A Review on Big Data Visualization Tools

DR D. SRIDEVI[1], DR A. KUMARAVEL[2], DR S. GUNASEKARAN[3]
[1]Assistant Professor, SRM Valliammai Engineering College
[2]Professor, Bharath Institute of Higher Education
[3]Professor, Gandhigram Rural Institute

Abstract- Big Data is tremendously large data sets that may be evaluated computationally to reveal patterns, trends, and associations, specially relating to human behavior and interactions. Simply large volume of data analyzed and the visuals can vary based on the analyzing aspect in related to their dimensions. The tremendous increase in volume of data day by day makes data visualization an essential part of research. Recently more number of tools is available for data visualizing process. In this paper, a detailed study about big data, data visualization, techniques and tools are discussed.

Index Terms— Big Data, Big Data Analytics, Data Visualization, Visual Analytics,

I. INTRODUCTION

The rate of data growth over years is amazing: according to Science Daily, a full 90% of all the data in the world has been generated over the last two years [1]. All of this represents a real tsunami and requires a paradigmatic shift respect to the past as for theories, technologies or approaches in data management and more attention to survive it [2].

In the Big Data, scientists need not construct a complex model and explain its rules through complex logic-based languages. However the only need is to tune statistical analysis or machine learning techniques to get more insights from them, and very quickly. A study by the Economic Times suggests that large organizations using Big Data analytics outperform competitors, who do not utilize this [3].

The challenges that the Big Data imperative [4] imposes to data management severely impact on data visualization.

The "bigness" of large data sets and their complexity in term of heterogeneity contribute to complicate the representation of data [5] [6], making the drawing algorithms quite complex.
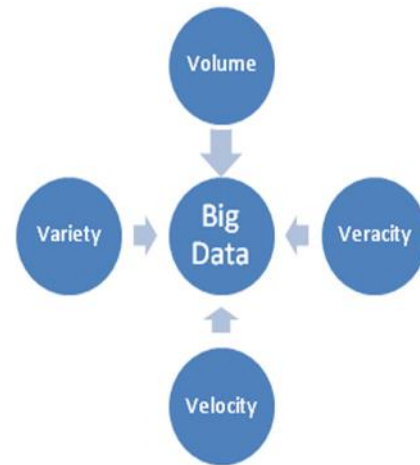

Figure : Big Data

The Infographics become Infonomic, a composite term between the term Information and Economics that wield information as a real asset, a real opportunity to make business and to discover the world [7].

The paper is structured as follows. The second section presents the techniques in visual data representation and the third section introduces the factors that influence data visualization choices. The fourth section describes about the types of data visualization whereas the fifth section describes about visualization tools for big data. Finally, the last section draws the conclusions, summarizing the major findings.

II. TECHNIQUES IN VISUAL DATA REPRESENTATION

Data can be represented in different forms like simple line diagrams, bar graphs, tables, matrices etc. Some visual techniques are generally used for presentation of data are as follows.

2.1 Isoline – It is basically a 2D data representation of a curved line that generally transfers constantly on the surface of the graph, the plotting of line generally drawn on the basis of data arrangement instead of data visualization [8].

2.2 Isosurface – It is a 3D representation of an Isoline. Isosurface are designed to present points that are bound by a constant value in a volume of space i.e. in a domain that covers 3D space [9].

2.3 DVR (Direct Volume Rendering) In this method we can obtain a 2D projection for a 3D data set and it is very clear and transparent in visualization process [10].

2.4 Streamline – It is a field that is generated from the description of velocity vector field of the data flow. It is now-a-days widely used in data visualization process [11].

2.5 Map: It is generally used to represent the location of different areas of a country. It is generally drawn on a plain surface. Google maps are widely used for data visualization. Now –a-days it is widely used for finding the location in different domains of country [12], [13].

2.6 Venn Diagram : Generally it is used to represent the logical relations that is existing in between finite collection of sets and Venn diagram generally are of many types i.e intersection of two sets and union of two sets etc [14]

In the above section the paper has described number of ways by which the data can be visualized which are in the form of 1D,2D or in 3D, hence in the below section the paper has described different types of data visualization.

III.   FACTORS THAT INFLUENCE DATA VISUALIZATION CHOICES

3.1 Audience. It's important to adjust data representation to the target audience. If it's end customers who browse through their progress in a fitness app, then simplicity is the key. On the other hand, if data insights are intended for researchers or experienced decision-makers, you can and often should go beyond simple charts.

3.2 Content. The type of data determines the tactics. For example, if it's metrics that changes over time, you most probably will use line charts to show the dynamics. To show the relationship between two elements you will use a scatter plot. In turn, bar charts are perfect for comparison analysis.

