

## DATA VISUALIZATION: A STUDY OF TOOLS AND CHALLENGES

<sup>1</sup>Dr Rasheedul Haque\*, <sup>2</sup>Dr Ahmad Tasnim Siddiqui, <sup>1</sup>Lee Har San, and <sup>1</sup>Fakir Mohamed  
Bin Omar Din

<sup>1</sup>MAHSA University, Malaysia.

<sup>2</sup>Sherwood College of Professional Management, India.

\*Email: [rasheedul@mahsa.edu.my](mailto:rasheedul@mahsa.edu.my)

### ABSTRACT

In recent years, scientific research and publication are getting more advanced and accurate, and publishing is getting more professional and clear. Now a day everyone is talking about data visualization. Data visualization is a very encouraging area. Almost every company and educational institution is using data visualization tools and techniques to display beautiful and attractive data in the pictorial format. They are also extensively using data visualization for annual reports, research reports, and grants, etc. Data visualization helps to elucidate facts, and figures and decide sequences of actions. Any field of study will be benefited which requires advanced techniques for presenting large and complex information. There is a number of good data visualization tools available which provide an accessible technique to see, understand and explain trends and patterns in data. This paper summarizes the tools and the major challenges of data visualization.

**Keywords** – Data visualization, Graphics, Information visualization, Reports, Tools

### I. INTRODUCTION

We live in a society of information technology and there has been the requirement of presenting enormous amounts of data in such a manner that can be comprehensive and easily available. Visualization is a pictorial representation of data that aims clear understanding of the user. Every organization is producing a large amount of data every day. Due to this reason, the amount of data storage on the web has increased intensely. It is very challenging for users to explore, visualize, and use this huge amount of data. Data visualization is directly related to the design implementation, development, and applications of computer-made pictorial representations of the data. This permits management people to see the analysis in visualized format and makes it easy to understand the data. It also supports them to understand patterns, and information, and make an opinion (Sadiku, 2016).

These days data visualization contributes a clear idea of what the data and information mean by giving it a visual look through the chart, maps or graphs. This makes the data more understandable and natural for our minds. It makes it easier to find out the patterns and trends within huge data sets. Showing information in a beautiful manner is an art. Our eyes are made

to see and identify colorful and beautiful patterns around us. Unfortunately, across the world, many facts and figures are presented incorrectly or partially correct and they use data visualization practices (Midway, 2020). It is very much important to be aware of the best tactics to create the most effective data visualization. Visualized learning is very effective and one of the basic forms of information interpretation, which generally combines the images and texts in the form of graphs and charts with reading text (Stirling, 1987). The new data visualization techniques should come with better methods to analyze, process and visualize the data. In recent times, technology has provided an advancement to create better visual presentations in terms of quick and complex creation of data visualization.

Data visualization or information visualization enables you to interact with data, it also gives life to our data by making it visual. Over the last couple of decades, research and publication is getting more advanced, accurate and clear. Especially, publishing is getting more professional and clear (Goodman et. al. 2018). To make a beautiful and effective visual normally zero or very less coding skill is required. We are lucky enough that we have different types of data visualization tools available which are being used to create beautiful and effective visuals such as charts, graphs, maps, dashboards and infographics, etc. We can use Microsoft Excel as a data visualization tool. People are not aware that MS-Excel has enormous capability to visualize the data. It has properties like PivotTables, Charts, slicers and more. But for more advanced visualization there are tools like Power-BI, Tableau, Info gram, Chart Blocks, Google Charts, Datawrapper, Plotly, SAP Analytics Cloud, D3.js, and NVD3, etc. These are the tools that can help you to get a clean and interactive visualization of all the data. They can produce an effective visual within a minute but we need to choose the correct tool according to our data (L' Astorina, (2017).

