

Cambridge Crime Analysis

Pankaj Shah

12/24/2018

- **Introduction**
- **Data**
- **Duplicated data**
- **Date of Report**
- **Crime date time**
- **Shift**
- **Reporting Area**
- **Neighborhood**
- **Top 10 crimes**
- **Crimes/Year**
- **Location**
- **Hour**
- **Crimes | Day | Month**
- **Hit and Run**
- **Comparing top two crime**
- **HeatMap**
- **Crimes by Month %**
- **Corplot**
- **Summary: Time series Breakpoint**
- **Crime_Code/Year**
- **Crime_code/Month**
- **Top crime in each Neighborhood**
- **Top Crime in each Reporting Area**
- **Top Crime in each Neighborhood/Reporting Area**
- **Top Two Crime Neighborhoods**

Introduction

The Cambridge Police Department have announced that crime in Cambridge dropped for the sixth consecutive year to unprecedented levels not recorded since 1961. The results were released after the Cambridge Police Department's Crime Analysis Unit recently finalized figures. Analysis of the overall statistics shows declines in both property and violent crime - 8 percent and 7 percent, respectively for the year 2016. Violent crime has declined by 75 percent since 1990, the most violent year on record in Cambridge. The total number of reports of murder and non-negligent homicide, rape, robbery, aggravated assault, burglary, motor vehicle theft, larceny-theft, and arson was at its lowest since 1961 [Comparing to External Data Source]

These historically low crime rates can be reflect as the high quality and excellent service from the Cambridge Police Department delivers to their community and I believe data analysed and used in most efficient way to work. Cambridge crimes considered serious went down 10 percent in the first four months of this year 2018 compared with the same period last year. We have data for only 6 momths of 2018. While 42 fewer crimes were reported this year when compared with the same time period last year. According to police in the year 2017, the City of Cambridge recorded its lowest crime index total since 1963. That downward spiral of serious criminal activity in the city has continued throughout the first four months of 2018 as we have seen on our datasets.

According to the cambridge Police Department, most of the people who commit crimes in Cambridge don't live there, so it is probably not simply a story that criminals were priced out and moved away because of housing market taking a spike. Upon inquiring I found out that The Cambridge Police Department works very closely with BAIR Analytics Inc in controlling crimes in Cambridge Neighborhood. They have recently partnered to provide a new way for the public to stay informed about crime in Cambridge. The Cambridge Police Department now has an online crime map called RAIDS Online (www.raidsonline.com) that maps and analyzes crime data, alerts Cambridge citizens about crimes in their area, and allows the Cambridge Police Department to quickly alert the public about crimes as they occur.

I really enjoyed diving into the datasets and here's why I think you will too. The datasets is very limited with 68,000 Observations and 7 variables. Like other project report which I have done in Boston Crime Analysis, this datasets was limited in multi layered fators like latitude, longitude. Anyways I try to do best Whatever I could with this Datasets as alternative approach and different Exploratory Analysis.

Executive Summary Information Description

- Top most Crimes that are reported in Cambridge are Shoplifting which is more prevalant in East Cambridge, Larcency from Motor Vehicles, Hit and Run, Domestic dispute, Larcency from Person are other types of crimes which are reported most.

- Crime in Cambridge dropped for the sixth consecutive year to unprecedented levels not recorded since 1961.
- In East Cambridge Area we see more shop lifting. Presence of Galleria Mall.
- In CambridgePort, West Cambridge & Mid Cambridge Area most of the crimes are related to Larcency from Motor Vehicle.
- Most of the crimes related to Hit and Run occurs in North Cambridge Area.
- Domestic Dispute are more prevalent in Area 4, Inman/Harrington neighborhood.
- Hit and Run is top most crime in MIT, Highlands area but very few compare to other neighborhoods.
- Top most crime in Strawberry Hill is Domestic Dispute which is again very less compare to other Neighborhoods.
- Larcency from Person are more prevalent in Riverside Neighborhood.
- Most of the crimes are reported from Cambridgeport followed by East Cambridge Neighborhood and Area 4.
- 100 Cambridgeside place is mostly reported location followed by 600 Massachusetts Ave and then 500 Massachusetts Avenue.
- When breaking down 13 Neighborhood we find 4 groups which shows similar pattern throughout the time period. Similar is the case in half year of 2018 data.
- The crime is lowest at 5AM whereas highest is reported around 5PM.
- Major crimes account for 70% of crime datasets and 30% of crime datasets comes from minor crime.
- As per the report I did find truth that 2017 was lowest reported crime compare to rest of the year.
- Crimes picks up in Summer month and drops down during the winter month.
- Larcency from motor vehicle is high during the summer months in west Cambridge as well as Cambridgeport Area.
- On our heatmap we did find out that Feb 23rd around 4'o clock we have more Hit and Run cases than any others. It could be one single case or more multiple case on that day.
- Around December the Larcency from property is reported more than any other crime in some neighborhoods.
- There are many insights inside each plot and bar graph to know more.

Data

How I define all these datasets

- df : Full datasets from Jan 2009 to June 2018.
- df_2018 : Datasets from Jan 2018 to June 2018.
- df_2017 : Datasets from Jan 2017 to Dec 2017.
- ca_crime_df: Datasets from Jan 2009 to Dec 2017.

LIBRARY Map of Cambridge Data Info

[Code](#)

Let's import our csv file

[Code](#)

Duplicated data

Lets check for any duplicates

[Code](#)

```
## [1] 68008
```

[Code](#)

```
## # A tibble: 3 x 8
##   file_number dupe_count date_of_report crime_date_time crime
##   <chr>          <int> <chr>           <chr>
## 1 2011-09725      3 12/09/2011 06... 12/09/2011 18:... Homi...
## 2 2011-09725      3 12/09/2011 06... 12/09/2011 18:... Homi...
## 3 2011-09725      3 12/09/2011 06... 12/09/2011 18:... Homi...
## # ... with 3 more variables: reporting_area <int>, neighborhood <chr>,
## #   location <chr>
```

[Code](#)

As out of 68010 we have only 3 duplicated data. We kept only non duplicated data into our datasets. Although it doesn't affect our results but when we see something we should do something.

Glimpse of Unique datasets

[Code](#)

```
## # Observations: 68,008
## # Variables: 7
## $ file_number      <chr> "2018-04386", "2018-04383", "2018-04382", "201...
## $ date_of_report   <chr> "06/30/2018 10:50:00 PM", "06/30/2018 06:30:00...
## $ crime_date_time <chr> "06/30/2018 22:46 - 22:49", "06/30/2018 18:30"...
## $ crime            <chr> "Commercial Robbery", "Larceny of Services", "...
## $ reporting_area   <int> 101, 708, 507, 1002, 1113, 402, 704, 1110, 106...
## $ neighborhood     <chr> "East Cambridge", "Riverside", "Cambridgeport"...
## $ location          <chr> "200 MONSIGNOR O'BRIEN HWY, Cambridge, MA", "0 ...
```

Date of Report

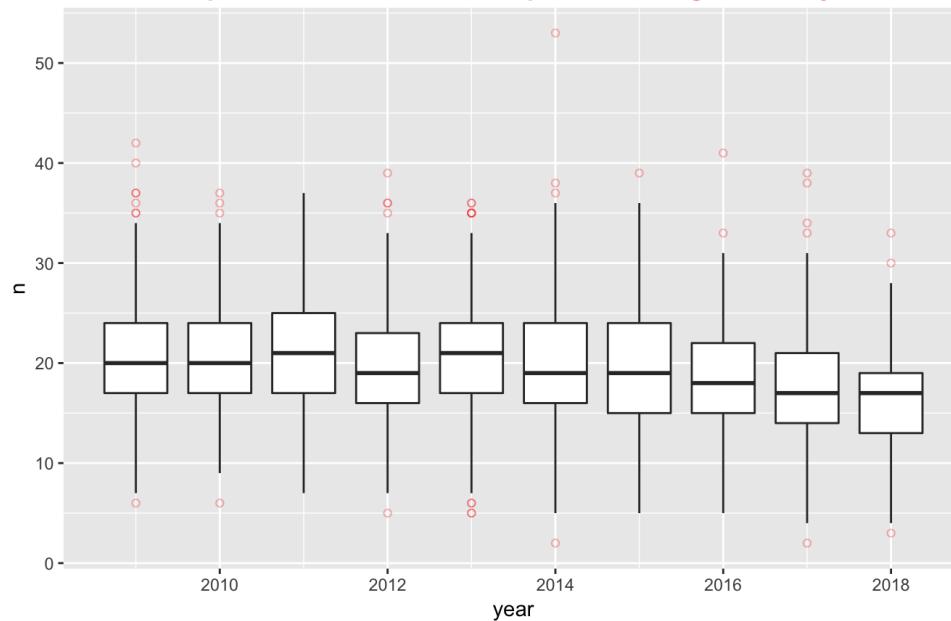
Lets starts to work with Date of Report Column

[Code](#)

```
## [1] "2018-06-30 22:50:00 UTC" "2018-06-30 18:30:00 UTC"
## [3] "2018-06-30 18:16:00 UTC"
```

[Code](#)

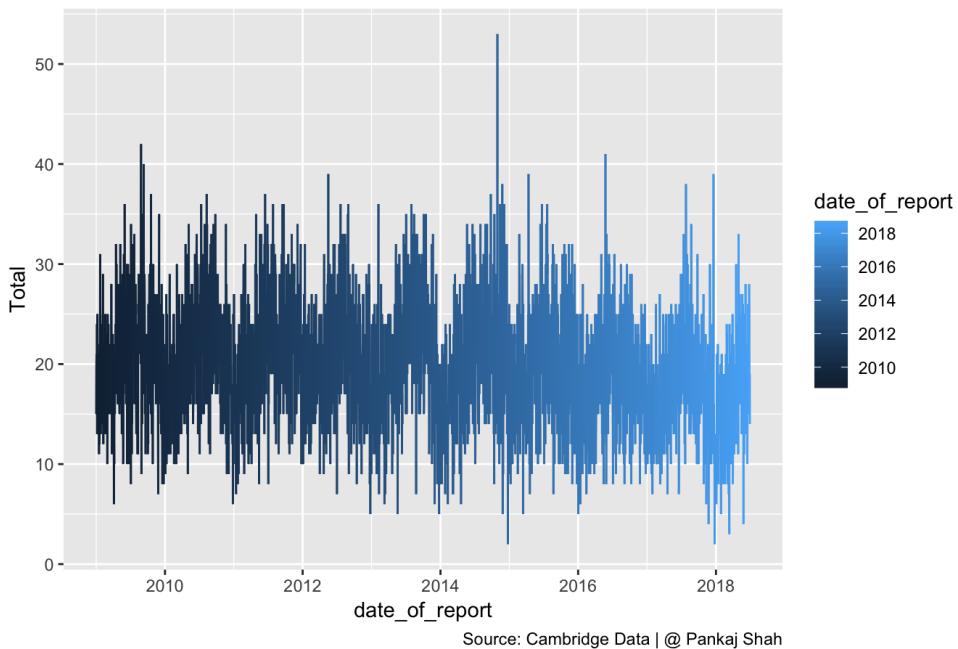
Box plot of Crime Incident Reported throughout the year



Source: Cambridge Data | @ Pankaj Shah

[Code](#)

Time Series Pattern for Distribution of crime

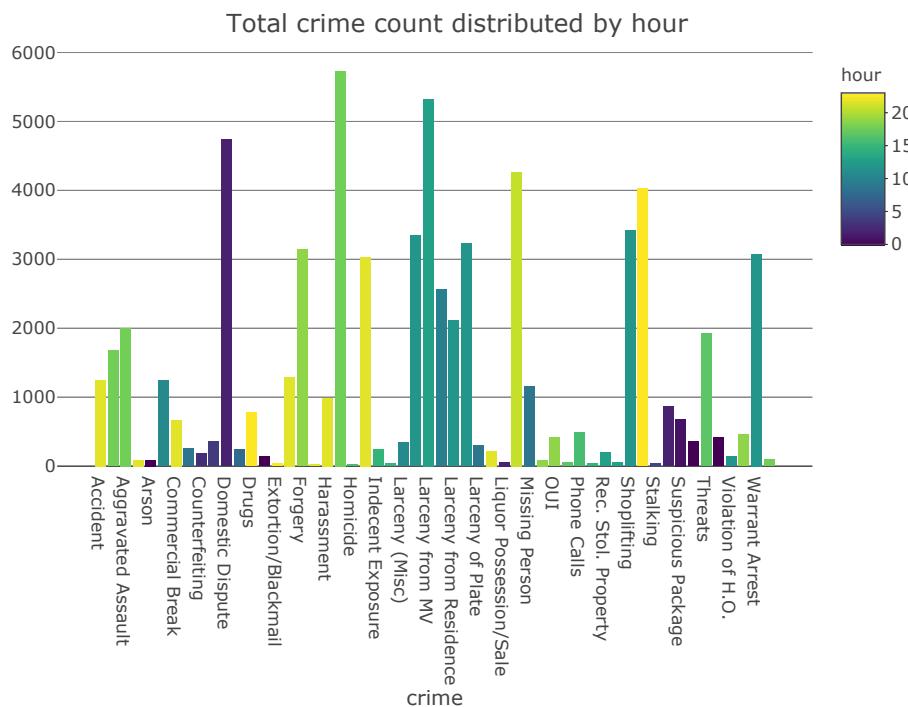


Crime date time

Code

```
## [1] "06/30/2018 22:46 - 22:49" "06/30/2018 18:30"
## [3] "06/30/2018 18:00 - 18:45"
```

Code



Here we have two date observation in same column seperated by space. These two variables are wrapped around quotes. By observing in detail I have found out that one timestamp is when the crime has been reported and another time stamp is when the case was closed/dismissed.Lets seperate these one column into two column. If the crime was closed on same day then we will see the end date column is missing.In the beginning we saw 11 missing values in “crime_report_date” those are the same column which have missing crime_report_time. We can fill these missing values from date of report which is slightly different from the crime report date but doesn't vary more.

Highlights

- Hit and Run happen During the day.
- Mostly Domestic dispute are reported either in evening or during the day.
- Shop Lifting happens during the day.
- Warrant Arrest happens mostly during the day.
- Threats call are reported mostly in day.
- Very less weapon Violations.
- Larceny from Motor Vehicles happens around Evening.
- Larceny from Building happens during the day when people are not in house.
- Forgery are reported mostly in morning and day.
- Simple Assault are reported late at night.

Shift

Lets breakdown our crime data by shift so that we know what types of crime occur through out the day. I believe the officer works may be in 3 shift pattern but due to lack of validity of our source I will break down the data into 4 shifts so that its easy for us to visualise the patterns and crime easier.

Code

```
##  
## 00-06 06-12 12-18 18-24  
## 9120 17765 24686 16409
```

Code

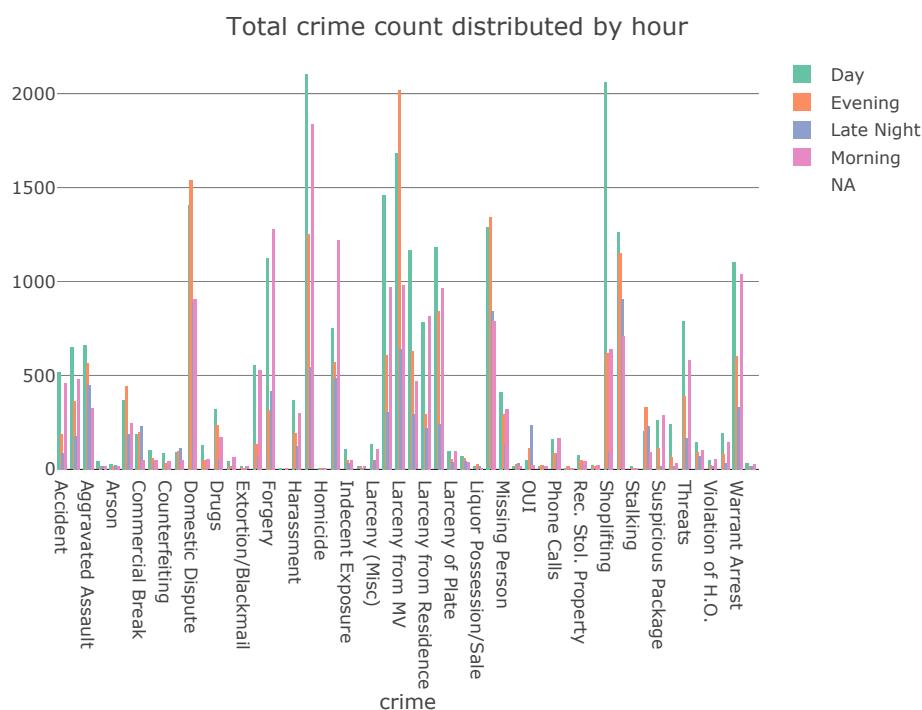
```
##  
##      Day    Evening Late Night   Morning  
## 0.3631362 0.2413798 0.1341571 0.2613269
```

Most of the crimes are reported during the day(12pm-18pm) almost 36 % followed by Morning (6am-12pm) crimes around 26% and least happens between 12am-6am.

- Hit and Run & Shopliftings are reported during the day.
- Larcency and Domestic Dispute are reported during the Evening.
- Forgery and House breaks are reported in the Morning.
-

Hour

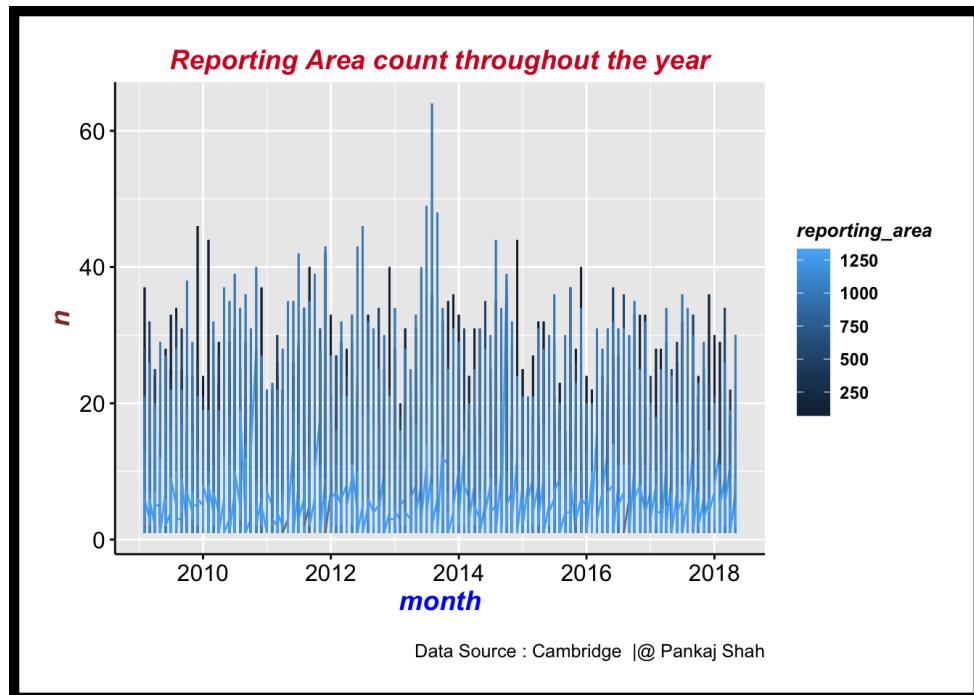
Code



you can Zoom in/out using hoover tooltips to see details

Reporting Area

Code



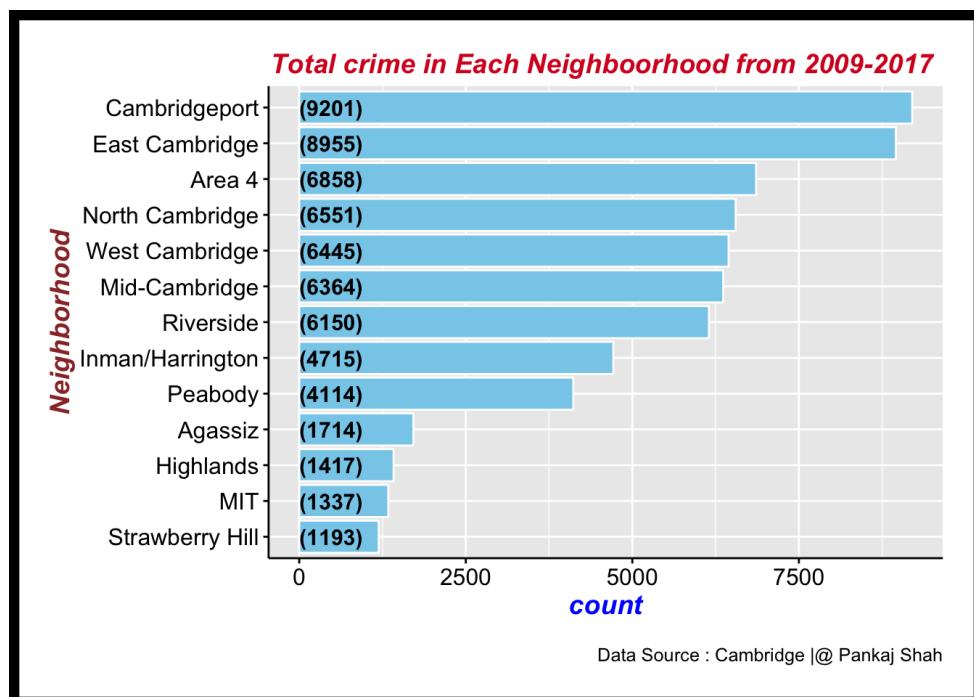
Neighborhood

Recent dataframe_df refers to datasets from 2009 to 2017. Would like to call Recent as it is complete datasets that why. Don't confuse with 2018 partial dataframe. only for some comparison of Crimes, Neighborhood and some other purpose we will use 2018 Partial data. Most of the analysis will be done on full datasets from 2009 to 2017.

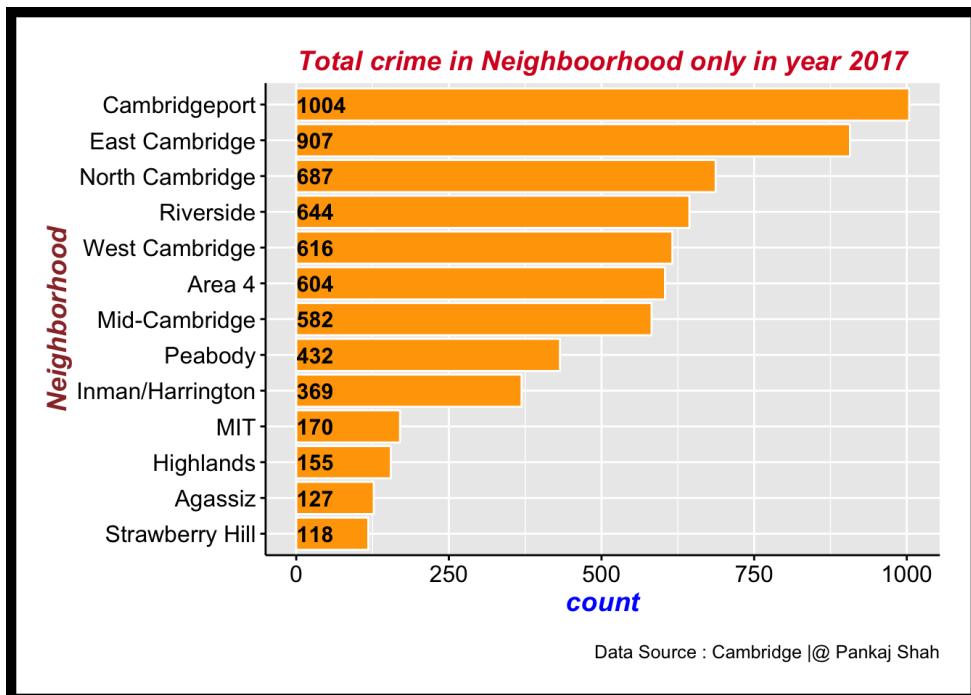
Check the Cambridge Neighborhood Link Below:

https://commons.wikimedia.org/wiki/File:Neighborhood_Map_of_Cambridge,_MA.png#/media/File:Neighborhood_Map_of_Cambridge,_MA (https://commons.wikimedia.org/wiki/File:Neighborhood_Map_of_Cambridge,_MA.png#/media/File:Neighborhood_Map_of_Cambridge,_MA)

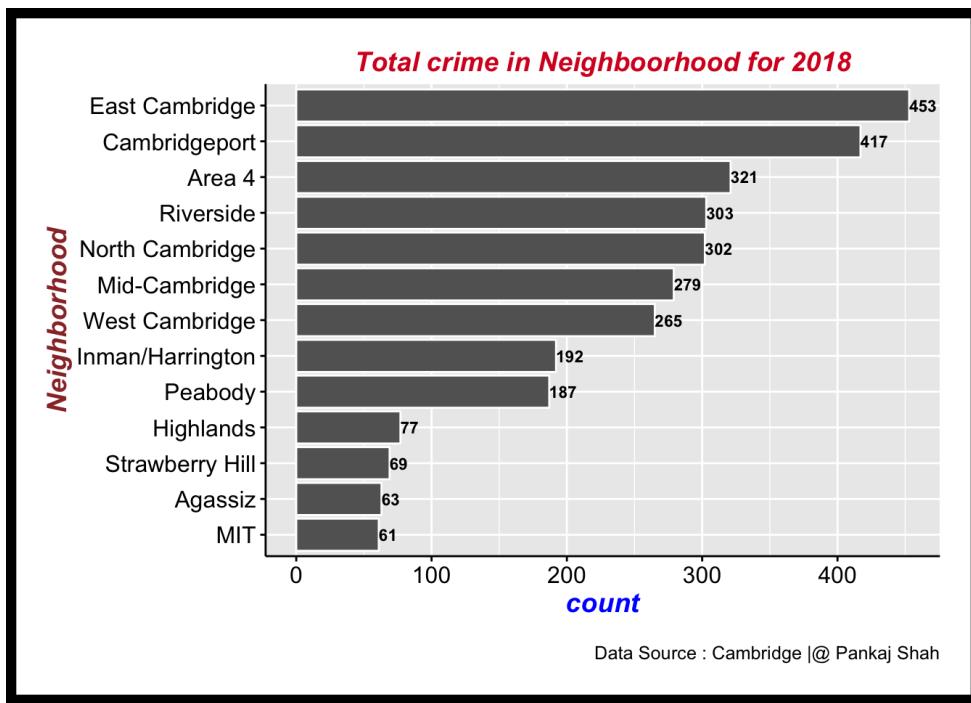
Code



Code



Code



Code

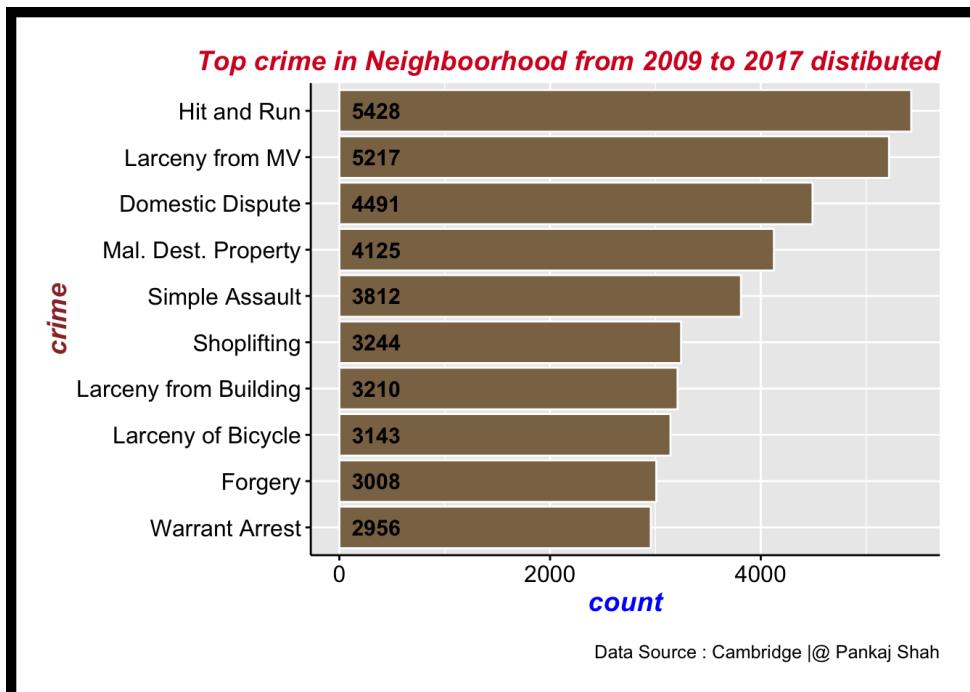
We can see there is slight change in the position but overall the number of incident being report is almost the same. Here we compare past data from 2009 to 2017, year alone 2017 and half a year of datasets 2018 and found out that the pattern remains the same. Same trend apart from little ups and downs which is good thing for Cambridge Police as they can focus on same pattern rather than being surprised and need to use more force or more planning. It is easy to control crime when patterns remain constant through the cycle. It is easy to predict.

Both East Cambridge and Cambridgeport reports most of the crimes.

External Data source: If I could have external validated data source with population census. I could have normalise these data by population census count to see if geography was the reason for the counts in crime. Hypothetically the probability of crime increases as population density per square foot increases. Another theory could be exploring the datasets with poverty/education rate for causal inferences. For now we will just concentrate on the data that is available through Cambridge Data Portal.

Top 10 crimes

Code



Code

Show 10 ↑ entries

	crime	count
1	Hit and Run	5428
2	Larceny from MV	5217
3	Domestic Dispute	4491
4	Mal. Dest. Property	4125
5	Simple Assault	3812
6	Shoplifting	3244
7	Larceny from Building	3210
8	Larceny of Bicycle	3143
9	Forgery	3008
10	Warrant Arrest	2956

Code

You can choose how many entries.

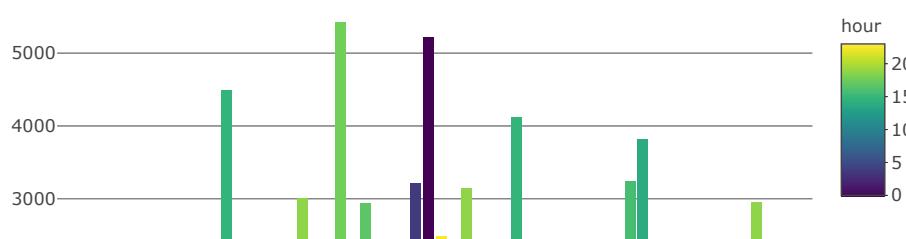
We have Hit and Run related to Motor Vehicle crime on top of List followed by Larceny from motor vehicle and then to domestic dispute from year 2009 to end of 2017. In our Boston Datasets we also saw crimes related to Motor Vehicle coming on top of the list. Seems like Motor Vehicel related crimes are more frequent in two close neigborhood.

