

Crimes in Boston: Data Analysis using R

Prepared By:

Mir Rafiul Ahmed (2017231005)

About Dataset

Context

We have chosen the dataset 'Crimes in Boston' which is basically crime incident reports, provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred.

Content

Records begin in June 14, 2015 and continue to September 3, 2018. This dataset contains 327821 records over 2015-2018 timeframe

Acknowledgements

The data is a public source provided by Analyze Boston. The most up-to-date version can be found [here](#)

Types of Data Mining/ Econometric Techniques used in Project

- Descriptive Statistics (Explanatory Data Analysis)
- Regression (Single and Multivariate Regression)
- Hypothesis (F test)
- Forecasting (ARIMA)
- Heteroscedasticity (Breusch-Pagan Test and White Test)
- Multicollinearity (Variance Inflation Factor)

Theoretical Area for Topic

This data analysis would fit in the **macroeconomic area** as this dataset is concerned with crime records of Boston from 2015-18 with more than 0.3 million records.

Package Used in this Project

`install.packages("ggplot2")` – General Scheme for Data Visualization

`install.packages("stargazer")` – For creating quality tables and summary statistics

`install.packages("aggr")` – For calculating or plot the amount of missing/imputed values in each variable and the amount of missing/imputed values in certain combinations of variable

`install.packages("dplyr")` - a package for making data manipulation easier

`install.packages("tidyverse")` - an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

`install.packages("janitor")` - expedite the initial data exploration and cleaning that comes with any new data

`install.packages("forecast")` - provides methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling

`install.packages("magrittr")` – provides a mechanism for chaining commands with a new forward pipe operator `%>%`

`install.packages("wooldridge")` – aims to lighten task by efficiently loading any dataset found in the text with a single command

`install.packages("skedastic")` – make diagnostic methods for detecting heteroscedasticity in linear regression model

`install.packages("lmtest")` – collection of tests datasets and examples for diagnostic checking in linear regression model

