IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⫻ Peer-reviewed & Refereed journal ⫻ Vol. 11, Issue 8, August 2024

DOI: 10.17148/IARJSET.2024.11829

The Art of Data Visualization: Matching graphs to data

Syeda Neha[1], Dr Mohan Aradhya[2]

PG Student, Master of Computer Applications, Rashtreeya Vidyalaya College of Engineering, Bengaluru- 560059[1]

Associate Professor, Department of Master of Computer Applications, Rashtreeya Vidyalaya College of Engineering, Bengaluru- 560059[2]

Abstract: As data is being generated each and every time in the world, the importance of data mining and visualization will always be on increase. Mining helps to extract significant insight from large volume of data. After that we need to present that data in such a way so that it can be understood by everyone and for that visualization is used. Most common way to visualize data is chart and table. Visualization is playing important role in decision making process for industry. Visualization makes better utilization of human eyes to assist his brain so that datasets can be analyzed and visual presentation can be prepared. Visualization and Data Mining works as complement for each other. In this paper, we present which type of chart is suited for which type of data.

Keywords: Data integrity, Information Visualization, Scientific Visualization, Decision Making, Graph, Chart, Xmdv tool.

I. INTRODUCTION

In simple terms, visualization is the process of creating images to make data easily understandable for others. When combined with data mining and human-computer interaction, visualization enhances the analysis of visual data. Initially, visualization was categorized into two types: Information Visualization, which dealt with abstract and non-spatial data, and Scientific Visualization, which focused on scientific data with spatial components [12]. Today, visualization encounters challenges such as mapping, handling high dimensionality, and balancing design tradeoffs [13]. Visualization plays a crucial role in uncovering patterns, trends, and relationships within datasets. As David McCandless, an author, data journalist, and information designer, aptly puts it [32]: "By visualizing information, we turn it into a landscape that you can explore with your eyes, a sort of information map. And when you're lost in Information, an information map is kind of useful."Figure 1 illustrates the general steps involved in the visualization process. Firstly, data is collected from various sources. Next, potential patterns and meanings are aggregated. The data is then analyzed, followed by its graphical interpretation. Finally, users interact with these graphical representations.To achieve optimal visualizations for any given dataset, certain fundamental principles should be considered [27].

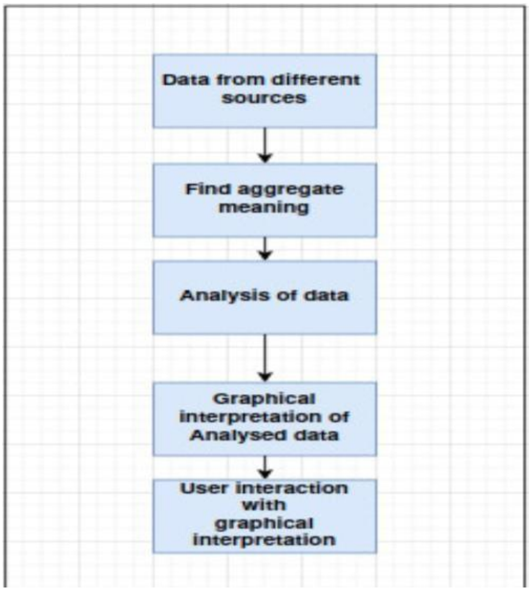

Fig.1. General steps in the process of Visualization.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 11, Issue 8, August 2024

DOI: 10.17148/IARJSET.2024.11829

a) Begin by comprehending the size and cardinality of the dataset provided. Assess the volume of data points, the number of variables or dimensions involved, and any specific characteristics such as temporal aspects or spatial dimensions.

b) Identify the specific type of information that needs to be communicated through visualization. This involves understanding the key insights, patterns, or relationships within the data that are most relevant to the audience.

c) Tailor the visual information processing to suit the targeted audience. Consider factors such as their domain knowledge, technical expertise, and the level of detail they require. Adjust the complexity and depth of the visualization accordingly to ensure it is meaningful and accessible.

d) Select the most effective and intuitive visual representation for the given dataset and audience. Choose visual forms that enhance understanding and facilitate interpretation, such as charts, graphs, maps, or diagrams. Ensure clarity, simplicity, and relevance in the visual portrayal to effectively convey the intended message.

