IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS ON VISUALIZATION OF DATA SCIENCE INTEGRATED TO DIFFERENT RESEARCH WITH NETWORKS THEORY

[1]Sai Arjun Madikonda

Btech, Gitam deemed to be University Hyderabad

ABSTRACT

Data complexity is a major consideration when scaling the data. Integration of heterogeneous data sets helps solve big data difficulties. Large-scale databases are difficult to comprehend and analyze since they require a lot of processing and storage space. With the rapid growth of data, extracting it in a fashion that the human mind can understand is becoming increasingly difficult. An introduction of Big Data research issues and triumphs, as well as the methodology and tools used to visualize it, will be discussed in this session. As a result, the primary purpose of this work is to summarize and provide novel solutions for the present state-of-the-art in Big Data visualization. Data types, analytical methodologies, and visualization techniques and tools are categorized in this study, with a special focus on the evolution of visualization methods over the last few years. The findings of the study show that there is still a disparity between the two groups. With regard to data visualization, the tools and applications we tested didn't have much support for advanced techniques like radial or temporal data display. Professional data analysts are eager to learn and use new tools and approaches, according to interviews. Exploratory analytic approaches are of particular relevance to users, as they can assist in their workflow.

1. INTRODUCTION

Visualization researchers have been extremely successful over the past few decades, developing a wide range of new approaches for presenting data in a visually appealing way. These methods include everything from data representation techniques (such as parallel coordinates) to procedures for user involvement and instruction (e.g., overview-first, details-on-demand). The wide range of visualization methods is evident in current surveys. Research in information visualization has already uncovered over 80 survey papers that describe pertinent state-of-the-art approaches (McNabb and Laramee, 2017), and another recent assessment of books in the field found the same quantity and variety (Rees and Laramee, 2019).

Humanity's whole history is a massive collection of data. Since the beginning of time, people have been storing information. We now live in a world where data is woven into every aspect of society, from politics to economics to science. When you look at social media sites like Facebook and Twitter, you'll see a huge amount of information being generated every day by users. Public access to data from government, scientific, and technical labs is now possible, as is data from space research projects.

260 gigabytes of human genome data are available through the 1000 Genomes Project. At the Internet Archive, ClueWeb09, and other sites, there are more than 20 terabytes of publicly accessible material.

More complex visualization techniques (such as chord diagrams, horizon graphs) are scarce among the tools and apps, and recent breakthroughs in visualization seldom make their way into the tools as new features or upgrades. An updated assessment of popular visualization techniques in 13 open source and six commercial programmes revealed that this issue persists. We were able to corroborate these findings.

1.1 Complexity as a Challenging Parameter to Integrate in Data Visualization

In recent years, there has been a considerable growth in research on visualization and complexity, as well as a significant increase in publications on these two issues. Complexity emerges in organisations as a result of interactions among their constituent parts according to systems theory (Gibb et al., 2019). Organizations often respond to environmental changes by adding complexity through the use of both controllable and uncontrollable factors. Scale plays a role in complex systems because emergent features are supplied via a network of internal and hyper-processes to attain a specific goal. For data visualization, the complexity of organisational dynamics is a major difficulty. It follows that every organisation or phenomenon that has a complicated structure or dynamic must have its visualization methods and procedures altered to portray the object's dimensional structure and scaled dynamics.

1.2 The Lack of an Integral Data Visualization Taxonomy to Tackle Complexity

Consolidation of scientific issues such as data visualization and complexity is taking place as more and more scientific papers and experts are working in these areas. In spite of this pleasant impression, a closer look reveals that associated issues are still unresolved:

1) A consistent pattern of complexity cannot be found in any scientific discipline that uses the term "complexity" as a synonym for "layered networks" from both an organisational and analytical perspective.

2) There are currently no boundaries to how complex an object can be represented analytically from the perspective of the subject. Instead, the complexity of an object forces data visualization approaches to be customised and sophisticated in order to capture its full range of analytical capabilities.

3) Finally, the standards for data visualization are extremely high: it has a wide range of applications as a tool, is specialised and versatile, and outcomes must be swift and effective.

