

# CRIME DATA INVESTIGATION AND VISUALIZATION USING R

**1st Author**

**Swetha. G**

**KMIT, MTech - CSE**

**gangapuramswetha@gmail.com**

## ABSTRACT

This project explores Crime data of a location during a time period dynamically allowing the user to help in visualizing the density of Crime Map using R and Shiny Given Location and first month of data collected and length of analysis period, the graphs come up within a second giving an insight to the investigators from 3 dimensions:

-Crime Data: shows the original crime data records downloaded from the data.police.uk.

-Crime Map: density plot of the crime data.

-Trends: bar charts of crime records over time in different crime categories.

## General Terms

Data Analysis, Statistical Computing, Data visualization, data mining.

## Keywords

Keywords R, Shiny, Crime data, data analysis

## 1. INTRODUCTION

Understanding the temporal development and patterns of criminal networks is important for law enforcement and intelligence agencies to investigate and prevent crimes. Extracting and visualizing criminal networks from a large amount of crime data has been a challenge over the past years. In particular, the visualization of the dynamic development of such networks over time has been difficult in many ways. Recent advancement of data analysis provides new analytical reasoning tools to explore and analyze a large amount of data with interactive visual interfaces. By employing the ideas of data analysis, we propose here a framework to visualize criminal networks using the R environment called as CrimeMap. The interactive and visual features of *Crimemap* can be useful in discovering and analyzing both relational and temporal patterns of criminal networks.

**Crime mapping** is used by analysts in law enforcement agencies map, visualize, and analyze crime incident patterns. It is a key component of Crime Analysis and the CompStat policing strategy. Mapping crime, using Geographic Information Systems (GIS), allows crime analysts to identify crime hot spots, along with other trends and patterns.

**Crime Analysts** use **Crime mapping and analysis** to help law enforcement management (e.g. the police

chief) to make better decisions, target resources, and formulate strategies, as well as for tactical analysis (e.g. crime forecasting).

*Crimemap* helps investigators to explore the cooffending relationships between criminals in a social network context. In order to test the functionalities of *this project*, we have used the crime database from the data.police.uk as test data. This project explores Crime data of a location during a time period dynamically allowing the user to help in visualizing the density of Crime Map using R and Shiny.

## 2. EXISTING SYSTEM

With the advancement of technology used by law enforcement and intelligence agencies, there is a critical need for new and fundamental understanding of the structure and dynamics of criminal networks.

The present system does not give the deep insights for the visualization of crime data. The major issue in the existing system is that it will take longer time to analyze and visualize the data.

## 3. PROPOSED SYSTEM

In the proposed system Crime data of a location is explored during a time period dynamically allowing the user to help in visualizing the density of Crime Map using R and the Shiny package in R.

Given Location and first month of data collected and length of analysis period, the graphs come up within a second giving an insight to the investigators from 3 dimensions:

- Crime Data : shows the original crime data records downloaded from the data.police.uk
- Crime Map : density plot of the crime data.
- Trends: bar charts of crime records over time in different crime categories.

Crime Maps are still more customized to an extent of drawing statistics according to

- Crime type and
- Crime Category

## 4. TECHNOLOGIES/ PLATFORMS USED

### 4.1 Introduction to R

R is a programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years.

R is an implementation of the S programming language combined with lexical scoping semantics inspired by Scheme. S was created by John Chambers while at Bell Labs. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team, of which Chambers is a member. R is named partly after the first names of the first two R authors and partly as a play on the name of S.

R is a GNU project. The source code for the R software environment is written primarily in C, Fortran, and R. R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems. R uses a command line interface; there are also several graphical front-ends for it.

R is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in R are essentially ephemeral, written for a single piece of data analysis.

### 4.2 RStudio

RStudio is a free and open source integrated development environment (IDE) for R, a programming language for statistical computing and graphics.

RStudio is available in two editions: RStudio Desktop, where the program is run locally as a regular desktop application; and RStudio Server, which allows accessing RStudio using a web browser while it is running

RStudio is written in the C++ programming language and uses the Qt framework for its graphical user interface.

## 5. IMPLEMENTATION

### 5.1 R and Shiny Package Implementation

#### 5.1.1 Shiny Package

Shiny makes it super simple for R users to turn analyses into interactive web applications that anyone can use.

