iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6050

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

A Statistical Approach to Analyse and Visualize Cybercrimes

S. Venkata Krishna[1], T. Mahendra Reddy[2], R. Narendra Kumar[3], P. Praveen Kumar[4], Syed Jani Basha[5]
[1, 2, 3, 4]*UG Students,* [5]*Assistant Professor, Kallam Haranadhareddy Institute of Technology, Guntur*

Abstract: ***Cybercrime incidents are computer related crimes which occur around the world and are a serious threat to the world. Due to cybercrime incidents, lots of people facing security issues. In relevant works, there are different methodologies to identify existing cybercrimes and some are put under unfounded if they are not in the list of methodologies. Due to this, most of the new cybercrime records are neglected and were not recoded. The proposed work analyse all cybercrime elements which were recorded and put them in a schema. This schema facilitates an opportunity to combine various cybercrime characteristics and defining them in a two-level classification enables for a better understanding and will be easy for analysis and visualization. This work presents a statistical analysis with sampling data can be visualized in graphical representations where user can analyse the crime easily. This has been implemented using R tool.***

Keywords: ***cybercrime, analyse, visualize***

I. INTRODUCTION

Cybercrime includes a blend of various ordinary violations with new illicit acts. Singular cybercrime episodes are events of specific criminal offenses and, as numerous national wrongdoing measurements and overviews illustrate, are consistently expanding. As indicated by the Federal Bureau of Investigation, the Internet Complaint Center got 269422 grumblings of Internet wrongdoing in 2014, which shows an ascent of 1600% in contrast with the 16838 whines remembered for the underlying report [1, 2]. What might be compared to 117339 assaults for each day. Correspondingly, the German Crime Statistics showed a 23.6% expansion in the quantity of cybercrime occurrence from 2007 to 2008. This paper means to contribute toward better understanding cybercrime by proposing a diagram based cybercrime episode depiction that: 1) identifies the highlights of a cybercrime occurrence and their potential components and 2) gives a two-level offense classification framework dependent on specific rules. The proposed diagram can be stretched out with a rundown of suggested activities, comparing measures and viable strategies that neutralize the offense type and accordingly the specific episode. This coordinating will empower better observing, dealing with, and directing the different cybercrime offenses and their manifestation as specific incidents [3]. This paper presents the related work in Section II, issues distinguished in the Section III, proposed approach is introduced in Section IV, Architecture for proposed approach in Section V, Decision Tree calculation to anticipate phishing vulnerabilities is introduced in Section V, Implementation and Deployment is introduced in Section VI, Data Visualization is talked about in Section VII with Concluding comments.

II. RELATED WORK

The various understandings of what cybercrime involves alongside non-orderly classification of the relating offenses and absence of suggested activities are not contributing toward overseeing and coordinating successful orders, arrangements, and authoritative activities at nearby, national, or global level and result in inadequate treatment of cybercrime incidents[4,5,6].

Table 1: Identified Features Of Cybercrime Incidents

No.	Feature	Feature Description	Answers the question
1	INCIDENT	Description of the incident	What happened?
2	IDENTIFIED OFFENCE	Criminal offence that occurred	Is it considered criminal activity? Which one?
3	OFFENDER	Individual or entity that is responsible for the incident	Who is responsible?
4	ACCESS VIOLATION	Computer/network violation approach	How it occurred?
5	TARGET	Values that are the desired target	What was targeted?
6	VICTIM	Individual or entity that has suffered	Who has suffered?
7	HARM	The cause harm	What was the harm induced?
8	ACTIONS, MEASURES & POLICIES	Recommendations for the particular incident	What can be done to tackle and prevent it?

©IJRASET: All Rights are Reserved

Gordon et.al.[7] have proposed a typology comprised of two classes. Type I offenses portray solitary or discrete occasions encouraged by the presentation of malware projects, for example, keystroke lumberjacks, infections, and rootkits. Type II offenses are encouraged by programs that are not classified as wrongdoing product, and they are commonly rehashed contacts or occasions from the point of view of the client. Ian Walden et.al.[8] gave a through legitimate assessment of the considerable and procedural guidelines identifying with PC wrongdoing in the entirety of its indications. He offers an away from of the applicable specialized parts of cybercrime examination and Combines itemized investigation with broad case referencing and genuine models from training.

