

# Data visualization: foundations, techniques, and applications

Hans-Christian Hege  
hege@zib.de / Zuse Institute Berlin, Berlin, Germany

Receipt: 13-8-2018 / Acceptance: 4-9-2018

**ABSTRACT.** The idea that there is no precedence for the amount of data that is being generated today, and that the need to explore and analyze this vast volumes of data has become an increasingly difficult task that could benefit from using Data visualization is presented. It is pointed that the goals of data visualization are data-driven and depend largely on the type of application, but the final objective is to convey to people information that is hidden in large volumes of data. Finally, the visualization pipeline is presented to review aspects of visualization methodology and visualization tool design, to conclude that the true potential of visualization emerge from the interaction of the user with the visualization model. The paper concludes establishing that the current processes of digital transformation will increase the need for visual analytics tools.

**KEYWORDS:** data visualization, visualization evolution, computer graphics basics, visualization pipeline, interactive visualization

## Visualización de datos: fundamentos, técnicas y aplicaciones

**RESUMEN.** La cantidad de datos que se genera hoy no tiene precedentes y la necesidad de explorar y analizar estos vastos volúmenes de información se ha convertido en una tarea cada vez más difícil, ante la cual se presenta el beneficio del uso de su visualización. Se señala que los objetivos de dicha visualización son impulsados por los propios datos y dependen en gran medida del tipo de aplicación, pero el objetivo final es transmitir a las personas la información que está oculta en grandes volúmenes de datos. Finalmente, se presenta el canal de visualización para revisar aspectos de la metodología de visualización y el diseño de la herramienta de visualización, para concluir que el verdadero potencial de la visualización surge de la interacción del usuario con el modelo de visualización. El documento termina estableciendo que los procesos actuales de transformación digital aumentarán la necesidad de herramientas de análisis visual.

**PALABRAS CLAVE:** visualización de datos, evolución de la visualización, conceptos básicos de gráficos por computador, flujo de visualización, visualización interactiva

## 1. EVOLUTIONARY BASIS OF VISUALIZATION

Visualization means the development and application of techniques for creating images, diagrams, or animations in order to communicate information.

Humans have a long and intense relationship with pictures: starting from cave paintings to pictorial symbols and writing systems, da Vinci's revolutionary techniques for conveying technical or scientific information, visual and cinematic arts, and computer-generated imagery and social networks that are flooded with pictures.

Why do humans have such a preference for visual information? This is due to two evolutionary reasons: During evolution, the human visual system has developed a tremendous capacity, making pictorial transmission and processing of visual information the most effective way of evaluating sensory data. Therefore, do images fit perfectly to human perception. Furthermore, due to evolutionary conditions, spatial thinking has developed into one of the strongest components of human cognition. Since images encode information (in the first place) spatially, images also fit perfectly to human cognition.

It is the combination of these two factors, namely, the exploitation of the most efficient information channel and the most advanced cognitive functionality, that explains the significance of pictorial representations to human beings.

## 2. NEED OF DATA VISUALIZATION

For centuries, the amount of quantitative information collected, e.g., by census, surveys, and measuring techniques, increased. And, in the digital century, it exploded. The need to make an ever-increasing amount of data comprehensible to humans led to the development of Data Visualization. The aim of Data Visualization is to make 'hidden' information in the data accessible and understandable to humans by turning it into visual information.

Historically, the roots of data visualization are old: Examples from ancient cultures are known and important milestones happened in the Renaissance, in particular the discovery of the central perspective and the invention of technical drawings. Technical drawings have become an integral part of the technological transformations in the early modern age. A prominent example of an old and very advanced visualization technology is cartography: It has gone through thousands of years of development and reached such a complexity that it became an academic discipline itself (at the beginning of the 20th century). Before computers were available, all these depictions were created manually. With the advent of computer graphics, also visualization has been revolutionized.

### 3. HISTORY OF COMPUTER GRAPHICS

After the invention of the computer in the 1940s, it did not take long until the idea to create images with the computer came up. Already in the 1950s, the first graphical output devices (plotter and cathode ray tube) and a graphical input device (light pen) were developed. In the 1960s, the development of computer graphics started in universities, national laboratories, and industries. Plotters were used in geodesy, road construction, and meteorology but also by pioneers in computer arts.

