# **Exploratory Review of Data Visualization and Time Series**

Varun P. Patil<sup>1</sup>, Ajit S. Mali<sup>2</sup>

<sup>1</sup> Research Scholar, <sup>2</sup> Assistant Professor <sup>1, 2</sup> Department of Computer Science and Engineering, RIT College, Maharashtra, India.

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 2, March-April 2023

License: Creative Commons Attribution-ShareAlike 4.0 International License



## Abstract

Graphical representation of data, which helps in understanding complex data like "Big Data" is called "Data Visualization". As the data is growing rapidly, it becomes difficult to interpret the patterns and trends in data. Apart from this, data visualization is used for monitoring data for data cleaning, evaluating models, and representing the results of models. Time series data visualization is an instrument, where a time series is a group of observations on a single entity over some time, it also helps in the recognition of important features of the time. Time Series Visualization is the first step toward getting trending patterns in from the time series. There are 2 types of time series, uni-variant and multivariant. These 2 types can be used for analysis of time series. The main objective of time series modeling or analysis is to analyse the past data and to develop the model which could find trends, patterns, etc. This model could predict future values, which can be used for forecasting. This paper will provide information about various common data visualization techniques, libraries and tools. It also illustrates the importance of "Time series".

# 1. Introduction to Data Visualization

Data visualization is a field of study that deals with an enormous amount of data, which can be in the form of images, texts, PDFs, PPTs, and many more. To find useful information (meta-data) and patterns or understand them in an effortless way data visualization comes into play. Data visualization represents this complex data in form of various forms like [1] Graphs, Charts, Venn diagrams, Histograms, Points, Bars, etc. To access these there are various data visualization tools as well as libraries in python titled "Seaborn", "Matplotlib" and many more, and tools like Grafana, Fusion Charts, and Google Charts.

# 1.1. Various Types of Data Visualization Techniques

# 1.1.1. Graphs and Charts

Graphs are the most basic form of representation in data visualization. The information provided by these is fast and easy. It provides an understanding of complex and large data in a simple and in less amount of time. In most cases, graphs are preferred over tables as they show the overall shape of the data along with the message in the data that is patterns, trends, etc. Charts are same as graphs the difference is that it provides information that is form of graph, tables, etc. For Example, the Figure 1

represents a graph which displays the most liked fruit, and Figure 2 displays pie chart which provide information of most used visualization techniques.



Figure 1: Graphical Representation of Most Liked Fruit



Figure 2: Pie Chart of Most Used Visualization Technique

# 1.1.2. Venn Diagrams

A Venn diagram is used to represent the relationship between 2 or more sets, groups, and many more with the help of circular shapes like circles, ovals, and ellipses. A Venn diagram is also called a Logic diagram, Set diagram, or primary diagram. It is very much useful when we want to study intersections between 2 or 3 sets. The representation of 3 sets can be achieved by using an upset plot. Although we can do 4 sets at a time, it is highly recommended not to exceed more than 3 sets as it becomes extremely hard to read and interpret. Figure 3 shows the number of people liking tea, coffee and both in the form of Venn diagram.



Figure 3: Venn Diagram of People Liking Tea and Coffee

## 1.1.3. Histograms

A histogram is a frequency distribution graph. A histogram is used when there is numerical data only, it is used while analyzing the requirements of the consumer. Used when we want to see the difference between 2 or more results. At least 50 data entries are required to draw histogram.

# **5 Most Common Histogram Shapes**

## (1) Bell Shaped

It represents a normal distribution. It has a shape like bell so it is called bell-shaped histogram. Figure 4 represents bell-shaped histogram.



Figure 4: Bell Shaped Histogram

## (2) Bi-modal

This type has 2 peaks in the graph which mean that the data is from 2 different systems, so they should be worked out separately. Figure 5 represents bi-modal histogram.



Figure 5: Bi-modal Histogram

## (3) Skewed

It gives an asymmetric shape as a result due to the natural limit which prevents result to appear on either side. There are many types of skews. Figure 6 shows right-skewed histogram, and Figure 7 shows left-skewed histogram.



Figure 6: Right Skewed Histogram



Figure 7: Left Skewed Histogram

## (4) Uniform Histogram

In a uniform histogram, the frequency of the data is evenly distributed across all potential values. This indicates that the data is evenly distributed across all values and that there are no significant gaps or outliers. Figure 8 shows Uniform Histogram.



