

An Analysis of Data Visualization Tools

Vijay Gupta, PhD
Associate Professor

International School of Informatics & Management Technical Campus
Sector-12, Mahaveer Marg, Mansarovar, Jaipur – 302020, Rajasthan

ABSTRACT

Due to increase in information data is being generated very fast in day to day life. Large amount of data get collected from different organizations, which is very difficult to analyze and visualize. Data visualization is the way to present the data in a pictorial or graphical format. It helps decision makers to see analytics visually. New patterns and difficult concepts can be easily understood by data visualization. By using visual elements like charts, graphs, and maps, data visualization tools provide a convenient way to see and understand trends, outliers, and patterns in data. This paper focuses on data visualization using different tools. It also analyses and visualize the data by using R language with different data sets.

Keywords

Data Analytics, Data Visualization, R Language

1. INTRODUCTION

Data visualization is a method which aims to converse data efficiently and clearly to the user through graphical depiction. The discovery process is due to effective and efficient data visualization. It is the transitional between the human perception and quantitative context of the data. Data visualization is an indispensable component of the systematic path from data into facts and understanding. This technology is a powerful technology having a great prospective to help researchers as well as companies for building revenue decision [1].

Jacques Bertin who wrote the classic works of graphical visualization “Semiology of Graphics” states that the “transformation from numbers to insight requires two stages.”

There are seven visual variables given by Bertin to visualize the images are

- (a) Position
- (b) Form
- (c) Orientation
- (d) Color
- (e) Texture
- (f) Value
- (g) Size

Visual Processing happens in three steps

- (1) Formation of the retinal image,
- (2) Decomposition of the retinal image info into an array of specialized representations and
- (3) Refabrication of the information into object perception.

Bertin's key concept is that the image, from that the idea derives its name. An image is the elementary perceptual unit of visualization.

2. BENEFITS OF DATA VISUALIZATION

Data visualization allows users to see numerous diverse perspectives of the data. It makes it possible to interpret vast amounts of data .It also offers the ability to note exceptions in the data.

To analyze visual patterns in data, data visualizations plays very important role. Exploring trends within a database through visualization by letting analysts find the way through data and visually familiarize themselves to the patterns in the data. Data visualization can help translate data patterns into insights, making it a highly effective decision-making tool It equips users with the ability to see influences that would otherwise be difficult to find. By simplifying the appearance, Data Visualization can reduce the time and difficulty it takes to move from data to decision making.

3. TOOLS AND LANGUAGES FOR DATAVISUALIZATION

In big data and advanced analytics projects, data visualization plays very significant role.The business intelligence software has the excellent data visualization functionality, and they are very easier than conventional statistical analysis software. In conventional software analyzing any value from big data becomes difficult. Through business intelligence software maintaining data through graphs and the different chart is simple, so the analyzing of data is easy.

Tools

(a) R Shiny

To create data visualizations, R Shiny provides the web framework. It is an open source package. It develops interactive charts and applications. To create charts, interactive applications, there is no need to have a good knowledge of HTML, CSS or JavaScript. It also provides Web visuals, which are very interactive.

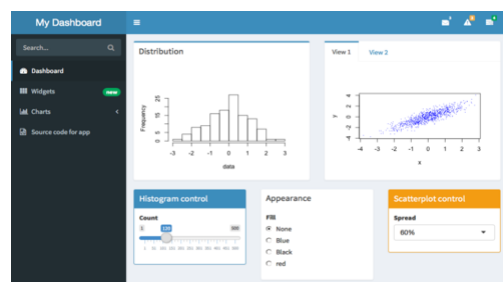


Fig.1 : R Shiny

(b) Tableau Public

It requires 10 GB storage. IT has very good drag and drop interface. Users can watch their data update in real time. New version (Tableau 2018.3) is optimized with mobile devices.

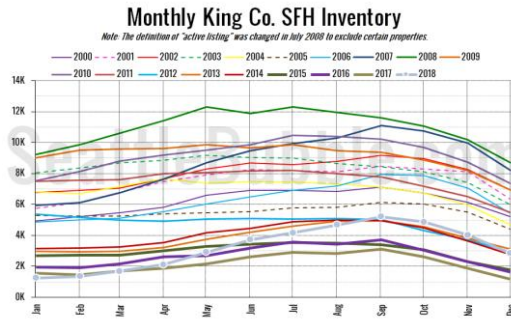


Fig. 2 : Tableau Public

(c) Datawrapper

This data visualization tool is an open source tool. It uploads data in csv format. It is used in small business organizations. Datawrapper allows 10,000 views per chart.

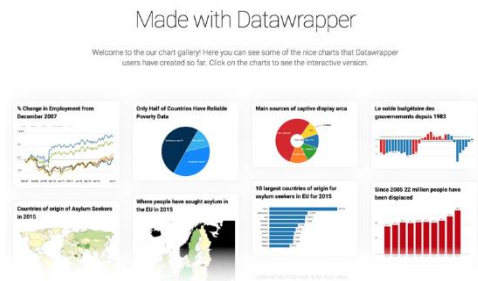


Fig. 3: Datawrapper

(d) Pivot

Pivot has an intuitive interface. It is used for exploratory analytics. It has two operations: Filter and Split. Filter narrows the view of data and is equivalent to the “WHERE” clause in SQL. Split is very similar to SQL’s “GROUP BY” function. Grocery price/Promotional analysis and optimization can be done exceptionally good by this tool.

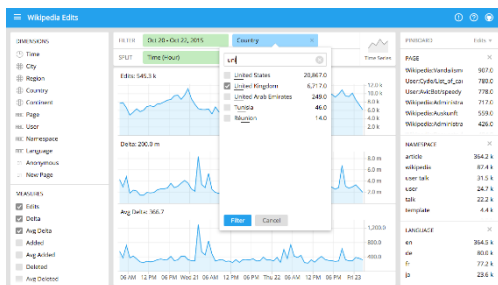


Fig. 4: Pivot

(e) D3

This tool binds arbitrary data to a Document Object Model (DOM) through JavaScript library. Document will be transformed into data-driven documents. This tool is useful for programmers. D3 is capable to create an array of maps, charts and diagrams.

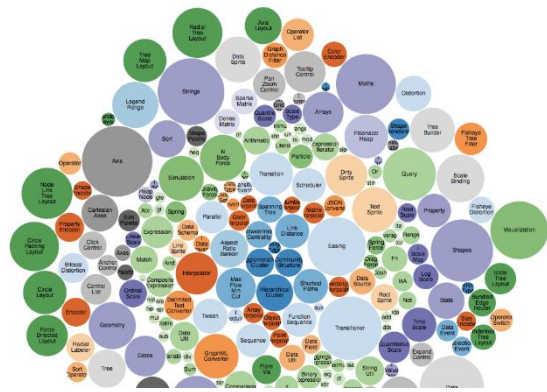


Fig. 5 : D3

(f) ChartBlocks is a user-friendly online tool. By this tool data can be visualized from spreadsheets, databases and live feeds. It does not require any coding. There is a chart building wizard which is used for creating diagrams. By using D3.js JavaScript library visualizations will be receptive and well-suited with any screen size and device.

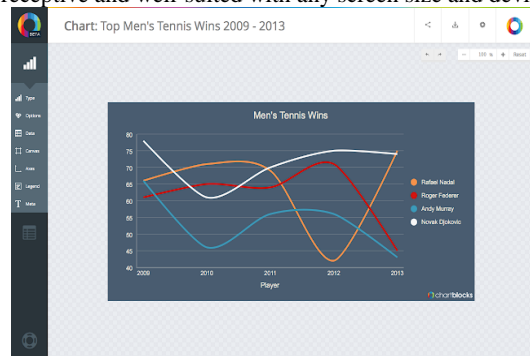


Fig. 6: ChartBlocks

(g) NVD3

To build reusable charts and components, this tool runs on the top of D3.js. The main aim of the project is to keep all charts precise and customizable. NVD3 is a simpler interface on top of D3.js and keeps all its powerful features. Front end engineers at Novus Partners have developed this tool. This tool is very useful during data visualization and in charting technology.

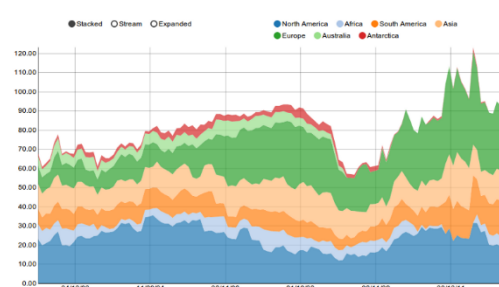


Fig. 7 : NVD3

(h) Google Charts

This visualization tool runs on HTML5 and SVG. It has cross-browser compatibility that can run on Android, iOS or any other platform. Google Charts are very user friendly and their site features are very good and comprehensive gallery where a person can see different kind of visualizations and interactions.

