Chapter 4

VISUALIZING BIG DATA

Ekaterina Olshannikova, Aleksandr Ometov, Yevgeni Koucheryavy and Thomas Olsson, Springer, Vol. 2, No. 22, October 2015.

Abstract

This chapter provides a multi-disciplinary overview of the research issues and achievements in the field of Big Data and its visualization techniques and tools. The main aim is to summarize challenges in visualization methods for existing Big Data, as well as to offer novel solutions for issues related to the current state of Big Data Visualization. This paper provides a classification of existing data types, analytical methods, visualization techniques and tools, with a particular emphasis placed on surveying the evolution of visualization methodology over the past years. Based on the results, we reveal disadvantages of existing visualization methods. Despite the technological development of the modern world, human involvement (interaction), judgment and logical thinking are necessary while working with Big Data. Therefore, the role of human perceptional limitations involving large amounts of information is evaluated. Based on the results, a non-traditional approach is proposed: we discuss how the capabilities of Augmented Reality and Virtual Reality could be applied to the field of Big Data Visualization. We discuss the promising utility of Mixed Reality technology integration with applications in Big Data Visualization. Placing the most essential data in the central area of the human visual field in Mixed Reality would allow one to obtain the presented information in a short period of time without significant data losses due to human perceptual issues. Furthermore, we discuss the impacts of new technologies, such as Virtual Reality displays and Augmented Reality helmets on the Big Data visualization as well as to the classification of the main challenges of integrating the technology.

Introduction

The whole history of humanity is an enormous accumulation of data. Information has been stored for thousands of years. Data has become an integral part of history, politics, science, economics and business structures, and now even social lives. This trend is clearly visible in social networks such as Facebook, Twitter and Instagram where users produce an enormous stream of different types of information daily (music, pictures, text, etc.) [1]. Now, government, scientific and technical laboratory data as well as space research information are available not only for review, but also for public use. For instance, there is the 1000 Genomes Project [2, 3], which provide 260 terabytes of human genome data. More than 20 terabytes of data are publicly available at Internet Archive [4, 5], ClueWeb09 [6], among others.

Lately, Big Data processing has become more affordable for companies from resource and cost points of view. Simply put, revenues generated from it are higher than the costs, so Big Data processing is becoming more and more widely used in industry and business [7]. According to International Data

Corporation (IDC), data trading is forming a separate market [8]. Indeed, 70 % of large organizations already purchase external data, and it is expected to reach 100 % by the beginning of 2019.

Simultaneously, Big Data characteristics such as volume, velocity, variety [9], value and veracity [10] require quick decisions in implementation, as the information may become less up to date and can lose value fast. According to IDC [11], data volumes have grown exponentially, and by 2020 the number of digital bits will be comparable to the number of stars in the universe. As the size of bits geminates every two years, for the period from 2013 to 2020 worldwide data will increase from 4.4 to 44 zettabytes. Such fast data expansion may result in challenges related to human ability to manage the data, extract information and gain knowledge from it.

The complexity of Big Data analysis presents an undeniable challenge: visualization techniques and methods need to be improved. Many companies and open-source projects see the future of Big Data Analytics via Visualization, and are establishing new interactive platforms and supporting research in this area. Husain et al. [12] in their paper provide a wide list of contemporary and recently developed visualization platforms. There are commercial Big Data platforms such as International Business Machines (IBM) Software [13], Microsoft [14], Amazon [15] and Google [16]. There exists an open-source project, Socrata [17], which deals with dynamic data from public, government and private organizations. Another platform is a JavaScript library D3 [18] for dynamic data visualizations. This list can be extended with Cytoscape [19], Tableau [20], Data Wrangler [21] and others. Intel [22] and Statistical Analysis System (SAS) [23] are performing research in data visualization as well but more from a business perspective.

Organizations and social media generate enormous amounts of data every day and, traditionally, represent it in a format consistent with the poorly structured databases: web blogs, text documents, or machine code, such as geospatial data that may be collected in various stores even outside of a company/organization [24]. On the other hand, information stored in a multitude repository and the use of cloud storage or data centers is also widely common [25]. Furthermore, companies have the necessary tools to establish the relationship between data segments in addition to the process of making the basis for meaningful conclusions. As data processing rates are growing continuously, a situation may appear when traditional analytical methods would not be able to stay up to date, especially with the growing amount of constantly updated data, which ultimately opens the way for Big Data technologies [26].

This paper provides information about various types of existing data to which certain techniques are useful for the analysis. Recently, many visualization methods have been developed for a quick representation of data that is already preprocessed. There has been a step away from planar images towards multi-dimensional volumetric visualizations. However, Big Data visualization evolution cannot be considered as finished, inasmuch as new techniques generate new research challenges and solutions that will be discussed in the following paper.

Current activity in the field of Big Data visualization is focused on the invention of tools that allow a person to produce quick and effective results working with large amounts of data. Moreover, it would be possible to assess the analysis of the visualized information from all the angles in novel, scalable ways. Based on Big Data related literature, we identify the main visualization challenges and propose a novel technical approach to visualize Big Data based on the understandings of human perception and new Mixed Reality (MR) technologies. From our perspective, one of the more promising methods for improving upon current Big Data visualization techniques is in its correlation with Augmented Reality (AR) and Virtual Reality (VR) that are suitable for the limited perception capabilities of humans. We identify important steps for the research agenda to implement this approach.

This paper covers various issues and topics, but there are three main directions of this survey:

- Human cognitive limitations in terms of Big Data Visualization.
- Applying Augmented and Virtual reality opportunities towards Big Data Visualization.
- Challenges and benefits of the proposed visualization approach.

The rest of paper is organized as follows: The first section provides a definition of Big Data and looks at currently used methods for Big Data processing and their specifications. Also it indicates the main challenges and issues in Big Data analysis. Next, in the section Visualization methods, the historical background of this field is given, modern visualization techniques for massive amounts of information are presented and the evolution of visualization methods is discussed. Further in the last section, Integration with Augmented and Virtual Reality, the history of AR and VR is detailed with respect to its influence on Big Data. These developmental processes are supported by the proposed oncoming Big Data visualization extension for VR and AR, which can solve actual perception and cognition challenges. Finally, important data visualization challenges and future research agenda are discussed.

Big Data: an overview

Today large data sources are ubiquitous throughout the world. Data used for processing may be obtained from measuring devices, radio frequency identifiers, social network message flows, meteorological data, remote sensing, location data streams of mobile subscribers and devices, and audio and video recordings. So, as Big Data is more and more used all over the world, a new and important research field is being established. The mass distribution of the technology and innovative models that utilize these different kinds of devices and services, appeared to be a starting point for the penetration of Big Data in almost all areas of human activity, including the commercial sector and public administration [27].

Nowadays, Big Data and the continuing dramatic increase in human and machinegenerated data associated with it are quite evident. However, do we actually know what Big Data is, and how close are the various definitions put forward for this term? For instance, there was an article in Forbes in 2014 which is related to this controversial question [28]. It gave a brief history of the establishment of the term, and provided several existing explanations and descriptions of Big Data to improve the core understanding of the phenomenon. On the other hand, Berkeley School of Information published a list with more than 40 definitions of the term [29].

As Big Data covers various fields and sectors, the meaning of this term should be specifically defined in accordance with the activity of the specific organization/person. For instance, in contrast to industry-driven Big Data "V's" definitions, Dr. Ivo Dinov for his research scope listed another data's multi-dimensional characteristics [30] such as data size, incompleteness, incongruency, complex representation, multiscale nature and heterogeneity of its sources [31,32].

In this paper the modified Gartner Inc. definition [33, 34] is used: Big Data is a technology to process high-volume, high-velocity, high-variety data or data-sets to extract intended data value and ensure high veracity of original data and obtained information that demand cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control [35].

Big Data processing methods

Currently, there exist many different techniques for data analysis [36], mainly based on tools used in statistics and computer science. The most advanced techniques to analyze large amounts of data include: artificial neural networks [37–39]; models based on the principle of the organization and functioning of
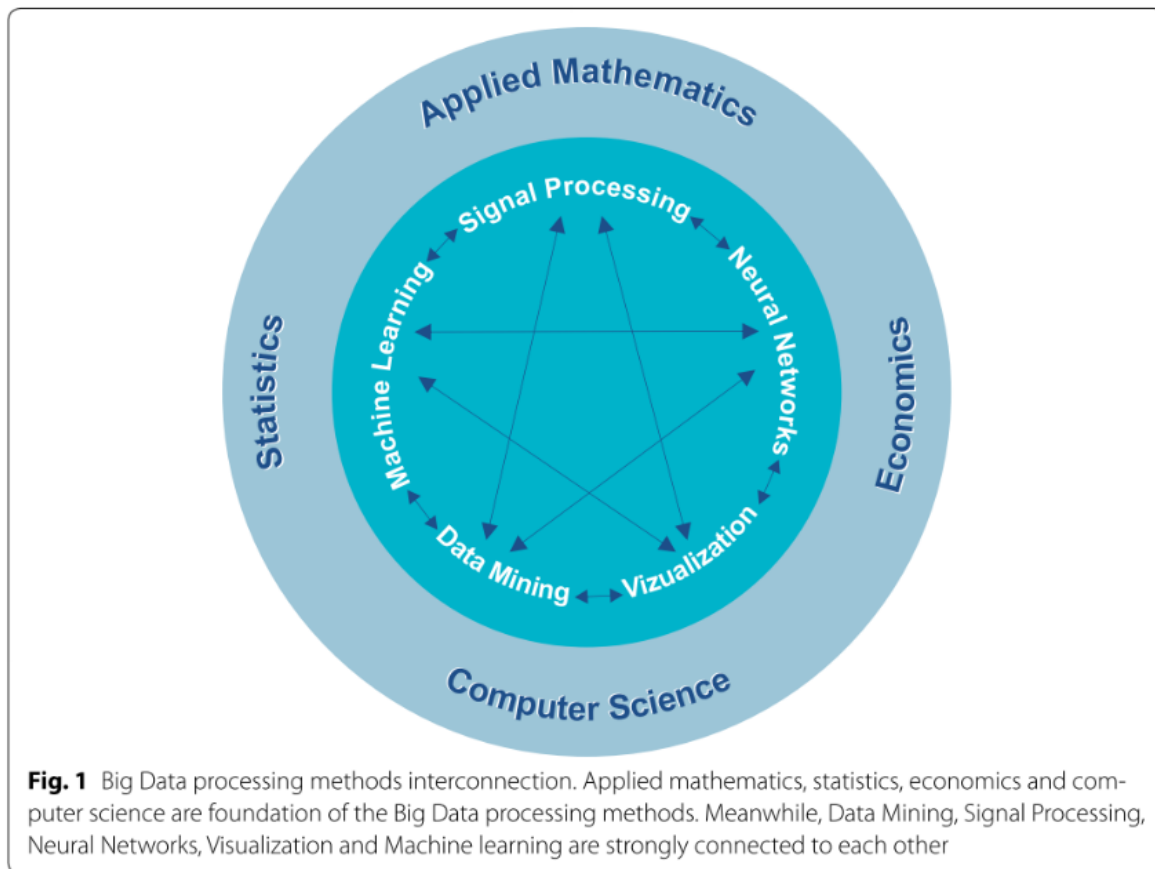
biological neural networks [40, 41]; methods of predictive analysis [42]; statistics [43, 44]; Natural Language Processing [45]; etc. Big Data processing methods embrace different disciplines including applied mathematics, statistics, computer science and economics. Those are the basis for data analysis techniques such as Data Mining [39, 46–49], Neural Networks [41, 50–52], Machine Learning [53–55], Signal Processing [56–58] and Visualization Methods [59–61]. Most of these methods are interconnected and used simultaneously during data processing, which increases system utilization tremendously (see Fig. 1).

We would like to familiarize reader with the primary methods and techniques in Big Data processing. As this topic is not a focus of the paper, this list is not exhaustive. Nevertheless, the main interconnections between these methods are shown and application examples are given.

Optimization methods are mathematical tools for efficient data analysis. Optimization includes numerical analysis focused on problem solving in various Big Data challenges: volume, velocity, variety and veracity [62] that will be discussed in more detail later. Some widely used analytical techniques are genetic programming [63–65], evolutionary programming [66] and particle swarm optimization [67, 68]. Optimization is focused on the search of the optimal set of actions needed to improve system performance. Notably, genetic algorithms are also a specific part of machine learning direction [69]. Moreover, statistical testing, predictive and simulation models are applied also as for Statistics methods [70].

Statistics methods are used to collect, organize and interpret data, as well as to outline interconnections between realized objectives. Data-driven statistical analysis concentrates on implementation of statistics algorithms [71, 72]. A/B testing [73] technique is an example of a statistics method. In terms of Big Data there is a possibility to perform a variety of tests. The aim of A/B tests is to detect statistically important differences and regularities between groups of variables to reveal improvements. Besides, statistical techniques contain cluster analysis, data mining and predictive modelling methods. Some techniques in spatial analysis [74] originate from the field of statistics as well. It allows analysis of topological, geometric or geographic characteristics of data sets.

