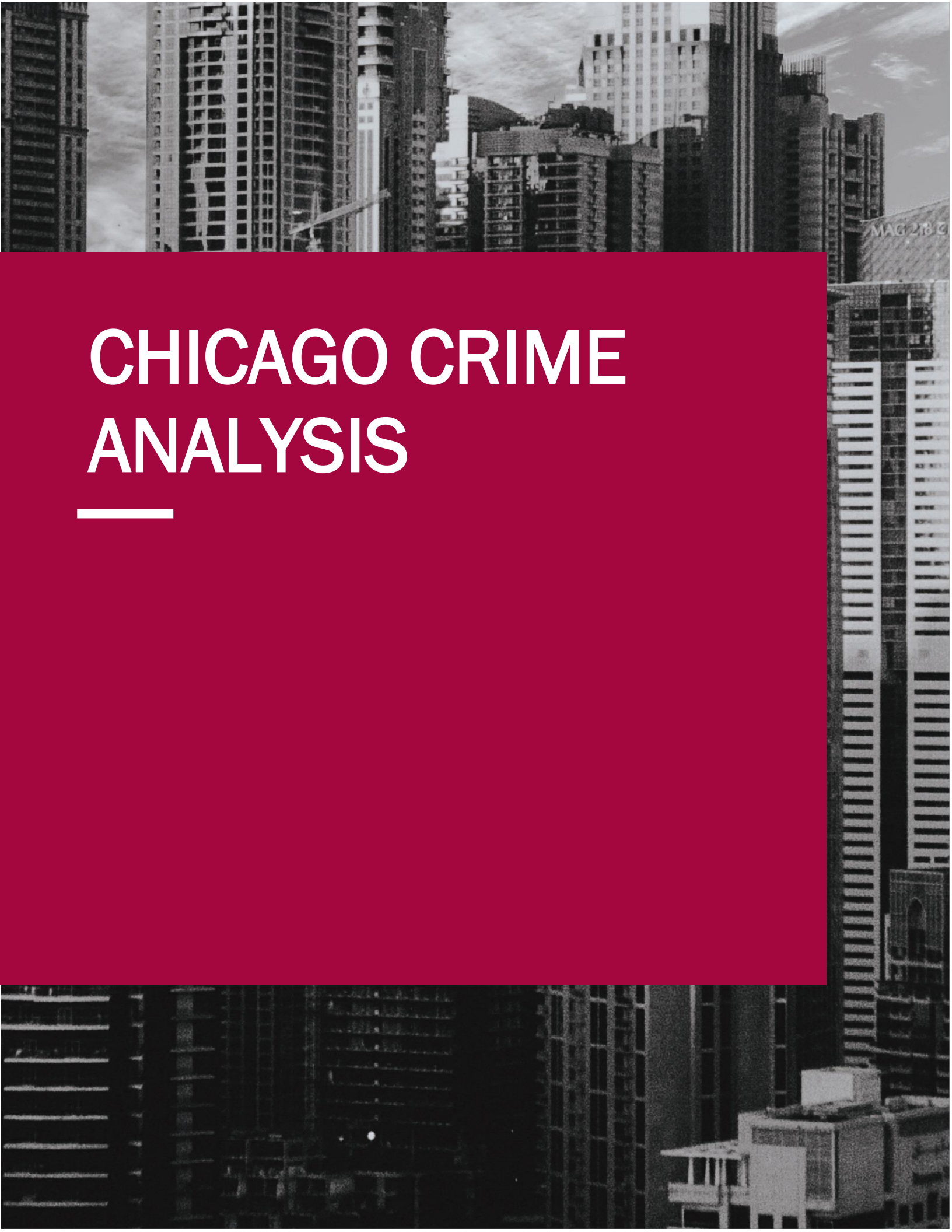


CHICAGO CRIME ANALYSIS

Introduction

Crime in Chicago has been tracked by the Chicago Police Department's Bureau of Records since the beginning of the 20th century. The city's overall crime rate, especially the violent crime rate, is higher than the US average. Chicago was responsible for nearly half of 2016's increase in homicides in the US, though the nation's crime rates remain near historic lows. The reasons for the higher numbers in Chicago remain unclear. An article in The Atlantic detailed how researchers and analysts had come to no real consensus on the cause for the violence.

Chicago experienced a major rise in violent crime starting in the late 1960s, a decline in overall crime in the 2000s, and then a rise in murders in 2016. **Murder**, **rape**, and **robbery** are common violent crimes in the city, and the occurrences of such incidents are documented by the **Chicago Police Department** and indexed in annual crime reports.

Besides, The Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system is a web application enabling the public to search the Chicago Police Department's database of reported crime. Individuals can see maps, graphs, and tables of reported crime. The database contains 90 days of information, which can be accessed in blocks of up to 14 days. Data is refreshed daily. However, the most recent information is always six days old.

So, I chose this topic in consideration of **Chicago Police Department** and the **public** (mostly, residents of Chicago and United States). Chicago police department has a tool that can visualize the crime data but even though I chose this topic because moving forward, I would like to perform some predictive analysis. This analysis helps the police to predict the crimes based on the time series forecasting and sentimental analysis in that area.

Moreover, my analysis presents a clear picture on type of crimes that are prevailing more in the city of Chicago and what is their impact on the public. Also, it clearly states the responsibility of Chicago Police to predict and eradicate crimes.

Design

Going forward with this idea, my visualization design process rises some of the key points like;

- What is the best **suitable plot** for rendering data and visualizing it in an informative way.

  To find the best plot, I rendered data in Time Series Plot, Heat Maps, Bar Plots, Multivariate Line Plots, and finally Histograms. But each plot has its own advantages and disadvantages. However, I found Heat Map and Time series plots are best among others for visualizing this data as it's Time Series and Multivariate.

- What could be the parameters to be plotted.

  In the recent years, the crimes related to Homicides are prevailing the most. So, I have decided to plot it along with the yearly crimes and arrests. Besides, the Heat Map picturizes he crime intensity in an area irrespective of the parameters.

- What kind of ideology should be behind this project and what is its importance.

  Some of the key points and beliefs were like Chicago crime frequency had been in media for wrong reasons, and Crimes are impacting the people in the country. So, to judge these beliefs I visualized crime frequency in the past and present along what kind of crimes are more prevalent and what kind of impact they have on the people. Moreover, in this scenario I have also considered arrest rates.

Finally, I chose Heat Map, Circle Map and Time Series Plotting for this project to answer my questions.

Implementation

I have implemented this project using Python and R along with Shiny Dashboard, Plotly, Leaflet and Highcharter libraries. Python is used for wrangling the data, Feature engineering and doing some predictive analysis using **Facebook's Prophet library** for Time Series Forecasting. R is used to render Web application using shiny dashboard and some CSS.

To increase the **complexity**,

- I chose a dataset that has about **1.5 Million** rows with 22 columns.
- I used R Shiny Dashboard because it has more advantages than D3 when coming to Time Series Plotting. R can be directly interfaced to **MongoDB** for streaming Real Time Data with Mongo connector. So, I chose R to D3.
- For improving **UX** and **UI**, I used libraries like **Leaflet** for **Maps**, **DT** for **Data Table**, **Highcharter** for **Time Series Plotting**, **Plotly** for **bar plots** and **xts** for **Extensible Time Series plots**.

Leaflet is one of the most popular open-source JavaScript libraries for interactive maps. It's used by websites ranging from The New York Times and The Washington Post to GitHub and Flickr, as well as GIS specialists like OpenStreetMap, Mapbox, and CartoDB.

Some of the Leaflet features are;

- Interactive panning/zooming.
- Create maps right from the R console or RStudio.
- Embed maps in knitr/R Markdown documents and Shiny apps.
- Easily render spatial objects from the sp or sf packages, or data frames with latitude/longitude columns.
- Use map bounds and mouse events to drive Shiny logic.
- Display maps in non-spherical Mercator projections.

Highcharter is a R wrapper for Highcharts JavaScript library and its modules. Highcharts is very mature and flexible JavaScript charting library and it has a great and powerful API.

Some of the Highcharter features are;

- Chart various R object with one function. With hchart(x) you can chart **xts**.
- Support Highstock charts.
- Themes: you configurate your chart in multiples ways. There are implemented themes like economist, financial times, google, 538 among others.
- Plugins: motion, drag points, fontawesome, URL-Pattern, annotations.

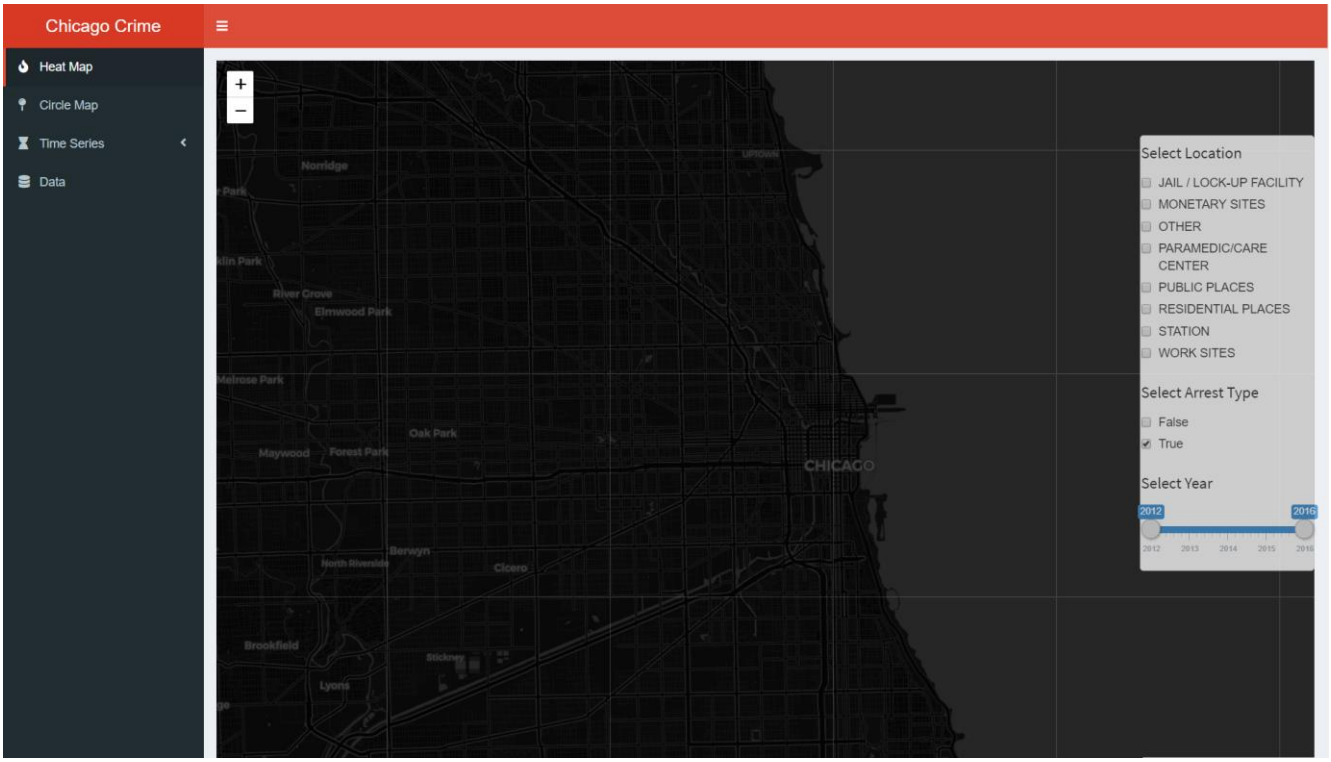
User Guide

To get the most out of this project, please follow the instructions for viewing visualizations;

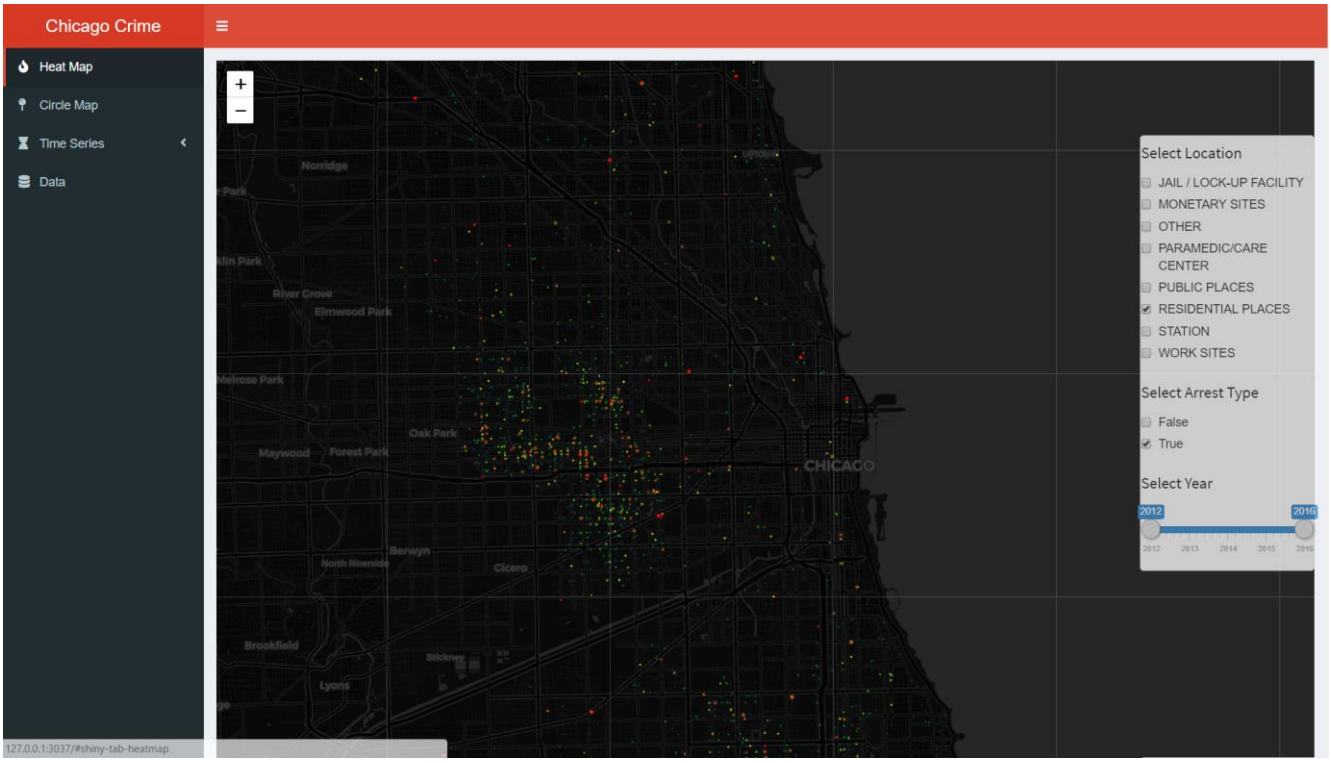
1. This application consists of three files global.R, server.R, and ui.R. The global.R is for declaring global variables.
2. Now run the application after installing all the libraries;

```
library(shiny)
library(shinydashboard)
library(leaflet.extras)
library(xts)
library(leaflet)
library(plotly)
library(DT)
library(highcharter)
library(shinythemes)
```

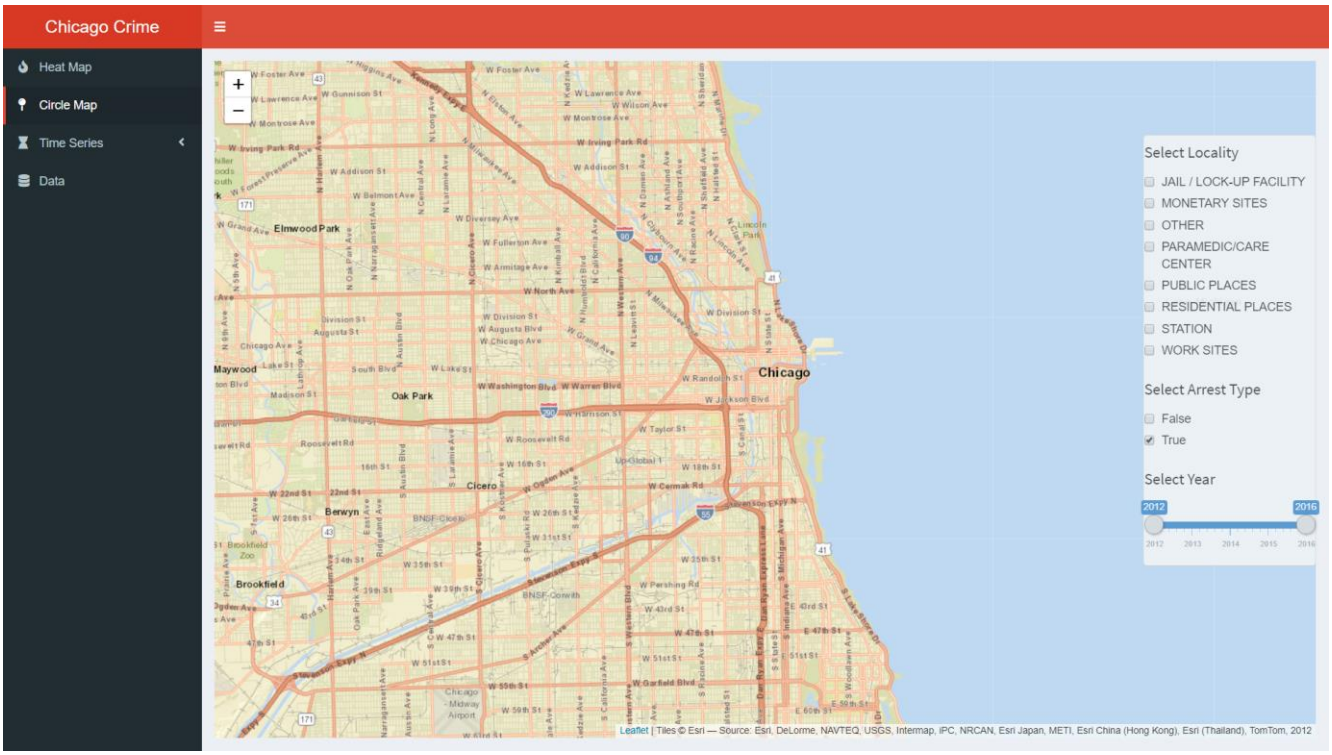
3. Open the application in web browser.



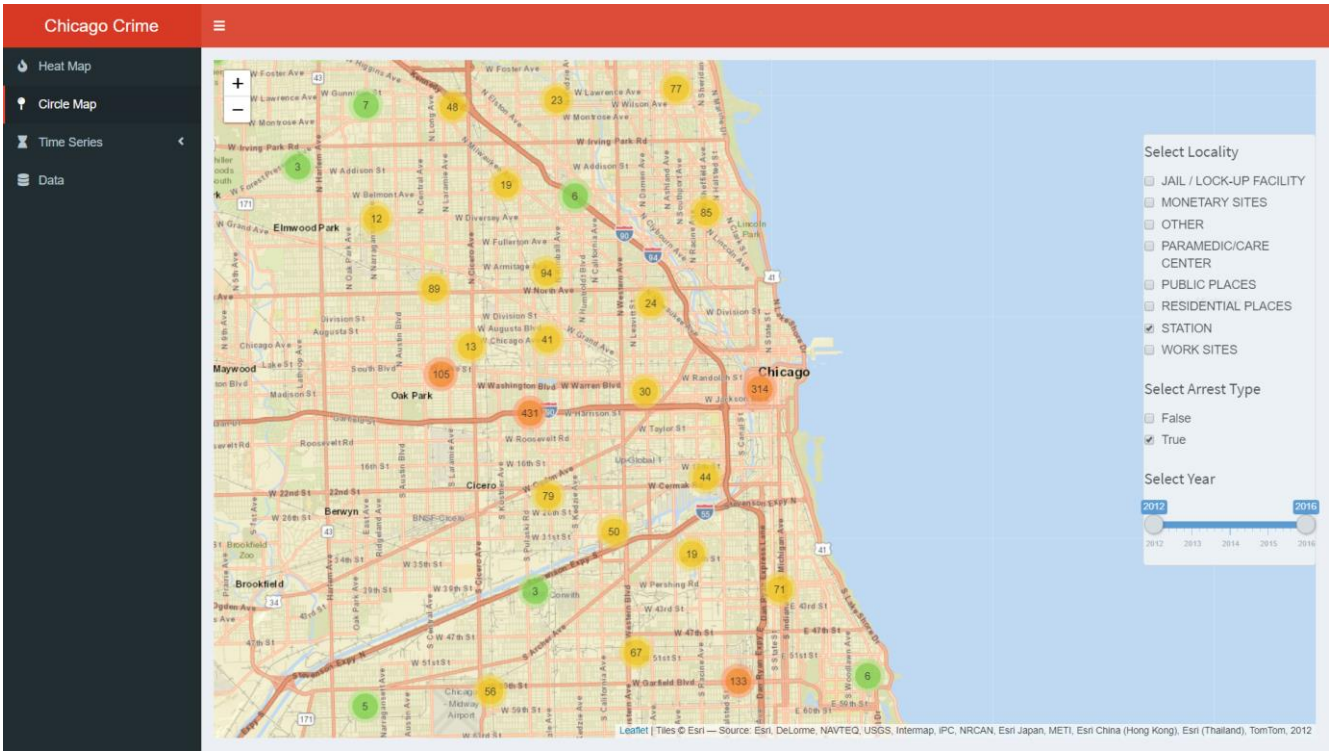
4. Now select the Locations as many as required along with the Arrest Type. Change the Year range if required.



