GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
DATA VISUALIZATION TECHNIQUES: A REVIEW

Rajvir Kaur[*1] & Birpal Kaur[2]
*[1]Research Scholar, Faculty of Engineering, Design and Automation, GNA University, Punjab, India
[2]Assistant Professor, Faculty of Engineering, Design and Automation, GNA University, Punjab, India

ABSTRACT

Data visualization involves representing data and information in a graphical or pictorial form so that it can be easily understandable. At Present time, data is increasing at a very fast rate so, it is important to visualize and analyze the massive amount of data by using various visualization techniques. Data Visualization techniques are very helpful to visualize and understand outliers, trends, and patterns in data and thus helpful in decision making. This paper presents a review of the basic concepts of data visualization and various techniques and tools used for visualizing data. Some big data visualization techniques, which are the need of the hour, are also being discussed.

Keywords: Data visualization, Information Visualization, Data visualization Techniques and tools, Big data visualization.

I. INTRODUCTION

Data visualization is the procedure of translating data into visual elements like graphs, charts, and maps [1]. It is important to understand "Why do we need data visualization?" before visualization of any type of data. As we all know, at the present time data is growing very exponentially. Enormous amount of data is generated by social media and by various organizations in every minute of the day. According to the World Economic Forum, the world produces 2.5 quintillion bytes of new data every day and 90% of all the data has been generated over the last two years [2]. It is very difficult for users to explore, visualize, and manage this enormous amount of data. So to visualize complex data, we need data visualization techniques. In this paper, we will discuss various kinds of data visualization techniques.

Nowadays, Data visualization techniques are very useful because the visual context of information makes it easier to identify trends and patterns than looking through thousands of rows and columns on a spreadsheet. Data visualization make use of computer graphics to show patterns, trends, and relationships among a huge amount of data. It deals with the design, development, and application of computer-generated pictorial and graphical representation of the data [3]. To visualize a large amount of complex data, data visualization techniques make use of graphs and charts. There are various visualization techniques like Line Charts, Bar Charts, Pie Charts, Scatter Plots, Maps, Word Cloud, Network Diagrams, and Tree Diagrams for representing one dimensional as well multi-dimensional data [4]. We will also discuss various data visualization tools used for visualizing the large amount of complex data generated by people every day.

II. VISUALIZATION TECHNIQUES

Data Visualization is a new field in Information technology. Visualization is the graphical representation of information, with the aim of providing the viewer with a qualitative understanding of the information contents [5]. It is an easy way to represent more complicated data in the form of graphics by using various visualization techniques. Data visualization techniques are used to show the relationship among datasets. Most Common visualization techniques are as shown:

1. Charts

Line chart is used to show the dependence or relationship of one variable to another variable. It helps us to understand patterns, trends, and fluctuations in your data. It can also be used to compare changes over a period of time. If you

want to show a trend over time such as days, months, or years then you can use a line chart. Figure1 showing the trend for three products (ie: Laptops, Desktops, and Tablets).
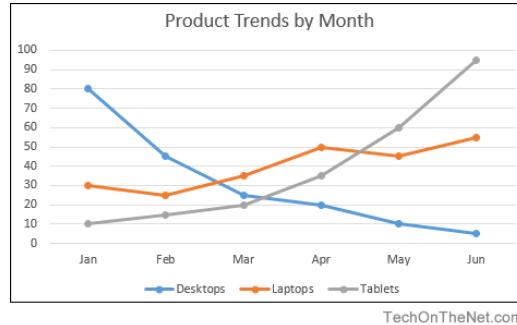

Fig 1: Line chart [6]

Pie chart is most commonly used for comparing the parts of a whole. It is most effective when percentages are included to describe the content and there are limited number of components. By using a pie chart, you can easily compare which items are most popular like below Figure2 shows which social media platform is most used by the users.
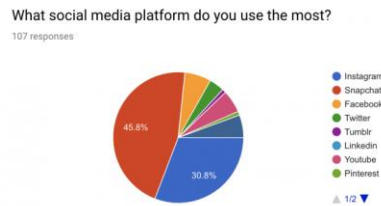

Fig 2: Pie chart [7]

Bar chart is useful in comparing the values of different groups or categories. By using rectangular bars, we can represent the category values, and they can be constructed with horizontal or vertical bars. Following Figure3 represents the number of purchases made on a site by different types of users.
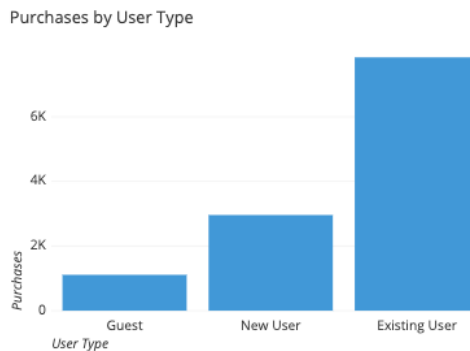

Fig 3: Bar chart [8]

2. Plots

Plot visualization allows distributing two or more data sets over a two dimensional or even three dimensional space to show the relationship between these data sets.

