International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 8, Issue 3 Ser I (March 2019), PP. 01-08

Geospatial Analysis and visualization of Crime Data using 'R' Open source Software

Murugasan. R[*], Shanmathi. B[**], Lokesh. M[**] and Venkatesan. C[***]
** (Professor, Institute of Remote Sensing, Department of Civil Engg., Anna University, Chennai-600025)*
*** (B.E student, Institute of Remote Sensing, Department of Civil Engg., Anna University, Chennai-600025)*
**** Ph.D. Scholar, Institute of Remote Sensing, Department of Civil Engg., Anna University, Chennai-600025)*
Corresponding Author:Murugasan. R;E

ABSTRACT: *The Exploratory Spatial Data Analysis (ESDA) is required to be carried out for the geospatial analysis and better visualization of crime data. Normally for geospatial analysis and visualization, Geographic Information System (GIS) software like ArcGIS, QGIS etc. are used. But conventional GIS software are not found to be effective for ESDA. Recently 'R' Software has been identified as alternate to GIS software which can perform all the functions similar to GIS software, but at the same time it performs ESDA also which is preferred for crime data analysis.In this study, ESDA was carried out in R Software with six types of crime data. The dummy crime data of Thiruvallur District in Tamilnadu was considered in this study for the purpose of analysis and visualization. Variety of visualisation techniques were attempted for the crime data and it was found that R Software had the capability of exploratory spatial data analysis and better visualization of crime data.*
Keywords*: ESDA, Crime Data, Analysis, Visualization*

Date of Submission: 26-06-2019 Date of acceptance: 11-07-2019

I. INTRODUCTION

Monitoring crimes by analyzing the spatial distribution its occurrences is useful for the prediction of crimes for police peoples. Effective analysis of non-spatial crime data along with spatial information has become challenge for the police peoples. The Exploratory Spatial Data Analysis (ESDA) is required to be carried out for the geospatial analysis and better visualization of crime data [1]. ESDA is an extension of Exploratory Data Analysis (EDA) which is used to extract and detect spatial properties of data [2,3,4]. There is a need of additional techniques to those found in EDA for carrying out spatial analysis and modelling. It is also important to be able to link numerical and graphical procedures with the map [5,6]. Normally for any kind of spatial analysis, Geographic Information System (GIS) software like ArcGIS, QGIS etc. are employed. GIS has not been developed with ESDA in mind but rather in terms of data management, cartographic modelling and some selected forms of spatial analysis (ex. network analysis). Therefore, traditional GIS software alone is not sufficient for Exploratory Spatial Data Analysis. To complete a task of Exploratory Spatial Data Analysis, most of the traditional GIS software were in need of supporting software or two or more GIS software were employed in a single ESDA application project. Recently 'R' Software has been identified as alternate to GIS software which can perform all the functions similar to GIS software, but at the same time it performs ESDA also which is preferred for crime data analysis [7,8].

R is a free and open source computer program for processing data. It runs on all major operating systems and relies primarily on the command line for data input (www.r.project.org) [9]. This means that instead of interacting with the program by clicking on different parts of the screen via a graphical user interface (GUI), users type commands for the operations they wish to complete. A key benefit is that commands sent to 'R' can be stored and repeated from scripts. This facilitates transparent and reproducible research by removing the need for software licenses and encouraging documentation of code. Furthermore, access to R's source code and the provision of a framework for extensions has enabled many programmers to improve on the basic, or 'base', R functionality. As a result, there are now more than 5000 official add-on packages, allowing R to tackle almost any numerical problem. If there is a useful function that R cannot currently perform, it is likely that someone is working on a solution. One area where extension of 'R's basic capabilities have been particularly successful in recent years is the addition of a wide variety of spatial analysis and visualization tools [8].

Hence it has been decided in this study to demonstrate the capability of R software by analyzing and visualizing the dummy crime data of Thiruvallur District in Tamilnadu.

II. LITERATURE REVIEW

The literature review was carried out in detail to understand the present status in analysis and visualization of crime data and the abstract of few selective literatures are given below.

Hirschfield et al carried out GIS analysis of spatially-referenced crime data for Merseyside, U.K. Their project had demonstrated the flexibility of PC ArcView as a framework for integrating and cross-referencing a range of disparate data sets [10].Anderson and Joaquin suggested control chart techniques to analyse crime rates using ARIMA (Autoregressive Integrated Moving Average) time series models and statistical process control techniques to determine if changes occurred in the process that generated a series of observations [11].

Brown et al proposed ReCAP (Regional Crime Analysis Program), a software system which provides crime analysts with the capability to mine data to catch criminals. This software provides spatial, temporal, and attribute matching methods for mining data [12]. Messner carried out ESDA with homicide crime data of 78 countries in and around St.Louis city and observed positive spatial autocorrelation [13].

Murray et al used GIS for the analysis of crime data in Brisbane city and demonstrated the potentiality of spatial analysis in crime occurrence prediction [14]. Ratcliffe discussed the future training needs using a simple model of intelligence-led crime reduction. This model suggested that training for managers to enable a greater understanding of the analyses presented to them, and how to use mapping to further crime prevention and reduction, may be as important as increasing the technical ability of crime analysts [15].Grubesic outlined several problematic aspects of optimization-based cluster analysis for crime hot spot detection. This study demonstrates the capability of the statistical techniques for the analysis of spatio-temporal crime data of Ohio city [16].

R was conceived and is still primarily known for its capabilities as a 'statistical programming language'. Spatial data analysis and visualization is an important growth area within this increased functionality [17].Kelly et al illustrated 'R' as a valuable tool for performing a wide variety of statistical analyses. They also suggested that statisticians and quantitative methodologists can use 'R' for their research [18].

Miller et al proposed RTM as an approach to risk assessment to aid in crime forecasting by incorporating underlying causes of crimes and standardizing all of these factors to common geographic units over a continuous surface. Essentially, RTM assigns a (weighted or un-weighted) value signifying the presence, absence or intensity of each factor at every place throughout a given geography. Each factor was represented by a separate coverage (raster) map of the same geography. All map layers were combined in a Geographic Information System (GIS), they produce a composite map - a risk terrain map [19].

Dyga and Sławinska proposed histograms and mosaic plots to show the crime rates for greater clarity, presentation of findings and interpretation of them followed by maps depicting crime rates in particular areas of Poland [20].Sengupta et al suggested a method to pull data directly from web, handle, clean, and process crime data, utilize geo-coded information through shape files, retrieve hidden information through visualizations, create new variables from limited data, build a predictive engine for crime, and test how good it is [21].

2.1 Observations from Literature Review

The following observations were made from literature review.

1. Most of the literatures and journals on crime analysis highlighted Exploratory Spatial Data Analysis (ESDA) concerning with the description and exploration of spatial data.
2. Geographic Information System has not been developed with ESDA in mind but rather in terms of data management unlike R.
3. Geographic Information System graphics tend to be strong on presentation of data rather than exploration.
4. It is worth noting that there are a few key differences between 'R' and traditional desktop GIS software. While dedicated GIS programs handle spatial data by default and display the results in a single way, there are various options in R that must be decided by the user.
5. More exposure to R is required for better implementation of results of crime analysis.

III. DATA SET CREATION FOR CRIME ANALYSIS

Adrian A command set in the R programming language was used to extract rows and columns from a given data frame. So, a dataset of crime in the form of Comma Separated Values (CSV) format was found to be appropriate for this study. A dummy spatial data set of crime in CSV format, which includes location (latitude and longitude), date of crime occurrence, month, day and type of offence in excel sheet was created for Thiruvallur district for exploring visualization techniques in R.

IV. BASE MAP PREPARATION AND SCATTER PLOTS

R is able to import a very wide range of spatial data formats by linking with the Geospatial Data Abstraction Library (GDAL). As a library, it presents a single raster abstract data model and single vector abstract data model to the calling application for all supported formats. The basic map of the study area i.e.,

Thiruvallur district generated by R software is shown in Fig. 1. Before exploring the spatial components of data, scatter plots prepared by R software indicating spatial distribution of crime locations are shown in Figure 2 and Figure 3.

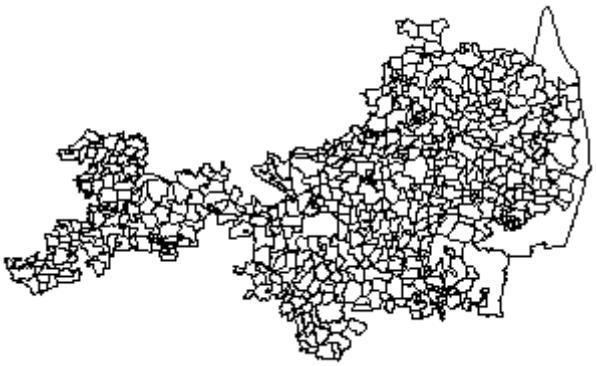

Fig. 1.Basic map of the study area

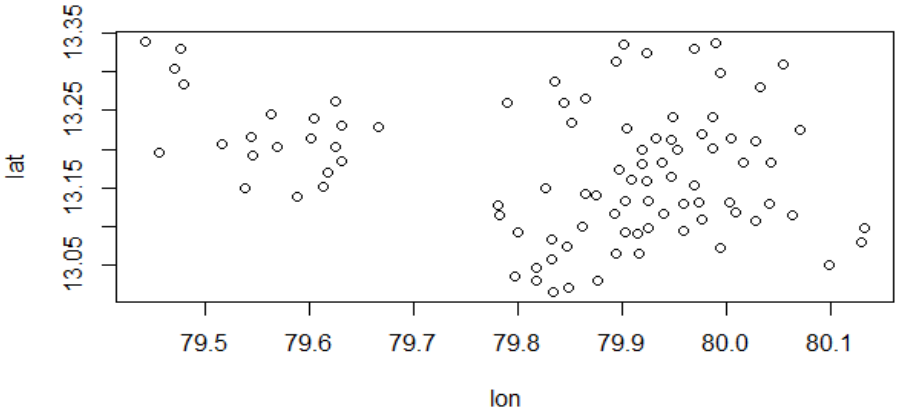

Fig. 2.Spatial distribution of total crimes

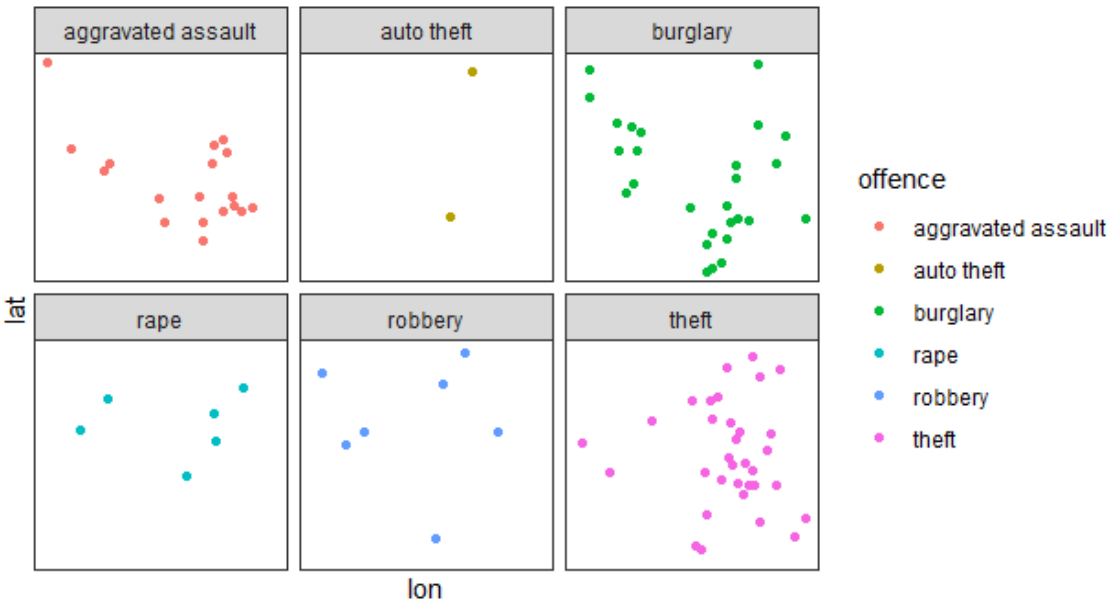

Fig. 3. Spatial distribution of each type of crime

R software also allows to download and plot basic maps from Google maps (and other maps). These maps can then be used as layers within R software for quick geospatial visualization. The spatial distribution of crime locations on Google maps generated by R software are shown in Fig. 4.

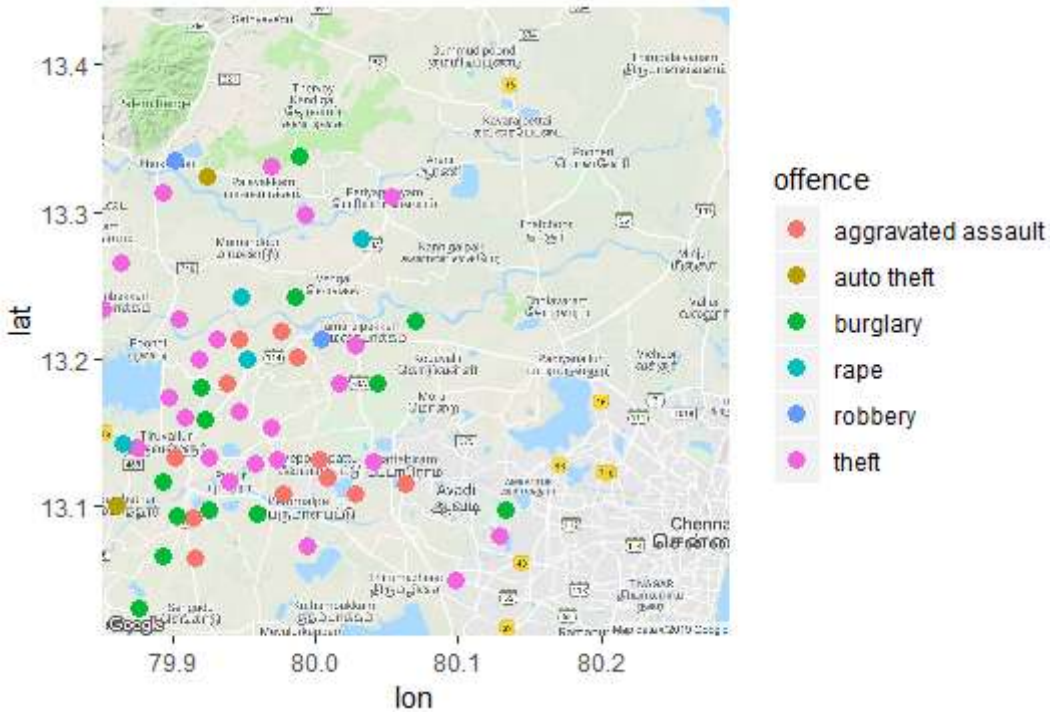

Fig.4. Spatial distribution of each type of crime on Google Map

V. CHOROPLETH MAP

Choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map to show the intensity of distribution of a particular variable. Fig. 5 shows the crime locations in shapefile and Fig. 6 shows the Choropleth map for crime intensity generated using R commands.

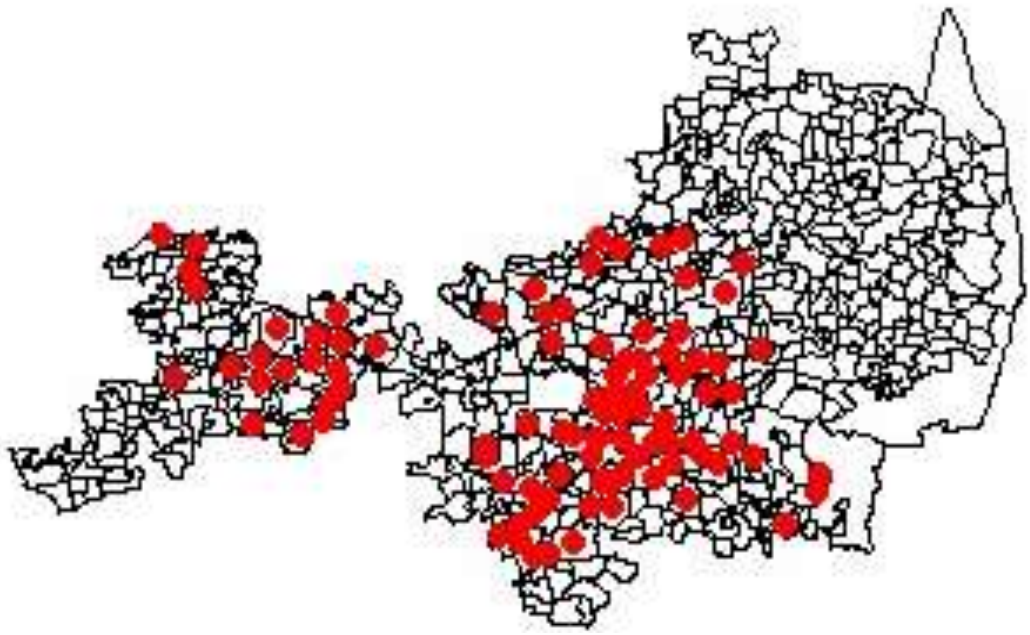

Fig. 5. Shapefile with crime locations

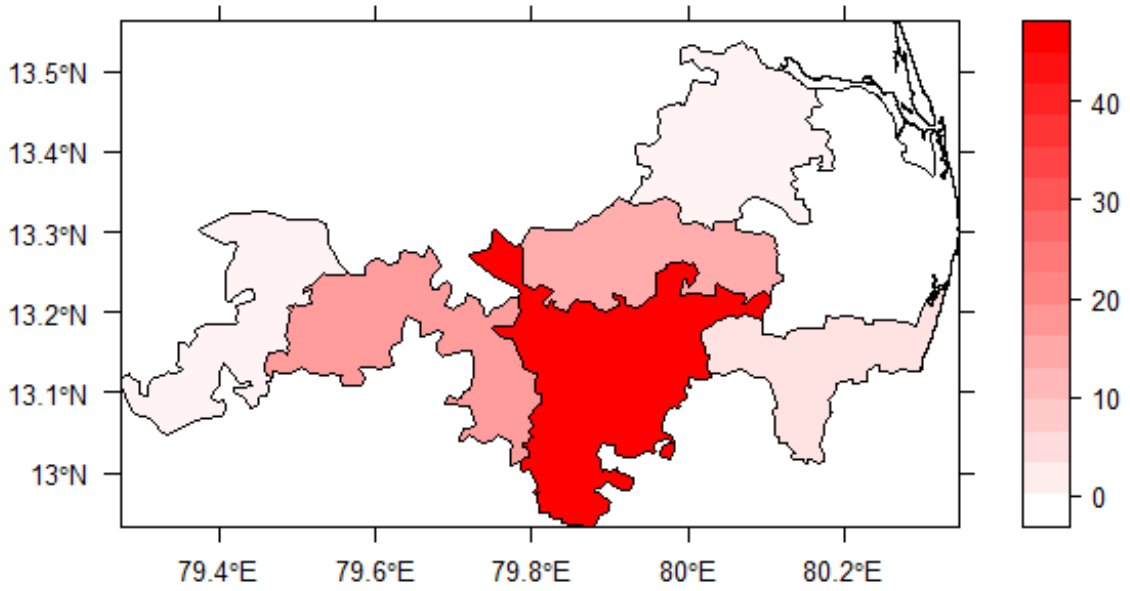

Fig. 6. Choropleth map for crime intensity

VI. CONTOUR HEATMAPS

A contour heatmap is another way to demonstrate data density, especially when choropleth map over a shape file (like the country's provinces) is not desired. The contour heatmap uses two-dimensional kernel density estimation in order to create the contoured heatmap of the spatial point data. The contour heatmap of crime data on Google map generated by R Software is shown in Fig.7.

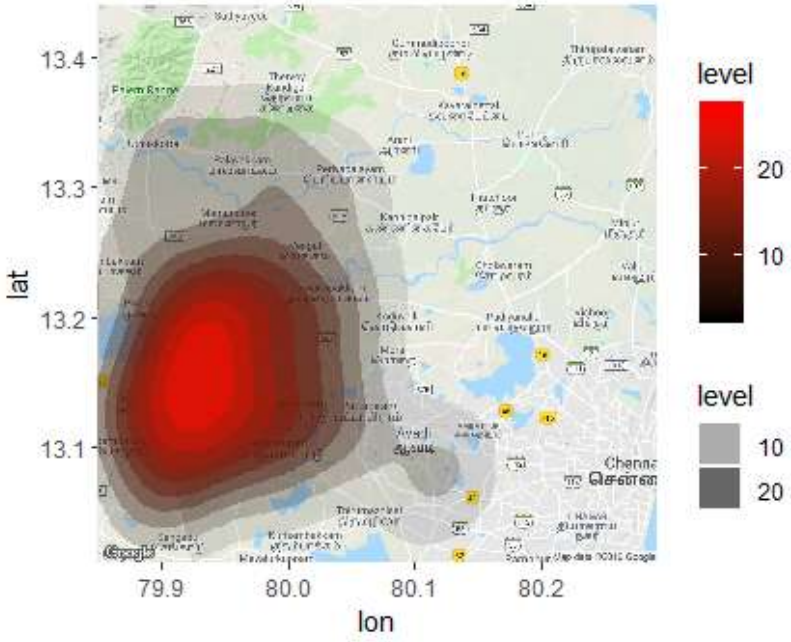

Figure 7 Contour Heatmap of crime data on Google Map

VII. INTERACTIVE MAPS

To make a map, you need the information of the shape of the area that interests you: we call it a spatial object. Several options exist to acquire this info. Leaflet and **ggmap** will provide a background coming from google. A few R libraries provide the information for common boundaries like world countries. For studying a specific area, we need to find the geospatial information somewhere on the web, under the **Shapefile** or the **GeoJSON** format. Once we get this shape information, we use **ggplot2** for drawing, or leaflet for interactive representation. Note that dozens of libraries exist to make map with R, here in this chapter just a curated selection of the best performing tools were explored.

An interactive map is an image with hotspots. A hotspot is a location on the map that responds when the mouse moves over it, off of it, or click it. Interactive mapping involves using maps that allow zooming in and out, panning around, identifying specific features, querying underlying data such as by topic or a specific indicator (e.g., socioeconomic status), generating reports and other means of using or visualizing select information in the map. Such Interactive maps can be obtained using R through Libraries like **sp**, **sf** and **ggplot2**. Fig. 8 shows the interaction map of Crime occurrence based on block ID.

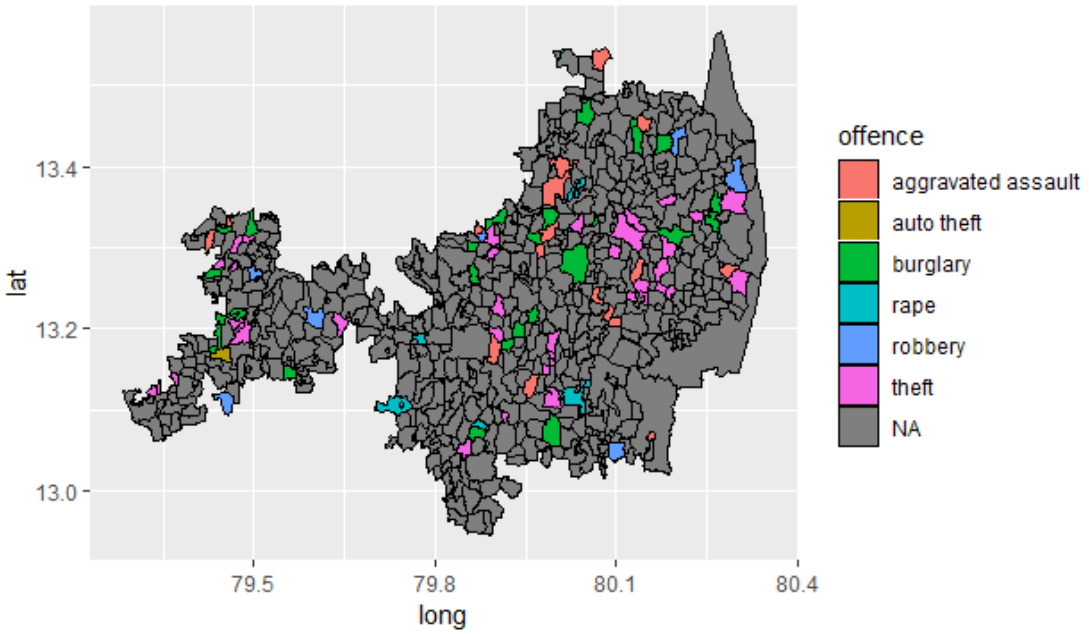

Fig. 8. Identification of Crime occurrence based on block ID

VIII. VISUALISATION OF KML IN GOOGLE EARTH

Keyhole Mark-up Language (KML) is an XML language focused on geographic visualization. There is a growing interest in using Google Earth to visualize spatio-temporal data produced in the R environment for statistical computing. As KML provides a diversity of visualization options, it is one of the more attractive platforms for scientific visualization of geographical phenomena as it can enable scientists to detect pattern in their data not visible in other software. R allows to write geospatial data to a KML file. Fig. 9 shows the KML visualisation of crime data generated by R Software.

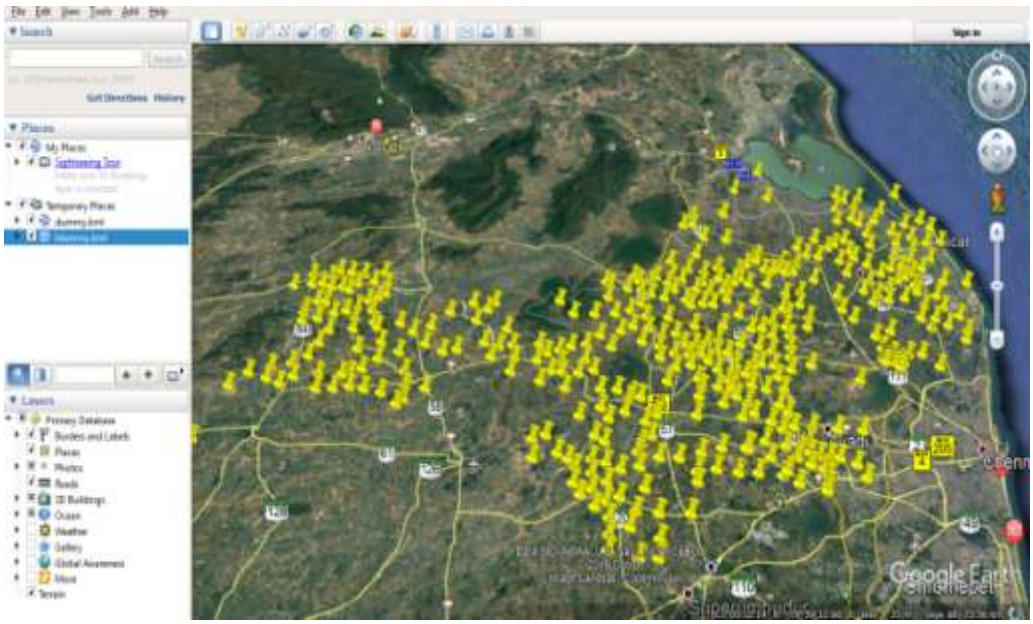

Fig. 9. KML Visualization of crime data on Google Earth

IX. TIME WISE PLOTS

Through the use of R's summary function and the ggplot2 library, breaking down a large data set and looked for various insights in this ggplot2 in R was achieved. Bar chart representing types of crimes with frequency and time wise distribution of crime occurrence with respect to day, week and month in the study area were acquired using R and shown in Fig. 10, Fig. 11 and Fig. 12. These figures are showing the distribution of all the crimes without any classification. Similar plots can be prepared for the individual crimes also showing temporal variation.

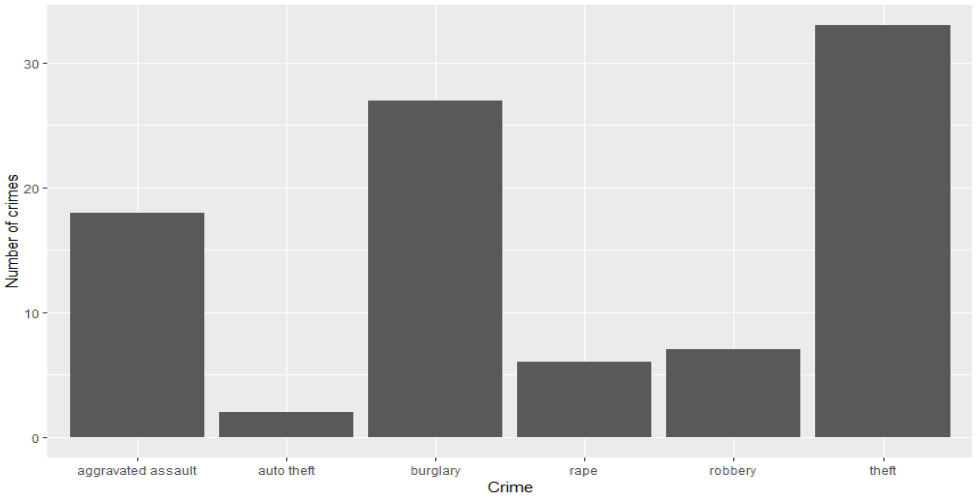

Fig. 10. Distribution of crime occurrence in the study area

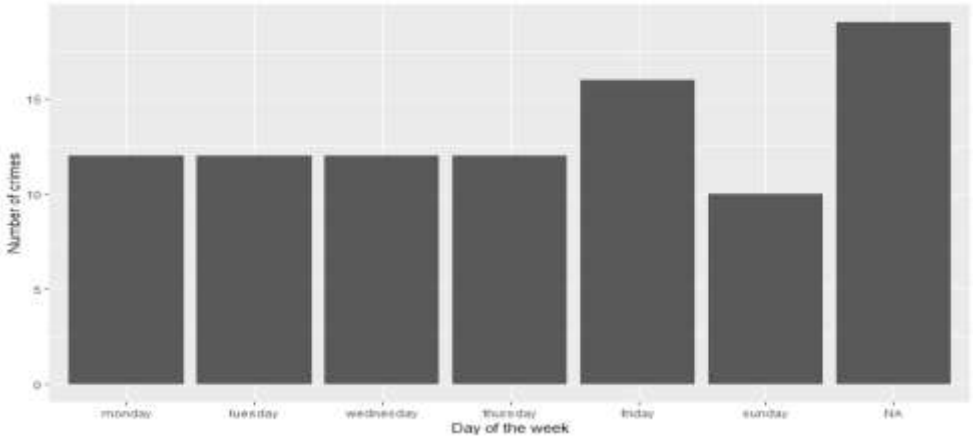

Fig. 11. Day wise distribution of crime occurrence

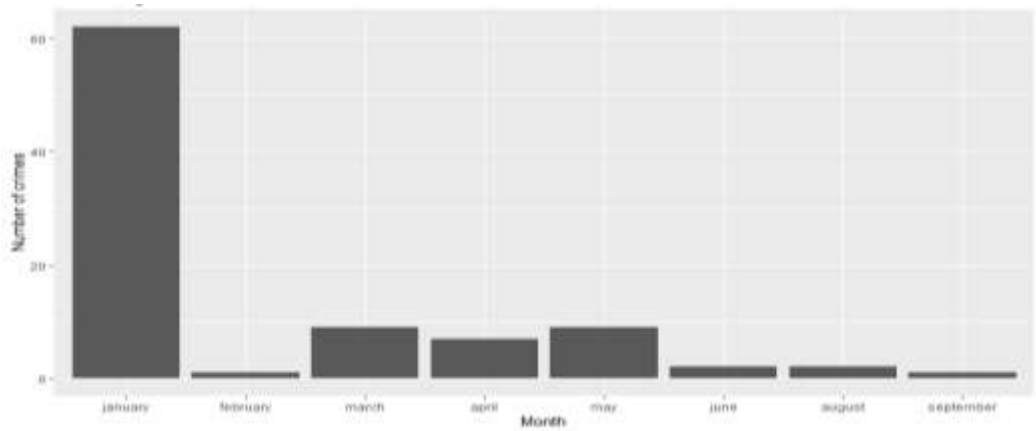

Fig. 12. Month wise distribution of crime occurrence

X. CONCLUSIONS

In this study, a variety of exploratory spatial analysis techniques were performed in R-Software for the analysis and visualization of spatial distribution of crime data. The data handling and interaction was found to be easy and there are multiple choices available in R software for the analysis and visualization of crime data. It can be concluded that better clarity in the analysis and visualization of crime data in customized manner is possible with 'R' Software.

REFERENCES

[1]. S. Shekhar, H. Xiong, and X. Zhou, "Crime Mapping and Analysis by R.E. Wilson, and K.M. Filbert," in Encyclopedia of GIS, Springer, Cham, 2017.
[2]. M. Fischer, and P. Nijkamp, "Exploratory Spatial Data Analysis by J. Symanzik," in Handbook of Regional Science, Springer, Berlin, Heidelberg, 2014.
[3]. N. Andrienko, and G. Andrienko, Exploratory analysis of spatial and temporal data: A systematic approach, New York: Springer, 2005.
[4]. M. Fischer, and J. Wang, Spatial data analysis: models, methods and techniques, Springer, Berlin/Heidelberg/New York, 2011.
[5]. N. Andrienko, G. Andrienko, A. Savinov, H. Voss and D. Wettschereck, "Exploratory Analysis of Spatial Data Using Interactive Maps and Data Mining," Cartography and Geographic Information Science, 28(3), 2001,151-166.
[6]. L. Anselin, I. Syabri, and Y. Kho "GeoDa: an introduction to spatial data analysis," Geographical Analysis, 38(1), 2006, 5-22.
[7]. M. Fischer, and A. Getis, "Exploratory spatial data analysis by R.S. Bivand," Handbook of applied spatial analysis: software tools, methods and applications. Springer, Berlin/Heidelberg, 2010, 219-254.
[8]. R.S. Bivand, P. Edzer, and G.R. Virgilio, Applied Spatial Data Analysis with R, Springer, 2013.
[9]. R Development Core Team, "R: a language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2011. [Online]. Available: http://www.R-project.org/ [Accessed: 22 January 2019]
[10]. A. Hirschfield, P. Brown and P. Todd "GIS and the analysis of spatially-referenced crime data: Experiences in Merseyside, U.K.," International Journal of Geographical Information Systems, 9(2), 1994, 191-210.
[11]. E. Anderson and D. Joaquin, "Using process control chart techniques to analyse crime rates in Houston, Texas," Journal of the Operational Research Society, 47(7), 1996, 871-881.
[12]. D.E. Brown, "The Regional Crime Analysis Program (RECAP): A Framework for to Catch Criminals," Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 3, 1998, 2848-2853.
[13]. S.F Messner, L. Anselin, R. D. Baller, D.F. Hawkins, G. Deane, and S.E. Tolnay, "The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis," Journal of Quantitative Criminology, 15(4), 1999, 423-424.
[14]. A.T. Murray, M. Ingrid, S.W. John and P. Mullins, "Exploratory Spatial Data Analysis Techniques for Examining Urban Crime: Implications for Evaluating Treatment," The British Journal of Criminology,41(2), 2001, 309-329.
[15]. J. Ratcliffe, "Crime Mapping and the Training Needs of Law Enforcement," European Journal on Criminal Policy and Research, 10, Kluwer Academic Publishers, Netherlands, 2004, 65-83.
[16]. H. Grubesic, T. Murray, and Alan "Detecting Hot spots using cluster analysis and GIS," Proceedings from the Fifth Annual International Crime Mapping Research Conference, 26, 2001.
[17]. T. Laurent, A.R. Gazen, and C.T.Agnan, "GeoXp: an R package for interactive exploratory spatial data analysis," Journal of Statistical Software, 47(2), 2012.
[18]. K. Kelly, K. Lai, and P.J. Wu, "Using R for Data Analysis-A Best Practice for Research," Book of Best Advanced Practices in Quantitative methods, Sage publications, 34, 2008, 535-570.
[19]. J. Miller, L. W. Kennedy and J. M. Caplan, "Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting," Journal of Justice quarterly, 28(2), 2011, 360-381.
[20]. A. Dyga, and M. Sławioska "Application of R in crime data analysis", Kielce University of Technology, Poland, 2015.
[21]. Y. Zhao, and Y. Cen, "Crime Analyses Using R by A. Sengupta, M. Kumar and S. Upadhyay," Data Mining Applications with R, Academic Press, ch.13, 2014, 367-395.

Murugasan. R;E" Geospatial Analysis and visualization of Crime Data using 'R' Open source Software" International Refereed Journal of Engineering and Science (IRJES), vol. 08, no. 03, 2019, pp 01-08