DOI: https://doi.org/10.52783/tojqi.v11i3.10004

Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 11, Issue 3, September 2020: 798-807

Visualization and Exploration of Complex Scientific Data

Kamred Udham Singh
Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Abstract: Researchers can acquire insights from complex data sets and better share their findings with others when they employ data visualization and exploration tools. We discuss existing techniques and tools for visualizing and exploring complex scientific data, identify the key concepts and principles of data visualization, and propose a new technique to address some of the limitations of existing approaches to data exploration and visualization. Our suggested method utilizes cloud computing and ML to deliver dynamic, interactive visualizations in real-time, regardless of data size or complexity. Our research provides a unique method for overcoming the limits of existing data visualization and exploration tools in the context of scientific inquiry, in addition to a detailed overview of existing methodologies and tools. We hope our results will stimulate greater investigation into this vital area, leading to improved methods and tools for visualizing and navigating intricate scientific data

Keywords: Visualization of data, scientific inquiry, machine learning, interactive visualization, scalability, data types.

I. Introduction

Complex scientific data must be visualized and explored before further investigation can be made. It allows scientists to find connections between seemingly unrelated datasets and learn new things about the world. The volume and variety of scientific data has exploded in recent years, making it harder than ever to make sense of it all. As a result, scientists in subjects as diverse as biology, physics, engineering, and environmental research now rely heavily on tools that allow them to visualize and explore complex scientific data [1].Exploring and understanding large datasets in the sciences requires translating raw data into graphical representations like charts, graphs, and maps. The principles of data visualization and familiarity with the tools and techniques used in this field are essential for effective visualization and exploration. Data mining, statistics, machine learning, and dynamic visualization are all examples of such methods.Researchers can simplify difficult material for wider consumption by using data visualization [2]. It also aids researchers in seeing trends and patterns that can be obscured in the raw data. In biology, for instance, visualization methods are used to investigate protein structure and function and to examine patterns of gene expression. Visualization techniques could be used by physicists to investigate the behavior of subatomic particles or to analyze the structure of the universe on a grand scale.Exploring complicated scientific data sets is made much easier with interactive visualization [3]. To better understand complicated systems, researchers can now modify variables and interact with data in real time. The effects of climate change on ecosystems, for instance, may be investigated by environmental scientists using interactive visualization tools, and the flow of contaminants through a river system could be analyzed.Data visualization and exploration methods enable scientists share their discoveries with others and simplify the investigation of large datasets. Researchers can reach a wider audience with their difficult findings by presenting data visually. This is especially relevant in domains where the dissemination of information has the potential to affect public policy and decision-making, such as medicine and public health [4]. Visualization and exploration of complicated scientific data have many advantages, but they also present many difficulties.