In 1963, the first computer-generated film in space research was created. While early computer graphics were vector graphics, at end of the 1960s, raster graphics were developed. In 1968, ray casting was invented and the pursuit of photorealism started. In 1969, the first SIGGRAPH conference took place, which from the beginning was and still is the major annual event in computer graphics. Computer graphics became a recognized subdiscipline of computer science at universities. In the 1970s, many early breakthroughs happened, like hidden surface and texture mapping. In 1979, ray tracing was invented, which recursively traced individual light rays according to optical laws. In the 1980s, graphics hardware development started, which brought the costs of image synthesis down to affordable prices. The pursuit of photorealism continued by developing global illumination models. The first fully computer-animated movies were created and videogames started. In the 1990s, photorealism came a step closer, and OpenGL and DirectX APIs were developed, whose major functionality could be implemented in hardware. Powerful workstations with graphics hardware became then ubiquitous in industry and research.

The world's first GPU, realizing the whole graphics pipeline in hardware, appeared in 1999. High-end graphics became now possible not only on expensive workstations but also on consumer-level PCs. This development continued in the 2000s with improved texture and shading in computer graphics via the widespread adoption of normal mapping, bump mapping, and a variety of other techniques. The movie industry increasingly used computer graphics to produce more and more realistic animations and special effects in films. Scientific computing started to utilize the functionalities of GPUs for general-purpose computing (GPGPU – general-purpose computing on graphics processing units). Shading languages were introduced for real-time rendering, enabling the computation of complex optical effects and demonstrating the first real-time ray tracers on highly-parallel computer systems.

At the end of this decade, Nvidia announced an API for real-time ray tracing on GPUs. In the 2010s, computer-generated imagery became ubiquitous and real-time photorealistic rendering became possible. Graphics hardware, supporting the multistage process of rendering, became even more flexible by offering programmable shaders that shader languages. GPUs for mobile devices became available, bringing real-time computer graphics to tablets and smartphones. Ray tracing, which has been the Holy Grail for rendering realism since its

invention almost 40 years ago, became available on consumer platforms using GPUs specifically designed to support ray tracing. It has been recently announced that GPUs combine hardware-based ray-tracing functionality with deep learning units (recurrent autoencoders) for efficient denoising. This technology promises to bring real-time ray tracing on consumer platforms into the regime of image resolutions and frame rates that are required in practical applications<sup>1</sup>.

#### 4. DEVELOPMENT OF DATA VISUALIZATION

An important step towards the foundation of data visualization was taken in 1967 by cartographer Jacques Bertin with his book on graphic theory and visualization (Bertin, 1967). He developed the theory of “graphic variables” and showed that in cartography, depending on the scale level of the data to be displayed, specific uses of the graphic variables are favorable. In the 1970s, John Wilder Tukey, a pioneer in data analysis, developed new methods for visualizing statistical data, also of higher dimension, and introduced an explorative data analysis, i.e., the interactive, visually-supported data analysis (Tukey, 1977). The development of computer-aided designs started and, in sciences, e.g., in molecular crystallography, researchers started to utilize computer graphics techniques for depicting and understanding data. With the increasing volume and complexity of the data produced in science, the use of visualization became even more attractive. In the 1980s, supercomputers entered science, which significantly increased the volume of simulation data. This, together with the vision of using data visualization for understanding scientific data in general, led to the 1987 NSF visualization report (McCormick, 1987). This prophetic publication, based on recommendations of a panel of experts and pioneers, is often regarded as the birth of data visualization as a scientific field.

In the following year, volume rendering was proposed in two papers (Levoy, 1988; Drebin, Carpenter, & Hanrahan, 1988): a computer graphical technique for visualizing volumetric data. The promising early developments, the rapid technical progress in graphics hardware, the number and variety of new ideas, the number of interested users, and the orientation diverging more and more from pure computer graphics led to the foundation of a new dissemination platform: the IEEE Visualization Conference in 1990<sup>2</sup>. This conference, the most important forum for data visualization, has been held annually until today. As the field matured as a scientific one, chairs for visualization were set up at universities, establishing the new academic subject.

In the course of developing the field, three subareas emerged: Scientific Visualization (SciVis), Information Visualization (InfoVis), and Visual Analytics (VA). However, despite

---

1 For more details on computer graphics history, see Wikipedia, Computer Graphics – History, [http://en.wikipedia.org/wiki/Computer\\_graphics#History](http://en.wikipedia.org/wiki/Computer_graphics#History)

2 See Wikipedia, IEEE Visualization, [https://en.wikipedia.org/wiki/IEEE\\_Visualization](https://en.wikipedia.org/wiki/IEEE_Visualization) and <http://ieevis.org>.

many attempts to define a clear demarcation, there are no clear-cut borders. The reason is that the historical development was not only based on scientifically justifiable distinctions but also on non-scientific aspects like preference for / dislike of, mathematically-oriented research vs. perceptually- and cognitively-oriented research. Even aspects of funding opportunities and sociological contingencies played a role in the historical development.

It is well known that particularly in fields where numerical simulations are performed, SciVis is often associated with that kind of data. However, there is plenty of scientific data that is partially or not spatiotemporal. Since practical solutions require to consider all types of data in an application field, SciVis research also dealt with non-spatial data, making the classification doubtful.

InfoVis is more about visualizing ‘abstract’ data, such as higher-dimensional data, or data represented by graphs. The only fundamental difference, though significant, is that in spatiotemporal data the graphical variable ‘location’ is generally defined by the data. This, in turn, means that the design space of graphical representations is more limited. Consequently, in InfoVis, the design and evaluation of graphical representations have a more prominent place. Nevertheless, such kind of data also show up in scientific research. Furthermore, InfoVis is as scientific as SciVis, and SciVis is also related to visualizing information, which shows that the terminology is confusing.

Visual Analytics emphasizes a specific flavor that was already present in SciVis. In this subfield, there is a long history of combining visualization with automated data analysis, often in the form of ‘feature extraction’ (where feature means some patterns or structures characterizing the data, like the vortices in a fluid flow). In InfoVis, however, this combination of visualization and automated data analysis was less common. This led to the emergence of a new field, called Visual Analytics, which combines the power of automated data analysis with the ability of humans to quickly capture patterns or trends visually. A more extensive, perhaps somewhat overbearing, definition of Visual Analytics is “the science of analytical reasoning facilitated by interactive visual interfaces” (Cook & Thomas, 2005).

## 5. GOALS OF DATA VISUALIZATION

At the heart of the visualization there is data, which represents some section of reality. All kinds of data are possible: scientific, medical, technical, economic, environmental... As long as this data is processed automatically and the processing itself is not of concern, visualization is not needed. Visualization comes into play when information must be conveyed to humans.

The range of goals and applications is wide. For instance, natural scientists or economists are often concerned about creating a formal model for every process. To do this, they observe the process, i.e., acquire the data, determine structural and quantitative characteristics, and

connect these results with existing knowledge or established frameworks. Health professionals want to diagnose illnesses, make therapy decisions, and monitor therapies. To this end, they acquire patient-specific data, link it with their knowledge, interpret it, and finally, based on the results, make decisions. Other examples are everyday life applications, like navigation using interactive maps and routing algorithms, or analysis of relationships in social networks using visual graph representations.

Generally speaking, the goal of data visualization is the data-driven assistance in understanding complex situations or processes, making decisions, or developing models. As different as these goals are, the core task is always to convey to people information that is ‘hidden’ in potentially large data and to do this in a perceptually effective and cognitively favorable manner. The overall goal is to help people to carry out tasks more effectively.

## 6. SOME ASPECTS OF THE METHODOLOGY OF VISUALIZATION

The science of visualizing data uses knowledge on the theory of colors, the human visual system, and the human cognition to systematically derive visual metaphors for the correct, efficient and comprehensive recognition of data patterns and data relations.

The activity of visualization is schematized by the visualization pipeline. The visualization pipeline specifies the process chain by means of which data is converted into images. It consists of a sequence of functions for generating, filtering, and cleaning data; mapping data to geometries, colors, and textures; rendering those objects; and displaying the rendered image. All levels in the pipeline can be interactively performed and controlled. This means data visualization is not only about creating images but also about how to interact with visual representations. The design space is therefore huge.

When designing a visualization software, three kinds of resource limitations must be taken into account: those of humans, displays, and computers. Though the two latter items, in principle, can be scaled a bit further, in practical situations there are often hard limits. When designing a visualization tool, two major types, explanation and exploration must be considered.

Regarding the type of data: In visualization, abstraction from semantics (i.e., the real-world meaning of the data) and focusing on their structural properties must be done. There are different classifications in use; the following is just one example.

Most data sets are mappings. That map ‘points’ in a domain to ‘points’ or ‘attributes’ in some space. A simple example of domain is  $D = [1,2,3,\dots,m] \times [1,2,3,\dots,n]$ . If we assign to each  $(i,j) \in D$  some value, we get a table.

If we assume the intervals that build up  $D$  as (conceptually) continuous, we get a field. Of course, fields cannot be represented on the computer by enumerating their infinite values.

Therefore, fields are sampled in a finite number of sampling points, which together define some ‘grid’ in the domain. Depending on the ordering of the points in the domain, or the grid type, which ranges from an orthogonal equidistant Cartesian grid to an unstructured multiresolution grid, there is a large variety of fields.

Furthermore, the attributes can be more complex than just real values: they can be vectors (comprising a vector field, e.g., a flow field) or tensors (comprising a tensor field). Additionally, the attributes can be static or dynamic (time-dependent). Another typical domain is a graph which consists of nodes and edges. If values are assigned to the nodes and/or edges of the graph, we get an attributed graph.

Regarding the ‘attribute types’, categorical values (the data values represent names or categories) and ordered attributes may be distinguished. Within the ordered type, we find ordinal attributes (e.g., rankings) and quantitative attributes (e.g., integers and real numbers).

All these dimensions span a large space of possible data types. And, accordingly, an array of techniques have been developed to visualize data of these different types. For examples, see recent textbooks (Munzner, 2014; Mayer, 2018; Telea, 2015).

However, a visualization technique that serves well for one task can be poorly suited for another one. Therefore, when designing a visualization or ‘visual encoding’, the second dimension ‘task’ has also to be considered. Here again, an abstraction from the specific application must be done, aiming at singling out the basic types of elementary cognitive and productive tasks that users have to perform.

There are many suggestions on how to sort and categorize in the ‘task space’. Since no full agreement has been achieved, this issue will not be detailed herein. An overview on the categorizations of data and tasks, and the related visual encodings is provided in textbooks (Munzner, 2014; Mayer, 2018), particularly for spatial and spatiotemporal data (Telea, 2015).

Another important point not presented in this research is the fact that the true potential of visual data analysis can be exploited only with interaction. Users must have the possibility to interact with visual representations. Therefore, the design task consists not only in finding a suitable visual encoding but finding a suitable combination of interaction and visual encoding.

Powerful techniques that strongly rely on interaction have been developed in visualization. One specific example is linked to views that employ individual visualizations (displaying different aspects of the data) and are tied together through user interaction. Combined with ‘brushing’ (i.e., the selection of data items by simply brushing them with a virtual brush in one view) and then highlighting the selected data in other views, a sophisticated exploration of even high-dimensional data becomes possible.

## 7. DIGITAL TRANSFORMATION & VISUALIZATION

In order to achieve higher processing speed, lower error rates, and lower costs, processes in business, logistics, traffic, administration, and everyday life are delegated more and more to networked computer systems equipped with sensory systems (Industry 4.0, IOT, among others).

However, it must be avoided that humans are excluded from or become dependent on mechanisms and algorithms that are opaque to them. This means that the aim should be to make these computer systems transparent to the people who depend on them. In this context, visualization can surely help.

On the other side, experts who are implementing these systems will remain 'in the loop'. They need to gain a deep insight into these systems of growing complexity and the processes running on them. Therefore, it is to be expected that, similarly to computer security where data visual analytics has become an essential tool, digital transformation will also increase the need for data visualization.

## REFERENCES

- Bertin, J. (1967). *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*, Paris, Mouton/Gauthier-Villars
- Cook, K. A., & Thomas, J. J. (2005). *Illuminating the path: The research and development agenda for visual analytics* (No. PNNL-SA-45230). Pacific Northwest National Lab. (PNNL), Richland, WA (United States).
- Drebin, R. A., Carpenter, L., & Hanrahan, P. (1988). Volume rendering. In *ACM Siggraph Computer Graphics* (Vol. 22, No. 4, pp. 65-74). ACM.
- Levoy, M. (1988). Display of surfaces from volume data. *IEEE Computer graphics and Applications*, (3), 29-30.
- McCormick, B.H., DeFanti, T.A., and Brown, M.D. (1987). Visualization in Scientific Computing. *Computer Graphics* 21(6). New York: ACM SIGGRAPH.
- Meyer, M. & Fisher, D. (2018). *Making Data Visual*. O'Reilly Media
- Munzner, T. (2014). *Visualization analysis and design*. AK Peters/CRC Press.
- Telea, A. C. (2015). *Data visualization: principles and practice*. AK Peters/CRC Press.
- Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley.