Data Visualization Revisited

Kai A. Olsen[1,2], James G. Williams[3], and Michael B. Spring[3](✉)

[1] Molde University College, Molde, Norway
[2] Department of Informatics, University of Bergen, Bergen, Norway
[3] School of Computing and Information, University of Pittsburgh, Pittsburgh, USA
spring@pitt.edu

Abstract. More than 25 years ago we developed a data visualization system called Vibe. During this same period we developed a system for collaborative authoring – CASCADE – that made heavy use of visualization. These were but a few of many efforts at that time to develop new methods for understanding data, stimulated by improved hardware - faster CPUs, more memory and high resolution graphical displays that made it possible to perform advanced visualization on ordinary PCs. In this paper we revisit some of these efforts and then discuss where visualization is today. We briefly examine big data and scientific visualization where many of the issues we explored 25 years ago are being revisited. Our focus however is on general visualization. What we find is that advanced visualization systems for data presentations have not come into general use. We explore some of the reasons this may be the case.

Keywords: Visualization · Systems · Data analysis

1 Introduction

Through the 80's and 90's, Vibe, CASCADE, and many other systems, introduced novel and advanced visualization techniques for what we would call general visualizations, i.e., where the idea is to present the data in meaningful ways. Our objective was to extend the traditional range of visualization techniques beyond such techniques as time series, bar code diagrams, pie charts, etc., i.e., visualization methods that have been here for hundreds of years.

While our focus is on general visualization techniques we see that scientific visualization and visualization of c data continue to receive attention [1]. We are in a highly data-driven environment, in which data are acquired on a continuous basis for a variety of purposes. The ability to make accurate and timely decisions based on available data is crucial for business decisions, medical treatments, national security, crime resolution, disaster management and many other areas. The traditional statistical and visualization techniques used to extract knowledge from these large and complex datasets have not developed significantly.

Visualization has proven effective not only for presenting essential information in vast amounts of data but also for driving complex analyses. Big data exploration/analytics and

G. Chowdhury et al. (Eds.): iConference 2018, LNCS 10766, pp. 292–302, 2018.
https://doi.org/10.1007/978-3-319-78105-1_33

presentation create new research issues for the computer graphics and visualization community. Big data characteristics such as volume, velocity, variety, value and veracity require quick decisions for implementation, as the information can lose value very fast. Such fast data expansion will result in challenges related to a human's ability to manage the data, extract information and gain knowledge from it using traditional statistical and visualization methods. 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. Big data related visualization challenges are based on an understanding of human perception and the correlation between Augmented Reality and Virtual Reality that are suitable for the perception capabilities of humans. To visualize Big data, feature extraction and geometric modeling can be implemented. A visual representation is more likely to be accepted by a human in comparison with unstructured textual information or statistics. Likewise, an analyst is likely to find relationships in data using advanced visualization tools such as Vibe that would not be discovered otherwise.

While scientific visualization, geo-coded visualization and big data may provide us with many new and advanced visualization techniques, which may be introduced into mainstream data visualization sometime in the future, we shall concentrate here on general visualization. Here the objective of visualization is to present the data to the reader. The idea is to show the information that may be hidden in the data in a way such that the reader gets an immediate understanding. While the traditional visualizations are a part of most reader's vocabulary, the aim of many of the visualization methods and their prototypes developed in the early nineties, such as Vibe and CASCADE, was to extend this vocabulary with more advanced methods.

2 General Data Visualization – The Vibe System

Vibe is a multidimensional visualization system [2, 3]. With Vibe the users define POIs (points of interest). These work similar to axes in a two dimensional visualization, however, in a Vibe display one could have many POIs. Data objects are attracted to POIs. If an object has a value on only one POI a symbol for the object will be placed on top of the POI. If it has an equal value for two POIs it will be placed between these, etc.

Figure 1 shows a VIBE diagram with four POIs, each shown as a small circle. One data object has only values on POI A, and the icon representing this object is placed on top of A. We see two objects that have a value on POIs A and B, one with a stronger value on A than B, and another with a stronger value on B and A. The objects in the middle have a value on all POIs. Vibe allows the user to reposition the POIs and to block the influence from one or more POIs. The system can then visualize how the data objects move by drawing a line from the former to the new position. The user can click on any object and will then get additional information on this data object. There are functions for zooming in and out and to use color, for example to distinguish different types of data objects.