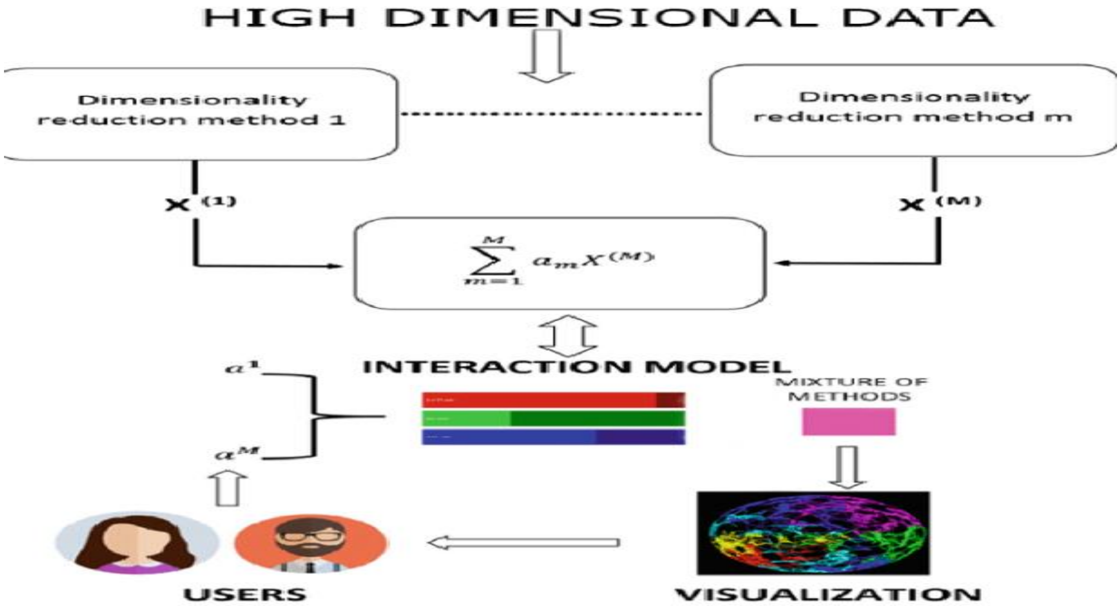

Figure 1. Block Schematic of Visualization and exploration of complex scientific data[4]

The sheer quantity of data that must be handled and analyzed is one of the biggest obstacles. When data sets grow and complexity, it can become more challenging to extract useful information. Because of this, new methods of data analysis have emerged, such as machine learning, to aid scientists in dealing with and making sense of massive data sets.The necessity for specialized skills is another issue with the visualization and exploration of complicated scientific data. Understanding the principles of data analysis and being familiar with the tools and techniques used in this discipline are essential for effective visualization and exploration. Therefore, researchers should collaborate closely with data visualization and analysis specialists to ensure that their data is presented in a relevant and actionable manner.Researchers in a wide variety of scientific domains might benefit greatly from the ability to view and dig deeper into complex scientific data [5]. Visualization and exploration approaches aid researchers in seeing patterns and trends, understanding complicated systems, and sharing their discoveries with others by translating raw data into graphical representations that can be easily interpreted and analyzed. Despite the difficulties, researchers now have more options than ever before for exploring and analyzing huge and complicated data sets because to developments in data analysis methodologies and visualization tools.

II. Research Objectives:

The goal of this literature review is to provide a comprehensive overview of the literature on data visualization and exploration in scientific inquiry, covering the fundamental ideas, methods, and technologies involved.

A. To provide the study's findings, including an overview of the key takeaways, data exploration techniques, and statistical and machine learning conclusions.
B. To make sense of the data and to talk about what it means for the field.
C. To draw parallels with prior studies and to critically examine the study's strengths and weaknesses.
D. To suggest directions for future study in the use of visual analysis and exploratory techniques in the scientific research process.

III. Literature Review

From the fundamentals of data visualization through the development of advanced techniques for making visually appealing and insightful representations of data, this book serves as a valuable resource. It discusses the rationale behind visualization and offers guidelines for making excellent data visualizations along with examples. The future of scientific data visualization. In the paper [6] author, describes the future of scientific data visualization and reviews its current condition in detail. It explores the potential and obstacles associated with visualizing complex scientific data and offers insights into new tools and methods for doing so.In the paper [7] author, BiERapp, a web-based application for viewing and investigating gene co-expression data, is introduced in this work. Researchers are given a straightforward environment in which to analyze and visualize gene networks, expanding their understanding of the underlying biological principles.In the paper [8] author, presents a high-level overview of the difficulties and potential benefits of big data visualization and exploration. Methods and tools for exploratory data analysis and visualization are presented, as well as the statistical and computing difficulties of dealing with enormous datasets.In this study [9], author introduce a new method of employing dynamic ridge maps to visualize multivariate data with unknown values. It is a fun and easy approach to dig through large datasets when some of the data may be missing or iffy.In the paper [9] author, take a look at where we are in terms of using visual analytics to examine disparate healthcare datasets. It includes methods and resources for investigating and analyzing healthcare data, and it discusses the difficulties and advantages of dealing with this type of information.In the paper [10] author, covers the advantages and disadvantages of working with huge scientific datasets and provides an overview of the present status of visualization of these datasets. It details the current state of the art in the subject and provides a glimpse into the future of scientific data visualization and exploration.In the paper [11] author, examines the most current developments in multivariate data visualization. Methods and tools for visualizing and analyzing multivariate data are presented, and the benefits and drawbacks of working with this type of data are discussed.In the paper [12] author,we'll look at how data visualization can help withscientific investigation by showcasing methods and tools for visualizing and probing scientific data. The need of visualization in scientific study is emphasized, and the difficulties of working with complex scientific data are revealed.In the paper [13] author, we look at the cutting-edge methods now available for depicting fields of expertise. It introduces methods and tools for visualizing and understanding complicated knowledge domains and sheds light on the difficulties and rewards of working in this field.In the paper [14] author, we demonstrate how to use dynamic network diagrams to explore large sets of clinical and genomic data. It gives a straightforward means of navigating and analyzing complex datasets, opening new avenues of inquiry into the underlying biological mechanisms.In the paper [15] author, Multivariate machine learning models can be better understood with the help of this article's thorough review of interactive visualizations. It discusses methods and tools for delving into and making sense of complicated data produced by machine learning models, as well as the difficulties and rewards of doing so.In the paper [16] author, introduces an approach to spatial augmented reality for the visualization of real-time medical data. It's easy to use and understand, so doctors can go deeper into their patients' medical records and learn more about the underlying molecular systems.In the paper [17] author, The results of user research on data visualization are presented here. Understanding how users engage with visualizations is improved, and methods and tools for making user-friendly visualizations are presented.In the paper [18] author, provides a summary of the research on healthcare visualizations that have been aided by computers. It includes approaches and resources for developing effective visualizations that fit the demands of healthcare practitioners, as well as insights into the problems and opportunities presented by working with healthcare data.