Let your users choose input parameters using friendly controls like sliders, drop-downs, and text fields. Easily incorporate any number of outputs like plots, tables, and summaries.

#### 5.1.2 Loading data from Police API

The street-level crime data is one of the 9,000 datasets available from [data.gov.uk](http://data.gov.uk). The data can be downloaded systematically via the Police API.

#### 5.1.3 Wrapper R function

Wrapper function takes a location within England and Wales, downloads crime data around that location over a certain period of time and creates crime weather plots.

#### 5.1.4 Data

The street-level crime data is one of the 9,000 datasets available from [data.gov.uk](http://data.gov.uk). The data can be downloaded systematically via the Police API. The latest version of the API no longer requires authentication.

The following URL can be used to obtain crime records at street-level within a one-mile radius of a single point. The parameters required are **latitude**, **longitude** and **month**. The downloaded data is in JSON format which can be converted into R's data format using the `{RJSONIO}` package.

**Example URL:** <http://data.police.uk/api/crimes-street/all-crime?lat=52.629729&lng=-1.131592&date=2012-04>

#### 5.1.5 Methodology

The methodology can be summarised in the following six steps:

1. Obtaining latitude and longitude of a user-defined location using `ggmap::geocode`.
2. Downloading crime data via the Police API as discussed above.
3. Converting JSON into a list and then a data frame.
4. Downloading a base map from Google using `ggmap::get_googlemap`.
5. Coverting the base map into a ggplot object using `ggmap::ggmap`.
6. Add multiple layers on top of the base map using the data frame like a normal ggplot.

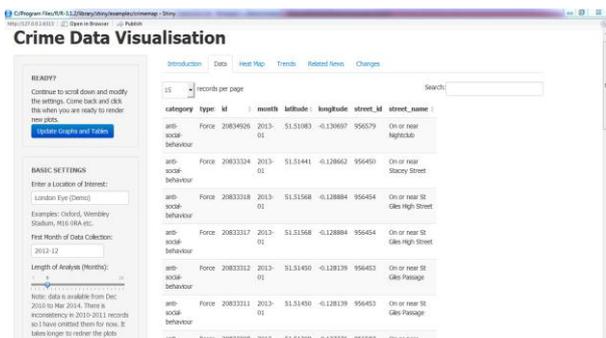


Fig 1: The Data tab shows the original crime data records downloaded from the data.police.uk

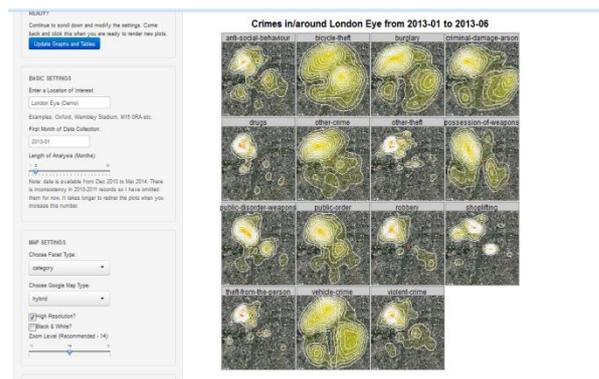


Fig 4: Crimes by categories in London from Jan-2013 to Apr-2013 on a hybrid map



Fig 2: All crimes around London Eye from Jan-2013 to Apr-2013

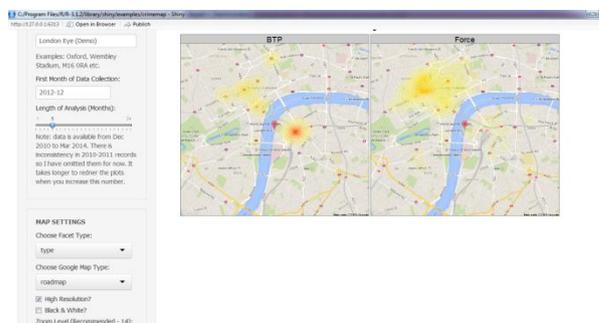


Fig 5: Typical crimes and traffic incidents around London Eye from Jan-2013 to Apr-2013

