

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337414019>

# Crime Analysis Using R

Article · December 2017

CITATIONS

0

READS

957

2 authors, including:



[R Madana Mohana](#)

Chaitanya Bharathi Institute of Technology

45 PUBLICATIONS 89 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



"Intelligent Fish Stock Prediction" using Image Processing, Machine Learning and Deep Learning. [View project](#)



R Software Online Course [View project](#)

# CRIME ANALYSIS USING R

GOUNI AKILA

M.Tech Scholar, Computer Science and Engineering  
Bharat Institute of Engineering and Technology  
Hyderabad, Email: [akilagouni94@gmail.com](mailto:akilagouni94@gmail.com)

Dr. R. MADANA MOHANA

Associate Professor, Department of CSE  
Bharat Institute of Engineering and Technology  
Hyderabad, Email: [madanmohanr@biet.ac.in](mailto:madanmohanr@biet.ac.in)

**Abstract:** - In this paper investigates the Crime information of a day and age progressively enabling the client to help in picturing the thickness of Crime Map utilizing R and month of information gathered and length of examination period, the charts come up inside a moment giving knowledge to the agents. Understanding the fleeting advancement and examples of criminal systems is essential for law authorization and insight offices to research and anticipate violations. Removing and envisioning criminal systems from a lot of wrongdoing information has been a test over the previous years. Specifically, the perception of the dynamic improvement of such systems after some time has been troublesome from multiple points of view. Late headway of information examination gives new scientific thinking devices to investigate and break down a lot of information with intelligent visual interfaces. By utilizing the thoughts of information investigation, we propose here a system to envision criminal systems utilizing the R condition called as CrimeMap.

**Keywords:** Crime Data, Crime Map, Trends.

## 1 Introduction

The intelligent and visual highlights of Crimemap can be valuable in finding and Wrongdoing Analysts utilize Crime mapping and examination to help law authorization administration (e.g. the police boss) to settle on better choices, target assets, and plan methodologies, and in addition for strategic examination (e.g. crime forecasting). Crimemap causes agents to investigate the co affronting in the proposed framework Crime information of year and month date and day and age progressively enabling the client to help in envisioning the thickness of Crime Map utilizing R. Given month of information gathered and length of examination period, the diagrams come up inside a moment giving an understanding to the specialists from 3 measurements:

*Crime Data:* shows the original crime data records downloaded from the [data.police.uk](http://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.

breaking down both social and transient examples of criminal systems.

More learning about particular range of a subject. Wrongdoing examination and aversion is an efficient approach for distinguishing and breaking down examples and patterns in wrongdoing. This framework can anticipate locales which have high likelihood for wrongdoing event and can imagine wrongdoing inclined ranges. With the expanding appearance of automated frameworks, wrongdoing information investigators can help the Law implementation officers to increase speed of the method to solve violations.

### Hindrance:

This product predicts wrongdoing apt areas in India on a particular day. Additionally different problem is that we are not predicting the time in which the wrongdoing is going on. Since time is a vital factor in wrongdoing we need to foresee the wrongdoing inclined locales as well as the best possible time. [1]

## 2 Related Work

Writing study is most imperative for comprehension and increasing substantially

### Favorable position:

Utilizing the idea of information mining we can separate already obscure, valuable data from

unstructured information. Rather than concentrating on reasons for wrongdoing event like criminal foundation of guilty party, political ill will and so forth we are concentrating chiefly on wrongdoing variables of every day. [1]

Through information preparing of enormous populace wrongdoing, K-implies mining calculation is connected to the general population security framework. When managing expansive measure of information, Kmeans calculation grouping comes about are incredibly impacted by the decision of introductory bunching focus. In the event that the choice of starting bunching focus is inappropriate, the grouping results might be caught in nearby ideal arrangement, and show signs of improvement bunching impact. In this way, versatile hereditary calculation is utilized to enhance K means calculation. The examination comes about demonstrate that the proposed plan can get the significant data and perfect bunching impact. Utilizing object metaphysics to characterize abnormal state ideas:

