

BIG URBAN DATA VISUALIZATION APPROACHES WITHIN THE SMART CITY: GIS-BASED OPEN-SOURCE DASHBOARD EXAMPLE

Rabia BOVKIR and Arif Cagdas AYDINOGLU

Dept. of Geomatic Engineering, Gebze Technical University, Kocaeli –Turkey
rbovki@gtu.edu.tr; aydinoglu@gtu.edu.tr

KEY WORDS: Big Data, Visualization, Dashboard, Smart City, City Dashboard

ABSTRACT:

The change and rapid development in information and communication technologies in recent years has caused increasing the data sources and the volume of data. With the smart city concept, big and complex data collected from the internet of things (IoT), smart sensors, and wearable technologies has been introduced into our lives. Such big data can provide various opportunities and possibilities in urban management, however, to obtain meaningful information from big volume complex data is challenging. Visual analytics techniques provide an effective way for extracting significant information from big data. With the integration of geographic information systems (GIS) and its ability on geographic analysis, big data produced by the IoT and sensors can be analysed and represented in real time via city dashboards. In this study, literature research about big data visualization approaches within the smart cities is examined. Dashboard examples with the integration of GIS is examined and an open-source GIS-based dashboard example is designed in Apache Superset. The relationship between air quality and traffic density is demonstrated with different visualization methods.

1. INTRODUCTION

In recent decades, within the sustainability of the resources and efficiency of the services, the smart solutions and applications emerged and evolved rapidly. The problems, inadequacies and contradictions have agglomerated in the urban areas due to the uncontrolled and sudden urbanization (Cao et. al., 2020). Smart city is a concept emerged with the aim of mitigating the negative effects of rapid urbanization, industrialization and consumption and increasing the efficiency of urban services (Taamallah et. al., 2017). Since their initial adaptation, smart cities have evolved over the years, defining almost any kind of technological advancement and innovation with the development of the internet of things (IoT), smart sensors and devices (Allam and Dhunny, 2019). One of the most important ability of smart cities is to provide the real-time data in order to enable better decision-making.

The emerging of smart city concept strongly promotes the big data production through the implication of IoT, smart devices and sensors. Big data can provide new opportunities and possibilities in the management and development of cities (Allam and Dhunny, 2019). However, it is quite difficult to reach meaningful information from this such huge volume of data. According to International Data Corporation (IDC), only in 2020, 64.2 ZB (1 ZB=10²¹ byte) of data was created or replicated (URL-1). Therefore, big data storage, management, analysis, and representation play a key importance for successful implementations. The most efficient way to extract meaningful information from big volume of complex data is to use visual analytics techniques (Balzer et. al., 2020; Wang and Lu, 2020).

Visualization is the analysis and presentation of the internal structure of data, datasets, and information by using different methods such as computer graphics, digital image and signal processing (Wang and Lu, 2020). Visualizing technology make possible to extract useful information from large, unstructured,

and complex datasets via its ability to fast explore and interpret the big data. With the integration of geographic information systems (GIS) and its ability on geographic analysis and geo-statistics, big geographic data produced by the IoT and sensors can be analysed and represented in real time via city dashboards. Optimized data on dashboards is presented as meaningful information, increasing efficiency by helping citizens to focus more easily on the results of implementations within the scope of smart cities (Lavalle et. al., 2020).

In this study, literature research about big data visualization approaches within the smart cities is examined and most used visualization techniques is listed with examples. In addition, dashboard approaches with the integration of GIS for smart cities are reviewed. As an example of visualization approach for smart cities, an open-source GIS-based dashboard is designed in Apache Superset. The relationship between air quality parameters and traffic density is demonstrated with different visualization methods via designed GIS-based open dashboard example.

2. BIG DATA VISUALIZATION APPROACHES

In data mining applications, it is very important to understand the relationships of data with each other. The best way to provide this understanding during the data discovery is to visualize the data and build creative bridges between human perception and computer systems (Keim, 2002). Data visualization techniques consist of a combination of many disciplines such as computer graphics, image processing, computer vision, and user interface design.

Visual data mining techniques are classified under different names in the literature (Figure 1). Visual data mining is classified according to the *Data Type to Be Visualized*, *Visualization Techniques*, and *Interaction and Distortion Techniques* (Keim, 2002).