3.3 Context. You may use different approaches to the way your graphs look and therefore read depending on the context. To emphasize a certain figure, for example serious profit growth compared to other years, you may want to use the shades of one color and pick the bright one for the most significant element on the chart. On the contrary, to differentiate elements, you'll use contrast colors.

3.4 Dynamics. There are various types of data, and each of it implies a different rate of change. For example, financial results can be measured monthly or yearly, while time series and tracking data is constantly changing. Depending on the type of change, you may consider dynamic representation (steaming) or a static visualization.

3.5   Purpose. The goal of data visualization also has serious influence on the way it is implemented. In order to make a complex analysis of a system or combine different types of data for a more profound view, visualizations are compiled into dashboards with controls and filters. However, dashboards are not necessary to show a single or occasional data insight.[15]

IV.   TYPES OF DATA VISUALIZATION

Data visualization can be done in different ways that the paper has described in the above in this section various data visualization techniques will be discussed.

4.1 Multidimensional data visualization- In this approach numerous data are generally used to represent the data. Generally pie charts, histograms, bar charts etc are generally used to multidimensional data visualization. Google charts tool is used to create multidimensional data visualization.

4.2 Network Data Visualization- This approach is generally used to represent the relations that are too complex in the form of hierarchies. Some of the basic tools used for network data visualization are hive plot, Pajek, Gephi, NodeXL etc.

4.3 2D/3D standard figure- This may be implemented as bars, line graphs, various charts, etc. The main drawback of this type is the complexity of the acceptable visualization for complicated data structures [16].

4.4 Geometric transformations- This technique represents information as scatter diagram. This type is geared towards a multi-dimensional data set's transformation in order to display it in Cartesian and non-Cartesian geometric spaces. This class includes methods of mathematical statistics [17].

4.5 Display icons- Ruled shapes (needle icons) and star icons. Basically, this type displays the values of elements of multidimensional data in properties of images. Such images may include human faces, arrows, stars, etc. Images can be grouped together for holistic analysis. The result of the visualization is a texture pattern, which varies according to the specific characteristics of the data [18].

4.6 Methods focused on the pixels- Recursive templates and cyclic segments. The main idea is to display the values in each dimension into the colored pixel and to merge some of them according to specific measurements Since one pixel is used to display a single value, therefore visualization of large amounts of data can be reachable with this methodology [19].

4.7 Hierarchical images- Tree maps and overlay measurements .These type methods are used with the hierarchical structured data. [20].

However there are number of data visualization techniques are present. These are the some of the techniques described in the above.

V. VISUALIZATION TOOLS FOR BIG DATA

5.1 Dygraphs
Dygraphs is a fast, versatile open source JavaScript charting library .It is highly personalized and designed to interpret dense data sets, it works in all browsers and it can be zoomed on mobile and tablet devices. Few characteristics of Dygraphs they are used to handle huge datasets, highly customizable; highly compatible and they give strong support for error bar or confidence interval. Dygraph chart is displayed in figure 1 [21].
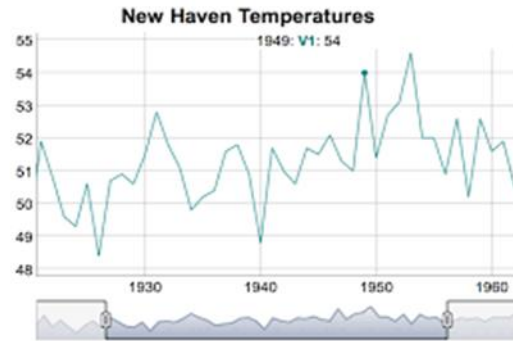

Figure 1 Dygraphs

5.2 ZingChart
ZingChart is a powerful charting library and they have ability to create charts dashboards and infographics. It is featured with -rich API set that allow user to built interactive Flash or HTML5 charts. It provide hundreds of chart variation and many methods For Example Bar, Scatter, Radar, Piano, Gauge, Sparkline, Mixed, Rank flow and word cloud. Figure 2 shows Zing chart.[22]

Figure 2 ZingChart

5.3. Polymaps
Polymaps is a free java script charting library for image and vector- tiled maps using Scable Vector Graphics (SVG).They provide dynamic and interactive maps in web browsers. Complex data sets can be visualized using polymaps and offers multi-zoom functionality. The characteristics of polymaps are it uses Scalable Vector Graphics (SVG) and the

Basic CSS rules are used and its imagery in spherical Mercator tile format. Figure 3 shows the layout of Polymaps. [22]

Figure 3 Polymaps

5.4. Timeline

Timeline is a different tool which delivers an effective and interactive timeline that responds to the user's mouse, it delivers lot of information in a compressed space.