III. ISSUES IDENTIFIED

Various Issues identified from the above literature are as follows:

A. Implementing algorithms in java required good programming knowledge. Data mining applies filter mechanism in the data implementation and analysing.
B. While not explicitly stated, the literature implies that imitation attacks are superior anonymity strategies to obfuscation attacks.
C. No perfect visualisation.
D. Parallel processing is not possible in clustering Analysis
E. Mining techniques does not work well with the big data in analysis when we occurred with multiple column analysis.

Out of those identified issues, it is analysed that the Programming constraints and Visualization have been carried as a proposed work. In this regard, a statistical approach has been proposed.

IV. PROPOSED APPROACH

The need of the proposed work has been discussed in the earlier section with various issues identified. So, the proposed work includes an approach which will get a proper solution for the identified issues. The approach includes the following steps.

Step 1: Gather the dataset from City of Denver Country in the CSV (Comma Separated Format). The dataset will be available in many formats like .xml, .xlsx etc. Here, the CSV file is used for data processing.

Step 2: Store / save the data in the computer which has an Operating System Interface.

Step 3: Download and install R for Windows 3.6.3, RStudio 1.3.959-1 and latest R tools.

Step 4: Buffer the CSV file into RStudio. If there is successful buffer, then Run it by clicking on RUN button

Step 5: After successful compilation, generated graphs are visualized in the console.

V. ARCHITECTURE FOR PROPOSED APPROACH

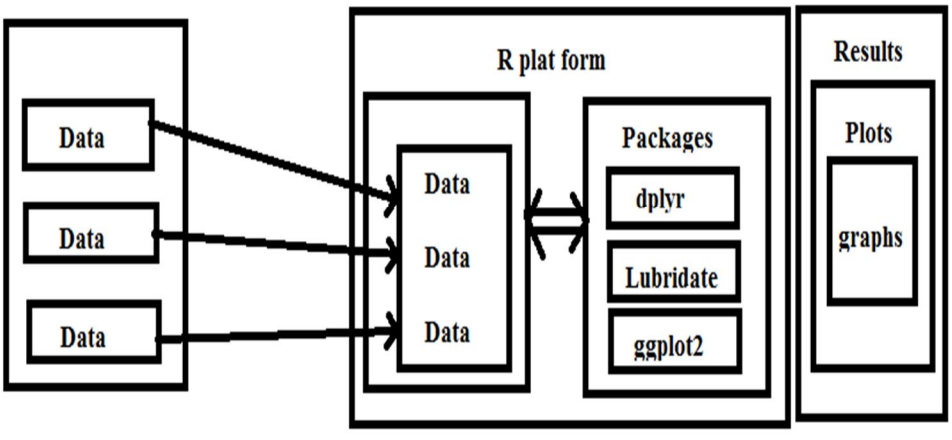


Figure 1: Proposed Architecture

The data is collected from different databases. The data which is collected processed into R platform. First, the date is loaded into the RStudio v1.3.959-1 software. Later all the respective packages will be loaded. Various visualization results includes Graphs, Plots, Results may be analysed.

©IJRASET: All Rights are Reserved

VI. IMPLEMENTATION AND DEPLOYMENT

Inorder to visualize the respective Denver dataset, the key commands needs to be implemented to analyze the data.The following R tool packages needs to be executed for the visualization of data as follows:

A. Dplyr

It is a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges:
mutate() adds new variables that are functions of existing variables
select() picks variables based on their names.
filter() picks cases based on their values.
summarise() reduces multiple values down to a single summary.
arrange() changes the ordering of the rows.

B. Lubridate

Getting R to agree that your data contains the dates and times you think it does can be tricky. Lubridate simplifies that. Identify the order in which the year, month, and day appears in your dates. Now arrange “y”, “m”, and “d” in the same order. This is the name of the function in lubridate that will parse your dates.

C. ggplot2

ggplot2 is a data visualization package for the statistical programming language R. Created by Hadley Wickham in 2005, ggplot2 is an implementation of Leland Wilkinson's Grammar of Graphics—a general scheme for data visualization which breaks up graphs into semantic components such as scales and layers.

