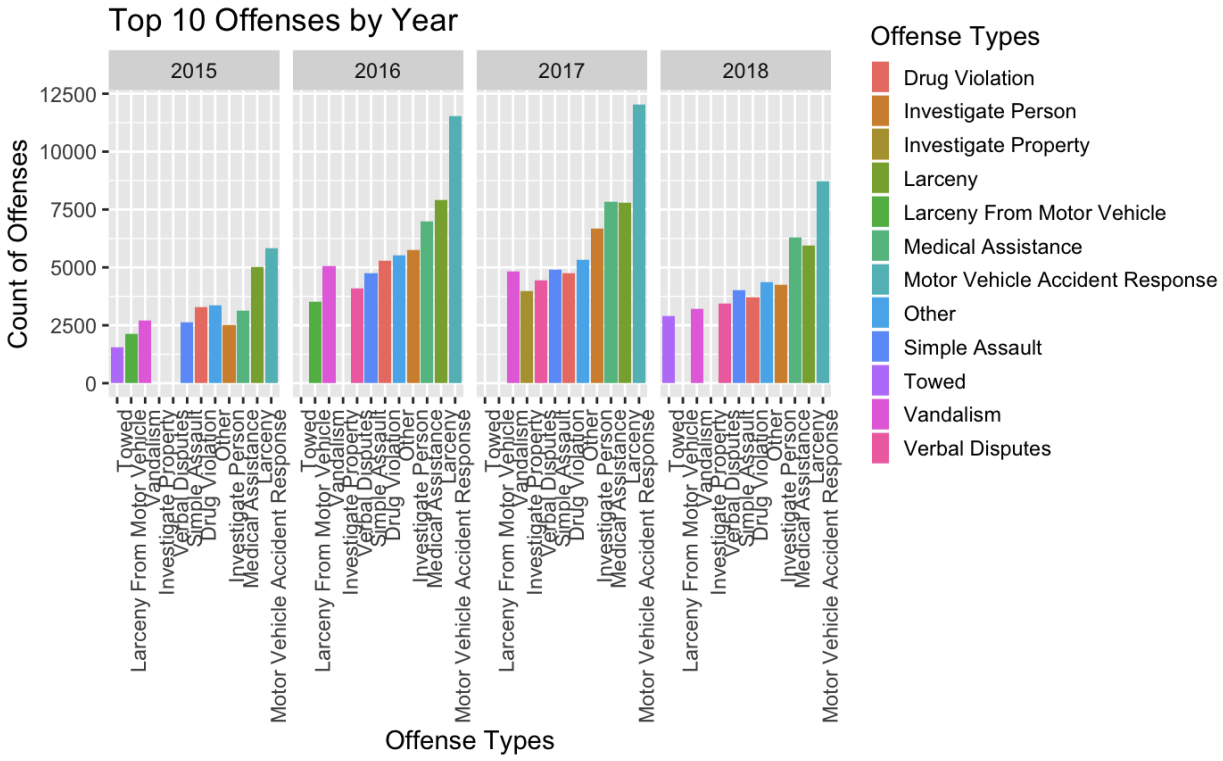

Figure 21. Top 10 crimes by year

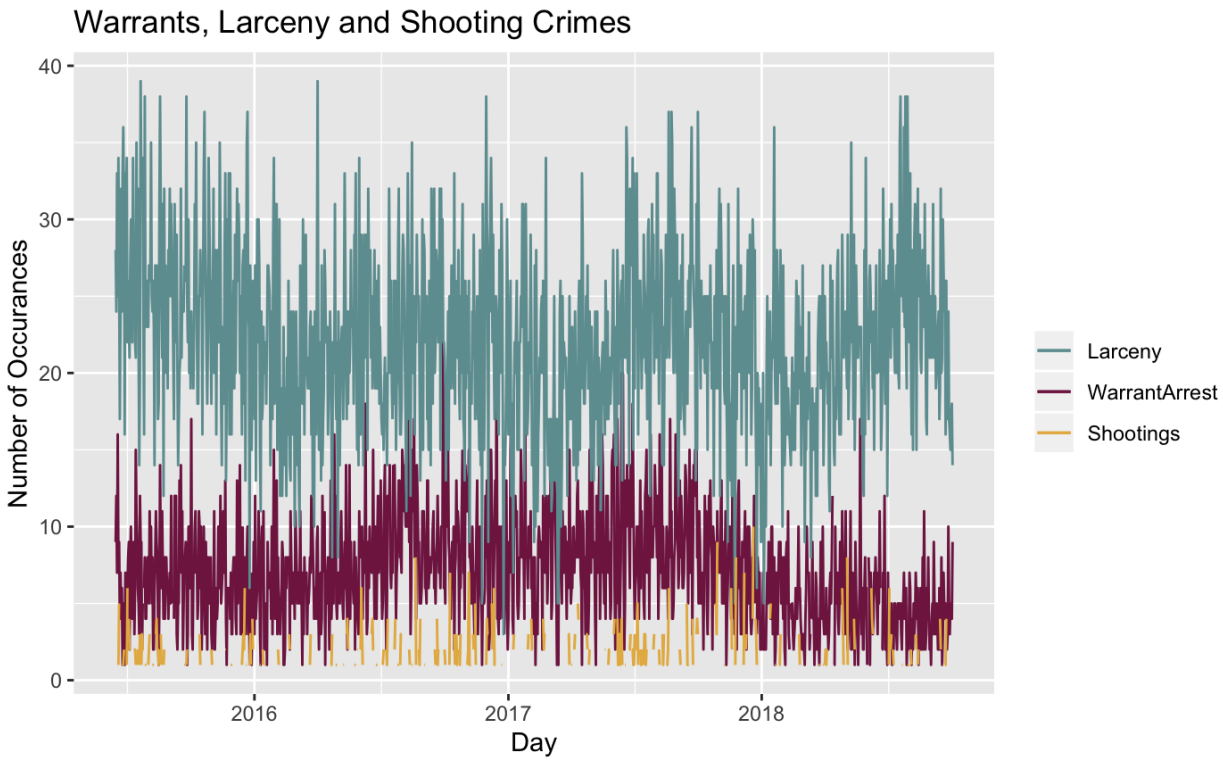

Figure 22. Number of occurrences of warrant, larceny and shooting crimes

Modeling

In order to answer the questions presented, the group needed to conduct basic modeling to identify trends that could be used to solve the problems presented in the summary.

Daily Temperature vs. Daily Crime Rate

To begin, the *lm()* function in R was used to calculate if there was any relationship between daily average temperature and daily crime rate.

```
model_avgday <- lm(formula=overlapday$Average~overlapday$Average_Temp,
data=overlapday)
summary(model_avgday)
plot(model_avgday)
plot(overlapday$Average~overlapday$Average_Temp, data=overlapday)
avgdayplot <- ggplot(overlapday,aes(overlapday$Average_Temp, y = overlapday$Average)) +
geom_point() + stat_smooth(method="lm", color="red") + labs(x="Average Daily Temperature",
y="Average Daily Crime") + ggtitle("Temperature vs. Crime Rate")
avgdayplot
```

The formula returned from R is shown below.

```
Call:
lm(formula = overlapday$Average ~ overlapday$Average_Temp, data = overlapday)

Residuals:
     Min       1Q   Median       3Q      Max
-115.218  -12.046    0.174   11.955   59.212

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              226.85893    3.57241   63.50   <2e-16 ***
overlapday$Average_Temp    0.82368    0.06459   12.75   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.37 on 364 degrees of freedom
Multiple R-squared:  0.3088,    Adjusted R-squared:  0.3069
F-statistic: 162.6 on 1 and 364 DF,  p-value: < 2.2e-16
```

Given the output, the following can be interpolated from the formula.