2. LITERATURE REVIEW

When doing multidisciplinary research, literature reviews are critical because they provide a clear picture of current knowledge and help guide researchers from a wide range of disciplines down the path of investigation. Because of this, we decided to do a scoping review (Munn et al., 2018). Scoping reviews use a methodical approach to identifying the scope of a research problem, displaying the volume of literature, the study focus, and the key gaps in coverage, among other things (Armstrong et al., 2011).

Despite the fact that big data analytics have been extensively studied in academia, some industrial advancements and new technologies have mainly been explored in industry publications thus far (Elgendy and Elragal, 2014; Elragal and Klischewski, 2017). By critically summarising research in academia and industry, a literature review serves as the foundation for any additional study in information systems, hence it can be considered either a part of such research or research in itself. But this is more than a literature review, as it must show the connections between various publications and find connections between ideas and practise.

In our study, we conducted a literature review on the organisational changes, drivers, and activities linked to the realisation of big data value at various stages of analysis in the context of our study. Organizations confront challenges in realising value from big data, and we are responding to the call to focus on those challenges (Galliers et al., 2015). Threefold is the sum total of our effort. At the organisational, supra-organizational, and work-practice levels, we first outline six debates that are critical to how firms realise value from big data. It is necessary to do an empirical analysis of the competing viewpoints in order to determine whether or not they have any significance. When it comes to extracting value from large data, two sociotechnical variables are critical: portability and interconnectedness. Lastly, we believe that empirical research is needed to demonstrate how cross-level interactions play a role in realising the full potential of big data. Thus, it is a continuation of calls for research into how organisations' use of big data impacts their bottom lines (e.g., George et al., 2014, Markus and Topi, 2015).

As a result of their interest in novel visualization techniques, interviews with data scientists (Meeks, 2019) show that they are eager to experiment with new methods (Liu et al., 2019). Lack of time, documentation, and integration with current data science platforms were the top obstacles to data scientists learning and mastering these new visualization methodologies, according to the study.

All aspects of society can be affected by big data, from the social to the educational. It is increasingly important to manage raw data in companies that rely on technology since data volumes are increasing. For big data to be useful, new approaches must be devised to cope with the complexity and hidden facts they present. Because of this, big data analytics have been suggested as a tool for a variety of tasks including experimentation, simulation, data analysis, and monitoring. BDA technology can be used to do predictive analysis based on supervised and unstructured data input given by machine learning. The more precise and accurate the data input, the higher the analytical performance of machine learning analytics. It is a branch of machine learning known as deep learning that focuses on finding patterns and trends in data that are otherwise invisible.

2.1 Visualization and big data and characteristics

The term "visualization" generally refers to the use of images or graphics to represent data. Big data needs to be analysed and interpreted in a systematic way to have a deeper understanding of it. Data visualization aids in bringing together a variety of data sources, making connections between them, discussing issues as they arise, and figuring out where to focus analysis efforts (Abdullah et al., 2020). Researchers can use this approach to uncover previously unobserved patterns and learn how they are preserved. Data visualization technologies can be used by business analysts to identify areas for improvement, identify characteristics that influence consumer behaviour, and predict revenue volumes. (Ali et al., 2016).

2.2 Big Data Visualization Process

Figure 1 shows the steps involved in the visualization process:

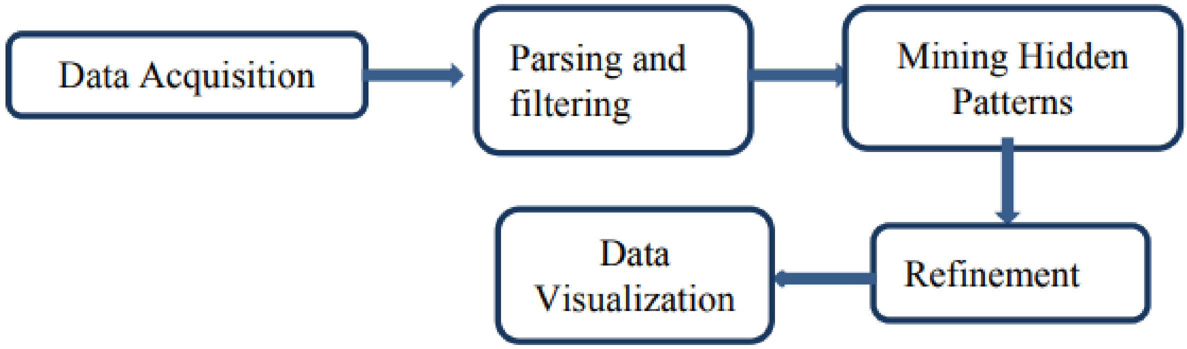

Figure1: Big Data Visualization Process

To begin the process of visualising data, one first need to gather information from a variety of sources. From many sources, there may be unstructured or semi-structured data, thus it must be parsed in a standardised format (Zeebaree, 2020). The next step is to remove the unnecessary data from the visualization process. Using diagrams and charts, useful patterns can then be discovered and depicted. The user's basic comprehension of secret knowledge is revealed through the extraction and visualization of useful patterns.

3. METHODOLOGY

3.1 Big data visualization Methods

There are a variety of ways to visualise huge data. In terms of data amount, diversity, and dynamics, these methods are graded. The following are a few different methods to display data:

3.1.1. Tree map

Hierarchical data can be represented as a nested rectangle collection in this technique. The parent rectangle is divided into sub-rectangles by a tiling technique. A taught approach is typically used. Rectangular regions are used to give numbers to categories. Consequently, the restriction of zero and negative values is only applicable to treemaps. ' In addition, the inclusion of extra pixels skews the hierarchy.

3.1.2. Circle Packing

It's a different kind of treemap that employs circles to indicate different hierarchical layers. The number of a type is determined by the circle region. The treemap, for example, uses several colours in different groupings. In comparison to the treemap, this method takes up more space.

3.1.3. Parallel Coordinates

This method is used to present vast volumes of information. A parallel coordinate system can be used to depict both the forest and the tree, as well as individual data components. To acquire consistent results, line trends are drawn. Individual data items can be highlighted to examine their precise output. However, a large number of data objects contribute to overplotting. This strategy isn't used with categorical data.

3.2 C.4 Stream Graph

This method is used to show how values change over time on a central timeline. It demonstrates how the quality of data from various categories has increased over time. The size of each stream form in a stream graph corresponds to the values of each category. It's good for displaying a lot of information. You can quickly become aware of a vast amount of data using data visualization tools. People may learn stuff they didn't know before with the correct data visualization tool (outliers, hidden patterns, or groups). These tools also allow you to explore into data sets that are constantly changing. Table 1 lists the most important features for big data visualization applications. Another example of a tool that aids in the study of the entire solar system is Space Titans 2.0. The goal was to obtain a new perspective on how our world seems as a result of today's greater spatial awareness provided by VR. From the standpoint of Large Data Visualization, skilling is a significant challenge caused by multidimensional structures, which necessitates scanning an information branch to gain any particular meaning or knowledge. Researchers are also interested in how virtual items and real-world scene vision are merged. This mapping has the potential to misrepresent the genuine scene while also slowing down the device. Even actual and virtual distances differ; as a result, a suitable structure system has been created to improve the link. In addition, palaeontology, type interpretation, MRI, and physics must all be thoroughly investigated.

Table 1: Big Data Tool Characteristics

Tools	Applications	Characteristics
Tableau	Market intelligence platform for the visual data collection used by scholars and public bodies	Can manage huge amounts of data, filter several data sets concurrently, users can generate and share dynamic and sharingable, dashboards depicting patterns and variants, develop interactive dashboards, built-in R support, Google Big Data Query API.
Plotly	online graphing, analysis, and static tools in both Python, R, MATLAB, Perl, J Arduino, and Restate graphics libraries	New open-access agile framework for data analytics and market research.
SAS Visual Analytics	Design tool; report, dashboard, and analytical distribution	Full research tool to allow users to recognize trends and relationships in data that are not clear initially
Microsoft Power BI	Using natural language questions on a dashboard to create immersive graphics, graphs and dashboards	For business users with their most important measurements in a single place, updated in near real - time, and available on all of their devices, power dash boards include a 360 ° view
D3.js	Using SVG, CSS specification, and HTML5 that are commonly applied	JavaScript library for immersive, collaborative web browser visualization

Visualization techniques have progressed significantly in recent decades (see Fig. 2), with human imagination serving as the only restriction to new techniques. While anticipating the next steps in data visualization evolution, it's vital to reflect on previous victories. In the discipline of statistics and analytics, quantitative data visualization is claimed to have only recently emerged. Prior to the nineteenth century, however, cartography and statistical drawings were created to aid in statistical reasoning, commercial planning, and other endeavours. Mathematical and statistical advantages, as well as improvements in drawing and reproducing visuals, have all resulted from the progress of visualization approaches.