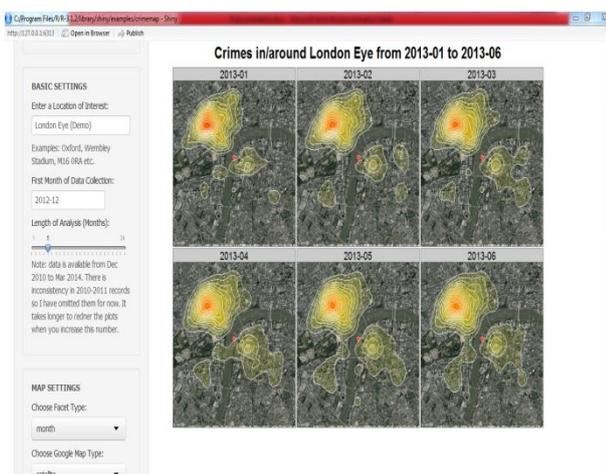


Fig 3: Monthly crimes for the year 2013 on a satellite map



Fig 6: Trends in London from Jan to June

## 6. CONCLUSION

In this project, we propose a framework to visualize criminal networks using the R environment, called as CrimeMap using the R shiny package. This project makes it super simple for R users and others to turn analyses into interactive web applications that anyone can use. The interactive and visual features of Crimemap can be useful in discovering and analyzing both relational and temporal patterns of criminal networks.

Crime data of a location is explored during a time period dynamically allowing the user to help in visualising the density of Crime Map using the the controls, Map and trends. Crime Maps are still more customized to an extent of drawing statistics according to Crime type and Crime Category.

## 7. REFERENCES

- [1] Ihaka, Ross (1998). R : Past and Future History. Interface98 (Technical report). Statistics Department, The University of Auckland, Auckland, New Zealand.
- [2] Fox, John and Andersen, Robert (January 2005). "Using the R Statistical Computing Environment to Teach Social Statistics Courses" (PDF). Department of Sociology, McMaster University. Retrieved 2006-08-03.
- [3] Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.
- [4] Vance, Ashlee (2009-01-06). "Data Analysts Captivated by R's Power". New York Times. Retrieved 2009-04-28. R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca...
- [5] David Smith (2012); R Tops Data Mining Software Poll, Java Developers Journal, May 31, 2012.
- [6] Karl Rexer, Heather Allen, & Paul Gearan (2011); 2011 Data Miner Survey Summary, presented at Predictive Analytics World, Oct. 2011.
- [7] Robert A. Muenchen (2012). "The Popularity of Data Analysis Software".
- [8] Tippmann, Sylvia (1 Jan 2015). "Programming tools: Adventures with R". *Nature* (517): 109–110. doi:10.1038/517109a.
- [9] Morandat, Frances; Hill, Brandon (2012). "Evaluating the design of the R language: objects and functions for data analysis". *ECOOP'12 Proceedings of the 26th European conference on Object-Oriented Programming*.
- [10] Vance, Ashlee (2009-01-07). "Data Analysts Are Mesmerized by the Power of Program R: [Business/Financial Desk]". *The New York Times*.
- [11] Vance, Ashlee (2009-01-08). "R You Ready for R?". *The New York Times*.
- [12] Wickham, Hadley. "Style Guide". Retrieved 14 June 2012.
- [13] Bengtsson, Henrik (January 2009). "R Coding Conventions (RCC) – a draft". Retrieved 14 June 2012.
- [14] Tippmann, Sylvia (29 December 2014). "Programming tools: Adventures with R". *Nature* (517): 109–110. doi:10.1038/517109a.
- [15] Morandat, Frances; Hill, Brandon (2012). "Evaluating the design of the R language: objects and functions for data analysis" (PDF). *ECOOP'12 Proceedings of the 26th European conference on Object-Oriented Programming*.
- [16] Clark, Mike (October 2007). "Introduction to SPSS 16". University of North Texas Research and Statistical Support. Retrieved 19 September 2013.
- [17] StatSoft (n.d.). "Using the R Language Platform". StatSoft Inc. Retrieved 20 September 2013.
- [18] Parmar, Onkar (31 March 2011). "'R' integrated with Symphony". Platform Computing Corporation. Retrieved 20 September 2013.
- [19] SAS (11 November 2010). "Calling Functions in the R Language (SAS/IML)". Retrieved 20 September 2013.
- [20] Tableau (17 December 2013). "R is Here!". Retrieved 29 January 2015.