



Big Data Visualization: Tools and Technique

Kiran Choudhary, Gajanand Sharma

Kiran Choudhary, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

Choudharykiran004@gmail.com

Gajanand Sharma, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

Gajanan.sharma@gmail.com

DOI 10.5281/zenodo.1186152

Abstract: Data Visualization is the process of presenting raw data into graphical format. It speeds up the process of data analysis. Visualization provide well designed visual encoding can exterminate cognitive calculation with ingenious perceptual conclusions and improve comprehension, memory, and decision making. Visual representations help people analytic thinking, because our brains process visual information efficiently. Data Visualization help to quickly interpret the complication of understanding data in numerical form.

Keywords: Data, Data Visualization, Data into Graphical Form.

I. BIG DATA VISUALIZATION

Big Data Visualization allows visual representation of data patterns, enabling user for more information from data. Big Data Visualization involves the presentation of data in graphical format that makes it easy to understand and interpret. Big Data Visualization technique seize the opportunity this fact :they are all about turning data into pictures by presenting data in pictorial or graphical format. This make it easy to understand for decision makers to take vast amount of data at a glance to see what is going on, what it is that the data has to say. There are three factor on the basis of which we can classify the data visualization techniques today. Kind of data, technique used and inter-functionality.

II. BIG DATA VISUALIZATION TOOLS

Big Data Visualization Tools to make sense from clustered data.

Tableau

Tableau is the most popular and most innovative data visualization tool available today, Tableau is data visualization software that help people to see and understand their data. Tableau is one of the fastest involving Business Intelligence and data visualization tool. It is very fast to deploy, easy to learn and very intuitive. Tableau is for manipulating big data and its available to everyone.

It has two other variants 'Tableau Server' and cloud based 'Tableau Online', which are dedicatedly designed for big data related organization. You don't have to be a ode to use this tool. This tool is very handy and provide fast speed.

QlikView

QlikView is a business discovery platform that provides self-services BI for all business users in organizations. With QlikView You can analyze data and use your discoveries to support decision making. QlikvView lets you ask and answer

your questions and follow your ownpaths to insight. Qlikview enables you and your colleagues to reach decisions collaboratively.

At the core of QlikView is our patented software engine, which generates new views of data on the fly. QlikView compresses data and holds it in memory, Where it is available for immediate exploration by multiple users. For data sets too large to fit in memory, QlikView connects directly to the data source. QlikView delivers an associative experience across all the data used for analysis, regardless of where it is stored. You can start anywhere and go anywhere; and are not limited to pre-defined drill paths or pre-configured dashboards.

FusionCharts

FusionCharts is a software service provider creating data visualization products. Its flagship product, FusionCharts Suite XT, is a comprehensive JavaScript-based charting solution that focuses on easing the process of data visualization through charts. With over 90 charts and 1000 maps, FusionCharts Suite XT covers a wide range of data visualization requirements.

Highcharts

Like FusionCharts this also requires a license for commercial use, although it can be used freely as a trial, non-commercial or for personal use. Its website claims that it is used by 72 of the world's 100 largest companies and it is often chosen when a fast and flexible solution must be rolled out, with a minimum need for specialist data visualization training before it can be put to work. A key to its success has been its focus on cross-browser support, meaning anyone can view and run its interactive visualizations, which is not always true with newer platforms.

Datawrapper

Datawrapper is increasingly becoming a popular choice, particularly among media organizations which frequently use it to create charts and present statistics. It has a simple, clear interface that makes it very easy to upload csv data and create straightforward charts, and also maps, that can quickly be embedded into reports.

Plotly

Plotly enables more complex and sophisticated visualizations, thanks to its integration with analytics-oriented programming languages such as Python, R and Matlab. It is built on top of the open source d3.js visualization libraries for JavaScript, but this commercial package (with a free non-commercial license available) adds layers of user-friendliness and support as well as inbuilt support for APIs such as Sales force.



Sisense

Sisense provides a full stack analytics platform but its visualization capabilities provide a simple-to-use drag and drop interface which allow charts and more complex graphics, as well as interactive visualizations, to be created with a minimum of hassle. It enables multiple sources of data to be gathered into one easily accessed repositories where it can be queried through dashboards instantaneously, even across Big Data-sized sets. Dashboards can then be shared across organizations ensuring even non technically-minded staff can find the answers they need to their problems.

III. HANDLING LARGE DATA VOLUMES

Data volume can become an issue because traditional architectures and software may not be able to process huge amounts of data in a timely manner, thus requiring you to make compromises and aggregate the details you want to visualize. Even the most common descriptive statistics calculations can become complicated when you are dealing with big data and don't want to be restricted by column limits, storage constraints and limited support for different data types. The SAS in-memory platform solves these issues by speeding up the task of data exploration, and a visual interface displays the results in an easy-to-understand visualization.

IV. VISUALIZING SEMI STRUCTURED AND UNSTRUCTURED DATA USING WORD CLOUDS AND NETWORK DIAGRAMS

The variety of big data brings challenges because semi structured and unstructured data require new visualization techniques. A word cloud visual (where the size of the word represents its frequency within a body of text) can be used on unstructured data as a way to display high- or low-frequency words (see Figure 7). SAS Visual Analytics takes the concept of word clouds a step further by taking advantage of taxonomies and ontologism to make associations. Words are then organized into topics based on how the words are used. SAS Visual Analytics word clouds can display the hot topics of the day gleaned from such text analysis. Users can drill down by clicking on an individual topic to see exactly what words or phrases comprise that topic.

V. VISUALIZATION WITH CORRELATION MATRICES

Velocity is all about the speed at which data is coming into the organization. The ability to access and process varying velocities of data quickly is critical. A correlation matrix combines big data and fast response times to quickly identify which variables are related. It also shows how strong the relationships are between variables. SAS Visual Analytics makes it easy to assess the relationships. Simply select a group of variables and drop them into a visualization pane. The intelligent auto charting function displays a color-coded correlation matrix that quickly identifies strong and weak relationships between the variables. Darker boxes indicate a stronger correlation; lighter boxes indicate a weaker correlation (see Figure 9). If you hover over a box, a summary of the relationship is shown. You can double-click on a box in the matrix for further details.

VI. FILTERING BIG DATA

When working with large amounts of data, being able to quickly and easily filter your data is important. What if you only want to view data for a certain region, product line or some other variable? SAS Visual Analytics has filtering capabilities that make it easy to refine the information you see. Simply add a measure to the filter pane or select one that's already there, and then select or deselect the items on which to filter. But what if the filter isn't meaningful or it skews the data in undesirable ways? One way to better understand the composition of your data is through the use of histograms. Histograms provide a visual distribution of the data along with cues for how the data will change if you filter on a particular measure. Histograms save time by giving you an idea of the effect the filter will have on the data before you apply it. Rather than relying on trial and error or instinct, you can use the histogram to help you decide what to focus on.

CONCLUSION

Visualizing your data can be both fun and challenging. It is much easier to understand information in a visual compared to a large table with lots of rows and columns. However, with the many visually exciting choices available, it is possible that the visual creator may end up presenting the information using the wrong visualization. In some cases, there are specific visuals you should use for certain data. In other instances, your audience may dictate which visualization you present. In the latter scenario, showing an alternative visual that conveys information more clearly may provide just what's needed to easily grasp the most relevant factors affecting important decisions. You can choose the most appropriate visualization by understanding the data and its composition, what information you are trying to convey to your audience, and how viewers process visual information. Ease of use is key to getting the most from data visualization tools, which in turn can help you realize value from your other BI and analytics investments.

REFERENCES

- [1]<https://www.ijarce.com/upload/2016/february-16/IJARCE%2015.pdf>
- [2]<https://curate.nd.edu/show/zs25x636372>
- [3]<https://www.datamation.com/big-data/big-data-visualization.html>
- [4]Kiran kumara Reddi & DnvsI Indira "Different Technique to Transfer Big Data : survey" IEEE Transactions on 52(8) (Aug.2013) 2348 { 2355 }
- [5]Umasri.M.L, Shyamalagowri.D ,Suresh Kumar.S "Mining Big Data:- Current status and forecast to the future" Volume 4, Issue 1, January 2014 ISSN: 2277 128X
- [6]www.sas.com
- [7]MarcinJedyk, MAKING BIG DATA, SMALL, Using distributed systems for processing, analysing and managing large huge data sets, Software Professional s Network, Cheshire Data systems Ltd
- [8]<http://www.forbes.com/sites/bryanpearson/2015/04/10/exercise-inservice-fitbit-omni-channel-begs-for-omni-prescience/>
- [9]http://en.wikipedia.org/wiki/Parkinson's_law
- [10]<http://www.news-sap.com/sentiment-analysis-with-big-data/>