

A Study on Data Visualization

*S.Sravani¹, M.Vinitha²

¹Dept. of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India.

²Dept. of Computer Science and Engineering, RGUKT, APIIT, Kadapa, India.

Corresponding Author: S.Sravani

Abstract: Data visualization is presenting the data in pictorial or graphical format. With the rapid growth of data, the importance to extract useful information from the data is increasing. With ever growing data it takes hours or days to get answers for questions. More users are expecting self access to information in a form they can easily understand and share with others. Data visualization for text is an effective way to see and analyze what a designated text is saying.

Keywords: Visualization, Histogram, Scatter plot, Network, Gantt Chart

Date of Submission: 04-09-2017

Date of acceptance: 20-09-2017

I. Introduction

With data generated rapidly by the Internet and bioinformatics and sensors and so on, the data visualization and the big data with distributed processing have become a more popular topic today. At the same time, the analysis of big data, particularly mining valuable information from big data, and the data visualization with a very good tool are the research trend in the big data and visualization field. However, the kind of product, which is combined by the big data processing framework and visualization tools, is rare. In the information age, all kinds of data are just like natural resources, and the demand for the product becomes more imminent.

A primary goal of data visualization is to communicate information clearly and efficiently via statistical graphics, plots and information graphics. Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message. Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable.

II. History Of Data Visualization:

The concept of using pictures to understand data has been around for centuries, from maps and graphs in the 17th century to the invention of the pie chart in the early 1800s. Several decades later, one of the most cited examples of statistical graphics occurred when Charles Minard mapped Napoleon's invasion of Russia. The map depicted the size of the army as well as the path of Napoleon's retreat from Moscow – and tied that information to temperature and time scales for a more in-depth understanding of the event. It's technology, however, that truly lit the fire under data visualization. Computers made it possible to process large amounts of data at lightning-fast speeds. Today, data visualization has become a rapidly evolving blend of science and art that is certain to change the corporate landscape over the next few years.

III. Importance Of Data Visualization:

With big data there's potential for great opportunity, but many retail banks are challenged when it comes to finding value in their big data investment. For example, how can they use big data to improve customer relationships? How – and to what extent – should they invest in big data? Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

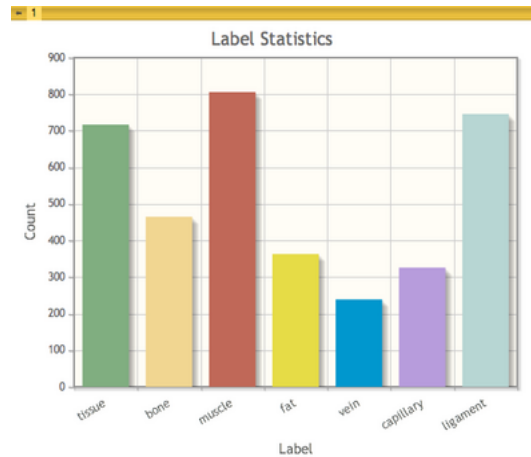
Data visualization can also:

- Identify areas that need attention or improvement.
- Clarify which factors influence customer behavior.
- Help you understand which products to place where.
- Predict sales volumes.

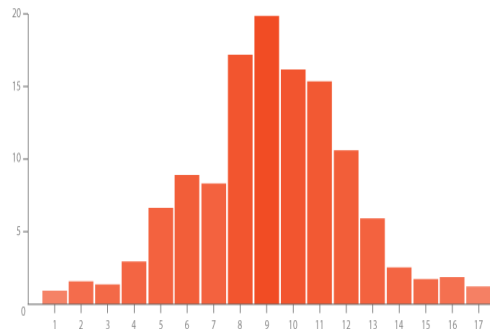
Data visualization is going to change the way our analysts work with data. They're going to be expected to respond to issues more rapidly. And they'll need to be able to dig for more insights – look at data differently, more imaginatively. Data visualization will promote that creative data exploration.

IV. Types Of Representation Of Data:

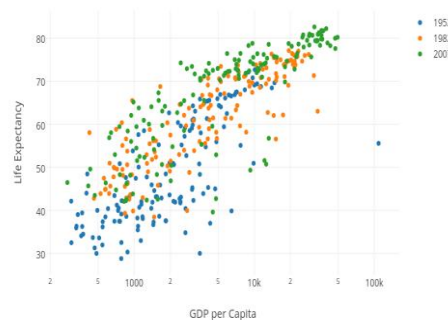
Bar chart: The classic Bar Chart uses bars to show discrete, numerical comparisons between categories of data, which can be displayed horizontally or vertically. A bar graph will have two axis. One axis will have numerical values, and other will describe the types of categories being compared. Length, category and color are taken as visual Dimensions to represent data.



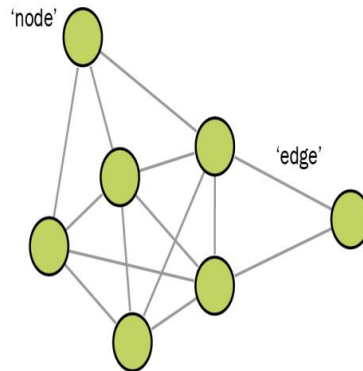
Histogram: Histogram is used to plot the frequency of score occurrences in a continuous data set that has been divided into classes, called bins. Each bar in the histogram represents the tabulated frequency at each interval/bin. Histograms gives an estimate as to where values are concentrated, what the extremes are and whether there are any gaps or unusual values. For histogram, bin limits, length/count and color are taken as visual dimensions to represent data.



