Requirements of Data Visualisation Tools to Analyse Big Data: A Structured Literature Review

Joy Lowe, Machdel Matthee

▶ **To cite this version:**

Joy Lowe, Machdel Matthee. Requirements of Data Visualisation Tools to Analyse Big Data: A Structured Literature Review. 19th Conference on e-Business, e-Services and e-Society (I3E), Apr 2020, Skukuza, South Africa. pp.469-480, 10.1007/978-3-030-44999-5_39 . hal-03222857

HAL Id: hal-03222857

https://inria.hal.science/hal-03222857v1

Submitted on 10 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Requirements of Data Visualisation Tools to Analyse Big Data: A Structured Literature Review

Joy Lowe[0000-0003-4937-4313] and Machdel Matthee[0000-0002-6973-1798]

Department of Informatics, University of Pretoria
u16331878@tuks.co.za, machdel.matthee@up.ac.za

Abstract. The continual growth of big data necessitates efficient ways of analysing these large datasets. Data visualisation and visual analytics has been identified as a key tool in big data analysis because they draw on the human visual and cognitive capabilities to analyse data quickly, intuitively and interactively. However, current visualisation tools and visual analytical systems fall short of providing a seamless user experience and several improvements could be made to current commercially available visualisation tools. By conducting a systematic literature review, requirements of visualisation tools were identified and categorised into six groups: dimensionality reduction, data reduction, scalability and readability, interactivity, fast retrieval of results, and user assistance. The most common themes found in the literature were dimensionality reduction and interactive data exploration.

Keywords: Big Data Visualisation, Visualisation Tools, Visual Analytics.

1 Introduction

Big data is growing at an exponential rate due to unprecedented data capture from smart devices, the internet of things, measurement and sensor technology, transaction data, metadata, and social media (Ali, Qadir, Rasool, Sathiaseelan, & Zwitter, 2016; Gisbrecht, 2013; Resnyansky, 2019). With an estimated 3.7 billion internet users globally (Marr, 2018) the internet is a significant factor in the generation of data. Google, the most popular search engine, receives at least 5.5 billion search queries per day (Sullivan, 2016) and processes 24 petabytes of data per day (Ali et al., 2016). User generated social media data is another significant contributor to big data, with Facebook reporting 2.32 billion active monthly users (Noyes, 2019a) and Twitter reporting 321 million active monthly users with 500 million tweets posted daily (Noyes, 2019b)

Big data is defined as large, complex, high velocity data sets (Seokyeon et al., 2015). Big data is characterised by volume, velocity, value, variety, variability, and complexity (Katal, Wazid, & Goudar, 2013). Due to the nature of these large, heterogeneous data sets, interpretation and comprehension of the data is often difficult, and current data storage and analysis technologies are no longer able to effectively store and analyse such large datasets (Genender-Feltheimer, 2018). However, using big

data brings new opportunities for research and new insights across a wide variety of fields (Resnyansky, 2019) and by visually exploring and analysing the data, unexpected discoveries can be found.

Data visualisation is a solution to aid meaningful analysis, accessibility and interpretation of big data as it relies on human cognitive capabilities to process visual information (Gisbrecht, 2013). Visualising data creates an image that the user is able to parse in three subtasks: perceptual grouping, image segmentation and object recognition (Zhu, 2003). Image parsing is an efficient means of comprehending large datasets because grouping or clustering of data points, as well as outliers, become apparent. Further, the inclusion of interactive technology and visual analytics allows users to gain more information about specific data points and areas of interest through visual exploration, which facilitates the formation, testing and validation of hypotheses (Elmqvist & Fekete, 2010).

Increasingly complex and sophisticated data analytics and visualisation algorithms are required in response to the increasing significance of big data (Gisbrecht, 2013). The velocity at which the data is generated produces the need for real time analysis. This paper will explore and compare various techniques which have been developed in response to these requirements to effectively visualise and comprehend big data.