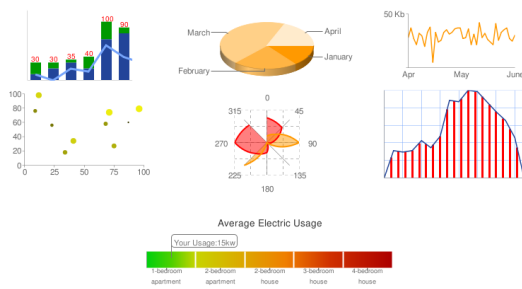


Fig. 8: Google Charts

(i) Chartbuilder

Chart builder is a well-known chart-creation tool that was made publicly available by financial news website Quartz in 2013. This tool was developed in-house. It is used to render numerical data quickly so that journalists make their stories stand out.

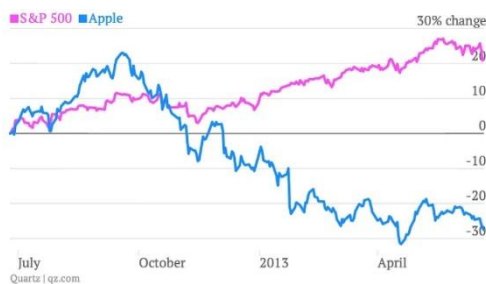


Fig. 9: Chartbuilder

4. PROGRAMMING LANGUAGES

Python language plays a significant role in data visualization because of its interesting library structure. There are a large number of inbuilt libraries, but the two exclusive libraries for data visualizations are Matplotlib and Seaborn. Matplotlib is the first data visualization library. This library has complete 2D support and unlimited 3D graphics support. This library is used to create extraordinary quality figure in an interactive environment across platforms. Animation can also be performed by this library. To create the informative and attractive statistical graphics in Python, Seaborn library can be used to perform this function. It offers different features such as colour palettes, functions and tools for visualizing the linear regression, metric of data, and statistical time series. Complex data visualization can also be performed by this library.

Language R: R is another programming language for data visualization. R language supports 4 graphics systems viz. base graphics, lattice graphics, grid graphics and ggplot2. Large data can be easily visualized using these graphics system. Base graphics is the simple to understand and provides a various useful tools for creating the descriptive tools. By using R programming language, complex plots can be developed which saves a lot of time. The R language has superb capability of building 3D graphs.

Java: The data visualization through Java language become simple because of the different inbuilt library available in Java such as Java 2D, Java 3D, and Java advanced imaging. Because of an inbuilt library, an application for data visualization takes very less amount of time. Interactive charts can be added to web-based application. It designs thousands of possible chart and style combination easily, and powerful chart definition language allows for unlimited customization. For interactive and joint visualization analysis of numerical data VisAD, a Java component Library plays an important role.

Language C#: C# is the object-oriented programming language developed by Microsoft with .NET framework. This language has 2d, and 3d API which allows to represents a large amount of data in a graphical structure. Through this API you can represent memory buffer sharing and optimise for fast general 3D visualization. You can manage some options using this API such as 3D scatterplot maps, charts, and histograms. You can also optimise for application development, interactivity, and flexibility. You can find freelance jobs online in this sector.

5. CASE STUDY USING R LANGUAGE

To visualize the data using R , a data set has been used .This data set is of ‘ Car Road Test’.

This data frame has 11 numeric variables.

kmp1	Km/liter
cyl	Number of cylinders
disp	Displacement
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (1000 Kg)
qsec	1/4 Km time
vs	Engine (0 = V-shaped, 1 = straight)
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

There are two research questions

- (a) Which Car with automatic or manual transmission is better in term of Km per liter (Kmpl)?
- (b) Quantify the kmp1 difference between automatic and manual transmission.

To find the answers of these questions, Univariate analysis on target variable (kmp1) has been applied

6. UNIVARIATE ANALYSIS

- (a) This analysis focus on investigating the target variable (kmp1) alone by splitting the observations into two groups, i.e. cars with automatic or manual transmission.
- (b) Compute sample means by group (i.e. transmission automatic vs. manual).
- (c) Create a subset of the data by group.

```
cars_automatic = subset(mcars, am == 0)
cars_manual = subset(mcars, am == 1)
dim(mcars)
-> [1] 32 11
dim(cars_automatic); dim(cars_manual)
[1] 19 11
[1] 13 11
# sample means kmp1 by group
```

```
mean(cars_automatic$kmpl); mean(cars_manual$kmpl)
```

```
[1] 17.147
```

```
[1] 24.392
```

```
# Standard Deviation kmpl by group
```

```
sd(cars_automatic$kmpl)
```

```
sd(cars_manual$kmpl)
```

```
[1] 3.833
```

```
[2] 6.166
```

```
# % increase in kmpl based on sample mean
```

- $(\text{mean}(\text{cars_manual\$kmpl}) - \text{mean}(\text{cars_automatic\$kmpl})) / \text{mean}(\text{cars_automatic\$kmpl})$

- [1] 0.4225

Results

- kmpl empirical mean of cars with manual transmission is greater than cars with automatic transmission, however has also higher variance.
- On average cars with manual transmission have 42% more mileage than cars with automatic transmission

7. VISUALIZATION USING R

Distribution Kmpl-automatic transmission

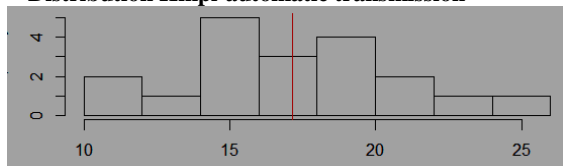


Fig. 10 Distribution Kmpl-manual transmission



Fig. 11 Kmpl vs hp-by transmission type

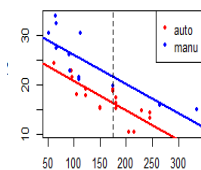


Fig.12 Kmpl vs weight- by transmission type

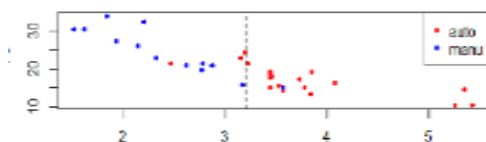


Fig.13

8. CONCLUSION

Data visualization is the conception which is useful for displaying data and information in the form of graphical charts, figures and bars. Data visualization helps in business intelligence. This paper describes how the data can be visualized by using different tools. By these tools and languages the organizations can improve their performance and also find the place where there is possibility of increasing business and also give analysis.

9. REFERENCES

- [1] Jiawei Han, Micheline Kamber and JianPei(2009). Data Mining Concepts and Techniques, Third edition, MK Publications.
- [2] <https://seattlebubble.com/blog>
- [3] <https://rstudio.github.io/shinydashboard/>
- [4] <http://nvd3.org/examples/>
- [5] <https://www.freelancinggig.com/blog/2017/10/03/best-programming-languages-data-visualization/>
- [6] Gupta, R., Gupta, S., & Singhal, A. (2014). Big data: Overview. International Journal of Computer Trends and Technology, 9(5), 266-268. Retrieved from <http://www.ijcttjournal.org>
- [7] Lima, M. (2011). Visual complexity: Mapping patterns of information. New York, NY: Princeton Architectural Press.
- [8] Mackinlay, J. D., & Winslow, K. (2014). Designing great visualizations. Seattle, WA: Tableau. Retrieved from <https://www.tableau.com/whitepapers/designing-great-visualizations>
- [9] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2012). Empirical studies in information visualization: Seven scenarios. IEEE Transactions on Visualization and Computer Graphics, 18(9), 1520-1536. doi: 10.1109/TVCG.2011.279
- [10] Keim, D., Qu, H., & Ma, K.-L. (2013). Big-data visualization. Computer Graphics and Applications, IEEE, 33(4), 20-21. doi: 10.1109/MCG.2013.54.
- [11] Endert, A., Fiaux, P., & North, C. (2012). Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. Visualization and Computer Graphics, IEEE Transactions on, 18(12), 2879-2888. doi: 10.1109/TVCG.2012.260
- [12] Few, S. (2015). Visual business intelligence. Visual Business Intelligence Blog. Retrieved from <http://www.perceptualedge.com/blog/?p=1277>
- [13] Lima, M. (2015). Manuel Lima: A visual history of human knowledge. Retrieved from https://www.ted.com/talks/manuel_lima_a_visual_history_of_human_knowledge#t-371736