#### **Preferred standpoint:**

It can deal with information effectively synopsis systems are normally connected to the first information. [2], [5]

Iniquity against ladies nowadays has moved toward becoming a problem of every country over the world various nations are making effort to control the problem. Preventive measures are taken to reduce the widen number of instances of Iniquity against ladies. A gigantic measure of informational collection is produced each year on the premise of revealing of wrongdoing. This information can demonstrate extremely valuable in breaking down and anticipating wrongdoing and enable us to keep the wrongdoing to some degree. Wrongdoing examination is a territory of indispensable significance in police office. Investigation of wrongdoing information can enable us to break down wrongdoing design, between related clues& critical concealed relations between the violations. That is the reason information mining can be extraordinary guide to dissect, imagine and anticipate wrongdoing utilizing wrongdoing informational index.

#### **Hindrance:**

Wrongdoing can likewise be associated on the premise of age gathering, area of wrongdoing and kind of wrongdoing. Expectation of the wrongdoing will be shown utilizing different outlines pie diagrams, warm maps, spikes and charts.

#### **Favorable position:**

A perfect wrongdoing investigation device ought to have the capacity to distinguish wrongdoing designs rapidly and in a proficient way for future wrongdoing design identification and activity.[3], [6]

### **3 Methodology**

With the advancement of modernization applied by law implementation and knowledge offices, there is a fundamental requirement for new and basic comprehension of the structure and elements of criminal systems. The present framework does not give the intellectual experiences for the perception of wrongdoing information. The significant issue in the current framework is that it will require longer investment to examine and envision the information.

In the proposed framework Crime information of year and month date and day and age are progressively enabling the client to help in imagining the thickness of Crime Map utilizing R. Given month of information gathered and length of examination period, the charts come up inside a moment giving an understanding to the specialists from 3 measurements:

*Wrongdoing Data:* demonstrates the first wrongdoing data records downloaded from the data.police.uk Crime Map: thickness plot of the wrongdoing data.

*Patterns:* bar outlines of wrongdoing records over time in various wrongdoing classes. Wrongdoing Maps are still more modified to a degree of attracting measurements as indicated by Crime sort and Crime Category.

#### **3.1.1 Existing System:**

With the headway of innovation utilized by law implementation and knowledge organizations, there is a basic requirement for new and crucial

comprehension of the structure and elements of criminal systems. The present framework does not give the profound bits of knowledge for the representation of wrongdoing information. The significant issue in the current framework is that it will require longer investment to break down and picture the information.

### 3.1.2 Proposed System:

In the proposed framework the Crime information of year and month date and day and powerfully enabling the client to help in imagining the thickness of Crime Map utilizing R. Given month of information gathered and length of examination period, the diagrams come up inside a moment giving the knowledge to the specialists from 3 measurements:

*Wrongdoing Data:* demonstrates the first wrongdoing data records downloaded from the data.police.uk

*Wrongdoing Map:* thickness plot of the wrongdoing data.

*Patterns:* bar graphs of wrongdoing records over time in various wrongdoing classifications. Wrongdoing Maps are still more redid to a degree of attracting insights as per Crime sort and Crime Category.

#### 3.1.2.1 Architecture

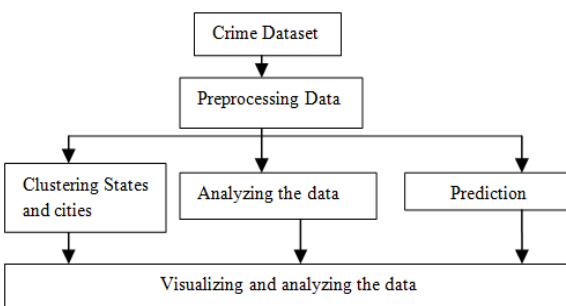


Fig. 1: Architecture for Crime Analysis Using R

#### 3.1.2.2 K-Means Clustering Algorithm

K-means Algorithm is a type of partitioning algorithm. Here the mean value of objects in a cluster represents a cluster. This algorithm is an iterative clustering algorithm where objects are moved among sets of clusters just before the chosen set is gained. The algorithm is constructed based on the idea of user stated input parameter (k). A set of 'n' objects are

partitioned into 'k' clusters. A high degree of comparison surrounded by the elements in clusters is obtained, while a high degree of dissimilarity among elements in different clusters is achieved simultaneously. The cluster's centroid gives the measure of cluster's similarity.

