

Visualization Tools and Techniques in Big Data

Aashlesha Bhingarde¹, Prof. Deepali Vora²

¹Dept. of IT Engineering, Vidyalankar Institute of Technology, Maharashtra, India

²Dept. of IT Engineering, Vidyalankar Institute of Technology, Maharashtra, India

Abstract - Day by day, data is rapidly increasing at a faster rate in every field. Data which are very large in size and cannot be processed using traditional database systems are known as Big data. In the era of Big Data, a great attention deserves the visualization of large data sets. The value of data becomes an important factor in every aspect. The Data exploration and visualization systems play a vital role in the Big Data era. As the With availability of enough visualization techniques it can be very confusing to know what and when should be appropriate technique to use in order to convey maximum possible understanding. Different available visualization techniques are use for different situation which convey different level of understanding. This document discusses about big data Visualization, its challenges and tools and techniques.

Key Words: Visualization, Big data visualization, Big data visualization techniques, visualization tools.

1. INTRODUCTION

The information is the essential part of human life now, which encourage new ways to think and evolve new advancement. As data is raw fact are gathered together through different fields which requires immediate need for the new generation of different tools and techniques which will help to derive useful and authentic knowledge from the fast growing data volumes and these techniques forms the subject matter of the newly independent knowledge discovery field in database (KDD). Hence, the need for Big data visualization. In section first two of this we discussed what is Big Data Visualization and why we need, Importance of Big Data Visualization, Its challenges, features and the uses of it. In section three we focused on visualization techniques in the discovery of knowledge and how they can be applied. Section four contains description about tools. Section five contains comparative study about Visualization Tools which carrying out a research about most commonly used platforms and analyze their main features which will help to choose one or several platforms. Section six is about discussion that has properties of visualization techniques that can handle in big data and this was represented in a table Also. In section seven we concluded our study.

1.1 Visualization

Visualization is a powerful technique for refining your self-image and making important changes in your life. The term involves placing yourself in a trance state and imagining positive scenes. It is a graphical representation that best conveys the complicated ideas clearly, precisely, and efficiently.

2. BIG DATA VISUALIZATION

Big Data visualization term involves the presentation of data of almost any type in a graphical format that makes it easy to understand and interpret. Today's enterprises collect and store vast amounts of data that would take years for a human to read, and let alone to understand. But researchers have determined that human retina can transmit data to the brain at a rate of about 10 megabits per second. Big Data visualization relies on powerful computer systems and to ingest raw corporate data and process it to generate graphical representations that allow humans to take in and understand vast amounts of data in seconds.

One of the most important benefits of data visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Also Well designed data graphics are usually the simplest and at the same time, the most powerful.

2.1 Challenges of Big Data Visualization

Big Data visualization can be an extremely powerful business capability, but before an organization can take advantage of it some key issues need to be addressed. These include:

- **Availability of visualization specialists:** Many Big Data visualization tools are designed to be easy enough for anyone in an organization to use, often suggesting appropriate Big Data visualization examples for the data sets under analysis. But to get the most out of some tools it may be necessary to employ a specialist in big data visualization techniques who can select the best data sets and visualization styles to ensure the data is exploited to the maximum. Which deals with **Volume**-the methods are developed to work with large no of datasets.[2]
- **Visualization hardware resources:** Under the hood, Big Data visualization is essentially a computing task, and the ability to carry out this task quickly – to enable organizations to make decisions in a timely manner using real-time data – may require powerful computer hardware, fast storage systems, or even a move to cloud. That means Big Data visualization initiatives are as much an IT project as a management project. Which deals with **Velocity**- methods, business can replace batch processing with real time processing.
- **Data quality:** The insights that can be drawn from Big Data visualization are only as accurate as the data that is being visualized: if it is inaccurate or out

of date then the value of any insights is questionable. That means people and processes need to be put in place to manage corporate data, metadata, data sources, and any transformations or data cleaning that are performed before storage. Which deals with **Variety**- the methods developed to combine as many data sources.[2]

2.2 Problems and Solution for big data visualization

- Visual noise: Most of the objects in dataset are too relative to each other. Users cannot divide them as separate objects on the screen.
- Information loss: Reduction of visible data sets can be used, but leads to information loss.
- Large image perception: Data visualization methods are not only limited by aspect ratio and resolution of device, but also by physical perception limits.
- High rate of image change: Users observe data and cannot react to the number of data change or its intensity on display.
- High performance requirements: It can be hardly noticed in static visualization because of lower visualization speed requirements--high performance requirement.[1]

Potential solutions to some challenges or problems about visualization and big data were presented [1]

- Meeting the need for speed: One possible solution is hardware. Increased memory and powerful parallel processing can be used. Another method is putting data in-memory but using a grid computing approach, where many machines are used.
- Understanding the data: One solution is to have the proper domain expertise in place.
- Addressing data quality: It is necessary to ensure the data is clean through the process of data governance or information management.
- Displaying meaningful results: One way is to cluster data into a higher-level view where smaller groups of data are visible and the data can be effectively visualized.
- Dealing with outliers: Possible solutions are to remove the outliers from the data or create a separate chart for the outliers.

3. VISUALIZATION TECHNIQUES IN BIG DATA

Line Chart

A line chart that displays the relationship between each variable on the chart. These charts are frequently used to make comparison between lots of items at the same time. The stacking lines are being used to also make comparison between the trends for multiple variables in Line chart. One can decide to make use of the line charts when a variable change needs to be displayed.

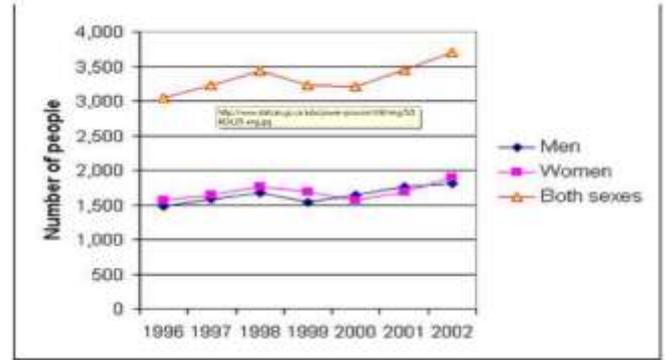


Figure.1. Simple Line chart [11]

Pie charts

It is as well-known as a circle graph. A pie chart shows information statistics and also data in a way that is not difficult to read called "pie-slice" form and the various sizes of slice shows how much of an element is in existence. When the slice is big enough, then it shows of the data was gathered.

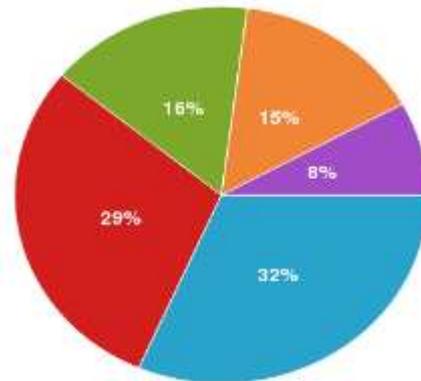


Figure.2. Simple Pie chart [11]

Bar chart

Is as well referred to as column chart and bar charts used for comparison of items of different groups. Bars are used to represent the various values of a group and charts makes use of both horizontal bars and vertical bars . When the values to be represented are clearly different and such differences in the bar charts are been seen by human eye, then only one can decide to make use of a bar chart.

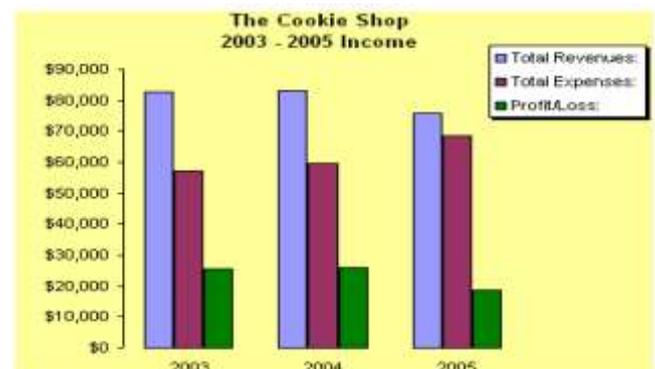


Figure.3. simple Bar chart [11]

Table

Table is the arrangement of data using rows and column. To conducting research and analysis of data, the role of table is very important. Tables are simple to understand and analyze and its simple to interpret the method of data representation.

