Visualization for Data Analytics and Data Science

Raghunandan Alugubelli,

University of South Florida.

Abstract

We are living in a world that generates large amounts of data daily. Corporations are making data-driven decisions basing on their data and data visualization is an essential part of it. Organizations store essential data for a better competitive position and save it in future and decision-making processes. We have various tools available in the market- some of them are Sisense, Tableau desktop, Whatagraph, and Hubspot. These tools assist in making data presentation easy. This paper focuses on the importance of data visualization, primary tools, and software for data visualization, and theoretical architectural framework for data visualization. Finally, the article will focus on the critical challenges faced by data visualization and steps for mitigation.

Keywords: *Data Visualization, Data Analytics, Architecture*

Importance of Data Visualization

Looking at the data by tables, datasets can be cumbersome. Visualization can help users grasp the relations in data quickly, which allows them to make data-driven decisions. Several data visualization tools can help filter the information so that users can look at the data dynamically. On the other hand, traditional visualization practices, such as graphs, column charts, and tables, have been replaced by 3D visualization patterns in the modern world (Kot, Wuensche, Grundy & Hosking, 2005). Data visualization efficiency depends primarily on the data itself. Ingesting incorrect data into the visualization engine will result in erroneous, half-baked, or obsolete data visualization (Few & Edge, 2007).

Kumar & Kirthika (2017) states that with the increased demand for innovative data visualization, several standards and guidelines have also been put forth. In addition, research establishes that programming knowledge in a broader way can improve the data visualization process.

Data visualization and programming are related because most visualizations programs and patterns are programmed to attain the desired output. On the other hand, most programming knowledge has advantages but has left many gaps regarding its applications. Therefore, data visualization and programming languages are more feasible for data analytics than pure programming. Human beings process information and data using graphs or charts to visualize a considerable amount of data. It is easier to present information than poring over reports and spreadsheets. Furthermore, data visualization as a method of representing data is an easy way to convey the various concepts universally.
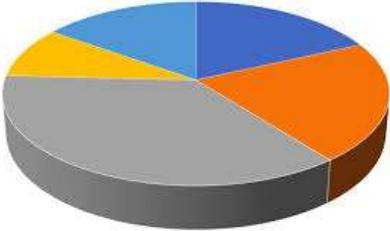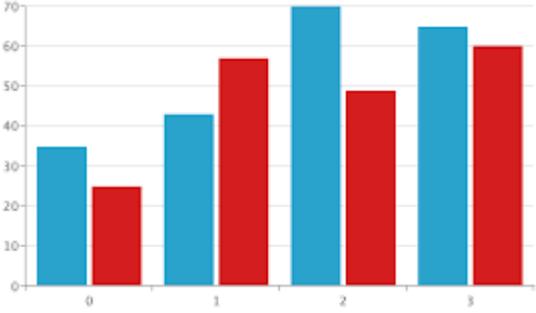
Data Visualization Architecture