Dataset in a Glimpse

Group 6.R* crime																	
Filter																	
	INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long	Location
1	1182080058	2403	Disorderly Conduct	DISTURBING THE PEACE	E18	495	NA	2018-10-03 20:13:00	2018	10	Wednesday	20	Part Two	ARLINGTON ST	42.26261	-71.12119	(42.26260773, -71.12118637)
2	1182080053	3201	Property Lost	PROPERTY - LOST	D14	795	NA	2018-08-30 20:00:00	2018	8	Thursday	20	Part Three	ALLSTON ST	42.35211	-71.13531	(42.35211146, -71.13531147)
3	1182080052	2647	Other	THREATS TO DO BODILY HARM	B2	329	NA	2018-10-03 19:20:00	2018	10	Wednesday	19	Part Two	DEVON ST	42.30813	-71.07693	(42.30812619, -71.07692974)
4	1182080051	413	Aggravated Assault	ASSAULT - AGGRAVATED - BATTERY	A1	92	NA	2018-10-03 20:00:00	2018	10	Wednesday	20	Part One	CAMBRIDGE ST	42.35945	-71.05965	(42.35945371, -71.05964817)
5	1182080050	3122	Aircraft	AIRCRAFT INCIDENTS	A7	36	NA	2018-10-03 20:49:00	2018	10	Wednesday	20	Part Three	PRESCOTT ST	42.37526	-71.02466	(42.37525782, -71.02466343)
6	1182080049	1402	Vandalism	VANDALISM	C11	351	NA	2018-10-02 20:40:00	2018	10	Tuesday	20	Part Two	DORCHESTER AVE	42.29920	-71.06047	(42.29919694, -71.06046974)
7	1182080048	3803	Motor Vehicle Accident Response	M/V ACCIDENT - PERSONAL INJURY	NA	NA	NA	2018-10-03 20:16:00	2018	10	Wednesday	20	Part Three	NA	42.32073	-71.05676	(42.32073413, -71.05676415)
8	1182080047	3301	Verbal Disputes	VERBAL DISPUTE	B2	603	NA	2018-10-03 19:32:00	2018	10	Wednesday	19	Part Three	TREMONT ST	42.33381	-71.10378	(42.33380683, -71.10377843)
9	1182080045	802	Simple Assault	ASSAULT SIMPLE - BATTERY	E18	543	NA	2018-10-03 19:27:51	2018	10	Wednesday	19	Part Two	AVILA RD	42.25614	-71.12803	(42.25614494, -71.12802506)
10	1182080044	3410	Towed	TOWED MOTOR VEHICLE	D4	621	NA	2018-10-03 20:00:00	2018	10	Wednesday	20	Part Three	COMMONWEALTH AVE	42.34887	-71.08936	(42.34886600, -71.08936284)
11	1182080043	3803	Motor Vehicle Accident Response	M/V ACCIDENT - PERSONAL INJURY	D14	750	NA	2018-10-03 19:33:00	2018	10	Wednesday	19	Part Three	FOSTER ST	42.34432	-71.15778	(42.34432328, -71.15778368)
12	1182080042	706	Auto Theft	AUTO THEFT - MOTORCYCLE / SCOOTER	E13	582	NA	2018-10-01 20:00:00	2018	10	Monday	20	Part One	BYNNER ST	42.32324	-71.10892	(42.32324363, -71.10892316)
13	1182080041	3006	Medical Assistance	SICK/INJURED/MEDICAL - PERSON	E18	484	NA	2018-10-03 17:18:00	2018	10	Wednesday	17	Part Three	BELNEL RD	42.26060	-71.10306	(42.26059891, -71.10306140)
14	1182080040	3115	Investigate Person	INVESTIGATE PERSON	B3	427	NA	2018-10-03 08:00:00	2018	10	Wednesday	8	Part Three	ELIZABETH ST	42.27987	-71.08798	(42.27986526, -71.08798275)
15	1182080039	3006	Medical Assistance	SICK/INJURED/MEDICAL - PERSON	B3	469	NA	2018-10-03 19:58:30	2018	10	Wednesday	19	Part Three	FOTTLER RD	42.27792	-71.09641	(42.27791927, -71.09640610)
16	1182080038	3831	Motor Vehicle Accident Response	M/V - LEAVING SCENE - PROPERTY DAMAGE	NA	NA	NA	2018-10-03 19:30:00	2018	10	Wednesday	19	Part Three	NA	42.31596	-71.09043	(42.31596119, -71.09042564)
17	1182080037	2647	Other	THREATS TO DO BODILY HARM	C11	385	NA	2018-10-03 18:35:00	2018	10	Wednesday	18	Part Two	HALLET ST	42.28077	-71.04736	(42.28076737, -71.04736497)
18	1182080035	2647	Other	THREATS TO DO BODILY HARM	B2	326	NA	2018-10-03 19:56:00	2018	10	Wednesday	19	Part Two	DUNKELD ST	42.31278	-71.07563	(42.31277782, -71.07562922)
19	1182080034	3115	Investigate Person	INVESTIGATE PERSON	D4	626	NA	2018-10-03 18:41:00	2018	10	Wednesday	18	Part Three	KILMARNOCK ST	42.34268	-71.09938	(42.34268073, -71.09937714)
20	1182080031	3108	Fire Related Reports	FIRE REPORT - HOUSE, BUILDING, ETC.	C11	338	NA	2018-10-03 18:18:00	2018	10	Wednesday	18	Part Three	NAVILLUS TER	42.30999	-71.06213	(42.30998781, -71.06213111)
21	1182080030	3831	Motor Vehicle Accident Response	M/V - LEAVING SCENE - PROPERTY DAMAGE	C6	234	NA	2018-10-02 20:00:00	2018	10	Tuesday	20	Part Three	COLUMBIA RD	42.33001	-71.03851	(42.33001489, -71.03851306)
22	1182080029	613	Larceny	LARCENY SHOPLIFTING	D4	146	NA	2018-10-03 19:09:00	2018	10	Wednesday	19	Part One	HUNTINGTON AVE	42.34797	-71.07850	(42.34797240, -71.07849582)
23	1182080028	3114	Investigate Property	INVESTIGATE PROPERTY	B2	295	NA	2018-10-03 18:24:00	2018	10	Wednesday	18	Part Three	HANSFORD ST	42.32528	-71.08352	(42.32527567, -71.08351579)
24	1182080027	3301	Verbal Disputes	VERBAL DISPUTE	B2	294	NA	2018-10-03 18:53:00	2018	10	Wednesday	18	Part Three	AKRON ST	42.32579	-71.08564	(42.32579408, -71.08563758)
25	1182080027	619	Larceny	LARCENY ALL OTHERS	B2	294	NA	2018-10-03 18:53:00	2018	10	Wednesday	18	Part One	AKRON ST	42.32579	-71.08564	(42.32579408, -71.08563758)
26	1182080026	2629	Harassment	HARASSMENT	C6	256	NA	2017-08-31 00:00:00	2017	8	Thursday	0	Part Two	MOUNT VERNON ST	42.31779	-71.04211	(42.31779354, -71.04210959)
27	1182080025	311	Robbery	ROBBERY - COMMERCIAL	NA	NA	NA	2018-10-03 19:05:00	2018	10	Wednesday	19	Part One	NA	42.34952	-71.07949	(42.34952402, -71.07949284)
28	1182080024	1842	Drug Violation	DRUGS - POSS CLASS A - HEROIN, ETC.	E13	640	NA	2018-10-03 17:30:00	2018	10	Wednesday	17	Part Two	BOYLSTON ST	42.31694	-71.10395	(42.31694189, -71.10394558)
29	1182080023	3115	Investigate Person	INVESTIGATE PERSON	D4	165	NA	2018-05-12 00:00:00	2018	5	Saturday	0	Part Three	HARRISON AVE	42.34285	-71.06516	(42.34285014, -71.06516235)
30	1182080022	3006	Medical Assistance	SICK/INJURED/MEDICAL - PERSON	D14	779	NA	2018-10-03 19:19:00	2018	10	Wednesday	19	Part Three	MONTCALM AVE	42.35321	-71.14841	(42.35321122, -71.14841124)
31	1182080021	3006	Medical Assistance	SICK/INJURED/MEDICAL - PERSON	A1	80	NA	2018-10-03 18:49:00	2018	10	Wednesday	18	Part Three	MARGARET ST	42.36570	-71.05580	(42.36570319, -71.05580301)
32	1182080020	2647	Other	THREATS TO DO BODILY HARM	A7	15	NA	2018-10-03 18:46:00	2018	10	Wednesday	18	Part Two	CHAUCER ST	42.38306	-71.02118	(42.38305542, -71.02117632)
33	1182080019	802	Simple Assault	ASSAULT SIMPLE - BATTERY	E13	948	NA	2018-10-03 18:07:00	2018	10	Wednesday	18	Part Two	AMORY ST	42.31916	-71.10068	(42.31915678, -71.10067583)
34	1182080018	2900	Other	VAL - VIOLATION OF AUTO LAW - OTHER	C11	255	NA	2018-10-03 19:05:00	2018	10	Wednesday	19	Part Two	GREENMOUNT ST	42.31041	-71.05918	(42.31041438, -71.05918086)
35	1182080018	2905	Violations	VAL - OPERATING WITHOUT LICENSE	C11	255	NA	2018-10-03 19:05:00	2018	10	Wednesday	19	Part Two	GREENMOUNT ST	42.31041	-71.05918	(42.31041438, -71.05918086)
36	1182080016	1402	Vandalism	VANDALISM	C11	405	NA	2018-09-30 00:00:00	2018	9	Sunday	0	Part Two	WASHINGTON ST	42.27235	-71.06872	(42.27234506, -71.06871774)
37	1182080015	2646	Liquor Violation	LIQUOR - DRINKING IN PUBLIC	B2	279	NA	2018-10-03 19:18:00	2018	10	Wednesday	19	Part Two	DEARBORN ST	42.32856	-71.07928	(42.32855807, -71.07927600)
38	1182080013	1402	Vandalism	VANDALISM	B2	588	NA	2018-10-02 22:00:00	2018	10	Tuesday	22	Part Two	GURNEY ST	42.33201	-71.09599	(42.33201285, -71.09598982)