Figure 4 Timeline

Each element can be clicked to reveal more in-depth information; it gives a big-picture view with full detail. Timeline is demonstrated in figure 4. [22]

5.6 Modest Maps

Modest Maps is a lightweight, simple mapping data visualization tool for web designers.

Figure 5 Modest Maps

It has highest performance and compatibility with new technology and has well designed codes which are tested and deployed widely. Modest Map is given in figure 5 [22]. There are various tools are present for big data visualization and present the best analysis result but in the paper some of them are discussed in the above section.

VI.   CONCLUSION

To implement Big data analytics needs   improvised and advanced tools and methods. All the organizations are unable to buy the domains which are important for data analysis due to economic and infrastructural constraints. So in order to solve the requirement of advanced tools and technologies, open source data visualization tools are used for analyzing the big data. The tools are more secure and provide the exact results after analysis of big data.

REFERENCES

[1] Dragland, A. (2013). Big data? For better or worse. ˚ ScienceDaily.

[2] Caldarola, E. G., Sacco, M., and Terkaj, W. (2014). Big data: The current wave front of the tsunami. ACS Applied Computer Science, 10(4):7–18.

[3] Bhanu, S. (2013). Companies adopting big data analytics to deal with challenges. The Economic Times.

[4] Caldarola, E. G., Picariello, A., and Castelluccia, D. (2015a). Modern enterprises in the bubble: Why big data matters. ACM SIGSOFT Software Engineering Notes, 40(1):1–4

[5] Caldarola, E. G. and Rinaldi, A. M. (2016). Improving the visualization of wordnet large lexical database through semantic tag clouds. In

Big Data (BigData Congress), 2016 IEEE International Congress on, pages 34–41. IEEE.

[6] Caldarola, E. G., Picariello, A., and Rinaldi, A. M. (2015b). Experiences in wordnet visualization with labeled graph databases. In International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management, pages 80–99. Springer

[7] Caldarola, E. G. and Rinaldi, A. M. (2016). Improving the visualization of wordnet large lexical database through semantic tag clouds. In Big Data (BigData Congress), 2016 IEEE International Congress on, pages 34–41. IEEE

[8] F. Mu, W. Xue, J. W. Shu and W. M. Zheng, "Metadata management mechanism based on route directory", Journal of Tsinghua University (Sci & Tech), vol. 49, no. 8, (2009), pp. 1229-1232.

[9] S. Yu, X. L. Gui, R. W. Huang and W. Zhuang, "Improving the storage efficiency of small files in cloud storage, Journal of Xi'An Jiao Tong University, vol. 45, no. 6, (2011), pp. 59-63.

[10] Bibhudutta Jena"A review on Big Data Challenges and Techniques Used to Overcome from it" International Journal of Science and Research (IJSR), Volume 6 Issue 1, January 2017, 201 – 204.

[11] Bibhudutta Jena"A review on Joining Approaches in Hadoop Framework and Skewness Associate to it " Vol. 2 - Issue 6 (Nov - Dec 2016), International Journal of Engineering and Techniques (IJET) , ISSN: 2395 – 1303.

[12] Richard K. Lomotey and Ralph Deters , " Terms Mining in Document-Based NoSQL: Response to Unstructured Data " , 2014 IEEE International Congress on Big Data.

[13] S. Ghemawat, H. Gobioff and S. T. Leung, "The Google file system", Proceedings of the 19th ACM symposium on Operating systems principles, Bolton Landing, NY, USA, (2003) October 19–22, pp. 29- 43.

[14] S. Yu, X. L. Gui, R. W. Huang and W. Zhuang, "Improving the storage efficiency of small files in cloud storage, Journal of Xi'An Jiao Tong University, vol. 45, no. 6, (2011), pp. 59-63.

[15] https://www.digiteum.com/data-visualization-techniques-tools

[16] Tory M, Kirkpatrick AE, Atkins MS, Moller T. Visualization task performance with 2D, 3D, and combination displays. IEEE Trans Visual Comp Graph. 2006; 12(1):2–13.

[17] Stanley R, Oliveria M, Zaiane OR. Geometric data transformation for privacy preserving clustering. Departament of Computing Science; 2003.

[18] Healey CG, Enns JT. Large datasets at a glance: combining textures and colors in scientific visualization. IEEE Trans Visual Comp Graph. 1999;5(2):145–67.

[19] Keim DA. Designing pixel-oriented visualization techniques: theory and applications. IEEE Trans Visual Comp Graph. 2000;6(1):59–78.

[20] Kamel M, Camphilho A. Hierarchic Image Classification Visualization. In: Proceedings of Image Analysis and Recognition 10th International Conference, ICIAR; 2013.

[21] http://www.creativebloq.com/design-tools/data-visualization-712402