Northeastern University

REPORT

Boston: Is it safe?

Course ALY6015 : **Intermediate Analytics**

CRN: 80797

Term: Winter 2019 Quarter

Student: Manpreet Kaur

Submitted to: Valeriy Shevchenko

Dated: May 14, 2019

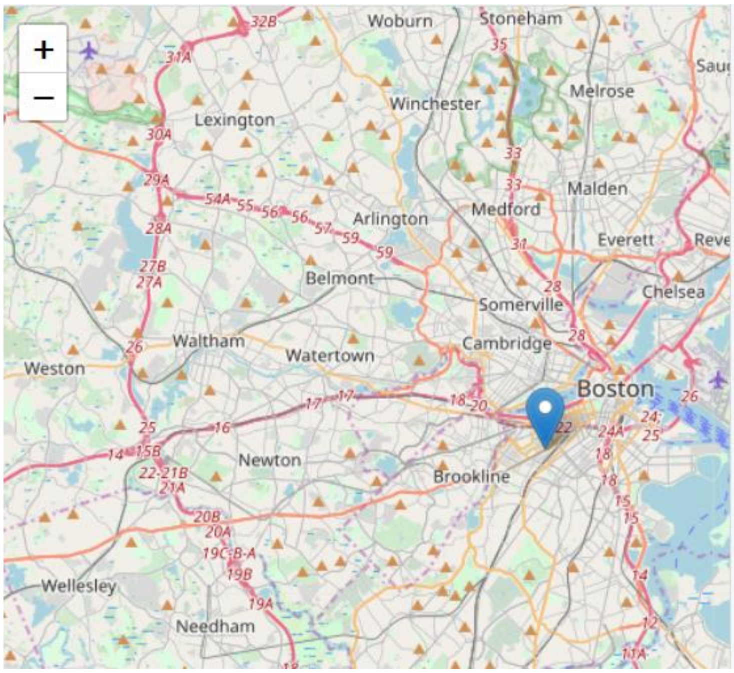
Abstract

We started the project wanting to know that this Boston city, although we know is greater than most cities in the US, how safe it was. Is it true to say that Boston was more violent than New York and Seattle, but less violent than Chicago and Las Vegas, according to numbers from the FBI, based on crimes committed back in 2015.As of 12/21/18 Nationally, Boston ranked 14 out of 50 according to Us News. Our goal was to dig into the dataset of Boston Crimes, collected from Kaggle and analyze the data to throw light about the crimes expected in the year 2019. Also, we did our analyses regarding the crime rates during weekends and weekdays. A lot of other interesting analysis will be presented through this report.

Exploratory Data Analysis

Displaying the code of location Boston through leaflet package by using the following code:

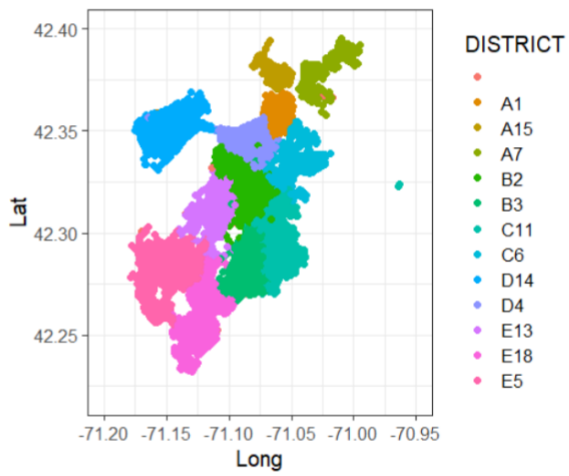
```
#displaying the map of boston through leaflet package
leaflet() %>%
  setView(lng=-71.0892, lat=42.3398, zoom = 10) %>%  #using setView method to set the view of map (center and zoom level)
  addTiles() %>%
  addMarkers(lng=-71.0892, lat=42.3398, popup="Boston")
```



Following is the map of Boston distributed among crime districts. We used the following code snippet for the same.

```
#displaying the map of boston distributed among crime district by latitude and longitude
#crime mapping
qplot(Long, Lat, data= pdata, color=DISTRICT, geom='point', xlim = c(-71.2,-70.95), ylim= c(42.22,42.4))+
  theme_bw(base_size=15)+
  geom_point(size = 2)
```

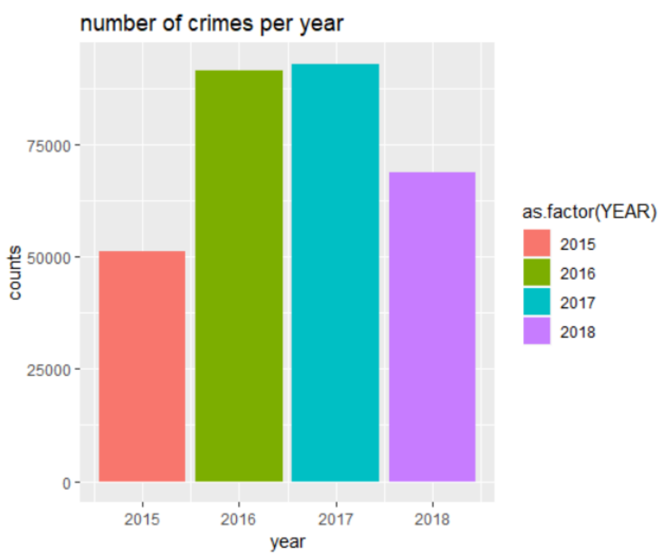
We get the following output:



Now, we are displaying the crimes count per year by using the following code:

```
#number of crimes count per year
ggplot(pdata, aes(x=YEAR, fill= as.factor(YEAR))) +
  labs(x="year", y="counts", title = "number of crimes per year")+
  geom_bar()
```

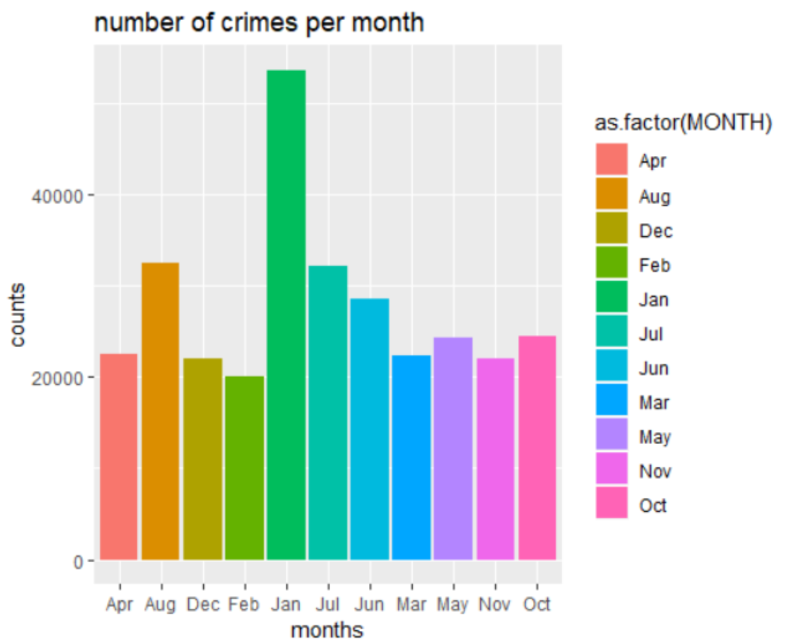
We get the following output:



As we can see the crime rate has increased from 2015 and is highest in 2017. Fortunately, 2018 has experienced less crime rate compared to the previous 2 years.

