

INTERACTIVE VISUAL ANALYTICS ON BIG DATA: TABLEAU VS D3.JS

Lekha R. Nair
Sujala D. Shetty
Siddhanth D. Shetty

BITS Pilani, Dubai Campus, UAE
lekharnair@gmail.com, sujala@dubai.bits-pilani.ac.in,
siddaredevill@gmail.com.

Keywords: Big Data; Big Data Visualization; Visualization Tools; Tableau; D3.js.

Big Data from innumerable sources offers immense prospects for intelligent decision making in every single field, in spite of its challenging factors like huge volume, extreme variety or varying velocity. Big Data analytics can provide new perceptions from data, which would have otherwise unnoticed, while proper visualization of the data itself may result in solutions to Big Data problems through visual means by presenting the vital elements in the vast data in a human-interaction friendly format. But most of the traditional graphics and charting tools fail to draw the big picture from the big data that spans millions of records. This paper exhibits the effectiveness of visualization in exposing and communicating the essential facts in a large dataset using the two currently popular but divergent visualization tools, D3.js and Tableau, and performs a comparative analysis of the two.

for citations:

Nair L., Shetty S., Shetty S. (2016), *Interactive visual analytics on Big Data: Tableau vs D3.js*, Journal of e-Learning and Knowledge Society, v.12, n.4, 139-150. ISSN: 1826-6223, e-ISSN:1971-8829

1 Introduction

In any field, whether it is healthcare, business, government, science, security or disaster management, taking the right decision based on the available data at the right time matters much in ensuring its success. According to (Keim, Huamin & Kwan-Liu, 2013), abundant data is available now which is mostly with a short life span, and with proper visualization, essential information contained in this huge data can be clearly presented and complex analyses can be performed on them that aids in fast decision making. Visualization tools are supposed to supplement the domain expertise and deliver a big image so that users are able to frame critical questions and later propose heuristic and insightful answers to these questions. (Kung, 2015).

Performing visual analytics on Big Data involving millions of records is hard to achieve using traditional tools like excel as the complexity of big data demands new visualization tools or enhancements in existing tools to handle the same. This paper implements visualization of a large crime dataset of the Chicago city where the facts and trends regarding crimes over the years are exposed by making use the two leading visualization tools, the Tableau and D3.js and a comparative analysis of the foresaid tools is carried out.

1.1 Role of Visualization in Education and Learning

Computer based visualizations supports teaching and learning as they promote understanding and generate interest and engagement among students while the same can be used for computational purposes (Vavra, 2011). Researchers have agreed that visualization supports the description, development and learning of concepts in science and allied fields (Philips *et al.*, 2010).

According to (Pylyshyn, 2003), visualization objects can portray larger concepts cracked down into understandable smaller ones and can track relationship among variables aiding students to seek alternative solutions. As visualization provides an image of the data, it aids in learning process by offering a comprehensive solution to refer, review and recall the concepts.

In computer science education, algorithm visualization has greater impact on learning as students are engaged more in further activities structured around visualization (Grissom *et al.*, 2003). Web based visualization tool based on Python for teaching introductory courses in computer is the topic of discussion of (Guo, 2013).

Visualization aids learning process in every field, but there is no single tool that fits for all functions. Visualization will be effective only when the appropriate tool is selected and used in the right way.

2 Related Works

Visual depiction of data has its history stretching to ancient times wherein (Friendly, 2008) gives a comprehensive history and advancements in the field of data visualization from primitive to modern age. As stated by (Kosara & Jock, 2013), now that researchers are having clear idea of visual design space, the next logical step in visualization research is to effectively utilize the elements of storytelling in visualization.

Hundreds of visualization tools are available today that meant for diverse usages, as the term visualization itself is used in varied perspectives. It may involve direct visualization of images where images themselves serves as the data, like the electron microscopy data, which requires specialized tools for knowledge discovery. (Beyer, 2013) explains the system for interactive analysis and concurrent visualization of electron microscopy data employed on mouse cortex volume of one teravoxels.

