

Results:

Consider a use case where the user has selected below options from the GUI: For the below example user has made selection as

City: Chicago

Zip code: 60601

Type of Crime: Battery

Year: 2015

Month: January

Date: 01 &

Time: 00-01 AM

Once user submit these option is gets reflected on the terminal as shown on the left side of the screen and also gets saved as a .csv file that would further be used for the analysis process.

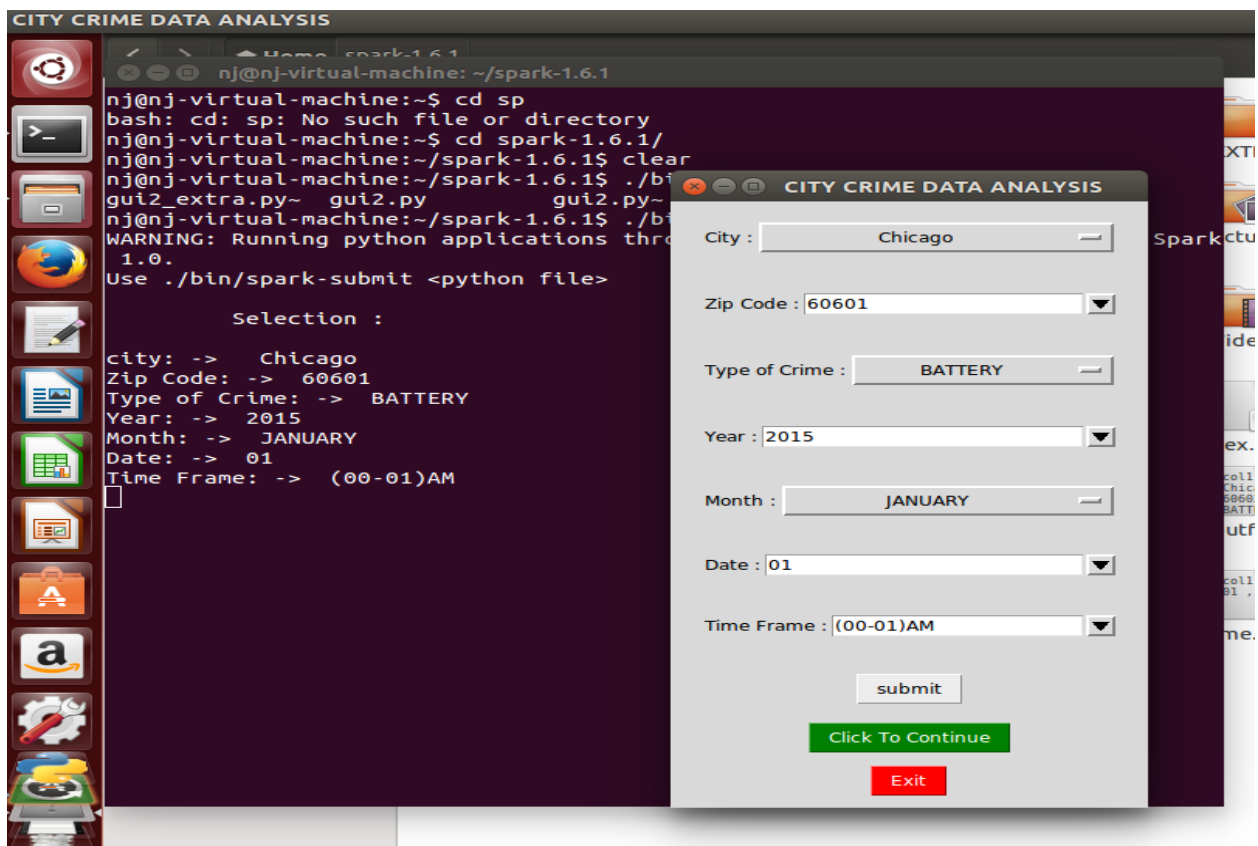


Figure : GUI ;- Graphical User Interface

- After user submit the selection and click on the further process, the code in R runs the analysis on the data and present the end results on GUI as shown below.

Here it shows that the particular crime during the chosen time period happened for 208 times.

