MÜHENDİSLİK FAKÜLTESİ

Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi

International Journal of Engineering Research and Development

Cilt/Volume:10 Sayı/Issue:2 Haziran/June 2018 https://doi.org/10.29137/umagd.354768

Araştırma Makalesi / Research Article

Applıcation of Big Data Visualizatıon with Agricultural Value-Added of All Countries

Merve DOĞRUEL ANUŞLU[1], Timuçin ANUŞLU[2]

[1]*İstanbul Gedik University, Department of Industrial Engineering, 34876, İstanbul / TURKEY*
[2]*MEF University, Big Data Analytics, 34396, İstanbul / TURKEY*

Başvuru/Received: 16/11/2017 **Kabul/Accepted**: 11/06/2018 **Son Versiyon/Final Version**: 29/06/2018

Abstract

In recent years, especially with the developing technology, the concept of "big data" has emerged and it has become very important to analyse big data in both academic and sectoral studies. "Big data visualization" is a valuable and useful method in order to make it easy to understand information, and even interpretation by decision-makers who are not experts in data analysis. The agricultural sector, which is important for all humanity and plays an important role in the world economy, is a sector that needs to be examined and developed. With this study, the visualization of the agricultural value-added percentages of the countries between 1960 and 2016 in the world has been carried out. With this visualization, the most influential and non-influential countries in agriculture and changes of their value added percentage over the years can be examined with an interactive map. In order to make this study, all the steps in visualization of big data have been applied and it is aimed to be a case study for the studies of big data visualization studies which are still very rare in the literature.

Key Words

Data Visualization, Big Data, Agricultural Value Added, Choropleth Maps

merve.anuslu@gedik.edu.tr

I.BIG DATA VISUALIZATION

There are no generally accepted a definition of big data concept. However, big data includes high volume (volume), high-speed (velocity), diversity (variety) concepts (Oussous et al., 2017: 3). These three concepts are known as the 3V of big data. Veracity, value, viability, and visualization have been added these concepts recently (Fernández et al., 2016: 71).

The big data may be defined as a cultural, technological and academic phenomenon on the interaction of the following concepts (Boyd and Crawford, 2012: 663):

Technology: To maximize the calculation power and algorithm accuracy to collect, analyse, establish connection and compare the big data sets
Analysis: Process big data sets to identify the pattern with a view to making economic, social, technical and legal claims
Myth: A common belief that big data sets can present a higher level of intelligence and information and can produce understandings with reality, objectivity and accuracy aura, which was impossible before.

It is necessary to use the aesthetic shapes and functionality together to reach the information, which is hidden from the informatics in the complex and big-sized data sets (Chen and Zhang, 2014: 321). Big data visualization is important for it provides a clear information on the data patterns, reveal the secret structures in data and provides an ability to define the to-be-developed areas (Patil, 2016: 437).

Visualization, which has different definitions in the literature, may be defined as the computer-aided, interactive and visual presentation of the data to strengthen the cognitive. Cognitive may be explained as obtaining or using the perception power of human (Khan and Khan, 2011: 1).

Visualization helps to increase a person's cognitive processes. Data visualization is found to provide advantages in knowledge acquisition process for people such as understanding the big data, unexpectedly revealing the interesting structures of perceived information, quickly recognizing errors and extremeness in the data set, defining the patterns in the data, and facilitating the creation of hypothesis from the data (Tam and Song, 2016: 399-400). The main purpose of the Data visualization may be defined as discovering, analysing and describing the data for data to be easily understandable, interpretable and accelerating the decision-making processes by combining the different areas such as technology, statistics, and design.