Expenditure	Homeowners			Renters		
	1986	2010	Percent change	1986	2010	Percent change
Annual expenditures	\$50,050	\$50,790	0	\$33,524	\$33,460	0
Food	7,907	6,820	-14	5,169	4,802	-7
Food at home	4,380	4,000	-13	2,972	2,902	-2
Food away from home	3,527	2,820	-15	2,196	1,900	-13
Housing	16,637	18,503	11	11,038	12,643	16
Apparel and apparel services	3,030	1,781	-41	2,111	1,544	-27
Transportation	11,809	9,058	-22	8,408	5,048	-21
Gasoline and motor oil	2,147	2,458	15	1,285	1,511	18
Healthcare	2,837	4,016	42	1,313	1,518	16
Health insurance	943	2,314	145	400	909	124
Entertainment	2,775	3,068	11	1,488	1,590	8
Personal insurance and pensions	5,388	6,665	24	2,374	2,907	22

NOTE: For comparison, 1986 dollars were converted to 2010 dollars.
SOURCE: U.S. Bureau of Labor Statistics.

Figure.4. Simple Table [11]

Scatter Plot

A scatter plot is type of plot which described as a 2-dimensional plot which displays the joint variation of two data items. It is also called a scatter chart, scatter diagram, scatter graph. For this, observations are being represented by each marker and the marker position usually indicate value for the observations. scatter plot shows the data in Cartesian coordinate in a graphical display which displays the relationship that exist between two variables in which one is represented as a vertical distance and the other as horizontal distance.

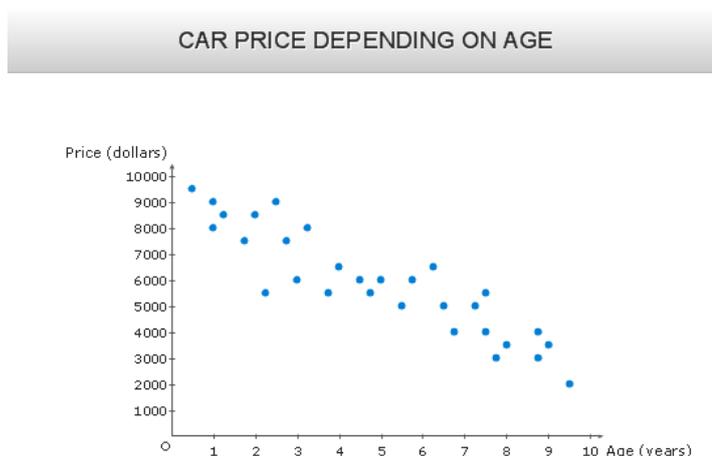


Fig.5. Basic scatter plot of two variables[11]

Bubble Chart

A bubble plot is some degree of difference of a scatter plot ,the markers in it are being substituted with bubbles and this is possible only we have a set of data points which has three values contained in each data item It shows the relationship

that exists between the minimum of three variables.[11] Two of them gets represented by the plot axes i.e. x-axis and y-axis, while the third one axis is by the bubble size and each bubble is a representation of an observation.



Figure.6. Showing a simple bubble plot [11]

Parallel Coordinates

The parallel coordinate technique which makes use of the concept of networking a multi-dimensional point to some axes and all of these are parallel to each other. In these technique, single data elements are being plotted across many dimensions and these dimensions are connected unto a y-axis and each object of the data is shown along the axes as a series of connected points.[11] The parallel coordinate is important if you want to show a multidimensional data and a lot of these dimensions are being organized and expanded by this technique.

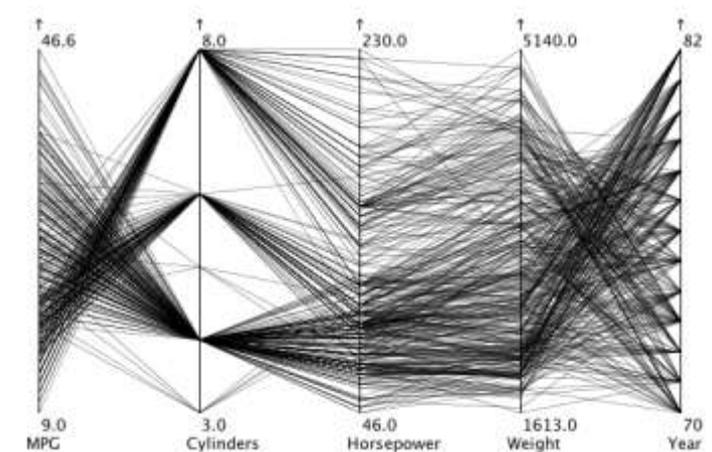


Figure.7. Showing a parallel coordinate [11]

Symbol maps

These are maps with symbols on them. The symbols differ in size, which makes easy to compare them. Just Imagine a US manufacturer who has launched a new brand recently and the manufacturer is interested to know which regions liked the brand particularly.