Figure 8: Uniform Histogram

# (5) Non-uniform Histogram

The frequency of the data is not evenly distributed across all possible values in a non-uniform histogram, also referred to as an uneven or non-uniformly distributed histogram. This indicates that the distribution of the data is not roughly uniform across all values and that there are significant gaps or outliers in the data. Figure 9 is a perfect example of Non-uniform Histogram.



Figure 9: Non-uniform Histogram

# **1.2.** Common Visualization Libraries

There are 2 most famous libraries that are used to visualize data:

# 1.2.1. Matplotlib

Matplotlib [4] is a standard library used to derive 2-D graphs it is also an extension to Numpy. Many other libraries have been developed with the help of Matplotlib so we can say it is a "Parent Library". There are a lot of functions in Matplotlib which are used at the backend, it is also used to create new axes and layouts of plotting. Matplotlib helps in plotting various graphs, pie charts, heatmaps, histograms, non-cartesian coordinates in graphs.

# 1.2.2. Seaborn

Seaborn [4] is another data visualization library that is based on Matplotlib it covers better handling of some of the data frames of pandas as compared to Matplotlib. It uses not more complex procedures to create an attractive graph. Another reason for Seaborn being popular that it is specialized in [5] statistical visualization and also it uses very less syntax. Much clean API is provided by Seaborn as compared to Matplotlib. Seaborn is also called as "Statistics library".

# 1.3. Popular Data Visualization Tools

Data visualization tools are software API that converts massive amount of data into graphical format. These tools take input from a large amount of data and the output produced is in the form of [1] charts, graphs, heatmaps, etc. Grafana, Fusion Charts, Google Charts, etc. Are some of the popular data visualization tools.

# 1.3.1. Grafana

Grafana is an open-source solutions tool that is used for data analytics, helping one to monitor applications with the help of Grafana's customizable dashboard. Grafana can establish connections with most databases like MySQL, Influx DB, Graphite, and many more. Grafana offers monitoring and analysis over some period of time this is called Time Series analysis. Besides offering open-source solutions Grafana also provides Grafana-Cloud and Grafana-Enterprise.

# **1.3.2.** Fusion Charts

Fusion charts is a charting library which provides data visualization products like JavaScript charts, Widgets, Dashboard, etc. The main idea of fusion charts is to satisfy the charting capabilities which Microsoft Excel could not satisfy. It offers us to display information on the website in a clean and dynamic way.

There are various reasons why one should use fusion charts:

- (1) Supports variety of charts.
- (2) Customizable.
- (3) Highly user interactive.
- (4) Supports various frameworks.

Figure 10 is simple decision tree architecture made by using Fusion Charts.



Figure 10: Decision Tree Architecture

## 1.3.3. Microsoft Excel

Microsoft Excel offers a multitude of choices for creating charts and graphs to visually represent data, making it a useful tool for data visualisation.

Column charts, bar charts, line charts, scatter plots, and pie charts are just a few of the numerous chart styles Excel offers [1]. Numerous other elements of your chart or graph, like the colours, labels, and data series, are all customizable. To further aid in data analysis and interpretation, Excel also offers tools for adding trendlines, error bars, and other characteristics to your chart.

## 1.3.4. Microsoft Power-BI

Microsoft Power BI is a robust data visualisation and business intelligence application that lets you connect to, view, and analyse data from a variety of sources. You can build interactive dashboards, reports, and charts using Power BI that will aid in providing you with insights and assisting you in making better business decisions. Additionally, you can collaborate in real-time on projects and share your work with others [1].

## **Features of Microsoft Power-Bi**

- (1) Advanced analytics, such as predictive modeling and machine learning.
- (2) Options for customization include the ability to create custom visuals and incorporate your work in other programmes.
- (3) Real-time data refresh and collaboration capabilities.
- (4) Connectivity to a wide range of data sources, including Excel, CSV, SQL databases, and cloudbased services like Google Analytics, Salesforce, and more.

## 1.3.5. Tableau

Tableau [1] is widely used to analyse and comprehend data in corporations, academia, and government organisations. It is renowned for its simplicity of use and user-friendly interface, which enables even non-technical people to produce images of a high calibre. Overall, Tableau is a useful tool for data

analysis and visualisation that can provide you new perspectives and enable you to make wiser business decisions.

# 1.3.6. IBM Watson Analytics

IBM Watson Analytics [1] is a cloud-based data visualisation and business intelligence software that assists you in analysing and comprehending your data. It automatically finds insights and creates predictive models using machine learning algorithms and natural language processing.

Watson Analytics is made to be simple to use, even for individuals without any prior data analysis knowledge. It provides a selection of pre-built dashboards and templates to assist you in getting started quickly, as well as a number of tools for cleaning and exploring data.

# 2. Introduction to Time Series

Time series [3] is an orderly sequence of values that are uniformly spaced over a time stamp. There are 2 types of time series one is uni-varient time series and other is multi-varient time series. If there is a single value produced at each time stamp then it is called uni-varient time series; and if there are multiple values produced at each time stamp then it is called multi-varient time series. Multi-varient time series is used to understand related data. For example, male vs female population over a period of 1 year. Time series is a small part of sequence modelling. It can be continuous and discrete, continuous values are measured at every instance and discrete values are measured at discrete point of time. Sequence modelling means we need to produce a sequence of values from a certain input this input can be time series.

## 2.1. Components of Time Series

There are 4 components of which affect the time series:

## (1) Trends

The tendency of time series to fluctuate that is increase, decrease and stay stagnant over a long period of time is called Simple trend or Secular trend.

## (2) Seasonal

The seasonal variation in the time series are increasing and decreasing over period a year then it is seasonal component. Example of the seasonal component can be understand by simple change in weather, sales of products that is ice-cream sold in summer is more as compared to winter.

## (3) Cyclical

Cyclical fluctuations are the period changes in the time series which are repeated again and again. This is completely different from seasonal variations. For example, fluctuations in stock prices are cyclic not seasonal.

## (4) Irregular

Irregular variations are the random fluctuations in time series which are unpredictable events, these events are not periodic, do not occur in cycle or they are not seasonal. The best example of this is the stock price increase or decrease due to a good or bad news. Another example is of occurrence of tsunami, earthquakes, etc.

#### **IJIRMPS**

# 2.2. Need of Time Series Analysis

Time series analysis is a specific way of fitting the time series data to a model. This analysis is use to do future prediction and to predict this one needs to know the nature of time series. Time series analysis is used to develop a model which could get data from the existing time series. This model is used to predict future events. This type of approach is used when we do not have proper knowledge about statistical patterns and its successive observations. Apart from future prediction time series is also used for cleaning data and understanding data that is patterns.

# **2.3.** Application of Time Series

- (1) **Finance:** Time series models are used in financial analysis to forecast stock prices, interest rates, and other financial metrics.
- (2) Weather Forecasting: Time series models are used to predict weather patterns and forecast future temperatures, precipitation, and other meteorological data.
- (3) **Demographics:** Time series analysis is used to study population trends and forecast future population growth.
- (4) **Healthcare:** Time series analysis is used to study patient outcomes and to predict future health trends.
- (5) **Marketing:** Time series analysis is used to analyze sales data and forecast future demand for products and services.

# 3. Conclusion

Data play a crucial role in today's most of the business decisions, future predictions, pattern recognition and many more. For this, data visualization and time series are very much important as data visualization displays complex data in the simplest form using various techniques, tools and libraries. Whereas, in time series data cleaning, understanding patterns and predicting future values are carried out which is affected by the 4 main components that are Trends, Seasonal variation, Cyclical variation and Irregular variation. This paper summarize all the information that are necessary for Data Visualization and Time Series analysis.

# References

- Diamond, M.P., & Mattia, A. (2015). Data Visualization: An Exploratory Study into the Software Tools Used by Businesses. Journal of Instructional Pedagogies, 17. <u>http://files.eric.ed.gov/fulltext/EJ1151731.pdf</u>
- 2. Research on Various Tools in Big Data. (2019). International Journal of Innovative Technology and Exploring Engineering, 8(6S4), 1097–1100. <u>https://doi.org/10.35940/ijitee.f1228.0486s419</u>
- Fang, Y., Xu, H., & Jiang, J. (2020). A Survey of Time Series Data Visualization Research. IOP Conference Series: Materials Science and Engineering, 782(2), 022013. <u>https://doi.org/10.1088/1757-899x/782/2/022013</u>
- Oberoi, A., & Chauhan, R. (2019). Visualizing data using Matplotlib and Seaborn libraries in Python for data science. International Journal of Scientific and Research Publications (IJSRP), 9(3), 8733. <u>https://doi.org/10.29322/ijsrp.9.03.2019.p8733</u>
- Waskom, M. (2021). Seaborn: Statistical Data Visualization. Journal of Open Source Software, 6(60), 3021. <u>https://doi.org/10.21105/joss.03021</u>
- Khalid, Zhwan & Zeebaree, Subhi. (2021). Big Data Analysis for Data Visualization: A Review. International Journal of Science and Business, 5(2), 64–75. <u>https://doi.org/10.5281/zenodo.4462042</u>