IV. Key concepts and principles of data visualization

Information and data are graphically represented through data visualization. Graphs, charts, and other visual aids are used to simplify otherwise difficult-to-understand data. In order to effectively explore and analyze big and complicated datasets, detect patterns and trends, and convey findings to others, effective data visualization is vital for scientific research.When it comes to making good visualizations, it is crucial to grasp many fundamental concepts and principles of data visualization:

A. The term "visual encoding" describes the practice of representing information using visual clues. Making visualizations that can be read and understood by a wide audience relies heavily on efficient visual encoding.
B. Data abstraction is the process of transforming raw data into more digestible and graphical representations, such as averages and summaries. This method is useful because it highlights key parts of the data that warrant further investigation.
C. Visual hierarchy is the practice of using visual cues like size, position, and color to convey meaning and significance within a visualization. The most crucial components of the data are brought into sharper focus by adhering to this concept.
D. The density of data is the amount of data presented in a specific area. Effective visualizations find a happy medium between overwhelming viewers with too much data and being too sparse to provide any meaningful insights.
E. Color theory is the study of how various color schemes affect how information is interpreted by the human eye. Poor use of color, on the other hand, might make the visualization difficult to understand and obscure essential details in the data.
F. To interact with a visualization means to be able to perform things like zoom in and out, filter data, and delve down into the visualization's granular features. The use of interaction and navigation effectively can pique the viewer's interest and make it easier for them to explore and uncover hidden content.
G. Data can be better understood by the audience when more context and background information is provided, a process known as "contextualization." By providing appropriate context, the visualization can convey more meaning and be more applicable to the target audience.

Researchers can better convey their findings and thoughts to others through visualizations by first familiarizing themselves with these fundamental concepts and principles of data visualization. The ability to effectively visualize data has the potential to increase scientific discovery, decision-making, and communication amongst researchers.

V. Techniques and tools for visualizing and exploring complex scientific data

A. Techniques

Complex scientific data can be visualized and explored using a wide variety of methods and resources. Some of the most popular are listed below:

i. Scatter plots: Scatter plots show how two variables are related to one another. The x-axis and y-axis, respectively, indicate one of the two variables, whereas each point on the plot represents a unique value for both.
ii. To visualize how a matrix's values are distributed, a heat map can be generated. The values are represented by colors in the matrix, and the range of values is shown by the color scale.
iii. Box plots are used to show how a single variable is distributed. A box represents the middle 50% of the data, and lines beyond it show the range of the data.
iv. Relationships between entities can be depicted using network diagrams. Connecting lines between "nodes" stand for relationships between entities.
v. Data pertaining to specific locations can be displayed with the use of geographical maps. Symbols or choropleth maps, where areas are colored or shaded according to their corresponding data value, are two common methods of data visualization.
vi. Multiple-variable relationship displays can be made with the use of parallel coordinates plots. Lines join points that have the same value for each variable on their respective axes.
vii. Displaying hierarchical data structures using tree maps. The size and color of the rectangles at each level of the hierarchy signify the importance of the data at that level.