The data visualization architecture is instrumental in enterprise business analytics. When a data architecture encompasses machine learning algorithms or advanced predictive analytics, visualizing the outputs becomes essential for monitoring results and ensuring that model performance is accurate. In addition, complex algorithm visualizations are usually easy to interpret than numerical outputs.
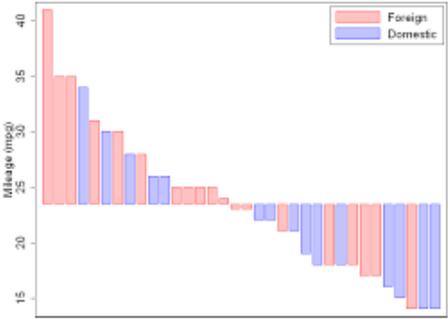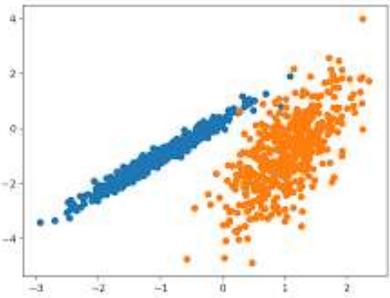
Data visualization cycle
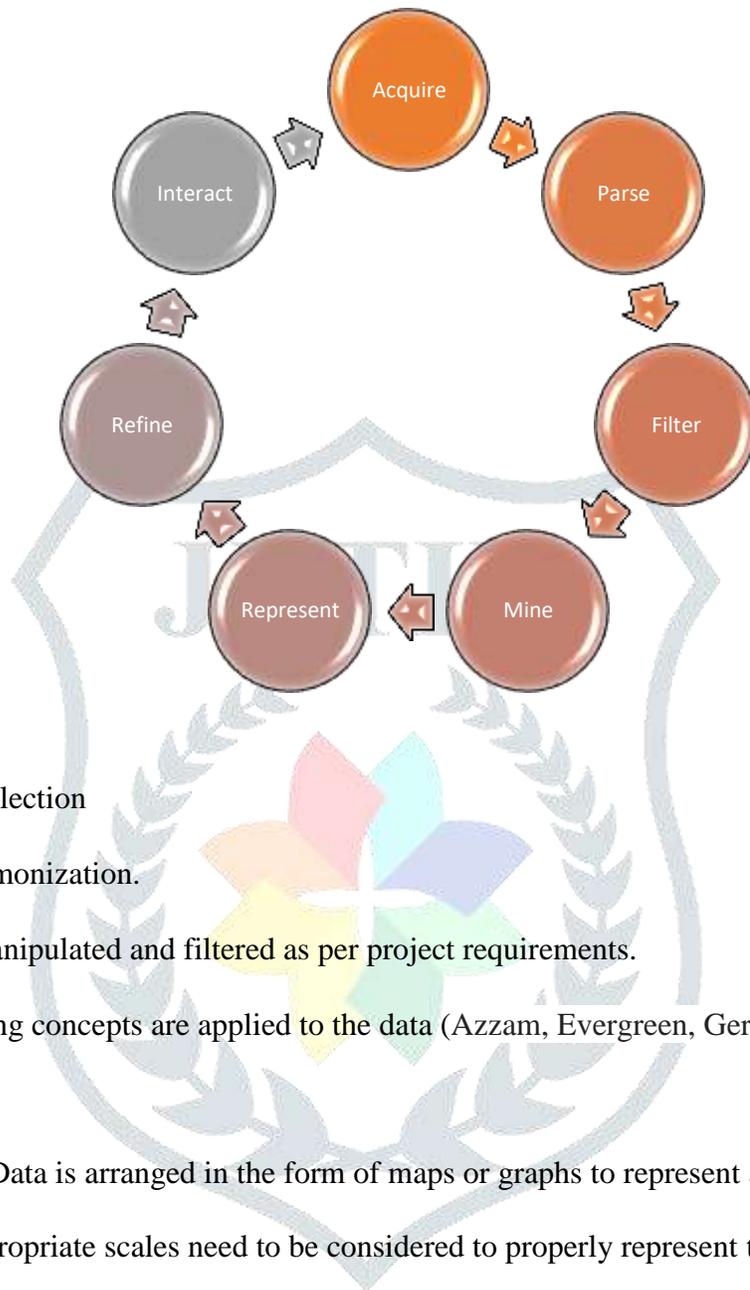
- Acquire-Data collection

- Parse – Data harmonization.

- Filter- Data is manipulated and filtered as per project requirements.

- Mine- Data mining concepts are applied to the data (Azzam, Evergreen, Germuth & Kistler, 2013).

- Representation: Data is arranged in the form of maps or graphs to represent analysis to the stakeholder. Appropriate scales need to be considered to properly represent the graphs with proper labels, as stakeholders must understand what has been analyzed with visualization forms.

- Refine: In this stage, the graphs are refined for better visualization & efficiency.

- Interact: Here, stakeholders interact with dashboards and visualizations.