Figure.9. Showing a Symbol Maps[10]

Connectivity charts

Connectivity charts Show the links between events or phenomena.

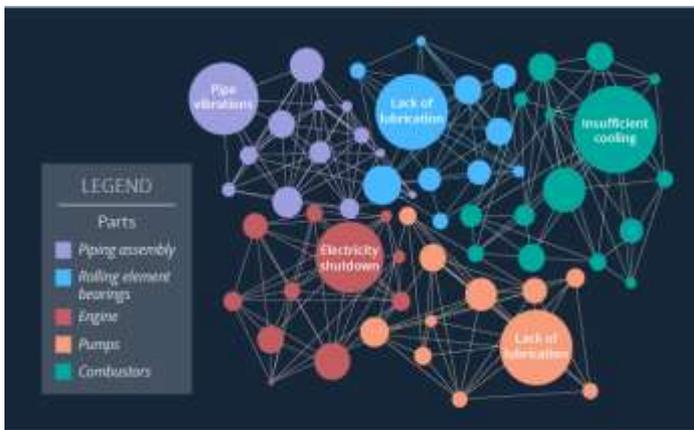


Figure.10. Showing a Connectivity Charts[10]

Word clouds

In this word clouds give companies the idea of how often a word is used. The words are used in the cloud are of different sizes. The more the size of word– the higher the frequency. Word cloud technique may be helpful, for example, for sentiment analysis of the customers’ social media posts.



Figure.11. Showing a Word Cloud[10]

4. VISUALIZATION TOOLS IN BIG DATA

Qlik View

It is a solution that focuses on the user as the receiver of data which allows users to explore and discover your data in a workflow similar to the way developers work when processing data. The software strives to maintain the association between data.



Figure .12. Qlik View Tool[7]

Klipfolio

Klipfolio is a BI solution that resides 100% in the cloud (for that no desktop application is required) providing a genuinely insightful tool for data visualization and also for dashboard composition. This enables to process data most efficiently, boosts the real-time solutions and optimization, rather than the relying on a periodical regression model. In this way, To summarize, Klipfolio you can create sheets with any type of visualization, but you need to think about how you can combine and change the data first.



Figure .13. Klipfolio Tool[7]

Tableau

In summary, Tableau features, its most convenient use case is the discovery of structured data through charts, graphs, and other visualization types. No other BI solution make it easier than Tableau, giving amazing power to its users. One of the important tool and most usable data visualization tool in big data environment is Tableau .



5. CONCLUSIONS

Big data is large in volume, variety, velocity, value, and veracity. Data value is closely related to the data volume and variety of the data. Data visualization may not be an exact solution for analyzing the large volume of data, where they need to carry preprocess like proper extraction of data from variety of data sources. The use of the data visualization techniques used in big data could be interesting and at times challenging as well, it all depends on how effective you put it to use but for you to be able to choose the best visualization technique to display your data effectively, you must first of all understand the data want to visualize with its size and cardinality Also you should determine what you are trying to visualize and the type of information to be communicated.

Due to economic and infrastructural constraints every organizations are not able to buy all the domains required for analyzing the data. Therefore to fulfill the requirement of advanced tools and technologies organizations are used open source data visualization tools for analyze the big data.

We discussed about different data visualization techniques and different data visualization tools used for representing the analyzed results of big data. Also we talked about big data visualization basics, challenges, features, problems and its solution, comparative study of frequently used visualization tools which will benefit future researchers of big data visualization to continue their work.

6. REFERENCES

[1] Sagargouda S. Patil, "Overview of Big Data Visualization", 1st International Conference on Innovations in Computing & Networking (ICICN16), CSE, RRCE.