1) As temperature increases by 1 degree, daily total occurrences of crimes increases by 0.82368.
2) The p-value = <2e-16 and is therefore less than 0.5 which represents a statistically significant value.
3) The equation accounts for 30.7% of y values given the adjusted R-squared.

The *plot()* function shows some of the large residuals though involved in the model.

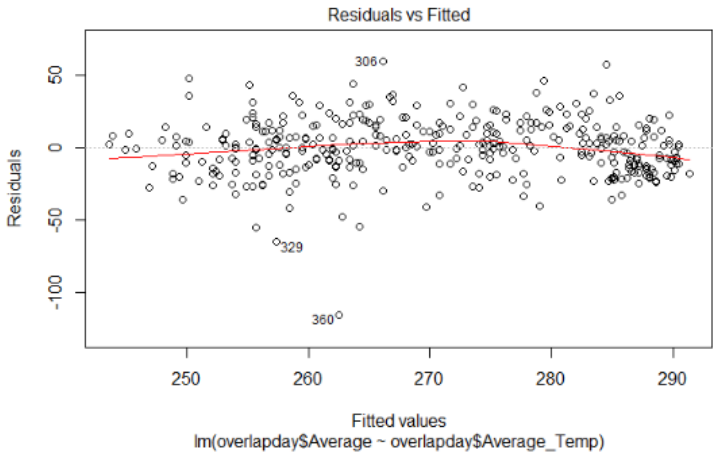

Figure 23. Residuals vs. Fitted for Daily Temperature vs. Daily Occurrences of Crime

Plotting the formula over the data, Figure 23, shows there appears to be a positive linear relationship indicating that as temperature increases, so does the total number of crimes committed daily.

Daily Temperature and Precipitation vs. Daily Crime Rate

Due to the low Adjusted R^2 value above (0.3088), precipitation totals for the day was added.

```
Call:
lm(formula = overlapday$Average ~ overlapday$Average_Temp + overlapday$Average_Prcp,
    data = overlapday)

Residuals:
     Min       1Q   Median       3Q      Max
-114.752  -12.260    0.369   11.477   57.185

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             229.53767    3.65413   62.82  < 2e-16 ***
overlapday$Average_Temp   0.81001    0.06411   12.63  < 2e-16 ***
overlapday$Average_Prcp -17.40434    5.98032   -2.91  0.00383 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.18 on 363 degrees of freedom
Multiple R-squared:  0.3246,    Adjusted R-squared:  0.3209
F-statistic: 87.22 on 2 and 363 DF,  p-value: < 2.2e-16
```

As shown above, the Adjusted R-squared value did increase to 0.3209, which is still very low. Again, due to the amount of multi linearity that is involved in the data (multiple y values to a singular x value), the low Adjusted R-squared does not surprise.

Modeling Conclusions

- Modeling was conducted on a limited basis for this project given the late introduction of the concept to the syllabus. Regardless, the *lm()* proved useful in determining if a relationship existed between temperature and crime.
- A low Adjusted R-squared value was seen when crime was predicted a function of temperature (0.3069). An increase in adjusted R-squared was seen when adding in precipitation (0.3209).

- Backed up by visualizations, there appears to be a statistically significant relationship showing a positive linear relationship between temperature, precipitation, and crime.

Answering the Questions

Has there been an increase of crime from 2015-2018?

- To answer this question, descriptive statistics and visualizations were used.
- Due to incomplete year sets, a weighted average was calculated per year (Table 3) and was added to 'year' data frame.
- Executing the 'DescStat()', the following analysis was done on the 'Average Occurrences Per Day' with the following results shown in Table 4.
- To visualize this data, a plot was created showing the number of crimes recorded each day and the min (green), mean (blue), and max (red) values recorded (Figure 16).
- As shown by the descriptive statistics, the weighted average of daily crime occurrence all fall with 1.8% of the mean value and represent only 15% of the 33.0 standard deviation data. A key distinction is that we are not evaluating a specific type of crime for reduction (i.e. targeting effort to decrease assault, speeding, etc.), but rather overall crime rates were analyzed for this question. Therefore, even though budget increased for Boston every year, there appears to be no significant statistical evidence that the overall frequency of crime has decreased over time.