II. DATA VISUALIZATION TECHNIQUE CATEGORIES

According to [2], data visualization techniques can be categorized into several classes:

a) **Describing Data:** This class focuses on providing an overview of the dataset itself. It includes techniques that summarize and present basic statistics, distributions, and characteristics of the data.

b) **Viewing Relationships:** Techniques in this class emphasize illustrating relationships between observations and variables within the dataset. This involves methods like scatter plots, network graphs, and correlation matrices to visualize connections and dependencies.

c) **Picturing Data (Icons, Glyphs, Color Coding):** This class involves mapping data items into easily recognizable shapes, icons, glyphs, or using color coding to convey information effectively. It aims to enhance visual understanding and facilitate comparisons.

d) **Temporal Visualization:** This class is dedicated to visualizing data that changes over time. Techniques here include time series plots, line graphs, calendars, and animations, which are suitable for tracking trends, patterns, and fluctuations across temporal dimensions.

e) **Spatial Visualization**: Techniques in this class are used to visualize spatial datasets that are associated with geographical or physical locations. Maps, cartograms, choropleth maps, and heatmaps are commonly used to represent spatial relationships and distributions of data.

f) **Spatio-Temporal Visualization:** This class combines spatial and temporal properties, focusing on datasets that evolve both geographically and over time. Examples include animations of weather patterns, tracking disease outbreaks, and analyzing movement data, where understanding both spatial and temporal aspects is critical for insight generation.

Each class of visualization technique serves specific purposes in data analysis, helping to uncover patterns, trends, and relationships that are essential for decision-making and understanding complex datasets.

III. STATE OF THE ART

Recent advancements in data visualization research highlight various innovative approaches and techniques:

1. **Ming C. Hao [14]:** Proposed "Visual Analysis of Multi-Attribute Data Using Pixel Matrix Displays," particularly beneficial for handling detailed information.

2. **Adam Perer [15]:** Advocated for the tight integration of statistical and visualization techniques to accelerate insight development.

3. **Zhao Kaidi [4]:** Introduced a novel algorithm for visualizing 4D data, addressing the challenges of representing data in four dimensions.

ISSN (O) 2393-8021, ISSN (P) 2394-1588
IARJSET
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 8, August 2024
DOI: 10.17148/IARJSET.2024.11829

4. **Martin Wattenberg [6]:** Introduced the "arc diagram" method for visualizing relationships, offering a clear and intuitive way to depict connections between entities.

5. **Ming C. Hoa [3]:** Developed two new techniques: "cell-based visual time series" and "visual content query" for enhanced visual analytics capabilities.

These contributions highlight the diverse applications of data visualization in various domains, from enhancing analytical capabilities to supporting decision-making processes and understanding complex datasets.

IV. METHODS OF VISUALIZATION

4.1 Arc Diagram
The arc diagram is primarily used to visualize complex data structures within strings such as text, music compositions, and compiled code. These structures often feature repeated substrings, which are advantageous for visualization as they serve as predictive units. Martin Wattenberg [6] described the arc diagram technique, which employs pattern matching algorithms to identify and visually represent these repeated substrings as translucent arcs. Arc diagrams find significant application in music to reveal composition structures [6], as well as in other domains like web pages, compiled codes, and DNA nucleotide sequences. Future advancements may leverage additional pattern matching algorithms to uncover deeper insights using this method.

4.2 Flow Maps
Flow maps visualize the movement or flow of processes, such as migration patterns between countries. Doantam Phan [9] introduced a method for generating flow maps using hierarchical clustering inspired by graph layout algorithms. Key characteristics of flow maps include intelligent distortion of positions, merging of edges with shared destinations, and intelligent edge routing. Phan's approach involves:

- Layout Adjustment
- Primary Hierarchical Clustering
- Rooted Hierarchical Clustering
- Spatial Layout
- Edge Routing
- Handling Multiple Layers

These steps ensure that flow maps effectively represent complex flow patterns in a clear and informative manner.