The user interface is shown in Fig. 2. In the example shown here the POIs are data types that characterize countries, such as life expectancy, literacy and infant mortality. Each is offered in two variations, one that attracts higher than average values, and one

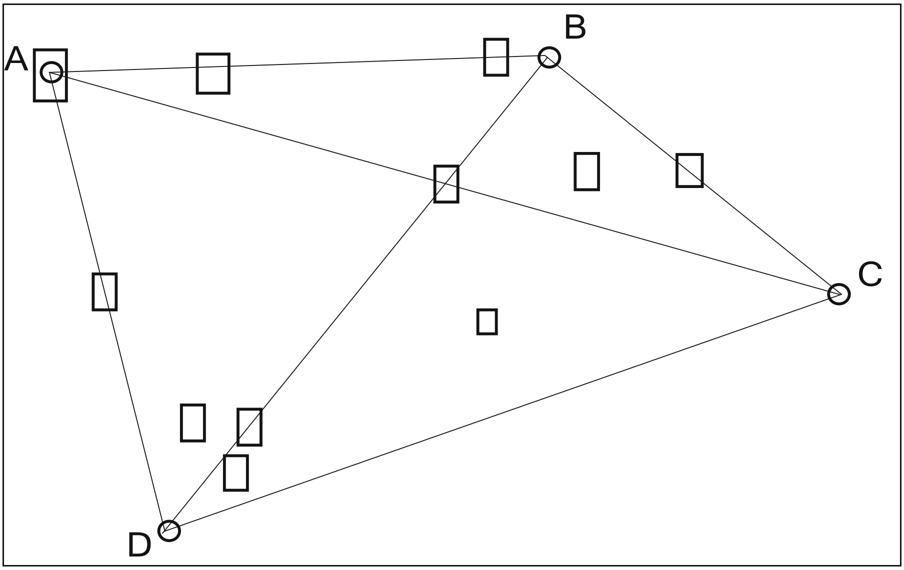

Fig. 1. A VIBE diagram with four POIs

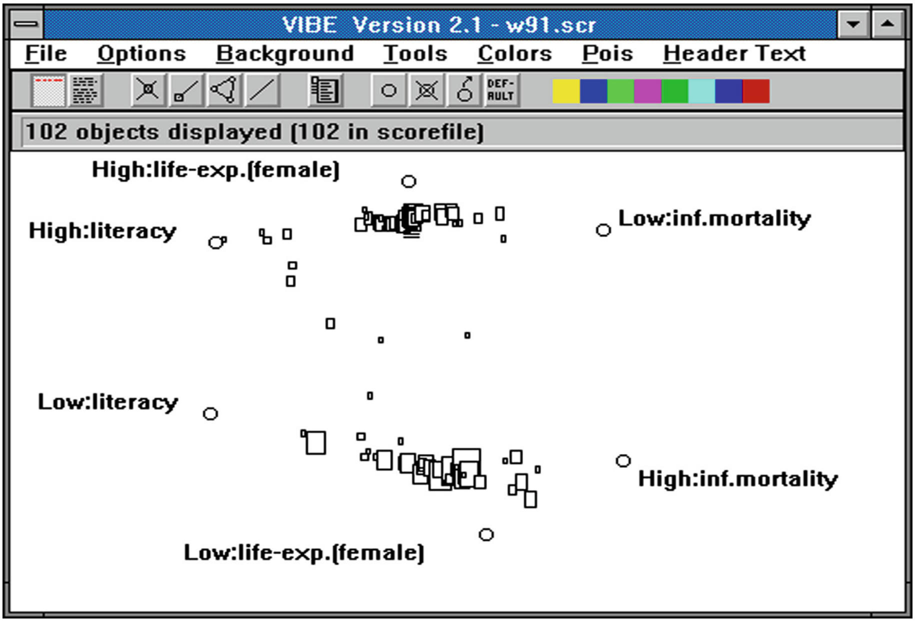

Fig. 2. VIBE user interface, here displaying countries of the world

that attracts lower than average values. The data objects are the countries of the world. We see that these fall into two groups, the ones that have positive values on these POIs and the others. Some countries fall outside of these groups.

