

Data Visualization

Pratik Sanghani¹, Sanket Bhang², Rajas Telang³, Atharva Atre⁴, Manali Patki⁵,
Aatul Palandurkar⁶

^{1,2,3,4}Student, Dept. of Information Technology, St. Vincent Pallotti College of Engg. and Tech., Nagpur, India

⁵Asst. Professor, Dept. of Information Technology, St. Vincent Pallotti College of Engg. and Tech., Nagpur, India

⁶Director, Kloud Ancyber Technologies Pvt. Ltd., Pune, India

Abstract: Data Visualization is a newly area in Information technology. Data visualization has many gains in technologies because it is an easy way to represent more complex data in the form of infographs. Data visualization is used to show the relationship among complex datasets. Data Visualization makes it potential for businessman and the technologist to get insight in these data in an effectual and good manner. There are wads techniques to represent data in the form of graphics such as Graphs, Charts, Maps and Images etc. We propose a data visualization basic concept, techniques, and comparison of techniques and research on the New Data Visualization techniques.

Keywords: Data visualization, Complex data, Insight, Data comparison.

1. Introduction

Data visualization is a general term that explain any effort to help people understand the importance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text based data can be exposed and recognized easily with data visualization software [1] To understand information clearly and efficiently, data visualization uses statistical graphics, plots, information graphics and other tools. Numerical data may be encoded using dots, lines, or bars, to visually understand the quantitative message. Effective visualization helps users to analyze the data. It makes complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic follows the task. Tables are generally used where users will look up a specific measurement, while charts of various types are used to show patterns or relationships among various variables [2]. Data visualization is an art and a science. It is viewed as a branch of statistics by some, and also as a grounded theory development tool. Increased amounts of data created by Internet activity and number of sensors in the environment are referred to as "big data" or Internet of things. Processing, analyzing and communicating this data present analytical challenges for data visualization. The field of data science and practitioners called data scientists help to solve this challenge [3].

2. Literature survey

Data Visualization is an emerging area in Information technology. It is presented by Human Computer Interaction field. Enterprises today are beginning to realize the important role data plays in achieving business goals. Concepts that used to be difficult for companies that influence a customer to make a purchase, behavior patterns that point to fraud or misuse, inefficiencies slowing down business process now can be understood and explain by collecting and analyzing Data. The insight from such analysis helps organizations improve operations and identify new product and service opportunities. Data visualization promise to provide the advantages that companies need to run revenue growth and gain a competitive edge [2].

Great technological have enabled researchers to create a big and effective amount of data. Data analysis is replacing data generation as the rate-limiting step in data science. With the help of this information, we have an opportunity to understand the causes of human diseases. However, the unprecedented scale, resolution, and variety of data create new analytical challenges. Visual representation of data offers insights that can lead to new understanding, whether the purpose is analysis or communication [3].

Visualization can be used for decision making and data analysis. People interaction with visualization tool has strongly effect on the understanding of data and system functions. Therefore, human interaction contribute significantly role in the valuation and design of visualization tool. Visualization help human to display data as a graphics. Navigation techniques and graph visualization which are used in information visualization. Information demonstrate by the graph visualization. It handling the information visualization by using graph visualization techniques.

Data visualization using python is used to provide a stronger foundation in data visualization. It provides a broader coverage of the matplotlib library and an overview of seaborn which is package for statistical graphics. It covered customizing graphics, plotting two-dimensional arrays (such as pseudo color plots, contour plots, images etc.), statistical graphics (such as visualization distribution & regressions), and working with time series and image data.

As we all know that there is a huge amount of data available

in different formats and it is not possible to understand that data so we are trying to present the complex datasets in diagrammatic way i.e. Info graphs. Info graphs is a chart that represents data in the form of visuals. It helps us to understand the data. It is very important in decision making, identifying the affected areas and future prediction.

3. Proposed system

Data visualization is the presentation of data pictorial or graphical format. It enables decision makers to see analytics presented visually, so they can understand difficult concepts or identify new patterns. With interactive visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more information. It is possible because of computers to process large amounts of data at lightning-fast speeds. Today, data visualization has become a rapidly evolving mixture of science and technology that is certain to change the corporate scenario over the next few years.

4. System description

Companies are using tools like tableau, Sisense etc. to visualize the data, but we are referring python and its libraries to visualize the same data. The major advantage of using python because of its fast extraction of data and showing them into the graph format which will help in gaining faster and efficient results. This project uses the coding to describe the various factors of a particular topic and represent them in the graphical format which will be clearly understood to the end user clearly.

5. Scope

Our visualization system will give the idea about different type of data, get the real and accurate picture of it, maintain an inventory details which will help in estimating the raw material required for future goals. On the basis of visualization our system which will help in planning. Our system will help to put the data into the users' hands allowing them to more quickly identify issues and improve response times. Using visualizations allows users to get the big picture and see the details at the same time. It simplifies the data by allowing the users to interact only with relevant data. allows users to better absorb the data and see new paths. This enables users to identify new patterns and trends that were impossible to see using tabular data. It allows decision-makers to view data using graphical representations including charts, fever charts, and heat maps. It also allows users to better absorb the data and see new paths.