[2] Bibhudutta Jena, "A Review on Big Data Challenges and Techniques Used to Overcome from IT", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391, Volume 6 Issue 1, January 2017.

[3] Muzammil Khan, "Data and Information Methods, and Interactive Mechanisms: A Survey", International Journal of Computer Applications (0975 - 8887), Volume 34- No.1, November 2011.

[4] Leishi Zhang, Andreas Stoffel, Michael Behrisch, "Visual Analytics for the Big Data Era - A Comparative Review of State-of-the-Art Commercial Systems", IEEE Conference on Visual Analytics Science and Technology 2012 October 14 - 19, Seattle, WA, USA 978-1-4673-4753-2/12/\$31.00 ©2012 IEEE.

[5] R.S. Raghav, Sujatha Pothula, T. Vengattaraman, Dhavachelvan Ponnuram, "A Survey Of Data Visualization Tools For Analyzing Large Volume Of Data In Big Data Platform".

Power BI

Power BI is a business analytics service provided by Microsoft which provides interactive visualizations with self-service business intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on information technology staff or database admin. Power BI is the software solution, which is developed and supported by Microsoft, for business intelligence and analytics needs.

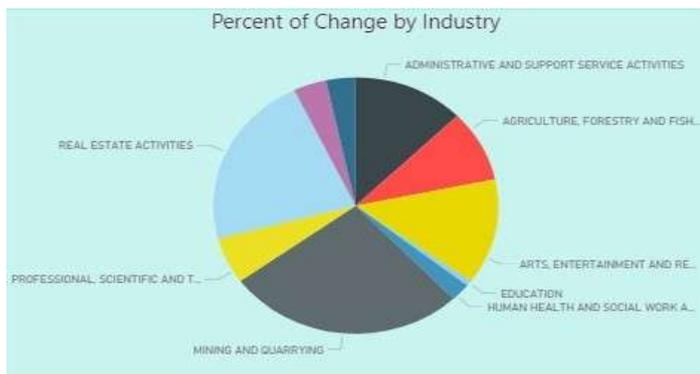


Figure.15. Power BI Tool [7]

Google Data Studio

The youngest tool on our list today is the part of Google's analytics solutions—Google Data Studio. Being relatively new to this field, it strives to take its position among many competitors via ease of usage, simple yet beautiful design, innovative problem-solving, and straightforward, habitual ways to share dashboards



Figure.16. Google Data Studio Tool [7]

[6] BibhuduttaJena, "A Review on data visualization tools Used for Big Data", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056, Volume: 04 Issue: 01 | Jan -2017.

[7] MilošKrstajić, Daniel A. Keim , "Visualization of Streaming Data: Observing Change and Context in Information Visualization Techniques".

[8] Big Data Visualization Tools in 2018, "https://medium.com/activewizards-machine-learning-company/a-comparative-analysis-of-top-6-bi-and-data-visualization-tools-in-2018-658490665973", accessed on 28th September 2018

[9] Enrico Giacinto Caldarola^{1,2} and Antonio Maria Rinaldi^{1,3}, "Big Data Visualization Tools: A Survey The new paradigms, methodologies and tools for large datasets visualization", <https://www.researchgate.net/publication/318679685>.

[10] J.Jabanjalin Hilda and C.Srimathi , "A Survey on Data Visualization Applications Challenges Techniques and Technologies" , Asian Journal of Computer Science and Technology ISSN 2249-0701 Vol. 4 No. 1, 2015, pp.29-33.

[11] Big Data Visualization techniques, "https://www.scnsoft.com/blog/big-data-visualization-techniques", accessed on 13th June 2018.

[12] Samuel Soma Ajibade¹, Anthonia Adediran², "An Overview of Big Data Visualization Techniques in Data Mining", International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 4, Issue 3, pp: (105-113), Month: July - September 2016

[13] Enrico GiacintoCaldarola and Antonio Maria Rinaldi, "Big Data Visualization Tools: A Survey -The new paradigms, methodologies and tools for large data sets visualization", <https://www.researchgate.net/publication/318679685>.

[14] ArockiaPanimalar.S¹, Komal M.Khule², Karthika.S³, Nirmala Kumari.T⁴, "Data Visualization Tools and Techniques For Datasets In Big Data", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 04 Issue: 08 | Aug -2017.