Scatter Plot is a two-dimensional plot for showing the relationship between two data items. All symbols such as plus signs , dots, and squares represents an observation. The value for each observation indicates the marker position. Scatter plots are used for determining the relationship, or correlations, between X and Y variables. Figure4 shows the heights and diameters for a sample of fictional trees.
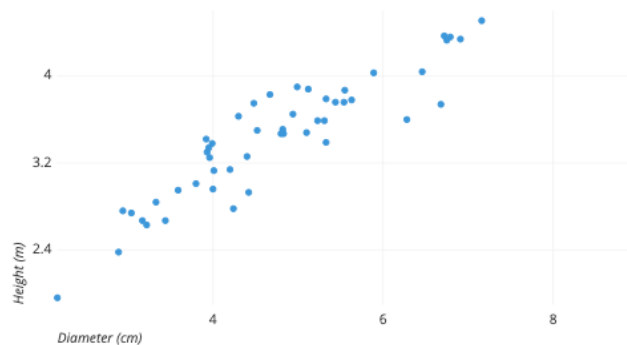

Fig 4: Scatter Plot [9]

3. Maps

Map visualization is used to display and analyze the geographically related data and present it in the form of maps. It is generally used to represent the location of different areas of a country [1]. This type of data expression is clearer and more intuitive. We can visually see the distribution of data in every region.

Heat map visualization is a graphical representation of numerical data, where every data point values are shown using colors [10]. The warm-to-cool is the most commonly used color scheme in heat visualization with the warm colors showing high-value data points and cool colors showing low-value datapoints. Below Figure5 represents the daily precipitation distribution, grouped by month, and recorded over eleven years in Seattle, Washington.
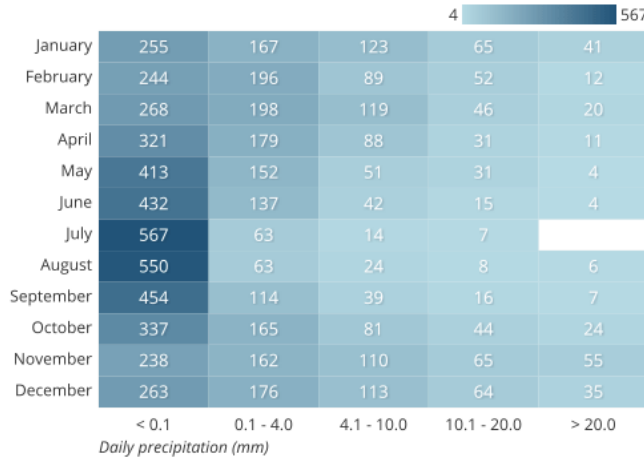
Seattle precipitation by month, 1998-2018

Fig 5: Heat Map [11]

III.   VISUALIZING BIG DATA TECHNIQUES

Data is generated very exponentially everyday**. This huge amount of generated data is known as Big Data. To visualize large data, new and more** interactive data visualization techniques **are proposed [12]. Following are some techniques to visualize large data:**

1.   Box Plot

Box Plot is used to analyze or display the distribution of a large amount of data. It is a graphical way of displaying the distribution of a set of data depends on the five features - the minimum, lower quartile, upper quartile, median, and maximum [13]. Median (50th percentile) represents the middle line that breaks the box into parts. Lower quartile (25th percentile) represents the lower boundary of the box. Upper quartile (75th percentile) represents the upper boundary of the box. Below Figure6 represents the Box Plot.
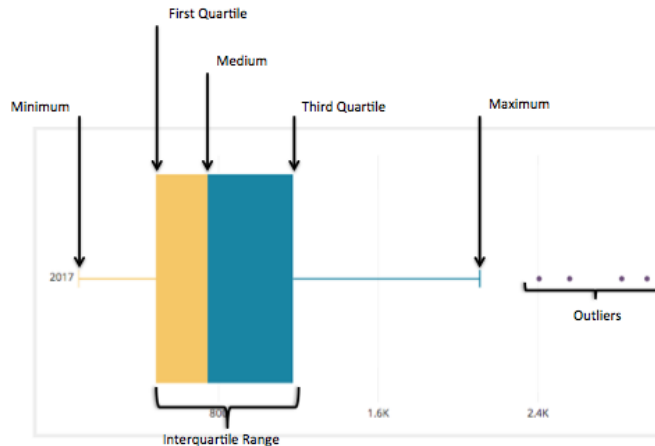