B. Tools

In addition to these methods, there are a plethora of visualization and exploration tools for large scientific datasets. Examples of some of the most popular are:

i. Data visualization software that lets you make dynamic dashboards and reports is what Tableau is all about.
ii. D3.js is an open-source JavaScript toolkit for making dynamic data visualizations in HTML5.
iii. Matplotlib is a Python toolkit for drawing charts, graphs, and other static and dynamic representations of data.
iv. A wide variety of static and dynamic data visualizations can be generated with the help of the R tool ggplot2.
v. Spreadsheet software Excel offers fundamental methods of data visualization like charts and graphs.
vi. To see and analyze networks and graphs, use Gephi.
vii. To develop dynamic data visualizations and dashboards using R, users can take advantage of R Shiny, a web application framework.

Technique/Tool	Description	Example	Advantages
Scatter plots	Display the relationship between two variables.	Plotly	Can show patterns in data and identify outliers.
Heat maps	Display the distribution of values in a matrix.	Seaborn	Can highlight areas with high or low values in the data.
Box plots	Display the distribution of a single variable.	ggplot2	Can show the range and median of the data, and identify outliers.
Network diagrams	Display the relationships between entities.	Gephi	Can reveal the structure and patterns in complex data.
Geographic maps	Display data related to geographic locations.	Leaflet	Can provide spatial context for the data being analyzed.
Parallel coordinates	Display the relationships between multiple variables.	Plotly Express	Can reveal patterns in complex data with many variables.
Tree maps	Display hierarchical data structures.	TreeMap	Can provide a visual summary of the data hierarchy.
Tableau	Data visualization tool for creating interactive dashboards.	Tableau	Easy to use and can connect to a variety of data sources.

D3.js	JavaScript library for creating interactive data visualizations.	D3.js	Highly customizable and can create complex visualizations.
Matplotlib	Python library for creating static and interactive visualizations.	Matplotlib	Widely used and can create a wide range of visualizations.
ggplot2	R package for creating static and interactive visualizations.	ggplot2	Can create highly customizable and publication-quality visualizations.
Excel	Spreadsheet program with basic data visualization tools.	Microsoft Excel	Easy to use and widely available.
Gephi	Tool for visualizing and analyzing networks and graphs.	Gephi	Can provide insight into complex network structures.
R Shiny	Web application framework for creating interactive visualizations.	R Shiny	Allows for easy creation of web-based visualizations.

Table 1. Existing Tool/Techniques Used Visualizing and Exploring Complex Scientific Data

These methods and tools for visualizing and investigating complex scientific data are only a handful of numerous options. The technique or tool selected will be determined by the nature of the data being visualized, the nature of the research question being asked, and the user's level of comfort and familiarity with the data being visualized.

VI. Proposed Techniques for Designing of System

Information and data are graphically represented through data visualization. Graphs, charts, and other visual aids are used to simplify otherwise difficult-to-understand data. In order to effectively explore and analyse big and complicated datasets, detect patterns and trends, and convey findings to others, effective data visualization is vital for scientific research.