4.3 Graph Analytics
Graph analytics focuses on analyzing and visualizing graph structures to address real-world challenges across various applications. Success in graph analytics is measured not only by algorithmic criteria but also by practical application outcomes [11]. Common applications include:

- Electric Power Grid Analytics
- Social Network and Citation Analytics
- Text and Document Analytics
- Knowledge Domain Analytics

Graph analytics techniques aim to derive actionable insights from interconnected data points and structures.

4.4 Voronoi Treemaps
Voronoi Treemaps are effective for visualizing hierarchical and attributed data, overcoming the rectangular shape limitations of traditional Treemaps. Michael Balzer [37] introduced Treemaps based on subdivision into arbitrary polygons, allowing for flexible and efficient data representation. The process involves:

- Creating polygonal subdivisions according to the top hierarchy level
- Outputting a set of polygons representing nodes at each hierarchy level

This approach enhances the clarity and usability of Treemaps for complex datasets.

4.5 Geometric Projection
Geometric Projection is utilized for multidimensional and multivariate visualization, capable of mapping data onto Cartesian planes or arbitrary spaces. Key techniques include:
- Scatterplot Matrix: Displays joint variations of multiple data items using a matrix of scatter plots

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 11, Issue 8, August 2024

DOI: 10.17148/IARJSET.2024.11829

- Parallel Coordinates: Represents attributes with parallel vertical axes, facilitating correlation analysis among attributes.

These techniques are essential for detecting outliers, handling large datasets, and exploring relationships between variables.

4.6 **Pixel-Oriented Techniques**

Pixel-oriented techniques are employed in visualizing multivariate data, where each attribute is represented by a colored pixel. These techniques include:

- Recursive Patterns: Influence data arrangement through recursive processes, enhancing query independence
- Pixel Bar Charts: Directly represent data values without aggregation, suitable for high-dimensional datasets

Pixel-oriented approaches offer granularity and flexibility in representing complex data structures effectively.

Each of these visualization methods serves distinct purposes and is tailored to address specific challenges in understanding and interpreting diverse datasets across various domains.

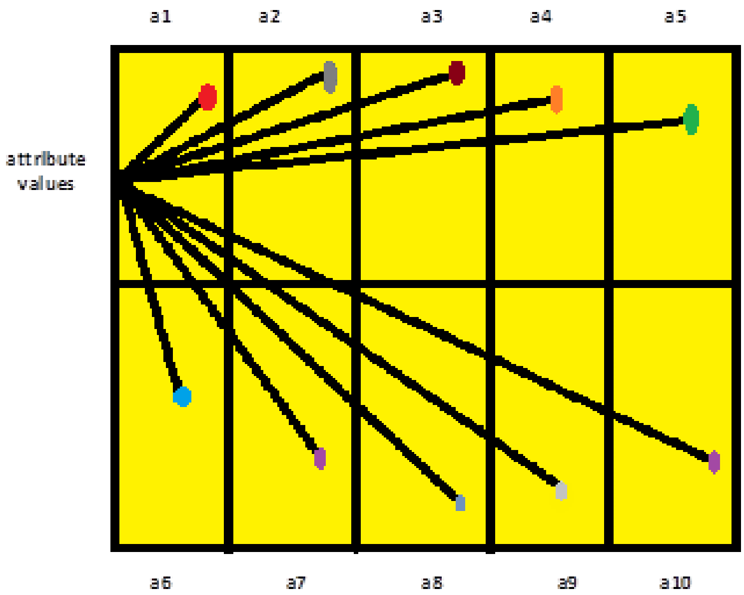

Fig. 2 Pixel Visualization if 10 Dimensional Data (a is attribute)

4.7 Hierarchical Display

Hierarchical display techniques focus on visualizing hierarchical data structures, where data spaces are subdivided into nested levels to reveal relationships and hierarchies.

Dimensional Stacking (General Logic Diagrams): This technique modifies hierarchical axis representation by separating the data space into stacked 2D subspaces [13].

TREEMAPS: TREEMAPS PARTITION THE SCREEN INTO RECTANGLES OR POLYGONS BASED ON THE HIERARCHICAL STRUCTURE OF THE DATA. EACH REGION'S SIZE REPRESENTS THE VALUE OF AN ATTRIBUTE, FACILITATING THE VISUALIZATION OF HIERARCHICAL DATA PARTITIONS [13]. THIS METHOD IS PARTICULARLY EFFECTIVE FOR DISPLAYING NESTED STRUCTURES AND HIERARCHICAL RELATIONSHIPS ACROSS DIFFERENT LEVELS OF DATA GRANULARITY.

These hierarchical display techniques are essential for organizing and visualizing complex hierarchical data in a manner that is intuitive and informative, aiding in better understanding and analysis of hierarchical relationships within datasets.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 8, August 2024

DOI: 10.17148/IARJSET.2024.11829

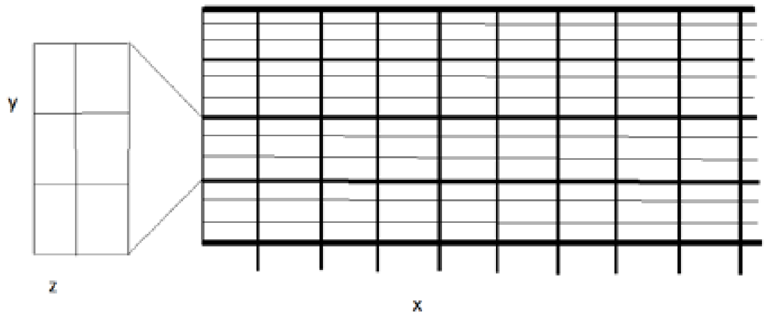

Fig. 3 Dimensional Stacking.

4.8 Iconography

Iconography techniques map multidimensional data items to visual icons, utilizing different visual features to represent varying data attribute values [13].

Chernoff Faces: One of the most popular techniques in iconography, Chernoff Faces visualize data items by mapping dimensional positions of a face (such as mouth, eyes, nose, etc.) to data properties [13]. Each feature of the face represents a different attribute, allowing for the simultaneous representation of multiple dimensions within a single visual icon.

Star Glyph: A variant of the glyph family, star glyphs are widely used in iconography to represent data items. They use star shapes where each point of the star encodes a different attribute. However, their effectiveness diminishes when the number of data items increases significantly. Star glyphs can also be enhanced by combining them with other glyphs to encode additional information [13].

Shape Coding: This technique uses very small arrays of pixels to visualize data, where each array represents one item of data. Shape coding is particularly useful for compactly representing data items in a visually distinct manner, allowing for easy comparison and analysis.

These iconography techniques provide versatile means to visually encode and interpret multidimensional data, offering insights into complex datasets through intuitive and informative visual representations.

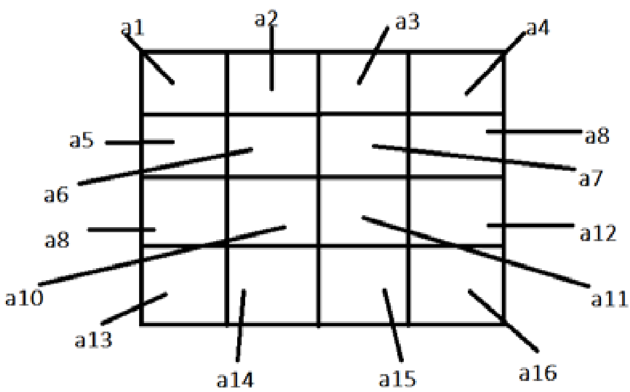

Fig. 4 Array for Shape Coding.(a1 to a16 are attributes)

4.9 Charts and Graphs

Charts and graphs are among the most common and widely used methods for visualizing information, offering straightforward visual representations that are easily understandable for audiences. Here are some key types:

Line Graphs (Line Charts): Line graphs depict the relationship between variables over a continuous interval. They are ideal for tracking trends and comparing multiple items over the same period of time [27]. Line graphs use lines to connect data points, emphasizing the change or trend in values over time.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 8, August 2024

DOI: 10.17148/IARJSET.2024.11829

Bar Charts: Bar charts compare quantities or values of different categories. Bars, which can be either vertical or horizontal, represent the values, making it easy to visually compare the heights or lengths of the bars to understand the differences between categories [27]. Different colors are often used to differentiate between bars when they are closely positioned or when multiple datasets are presented. Bar charts are effective when comparing discrete categories with varying ranges of values.