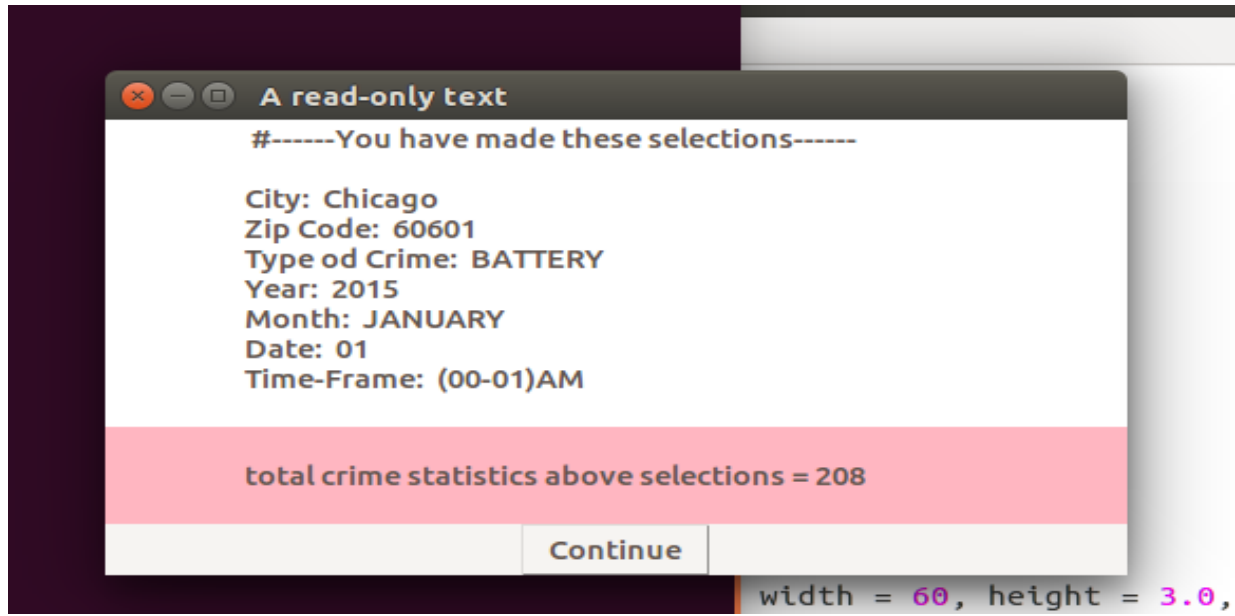


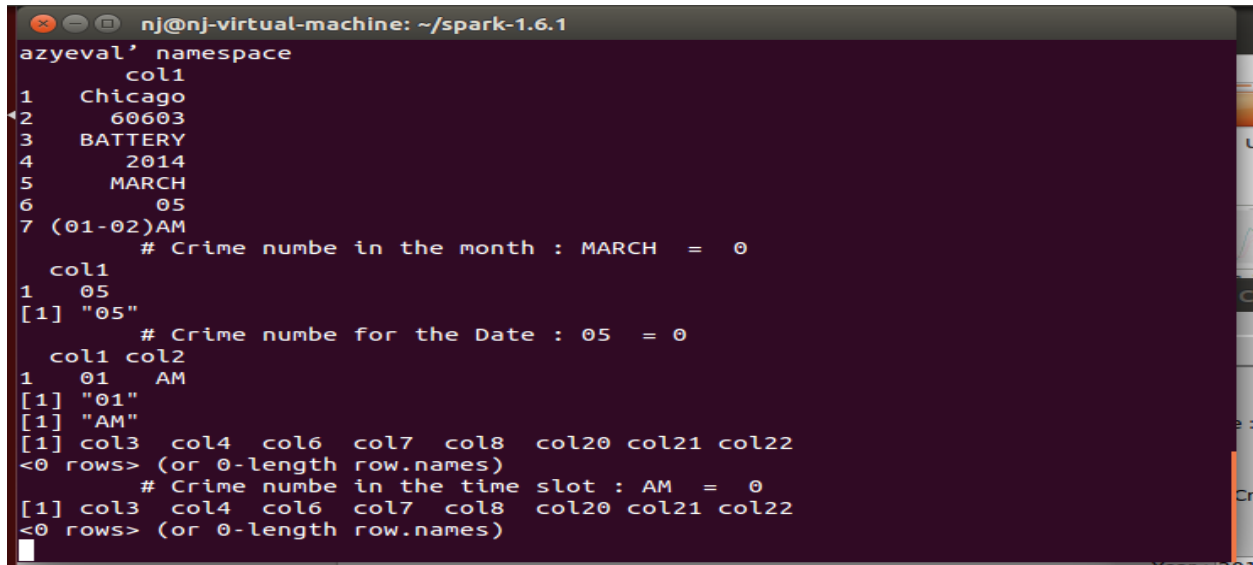
Figure: GUI selection and crime statistic result

- If we remove the nice GUI presentation, we can see actual process running at the background. This process sample for one of the option selection is shown here in the below figure.

```
nj@nj-virtual-machine: ~/spark-1.6.1  
The following object is masked from 'package:stats':  
  nobs  
The following object is masked from 'package:utils':  
  object.size  
The following object is masked from 'package:base':  
  startsWith  
# City : Chicago  
col1  
1 60603  
[1] "data.frame"  
[1] 1511  
# Over all Crime Number from the Chicago City data File = 1511  
# Total Data Logs from the file = 33242  
[1] 35  
# Total crime for the selected Zip code: 60603 = 35  
col1  
1 BATTERY  
[1] "BATTERY"  
# Type of Crime Selected: BATTERY  
# Crime numbe for : BATTERY = 2  
# Crime numbe in : 2014 = 0  
Attaching package: 'dplyr'  
The following objects are masked from 'package:gdata':  
  combine, first, last  
The following objects are masked from 'package:stats':  
  filter, lag  
The following objects are masked from 'package:base':  
  intersect, setdiff, setequal, union
```

Figure : Screenshot of the process

- This is another screenshot that shows the subset process in R at the backend. Simplified version of these process has been taken out to the GUI for nice user visual analysis.



```
nj@nj-virtual-machine: ~/spark-1.6.1
azyeval' namespace
  col1
1   Chicago
2   60603
3   BATTERY
4   2014
5   MARCH
6   05
7 (01-02)AM
# Crime numbe in the month : MARCH = 0
  col1
1   05
[1] "05"
# Crime numbe for the Date : 05 = 0
  col1 col2
1   01   AM
[1] "01"
[1] "AM"
[1] col3 col4 col6 col7 col8 col20 col21 col22
<0 rows> (or 0-length row.names)
# Crime numbe in the time slot : AM = 0
[1] col3 col4 col6 col7 col8 col20 col21 col22
<0 rows> (or 0-length row.names)
```

Figure : Screenshot of the process

- One of the key feature of this project is to represent the data analysis through map. As a part of this feature, for the visual analysis, based on the city selection from the GUI user will see the city map for the chosen option. In this case it is Chicago.