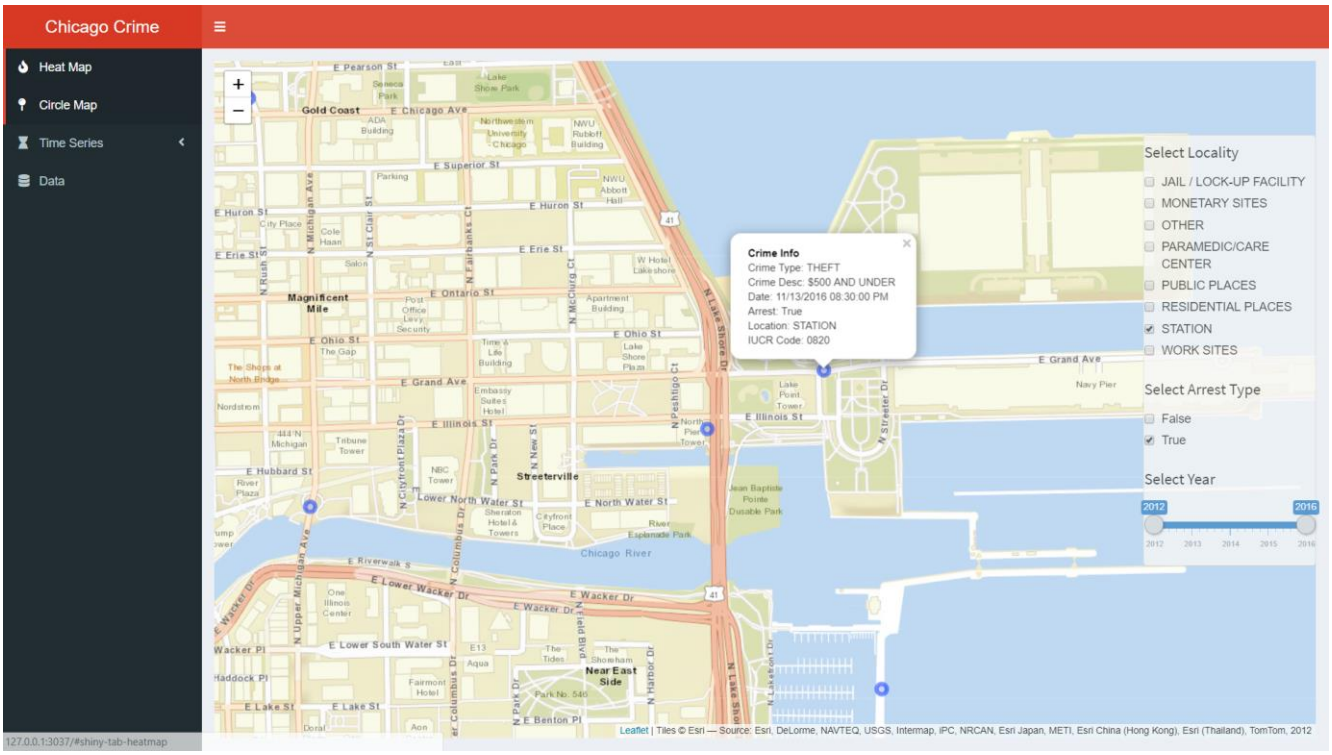
5. Now zoom in heat map for clearer look.
6. After Heat Map select Circle Map from Navigation Pane.



7. Select the Locality as many as you want.



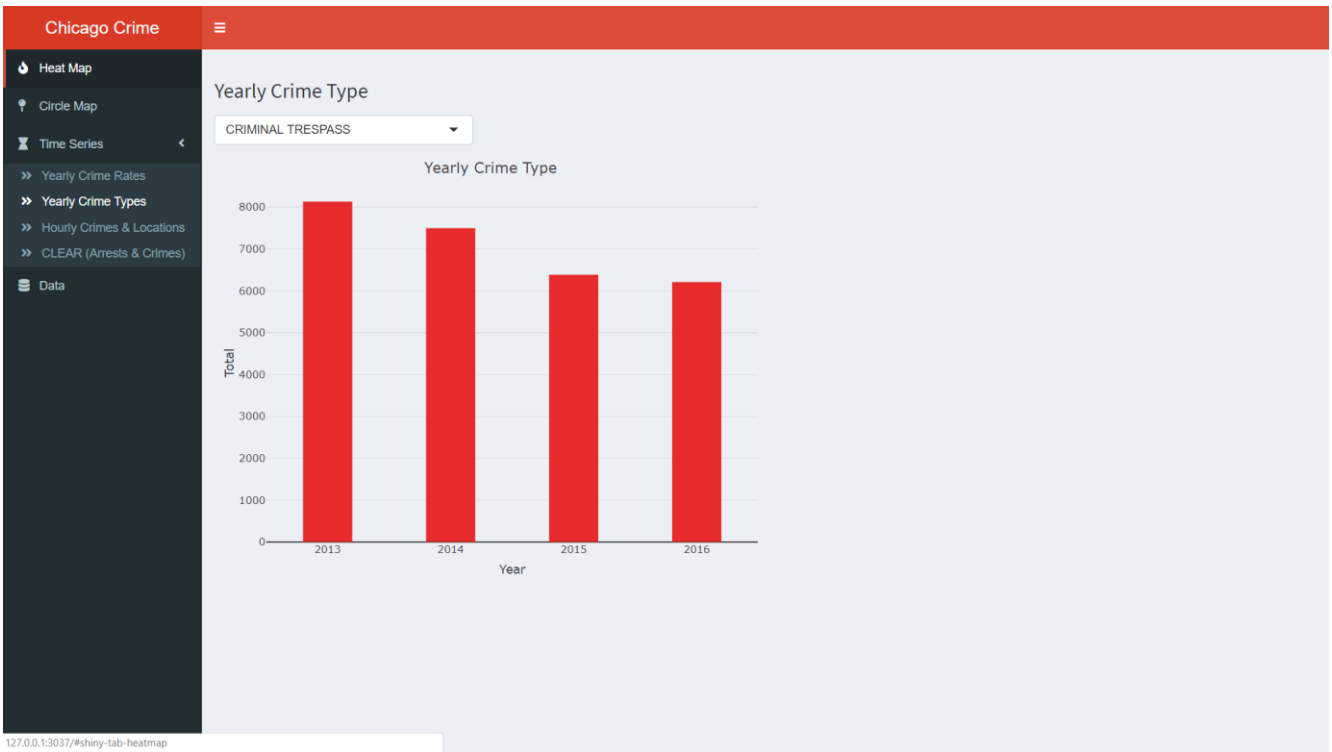
8. Now Click on the bubble / circle to view the Tool Tip Information.



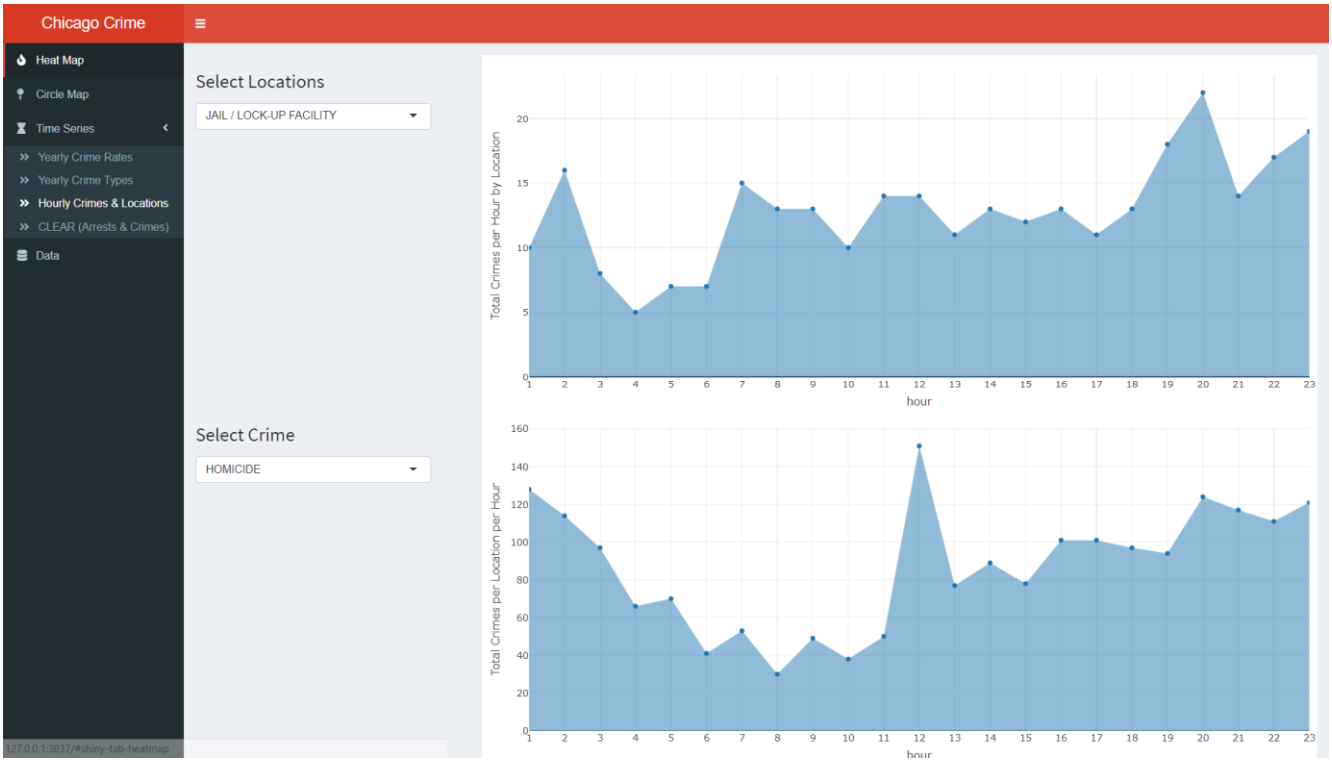
This tooltip info is helpful for Chicago Police Department.

9. Coming to Time Series in the Navigation Pane, Click on Yearly Crime Rates and select the parameter.

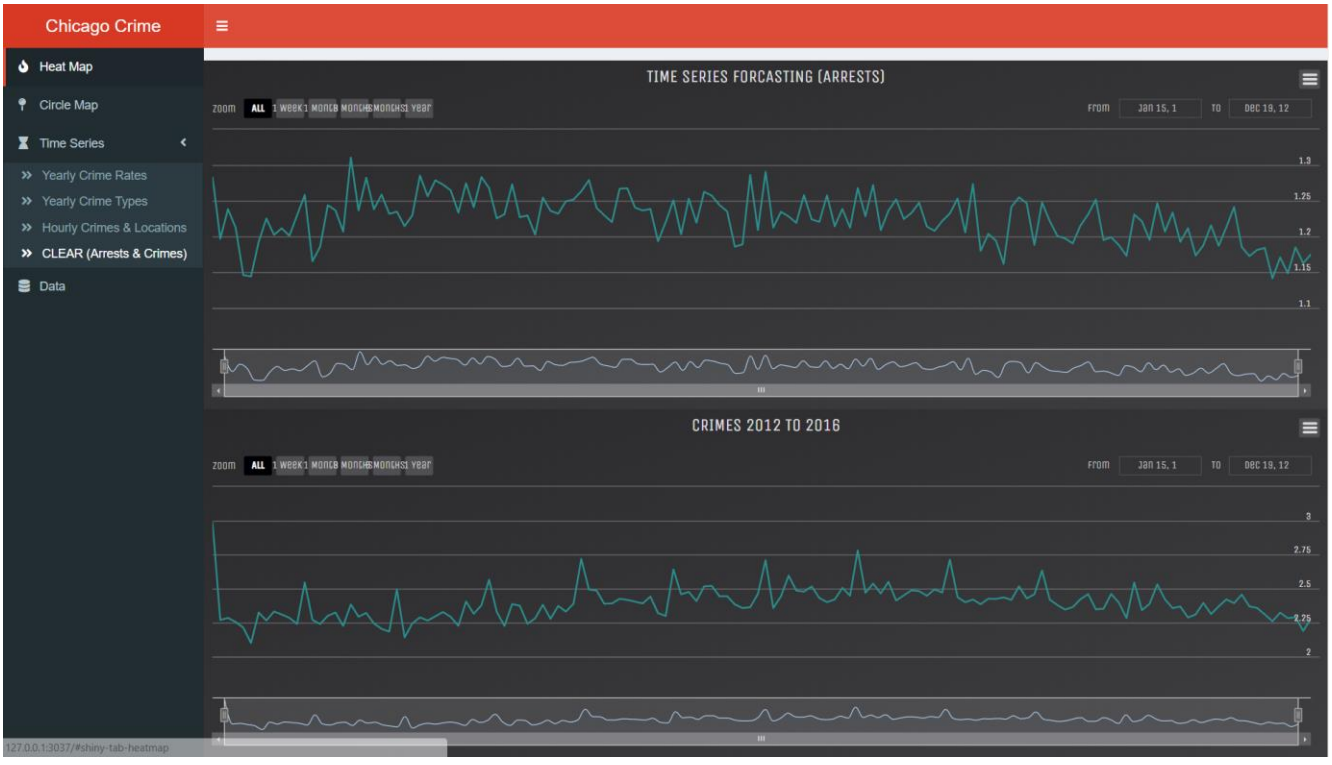
10. Hover pointer on the plot for tooltip information.
11. Next select Yearly Crime Types.



