Analysis of Current Visualization Techniques and Main Challenges for the Future

Manuel Pérez Cota[1]*, María Díaz Rodríguez[1], Miguel Ramón González-Castro[1], Ramiro Manuel Moreira Gonçalves[2]

[1] *University of Vigo*, SPAIN
[1] *University of Tras-os-Montes e Alto Douro*, PORTUGAL

***Corresponding Author:** mpcota@uvigo.es

ABSTRACT

The big amount of data generated nowadays are being used by Big Data tools to generate knowledge and to facilitate the decision-making. However, this situation creates a new challenge: how to visualize all these data without losing mid/long term crucial information. The purpose of this article is to analyze the state of the art on massive data visualization, main problems and challenges of information representation current techniques as well as the evolution of the tools and the future of them, in other words, new functionalities to offer.

Keywords: masive data visualization, big data, 3D, interactive visualization, visualization tools

INTRODUCTION

Big Data new technologies are able to handle big amounts of data with the goal of obtaining new knowledge that facilitates the decision-making. However, this process that starts from data to obtain information and, from here, to achieve knowledge, is not only focused on the treatment of the data, but that all this has no value if they cannot be displayed in the right way so that the user understands what is being shown and not be overwhelmed by complex and difficult to understand visualizations.

This paper analyses, therefore, what are the current visualization tools explaining, first, which are the current techniques of visualization, focused mainly on static representations of the information. Next, we analyze the main problems and challenges in the field of data visualization, where we have examined the main tools of the market describing four basic points: data loading, visualization of this same data, advantages and disadvantages from the users' perspective.

Once the main problems with which data visualization have been studied, where information loss is a crucial point, the direction of the market and how the 3D tools can be used to obtain a new knowledge that the 2D visualization techniques are not able to display, is analyzed. The article finishes by commenting on the advantages and disadvantages of data representation in three dimensions.

CURRENT TECHNIQUES FOR VISUALIZING INFORMATION

Visual representation of the data is a crucial step in Big Data to be able to obtain all the benefits that this technology provides. Among the main benefits of visualization tools (Wang et al., 2015; Gonçalves et al., 2015; Branco et al., 2015; Leite et al., 2016), improved decision-making is the main one, but other important aspects to
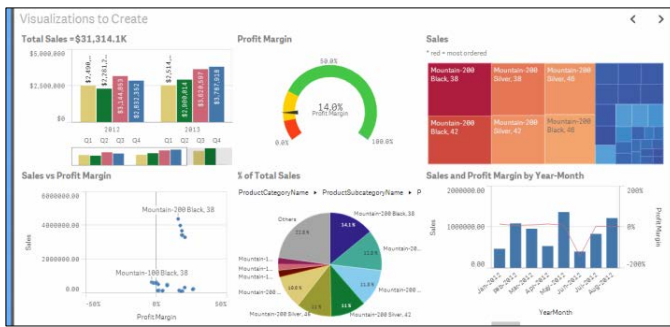
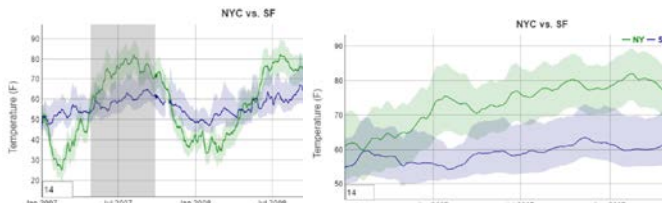**Figure 1.** Several visualization techniques (HayCanal.com, 2014)

Figure 2. Selection technique (Dygraphs, 2015)

consider are better ad-hoc analysis of data, improved collaboration between users and increased return on investment (ROI).

There are a number of common techniques for visually representing data, including tables, histograms, storyboards or panels as shown in **Figure 1**. This does not mean that at the time of making a visualization any type of representation available can be chosen, but, each method of visualization has a specific objective and therefore it is necessary to determine which type of information is wanted to display and then, choose the most appropriate graphic representation. For example, if you want to display numeric portions you can use pie charts, while if you want to show comparisons, bar charts are more useful, or if you want to display hierarchical relationships, treemap graphics would be very suitable for this type of data.