Visualizations

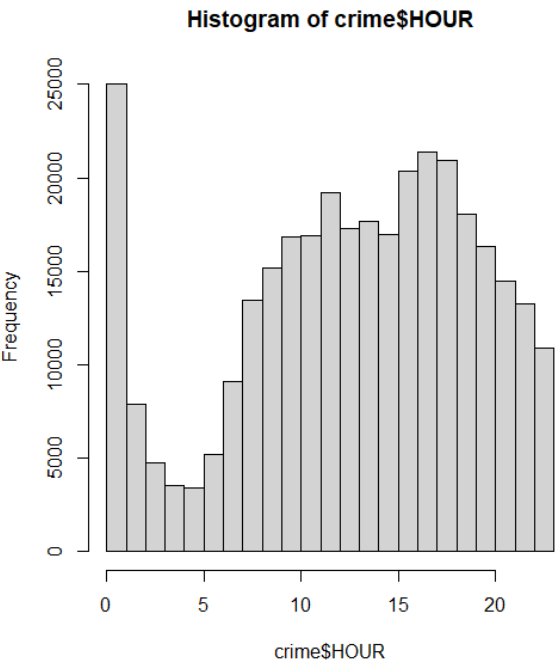
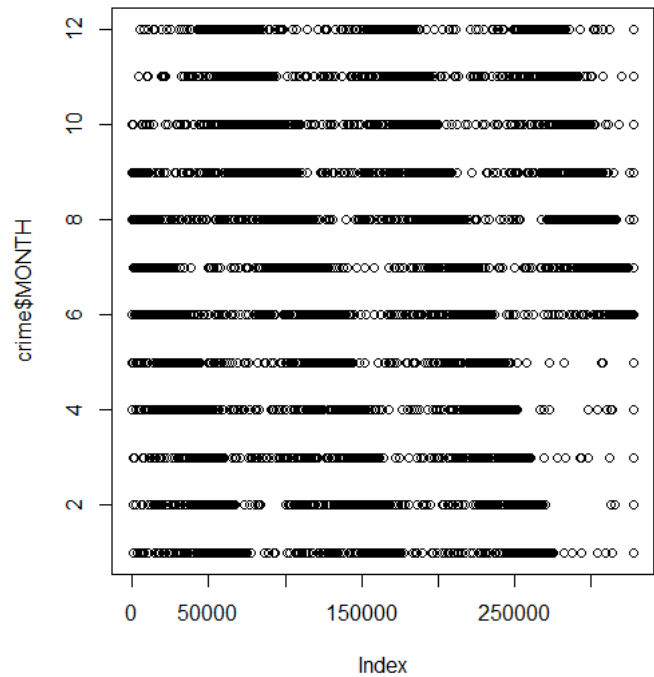
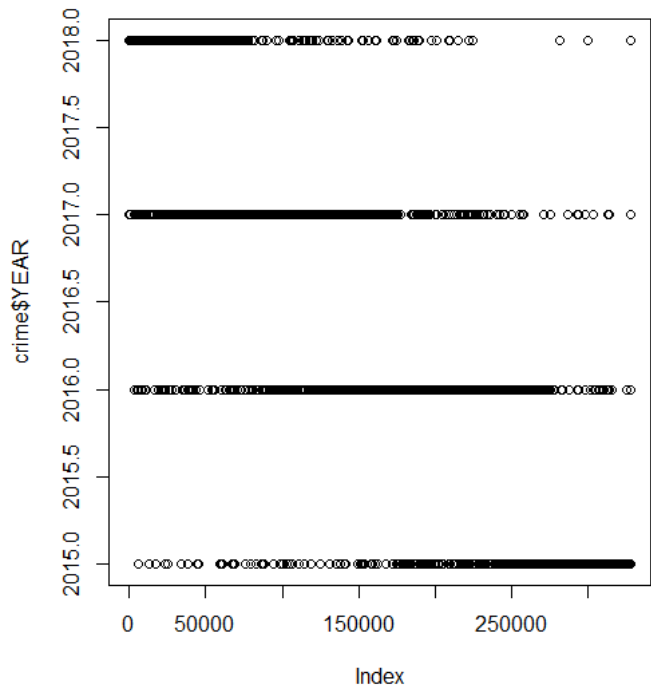
Summary of crime data file

```
> summary(crime)
```

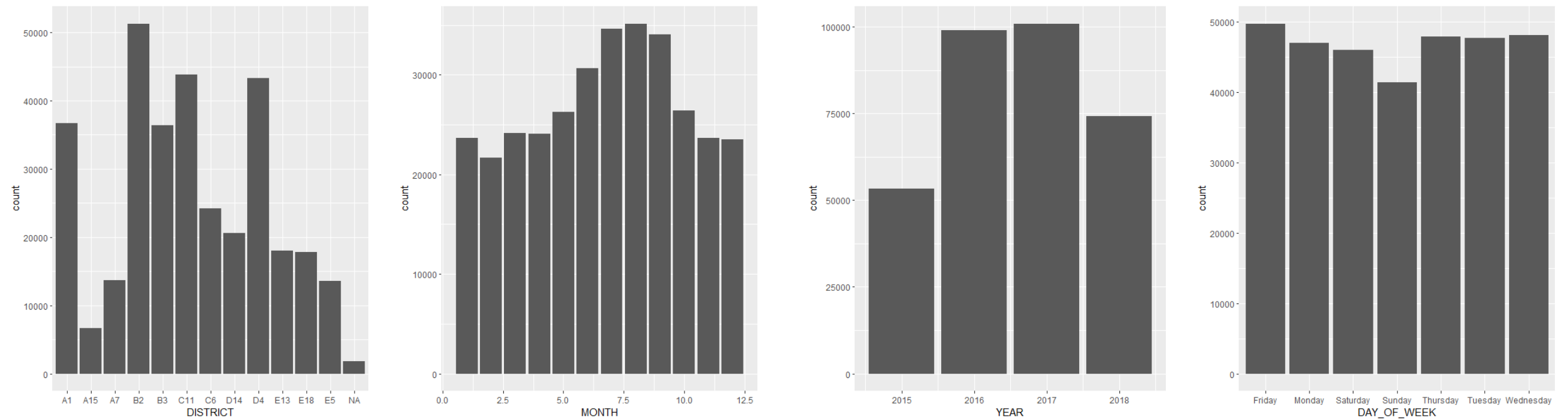
INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	
Length:327820	Min. : 111	Length:327820	Length:327820	Length:327820	Min. : 0.0	Mode:logical	Min. :2015-06-15 00:00:00.00	
Class :character	1st Qu.:1001	Class :character	Class :character	Class :character	1st Qu.:177.0	NA's:327820	1st Qu.:2016-04-20 09:43:45.00	
Mode :character	Median :2907	Mode :character	Mode :character	Mode :character	Median :343.0		Median :2017-02-14 15:49:00.00	
	Mean :2318				Mean :383.2		Mean :2017-02-10 07:26:53.98	
	3rd Qu.:3201				3rd Qu.:544.0		3rd Qu.:2017-11-30 18:23:45.00	
	Max. :3831				Max. :962.0		Max. :2018-10-03 20:49:00.00	
					NA's :20920			
YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long	Location
Min. :2015	Min. : 1.000	Length:327820	Min. : 0.00	Length:327820	Length:327820	Min. : -1.00	Min. : -71.18	Length:327820
1st Qu.:2016	1st Qu.: 4.000	Class :character	1st Qu.: 9.00	Class :character	Class :character	1st Qu.:42.30	1st Qu.: -71.10	Class :character
Median :2017	Median : 7.000	Mode :character	Median :14.00	Mode :character	Mode :character	Median :42.33	Median : -71.08	Mode :character
Mean :2017	Mean : 6.672		Mean :13.11			Mean :42.21	Mean : -70.91	
3rd Qu.:2017	3rd Qu.: 9.000		3rd Qu.:18.00			3rd Qu.:42.35	3rd Qu.: -71.06	
Max. :2018	Max. :12.000		Max. :23.00			Max. :42.40	Max. : -1.00	
						NA's :20632	NA's :20632	

```
> |
```