Hidalgo (2014), describes five visualization engines developed at MIT including *Dataviva*, meant for data from Brazil economy and *Pantheon* meant for investigating global cultural production. Extension of visualization to the personal domain is discussed in (Huang *et al.*, 2015) which shows how personal visualization helps individuals to gain insights about themselves, their relations and the community they are involved in. Visualization of streaming data is the topic of work by (Li & Baciu, 2014). Big Complex data sets are posing visualization challenges and (Paul, Randall & Bohdan, 2015) suggests a design-first approach to have meaningful abstractions from Big Data for effective visualization.

Several research works based on general purpose visualization tools have been carried out in recent years. A comparative case study of JavaScript libraries D3.js and HighCharts on visualization is performed in (ElTayeby, 2013). A decision making system based on evidence is proposed in (Riley, 2015) which uses Tableau visualizations on a criminal justice system database. A visualization oriented data separation and aggregation to transform large dataset into small dataset with D3 based rendering implementation is discussed in (Li, 2015). Visualization and analysis of graphs over time is presented in (Schneider, 2015), while visualization of movement in theme parks is demonstrated in (Cai, 2015), where Tableau is utilized for creating visualizations in both cases.

3 Visualization Tools

Among the myriad visualization tools available today, we are considering only the general purpose visualization tools meant for computer based graphical

representation of numerical or textual data that could be stored in a spreadsheet or database, in this paper. A review of commercial systems for visual analytics has been given in (Zhang 2012). Table 1 gives the list of popular visualization tools available now.

3.1 Classification and Selection of Visualization Tools

It is hard to adapt a single criterion for classifying general purpose visualization tools. One way is to group them as open source tools and proprietary tools. Proprietary tools like Tableau, Qlicksense, Spotfire or FusionCharts offer customer support, but tend to be expensive compared to open source tools like D3.js, Datawrapper and DyGraphs, though there are exceptions. Google charts, a proprietary tool is free, while the open source HighCharts is not free for commercial use. Proprietary tools are often available as a bundled package which make large footprint in user’s system. Tailor made visualizations are possible using open source tools which can be extremely creative.

Another way to categorize tools is on the basis of whether they are drag and drop type or whether coding is required for creating visualizations. Most of the proprietary visualization tools are drag and drop type, which requires no coding. Hence it is easy to learn and visualizations can be created in minutes, but the drawback is that out-of-box visualizations are not possible. Many open source tools require heavy or light coding where Java Script being the primary coding language to create visualization, which requires expertise and more development time.

Visualization tools can also be grouped in terms of possible output types like maps, timelines, bubble charts etc. or in terms of output formats like SVG, PNG, Flash etc.

Table 1
POPULAR VISUALIZATION TOOLS AVAILABLE NOW

Tool	Type	Coding Requirement	Visualization techniques and Remarks
Tableau Desktop	Proprietary, expensive	Drag and drop	Variety of built-in charts and maps; BI tool, scalable
Qlick sense	Proprietary	Drag and drop	Adequate variety of charts; BI tool; memory requirement based on data volumes
D3.js	Free & Open source,	Heavy coding required	All types of charts possible, but no built-in charts; Java Script library
Google-Charts	Free	Light coding required	Limited number of charts; Java Script based
Visual.ly	Proprietary	Easy	Infographic generation tool

Tool	Type	Coding Requirement	Visualization techniques and Remarks
HighCharts	Open source, not free for commercial use	Coding required	Adequate variety of charts and maps JavaScript based, output SVG and VML to support old browsers
Dygraphs	Open source	Heavy Coding	JavaScript based, capable of handling huge datasets
Leaflet	Open source	Coding Required	For geomaps only; Light, suitable for mobile applications
Timeline	Open source	Easy	Only for timelines
Fusion-Charts	Proprietary, expensive	Light coding	Wide range of charts and maps in 2D & 3D; Javascript based
Data-wrapper	Open Source	Drag and drop	Limited range of maps and charts; Built on top of D3.js
Spotfire	Proprietary	Drag and drop	Adequate variety of charts and maps;