Fig 6: Box Plot [14]

2.   Word Cloud

Word Cloud visualization represents how words frequently appear in the given body of text with its relative size in the cloud. This technique is used to show low-frequency and high-frequency words as shown in Figure7. It is an
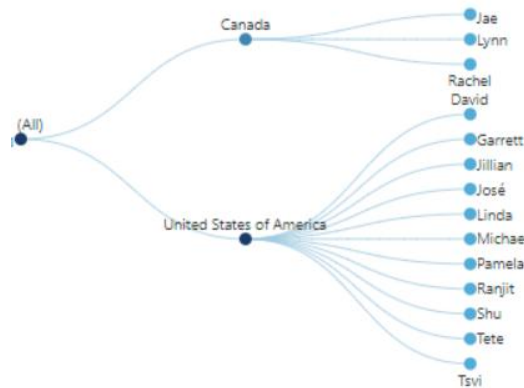
21

effective way to understand the dataset in view of the most popular words. Word cloud is also known as tag cloud or text cloud.

Fig 7: Word Cloud [15]

3. Network Diagram

Network Diagram is also known as *Network Map, Network Graph, and Node-Link Diagram.* This type of visualization displays the interconnections between a set of entities. The entities are called nodes or vertices and the connections between vertices are called links or edges. They are used in many applications, for example, mapping product sales across geographic areas or for analysis of social networks. Below Figure8 shows network visualization.

Fig 8: Network Diagram [13]

4. Tree Diagram

Tree Diagram is used for representing the data in a hierarchical structure or tree-like structure. It consists of a single root node that has no parent. Links or edges are used to connect the nodes with each other that show the relationship between nodes as shown in Figure9. The end nodes or leaf nodes have no child nodes. This type of visualization is used in businesses and organizations for managerial purposes [16].

Fig 9: Tree Diagram [17]

5. Matrix Visualization

Matrix is a visualization technique that allows mirroring the correlations between various constantly refreshing (steaming) data sets [18].

A correlation matrix is a table demonstrating the correlation coefficients between factors. This allows the identification of relationships between variables by linking the fast response times and big data. Every cell in the table shows the relationship between two variables as shown in Figure 10.
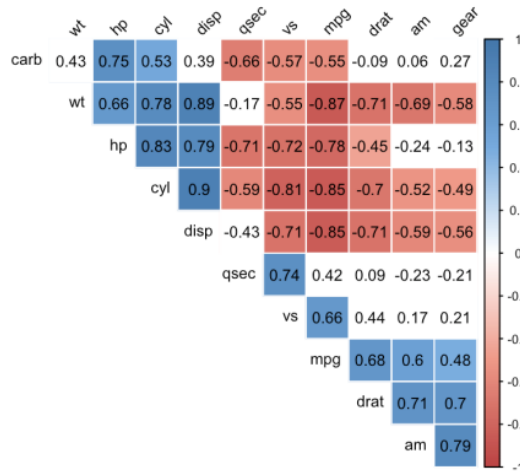

Fig 10: Correlation Matrix [19]

IV. DATA VISUALIZATION TOOLS

Data visualization tools are very helpful to transform every component of the data into interactive charts and pictures [20]. Depend on these interactive charts and pictures, the visualization of the generated charts and graphs is more effective than the raw data. Common tools used for visualization are as follows:

1. Tableau

It is the most powerful tool which is commonly used in the business intelligence industry. It helps in presenting raw data in the form of dashboards and worksheets by creating interactive graphs and charts. It supports a variety of data

file formats such as txt, xlsx, and csv. Extensive research has done into empowering Tableau to create graphical visualizations as precisely as possible and to make them easy for people to understand [21].

2. D3

D3 stands for **Data-Driven Documents, is the most well-known data visualization tool today.** It is built on JavaScript library and is an open-source tool that is used to create custom interactive data visualizations using SVG, HTML, and CSS [22]. By using this tool, we can create even the most complex charts and graphs.