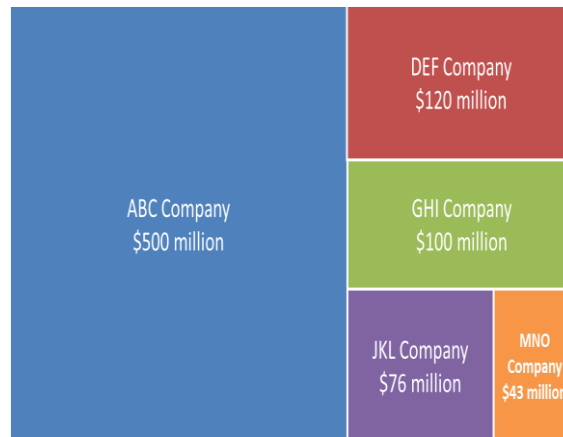
Scatter plot: scatter plot is a mathematical diagram which uses Cartesian coordinates to display values of two variables for a set of data. The data is displayed as a collection of points, each having the value of one variable showing the position on the horizontal axis and the value of other variable determining the position on the vertical axis. By displaying a variable in each axis, you can detect if a relationship or correlation between the two variables exist.



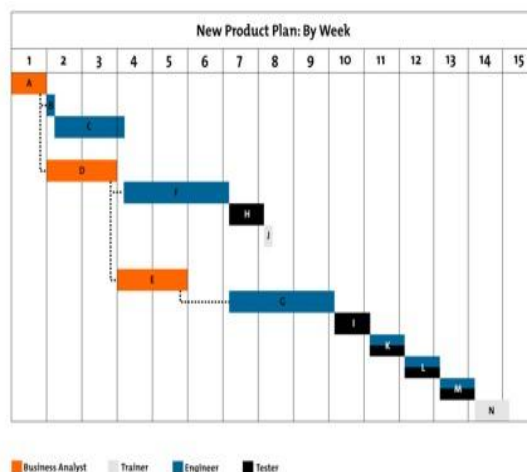
Network: Network visualization helps us to find clusters in the network like grouping facebook friends into different clusters and discovering bridges between clusters in the network. Network Visualization is useful to determine the most influential nodes in the network. This can also be useful for finding outlier actors who does not fit in any cluster or in the periphery of a network.



Treemaps: Treemaps are an alternative way of visualizing the hierarchical structure of a tree diagram while also displaying quantities for each category via area size. Each category is assigned a rectangle area with their subcategory rectangles nested inside of it. The area size of the parent category is the total of its subcategories. Ben shneiderman originally developed Treemaps as a way of visualizing a vast file directory on a computer, without taking up too much space on the screen. Treemaps are compact and space-efficient option for displaying hierarchies, that gives a quick overview of the structure.



Gantt chart: Gantt charts are commonly used as an organizational tool for project management. Gantt chart displays a list of tasks with their duration over time showing when each task starts and ends. This makes Gantt charts useful for planning and estimating how long an entire project might take. You can also see what activities are running in parallel to each other. Connecting arrows can be used to show which tasks are dependent on each other.



V. Conclusion:

In this paper we discussed importance of data visualization and uses. Different types of representation of data are discussed. In future algorithms can be developed for data visualization by which we can generate visuals of textual data automatically.

References:

- [1] A. Paterek, "Improving regularized singular value decomposition for collaborative filtering," in Proceedings of the KDD Cup and Workshop (KDDCup '07), pp. 5–8, San Jose, Calif, USA, August 2007.
- [2] Q. Liu, E. Chen, H. Xiong, C. H. Q. Ding, and J. Chen, "Enhancing collaborative filtering by user interest expansion via personalized ranking," IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics, vol. 42, no. 1, pp. 218–233, 2012.
- [3] C. Birtolo, D. Ronca, and R. Armenise, "Improving accuracy of recommendation system by means of item-based fuzzy clustering collaborative filtering," in Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA '11), pp. 100–106, IEEE, November 2011.
- [4] S. Zhang, W. Wang, J. Ford et al., "Learning from incomplete ratings using non-negative matrix factorization," in Proceedings of the 6th SIAM Conference on Data Mining (SDM '06), 2006.
- [5] S. Wei, N. Ye, S. Zhang, X. Huang, and J. Zhu, "Collaborative filtering recommendation algorithm based on item clustering and global similarity," in Proceedings of the 5th International Conference on Business Intelligence and Financial Engineering (BIFE '12), pp. 69–72, IEEE, Lanzhou, China, August 2012.
- [6] Z. Liu, W. Qu, H. Li, and C. Xie, "A hybrid collaborative filtering recommendation mechanism for P2P networks," Future Generation Computer Systems, vol. 26, no. 8, pp. 1409–1417, 2010.
- [7] R. Pan and M. Scholz, "Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering," in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09), pp. 667–675, July 2009.
- [8] A. Bahga and V. K. Madiseti, "Analyzing massive machine maintenance data in a computing cloud," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1831–1843, 2012.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with SI. No. 5019, Journal no. 49102.

S. Sravani. "A Study on Data Visualization." IOSR Journal of Computer Engineering (IOSR-JCE), vol. 19, no. 5, 2017, pp. 14–17.