Further, we are displaying the crimes count per month by using the following code:

```
#number of crimes count per month
ggplot(pdata, aes(x=MONTH, fill= as.factor(MONTH))) +
  labs(x="months", y="counts", title = "number of crimes per month")+
  geom_bar()
```

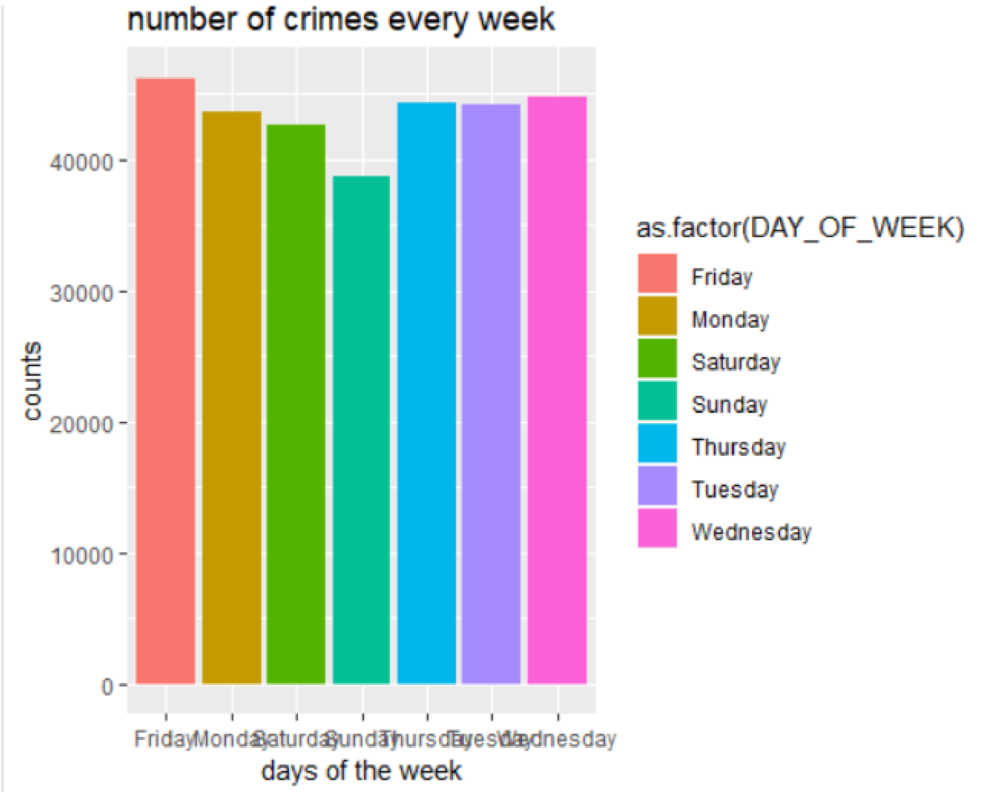


From the output, we can say that the maximum crimes occurred was in the month of January. May be it's because of the new year. And maximum public or tourist come around.

Further, we are displaying the crimes count per week by using the following code:

```
#number of crimes count per day of the week
ggplot(pdata, aes(x=DAY_OF_WEEK, fill= as.factor(DAY_OF_WEEK))) +
  labs(x="days of the week", y="counts", title = "number of crimes every week")+
  geom_bar()
```

We get the following output:



From the graph, it’s obvious to note that on Friday maximum crime occurs. The reason is inevitable that it's weekend and maximum public roam around outside.

Further, displaying the crimes as per hours shift, on which time the maximum crimes occurred.

For that, we divided the 24 hours time slot into 4 parts each of 6 hours by using the following code:

```
#dividing the shift into 4 groups and generating six points of the day to bin the day into four equal segments
time_diff<- c("0","6","12","18","24") #breaking day into 6 interval period
pdata$time_diff <- cut(pdata$HOUR,
                       breaks = time_diff,
                       labels = c("00-06","06-12","12-18","18-24"),
                       include.lowest = TRUE)
table(pdata$time_diff) #displaying the crime counts as per hour shift
```

we have got the counts of crimes as per hours shift:

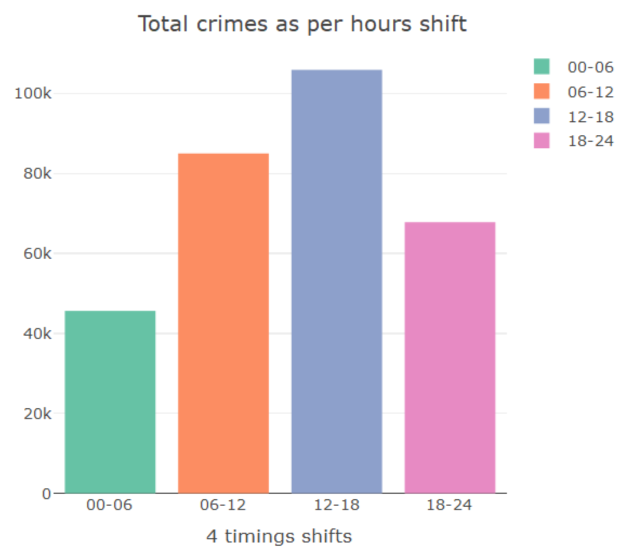
```
> table(pdata$time_diff) #displaying the crime counts as per hour shift

 00-06  06-12  12-18  18-24
 45628  85001 105915  67821
>
```

From the above output we can see that ,from 12:00am to 6:00am 45,628 crimes occured. From 6:00am to 12:00pm 85,001 crimes occured. From 12:00pm to 6:00pm the highest number of crimes occured of 105,915. Lastly in evening from 6:00pm to 12:00 am 67,821 crimes occured.

Next, we are plotting the crimes as per hours shift.

```
plot_shift #displaying the crimes counts as per hours shift

#plotting the crimes according to the days of week
plot_crime_offense_day<- plot_ly(pdata, x= ~ DAY_OF_WEEK, color= ~ DAY_OF_WEEK) %>%
  add_histogram() %>%
  layout(
    title = "Total district count by the crime during the day",
    xaxis = list(title = "Day of week",
                 yaxis = list(title = "Count"))
```

We get the following output:



The highest number of crimes has occured in 12:00pm-6:00pm

Next, we are counting the street crimes and displaying top 10 crimes

```
#counting street crimes
street_crime<- sort(table(pdata$STREET), decreasing = TRUE)
head(street_crime, 10)
```

We get the following output:

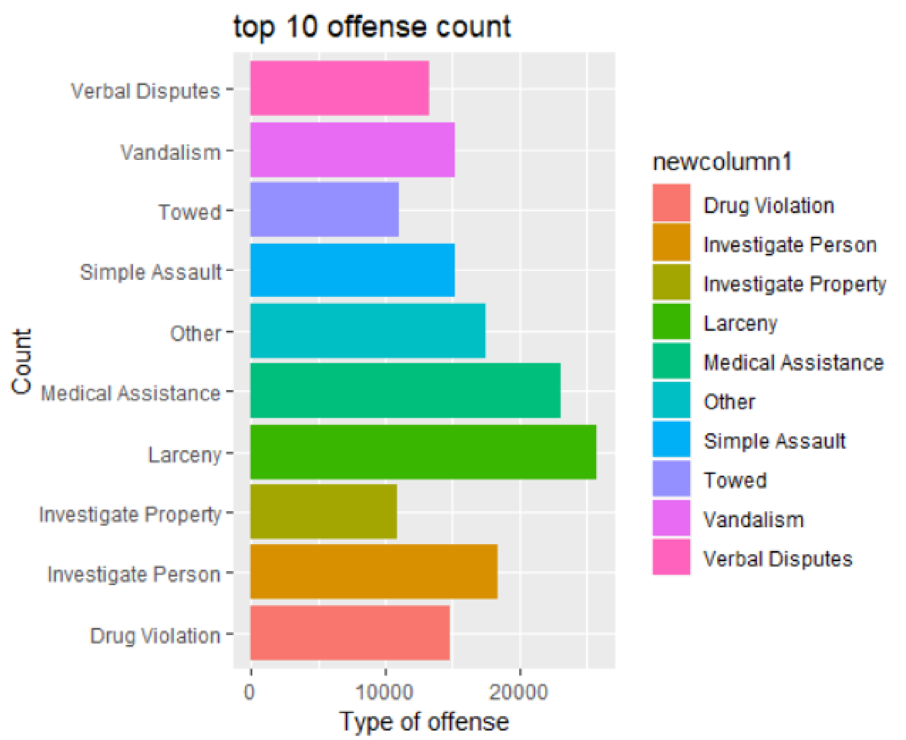
```
> #counting street crimes
> street_crime<- sort(table(pdata$STREET), decreasing = TRUE)
> head(street_crime, 10)

    WASHINGTON ST      BLUE HILL AVE        BOYLSTON ST     DORCHESTER AVE          TREMONT ST MASSACHUSETTS AVE
            14237               7156               7131               5146                4783              4528
     HARRISON AVE          CENTRE ST  COMMONWEALTH AVE      HYDE PARK AVE
             4511               4386               3899               3501
>
```

The highest crime has occured on washington crime of 14,237. The least number of crimes has occured on hyde park ave of 3501.

Further, displaying top 10 offense count by using the following code:

```
#displying top 10 offense count
ggplot(headdata, aes(x=newcolumn1, y=newcolumn, fill= newcolumn1))+
  geom_bar(stat = "identity") +
  coord_flip()+ #flipping the cartesian coordinates
  labs(y = "Type of offense", x = "Count",title ="top 10 offense count")
```

We get the following output:

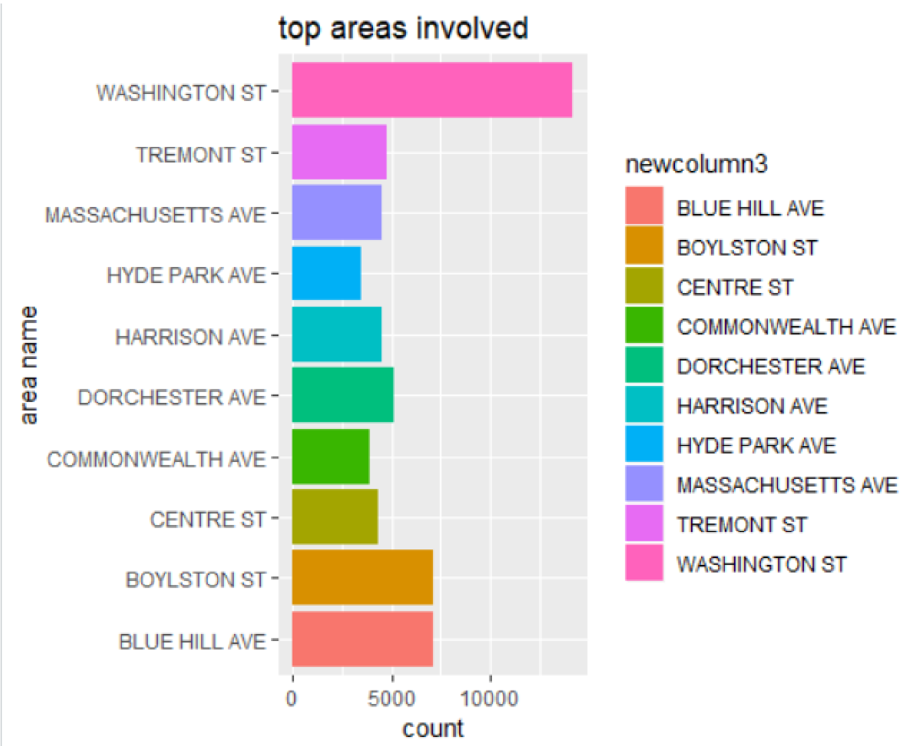


The highest crime that has occured is Larceny of about 21,000. Second highest is Medical Assistance. The least is Vehicle of towing about 11,000.

Further displaying the crime counts of top areas involved by using the following code:

```
#displaying crime counts of top areas involved
ggplot(headdata,aes(x=newcolumn3,y=newcolumn2, fill=newcolumn3))+
  geom_bar(stat="identity")+
  coord_flip()+  #flipping the cartesian coordinates
  labs(y="count", x="area name",title="top areas involved")
```

We get the following output:



As we can see, the washington street tops the lists of more than 15,000 crimes.

The Boylston street and Blue Hill avenue are same with rate of more than 5000 crimes.

The least is Hyde park ave with count of less than 5000 crimes.

Generating word cloud in order to understand and visualize crimes as per streets.

Our EDA supports the fact that the most dangerous street is washington street. Higher the font, more that city is likely to be in danger of crime.

Now, considering 2016 data in order to analyse crime rates:

```
#considering only 2016 data and showing highest offense codes reported
graph1<-filter(pdata,YEAR==2016) #filtering 2016 year offense codes
table(graph1$OFFENSE_CODE_GROUP) #displaying counts of all offense codes
ocg1<-sort(table(graph1$OFFENSE_CODE_GROUP),decreasing = TRUE)[2:11] #taking top 10 offense code group
```

We get the counts of types of crimes occured.

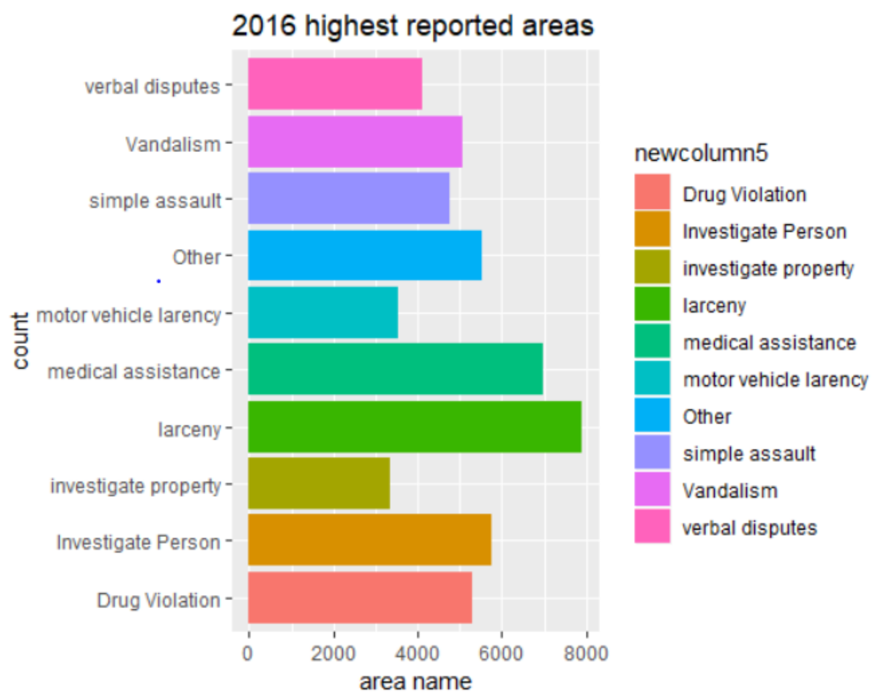
```
                Aggravated Assault                          Aircraft                                     Arson
                              2149                                 4                                        33
    Assembly or Gathering Violations                      Auto Theft                        Auto Theft Recovery
                               312                              1395                                       285
                        Ballistics                 Biological Threat                                 Bomb Hoax
                               283                                 0                                        36
        Burglary - No Property Taken              Commercial Burglary                          Confidence Games
                                 1                               425                                      1050
                    Counterfeiting               Criminal Harassment                         Disorderly Conduct
                               479                                35                                       726
                    Drug Violation                      Embezzlement                              Evading Fare
                              4580                                84                                       124
                        Explosives               Fire Related Reports                         Firearm Discovery
                                 6                               593                                       179
                 Firearm Violations                            Fraud                                  Gambling
                               489                              1768                                         0
                        Harassment           Harbor Related Incidents                             HOME INVASION
                              1351                                25                                        29
                          Homicide                 HUMAN TRAFFICKING HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
                                43                                 2                                         1
                 Investigate Person               INVESTIGATE PERSON                      Investigate Property
                              5509                                 2                                      3131
           Landlord/Tenant Disputes                          Larceny                Larceny From Motor Vehicle
                               294                              7588                                      3275
    License Plate Related Incidents                 License Violation                          Liquor Violation
                               139                               576                                       256
                      Manslaughter                Medical Assistance                     Missing Person Located
                                 4                              6615                                      1670
            Missing Person Reported      Motor Vehicle Accident Response            Offenses Against Child / Family
                              1264                              9307                                       159
         Operating Under the Influence                            Other                            Other Burglary
                               151                              5153                                       132
              Phone Call Complaints          Police Service Incidents                 Prisoner Related Incidents
                                 9                               790                                        66
                    Property Found                     Property Lost                    Property Related Damage
```

```
                 Firearm Violations                            Fraud                                  Gambling
                               489                              1768                                         0
                        Harassment           Harbor Related Incidents                             HOME INVASION
                              1351                                25                                        29
                          Homicide                 HUMAN TRAFFICKING HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
                                43                                 2                                         1
                 Investigate Person               INVESTIGATE PERSON                      Investigate Property
                              5509                                 2                                      3131
           Landlord/Tenant Disputes                          Larceny                Larceny From Motor Vehicle
                               294                              7588                                      3275
    License Plate Related Incidents                 License Violation                          Liquor Violation
                               139                               576                                       256
                      Manslaughter                Medical Assistance                     Missing Person Located
                                 4                              6615                                      1670
            Missing Person Reported      Motor Vehicle Accident Response            Offenses Against Child / Family
                              1264                              9307                                       159
         Operating Under the Influence                            Other                            Other Burglary
                               151                              5153                                       132
              Phone Call Complaints          Police Service Incidents                 Prisoner Related Incidents
                                 9                               790                                        66
                    Property Found                     Property Lost                    Property Related Damage
                              1004                              2699                                       277
                       Prostitution          Recovered Stolen Property                       Residential Burglary
                                62                               387                                      1773
        Restraining Order Violations                          Robbery                           Search Warrants
                               527                              1361                                       286
                           Service                    Simple Assault                                     Towed
                                77                              4413                                      3056
                         Vandalism                   Verbal Disputes                                Violations
                              4840                              4041                                      1523
                   Warrant Arrests
                              2530
```

