Alexandru C. Telea Alexandru C. Telea Telea, Alexandru C.

Data Visualization Methods

Data visualization methods

461

464

Data Visualization Methods

The term *data visualization* refers to the process of transforming data, which is generated by means of measurements of various processes taking place in the physical world or created by computer applications such as simulations, to pictures. The purpose of this transformation is 2-fold: (1) to help users understand their data better and easier and (2) to let them discover unknown facts about the underlying phenomena from which data are derived. To be able to do this, several classes of visualization methods have been developed, each of them focusing on a specific type of data, users, or application domain. This entry describes the landscape of data visualization methods and their relationships with science and engineering. Particular attention is devoted to the increasing role of visualization in education. The entry concludes with an overview of the challenges of teaching data visualization as well as a list of resources on data visualization methods.

Aims and Scope of Data Visualization

The core aim of data visualization is to provide ways to handle information by means of the study of graphical representations thereof. This serves several goals, as follows. Large amounts of data can be compactly presented, therefore easing the user's burden of separating important aspects from details and also reducing the time and effort required to study a given process (*scalability*).

Nontechnical users can be shielded from complex aspects related to data acquisition and processing, so they can focus on those high-level aspects that the data capture (*simplicity*). Different types of users having different backgrounds are enabled to communicate and learn about a data-intensive problem based on the same (visual) medium (*communication*). Finally, the visual depiction of data enables finding complex patterns that one is not aware of and which are hard to find when using solely traditional data analysis methods (*discovery*).

Data visualization serves two main purposes. First, data can be *presented* to interested audiences, such as scientists, professionals, or students. In this context, the presenter uses visualization as an enabling instrument to facilitate communication, by reducing the complexity of the exposition and focusing on the essential details. This usage of visualization is the most widespread and covers presentation means as diverse as infographics, PowerPoint presentations, narrated videos, and web-based dashboards. Central to this use-case is that the presenter already knows the information to be communicated and chooses the visual instruments and presentation techniques that best suit an efficient and effective presentation. Consumers of such visualizations range from professionals and specialists to students at all education levels and, more generally, the grand public. This usage of visualization is also known as "presenting the known."

A second role of data visualization addresses data *exploration*. The aims of the audience for this role of visualization differ from the previous one: Both presenter and public are now interested in discovering previously unknown aspects embedded in a given data collection. As such, there is a less clear difference between presenters and public, so one typically speaks about visualization *users*. This usage of visualization relies upon more specialized instruments, such as advanced computer programs that allow users to interactively drill down in large data collections, select specific data subsets of interest, and depict them visually by a wide range of techniques, each of which focuses on the discovery of a different kind of pattern present in the data.

Consumers of such visualizations are professionals who are intimately familiar with the application domain from which the data at hand emerge and also with the aforementioned specialized visualization instruments. This usage of visualization is also known as "discovering the unknown."

Types of Data Visualization Methods

Before the advent of modern computing technology, collecting and visually depicting data has been done manually, leading to a variety of maps, graphs, and charts. The study of these depictions has led to the formulation of several key design principles for the creation of effective visualizations. As described by Jacques Bertin and Edward Tufte, these involve the choice of suitable *visual variables* (such as position, shape, size, color, brightness, texture, and animation speed) to encode variables in the data to be displayed, the maximization of ratio of displayed data to ink being used in the display, the avoidance of visual clutter, and the consistent use of legends and annotations that explain how data have been visually encoded. Data visualization can be seen as a blend between graphics design, visual perception, cognitive science, and data science.

The increasing computing power available to generate, collect, and analyze data and the increase in resolution and color quality of computer displays have shifted the focus from hand-crafted visualizations to computer-generated ones. Following these developments, data visualization has evolved and diversified since the 1970s into a number of subfields. These are outlined in the following sections.

Scientific Visualization

A first main use-case for computer-generated visualizations has been to explore the increasing amounts of data generated by numerical simulations of physical phenomena. The resulting discipline has been called *scientific visualization* due to its original presence in scientific, engineering, and research laboratories. Since its appearance in the 1970s, scientific visualization has evolved to cover the

visual analysis of data collections from geosciences, weather and climate science, and medical science. The main characteristic of scientific visualizations is that they target data that have a natural spatial embedding, such as surfaces and shapes that exist in two or three dimensions, and whose measured attributes are described on a continuous scale, such as length, density, or temperature. Key methods developed in the context of scientific visualization include techniques for the display of scalar data, such as temperature or pressure; vector data, such as force or velocity; and volume data, such as three-dimensional computer tomography scans.

Information Visualization

A subsequent main use-case for data visualization has been created by the explosion of data generated by developments in information technology, sensor devices, and the Internet with its various data-intensive facets such as social networks, online commerce, and e-governance. The rapid increase in volume, velocity of change, and variability of such data is currently known as the "3V" characteristics of so-called *big data*. To cope with this, new visualization methods have been developed. In contrast to scientific visualization, these aim at displaying data that do not have a natural embedding in two or three spatial dimensions and whose measured attributes are not necessarily continuous quantities. Examples of such data collections include tables stored in large data warehouses, document archives, Twitter feeds, and communication networks. Key methods developed in information visualization address the display of networks and hierarchies, tables having hundreds of columns and hundreds of thousands of rows, and trends and similarity relations in large time series such as stock exchange data. Used originally in business intelligence and homeland security contexts, information visualization has spread to many other application domains in which large collections of nonspatial data need to be explored, such as software maintenance, medicine, and bioinformatics.

Visual Analytics

Following the establishment of both scientific and information visualization, both practitioners and researchers have observed that the tools and techniques offered by these two subdisciplines are necessary, but not sufficient, for the end-to-end process of getting insight in complex data-intensive phenomena. As increasingly sophisticated tools have been provided by both scientific and information visualization, the complexity of the problems that such tools have to address has also increased. Getting insight into such problems has become more intricate than simply using one or a few suitable techniques to display the available data. For a given use-case, many exploration paths are possible, many hypotheses on the causes behind the observed data have to be investigated in parallel, and each such investigation can be done using multiple techniques.

Visual analytics has emerged as the answer to the explosion of the "search space" and "tool space." Rather than imagining novel visualization techniques, visual analytics focuses on ways to combine techniques of different kinds, such as data mining and searching, statistics, machine learning, and visualization, into realizing end-to-end pipelines that help users answer complex questions on the data at hand. Typical to visual analytics is the iterative refinement of knowledge and insight into the problem under study, which is done by the use of interactive visualization techniques. As such, and in contrast to earlier scientific and information visualization solutions, visual analytics focuses on the needs of an analyst to make sense of data, atop of the needs of a user interested solely in seeing the data. Following this analogy, visual analytics focuses on supporting the process followed by the user to extract knowledge, of which the images that depict data are just one component.

Data Visualization in Education

Visualizing data is a crucial part of technical and scientific communication, starting from the undergraduate level and ranging up to the senior researcher and practitioner level. Its tools of the trade range from relatively simple infographics like bar, line, and pie charts to sophisticated tools for the visual exploration of dynamic networks and data tables of millions of elements.

As such, visual communication has become an increasingly important element of technical and scientific education. Typical elements covered by this educational process are the visual design of slide presentations, infographics, and illustrations for scientific articles and of the associated narratives. Typical instruments supporting this process are presentation tools such as PowerPoint and Keynote, but also more advanced data visualization tools such as Matlab and R.

While constructing scientific and information visualizations has become increasingly easy due to the diversification of available software tools, important challenges still exist in educating the upcoming generation of scientists and practitioners to create effective visualizations. One such key challenge relates to the scope, or focus, of visualization education forms, which can be roughly divided into two types. The first type focuses on visual design, perceptual factors, and presentation techniques that are required by an effective visualization—or in other words how to design (but not how to implement) visualizations, so as to avoid "chart junk," or the creation of graphically rich, but information-scarce, visualizations. Such knowledge is provided in the context of many studies, ranging from exact sciences to cognitive and communication sciences and graphics design. This education type focuses heavily on the use of existing ready-to-use tools of low to moderate sophistication, such as PowerPoint, Keynote, and Tableau. As such, it serves a wide spectrum of users but cannot deliver customized visualizations for certain problems, data types, and user questions.

Conversely, the second type of visualization education focuses mainly on technical issues of all types of visualizations (scientific, information, and visual analytics)—or, in other words, how to implement a given visual design. This approach to visualization education covers aspects such as the choice and combination of data representation; data storage, processing, and data mining algorithms; and computer graphics and interaction techniques involved in the realization of an end-to-end visualization software application. This education type focuses on extending and adapting visualization software coming in the

form of libraries and frameworks and creating novel research-grade visualization methods. As such, it addresses a narrower spectrum of users than the first type, as a solid background in mathematics, statistics, and software engineering is required. The drawback of this approach is its relatively specialized and narrow focus in terms of addressing only a subset of visualization problems and requiring advanced technical skills from its students. An optimal formula for visualization education combines an entry-level course in visualization design general principles, followed by an in-depth, more specialized course covering the implementation of visualization techniques for a selected application area.

Many resources are available to study data visualization. On the design side, books by Edward Tufte and Jacques Bertin give an excellent overview of the requirements and guidelines for creating effective, compelling infographics. The books by Tamara Munzner and Colin Ware cover information visualization design. Andy Kirk offered a communication-sciences view on data visualization. Stephen Few's book focused treatment of information visualization with the accent on tables and graphs. Finally, Alexandru Telea's book covered the technical implementation aspects of data visualization, with a focus on scientific visualization, and also provides a survey of popular visualization software tools and applications.

Alexandru C. Telea

See also Active Learning; Data-Driven Decision Making; Data Mining; Graphical Modeling; Multidimensional Scaling; R

Further Readings

Bertin, J. (2010). Semiology of graphics: Diagrams, networks, maps. Esri Press.

Dill, J., Earnshaw, R., Kasik, D., Vince, J., & Wong, P. C. (2012). Expanding the frontiers of visual analytics and visualization. Springer.

Few, S. (2012). Show me the numbers—Designing tables and graphs. Analytic Press.

Kirk, A. (2012). Data visualization: A successful design process. Packt Publishing.

Munzner, T. M. M. (2015). Visualization analysis and design. CRC Press.

Telea, A. C. (2014). Data visualization—Principles and practice (2nd ed.). CRC Press.

Tufte, E. (1983). The visual display of quantitative information. Cheshire, CT: Graphics Press.

Ware, C. (2012). Information visualization: Perception for design. Morgan Kaufman.