Big Data in Health and the Importance of Data Visualization Tools
Sağlıkta Büyük Veri ve Veri Görselleştirme Araçların Önemi

Fehmi Skender[1,*], Ilker Ali[1]
[1]Department of Computer Engineering, International Vision University, Gostivar, North Macedonia
ORCIDs: 0000-0001-6046-5772, 0000-0002-2111-415X
E-mails: fehmi.skender@vizyon.edu.mk, ilker@vizyon.edu.mk
*Corresponding author.

***Abstract*—Big data concepts are increasing with their spatial speed, from personal information to extensive volume data. Since the human brain perceives visual data faster, the data must be processed and displayed appropriately. As in all areas of life, the size of the data obtained in the health sector has increased rapidly. Data storage and security have gained importance with the excessive increase in data. Big data, data mining, and visualization tools have become increasingly important to process and use data for valuation purposes. Therefore, the visualization of data and the use of analysis tools play a significant role in data processing and decision-making in the development of the health sector. The importance of data visualization tools in the health sector will become increasingly indispensable. There are many software tools developed for these purposes. This study's literature review explained the basic concepts of big data and data visualization. Research in the health sector around the world was summarized. In addition to this literature review, analyses with comparison and deduction research methods were also carried out. As a result, suggestions were made by making predictions for future studies in the health sector.**

***Keywords*—big data; data visualization; health sector**

***Özetçe*—Büyük veri kavramları kişisel bilgilerinden başlayarak büyük hacimli verilerine kadar uzaysal hızı ile artış göstermektedir. İnsan beyni görsel verileri daha hızlı bir şekilde algıladığından dolayı verilerin de uygun olarak işlenme ve gösterilmesi gerekmektedir. Yaşamın her alanında olduğu gibi sağlık sektöründe de elde edilen verilerin boyutu hızlı bir şekilde artmıştır. Verilerin aşırı artışı ile birlikte, verilerin saklanması ve güvenliğinin sağlanması da önem kazanmıştır. Verilerin işlenmesi ve faydalı amaçlar için kullanabilmesi amacıyla büyük veri, veri madenciliği ve veri görselleştirme araçları giderek önem kazanmıştır. Dolayısıyla, sağlık sektörün gelişiminde verilerin işlenmesi ve karar verilmesi konularında, verilerin görselleştirilmesi ve analiz araçların kullanılması büyük rol taşımaktadır. Sağlık sektöründe veri görselleştirme araçların önemi gittikçe vazgeçilmez olacağı aşikardır. Bu amaçlarla geliştirilmiş birçok yazılım aracı bulunmaktadır. Bu çalışmada, büyük veri ve veri görselleştirme temel kavramlarının açıklanması için literatür taraması yapılarak dünya genelinde sağlık sektöründe yapılan araştırmalar özetlenmiştir. Bu literatür taramasının yanı sıra, karşılaştırma ve dedüksiyon araştırma yöntemleri ile analizler de gerçekleştirilmiştir. Sonuç olarak, sağlık sektöründeki ilerde yapılması gereken çalışmalar için tahminlerde bulunularak öneriler yapılmıştır.**

***Anahtar Kelimeler*—büyük veri; veri görselleştirme; sağlık sektörü**

I. INTRODUCTION

Big data or "Big data" is an essential concept and has become a term very close to experts and ordinary citizens every day. Big data has recently become not just a popular term but a necessary and demanding term. It is mentioned that there is no end to the continuous interaction of software developers on the one hand and the user and business sector on the other. Along with big data, data processing and visualization is now an issue that interests everyone [1].

As in all areas of life, everyone is witnessing an increase in the volume of data, both personal and social. In the age of computers and smartphones, not only personal data but also the data of companies and organizations are increasing rapidly [2].