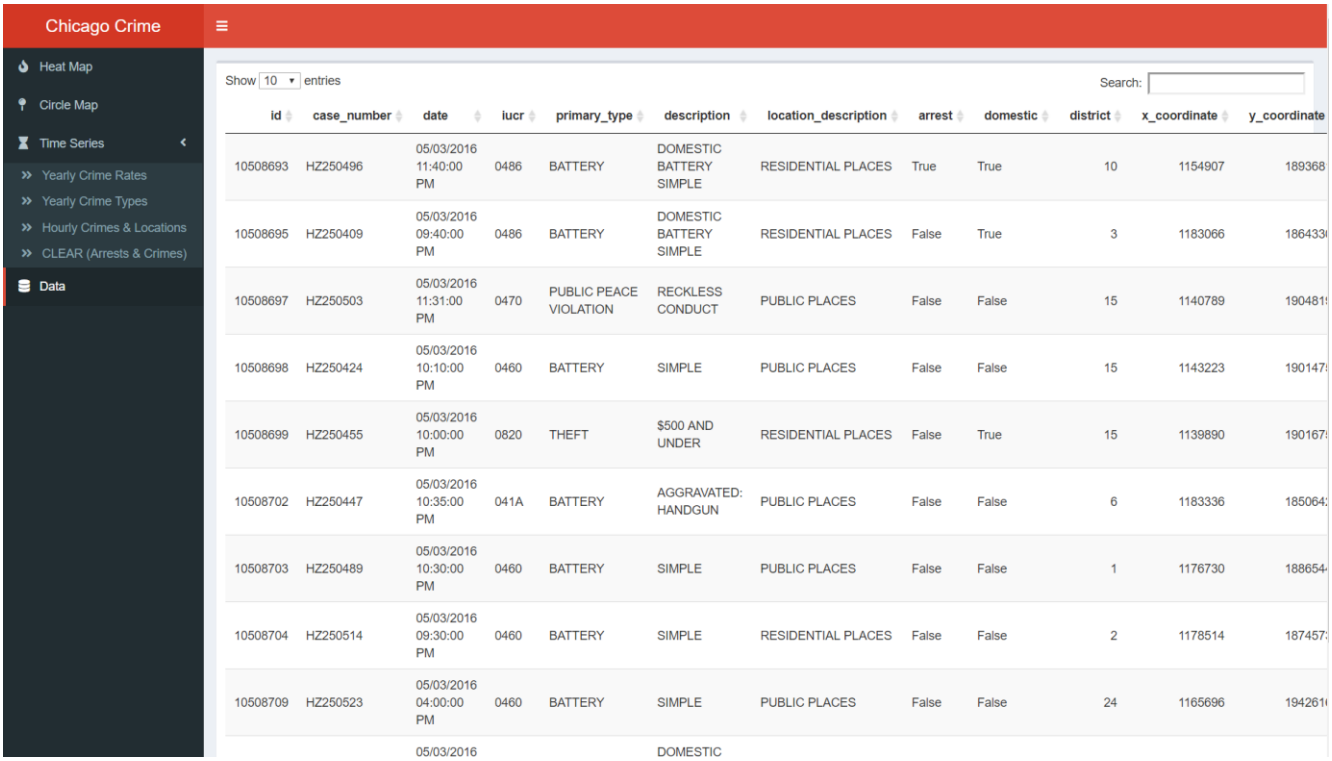
12. Hover pointer on the plot for tooltip information.
13. Select Hourly Crimes & Locations.



14. Hover pointer on the plot for tooltip information.
15. Select CLEAR (Arrests & Crimes) for visualizing Time series data. This will take some time to load because of huge data. So please be patient.



16. Hover pointer on the plot for tooltip information and select the information by dragging the bar at the bottom or selecting the zoom parameter or entering from and to.
17. Finally, select data to view the data that is used for this project. For this Data Table is used, so dynamic searching is possible.



id	case_number	date	iucr	primary_type	description	location_description	arrest	domestic	district	x_coordinate	y_coordinate
10508693	HZ250496	05/03/2016 11:40:00 PM	0486	BATTERY	DOMESTIC BATTERY SIMPLE	RESIDENTIAL PLACES	True	True	10	1154907	189368
10508695	HZ250409	05/03/2016 09:40:00 PM	0486	BATTERY	DOMESTIC BATTERY SIMPLE	RESIDENTIAL PLACES	False	True	3	1183066	186433
10508697	HZ250503	05/03/2016 11:31:00 PM	0470	PUBLIC PEACE VIOLATION	RECKLESS CONDUCT	PUBLIC PLACES	False	False	15	1140789	190481
10508698	HZ250424	05/03/2016 10:10:00 PM	0460	BATTERY	SIMPLE	PUBLIC PLACES	False	False	15	1143223	190147
10508699	HZ250455	05/03/2016 10:00:00 PM	0820	THEFT	$500 AND UNDER	RESIDENTIAL PLACES	False	True	15	1139890	190167
10508702	HZ250447	05/03/2016 10:35:00 PM	041A	BATTERY	AGGRAVATED: HANDGUN	PUBLIC PLACES	False	False	6	1183336	185064
10508703	HZ250489	05/03/2016 10:30:00 PM	0460	BATTERY	SIMPLE	PUBLIC PLACES	False	False	1	1176730	188654
10508704	HZ250514	05/03/2016 09:30:00 PM	0460	BATTERY	SIMPLE	RESIDENTIAL PLACES	False	False	2	1178514	187457
10508709	HZ250523	05/03/2016 04:00:00 PM	0460	BATTERY	SIMPLE	PUBLIC PLACES	False	False	24	1165696	194261
		05/03/2016			DOMESTIC						

So, these are the brief instructions for viewing and exploring the narrative visualization using a standard web browser.

Conclusion

In doing this project I learnt various phases in visualizing the data, starting from finding the topic of interest to molding it in to a useful information. On the other hand, I learnt using various libraries that used for prediction and time series analysis like Highcharter, xts and leaflet. Moreover, to improve result I can use some predictive analysis using TensorFlow or Keras and plot the result accordingly. Also, I can link this application with Database like MongoDb for simulating real-time data.