Displaying the 2016 crimes of highest reported areas:

```
#generating the bar chart of 2016 highest reported areas
ggplot(headdata,aes(x=newcolumn5,y=newcolumn4, fill=newcolumn5))+
  geom_bar(stat="identity")+
  coord_flip()+ #flipping the cartesian coordinates
  labs(y="area name", x="count",title="2016 highest reported areas")

sum(headdata$newcolumn4) #displaying the total sum of top 10 2016 offense codes

#2017
```

We get the following output:



As we can see that larceny tops the list amongst highest crimes in 2016 with count of nearly 8000. The second highest is medical assistance. The least is investigation of property with count of more than 3000.

Total 52,256 crimes occured in 2016.

```
> sum(headdata$newcolumn4) #displaying the total sum of top 10 2016 offense codes
[1] 52256
```

Similarly, amalyzing crime counts and data for 2017 and 2018

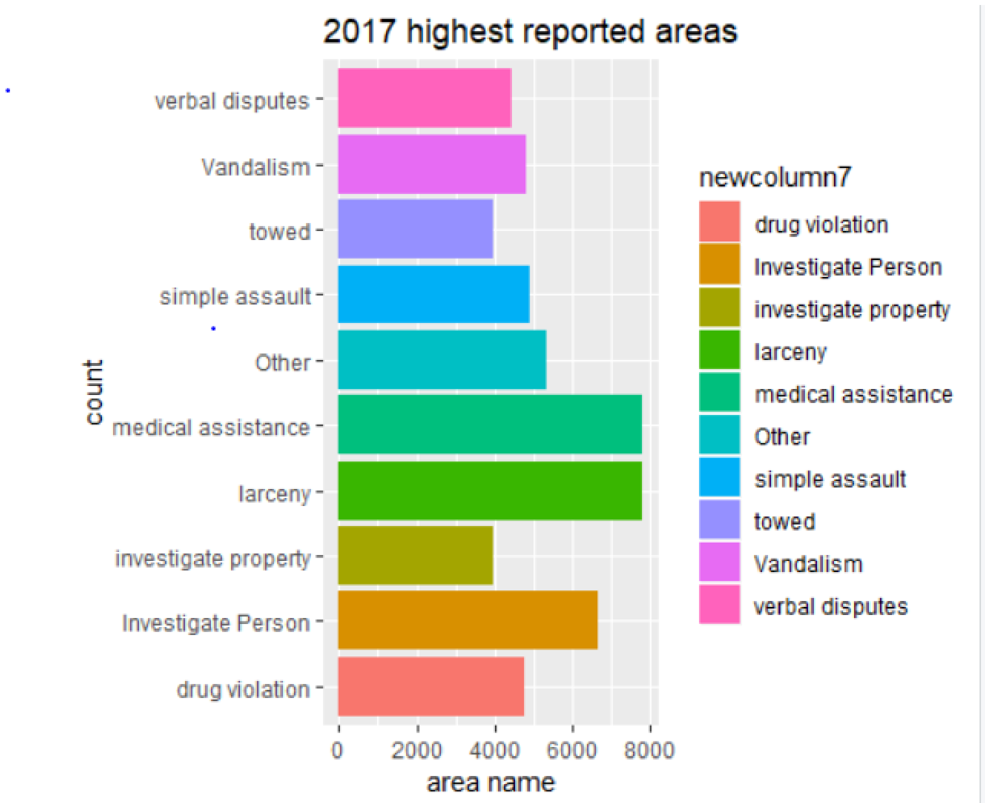
```
#generating the bar chart of 2017 highest reported areas

graph2<-filter(pdata,YEAR==2017) #filtering 2017 year offense codes
table(graph2$OFFENSE_CODE_GROUP)
ocg2<-sort(table(graph2$OFFENSE_CODE_GROUP),decreasing = TRUE)[2:11]
```

Following are the crime counts of the crimes that occured in 2017.

```
                Aggravated Assault                              Aircraft                                  Arson
                              2218                                    17                                     31
     Assembly or Gathering Violations                        Auto Theft                    Auto Theft Recovery
                               232                                  1299                                    344
                         Ballistics                     Biological Threat                             Bomb Hoax
                               325                                     2                                     10
        Burglary - No Property Taken                    Commercial Burglary                     Confidence Games
                                 0                                   426                                    864
                     Counterfeiting                    Criminal Harassment                    Disorderly Conduct
                               452                                    28                                    775
                     Drug Violation                            Embezzlement                          Evading Fare
                              4000                                   107                                    110
                         Explosives                   Fire Related Reports                    Firearm Discovery
                                 5                                   570                                    208
                  Firearm Violations                                 Fraud                              Gambling
                               419                                  1693                                      6
                         Harassment                Harbor Related Incidents                         HOME INVASION
                              1452                                    37                                     32
                           Homicide           HUMAN TRAFFICKING HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
                                50                                     5                                      1
                 Investigate Person                     INVESTIGATE PERSON                  Investigate Property
                              6332                                     1                                   3811
            Landlord/Tenant Disputes                               Larceny             Larceny From Motor Vehicle
                               277                                  7537                                   2982
     License Plate Related Incidents                      License Violation                     Liquor Violation
                               188                                   484                                    324
                       Manslaughter                     Medical Assistance                 Missing Person Located
                                 3                                  7381                                   1470
             Missing Person Reported         Motor Vehicle Accident Response          Offenses Against Child / Family
                              1145                                  9604                                    153
        Operating Under the Influence                                Other                        Other Burglary
                               125                                  4986                                    130
               Phone Call Complaints                 Police Service Incidents               Prisoner Related Incidents
                                 4                                   380                                     67
                     Property Found                          Property Lost                Property Related Damage
                              1201                                  2976                                    271
                       Prostitution               Recovered Stolen Property                  Residential Burglary
                                85                                   419                                   1519
```

```
                     Property Found                          Property Lost                Property Related Damage
                              1201                                  2976                                    271
                       Prostitution               Recovered Stolen Property                  Residential Burglary
                                85                                   419                                   1519
         Restraining Order Violations                               Robbery                       Search Warrants
                               510                                  1234                                    331
                            Service                         Simple Assault                                 Towed
                                77                                  4534                                   3713
                          Vandalism                         Verbal Disputes                            Violations
                              4643                                  4387                                   1296
                     Warrant Arrests
                              2747
> ocg2<-sort(table(graph2$OFFENSE_CODE_GROUP),decreasing = TRUE)[2:11] #taking top 10 offense code groups for 2017 year
```

```
ggplot(headdata,aes(x=newcolumn7,y=newcolumn6, fill=newcolumn7))+
  geom_bar(stat="identity")+
  coord_flip()+ #flipping the cartesian coordinates
  labs(y="area name", x="count",title="2017 highest reported areas")

sum(headdata$newcolumn6) #displaying the total sum of top 10 2017 offense codes
```