In another visualization of the data on all the countries of the world (Fig. 3) we also introduced POIs for GDP (Gross Domestic Product). Back in 1991 we were able to show that there were two countries that fell outside of the "normal" groups: Cuba had a low GDP per person but scored well on the POIs for literacy, life expectancy and infant mortality while Kuwait had a high GDP but scored low on these POIs. That is, by visualizing these data one was able to see combinations that were not apparent from the data directly.

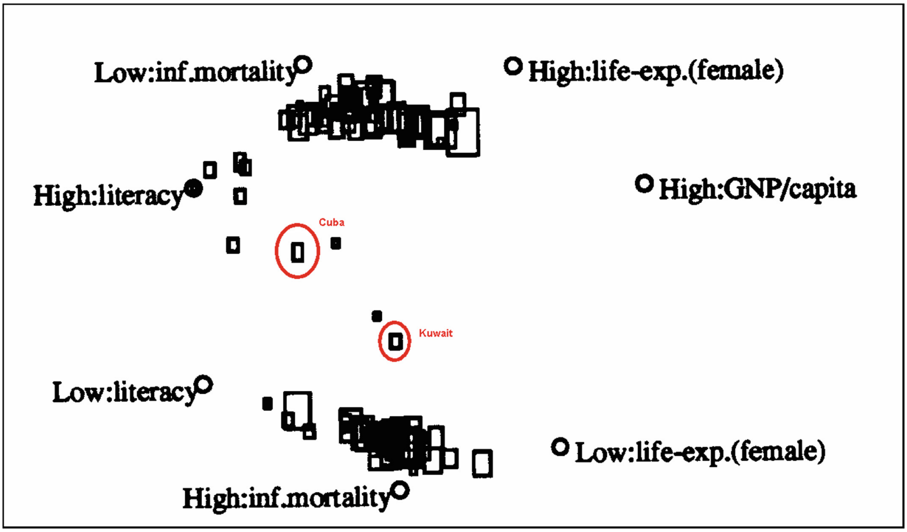

Fig. 3. Countries of the world fall into one of two groups

3 Document Visualization: CASCADE

During this same period we worked on the development of a collaborative authoring software system called CASCADE – "Computer Aided Support for Collaborative Authoring and Document Editing" [4, 5]. The system made heavy use of visualization to augment the authoring and editing process.

Figure 4 shows a document with comments. In the upper left corner, is a dialog box that controls how comments are rendered within the document. The classification of comments and the colors associated with them can be set per project. In this case, comments are rendered and shown in the document and the "mural" based on the status of the comment. The comment dialog appears when the user clicks on the document and allows the user to make a comment and classify it – in this case over three dimensions. The "mural" appears to the right of the scroll bar which shows a thumb proportional to the amount of the document displayed. Three comments are visible in the portion of the document displayed, one pending, one open, and one settled. Four comments are shown in the mural. The user can see that there is another open comment towards the end of the document as depicted in the mural.

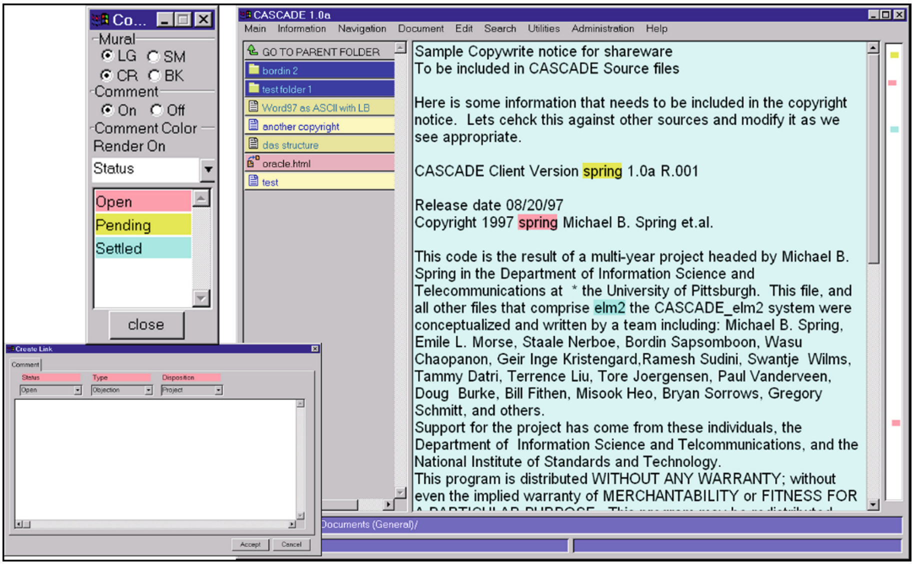

Fig. 4. CASCADE showing comments, comment control, and mural (Color figure online)

Note that the folder listing is also color coded with files shown as dark type on pastel backgrounds – yellows for text, blues for images, etc. and folders displayed as light type on dark backgrounds. Folders could be of various types as well – ordered versus unordered as an example.