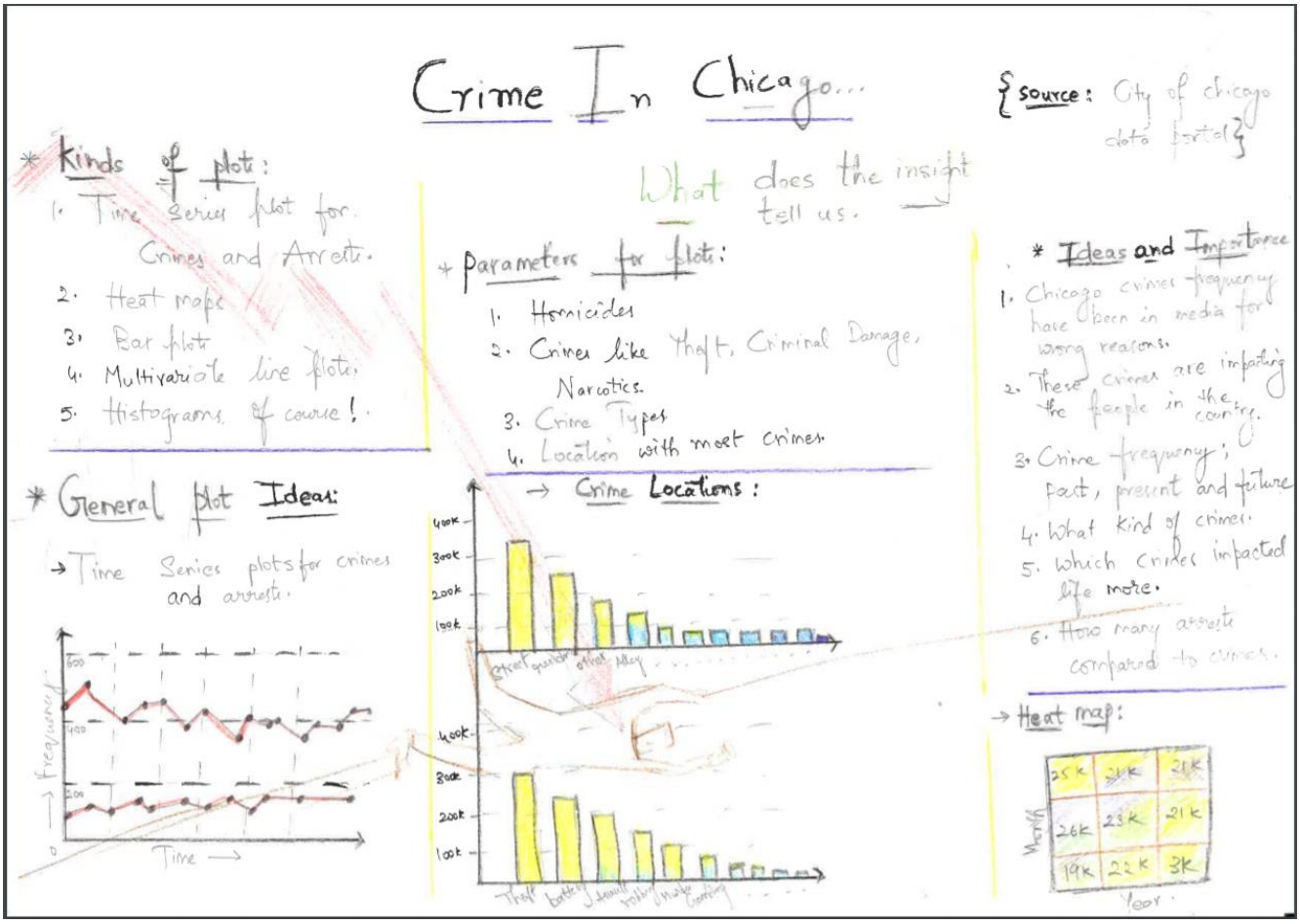
References

(https://rstudio.github.io/leaflet/, n.d.)
(http://jkunst.com/highcharter/index.html, n.d.)
(https://en.wikipedia.org/wiki/Crime_in_Chicago, n.d.)

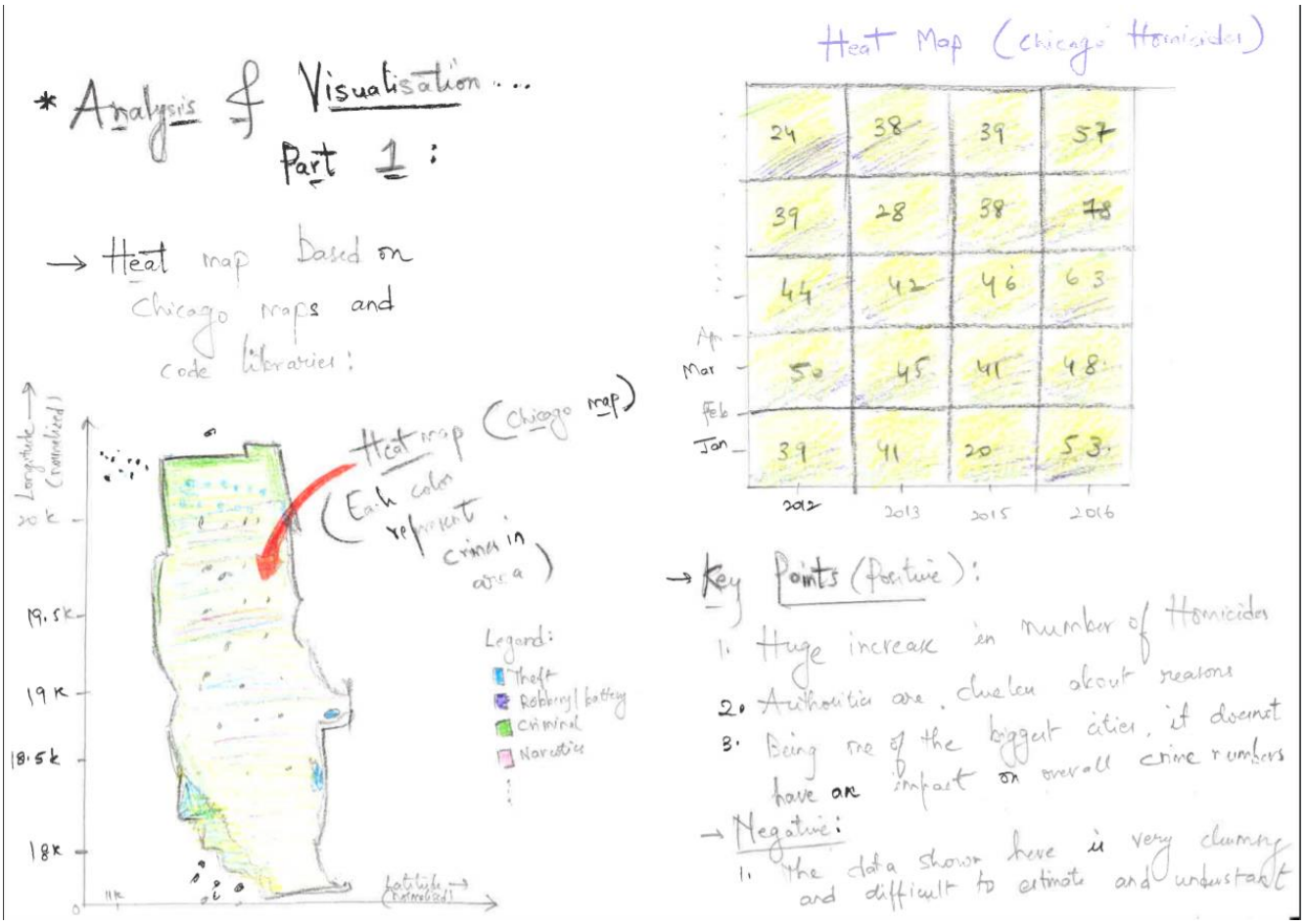
Appendix

Five Design Sheets;