Data, the new trend of the 21st century, has made it difficult to extract meaningful information from it due to its increasing volume and diversity. Data analytics perspectives are broadly related to Big Data, wireless sensors, blogs, social media, emails, etc. collected from non-traditional sources such as In addition, companies can better understand large datasets by using big data analysis mechanisms. In this study, general information about the definition of big data, its relationship with other technologies, technologies used in the field of big data, and analysis techniques are given [3].

Big Data analytics can reveal bits of information hidden in too much data processing [4]. Big data has volume, variety, speed, variability, and reliability characteristics. Big Data evaluation consists of two classes: systematic use and reinforcement of new elements. Big data can be characterized by 6V data: value, volume, velocity, variety, veracity, and variability (Fig. 1) [5].

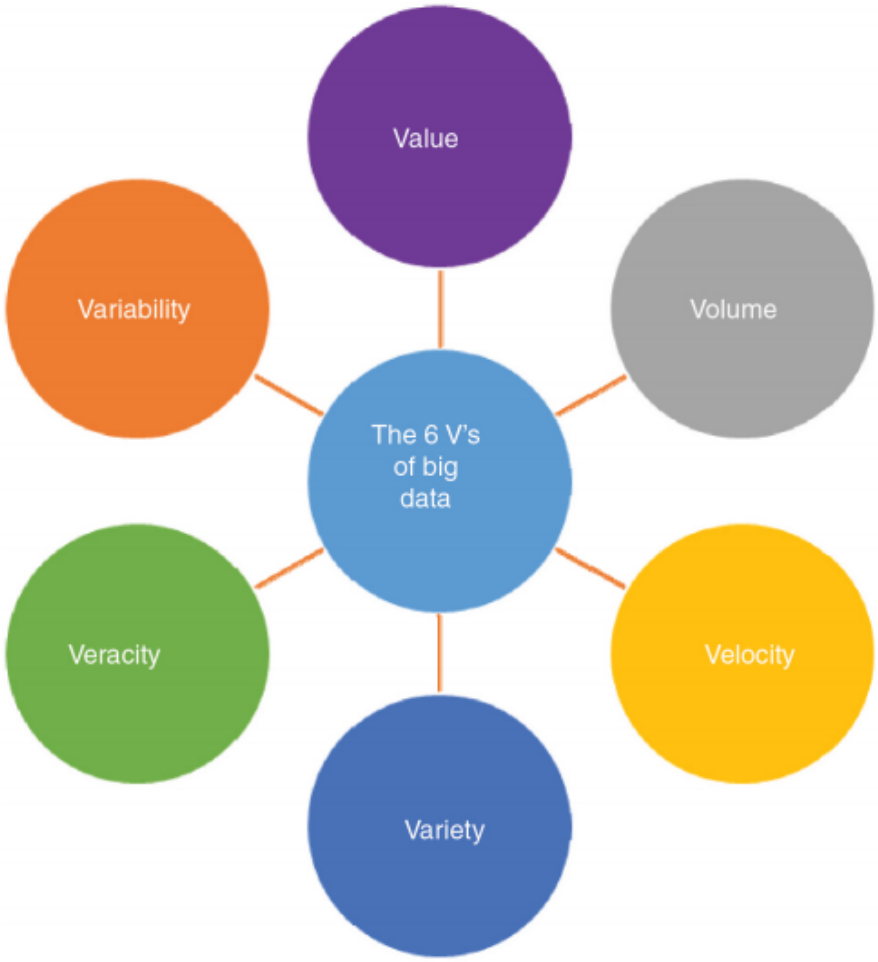

Figure 1: The big data characteristics of 6V notation.

Just as the term shading favors innovation, Big Data can be truly undefined. Big Data has become an indispensable concept for information data from social networks, web server logs, government reports, health sector data, budget business data, and every aspect of life [6], [7]. In this study, we reviewed the data visualization tools to emphasize their potentials in the healthcare sector.

II. Data Visualization Tools

Tableau Public is a very good option for data visualization. The tool attracts attention with its simple use. With Tableau Public, you can perform all kinds of visualizations easily and quickly. No coding knowledge is required for this. You can use it for free, or you can choose the paid version. In the paid version of Tableau Public, the limits are extremely wide. You can bring together different data such as Excel and PDF, as well as create tables, graphs, and maps and visualize your data without any problems [8].

You can combine or collate data and tables among themselves. You can organize your data using group and cluster properties. It is known that Tableau Public is used in many sectors, from private companies to public institutions. At the same time, you can share the data you visualize in Tableau Public with others on the internet. It is even possible to communicate with compatible mobile devices [9].

Stata is a complete toolbox that provides data management capability, analysis, and a colorful graphical interface. Stata can be referred to as statistical policy software common to international organizations such as the United Nations, governments, and institutions, including academics in Public health, Economics, Social Work, and Medicine. It remains the most powerful software available in analytics. The name Stata is a syllabic abbreviation of the words statistics and data, and it was released in 1985, followed by the graphical user interface option in 2003 [10].

Many statistical features range from descriptive and cross-tabulation analysis to more advanced techniques such as struc-

tural equation modeling, probability models, survival analysis, time series, and multilevel models. Stata gives users control over the statistical compilation of data, variables as well as groups. Stata works well with longitudinal data but can only hold a dataset in memory that needs to be rewritten to add or access a new dataset [11].

Yellowfin BI has several options. "Streamline Business Operations" can drive your company's analytical, operational and strategic planning by monitoring critical metrics through up-to-date visualizations. "Get a 360-degree View" impress your customers with your app's white-label, customized, actionable dashboards. You can embed the dashboard module and individual visualizations in wiki pages, web pages, and host applications. The tool allows you to import and embed reports from Tableau and Qlik [12]. "Customize" expands dashboards by combining widgets, action buttons, and low-code/no-code functionality with the Javascript API. Create floating panels, custom filter controls, and more with HTML, CSS, and Javascript. "Augmented Analytics" identifies critical metrics with assisted insights, automated alerts, and storytelling. "Data Stories" combines the power of data with business narratives for long-form analysis to provide context for stakeholders and customers. You can add live and static reports and add images and videos to explain the story behind the data.

R is a well-equipped open-source statistical software tool for handling the features, visualization, analysis, and features of machine learning 'heavy computing,' and although relatively new in the user space, it is strictly a programming' command-line interface (CLI) software tool, R now, cancer Research has a strong fan base of more than 6,000 packages, with contributions from data scientists, bioinformatics, and medical researchers spanning a wide range of disciplines from clinical analysis, molecular biology, phylogeny, and meta-analysis [13].

The R-studio Integrated Development Environment (IDE) with R tools works like the Oracle Database Engine using SQL [14]. The previous version was released in 1993, and the IDE was released in 2011. R, Excel files (.xls, .xlsx), Text files (.txt, .dat, .csv), SPSS (.sav), Stata (.dta), SAS (.sas7bdat), Other (.xml, JSON) . R interacts well with other software with a fairly steep learning curve given the different data types. Specifically, Metaphor is one of the many R packages available for meta-analyses and includes the most comprehensive analysis tools. Their website includes useful analysis, drawing examples, and relevant code. However, because the package requires the R environment, it may be difficult for those who have never used R to get used to the package so quickly. JASP or Jamovi packages are worth mentioning [15].

GraphPad Prism is very popular in academia and industry. It also has functions that allow researchers to perform laboratory research and clinical trial testing using probability models such as t-test, one-way ANOVA, contingency table, survival analysis, and logistic regression model. The software, unlike others, comes with an interpreted results analysis page after the predictions are generated. The language is easy to understand with less technique. The software also has a built-in automatic feature that combines analysis and graph output into a single snapshot - this, in addition to the behavior of automated reanalysis of data when any of the data points are changed, analysis or graph plotted, all performed at runtime, without the need for any user to remake [16].