These types of charts and graphs are essential tools for visual communication, providing clear and concise representations of data that facilitate quick understanding and interpretation for a wide range of audiences.

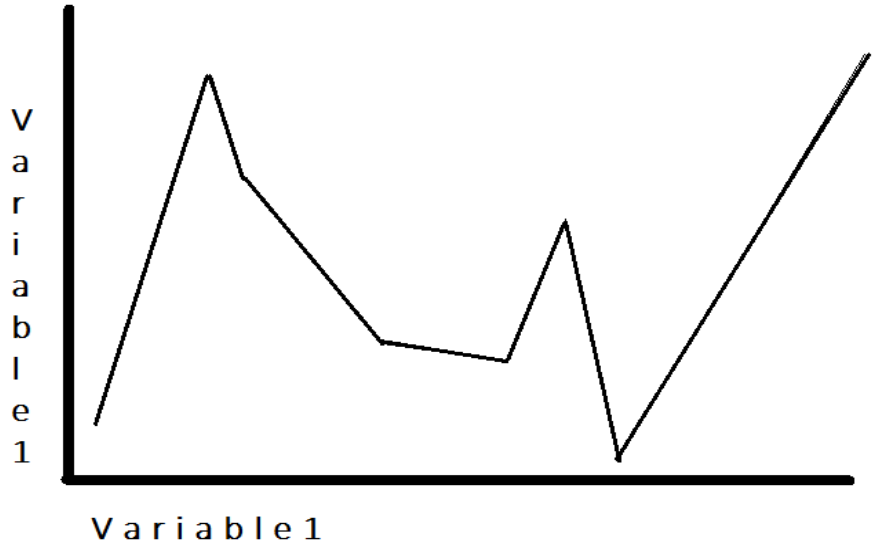

Fig. 5 Line Graph

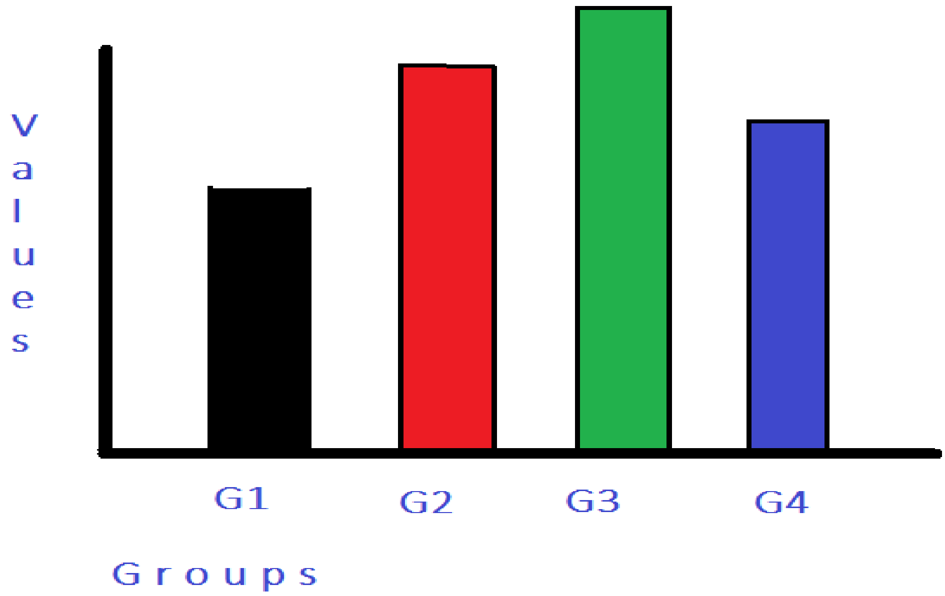

Fig. 6 Bar Chart (G1 to G4 are group)

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 8, August 2024

DOI: 10.17148/IARJSET.2024.11829

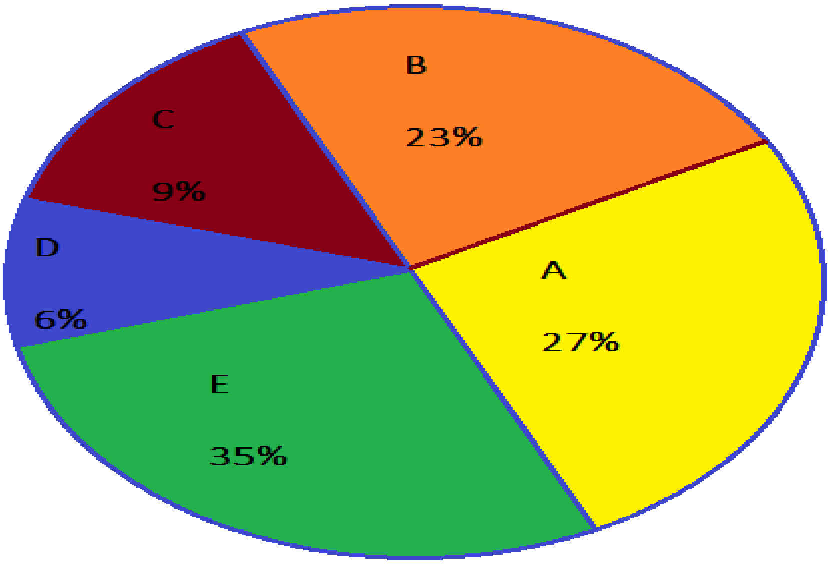

Fig. 7 Pie Chart

Pie Charts

Pie charts are a subject of discussion due to challenges in interpreting their angles and areas accurately by the human eye. While they can effectively convey proportions when accompanied by additional information such as percentages, they have specific considerations:

1. Interpretation Challenges: Pie charts rely on visual angles and areas to represent proportions. However, accurately estimating these angles and comparing areas can be difficult, particularly when the differences between segments are small.
2. Usefulness with Additional Information: When percentages or numerical values are provided alongside a pie chart, it enhances clarity and aids in understanding the exact proportions represented by each segment.
3. Screen Size Considerations: Pie charts may not be ideal for small screens or when designing dashboards for mobile devices. The intricate details of segment labels and the reliance on precise visual distinctions can be challenging to display effectively on smaller screens.

Despite these considerations, pie charts remain a popular choice for visualizing proportions and relative contributions within a dataset, especially when combined with supplementary information to improve comprehension and accuracy of interpretation.

ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of various researchers and professionals in the field of cloud security and digital payments. Special thanks to the developers and maintainers of the IEEE LaTeX style files, which facilitated the preparation of this paper.

REFERENCES

[1] Amirtharajan, Rengarajan and John Bosco Balaguru Rayappan. “Steganography-Time to Time: A Review.” *Research Journal of Information Technology* 5 (2013): 53-66.
[2] Abdul-wahab, Shams N., Mostafa Abdulghafoor Mohammed, and Omar A. Hammood. "Theoretical Background of steganography." *Mesopotamian Journal of CyberSecurity* 2021 (2021): 22-32.
[3] Mandal, Pratap Chandra, Imon Mukherjee, Goutam Paul, and B. N. Chatterji. "Digital image steganography: A literature survey." *Information sciences* 609 (2022): 1451-1488.
[4] Cheddad, Abbas, Joan Condell, Kevin Curran, and Paul Mc Kevitt. "A comparative analysis of steganographic tools." *In Proceedings of the Seventh IT&T Conference. Institute of Technology Blanchardstown, Dublin, Ireland. 25th*, pp. 29-37. 2007.

DOI: 10.17148/IARJSET.2024.11829

[5] Arora, Himanshu, Cheshta Bansal, and Sunny Dagar. "Comparative study of image steganography techniques." *In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN),* pp. 982-985. IEEE, 2018.

[6] Watni, Dipti, and Sonal Chawla. "A comparative evaluation of jpeg steganography." *In 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC)*, pp. 36-40. IEEE, 2019.