3. Datawrapper

Data Wrapper is used for creating visualizations like maps, data tables and charts. It is easy to use tool which is increasingly popular among publishers, journalists, and media organizations. They often make use of datawrapper to embed live charts into their news articles [23].

4. Google charts

It is a powerful interactive data visualization tool that is easy to use and compatible with all platforms and browsers. Rendering of charts is purely based on HTML5/SVG [24].

5. Python-Based Tools

Data visualization with python helps the data analysts for the creation of interactive and highly customized graphs. Python libraries help users to visually represent different data sets and their relations. The following are some of the visualization libraries for python.

Bokeh is an interactive visualization library that concentrated on developing interactive charts and targets modern web browsers for presentation [25]. It is an open-source, versatile, and integrates highly with javascript.

Seaborn library produces a high-level interface for drawing attractive demographic graphics in Python. It is greatly integrated with PyData stack, including support for pandas and numpy data structures [26].

Altair is a declarative statistical visualization library in python and based on Vega and Vega-Lite visualization [25]. Because it contains a declarative library, it is necessary to mention the relationship between the data sets and encoding channels once, and the rest plotting is handled automatically.

Plotly is a javascript based library toolkit to form data visualizations. This library is very helpful to create some unique charts such as 3D charts, contour plots, and dendrograms [27]. Plotly provides one of the highest interactive interfaces to its users.

V.   COMPARISON TABLE

Techniques	Features
Line Chart	• Quick analysis of data • Used for displaying changes over periods of time.
Pie Chart	• Compare parts of a whole • Shows percentage of total for each category
Bar Chart	• Easy to prepare • Compare two or three data sets easily
Scatter Plot	• Show the relationship between two variables • Retains exact data values and sample size
Map	• Used to display and analyze the geographically related data • Overall trends show well
Box Plot	• Displays 5-point summary and outliers • Compares two or more data sets very easily

	• Easily handles extremely large data sets
Word Cloud	• Brand names pop and keywords float to the surface
	• Easy to create
	• Inaccurate
Network Diagram	• Show precedence well
	• Ability to calculate the critical path
	• Reveal interdependencies not shown in other techniques
Tree Diagram	• Represents data in a hierarchical structure
	• Tree diagrams force careful analysis of the problem
Correlation Matrix	• Used to summarize a huge amount of data where the target is to see patterns
	• Used as an input into a more advanced analysis.

VI.   CONCLUSION

Data Visualization is the technique of displaying more complex data in the form of graphics so that, it can be easily understandable. There are different data visualization techniques that are used to represent data in visual form for its better understanding and draw useful insights. In this paper, we have discussed the various visualization techniques for representing one dimensional as well as the new emerging multi-dimensional data i.e. big data. We will always need different tools and techniques for analyzing and interpreting this every day generated large and complex data. To promote the future, there will always be a need for better analysis using better techniques that can help the managers and the entrepreneurs to make big decisions in their fields. The coming decade will largely depend on the power and efficiency of the visualization methods and tools for getting valuable insights to solve present human problems and finding new business paths.

REFERENCES

1. B. Jena, "A Review on data visualisation tools Used for Big Data," *Int. Res. J. Eng. Technol.,* vol. 4, no. 1, pp. 492–495, 2017.
2. "Data Visualization: What it is and why we use it. | MicroStrategy." [Online]. Available: https://www.microstrategy.com/us/resources/introductory-guides/data-visualization-what-it-is-and-why-we-use-it. [Accessed: 15-May-2020].
3. D. S. Ebert, J. M. Favre, and R. Peikert, "Data visualization," *Comput. Graph.,* vol. 26, no. 2, pp. 207–208, 2002, doi: 10.1016/S0097-8493(02)00051-1.
4. F. Nazeer, N. Nazeer, and I. Akbar, "Data Visualization Techniques – A Survey," *Int. J. Res. Emerg. Sci. Technol.,* no. 4, pp. 4–8, 2017.
5. Z. Kaidi, "Data Visualization in the Geosciences," *Technometrics,* vol. 47, no. 3, pp. 382–382, 2005, doi: 10.1198/tech.2005.s311.
6. "MS Excel 2016: How to Create a Line Chart." [Online]. Available: https://www.techonthenet.com/excel/charts/line_chart2016.php. [Accessed: 04-May-2020].
7. "What social media platform do you use? – Social (?) Media." [Online]. Available: https://wp.wwu.edu/socialmedia18/findings/what-social-media-platform-do-you-use-2bql556/. [Accessed: 04-May-2020].
8. "A Complete Guide to Bar Charts | Tutorial by Chartio." [Online]. Available: https://chartio.com/learn/charts/bar-chart-complete-guide/. [Accessed: 02-May-2020].
9. "Scatter Plots | A Complete Guide to Scatter Plots." [Online]. Available: https://chartio.com/learn/charts/what-is-a-scatter-plot/. [Accessed: 04-May-2020].
10. "What is Heatmap Visualization? When & How to Use? [2020]." [Online]. Available: https://vwo.com/blog/heatmap-visualization/. [Accessed: 02-May-2020].
11. "A Complete Guide to Heatmaps | Tutorial by Chartio." [Online]. Available: https://chartio.com/learn/charts/heatmap-complete-guide/. [Accessed: 04-May-2020]. P. photo, "Title Data Visualization Techniques From Basics to Big Data With SAS ® Visual Analytics."