SAS is the foundation of advanced analytics with functionalities spanning a variety of scientific and engineering businesses and organizations. The development of SAS (System of Statistical Analysis) was started in 1966 by Anthony Bar of North Carolina State University and later joined by James Goodnight. The National Institute of Health-funded this project to analyze agricultural data to increase crop yields. SAS is compatible with Excel files (.xls), text files (.txt, .dat, .csv), IBM SPSS (.sav), Stata (.dta), JMP (.jmp) and Other file extensions (.xml). This allows data to be easily imported and exported without resorting to error-prone manual actions. SAS also comes with a good interactive Graphical interface. But SAS can be cumbersome at times to create excellent graphics with its syntax [17].

SAS's downstream features or benefits depend on its size and proprietary license ownership. Key among these is the time to implement new ideas and methods and the techniques in the documentation process. SAS has gained popularity in Financial Services, Government, Manufacturing, Health, and Life Sciences.

MATLAB (from the Mathworks) was launched in 1984. MATLAB is a complete command-line interface (CLI) or programming language used by scientists and engineers. As with R, the learning path is steep, and at some point, you will have to create your own code. Numerous toolboxes are also available to help answer your research questions (such as EEGLab for analyzing EEG data). The difficult-to-use feature is complemented by a wide range of statistical methods and flexibility in what the software can do. MATLAB has gained popularity among scientists in engineering, numerical analysis, linear algebra, and image processing. MATLAB is compatible with Excel files (.xls, .xlsx), Text files (.txt, .dat, .csv), Other (.xml, JSON). MATLAB has good graphics and integrates easily with high-end programming software such as Python and C++ [18].

JMP combines powerful statistics with dynamic graphics in memory and on the desktop. Its interactive and visual paradigm enables JMP to reveal insights that are impossible to obtain from raw number tables or static graphs [19].

JASP is a free and open-source graphing program for statistical analysis sponsored by the University of Amsterdam. It is designed to be easy to use and familiar to SPSS users. Offers standard analysis procedures in both classical and Bayesian format [20].

III. Results

Even if the number of data visualization tools does not increase, it is constantly updated and supplemented. If it is known that the human eye is affected by colors and patterns, more than 90% of the information presented in the brain is visual. Data, which is part of daily life, no longer focuses on data structures but the purposes used. As in every field, visualization of data in health is used as if it had never been before, especially in data processing and decision making. It

has been processed and analyzed with more examples of using some of the research done in the field of health. By means of the analysis of data visualization tools used in the health sector, many researchers see the greatest benefits of the rapid progress of information and technology.

Rushton explained the first steps were taken to establish the spatial analysis basis for Geographical Information Systems (GIS), for further data processing and for more effective data processing and decision making in the health sector. In particular, the systems that contribute to the general public health are geographically mapped for the general health of the USA and then the whole world, and it can be considered as if the basis for the visualization of contemporary data has been laid [21].

Ouatik and colleagues briefly touched upon the tools for processing heterogeneous data as big data, analysis, and decision-making techniques in healthcare on large amounts of structured, unstructured, and semi-structured data. The impact of big data on healthcare services and various tools available in the Hadoop ecosystem are examined, and the conceptual architecture of data analytics is explored. In the research, attention was paid to big data, Hadoop, and MapReduce tools [22].

In another study, the research outlines various big data challenges, opportunities, and modeling methods. Consideration has been given to distributed cloud services, and automated and semi-automatic classification techniques, with examples of processing heterogeneous datasets using imaging, genetics, and health data. Despite significant advances in managing and processing large, complex, and heterogeneous data, the need to develop new innovative technologies that further develop, scale and optimize has been presented. The research has also proven that big data will affect every sector of the economy, and their distinguishing feature is "team science" [23].