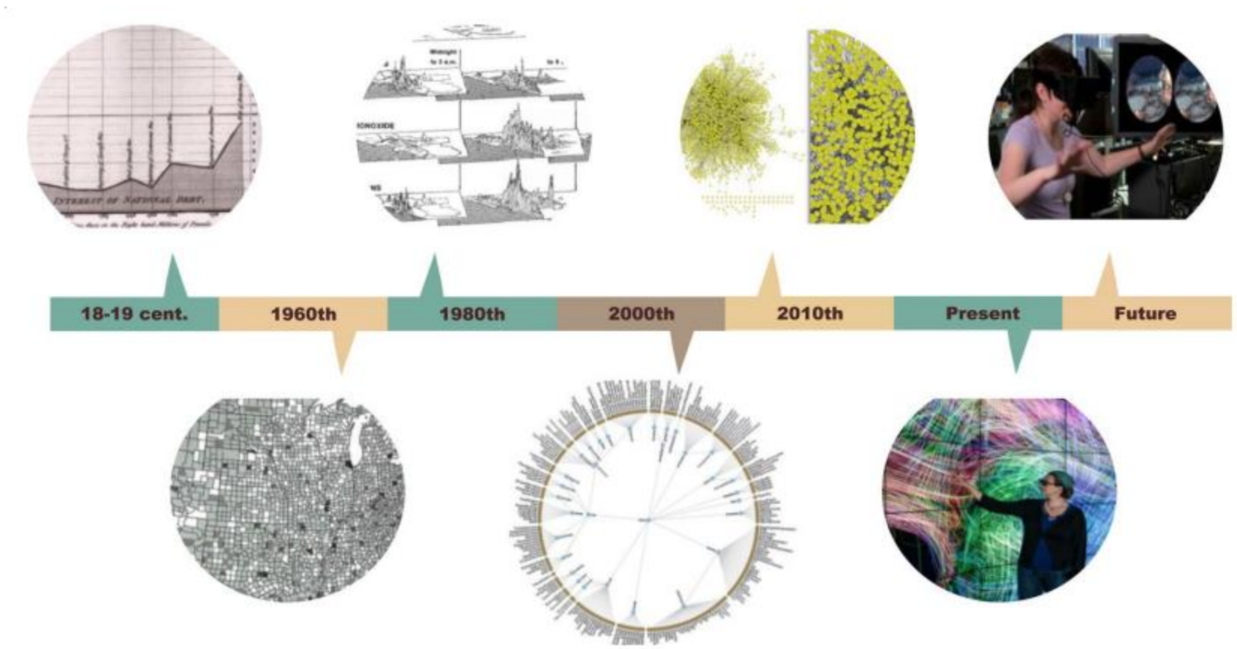

Fig. 2 The development of visualization techniques. The development of visualization methods dates back to the 18th century, and thanks to technological advancements, it is rapidly increasing today.

Tools for precise observation and measuring had been established by the 16th century. Specifically, it was at that time that the initial steps in the development of data visualization were performed.

The paradox of measuring space, time, and distance dominated the 17th century. Furthermore, demographic and economic data were being collected around the world.

Data graphical representation, statistical theory, and new graphic forms were all developed in the 18th century. Geological, pharmacological, and economic data were first depicted on themed maps at the turn of the century. Cartograms and geometric figures were employed by Charles de Fourcroy to compare regions and populations. As a revolutionary, Johann Lambert (1728–1777) was known for his use of tables and line graphs to show how data changed over time. Initially, basic charts and one-dimensional histograms were used, with the latter being the preferred option. Those examples, however, only apply to tiny data sets. This style of diagram would become useless if more information was added to it.

At the turn of the twentieth and twenty-first centuries, progress was made in the creation of interactive statistical computers and novel data analysis paradigms. Technological advances surely facilitated the quick development of visualization approaches, methodologies, and tools. Large-scale statistical and graphical software engineering was developed, considerably boosting the speed and capacity of computer processing.