*Input:* Number of chosen clusters are denoted by k and A data set containing 'n' objects is denoted by 'D'.

*Step 1:* At first select 'k' objects randomly from D, as initial cluster centers.

*Step2:* Based on the distance between the objects and cluster means, each remaining object is assigned to the cluster to which it is most similar or near.

*Step 3:* Calculate new mean value of the object for each cluster.

*Output:* A set consisting of k clusters.

#### 3.1.2.3 Tree Based Algorithm

Tree based learning computations are thought to be exactly important and for the most part applied handled researching methods. Tree based approaches use perceptive models with high correctness, strength and integrity of accepting. Different to direct models, they blueprint aligned connections great. They are talented at taking care of any sort of problem nearby.

Methods like choice trees, arbitrary woods; tendency boosting is in issue immensely handled as a part of an immense range of info science problems. Therefore, for each proficient, it's necessary to take in these computations and apply them for demonstrating.

This instructional exercise is engaged to allow beginners to take in tree based publishing opening with not outside help. Later the profitable completion of this informative exercise, one is depended upon to finish up definitely suitable at handling tree based computations and assemble prescient models. Choice tree is a kind of handled learning computation that is generally applied as a part of grouping problems. It works for both aligned out and fixed info and output factors. In this policy, we separate the commonality or check into at least two identical sets in view of most important plotter/differentiator in input aspects.

## 4 Results

### 4.1 Dataset:

The dataset used here is UK Crime dataset of year 2011-2012. The dataset consists of 106,084 crime records.

Crime_Type	Station_Name	Month	Year	Crime Count	Location_Type	Operator
1 Criminal damage and arson	Abbey Road	9	2012	1	STATION	Transport for London
2 Anti Social Behaviour	Abbey Road	12	2011	1	STATION	Transport for London
3 Public Disorder and Weapons	Abbey Road	2	2012	1	STATION	Transport for London
4 Public Disorder and Weapons	Abbey Road	5	2012	1	ON TRAIN	Transport for London
5 Robbery	Abbey Road	5	2012	1	ON TRAIN	Transport for London
6 Vehicle crime	Abbey Road	1	2012	1	STATION	Transport for London
7 Violent crime	Abbey Road	10	2012	1	ON TRAIN	Transport for London
8 All crime and ASB	Abbey Road	1	2012	1	STATION	Transport for London
9 All crime and ASB	Abbey Road	2	2012	1	STATION	Transport for London
10 All crime and ASB	Abbey Road	5	2012	2	ON TRAIN	Transport for London

Fig. 2: Dataset

### 4.2 Results of K-means Algorithm:

*Use of K-means Algorithm:* K-means furnishes logical sections of dataset contributed for clustering. It benefits adjacency to select data points to a particular cluster; here the basis of allocation is the slightest distance from the cluster midpoint. Unsuccessfully, the amount of data gained has not been met by the quickness of the algorithms. A number of tools located the decisions have been advised to advance the handling power of distinct algorithms. In this paper, we present a k-mean algorithm for clustering which accomplishes the data repetition appearing in the dataset.

*Input:* UK Crime data (2011-2012)

*Output:*

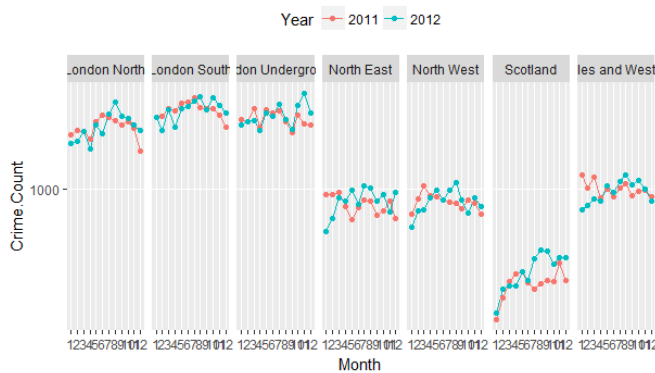


Fig. 3: Result of K-Means Algorithm

### 4.2.1 Crime Analysis based on Month and Area

*Input:* UK Crime data (2011-2012)