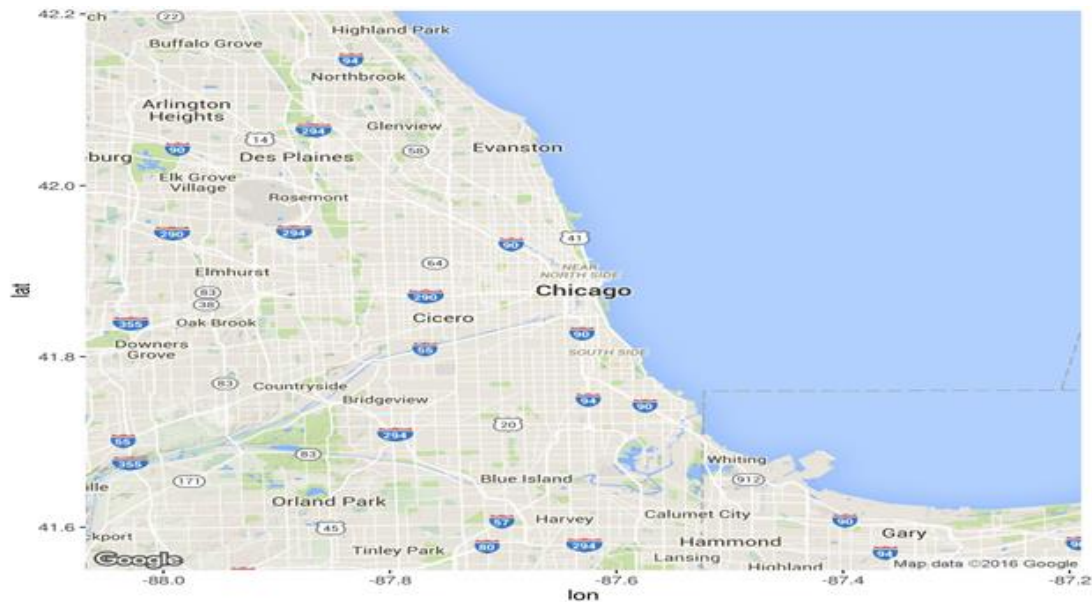


Figure: Chicago map

- After this, the data analysis for the chosen city is processed that takes whole data of the crime happened over the year from 2001 to 2016 and that will be represented through frequency graph as followed, to read the color code it is taken those area dark right to light right based on more crime to less crime frequency for the different parts of the city.

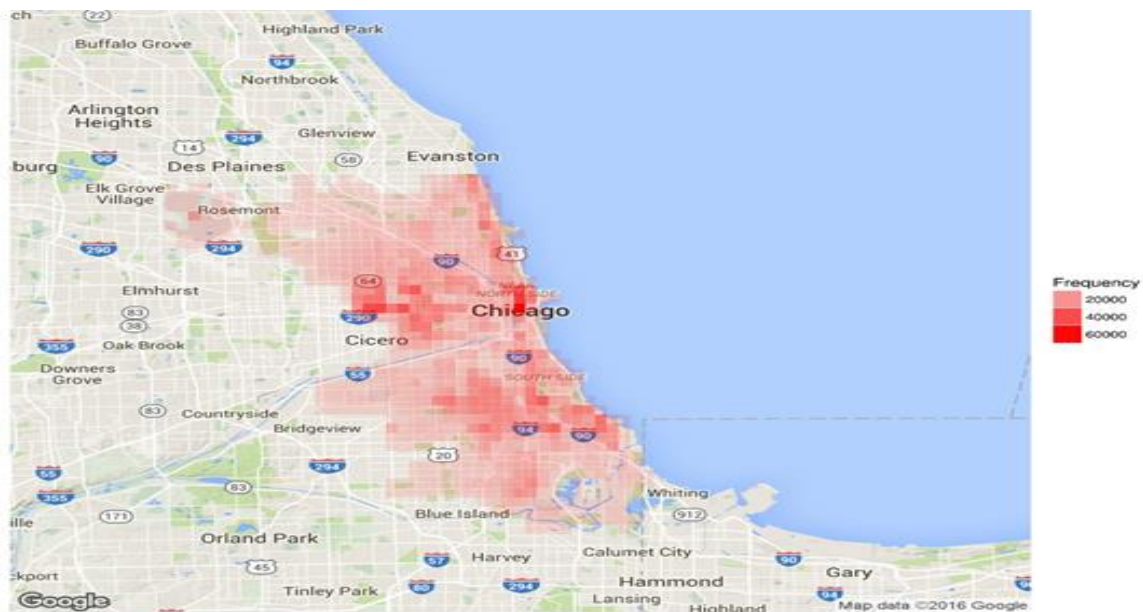


Figure: Crime Frequency over the years form 2001-2016

- Next result on the screen you would see is for overall crime data analysis line graph over the year from 2001 to 2016. I have implemented different color coding for different days of the week for the easy understanding of the graph: X-axis is the data for the Hour from 0 to 24, Y-axis shows total number of crimes for particular day of the week.
- You can see huge increased spike on Friday to Sunday that shows that during these days of the week crime rate is high.
- It's also interesting to notice that crime rate is high during the time of 8:00PM to 1:00 AM and it's very low from 5:00AM till 10:00AM.

