March-April 2020
ISSN: 0193-4120 Page No. 5999 - 6006

Data Visualization Tools and Techniques for Big Data Analytics

[1]S.Shanthi, [2]K.Nirmaladevi, [3]M.Pyingkodi
[1,2,3] Kongu Engineering College, India
[1]shanthis@kongu.ac.in, [2]k_nirmal@kongu.ac.in , [3] pyingkodi@kongu.ac.in

Article Info
Volume 83
Page Number: 5999 - 6006
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 01 April 2020*

Abstract:

The growth of data in the present world is increasing drastically, since large amount of data is generated from different domains. Due to this enormous growth of data, data investigation and visualization plays a vital role in the Big data era. Data visualization is the representation of information in the form of diagram, chart, picture etc. In the life cycle of data analytics, data visualization is an important phase, because if the outcome of the analytics is communicated through visualization rather than the text, the extracted knowledge of the user from the outcome of the analytics will be high. In this paper different types use case for visualization techniques, different types of visualization tools and the real time applications of data visualization in different domains are discussed. Also the guidelines for selecting the appropriate tools and techniques are discussed.

Keywords: Visualization, Data analytics, Tools, Visualization Techniques, Big Data, Charts, Graph, Tableau

I. INTRODUCTION

HE rate of data growth over the year is amazing. In day today life a volume of different types of big data are generated from different domains. Everybody from officials and departmental leaders to examination, call focus laborers and creation line workers – envisions taking in things from those different arrangements of information that can assist them with settling on more advantageous choices, take exquisite activities and work all the more productively. Concerning proficient development, the headway, utilizing information driven bits of knowledge t to devise significant methodologies and utilize important innovativeness is vital. The analysts were utilized diverse insightful methods to mine the concealed example from the large information. Different methodologies like data mining, machine learning, artificial intelligence, data analytics etc are used to analyze the big data and extract hidden knowledge from the volume of data. The hidden knowledge not only provides astute insights into critical elements of the problem domain but if presented in an inspiring, digestible, and consistent format, it can tell a tale that everybody within the organization can get behind. Visualization is a process of mapping information to visuals.

Data visualization techniques refer to the design of graphical representations of information. Data visualization techniques play a vital role to understand the big-data in real time applications by utilizing complex sets of numerical and factual data. Visualization is the process of generating images by filtering, mapping and rendering of data [1]

Visualizations can make complex datasets clear in an moment by presenting information in instinctive and

Published by: The Mattingley Publishing Co., Inc.

March-April 2020
ISSN: 0193-4120 Page No. 5999 - 6006

user-friendly ways, opening up the possibility to dive extremely into existing data resources and expose novel insights into opportunities for improvement. Undoubtedly, the perception is the natural stage to the clients inside the information life cycle's stage, along these lines a successful, productive and noteworthy portrayal of the concealed information or investigated data may result as significant as the logical procedure itself. Figure 1 speak to the significance of perception in changing information to information. Table 1 delineates the Air Quality Index (AQI) esteem range, conditions and hues to symbolize the air quality conditions.Though both color and text are used in Table 1 to define the air quality conditions, human cognition observes it by using colors rather than the text [2].

II. Big Data Dimensions and Data Visualization Techniques

Big data brings new challenges to visualization due to the speed, size and diversity of data that must be considered for analysis. The cardinality of the attributes are trying to visualize should also be look into it.

2.1 Line Charts

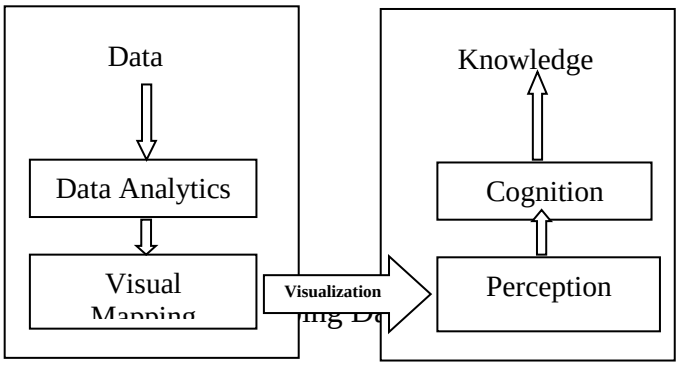
Line charts demonstrate the relationship of one variable to another. They are most frequently used to track changes or trends over time. Line charts are also helpful when comparing multiple items over the same time period.