The most common used techniques are based on static methods of information representation, but also interactive tools are available that allow the user to modify the graphs by changing multiple variables. Within this type of visualization, there are a number of key functionalities to take into account (Khan and Khan, 2011):

Selection

Thanks to this functionality, the user will have the ability to select areas of the visualization, zoom them and know, in greater detail, the information located in the selected set.

This allows focusing on specific areas, to study them in greater detail and locate information that remains hidden in the overview. It will be important to take into account the degree of magnification available to avoid loss of quality in the visualizations, as well as loss of meaning due to a disorientation of the user.

Figure 2 shows two images, on the left you can see the selection of a certain portion of graph (grey rectangle) that will be expanded to show more detailed information as it can be shown on the right image.

Link and Brush

This functionality refers to the ability to link different views of the same data, thus providing multiple perspectives on the same piece of information. With this, the user will obtain different data about a certain collection and, in this way, can detect relationships between different markers.

Actually, this functionality consists of two steps as its name indicates. On one hand we have the technique of linking, that establishes a connection between different views in such a way that when performing actions, such as filtering (next point), will allow changing all the visualizations at the same time so that the user can determine links between indicators. On the other hand, we have the technique of brushing that consists of the selection of a set of data.

Figure 3 shows 4 different graphs linked so that crime information is displayed (Columbus, Ohio) and, when selecting an area of the first chart, the others are updated to display the information related to that selection.
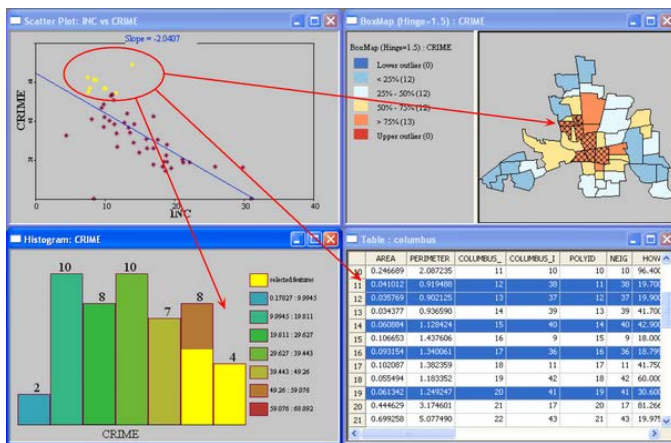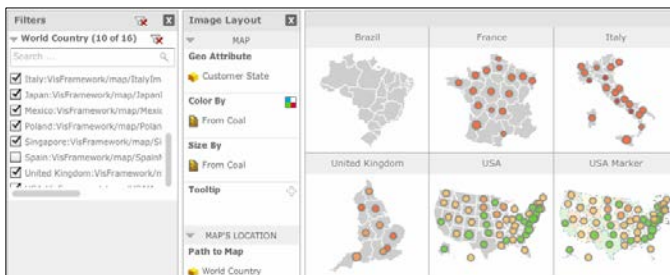
Figure 3. Link and Brush technique

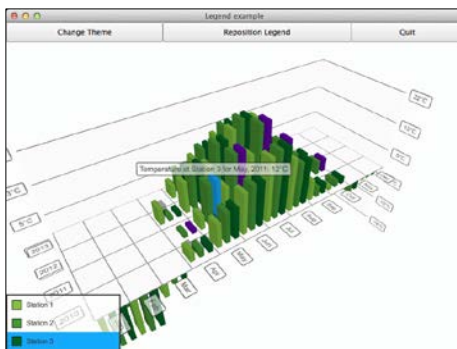**Figure 4.** Filter technique (MicroStrategy Web, 2015)

Figure 5. Reorganization technique: Space visualization (Qt, 2016)