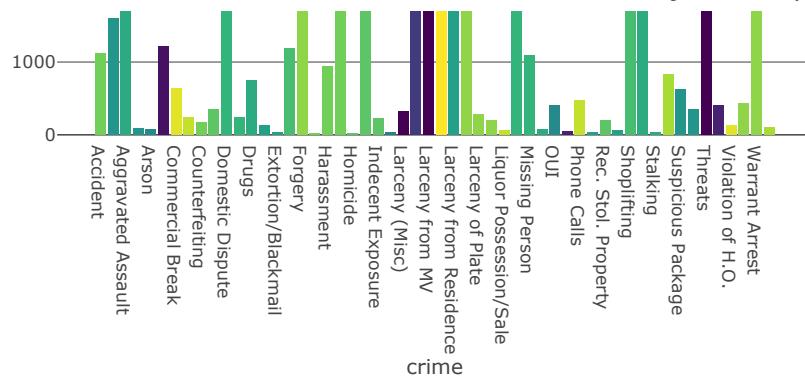
Lets see if the crimes have been decreasing over the year or did at some point the crimes have different pattern. Lets revisit the count to see if there was decreasing crimes year after year.

other crimes

Code

Total crime count distributed by hour

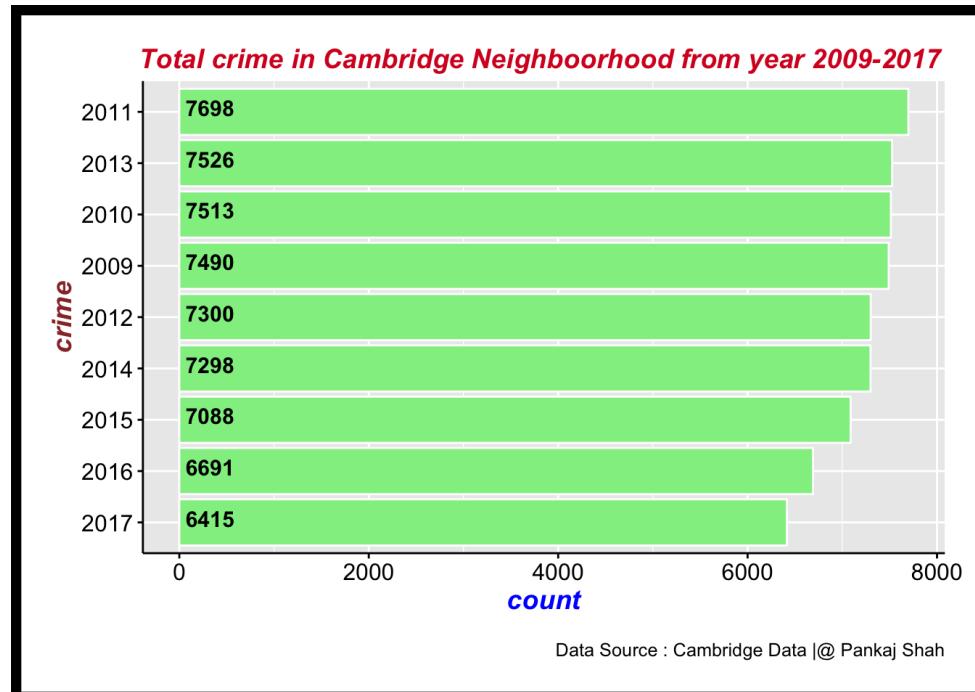




We can see other minor crimes. Seems like almost 80% crimes are coming from top 10 crimes and rest 20% crimes are made up of all small minor crimes.

Crimes/Year

Code



Code

Show **10** entries

year	count
1	7698
2	7526
3	7513
4	7490
5	7300
6	7298
7	7088
8	6691
9	6415

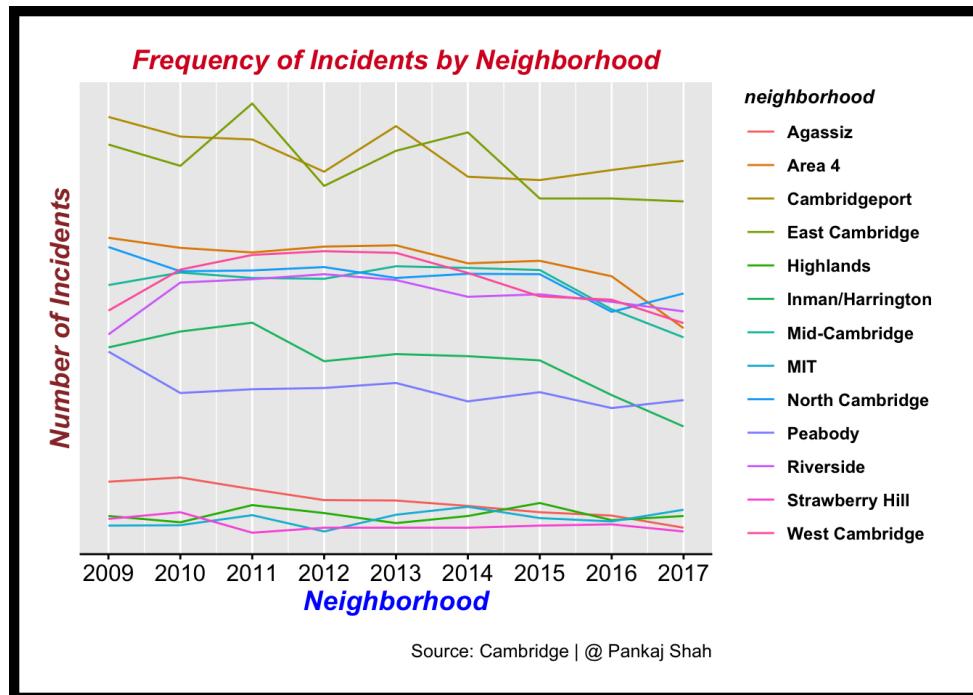
As we can see that from 2009 till 2017, the lowest was reported in 2017. If you compare the numbers they are not so much different, they lie in same ball park. If we compare year 2011 and year 2017 we can see the difference is almost drop in 16% but that is difference between having more crime reported and recent years. Similarly if we compare when they start collecting digital data the change is almost 14% drop which can be said remarkable.

Neighborhood Crime/year

Crime breakdown report by Neighborhood over the years

Lets see how the neighborhood reported crime Over the years have shifted. We will cut off the year 2018 otherwise we will see all the crimes dropping down as we don't have full year data.

Code



Lets break down the anlaysis part by neigborhood. We can see there are 4 different groups formed out of 13 neigborhood. East Cambridge and Cambridgeport remains on top of the list throughout the year. I also found out that Cambridge Police headquarter is also situated in East Cambridge neigborhood which makes perfect sense looking at the crime distributions.

East Cambridge

In year 2011 where most of crimes was reported came mostly from East Cambridge, as we can see spike of crimes and then next year it was drop down even less than 2010. In the year 2014 we can see again Crimes are picking up. Although it remains on top of the list compare to rest of the neigborhood overall the crime patterns seems to drop.

CambridgePort

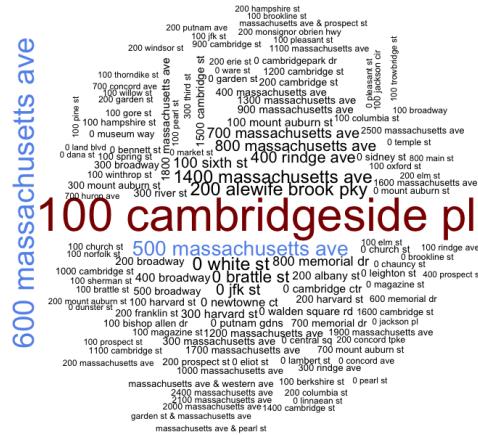
CambridgePort follows different path than East Cambridge taking a peek in year 2013 and dropping down afterwards and staying flat. Both of these two district reports heavily in crime than other neigborhood.

MIT, Strawberry Hill, Highlands stays & Agassiz fall in one group where crimes are reported very less. MIT area is mostly inhabitaed by MIT professionals and campuses. Strawberry Hill resides very close to Belmont and Watertown line where crime rates are reported less than any other cities in Massachusetts(External Data). Cambridge Highlands resides next to strawberry Hill but apart from that it is hard to say anything about it. May be if we have population density datasets and other couple validate source datasets to determine the casual infrence of these 4 neigborhoods crime.

Location

We need to split and clean up the location

Code



So most of the arrest are made in 100 Cambridgeside Place, Upon diagnosing I found out the 100 Cambridge place address to CambridgeSide Gallaria Mall. Although Cambridge Police Department indicates the location is around the vicinity of 100 blocks range. 100 Cambridgeside Place resides in East Cambridge Neighboorhood. So most of the crimes that were reported on East Cambridge were comming from 100 Cambridgeside Place.

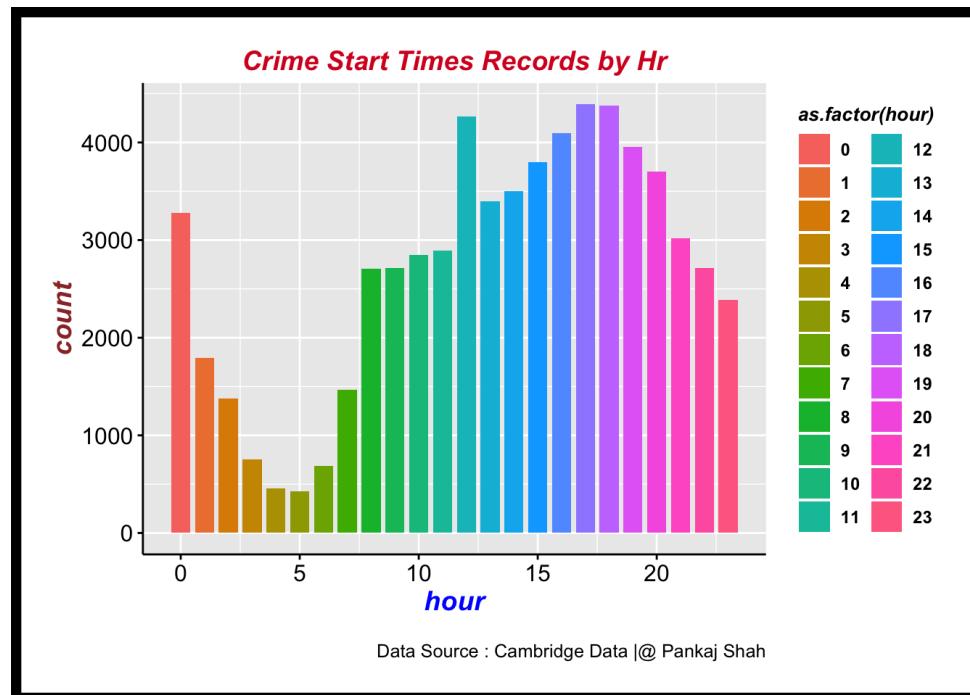
Next was 600 Massachusetts Avenue and 500 Massachusetts avenue which falls in Cambridgeport neightborhood. These two address combines brings up the Cambridgeport area with most of the crimes.

Hour

Breakdown Crime by Hour

Code

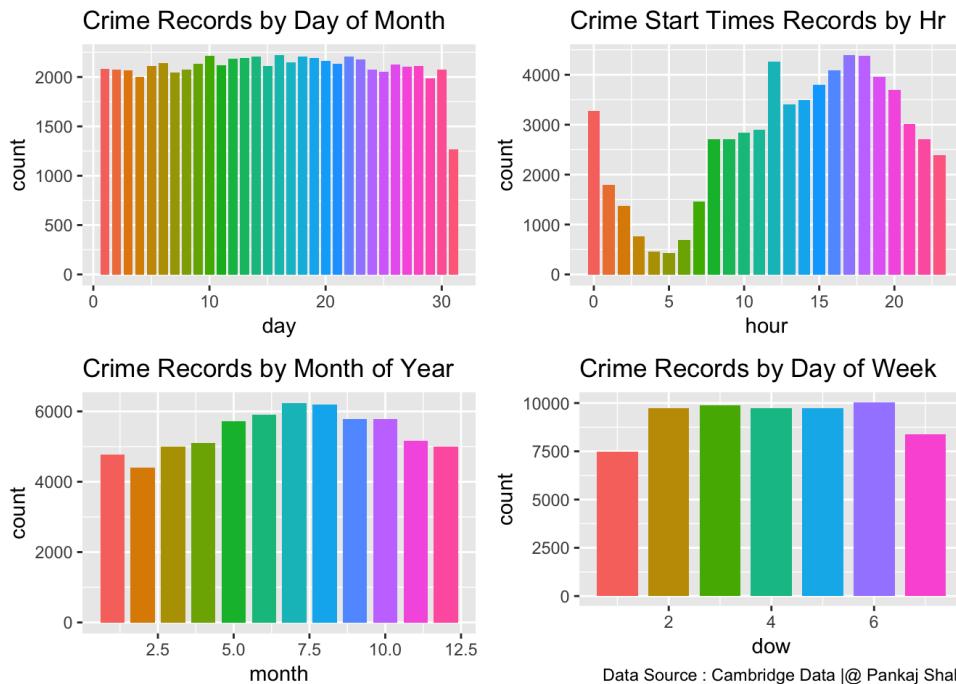
```
## Warning: Removed 28 rows containing non-finite values (stat_count).
```



Interesting the lowest Number of Crime is at 5 am whereas Highest numbers of crime are reported around 5pm.As the day progress the reporting of crime increases and peaks at 12 then falls back and again rise upto 6 and then we see it is dropping back.

Lets see the patterns breaking down by year, month, day of week, day

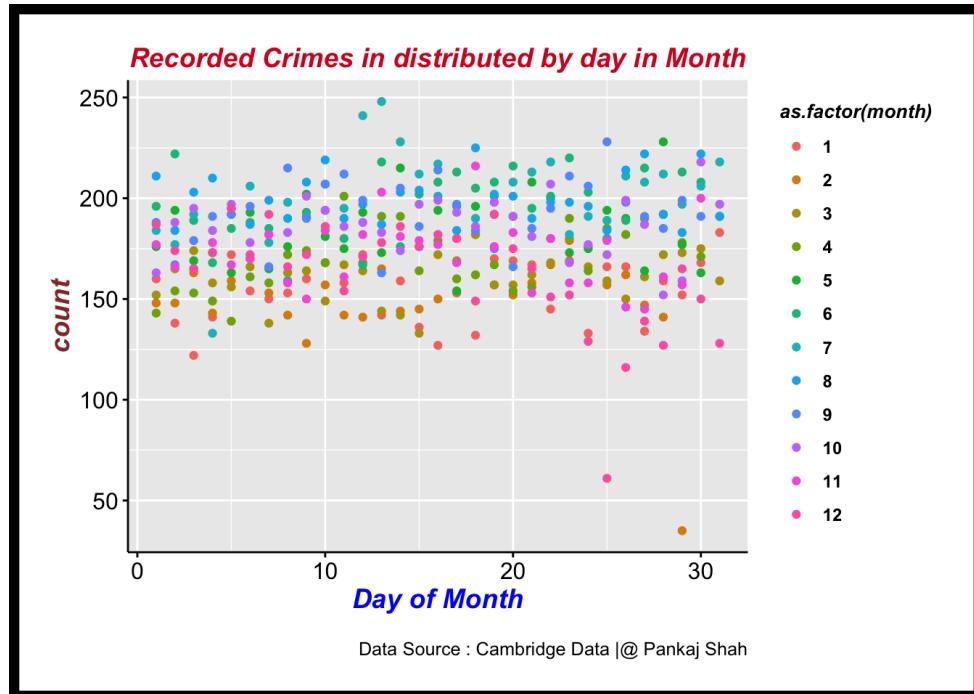
[Code](#)



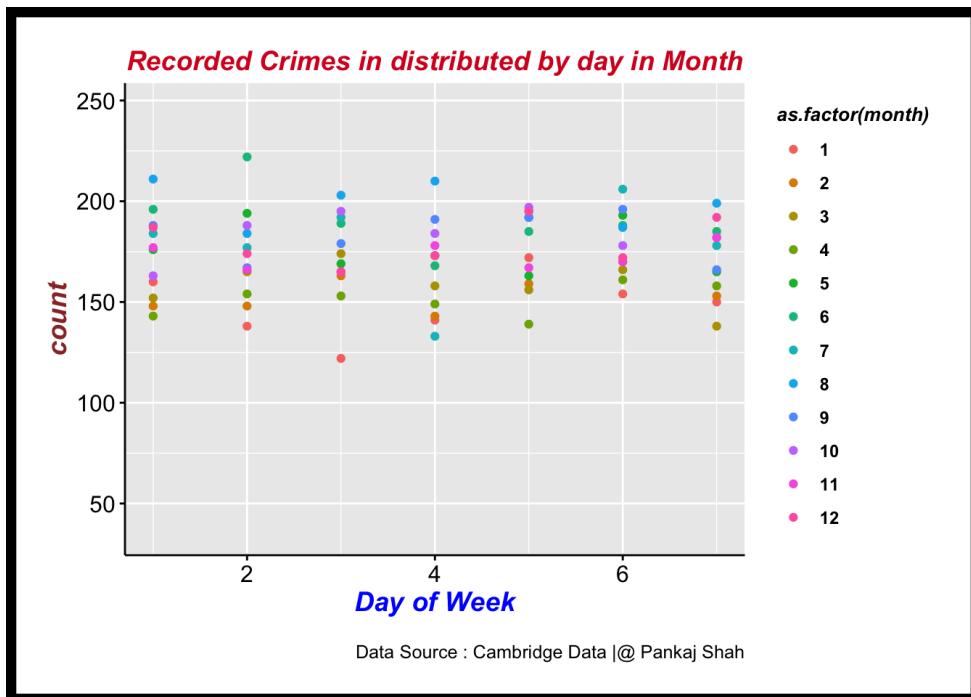
Crimes tend to rise in Summer months mostly around July, August. Crimes records stay same throughout the month. Crimes tend to be high around evening rush hour commute, mostly hit and run is coming top on the list which makes complete sense. Sunday seems to have less crime and Friday seems to have more crime reported. Understand night life as well Sunday being quite Church Day.

Crimes | Day | Month

[Code](#)



[Code](#)

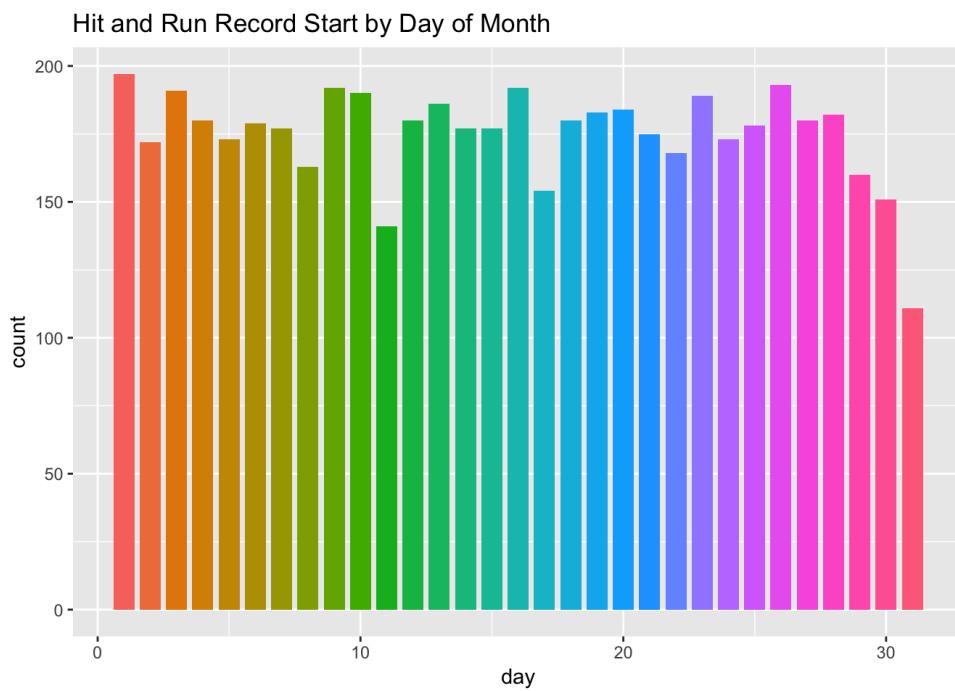


I thought around first week of every month the crime should peak but seems like the crime is distributed throughout the week in same pattern. Range of crime reports stays the same. Only difference is coming around the seasons.

Hit and Run

You can scroll the axis to see all features.

[Code](#)



[Code](#)

Show 10 entries

crime	month	Counts
1 Hit and Run	1	514
2 Hit and Run	2	569
3 Hit and Run	3	406
4 Hit and Run	4	394

	crime	month	Counts
5	Hit and Run	5	479
6	Hit and Run	6	456
7	Hit and Run	7	432
8	Hit and Run	8	429
9	Hit and Run	9	423
10	Hit and Run	10	448

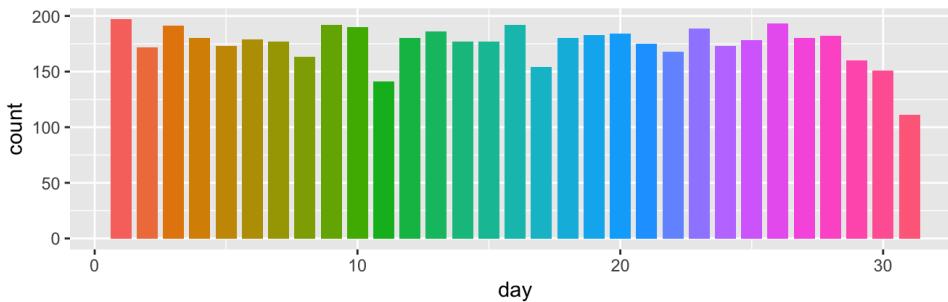
Code

Show 10 ↑ entries

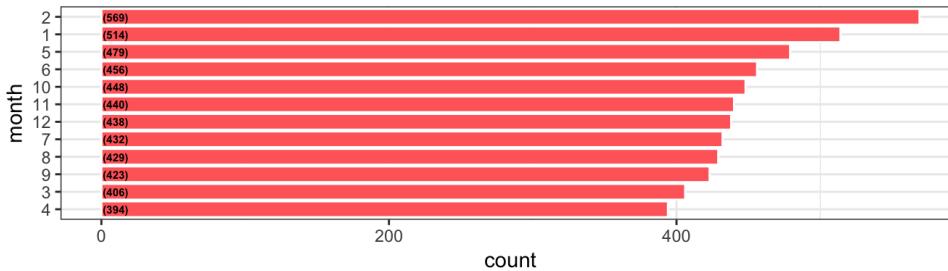
	crime	season	Counts
1	Hit and Run	Fall	1311
2	Hit and Run	Spring	1279
3	Hit and Run	Summer	1317
4	Hit and Run	Winter	1521

Code

Hit and Run Record Start by Day of Month



Total Hit and Run in Cambridge Neighbourhood by Month



Data Source : Cambridge Data |@ Pankaj Shah

11th and obviously 31st seems to have less hit and run but apart from that we have normal distribution of hit and run. Interesting is April seems to have less Hit and Run Cases and February accounts for most Hit and Run cases. Weatherwise Februaury is Worst month in Northeast Region accumulating lots of snow and making road narrower. Thats why we see lots of Hit and Run Spike around jan and Feb. Also March and April is rainy season, does that account for less hit and run as drivers are driving slow and because of poor visibility. Also the cyclist and pedestrians are less because of weather. Hard to say but it is easy to assume. Around Summer months I would have guess more hit and run but seems like the report for crime doesnt stays around range of 400-480 and similar in fall.Only thing we could observe is first two months of winter are bad leaving December apart.

Comparing top two crime

Compare two most reported crime side by side

Code

```
## Using 'Counts' as value column. Use 'value.var' to override
```

Code

Show 10 ↑ entries

	1	2	3	4	5	6	7	8	9	10	11	12	NA
Accident	3	0	3	4	8	7	6	9	2	2	4	1	0
Admin Error	10	11	3	6	22	14	10	11	10	16	5	9	0
Aggravated Assault	7	9	14	9	11	12	14	13	11	15	11	6	0
Annoying & Accosting	0	1	0	0	0	1	0	0	0	0	1	1	0
Arson	0	1	3	0	0	0	4	0	0	0	1	0	0
Auto Theft	5	7	7	9	9	8	9	9	6	3	10	5	0
Commercial Break	4	1	5	3	4	1	3	3	6	3	4	3	0
Commercial Robbery	1	2	0	0	4	2	0	3	0	1	1	0	0
Counterfeiting	1	1	0	0	3	0	2	2	1	2	0	0	0
Disorderly	2	1	1	1	3	2	3	0	1	2	0	2	0

Larcency from Motor Vehicle is low in February Month and month of December. One Possiblty is February is Winter Month. May be most of the vehicles might be burried or covered in snow. People doesnt leave or become careless when they have to leave the motor vehicle for longer time. It also takes lot of effort to break in when the weather condition is not favorable. As Expeeted August 7 october being highest as weather is warmer in August and around October its getting colder.

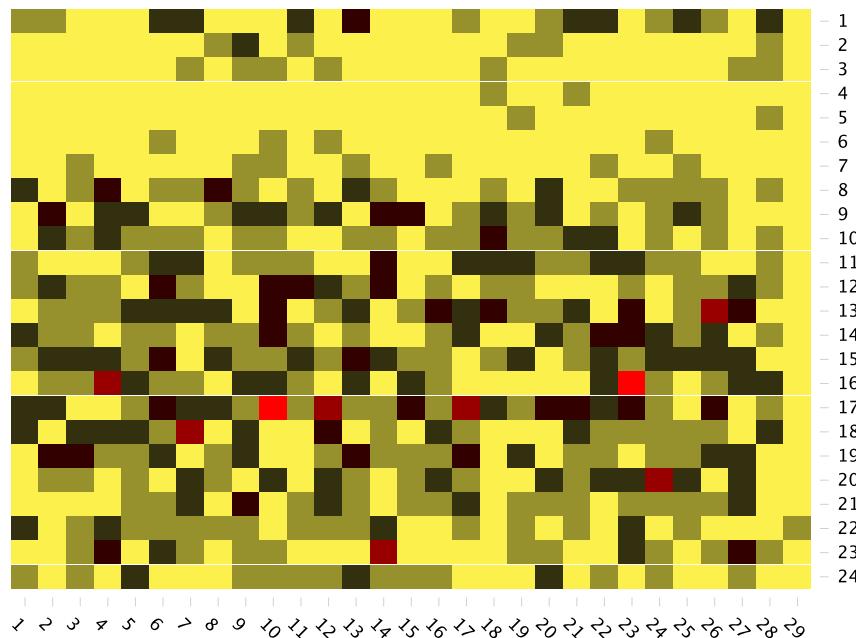
HeatMap

HeatMap Breakdown by the hour in a day within a month of February

[Code](#)

```
## Using 'Counts' as value column. Use 'value.var' to override
```

[Code](#)



Color Strength : Yellow < Black < Red

On the Heatmap if you notice we can see brightest Red dots which falls on 23rd around 16 HRS which is 4'oclock. Most of the crimes related to Hit and Run happens. So its more probable that on 23rd around 4'o clock the cambridge Police department can expect Hit and Run Cases. There are some others falling on 10th of month around 5 o clock. But most of them we can see happening around rush hour 4-6 pm with couple of execption one occuring at 11 Am and other one occuring at 11 PM on same day 14th. How strange is that.

Crimes by Month %

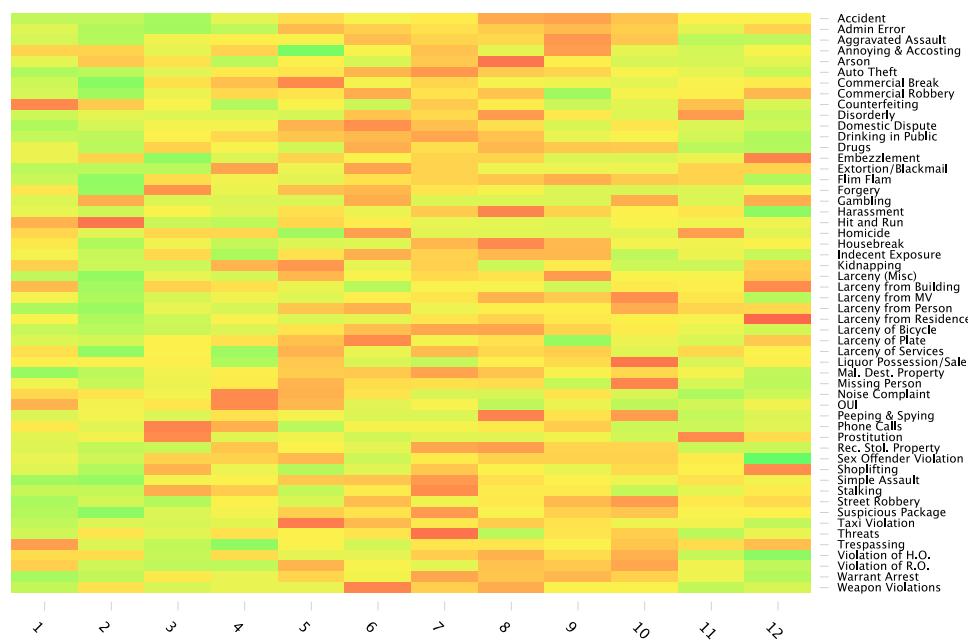
Heatmap of All Percentages

Code

Show 10 entries

	1	2	3	4	5	6	7	8	9	10	11	12
Accident	5.42	5.42	4.27	7.11	9.42	8	8.62	12	12.36	10.76	8.36	8.27
Admin Error	7.3	5.86	5.49	6.24	10.23	9.54	8.92	9.67	10.29	9.61	7.42	9.42
Aggravated Assault	7.24	6.46	8.02	8.39	8.23	9.79	9.17	9.01	10.83	9.53	6.61	6.72
Annoying & Accosting	10.34	10.34	6.9	10.34	1.15	8.05	11.49	6.9	13.79	6.9	5.75	8.05
Arson	6.1	12.2	9.76	2.44	8.54	4.88	12.2	19.51	8.54	4.88	4.88	6.1
Auto Theft	5.21	5.62	7.11	8.68	8.93	11	12.24	10.01	9.59	8.11	7.36	6.12
Commercial Break	6.99	5.43	8.85	9.78	11.34	8.07	9.16	8.07	7.92	7.61	8.23	8.54
Commercial Robbery	6.78	5.08	7.63	9.32	8.9	11.02	8.9	10.17	5.08	8.05	8.47	10.59
Counterfeiting	13.87	10.4	8.09	4.62	8.09	5.78	10.4	8.67	5.78	6.94	10.98	6.36
Disorderly	6.63	7.49	7.2	7.2	6.92	10.09	9.22	11.53	8.65	7.2	11.53	6.34

Code

**Strength: Green < Yellow < Red**

In this Heat Map we can see what kind of are most occurred breaking down by crime and the month. Most of the crimes seems to cool off around winter and pickup around summer month. In our table we saw that crimes that is reported stays the same so around winter month we have more hit and run cases and during summer we have more Larceny and other cases coming through. Accidents are reported around summer and very less in winter which is quite a strange. We might have guessed because of narrow street and snow playing major role in accident there might be more cases but does not seem so. Homicide peaks around June and then around November. Arson are reported around month of August. Larcency from the Residence spikes around December which makes complete sense as lot of people are buying gifts and expecting deliveries. Prostitution are reported more in March and November. Like these there are various variables to look into and make analysis.

Lets Scale these by column percentage wise as we have seen row percentage wise across. Scale by Column

Code

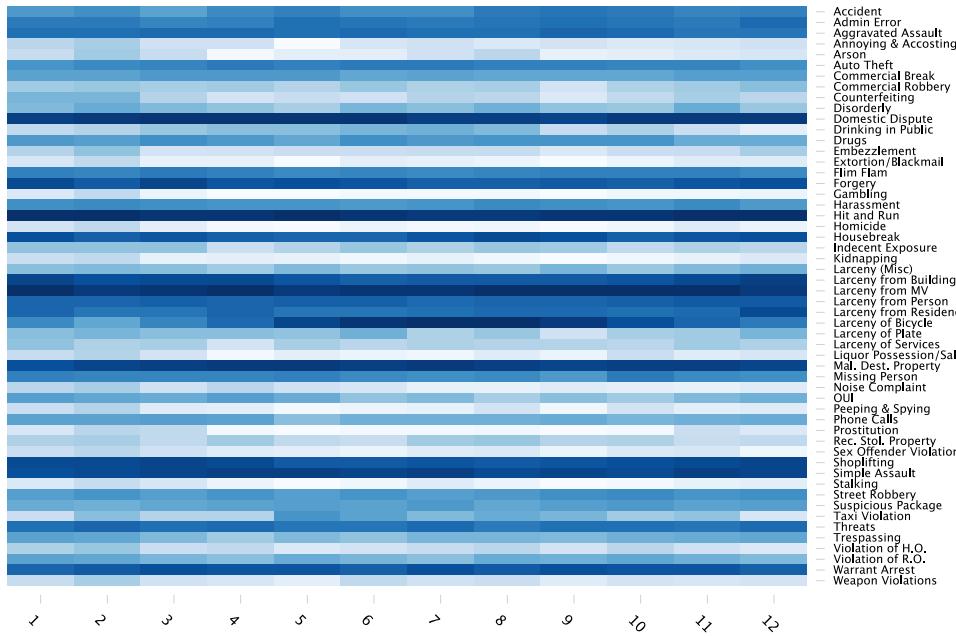
Show 10 entries

	1	2	3	4	5	6	7	8	9	10	11	12
Accident	1.28	1.39	0.96	1.57	1.86	1.53	1.56	2.18	2.4	2.09	1.82	1.86
Admin Error	2.45	2.14	1.76	1.96	2.87	2.59	2.3	2.51	2.85	2.67	2.31	3.02
Aggravated Assault	2.91	2.82	3.08	3.16	2.77	3.19	2.83	2.8	3.6	3.17	2.46	2.58

Cambridge Crime Analysis

	1	2	3	4	5	6	7	8	9	10	11	12
Annoying & Accosting	0.19	0.2	0.12	0.18	0.02	0.12	0.16	0.1	0.21	0.1	0.1	0.14
Arson	0.1	0.23	0.16	0.04	0.12	0.07	0.16	0.26	0.12	0.07	0.08	0.1
Auto Theft	1.32	1.55	1.72	2.06	1.89	2.25	2.38	1.96	2.01	1.7	1.72	1.48
Commercial Break	0.94	0.8	1.14	1.24	1.28	0.88	0.95	0.84	0.88	0.85	1.03	1.1
Commercial Robbery	0.33	0.27	0.36	0.43	0.37	0.44	0.34	0.39	0.21	0.33	0.39	0.5
Counterfeiting	0.5	0.41	0.28	0.16	0.25	0.17	0.29	0.24	0.17	0.21	0.37	0.22
Disorderly	0.48	0.59	0.5	0.49	0.42	0.59	0.51	0.65	0.52	0.43	0.78	0.44

Code



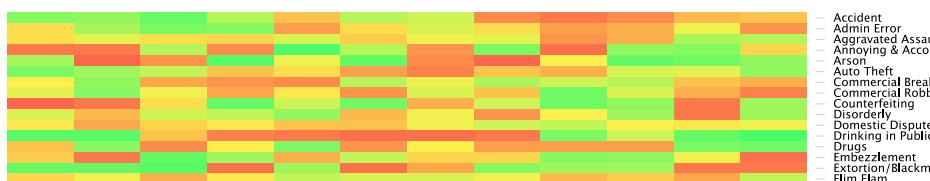
Lets mix both Rows and Column to get closer Look at Scale.

Code

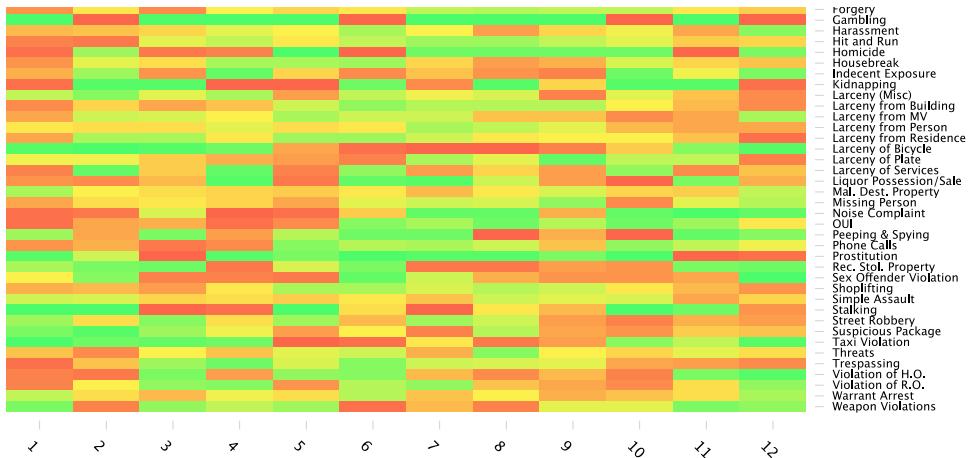
Show **10** entries

	1	2	3	4	5	6	7	8	9	10	11	12
Accident	1.41	1.46	1	1.62	1.98	1.6	1.68	2.35	2.61	2.31	2	1.96
Admin Error	1.89	1.58	1.29	1.42	2.15	1.9	1.74	1.89	2.18	2.06	1.78	2.24
Aggravated Assault	1.88	1.75	1.88	1.91	1.73	1.95	1.79	1.76	2.29	2.05	1.58	1.6
Annoying & Accosting	2.68	2.79	1.62	2.35	0.24	1.61	2.24	1.35	2.92	1.48	1.38	1.91
Arson	1.58	3.3	2.29	0.55	1.79	0.97	2.38	3.82	1.81	1.05	1.17	1.45
Auto Theft	1.35	1.52	1.67	1.97	1.88	2.19	2.39	1.96	2.03	1.74	1.76	1.45
Commercial Break	1.81	1.47	2.07	2.22	2.38	1.61	1.79	1.58	1.68	1.63	1.97	2.03
Commercial Robbery	1.76	1.37	1.79	2.12	1.87	2.2	1.74	1.99	1.07	1.73	2.03	2.52
Counterfeiting	3.6	2.81	1.9	1.05	1.7	1.15	2.03	1.7	1.22	1.49	2.63	1.51
Disorderly	1.72	2.02	1.69	1.64	1.45	2.01	1.8	2.26	1.83	1.55	2.76	1.51

Code

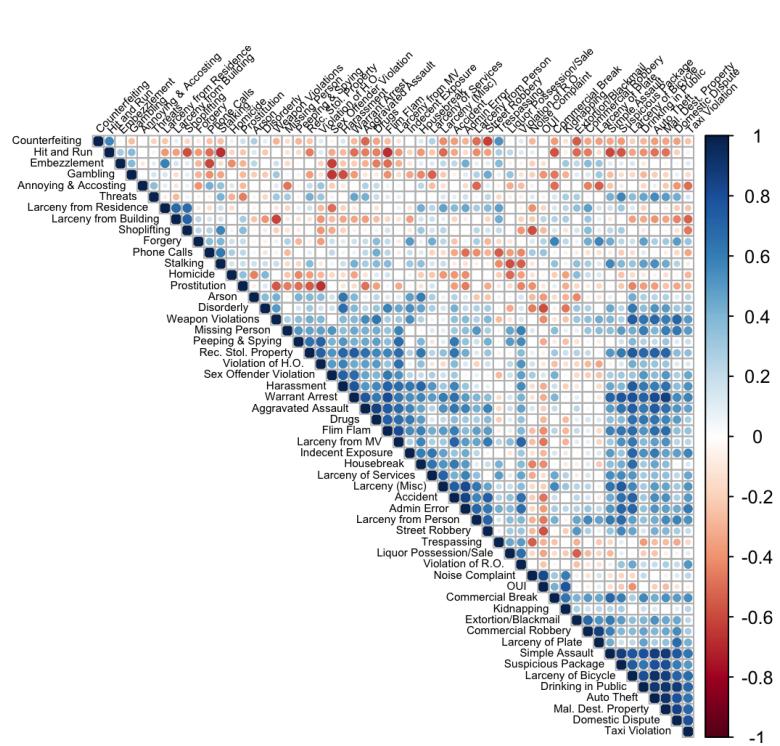


Cambridge Crime Analysis



We can take a see lots of similarity and how things have changed in the heatmap as I have mentioned earlier.

Corplot



Code

Code

	Accident	Admin Error	Aggravated Assault	Annoying & Accosting	AI
Accident	1	0.844462760456648	0.638297214615918	-0.021487700023129	0.25020543701
Admin Error	0.844462760456648	1	0.581511623829433	-0.162964169716148	0.12403491561
Aggravated Assault	0.638297214615918	0.581511623829433	1	0.268783927201716	0.085640015551
Annoying & Accosting	-0.021487700023129	-0.162964169716148	0.268783927201716	1	0.00421216638517261
Arson	0.250205437008914	0.124034915698707	0.0856400155566242	0.00421216638517261	
Auto Theft	0.544348019358957	0.527308688071454	0.759340363542952	0.076812315634118	0.24161944401

Commercial Break	0.188445334586112	0.289083541775725	0.206209814626843	-0.497994199591567	-0.1600085185
Commercial Robbery	0.1532770418513	0.325596009695105	0.04304752704115	-0.441561065415815	-0.07642437091
Counterfeiting	-0.367033776227961	-0.299535382755226	-0.516302162409144	-0.015560636379825	0.2716458506
Disorderly	0.421215888448052	0.214997259101765	0.228841405283682	-0.0336884082708282	0.377847807

Previous 1 2 3 4 5 6 Next**You can scroll the x-axis to see all features.**

- If you are Stalking it negatively correlates with Liquor Possession, Tresspassing, Violation of R.O
- Prostitution is negatively corelated with violation of H.O
- Sex offender Violation is negatively correlates with Larcency from Residence, Embezzlement.

Red Dots are Negatively Correlated Blue dots are positively Correlated

Lets see how likely person which is involved in Hit and Run and another column Accidents correlates to other crime.

Code

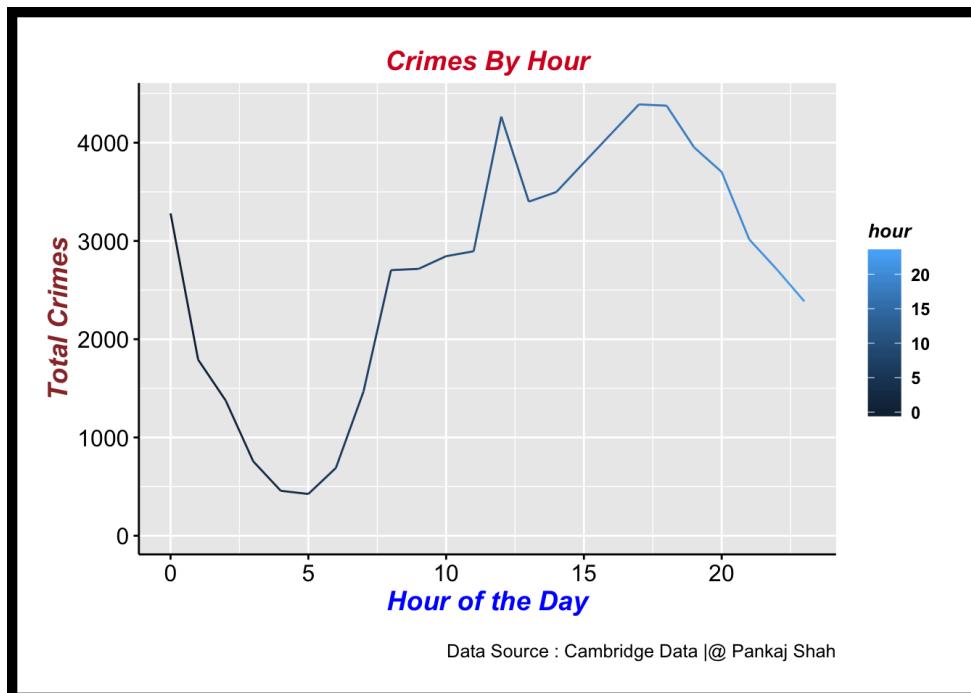
	Accident	Admin Error	Aggravated Assault	Flim Flam	Hit and Run
Accident	1	0.844462760456648	0.638297214615918	0.604852231950657	-0.330591201493538
Hit and Run	-0.330591201493538	-0.162733847399132	-0.456200565767977	-0.674806389689029	1 -

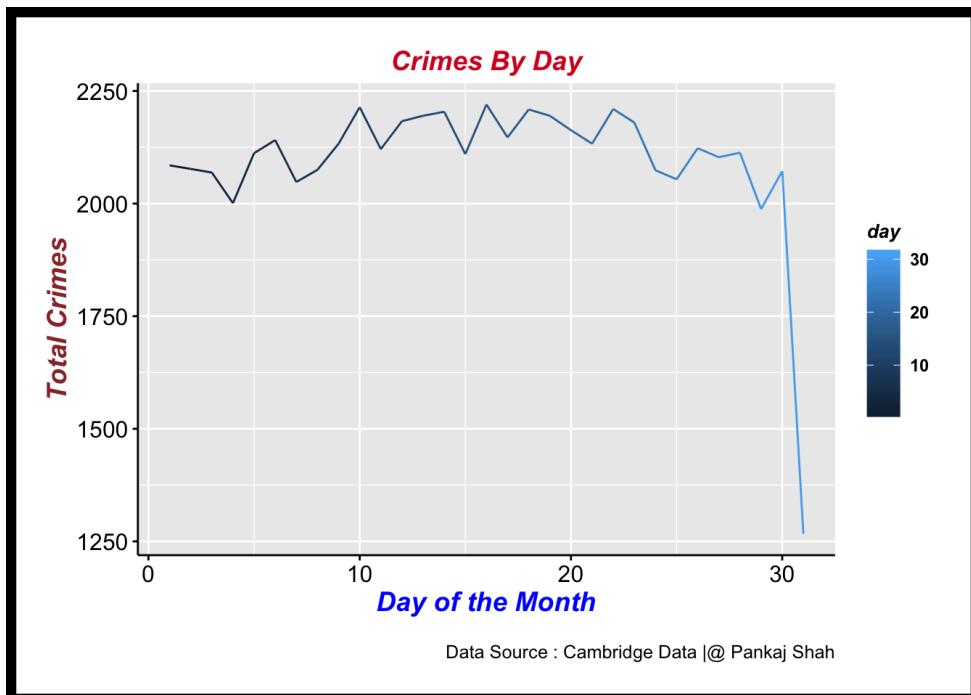
Person who is involved in accident is negatively correlated to hit and run.if you are involved in accident then there is 81% chance that it will have Admin Error. Thats Funny how things correaltes.

Overall what we discuss above I would like to show the summary in one color from 1st Jan 2009 to 31st Dec 2017.

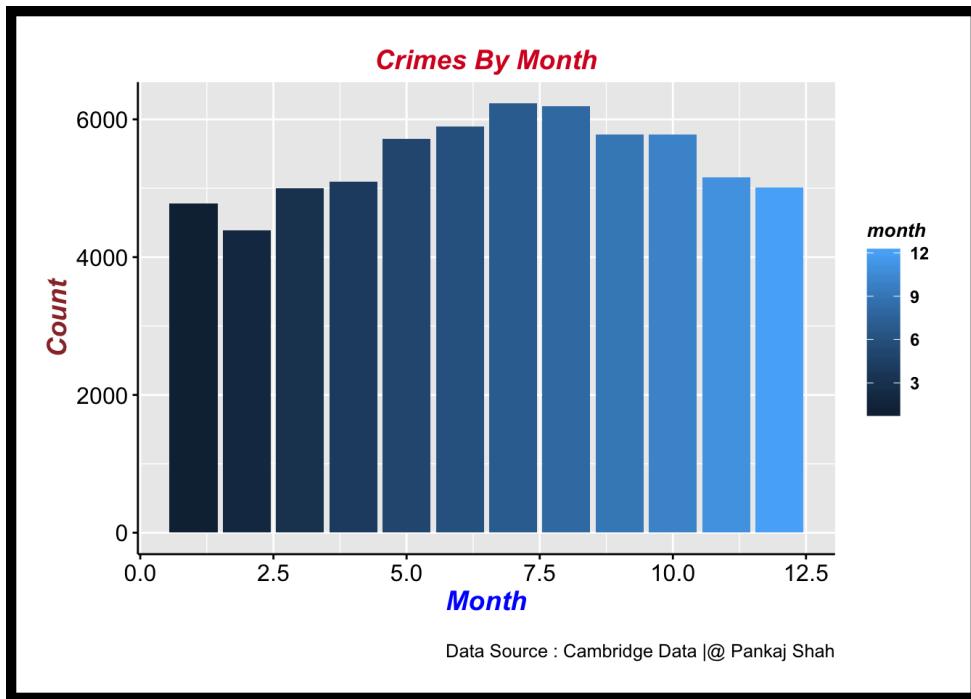
Summary: Time series BreakpointCode

```
## Warning: Removed 1 rows containing missing values (geom_path).
```

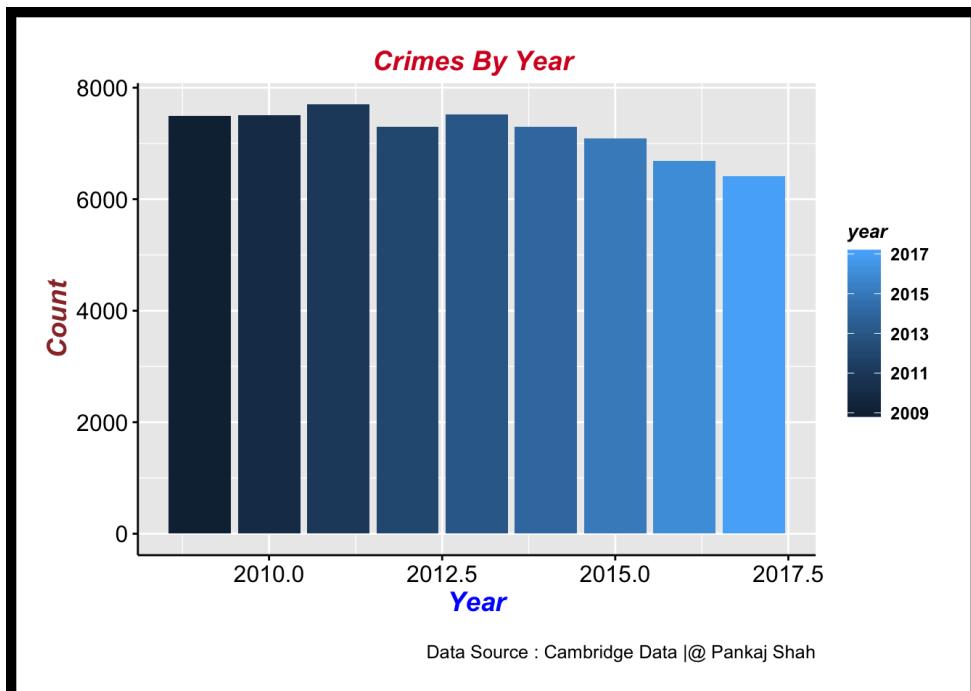
Code



Code

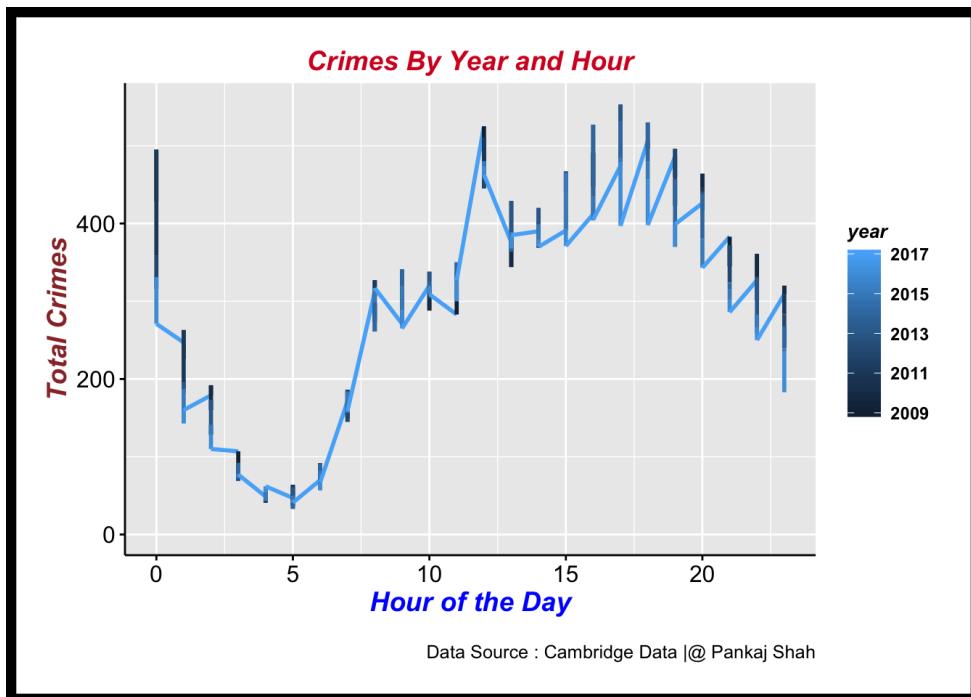


Code



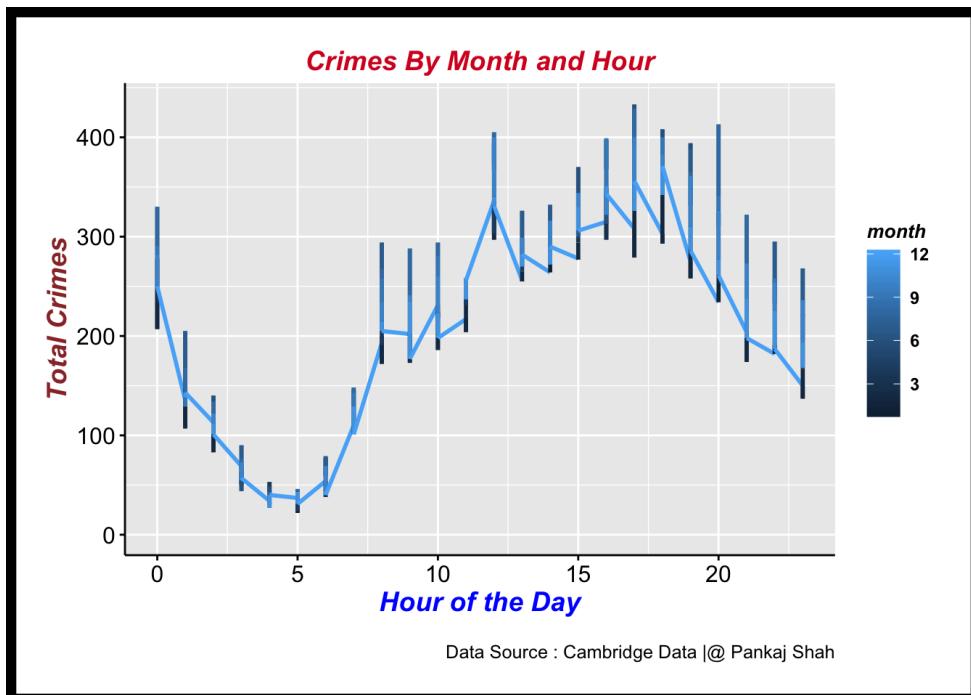
Code

```
## Warning: Removed 5 rows containing missing values (geom_path).
```

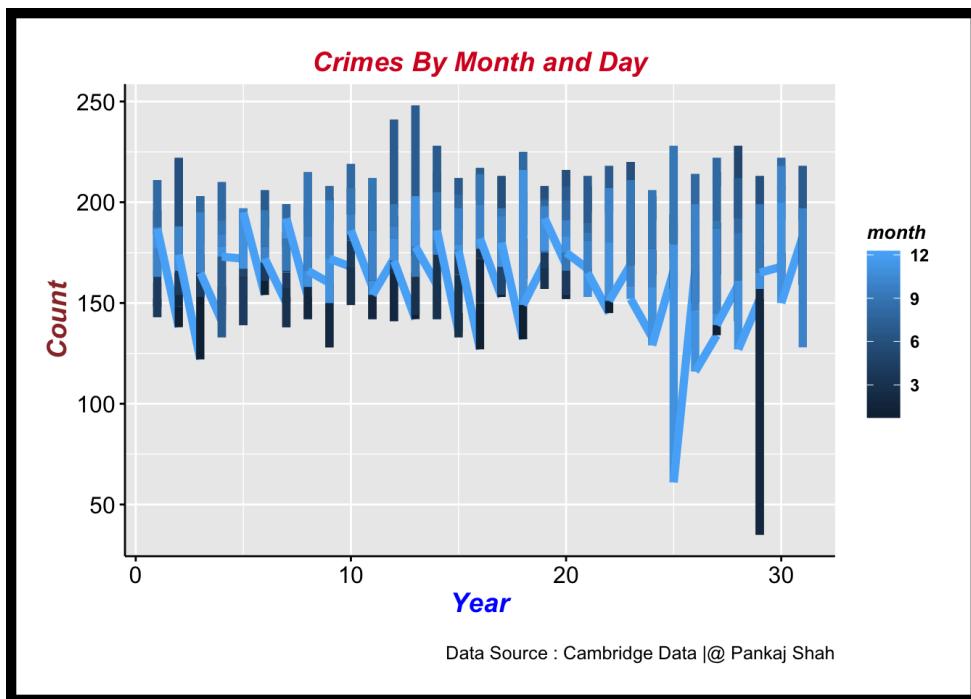


Code

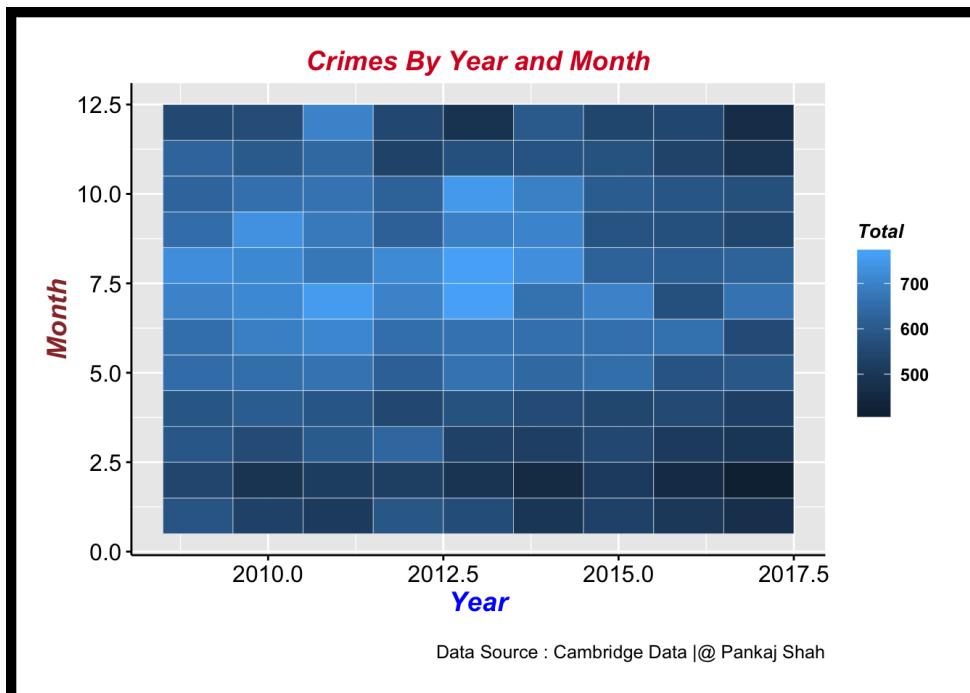
```
## Warning: Removed 10 rows containing missing values (geom_path).
```



Code

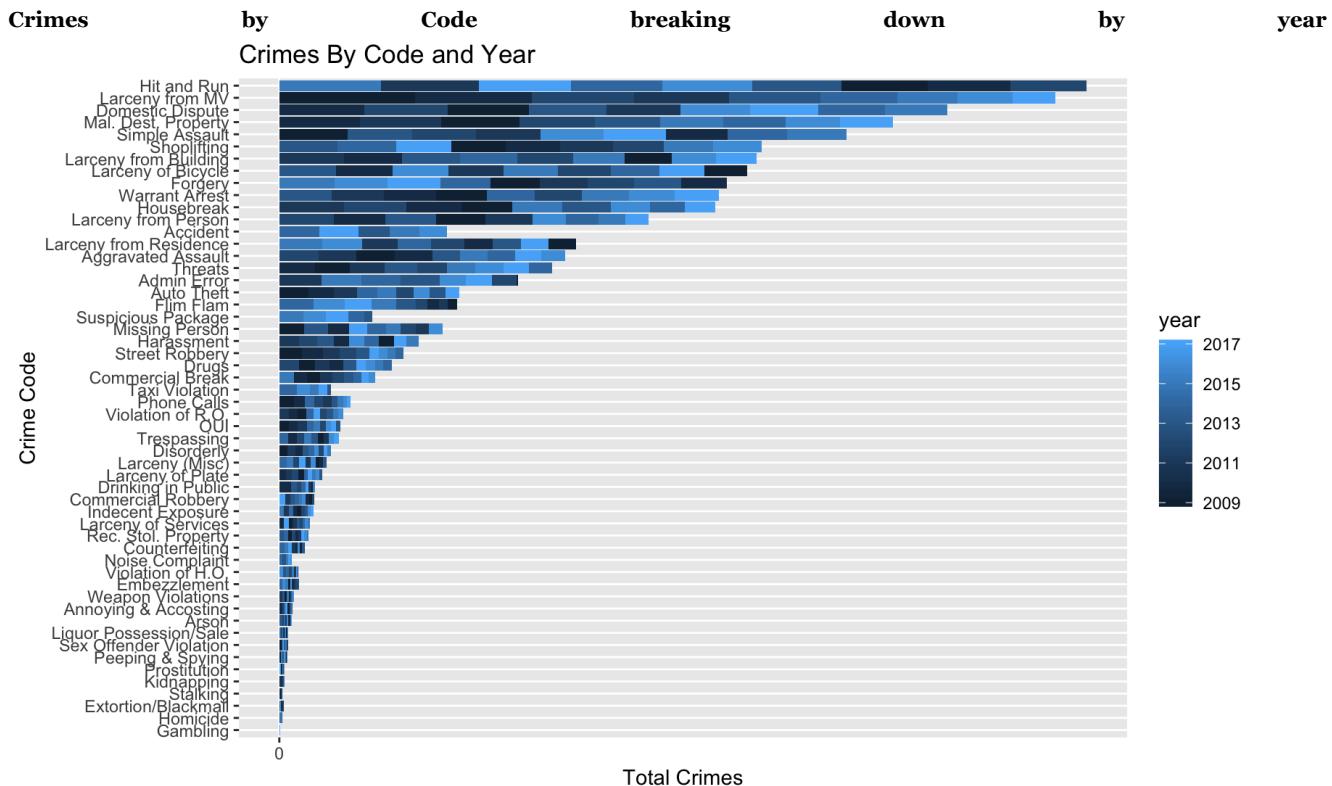


Code



Crimes are less during early morning hours but as the day progresses they continue to grow, peak during early evening time and slow down as the day ends. Crimes are more during months which fall in the middle of the year. Similar trend was seen in previous visual. Crimes are most starting from May to August. Is it due because of summer? Leaving out 2016 as it is not completed, if you check other years you will clearly find the reduction in the number of crimes over the last 10 year period. Crimes are decreasing over the years * Most noticeable trend is decline in crimes during Christmas. There is huge decline in crimes.

Crime_Code/Year



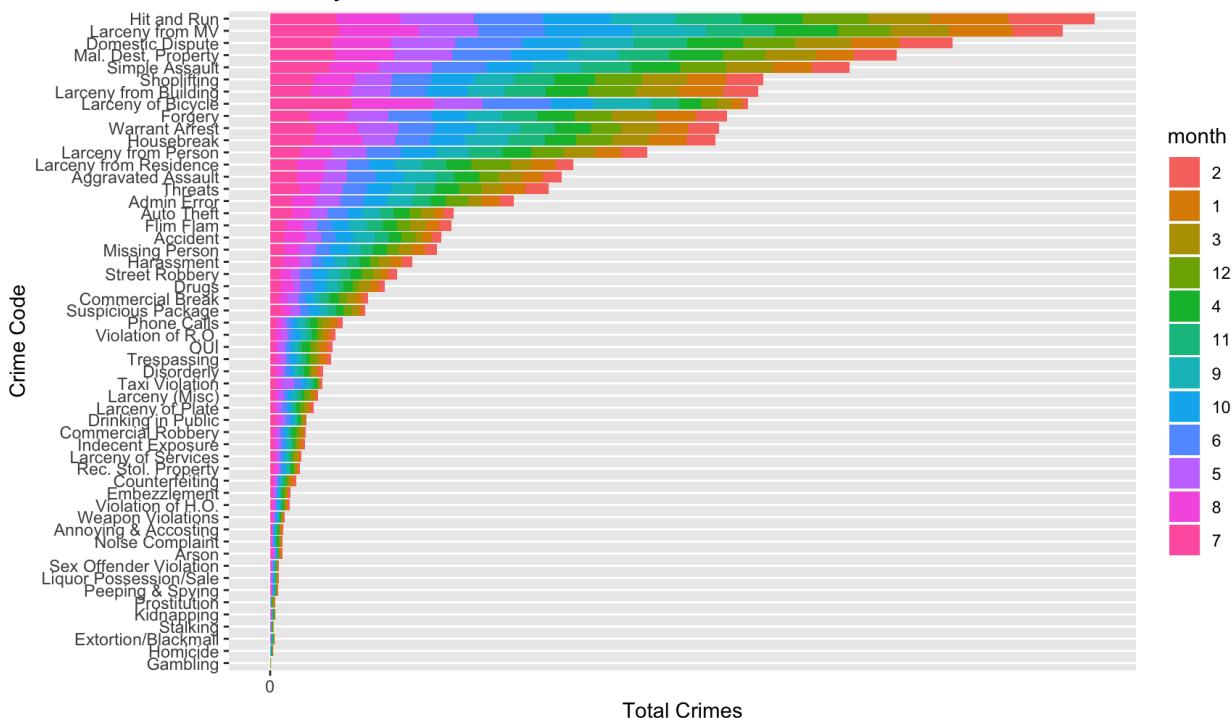
Show 10 entries

	year	crime	Total
1	2009	Larceny from MV	913
2	2010	Larceny from MV	784
3	2012	Larceny from MV	686
4	2015	Hit and Run	681
5	2011	Hit and Run	659
6	2011	Larceny from MV	640
7	2017	Hit and Run	619
8	2013	Larceny from MV	615
9	2014	Hit and Run	614
10	2016	Hit and Run	606

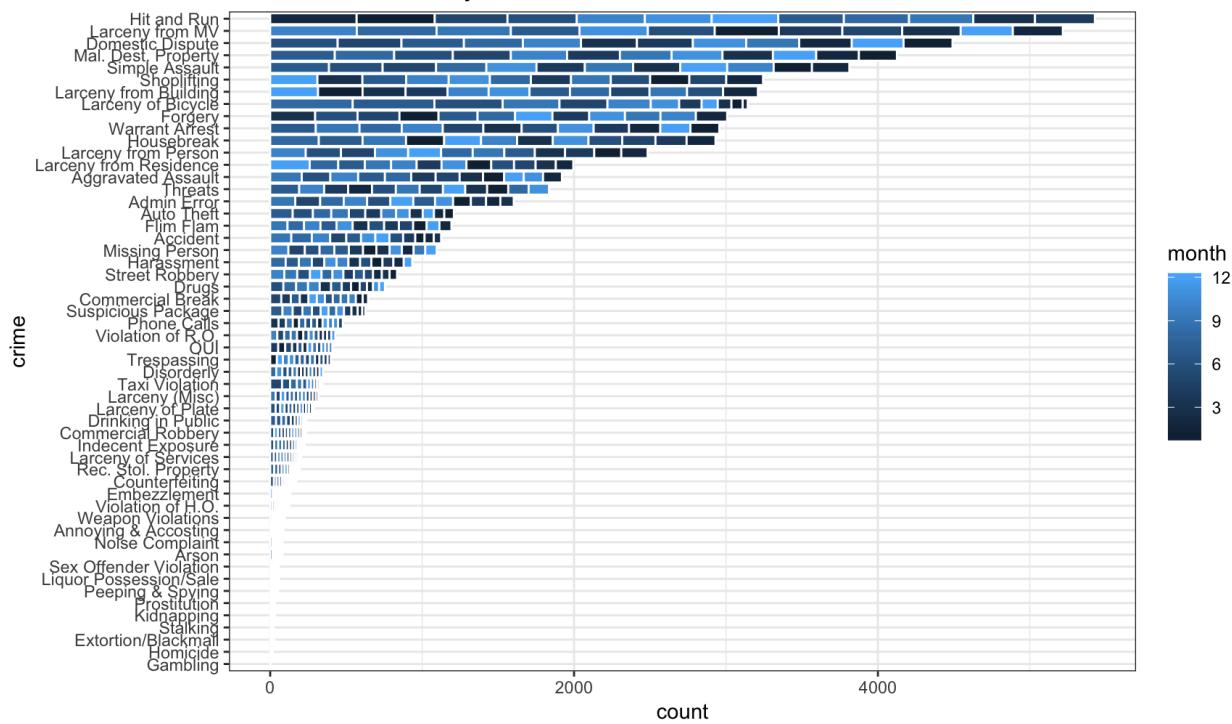
Crime_code/Month

Crimes by Code breaking down by month

Crimes By Code and Month



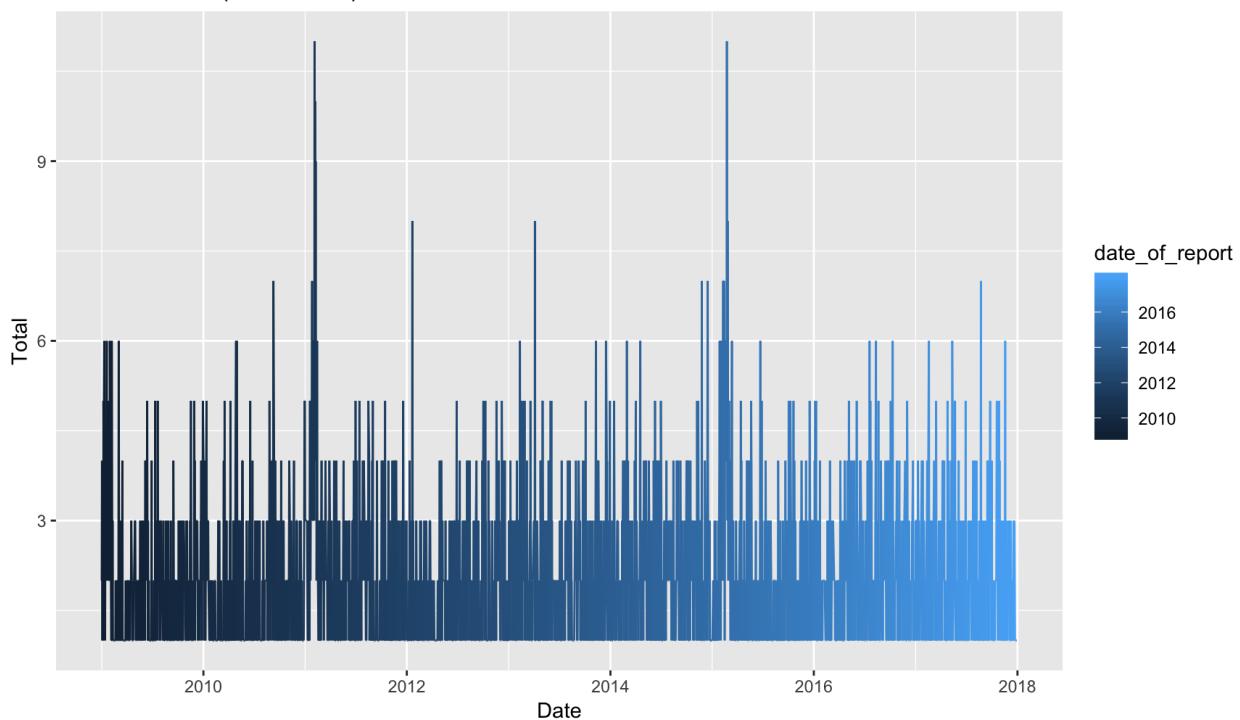
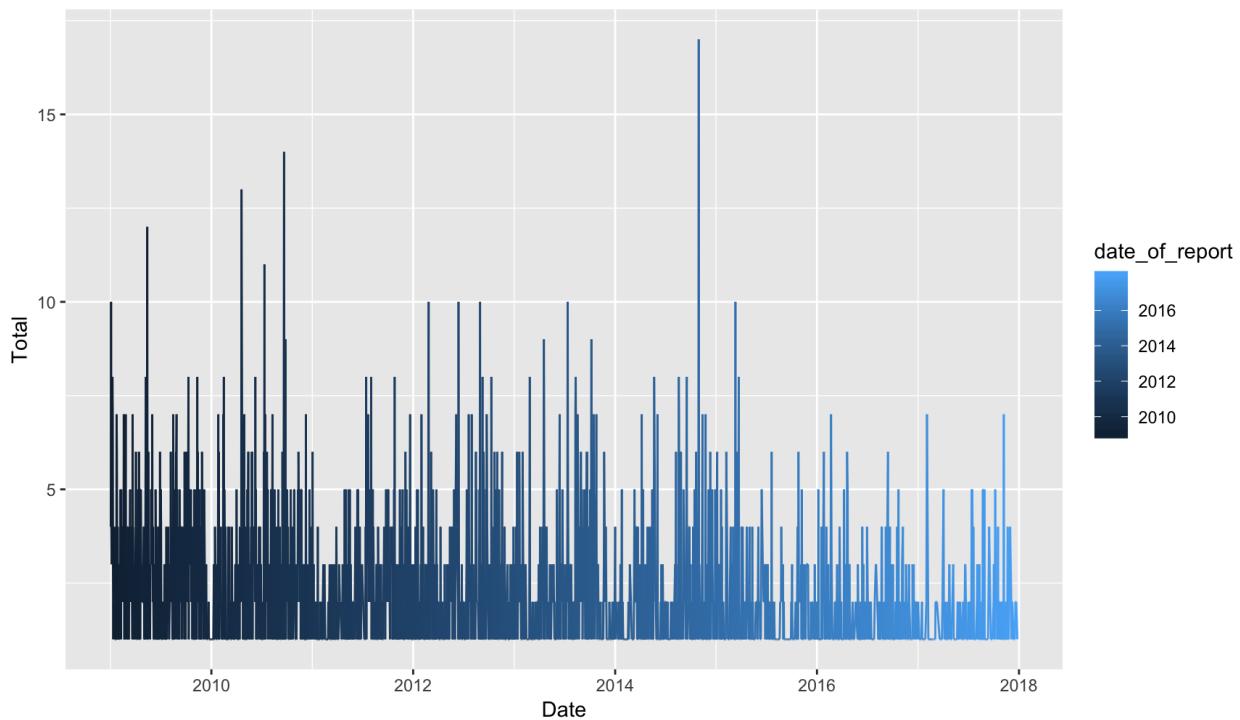
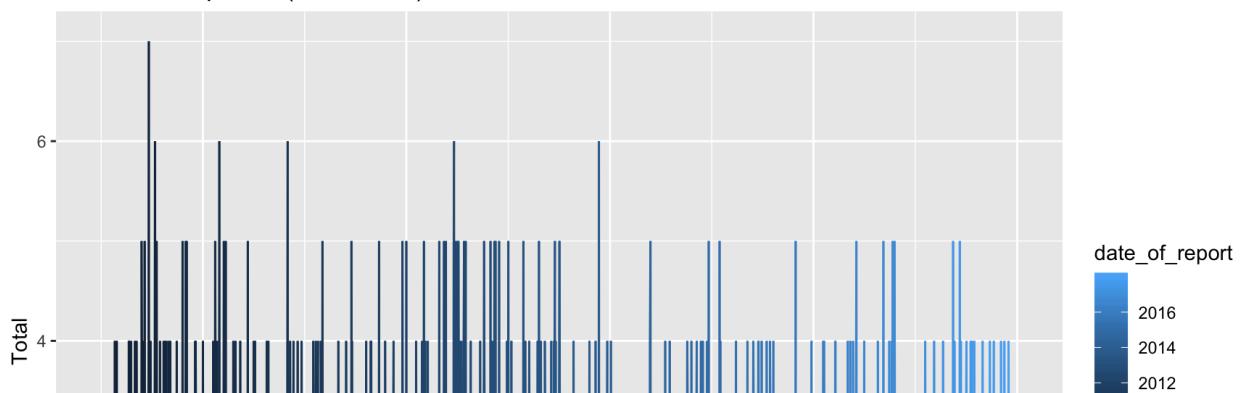
Total crime divided by Month