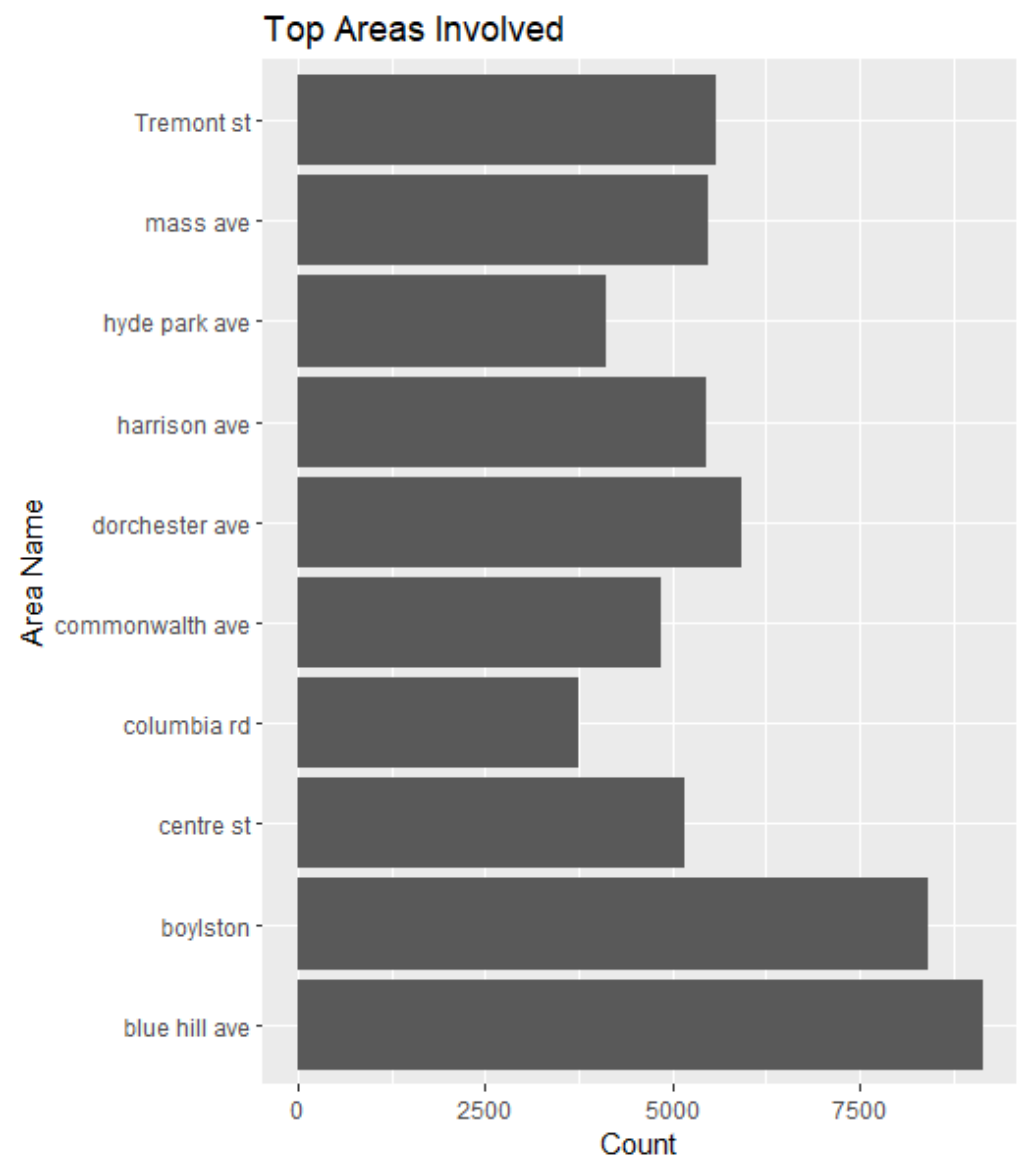
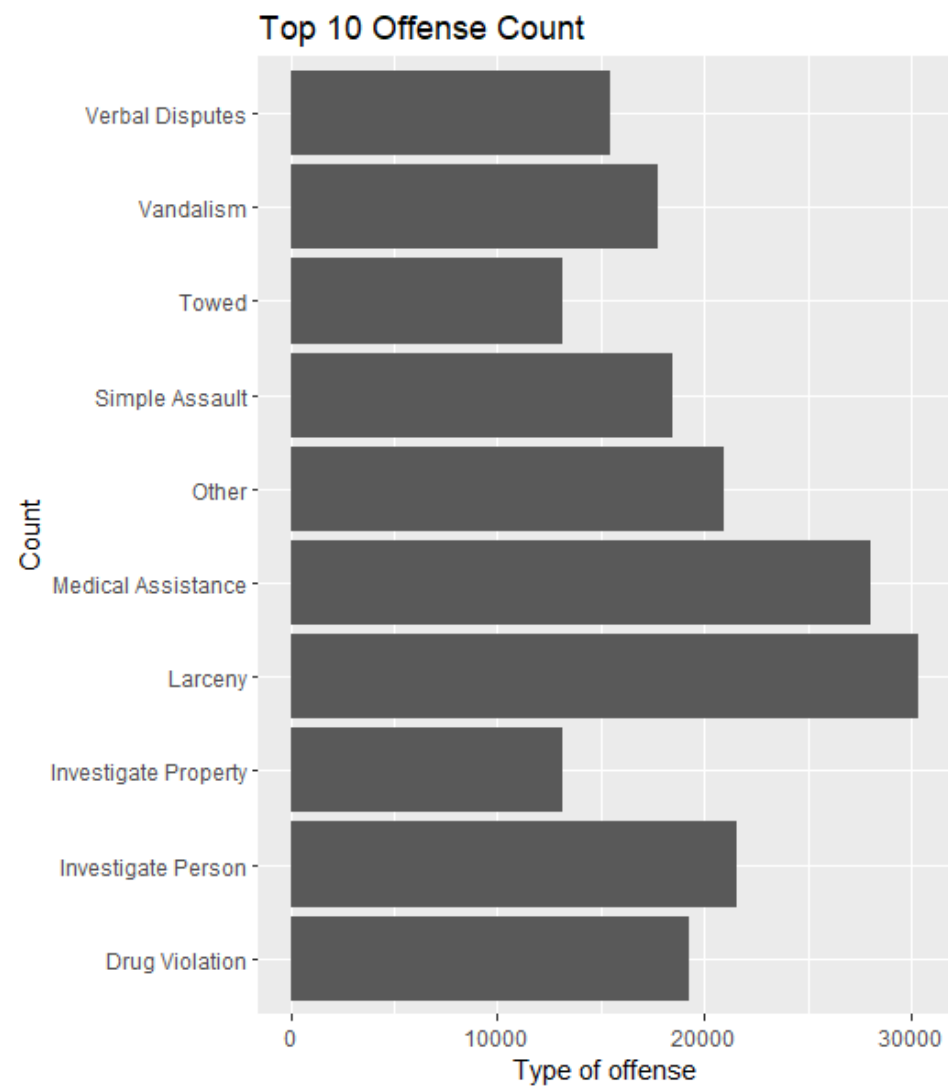
Crime Data Plotting