From the above graph we can say that larceny and medical assistance has the same crime count of nearly 8000. The least was towing and the second largest was investigation of the property with count of approx. 6600.

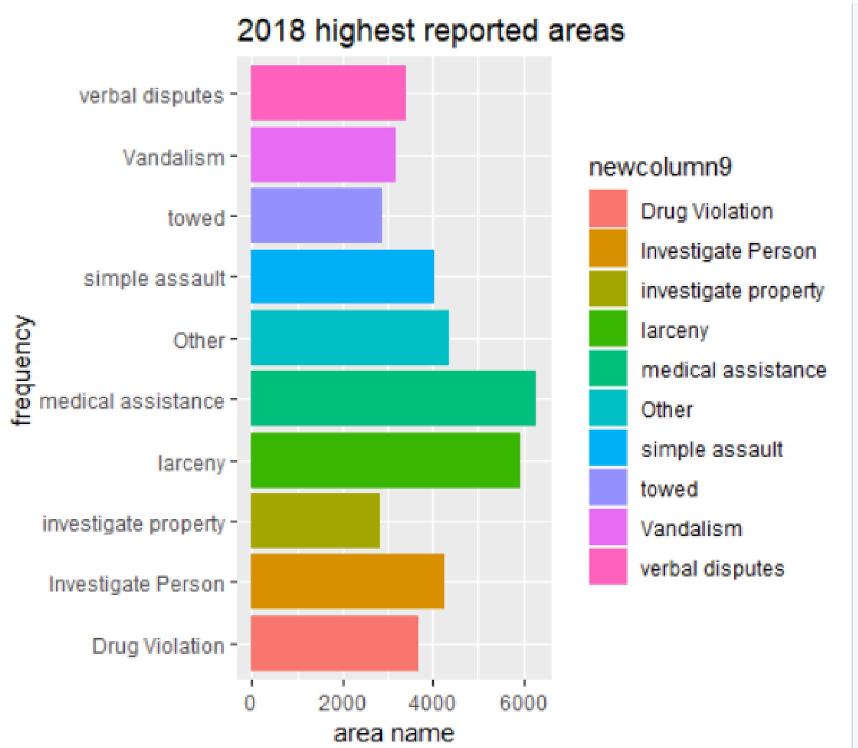
```
> sum(headdata$newcolumn6) #displaying the total sum of top 10 2017 offense codes
[1] 54469
>
```

Total number of crimes that occured in 2017 were 54,469.

Following is the code snippet for 2018 crime analysis:

```
#2018
graph3<-filter(pdata,YEAR==2018) #filtering 2018 year offense codes
table(graph3$OFFENSE_CODE_GROUP)

ocg3<-sort(table(bbb$OFFENSE_CODE_GROUP),decreasing = TRUE)[2:11]#taking top 10 offense code groups for 2018 year

headdata$newcolumn8<-c(6292,5949,4372,4241,4013,3692,3433,3208,2847,2899)
headdata$newcolumn9<-c("medical assistance","larceny","Other","Investigate Person","simple assault","Drug Violation",

ggplot(headdata,aes(x=newcolumn9,y=newcolumn8, fill=newcolumn9))+
  geom_bar(stat="identity")+
  coord_flip()+ #flipping the cartesian coordinates
  labs(y="area name", x="frequency",title="2018 highest reported areas")


sum(headdata$newcolumn8) #displaying the total sum of top 10 2018 offense codes
```

We get the following output:



We can see that in 2018, the highest crime occured was medical assistance with count of more than 6000 followed by larceny with count of almost 6000. The least was towing with count of nearly 3000.

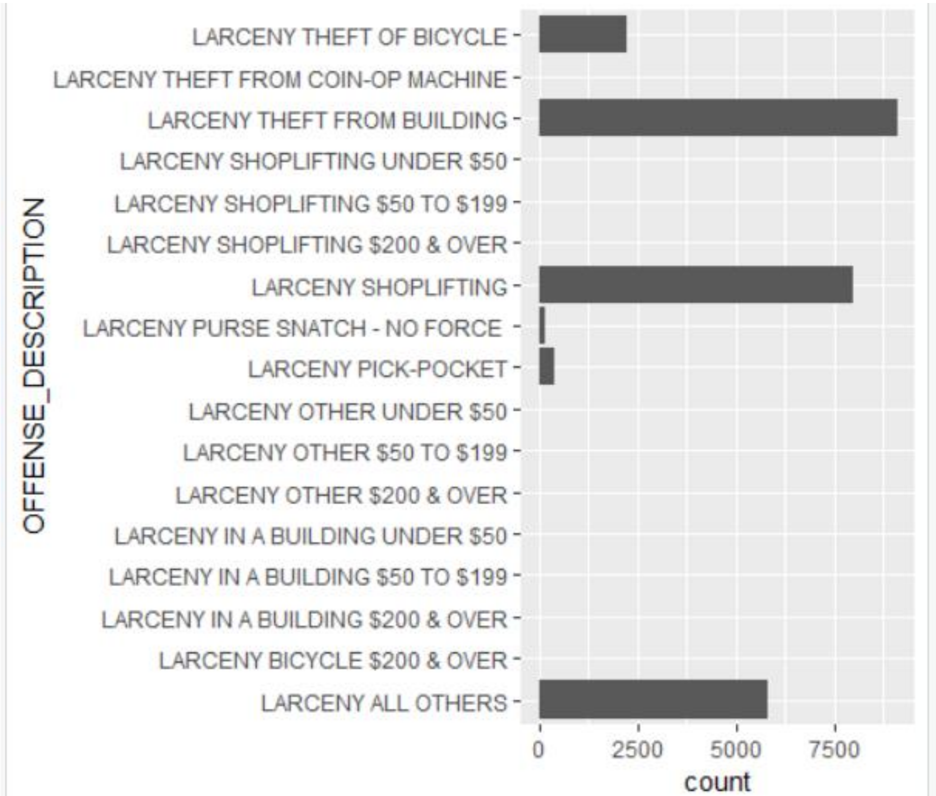
In 2018 there were 40,946 crimes that occured.

```
> sum(headdata$newcolumn8) #displaying the total sum of top 10 2018 offense codes
[1] 40946
```

From the above analysis of 3 years graph we can conclude that larceny was highest crime that occured. So we digged further into it what type of larceny occurs.

```
#displaying the count of larceny offense description
ggplot(ldff, aes(x=OFFENSE_DESCRIPTION),fill=as.factor(OFFENSE_DESCRIPTION)) +
  coord_flip()+  #flipping the coordinates
  geom_bar()
```

We get the following output:



On analyzing we come to know that larceny theft from the building tops the list with count of almost more than 8000. The second crime occured is Larceny shoplifting with count of almost more than 7500.