6. System analysis

A. Flow Chart

A flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution model to a

given problem. Flowcharts are used in analyzing, designing, documenting or managing a process or program in various fields. Flowchart given below illustrates the flow of the website all the way from the homepage to the various sections of the website and then ending at a common point.

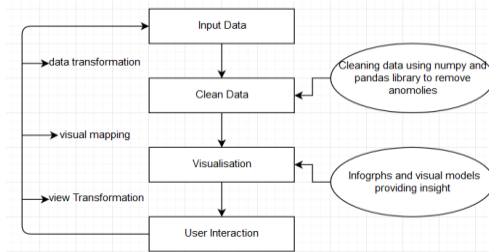


Fig. 1. Flowchart

7. System requirement

In this section we describe the system requirements of the Data Visualization. The proposed system needs to visualized the complex datasets (stock market, election), fast extraction of data into insightful graphs.

A. Hardware requirement

- Processor – i3, 5th Generation and above.
- Speed - 1.5 GHz and above.
- Speed - 1.5 GHz and above.
- RAM - 2 GB (min) and above.
- Hard Disk – 256 GB and above.

B. Software requirement

- Operating System – Windows 7/8/10 /Linux
- Tools – Anaconda, Python, Jupyter.

Anaconda – Anaconda is a free and open source distribution the Python and R programming languages large-scale data processing, predictive analytics, scientific computing, that aims to simplify package management and deployment. Package versions are managed by the package management system. Jupyter - Project Jupyter supports execution environments in various languages. Project Jupyter's name is an address to the three core programming languages supported by Jupyter, which are Julia, Python and R. Python - Python is an interpreted high-level programming language for general-purpose programming. It provides constructs that enable clear programming on both small and large scales.

8. System Architecture

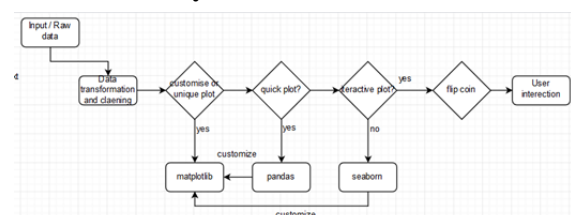


Fig. 2. System architecture

9. Implementation

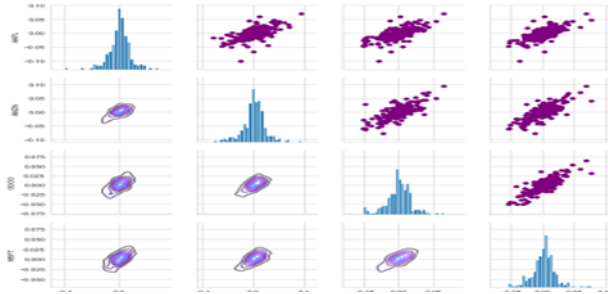


Fig. 3. Series of histogram to represent the daily return of the various companies

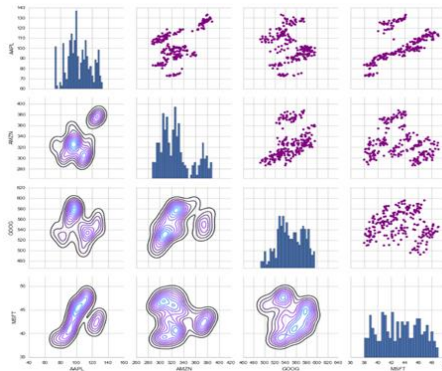


Fig. 4. Series of histogram to represent the daily closing value of the various tech companies

10. Conclusion

This paper presented the visualization of complex datasets using python.

References

- [1] Faiza Nazeer, Nadia Nazeer and Iqra Akbar "Data Visualization Techniques – A Survey" IEEE, vol. 4, Issue 3, Mar-2017.
- [2] Zhicheng Liu, Biye Jiang, Jeffrey Heer, "Mens: Real-time Visual Querying of Big Data," Computer Graphics Forum (Proc. Euro Vis), vol. 32, no. 3, 2013.
- [3] M. Trutschl, G. Grinstein, U. Cvek, "Intelligently resolving point occlusion", Proceedings of Info Vis, pp. 131-136, Oct. 2003.
- [4] A. KEIMD, "Designing pixel-oriented visualization techniques: Theory and applications", IEEE TVCG, vol. 6, no. 1, pp. 59-78, 2000.
- [5] B. Wong, "Visual representation of scientific information", Sci. Signaling, vol. 4, 2011.
- [6] P. Shah, E. G. Freedman, "Bar and line graph comprehension: An interaction of top-down and bottom-up processes", Topics in Cognitive Sci., vol. 3, pp. 560-578, 2011.
- [7] Shaun Bangay, "Visview: A system for the visualization of Multi-dimensional data", in "Visual Data Exploration and Analysis V," 1998.