Data mining includes cluster analysis, classification, regression and association rule learning techniques. This method is aimed at identifying and extracting beneficial information from extensive data or datasets. Cluster analysis [75, 76] is based on principles of similarities to classify objects. This technique belongs to unsupervised learning [77, 78] where training data [79] is used. Classification [80] is a set of techniques which are aimed at recognizing categories with new data points. In contrast to cluster analysis, a classification technique uses training data sets to discover predictive relationships. Regression [81] is a set of a statistical techniques that are aimed at determining changes between dependent and independent variables. This technique is mostly used for prediction or forecasting. Association rule learning [82, 83] is set of techniques designed to detect valuable relationships or association rules among variables in databases.

Fig. 1 Big Data processing methods interconnection. Applied mathematics, statistics, economics and computer science are foundation of the Big Data processing methods. Meanwhile, Data Mining, Signal Processing, Neural Networks, Visualization and Machine learning are strongly connected to each other

Machine Learning is a significant area in computer science which aims to create algorithms and protocols. The main goal of this method is to improve computers' behaviors on the basis of empirical data. Its implementation allows recognition of complicated patterns and automatic application of intelligent decision-making based on. Pattern recognition, natural language processing, ensemble learning and sentiment analysis are examples of machine learning techniques. Pattern recognition [84, 85] is a set of techniques that use a certain algorithm to associate an output value with a given input value. Classification technique is an example of this. Natural language processing [86] takes its origins from computer science within the fields of artificial intelligence and linguistics. This set of techniques performs analysis of human language. Sometimes it uses a sentiment analysis [87] that is able to identify and extract specific information from text materials evaluating words, degree and strength of a sentiment. Ensemble learning [88, 89] in automated decision-making systems is a useful technique for diminishing variance and increase accuracy. It aims to solve diverse machine learning issues such as confidence estimation, missing feature and error correction, etc.

Signal processing consists of various techniques that are part of electrical engineering and applied mathematics. The key aspect of this method is the analysis of discrete and continuous signals. In other words, it enables the analog representation of physical quantities (e.g. radio signals or sounds, etc.). Signal detection theory [90] is applied to evaluate the capacity for distinguishing between signal and noise in some techniques. A time series analysis [91, 92] includes techniques from both statistics and signal processing. Primarily, it is designed to analyze sequences of data points with a demonstration of data values at consistent times. This technique is useful to predict future data values based on knowledge

of past ones. Signal processing techniques can be applied to implement some types of data fusion [93]. Data fusion combines multiple sources to obtain improved information that is more relevant or less expensive and has higher quality [94].
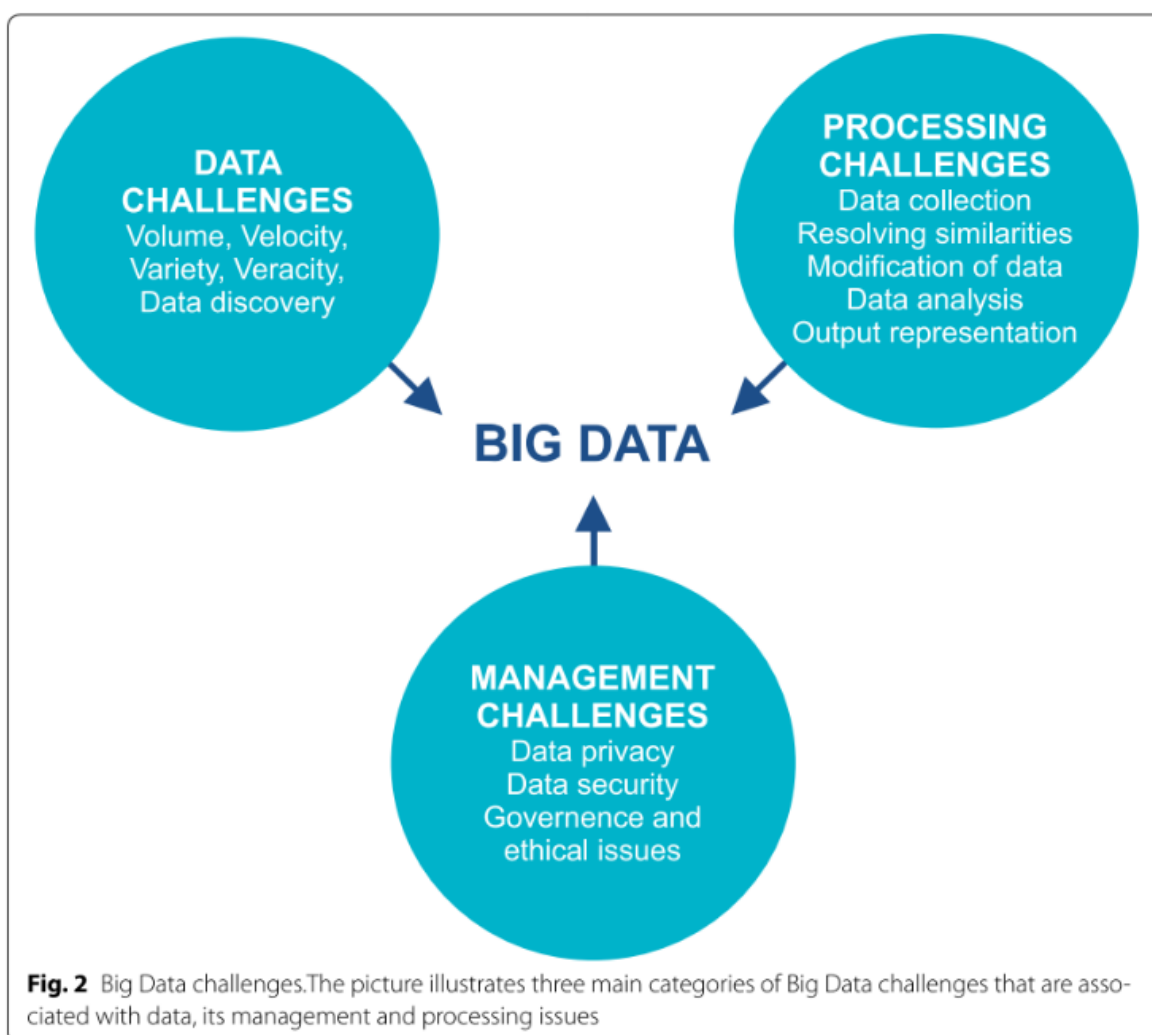
Visualization methods concern the design of graphical representation, i.e. to visualize the innumerate amount of the analytical results as diagrams, tables and images. Visualization for Big Data differs from all of the previously mentioned processing methods and also from traditional visualization techniques. To visualize large-scale data, feature extraction and geometric modelling can be implemented. These processes are needed to decrease the data size before actual rendering [95]. Intuitively, visual representation is more likely to be accepted by a human in comparison with unstructured textual information. The era of Big Data has been rapidly promoting the data visualization market. According to Mordor Intelligence [96] the visualization market will increase at a compound annual growth rate (CAGR) of 9.21 % from $4.12 billions in 2014 to $6.40 billions by the end of 2019. SAS Institute provides results of an International Data Group (IDG) research study in the white paper [97]. The research is focused on how companies are performing Big Data analysis. It shows that 98 % of the most effective companies working with Big Data are presenting results of the analysis via visualization. Statistical data from this research provides evidence of the visualization benefits in terms of decisionmaking improvement, better ad-hoc data analysis, improved collaboration and information sharing inside/outside an organization.

Nowadays, different groups of people including designers, software developers and scientists are in the process of searching for new visualization tools and opportunities. For example, Amazon, Twitter, Apple, Facebook and Google are companies that utilize data visualization in order to make appropriate business decisions [98]. Visualization solutions can provide insights from different business perspectives. First of all, implementation of advanced visualization tools enables rapid exploration of all customers/ users data to improve customer-company relationships. It allows marketers to create more precise customer segments based on data from purchasing history or life stage and other factors. Besides, correlation mapping may assist in the analysis of customer/ user behavior to identify and analyze the most profitable of them. Secondly, visualization capabilities allow companies opportunities to reveal correlations between product, sales and customer profiles. Based on gathered metrics, organizations may provide novel special offers to their customers. Moreover, visualization enables tracking of revenue trends and can be useful for risk analysis. Thirdly, visualization as a tool provides better understanding of data. Higher efficiency is reached by obtaining relevant, consistent and accurate information. So, visualized data could assist organizations to find different effective marketing solutions. In this section we familiarized the reader with the main techniques of data analysis and described their strong correlation to each other. Nevertheless, the Big Data era is still in the beginning stage of its evolution. Therefore, Big Data processing methods are evolving to solve the problems of Big Data and new solutions are continuously being developed. By this statement we mean that big world of Big Data requires multiple multidisciplinary methods and techniques that lead to better understanding of the complicated structures and interconnections between them.

Big Data challenges

Big Data has some inherent challenges and problems that can be primarily divided into three groups according to Akerkar et al. [36]: (1) data, (2) processing and (3) management challenges (see Fig. 2). While dealing with large amounts of information we face such challenges as volume, variety, velocity and veracity that are also known as 5V of Big Data. As those Big Data characteristics are well examined in scientific literature [99– 101] we will only discuss them briefly. Volume refers to the large amount of data, especially, machine-generated. This characteristic defines a size of the data set that makes its storage

and analysis problematic utilizing conventional database technology. Variety is related to different types and forms of data sources: structured (e.g. financial data) and unstructured (social media conversations, photos, videos, voice recordings and others). Multiplicity of the various data results in the issue of its handling. Velocity refers to the speed of new data generation and distribution. This characteristic requires the implementation of real-time processing for the streaming data analysis (e.g. on social media, different types of transactions or trading systems, etc.). Veracity refers to the complexity of data which may lead to a lack of quality and accuracy. This characteristic reveals several challenges: uncertainty, imprecision, missing values, misstatement and data availability. There is also a challenge regarding data discovery that is related to the search of high quality data in data sets.

Fig. 2 Big Data challenges. The picture illustrates three main categories of Big Data challenges that are associated with data, its management and processing issues

The second branch of Big Data challenges is called processing challenges. It includes data collection, resolving similarities found in different sources, modification data to a type acceptable for the analysis, the analysis itself and output representation, i.e. the results visualization in a form most suitable for human perception.

The last type of challenge offered by this classification is related to data management. Management challenges usually refer to secured data storage, its processing and collection. Here the main focuses of

study are: data privacy, its security, governance and ethical issues. Most of them are controlled based on policies and rules provided by information security institutes on state or international levels.

Over past generations, the results of analyzed data were represented as visualized plots and graphs. It is evident that collections of complex figures are sometimes hard to perceive, even by well-trained minds. Nowadays, the main factors causing difficulties in data visualization continue to be the limitations of human perception and new issues related to display sizes and resolutions. This question is studied in detail further in the section "Integration with Augmented and Virtual Reality". Preparatory to the visualization, the main interaction problem is in the extraction of the useful portion of information from massive volumes. Extracted data is not always accurate and mostly overloaded with excrescent information. Visualization technique is useful for simplifying information and transforming it into a more accessible form for human perception.

In the near future, petascale data may cause analysis failures because of traditional approaches in usage, i.e. when the data is stored on a memory disk continuously waiting for further analysis. Hence, the conservative approach of data compressing may become ineffective in visualization methods. To solve this issue, developers should create a flexible tool for the practice of data collection and analysis. Increases in data size make the multilevel hierarchy approach incapable in data scalability. Hierarchy becomes complex and intensive, making navigation difficult for user perception. In this case, a combination of analytics and Data Visualization may enable more accessible data exploration and interaction, which would allow improving insights, outcomes and decision-making.

Contemporary methods, techniques and tools for data analysis are still not flexible enough to discover valuable information in the most efficient way. The question of data perception and presentation remains open. Scientists face the task of uniting the abstract world of data and the physical world through visual representation. Meanwhile, visualization-based tools should fulfill three requirements [102, 103]: expressiveness (demonstrate exactly the information contained in the data), effectiveness (related to cognitive capabilities of human visual system) and appropriateness (cost-value ratio for visualization benefit assessment). Experience of previously used techniques can be repurposed to achieve more beneficial and novel goals in Big Data perception and representation.

Visualization methods

Historically, the primary areas of visualization were Science Visualization and Information Visualization. However, during recent decades, the field of Visual Analytics was actively developing.