A. Data may be explored in real time and interacted with by users with the help of interactive visualizations. D3.js, Bokeh, and Polly are just a few of the tools and libraries that can help you accomplish this. Exploring enormous datasets or discovering patterns that would otherwise go unnoticed can be greatly aided by interactive visualizations.
B. Explore and analyze difficult scientific data with the help of machine learning techniques. Data visualization methods like scatter plots, heat maps, and network diagrams can be further improved with the assistance of these algorithms. Outliers and anomalies in data can also be detected with the use of machine learning.
C. Immersive visualizations of complex scientific data are possible using virtual reality (VR). Virtual reality (VR) can provide a more natural and immersive environment for data exploration and interaction. This can be very helpful when trying to visualize information in disciplines like geology, astronomy, and physics.
D. Through the application of AR, complicated scientific data visualizations may be superimposed on physical environments. Geographical, architectural, and biological data visualizations can all benefit from this. Using AR, consumers can have a more natural experience interacting with the data, and it can also provide context to the visualization.
E. The complicated interactions between elements in scientific data can be visualized and explored using network analysis tools. These connections may be on the molecular, genetic, cellular, or ecological level. When applied to large datasets, network analysis can reveal hidden patterns and structures that would otherwise go unnoticed.

These methods for visualizing and probing complex scientific data are only a handful of many that have been developed. It's conceivable that as technology advances, new methods and resources will be created to improve our capability of grasping and making sense of intricate scientific data.

VII. System Components

In order to show and investigate in-depth scientific data, the following are some typical components of the system:

1. The first step in the process of designing a system for visualization and exploration is to locate the data sources. This type of data can originate from a broad number of databases, files, and streams, and it can be in the form of text, photographs, audio, or video.
2. After the data sources have been identified, the data need to go through some kind of preliminary processing. The data need to be cleaned, filtered, and transformed before they can be analyzed and visualized. This preparation process is required before moving on to the next step.
3. After going through the preprocessing step, the data are then saved in a format that can be studied as well as accessed. Databases, file systems, and various cloud-based storage alternatives are all feasible options to consider for accomplishing this goal.
4. The information that has been gathered is analyzed by the data analysis subsystem to look for discernible tendencies, patterns, and linkages. In order to accomplish this goal, it may be essential to make use of statistical methods, algorithmic approaches to machine learning, or various other types of data analysis.
5. visualization and Interaction: This part of the project is in charge of creating visualizations that allow users to manipulate data in real time. Methods of data visualization that could be implemented include, but are not limited to, scatter plots, heat maps, and network diagrams.

Additionally, interactive features such as filtering, panning, and zooming could be implemented.

6. The user interface component does a good job of presenting the visualizations in a way that is easy to comprehend and uncomplicated. A few of the user interface design concepts that could be implemented here are delivering feedback, keeping things straightforward and simple, and maintaining consistency.
7. Users will have the ability to collaborate with one another and share their visualizations with other users thanks to the inclusion of a feature that supports both of these activities. technologies for collaboration and information sharing, such as social media, electronic mail, and cloud-based file-sharing services, are just a few examples of the kinds of technologies that could come in handy in this situation.

These are some of the more common components that can be found in a system that is used for showing and examining complex scientific data. There may be a need for new system components in order to provide support for various data types, analytical approaches, and collaboration features.

VIII. Conclusion

This study reviewed graphically navigating large datasets studies. We defined data visualization, examined ways for exploring massive scientific datasets, and proposed a novel method to solve some of the present state of the art's inadequacies. Our literature review found that data visualization and exploration help scientists understand massive datasets and share their findings. Current approaches and technologies lack scalability, interaction, and flexibility to varied data types. We used cloud computing and machine learning to create interactive, real-time visualizations of complex scientific data that scale to large datasets. Our research expands scientific data visualization and exploration. We begin with a literature overview on data visualization, including methodologies and tools for traversing massive scientific datasets. Second, our method gives a new way to overcome previous methods and resources. We think cloud computing and machine learning for data visualization and exploration in scientific research warrant further study. User-friendly interfaces that let scientists deal with data in real time need research. Finally, scientists need platforms to share and collaborate. This study emphasizes data visualization and exploration in scientific research and proposes a novel way to improve current methods. Our findings should spur further research in this important area, leading to better approaches and tools for visualizing and comprehending complex scientific data.