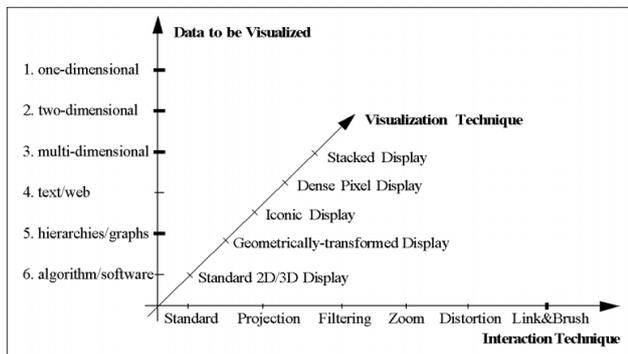


Figure 1. Classification of Data Visualization Techniques by Keim, 2002.

Data Type to Be Visualized: Visualization techniques vary according to the processed data types.

- **One-Dimensional Data:** One-dimensional data has only one variable. Temporal data can be an example to one-dimensional data (Keim, 2002).
- **Two-Dimensional Data:** Two-dimensional data has two different dimensions. Geographic data with two different dimensions (latitude and longitude) is a typical example of two-dimensional data. Maps with X-Y plots are a typical method for displaying two-dimensional data (Keim, 2002).
- **Multidimensional Data:** Many datasets contain more than three dimensions, so they cannot be observed with simple visualization techniques such as 2D or 3D graphs. Sophisticated visualization techniques are used in order to represent this type of data (Keim, 2002).
- **Text and Hypertext:** It is not possible to demonstrate all data types in sizing terms. Multimedia sources like web pages consist of text and hypertexts. Principle component analysis (PCA) and multi-dimensional scaling (MDS) are two of the leading methods applied for dimensional reduction of this data type.
- **Hierarchies and Graphs:** Data records are often associated with other pieces of information and graphs are often preferred to show these types of ordered, hierarchical or arbitrary relationships. A graph compose of a set of objects, called “node”s, and connections between these nodes, called “edge”. File structures in computers, e-mail relations between people or shopping behaviours can be given as examples (Keim, 2002).
- **Algorithms and Software:** The purpose of software visualization is to enable the development of the software by helping to understand algorithms. Graphical techniques can be classified according to the graphical possibilities of the software (Keim, 2002).

Visualization Techniques: There are various visualization techniques for visualizing data. In addition to standard 2/3D techniques, there are special visualization techniques developed for multi-dimensional and complex data.

- **Standard 2/3D Displays:** The best known of such techniques are scatterplots, which plot two and three variable datasets along the x, y, and z axes into the Cartesian coordinate system. In addition, line, bar, histogram, and pie graphics are also available between standard 2D/3D display graphics.
- **2D-3D Scatterplots:** The scatter plot is a visual technique that helps to determine the direction, type, and magnitude of the relationship between two or three variables. In the 2D scatter plot, the pair of values (x_i, y_i) obtained from n units are shown on

the xy coordinate plane; in the 3D scatter plot, the trio of values (x_i, y_i, z_i) obtained from n unit are shown on the xyz coordinate plane, in such a way that all values are shown on the relevant axes (Keim, 2002).

- **Box Plots:** Box plots are suitable to be drawn in terms of recognizing the data before starting analytical studies. The long lines at the top and bottom of the chart correspond to the 25% and 75% percentiles of the data sorted by size. If the median is not in the middle of the graph, this graph indicates that the data is skewed. The hyphens at the bottom and the top of the chart show the rest of the data, assuming there are no outliers in the data. Plus signs at the top and bottom of the graph indicate outliers in the data (Vatansever, 2008).
- **Line and Multiple Line Graphs:** Line charts are used to show single-valued or piecewise continuous functions of a variable. Line charts are generally preferred to show two-dimensional data (x, y). On the other hand, multi/multiple line charts are used for visualizing data with more than two variables, such as $(x_1, y_1, y_2, y_3, \dots)$. However, there can be some problems in multi-line charts used in the display of multi-dimensional data because of size separations (Keim, 2002) (Figure 2).



Figure 2. Box Plot (upper-left), Multiple Line (upper-right) and Scatterplot Graphs (bottom) in 2D (URL-2)

- **Geometrically Transformed Displays:** Geometrically transformed representations aim to highlight the interesting information within multidimensional datasets. Most common techniques in this class are scatterplot matrixes, RadViz, survey plots, parallel coordinates, Andrews curves, principal component analysis (PCA), multidimensional scaling (MDS), and self-organizing maps (SOM). *Scatterplot matrixes* are one of the most used multidimensional visualization technique, thanks to their convenience in detecting correlations between binary variables by showing the interactions/correlations of two or more variables at the same time (Vatansever, 2008). However, it is difficult to obtain information from scattering matrixes for analysing large datasets. *Parallel coordinates* allow to represent many axes in the same two-dimensional plane by drawing axes parallel to each other (Figure 3). As the number of objects increases, a large number of overlapping lines render the image uninterpretable (Vatansever, 2008). *PCA* is a commonly used method of expressing the variance structure of the original variables with fewer new variables that are linear components

of these variables. It can be used for data reduction, visualization, clustering and estimation.

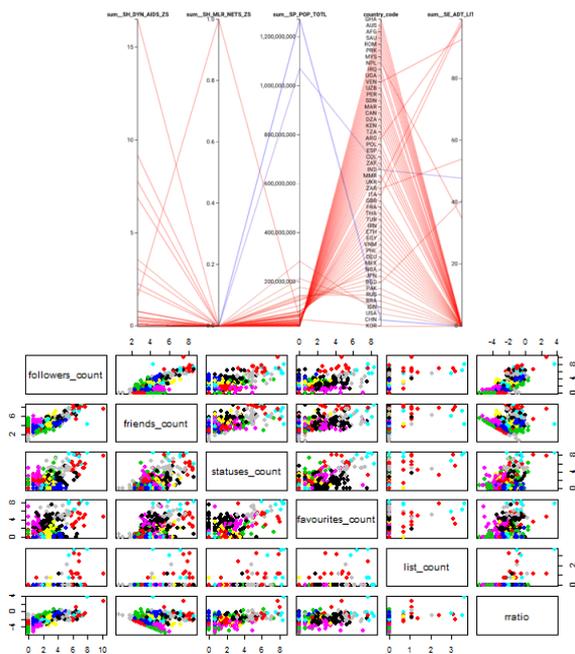


Figure 3. Parallel Coordinates (upper) and Scatterplot Matrix (bottom) Examples (URL-2)

- **Iconic Displays:** The idea in iconic/symbolic representation methods is to map the features of multidimensional data using symbolic features. Each visual property of the icon changes according to the values contained in the data. Typical examples of symbolic representations are Chernoff faces and star plots (Figure 4).

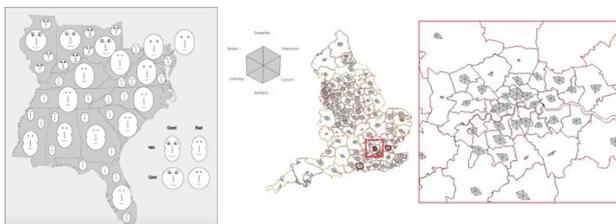


Figure 4. Map Visualizations with Chernoff Faces (left) (Nelson, 2007) and Star Plots (right) (McNabb and Laramee, 2019)

- **Dense Pixel Displays:** This representation is based on the principle that values of different sizes are mapped to different coloured pixels and the pixels of each size are assigned to new groups, considering their neighbourhood. This technique is suitable for the visualization of large multidimensional datasets (Figure 5).

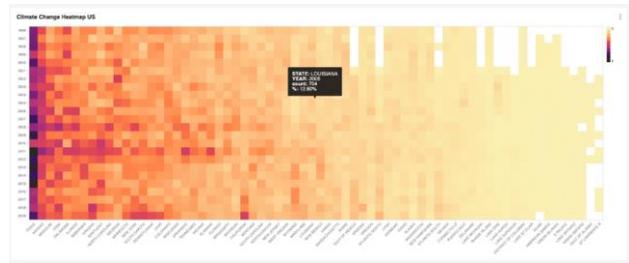


Figure 5. Pixel Display Examples (URL-2)

- **Stacked Displays:** The most used method in this demonstration is dimensional stacking. This method is based on the principle of embedding one coordinate system into another. The best-known examples of this type is N-Vision, or the system called "World-within-Worlds", Treemap and Cone Trees (Keim, 2002). (Figure 6).

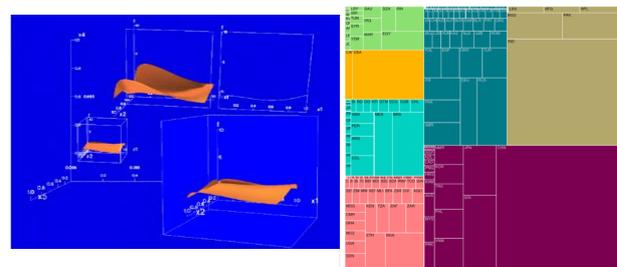


Figure 6. World-within-worlds (left) (Beshers and Feiner, 1993) and Treemap (right) (URL-2) Examples

Interaction and Distortion Techniques: Some tools and methods are needed to use data visualization techniques effectively. Such tools that increase the effectiveness of visualization techniques are (Keim, 2002):

- **Dynamic Projections:** The basis of this method is to search for meaningful information by dynamically changing the projections of the multidimensional data set. The classic example of dynamic projections is the GrandTour system showing all the two-dimensional projections of the multidimensional dataset as a series of scatterplots.

- **Interactive Filtering and Zooming:** When searching for information in large data sets, segmenting the data set interactively and dealing with interesting subsets is important in terms of saving time and making information easier to find. Similarly, zooming is a widely used and well-known technique that enables the data to be viewed in different resolutions to get useful information. In addition, in very large datasets, when all data is highly compressed, it provides a general view dominance for all of them.

- **Interactive Distortion:** Interactive distortion techniques ensure the preservation of the general structure of the data while scanning the subfiles of the data stack hierarchy, called drill-down, during the information search process. With this technique, some of the data can be observed in full detail, while the rest can be examined in less detail.

- **Interactive Linking and Brushing:** The purpose of using interactive linking and brushing is the use of different visualization methods together to overcome the shortcomings of the techniques used alone. The process of showing subsets of a graph with different symbols and colours is called "brushing", and the automatic marking of observations marked on one graph in the other graph is called "linking". In interactive linking and

brushing, data subsets marked in different colours and symbols in one graph are displayed simultaneously in different colours and symbols in the other graph (Figure 7).

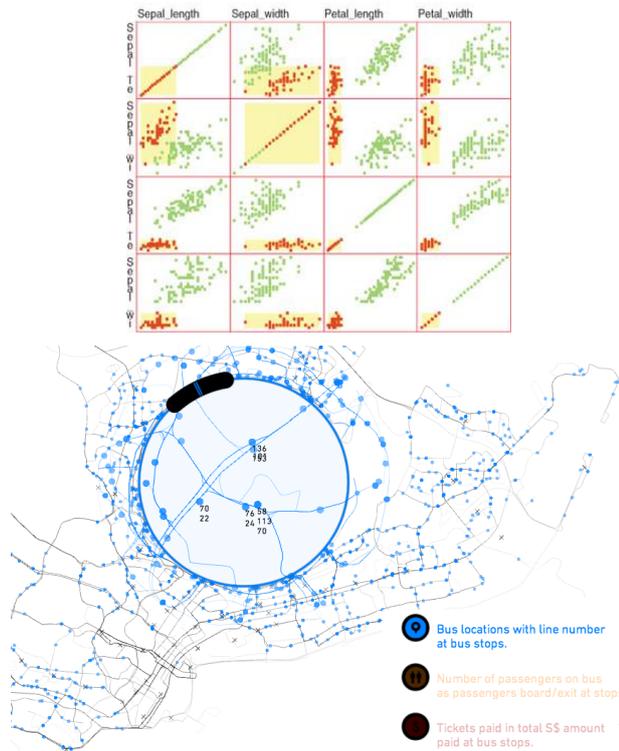


Figure 7. Interactive Linking/Brushing (upper) (Ward, 2009) and Distortion (bottom) (URL-3) Examples

3. GIS-BASED OPEN-SOURCE DASHBOARD EXAMPLE FOR SMART CITIES

Dashboards are popular information systems elements that provides advantages in terms of managing big data. Using a dashboard for providing meaningful information in graphical and other forms of visualization has been increasingly used within the smart city applications (Kitchin et al., 2015). Web dashboards can be configured in three different ways, as operational, tactical, and strategic dashboards. Urban dashboards designed clearly with open data strategy make available the users to understand and focus on big data in a wider perspective (Pettit et al., 2017).

In recent years, the exponential growth in urbanization and population in cities has brought various problems such as air, water, soil pollution due to the dense traffic, unconscious and excessive consumption of resources (Lu and Cao 2015; Sabrin, et al., 2020). Air pollution is accepted as one of the most important factors affecting human health negatively by the World Health Organization (WHO) and it is stated that the highest urban air pollution is coming from road traffic due to the the motor vehicles (WHO, 2021). Therefore, it is important to monitor the relation of air quality and traffic in smart cities for providing interventions when needed.

In this study, the relation between air quality and traffic density is visualized and presented via open-source dashboard for Istanbul. Air quality data includes parameters such as PM_{2.5}, PM₁₀, O₃, SO₂, NO₂ and CO is obtained hourly for January 1st

to December 31st, 2020; from sensors in National Air Quality Monitoring Network (www.havaizleme.gov.tr) established by the Ministry of Environment and Urbanization. Traffic data is obtained hourly for same date range from the Open Data Portal (data.ibb.gov.tr) of Istanbul Metropolitan Municipality. NoSQL database environment MongoDB is used for the management and querying of air quality data and PostgreSQL is used for the management and querying of traffic data. Apache Superset is used for the visualization and dynamic presentation of air quality parameters and traffic density.

Apache Superset is a fast and lightweight, modern enterprise-ready business intelligence web application with many visualization tools from simple graphs to detailed charts (Superset, 2021). It supports connection with various database environment such as PostgreSQL and with this integration, traffic density data located 250 m away from the air quality data was merged in PostGIS and transferred to Superset. Stations are visualized with the Mapbox map tool and OpenStreet satellite imagery was used as a basemap. In addition, average driving speed and pollutant value are visualized with bar graphics (Figure 8).

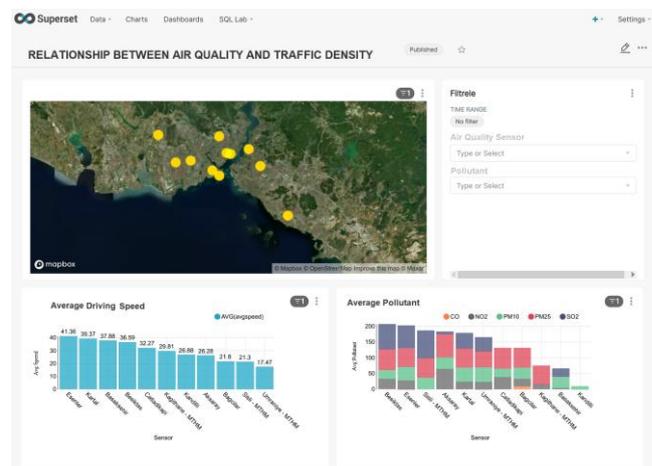


Figure 8. Designed GIS-based Dashboard Example for Istanbul (1)

In Figure 9, sensor-based air quality according to active pollutants is visualized as a line time series graph; the concentration ratios of active pollutants in 2020 is visualized as a pie chart graph; the ratio of the active pollutant at the air monitoring stations and the average number of vehicles close to the relevant station is visualized as a sunburst graph; the proportions of pollutants according to the air monitoring stations as a treemap graph; the most effective pollutant recorded by air monitoring stations as a heatmap graph; air quality, average, minimum and maximum speeds and average number of vehicles are visualized as box graphs.



Figure 9. Designed GIS-based Dashboard Example for Istanbul (2)

4. CONCLUSION

The recent developments and implementations in smart cities promote the big data production and analysis via IoT and smart sensor networks. Effective big data management in smart city implementations can provide successful management and decision making in city policy. However, extracting meaningful and useful information from big data needs extra effort and methods. The most effective way to obtain significant information from big data is to apply data mining and visualization techniques. With data visualization, the relationships between data can be revealed and flexible and creative bridges can be built between human perception and computer systems for deeper understanding.

Dashboards are very useful tools to focus on the data produced from smart city implementations and to understand data more easily with dynamic representations. Successful smart cities adopt their own smart dashboards, like London (URL-4), Amsterdam (URL-5), Barcelona (URL-6), Dublin (URL-7). In this perspective, GIS-based dashboard is designed as an example big data visualization platform for smart cities.

In this study, the relationship between air quality parameters and traffic density is demonstrated in Istanbul. By using different visualization methods like time series graphs, pie and bar chart, heatmap and boxplot, more easy and deeper understanding is provided to users via the designed example. The biggest challenge with big data visualization in this study is missing data and visual noise problems. Missing data problem causes information loss and corrupt the data integrity. Visual noise occurs when the data is too relative to each other in the dataset and it makes the data division is difficult. Therefore, pre-process and data mining steps before the visualization is extremely important.

REFERENCES

Allam, Z. and Dhunny, Z. A., 2019. On big data, artificial intelligence and smart cities. *Cities*, 89, 80-91.

Balzer, C., Oktavian, R., Zandi, M., Fairen-Jimenez, D., Moghadam, P. Z., 2020. Wiz: A Web-Based Tool for Interactive Visualization of Big Data. *Patterns*, 1(8). <https://doi.org/10.1016/j.patter.2020.100107>

Beshers, C. and Feiner, S., 1993. AutoVisual: Rule-Based Design of Interactive Multivariate Visualizations. *IEEE Computer Graphics and Application*, 13(4), 41-49.

Cao, X., Wang, M., Liu, X., 2020. Application of Big Data Visualization in Urban Planning. *IOP Conf. Ser.: Earth Environ. Sci.* 440 042066.

Keim D. A., 2002. Information Visualization and Visual Data Mining. *IEEE on Transactions on Visualizations and Computer Graphics*, 8, 100-107.

Kitchin, R., T. Lauriault, and G. McArdle, 2015. Knowing and governing cities through urban indicators, city benchmarking, and real-time dashboards. *Regional Studies, Regional Science, Journal of the Regional Studies Association*, (1), 6-28.

Lavalle, A., Teruel, M. A., Maté, A., Trujillo, J., 2020. Improving Sustainability of Smart Cities through Visualization Techniques for Big Data from IoT Devices. *Sustainability*, 12(14). <https://doi.org/10.3390/su12145595>

Lu J. ve Cao X., 2015. PM_{2.5} pollution in major cities in China: Pollution status, emission sources and control measures. *Fresenius Environ. Bull.*, 24, 1338–1349.

McNabb, L. and Laramée, R. S., 2019. Multivariate Maps—A Glyph-Placement Algorithm to Support Multivariate Geospatial Visualization. *Information*, 10(10), 302. <https://doi.org/10.3390/info10100302>

Nelson, E. S., 2007. Face Symbols: Research Issues and Cartographic Potential. *Cartographica*, 42(1), 53-64.

Pettit C., Lieske S.N., Jamal M., 2017. *CityDash: Visualising a Changing City Using Open Data*. In: Geertman S., Allan A., Pettit C., Stillwell J. (eds) *Planning Support Science for Smarter Urban Futures*, Springer, Cham.

Prandi, F., Soave, M., Devigili, F., Andreolli, M., De Amicis, R., 2014. Services oriented smart city platform based on 3D city model visualization. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, 2, 59-64.

Sabrin, S., Karimi, M., Fahad, Md G.R. and Nazari, R., 2020. Quantifying environmental and social vulnerability: Role of urban Heat Island and air quality, a case study of Camden, NJ. *Urban Climate*, 34, 100699.

Superset, 2021: *What is Apache Superset?* Accessed at: <https://superset.apache.org/docs/intro#what-is-apache-superset> (07.07.2021).

Taamallah, A., Khemaja, M., and Faiz, S., 2017. Strategy ontology construction and learning: Insights from smart city strategies. *International Journal of Knowledge-Based Development*, 8(3), 206–228.

Tecuci, G., 2012. Artificial intelligence. *Wiley Interdisciplinary Reviews. Computational Statistics*, 4, 168–180.

Vatansever, M., 2008. Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı ve Uygulanması. MSc Thesis, Yıldız Technical University.

Wang, W. and Lu, C., 2020. Visualization analysis of big data research based on Citespace. *Soft Comput.*, 24, 8173–8186. <https://doi.org/10.1007/s00500-019-04384-7>

Ward, M. O., 2009: *Linking and Brushing*. In: Liu, L. and Ozsu, M. T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_1129

WHO, 2021: *Official website of WHO*. Accessed at: <https://www.who.int/news-room/air-pollution> (07.07.2021).

WEB REFERENCES

URL-1:

<https://www.idc.com/getdoc.jsp?containerId=prUS47560321>
(Accessed: 07.07.2021).

URL-2:

<https://apache-superset.readthedocs.io/en/0.28.1/gallery.html> (Accessed: 07.07.2021).

URL-3:

<http://pmcruz.com/information-visualization/data-lenses> (Accessed: 07.07.2021).

URL-4:

<https://citydashboard.org/london/> (Accessed: 07.07.2021).

URL-5:

<https://citydashboard.waag.org/> (Accessed: 07.07.2021).

URL-6:

<http://perceptuali.com/Production/Smart-City.php>
(Accessed: 07.07.2021).

URL-7:

<https://www.dublindashboard.ie/#:~:text=Dublin%20Dashboard%20provides%20citizens%2C%20public,all%20aspects%20of%20the%20city.&text=The%20underlying%20data%20is%20fr eely,their%20own%20applications%20and%20visualisations>
(Accessed: 07.07.2021).