VII. DATA VISUALIZATION

The Denver dataset for the various years such as 2015 to 2019 is implemented and deployed for execution then the respective data formats are visualized as follows from Figure 2 to Figure 8.

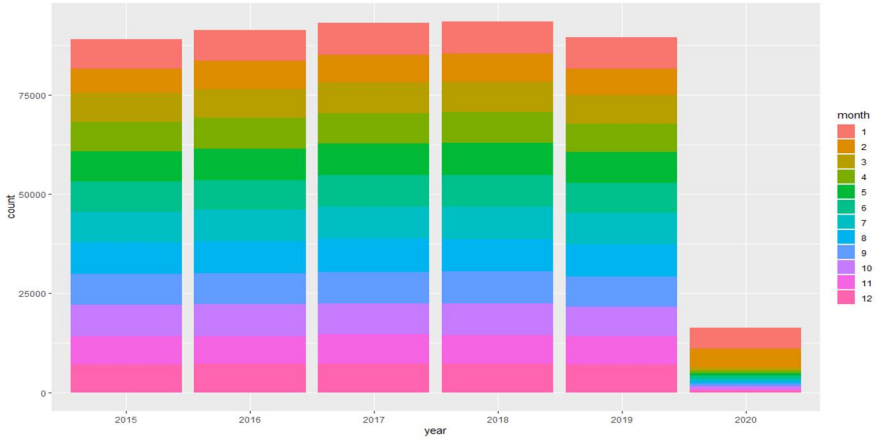


Figure 2: Bar Graph of month wise crimes from 2015 to 2020

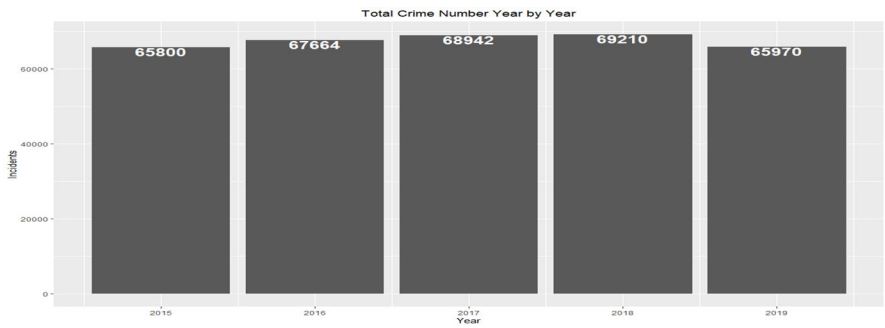


Figure 3: Bar Graph for Total Crime Number year by year from 2015 to 2019

©IJRASET: All Rights are Reserved

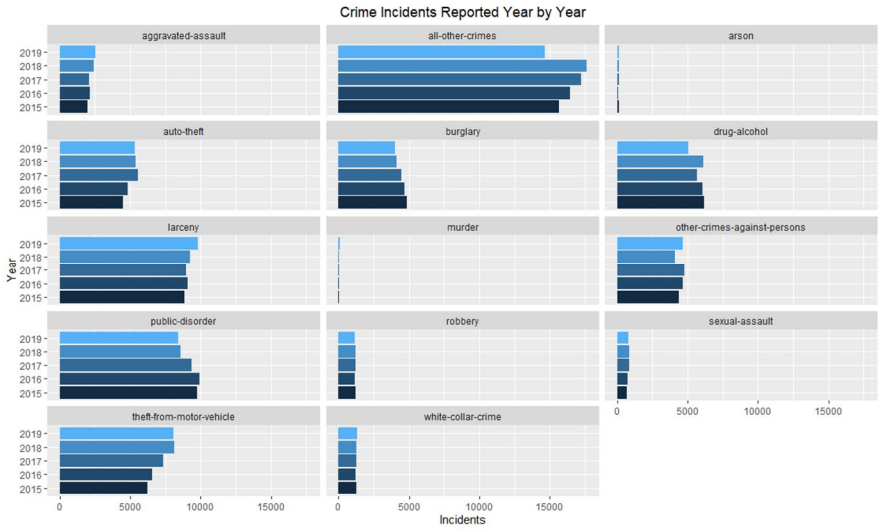


Figure 4: Bar Graph for Crime incidents reported year by year

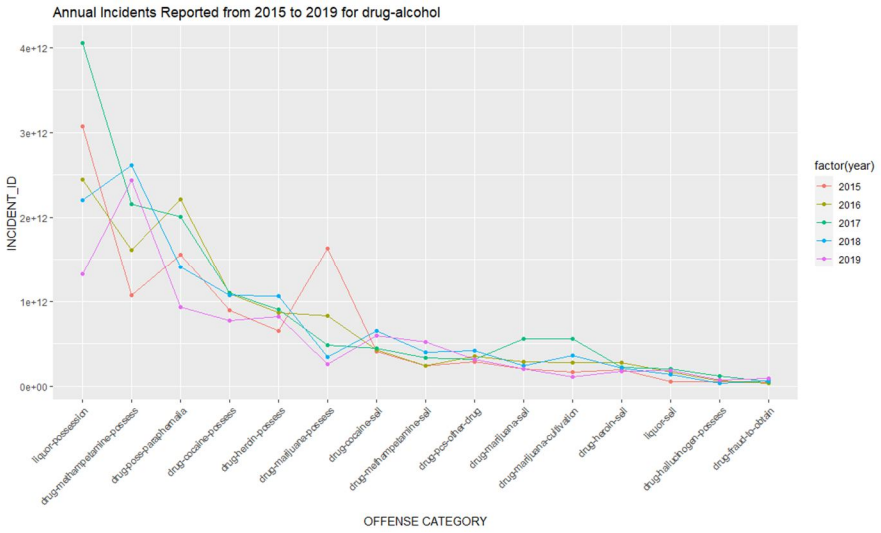


Figure 5: Annual incidents reported from 2015 to 2019 for drug-alcohol

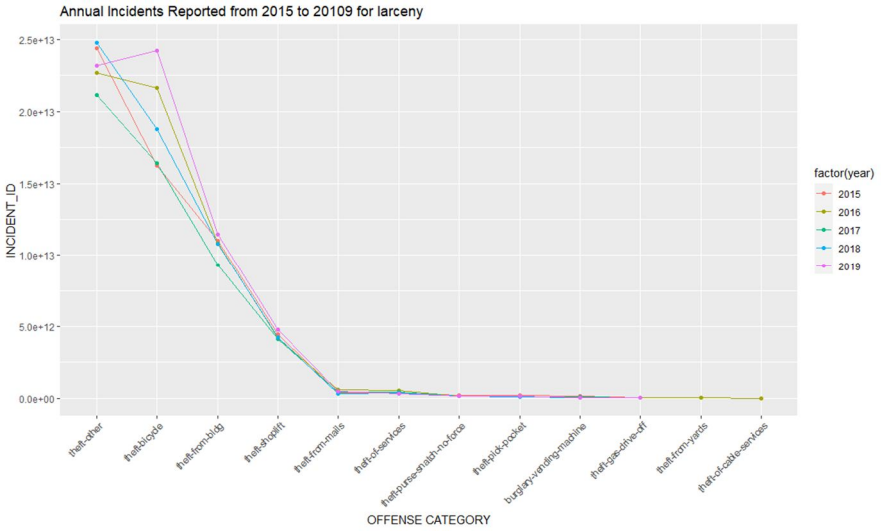


Figure 6: Annual incidents reported from 2015 to 2019 for larceny

©IJRASET: All Rights are Reserved

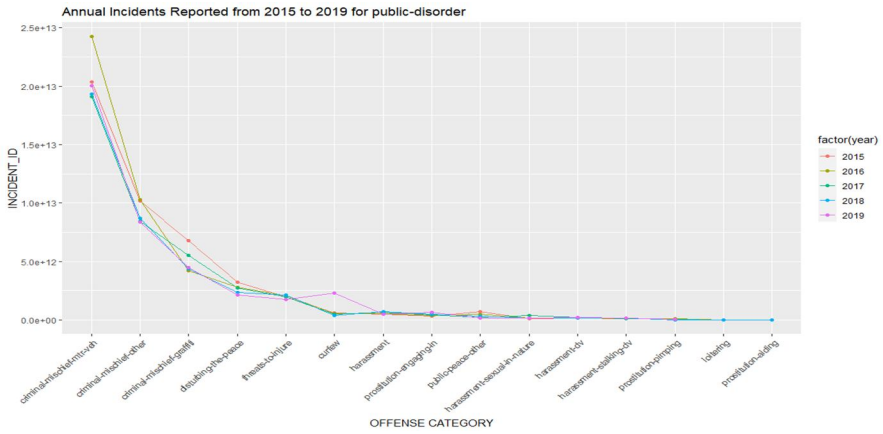