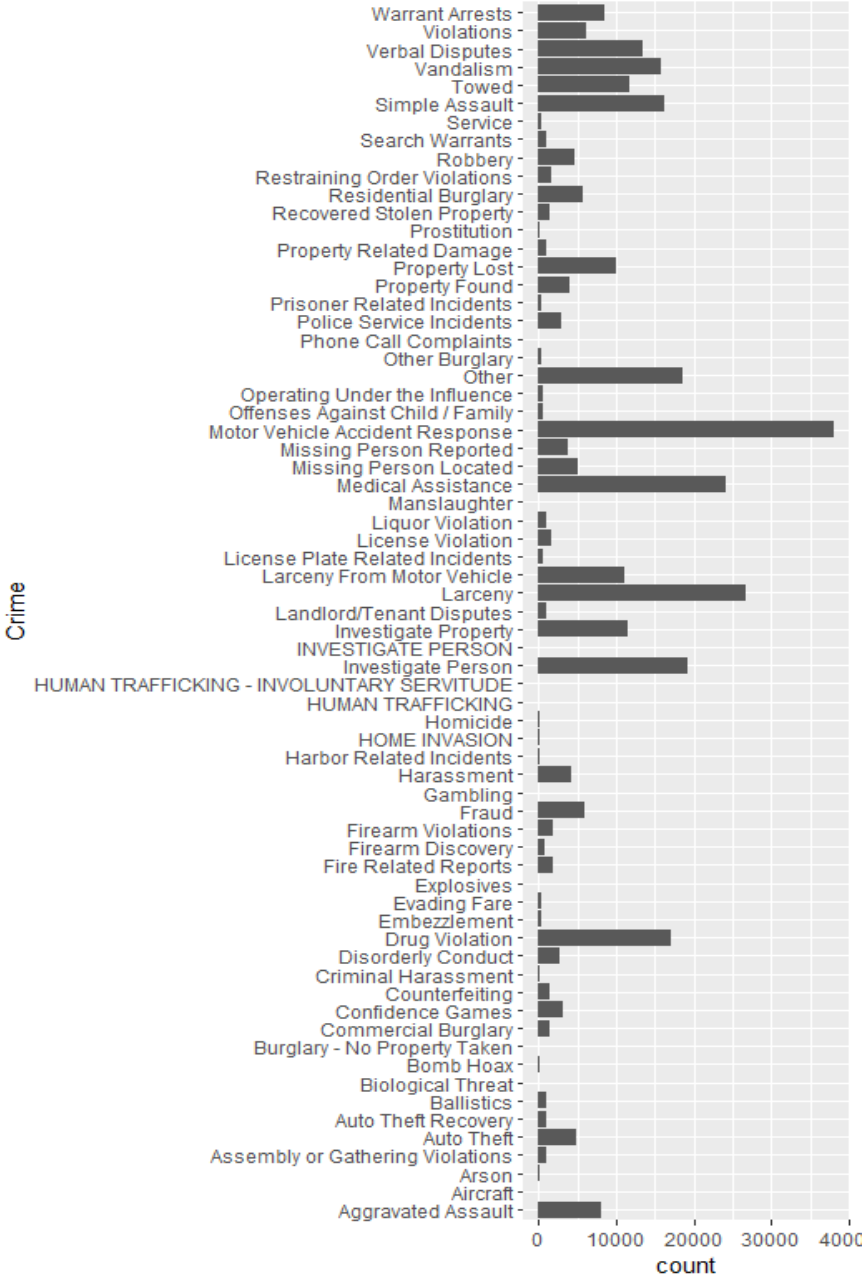
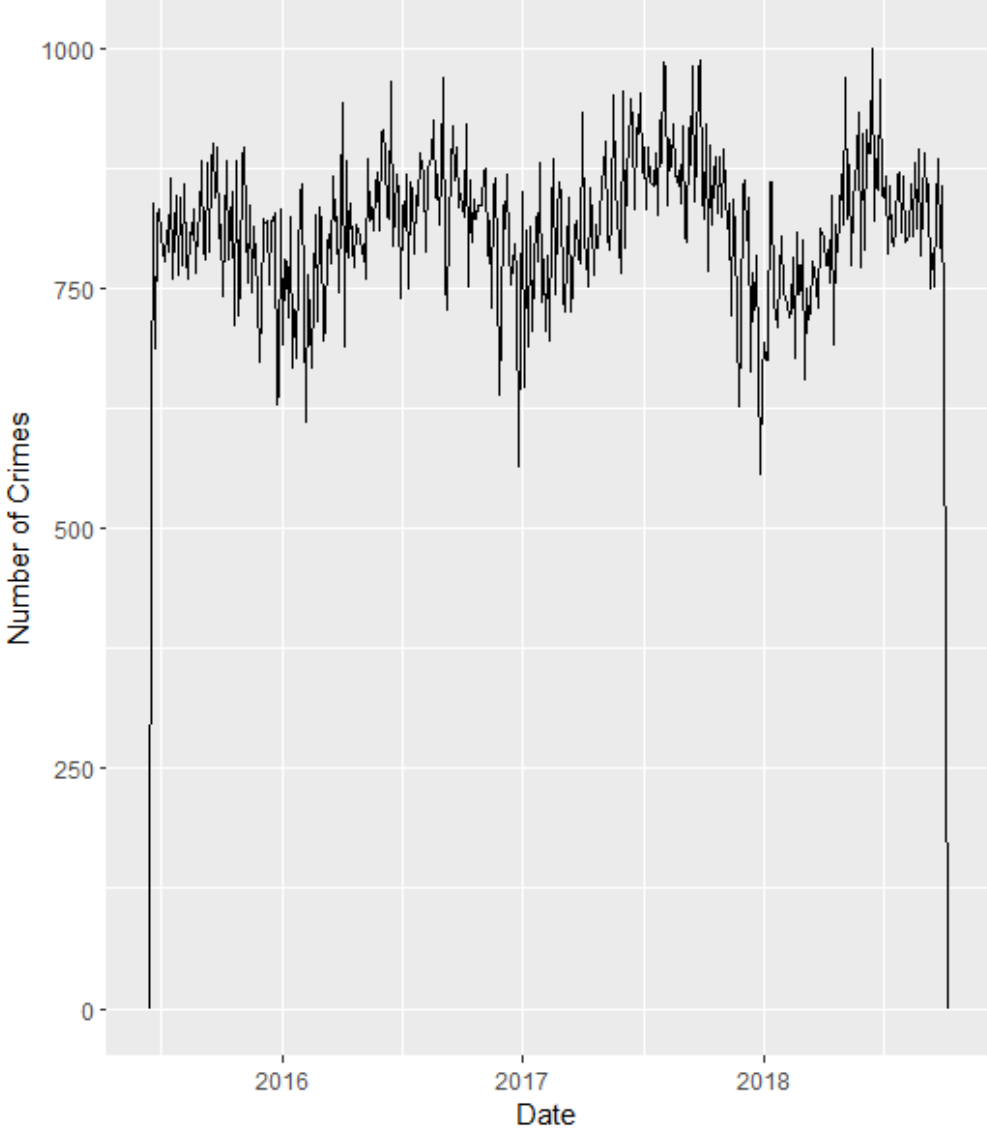
Crime Analysis over District, Month, Year and Day of the Week