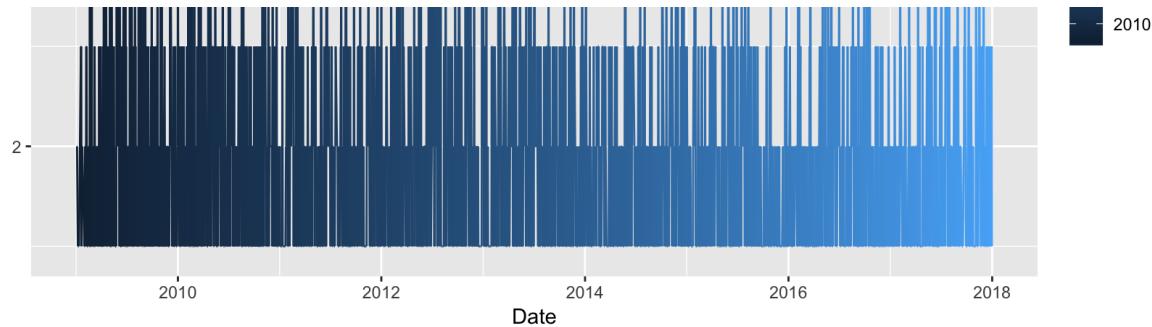


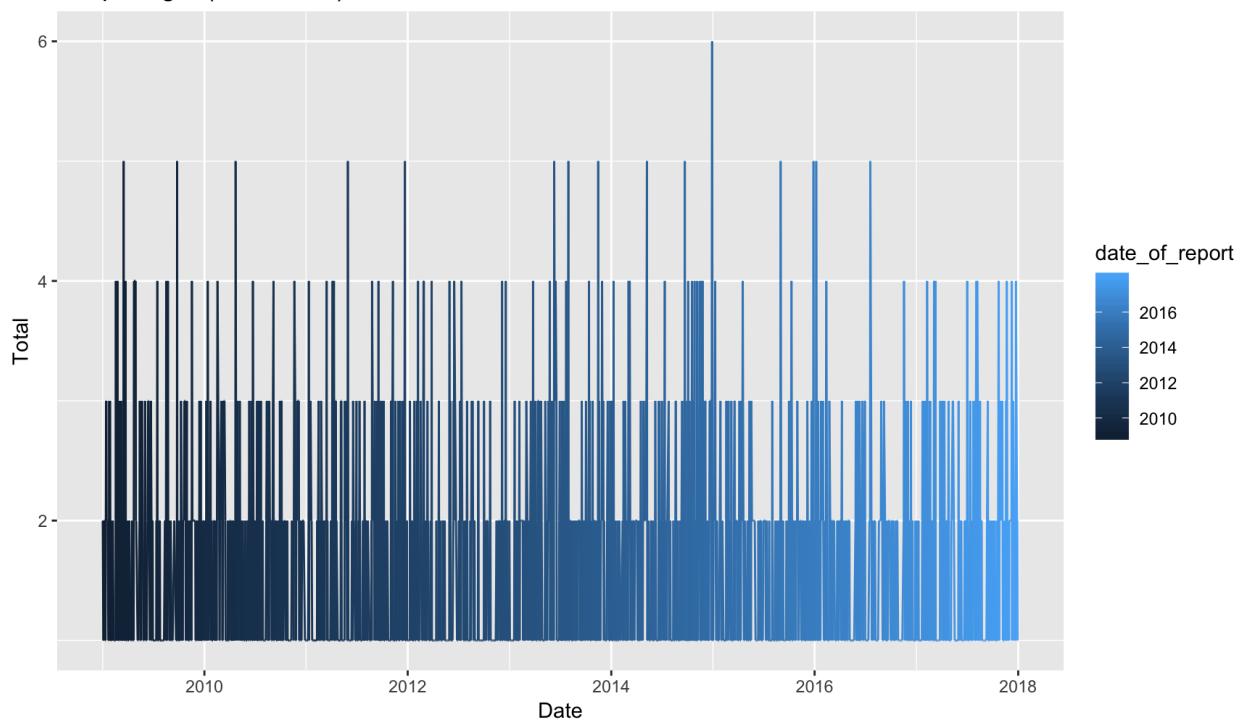
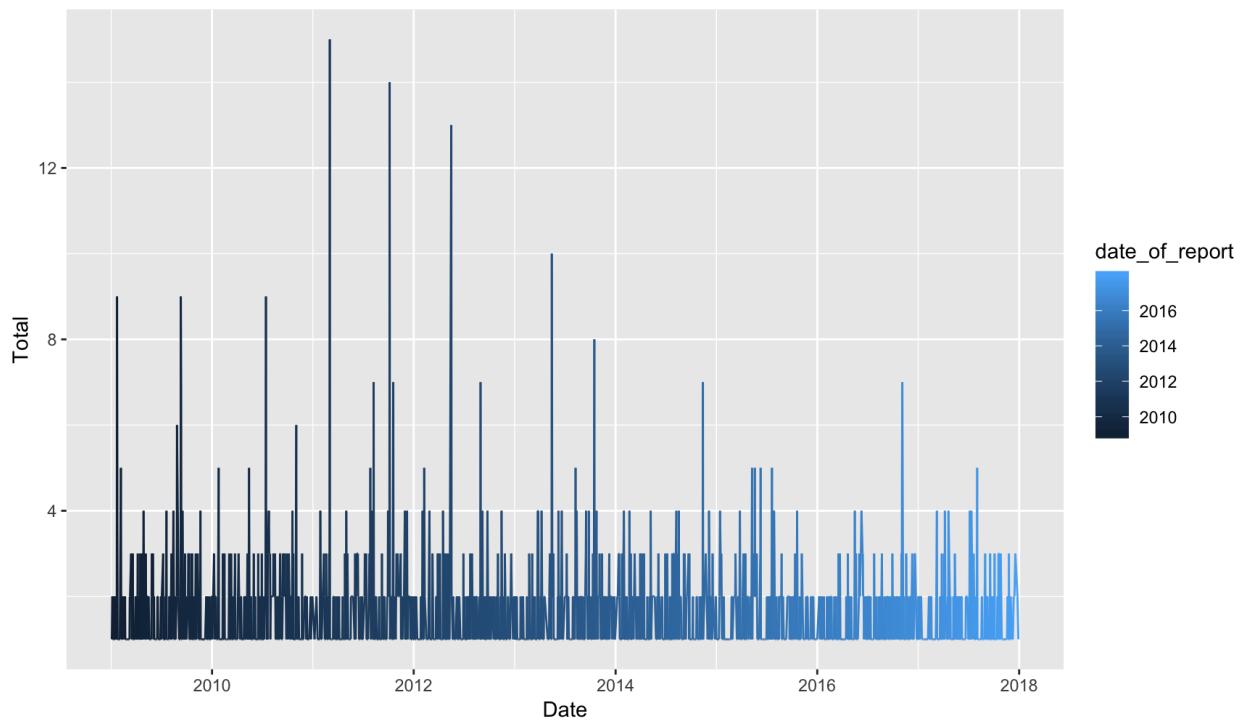
Show 10 ↑ entries

month	crime	Total	
1	10	Larceny from MV	570
2	2	Hit and Run	569
3	8	Larceny of Bicycle	543
4	7	Larceny of Bicycle	538
5	8	Larceny from MV	525
6	1	Hit and Run	514
7	9	Larceny from MV	489
8	5	Hit and Run	479

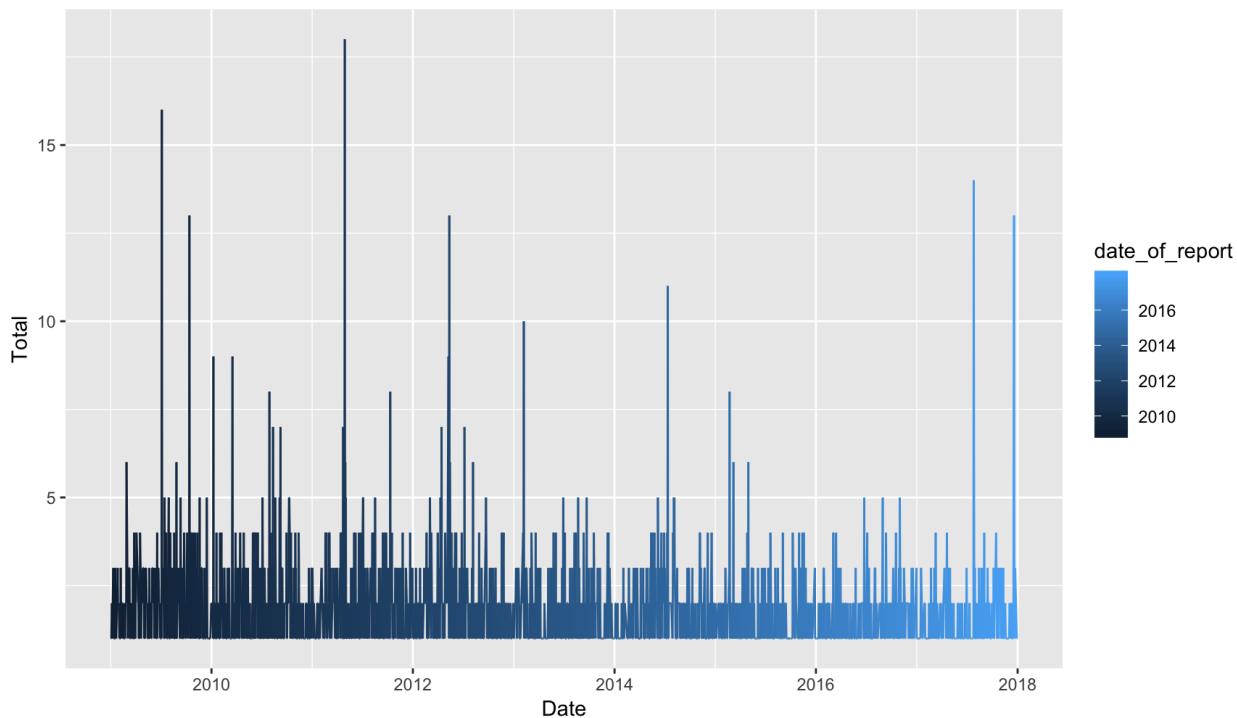
month		crime	Total
9	6	Hit and Run	456
10	7	Larceny from MV	452

Hit and Run - (2009-2018)**Larceny from MV - (2009-2018)****Domestic Dispute - (2009-2018)**



Shoplifting - (2009-2018)**Warrant Arrest - (2009-2018)**

Mal. Dest. Property - (2009-2018)

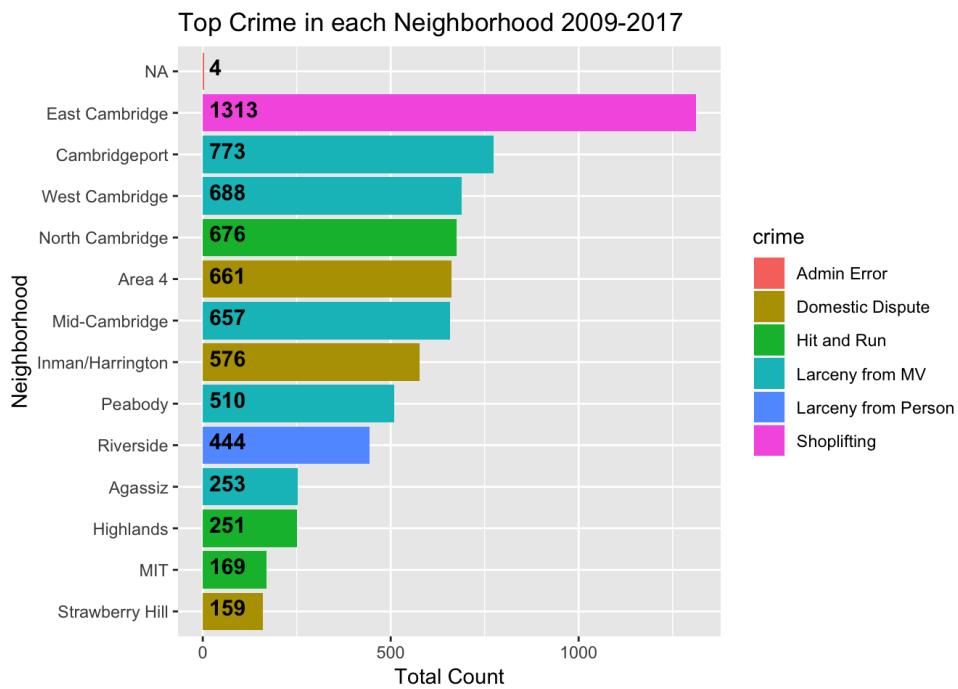


Top crime in each Neighborhood

Code

```
## Selecting by Total
```

Code



Code

Show 10 entries

	neighborhood	crime	Total
1	East Cambridge	Shoplifting	1313
2	Cambridgeport	Larceny from MV	773

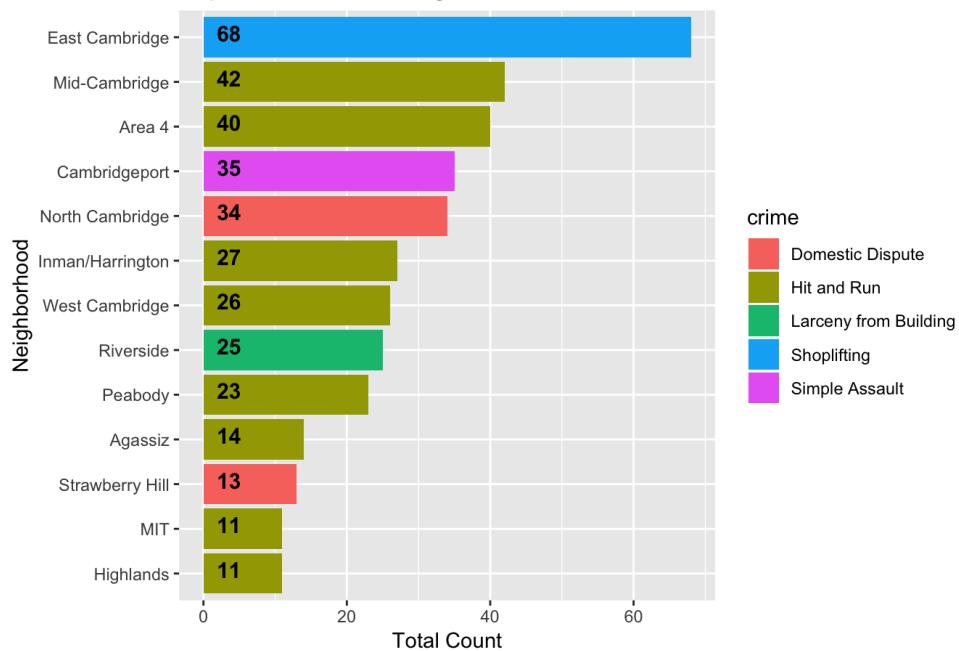
neighborhood	crime	Total
3	Larceny from MV	688
4	Hit and Run	676
5	Domestic Dispute	661
6	Larceny from MV	657
7	Domestic Dispute	576
8	Larceny from MV	510
9	Larceny from Person	444
10	Larceny from MV	253

Code

Selecting by Total

Code

Top Crime in each Neighborhood 2018



Top Crime in each Reporting Area

Code

Selecting by Total

Code

Show 10 entries

reporting_area	crime	Total
1	Shoplifting	1289
2	Larceny from Person	457
3	Warrant Arrest	336
4	Larceny from Person	310
5	Shoplifting	282
6	Warrant Arrest	239
7	Domestic Dispute	236
8	Domestic Dispute	208

reporting_area	crime	Total
9	305	Domestic Dispute
10	403	Domestic Dispute

Top Crime in each Neighborhood/Reporting Area

Code

```
## Selecting by Total
```

Code

Show 10 ↑ entries

neighborhood	reporting_area	crime	Total
1	East Cambridge	102 Shoplifting	1289
2	West Cambridge	1002 Larceny from Person	457
3	Cambridgeport	501 Warrant Arrest	336
4	Riverside	708 Larceny from Person	310
5	North Cambridge	1101 Shoplifting	282
6	East Cambridge	108 Warrant Arrest	239
7	North Cambridge	1109 Domestic Dispute	236
8	East Cambridge	105 Domestic Dispute	208
9	Area 4	403 Domestic Dispute	205
10	Inman/Harrington	305 Domestic Dispute	205

Top Two Crime Neighborhoods

- 101-109 reporting area falls in East Cambridge
- 501- 510 reporting area falls in Cambridgeport

Code

Show 10 ↑ entries

	reporting_area	n
1	101	736
2	102	2967
3	103	607
4	104	388
5	105	1850
6	106	322
7	107	1231
8	108	852
9	109	2
10	501	2594

Conclusion

See the Executive Summary for detail Analysis Finding at the top of page.

Thank you for Reading the Post.Hope you enjoyed reading as much as “fun” I had making it.