*Output:*



Fig. 4: Results of Crime Analysis based on Month and Area

TABLE 1  
CRIME ANALYSIS RESULTS BASED ON MONTH

	Crime_Type	dates	Crime.Count
1	Anti Social Behaviour	15005	801
2	Burglary	15005	29
3	Drugs	15005	299
4	Other Crime	15005	154
5	Other Theft	15005	1411
6	Public Disorder and Weapons	15005	653

## 4.2.2 Crime Analysis based on Crime Type

Input: UK Crime data (2011-2012)

TABLE 2  
TEST CASE OF CRIME ANALYSIS

S.No.	Test Case	Input	Expected Output	Actual Output	Status
1	Crime Type	UK Crime Data	Crime Type and Crime count will be displayed on the graph	Crime Type and Crime count will be displayed on the graph	Displayed

Output:

TABLE 3  
OUTPUT OF CRIME ANALYSIS

Crime_Type	Station_Name
OtherTheft : 20066	Leeds : 518
Anti Social Behaviour : 11152	KingsCrossSt Pancras: : 439
Public Disorder and Weapons: : 8814	Manchester Piccadilly : 353
Drugs : 3382	Victoria : 344
Other Crime : 2632	BirminghamN ew Street : 337
(Other) : 2649	Baker Street : 335
NA's : 57389	(Other) : 103758

TABLE 4  
OUTPUT OF CRIME ANALYSIS URLS

Operator	OperatorURL
Transport for London : 26944	http://www.tfl.gov.uk/ : 26944
South Eastern Trains : 9130	http://www.southeasternrailway.co.uk/ : 9130
South West Trains : 8745	http://www.southwesttrains.co.uk : 8745
Southern Railway : 6818	http://www.southernrailway.com : 6818
Northern Rail : 6611	http://www.northernrail.org/ : 6611
First Great Western : 6342	http://www.firstgreatwestern.co.uk/ : 6342
(Other) : 41494	: 41494

Min. : 1  
1st Qu.: 682  
Median : 1273  
Mean : 1296  
3rd Qu.: 1924  
Max. : 2510

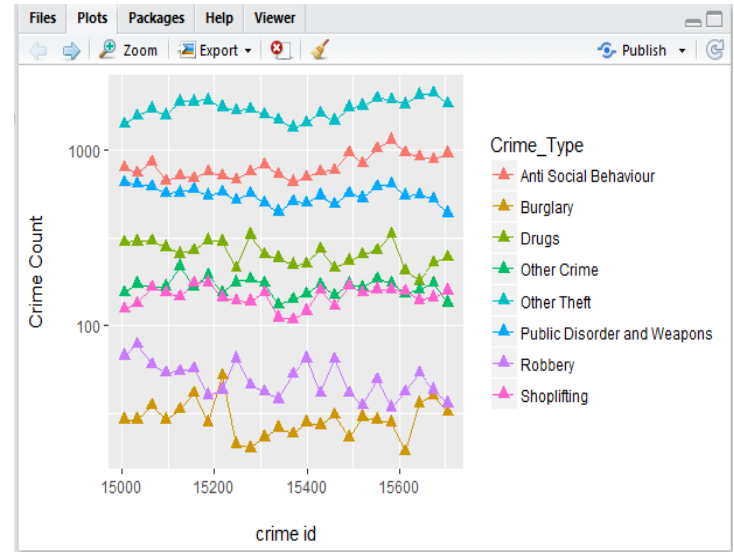


Fig. 5: Results of Crime Analysis based on Crime Type

## 4.2.3 Crime Analysis based on Location

TABLE 5  
OUTPUT OF CRIME ANALYSIS BASED ON LOCATION

S.No.	Test Case	Input	Expected Output	Actual Output	Status
1	Crime Analysis based on Location	UK Crime Data 2011-2012	Crime count will be displayed based on the Location	Crime count will be displayed based on the Location	Displayed