[7] Aslantaş, Funda, and Cemal Hanilçi. "Comparative Analysis of Audio Steganography Methods." *Journal of Innovative Science and Engineering* 6, no. 1 (2022): 122-137.

[8] Maiti, Anamitra, Subham Laha, Rishav Upadhaya, Soumyajit Biswas, Vikas Choudhary, Biplab Kar, Nikhil Kumar, and Jaydip Sen. "Boosting Digital Safeguards: Blending Cryptography and Steganography." *arXiv preprint arXiv:2404.05985* (2024).

[9] Strasser, Gerhard F. "The rise of cryptology in the European Renaissance." *In The History of Information Security*, pp. 277-325. Elsevier Science BV, 2007.

[10] Berezin, Casey-Tyler, Samuel Peccoud, Diptendu M. Kar, and Jean Peccoud. "Cryptographic approaches to authenticating synthetic DNA sequences." *Trends in Biotechnology* (2024).

[11] Butora, Jan, Yassine Yousfi, and Jessica Fridrich. "How to pretrain for steganalysis." *In Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security*, pp. 143-148. 2021.

[12] Pevný, Tomáš, and Andrew D. Ker. "The challenges of rich features in universal steganalysis." *In Media Watermarking, Security, and Forensics* 2013, vol. 8665, pp. 203-217. SPIE, 2013.

[13] Böhme, Rainer, and Rainer Böhme. "Principles of modern steganography and steganalysis." *Advanced Statistical Steganalysis* (2010): 11-77.

[14] Chutani, Shaveta, and Anjali Goyal. "A review of forensic approaches to digital image Steganalysis." *Multimedia Tools and Applications* 78, no. 13 (2019): 18169-18204.

[15] Bateman, Philip, and Hans Georg Schaathun. "Image steganography and steganalysis." *Department Of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, United Kingdom,* 4th August (2008).

[16] Rafat, Khan Farhan, and Muhammad Junaid Hussain. "Secure steganography for digital images." *International Journal of Advanced Computer Science and Applications* 7, no. 6 (2016): 45-59.

[17] Hassaballah, M., Mohamed Abdel Hameed, and Monagi H. Alkinani. "Introduction to digital image steganography.*" In Digital Media Steganography*, pp. 1-15. Academic Press, 2020.

[18] Cheddad, Abbas, Joan Condell, Kevin Curran, and Paul Mc Kevitt. "Digital image steganography: Survey and analysis of current methods." *Signal processing* 90, no. 3 (2010): 727-752.

[19] Abduallah, Wafaa Mustafa, and Abdul Monem S. Rahma. "A review on steganography techniques." *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)* 24, no. 1 (2016): 131-150.

[20] Oolo, Egle, and Andra Siibak. "Performing for one's imagined audience: Social steganography and other privacy strategies of Estonian teens on networked publics." *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 7, no. 1 (2013): 7.

[21] Nicolás-Sánchez, Alejandro, and Francisco J. Castro-Toledo. "Uncovering the social impact of digital steganalysis tools applied to cybercrime investigations: a European Union perspective." *Crime Science* 13, no. 1 (2024): 11.

[22] Hamid, Nagham, Abid Yahya, R. Badlishah Ahmad, and Osamah M. Al-Qershi. "Image steganography techniques: an overview." *International Journal of Computer Science and Security (IJCSS)* 6, no. 3 (2012): 168-187.

[23] Artz, D., 2001. "Digital steganography: hiding data within data." *IEEE Internet computing*, 5(3), pp.75-80.

[24] Nicolás-Sánchez, Alejandro, and Francisco J. Castro-Toledo. "Uncovering the social impact of digital steganalysis tools applied to cybercrime investigations: a European Union perspective." *Crime Science* 13.1 (2024): 11.

[25] Haji, Mohammed Suliman, et al. "A Survey on Digital Image Steganography and Steganalysis." *Journal of Computational and Theoretical Nanoscience* 17.7 (2020): 3256-3263.

[26] Balogun, Aishat. "Cybersecurity: Introduction to Steganography.*" Journal of Technology-Integrated Lessons and Teaching* 1.2 (2022): 18-23.