Data Visualization types

Data Visualization types	Usefulness
Pie Chart	Slices of a pie represent the Pie chart, which is in a circular graph. The slices represent the categories and their respective values. The pie chart displays the whole proportion when they are very close to each other. Therefore, the pie chart can showcase sales, expenses and profits (Christensen, 2017).
t-SNE algorithm	The key benefits of t-SNE are the capability of preserving local structure. It indicates that points closer to each other in the high-dimensional data set will likely be closer to each other in the chart. It also generates attractive visualizations (Sadiku, M., Shadare, Musa, Akujuobi, & Perry, 2016).
Bar Chart	A bar chart can categorize the data by displaying it on the x-axis and the y-axis. Furthermore, each bar can have its colors representing the values (Keim, Hao, Dayal, & Hsu, 2002). In bar charts - total, average, and percentages can be easily

	showcased, and the values can be easily categorized(Theus, 2002).
Waterfall Plot	It assists in comprehending the cumulative impact of sequentially introduced negative or positive values.
Clustering Algorithm	It assists in generalizing the clusters of dissimilar sizes and shapes, for example, elliptical clusters.

Various Visualization Softwares

Names	Features
Hubspot	Most significantly, it assists in serving as search engine optimization. In addition, the Hubspot CRM suite is a significant provider of sales analytics.
Whatagraph	Whatagraph provides marketing analytics using social media data. It integrates data from various social media servers such as Facebook, Linkedin, Google, etc.
Tableau Desktop	Tableau Desktop is a reporting software that offers dynamic dashboards and reporting across several databases and data source systems. Tableau is one of the top companies in the space of data visualization. The Tableau server provides a secure way of dealing with the scale-up and the scales down

	multithreaded and multiprocess. Tableau can also use the other scripting language that enhances the Tableau server procedures.
Sisence	Sisence has embedded visualizations that interface business applications, analytics, and BI platforms.
Power BI	Power BI is an analytical tool from Microsoft Corporation. Power BI integrates with several databases, cloud systems. Currently in the top three by market capitalization in the area of data visualization software.
The Teammate	The Teammate is a relatively newer tool. Teammate provides services in the area of audit analytics.
Jupyter	Jupyter is a notebook that visualizes data in python. Jupyter utilizes multiple libraries for data manipulation and also for building dashboards. Jupyter is an open-source tool. One downside of the open-source tool is that they are prone to more attacks.
Google charts	Google charts give a vivid vision from the data(Varela-González et al. 2014). The Charts are readily available to draw. It distributes much information on a page efficiently.
Oracle Analytics Cloud	Oracle Analytics Cloud is a relatively new tool. It is from Oracle corporation.

Data Visualization Challenges and Improvements

The main challenge in evaluating the data visualization processes is the complexity of the process and the difficulty of replicating the conditions(Kumar, & Kirthika, 2017). Information visualization represents abstract data like text and numerals and different data types into a visual form that humans can understand. Another challenge lies in data importing and cleaning, visual representations combining textual labels, and accessing many data to achieve usability. These are some, but there might be more problems (Archambault, Helouvry, Strohl & Williams, 2015). These challenges can somehow be conquered by

techniques and applying techniques. Data importing and cleaning by viewing and maintaining data in a particular format will help to some extent efficiently process the data.

Moreover, by using labels and screen tips, one can overcome the challenges. Furthermore, while accessing a large amount of data by using data mining techniques, we can frame and show our data in whichever way we want. Many data mining techniques can apply to raw data and reduce formatting this data (Khan, & Khan, 2011).

Latest Developments in Data Visualization

With the development of science and technology, it is expected that significant changes can be seen when it comes to data visualization. Likewise, with the advancement in technology, data visualization has also seen significant developments. Today, we are witnessing the increasing popularity of 3D data visualization tools for geolocation information.