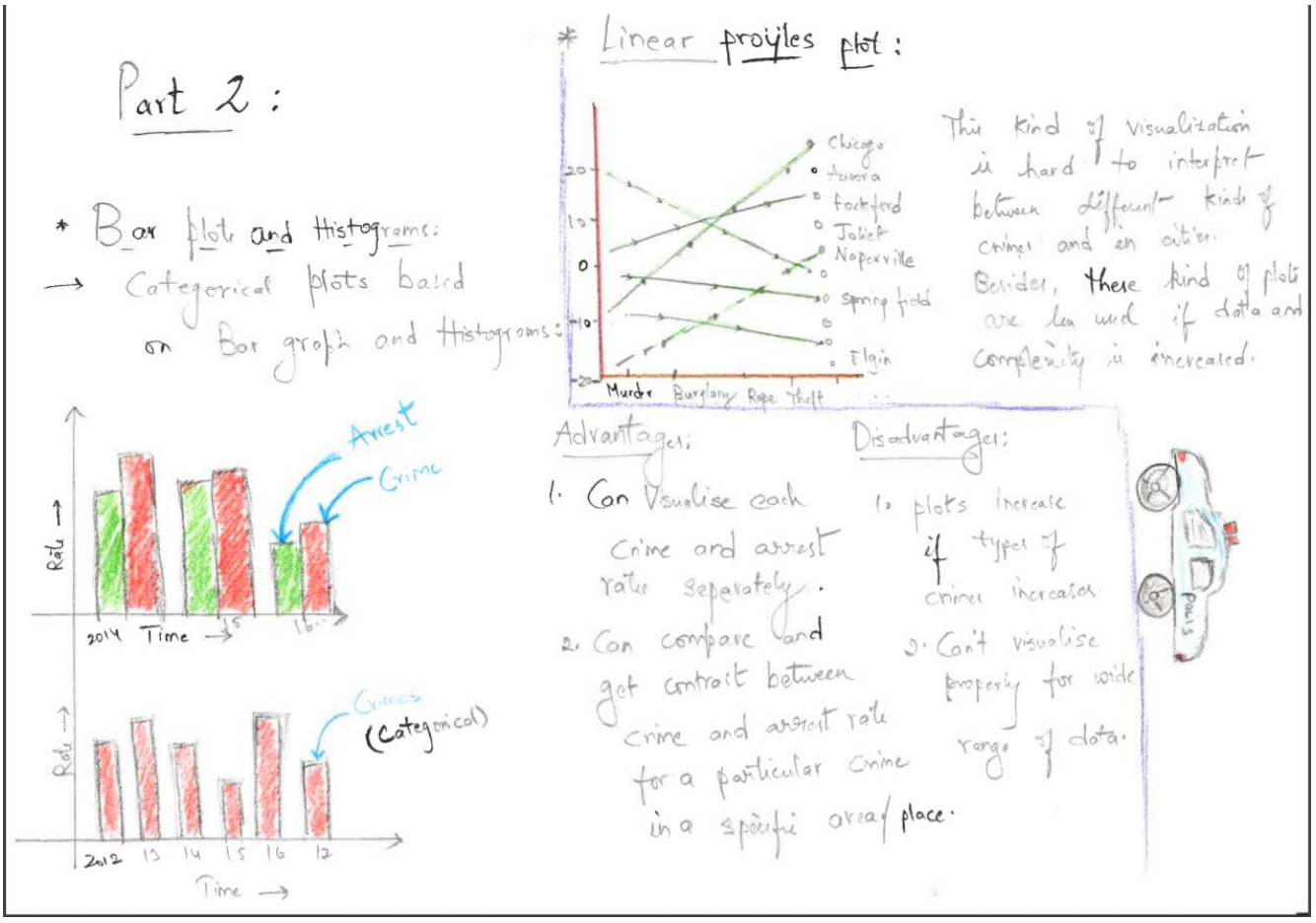
1. Sheet 1:



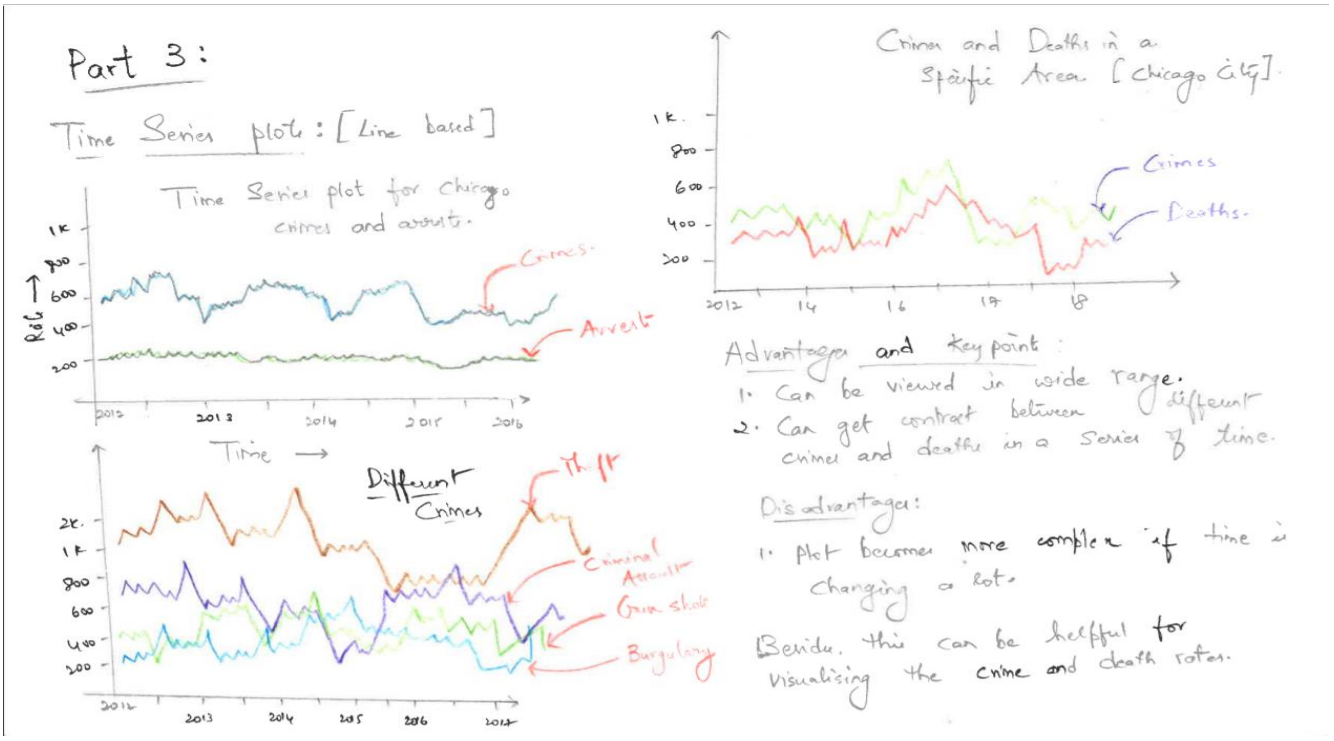
2. Sheet 2:


* Analysis & Visualisation ...
Part 1 :
→ Heat map based on Chicago maps and code libraries :
Heat map (Chicago map)
(Each color represent crimes in area)
Legend:
Theft
Robbery/ battery
criminal
Narcotics
Heat Map (Chicago Homicides)
24 38 39 57
39 28 58 78
44 42 46 63
50 45 41 48
39 41 20 53
Apr
Mar
Feb
Jan
2012
2013
2015
2016
→ Key Points (Positive):
1. Huge increase in number of Homicides
2. Authorities are clueless about reasons
3. Being one of the biggest cities, it doesn't have an impact on overall crime numbers
→ Negative:
1. The data shown here is very clumsy and difficult to estimate and understand

3. Sheet 3:


Part 2 :
* Bar plot and Histograms:
→ Categorical plots based on Bar graph and Histograms:
Rate
Arrest
Crime
2014 Time →
Crimes (Categorical)
2012 13 14 15 16 17
Time →
* Linear profiles plot :
Chicago
Aurora
Rockford
Joliet
Naperville
Springfield
Elgin
20
10
0
-10
-20
Murder Burglary Rape Theft
This kind of visualization is hard to interpret between different kinds of crimes and en cities.
Besides, these kind of plots are less used if data and complexity is increased.
Advantages:
1. Can visualise each crime and arrest rate separately.
2. Can compare and get contrait between crime and arrest rate for a particular crime in a specific area/place.
Disadvantages:
1. plots increase if types of crimes increases
2. Can't visualise properly for wide range of data.
Police

4. Sheet 4:



5. Sheet 5: