❒   1340

Metrics and Benchmarks for Empirical and Comprehension Focused Visualization Research in the Sales Domain

Loo Yew Jie, Doris Hooi-Ten Wong, Zarina Mat Zain, Nilam Nur Amir Sjarif, Roslina Ibrahim, Nurazean Maarop
Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

Article Info	ABSTRACT
	Data visualization is an effort which aims to communicate data effectively and clearly to the audience through graphical representation. Data visualization efforts must be coordinated with an understanding into the Cognitive Learning Theory (CLT). In the sales domain, sales data visualization are made possible with the available Business Intelligence (BI) tools such as Microsoft Power BI, Tableau, Plotly, and others. These tools allow seamless interaction for the top management as well as the sales force with regard to the data. Sales data visualization comes with an array of advantages such as self-service analysis by business users, rapidly adapt to changing business conditions, and enable continuous on-demand reporting among others. The advantages of sales data visualization also comes with the challenges such as difficulty in identifying visual noise, high rate of image change, and high performance requirements. In an effort to reduce cognitive activity that does not enhance learning, sales visualization dashboard must be designed in a way that is neither too simplistic nor too complex to ensure that the Intrinsic Cognitive Load (ICL), Extrinsic Cognitive Load (ECL), and Germane Cognitive Load (GCL) are in sync with the audience. With the combination of sales data visualization and CLT, understanding complex sales details quickly is made possible by not only the top management of the organization, but also the sales force of the organization.

Corresponding Author:

Doris Hooi-Ten Wong,
Universiti Teknologi Malaysia,
54100 Kuala Lumpur, Malaysia.
Email: doriswong@utm.my

1.   INTRODUCTION

The purpose of data visualization is to project the data clearly and effectually to the spectators by using graphical illustration. It is a crucial part of the process to uncovering the key points within the process. With multiple source of data available, visualization is important and is being fully utilized by many organization worldwide in making day to day decision until it is regarded as an vital process in Business Intelligence. Influx of data occurs commonly in today's data driven ecosystem and the challenge is to present a metric and benchmarks for empirical and comprehension focused visualization. In reality, three other important topics as suggested by Singh and Wajgi [1] that the decision makers will faces such as:
1) The procedure of visualization can be flexible and versatile.
2) Supporting evidences are transparent to acquire; and
3) The speed of computing and the cost of processing.

This paper presents a research on how what are the appropriate metric and benchmarks in producing effective visualization in the sales domain.

Traditionally, visualization has been the domain of statistics. A standard textbook in statistics [2] has a chapter on creating bar charts, pie charts, line charts, histograms, etc. These are simple representations of

data that require standard input. However, with the proliferation of types and variety of data, there is a need for more types of analyses and presentations that a) bring out the relationships between different elements b) summarize complex data with simple and easily understood visuals c) simplify the visualization without the loss of the many dimensions of the data and d) at the same time, achieve all this quickly with easy to use analytical tools. Visualization is particularly important for hierarchical data, where the individual data points are connected in a tree-like structure, with large clusters of data broken into sub-categories. The hierarchical analyses of data suggested here can help people to see relationships between variables and groups, while making it easy to check on data veracity. The visualization helps to understand the break-up of sales data into categories, subcategories etc.

Benchmarking enables companies to see their positions relative to their competitors in order to explore the opportunities to improve their market position. This is taken by its definitions: "Benchmarking is the process of continuously measuring and comparing one's business processes against comparable processes in leading organizations to obtain information that will help the organization identify and implement improvements" [3]. While metrics can be defined as "Standards if measurement by which efficiency, performance, progress, or quality of a plan, process or product can be assessed [4].

2.   RESEARCH METHOD
Meloncon and Warner [5] reviewed the major categories found in data visualization includes comparison of types of visualizations, graphs, icons, other and online.