Does crime happen during a certain time of day, day of week, season, or time of year?

- To answer this question, basic visual analysis was performed; depending on if there was likely a difference based on the visual, 2 sample t tests were performed to determine if a particular day, week, or season differs from the sample subset of data.
- Boxplots were created to visualize the data. Plots were made for each day and month. There was insufficient evidence to require further examination via t tests.

Do warrants act as a precursor to a specific crime?

- First we looked at the top 10 crimes by year to see if there was any visual shifts in the frequency of a specific type of crime from year to year, Figure 21.
- We can see from Figure 8, the frequency of warrant arrest pales in comparison to the other top offenses across districts.
- Next we counted how many warrant arrests, larceny and crimes that involve a shooting occurred each day over the 4-year period. We graphed these using a line chart to see if there was any pattern to each of these crimes that would indicate warrant arrest was a precursor to larceny or crimes that involved shooting.
- We can see in Figure 22, that the shapes of the three graphs are very similar, however not exact. You can see that towards the end of 2018, there is a noticeable spike in larceny, where there isn't much change in the natural variation of warrant arrest. We came to the conclusion that having more than 2 full years of data would be required to analyze if a lesser crime could be used to determine a more severe crime, as it would allow us to determine more of a pattern.

Does temperature correlate to a reduction or increase of crime?

- To determine if temperature correlates to crime, descriptive statistics, visualizations, and modeling was used.

- To conduct descriptive statistics, Figure 2 and Table 4 shows a break down of how temperature and crime varies on a daily basis. This break down provided the understanding to how best to visualize the data.
- For visualizations, Figure 14 shows a breakdown of temperature from 2015-2018, while Figure 16 shows a breakdown of Daily Crime Occurrence from 2015-2018. On a very high level, it appeared the two were correlated by seasons.
- Given an apparent relationship from visualizations, the *lm()* function in R was used to examine if there was a correlation. Statistically, it appeared a relationship existed given the small p-value, but the overall Adjusted R-Squared was only 0.3069. This is as expected though, given that for any given temperature, multiple occurrences of crime paired with a given temperature, as shown in Figure 19. Regardless, Figure 19 appears to show a positive linear relationship between temperature of crime. This logically makes sense given that citizens are generally more active given higher temperatures.
- With the addition of precipitation, an increase in Adjusted R-squared was observed.
- In summary, a relationship appears to exist between temperature + precipitation and daily occurrences of crime.

Is there a statistical difference between the neighborhoods and does it change over time?

- To answer this question, boxplots of crimes per day were plotted. We looked at the medians and quartiles of the different districts. See Figure 20.
- Districts B2, C11, and D4 had higher than average crime when compared to the other districts in Boston. When compared to districts on the low end of the spectrum such as A15, A7, and E5 there was a statistical difference between these districts. This is particularly evident when analyzing crimes that involved a shooting. The aforementioned districts with low crime rates had particularly low amounts of shootings. The upper quartiles for the high crime districts dwarfs that of the low crime districts.
- That is, when performing a 2-sample t test on these districts, the alpha value was less than .05 on a 2-sided test. Statistically speaking, one could say that the districts on the low end come from a different sample than the districts on the low end.
- To answer the second portion of the question, does it change over time, plots were made with the total number of crimes per day on the y axis and the days on the x axis. Each of the districts had a pretty steady amount of total crime. The steepest increase occurred in district B3 where an increase of about 2 crimes per day was observed. This is not a statistically significant difference.

References

[1]	"Boston Police Department," Wikipdeia, 06 08 2019. [Online]. Available: https://en.wikipedia.org/wiki/Boston_Police_Department. [Accessed 07 08 2019].