2 Literature Background

2.1 The Growth of Big Data

Big data has been hailed as one of the greatest challenges of the 21st century due to the volume of data, the speed at which the data is being generated, as well as the inconsistency of the data (Katal et al., 2013). A number of factors contribute to the growth and complexity of big data including the internet of things (IOT), click stream data, and social media data. Big data has become an integral part of business, social media, scientific research and several other fields (Olshannikova, Ometov, Koucheryavy, & Olsson, 2015) and the data is predicted to grow even further, at faster rates.

2.2 The Need for Visualisations to Comprehend Big Data

Big data is now beyond human comprehension through simple exploration and investigation of the data (Long & Linsen, 2011). People are able to comprehend the world more quickly and easily through visual cues, and this extends to visualising data. Vision is the most valuable human sense and most people prefer visual representations to other sensory information.

In order to capitalise on the big data revolution and gain as much insight as possible, it is imperative that the data is analysed efficiently and accurately, and thereafter visualised to allow for the fast interpretation of the data, much of which is generated in real time.

Furthermore, visualisations of data make trends and patterns in the data much more apparent such as clustering, the distribution of the data, and correlation within the data

(Long & Linsen, 2011). The goal of many businesses regarding data analysis is to recognise such patterns, emphasizing the need for data visualisation to achieve strategic objectives.

Visualising data is a means of "uniting the abstract world of data with the physical world through visual representation" (Olshannikova et al., 2015). Visualisations can be likened to the front end of big data (L. Wang, Wang, & Alexander, 2015) and can be used to access and interpret the data, making the insights and trends more apparent.

2.3 The Need for Awareness of Data Visualisation Techniques Among Professionals

Many business professionals now need to be educated in data related information and become "data literate" in order to generate business reports as well as interpret reports and business intelligence executive dashboards. These business professionals must be made aware of the features and limitations of each visualisation technique and visualization software or tool, particularly with interactive dashboards becoming more popular. In this paper some visualization tools and techniques will be discussed.

2.4 Commercially Available Business Intelligence and Visualisation Tools

There are a number of currently available commercial Business Intelligence tools that include the ability to create visualisations, which can be utilized by professionals in a number of different fields and industries. Figure 1 below shows the Magic Quadrant for Analytics and Business Intelligence Platforms published by Gartner in February 2019. This quadrant displays the 'leaders', 'visionaries', 'challengers', and 'niche players' of business intelligence platforms and provides a good overview of the landscape of the industry at present.

Fig. 1. Gartner Magic Quadrant for Analytics and Business Intelligence Platforms (Orad, 2019)

3 Research Method

A systematic literature review was carried out during the course of this research. The following eight steps are suggested when conducting a systematic review of Information Systems research (Okoli & Schabram, n.d.):

1. Identifying the purpose and intended goals of the literature review
2. Outlining the protocol and detailed procedure to be followed
3. Searching for the literature
4. Practical screening/screening for inclusion
5. Quality appraisal/screening for exclusion
6. Data extraction
7. Synthesis/analysis of studies
8. Writing the review

3.1 Research Question

What are the requirements of data visualisation tools used to analyse big data?

3.2 Search Terms

Title(data) AND title(visual*) AND abstract(big data).

The wild card search term "visual*" was used to cater for all variations of the words that derive from the common stem 'visual' including visualise, visualising, visualisation, visualisations, as well as being inclusive of both variations of spelling for each of these words: the American spelling which includes a 'z' and the English spelling which contains an 's'. In the case where a search engine did not support wild cards, all possible variations of the word were specified using OR Boolean clauses so that any of the spellings would be included in the search results.

3.3 Selection Criteria

Inclusion criteria

- Articles published between January 2014 and October 2019 were included to ensure the most up to date research was evaluated since the field of big data visualisation is changing rapidly.
- Peer reviewed journal articles and conference papers were included.
- Only articles written in English were included.

Exclusion criteria

- Articles published prior to January 2014 and after October 2019 were excluded.
- The following document types were excluded: periodicals, editorials, books, and book chapters.

3.4 Source Selection

The following sources and databases were consulted during the course of this research:

- ABI/Inform
- ACM Digital Library
- EBSCOhost
- ScienceDirect

The Prisma flow chart below shows the number of articles identified through each database and the process of exclusion that was followed to reach the final 31 articles that were included in the literature review.

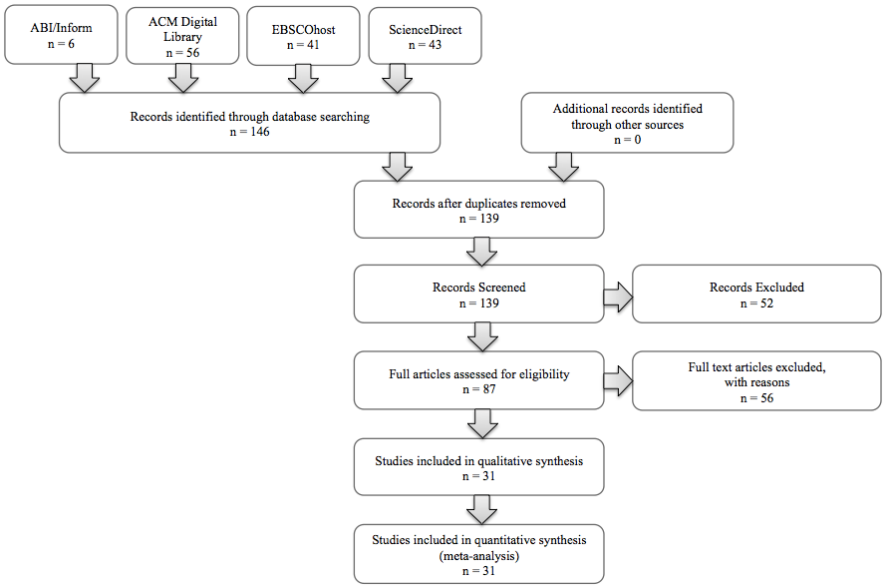

Fig. 2. Prisma Flow Chart

4 Analysis of Findings

After reviewing the literature, many authors are in agreement that data visualisation will become increasingly important as big data becomes more prevalent in more organisations and in order to interpret the data, effective visualisation techniques will be required (S. Wang & Li, 2019). There has also been a call to improve current visualisation techniques (Li, Kuroda, & Matsuzaki, 2016; Olshannikova et al., 2015) in order to handle the challenges of visualising big data.

A combined list of requirements of data visualisation tools to interpret big data found in the literature are shown in Figure 3 below. The bar graph shows the number of

papers included in the literature which fall into each category. Note that some articles address more than one category, thus the total of the values of the bars in figure 3 is greater than the total number of articles included in the literature review.