2.1. Comparison of Types of Visualization
1) Animations and static visualizations - Animations did not greatly promote positive learning outcomes, and even resulted in performance degradations.
2) Text, tables, and bar graphs - Graphs are great ways to express risk communication practice due to their ability to capture attention and elicit information extraction with minimal cognitive effort, and will improve comprehension.
3) Tables adnd bar graph - When data is presented in these formats, audience with experience and knowledge with bar graphs preferred bar graphs, while those with experience and tables found graphs equally easy to use. When examining tests with borderline result, bar graphs is still the preferred medium of visualization.
4) Numbers and icons - Graphics and icons were the only discrepancy between impacted comprehension and recall; but not impacted by the actual level of iconicity of graphic.

2.2. Graphs
Generally, graphs are excellent when it comes to data visualization, although there exist a debate between using graphs and lines. However, it is subjected to the audiences literacy background. It is also discovered that graph conventions (titles, legends, orientation and colors) and literacy rates are important and should be taken into account.

2.3. Icons
Icons are an effective method to show information since they boosted recall of information and effective in improving understanding.

2.4. Others
Other types of visualization includes pie chart, maps and photographs. In brief, pie charts were preferred when displaying genomic risk information due to the similarity to common object and the seeming simplicity of basic percentages and allowed simpler visualization. Studies showed that domain knowledge can influence information selection and understanding of complex graphics, and they offer empirical support for the data visualization concept that the display should avoid include any more information that is required.

2.5. Online
The three most widely discussed ones includes personal health record, patient information website and electronic health record. While they are subject to their respective interface designs, the major concern of these visualizations is that the graphical information were too complex and included excessive information to absorb and understand.

2.6. Visual Analytics (VA)

Figure 1 Visual Analytics (VA) could be a teach appearing critical guarantee in making a difference clients pick up knowledge into complex information. VA devices use human perceptual and subjective capacities by utilizing intelligent representations as interfaces amongst clients and their information, in this way making information related undertakings more compelling and effective.
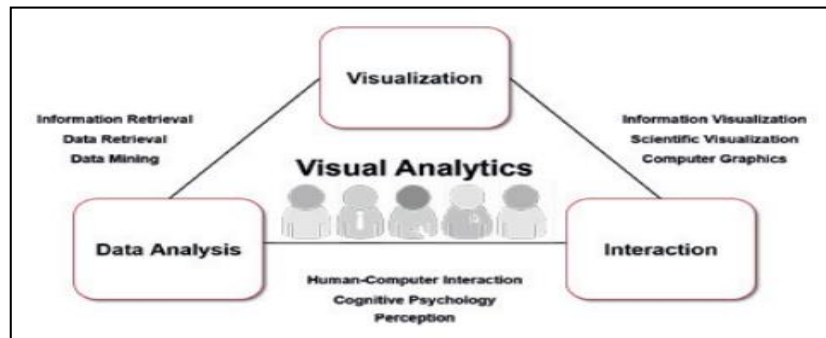

Figure 1. Overview of visual analytics

Visual Analytic expects to decrease complex intellectual work to process huge amount of data sets towards an answerable information [6].

The information from company's operation with customers' interaction are very rich. There are some structured data where can be stored, retrieve and analyze in spreadsheets or in relational database. There are also semi-structured data like email data or website traffic date where need extra effort to process and analyze then summarize it in significant ways. For unstructured data where it is known as a very wealth of data which are related to company; customers, reviews, testimonials, and social media. It is important for the company to handle the data, storing, retrieving and managing all different type of data because it is help the company to prioritize the performance measures based on these data.

To perform the benchmark and measurements for information representation, organization utilizes excel spreadsheet and tableau tool income information:

1) Develop a period arrangement plot of number of requests put for consistently in the informational index.
2) Visualize the total number of requests put for every day of the month.
3) Show a guide representation with every one of the states in the US and qualities for the quantity of requests put in each state and the average income per arrange in that state.
4) Graph the quantity of site visits every day for all dates in the dataset.
5) Graph the quantity of site hits for all dates in the dataset.
6) Design a dashboard that all the while shows the guide perception and the diagram with the quantity of site hits for all dates in the dataset.
7) Design Strategy map and balanced scorecard. A strategy map is a supportive representation apparatus worked around the balanced scorecard ideas that outlines circumstances and end results connections between key activities displayed close by benchmark and measurements markers. Regularly, a technique outline four particular territories for measurements and benchmark assessment.
8) Financial point of view – demonstrates approaches to accomplish economic development to fulfill investors (slack markers).
9) Customer point of view – portrays accomplishment with clients and characterizes client sections (a blend of slack and lead markers).
10) Internal process point of view – exhibits how esteem is conveyed to clients (lead markers).
11) Learning and development point of view – centers around individuals, innovation, and hierarchical atmosphere (lead markers).
12) Properties of metrics. To outline KPIs, it is useful to remember that all together for a metric to be fruitful, it ought to be Simple to comprehend and benchmark against; Map to key business exercises, activities to comes about; Actionable – center consideration and guide right conduct; Reliable and substantial; and Timely (SMART).
13) Dashboards - Orgaizaition regularly utilize electronic dashboards to see KPIs. A dashboard viably portrays markers utilizing designs which makes it considerably less demanding to recount a story and

convey it all through the organization. It can likewise be furnished with notice signs or alarms conveyed when a metric is outside of preset parameters.

Sales forecasting is among the fundamental inputs for planning decisions throughout the supply chain. Estimating future demand more accurately is critical for meeting it, while minimizing inventory and other related costs. These demand estimates are often modelled based on historical patterns in the data [7].

3. FINDINGS

According to Havemo [8], from a reporting perspective, visual means such as graphs and other visualisation is essential in increasing business models presentation. According to the 2015 Gleansight Benchmark Report [9] on data visualization, there are a number of reasons why data visualization will be implemented.

3.1. Empower non-IT Professionals

If available tools are too complex, it's very common for organizations to depend heavily on IT for running queries, customizing reports, and conducting analysis. All these things create bottlenecks for users. More and more companies are looking to increase adoption of self-service BI to support their goal in empowering non-IT professionals.

3.2. Rapidly Adapt to Changing Business Conditions

Data visualization is ideal for articulating qualitative changes in business data sets such as an acquisition, merger, new business unit, or change in the data hierarchy. Data visualization may provide a great way to understand variances in the numbers with greater ease.

3.3. Encourage Data Exploration

Giving users visually stimulating, and simple interfaces minimizes the skills required to conduct analysis. Top Performers recognize that the best thing they can do for the business is give users with context about how to interpret data trends easy access to the data [9].

Ali et.al. [10] emphasized some of the big data visualization problem, which include:
1) Visual noise: High relativity between each objects in the dataset, resulting high difficulty to separate them.
2) Information loss: Some information are sacrificed in the effort to improve dataset visibility and increase response time.
3) Personal perception and interpretation of the visualisation.
4) Highly dynamic data requires constant visualisation updating increases difficulty for user to react to the figures shown.
5) High performance requirements: Dynamic visualization demands for more requirements compared to static visualization.

Many tools have been invented to help us out from the above problem. The most crucial feature that a visualization must have is interactivity. In the business world, many organization have opted for visualization tools to make interesting dashboard and attractive presentations. Among the most popular visualization tools are summarized in Table 1. Ali et. al. [10] compared these tools on the basis of various attributes. Some of the considerations when choosing the right visualization tools are listed below:
1) Tool is open source or not.
2) Visualisation created allows user to interact with them.
3) Suitable client type or packages to create the visualisation.
4) Readiness to integrate with data sources such as Hadoop Hive, Google Analytics, etc.
5) Availability of tutorials through Massive Open Online Courses (MOOCs).
6) Accessibility and availability of Application Programming Interface (API).

Table 1. Comparison of software attributes used in data visualization sales domain

	Tableau	Power BI	Plotly	Gephi	Excel 2016
Open Source	N	N	Y	Y	N
Interactive	Y	Y	N	N	Y
Desktop Client	Y	Y	N	Y	Y
Online Client	Y	Y	Y	N	Y
Mobile App.	Y	Y	N	N	Y
Integration	Y	Y	N	N	Y
MOOCs	Y	Y	Y	Y	Y
API	Y	Y	Y	Y	Y

Although the aforementioned tools offers powerful features and often used by businesses, however they also come with limits/demerits as highlighted by the authors.

1) Tableau: Tableau Public only comes with a 1GB storage and for larger work requirements, license of the server and Tableau Desktop will be required.
2) Microsoft Power BI: It comes with a free version but users must have a Work account and it is limited to 250 MB of storage for workbook. It is also slower if compared to Tableau.
3) Plotly: Pro users have limited to only 500 KB for upload size. Even if professional version, you will get unlimited charts but upload size of files will be limited to only 5 MB. Programming skills are required and no official offline client for Plotly is available.
4) Gephi: Only specializes in graph visualization, cannot be applied for other types of visualizations.
5) Excel 2016: Microsoft Office is a paid application and the only the Office 365 subscribers will gain access to the API.

Referring to Magee et.al. [11] proper data visualization increases the ability of the salesperson to interpret the data visualization presented. The paper also states that the human brain is hard-wired to narrative and visual patterns and not mathematical ones. Proper identification of salesperson interest or focus is also important in presenting the sales data to the salesperson. Besides that, data visualization in this context is also an organizational change agent. Salesperson were able to identify their key focus in order to identify the right leads to bring in the sales as discussed in the same paper.

4.   PROPOSED SOLUTION

It is undeniable that organizations normally have a collection of database, where each database storing different piece of information. However, visualizing these huge chuck of information is usually challenging and might lead to confusion if not presented appropriately. Generally, numbers and figures by themselves do not carry much meaning unless represented using the right visual. The question here is, when it comes to visualization, especially in the sales domain, what are the metric and benchmark one should follow to make reporting work effective and easily understood when presented to the management or stakeholders? This section discusses some of the proposed metric and benchmark for empirical and comprehension focused visualization in the sales domain.

This paper adopts the foundation for the design of instruction and assessment as proposed by Leppink [12] , which aims to keep cognitive activity to its minimal since it will jeopardize learning.This framework revolves around the Cognitive Load Theory as the development and automation of cognitive schemas regarding content to be delivered and learnt by the audience. The three types of cognitive load are: Intrinsic Cognitive Load (ICL), Extraneous Cognitive Load (ECL) and Germane Cognitive Load (GCL).

When preparing a presentation deck to report number and figures, it is important to ensure it is designed in such a way that only a minimum of working memory power is required for cognitive processes that do not contribute to learning as much. Balance is the key in this situation where the presentation deck should consist of elements that are clear and easily understood. Moreover, in reporting numbers and figures, it is not a good practice to merely learn the steps of a procedure. Rather, they have to be undertaken in a particular sequence to ensure a correct solution for a given situation. The sequence matters and that interactivity adds to ICL. Take the case of a business analyst, in such a situation, having to address a root-cause analysis on a drop in sales, where there are many possible diagnoses may take the ICL for less experienced analyst to the limits of their working memory. This will leads to creating a visualization dashboard that is over simplistic and fail to deliver up to the benchmark. On the contrary, a more advanced analyst will eventually experience a lower ICL in such a situation because they can activate more developed and perhaps already more automated cognitive schemas than their less experienced colleagues. This will leads to creating a visualization dashboard that is over complex and hard to be understood if the audience does not have the same level of ICL. Hence, careful reflection on this ICL factor is of paramount importance.

Hernando et. al. [13] concluded that it is not appropriate to show all the dependencies and interrelationships that exist in big data domains, because there would be an excess of information that would make it impossible to detect the relevant results. In general, an organization can consider the following pipeline proposed by Singh and Wajgi [1] as depicted in Figure 2.
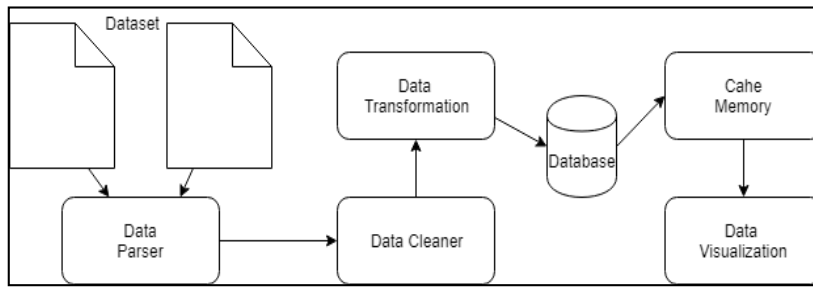
Figure 2. Sales data visualization pipeline

4.1. Data Parser

Depending on situations, user may find data set with multiple entries to be relevant or irrelevant. Therefore, parsing will be performed in java using java.util.Iterator class to examine the features that exist in the data set.

4.2. Data Cleaner

It is necessary to be removed and cleaned from the dataset to keep them relevant to the situation and reduce unnecessary computation resources.

4.3. Data Transfer HSSF

Workbooks were chosen for storing the FileInputStream provided by the user for change the feature name exist in the data set. The names of the feature may need further effort to rectify so there are in proper format may not be in proper format. For instance, Purchase Id will be expressed as PuID which may cause confusion.

4.4. Database

Once the data are properly processed, it will be imported into the database which contain appropriate data relevant to the user in the proper format.

4.5. Cache

Cache is often used to frequently used data that is extracted from the database to reduce time and effort to repeatedly performing the same extraction.

4.6. Visualization

Time duration provided by the end user is usually specified when it comes to data visualization. High value customers, regional sales and top products can be visualized. By using this practice, end user will then carried out their respective decision making process However, when designing the visualization dashboard, it is important to remember not to incorporate unnecessary ECL and audiences' level of knowledge should be taken into account.

5. APPLICATION

In this section how user can apply this theory in sales data visualization is discussed. The appropriate graphs/charts must be applied in the suitable context to increase ICL and eventually boost GCL. Abela [14] summarized a chart suggestions that is compact for visualization use. In general, there are four categories of chart, which includes: comparison, relationship, distribution an composition.

Comparison graphs can be further subdivided into two smaller group, either they are comparing among items or over time. For comparison among items, user can consider variable width column chart, table with embedded charts, bar chart or column chart. For comparisons over time, user can consider circular area chart, line chart, column chart or line chart. Relationship between two variables can be expressed in scatter plot. Meanwhile, relationship with three variables can be expressed in bubble chart. Distribution with single variable can be expressed in column histogram for few data points and line histogram for many data points. Distribution with two variables can be expressed in scatter plot. Finally, distribution with three variables can be expressed in 3D Area Chart. Composition which are changing over time can be expressed using stacked 100% column chart, stacked column chart, stacked 100% area chart or stacked area chart. While static

composition can be expressed using pie chart, waterfall chart or stacked 100% column chart with subcomponents.

The appropriate chart must be applied to the right context to improve GCL and avoid creating unnecessary ECL. Figure 3 summarizes the chart suggestion and their criteria.

In another context, visualizing sales related geographic or demographic data in maps does not necessary leads to better ICL since not everyone has the same level of geographical knowledge. In Figure 4, the average birth rate for countries in the region of Asia and The Americans are compared using maps.
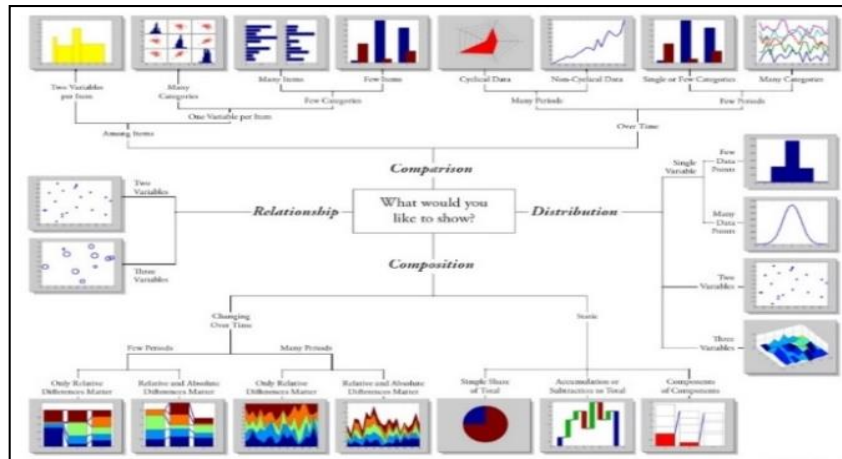

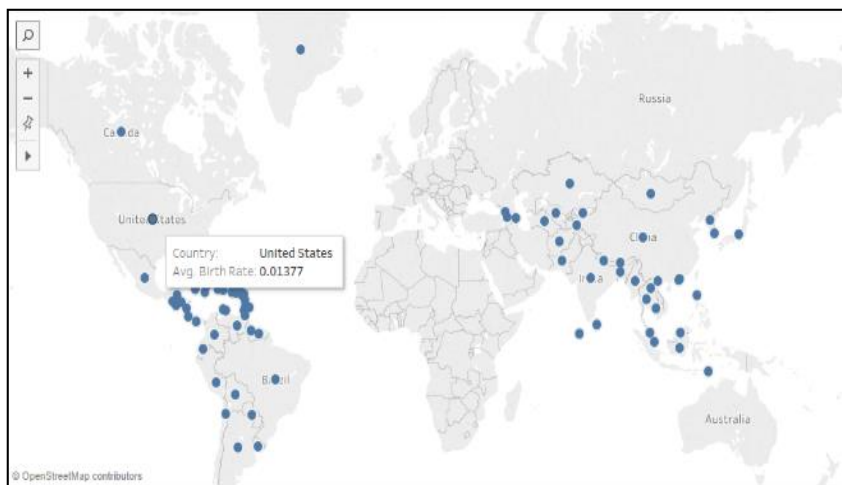Figure 3. Chart Suggestion

Figure 4. Visualizing birth rate by region using maps

Although aesthetic and appealing, however, this will create ECL if the audience geographical knowledge is limited. It is better to represent the comparison of average birth rate between two regions using a simple bar graph since it is clear and easily understood. In Figure 5, one can easily conclude that Asian countries have a better average birth rate than The Americans countries. Audience without much geographical knowledge can easily visualize the number and figures, hence leading to better ICL and GCL and avoiding ECL.
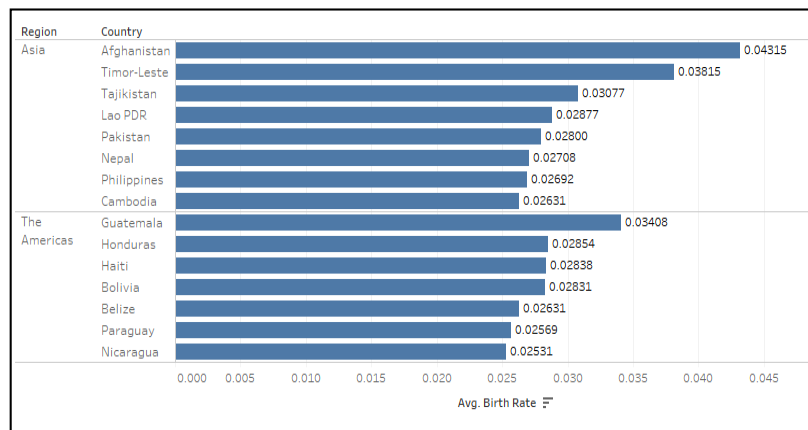
Figure 5. Visualizing birth rate by region using bar charts

6.   CONCLUSIONS AND FUTURE WORKS

As the need for quick decision-making keeps rising in marketing, particularly with the advent of the Internet, rapid understanding through visual representation of the effect of marketing variables on strategy will help in improving profitability. Spreadsheets and other non-visual data are very important and cannot be done away with. It is best to provide a marketing analyst both visual and non-visual data so that sound marketing decisions can be made. Some managers are best at understanding numbers and others are mostly visual; hence both must be provided to managers for making sound decisions. Heer and Shneiderman [15] state that multiple, linked visualizations are important for providing meaningful insights into multidimensional data rather than isolated visualization of the same data since the quantity of data that can be presented in a single image is limited and inter-relationships between variables and data sets cannot be entirely presented with a simple image. Effective data visualisation and understanding the audience of the data visualisation is crucial in the sales environment as it allows for sales personnel to understand the internalize the visualisation that suits the sales personnel style and it also allows the operational personnel to understand the internalize the visualisation that suits to their style.

Visualization unearths topics that are hidden due to the complexity of the issue, driving simplification of the topic, creating urgency and an effective sense of the opportunity cost to not take corrective action [11].Lastly, the implementation process of a data-driven project in a sales environment must ensure that effective data visualisation is in place to ensure the audience are fully engaged. The three common issues that must always be taken into consideration are: (i) The TIME taken or data gathering, (ii) The individual's ABILITY in any individual to synthesize, analyze understand the data visualisation and (iii) The ability to COMMUNICATE the acquired insights to others within their team down the line.

REFERENCES

[1]   Singh K, Wajgi R. Data Analysis and Visualization of Sales Data. World Conference on Futuristic Trends in Research and Innovation for Social Welfare, 2016.
[2]   Berenson M, Levine D, Szabat. Data Visualization in Marketing. Journal of Marketing Management. 2015; 3 (2): 36-49.
[3]   Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A Method for Continuous Quality Improvement in Health. Healthc Policy. 2012; 7 (4): 101-109.
[4]   Business   Dictionary,   "www.businessdictionary.com,"   [Online].   Available: http://www.businessdictionary.com/definition/metrics.html. [Accessed 6 May 2018].
[5]   Meloncon L, Warner E. Data Visualizations: A Literature Review and Opportunities for Technical and Professional Communication. Professional Communication Conference (ProComm), 2017.
[6]   Kaluza A, Gellrich S, Cerdas F, Thiede S, Herrmann C. Life Cycle Engineering Based on Visual Analytics. Procedia CIRP. 2018; 69: 37-42.
[7]   Sagaert YR, Aghezzaf EH, Kourentzes N, Desmet B. Tactical sales forecasting using a very large set of macroeconomic indicators. European Journal of Operational Research. 2018; 264 (2): 558-569.

[8]   EmelieHavemo. A visual perspective on value creation: Exploring patterns in business model diagrams. European Management Journal. 2017.
[9]   Gleanster. Gleansight Benchmark Report - Data Visualization. 2015.
[10]   Ali SM, Gupta N, Nayak GK, Lenka RK. Big data visualization: Tools and challenges. Contemporary Computing and Informatics (IC3I). 2016: 656-660.
[11]   Magee B, Sammon D, Nagle T, O'Raghallaigh P. Introducing data driven practices into sales environments: examining the impact of data visualisation on user engagement and sales results. Journal of Decision Systems. 2016; 25 (1): 313-328.
[12]   Leppink J. Cognitive load theory: Practical implications and an important challenge. Journal of Taibah University Medical Sciences. 2017. 12 (5): 385-391.
[13]   Hernando A, Bobadilla J, Ortega F, Gutiérrez A. Method to interactively visualize and navigate related information," Expert Systems with Applications, 2018.
[14]   Abela. The Extreme Presentation(tm) Method. September 2006. [Online]. Available: http://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf. [Accessed 6 May 2018].
[15]   Heer J, Shneiderman B. Interactive Dynamics for Visual Analysis: A taxonomy of tools that support the fluent and flexible use of visualizations. ACM Queue. 2016; 10 (2): 1-26.
[16]   Jovanovic J, Bagheri E, Gasevic G. Comprehension and Learning of Social Goals through Visualization. IEEE Transactions on Human-Machine Systems. 2015: 478-489.