## II. DATA VISUALIZATION TECHNIQUES

Data visualization is achieved by computer-supported tools to represent the visuals. Visualized data supports interactive data visualization which allows users to choose the format of their choice to display data. Commonly used types of data visualizations are (Data visualization beginner's guide, tableau.com):

- Line graph: This type of technique displays relationships between items.
- Bar chart: It is used to compare the quantities
- Pie chart: Is used to compare a portion of the whole chart
- Scattered plots: It is 2-D plot that shows variations.
- Tables: Represents normal and interactive table
- Maps: It displays maps such as heat map
- 3-D surfaces and solids: It shows surfaces or shapes such as cone
- Gantt chart

- Box-and-whisker Plots
- Bubble Cloud
- Bullet Graph
- Cartogram
- Circle View
- Histogram
- Timeline etc.

So, the layout of charts and graphs can take the form of a line graph, bar chart, scattered plot, pie chart, etc. It is very important to understand what type of chart or graph is suitable to use visualization for your data. Data is abstracted and summarized in the data visualization process. Key elements in the data are spatial variables such as position, size, and shape. A visualization system should follow the steps like preparation, cleaning, organization and execution to get the dataset. It can visualize the result in the form of graphs and charts and display the output in a user-friendly manner.

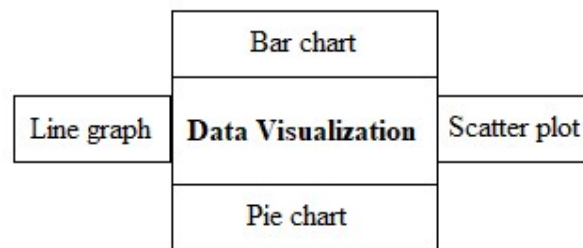


Fig. 1: Common techniques of Visualization

Most of the visuals are commonly used for decision-making. Data visualization is not only the representation of numbers, but it involves selecting the correct and cleaned data on which we are thinking to visualize. We have a wide range of application areas for data visualizations. It can be used in producing research-related data, health data, energy consumption, educational budgetary analysis, fraud detection representation and many more (Sadiku, 2016). Many data visualization algorithms and software tools have been developed. These tools provide a facility for users to interpret data more swiftly than earlier.

These tools include ManyEyes owned by IBM, Smart Money for the stock market, Insights owned by Facebook Corporation, Visual Analytics by SAS, and Thoth developed by California Institute of Technology, Tableau, Power-BI developed by Microsoft, Google charts by Google and TOPCAT, etc (Chen, 2010, Laher, 2016). These tools make data visualization very easy to understand, interpret and produce rapidly. Each and every tool has its own unique features and limitations.

Apart from all the above tools, we can also create a visualization using Python programming. Python provides multiple great graph designing libraries that are packed with large numbers of

different features. If you are well versed in Python programming and you want to produce live, interactive and highly customized plots python has excellent libraries for you. Here are a few popular plotting libraries:

- Area: It is for area plots
- Bar: It is for vertical bar charting
- Barh: Used for horizontal bar charting
- Matplotlib: It is low level and it provides lots of freedom
- Pandas Visualization: It is easy to use interface, built on Matplotlib
- Seaborn: It is a very high-level interface and has great default styles
- Ggplot: It is based on R's ggplot2. It uses the Grammar of Graphics
- Hist: Can be used for histogram
- Plotly: It can create interactive plots
- Scatter Plot: To create scatter plots in Matplotlib

The code snippet below is responsible to generate the visualization:

#### A. Examples with Python

```
#Example to draw line plot
```

```
import sys
```

```
import matplotlib
```

```
matplotlib.use('Agg')
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
ypoints = np.array([3, 6, 1, 9])
```

```
bpoints = np.array([5, 5, 1.5, 8])
```

```
plt.title('Line by Ahmad Tasnim')
```

```
plt.plot(ypoints, color = 'g')
```

```
plt.plot(bpoints, color = 'b')
```

```
plt.show()
```

```
#These lines to make our compiler able to draw:
```

```
plt.savefig(sys.stdout.buffer)
```

```
sys.stdout.flush()
```

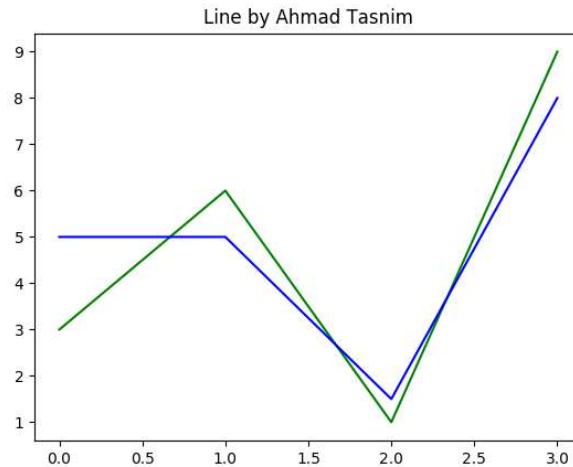


Fig. 2. Line plot created by Matplotlib using Python

*Example 2:*

```
# create color dictionary
colors = {'Iris-setosa':'r', 'Iris-versicolor':'g', 'Iris-virginica':'b'}
# create a figure and axis
fig, ax = plt.subplots()
# plot each data-point
for p in range(len(iris['sepal_length'])):
    ax.scatter(iris['sepal_length'][p], iris['sepal_width'][p], color=colors[iris['class'][p]])
# set a title and labels
ax.set_title('Iris Dataset')
ax.set_xlabel('sepal_length')
ax.set_ylabel('sepal_width')
```

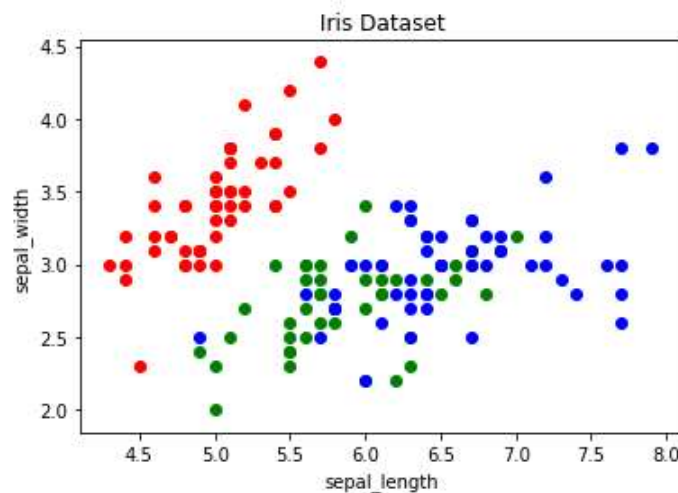


Fig.3. Scatter plot created by Matplotlib using Python (Tanner, 2019)

We can use data visualizations for:

- Identifying trends and outliers inside a data set
- Making data attractive and easily understandable
- Emphasizing important parts of a data set
- Telling a story found within the data
- Highlighting an argument or belief
- Make more interactive books, blog posts, reports, etc.

### B. Creating Report in Power-BI

After finishing with reports in Power BI Desktop, I would like to install the version of Power BI Desktop which is optimized for Power BI Report Server, because you know the server and the app are always in sync. We can have both versions.

- We can connect to a variety of data sources. From the welcome screen, select the Get Data option and import the file/files you want to. I'm using an On-Street Crime dataset which is downloaded from Kaggle.
- Now, get to the Data View and make the necessary alterations in the data we've imported.
- We can now begin designing the report.
- After things are done, we need to publish the report on the Power BI website.



Fig 4. Report generated by Power-BI (Komal, 2020)

### III. COMMON FEATURES IN BEST DATA VISUALIZATION TOOLS

We know that a lot of tools for data visualization are available, but there are few things in common in the best data visualization tools. First and foremost is the ease of use. There are many typically complicated apps present to visualize data. Some tools have outstanding tutorials and documentation and they are designed in such a way that users feel natural to use.

The best tools can handle massive data sets. In fact, they can even perform multiple and complex sets of data in a single visualization process.

They can provide an output of different charts, graphs, and map types. Many tools can produce output both images as well as interactive graphs. There are some exemptions to the range of output criteria; however, some visualization tools emphasize an explicit type of graph, chart or map and produce it superbly. These tools are also referred to as among the “top” tools.

Finally, it is necessary to be cost-effective. A higher price range does not mean eliminating a tool, but the higher price label should be justified in relation to better features, better support, and overall great value. Some crucial features any data visualization tool possesses are:

- Decision-making Ability
- Integration Capability
- Highly-effective Infrastructure
- Prompt Discovery of Results or Insights
- Real-time Team Collaboration

#### **IV. CHALLENGES IN DATA VISUALIZATION**

“What anybody can think the most significant issue in data visualization is?”

The answers mostly landed in three major categories:

- Issues around data,
- Visualization in practice, and
- The general profession.

Data visualization has changed the scenario and it is facing a great challenge due to the massive amount of data. Data visualization of large and complex data has become a demanding task (Aggarwal, et. al. (2020). From the very simple project to complex projects there are lots of things changed. The approach and understanding of data are changed. Due to changes, we need technically sound users who can tackle a large amount of complex data to produce better visualization. Finding the correct and useful data and preparing the data set is a challenge in visualizing the data. It is not easy to enumerate the quality of the data visualization technique. Erroneous, missing and duplicate entries should be identified and corrected data sets should be available to process. We should also think about the size, speed, real-time scalability, interactive scalability, perceptual scalability and the variety of the data (Sadiku, 2016). Recent years have seen a tremendous rise in real-time complex data. It is a huge challenge to cope with increasing data in static or dynamic form.

## V. CONCLUSION

This paper provides information on the field of data visualization. After the discussion, we can say that data visualization is the method to represent data pictorially, clearly and effectively. It emphasizes not only how to visualize but also how visualization can be helpful (Kaidi, 2004). It has appeared as a very powerful, widely acceptable and applicable tool to analyze and understand huge and complex data. Data visualization has come up as a rapid, easy method of conveying ideas in a universally accepted format. It communicates complex ideas with ease, clarity, accuracy, and effectiveness. These benefits make data visualization to be useful in various fields of study. Data visualization is the need of the future to reduce the report generation effort.

## REFERENCES

- Aggarwal, R. et al. (2020). Challenges and opportunities with big data visualization. *ACM Proceedings of the 7th International Conference on Management of computational and collective intelligence in Digital EcoSystems*
- Chen, C. (2010). Information visualization. *WIRES Computational Statistics*. Vol. 2, July/August, pp. 387-403.
- Data visualization beginner's guide: A definition, examples, and learning resources. Retrieved March 10, 2021, from [www.tableau.com/learn/articles/data-visualization](http://www.tableau.com/learn/articles/data-visualization)
- Goodman, A. A., Borkin, M. A. M. A. and Robitaille, T. P. (2018). New Thinking on, and with, Data Visualization. *2018 Sackler Colloquium*.
- Kaidi, Z (2004). Data Visualization. *National University of Singapore*. Document version 1.0. Retrieved March 10, 2021, from [https://www.cs.uic.edu/~kzhao/Papers/00\\_course\\_Data\\_visualization.pdf](https://www.cs.uic.edu/~kzhao/Papers/00_course_Data_visualization.pdf)
- Komal (2020). Everything You Need To Know About Power BI Charts. Retrieved March 11, 2021, from <https://www.edureka.co/blog/popular-power-bi-charts/>
- L'Astorina, E. (2017). Review of 20 best big data visualization tools. Retrieved March 9, 2021, from <https://bigdata-madesimple.com/review-of-20-best-big-data-visualization-tools/>
- Laher, R. R. (2016). Thoth: software for data visualization and statistics. *Astronomy and Computing*. Vol. 17, pp. 177-185.
- Midway, S. R. (2020). Principles of Effective Data Visualization. *Cell Press USA*. Retrieved March 9, 2021, from <https://doi.org/10.1016/j.patter.2020.100141>
- Sadiku, M. N. O. et al. (2016). Data Visualization. *International Journal of Engineering Research and Advanced Technology (IJERAT)*. ISSN: 2454-6135. 02(12).
- Stirling, P. (1987). Power lines. *NZ Listener*, 13–15.
- Tanner, G. (2019). Introduction to Data Visualization in Python. Retrieved March 10, 2021, from <https://towardsdatascience.com/introduction-to-data-visualization-in-python-89a54c97fbcd>