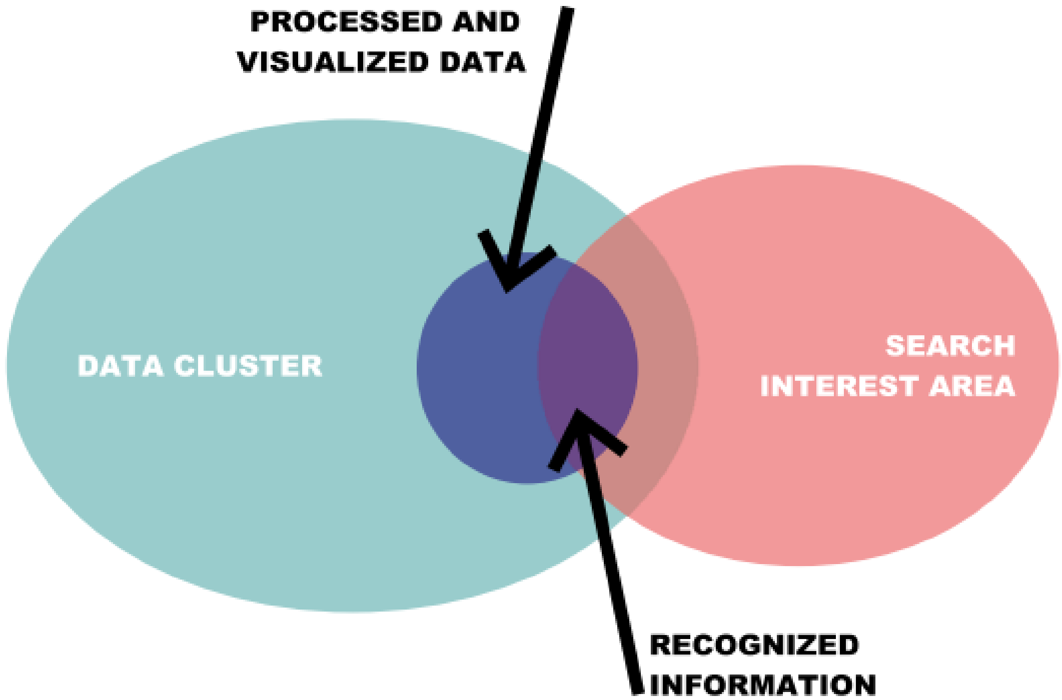

Fig. 3 A problem with human perception. The human ability to perceive enormous amounts of data is insufficient.

To depict multiple data types, the second aspect relies on display techniques and samples. On the other side, data visualization is speculative and highly limited by one's perceptual abilities and desires (see Fig. 3).

4. RESULTS AND DISCUSSION

The following are some of the key study conceptual findings:

(1) There are six main elements in communication sciences that can be applied to data visualization: the message, the form, the encoder/context/channel/decoder.

(2) Information can be represented graphically, numerically, visually, and thematically through the use of a variety of different mediums.

(3) The complexity of data visualization must be taken into account and organised in six levels: basic, extended, dynamic, synthetistic and integrative interactive levels of complexity.

Layers specified by analytical criteria reveal the level of complexity of the organised entity or event being represented and are used to make object-oriented data visualization more consistent.

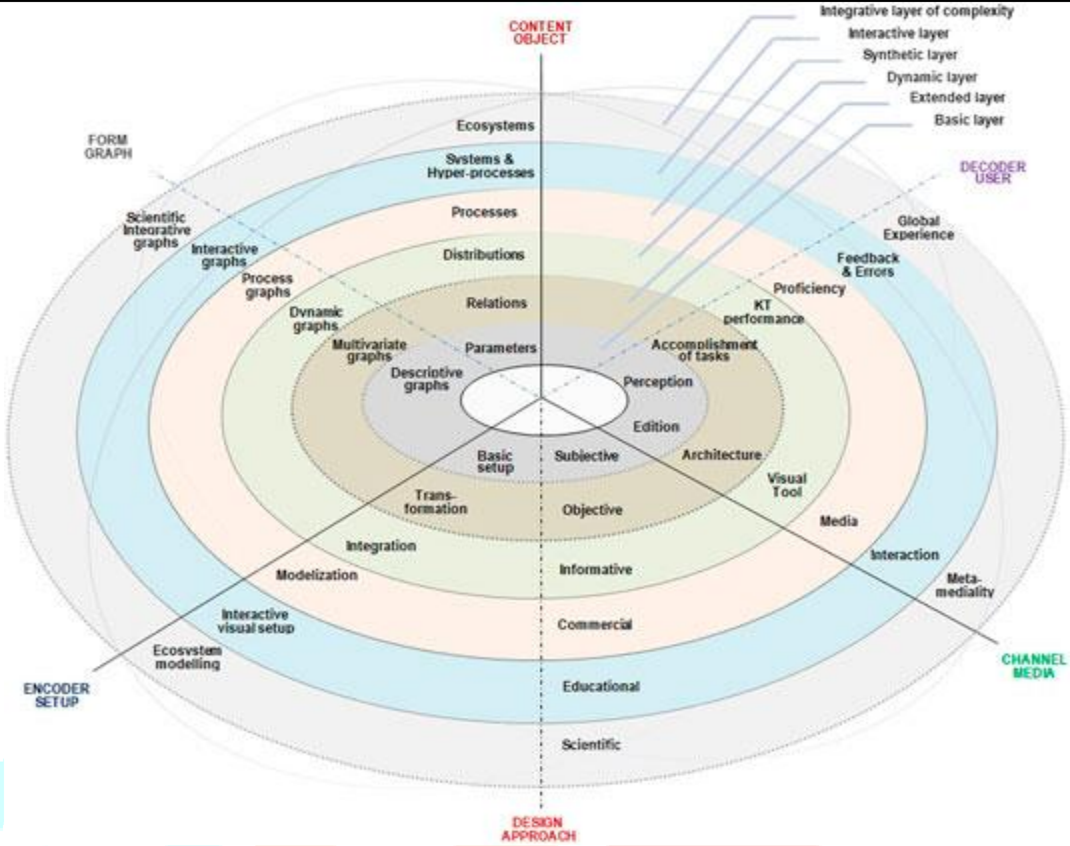

FIGURE 4: Examples of dimensional taxonomy for object-oriented data visualization from the standpoint of communication science: elements-axes as factors of completeness, and layers spheres as factors of complexity.

Once the possibilities of the combination of these elements have been explored, the data visualization process must be handled, resulting in the observed successes at each crossroads between communication component x organisational complexity (see Figure 4).