Among these varied visualization tools, Tableau, the proprietary drag and drop tool, and D3.js, the free open source tool which makes any visualization possible through coding, stand out among their respective groups. Many of the proprietary visualization tools are basically business intelligence (BI) tools and according to Gartner annual report 2016 on BI and analytics, Tableau is ranked as the Leader for the fourth consecutive year with only Qlick and Microsoft are closer to it, which is one of the reason to choose Tableau to perform visual analytics from among other similar tools like Qlicksense or Spotfire. On a similar line, D3.js outperforms all other Java Script based tools as it offers versatile functionalities like data manipulation and transformation and makes effective use of the power of new browser and web technologies. In an assessment of design characteristics of visualization tools done by Lensing T. J (2014), D3.js and Tableau were referred as the best visualization tools available.

Though visualization tools are plentiful, this paper aims at the comparison of two tools, Tableau and D3.js, while performing visual analytics. Both tools are described in detail in the following sections

3.2 Tableau

Tableau is a proprietary business intelligence solution, highly preferred as a visualization tool capable of delivering interactive visualizations in no time with its drag and drop nature. It offers a wide variety of options including pie, bar and bubble charts, maps, heat maps, scatter plots making use of which informative dashboards can be created instantly from diverse datasets. It performs aggregations, highlighting or drilling down in charts with much ease that even novice users can create visualizations to illuminate facts in a huge

dataset. Tableau can effortlessly connect to the data stored in excel, csv and text files and recognize fields and formats. Still, to access data stored in many databases, Tableau users require appropriate database connectors and expertise.

Tableau can be used to define and calculate new variables and perform simple data manipulations with usage of mathematical formulae like excel, but it is much powerful and can handle millions of records where excel fails. The limited analytical ability of Tableau can be overridden by its capability to integrate with several platforms including the Big Data platform Hadoop, the statistical analytic tool R and its support for Google Big Query API. Nevertheless, Tableau provides native backing for proprietary Big Data players like Amazon Red Shift and TeraData, it is still lacking the direct support of many prevalent open source NoSQL databases like the MongoDB.

To cater varying needs of diverse users, Tableau Software offers selections comprising Tableau Desktop, Tableau Server, Tableau Mobile or Tableau Public to choose from. Tableau Desktop is appealing to individuals and small organizations, while Tableau Server is apt for organizations with many users. Tablet and Mobile phone consumers are supported by the Tableau Mobile which added great value to the software due to the proliferation of smartphone usage among individuals. Though Tableau public is free to use and can be used for small scale visualizations, it has restricted features compared to Tableau Desktop, making it unsuitable for big data visualizations.

With Tableau server, one can share visualizations among different users, at the same time allow restricted access to different views and underlying data by applying its user filter or row-level security feature. Tableau provides considerable amount of interactivity with changes in display on user clicks and mouse over certain parts of the visualization

3.3 D3.js

D3.js (Data Driven Documents) (Bostock, Ogievetsky & Heer, 2011), the java script library is now been extensively utilized to create powerful and highly interactive visualization. With D3, arbitrary data can be bind to a Document Object Model (DOM) and hence data manipulations as well as data-driven transformations can be performed on the document. D3 creates visualization using Scalable Vector Graphics (SVG), HTML and Cascaded Style Sheets (CSS) and handles different data formats including csv, JSON, and GeoJSON. More details are available at website. It is a free and open source tool, but requires basic knowledge in html, java script, CSS, SVG to create graphics. Coding of low level functions from scratch is required even to create simple visualizations which is relatively hard and extends the development time considerably as D3 is not having a prebuilt library of charts or graphs to choose

from. However, built-in reusable functions and samples of commonly used graphics are readily available.