The least larceny occured is pick pocketing.

Hypothesis Testing

To find out if the mean of the crimes during weekends and weekdays are similar or not, we set the null and alternate hypothesis as below

Null Hypothesis: Ho: crimes at weekday = crimes at weekend
Alternate Hypothesis: Ha: crimes at weekday != crimes at weekend

We set a constant sample selection and selected a sample of size 30

```
set.seed(7) #to set the sample selection
pdata.sample <- sample_n(pdata,30, replace = TRUE)#select random 30 samples
```

subset the table according to days

```
monday<-subset(pdata.sample,subset = DAY_OF_WEEK=="Monday")#subset table when day is monday
tuesday<-subset(pdata.sample,subset = DAY_OF_WEEK=="Tuesday")#subset table when day is tuesday
wednesday<-subset(pdata.sample,subset = DAY_OF_WEEK=="Wednesday")#subset table when day is wednesday
thursday<-subset(pdata.sample,subset = DAY_OF_WEEK=="Thursday")#subset table when day is thursday
friday<-subset(pdata.sample,subset = DAY_OF_WEEK=="Friday")#subset table when day is friday
saturday<-subset(pdata.sample,subset = DAY_OF_WEEK=="Saturday")#subset table when day is saturday
sunday<-subset(pdata.sample,subset = DAY_OF_WEEK=="Sunday")#subset table when day is sunday
```

Count of the number of weekdays

```
monday_count <- count(monday)#count of the mondays
tuesday_count <- count(tuesday)#count of the tuesdays
wednesday_count <- count(wednesday)#count of the wednesdays
thursday_count <- count(thursday)#count of the thursdays
friday_count <- count(friday)#count of the fridays
saturday_count <- count(saturday)#count of the saturdays
sunday_count <- count(sunday)#count of the sundays
```

Separately count the weekdays and weekends and convert it to numeric type

```
weekday.counts <- c(monday_count,tuesday_count,wednesday_count,thursday_count) #net weekdays values
weekend.counts <- c(friday_count, saturday_count, sunday_count) #net weekdends values

weekday.counts <- as.numeric(as.character(weekday.counts)) #weekdays numeric
weekend.counts <- as.numeric(as.character(weekend.counts)) #weekends numeric
```

Carry out the hypothesis 2 tailed t-test

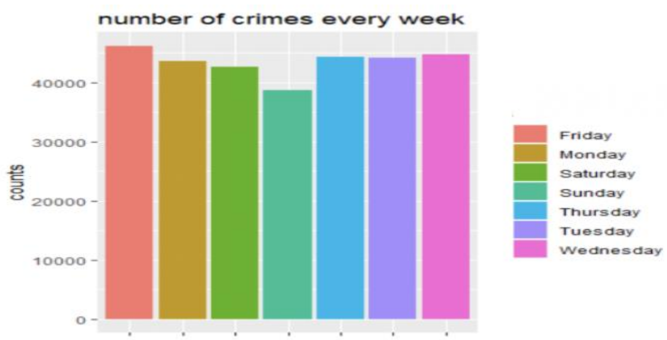
```
test.paired <- t.test(weekday.counts, weekend.counts, mu=0, alternative = "two.sided", paired = F, conf.level = 0.99) #confidence level 99%
test.paired # t - test
```

The result of the t-test is as below

```
        Welch Two Sample t-test

data:  weekday.counts and weekend.counts
t = 0.5412, df = 4.6858, p-value = 0.6131
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -7.273789  9.440456
sample estimates:
mean of x mean of y
 4.750000  3.666667
```

The p-value is 0.6131 which much higher 0.05. This shows that we refuse to reject the null hypothesis.

This is consistent with the following plot where the count of crimes in all the weekdays are almost similar.



With 99% confidence we can state that, at BOSTON, the mean crimes at weekdays is equal to that during weekends

Time Series Analysis

Converting it to Time series

```
# Finding the class of column OCCURED_ON_DATE
class(crimes$OCCURRED_ON_DATE)
```

```
> class(crimes$OCCURRED_ON_DATE)
[1] "Date"
>
```

```r
# As the class of column OCCURED_ON_DATE is factor we are converting it into date formart (year-month-day)
crimes$OCCURRED_ON_DATE <- as.Date(crimes$OCCURRED_ON_DATE, format="%Y-%m-%d")

# The column has both dates and times so now we are dividing it into only dates
dates<-cut(crimes$OCCURRED_ON_DATE, 'day')

# Now we are having the counts of each date which represents the number of times
tab.dates<- table(dates)

#converting it into data frame with its frequency
crimes.dates<-data.frame(Date=format(as.Date(names(tab.dates)), '%d/%m/%Y'),
          Frequency=as.vector(tab.dates))
# Having a look at the data frame
head(crimes.dates)
```

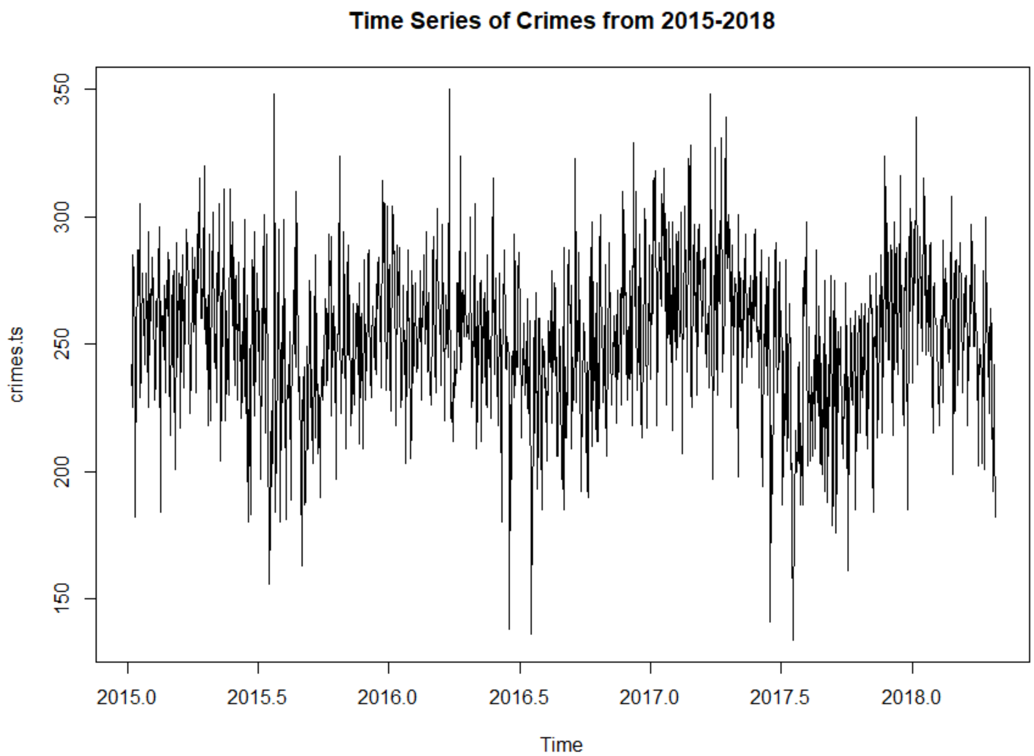
```
> head(crimes.dates)
        Date Frequency
1 15/06/2015       239
2 16/06/2015       242
3 17/06/2015       225
4 18/06/2015       285
5 19/06/2015       276
6 20/06/2015       246
```

```r
#converting it into the time series, starting at the year 2015-June and 167th day of the year with frequency 365 days
crimes.ts<- ts((crimes.dates$Frequency),start=c(2015,6,167),frequency =365 )

#having a look at the time series
head(crimes.ts)
```

```
> head(crimes.ts)
Time Series:
Start = c(2015, 6)
End = c(2015, 11)
Frequency = 365
[1] 239 242 225 285 276 246
```

In this time series, The starting date is 15th June 2015 and the frequency is 365.