Figure 7: Annual incidents reported from 2015 to 2019 for public-disorder

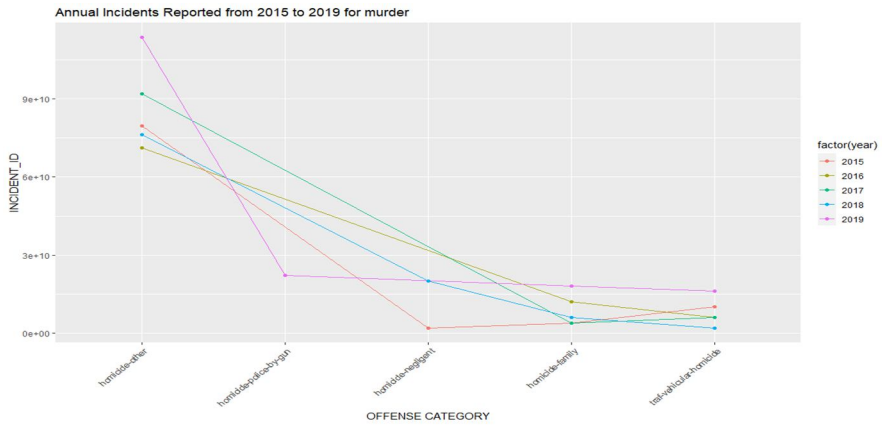


Figure 8: Annual incidents reported from 2015 to 2019 for murder

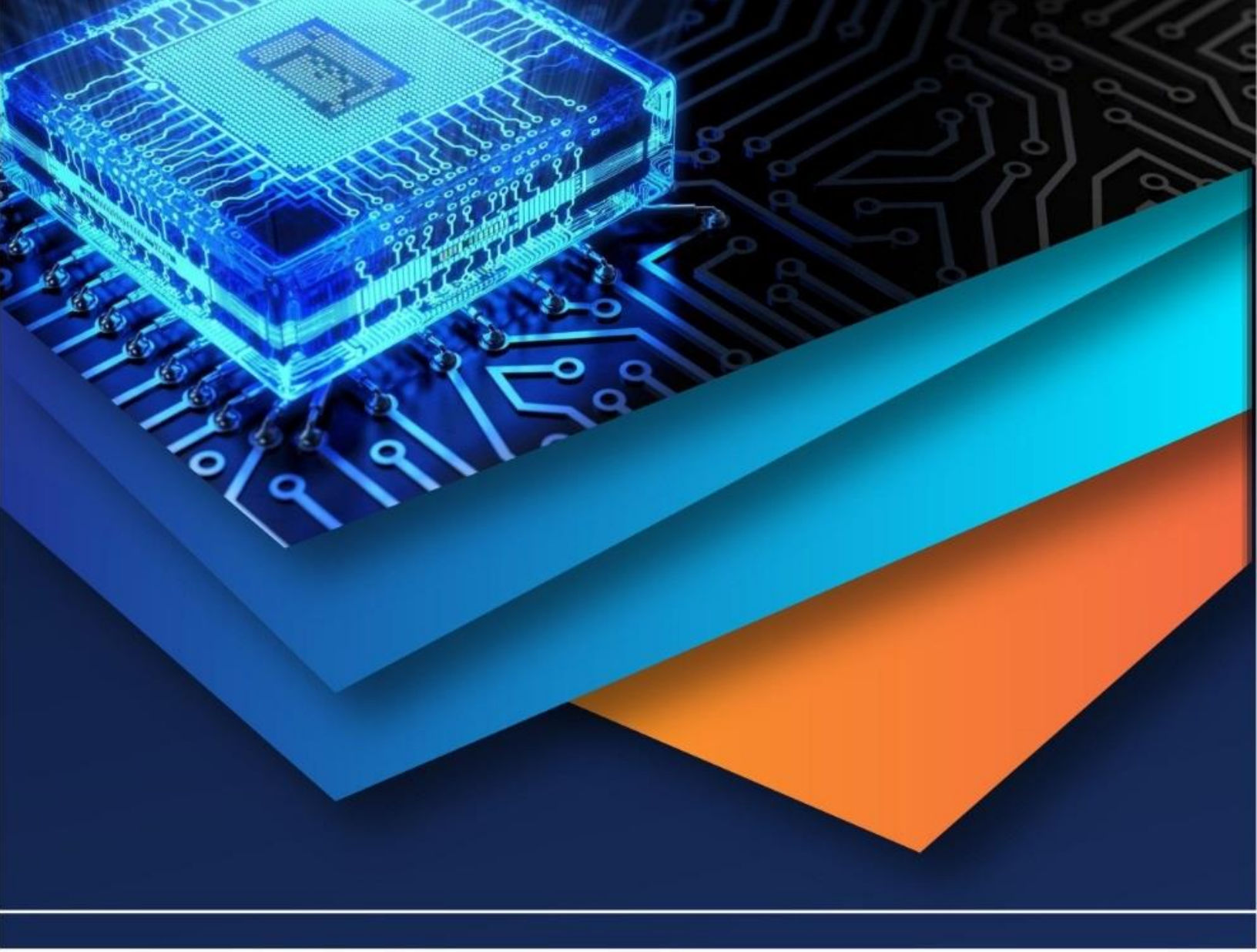
VIII. CONCLUSION

The need of cybercrimes investigation and the information perception is talked about in an adequate way. A methodology has been proposed to break down and envision the information of cybercrimes occurring in and around by considering the Denver dataset for test investigation. A similar methodology might be executed on any dataset for additional examination reason. The R instruments perform well and present the better representation by considering different R bundles. The outcomes introduced give away from of the Denver dataset. By dissecting the diagrams, robbery and open issue shows higher crime percentage from the taken years and every other-wrongdoing likewise has