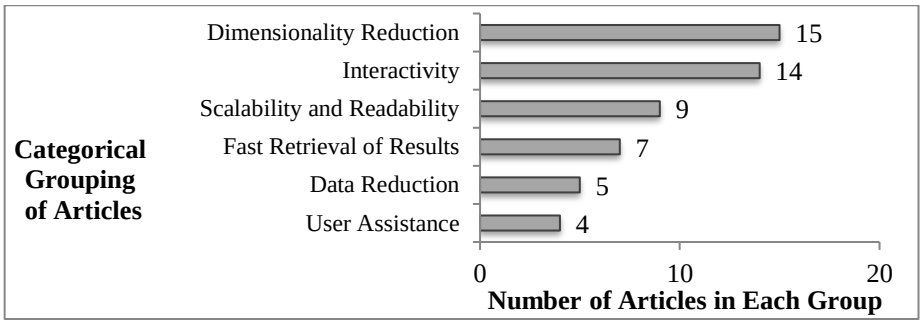

Fig. 3. Number of Articles in Each Category

4.1 Dimensionality Reduction

The dimensionality reduction category contributed the highest number of papers (15 of 31). Dimensionality reduction is a key requirement of big data visualisation (Zou, Zeng, Cao, & Ji, 2016) because many of the large datasets that exist today contain multiple dimensions, however, humans can only perceive three dimensions, giving rise to the need for algorithms that can reduce datasets to two or three dimensions which can be visualised. There are a number of statistical algorithms which can achieve this including Principal Component Analysis (PCA), t-distributed stochastic neighbour embedding (t-SNE) and diffusion maps (Agrawal, Dai, & Andres, 2015; Fernandez, Gonzalez, Diaz, & Dorronsoro, 2015; Genender-Feltheimer, 2018; Gisbrecht & Hammer, 2015; Shirota, Hashimoto, & Basabi, 2017). Mapping multidimensional datasets into clusters that are represented in two or three dimensions is also common and allows the partitioning of data into similar groups (Keck et al., 2017). It is useful to maintain the original structures of the data (Fernandez et al., 2015) so that further analysis can be carried out after identifying certain clusters or patterns of interest during the two- or three-dimensional exploration of the data (Keck et al., 2017; Xie, Chenna, Le, & Planteen, 2016). Advantages of dimensional reduction include easily understood visualisations, reduced data quality challenges, and improved computational efficiency (Genender-Feltheimer, 2018).

Dimensionality reducing algorithms are either linear, or non-linear. Linear algorithms are only able to learn linear underlying topological spaces, while non-linear algorithms are able to learn complex underlying topological spaces and focus on preserving neighbourhood geometry (Genender-Feltheimer, 2018). Dimensionality reduction algorithms can also be grouped into supervised and unsupervised algorithms. Supervised algorithms require a ‘training’ data set to ‘learn’ from, and subsequently the model is applied to a second ‘scoring’ data set. Unsupervised algorithms do not require a training dataset, but rather, are able to detect patterns in the data (Mwangi,

Soares, & Hasan, 2014). Examples of supervised algorithms include Linear Discriminant Analysis and Linear Regression; and examples of unsupervised algorithms include neural networks and other deep learning methods. Supervised algorithms have the potential to be tainted with human bias, as the training dataset contains user defined target labels (Mwangi et al., 2014).

4.2 Interactivity

By providing a visual interface which the user can interact with, the user is able to intuitively comprehend the data, perceive the underlying patterns (Lugmayr, Greenfeld, & Zhang, 2017) and query the data visually, without the need for programming knowledge. The ability for a user to interact with a visualisation makes visual exploration of the dataset possible and this has the major advantage of combining both human and machine intelligence (Chen, Guan, Zhang, Du, & Wang, 2019) to uncover unexpected and interesting phenomena within the dataset (Cho, Lee, Varma, & Lee, 2014). The benefits of visual analytics are "visual perception, interactive exploration, improved understanding, informed steering and intuitive interpretation" (Liu, 2019). This has the potential to uncover patterns in the data using a bottom-up approach (Ruan & Zhang, 2017) or to test theories and search for evidence within the data using a top-down approach (Genender-Feltheimer, 2018; Mwangi et al., 2014). Visualisation tools should enable the user to drill down into areas of interest within the data (Zhao, Zhang, Liu, Zhang, & Luke, 2017) to aid in the visual analysis.

Several authors write about the need to keep the user informed at each step of the visual analytics process. This gives the user situational awareness and keeps the user informed about the underlying processes being run in order to render the data or return results (Zhao et al., 2017). Currently, many visualisation tools are a "Black Box" and do not give enough feedback to the user. Results are produced without a description of how the results were calculated, and in some visualisation systems, the user is not given the option to select a statistical method of their choice (Seokyeon et al., 2015).