12. *"Best Data Visualization Techniques for small and large data." [Online]. Available: https://www.kdnuggets.com/2019/04/best-data-visualization-techniques.html. [Accessed: 31-Mar-2020].*
13. *"What Is a Box Plot and When to Use It | Tutorial by Chartio." [Online]. Available: https://chartio.com/resources/tutorials/what-is-a-box-plot/. [Accessed: 04-May-2020].*
14. *"Visualising Statistics: The importance of seeing not just describing data - Statistics Views." [Online]. Available: https://www.statisticsviews.com/details/feature/6314441/Visualising-Statistics-The-importance-of-seeing-not-just-describing-data.html. [Accessed: 01-Apr-2020].*
15. *"Tree Diagram - Learn about this chart and tools to create it." [Online]. Available: https://datavizcatalogue.com/methods/tree_diagram.html. [Accessed: 01-Apr-2020].*
16. *"Using a tree diagram | Data Visualizations | Documentation | Learning." [Online]. Available: https://www.dundas.com/support/learning/documentation/data-visualizations/using-a-tree-diagram. [Accessed: 01-Apr-2020].*
17. *"A QUICK INTRO OF BIG DATA VISUALIZATION TECHNIQUES AND TOOLS." [Online]. Available: https://datascience.foundation/sciencewhitepaper/a-quick-intro-of-big-data-visualization-techniques-and-tools. [Accessed: 02-May-2020].*
18. *"Visualize correlation matrix using correlogram - Easy Guides - Wiki - STHDA." [Online]. Available: http://www.sthda.com/english/wiki/visualize-correlation-matrix-using-correlogram. [Accessed: 02-May-2020].*
19. *J. Liu, T. Tang, W. Wang, B. Xu, X. Kong, and F. Xia, "A Survey of Scholarly Data Visualization," IEEE Access, vol. 6, no. c, pp. 19205–19221, 2018, doi: 10.1109/ACCESS.2018.2815030.*
20. *"The 7 Best Data Visualization Tools Available Today." [Online]. Available: https://www.forbes.com/sites/bernardmarr/2017/07/20/the-7-best-data-visualization-tools-in-2017/#60c27a426c30. [Accessed: 01-Apr-2020].*
21. *"Review of 20 best big data visualization tools." [Online]. Available: https://bigdata-madesimple.com/review-of-20-best-big-data-visualization-tools/. [Accessed: 01-Apr-2020].*
22. *"13 Free Data Visualization Tools." [Online]. Available: https://www.webfx.com/blog/web-design/free-data-visualization-tools/. [Accessed: 01-Apr-2020].*
23. *"15 Most Popular Data Visualization Tools." [Online]. Available: https://www.softwaretestinghelp.com/data-visualization-tools/. [Accessed: 02-May-2020].*
24. *S. K. A. Fahad and A. E. Yahya, "Big Data Visualization: Allotting by R and Python with GUI Tools," 2018 Int. Conf. Smart Comput. Electron. Enterp. ICSCEE 2018, no. July, 2018, doi: 10.1109/ICSCEE.2018.8538413.*
25. *"Top 5 Python Libraries For Data Visualization." [Online]. Available: https://analyticsindiamag.com/top-5-python-libraries-for-data-visualization/. [Accessed: 03-May-2020].*
26. *"6 Best Python Data Visualization Tools." [Online]. Available: https://www.knowledgenile.com/blogs/python-data-visualization-tools/. [Accessed: 03-May-2020].*