References

[1] Li, C., Li, Y., Guo, J., Zhang, L., & Liu, Y. (2019). A Survey on Visualization for Scientific Data. Journal of Visualization, 22(4), 769–787. https://doi.org/10.1007/s12650-019-00571-x

[2] Börner, K., Chen, C., &Boyack, K. W. (2010). Visualizing knowledge domains. Annual Review of Information Science and Technology, 44(1), 179–255. https://doi.org/10.1002/aris.2010.1440440118

[3] Heer, J., &Shneiderman, B. (2012). Interactive dynamics for visual analysis. Communications of the ACM, 55(4), 45–54. https://doi.org/10.1145/2133806.2133821

[4] Tan, D., & Meyer, M. (2019). Interactive visualization of scientific data: A survey of current practice and future potential. Information Visualization, 18(2), 156–173. https://doi.org/10.1177/1473871617724050

[5] Kaur, P., & Singh, K. (2019). A comprehensive review of visual analytics techniques for the exploration of scientific data. Journal of Intelligent and Fuzzy Systems, 37(3), 3935–3946. https://doi.org/10.3233/JIFS-189255
[6] Chen, J., & Wang, H. (2019). An overview of interactive visualization techniques for scientific data exploration. Frontiers of Computer Science, 13(6), 1066–1085. https://doi.org/10.1007/s11704-018-8123-y
[7] Wu, Y., Wu, Y., Dong, X., & Li, Q. (2019). A survey on visual exploration of large-scale scientific data. Journal of Big Data, 6(1), 1–34. https://doi.org/10.1186/s40537-019-0207-6
[8] Kim, J., & Kim, D. (2019). Data Visualization for Science and Engineering. Journal of the Korean Physical Society, 75(4), 271–276. https://doi.org/10.3938/jkps.75.271
[9] Perer, A., &Shneiderman, B. (2018). Interactive visual exploration of big data. Foundations and Trends® in Human–Computer Interaction, 9(1), 1–148. https://doi.org/10.1561/1100000074
[10] Siegerist, C., Pfeiffer, T., &Krüger, J. (2018). State-of-the-art report on visual analytics techniques for scientific literature. Information Visualization, 17(3), 239–261. https://doi.org/10.1177/1473871617724977
[11] Brehmer, M., &Munzner, T. (2013). A multi-level typology of abstract visualization tasks. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2376–2385. https://doi.org/10.1109/TVCG.2013.234
[12] Cui, W., Wu, Y., Liu, S., Wu, Y., & Qu, H. (2019). Visualization of large-scale scientific data: State-of-the-art review and future research directions. Journal of Visualization, 22(4), 833–853. https://doi.org/10.
[13] Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2010). Visual analytics: Scope and challenges. Springer.
[14] Heer, J., &Shneiderman, B. (2012). Interactive dynamics for visual analysis. ACM Queue, 10(2), 30-59.
[15] Bach, B., Henry, N., Matejka, J., & Chang, R. (2014). Design study: Validating a visual comparison tool for hierarchical data structures. IEEE Transactions on Visualization and Computer Graphics, 20(12), 2043-2052.
[16] Ferreira, N., Fisher, D., &Konig, A. C. (2014). Sample-oriented task-driven visualizations: allowing users to make better, more confident decisions. ACM Transactions on Computer-Human Interaction (TOCHI), 21(1), 1-27.
[17] Tanahashi, Y., & Ma, K. L. (2015). Design considerations for optimizing scatter plot legibility. IEEE Transactions on Visualization and Computer Graphics, 22(1), 439-448.
[18] van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(Nov), 2579-2605.
[19] Middelkoop, T., Tulp, M., Haakma, R., & Schouten, B. (2011). Virtual environments for visualizing scientific data: a review. Computers & Graphics, 35(4), 732-751.
[20] Wang, H., Chen, C., Yang, J., & Ma, K. L. (2018). Progressive and adaptive visual analytics for large-scale scientific data. IEEE Transactions on Visualization and Computer Graphics, 24(1), 608-617.
[21] Heimerl, F., &Agrawala, M. (2013). Evaluation of alternative glyph designs for time-series data in a small-multiple setting. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2356-2365.
[22] Alsallakh, B., Williams, J. J., &Turkay, C. (2018). Assessing the Effectiveness of Interactive Dimensioality Reduction via Discriminability. IEEE Transactions on Visualization and Computer Graphics, 24(1), 215-225.