4.3 Scalability and Readability

The displays on which the visualisations are rendered have a limited number of pixels, thus the lowest granularity possible is to plot one data point onto one pixel (Molina-Solana, Birch, & Guo, 2017; Olshannikova et al., 2015; Yang, Zhang, Wang, & Nguyen, 2015). Visualisation tools are required to "squeeze a billion records into a million pixels" (Bikakis, 2018) emphasizing the need for summarisation of data, particularly when creating visualisations. If the dataset is not sufficiently summarised, or if every data point is plotted, this may result in 'over plotting' which is inefficient and it can be difficult to discern the patterns and trends within the data (Li et al., 2016). The number of pixels is even lower on mobile device screens, thus the scalability of visualisations for mobile devices are even more important (Li et al., 2016; Molina-Solana et al., 2017). Furthermore, attempting to plot all data points overloads the users' cognitive capabilities, thus it is important to ensure readability and scalability

across all screen sizes and in all contexts (Eldawy, Mokbel, & Jonathan, 2015). In order to achieve this, visualisation tools should include zooming; overview and detail; and brushing of the visualisation (Li et al., 2016). The visualisation tools should also use the approach of providing a high level overview of the data initially, and subsequently load more detailed data as the user queries and drills down into areas of interest (Agrawal et al., 2015).

4.4 Fast Retrieval of Results

It is imperative that data visualisation and visual analytics systems retrieve data fast enough so that the user does not lose their focus and momentum when visually analysing a dataset. The intention is to hold the users concentration, enabling them to problem solve at a high level and creatively formulate queries to investigate further (Liu, 2019). Several authors agree that the result should be returned within one second, while some authors specify not more than ten seconds (Li et al., 2016).

One approach to address this is called approximate query processing, whereby an approximation of the data is returned quickly and the exact results are continuously processed in the background while the user moves onto other queries (Zhao et al., 2017). When the query has finished processing, which may take up to ten minutes, the user is informed and the discrepancies between the original estimate and the final results are displayed to the user (Moritz, Fisher, Ding, & Wang, 2017). Generally, it is more important to get a broad overview of the data, rather than exact figures, and in this regard, the approximate processing approach works well. Displaying the confidence interval of the estimate during the initial load of the data gives further benefit to this method, as the user is given an estimate and an indication of how accurate the estimate is. This is particularly useful when the results of simple queries are needed urgently (Masiane, 2019; Zhao et al., 2017).

Other approaches which may speed up data retrieval include catching or prefetching the results of frequent queries. It may also be advantageous to divide the data processing across several computers. Online Analytical Processing (OLAP) is another method which speeds up query processing (Moritz et al., 2017). Where possible, it can be useful to parallelise data processing and rendering so that they are performed simultaneously rather than sequentially (Hassan & Pernul, 2014).

4.5 Data Reduction

As mentioned in section 4.3, computer screens have a limited number of pixels onto which data points can be rendered. A possible solution is to make use of data reduction techniques to utilise the “screen real estate” efficiently (Yang, Zhang, Wang, & Nguyen, 2015). There are a number of ways to reduce the data, including filtering, clustering, and sampling techniques (Masiane, 2019). Examples of sampling techniques that could be used to reduce the dataset include stratified random sampling, systematic random sampling, and quota sampling. Probability sampling techniques select the data in such a way that each data point has an equal probability of being selected. In contrast, non-probability sampling techniques are useful in some cases for

the purposes of including at least one data point within each group or category that may be present in the data, meaning that data points which fall into a smaller group or category have a higher chance of being selected than data points which fall into larger categories or groups. This ensures each group is represented in the sample by the inclusion of at least one data point from each group or category (Genender-Feltheimer, 2018).

An inevitable disadvantage of data reduction is information loss (Abidi, Polys, Rajamohan, & Arsenault, 2018) resulting in the inability to identify nuances such as outliers as well as the true patterns and trends in the data (Keck et al., 2017).

4.6 User Assistance

Currently, the majority of visual analytics is carried out by domain experts, or mathematicians and statisticians with the appropriate analytical software and resources (Ko & Chang, 2018). However, with increasingly user friendly web-based visualisation tools, many more novice users or laymen could conduct data analysis in future (Seokyeon et al., 2015). According to Behrisch et al, however, the visualisation tools that are commercially available currently still require much improvement in order to provide users with appropriate assistance (Behrisch et al., 2019). Examples of methods to provide the user with assistance include recommending which visualisation techniques are most suitable, given a particular dataset as input (Seokyeon et al., 2015). Behrisch et al. (Behrisch et al., 2019) recommend calculating statistics as soon as the input data is loaded and displaying the results to the user immediately. The visualisation tool could further assist the user by predicting user queries based on use case, user profile, and the relevant information could be displayed based on these parameters.

Human Computer Interaction (HCI) principles are highly relevant in this domain and the user experience could be enhanced by incorporating principles such as gestault laws of grouping, and providing frequent and appropriate feedback (Seokyeon et al., 2015). Similarly, the use of colours is a simple way to highlight similarities or differences in the data (Elaiza, Khalid, Yusoff, Kamaru-zaman, & Izzati, 2014). As mentioned in section 4.2, visualisation tools must keep the user informed at each stage of the analysis to give the user situational awareness, inform the user of the progress of the query, and to give the user more control over the analysis (Behrisch et al., 2019; Zhao et al., 2017). Zhao et al (Zhao et al., 2017) argue that most visual analytic systems are designed too specifically, thus they are only suitable for a limited range of situations as they restrict data types and data structures. A potential solution to make visualisation systems more dynamic and flexible is through customisation and the use of plug-ins and dynamically linked libraries (Liu, 2019).

5 Conclusion

Big data and the visualisation of large datasets brings many promising opportunities for discoveries of patterns within the data. However, to fully capitalise on these op-

portunities, visualisation tools must be able to: perform dimensional reduction of multidimensional datasets; perform data reduction as it is not feasible to render each data point; ensure scalability and readability of visualisations; allow the user to interact with the data and visually analyse the data; retrieve results quickly so that the user does not lose their focus; and provide user assistance. Visualisation tools are currently too narrowly focused, and must be more flexible in terms of the data types they are able to process and the functions they are able to perform. As big data continually grows, data visualisation tools must improve rapidly to enable the exploration and analysis of the masses of data that are produced. Data visualization tools must improve at an equally rapid pace.

References

1. Abidi, F., Polys, N., Rajamohan, S., & Arsenault, L. (2018). *Remote high performance visualization of big data for immersive science.pdf*.
2. Agrawal, R., Dai, X., & Andres, F. (2015). *Challenges and Opportunities with Big Data Visualization*.
3. Ali, A., Qadir, J., Rasool, R. ur, Sathiaseelan, A., & Zwitter, A. (2016). Big Data For Development: Applications and Techniques. *Big Data Analytics*. https://doi.org/10.1186/s41044-016-0002-4
4. Behrisch, M., Streeb, D., Stoffel, F., Seebacher, D., Matejek, B., Pfister, H., … Mittelst, S. (2019). *Commercial Visual Analytics Systems – Advances in the Big Data Analytics Field*. *25*(10), 3011–3031. https://doi.org/10.1109/TVCG.2018.2859973
5. Bikakis, N. (2018). *Big Data Visualization Tools*. 1–11.
6. Chen, Y., Guan, Z., Zhang, R., Du, X., & Wang, Y. (2019). A survey on visualization approaches for exploring association relationships in graph data. *Journal of Visualization*, *22*(3), 625–639. https://doi.org/10.1007/s12650-019-00551-y
7. Cho, W., Lee, H., Varma, M. K., & Lee, M. (2014). *Big Data Analysis with Interactive Visualization using R packages*.
8. Elaiza, N., Khalid, A., Yusoff, M., Kamaru-zaman, E. A., & Izzati, I. (2014). Multidimensional Data Medical Dataset Using Interactive Visualization Star Coordinate Technique. *Procedia - Procedia Computer Science*, *42*, 247–254. https://doi.org/10.1016/j.procs.2014.11.059
9. Eldawy, A., Mokbel, M. F., & Jonathan, C. (2015). *A demonstration of HadoopViz- an extensible MapReduce system for visualizing big spatial data p1896-eldawy.pdf*.
10. Elmqvist, N., & Fekete, J. D. (2010). Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, *16*(3), 439–454. https://doi.org/10.1109/TVCG.2009.84
11. Fernandez, A., Gonzalez, A., Diaz, J., & Dorronsoro, J. (2015). *Diffusion Maps for Dimensionality Reduction and Visualization of Meteorological Data.pdf*.
12. Genender-Feltheimer, A. (2018). Visualizing High Dimensional and Big Data. *Complex Adaptive Systems Conference with Theme: Cyber Physical Systems and Deep Learning*. Chicago, Illinois, USA.
13. Gisbrecht, A. (2013). *Advances in dissimilarity-based data visualisation*. Bielefeld University.

14. Gisbrecht, A., & Hammer, B. (2015). *Data visualization by nonlinear dimensionality reduction*. *5*(April), 51–74. https://doi.org/10.1002/widm.1147
15. Hassan, S., & Pernul, G. (2014). *Efficiently Managing the Security and Costs of Big Data Storage using Visual Analytics Categories and Subject Descriptors*.
16. Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and Good practices. *2013 6th International Conference on Contemporary Computing, IC3 2013*, 404–409. https://doi.org/10.1109/IC3.2013.6612229
17. Keck, M., Kammer, D., Grunder, T., Thom, T., Kleinsteuber, M., Maasch, A., & Groh, R. (2017). Towards Glyph-based Visualizations for Big Data Clustering. *Proceedings of VINCI '17, Bangkok, Thailand, August 14-16*, 129.
18. Ko, I., & Chang, H. (2018). Interactive data visualization based on conventional statistical fi ndings for antihypertensive prescriptions using National Health Insurance claims data. *International Journal of Medical Informatics*, *116*(May), 1–8. https://doi.org/10.1016/j.ijmedinf.2018.05.003
19. Li, X., Kuroda, A., & Matsuzaki, H. (2016). *Polyspector ™ : An Interactive Visualization Platform Optimized for Visual Analysis of Big Data*. 109–111.
20. Liu, Z. (2019). Advances in Engineering Software A prototype framework for parallel visualization of large flow data. *Advances in Engineering Software*, *130*(December 2018), 14–23. https://doi.org/10.1016/j.advengsoft.2019.02.004
21. Long, T. Van, & Linsen, L. (2011). *Visualizing high density clusters in multidimensional data using optimized star coordinates*. 655–678. https://doi.org/10.1007/s00180-011-0271-3
22. Lugmayr, A., Greenfeld, A., & Zhang, D. J. (2017). *Selected Advanced Data Visualizations : " The UX-Machine ", Cultural Visualisation , Cognitive Big Data , and Communication of Health and Wellness Data*. 247–251.
23. Marr, B. (2018). How much data do we create every day? The mind-blowing stats everyone should read. Retrieved March 31, 2019, from Forbes website: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#5d18565f60ba%0Ahttps://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-st
24. Masiane, M. (2019). *Towards Insight Driven Sampling for Big Data Visualisation*.
25. Molina-solana, M., Birch, D., & Guo, Y. (2017). Improving data exploration in graphs with fuzzy logic and large-scale visualisation. *Applied Soft Computing Journal*, *53*, 227–235. https://doi.org/10.1016/j.asoc.2016.12.044
26. Moritz, D., Fisher, D., Ding, B., & Wang, C. (2017). *Trust but verify: Optimistic Visualizations of Approximate Queries for Exploring Big Data* (p. 2904). p. 2904.
27. Mwangi, B., Soares, J. C., & Hasan, K. M. (2014). Visualization and Unsupervised Predictive Clustering of High-Dimensional Multimodal Neuroimaging Data. *Journal of Neuroscience Methods*, 1–7.
28. Noyes, D. (2019a). The Top 20 Valuable Facebook Statistics – Updated December 2015. Retrieved March 31, 2019, from Zephoria, Digital Marketing website: https://zephoria.com/top-15-valuable-facebook-statistics/
29. Noyes, D. (2019b). Top 10 Twitter Statistics – Updated March 2019. Retrieved March 31, 2019, from Zephoria, Digital Marketing website: https://zephoria.com/twitter-statistics-top-ten/
30. Okoli, C., & Schabram, K. (n.d.). *A Guide to Conducting a Systematic Literature Review of Information Systems Research*. *10*(2010).

31. Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing Big Data with augmented and virtual reality : challenges and research agenda. *Journal of Big Data*, 1–27. https://doi.org/10.1186/s40537-015-0031-2
32. Orad, A. (2019). 2019 Gartner Magic Quadrant. Retrieved January 6, 2020, from https://www.sisense.com/gartner-magic-quadrant-business-intelligence/
33. Resnyansky, L. (2019). Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge. *Big Data & Society*, *6*(1), 205395171882381. https://doi.org/10.1177/2053951718823815
34. Ruan, G., & Zhang, H. (2017). Closed-loop Big Data Analysis with Visualization and Scalable. *Big Data Research*, *8*, 12–26. https://doi.org/10.1016/j.bdr.2017.01.002
35. Seokyeon, K., Jeong, S., Sung, U. A., Yoo, J. S., Han, S. M., Yeon, H., … Jang, Y. (2015). *Big Data Visual Analytics System for Disease Pattern Analysis p175-Kim.pdf*.
36. Shirota, Y., Hashimoto, T., & Basabi, C. (2017). *Visualization Challenge on Time Series Statistical Data* (p. 12). p. 12.
37. Sullivan, D. (2016). Google now handles at least 2 trillion searches per year. Retrieved March 31, 2019, from Search Engine Land website: http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247
38. Wang, L., Wang, G., & Alexander, C. A. (2015). *Big Data and Visualization : Methods , Challenges and Technology Progress*. *1*(1), 33–38. https://doi.org/10.12691/dt-1-1-7
39. Wang, S., & Li, W. (2019). Computers , Environment and Urban Systems Capturing the dance of the earth : PolarGlobe : Real-time scientific visualization of vector field data to support climate science. *Computers, Environment and Urban Systems*, *77*(June), 101352. https://doi.org/10.1016/j.compenvurbsys.2019.101352
40. Xie, Y., Chenna, P., Le, L., & Planteen, J. (2016). Visualization of Big High dimensional data in Three Dimensional space. *3rd International Conference on Big Data Computing, Applications and Technologies*.
41. Yang, Y., Zhang, K., Wang, J., & Nguyen, Q. V. (2015). Cabinet Tree : an orthogonal enclosure approach to visualizing and exploring big data. *Journal of Big Data*. https://doi.org/10.1186/s40537-015-0022-3
42. Zhao, H., Zhang, H., Liu, Y., Zhang, Y., & Luke, X. (2017). Journal of Visual Languages and Computing Pattern discovery : A progressive visual analytic design to support categorical data analysis. *Journal of Visual Languages and Computing*, *43*, 42–49. https://doi.org/10.1016/j.jvlc.2017.05.004
43. Zhu, S.-C. (2003). Statistical modeling and conceptualization of visual patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(6), 691–712. https://doi.org/10.1109/tpami.2003.1201820
44. Zou, Q., Zeng, J., Cao, L., & Ji, R. (2016). Neurocomputing A novel features ranking metric with application to scalable visual and bioinformatics data classi fi cation. *Neurocomputing*, *173*, 346–354. https://doi.org/10.1016/j.neucom.2014.12.123