Looking at the history of data visualization of (Friendly, 2008: 18-42), the pre-17th century is called "First Maps and Diagrams". The visualization resources in this period emerged with the maps that are prepared to help discovery and navigation with tables and geometric diagrams of the stars and other celestial bodies. The visual presentation of the information by developing new techniques and tools with the advancement of Europe in maritime in the 16th century had been done with new and more sensitive formats. 17th century was called as "Measurements and Theories". At the end of this century, which contains the beginning of the demographic statistics and political arithmetic and the rise of analytic geometry and coordinate system, measuring and forecast errors theory, the birth of possibility theory, the elements that initiate the visual thought are prepared. The 18th century was called as "The New Chart Forms." Cartographers started to show more than geographic location. Until the end of the century, geology, economics, demography and health data mapping initiatives had emerged. As the data volume progressed, the new visualization forms had emerged and the technological advancements such as colour, lithography, and press paved the way for new roads. The first half of the 19th century was called as the "Beginning of Modern Infography". All of the statistical graphics formats known today (bar and pie chart, histogram, line graph, the graph of a time series, contour graph) was discovered in that period. The simple maps in cartography transformed into data-driven complex atlas relying on various data. The second half of 19th century was called as "The Golden Age of Data Graphics) with a variety of innovation and aesthetic in thematic cartography and graphics. In this period, all the conditions for the rapid growth of visualization have been determined. The analysis offices have been established in Europe as the importance of digital information has been increased for transportation, trade, industrialization and social planning. The statistics theory initiated by Gauss and Laplace has mediated to have a better understanding of a wider and variety of data. The first half of the 20th century was called as "The Modern Dark Ages". There have been a small number of graphical innovations. The formal (commonly statistical) models and digitization in social sciences in the middle of the 1930's replaced the need for visualization at the end of the 1800's. Numbers, parameter forecasts and standard errors have especially become sensitive. Many statisticians did not interpret the pictures as real as the reality of three or more decimal number. However, this period may be interpreted as the application and dissemination area of graphical models rather than discoveries. Graphical models are available to the standards and curriculum of science, trade, and government. 1950-1975 was called as the "Rebirth of Visualization". Data visualization started to rise again in 1960's after inertia. John W Tukey in the USA has invented a variety of new, simple and effective graphical representations for "Exploratory Data Analysis – EDA". In France, the studies by Jacques Bertin and Jean-Paul Benzécri were important. At the end of this period, computer science research, data analysis developments, image and input technologies in the Bell Lab have paved the way for explosive development in new visualisation and techniques have transformed into a mature, living and multi-disciplinary research area. New paradigms of data manipulation, the discovery of graphical methods for categorical and discrete data, the discovery of new methods for visualization of multidimensional data visualization were the significant developments in that period. The large-scale statistical and graphical software engineering, whether commercial or non-commercial, the theoretical and technological facilities such as computer speed and processor capacity in line with the real data flow and big data problem enable development of the new visual methods.

Big data visualization applies to several fields such as scientific research, biology, crime analysis, education, and health and avails many benefits (Tam and Song, 2016: 401).

Converting raw data into visual format is a multi-phase procedure, containing data preparation, visualization pattern selection, visualization development, solution deployment and maintenance (Zhu et al., 2017: 30-31).

Data preparation: The preliminary processing of the raw data is usually the first step of data visualization. Basic preliminary processing; data formatting, missing value processing, outliers values processing, value normalization, and classification.

Visualization pattern selection: The second step is to select the most accurate and suitable processed data.

Visualization development: The third step is to convert the data set to the selected visual model. There is no difference between the static and interactive graphics; interactive graphics enable users to control and explore the data themselves. Visualization development depends on the software and package used.

Solution deployment and maintenance: Visualization solution is shared with the relevant stakeholders in the last step. The development must be made to keep up-to-date the solution. An update may manually be made with the newest data or a continuous update may be done with real-time data.

II.THE AGRICULTURAL INDUSTRY AND ECONOMIC IMPORTANCE OF VALUE ADDED CONCEPT

Agriculture is at the centre for reduction of poverty and hunger, improving the gender inequality and sustainable management of the environment (Byerlee et al., 2009: 17). Besides, the agriculture industry is of critical importance at economies with all development levels with its share in national income, employment, foreign trade, agricultural industry, aid and consumption expenses (Bayraç and Doğan, 2016: 24).

Playing an important role in country development, the agriculture industry has been increasingly integrated into the globalization process all across the world. The position of the agriculture industry in the overall economy is measured by its share of the value added created by the agriculture industry in the overall economy (Ege, 2011: 1). Value added refers to the balance which emerges after the deduction of purchased items to create that item from the sales price of that product. 1975 to present is called as "High-D, Interactive and dynamic data Visualization." The data visualization in the last quarter of the 20th century the national product is the total of the value added created by all the economic units in producing commodity and service in a country. There is a direct relationship between the national product with value added, and high value added means high national product (Güneş et al., 2015: 99).

According to the 2014 Agricultural Value-Added Data, it constitutes the 4,3% of the world GDP. The top 20 countries, which produce the 80% of total GDP, constitutes the 66% of agricultural GDP. The top 10 countries produce the 55% of total agricultural value added (Türkiye İhracatçılar Meclisi Tarım Raporu, 2016: 33).

III.APPLICATION

The objective of this study is to represent the agricultural value-added shares of the countries on a global map in a quick and apparent manner. To that end, to convert the raw data to visual format; data preparation, visualization pattern selection, and visualization development have been applied.

To visualize the agricultural value-added shares by countries on the map, choropleth maps which a kind of the thematic maps are suitable to use since it makes more understandable by colouring geographical areas or regions according to the variables.

Hadoop, R, Data Manager, D3, Tableau, Python, Splunk are some of the most popular tools for big data visualization (Miller, 2017: 23). R programming language is used, in which software libraries that are suitable for choropleth maps in this study.

The data preparation phase involves obtaining the agricultural value added of the countries, acquiring necessary data for mapping, completion of missing data and optimizing the data. Agricultural value-added value by countries is obtained from the WDI Library from the "The World Bank" in the currency of US Dollars. Mapping data is created by using ggplot library. Upon analysing the agricultural value-added data of the countries, we detected that some countries have a missing information in some years and such data are completed with spline interpolation among the interpolation methods, which enable calculation of unknown values by making use of the known values at some intervals. Small degree polynomials are created by dividing data interval to smaller intervals in spline interpolation and such polynomials are used to forecast the missing data according to their intervals. A linear spline interpolation is created in the first degree, quadratic one in the second degree and cubic in the third degree (Vatansever and Doğalı, 2011: 51). ImputeTS library has been used to complete the data with spline interpolation. It is required to use standardized data in the choropleth map (Muehlenhaus, 2014: 147), therefore, agricultural value-added data in USD is adjusted to be used in the application by proportioning with the total agricultural value added after the completion of missing data.

Sub-data set is created for 1960 from the resulting data and static choropleth map of the data has been generated as in Figure 1 with map feature of the ggplot library. Colour gamut has been distributed according to the log values of the data to have a clearer distinction between the agricultural value-added percentages in the intervals where the data has been intensely distributed. The visual generated, found suitable in line with the purpose of the study.

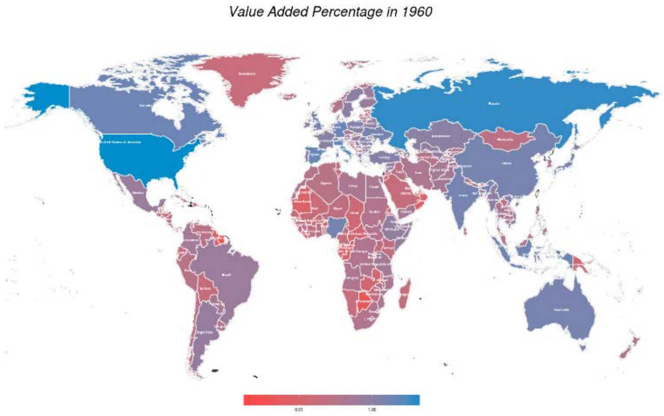

Figure 1: Static Choropleth World Map with Value-added Percentages by Countries-1960

By making the choropleth maps interactive in Figure 2, Figure 3 and Figure 4, it enables to monitor the countries of which agricultural value-added percentages are increased, decreased or remain stable on the basis of the year (http://www.anuslu.com/b3s17/map/BigDataAgricultureVisualization.html).

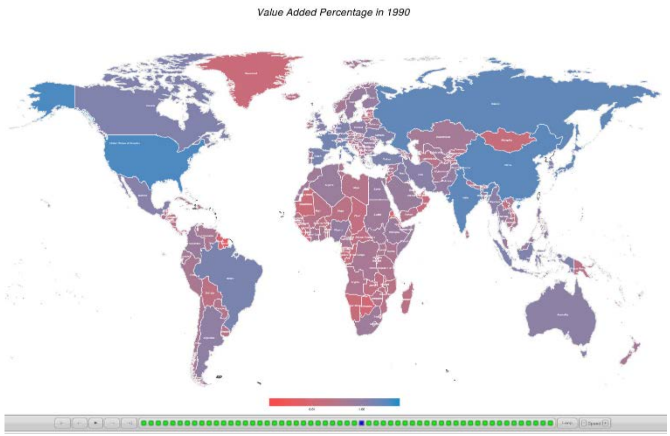

Figure 2: Interactive Choropleth World Map with Value-added Percentages by Countries-1990

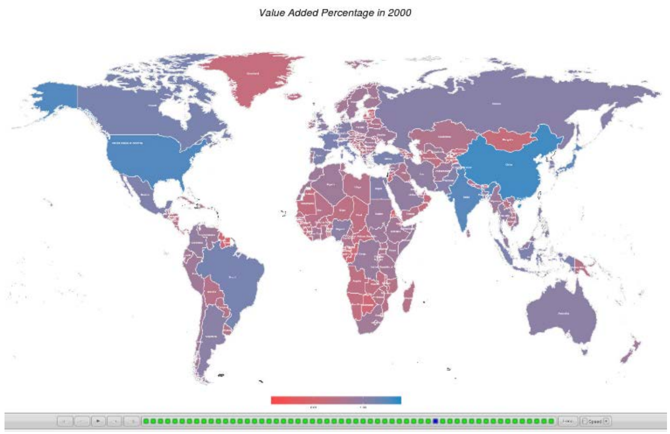

Figure 3: Interactive Choropleth World Maps with Value Added with Percentages by Countries-2000

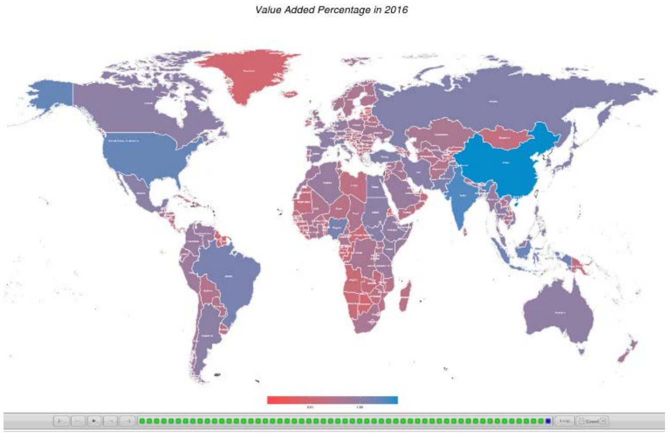

Figure 4: Interactive Choropleth World Maps with Value Added with Percentages by Countries-2016

In accordance with the development phase of the visualization, the interactive map software is designed to enable updating the following years.

IV.CONCLUSION

A visual has been carried out to monitor the agricultural value added and the change of countries' shares in the world between 1960-2016 and their agricultural value-added percentages and their positions in the world. The steps of the big data visualization have been applied through R software in the scope of this study and the interactive choropleth maps that show the value-added percentages by the countries are generated. By observing the colour changes in the map, we interpret the agricultural value-added percentages of the countries and their shares in the world by years.

Figure 5 illustrates the value-added percentages of the top 10 countries between 1960-2016 with agricultural value-added percentages in 2016.

Upon observing the interactive map, we can clearly see that China has the largest share in the world with agricultural value added in recent years. When we analyse the China with colour changes in the map, we can say that this impact started in 1990's and significantly accelerated in 2000. China is clearly the world leader in agricultural added value in the recent years and significantly improves its agricultural value-added percentage.

We can say that the USA has always a good role in the world with its agricultural value added. When we analyse the map colour transitions in the USA;

We can say that the USA is the most influential country in the world with its agricultural value added between 1960-1990, however, it lost its leadership despite taking place near the top with its agricultural value added after 1990's.

Russia was the second influential country in the USA in the world with its agricultural added value at the beginning of the 1960's, however, it had fluctuations in time and decreased its value-added percentages.

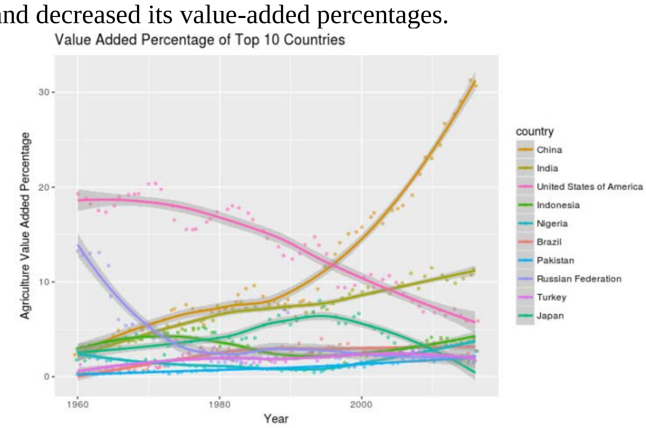

Figure 5: The change of top 10 countries in 2016 in years in comparison to other countries.

We can observe that India has a good place in the world with the agricultural added value in recent years and its agricultural value-added percentage increased in recent years.

We can say that Turkey did not have a significant colour transition in the map between 1960-2016 and therefore, it has a stable place in the world with agricultural value added.

We can generally identify the countries of which agricultural value-added percentages have increased, decreased and remain stable with this interactive visualization study at a short notice. This study illustrates the steps of the big data visualization with an application and remarks that data can be easily understood and interpreted with visualization. It is aimed that this study will be a reference to other studies in the big data domain for application of visualization.

REFERENCES

Bayraç, H. N. and Doğan, E. (2016). Türkiye'de İklim Değişikliğinin Tarım Sektörü Üzerine Etkileri. *Eskişehir Osmangazi Üniversitesi ˙IIBF Dergisi*. 11(1), 23-48.

Boyd, D. and Crawford, K. (2012). Critical Questions For Big Data. Information, Communication & Society. 15(5), 662-679.
Byerlee, D., Janvry, A. and Sadoulet, E. (2009). Agriculture for Development: Toward a New Paradigm. Annuel Review of Resource Economics.1, 15-31.

Chen, C.L. P. and Zhang C.-Y. (2014). Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. Information Science. 314-347.

Ege, H. (2011). Tarım Sektörünün Ekonomideki Yeri ve Önemi. Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü. 1-4.http://www.tepge.gov.tr/dosyalar/yayinlar/62889464af374f2fa8a14db901213d54.pdf Retrieved: 28 July 2017

Fernández, A., Carmona, C. J., Jesus, M. J. and Herrera, F. (2016). A View on Fuzzy Systems for Big Data: Progress and Opportunities. International Journal of Computational Intelligence Systems. 9(1), 69-80.

Friendly, M. (2008). A Brief History of Data Visualization. Chen, C., Härdle, W. and Unwin, A. (Ed.). In Handbook of Data Visualization. Berlin: Springer-Verlag, 2008, 15-56.

Güneş¸ S., Togay, A. and Güneş, Ç. (December2015). Katma Değer ve Kalkınma Bağlamında Ürün Tasarımı. Gazi Üniversitesi Sanat ve Tasarım Dergisi. 97-112.

Khan, M. and Khan, S. S. (November 2011). Data and Information Visualization Methods, and Interactive Mechanisms: A Survey. International Journal of Computer Applications. 34(1), 1-14.

Miller, J. D. (2017). Big Data Visualization. Birmingham: Packt Publishing.

Muehlenhaus, I. (2014). Web Cartography: Map Design for Interactive and Mobile Devices. Boca Raton: Taylor&Francis Group.
Oussous, A., Benjelloun, F-Z., Lahcen, A. A., and Belfkih S., (2017). Big Data Technologies: A Survey. Journal of King Saud University – Computer and Information Sciences. 1-18.

Patil, S. S. (12-13May 2016). Overview of Big Data Visualization. International Journal of Advanced Networking & Applications (IJANA), Bengalura: RajaRajeswari College of Engineering, 436-438.

Tam, N. T. and Song, I. (2016). Big Data Visualization: Application in Visualizing Learning Activities. Kim, K. J. and Joukov, N. (Ed.). In Information Science and Applications (ICISA) 2016. Singapore: Springer Science+Business Media, 2016, 399-409.

Türkiye İhracatçılar Meclisi Tarım Raporu 2016. http://www.tim.org.tr/files/downloads/Raporlar/Tarim_Raporu_2017.pdf Retrieved: 3 August 2017

Vatansever, F. and Doğalı, G. (16-18 May 2011). Klasik Enterpolasyon Yöntemleri ve Yapay Sinir Ağı Yaklaşımlarının Karşılaştırılması. 6th International Advanced Technologies Symposium (IATS'11), Elazığ: Fırat University, 51-54.

Zhu, Z., Heng, B. H. and Teow, K.-L. (2017). Interactive Data Visualization to Understand Data Better: Case Studies in Healthcare Systems. Information Resources Management Association (Ed.). In Decision Management: Concept, Methodologies, Tools and Applications. Hershey: IGI Global, 2017, 27-37.