Filter

Mentioned already in the previous point, this visualization technique allows the user to filter the elements of interest and, in this way, to reduce the amount of information shown to the necessary one in each moment.
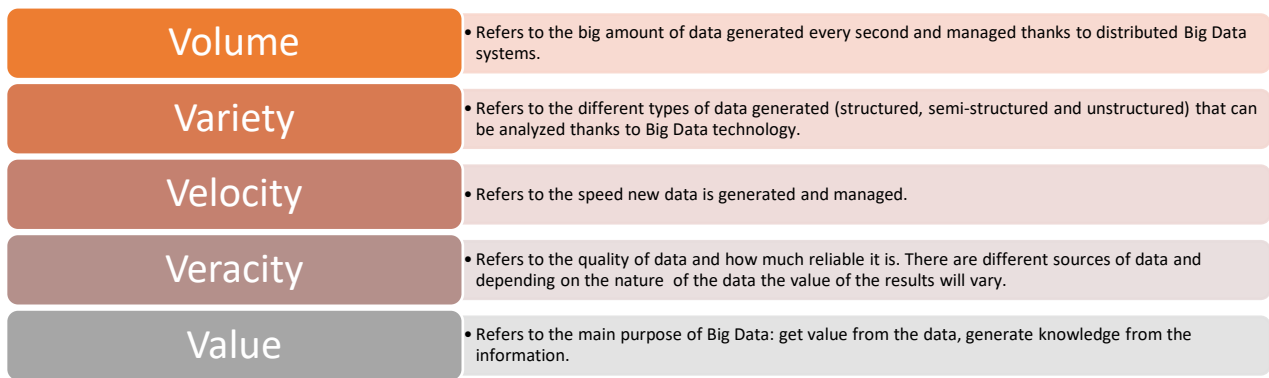
One of the myths of data visualization (Wang et al. 2015) is 'All data must be displayed', and some authors consider that this is not necessary since it only leads to overwhelm the user with a lot of information that is not able to process and that it is one of the main problems of data visualization nowadays.

Thus, as in the case of data analytics, the industry tends to develop software capable of displaying the most relevant information, discarding what is not of interest and thus freeing the user from all that information that is not necessary. However, all this filtering of information made by computer programs, is done from questions provided by users, so that if you do not ask the right questions, or do not know what questions to ask, you can lose very relevant information in the medium and/or long term.

Therefore, the ideal situation is to show all the data to the users and they, through all these functionalities described, determine through their experience, those that are important and which can be discarded.

Reorganization

Visualization dynamics are the main objective of the techniques discussed in here but, in many cases, it is not enough to provide the user with the option of zoom in and out, relating and filtering data, but it is necessary to

Volume	• Refers to the big amount of data generated every second and managed thanks to distributed Big Data systems.
Variety	• Refers to the different types of data generated (structured, semi-structured and unstructured) that can be analyzed thanks to Big Data technology.
Velocity	• Refers to the speed new data is generated and managed.
Veracity	• Refers to the quality of data and how much reliable it is. There are different sources of data and depending on the nature of the data the value of the results will vary.
Value	• Refers to the main purpose of Big Data: get value from the data, generate knowledge from the information.

Figure 6. Big Data: The 5 Vs

allow the reorganization of them so that each one better understands the information shown, especially when it comes to spatial visualizations.

As will be discussed later, spatial or 3D visualizations provide the user with a greater context of the information. Hence, the possibility of adding rotation to the list of more commented functionalities will allow the user to detect information that would otherwise remain hidden and to consider it in making decisions, thus avoiding the previously discussed problem of crucial information loss.

MAIN PROBLEMS AND CHALLENGES

The appearance of Big Data and the five Vs or five main attributes that define this new technology and which involve five new challenges in the field of information visualization will affect data visualization are shown in **Figure 6**.

Based on these properties, scalability and dynamism of graphic representations can be considered as two of the main challenges of data visualization. However, there are several functionalities that visualization tools are trying to provide nowadays to represent data, not only to show information but also to get value from it:

- Import data from different sources.
- Combine visual representation with text labels.
- Get information and knowledge form representations.
- Represent big volumes of data.
- Data mining integration.
- Collaborative work.
- Easy to use.
- Performance
- Limited cognition

There are new web-based visualization approaches that can help to reduce the generation costs and provide connections with different data sources allowing the representation to be up-to-date. These tools are very popular on business analytics environments but they are not used in scientific areas.

Analyzing the current market, it is possible to classify visualization tools in two main groups (Machlis, 2012; 2013; 2016):

Visualization tools for developers

These tools require technical skills to work with data and generate proper visualization. The most popular tools under this category because of the functionalities provided are:

- D3 Characteristics (**Table 1**) (D3, 2016)
- HighCharts Characteristics (**Table 2**) (HighCharts, 2016)
- Leaflet Characteristics (**Table 3**) (Leaflet, 2016)
- R Characteristics (**Table 4**) (R, 2016)

Table 1. D3 characteristics

Characteristic	Description
Data load	There are no limitations is this sense since it is necessary to develop a backend for each solution. This library allows connecting any type of data to a DOM (Document Object Model) and, once obtained, to apply different transformations to the document in function of the desired visualizations.
Data visualization	There are no limitations on types of visualizations since it is not a graphic library but a JavaScript library that uses web standards. It focuses on the manipulation of the DOM through visualization and interaction techniques in such a way that it has all the capabilities provided by browsers as well as the freedom to create the most appropriate visualizations.
User advantages	Perfect solution if someone is looking for high quality graphics and if the user knows exactly what he/she is looking for. Allows to be perfectly embedded in a portal/application. Although developers are needed for the use of this tool, this allows users to create almost any type of visualization and, therefore, get the most suitable for each situation, against other popular and easy to use tools where the user needs to limit the functionalities provided for each visualization.
User disadvantages	In order to create visualizations specialized developers are necessary, the development time is high and presents limitations in the performance as soon as the number of elements to visualize grows.

Table 2. HighCharts characteristics

Characteristic	Description
Data load	This tool loads the data into a JavaScript array that will be used later in the visualizations. A specialized developer must program this load. The data array can be found either in the local configuration, or in a separate file or even in a different path.
Data visualization	Large number of visualizations available both in desktop and mobile browsers.
User advantages	Simple and quick to use as long as what you want to create are not complex visualizations. It is a library that represents the different types of graphs as JSON objects so, changing the type of display, for example, from a bar chart to a circular one is as simple as changing a property in the JSON file from 'bar' to 'pie'. It allows to modify the visualizations dynamically and to zoom on the graphs.
User disadvantages	It does not allow the development of new graphics. Specialized developers are needed to create visualizations.

Table 3. Leaflet characteristics

Characteristic	Description
Data load	Since it is a JavaScript library, you can use data from multiple sources, but you need a specialized programmer to develop the code for loading them. The library uses GeoJSON files to load the data.
Data visualization	Focuses on interactive map display for mobile devices.
User advantages	Available for the most important browsers, both desktop and mobile. Light and without external dependencies. Developers do not need to have GIS knowledge.
User disadvantages	Required specialized developers. Only maps are displayed.

Table 4. R characteristics

Characteristic	Description
Data load	It allows importing data of files of multiple types like CSV, SAS, TXT, JSON and SPSS, or directly from tools such as Microsoft Excel, Microsoft Access, Oracle, MySQL and SQLite.
Data visualization	High quality graphics that can be exported to PDF, JPG, PNG and SVG format and even in tables for LATEX and HTML. Developers have specialized libraries for graphics.
User advantages	Focused on the statistical analysis, it has a wide range of statistical and graphical tools. Because it is a programming language, it allows users to develop their own functions apart from what it already provides by default as classification and grouping algorithms, statistical tests, time series analysis ... Since it is an open source (and open repository), all users can share packages that extend the functionality provided by R. Despite its complexity, there are many graphic interfaces, editors and IDEs that make the job easier. It works with all major operating systems on both 32-bit and 64-bit processors. It manages large volumes of data and the preparation of them (to be loaded) is easier than with other tools for developers thanks to the automation of many processes through scripts.
User disadvantages	Necessary specialized developers since it is a complex language and it has a strong learning curve.

Table 5. Tableau characteristics

Characteristic	Description
Data load	Supports most of the data sources.
Data visualization	Fast. Multiple visualizations are available for the user, not only atractive but also functional. Available for multiple platforms (desktop, mobile, browser) Most popular use is for internal dashboards which allow user to analyze data easily. However, it is important to indicate that it is not an analytic tool.
User advantages	Easy to use and manipulate to get more appropiate visualizations.
User disadvantages	If more complex visualizations are sought is necessary to know R (allows the connection with R to use a series of expressions). Certain types of visulizations, such as dashboards, are obtained through the Tableau Server and functions of formatting are only those provided by tableau. If the user needs to make certain visualizations not provided by Tableau, will not be possible to make them using this tool.

Visualization tools for general users

These tools are easy-to-use and it is not needed to have a technical background in order to create valuable visualizations.

- Tableau Characteristics (**Table 5**) (Tableau, 2016)
- Google Fusion Tables Characteristics (**Table 6**) (Google Fusion Tables, 2016)
- Quadrigram Characteristics (**Table 7**) (Quadrigram, 2016)
- Datawrapper Characteristics (**Table 8**) (Datawrapper, 2016)

Table 9 shows a comparison of the previous tools according to the areas analyzed and the main points to take into account when deciding the best visualization tool according to specific circumstances.

The market of visualization tools is very extensive and its main purpose nowadays is to get efficient tools with less visualization cost. However, there is one important factor to take into account in a world where predictive analytics is taking more and more importance: Information loss.

Visual Analytics Mantra says: "Analyze first, show the important, zoom, filter and analyze further, details on demand" (Keim et al., 2010). However, it also implies to trust on tools to analyze the data and, therefore, to determine which information is valuable and which does not have any importance. The result is knowledge loss that needs to be detected somehow.

Table 6. Google Fusion Tables characteristics

Characteristic	Description
Data load	To import data from a file, it is possible to use text files separed by a character, files encoded like UTF8 and the user can indicate a URL where to find data, among other options. It is a web service that, once loaded the data, automatically shows the user a map or a chart of it.
Data visualization	The user is limited to the types of available views and there is a visual programming language user must know before creating visualizations, but it is not complex. It is possible to add filters.
User advantages	This tool allows the use of fussion tables with other user, for example, to make data comparisons between same entities. Localizations tables can be converted in maps very quickly. Easy to use, it provides simple visualizations but the user has the option to use Google libraries or Javascript for more complex creations. Very powerful when managing the data, it is capable of detecting the same data even if they have different formats (Spain and SPAIN).
User disadvantages	The user may need to prepare the data before being imported to follow a format understandable by the tool (directions, percentages, formulae, ...)

Table 7. Quadrigram characteristics

Characteristic	Description
Data load	It uses Google Drive to store data and JSON files to load it.
Data visualization	There are more than 50 types of interactive visulizations, from bars to trees or animated maps.
User advantages	It is possible to create interactive visualizations without having programing skills. Its handling is very intuitive.
User disadvantages	The user options are restricted to the visualization types available in the tool, without the possibility of creating new classes of graphics. To create visualizations, it is necessary to know a specific programming language which is easy to learn.

Table 8. Datawrapper characteristics

Characteristic	Description
Data load	Data can be loaded through CSV files.
Data visualization	This tool provides a basic visualization which allows the user to get a result in a fast and easy manner.
User advantages	Fast and easy to use, it does not require a lot of user training.
User disadvantages	The tool provides a list of visulizations not very configurable and there is no option to create new types.

Table 9. Tools comparison

Analysis	Load data files	Visualization types	Development time	Easy to use	Configurable	Option to create	Easy to create
D3	Any	No limit	High	☒	☑	☑	☒
HighCharts	Any	Large	High	☒	☑	☑	☒
Leaflet	Any	Only maps	High	☒	☑	☒	-
R	Multiple	Any	High	☒	☑	☑	☒
Tableau	Multiple	Large	Low	☑	☑	☒	-
Google Fussion Tables	Multiple	Large	Med.	☑	☑	☒	-
Quadrigram	Json	Med.	Med.	☑	☑	☒	-
Datawrapper	CSV	Low	Low	☑	☒	☒	-

ADVANTAGES AND DISADVANTAGES OF 3D VISUALIZATION

When large sets of data have to be represented, it is common to use 2D graphic representations. However, the information to introduce to the user has a multidimensional sense not shown in these types of visualizations. The result of that is hidden information.

To improve the manner the data is introduced to the public without the risk of losing crucial knowledge, the tools must provide the ability of creating 3D visualizations.

The principal advantage of this type of representations is, therefore, the certainty that all the data is available because the information capacity of 3D graphs is enormous (Cellary and Walczak, 2012). Data visualization representations commonly use different techniques to indicate information, like colors, sizes, texts or shapes. Nevertheless, 3D representations extend them with new options like sizes orientations and rotation.

Thanks to 3D diagrams, the user is able to see all the data represented because of the ability to interact with the graph in two different ways: space and content. On one hand, the capacity to pivot the visualization allowing the user to analyze the information from different perspectives, permits the consumer to inspect all information, not only the first line. Hence, significant knowledge, undetectable in 2D graphs, is available for the user of this class of visualizations. On the other hand, the ability to work with the content can provide a better idea about how different variables can affect the global status. This advantage plus the three-dimensional perspective offers the user a more real information representation, as human eyes perceive everything in life. As a result, 3D graphs are more user-friendly providing the user a faster way to understand the information.

When you make a 3D representation or stereoscopic representation of an object, each user's eye perceives a different 2D perspective of the visualization due to the distance between both eyes. Subsequently, the 3D image is built on the brain, which receives information from two different points. For this reason, one 3D image contains more information that one in two dimensions and, in addition, it will allow a stereoscopic image to display areas otherwise hidden when using 2D representations.

When it comes to show the visualization, areas with more information must be within the screen, while those with fewer data can be located on the inside or outside of the screen. The location of the 3D object is very important since can generate discomfort to the user, as headaches, and it is important to avoid them. Therefore, if you want that the user understands easily the information displayed by an object in three dimensions, this must be placed inside the screen to generate less distress.

In addition to the location of the object, the distance where it is located is another important factor. The representation must be located at a distance equals to the screen to allow eyes to analyze the information in the same way they usually do. This distance allows that the focus and the rotation of the eyes, as well as the formation of the image in the retina, are done as in the real world.

There are autostereoscopic monitors allowing viewing images in 3D without using any additional devices such as 3D goggles or helmets. Those are monitors where the user is capable of appreciating the depth, thanks to each eye is analyzing images slightly different that allow the brain to calculate the third dimension.

However, this new visualization technique has also important disadvantages that must be improved in order to make them affordable. First, the complexity of this type of representations is extremely high which is why only very prepared professionals can develop them. In addition, it is necessary a lot of time to obtain them, not only in the first place, but also after user interaction. Therefore, scalability and dynamism, two of the main challenges of data visualization already mentioned in section II, are the principal points to improve to get efficient 3D data visualizations.

CONCLUSIONS

The main goal of this paper is to determine the future of visualization of massive data. For this, it is necessary to analyze the current state of the graphical representations and the characteristics of the main current tools. The common feature in all of them is the 2D representation of the information, trying at all times to show the most information without overwhelming the user. Very powerful tools are available not only to display the information, but also to make it interactive. Dynamism and scalability are the two factors to take into account because the growth of volume of data is higher and increases in real time.

However, it can be said that these tools are facing a paradox: on one hand, you have to represent enormous amounts of data, while on the other you have to visualize the information without confusing the user. As a result, high quality visualizations are got, from which knowledge is obtained, but they need to hide information.

In many cases the 2D visualizations are perfectly adequate and provide the advantages for which they were developed, however, the concealment of information can have important consequences in the medium and long term.

To avoid this problem, it is necessary to have visualizations that allow more data loading without risking user comfort when analyzing them. For this, 3D visualizations are available which, thanks to their rotation capacity, make the user perceive the information in the same way that He/She perceives the rest of the world. This not only allows to take into account all the data for the decision making but it does it in a friendly way for the user which makes the procedure more efficient. Therefore, the future of graphical representation of the data is directed towards the three-dimensional graphs although it is an arduous path where you have to deal with the two key points right now that are the dynamicity and the scalability of the representations.

REFERENCES

Branco, F., Gonçalves, R., Martins, J. and Cota, M. (2015). Decision Support System for the Agri-food Sector – The Sousacamp Group Case. New Contributions in Information Systems and Technologies. *Advances in Intelligent Systems and Computing*, 353. Springer.

Cellary, W. and Walczak, K. (2012). *3D rotation visualization advantages.* Springer-Verlag London Limited.

D3 (2016). [online] Available at: https://d3js.org/ [Accessed 1 Jul. 2017].

Datawrapper (2016). [online] Available at: http://docs.datawrapper.de/ [Accessed 1 Jul. 2017].

Dygraphs (2015). [online] Available at: http://dygraphs.com/ [Accessed 1 Jul. 2017].

Gonçalves, R., Casto, M., Cota, M., Branco, F., Martins, J. And Barroso, J. (2015). A new concept of 3D DCS interface application for industrial production console operators. *Universal Access in the Information Society*, 14(3).

Google Fusion Tables (2016). [online] Available at: https://developers.google.com/fusiontables/docs/v2/getting_started [Accessed 1 Jul. 2017].

HayCanal.com (2014). *Seis consejos para optimizar la visualización de datos en la empresa.* [online] Available at: http://www.haycanal.com/noticias/6868/seis-consejos-para-optimizar-la-visualizacin-de-datos-en-las-empresa [Accessed 1 Jul. 2017].

HighCharts (2016). [online] Available at: http://www.highcharts.com/docs [Accessed 1 Jul. 2017].

Keim, D.A., Mansmann, F. and Thomas, J. (2010) Visual analytics: How much visualization and how much analytics. *ACM SIGKDD Explorations Newsletter*, 11, pp. 5-8, December 2010. [online] Available at: https://bib.dbvis.de/uploadedFiles/39.pdf

Khan, M. and Khan, S.S. (2011). Data and information visualization methods and interactive mechanisms: A survey. *International Journal of Computer Applications.*

Leaflet (2016). [online] Available at: http://leafletjs.com/index.html [Accessed 1 Jul. 2017].

Leite, P., Gonçalves, J., Teixeira, P. and Rocha, Á. (2016). A model for the evaluation of data quality in health unit websites. *Health informatics journal*, 22(3), pp. 479-495.

Machlis, S. (2012). 8 cool tools for data analysis, visualization and presentation. *Computerworld*, March 2012.

Machlis, S. (2013). Six useful JavaScript libraries for dealing with data. *Computerworld*, March 2013.

Machlis, S. (2016). Chart and image gallery: 30+ free tools for data visualization and analysis, *Computerworld*, August 2016.

MicroStrategy Web (2015). *Creating an image layout visualization*, [online] Available at: http://www2.microstrategy.com/producthelp/9.5/WebUser/WebHelp/Lang_1033/Creating_an_image_layout_visualization.htm [Accessed 1 Jul. 2017].

Qt (2016). *Qt Quick 2 legend example.* [online] Available at: http://doc.qt.io/qt-5/qtdatavisualization-qmllegend-example.html [Accessed 1 Jul. 2017].

Quadrigram (2016). [online] Available at: http://www.quadrigram.com/tutorials/ [Accessed 1 Jul. 2017].

R (2016). [online] Available at: https://www.r-project.org/about.html [Accessed 1 Jul. 2017].

Smith, M.J.d., Goodchild, M.F. and Longley, P.A. (2015). *Geospatial analysis.* [online] Available at: http://www.spatialanalysisonline.com/HTML/index.html?eda__esda_and_estda.htm

Tableau (2016). [online] Available at: http://www.tableau.com/support/help [Accessed 1 Jul. 2017].

Wang, L., Wang, G., Alexander, C.A. (2015). Big data and visualization: Methods, challenges and technology progress. *Digital Technologies*, 1(1), pp. 33-38. [online] Available at: http://pubs.sciepub.com/dt/1/1/7/