In a recent research, detailed information and development opportunities about the main reasons and importance of Data Visualization and Power BI dashboards were presented. It explains the main reasons, and when the data goes through all the necessary processes, it reaches the crux with the visualization phase. Advantages of Power BI in research include a User-friendly interface, Consolidation of multiple data sources (Excel, CSV, XML, Text,...), Interactive reports and geomap visualizations, R, Python, and SQL integration, Consistent reporting and analysis, Multiple reports It is explained that it is displayed on many platforms and devices and is in continuous development. As the downsides, SAS Datasets have also proven no direct import. It has been shown with different examples how useful it is in data analysis and decision making with its use in the health sector [24].

Sparapani reviewed the use of SAS in medical data. SAS emphasizes the skills necessary to perform statistical analysis and lays the foundations for how important the data is to statistics, as well as how important it is to visualize it as a result. It was given importance to analyze the current and common old versions of SAS and to show both the basis and processing of the big data. In short, while researchers try to explain statistical concepts and methodology without making too many assumptions, it is important to ensure that not everyone can be an expert but to ensure that the software is used effectively. In the research, explanations were made from scattering charts, correlation, simple regression, and smoothing operations according to the possibilities offered by SAS [25].

Celik and Akdamar conducted a review on extracting meaningful information, presenting large amounts of data in plain text or traditional tabular format, effective graphical representations of data, visualizing data dependent on human visual perception abilities, and using computer-based interactive visual representations of abstract and non-physical data to strengthen human knowledge [26]. Examined the samples. It has been explained that multivariate data visualization, which is a type of information visualization, is indispensable to use in every field of daily life. In the study, basic information about big data is given, and general information about commonly used visualization techniques is given.

IV. Discussion and Conclusion

As the concept of Big Data is an additional part of life, its volume is growing from personal data to large companies and organizations. In the study a study was also conducted on a sample of 150 randomly selected people working in education, and their answers were aimed at both themselves and their children. Thus, it is observed that preschool children accept and send 50-70 messages per day on average. While the young and middle-aged people spend an average of 25% - 35% of their time on social networks, it can be said that an average person spends an average of 5-6 years on social networks throughout his or her life. According to the same research, while a person spends an estimated 7-9 years in front of the television in a lifetime, we witness how much data is in the loop on a daily basis; since televisions do not usually cause much data formation, almost double interaction with the Internet. The health sector means the life sector. Software developers are in constant contact with healthcare professionals. Data warehouses, OLTP, OLAP, Data centers, and similar concepts seem to fall into the shadow of Big Data.

In health, not only the organization, personnel, and financial database, but also the use of patients and medical devices and the resulting database must comply with the rules of contemporary life. Thus, while there are no problems in health management, it also provides opportunities for great progress in the diagnosis and treatment of the patients that concern them most.

Data visualization takes care of easy data recognition, processing, and transporting of ordinary citizens. Not everyone may be a computer or expert, but everyone has a natural right to be understood and benefited. In short, it contributes greatly to making the data in the graphs more understandable by coding them in the form of visual objects such as points, lines, and bars. In the health sector or in the treatment of patients, visual units in data visualization tools greatly contribute to seeing trends or outliers in the data. VDA in business intelligence tools - plays a big role in data visualization analytics. Data visualization forms the basis for the rapid assimilation of information and timely decision-making. Data is very important for everyone to access, use and understand, not just experts.

In healthcare, every second counts for patient treatment most of the time. Big data and data visualization is used in every field. However, its widespread use in health concerns everyone closely. No matter how important health is for human life, when interacting with Big Data, Data Visualization and Visual Data Analytics can have a positive impact on many people's lives.

REFERENCES

[1] Inan E, Yonyul B, Tekbacak F. A domain specific entity linking approach consuming multistore environment. Journal of Intelligent Systems with Applications 2018; 1(1): 46-52.

[2] Yigit E, Bilgin MZ, Oner AE. Predictive maintenance studies applied to an industrial press machine using machine learning. Journal of Intelligent Systems with Applications 2020; 3(2): 57-63.

