

# Big Data And Data Visualization

Pravin Deshmukh  
Indiana University  
300 N. Jordan Avenue  
Bloomington, Indiana 47405-1106  
praadesh@iu.edu

## ABSTRACT

This article provides an overview on importance of data visualization in presenting findings of Big Data solutions. Data visualization is very important to understand Big Data analytic.

## KEYWORDS

Data Visualization, Big Data

## 1 INTRODUCTION

Big data is widely used technology to consume huge amount of data. While there are various technologies available to process this data, for human it becomes very difficult to extract meaningful information when data becomes extremely large. Not many conventional visualization tools are designed to present meaningful and quality information for human perception. Hence it becomes very important to have interactive, intuitive and user friendly data visualizations in place so that decision makers, business users will have clear understanding of findings of big data solutions. These visualizations will help business users to make informed decision looking at various trends over the period of time.

## 2 BIG DATA

Big data by definition refers to the any large quantity of raw data that can be collected, stored and analyzed through various means. Big Data is used for data sets that are so large and complex that traditional data processing tools are inadequate to store and process them. Big data challenges includes capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying updating etc.

Big data has become more relevant recently because of the data exponential data growth happened in last decade or so. 90% of the data exists today is created in last 2 years[7]. Following are some interesting facts regarding this data explosion:

- 2.7 Zetabytes of data exist in the digital universe today
- Facebook stores, accesses, and analyzes 30+ Petabytes of user generated data
- Brands and organizations on Facebook receive 34,722 Likes every minute of the day
- Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes of data
- More than 5 billion people are calling, texting, tweeting and browsing on mobile phones worldwide
- 571 new websites are created every minute of the day

[4]

With this kind of data growth, it has become challenging for companies to process data to cleans data, identify good or bad data and produce meaningful outcome.

## 3 DATA VISUALIZATION

Data Visualization is representation of the data in a visual context which helps users to easily understand the significance of the data. A primary goal of data visualization is to communicate information clearly and efficiently using visuals like statistical graphs, plots and information graphics[6]. Trends, patterns and correlation between entities which are not evident in in text-based data or numbers can be exposed and recognized with the help of data visualization. It makes complex statistics more accessible, understandable and usable[6]

Big data visualization is not as simple as it used to be with traditional smaller datasets. In Big data data visualization, many data scientists use feature extraction and geometric modeling to drastically reduce data size before actual data rendering. Choosing proper data representation is also very important when visualizing big data.[5]

## 4 CHALLENGES OF DATA VISUALIZATION

Visualization of large data set is a quite demanding task. The conventional ways of presenting data have limitations because of the constant growth of the data. Modern Data visualization techniques have greatly evolved over the period of time. These modern visualization techniques and tools are designed to deal with following challenges posed by big data:

- Volume : As mentioned above, rate of data growth has increased significantly in last few years. These tools are designed to process huge amount of data and allows users to derive meaningful outcome from very large amount of data
- Velocity : With fast moving business dynamics, results are expected to be delivered real-time or near real time time frames hence tool are designed to process data in real time instead of batch processing
- Variety : With the introduction of social media sites, requirement of processing variety of data has increased like never before. These tools are developed to integrate with wide variety of data sources which includes structured data, semi-structured data and non-structured data.

## 5 DATA VISUALIZATION IS THE KEY TO ACTIONABLE INSIGHTS

There is a old saying “picture is worth a thousand words” That’s because an image can often convey “what’s going on”, more quickly,

more efficiently, and often more effectively than words. Big data visualization techniques exploit this fact: they are all about turning data into pictures by presenting data in pictorial or graphical format. This makes it easy for decision-makers to take in vast amounts of data at a glance to "see" what is going on.

Visualization allows business to take complex findings and present them in a way that is informative and engaging to all stakeholders. A strong understanding of data science is required for that visualization to be successful. We must all remember that in the end, the consumer of the product of all artificial intelligence or machine learning endeavors will be people. We should ensure results are delivered as actionable, impactful insights to act upon in business and in life. The human brain is only able to process two to three pieces of information at a time and many different aspects of consumer behavior are influenced by more than just two or three events. This means you have to utilize advanced analytics and statistical modeling to accurately predict consumer behavior and Key Performance Indicators (KPIs) for businesses.

Table 1 shows benefits of data visualization

Benefits	Percentage(%)
Improved decision-making	77
Better ad-hoc data analysis	43
Improved collaboration and information sharing	41
Provide self-service capabilities to business users	36
Improved return on investment (ROI)	34
Time savings	20
Reduced investment in IT	15

[5]

## 6 DATA VISUALIZATION TOOLS

Following are some popular and most widely used big data visualization tools :

- **Cognos Analytics:** Driven by their commitment to Big Data, IBM's analytics package offers a variety of self-service options to more easily identify insight. Cognos Analytics, an interactive way to search, explore, and share data-driven insights in a governed environment. Find precise and timely answers from your data or from content built by others. Create compelling reports and dashboards which you can easily distribute. Use automated alerts to monitor changes to key findings.
- **QlikView:** The Qlik solution proves its ability to perform the more complex analysis that finds hidden insights. Qlik data analytics platform can help companies gain the most leverage from a Big Data implementation by easing access and making Big Data both relevant and in-context for the organization's business users.
- **Microsoft PowerBI:** The Power BI tool enables you to connect with hundreds of data sources, then publish reports on the Web and across mobile devices. Power BI is a suite of business analytics tools that deliver insights throughout your organization. Connect to hundreds of data sources, simplify data prep, and drive ad hoc analysis. Produce beautiful reports, then publish them for your organization

to consume on the web and across mobile devices. Everyone can create personalized dashboards with a unique, 360-degree view of their business. And scale across the enterprise [1]

- **Oracle Visual Analyzer:** A web-based tool, Visual Analyzer allows creation of curated dashboards to help discover correlations and patterns in data.
- **SAP Lumira:** Calling it itself service data visualization for everyone, SAP Lumira allows you to combine your visualizations into storyboards. SAP Lumira is data visualization software that makes it easy to create beautiful and interactive maps, charts, and info-graphics. Import data from Excel and many other sources, perform visual BI analysis using intuitive dashboards, and securely share insights and data stories.
- **SAS Visual Analytics:** The SAS solution promotes its "capability and governance" along with dynamic visuals and flexible deployment options. It provides single, powerful in-memory environment which allows Interactive reporting, Visual data discovery, Self-service analytics, Scalability and governance.
- **Tableau Desktop:** Tableau's interactive dashboards allow users to "uncover hidden insights on the fly," and power users can manage metadata to make the most of disparate data sources. Tableau as an organization has been dedicated to data visualization for over a decade and the results show in several areas: particularly in usability they have edge over their other competitors. Tableau's functionality from an end-user perspective is much better than their closest rivals.
- **TIBCO Spotfire:** Offers analytics software as a service, and proves itself as a solution that scales from a small team to the entire organization. Spotfire represents the state of the art in visual data discovery, analytic applications and self-service dashboard creation - and also extends its capabilities into Mobile dashboards/apps.

[2]

## 7 CONCLUSION

Today we have access to the largest amount of data we ever had. But all data in the world is useless, in fact it will become a liability, if you cannot understand it. Data visualization is all about presenting data to the right people at the right time. As more and more businesses are analyzing their data with big data tools, data visualization is becoming an increasingly important component of analytics in the age of big data. The availability of new in-memory technology and high-performance analytics that use data visualization is providing a better way to analyze data more quickly than ever. Visual analytics enables organizations to take raw data and present it in a meaningful way that generates the most value. Nevertheless, when used with big data, visualization is bound to lead to some challenges. If we are prepared to deal with these hurdles, the opportunity for success with a data visualization strategy is much greater [3]

## ACKNOWLEDGMENTS

The authors would like to thank Prof. Gregor von Laszewski and entire team of TAs for all the help and support they provided during whole process of writing this paper.

## REFERENCES

- [1] Microsoft Corporation. 2017. Business intelligence like never before. *Microsoft Corporation* (2017), 1. <https://powerbi.microsoft.com/en-us/>
- [2] Paul Rubens. 2017. Big Data Visualization. *QuinStreet, Inc* (2017), 1. <https://www.datamation.com/big-data/big-data-visualization.html>
- [3] Ken Smyers. 2013. Data Visualization: Overcome the Five Biggest Challenges of Big Data! *Digital Technologies* (2013), 1. <http://controlltrends.org/by-event/event-coverage/in-the-community/04/data-visualization-overcome-the-five-biggest-challenges-of-big-data/>
- [4] Waterford Technologies. 2016. Big Data Statistics & Facts for 2017. Web. *Waterford Technologies* (07 2016). <https://www.waterfordtechnologies.com/big-data-interesting-facts/>
- [5] Lidong Wang, Guanghui Wang, and Cheryl Ann Alexander. 2015. Big Data and Visualization: Methods, Challenges and Technology Progress. *Digital Technologies* 1, 1 (2015), 33–38. <http://pubs.sciepub.com/dt/1/1/7>
- [6] Wikipedia. 2017. Data visualization. Web. *Wikipedia* (09 2017). [https://en.wikipedia.org/wiki/Data\\_visualization](https://en.wikipedia.org/wiki/Data_visualization)
- [7] Daniel Zeichner. 2016. The big data explosion sets us profound challenges - how can we keep up. Web. *Daniel Zeichner* (07 2016). <https://www.theguardian.com/science/political-science/2016/jul/02/the-big-data-explosion-sets-us-unprecedented-challenges-how-can-we-keep-up>

## 8 BIBTEX ISSUES

- Warning—no number and no volume in sept1007
- Warning—no number and no volume in sept1006
- Warning—no number and no volume in sept1005
- Warning—no number and no volume in sept1002
- Warning—page numbers missing in both pages and numpages fields in sept1002
- Warning—no number and no volume in sept1003
- Warning—page numbers missing in both pages and numpages fields in sept1003
- Warning—no number and no volume in sept1001
- Warning—page numbers missing in both pages and numpages fields in sept1001
- (There were 9 warnings)

## 9 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

### 9.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

### 9.2 Uncaught Bibliography Errors

- Missing bibliography file generated by JabRef
- Bibtex labels cannot have any spaces, \_ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

### 9.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

### 9.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use & but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

\section{Introduction} and NOT \section{INTRODUCTION}

### 9.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

### 9.6 Latex Errors

Erroneous use of quotation marks, i.e. use “quotes”, instead of " "

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % \_ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

## 9.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

## 9.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use `textwidth` as a parameter for `includegraphics`

Figures should be reasonably sized and often you just need to add `columnwidth`

e.g.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}
```