FAIRFIELD STREET, MANCHESTER, M1 2PB: 353 METROSTN: 391  
CORNHILL, EC3V 3LA: 346 STATIONS: 80146  
SMALLBROOK QUEESWAY, BIRMINGHAM, B2 4QA: 337 SUBWAY : 169  
MELTON STREET, EUSTON, NW1 2HS: 329 TRAMLINK: 1580



YORK ROAD, WATERLOO, SE1 8SW: 312  
(Other) :54564

Count.Prop

Min. :0.0003984

1st Qu.:0.0006418

Median :0.0012208

Mean :0.0053058

3rd Qu.:0.0028369

Max. :3.0000000

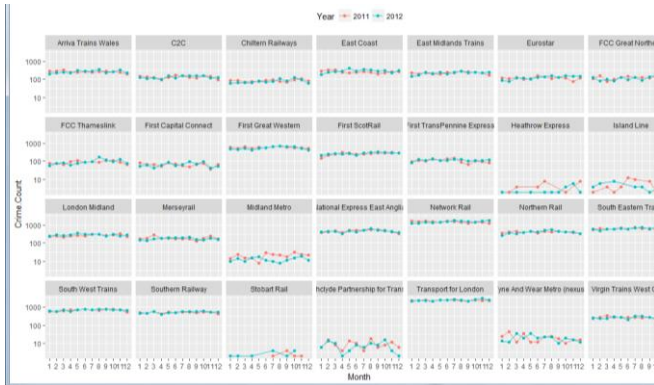


Fig. 6: Results of Crime Analysis based on Location Type

### 4.3 Result Analysis

Utilizing K-means determination tools can help analyzers accordingly to detect the familiar elements between distinct data sources and facts in numerous graphical reports, and the frizzy matching or physical matching method can be used to organize the connection between the distinct entities. Finally, it organizes connection and proof chain between distinct data linked to all cases, so we can discover additional information and unknown information. It can be seen that K-means has higher exactness and constancy.

TABLE 6  
COMPARISON OF RESULTS

Algorithm	Accuracy (%)	
	2011	2012
Decision Tree Algorithm	100	100
	70	80
	65	70
K Means Algorithm	100	100
	92	95
	75	80
Tree Based Algorithm	100	100
	90	97
	70	75

## 5 Conclusion

In this paper Crime Analysis is accomplished based on crime type, month, year and location. Visualization and Prediction are perfected so that the future crimes can be predicted before and precautions can be captured so that slightly we can end the crimes. Furthermore, for future research, we propose to expand this study to enhance the crime prediction methods in order to resolve the limitation of current methods to obtain more accurate results and good performance.

## References:

- [1] Shiju Sathyadevan, Devan M.S and Surya Gangadharan.S "Crime Analysis and Prediction Using Data Mining "978-1-4799-3486-71 14/\$31.00©201 4 IEEE
- [2] Teng Li "Criminal Behavior Analysis Method based on Data Mining Technology"2016 International Conference on Smart City and Systems Engineering 978-1-5090-5530-2/16 2016, DOI 10.1109/ICSCSE.2016.153
- [3] Tushar Sonawanev,Shirin Shaikh, Shaista Shaikh, Rahul Shinde, Asif Sayyad"Crime Pattern Analysis, Visualization And Prediction Using Data Mining" .
- [4] Vance, Ashlee (2009-01-06). "Data Analysts Captivated by R's Power".
- [5] David Smith (2012);" R Tops Data Mining Software Poll, Java Developers Journal, May 31, 2012".
- [6] Data Mining Concepts and Techniques Third Edition Jiawei Han University of Illinois at Urbana–Champaign Micheline Kamber Jian Pei Simon Fraser University.
- [7] Yanchang Zhao, R and Data Mining: Examples and Case Studies, Published by Elsevier, October 20, 2015.