Visual Analytics

Puyu Yang, Tuo Zhang, Fantuo Meng
Electrical and Computer Engineering
Western University
London, Ontario, Canada
Pyang64@uwo.ca, Tzhan332@uwo.ca, Fmeng33@uwo.ca,

***Abstract* – Excellent representation of data is always an essential step for data science to help the audiences to get an insight of data from the large explosive increasing size of information. This paper aims to help readers to understand the underlying and comprehensive knowledge of Visual Analytics(VA) from different perspectives. Those knowledge include definitions of Visual Analytics, the process of Visual Analytics, the scope of Visual Analytics and its related fields. Meanwhile, an ordinary confusion between Visual Analytics and Visualization is also discussed. Also, several popular Visual Analytics tools are also presented with specific using situation. Afterwards, we evaluate Visual Analytics from advantage side and challenge side. Lastly, a bioinformation Visual Analytics software is introduced.**

Keywords: Visual Analytics, Visualization, Perception, Cognition

I. Introduction and Motivation

With the rapid development of Internet and computer industry, our world has become an information-driven world, not limited to computer science and software engineering fields, like bioinformatics, business and politics, even everyone produces large amounts of data. Notably, in recent twenty years, the growths of data acquisition, data storage, data analytics and data processing technologies reduce the cost and time of data processing and storage. The acquisition of raw data is no longer the main problem, but the ability is still not sufficient enough to analyze those increasing information completely and deeply [1]. However, data is just managed and stored without filtering and refinement for later use. This phenomenon leads to the information overload problem that roughly tons of the raw data like an undiscovered gold mine currently have no value at all [1]. Hence, Data Analytic is the tool to discover this fortune.

To deal with the overload problem of information, Data Analytic needs to exploit and make use of to dig the knowledge hidden behind a significant amount of raw data. To achieve this goal, Data Analytics must provide answers to the following questions: [1]

- Is the information relevant and enough to support the given task?
- Is the information processed appropriately?
- Is the information presented appropriately?

Today, a series of software and programming languages appearing to help data analysts to get useful information by getting the overview of the ever-increasing data [2]. Apparently, this is much less than enough.

What we need is to find a way to interact appropriately with the data source, communicate the knowledge efficiently and represent the information to the audience [1]. **Visual Analytics** is the bridge between those gaps.

Visual Analytics aims to turn the information overload problem into an opportunity by presenting the filtered information and providing the interaction between individuals and the enormous amount of big data [2]. Thus, Decision makers can examine and interact with the extensive data information from various sources in different aspects [2]. Then, this ability would combine with the human factors like Perception, Cognition, background knowledge and related experiences to help perform better in decision and conclusion making [2].

We organize the rest of the paper as follows: Visual Analytics and its related fields would be defined and clarified in section 2. The process of the Visual Analytics discussed in section 3. Some Visual Analytics tools and libraries stated in section 4, and the evaluation of Visual Analytics is introduced from two perspectives in section 5. A real-world example of Visual Analytics software is introduced in section 6.

II. Definition and Related Fields

In general, Visual Analytics is "the science of analytical reasoning facilitated by interactive visual interfaces" [3]. The task for Visual Analytics is to corporate with human beings by providing the technology and information. To be more precise and specific, Daniel Keim's definition is "Visual Analytics combines the automated analysis techniques with interactive visualization for an effective understanding, reasoning, and decision making on the basis of very large and complex data sets" [1].

Nowadays, Visual Analytics has developed from just gathering data and drawing graphs. More importantly, Visual Analytics focuses more on how to accurately represent the data and the deep meaning, feature, and potential on the computer screen. However, unlike other computation fields like Data Mining and Machine Learning, the performance and effectiveness of the data transformation would not automatically improve with the continuous increasing computation speed. Still, there are a lot of methods and representation formats need to be chosen under consideration by the analysts [2].

To sum up, we could conclude the goal of Visual Analytics could as follows: Firstly, Visual Analytics needs to provide humane methods to dive into the vast and conflicting data and information to get insight. Secondly, Visual Analytics should have the ability to verify what we have already known and been able to discover the potential as we talked earlier. Lastly, Visual Analytics aims to have timely, defensible and understandable assessments and also could be used to communicate with decision-makers and audience [4].

From figure one, we could indicate that Visual Analytics is a combination of various related fields more than just Visualization; at the same time, Visual Analytics integrates Data Management, Data Mining, Human-Computer Interaction, Perception and Cognition. These five parts affect each other and contribute to the integration and design of Visual Analytics. It is important to combine those different areas to solve the real-world problem. Also, this could help to build and develop the efficient and reliable Evaluation methodology [1]. Additionally, for each piece that related to Visual Analytics is not likely to become a separate field or area, on the contrary, the influence of Visual Analytics will spread widely instead. In the following part, we will introduce each piece related to Visual Analytics that mentioned above.

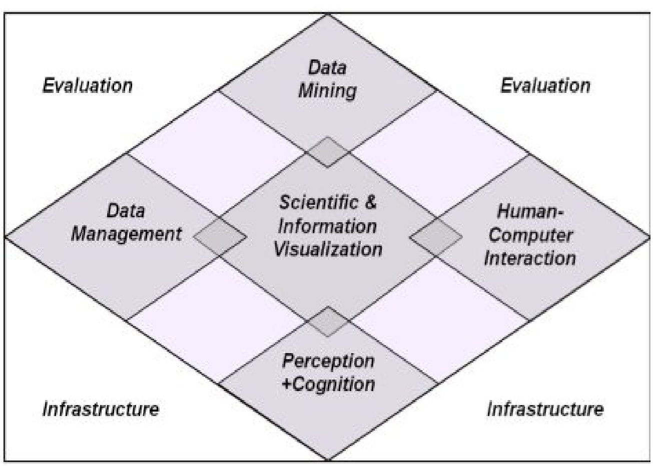

Fig. 1. Visual Analytics integrates several scientific fields [1].

A. Data Management

Data Management refers to the organization, catalog, location, storage, retrieval and maintenance of data, which is the target of data processing. The whole process consists of collecting, organizing, storing, processing, transmitting and retrieving different kinds of data. It is an essential field of computer application. The target of Data Management is to ensure the accessibility, reliability, and timeliness of the data. One of its aims is to extract and deduce valuable information from many raw data. The other goal is to use computer technology to preserve and manage complex, large amounts of data so that people can make use of the data. Moreover, a necessary precondition to performing any kinds of data analysis is an integrated and continuous data basis [5]. After data pre-processing steps like data cleaning, filtering and clustering the raw data, data management is a fundamental step to get the valid data that meets the requirements for the further analysis and processing to achieve the ultimate goal of interaction and Visual Analytics [1].

B. Data Mining

Data Mining, also called Data Analysis or Knowledge Discovery, is a step of Knowledge Discovery in Databases (KDD) and refers to the process of extracting and summarizing useful information through means of automatic analysis algorithms from the raw data [6]. Data Mining is associating with the statistics and online analytical processing, information retrieval, machine learning, expert system (depend on the past rule of thumb) and pattern recognition. The goal of Data Mining is to extract information from a data set and convert it into a new structure for further study. In addition to the original analysis steps, Data Mining also related to database and Data Management, data pre-treatment, model and deduce considerations, interest measurement, and complexity. To sum up, Data Mining is the process of extracting hidden prediction information and automatically discovers patterns from the dataset.

Therefore, Data Mining and Visual Analytics benefit from each other. Data Mining reduces redundancy and filters raw data for Visual Analytics and have a better understanding for users.

C. Human-Computer Interaction

Human-Computer Interaction (HCI), which is also called human-machine interface in manufacturing or process control system, focuses on human and machine interaction patterns among multidisciplinary research areas. In other words, the HCI discipline focuses on the problems of interface design and implementation between people and computers. Due to its nature and goal, HCI involves multiple disciplines of fields of computer science (image processing, computer vision and programming languages) and humanities (human body engineering, people, cognitive psychology and human factors engineering) [7].

It is well known that the purpose of Visual Analytics is to combine the advantage of computers and human to extract valuable knowledge from the data. To make users participate in the whole analysis process, developers need to develop an efficient user interface that minimizes the barrier between the human’s cognitive model and the computer’s view [8]. Therefore, human can improve their capabilities of understanding data, and the machine can not solve tasks without the human.

D. Perception and Cognition

Perception, with a wide range of definitions, is the reflection of objective things in human brain through the sensory system. All perception activities involve signals that go across the nervous system, which in turn causes the changes of physical and chemical stimulation of the body. Typically, there are two processes of perception: First, dealing with sensory signals, which transforms the low-level message to high-level message Second, perception processes which connect with a person’s knowledge affect themselves.

Cognition, which involves in processes such as knowledge gaining, evaluation, reasoning, and problem-solving, are the mental actions of getting required information through thought, experience, and sensory system. Usually, human uses existing knowledge and generates new knowledge in the cognitive process when remembering things from outside world.

When facing visualization system, users take advantage of a variety of perceptual methods to interact with the user interface to get enough information they care. The understanding and evaluating of the visualization systems are closely related to the cognitive characteristics of the human [9]. Research in this field is based on subjects, such as psychology, neuroscience and requirement engineering.

To design a visualization system with perfect functions, the software engineer should pay more attention to visual cognition and memory cognition. User's visual cognition is an important design factor, as visual is the primary method for the human to obtain information from the user interface. The factors of visual perception include two parts. The first part is the cognitive characteristics of human eyes to color and brightness; the second part is the resolution of time and spatial when facing different backgrounds. For example, the user will have calm mind if the background with cool color and small changes in light contrast. Memory cognition is also related to whether the user can efficiently master the usage of the system. The cognitive memory includes sensory memory, long-term memory and short-term memory [10]. These three kinds of memory play different roles when human remember things in real life. If system designer can plan the layout of the data visualization system according to various memory types of human, it will further reduce the user's’ memory burden.

On the technical side, the research of this area is influenced by two parts: (1) The relationship between data of required resources and display background (2) Improvement of an interactive algorithm to help the user remember the characters of data with less memory burden [11]. To deal with these two technical difficulties, perception-theory-based solutions for the statistics and image representations is proposed to let users adapt to data visualization automatically and effectively.

E. Visualization

As a new research discipline during the last two decades, visualization has evolved into two more specific fields, Scientific Visualization and Information Visualization [2]. In Scientific Visualization, the scientists obtain a massive amount of data from sensors, simulations or laboratory tests. Those data can usually be mapped into geographic coordinates or virtual 3D environments [1]. Being an essential aid for people to understand science, Scientific Visualization is widely applied in fields like aerospace engineering, natural science, and Biomedical.

Information Visualization is a more general term when people refer to Data Visualization. When it comes to the definition of Information Visualization, there are apparently a lot of possible answers to that. Firstly, Information Visualization is the process that can interpret and represent abstract raw data in visual terms [1]. To be more specific, Information Visualization is a kind of structured information representation form that takes data and figures out ways to translate into visual terms that are more readable and typically being applied in business, news and demographic fields [1]. Before visualizing, the raw data always includes hundreds of dimensions and lacks the natural mapping to the screen. Exploratory Data Analysis is born to deal with this problem by searching and analyzing databases to find implicitly useful information [2]. Therefore, different structured data could render in different visualization that we can see nowadays.

F. Visualization V.S. Visual Analytics

It is common that people are confused about the terms of **Visual Analytics** and **Visualization**. Unfortunately, people misuse them for most of the time. Indeed, these two areas are related to each other and both are based on scientific analytics methodology, Human-Computer Interaction, cognitive and perceptual science. Also, both of those fields build upon methods from scientific analytics, geospatial analytics and information analytics [2].

TABLE I indicates the differences between two “similar” conceptions from aspects of target and scope. The significant difference is that traditional Information Visualization does

TABLE I
DIFFERENCE BETWEEN VISUAL ANALYTICS AND VISUALIZATION

	Visual Analytics	**Visualization**
Target	Combines various analytical and intelligent methods to deal with Visualization.	Provides a view and interaction interface, does not involve data analysis.
Scope	HCI, Perception, Cognition, Data Analysis, Data Management and Visualization**.**	HCI, Perception and Cognition.

not involve the analysis tasks and algorithms. To be more specific, Visualization focuses more on producing the view and interaction interface from given information [1]. Visualization is used as a method to efficiently communicate and explore the information space [2].

On the contrary, Visual Analytics is the combination of Visualization and data analysis, especially human factors like interaction, cognition and perception, more than just Visualization [2]. Compared with Visualization, Visual Analytics focuses more on the intelligence that takes the user interaction as a parameter for the analytic process, which is the primary component and needs to integrate with the areas discussed earlier as well as visualization, human factors and decision-making process [1].

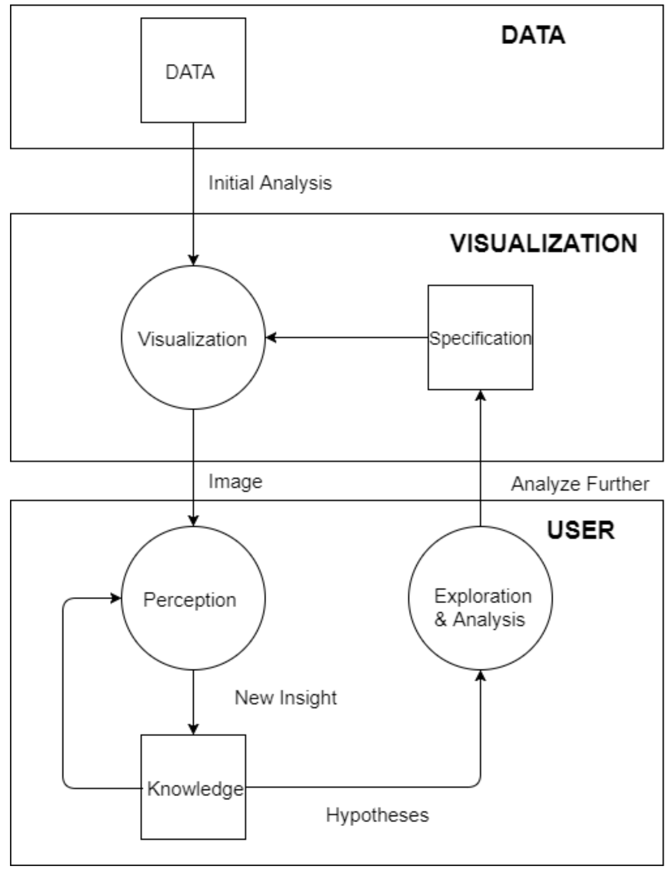

Fig. 2. Visual Analytics Process [1].

III. Visual Analytics Process

Visual Analytics is a process regarded as a whole data science process and is built based on the framework provided by analytical reasoning [4]. To assist human making assessment and decision, visual analysts should apply human judgment to conclude from a combination of evidence and assumptions [4].

To be more specific, the whole Visual Analytics process is the combination of analytical process and interactive visualization techniques [2]. To better describe this process, Keim rewrites a mantra to bring its focus toward Visual Analytics:'Overview first, Filter and zoom, Details on demand' to 'Analyze first, show the Important, zoom, filter and analyze further, Details on demand' [2]. This mantra means Visual Analytics need to allow users to enter a loop in each iteration. They can gain some insight and knowledge by manipulating and filtering, and then the visual representation renders again with less complexity but more useful data.

To better represent the whole process, van Wijk depicts it with a flowchart [12]. The process is about entire data science process. Firstly, we need the acquiring and wrangling means to get data through activities like web scraping, log collection, and database queries, which often takes the most time. In the initial analysis stage, the activities related to data science includes modeling, data analysis, data management and applying statistical and mathematical theory to discover insights. Then, data can be represented in visual form and mainly focus on how to describe the data graphically in an efficient manner. Next, the user enters a loop that they can get some knowledge and insight from the perception while that new knowledge would also aid in gaining new understanding. Afterwards, further automatic analysis like machine learning, statistics as well as Data Mining with more precise specification transforms the visualization to the original form which relies on by the system's interaction that enables the audience to potentially discover insights for themselves to confirm or make a new hypothesis. In each iteration, the audience may understand better what they have done in the previous iteration (e.g., realizing too many points to make sense, then go back and add more filters to data). After several iterations of these process, the user would ultimately confirm hypotheses made from previous iterations [2].

IV. Visual Analytics Tools And Libraries

We could compare Visual Analytics tools in a pyramid (Fig.3), and each tool has its strength area. Tools at the bottom of the pyramid are efficient and flexible when

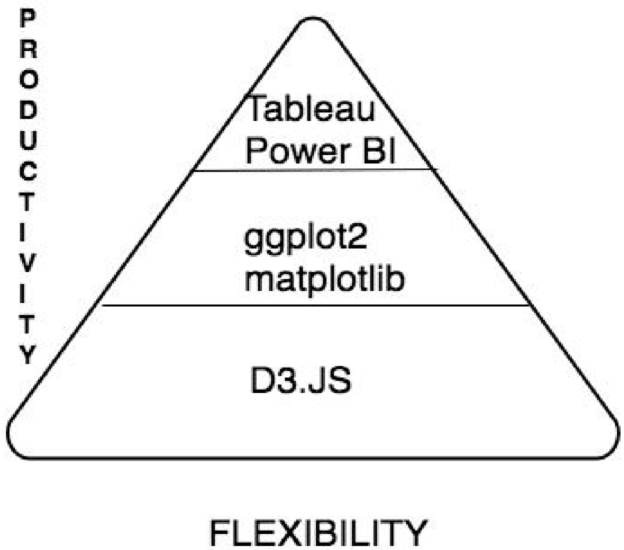

Fig. 3. Graphic tools' comparison of productivity and flexibility

choosing to create the visualization, but they have a low level of abstraction and overhead for learning and developing and vice versa. Users should select tools wisely depends on their requirements. The following part will show these tools sequentially from top to bottom (Fig.3).

A. Tableau

Tableau is created mainly for providing fast and easy-understand Visual Analytics. Without any advanced training, users could be familiar with all kinds of operations about creating interactive reports, dashboards, and visualizations [13]. Tableau is one of the simplest business intelligence tools without forcing users to write custom code and using the new console to change configurations. It could not only monitor information but also provides a complete analysis. Therefore, Tableau has powerful functions like Data connection, Data refreshment, collaboration, security management and multi-platform. It is convenient for Tableau Desktop to connect with Salesforce, databases, Google analytics, and JSON format data. For Data refreshment, it implements the update from metadata automatically. Apart from this, it can easily control users from its terminal and implement cross-platform in web pages, mobile phones, and tablets.

B. D3.js

Before 2011, the creation of visualizations lacks appropriate tools that can create vivid graphs. Things changed with the release of D3.js in 2011 that is an original representation-transparent method to visualization [14]. D3.js is the abbreviation of 'Data Driven Document' which is a JavaScript library deeply tied to HTML, JavaScript, CSS and SVG to display lively diagrams for users. HTML, CSS and JavaScript represent content, while SVG are for vector graphics [14]. D3 could bind raw data to a document object model (DOM) which shows the content and structure of the page, and visualize data by matching numeric values to attributes of visual elements, then transform document synchronously. The most powerful feature of d3 is data processing like mapping the data to graphics, data transformation, and interpolation calculation. It is also a free and open source that would have a burgeoning support for community growing. From technical part, D3 strictly follows Web standards to allow users easily to make programs compatible with modern mainstream browsers and avoid dependence on specific frameworks.

C. ggplot2/Matplotlib

Wilkinson proposed the conception of grammar graphics that was mainly to explain the conception of statistical graphic, and then Wickham made some adjustment based on the theory of grammar of graphics in R language and developed R ggplot2 drawing system [15]. ggplot2 in R language has the same role with Matplotlib in python. Matplotlib is the ancestor of python data visualization packages. Matplotlib is powerful but complex because there are many tools (especially the Pandas and Seaborn) encapsulated on Matplotlib like ggplot2. The whole frame of ggplot2 contains seven modules: Data, Aesthetic Theme, Geometries, Facets, Statistics, coordinates and Theme. Data represents the original data of data frame; aesthetic means the relationship of initial data; geometries indicate the library of drawing graphics; facets show how to distribute the data in multiple diagrams to draw and compare. In the end, using the layer function to combine these modules to construct graphic and show to users.

To sum up, Tableau and Power BI are the most productive tools with many predefined tools and layouts that could be implemented without programming, which are suitable for experienced and simpler tasks as well as non-programmers. On the contrary, D3.js is suitable for complex and less-experienced problems. It is essential to have strong programming (JavaScript, HTML, SVG) background to use D3.js. Also, D3.js is the most flexible tools for being just a library. There is no limitation for designers with JavaScript language supported. Additionally, the tools built by D3.js are hosted on the Internet everywhere like New York Times. For the tools in the middle like ggplot2 and matplotlib, they are used for visualizing the analysis results that required some programming base in R and Python, but still in a defined manner and less flexible than D3.js.

V. Evaluation

Evaluation of Visual Analytics is an important part of Visual Analytics system. The information gained from evaluation work is an important message for both designers and users. As mentioned before, Visual Analytics aims to combine and integrate the advantages of software and users into a process to transfer knowledge from raw data. In the meantime, the challenge of Visual Analytics is still a tough work for analysts and system designers.

A. Advantage of Visual Analytics

In this section, the discussion of advantage is illustrated from user and designer perspectives. As a bridge between users and data analysts, Visual Analytics can convey the law of data rather than simple figure to audiences. Besides, it can take advantage of vision and human memory to assist audiences in understanding and remembering data features.

Compared with traditional data science methods, Visual Analytics can convey more information to users and reduce the burden of them by taking advantage of the vision system and providing an overview of data features. Visual Analytics can also help audiences better understand the features of data structures and functions.

1). Convey more information than summary statistics: Nowadays, Visual Analytics is not just the substitution of summary statistics. Instead, it can spread more information than only statistics. In 1973, F. J. Anscombe constructed four sets of strange data [16]. The table two provide the whole dataset for this experiment. After calculating, these datasets share the same mean value, variance, correlation coefficient and linear regression model for x and y. For audiences, those data should look the same in the plot without deep thinking.

TABLE II
ANSCOMBE'S QUARTET DATASET [15]

A		B		C		D	
X	Y	X	Y	·X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8,14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12 .5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

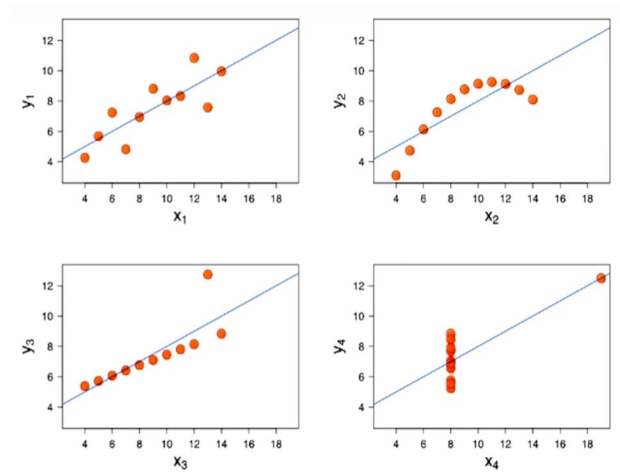

Fig. 4. Plot of Anscombe's quartet [15]

After drawing these four sets of data in the chart (Fig. 4), data analysts find they showed completely different situations. The reason is that summary statistics cannot cover

the curvature, outlier and trends. Observers could get more information than summary data from graphs. In the first plot, x and y have a weak linear relationship. While in the second plot, x and y have parabolic curve regression relationship. However, in the third plot, x and y show the strong linear regression with an outlier y value when x equals to 13. In the last plot, all value of x equals to 8 except one value at 19.

From this example, Visual Analytics can directly show the hidden relationship of data to audiences. If only showed the table of four sets of data, it is hard to find the relationship between them. The audiences may even have an illusion that these data share the same tendency and figures. Compared with traditional data science, Visual Analytics can present data in a more intuitive way by making the data more objective and persuasive.

2). Eyes accept much more information than other organs: Visual Analytics takes advantage of the critical position of vision in the cognitive system to help the audience to memorize data characteristics. The information caught from vision is far more than other sensory organs, and visualization system of Visual Analytics works in the same way.

According to the classification of perception, vision is the only one kind which helps human beings distinguish shape, size, color, space state and other properties of different objects.

The distinction of color, shape, size and space state which cannot be achieved by other perceptions except vision, play essential roles in object cognition of human life. As mentioned above, the percentage of external information obtained from the vision accounts for 80%, which is larger than any other sensory organs [17]. The target of Visual Analytics is to convey messages taking benefit of human version through graphical methods.

On the other hand, this feature could cause some side effects on the human cognition on receiving messages. It is essential to understand how human perceives the 3D world from the 2D image on our retina. When looking at a 3D image, eyes are going to be sensing data which will be processed by a perceptual processor. Typically, the perceptual processor is developed to understand the 3D world from three perspectives, including foreshortening, Linear and size constancy. Foreshortening indicates when showing a 3D cube, the edges in the depth direction are going to be shorter than the edges in the transverse directions, even though in the three-dimensional scene they are the same length [18]. Linear perspective explains that object further away appears smaller than we can see. The Early artist used this aspect to create an object from a long distance. The last one is size constancy, which indicates that objects do not change size, so smaller objects must farther away than larger objects. When we watch a 3D image, our perception of size is based on the three perspectives above to illustrate how far the object is away from us. Thus, the 3D image sometimes misleads us what the actual size of the object [19]. Usually, in Visual Analytics, 3D images are discouraged, as data analysts avoid producing illusion on audiences.

3). Human has some limitation on memories: There are three kinds of human memory: short-term memory, long-term memory, and working memory. Working memory contains information that is short, whereas, long-term memory is used to remember things for a long time through a process of studying [20]. Perceptual and cognitive processors connect working memory and long-term memory. While paying attention to one object in short time, short-term memory is used to store information. Nerves system will process the data through processors and put them in working memory. If one wants to remember something and keep details in mind, perception system will take data from working memory and put them in long-term ones.

Usually, Visual Analytics collects data characteristics of one image, which taking advantage of sensory image storage to reduce the burden of the brain by giving an overview and more clear structure. In this way, brain just needs to remember only one memory block rather than calling long-term memory. According to the research of Gorban, Alexander, and Donald, compared to the traditional information display, Visual Analytics helps receivers to form exciting studying experience with a better understanding of more memorable, accessible and comprehensible content [21].

4). Interaction enhances the process of Analytics: One of most significant differences between Visual Analytics and automated analysis is that Visual Analytics highlights interaction through the interactive user interface which could help users to have a better understanding of the data from different perspectives in the whole analysis process. A significant advantage of this method is that analytical methods of Data Mining and Machine Learning algorithms are combined with visualization techniques to analyze data more accurately and efficiently with new and complex problems. On the contrary, the fully automated analysis only works well with obvious and well-understand issues, but still has the problem of communicating and interpreting efficiently [22].

Another advantage of Visual Analytics is user participation could play an essential role in the analysis process. Analysts can discover (possibly unexpected) patterns, trends and correlations from vast, elaborate data sets with provided a flow of control over the whole process of analysis by the interactive user interface [22]. Hence, users could gain greater confidence and interests to explore data by themselves through performing further(visual/automatic) analysis [22], which would also incentivize them to understand better.

B. Challenge of Visual Analytics

The challenges for Visual Analytics focus on analyzing raw data and building effective visualization for targeted audiences. At the same time, data analysts and system designers should transfer accurate information and avoid using misleading visualization designs. In this section, challenges are illustrated according to technical and application aspects.

1). *Technical Challenge:* The primary target of Visual Analytics is to analyze a vast amount of data and display the result, which eventually helps users to make decisions. Visual presentation and human perception are two main aspects should be noteworthy. There are many critical technical challenges to be taken care of through improved data management and visual methods. Some designing and implementing trials will be listed below.

The first challenge is the extension ability of data and its dimension. Visual Analytics requires the technology to extend and reduce the input data according to the variables of the target system [23]. In many user scenarios, Visual Analytics designers have to integrate data of different types and qualities from various resources with the existing database. Therefore, Visual Analytics algorithms need to be competent enough when working with complex statistics.

The second challenge is to extract semantics from collections of documents. There are already some established systems in this fields which cannot achieve grammar correction automatically like the human [24]. Building a model of semantics, which is equipped with a large database of grammar and lexical examples, is a good way to deal with inadequate and wrong information.

The third challenge is to design user-interface and interactions for targeted audiences. Human interface is the tool which sends information from visual systems to users. The high-quality Visual Analytics system can reduce the burden of data analysts, so it is important to develop simple and convenient user interface rather than overly technical and complicated ones, which may divert programmers' attention. User feedback should also be taken as smartly as possible, as it is essential to help readers make fast and right decision with intuitive visualization systems.

The fourth challenge is the evaluation. It is difficult to evaluate a Visual Analytics system due to the complex and heterogeneous problems of Visual Analytics [25]. Traditional testing method, such as White-Box testing and Black-Box testing are not enough to deal with evaluation task. It is necessary to build an evaluation framework which can assess all requirements of users and data analysts.

The last challenge is to avoid misleading design. As mentioned above, 3D image can mislead us what the actual size of the object. The most famous example is Steve Jobs' presentation on 2008 Macworld (Fig 5). It is quite confusing because we do not have those annotations. The 19.5% part of a green pie which represents the Apple market share looks a bit bigger than the 21.2% part. This pie chart caused the illusion to audiences about the Apple market share.

2). *Application Challenge:* Visual Analytics is a promising research direction driven by application design requirements. In this section, we present five most significant application challenges that need to be mastered by data analysts and illustrate solutions of each from technical aspects.

Business analysis is a representative field of Visual Analytics. The primary challenge of this area is to analyze data from different perspectives and indicate assumptions to understand previous and current situations. Besides, to identify reappeared circumstances, it is essential for data analysts to predict future tendency by monitoring market [2].

Regarding biology area, Visual Analytics designers have been using tomography imaging and 3-dimensional digital reconstruction of visualization for recent years. Large amounts of data require fresh and efficient visualization systems, as traditional visualization techniques are incapable of processing so much data at the same time [2].

Visual Analytics also plays a vital role in the Socio-economic area. It is well-known that modern society is a complex system consists of economic, demographic effects, and political decisions [1]. To better understand the different phenomenon in society and make an appropriate choice, it is impossible to consider about interrelationships between analysis and visualization. From the technical aspect, the primary challenge is to find a solution to build a framework, which is the crucial issue for analyzing relationships of socio-economic.

Fig. 5. Steve Job's slide on 2008 Macworld[26]

Safety and security are two essential elements of human life. Analysts need to evaluate safety level based on vast amounts of heterogeneous information streams and warn the public when facing dangerous situations [1]. Data integration and interactive visualization are two practical tools for data analysis in this area.

The research of climate change often acquires the support of weather record and temperature curve. With the help of Visual Analytics, the human could observe, predict and analyze current and future changes through graphs. The difficulty of this field focuses on finding an accurate method to store data efficiently and show the right trend of the environment and climate change.

VI. An Example for Visual Analytics Applications

In bioinformatics research, cluster analysis has become an essential part. To choose the best clustering method and its parameters for a dataset, researchers always need to run different clustering algorithms and compare the result [27]. However, it is time-consuming, cognitively demanding and laborious to compare different clustering results [27].

XCluSim provides a solution for this problem. XCluSim is a Visual Analytics software that enables users to compare different clustering algorithms [27]. The authors build a taxonomy for categorizing existing techniques of clustering results visualization concerning the Gestalt principles of grouping [27]. Users can select the most appropriate interactive visualization regarding the Gestalt principles of grouping by using the taxonomy [27].

From the interface shown below(Fig. 6), we could see that there is three types of overviews: (A) parameter information view, (B) force-directed layout view, and (C) dendrogram view [27]. These views enable users to compare different clustering results simultaneously in a scalable way [27]. When some clustering results are selected in the overviews, they could render in the (D) enhanced parallel sets view for more in-depth comparison tasks [27], then users can access the detailed information of the selected clustering results with each result in each tab of (E) the tabular view [27].

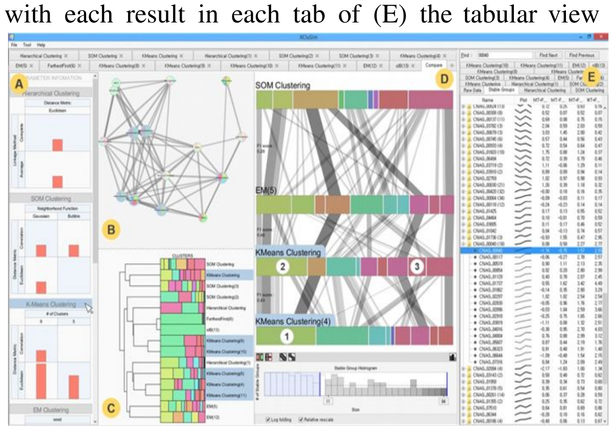

Fig. 6. Visualization for comparing multiple clustering results in XCluSim [27]

Moreover, XCluSim supports various clustering algorithms, dedicated visualizations and interactions for different types of clustering results which could offer the more efficient exploration of details on demand [27]. Those features are useful for researchers when they design new visualization for a clustering result [27].

VII. Conclusion

Visual Analytics is an uprising field during the recent decade of years. The first reason is that open source tools and practices are more accessible than before. The second reason is that people have realized the importance of gaining knowledge and insight from various data sources by automatic analysis [2]. The compressed knowledge from an enormous amount of data is transformed into a relatively small space since people pay more attention to the visual information due to the essential position of vision in sensory organs. The following part will introduce the discipline to build Visual Analytics system with tools mentioned above. As a bridge between raw data and readers, the system should achieve the fundamental goal of conveying messages. It is useless to spend too much time on data analysis and processing if users cannot derive data information by using Visual Analytics tools, even the model is entirely accurate. To convey useful and meaningful messages to users, designers should:

(1) Understand the data source deeply. It is essential to have a complete and comprehensive insight about what we should do behind raw data and how to maximize it with the explicit purpose.
(2) Make the visual form easily understandable. The audience of Visual Analytics is non-IT professionals, so it is essential to show the result in a readable form. “less is more ” is a useful principle, which means that it is not good to reveal all the data on the screen, but just needed.
(3) Express the hidden but essential information. Visual Analytics aims to efficiently express the unknown data from the visualization in the initial state.
(4) Reject misleading design that may cause misleading interpretations. Visual Analytics aims to make information transparent rather than unnecessary beautiful. For example, 3D visualizations referred to “chart-junk” is not encouraged by data designers.
(5) Encourage users to interact with data. Interaction can motivate users to explore the data from different perspectives. At the same time, annotations should always be triggered by interactions like mouse event.

In this paper, we introduce Visual Analytics by a short description of its main related fields like Visualization, Perception, Cognition, Data Mining, and Human-Computer Interaction, as well as some bias of Visual Analytics. By presenting the whole process of Visual Analytics, we figure out where those related fields are applied in the Visual Analytics and compare it with ordinary data science process. To better understand the science of Visual Analytics, we use several practical examples to demonstrate its advantages. Next, we introduce several modern Visual Analytics tool and several design principles based on personal experience. In the end, we discuss the application and technical challenge and foresee the future opportunity of the Visual Analytics and its related fields.

VIII. References

[l] D. Keim et al, "Visual analytics: Definition, process, and challenges," in Anonymous Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 154-175.

[2] D. A. Keim et al, "Visual analytics: Scope and challenges," in Anonymous Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 76-90.

[3] J. J. Thomas and K. A. Cook, "Illuminating the Path: Research and Development Agenda for Visual Analytics.", IEEE-Press, 2005

[4] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE Computer Graphics and Applications*, vol. 26, (1), pp. 10-13, 2006. . DOI: 10.1109/MCG.2006.5.

[5] J. Han and M. Kamber. "Data mining; concepts and techniques, 3d ed," Reference & Research Book News, vol. 26, (5), 2011.

[6] Maimon, O., Rokach, L. (eds.): The Data Mining and Knowledge Discovery Handbook. Springer, Heidelberg, 2005

[7] L. Little *et al*, *Perspectives on HCI Research with Teenagers.* (1st 2016 ed.) Cham: Springer International Publishing, 2016.

[8] W. Aigner et al, "Visual Methods for Analyzing Time-Oriented Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, (1), pp. 47-60, 2008.

[9] L. V. Cherenkova, L. V. Sokolova and A. Y. Naumova, "Characteristics of detection of the general direction of movement of visual objects by preschool-age children with typical and atypical cognitive development," *Human Physiology*, vol. 40, (5), pp. 490-496, 2014.

[10] J. J. Gedney, H. Logan and R. S. Baron, "Predictors of short-term and long-term memory of sensory and

affective dimensions of pain," *Journal of Pain*, vol. 4, (2), pp. 47-55, 2003.

[11] J. S. Yi et al, "Toward a Deeper Understanding of the Role of Interaction in Information Visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, (6), pp. 1224-1231, 2007.

[12] J. J. van Wijk, "The value of visualization," in 2005, . DOI: 10.1109/VISUAL.2005.1532781.

[13] D. G. Murray, Tableau Your Data: Fast and Easy Visual Analysis with Tableau Software. (Second ed.) Indianapolis, Indiana: Wiley, 2016.

[14] Bostock, M., V. Ogievetsky, and J. Heer. "D³ Data-Driven Documents." *IEEE Transactions on Visualization and Computer Graphics* 17, no. 12 (2011): 2301-2309.

[15] L. Wilkinson, "Comment on “A Model for Studying Display Methods of Statistical Graphics”," *Journal of Computational and Graphical Statistics*, vol. 2, (4), pp. 355-360, 1993.

[16] F. J. Anscombe, "Graphs in Statistical Analysis," *The American Statistician,* vol. 27, (1), pp. 17-21, 1973.

[17] Anonymous "Visualizing blood vessel abnormalities in living human eyes," *Proceedings of the National Academy of Sciences of the United States,* vol. 112, (18), pp. 5546, 2015.

[18] A. Koutsoudis et al, "Multi-image 3D reconstruction data evaluation," *Journal of Cultural Heritage*, vol. 15, (1), pp. 73-79, 2014.

[19] P. Wang et al, "Depth-Perception Preserved Content Remapping Based on Cornsweet Illusion for 3D Display," *Journal of Display Technology*, vol. 12, (11), pp. 1319-1328, 2016.

[20] G. A. Radvansky, *Human Memory*. (Third ed.) New York, NY: Routledge, 2017.

[21] A. N. Gorban’, "Principal manifolds for data visualization and dimension reduction," in 2008, .

[22] A. Kerren and F. Schreiber, "Toward the role of interaction in visual analytics," in 2012, . DOI: 10.1109/WSC.2012.6465208.

[23] J. Choo et al, "An interactive visual testbed system for dimension reduction and clustering of large-scale high-dimensional data," in 2013, . DOI: 10.1117/12.2007316.

[24] S. K. Bansal and S. Kagemann, "Integrating Big Data: A Semantic Extract-Transform-Load Framework," *Computer, vol. 48*, (3), pp. 42-50, 2015.

[25] S. Donahue et al, "Procedures for the Evaluation of the Visual System by Pediatricians,"*Pediatrics,* vol. *137*, (1), pp. e20153597, 2016.

[26] Block R. Live from Macworld 2008: Steve Jobs keynote. Engadget.https://www.engadget.com/2008/01/15/live-from-macworld-2008-steve-jobs-keynote/. Published July 14, 2016. Accessed December 4, 2017.

[27] S. L'Yi et al, "XCluSim: a visual analytics tool for interactively comparing multiple clustering results of bioinformatics data," *Bmc Bioinformatics,* vol. *16*, (11), pp. S5, 2015.