D3 can create any imaginable visualization within coding limits and can offer excellent interactivity. This is mainly due to its transformative nature that make use of web standards to create visualization where an upper hand on SVG controls can yield the best possible visualization compared to other tools. It allows animated transitions and graphical complexities at higher frame rates.

D3 supports all modern browsers including IE 9+, android and IOS, but certain features fail to display in older browsers. Concealing data for restricted access among different users is difficult to achieve in D3.

A comparison of Tableau and D3.js is given in table 2.

Table 2
TABLEAU VS. D3.JS

Tableau	D3.js
Business Intelligence tool; visualization software package	Java Script Library, not a visualization software package.
Proprietary, expensive	Free and Open Source
Drag and Drop; Learning very easy	Heavy coding required; learning very difficult
Development time of Dashboard is in minutes	Development time is from hours to days
Variety of built-in charts and maps to select from, but out-of-box visualizations are not possible	Any imaginable visualization(code-able) is possible, but every chart has to be built from scratch
Visualization format is proprietary, but allows export to JPEG, PNG,BMF and EMF formats	Output format is Scalable Vector Graphics(SVG)
By applying user filter or row-level security feature, restricted data access can be provided to different users	Concealing data to provide restricted access among different users is difficult to achieve
Able to identify dimensions and measures and can easily handle gigabytes of data	Struggle in handling large dataset in gigabytes.

4 Visual Analytics

4.1 Dataset

The dataset used for this work was the CSV file ‘crime 2001 to present’ from Chicago data portal, accessed on December 7th 2015, which holds 5,948,951 records having 23 attributes where each record mirrors crime incidents reported in the city of Chicago other than murders, taken from Chicago police department’s CLEAR (Citizen Law Enforcement Analysis and Reporting) system. Among the attributes, only the major ones were selected for this visualization, which were *Date*, *Primary Type* (of crime), *Location Description* and *Arrest*.

4.2 Visualization using Tableau

Tableau 9.1 desktop version in windows 8.1 environment running on an i5 machine with 4 GB RAM was used to create the dashboard shown in the Fig. 1. For this, Tableau was connected to the data from the 1.36 GB CSV file. It took a couple of minutes to read the data initially. It automatically recognized the data fields and formats, and dimensions and measures were properly identified. Using Tableau, Dashboard creation was able to be completed in less than ten minutes. It involved an area chart plotting number of crime records against each month ranging from 2001 January till 2015 November, a horizontal bar chart showing top 5 primary crime types from among 35 types recorded with theft at first position, a vertical bar chart displaying five most common location descriptions of crime occurrences and a pie chart of arrest statistics. On mouse over areas of charts, exact measures were displayed, providing interactivity

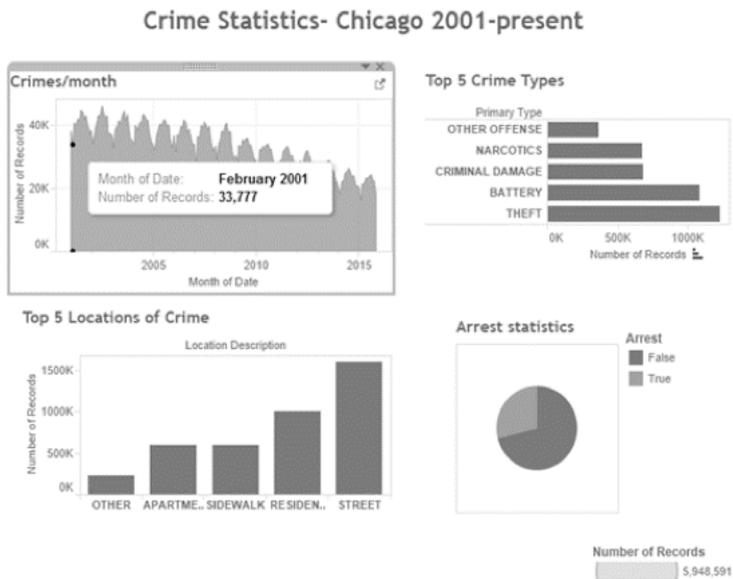


Fig. 1 - Visualization of Chicago crime dataset using Tableau.