[3] Zaslavsky A, Perera C, Georgakopoulos D. Sensing as a service and big data. Proceedings of the International Conference on Advances in Cloud Computing (ACC), July, 2012, Bangalore, India, pp. 21-29.

[4] Kumar A, Shankar R, Choudray A, Thakur LS. A big data MapReduce framework for fault diagnosis in cloud-based manufacturing. Loughborough University Institutional Repository, 2016.

[5] Ritevski B, Chen M. Big data analytics in medicine and healthcare. Journal of Integrative Bioinformatics 2018; 15(3): 20170030.

[6] Jee KY, Kim GH. Potentiality og big data in the medical sector: Focus on how to reshape the healthcare system. Healthcare Informatics Research 2013; 19(2): 79-85.

[7] Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ. Big data for health. IEEE Journal of Biomedical and Health Informatics 2015; 19(4): 1193-1208.

[8] Wikipedia. Tableau Software. 2021. Available at https://en.wikipedia.org/wiki/Tableau_Software.

[9] Nourani CF. Eco-morphic business digitization analytics. Researchgate 2020; available at https://www.researchgate.net/publication/342106614_Eco-Morphic_Business_Digitization_Analytics/link/5ee24b09a6fdcc73be705823/download.

[10] Gould W. The Stata blog. 2022. Available at https://blog.stata.com/.

[11] Linden A, Arbor A. Review of an introduction to Stata for health researchers, Fourth edition, by Juul and Frydenberg. The Stata Journal 2014; 14(3): 697-700.

[12] Yellowfin. Yellowfin Guide. 2022. Available at https://wiki.yellowfinbi.com/display/yfcurrent/Overview.

[13] Fox J, Andersen R. Using the R statistical computing environment to teach social statistics courses. Technical Note, in Arts Research Board of McMaster University, 2005. Available at https://socialsciences.mcmaster.ca/jfox/Teaching-with-R.pdf.

[14] Oracle. Oracle: Big-data for enterprise. An Oracle White Paper, 2011. Available at https://www.oracle.com/technetwork/database/bi-datawarehousing/wp-big-data-with-oracle-521209.pdf.

[15] Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 2010; 36(3): 1-48.

[16] Graphpad. Advice: When to plot SD versus SEM. GraphPad Statistics Guide, 2022, Available at http://www.graphpad.com/guides/prism/7/statistics/index.htm?statwhentoplotsdvssem.htm.

[17] Batko K, Slezak A. The use of big data analytics in healthcare. Journal of Big Data 2022; 9: 3.

[18] Radi B, El Hani A. Introduction to Matlab. Book chapter in Advanced Numerical Methods with Matlab 2, John Wiley & Sons, 2018.

[19] Wikipedia. JMP (Statistical Software). 2022, Available at https://en.wikipedia.org/wiki/JMP_%28statistical_software%29.

[20] Wikipedia. JASP (Jeffreys's Amazing Statistics Program). 2022, Available at https://en.wikipedia.org/wiki/JASP.

[21] Rushton G. Public health, GIS, and spatial analytic tools. Annual Review of Public Health 2003; 24: 43-56.

[22] Ouatik F, Erritali M, Ouatik F, Jourhmane M. Comparative study of MapReduce classification algorithms for students orientation. Procedia Computer Science 2020; 170: 1192-1197.

[23] Dinov ID. Methodological challenges and analytic opportunities for modeling and interpreting big healthcare data. GigaScience 2016; 5(1): 1-15.

[24] Roxane D. Power BI in clinical data exploration, powerful visuals that answer relevant questions. PHUSE Virtual EU Connect 2020, Available at https://www.lexjansen.com/phuse/2020/dv/PAP_DV09.pdf.

[25] Sparapani R. Statistical analysis of medical data using SAS. Journal of Statistical Software 2006; 16(5): 1-5.

[26] Celik S, Akdamar E. Big data and data visualization. Akademik Bakis Uluslararasi Hakemli Sosyal Bilimler Dergisi 2018; 65: 253-264.