2.2 Bar Charts

Bar charts are most generally used for evaluating the quantities of diverse groups or categories. Values of a category are denoted using the bars, and they can be configured with either vertical or horizontal bars, with the length or height of each bar representing the value.

Table 1 Air Quality Index Value, nomenclature and colour

(AQI) Values	Levels of Health Concern	Colors
When the AQI is in this range:	air quality conditions are:	as symbolized by this color:
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon



2.3 Pie and Donut Charts

Pie chat is the most widely used chart and the independent variable is plotted in clockwise or anticlockwise direction on the circular graph. The magnitude of the dependent variable is proportional to the length of the arc on the circumference of the graph. Radial lines are used to connect the arcs to

Published by: The Mattingley Publishing Co., Inc.

March-April 2020
ISSN: 0193-4120 Page No. 5999 - 6006

the center of the circle, thus dividing the pie into slices. Figure 2, Figure 3 and Figure 4 presents the average annual death rate in India due to air pollution from 1990 to 2016. It is observed that line and bar chart is more suitable than other types of visualization techniques for this scenario.

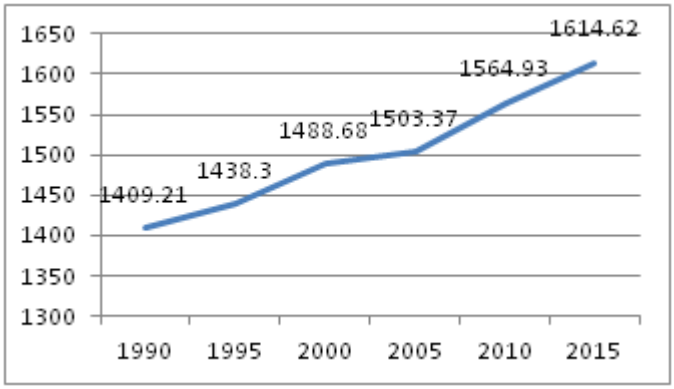


Figure 2 Average annual death in India due to air pollution 1990 to 2016 – Line Chart

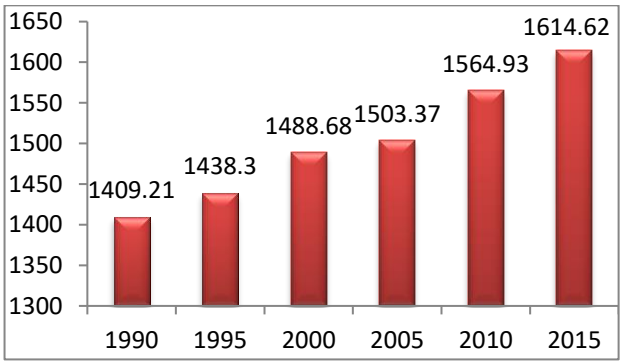


Figure 3 Average annual death in India due to air pollution 1990 to 2016 – Bar Chart

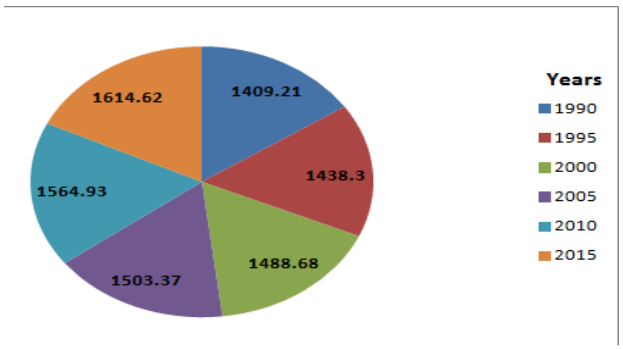


Figure 4 Average annual death in India due to air pollution 1990 to 2016 – Pie Chart

2.4 Bubble Plots: A Scatter Plot Variation

Bubble chat is muti-variable scatter plot, to plot a point along X and Y axis; and each plotted point represents a third variable by the area of its circle. As more number of bubble reduce the readability of the chart, this can be used for limited data size. Plotly is an easy way to create bubble plots. Figure 5 shows the population of different countries.

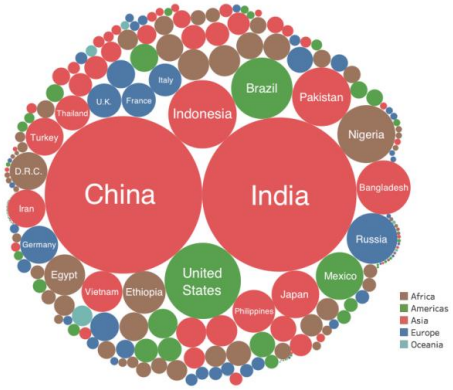


Figure 5 Countries by population using bubble Plot

2.5 Box Plot

A box plot is a graphical display of five statistics (the minimum, lower quartile, median, upper quartile and maximum) that review the distribution of a set of data. The lower quartile (25th percentile) is denoted by the lower edge of the box, and the upper quartile (75th percentile) is denoted by the upper edge of the box. The median (50th percentile) is represented by a central line that divides the box into sections. The inter quartile range (IQR) is the width of the box and box plots are used to find the outlier in the dataset which is shown in Figure 6.

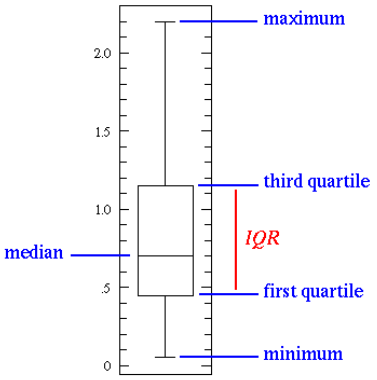


Figure 6 Box plot representation

2.6 Network diagram

Network diagram views relationship in terms of node and ties and used to visualize semi structured

Published by: The Mattingley Publishing Co., Inc.

March-April 2020
ISSN: 0193-4120 Page No. 5999 - 6006

and unstructured data. In social network like twitter, facebook, to analyze the interaction among the customers; in election,the voting information like who voted for whom, in an organization who has a relationship with whom the network diagram can be used. Figure 7 shows different organizations which supports in oil and gas development.

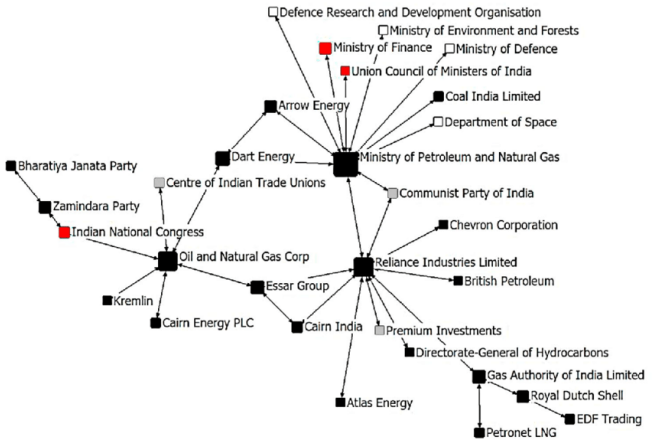


Figure7 Disagreement network. Note: Node color reflects organizational stance on shale oil and gas development. Pro: Black; Anti: White; Mixed: Red; Not Specified: Grey. Node size reflects centrality.

2.7 Correlation Matrices

Correlation matrix defines correlation coefficients among two variable. Table 2 lists the type of correlation coefficient for different types of variable.

Table 2 Data types vs correlation coefficient

	Quantitative	Ordinal	Nominal
Quantitative	Pearson	Biserial	Point Biserial
Ordinal	Biserial	Spearman rho/ Tetrachoric	Rank Biserial
Nominal	Point Biserial	Rank Biserial	Phi, Goodman & Kruskal's Lambda

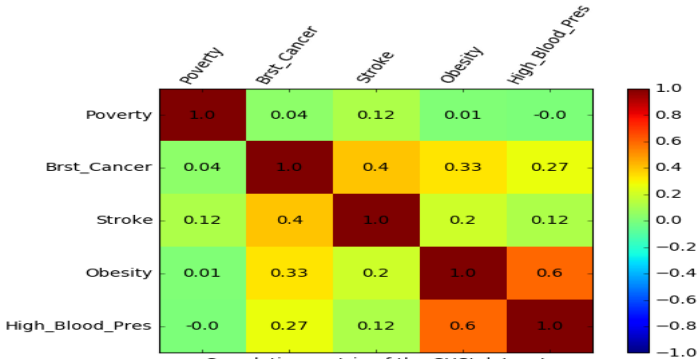


Figure 8 Correlation Matrix

2.8 Scatter Plots

Scatter plot suggests the correlation (positive correlation, negative correlation and no correlation) between the attributes with a certain confidence interval. Scatter plot also shows the missing value, unexpected gaps and logic errors in the data. Figure 9 shows the runs scored in 50 overs using scatter plot.

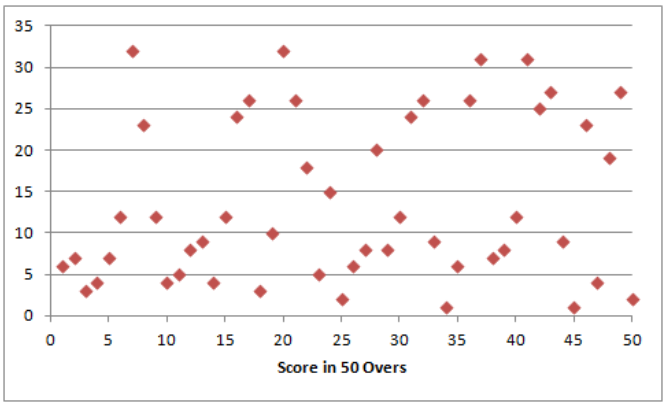


Figure 9 Scatter Plot

III.Data Visualization Tools

There are a large amount of open and licensed data visualization tools are available in the market. This section discuss some of the tools.

3.1 Tableau

Tableau is the big data visualization tool in the market and it requires zero coding knowledge, which allows the user to create charts, graphs, dashoards and many more. Barclays, Pandora and Citrix are using Tableau for visualization of their data analysis. The advantage of

Published by: The Mattingley Publishing Co., Inc.

using this, it provides interactive dashboards, its responsiveness and real time data analysis features. Also, it has the facilities for data extraction, processing, preparing visualization reports and sharing the reports and worksheets with others. Tableau is easy to use, because it has the drag and drop option to create visualization and connected with the different data source like Excel, Text file, MangoDB, SQL, Teradata, Hadoop, Oracle, Google cloud etc. The software released in Desktop, Server and online hosted model. Server permits dashboard sharing and collaboration among the different groups of organization. Online is web based and allows having visualization in mobile versions. Tableau Public and Tableau Reader is free for creating data visualization from spread sheet or file and display in web and Desktop. The demerits of Tableau are expensive for real time data analytics [3].

3.2Qlikview

Qlikview is the biggest competitor for the Tableau. It provides facilities for powerful business analytics and reporting capabilities. Qliksense is the sister package of Qlikview and supports a numbers of third party resources in online to integrate the users new projects in the existing [4].

3.3 Infogram

Infogram is fully featured drag and drop tool and limits to read data from limited data sources and create visualization of more than 35 chart types and more than 550 map types. Completed visualization can be exported into different formats like .pdf, .png, and .html etc [5].

3.4 Sisense

Provides visualization for anyone at anywhere with limited type of visualization. Also used to uncover underlying trends and patterns in the huge amount of business data from different source of data [5].

3.5 Whatagraph

This provides a platform for marketing agencies to report their marketing campaign to the clients.

3.6 Watson Analytics

IBM's cloud based data analytics and visualization tool and explore the data through natural language processing cognitive computing.

3.7 Datawrapper

It used to creates infographics, data tables, maps and responsive charts and used by Wasihington Post, Guardian, Wall Street Journal.

3.8 Microsoft Power BI

It is a business analytics tool, which support access to on premise and in cloud data. This software is free upto 1GB data limit. and uses D3.js for all types of graphics. This is integrated with Excel users, SQL database connectors, Python, R and MATLAB users.

3.9 Plotly

It is web based visualization tool, used to make interactive charts, dashboards and presentations. It's built on top of D3.js visualization libraries. This also allows presenting complex visualization by integrating with programming languages like Python, R and MATLAB. This is best suited for developers.

3.10 TeamMate Analytics

This tool is extensively used by the auditors to perform powerful data analysis and deliver significant outcome to their customers, teams and different levels of stakeholders. It includes more than 150 audit tools and gets data from Excel, assists the auditors in formatting, manipulating, extracting, transforming and analyzing all types of data.

3.11 Google Chart

Published by: The Mattingley Publishing Co., Inc.

March-April 2020
ISSN: 0193-4120 Page No. 5999 - 6006

It is a powerful, interactive visualization tool for web and mobile devices and Gets data from Excel, SQL database, CSV, Google spreadsheets, etc.

3.12 FusionCharts

It is a combination of the javascript, Angular JS, jQuery and React based platform integrated with wide range of devices and it is a good choice for mobile and web developers. This also works efficiently with server side languages like ASP.Net, Ruby on Rails, Java and PHP. It is very much appropriate for the developers looking to employ the visualization in their applications.

3.13 HighCharts

It is easy to use and used by 80% of the worlds' largest companies, including GitHub. Its cross-browser option adds to Highcharts' ease-of-use value, and its ability to mechanically find best possible placement for non-graph elements, such as legends and headings, does much the same.

3.14 CiteSpace

CiteSpace is a free java application for visualization of trends and patterns in scientific community.

Except for the above tools, there are still ample of constructive tools for us to attempt and explore.

IV.Visualization Techniques in Real world Applications

Visualization offer speedy, spontaneous, and simpler way of transmission critical concepts universally in various applications. The following are important guidelines for effective preparation of visualization of the data.

a) Know the audience
b) Set the goals
c) Select the appropriate visualization techniques
d) Take the right color combinations
e) Handle bigdata
f) Use ordering, Layout, and hierarchy to
g) Include comparisions
h) Tell your tale

Information-intensive disciplines such as industry, healthcare, transportation, environmental study, financial market, software maintenance, stock market, medicine, bioinformatics, the earth sciences, and computational fluid dynamics, criminal investigation and many more have adopted recent visualization methods to handle and explore their relevant billions of data [6].

Data visualization plots have been used to monitor the event data in industry namely alarm message, process data, and operator response for improved data management. Different data visualization techniques bar chart, pie chart, spider chart etc have been used to visualize the data [1]. The importance of visualization in understanding complex and large data have been reviewed and existing technical challenges were identified [7]. The big volume of heterogeneous business data is categorized as business intelligence, business ecosystem, and customer-centric. Different visualization techniques applied to address the critical and challenging part of the business process [8].

Hospitals collect a lot of information about the patient, treatment details, drug details etc. All the information are analyzed and used to improve treatment plans, reduce the cost, provide adequate facilities, if it is presented to the decision makers in a way that can be understood. Today, data visualization solutions can be begin everywhere in healthcare systems from hospital operations monitoring and patient profiling to demand projection and capacity planning [9].

The stock market is one field comes with thousands of companies with complex data arise, and the performance is conventionally measured by representing price fluctuation over time, as well as applying treemap for showing volume and relationship etc. Insight into stock data is important

Published by: The Mattingley Publishing Co., Inc.

in technical stock market analysis where exclusively the analyzed visualization representation is considered in decision making [10].

Scholarly information consists of large number attributes like authors, papers, citations, affiliations, and network to other scholars etc. The data analytics on this scholarly data produces a interesting and useful pattern among the scholars throughout the universe with respect to different dimensions. The well prepared information can assist the scientists to scrutinize scientific team formation and to have a inclusive learning about the research communications in the area of applied science, technology, engineering, management etc. Different tools for visualization have been studied and its challenges are addressed [11].

Transportation is another source for producing big data. Introduces the basic concept and pipeline of traffic data visualization, provides an overview of related data processing techniques, and summarizes existing methods for depicting the temporal, spatial, numerical, and categorical properties of traffic data [12].

Visualization of air quality data is important, not only for experts and managers, but also for the public in general. Air Pollutant in air is monitored, analyzed and Air Quality Index (AQI) is computed and visualization techniques used to display the findings in the web portal to easily understand by the public. Different types of visualization techniques are used to analyze time series data and show the correlation among various factors. PivotTable.js and D3.js tools were used to create visualization of the findings [13]. Using Google Earth to visualize urban air quality combines air quality information with respect to time and space. It visualizes huge amounts of air quality data in a spontaneous and vibrant method, flouting conventional patterns of data, formulas, and charts that articulate air quality. The imitation on the 3D virtual Earth platform can obtain the dynamic change of air quality to enhance the authenticity of data visualization of air quality. It is simple for users to recognize and helpful to providing decision support for public management and improving urban air quality [14].

In Urban city planning, the visualization can considerably increase the design and realization of sustainable future cities: visual illustrations efficiently map diverse mental worlds among different stakeholders like individual urban planners, national planning authorities, investors, the general public, local neighborhoods [15].

The IT industries are exploiting the data visualization techniques to realize the customer insights, boosting operational efficiency, and competitor's trends in order to improve the efficiency and continue competitive in the industry. Out of trillions of data, only 22% of data is useful, which is predicted to increase to 37% by end of 2020 [16].

CONCLUSIONS

The main idea behind this article is the overview of different types of visualization techniques and the applications where this visualization is required. Also summarize the software tools use to visualize the data. As the bigdata analytics is the emerging filed, the researchers needs to concentrate on visualization techniques to get the maximum insights from the analytics. Through this visualization, the business users can easily see what's important to focus on, and data scientists can use visualizations as a starting point for building models based on the most relevant variables.

REFERENCES

[1] Hu, W., Al-Dabbagh, A.W., Chen, T. and Shah, S.L., 2018. Design of visualization plots of industrial alarm and event data for enhanced alarm management. Control Engineering Practice, 79, pp.50-64.

[2] Shanthi, S. and Pyingkodi, M., 2019. Air Quality Index Prediction using Machine Learning Algorithms. International Journal of Recent

Technology and Engineering, 8(4), pp. 7489- 7492
[3] https://www.tableau.com/en-in last accessed March 2020
[4] https://www.qlik.com/us last accessed March 2020
[5] https://www.octoparse.com/blog/top-30-data-visualization-tools last accessed March 2020
[6] Gershon, N. and Eick, S.G., 1997. Information visualization applications in the real world. IEEE Computer Graphics and Applications, (4), p.66a.
[7] Zhou, F., Lin, X., Liu, C., Zhao, Y., Xu, P., Ren, L., Xue, T. and Ren, L., 2019. A survey of visualization for smart manufacturing. Journal of Visualization, 22(2), pp.419-435.
[8] Roberts, R.C. and Laramee, R.S., 2018. Visualising business data: A survey. Information, 9(11), p.285.
[9] Widanagamaachchi, W., Livnat, Y., Bremer, P.T., Duvall, S. and Pascucci, V., 2017. Interactive visualization and exploration of patient progression in a hospital setting. In AMIA Annual Symposium Proceedings (Vol. 2017, p. 1773). American Medical Informatics Association.
[10] Hua, J., Huang, M.L., Zreika, M. and Wang, G., 2018. Applying data visualization techniques for stock relationship analysis. Filomat, 32(5).
[11] Liu, J., Tang, T., Wang, W., Xu, B., Kong, X. and Xia, F., 2018. A survey of scholarly data visualization. IEEE Access, 6, pp.19205-19221.
[12] Chen, W., Guo, F. and Wang, F.Y., 2015. A survey of traffic data visualization. IEEE Transactions on Intelligent Transportation Systems, 16(6), pp.2970-2984.
[13] Li, H., Fan, H. and Mao, F., 2016. A visualization approach to air pollution data exploration—a case study of air quality index (PM2. 5) in Beijing, China. Atmosphere, 7(3), p.35.
[14] https://aqicn.org/contact/ accessed on November 2019
[15] Kunze, A., Burkhard, R., Gebhardt, S. and Tuncer, B., 2012. Visualization and decision support tools in urban planning. In Digital Urban Modeling and Simulation (pp. 279-298). Springer, Berlin, Heidelberg
[16] https://killervisualstrategies.com/blog/tech-industry-data-visualization.html last accessed March 2020