4.3 Visualization using D3.js

Since D3.js is a JavaScript library and not software package, it requires the support of other software components to create visualization. In this work, Node.js, an event triggered web server was used to function as the server for sending data to D3 and to host webpages and JavaScript libraries, while

supporting interactivity. The crime dataset file in CSV format was imported to open source NoSQL database MongoDB which acted as the data repository for this work. Also, DC.js, the data wrapper library for D3 and Crossfilter.js, the data manipulation library, were also utilized for this visualization. Aggregation and appropriate grouping of data was handled by DC.js and Crossfilter.js when interactive selective operations on visualization was done.

The development of visualization using D3.js involved several tasks including creating proper directory structure and writing hundreds of lines of code. A dashboard displaying count of crimes per month over a period from 2012 to 2015 and top five crime locations and crime types created in D3 is shown in Fig. 2, involving over a million records.



Fig. 2 - Visualization of Chicago crime dataset using D3.js over the years 2012-2015.

Other than displaying exact measures on mouse over chart areas, this visualization provided excellent interactivity such as selecting an area (time period) in the crimes per month chart, to observe the changes reflected in other charts as well. But the visualization failed to display the charts when the complete dataset of about 6 million records was involved.

5 Results and Discussion

Visual analytics of the crime records of city of Chicago shows that every year, crimes reported per month follows a Gaussian curve like pattern with its peak (most number of crimes) occurring in the month of July. This may be correlated to the climatic conditions in Chicago, where the average temperature

per month follows a similar pattern and July is the hottest month. It was also observable that there is a slight, but steady decrease in reported crimes over the years from 2001 to 2015. It is interesting to observe that though theft is the type of crime reported the most, narcotics crime types are significantly high and as far as location of the crime occurrence is concerned, Residence stands at second position next to Street.

While tableau was able to identify the dimensions and measures from the whole dataset and create visualization in minutes, without writing a single line of code, D3.js failed to create visualization when the full dataset was involved. Though with a big subset of dataset, D3.js was able to display a highly interactive visualization, several issues needs to be addressed for D3.js to be suitable to function as a Big Data visualization tool. Thus for quick and easy visualization involving commonly used maps or charts, Tableau is well suited and D3.js can be used when there is extraordinary charting requirements or high interactivity requisites.

Conclusion

In this work, interactive visualization of a big dataset, the Chicago crime dataset, was performed using the most popular and diverse visualization tools, the Tableau and D3.js and a comparative analysis of both tools had been carried out. Visual analytics of this dataset revealed interesting patterns which would have unnoticed otherwise. The comparative analysis of two tools revealed that open source visualization tools requires a lot of enhancements to cater the needs of emerging big data visualization issues. No single tool could be a perfect solution for all Big Data visualization problems, where it had to be chosen based on many factors including cost and time for development, expertise available, dataset volume and features, visualization requirements including its visual appeal, knowledge content, level of interactivity and sharing options.

REFERENCES

- Beyer, Justus, et al. (2013), *Exploring the connectome: Petascale volume visualization of microscopy data streams*. IEEE Computer Graphics and Applications 33 (4): 50-61.
- Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer (2011), *D³ data-driven documents*. IEEE Transactions on Visualization and Computer Graphics 17 (12): 2301-2309.
- Cai, Y., Tay, K., Winarta, B. and Kam, T.S., (2015), October, *Visualizing Movement in Theme Park*. In proceedings of IEEE Conference on Visual Analytics Science and Technology (VAST), (pp. 147-148). IEEE.