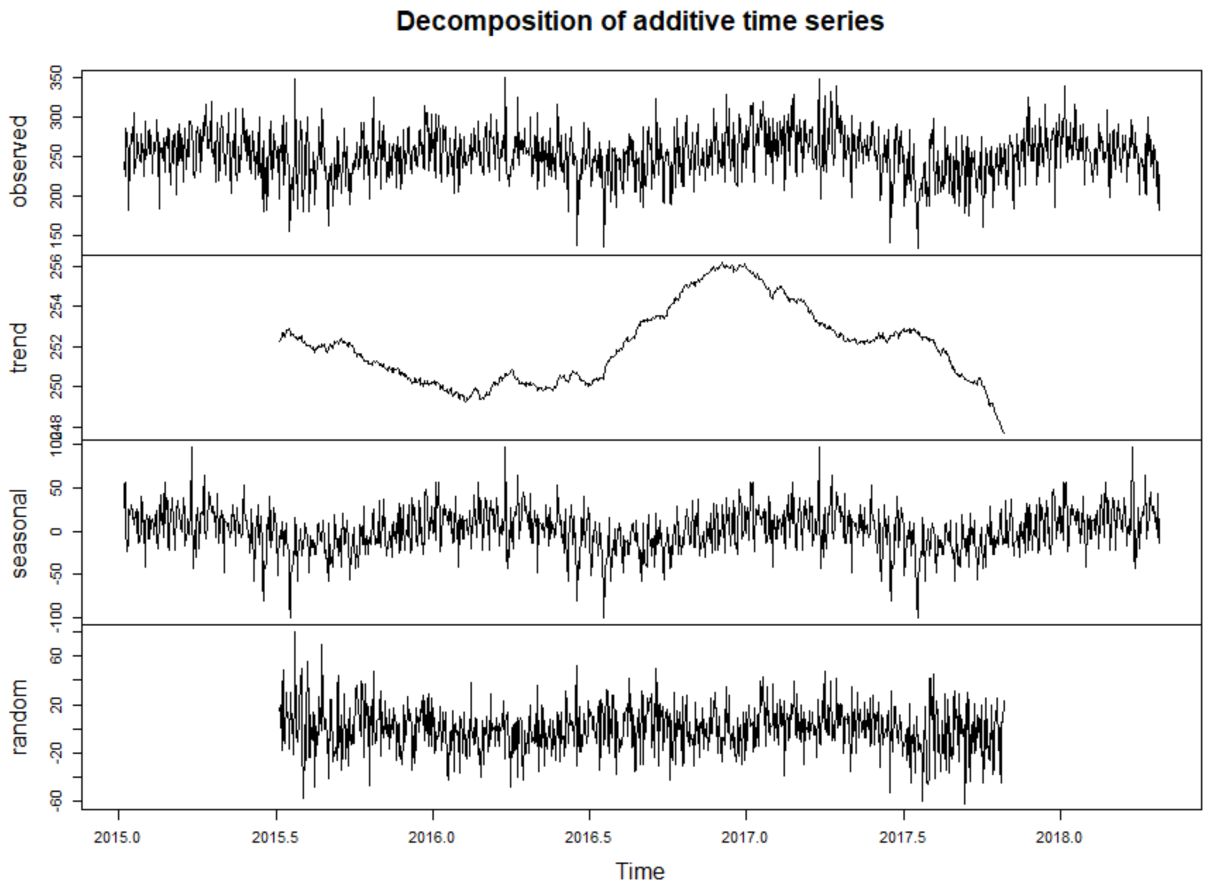
```r
#ploting the time series graph
plot(crimes.ts,main="Time Series of Crimes from 2015-2018")
```



We can see that the above times series fluctuates consistently and the mean and variance do not change over time so it's an additive time series. Now we can decompose this additive time series to find the trend.

Decomposing of additive time series

```
# Decomposing the time series into 3 other components trend,seasonal & random. To find how the trend shifts
plot(decompose(crimes.ts))
```



From the trend graph, we can see that we can observe that the Number of crimes per day in Boston has increased significantly from June 2017 and then gradually decreased from the year 2017.

Finding the best fit model for ARIMA

```
# Using auto arima to find the best order for the arima model with lowest aic value
# USing Tracr=TRUE reports the list of Arima models considered
mymodel<- auto.arima(crimes.ts,ic="aic",trace = TRUE)
mymodel
```

```
Fitting models using approximations to speed things up...

ARIMA(2,0,2)(1,0,1)[365] with non-zero mean : Inf
ARIMA(0,0,0)             with non-zero mean : 11671.93
ARIMA(1,0,0)(1,0,0)[365] with non-zero mean : Inf
ARIMA(0,0,1)(0,0,1)[365] with non-zero mean : Inf
ARIMA(0,0,0)             with zero mean     : 16794.37
ARIMA(0,0,0)(1,0,0)[365] with non-zero mean : Inf
ARIMA(0,0,0)(0,0,1)[365] with non-zero mean : Inf
ARIMA(0,0,0)(1,0,1)[365] with non-zero mean : Inf
ARIMA(1,0,0)             with non-zero mean : 11543.33
ARIMA(1,0,0)(0,0,1)[365] with non-zero mean : Inf
ARIMA(1,0,0)(1,0,1)[365] with non-zero mean : Inf
ARIMA(2,0,0)             with non-zero mean : 11544.82
ARIMA(1,0,1)             with non-zero mean : 11513.57
ARIMA(1,0,1)(1,0,0)[365] with non-zero mean : Inf
ARIMA(1,0,1)(0,0,1)[365] with non-zero mean : Inf
ARIMA(1,0,1)(1,0,1)[365] with non-zero mean : Inf
ARIMA(0,0,1)             with non-zero mean : 11554.75
ARIMA(2,0,1)             with non-zero mean : 11476.74
ARIMA(2,0,1)(1,0,0)[365] with non-zero mean : Inf
ARIMA(2,0,1)(0,0,1)[365] with non-zero mean : Inf
ARIMA(2,0,1)(1,0,1)[365] with non-zero mean : Inf
ARIMA(3,0,1)             with non-zero mean : 11480.66
ARIMA(2,0,2)             with non-zero mean : 11471.47
ARIMA(2,0,2)(1,0,0)[365] with non-zero mean : Inf
ARIMA(2,0,2)(0,0,1)[365] with non-zero mean : Inf
ARIMA(1,0,2)             with non-zero mean : 11471.15
ARIMA(1,0,2)(1,0,0)[365] with non-zero mean : Inf
ARIMA(1,0,2)(0,0,1)[365] with non-zero mean : Inf
ARIMA(1,0,2)(1,0,1)[365] with non-zero mean : Inf
ARIMA(0,0,2)             with non-zero mean : 11551.81
ARIMA(1,0,3)             with non-zero mean : 11471.13
```

