7 things you should know about... Data Visualization

Scenario

Olivia is a graduate student in kinesiology and also an assistant coach for the women's track team at her university. As part of her research, she has worked with faculty in the computer science department to customize a data visualization tool to represent a range of variables that affect maximum athletic performance.

Members of the track team wear heart rate monitors that record heart rate every five seconds when the runners train. The devices also record resting and maximum heart rates, and all of the data can be downloaded to a computer. The track team members are also tested in the kinesiology lab for factors such as lactate threshold, blood oxygenation, and many other data points. Olivia's data visualization tool uses this data to create a "picture" of an athlete's performance. Each image looks something like a 3D map of hilly terrain, with color differences that highlight the contours of the picture. Peaks in the imagery represent efficient performance. The runners can see in visual terms how their performance is affected by going above their aerobic thresholds, for example. During treadmill workouts, the athletes can watch a computer screen that displays visualizations based on data collected in real time. In this way, the runners can watch a representation of their efficiency and see how it changes as they modulate their effort, for example, or change their breathing patterns.

Over the course of a semester, Olivia collects data for 10 members of the track team, showing them visualizations of how they perform in numerous workouts, both on the track and on treadmills in the lab. The application generates custom images for each athlete, who can look at the images to "see" what her effort looks like and identify the factors that most benefit and hurt her performance. One runner discovers that in middle-distance events, her performance is maximized by a very even effort, whereas a teammate—whose body is able to recover faster—finds that in the same event, her best time comes from varying her intensity. The visualizations show correlations between weight training and aerobic capacity that surprise the runners and coaches, leading to changes in the training program for the whole team. By the end of the semester, each of Olivia's subjects has a better understanding of which training methods are most appropriate for her, and all of them have recorded faster times in their favorite events.

1 What is it?

Data visualization is the graphical representation of information. Bar charts, scatter graphs, and maps are examples of simple data visualizations that have been used for decades. Information technology combines the principles of visualization with powerful applications and large data sets to create sophisticated images and animations. A tag cloud, for instance, uses text size to indicate the relative frequency of use of a set of terms. In many cases, the data that feed a tag cloud come from thousands of Web pages, representing perhaps millions of users. All of this information is contained in a simple image that you can understand quickly and easily.

More complex visualizations sometimes generate animations that demonstrate how data change over time. In an application called Gapminder, bubbles represent the countries of the world, with each nation's population reflected in the size of its bubble. You can set the *x* and *y* axes to compare life expectancy with per capita income, for example, and the tool will show how each nation's bubble moves on the graph over time. You can see that higher income generally correlates with longer life expectancy, but the visualization also clearly shows that China doesn't follow this trend—in 1975, the country had one of the lowest per capita incomes but one of the longer life expectancies. The animation also shows the steep drop in life expectancy in many sub-Saharan African countries starting in the early 1990s (corresponding to the AIDS epidemic in that part of the world) and the plummeting of life expectancy in Rwanda at the time of that nation's genocide.

2 Who's doing it?

Data visualizations have long been used in academic settings, but many instructors are using new technologies to create rich, compelling visualizations to help students understand concepts more quickly and deeply. A history professor, for example, could use a visualization that shows which industries prospered and which suffered during the wars and economic cyclwes of the 20th century to explain demographic shifts and the social changes that followed. An economics professor might use the same visualization toexplain the financial connections in the national or global economy. Working with the visualization team at the Renaissance Computing Institute, a faculty member in the Department of Soil Science at North Carolina State University created an animated, interactive visualization that shows how fertilizer nitrates enter groundwater. The project combines data from a network of GIS

more ➪

EDUCAUSE Learning Initiative
www.educause.edu/eli

mapping systems and remote sensors and generates a visualization that shows where nitrates concentrate in soil and how different modes of fertilizer delivery—coupled with variables such as precipitation—affect the rates and locations of groundwater pollution. Faculty and researchers in a wide range of academic disciplines use visualizations to present data in ways that help generate new knowledge and understanding.

3 How does it work?

Visualizations encompass a wide and growing range of projects, reflecting creative ways of representing all sorts of data visually, with virtually no limit to what kind of information can be translated into an image. Visualizations have been created that represent the possible moves on a chess board, the structure of a piece of music, and the messages in a person's e-mail inbox, to name a few. The designer of a visualization determines which visual element (color, shape, size, motion, and so forth) will represent individual data points. Images can be 2D or 3D, can be fixed or dynamic, and can allow user interaction. One application, for example, shows political contributions to various candidates. You can select a state, a political race in that state, and a monetary threshold for contributions. The application builds a 2D image that shows who supported each candidate and at what level (based on the relative sizes of the circles that represent contributors), revealing interesting webs of political influence. Astrophysicists use visualizations to create 3D images that model the forces of a supernova. In each case, images or animations are the products of applications that render data in a visual form based on the design of the visualization.

4 Why is it significant?

Computer systems generate and store massive and growing amounts of data. At the same time, advanced networks, distributed processing, and other developments allow unprecedented access to data. Data visualizations offer one way to harness this infrastructure to find trends and correlations that can lead to important discoveries. Representing large amounts of disparate information in a visual form often allows you to see patterns that would otherwise be buried in vast, unconnected data sets. As opposed to the traditional hypothesis-and-test method of inquiry, which relies on asking the right questions, data visualizations bring themes and ideas to the surface, where they can be easily discerned. Visualizations allow you to understand and process enormous amounts of information quickly because it is all represented in a single image or animation. Moreover, virtually any kind of data from a broad range of academic disciplines can be represented visually, making data visualization a potentially valuable approach to learning for a large number of students and researchers.

5 What are the downsides?

Visualizations rely on accurate and matched data. If data are incomplete or faulty, or if data sets use different definitions or units, these issues must be resolved in order to create a valid visualization, and this can be time-consuming. Even if the data are reliable and consistent, a poorly conceived visualization might show nothing of consequence or exaggerate the significance of certain trends, resulting in flawed or misleading conclusions. In some cases, a lot of time and trouble go into a visualization that adds nothing to an understanding of the data that you wouldn't find in a simple table or even a textual description. Finally, users who prefer conventional ways of learning and processing information might be uncomfortable working with data visualizations, which require a different approach to understanding data.

6 Where is it going?

As data visualization tools become more powerful and more widely available, users will create representations of the relationships between data not previously considered together. Some of these experiments will show interesting relationships that can lead to new understanding about correlations and causality between a wide range of variables. Data visualizations will take greater advantage of animation tools, allowing users to control the parameters of those animations, and visualization projects will increasingly be cross-disciplinary, allowing experts in diverse fields to seek and express relationships between their respective areas of inquiry. One project, for example, compares data about individuals' stress levels with their geographic location. Many mashups employ data visualization to represent hidden connections, such as images that show the complex networks of users on social networking sites. These and other applications will broaden the general sense of how data can be read and interpreted.

7 What are the implications for teaching and learning?

Many data visualization projects allow users to develop as well as interpret images. Involving students in the formulation of visualizations requires them to think in creative ways about the material at hand and how the information can be effectively communicated through shapes, colors, animations, or other visual elements. By examining data and deciding how they are presented, students might have a stronger sense of ownership of that data. As consumers of visualizations, students have the opportunity to break out of traditional modes of thinking about data and challenge themselves to ask critical questions about what data do and do not show and about relationships between sets of data. Visual literacy is an increasingly important skill, and data visualizations are another channel for students to develop their ability to process information visually. For students and researchers, visualizations can be an important data-analysis tool, uncovering key findings that would otherwise be difficult or impossible to perceive.

EDUCAUSE Learning Initiative
advancing learning through IT innovation
www.educause.edu/eli
October 2007