- Data-Driven Documents. Accessed 2016. <https://d3js.org/>.
- ElTayeby, O., John, D., Patel, P. and Simmerman, S., (2013), October, *Comparative case study between D3 & Highcharts on Lustre metadata visualization*. In 2013 IEEE Symposium on Large-Scale Data Analysis and Visualization (pp. 127-128). IEEE.
- Friendly, Michael. (2008), *A brief history of data visualization*. In Handbook of data visualization, 15-56. Springer Berlin Heidelberg.
- Grissom, S., McNally, M.F. and Naps, T., (2003), June, *Algorithm visualization in CS education: comparing levels of student engagement*. In Proceedings of the 2003 ACM symposium on Software visualization (pp. 87-94)
- Guo, P.J., (2013), March, *Online python tutor: embeddable web-based program visualization for cs education*. In Proceeding of the 44th ACM technical symposium on Computer science education (pp. 579-584). ACM
- Hidalgo, Cesar. (2014), *Big data visualization engines for understanding the development of countries, social networks, culture and cities*. Proceedings of the 25th ACM conference on Hypertext and social media.
- Huang, Dandan et al. (2015), *Personal visualization and personal visual analytics*. IEEE Transactions on Visualization and Computer Graphics 21 (3): 420-433.
- Keim, Daniel, Qu Huamin, and Ma Kwan-Liu (2013), *Big-Data Visualization*. IEEE Computer Graphics and Applications 33 (4): 20-21.
- Kosara, Robert, and Mackinlay Jock (2013), *Storytelling: The next step for visualization*. Computer 46 (5): 44-50.
- Kung, Sun-Yuan (2015), *Visualization of big data*. IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC). pp. 447-448.
- Lensing, T.J., (2014), *Assessing The Design Characteristics of Effective Data Visualizations* (Doctoral dissertation, Arizona State University).
- Li, Chenhui, and G. Baciuc (2014), *VALID: A Web Framework for Visual Analytics of Large Streaming Data*. IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Beijing.
- Li, X., Kuroda, A., Matsuzaki, H. and Nakajima, N. (2015), October. *Advanced aggregate computation for large data visualization*. In proceedings of 2015 IEEE 5th Symposium on Large Data Analysis and Visualization (LDAV), (pp. 137-138). IEEE
- Paul, Celeste Lyn, Rohrer Randall, and Nebesh Bohdan (2015), *A Design First Approach to Visualization Innovation*. IEEE computer graphics and applications 35 (1): 12-18.
- Phillips, L. M., S. P. Norris and J. S. Macnab (2010), *Visualization in Mathematics, Reading and Science Education*. Dordrecht, Netherlands: Springer
- Pylyshyn, Z.W. (2003), *Seeing and visualizing: It's not what you think*. MIT Press.
- Riley, M., Harelson, D., Monte, M., Chong, M., O'Hern, M., Smith, M. and White, K.P. (2015), April. *Criminal justice system data analysis and visualization*. In Systems and Information Engineering Design Symposium (SIEDS), 2015 (pp. 295-299). IEEE.

- Schneider, B., Acevedo, C., Buchmuller, J., Fischer, F. and Keim, D.A. (2015), October. *Visual analytics for inspecting the evolution of a graph over time: Pattern discovery in a communication network*. In proceedings of IEEE Conference on Visual Analytics Science and Technology (VAST), 2015 (pp. 169-170). IEEE.
- Tableau Desktop. Accessed February 2016. <http://www.tableau.com/products/desktop>.
- Vavra, K. L. et al. (2011), *Visualization in science education*. Alberta Science Education Journal, 41(1): 22-30.
- Zhang, Leishi et al. (2012) *Visual analytics for the big data era—A comparative review of state-of-the-art commercial systems*. IEEE Conference on Visual Analytics Science and Technology (VAST). 173-182.