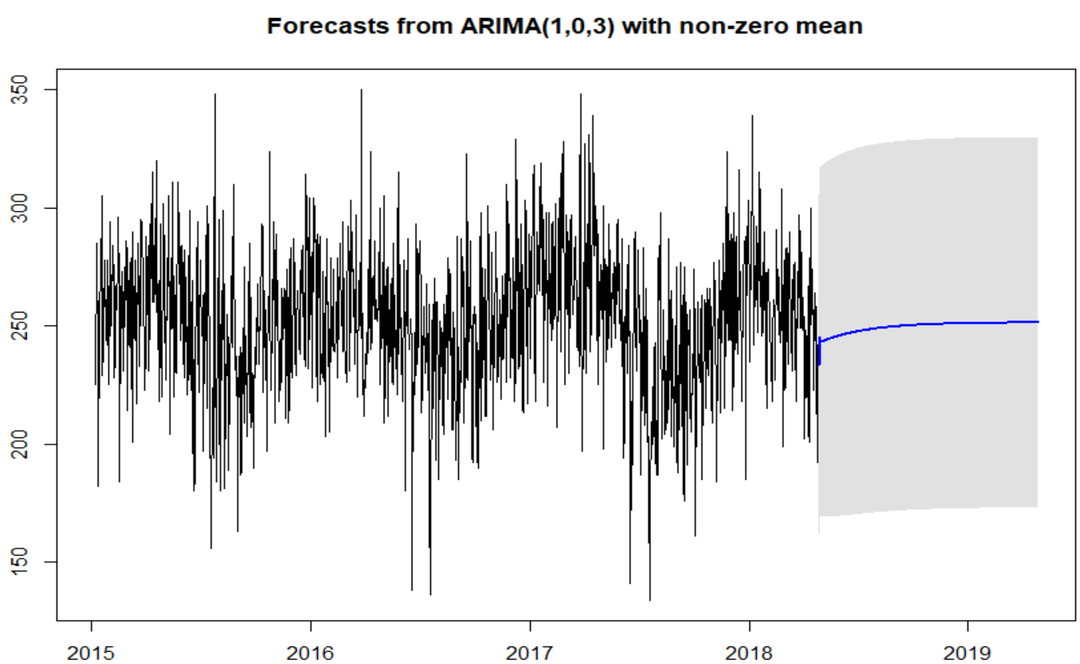
Through various combinations, we are trying to find the best fit model which has the lowest AIC value.

```
> mymodel
Series: crimes.ts
ARIMA(1,0,3) with non-zero mean

Coefficients:
         ar1      ma1      ma2     ma3      mean
      0.9875  -0.7370  -0.2317  0.0404  251.4547
s.e.  0.0059   0.0294   0.0355  0.0289    4.3894

sigma^2 estimated as 778.8:  log likelihood=-5728.42
AIC=11468.83   AICc=11468.9   BIC=11499.41
```

From the results, we found that the best-fitted model for ARIMA is of order ARIMA(1,0,3) with non-zero mean which has AIC=11468.83. So using this order we will forecast the crime rate for next 1 year.

Forecasting

```
#Using the best ARIMA model we are trying to forecast the next 365days trend in crime rate with 99% confidence interval
fut.crimes <- forecast(mymodel,level=c(99),h=365)

#plotting the graph of forecasting time series
plot(fut.crimes)
```



From the above forecasting graph, we could see that the trend in the number of crimes per day has slightly increased for the next 1 year compared to past year.

Holt-Winters Model

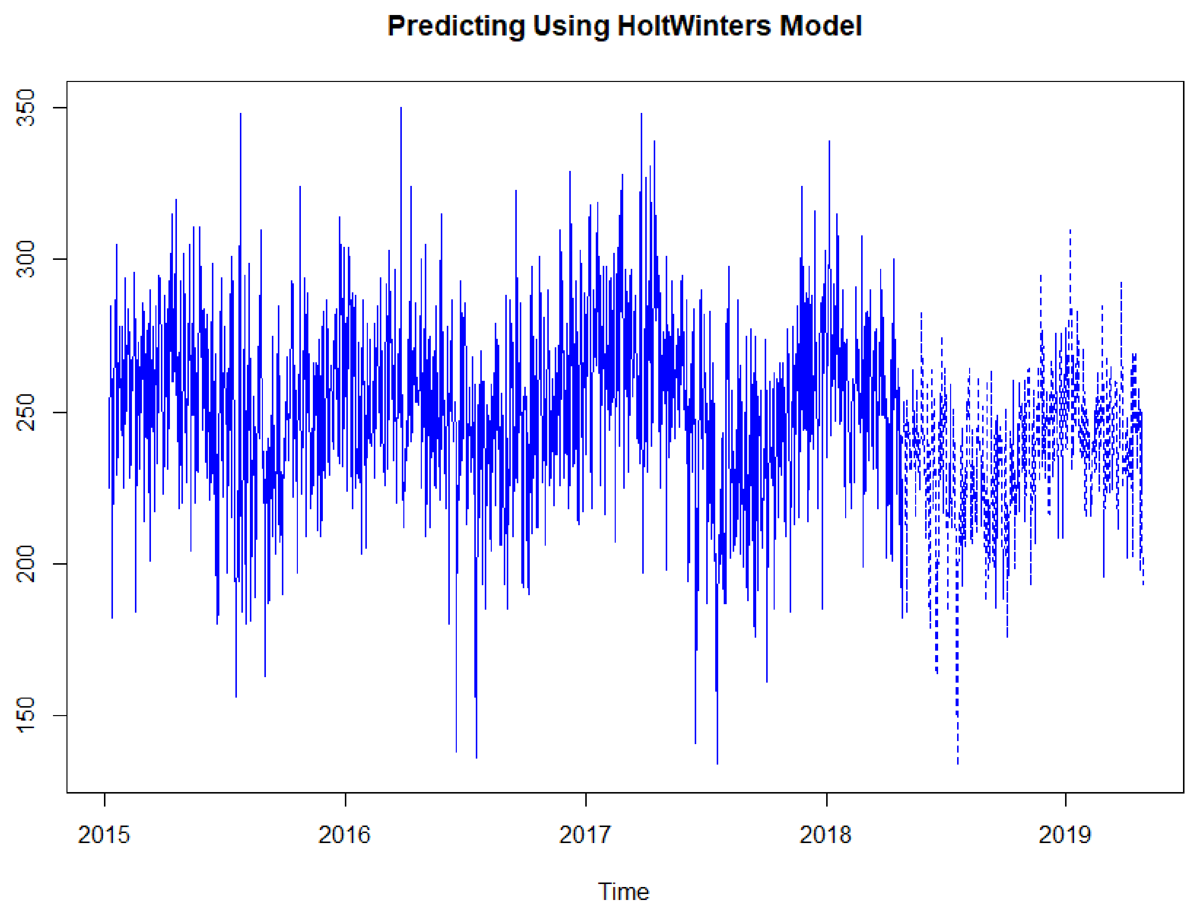
Now we are using Holt-winters model to predict the number of crimes for on a daily basis i n Boston

```
# Computing Holt-Winters Filtering of a given time series
hw.crimes<- HoltWinters(crimes.ts)

#using predict to fuction from results of model fitting, we are predicting next 356 daily crime rates
hw<- predict(hw.crimes,n.ahead = 365)
#having a loot at the predicted values
head(hw)
```

```
Time Series:
Start = c(2018, 118)
End = c(2018, 123)
Frequency = 365
          fit
[1,] 239.8115
[2,] 244.3430
[3,] 253.0010
[4,] 238.7308
[5,] 243.0833
[6,] 183.4944
```

Plotting the graph of Holt-Winters predicted values

```
# Ploting the predicted values of time series with the past years time series. with dotted line
ts.plot(crimes.ts,hw,col="blue",lty=1:3,main="Predicting Using HoltWinters Model")
```



From this plot we can have a predicted value of crimes on each day for the next 365days.This Analysis helps Boston police to act accordingly when the crime rate is high and try to reduce them.

Conclusion

After performing an in-depth analysis of crimes in Boston, we have analyzed the trends and patterns of the different locations at different times of the year, month and day. Some important points are mentioned below

1. **Washington street had the most crimes in the past 3 years.**
2. **January is the month when most crimes occur in Boston**
3. **Larceny is the most reported crime in Boston, and the building is the most popular among it.**
4. **Weekdays and weekends make no difference in the crime rates**
5. **There is a chance of a slight increase in crimes in the year 2019, based on the previous year's data.**