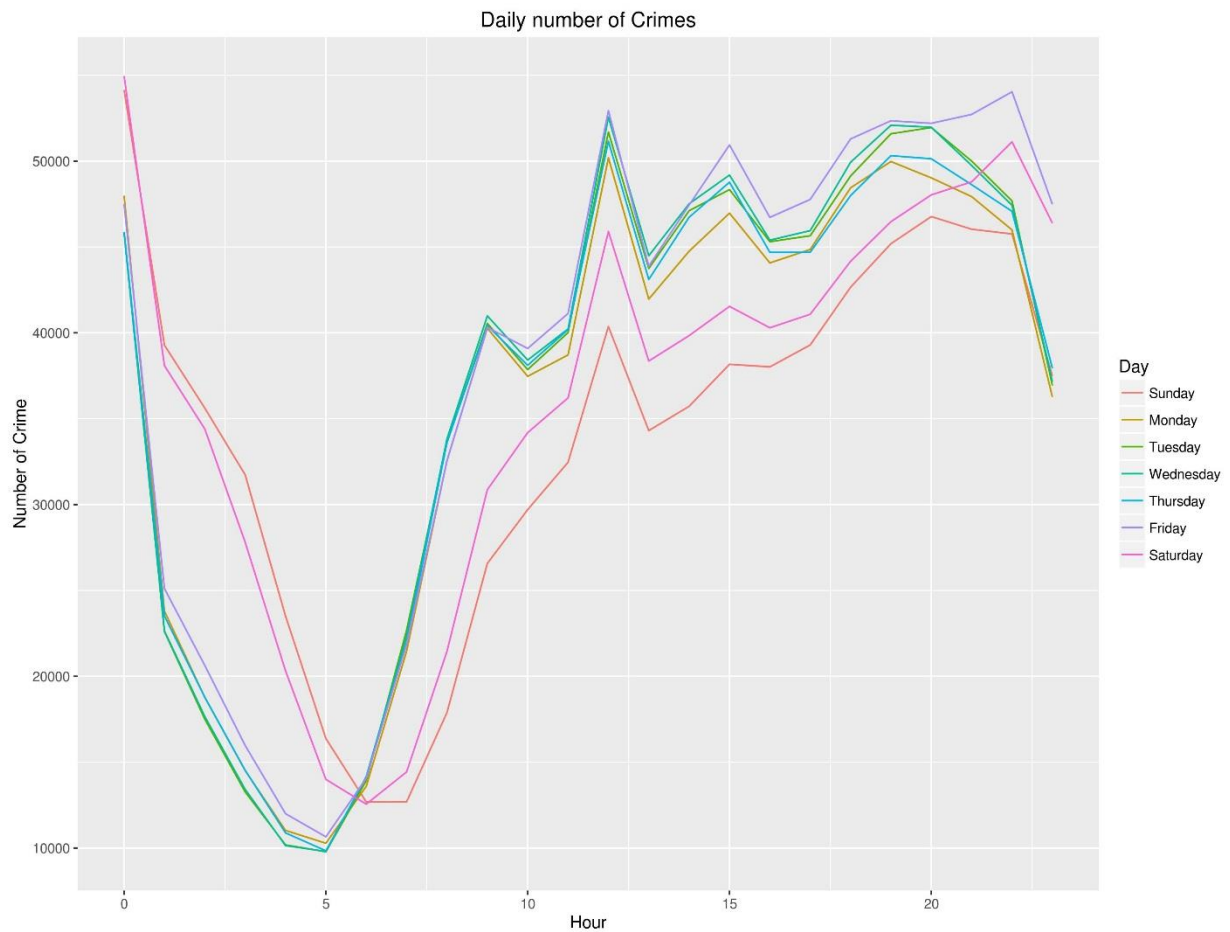


Figure : overall crime data analysis for seven days of week ranging from 0 to 24 hours

- In the next result, Chicago graph will be plotted for the crime data with the red dots (based on its locations) for the year 2015 as per the user selection earlier.00
- At the backend the code makes a call for the google map API through 'ggmap2' library in R, later it downloads the image as per the user selection and plotting requirements. These graphs may vary & it depends on the GUI selection, which in this case it's for 2015
- This is another way of visual analysis of the data we have in the file.

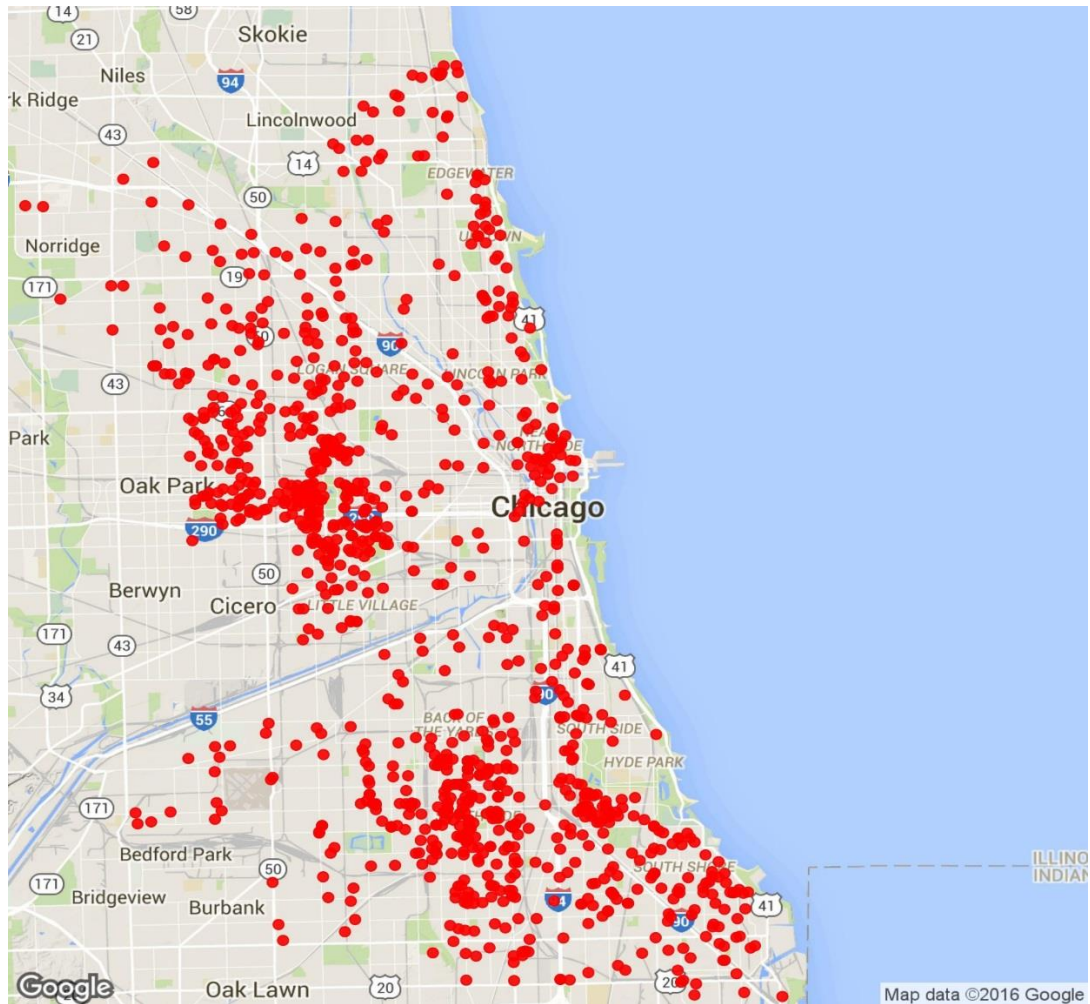


Figure : Google map crime data plotting for the year of 2015

- Below is the same graphical plotting for the above data, the only change is the “Hybrid view” of the map.

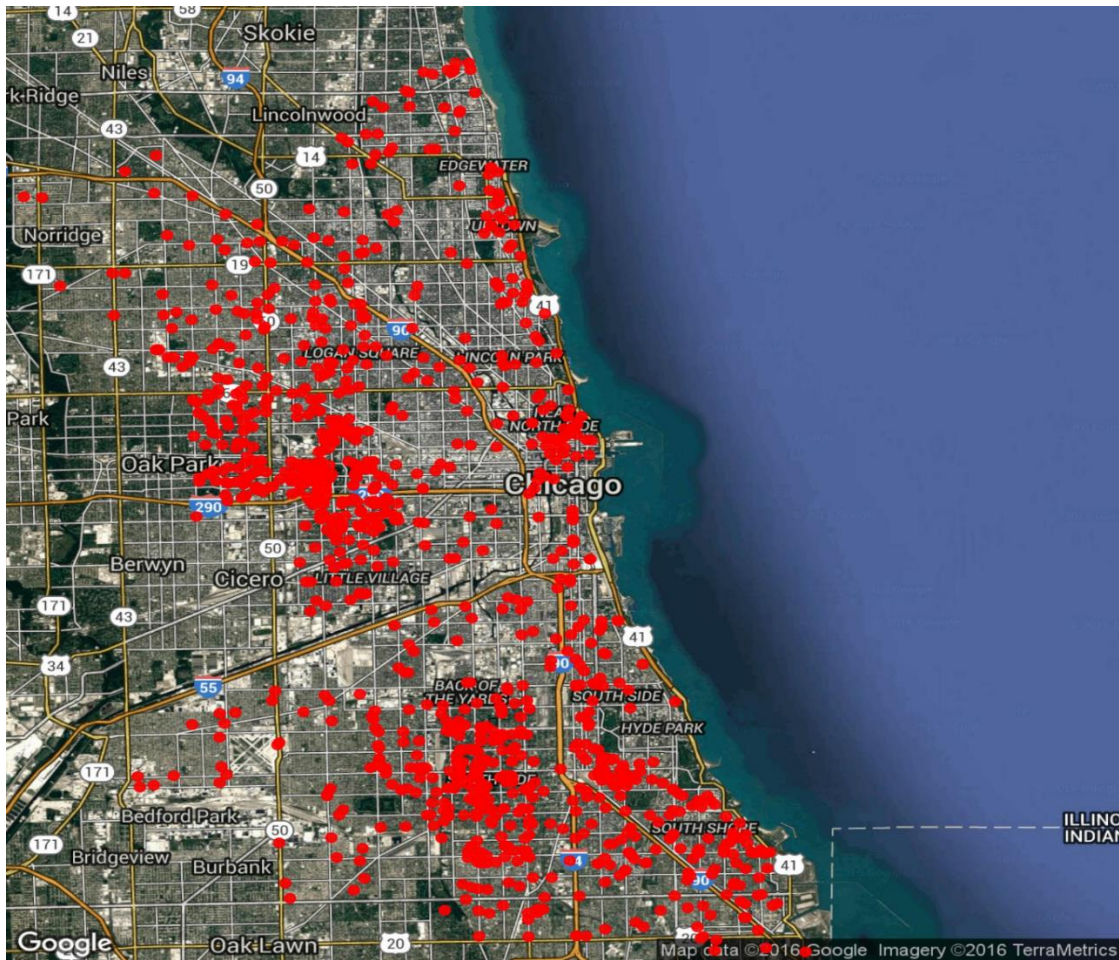


Figure : Hybrid View of Google map crime data plotting for the year of 2015

- Crime Density map is another way of representation of the overall crime data for the user selected year within the GUI option (in this case selection is made for 2015).
- As you can see areas with maximum crime rate is just covered with normal blue line and less crime prone areas show the orange and yellow color shade which means these sectors are comparatively safe during the year of 2015 compare to the others.

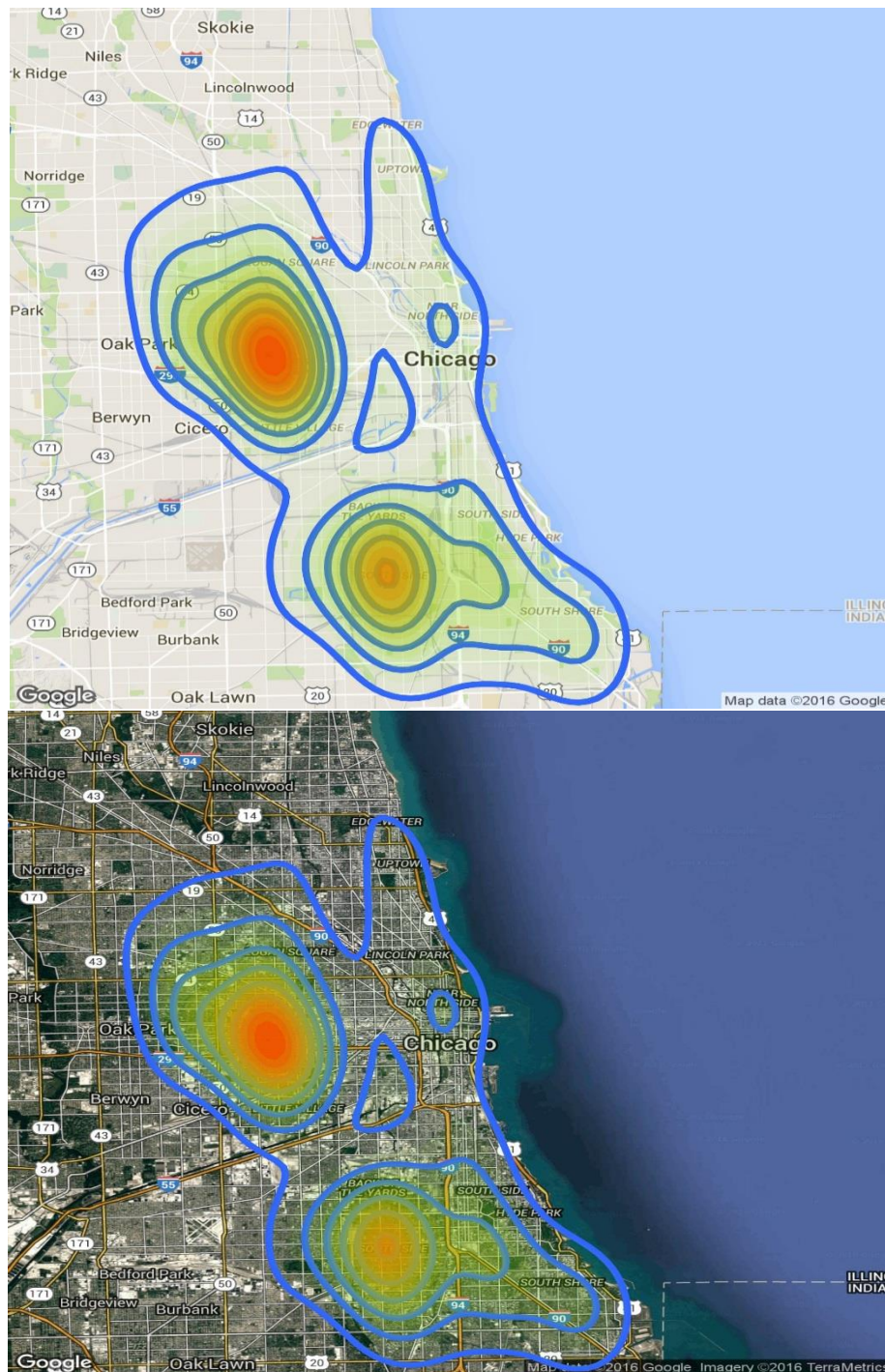


Figure : Cime Density map (Roadmap & Hybrid Map)

- One of the interesting analytical map has been abstracted for overall dataset is called “Heat Map”. Its interesting because it shows Hours Vs. Days Comparison for overall crime
- The color frequency is done such a way that less crime shown in black color and more is red; so it goes from black to red depends on the crime rate frequency.
- Reading the graph, we can say that on Saturday and Sunday during 11:00PM to 1:00AM crime frequency is up and so on Friday during 8pm to 12:00Am crime rate is up.
- Whereas in weekdays, from Monday to Friday and particularly in the early morning to afternoon the crime rate is comparatively too low !
- This may help to the local law agencies to keep tab on such days and this way they can make some better arrangements to serve people of the city.

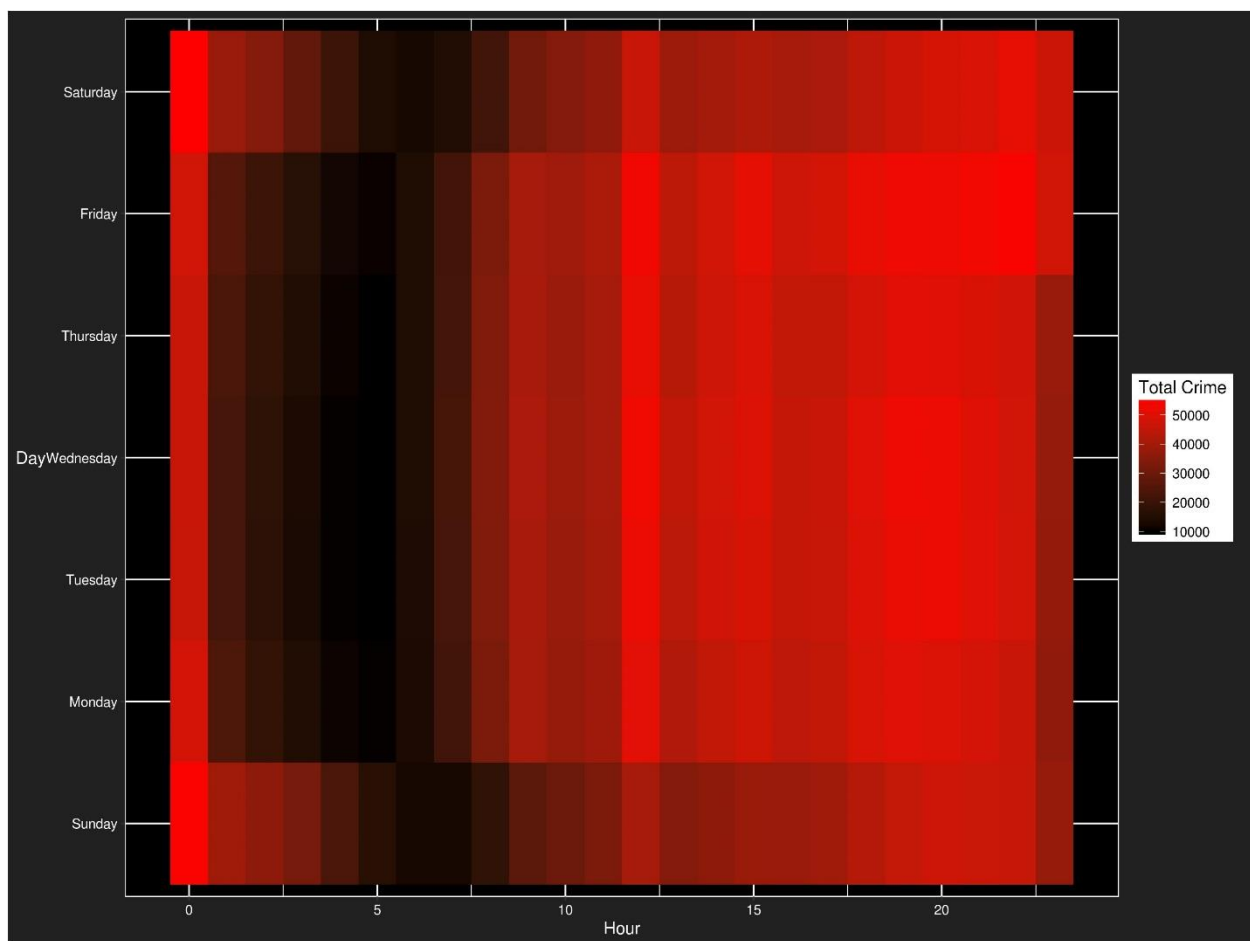


Figure : Heatmap

- Coming to the last results that is for GUI based Zip code selection; once user has made selection for the zip code, the script will run its customized data analysis and would produce the results with the map of that particular zip code area.
- In this case the selection is made for the zip code 60601 and hence for that results from the google map API is produced only for that particular zip code which is bounded with its latitude and longitude at the backend.



Figure : map for the zip code 60601



Figure : Hybrid map for the zip 60601

- The final results shows the highlighted user selected Zip Code area with the yellow color. In another way it distinguishes chosen option from the other and counts the crime data for that particular zip code during the R based GUI results process.

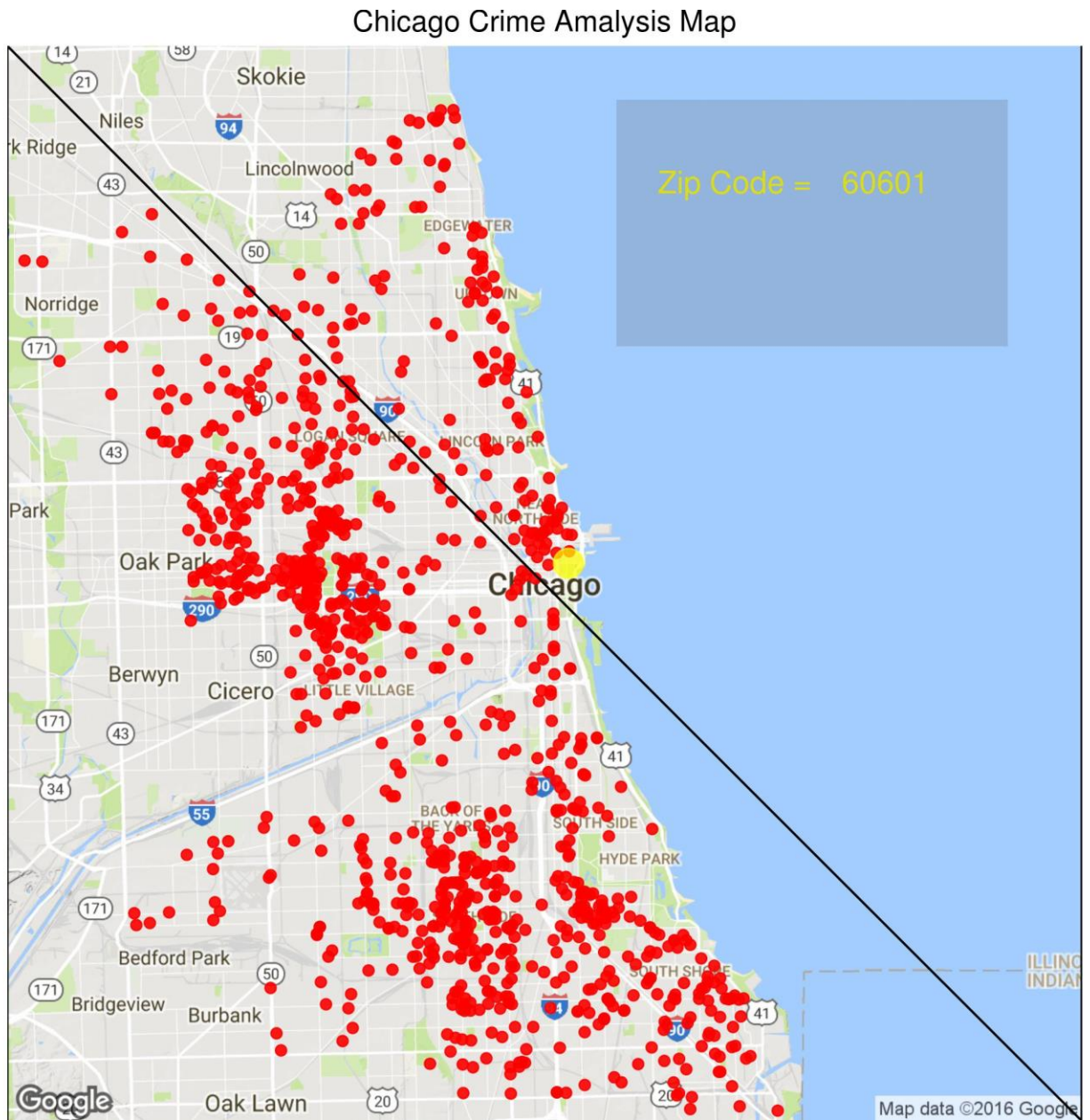


Figure : Zip code Highlight for the user selection

- Below image is the same as described the above with only difference in map type; here we have used “Hybrid” view of the same data presentation as above for the user selected zip code 60601.

Chicago Crime Analysis Map

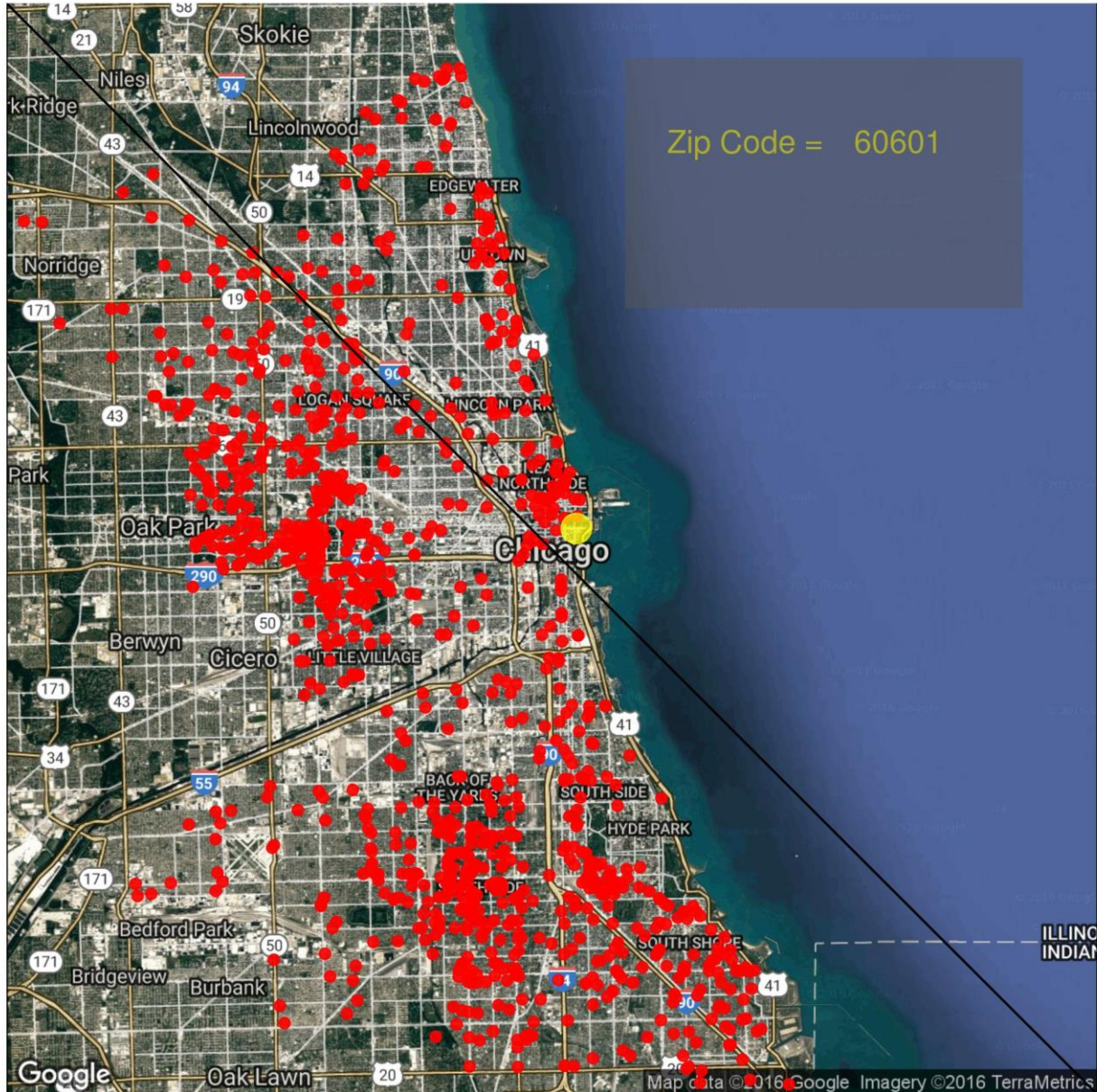


Figure : (Hybrid View) Zip code Highlight for the user selection