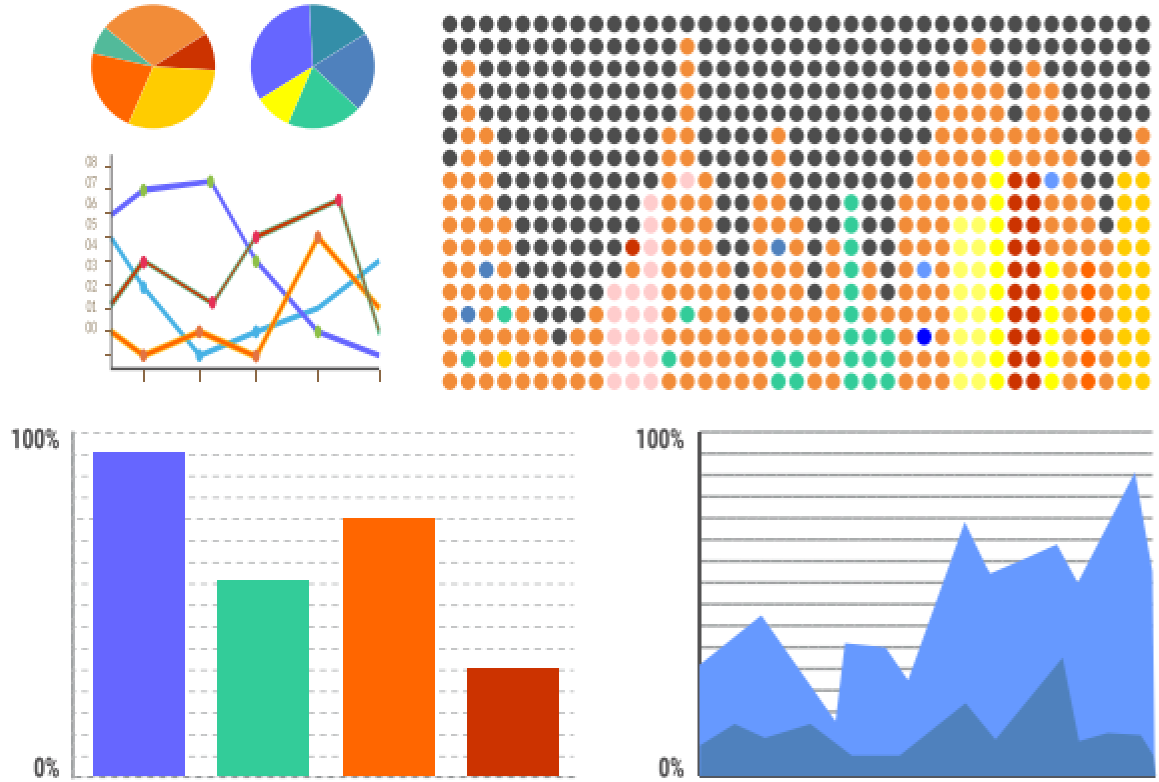

Fig. 5 The dashboard as an example. This image depicts a pie chart, a pixel-based data visualization, a line graph, and a bar graph.

To view log files in a number of formats, you can use the Dashboard's filters to narrow your search (see Fig. 5). It is possible to create dashboards with three different types of information: raw data, analysis, and visualization (graphical representation based on the analysis layer).

CONCLUSION

Processing and analysing Big Data presents numerous obstacles in practise. Data extraction, perception, and cognition are all growing more difficult as all data is currently visualised by computers. These duties require time, and the results aren't always correct or desirable. For users working with data, visual interfaces can be extremely helpful. For exploratory data analysis, however, the "Interactive Visualization Gap" still exists. Our investigation on the use of visualization approaches in standard data science tools, which is reported in this work, has also revealed this. In addition, this presentation examines some recent research on large data processing and data visualization. It also compares their outcomes based on the algorithms and procedures they use. This study was made as a result of the challenges and approaches provided in connected studies employing virtual reality based on big data visualization.

REFERENCES

1. Z. Munn, M.D. Peters, C. Stern, C. Tufanaru , A. McArthur, E. Aromataris
2. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach BMC Med. Res. Methodol., 18 (1) (2018), p. 143 Close
3. R. Armstrong, B.J. Hall, J. Doyle, E. Waters Scoping the scope of a Cochrane review
4. J. Public Health, 33 (1) (2011), pp. 147-150
5. R.D. Galliers, S. Newell, G. Shanks, H. Topi Special issue: the challenges and opportunities of 'datafication': Strategic impacts of 'big' (and 'small') and real time data—for society and for organizational decision makers
6. J. Strategic Inform. Syst., 24 (2) (2015), pp. II-III, 10.1016/S0963-8687(15)00033-5

7. G. George, M. Haas, A. Pentland From the editors: big data and management Acad. Manag. J., 57 (2) (2014), pp. 321-326, 10.5465/amj.2014.4002
8. Liu, J., Boukhelifa, N., and Eagan, J. R. (2019). Understanding the Role of Alternatives in Data Analysis Practices. IEEE Transactions on Visualization and Computer Graphics (Early Access).
9. Harfouchi F et al. Modified multiple search cooperative foraging strategy for improved artificial bee colony optimization with robustness analysis. Soft Computing. 2017;22(19)
10. Abdullah, P. Y., Zeebaree, S. R., Jacksi, K., & Zeabri, R. R. (2020). AN HRM SYSTEM FOR SMALL AND MEDIUM ENTERPRISES (SME) S BASED ON CLOUD COMPUTING TECHNOLOGY. International Journal of ResearchGRANTHAALAYAH, 8(8), 56–64.
11. Ali, S. M., Gupta, N., Nayak, G. K., & Lenka, R. K. (2016). Big data visualization: Tools and challenges. 656–660.
12. Dino, H., Abdulrazzaq, M. B., Zeebaree, S. R., Sallow, A. B., Zebari, R. R., Shukur, H. M., & Haji, L. M. (2020). Facial Expression Recognition based on Hybrid Feature Extraction Techniques with Different Classifiers. TEST Engineering & Management, 83, 22319–22329.