higher crime percentage. So the Government needs to take greater security cautions upon these wrongdoings so as to forestall in future years. For open issue, more guidelines ought to be purchased and increment fines if individuals defy the standards and guidelines.

IX. REFERENCES

[1] D. L. Shinder and M. Cross, Scene of the Cybercrime. Burlington, MA, USA: Syngress, 2008.
[2] FBI and NW3C. (May 22, 2015). 2014 Internet Crime Report. Accessed on May 17, 2016. [Online]. Available: https://pdf.ic3.gov/ 2014_IC3Report.pdf
[3] The Global State of Information Security Survey, PwC, London, U.K., 2016, accessed on Apr. 21, 2016. [Online]. Available: http:// www.pwc.com/gx/en/issues/cyber-security/information-security-survey/ download.html
[4] PricewaterhouseCoopers. (Sep. 30, 2014). The Global State of Information SecuritySurvey 2015—Managing Cyber Risks in an Interconnected World. Accessed on May 19, 2016. [Online]. Available: http://www.pwccn.com/webmedia/doc/635527689739110925_rcs_info_security_2015.pdf
[5] K. J. Higgins. (Feb. 10, 2015). Nation-State Cyber Espionage, Targeted Attacks Becoming Global Norm, Dark Reading. Accessed on May 19, 2016. [Online]. Available: http://www.darkreading.com/attacksbreaches/nation-state-cyber-espionage-targeted-attacks-becomingglobal-norm/d/d-id/1319025
[6] 2015 US State of Cybercrime Survey, PwC, London, U.K., Jul. 2015, accessed on May 18, 2016. [Online]. Available: http://www.pwc.com/us/en/increasing-it-effectiveness/publications/uscybercrime-survey-2015.html
[7] S. Gordon and R. Ford, "On the definition and classification of cybercrime," J. Comput. Virol., vol. 2, no. 1, pp. 13–20, 2006.
[8] I. Walden, Computer Crimes and Digital Investigations. Oxford, U.K.: Oxford Univ. Press, 2007.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)