Top Crime and Area Analysis



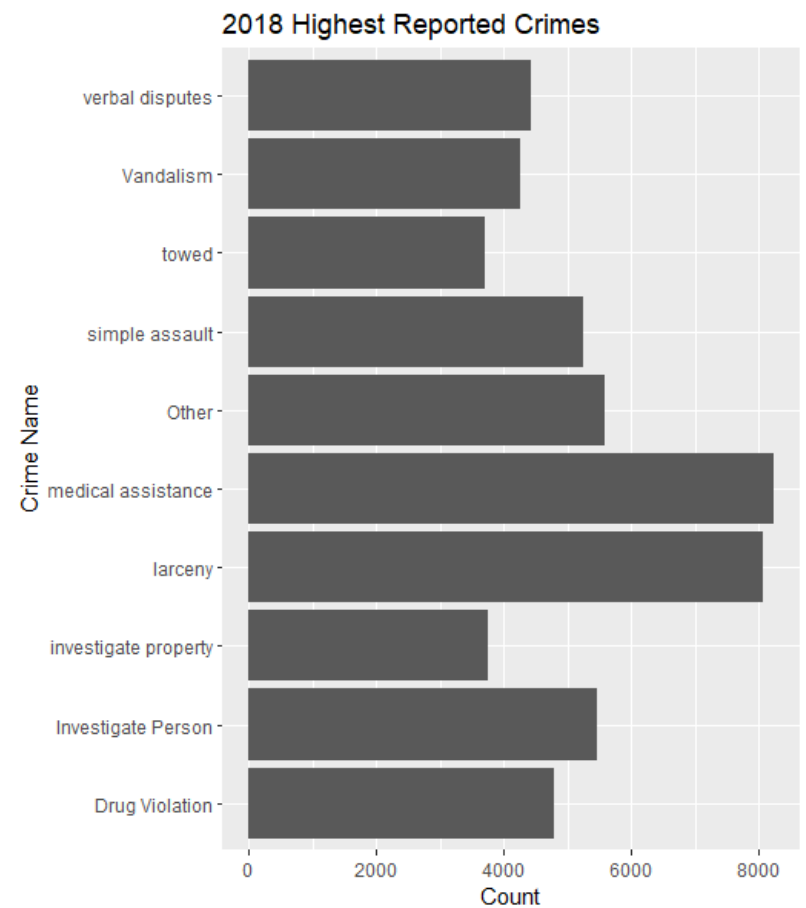
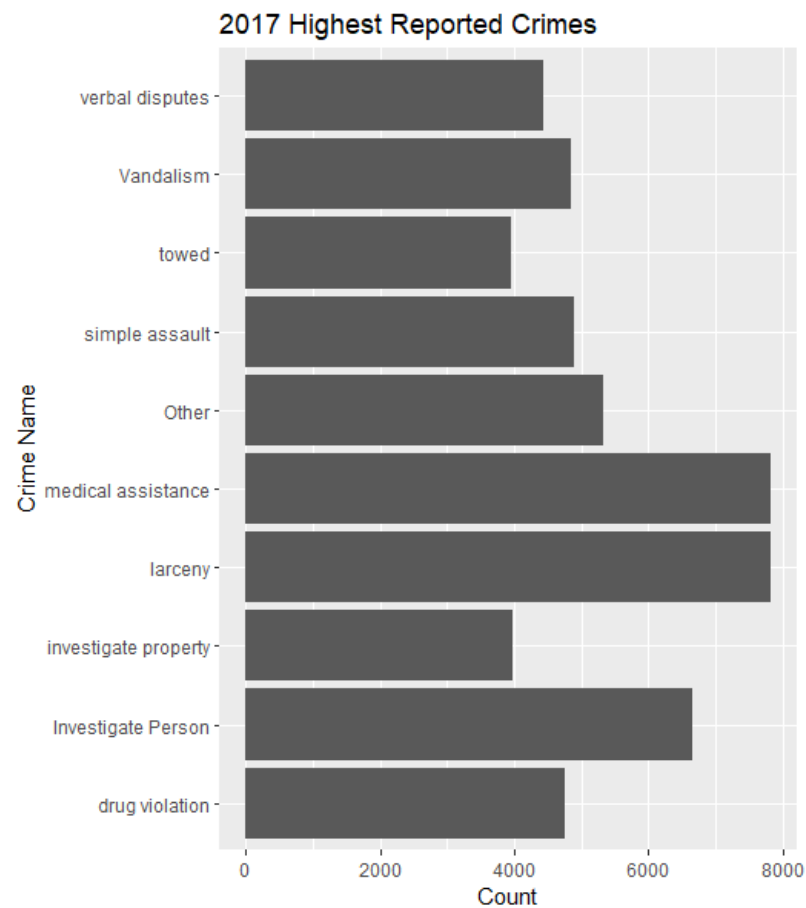
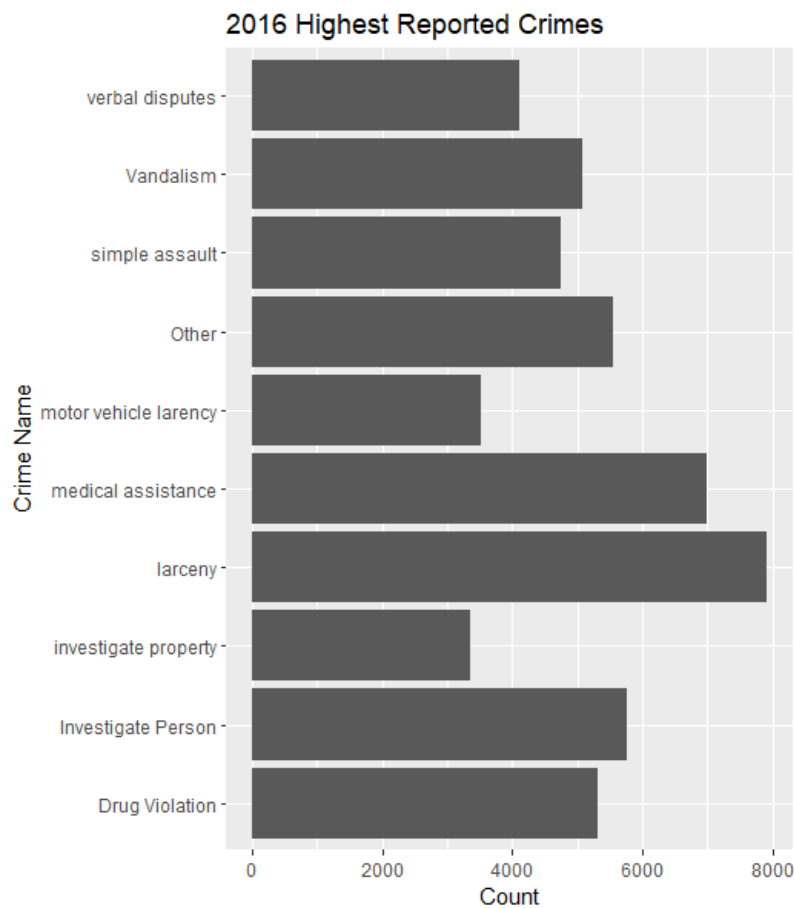
Crime Frequency