With the growth in sophisticated analytics tools and information sources, the business processes complexity has also increased. The demand for preventing these trends from turning out to be a hindrance paves the way for 3D visualization solutions' rapid adoption. Already, 3D visualization has proven to be particularly effective for businesses that combine classical analytics with monitoring of location and process geospatial information.

The 3D data visualization market, by the year 2025, is expected to develop to $6 billion, taking into account the interests in 3D data visualizations of both indoor and outdoor areas. Attempting to combine every layer comprising LiDAR (light detection and ranging) data, utility providers, government organizations, and companies are closely looking at more innovative solutions. One of those is virtual replicas, digital twins established with data technology. These are used to enhance environmental productivity and performance and investigate locations devoid of risking employees' well-being.

With traditional 2D data visualization, the issue is the restriction of the amount of data it can bear and the amount of data it can use in actuality. As a result, several insights remain hidden, untapped under

the pies, bars, and charts' layers. The display of data in 3-dimensions can expose geospatial information that can assist organizations in deriving insights and uncovering new business opportunities. It allows professionals and leaders to investigate "what is," along with investigating and predicting the answers to "what will" and "what if" questions.

Conclusion

Information technology has evolved to be one of the leading drivers of development in the business world. Every organization yearns to stay technologically updated. Among the most utilized techniques are data and information visualization tools. The application of these tools is on the rise as individuals and small organizations are steadily stepping in. However, some of the users find it challenging to select the best tool that fits their requirements. Therefore, every user must understand each tool based on its use, use it, and use it. After following, the challenges experienced during and after using data visualization tools will have been smothered or reduced to a great extent.

References

- Archambault, S. G., Helouvry, J., Strohl, B., & Williams, G. (2015). Data visualization as a communication tool. *Library Hi Tech News*.
- Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation*, *2013*(139), 7-32.
- Christensen, J. (2017). Effective Data Visualization: The Right Chart for the Right Data, and Data Visualization: A Handbook For Data-Driven Design. Technology|Architecture + Design, 1(2), 242-243. doi: 10.1080/24751448.2017.1354629
- Few, S., & Edge, P. (2007). Data visualization: past, present, and future. *IBM Cognos Innovation Center*.
- Henderson, S. G., & Mason, A. J. (2005). Ambulance service planning: simulation and data visualization. Operations research and health care (pp. 77-102). Springer, Boston, MA.
- Kot, B., Wuensche, B., Grundy, J., & Hosking, J. (2005, July). Information visualization utilizing 3D computer game engines case study: a source code comprehension tool.
- Keim, D. A., Hao, M. C., Dayal, U., & Hsu, M. (2002). Pixel bar charts: a visualization technique for very large multi-attribute data sets. Information Visualization, 1(1), 20-34.
- Khan, M., & Khan, S. S. (2011). Data and information visualization methods, and interactive mechanisms: A survey. International Journal of Computer Applications, 34(1), 1-14.
- Kumar, S., & Kirthika, M. (2017). Big Data Analytics Architecture and Challenges, Issues of Big Data Analytics. International Journal Of Trend In Scientific Research And Development, Volume1(Issue-6), 669-673. doi: 10.31142/ijtsrd4673
- Paneels, S., & Roberts, J. C. (2009). Review of designs for haptic data visualization. *IEEE Transactions on Haptics*, *3*(2), 119-137.

- Parbat, B., Dwivedi, A. K., & Vyas, O. P. (2010). Data visualization tools for WSNs: a glimpse. International Journal of Computer Applications, 2(1), 14-20.
- Sadiku, M., Shadare, A. E., Musa, S. M., Akujuobi, C. M., & Perry, R. (2016). Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, *2*(12), 11-16.
- Theus, M. (2002). Interactive data visualization using Mondrian. *Journal of Statistical Software*, *7*(11), 1-9.
- Varela-González, M., Solla, M., Martínez-Sánchez, J., & Arias, P. (2014). A semi-automatic processing and visualisation tool for ground-penetrating radar pavement thickness data. Automation in Construction, 45, 42-49.