As a separate discipline, visualization emerged in 1980 [104] as a reaction to the increasing amount of data generated by computer calculations. It was named Science Visualization [105–107], as it displays data from scientific experiments related to physical processes. This is primarily a realistic three-dimensional visualization, which has been used in architecture, medicine, biology, meteorology, etc. This visualization is also known as Spatial Data visualization, which focuses on the visualization of volumes and surfaces.

Information Visualization [108–111] emerged as a branch of the Human-Computer Interaction field in the end of 1980s. It utilizes graphics to assist people in comprehending and interpreting data. As it helps to form mental models of the data, for humans it is easier to reveal specific features and patterns of the obtained information.

Visual Analytics [112–114] combines visualization and data analysis. It has absorbed features of Information Visualization as well as Science Visualization. The main difference from other fields is the development and provision of visualization technologies and tools.
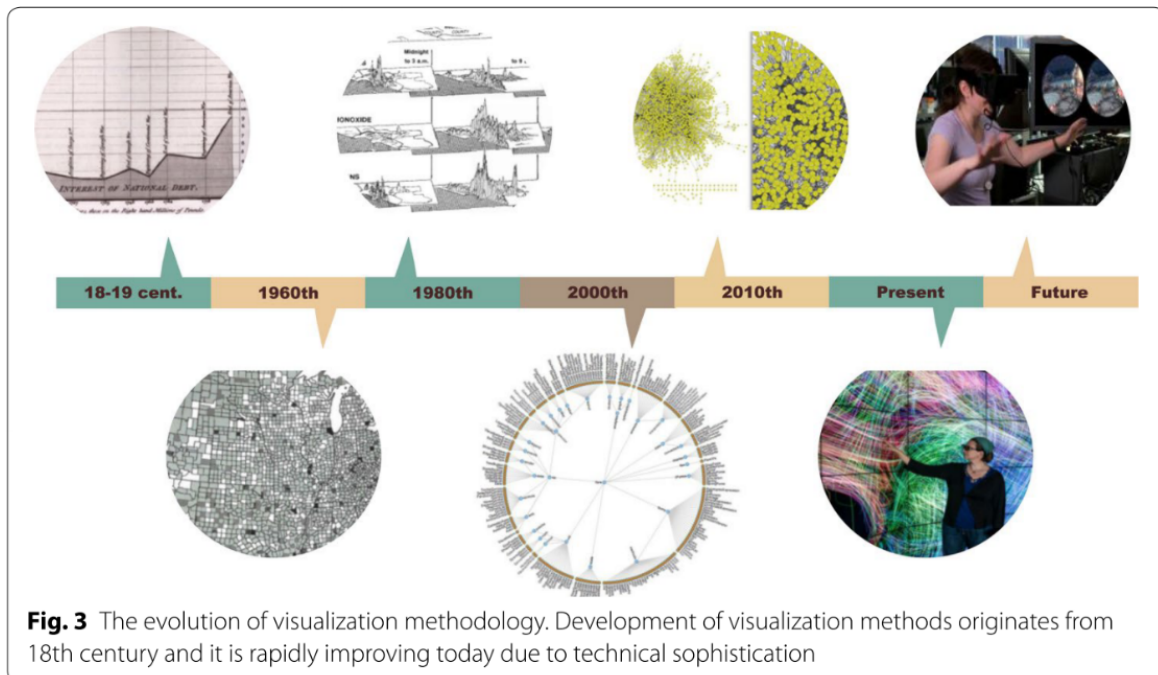
Efficient visualization tools should consider cognitive and perceptual properties of the human brain. Visualization aims to improve the clarity and aesthetic appeal of the displayed information and allows a person to understand large amount of data and interact with it. Significant purposes of Big Data visual representation are: to identify hidden patterns or anomalies in data; to increase flexibility while searching of certain values; to compare various units in order to obtain relative difference in quantities; to enable realtime human interaction (touring, scaling, etc.).

Visualization methods have evolved much over the last decades (see Fig. 3), the only limit for novel techniques being human imagination. To anticipate the next steps of data visualization development, it is necessary to take into account the successes of the past. It is considered that quantitative data visualization appeared in the field of statistics and analytics quite recently. However, the main precursors were cartography and statistical graphics, created before the 19th century for the expansion of statistical thinking, business planning and other purposes [115]. The evolution in the knowledge of visualization techniques resulted in mathematical and statistical advances as well as in drawing and reproducing images.

By the 16th century, tools for accurate observation and measurement were developed. Precisely, in those days the first steps were done in the development of data visualization. The 17th century was swept by the problem of space, time and distance measurements. Furthermore, the study of the world's population and economic data had started.

The 18th century was marked by the expansion of statistical theory, ideas of data graphical representation and the advent of new graphic forms. At the end of the century thematic maps displaying geological, medical and economic data was used for the first time. For example, Charles de Fourcroy used geometric figures and cartograms to compare areas or demographic quantities [116]. Johann Lambert (1728–1777) was a revolutionary person, who used different types of tables and line graphs to display variable data [117]. The first methods were performed as simple plots followed by onedimensional histograms [118]. Still, those examples are useful only for small amounts of data. By introducing more information, this type of diagram would reach a point of worthlessness.

At the turn of 20–21st centuries, steps were taken in the development of interactive statistical computing [119] and new paradigms for data analysis [120]. Technological progress was certainly a significant prerequisite for the rapid development of visualization techniques, methods and tools. More precisely, large-scale statistical and graphics software engineering was invented, and computer processing speed and capacity vastly increased [121].

Fig. 3 The evolution of visualization methodology. Development of visualization methods originates from 18th century and it is rapidly improving today due to technical sophistication

However, the next step presenting a system with the addition of a time dimension appeared as a significant breakthrough. In the beginning of the present century few dimensional visualization methods were in use as a part 2D/3D node-link diagram [122]. Already at this level of abstraction, any user may classify the goal and specify further analytical steps for the research, but unfortunately, data scaling became an essential issue.

Moreover, currently used technologies for data visualization are already causing enormous resource demands which include high memory requirements and extremely high deployment cost. However, the currently existing environment faces a new limitation based on the large amounts of data to be visualized in contrast to past imagination issue. Modern effective methods are focused on representation in specified rooms equipped with widescreen monitors or projectors [123].

Nowadays, there are a fairly large number of data visualization tools offering different possibilities. These tools can be classified based on three factors: by the data type, by visualization technique type and by the interoperability. The first refers to the different types of data to be visualized [124]:
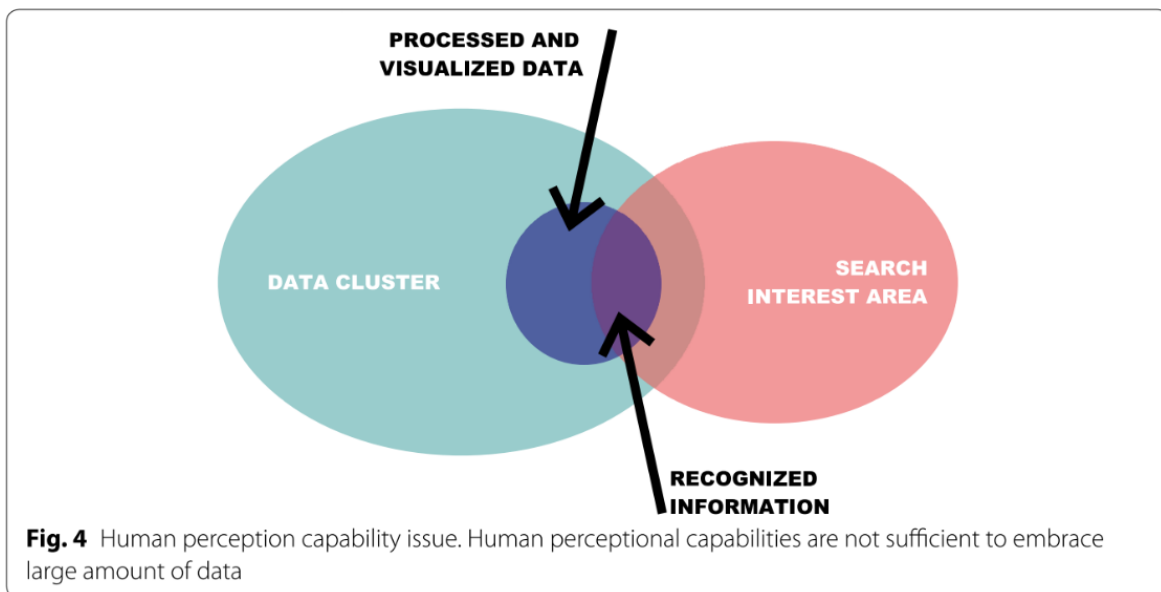
- *Univariate data* One dimensional arrays, time series, etc.
- *Two-dimensional data* Point two-dimensional graphs, geographical coordinates, etc.
- *Multidimensional data* Financial indicators, results of experiments, etc.
- *Texts and hypertexts* Newspaper articles, web documents, etc.
- *Hierarchical and links* The structure subordination in the organization, e-mails, documents and hyperlinks, etc.
- *Algorithms and programs* Information flows, debug operations, etc.
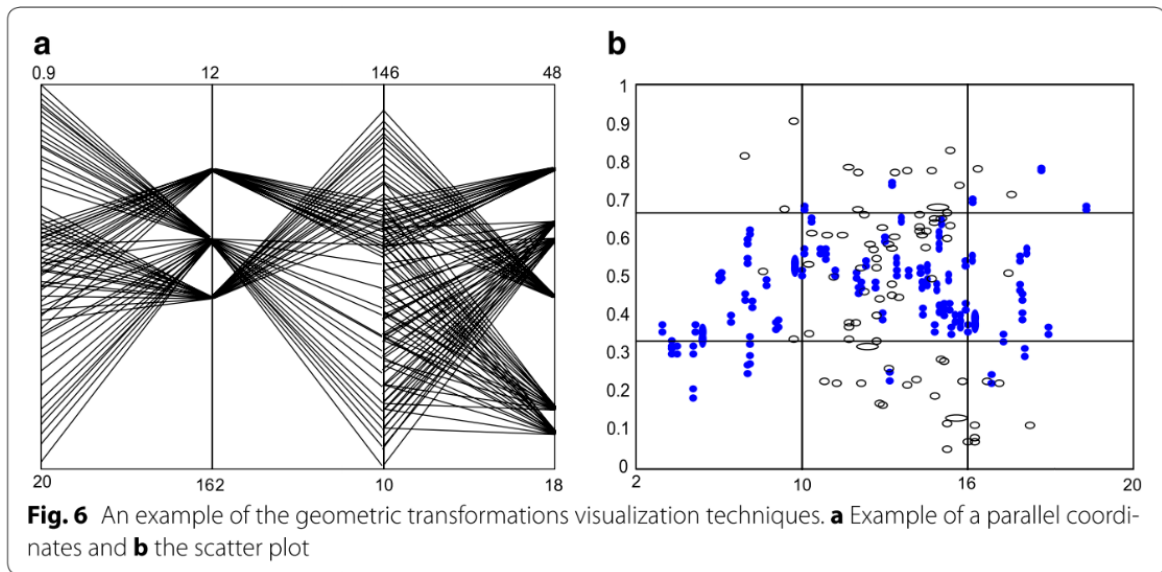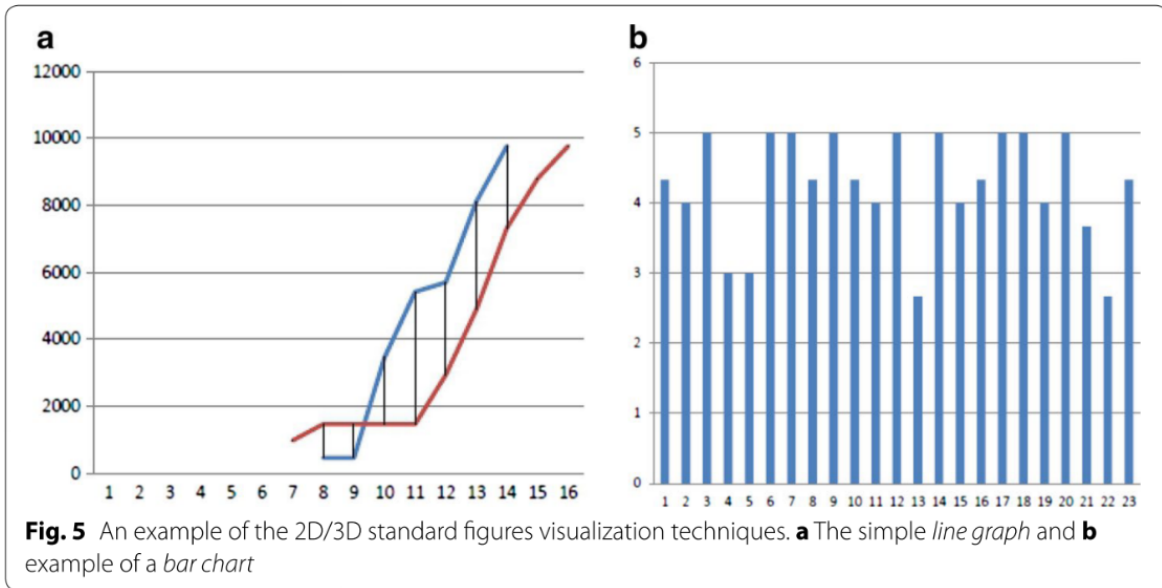
The second factor is based on visualization techniques and samples to represent different types of data. Visualization techniques can be both elementary (line graphs, charts, bar charts) and complex (based on the mathematical apparatus). Furthermore, visualization can be performed as a combination of various

methods. However, visualized representation of data is abstract and extremely limited by one's perception capabilities and requests (see Fig. 4).
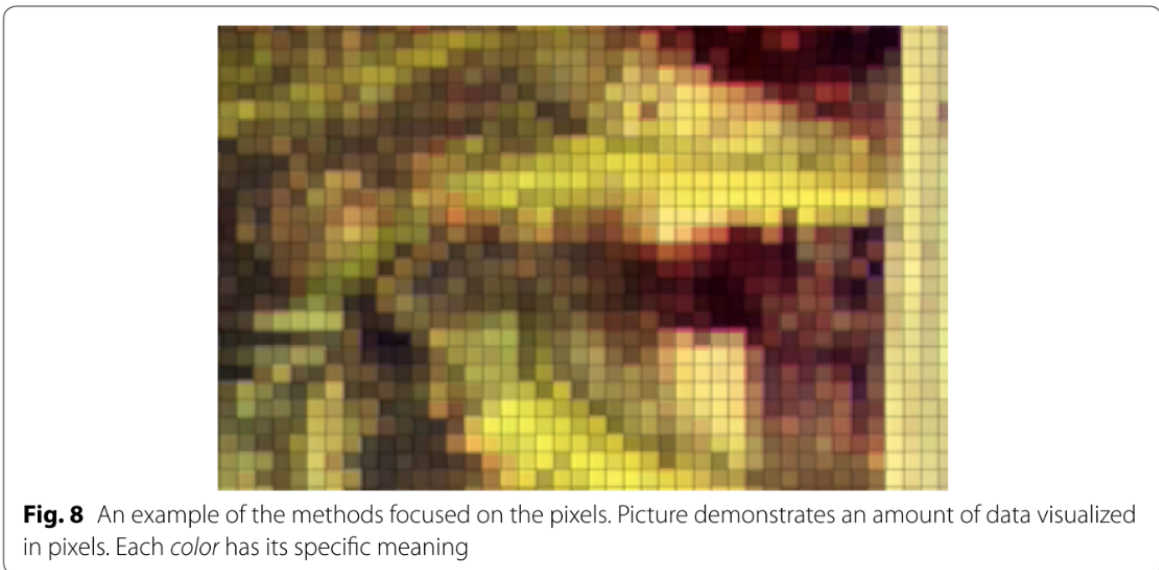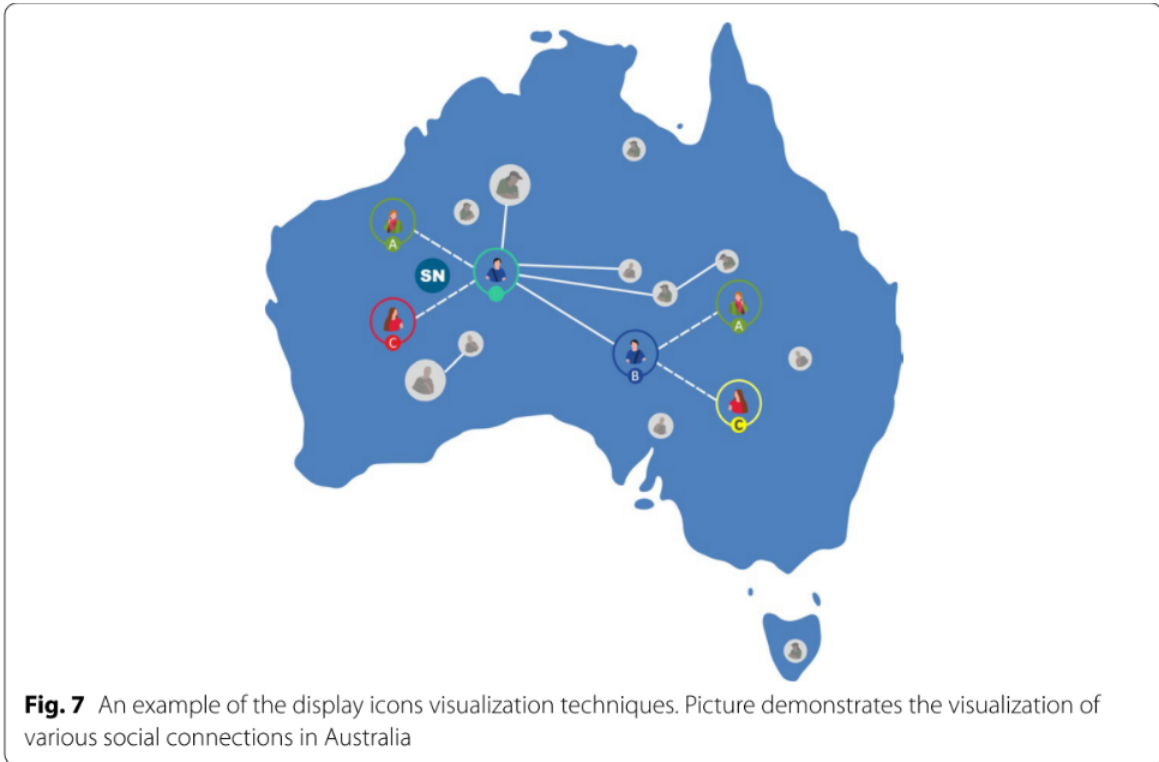
Types of visualization techniques are listed below:

1. 2D/3D standard figure [125]. May be implemented as bars, line graphs, various charts, etc. (see Fig. 5). The main drawback of this type is the complexity of the acceptable visualization for complicated data structures;
2. Geometric transformations [126]. This technique represents information as scatter diagram (see Fig. 6). This type is geared towards a multi-dimensional data set's transformation in order to display it in Cartesian and non-Cartesian geometric spaces. This class includes methods of mathematical statistics;
3. Display icons [127]. Ruled shapes (needle icons) and star icons. Basically, this type displays the values of elements of multidimensional data in properties of images (see Fig. 7). Such images may include human faces, arrows, stars, etc. Images can be grouped together for holistic analysis. The result of the visualization is a texture pattern, which varies according to the specific characteristics of the data;
4. Methods focused on the pixels [128]. Recursive templates and cyclic segments. The main idea is to display the values in each dimension into the colored pixel and to merge some of them according to specific measurements (see Fig. 8). Since one pixel is used to display a single value, therefore visualization of large amounts of data can be reachable with this methodology;
5. Hierarchical images [129]. Tree maps and overlay measurements (see Fig. 9). These type methods are used with the hierarchical structured data.

Fig. 4 Human perception capability issue. Human perceptional capabilities are not sufficient to embrace large amount of data

Fig. 5 An example of the 2D/3D standard figures visualization techniques. **a** The simple *line graph* and **b** example of a *bar chart*

Fig. 6 An example of the geometric transformations visualization techniques. **a** Example of a parallel coordinates and **b** the scatter plot

The third factor is related to the interoperability with visual imagery and techniques for better data analysis. The application used for the visualization should present visual forms that capture the essence of data itself. However, it is not always enough for a complete analysis. Data representation should be constructed in order to allow a user to have different visual points of view. Thus, the appropriate compatibility should be performed:

Fig. 7 An example of the display icons visualization techniques. Picture demonstrates the visualization of various social connections in Australia

Fig. 8 An example of the methods focused on the pixels. Picture demonstrates an amount of data visualized in pixels. Each *color* has its specific meaning

1. *Dynamic projection* [130]. Non-static change of projections in multidimensional data sets is used. An example of the dynamic projection in two-dimensional plane of multidimensional data in a scatter plots. It is necessary to note that the number of possible projections increases exponentially with the number of measurements and, thus, perception suffers more.
2. *Interactive filtering* [131]. In the investigation of large amounts of data there is a need to share data sets and highlight significant subsets in order to filter images. Significantly, that there should

be an opportunity to have a visual representation in real time. A subset can be chosen either directly from a list or by determining a subset of the properties of interest;

3. *Scaling images* [132]. Scaling is a well-known method of interaction used in many applications. Especially for Big Data processing, this method is very useful due to the ability to represent data in a compressed form. It provides the ability to simultaneously display any part of an image in a more detailed form. Nevertheless, a lower level entity may be represented by a pixel at a higher level, a certain visual image or an accompanying text label;

4. *Interactive distortion* [133] supports the research process data using distortion scale with partial detail. The basic idea of this method is that a part of the fine granularity displayed data is shown in addition to one with a low level of details. The most popular methods are hyperbolic and spherical distortion;

5. *Interactive combination* [134, 135] brings together a combination of different visualization techniques to overcome specific deficiencies by their conjugation. For example, different points of the dynamic projection can be combined with the techniques of coloring.

Fig. 9 An example of the hierarchical images. Picture illustrates a tree map of data
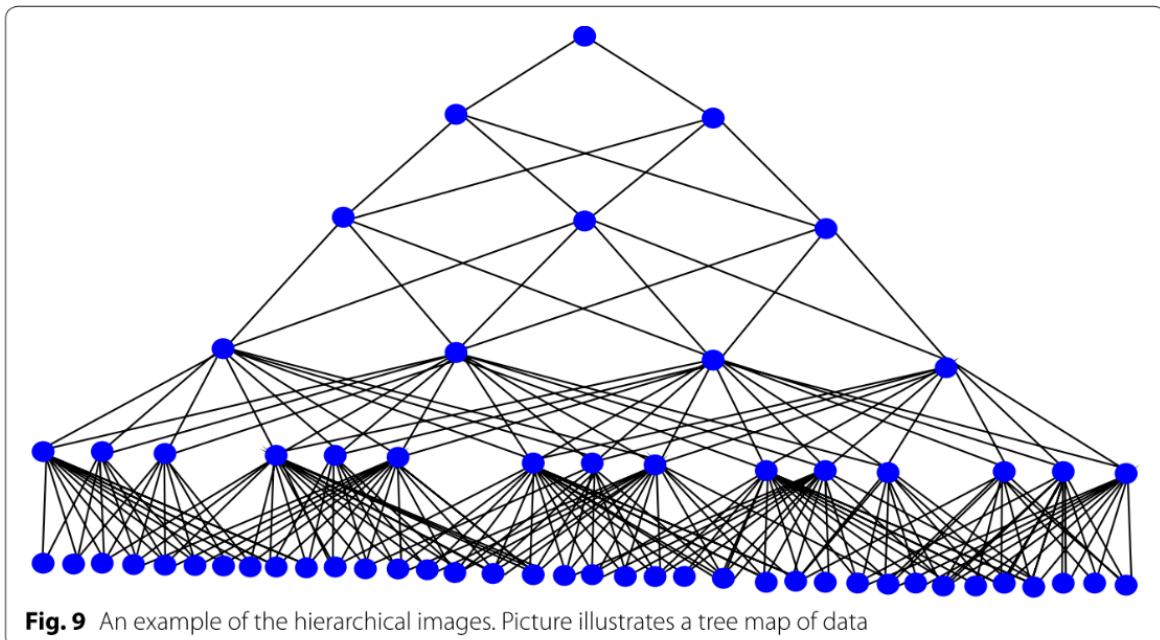
To summarize, any visualization method can be classified by data type, visualization technique and interoperability. Each method can support different types of data, various images and varied methods for interaction.

A visual representation of Big Data analysis is crucial for its interpretation. As it was already mentioned, it is evident that human perception is limited. The main purpose of modern data representation methods is related to improvement in forms of images, diagrams or animation. Examples of well known techniques for data visualization are presented below [136]:

- *Tag cloud* [137] is used in text analysis, with a weighting value dependent on the frequency of use (citation) of a particular word or phrase (see Fig. 10). It consists of an accumulation of lexical items (words, symbols or combination of the two). This technique is commonly integrated with web sources to quickly familiarize visitors with the content via key words.

- *Clustergram* [138] is an imaging technique used in cluster analysis by means of representing the relation of individual elements of the data as they change their number (see Fig. 11). Choosing the optimal number of clusters is also an important component of cluster analysis.
- *Motion charts* allow effective exploration of large and multivariate data and interact with it utilizing dynamic 2D bubble charts (see Fig. 12). The blobs (bubbles—central objects of this technique) can be controlled due to variable mapping for which it is designed. For instance, motion charts graphical data tools are provided by Google [139], amCharts [140] and IBM Many Eyes [141].
- *Dashboard* [142] enables the display of log files of various formats and filter data based on chosen data ranges (see Fig. 13). Traditionally, dashboard consists of three layers [143]: data (raw data), analysis (includes formulas and imported data from data layer to tables) and presentation (graphical representation based on the analysis layer)

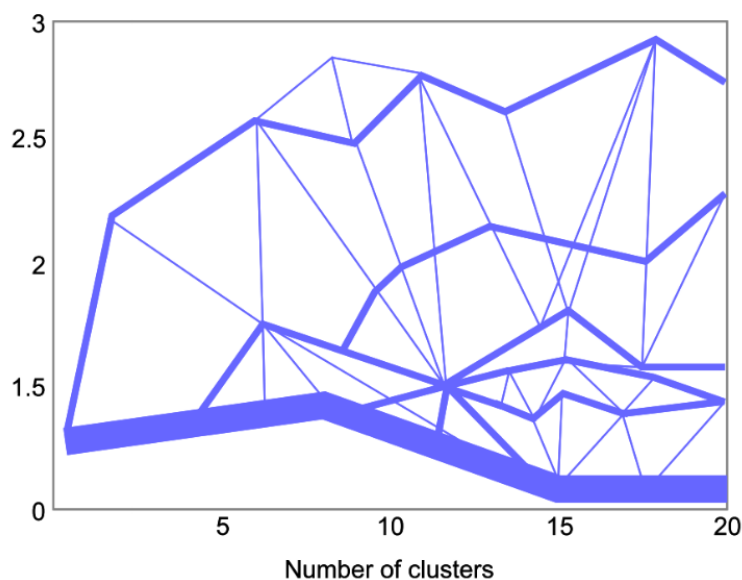**Fig. 10** An example of the tag cloud. This picture illustrates visualization of the paper abstract

Fig. 11 An example of the clustergram. This picture illustrates different state of data in several clusters
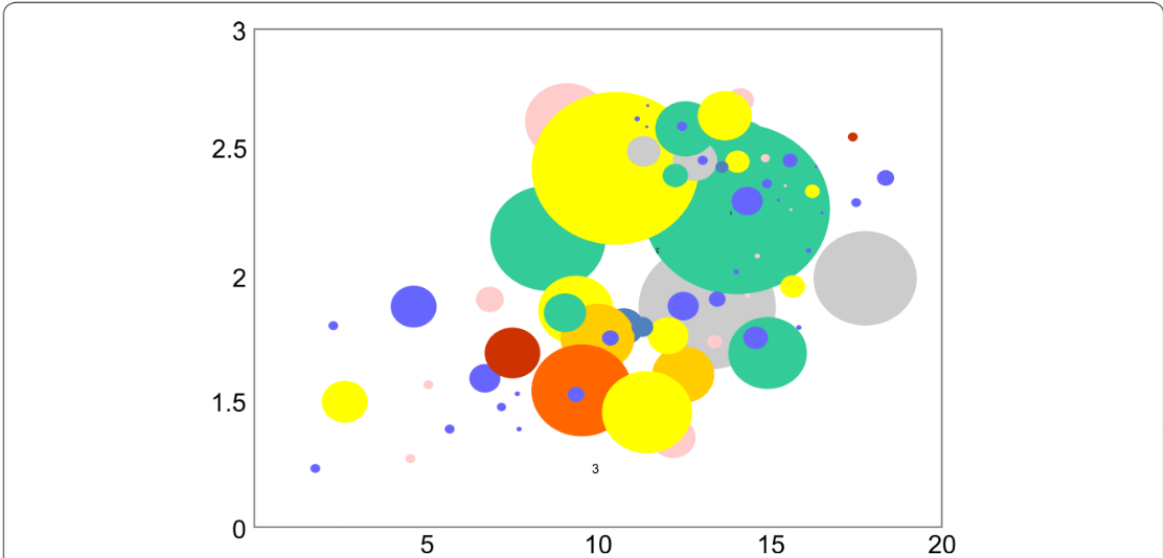
Fig. 12 An example of the motion chart. This picture illustrates the data in forms of bubbles that have various meaning based on *color* and size
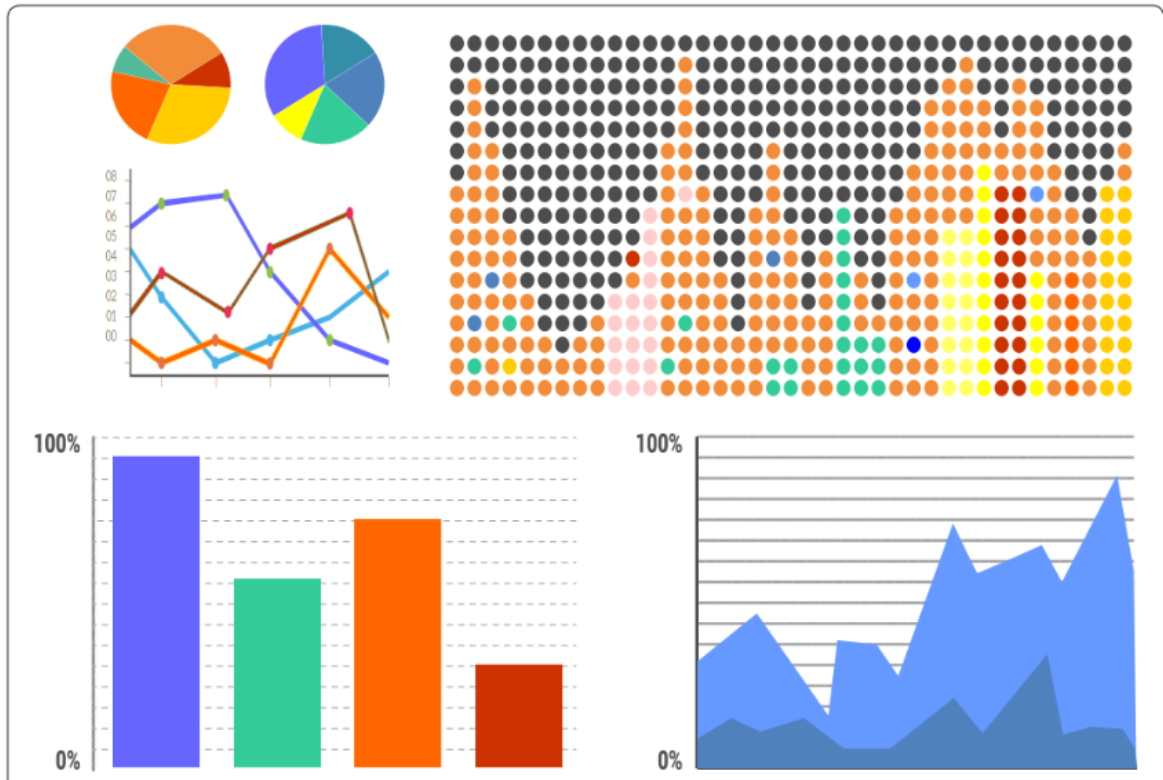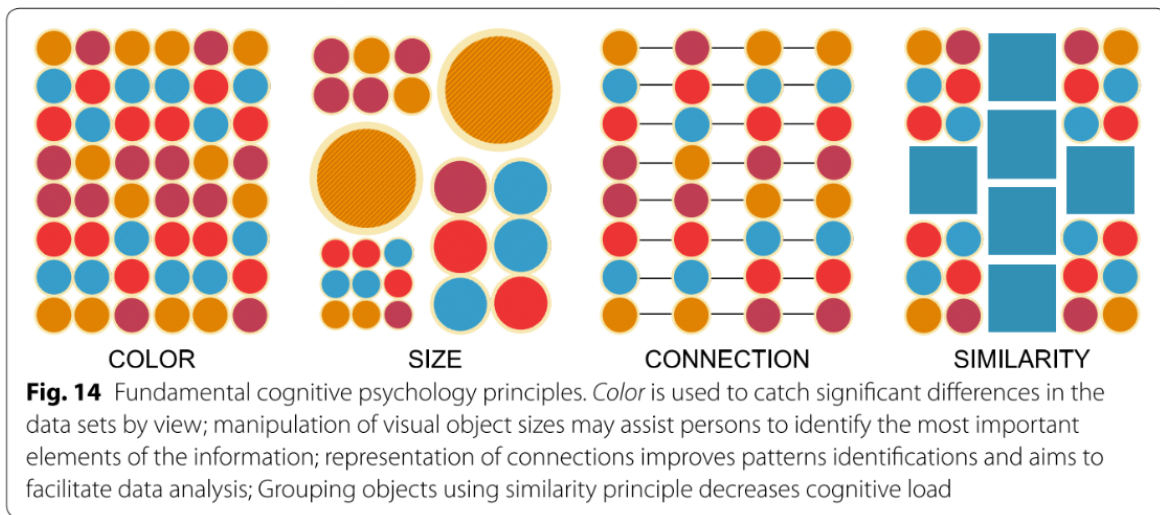

Fig. 13 An example of the dashboard. This picture illustrates pie chart, visualization of data in pixels, *line graph* and *bar chart*

Nowadays, there are many publicly available tools to create meaningful and attractive visualizations. For instance, there is a chart of open visualization tools for data visualization and analysis

published by Sharon Machils [144]. The author provides a list, which contains more than 30 tools from easiest to most difficult: Zoho Reports, Weave, Infogr. am, Datawrapper and others.

All of these modern methods and tools follow fundamental cognitive psychology principles and use the essential criteria of data successful representation [145] such as manipulation of size, color and connections between visual objects (see Fig. 14). In terms of human cognition, the Gestalt Principles [146] are relevant. The basis of Gestalt psychology is a study of visual perception. It suggests that people tend to perceive the world in a form of holistic ordered configuration rather than constituent fragments (e.g. at first, person perceives forest and after that can identify single trees as part of the whole). Moreover, our mind fills in the gaps, seeks to avoid uncertainty and easily recognizes similarities and differences. The main Gestalt principles such as law of proximity (collection of objects forming a group), law of similarity (objects are grouped perceptually if they are similar to each other), symmetry (people tend to perceive object as symmetrical shapes), closure (our mind tends to close up objects that are not complete) and figure-ground law (prominent and recessed roles of visual objects) should be taken into account in Big Data Visualization.

Fig. 14 Fundamental cognitive psychology principles. *Color* is used to catch significant differences in the data sets by view; manipulation of visual object sizes may assist persons to identify the most important elements of the information; representation of connections improves patterns identifications and aims to facilitate data analysis; Grouping objects using similarity principle decreases cognitive load

To this end, the most effective visualization method is the one that uses multiple criteria in the optimal manner. Otherwise, too many colors, shapes, and interconnections may cause difficulties in the comprehension of data, or some visual elements may be too complex to recognize.

After observation and discussion about existing visualization methods and tools for Big Data, we can clarify and outline its important disadvantages that are sufficiently discussed by specialists from different fields [147–149]. Various ways of data interpretation make them meaningful. It is easy to distort valuable information in its visualization, because a picture convinces people more effectively than textual content. Existing visualization tools aim to create as simple and abstract images as possible, which can lead to a problem when significant data can be interpreted as disordered information and important connections between data units will be hidden from the user. It is a problem of visibility loss, which also refers to display resolution, where the quality of represented data depends on number of pixels and their density. A solution may be in the use of larger screens [150]. However, this concept brings a problem of human brain cognitiveperceptual limitations, as will be discussed in detail in the section Integration with Augmented and Virtual Reality.

Using visual and automated methods in Big Data processing gives a possibility to use human knowledge and intuition. Moreover, it becomes possible to discover novel solutions for complex data visualization [151]. Vast amounts of information motivate researchers and developers to create new tools for quick and accurate analysis. As an example, the rapid development of visualization techniques may be concerned. In the world of interconnected research areas, developers need to combine existing basic, effective visualization methods with new technological opportunities to solve the central problems and challenges of Big Data analysis.

Integration with augmented and virtual reality

It is well known that the vision perception capabilities of the human brain are limited [152]. Furthermore, handling a visualization process on currently used screens requires high costs in both time and health. This leads to the need of its proper usage in the case of image interpretation. Nevertheless, the market is in the process of being flooded with countless numbers of wearable devices [153, 154] as well as various display devices [155, 156].

The term Augmented Reality was invented by Tom Caudell and David Mizel in 1992 and meant to describe data produced by a computer that is superimposed to the real world [157]. Nevertheless, Ivan Sutherland created the first AR/VR system already in 1968. He developed the optical see-through head-mounted display that can reveal simple three-dimensional models in real time [158]. This invention was a predecessor to the modern VR displays and AR helmets [159] that seem to be an established research and industrial area for the coming decade [160]. Applications for use have already been found in military [161], education [162], healthcare [163], industry [164] and gaming fields [165]. At the moment, the Oculus Rift [166] helmet gives many opportunities for AR practice. Concretely, it will make it possible to embed virtual content into the physical world. William Steptoe has already done research in this field. The use of it in the visualization area might solve many issues from narrow visual angle, navigation, scaling, etc. For example, offering a way to have a complete 360-degrees view with a helmet can solve an angle problem. On the other hand, a solution can be obtained with help of specific widescreen rooms, which by definition involves enormous budgets. Focusing on the combination of dynamic projection and interactive filtering visualization methods, AR devices in combination with motion recognition tools might solve a significant scaling problem especially for multidimensional representations that comes to this area from the field of Architecture. Speaking more precisely, designers (specialized in 3D-visualization) work with flat projections in order to produce a visual model [167]. However, the only option to present a final image is in moving around it and thus navigation inside the model seems to be another influential issue [168].

From the Big Data visualization point of view, scaling is a significant issue mainly caused by multidimensional systems where a need to delve into a branch of information in order to obtain some specific value or knowledge takes its place. Unfortunately, it cannot be solved from a static point of view. Likewise, integration with motion detection wearables [169] would highly increase such visualization system usability. For example, the additional use of an MYO armband [170] may be a key to the interaction with visualized data in the most native way. Similar comparison may be given as a pencil-case in which one tries to find a sharpener and spreads stationery with his/her fingers.

However, the use of AR displays and helmets is also limited by specific characteristics of the human eye (visual system), such as field of view and/or diseases like scotoma [171] and blind spots [172]. Central vision [173] is most significant and necessary for human activities such as reading or driving. Additionally, it is responsible for accurate vision in the pointed direction and takes most of the visual cortex in the brain but its retinal size is less than 1 % [174]. Furthermore, it captures only two degrees of

the vision field, which stays the most considerable for text and object recognition. Nevertheless, it is supported with Peripheral vision which is responsible for events outside the center of gaze. Many researchers around the world are currently working with virtual and AR to train young professionals [175–177], develop new areas [178, 179] and analyze the patient's behavior [180].

Despite the well known topics like colorblindness, natural field of view and other physiological abnormalities, recent research by Israel Abramov et al. [181] is overviewing physiological gender and age differences based on the cerebral cortex and its large number of testosterone receptors [182], as a basis for the variety in perception procedures. The study was mainly about the focused image onto the retina at the back of the eyeball and its visual system processing. We overview the main reasons for those differences, starting from prehistoric times, when African habitats in forest regions had limited distance for object detection and identification, thus obtained higher acuity for males may be explained. Also, sex differences might be related to different roles in the survival commune. So that males were mainly hunting (hunter-gatherer hypothesis)—they had to detect enemies and predators much faster [183]. Moreover, there are significant gender differences for far- and near-vision: males have their advantage in a far-space [184]. On the other hand, females are much more susceptible for brightness and color changes in addition to static objects in near-space [185]. However, we can conclude that male/ female differences in the sensory capacities are adaptive but should be considered in order to optimize represented and visualized data for end-uses. Additionally, there exists a research area focusing on the human eye movement patterns during the perception of scenes and objects. It can be based on different factors starting from particular culture peculiar properties [186] and up to specific search tasks [187] being in high demand for Big Data visualization purposes.

Further studies shall be focused on the usage of ophthalmology and neurology for the development of the new visualization tools. Basically, such cross-discipline collaboration would support decision making for the image position selection, which is mainly related to the problem of the significant information losses due to the vision angle extension. Moreover, it is highly important to take in account current hardware quality and screens resolution in addition to the software part. Nevertheless, there is a need of the improvement for multicore GPU processors besides the address bus throughput refinement between CPU and GPU or even replacement for wireless transfer computations on cluster systems. Never the less, it is significant to discuss current visualization challenges to support future research.

Future research agenda and data visualization challenges

Visualized data can significantly improve the understanding of the preselected information for an average user. In fact, people start to explore the world using visual abilities since birth. Images are often more easier to perceive in comparison to text. In the modern world, we can see clear evolution towards visual data representation and imagery experience. Moreover, visualization software becomes ubiquitous and publicly available for ordinary user. As a result, visual objects are widely distributed—from social media to scientific papers and, thus, the role of visualization while working with large amount of data should be reconsidered. In this section, we overview important challenges and possible solutions related to future agenda for Big Data visualization with AR and VR usage:

1. *Application development integration* In order to operate with visualized objects, it is necessary to create a new interactive system for the user. It should support such actions as: scaling; navigating in visualized 3D space; selecting sub-spaces, objects, groups of visual elements (flow/path elements) and views; manipulating and placing; planning routes of view; generating, extracting and collecting data (based on the reviewed visualized data). A novel system should allow multimodal control by voice and/or gestures in order to make it more intuitive for users as it is

shown in [188– 190] and [191]. Nevertheless, one of the main issues regarding this direction of development is the fact that implementing effective gestural and voice interaction is not a trivial matter. There is a need to develop a machine learning system and to define basic intuitive gestures that are currently in research for general [192–194] and more specific (medical) purposes [195].

2. *Equipment and virtual interface* It is necessary to apply certain equipment for the implementation of such an interactive system in practice. Currently, there are optical and video see-trough head-mounted displays (HMD) [196] that merge virtual objects into the real scene view. Both have the following issues: distortion and resolution of the real scene; delay of a system; viewpoint matching; engineering and cost factors. As for the interaction issue, for an appropriate haptic feedback in an MR environment there is a need to create a framework that would allow an interaction with intuitive gesture. As it is revealed in the section Integration with Augmented and Virtual Reality, glove-based systems [197] are mainly used for virtual object manipulation. The disadvantage of hand-tracking input is so that there is no tactile feedback. In summary, the interface should be redesigned or reinvented in order to simplify user interaction. Software engineers should create new approaches, principles and methods in User Interface Design to make all instruments easily accessible and intuitive to use.

3. *Tracking and recognition system* Objects and tools have to be tracked in virtual space. The position and orientation values of virtual items are dynamic and have to be re-estimated during presentation. Tracking head movement is another significant challenge. It aims to avoid mismatch of the real view scene and computer generated objects. This challenge may be solved by using more flexible software platforms.

4. *Perception and cognition* Actually, the level of computer operation is high but still not sufficiently effective in comparison to human brain performance even in cases of neural networks. As was mentioned earlier in the section Integration with Augmented and Virtual Reality, human perception and cognition have their own characteristics and features, and the consideration of this issue by developers during hardware and interface design for AR is vital. In addition, the user's ability to recognize and understand the data is a central issue. Tasks such as browsing and searching require a certain cognitive activity. Also, there can be issues related to different users' reactions with regard to visualized objects depending on their personal and cultural backgrounds. In this sense, simplicity in information visualization has to be achieved in order to avoid misperceptions and cognitive overload [198]. Psychophysical studies would provide answers to questions regarding perception and would give the opportunity to improve performance by motion prediction.

5. *Virtual and physical objects mismatch* In an Augmented Reality environment, virtual images integrate with real world scenery at the static distance in the display while the distance to real objects varies. Consequently, a mismatch of virtual and physical distances is irreversible and it may result in incorrect focus, contrast and brightness of virtual objects in comparison to real ones. The human eye is capable of recognizing many levels of brightness, saturation and contrast [199], but most contemporary optical technologies cannot display all levels appropriately. Moreover, potential optical illusions arise from conflicts between computer-generated and real environment objects. Using modern equipment would be a solution for this challenge.

6. *Screen limitations* With the current technology development level, visualized information is presented mainly on screens. Even a VR helmet is equipped with two displays. Unfortunately, and because of the close-to-the-eye proximity, users can experience lack of comfort while working with it. It is mainly based on a low display resolution and high graininess and, thus, manufacturers should take it into consideration for further improvement.

7. *Education* As this concept is relatively new, there is a need to specify the value of the data visualization and its contribution to the users' work. The value cannot be so obvious; that is why compelling showcase examples and publicly available tutorials can reveal AR and VR potential in visual analytics. Moreover, users need to be educated and trained for the oncoming interaction with this evolving technology. The visual literacy skill should be improved in order to have high performance while working with visualized objects. A preferable guideline can be chosen as Visual Information-Seeking Mantra: overview first, zoom and filter, then details on demand [200].

Despite all the challenges, the main benefit from the implementation of MR approach is human experience improvement. At the same time, such visualization allows convenient access to huge amounts of data and provides a view from different angles. The navigation is smooth and natural via tangible and verbal interaction. It also minimizes perceptional inaccuracy in data analysis and makes visualization powerful at conveying knowledge to the end user. Furthermore, it ensures actionable insights that improves decision making.

In conclusion, challenges of data visualization for AR and VR are associated not only with current technology development but also with human-centric issues. Interestingly, some researchers are already working on the conjugation of such complex fields as massive data analysis, its visualization and complex control of the visualized environment [201]. It is worthwhile to note that those factors should be taken into account simultaneously in order to achieve the best outcome for the established industrial field.

Conclusion

In practice, there are a lot of challenges for Big Data processing and analysis. As all the data is currently visualized by computers, it leads to difficulties in the extraction of data, followed by its perception and cognition. Those tasks are time-consuming and do not always provide correct or acceptable results.

In this paper we have obtained relevant Big Data Visualization methods classification and have suggested the modern tendency towards visualization-based tools for business support and other significant fields. Past and current states of data visualization were described and supported by analysis of advantages and disadvantages. The approach of utilizing VR, AR and MR for Big Data Visualization is presented and the advantages, disadvantages and possible optimization strategies of those are discussed.

For visualization problems discussed in this work, it is critical to understand the issues related to human perception and limited cognition. Only after that, the field of design can provide more efficient and useful ways to utilize Big Data. It can be concluded that data visualization methodology may be improved by considering fundamental cognitive psychological principles and by implementing most natural interaction with visualized virtual objects. Moreover, extending it with functions to exclude blind spots and decreased vision sectors would highly improve recognition time for people with such a disease. Furthermore, a step towards wireless solutions would extend device battery life in addition to computation and quality improvements.

References

1. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. Big Data: the next frontier for innovation, competition, and productivity. June Progress Report. McKinsey Global Institute; 2011.

2. 1000 Genomes: a Deep Catalog of Human Genetic Variation. 2015. http://www.1000genomes.org/.

3. Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: new opportunities for research and social challenges. Genome Med 2010;2(3)

4. Internet Archive: Internet Archive Wayback Machine. 2015.http://archive.org/web/web.php.

5. Nielsen J. Comparing content in web archives: differences between the Danish archive Netarkivet and Internet Archive. In: Two-day Conference at Aarhus University, Denmark. 2015

6. The Lemur Project: The ClueWeb09 Dataset. 2015. http://lemurproject.org/clueweb09.php/.

7. Russom P. Managing Big Data. TDWI Best Practices Report, TDWI Research; 2013.

8. Gantz J, Reinsel D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView IDC Anal Future. 2012;2007:1–16.

9. Beyer MA, Laney D. The importance of "Big Data": a definition. Stamford: Gartner; 2012.

10. Demchenko Y, Ngo C, Membrey P. Architecture framework and components for the big data ecosystem. J Syst Netw Eng 2013;1–31 [SNE technical report SNE-UVA-2013-02].

11. Turner V, Reinsel D, Gantz JF, Minton S. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. IDC Analyze the Future 2014.

12. Husain SS, Kalinin A, Truong A, Dinov ID. SOCR data dashboard: an integrated Big Data archive mashing medicare, labor, census and econometric information. J Big Data. 2015;2(1):1–18.

13. Keahey TA. Using visualization to understand Big Data. IBM Business Analytics Advanced Visualisation; 2013.

14. Microsoft Corporation: Power BI—Microsoft. 2015. https://powerbi.microsoft.com/.

15. Amazon.com, Inc. Amazone Web Services. 2015. https://aws.amazon.com/

16. Google, Inc. Google Cloud Platform. 2015. https://cloud.google.com/.

17. Socrata. Data to the People. 2015. http://www.socrata.com.

18. D3.js: D3 Data-Draven Documents. 2015.http://d3js.org.

19. The Cytoscape Consortium: Network Data Integration, Analysis, and Visualization in Box. 2015. http://www. cytoscape.org.

20. Tableau—Business Intelligence and Analytics. http://tableau.com/

21. Kandel S, Paepcke A., Hellerstein J, Heer J. Wrangler: interactive visual specification of data transformation scripts. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM; 2011. pp 3363–72.

22. Schaefer D, Chandramouly A, Carmack B, Kesavamurthy K. Delivering Self-Service BI, Data Visualization, and Big Data Analytics. Intel IT: Business Intelligence; 2013.

23. Choy J, Chawla V, Whitman L. Data Visualization Techniques: From Basics to Big Data with SAS Visual Analytics. SAS: White Paper; 2013.

24. Ganore P. Need to know what Big Data is? ESDS—Enabling Futurability. 2012.

25. Agrawal D, Das S, El Abbadi A. Big Data and cloud computing: current state and future opportunities. In: Proceedings of the 14th International Conference on Extending Database Technology, ACM; 2011. pp 530–3 (2011).

26. Kaur M. Challanges and issues during visualization of Big Data. Int J Technol Res Eng. 2013;1:174–6.

27. Childs H, Geveci B, Schroeder W, Meredith J, Moreland K, Sewell C, Kuhlen T, Bethel EW. Research challenges for visualization software. Computer. 2013;46:34–42.

28. Press G. 12 Big Data definitions: what's yours? Forbes; 2014.

29. Dutcher J. What is Big Data? Berkley School of Information; 2014.

30. Bashour N. The Big Data Blog, Part V: Interview with Dr. Ivo Dinov. 2014. http://www.aaas.org/news/ big-data-blog-part-v-interview-dr-ivo-dinov.

31. Komodakis N, Pesquet JC. Playing with duality: an overview of recent primal-dual approaches for solving largescale optimization problems. arXiv preprint. 2014. http://arXiv:1406.5429.

32. Manicassamy J, Kumar SS, Rangan M, Ananth V, Vengattaraman T, Dhavachelvan P. Gene suppressor: an added phase towards solving large scale optimization problems in genetic algorithm. Appl Soft Comp; 2015.

33. Gartner — IT Glossary. Big Data defintion. http://www.gartner.com/it-glossary/big-data/.

34. Sicular S. Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s, Gartner, Inc. Forbes; 2013.

35. Demchenko Y, De Laat C, Membrey P. Defining architecture components of the Big Data Ecosystem. In: Proceedings of International Conference on Collaboration Technologies and Systems (CTS), IEEE; 2014. pp 104–12 .

36. Akerkar R. Big Data computing. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2013.

37. Sethi IK, Jain AK. Artificial neural networks and statistical pattern recognition: old and new connections, vol. 1. New York: Elsevier; 2014.

38. Araghinejad S. Artificial neural networks. In: Data-driven modeling: using MATLAB in water resources and environmental engineering. Netherlands: Springer; 2014. pp 139–94.

39. Larose DT. Discovering knowledge in data: an introduction to data mining. Hoboken, NJ: John Wiley & Sons; 2014.

40. Maren AJ, Harston CT, Pap RM. Handbook of Neural Computing Applications. Academic Press; 2014.

41. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.

42. McCue C. Data mining and predictive analysis: intelligence gathering and crime analysis. Butterworth-Heinemann; 2014.

43. Rudin C, Dunson D, Irizarry R, Ji H. Laber E, Leek J, McCormick T, Rose S, Schafer C, van der Laan M et al. Discovery with data: leveraging statistics with computer science to transform science and society. 2014.

44. Cressie N. Statistics for spatial data. Hoboken, NJ: John Wiley & Sons; 2015.

45. Lehnert WG, Ringle MH. Strategies for natural language processing. Hove, United Kingdom: Psychology Press; 2014.

46. Chu WW, editor. Data mining and knowledge discovery for Big Data. Studies in Big Data, vol. 1. Heidelberg: Springer; 2014.

47. Berry MJ, Linoff G. Data mining techniques: for marketing, sales, and customer support. New York: John Wiley & Sons; 1997.

48. PhridviRaj M, GuruRao C. Data mining-past, present and future-a typical survey on data streams. Procedia Technol. 2014;12:255–63.

49. Zaki MJ, Meira W Jr. Data mining and analysis: fundamental concepts and algorithms. Cambridge: Cambridge University Press; 2014.

50. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems; 2014. 3104–12.

51. Rojas R, Feldman J. Neural networks: a systematic introduction. Springer; 2013.

52. Gurney K. An introduction to neural networks. Taylor and Francis; 2003.

53. Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning. Adaptive computation and machine learning series: MIT Press; 2012.

54. Murphy KP. Machine learning: a probabilistic perspective. Adaptive computation and machine learning series. MIT Press; 2012.

55. Alpaydin E. Introduction to machine learning. Adaptive Computation and Machine Learning Series. MIT Press; 2014.

56. Vetterli M, Kovačević J, Goyal VK. Foundations of signal processing. Cambridge University Press; 2014.

57. Xhafa F, Barolli L, Barolli A, Papajorgji P. Modeling and Processing for Next-Generation Big-Data Technologies: With Applications and Case Studies. Modeling and Optimization in Science and Technologies: Springer; 2014.

58. Giannakis GB, Bach F, Cendrillon R, Mahoney M, Neville J. Signal processing for Big Data. Signal Process Mag IEEE. 2014;31(5):15–6.

59. Shneiderman B. The big picture for Big Data: visualization. Science. 2014;343:730.

60. Marr B. Big Data: using SMART Big Data. Analytics and Metrics To Make Better Decisions and Improve Performance: Wiley; 2015.

61. Minelli M, Chambers M, Dhiraj A. Big Data, big analytics: emerging business intelligence and analytic trends for today's businesses. Wiley CIO: Wiley; 2012.

62. Puget JF. Optimization Is Ready For Big Data. IBM White Paper (2015)

63. Poli R, Rowe JE, Stephens CR, Wright AH. Allele diffusion in linear genetic programming and variable-length genetic algorithms with subtree crossover. Springer; 2002.

64. Langdon WB. Genetic programming and data structures: genetic programming + data structures = Automatic Programming!, vol. 1. Springer; 2012.

65. Poli R, Koza J. Genetic programming. Springer; 2014.

66. Kothari DP. Power system optimization. In: Proceedings of 2nd National Conference on Computational Intelligence and Signal Processing (CISP), IEEE; 2012; pp 18–21.

67. Moradi M, Abedini M. A combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems. Int J Elect Power Energ Syst. 2012;34(1):66–74.

68. Engelbrecht A. Particle swarm optimization. In: Proceedings of the 2014 Conference Companion on Genetic and Evolutionary Computation Companion, ACM; 2014. pp 381–406.

69. Melanie M. An introduction to genetic algorithms. Cambridge, Massachusetts London, England, Fifth printing; 1999. p 3.

70. Kitchin R. The data revolution: big data, open data. Data infrastructures and their consequences. SAGE Publications; 2014.

71. Pébay P, Thompson D, Bennett J, Mascarenhas A. Design and performance of a scalable, parallel statistics toolkit. In: Proceedings of International Symposium on Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), IEEE; 2011. pp 1475–84.

72. Bennett J, Grout R, Pébay P, Roe D, Thompson D. Numerically stable, single-pass, parallel statistics algorithms. In: International Conference on Cluster Computing and Workshops, IEEE; 2009. pp 1–8.

73. Lake P, Drake R. Information systems management in the Big Data era. Advanced information and knowledge processing. Springer; 2015.

74. Anselin L, Getis A. Spatial statistical analysis and geographic information systems. In: Perspectives on Spatial Data Analysis, Springer; 2010. pp 35–47.

75. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis, vol. 344. John Wiley and Sons; 2009.

76. Anderberg MR. Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks, vol 19, Academic press; 2014.

77. Hastie T, Tibshirani R, Friedman J. Unsupervised Learning. Springer; 2009.

78. Fisher DH, Pazzani MJ, Langley P. Concept formation: knowledge and experience in unsupervised learning. Morgan Kaufmann; 2014.

79. McKenzie M, Wong S. Subset selection of training data for machine learning: a situational awareness system case study. In: SPIE Sensing Technology + Applications. International Society for Optics and Photonics; 2015.

80. Aggarwal CC. Data classification: algorithms and applications. CRC Press; 2014.

81. Ryan TP. Modern regression methods. Wiley Series in Probability and Statistics. John Wiley & Sons; 2008.

82. Zhang C, Zhang S. Association rule mining: models and algorithms. Springer; 2002.

83. Cleophas TJ, Zwinderman AH. Machine learning in medicine: part two. Machine learning in medicine: Springer; 2013.

84. Bishop CM. Pattern recognition and machine learning. Springer; 2006.

85. Devroye L, Györfi L, Lugosi G. A probabilistic theory of pattern recognition, Vol. 31, Springer; 2013.

86. Powers DM, Turk CC. Machine learning of natural language. Springer; 2012.

87. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Mining Text Data, Springer; 2012. p. 415–63.

88. Polikar R. Ensemble learning. In: Ensemble Machine Learning, Springer; 2012. p. 1–34.

89. Zhang C, Ma Y. Ensemble machine learning. Springer; 2012.

90. Helstrom CW. Statistical theory of signal detection: international series of monographs in electronics and instrumentation, Vol. 9, Elsevier; 2013.

91. Shumway RH, Stoffer DS. Time series analysis and its applications. Springer; 2013.

92. Akaike H, Kitagawa G. The practice of time series analysis. Springer; 2012.

93. Viswanathan R. Data fusion. In: Computer Vision, Springer; 2014. p. 166–68.

94. Castanedo F. A review of data fusion techniques. Sci World J 2013. 2013.

95. Thompson D, Levine JA, Bennett JC, Bremer PT, Gyulassy A, Pascucci V, Pébay PP. Analysis of large-scale scalar data using hixels. In: Proceedings of Symposium on Large Data Analysis and Visualization (LDAV), IEEE; 2011. p. 23–30.

96. Report: Data Visualization Applications Market Future Of Decision Making Trends, Forecasts And The Challengers (2014–2019). Mordor Intelligence; 2014.

97. SAS: Data visualization: making big data approachable and valuable. Market Pulse: White Paper (2013)

98. Simon P. The visual organization: data visualization, Big Data, and the quest for better decisions. John Wiley & Sons; 2014.

99. Kaisler S, Armour F, Espinosa JA, Money W. Big Data: issues and challenges moving forward. In: Proceedings of 46th Hawaii International Conference on System Sciences (HICSS), IEEE; 2013. p. 995–1004.

100. Tole AA, et al. Big Data challenges. Database Syst J. 2013;4(3):31–40.

101. Chen M, Mao S, Zhang Y, Leung VC. Big Data: related technologies. Challenges and future prospects: Springer; 2014.

102. Miksch S, Aigner W. A matter of time: applying a data-users-tasks design triangle to visual analytics of time-oriented data. Comp Graph. 2014;38:286–90.

103. MiiIIer W, Schumann H. Visualization method for time-dependent data—an overview. In: Proceedings of the 2003 Winter Simulation Conference, vol. 1. IEEE; 2003.

104. Telea AC. Data visualization: principles and practice, Second Edition. Taylor and Francis; 2014.

105. Wright H. Introduction to scientific visualization. Springer; 2007.

106. Bonneau GP, Ertl T, Nielson G. Scientific Visualization: The Visual Extraction of Knowledge from Data. Mathematics and Visualization: Springer; 2006.

107. Rosenblum L, Rosenblum LJ. Scientific visualization: advances and challenges. Policy Series; 19. Academic; 1994.

108. Ware C. Information visualization: perception for design. Morgan Kaufmann; 2013.

109. Kerren A, Stasko J, Fekete JD. Information Visualization: Human-Centered Issues and Perspectives. LNCS sublibrary: Information systems and applications, incl. Internet/Web, and HCI. Springer; 2008.

110. Mazza R. Introduction to information visualization. Computer science: Springer; 2009.

111. Bederson BB, Shneiderman B. The Craft of Information Visualization: Readings and Reflections. Interactive Technologies: Elsevier Science; 2003.

112. Dill J, Earnshaw R, Kasik D, Vince J, Wong PC. Expanding the frontiers of visual analytics and visualization. SpringerLink: Bücher. Springer; 2012.

113. Simoff S, Böhlen MH, Mazeika A. Visual data mining: theory, techniques and tools for visual analytics. LNCS sublibrary: Information systems and applications, incl. Internet/Web, and HCI. Springer; 2008.

114. Zhang Q. Visual analytics and interactive technologies: data, text and web mining applications: data. Information Science Reference: Text and Web Mining Applications. Premier reference source; 2010.

115. Few S, EDGE P. Data visualization: past, present, and future. IBM Cognos Innovation Center; 2007.

116. Bertin J. La graphique. Communications. 1970;15:169–85.

117. Gray JJ. Johann Heinrich Lambert, mathematician and scientist, 1728–1777. Historia Mathematica. 1978;5:13–41.

118. Tufte ER. The visual display for quantitative information. Chelshire: Graphics Press; 1983.

119. Kehrer J, Boubela RN, Filzmoser P, Piringer H. A generic model for the integration of interactive visualization and statistical computing using R. In: Conference on Visual Analytics Science and Technology (VAST), IEEE; 2012. p. 233–34.

120. Härdle W, Klinke S, Turlach B. XploRe: an Interactive Statistical Computing Environment. Springer; 2012.

121. Friendly M. A brief history of data visualization. Springer; 2006.

122. Mering C. Traditional node-link diagram of a network of yeast protein-protein and protein-DNA interactions with over 3,000 nodes and 6,800 links. Nature. 2002;417:399–403.

123. Febretti A, Nishimoto A, Thigpen T, Talandis J, Long L, Pirtle J, Peterka T, Verlo A, Brown M, Plepys D et al. CAVE2: a hybrid reality environment for immersive simulation and information analysis. In: IS&T/SPIE Electronic Imaging (2013). International Society for Optics and Photonics

124. Friendly M. Milestones in the history of data visualization: a case study in statistical historiography. In: Classification—the Ubiquitous Challenge, Springer; 2005. p. 34–52.

125. Tory M, Kirkpatrick AE, Atkins MS, Moller T. Visualization task performance with 2D, 3D, and combination displays. IEEE Trans Visual Comp Graph. 2006;12(1):2–13.

126. Stanley R, Oliveria M, Zaiane OR. Geometric data transformation for privacy preserving clustering. Departament of Computing Science; 2003.

127. Healey CG, Enns JT. Large datasets at a glance: combining textures and colors in scientific visualization. IEEE Trans Visual Comp Graph. 1999;5(2):145–67.

128. Keim DA. Designing pixel-oriented visualization techniques: theory and applications. IEEE Trans Visual Comp Graph. 2000;6(1):59–78.

129. Kamel M, Camphilho A. Hierarchic Image Classification Visualization. In: Proceedings of Image Analysis and Recognition 10th International Conference, ICIAR; 2013.

130. Buja A, Cook D, Asimov D, Hurley C. Computational methods for high-dimensional rotations in data visualization. Handbook Stat Data Mining Data Visual. 2004;24:391–415.

131. Meijester A, Westenberg MA, Wilkinson MHF. Interactive shape preserving filtering and visualization of volumetric data. In: Proceedings of the Fourth IASTED International Conference; 2002. p. 640–43.

132. Borg I, Groenen P. Modern multidimensional scaling: theory and applications. J Educ Measure. 2003;40:277–80. 133. Bajaj C, Krishnamurthy B. Data visualization techniques, vol. 6. Wiley; 1999.

134. Plaisant C, Monroe M, Meyer T, Shneiderman B. Interactive visualization. CRC Press; 2014.

135. Janvrin DJ, Raschke RL, Dilla WN. Making sense of complex data using interactive data visualization. J Account Educ. 2014;32(4):31–48.

136. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute; 2011.

137. Ebert A, Dix A, Gershon ND, Pohl M. Human Aspects of Visualization: Second IFIP WG 13.7 Workshop on HumanComputer Interaction and Visualization, HCIV (INTERACT), Uppsala, Sweden, August 24, 2009, Revised Selected Papers. LNCS sublibrary: Information systems and applications, incl. Internet/Web, and HCI. Springer; 2009. p 2011.

138. Schonlau M. Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. Comput Stat. 2004;19(1):95–111.

139. Google, Inc.: Google Visualization Guide. 2015. https://developers.google.com.

140. Amcharts.com: amCharts visualization. 2004–2015. http://www.amcharts.com/.

141. Viégas F, Wattenberg M. IBM—Many Eyes Project. 2013. http://www-01.ibm.com/software/analytics/many-eyes/

142. Körner C. Data Visualization with D3 and AngularJS. Community experience distilled: Packt Publishing; 2015.

143. Azzam T, Evergreen S. J-B PE Single Issue (Program) Evaluation, vol. pt. 1. Wiley\

144. Machlis S. Chart and image gallery: 30+ free tools for data visualization and analysis. 2015. http://www.computerworld.com/

145. Julie Steele NI. Beautiful visualization: looking at data through the eyes of experts. O'Reilly Media; 2010.

146. Guberman S. On Gestalt theory principles. GESTALT THEORY. 2015;37(1):25–44.

147. Chen C. Top 10 unsolved information visualization problems. Comp Graph Appl IEEE. 2005;25(4):12–6.

148. Johnson C. Top scientific visualization research problems. Comp Graph Applications IEEE. 2004;24(4):13–7.

149. Tory M, Möller T. Human factors in visualization research. Trans Visual Comp Graph. 2004;10(1):72–84.

150. Andrews C, Endert A, Yost B, North C. Information visualization on large, high-resolution displays: Issues, challenges, and opportunities. Information Visualization; 2011.

151. Suthaharan S. Big Data classification: problems and challenges in network intrusion prediction with machine learning. ACM SIGMETRICS. 2014;41:70–3.

152. Field DJ, Hayes A, Hess RF. Contour integration by the human visual system: evidence for a local "association field". Vision Res. 1993;33:173–93.

153. Picard RW, Healy J. Affective Wearables. Vol 1, Springer. 1997; p. 231–40.

154. Mann S et al. Wearable technology. St James Ethics Centre; 2014.

155. Carmigniani J, Furht B, Anisetti M, Ceravolo P, Damiani E, Ivkovic M. Augmented reality technologies, systems and applications. 2010;51:341–77.

156. Papagiannakis G, Singh G, Magnenat-Thalmann N. A survey of mobile and wireless technologies for Augmented Reality systems. Comp Anim Virtual Worlds. 2008;19(1):3–22.

157. Caudell TP, Mizell DW. Augmented reality: an application of heads-up display technology to manual manufacturing processes. IEEE Syst Sci. 1992;2:659–69.

158. Sutherland I. A head-mounted three dimensional display. In: Proceedings of the Fall Joint Computer Conference; 1968. p. 757–64.

159. Chacos B. Shining light on virtual reality: Busting the 5 most inaccurate Oculus Rift myths. PCWorld; 2014.

160. Krevelen DWF, Poelman R. A survey of augmented reality technologies, applications and limitations. Int J Virtual Reality. 2010;9:1–20.

161. Stevens J, Eifert L. Augmented reality technology in US army training (WIP). In: Proceedings of the 2014 Summer Simulation Multiconference, Society for Computer Simulation International; 2014, p 62.

162. Bower M, Howe C, McCredie N, Robinson A, Grover D. Augmented Reality in education-cases, places and potentials. Educ Media Int. 2014;51(1):1–15.

163. Ma M, Jain LC, Anderson P. Future trends of virtual, augmented reality, and games for health. In: Virtual, Augmented Reality and Serious Games for Healthcare vol. 1, Springer; 2014. p. 1–6.

164. Mousavi M, Abdul Aziz F, Ismail N. Investigation of 3D Modelling and Virtual Reality Systems in Malaysian Automotive Industry. In: Proceedings of International Conference on Computer, Communications and Information Technology (2014). Atlantis Press; 2014.

165. Chung IC, Huang CY, Yeh SC, Chiang WC, Tseng MH. Developing Kinect Games Integrated with Virtual Reality on Activities of Daily Living for Children with Developmental Delay. In: Advanced Technologies, Embedded and Multimedia for Human-centric Computing, Springer, 2014. p. 1091–97.

166. Steptoe W. AR-Rift: stereo camera for the Rift and immersive AR showcase. Oculus Developer Forums; 2013.

167. Pina JL, Cerezo E, Seron F. Semantic visualization of 3D urban environments. Multimed Tools Appl. 2012;59:505–21.

168. Fonseca D, Villagrasa S, Marta N, Redondo E, Sanchez A. Visualization methods in architecture education using 3D virtual models and augmented reality in mobile and social networks. Procedia Soc Behav Sci. 2013;93:1337–43.

169. Varkey JP, Pompili D, Walls TA. Human motion recognition using a wireless sensor-based wearable system. Personal Ubiquitous Comp. 2011;16:897–910.

170. Nuwer R. Armband adds a twitch to gesture control. New Scientist. 2013;217(2906):21.

171. Timberlake GT, Mainster MA, Peli E, Augliere RA, Essock EA, Arend LE. Reading with a macular scotoma. I. Retinal location of scotoma and fixation area. Investig Ophthalmol Visual Sci. 1986;27(7):1137–47.

172. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol. 2002;86:238–42.

173. Deering MF. The limits of human vision. In: Proceedings the 2nd International Immersive Projection Technology Workshop; 1998.

174. Krantz J. Experiencing sensation and perception. Pearson Education (Us); 2012.

175. Rajanbabu A, Drudi L, Lau S, Press JZ, Gotlieb WH. Virtual reality surgical simulators-a prerequisite for robotic surgery. Indian J Surg Oncol. 2014;1–3.

176. Moglia A, Ferrari V, Morelli L, Melfi F, Ferrari M, Mosca F, Cuschieri A. Distribution of innate ability for surgery amongst medical students assessed by an advanced virtual reality surgical simulator. Surg Endosc. 2014;28(6):1830–7.

177. Ahn W, Dargar S, Halic T, Lee J, Li B, Pan J, Sankaranarayanan G, Roberts K, De S. Development of a Virtual Reality Simulator for Natural Orifice Translumenal Endoscopic Surgery (NOTES) Cholecystectomy Procedure. Medicine Meets Virtual Reality 21: NextMed/MMVR21 196, 1(2014).

178. Ma M, Jain LC, Anderson P. Virtual, Augmented Reality and Serious Games for Healthcare 1. Springer; 2014.

179. Wright WG. Using virtual reality to augment perception, enhance sensorimotor adaptation, and change our minds. Front Syst Neurosci. 2014; 8.

180. Parsons TD, Trost Z. Virtual reality graded exposure therapy as treatment for pain-related fear and disability in chronic pain. In: Virtual, Augmented Reality and Serious Games for Healthcare 1, Springer; 2014. p. 523–46.

181. Abramov I, Gordon J, Feldman O, Chavarga A. Biology of sex differences, p. 1–14.

182. McFadden D. Masculinization effects in the auditory system. Archiv Sexual Behav. 2002;31(1):99–111.

183. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychol Bull. 1995;117:250–70.

184. Stancey H, Turner M. Close women, distant men: line bisection reveals sex-dimorphic patterns of visuomotor performance in near and far space. Br J Psychol. 2010;101:293–309.

185. Rizzolatti G, Matelli M, Pavesi G. Deficits in attention and movement following the removal of postarcuate (area 6) and prearcuate (area 8) cortex in macaque monkeys. Brain. 1983;106:655–73.

186. Chua HF, Boland JE, Nisbett RE. Cultural variation in eye movements during scene perception. PNA. 2005;102(35):12629–33.

187. Zelinsky GJ, Adeli H, Peng Y, Samaras D. Modelling eye movements in a categorical search task. Phisiol Trans Royal Soc. 2013.

188. Piumsomboon T, Clark A, Billinghurst M, Cockburn A. user-defined gestures for augmented reality. In: HumanComputer Interaction–INTERACT 2013, Springer; 2013. p. 282–99.

189. Mistry P, Maes P, Chang L. WUW-wear Ur world: a wearable gestural interface. In: Extended Abstracts on Human Factors in Computing Systems, ACM; 2009. p. 4111–16.

190. Vanacken D, Beznosyk A, Coninx K. Help systems for gestural interfaces and their effect on collaboration and communication. In: Workshop on Gesture-based Interaction Design: Communication and Cognition (2014)

191. Mulling T, Lopes C, Cabreira A. Gestural interfaces touchscreen: thinking interactions beyond the Button from Interaction Design for Gmail Android App. In: Design, sser Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments, Springer, 2014; p. 279–88.

192. Piumsomboon T, Clark A., Billinghurst M. [DEMO] G-SIAR: gesture-speech interface for augmented reality. In: Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR), IEEE; 2014. p. 365–66.

193. Vafadar M, Behrad A. A vision based system for communicating in virtual reality environments by recognizing human hand gestures. Multi Tools Appl. 2014;1–21.

194. Roupé M, Bosch-Sijtsema P, Johansson M. Interactive navigation interface for Virtual Reality using the human body. Comp Environ Urban Syst. 2014;43:42–50.

195. Wen R, Tay WL, Nguyen BP, Chng CB, Chui CK. Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. Comp Method Program Biomed. 2014.

196. Rolland JP, Fuchs H. Optical versus video see-through head-mounted displays in medical visualization. Presence Teleoperators Virtual Environ. 2000;9(3):287–309.

197. Silanon K, Suvonvorn N. Real time hand tracking as a user input device. Springer; 2011. p. 178–89.

198. Keim DA, Mansmann F, Schneidewind J, Ziegler H. Challenges in visual data analysis. In: Proceedings of 10th International Conference on Information Visualization, IEEE; 2006. p. 9–16.

199. Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision. Cengage Learning. 2014.

200. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of IEEE Symposium on Visual Languages; 1996. p. 336–43.

201. Coffey D, Malbraaten N, Le T, Borazjani I, Sotiropoulos F, Keefe DF. Slice WIM: a multi-surface, multi-touch interface for overview+ detail exploration of volume datasets in virtual reality. In: Proc. of Symposium on Interactive 3D Graphics and Games, ACM. 2011. p. 191–98.