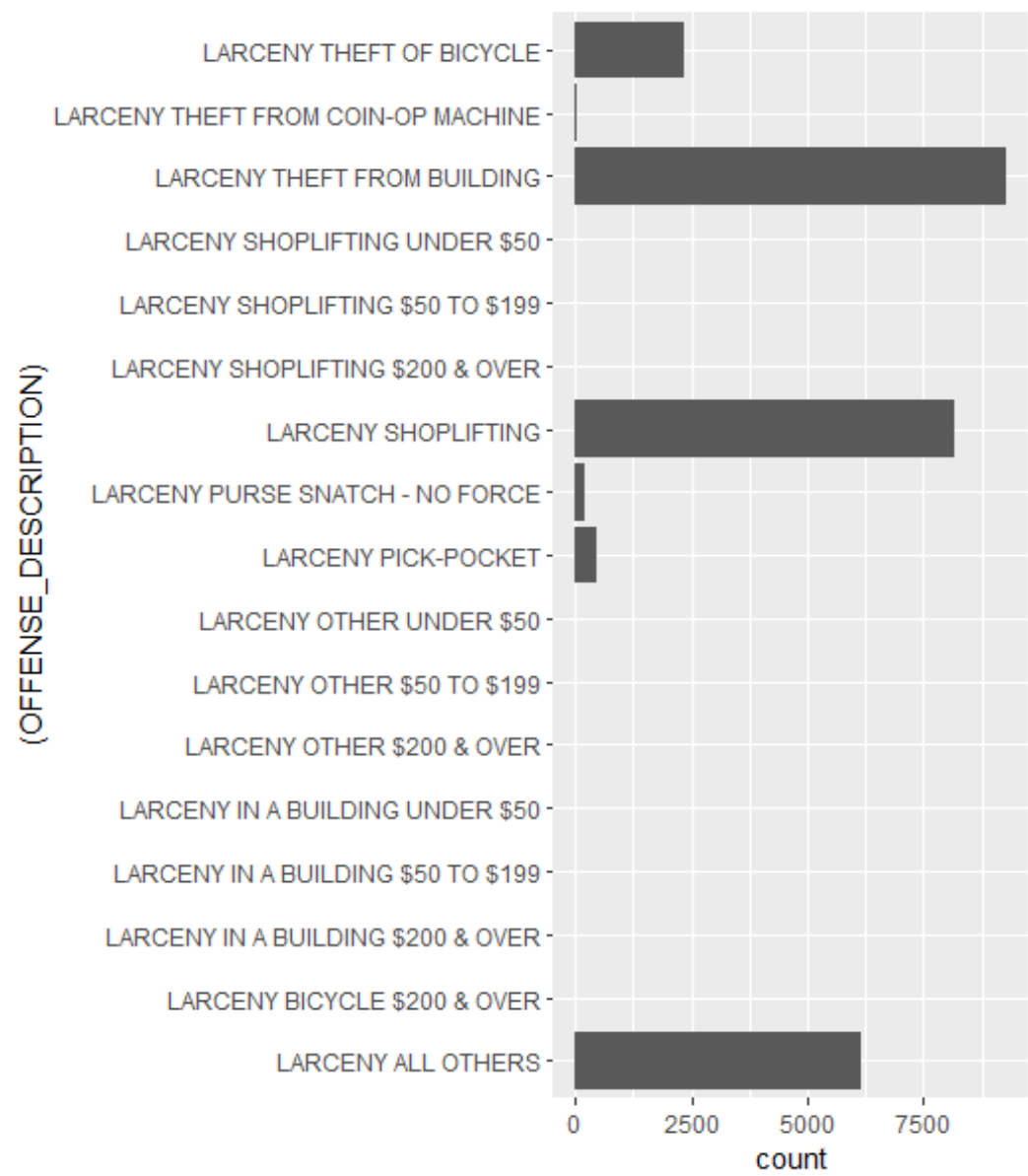
Heatmap of Crime By Location



Crime Analysis Over Time Period 2016-18

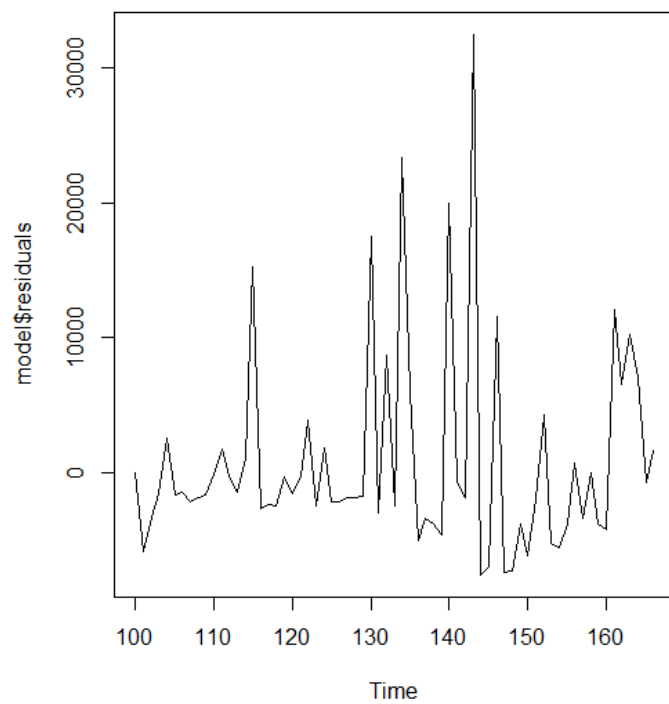


Deep Crime Analysis

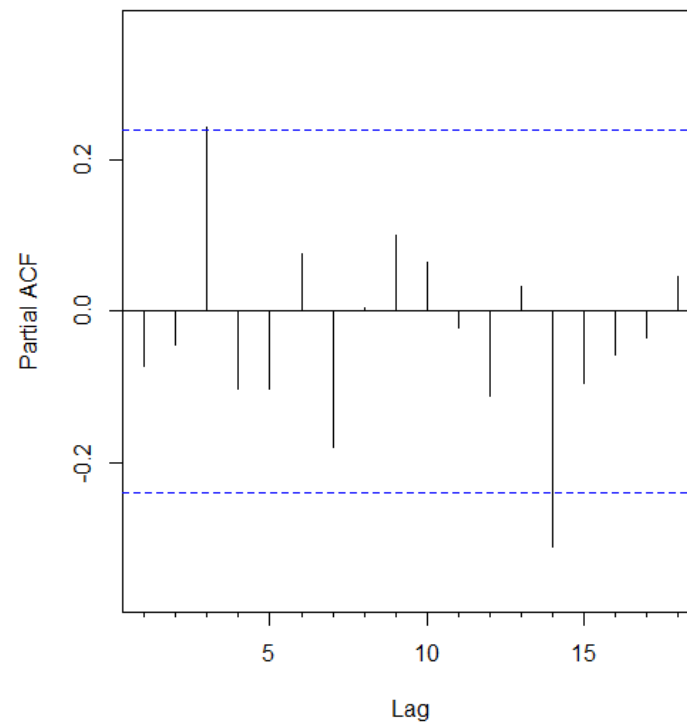


Time Series Forecasting

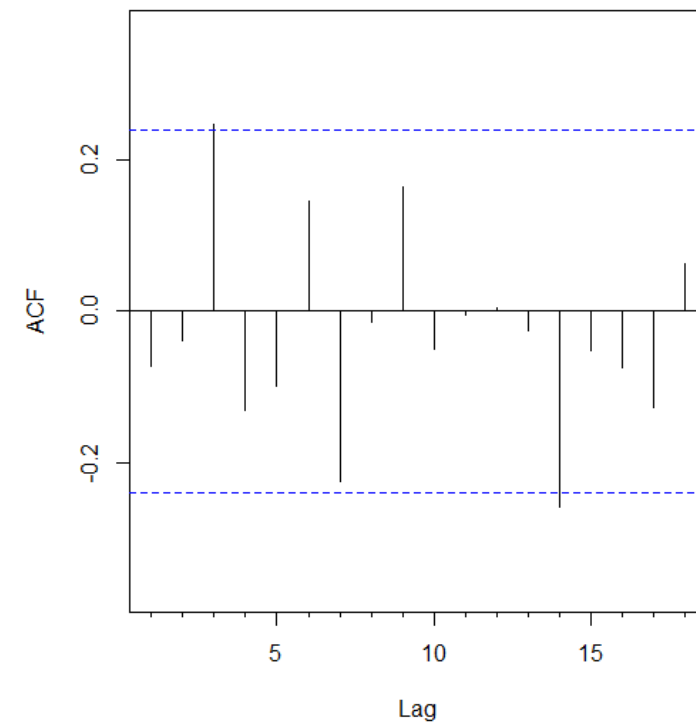
auto arima model



Series ts(model\$residuals)



Series ts(model\$residuals)



Regression Analysis

```
> summary(wage1)
      wage      educ      exper      tenure      nonwhite      female      married      numdep      smsa      northcen      south
Min.   : 0.530  Min.   : 0.00  Min.   : 1.00  Min.   : 0.000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.000  Min.   :0.0000  Min.   :0.000  Min.   :0.0000
1st Qu.: 3.330  1st Qu.:12.00  1st Qu.: 5.00  1st Qu.: 0.000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.000  1st Qu.:0.0000  1st Qu.:0.000  1st Qu.:0.0000
Median : 4.650  Median :12.00  Median :13.50  Median : 2.000  Median :0.0000  Median :0.0000  Median :1.0000  Median :1.000  Median :1.0000  Median :0.000  Median :0.0000
Mean   : 5.896  Mean   :12.56  Mean   :17.02  Mean   : 5.105  Mean   :0.1027  Mean   :0.4791  Mean   :0.6084  Mean   :1.044  Mean   :0.7224  Mean   :0.251  Mean   :0.3555
3rd Qu.: 6.880  3rd Qu.:14.00  3rd Qu.:26.00  3rd Qu.: 7.000  3rd Qu.:0.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:2.000  3rd Qu.:1.0000  3rd Qu.:0.750  3rd Qu.:1.0000
Max.   :24.980  Max.   :18.00  Max.   :51.00  Max.   :44.000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :6.000  Max.   :1.0000  Max.   :1.000  Max.   :1.0000

      west      construc      ndurman      trcommpu      trade      services      profserv      profocc      clerocc      servocc
Min.   :0.0000  Min.   :0.00000  Min.   :0.0000  Min.   :0.00000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000
1st Qu.:0.0000  1st Qu.:0.00000  1st Qu.:0.0000  1st Qu.:0.00000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000
Median :0.0000  Median :0.00000  Median :0.0000  Median :0.00000  Median :0.0000  Median :0.0000  Median :0.0000  Median :0.0000  Median :0.0000  Median :0.0000
Mean   :0.1692  Mean   :0.04563  Mean   :0.1141  Mean   :0.04373  Mean   :0.2871  Mean   :0.1008  Mean   :0.2586  Mean   :0.3669  Mean   :0.1673  Mean   :0.1407