Figure 5 shows a visualization of the activity of five users over a period of 29 days. The lines in each white bar show when the user was logged in. The histogram at the bottom shows times when the users were concurrently logged on over the period of

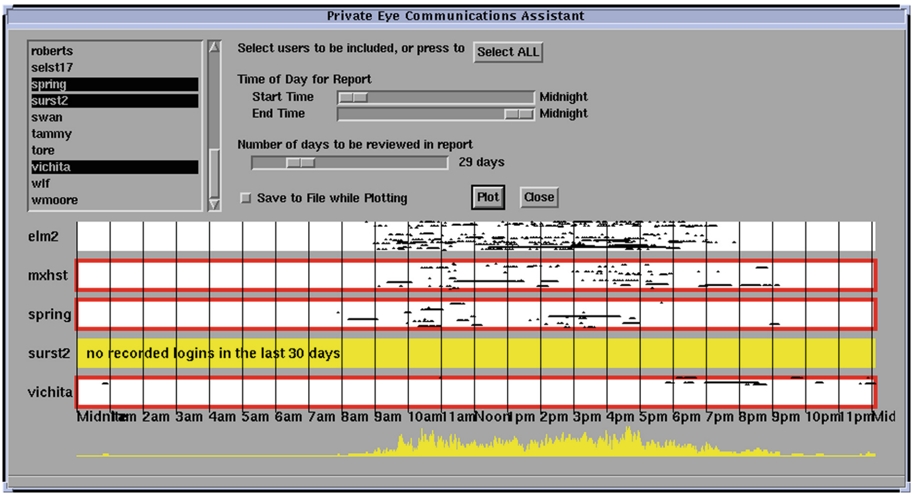

Fig. 5. Session information for users

review. The visualization shows the overall work effort of team members and may be used to find the best time to schedule synchronous meetings. Similar visualizations were developed to track keystrokes and mouse actions that could inform decisions about accelerator buttons that would ease editing and authoring activity.

CASCADE had a variety of static and dynamic hypertext displays. For example, a set of documents could be analyzed generating an ad hoc hypertext structure that could be used to access some or all of the comments on a set of documents. (One motivation for the development of CASCADE was the authoring of standards documents which involved tens to hundreds of editors working on large complex documents over a significant period of time). CASCADE was designed to make the processing of such documents more efficient. Figure 6 shows one of the more ambitious navigation tools – called "Docuverse". Docuverse was designed to show attributes of a large set of documents. In this case, it is displaying the size of about 400 documents. Beginning at the root, the user can choose any subdirectory on the left side to drill down. Docuverse had a half dozen attributes that could be visualized – age, size, number of comments, last access, last modification, etc. Double clicking on a folder would exit Docuverse into a more traditional directory listing.

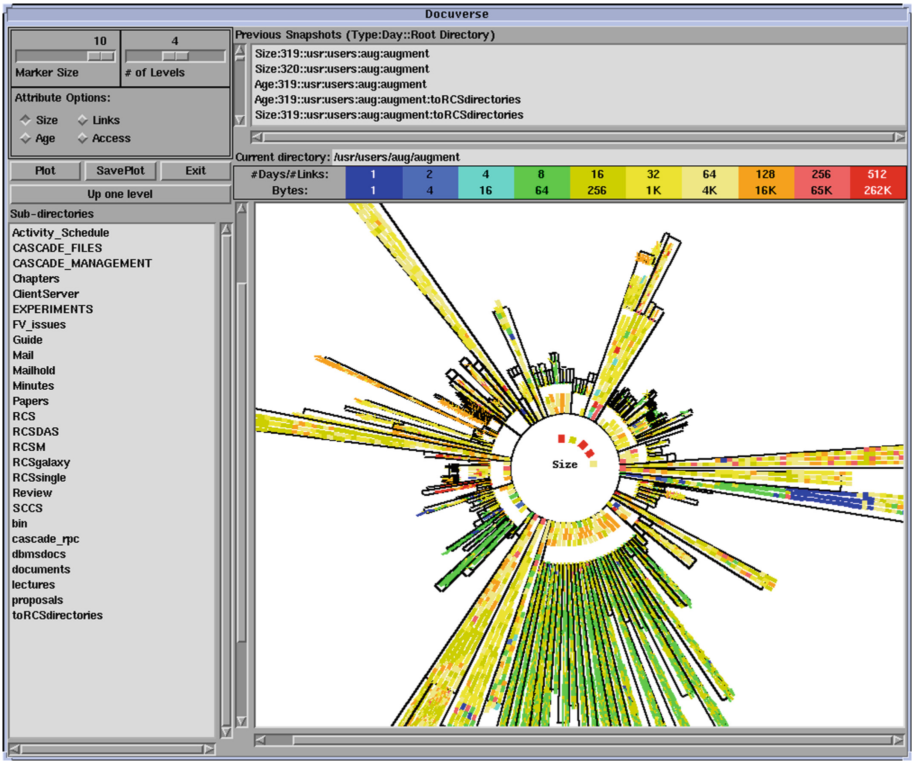

Fig. 6. Docuverse

4 Visualization Today

Novel visualization methods dedicated to abstract data have decreased in recent years. If we examine the visualizations that are used today in newspapers, reports and scientific papers we find only standard data visualizations as illustrated in Fig. 7. That is, two, perhaps three, dimensional figures, and charts. We find the traditional pie and bar diagrams. In practice, these are the type of diagrams that can be made by an Excel spreadsheet.

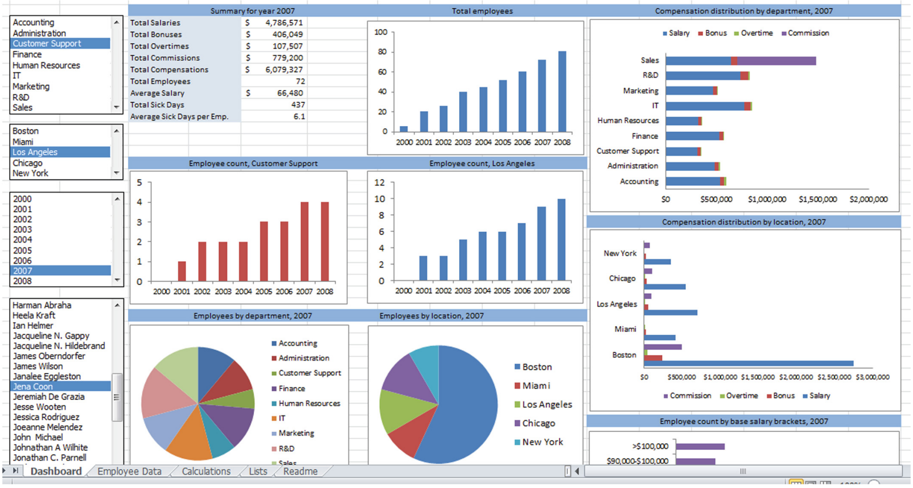

Fig. 7. General visualizations

We find some more "advanced" visualization, e.g. data visualized on maps and in three dimensional diagrams as shown in Fig. 8.

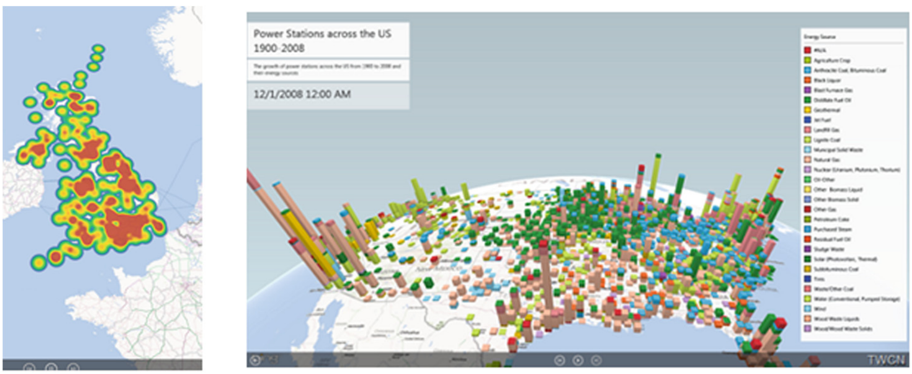

Fig. 8. Geocoding information

A particularly interesting geocoded example of a large data set visualization also makes use of a timeline as shown in the visualization of a distributed denial of service attack that took place on August 31, 2013 (Fig. 9). See www.digitalattackmap.com for more.

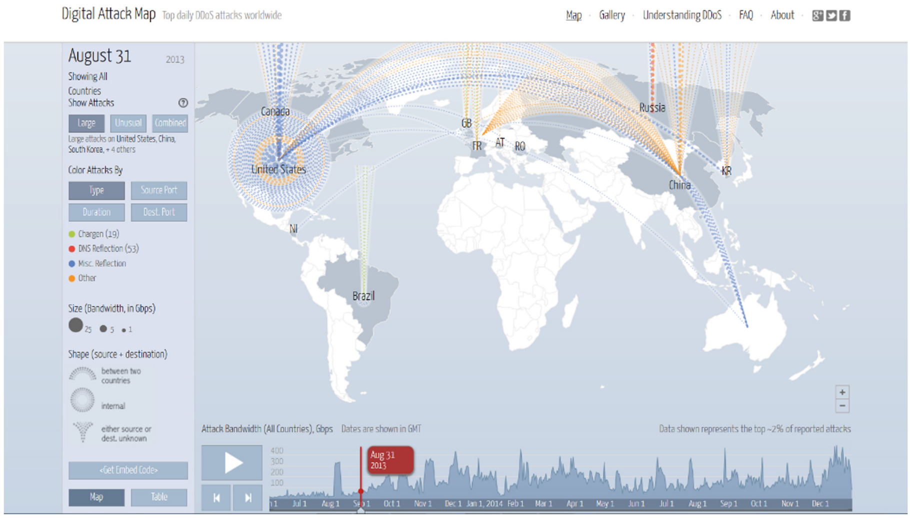

Fig. 9. DDOS attack on the US, China and others August 31, 2013

However, geocoding of data is nothing new. Minard's famous visualization of Napoleon's disastrous campaign in Russia is a well-known example. It was drawn in 1869. Today we may find more visualizations than at that time, since they both are easier to produce and easier to disseminate, but in practice there has been little development for abstract data visualizations – today we find, for the most part the same techniques used 150 years ago. The interesting question then is, why haven't Vibe, CASCADE and other experimental systems evolved?

5 Why Did General Visualization Tools Not Evolve?

One reason for not using advanced visualization methods is surely that in many cases the data will speak for itself. If the median income in the US is $865 per week for a full time worker there is no need for any visualization. If 33% of the population surveyed supports the president no visualization is needed. The number says everything, but instead of presenting the raw data that 330 out of a 1000 people surveyed confirmed that they supported the president, we offer the result as a percentage. In many cases it is sufficient to present the data as numbers or with simple diagrams.

However, if we want to study how income changes from year to year a time diagram will show this. With more advanced visualizations it is possible to see relationships not as apparent in multi-dimensional data sets. For example, using VIBE on the countries of the world we were able to find many facts that were not apparent from the data directly. One example was Cuba and Kuwait, another was how countries at war fell from the "good" group to the "bad" group, and how they managed to get back as soon as a war ended. We saw this when we visualized the Balkan countries over the years of conflict and peace. With other advanced visualization techniques it was possible to draw interesting information out of complex data. Still, nearly all the visualizations we see today are traditional.

We may criticize users for being too conservative – that they only do what they have always done. However, past experience shows that users are willing to learn something new if the advantages are high enough. Typewriters were replaced by word processing systems, users have been willing to learn how to produce high quality document layouts, PowerPoint presentations have grown in sophistication well beyond overheads, snail mail has been replaced by email that may include various kinds of emojis and pictures. In general users are willing to use computers for very many tasks and to develop new skills.

When working with complex data there are two basic considerations; data exploration/analysis and presentation to an audience. For data analysis/exploration, the primary audience is the data analyst. This is the person who is both attempting to analyze the data and to interpret the results. This person typically needs to work with feedback cycles of defining hypotheses, analyzing data, and visualizing the results. For data presentation, the audience is typically a group of end-users such as the readers of a publication. As far as we can see the novel and advanced methods failed both for exploration and for presentation.

Independent of the technique used the main idea of an abstract visualization is to translate numbers into graphical presentations. When communicating to a general audience it will be necessary for the visualizations to be intuitive. Traditional visualization techniques are intuitive. A pie chart shows part of the whole and connects to, yes, pies. Time series gives us the impression of something that goes up and down, such as a country road. Similarly, a bar chart may connect to the skyline of a large city, with high and low buildings. The more advanced visualization, especially the multi-dimensional, are less intuitive.

The visualizations for data exploration need to be easy to create and may often show multiple dimensions to unearth complex patterns. Many of the advanced data exploration tools require more hypothesis formation and concept creation than simpler visualization techniques such as a line graph or bar chart or basic statistical methods. For data presentation, it is critical that visualizations be simple and intuitive. The audience doesn't have the patience to decipher the meaning of complex presentations. For instance, presenting data in a chart as a visualization cannot stand-alone without an explanation to readers. With more complex visualizations produced by tools such as VIBE the users have a difficult time understanding the presentation without a comprehensive explanation of how to use them.

Tools such as Vibe for data exploration offer function over form where analytical, programming, data management, and business intelligence skills are more important than the ability to create presentable visualizations which may be why these tools are not used frequently. For visualizations, most users can understand the simple diagrams. These are used so many times that users get experience. There is no need to explain the basics of existing visualization techniques to readers. While VIBE, along with all the other advanced and novel methods, would need an explanation of how the visualizations are produced, this is not needed for simple charts. By using traditional visualizations we speak a language that the user knows, with the new advanced methods we introduce new "terms" and "concepts" that needs an explanation. The cost-benefit analysis may not favor new advanced visualizations.

6 Discussion

General visualization tools, tools that can be used on a wide range of data, are often favored over specialized tools, even if the specialized tools may be better for a specific application. That is, when we write reports, books or scientific papers we try to use the same tools all the time. This makes the production of the visualizations very effective. We also avoid the need for explaining the technique as long as we keep to visualizations that the readers have seen before. This holds true for general use of a technique. If we move into special areas such as scientific visualization users may find it worthwhile to invest in learning to use more advanced methods. In these cases the visualization is often used to let the researcher explore complex data to find correlations, and perhaps not to explain these to a general public. It is interesting to note the similarities between Spring and Jennings [6] and Olshannikova et al. [1]. Both look to rules for the use of virtual and augmented reality. On the other hand, with the type of visualizations that we consider here the goal is to communicate information clearly and effectively [7]. The basic rationale is to explain data to readers, the methods used are time series, ranking different values, showing correlation and deviation and methods to compare values.

The fact that general, simple techniques often are preferred over the more special and complex is seen in other areas as well. At one time there were expectations that 3D movies would conquer the market. While there have been some successful examples in few selected genres, 2D seems to do the job. Similarly, there may be interesting applications for virtual reality and augmented reality, but it seems that it is difficult to get these systems used generally. Speech recognition may be another example. There are applications where it is needed, for example where the user's hands are tied up, but it seems that most of us would still choose a keyboard for entering data. In general, reliability and simplicity seem to be what we want.

References

1. Olshannikova, E., Ometov, A., Koucheryavy, Y., Olsson, T.: Visualizing Big Data with augmented and virtual reality: challenges and research agenda. J. Big Data **2**, 1–27 (2015)
2. Olsen, K.A., Hirtle, S.C., Sochats, K.M., Williams, J.G.: Ideation through visualization: the VIBE system. Multimed. Rev. **3**(3), 48–59 (1992)
3. Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., Williams, J.G.: Visualization of a document collection: the VIBE system. Inf. Process. Manag. **29**(1), 69–82 (1993)

4. Sapsomboon, B., Spring, M.B.: Computer based collaborative authoring for standards development. Open Syst. Stand. Track. Rep. **5**(8), 4–6 (1996)
5. Spring, M.B., Vathanophas, V.: Peripheral social awareness information in collaborative work. J. Am. Soc. Inf. Sci. Technol. **54**(11), 1006–1013 (2003)
6. Spring, M.B., Jennings, M.C.: Virtual reality and abstract data: virtualizing information. Virtual Real. World **1**(1), c–m (1993)
7. Friedman, V.: Data visualization and infographics. In: Graphics, Monday Inspiration (2008)