3rd Qu.:0.0000  3rd Qu.:0.00000  3rd Qu.:0.0000  3rd Qu.:0.00000  3rd Qu.:1.0000  3rd Qu.:0.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:0.0000  3rd Qu.:0.0000
Max.   :1.0000  Max.   :1.00000  Max.   :1.0000  Max.   :1.00000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000

      lwage      expersq      tenursq
Min.   :-0.6349  Min.   : 1.0  Min.   : 0.00
1st Qu.: 1.2030  1st Qu.: 25.0  1st Qu.: 0.00
Median : 1.5369  Median :182.5  Median : 4.00
Mean   : 1.6233  Mean   :473.4  Mean   : 78.15
3rd Qu.: 1.9286  3rd Qu.:676.0  3rd Qu.:49.00
Max.   : 3.2181  Max.  :2601.0  Max.  :1936.00

> |
```

Regression (cont.)

Simple Regression

```
table1
=====
Dependent variable:
-----
(wage)
-----
educ          0.541***
              (0.053)
Constant      -0.905
              (0.685)
-----
Observations      526
R2                0.165
Adjusted R2       0.163
Residual Std. Error 3.378 (df = 524)
F Statistic      103.363*** (df = 1; 524)
=====
Note:            *p<0.1; **p<0.05; ***p<0.01
```

Multivariate Regression

```
table2
=====
Dependent variable:
-----
log(wage)
-----
educ          0.092***
              (0.007)
exper         0.004**
              (0.002)
tenure        0.022***
              (0.003)
Constant      0.284***
              (0.104)
-----
Observations      526
R2                0.316
Adjusted R2       0.312
Residual Std. Error 0.441 (df = 522)
F Statistic      80.391*** (df = 3; 522)
=====
Note:            *p<0.1; **p<0.05; ***p<0.01
```

Heteroscedasticity Test (Breusch-Pagan Test and White Test)

```
> lmtest::bptest(model1)
```

```
studentized Breusch-Pagan test
```

```
data: model1
```

```
BP = 15.306, df = 1, p-value = 9.144e-05
```

```
> lmtest::bptest(model2)
```

```
studentized Breusch-Pagan test
```

```
data: model2
```

```
BP = 10.761, df = 3, p-value = 0.01309
```

```
> skedastic::white_lm(model1)
```

```
# A tibble: 1 × 5
```

	statistic	p.value	parameter	method	alternative
	<dbl>	<dbl>	<dbl>	<chr>	<chr>
1	23.2	0.00000898	2	White's Test	greater

```
> skedastic::white_lm(model2)
```

```
# A tibble: 1 × 5
```

	statistic	p.value	parameter	method	alternative
	<dbl>	<dbl>	<dbl>	<chr>	<chr>
1	14.1	0.0287	6	White's Test	greater

Multicollinearity Test (VIF)

```
> vif(model2)
```

```
educ    exper    tenure  
1.112771 1.477618 1.349296
```



**THANK
YOU**