7 things you should know about...

Data Visualization II

Scenario

One requirement of Vera's master's program in sociology is a thesis, which includes a presentation of findings before her peers for review, discussion, and recommendations. In addition to her own clinical project, which traces correlations between environmental factors and learning disabilities in the United States, Vera gathers 12 years of data on learning disabilities from organizations in the United States, Canada, New Zealand, the Netherlands, South Africa, Australia, and Great Britain. As she begins looking over the paper with an eye toward presentation, she becomes convinced that her statistics, alone, will not clearly articulate the implications of her research to an audience of listeners, so she starts looking for ways to present the information visually.

She begins by using the Visual Understanding Environment, a concept-mapping tool to create a simple knowledge map that illustrates how complex clusters of symptoms characterize specific disabilities. Then she visits IBM's Many Eyes site, searching for examples in her field and ideas for presenting her statistics visually. There she finds a chart correlating a central auditory processing disorder with exposure to heavy metals. She initiates an online discussion with the creator of that chart, who makes suggestions about how she might clearly present her findings. Soon she has a map of the United States showing clusters of learning disabilities that have been associated with environmental toxicity.

For her conclusion, she hopes to convey her concerns about under-diagnosis of key disabilities. She enters the data she collected into a Google Spreadsheet and exports it into Google Motion Chart. The result is an animated chart that Vera runs several times during her presentation to contrast the extremely low number of dyspraxia diagnoses in the United States compared to the number in the six other countries in her data set.

In discussions with her classmates after the presentation, Vera learns that the charts, maps, and animations helped the audience understand the research and findings. What surprises her, however, is the extent to which working with various visual mapping tools to illustrate the data had clarified for her just what the data meant, further advancing her understanding of what was going on in the field.

1 What is it?

Data visualization is the use of tools to represent data in the form of charts, maps, tag clouds, animations, or any graphical means that make content easier to understand. The past two years has seen a blossoming of visualization applications, as well as of technologies and infrastructure to support increasingly sophisticated visual representations of data. The greatest change, however, may be in access to data. Electronic sensors, for example, have made weather information available on a previously unimaginable scale. While geographic information systems (GIS) have for years allowed individuals to gather, transform, and analyze data, new tools have become widely available that easily create unique mashups of disparate sources of data, as evidenced by the increasing number of applications that employ Google Earth and Google Maps. Growing access to information from education, government, astronomy, geology, medicine, and news offers an increasingly widening pool of data that can be combined to create impressive visuals ranging from cartography to cartoons.

2 Who's doing it?

Data visualization tools are popular among those who use social networking sites—on Facebook, for instance, users can create "friend maps" (digital ball-and-stick representations that show networks of friends), while a Flickr mapping function lets photographers easily show where they took photos. Twitter users can post from visual decks like Twittervision, where tweets appear on a world map, or they can use tools such as Stweet, a mashup of Twitter with Google Street View that shows a photo of the street from which the tweet is sent.

In academe, some users have turned to VUE (Visual Understanding Environment), a concept-mapping tool designed at Tufts University that facilitates creation of knowledge maps. At Columbia University, Professor Peter Eisenberger encourages students in his interdisciplinary course, The Earth/Human System, to use VUE to create maps linking the complex problems of sustainability to issues in their own fields of biology, physics, and social sciences. More widely known is the IBM project Many Eyes, which provides both a suite of visualization tools and a public forum for people to share the data visualizations they create. Google Maps and Google Earth have made their way into classes and other places where association of data with geography is valuable, as when

more ⇨

EDUCAUSE Learning Initiative
www.educause.edu/eli

students at the University of British Columbia created a Google Maps visualization to help others locate healthy local food sources in "food security" assessments of neighborhoods in Vancouver.

How does it work?

Creating a visual representation of statistics once involved compiling data, interpreting it, parsing it, and then determining what kind of visual presentation would best elucidate what the data meant. New data visualization tools provide a shortcut—a straight line from compiling data to illustrating it. The simplest visualization tools function as a kind of black box. The application presents an interface where users can paste data or link to a data source, spreadsheet, or RSS, and the tool then turns the data into a graphic, such as an animated map, an interactive chart, or a word cloud. More highly skilled users can use a wide range of technology tools to create almost any kind of visualization they can conceive, but the importance of newer tools is that they allow nontechnical users to present data from any discipline—or from multiple disciplines—without having to learn how to use complex modeling or multimedia software. Finding a tool that can turn data into an appropriate, informative visualization can be difficult, but a number of people use the library of applications at IBM's open-source venue Many Eyes or the Visualization Lab available from the *New York Times.*

Why is it significant?

Graphic representations of data are popular because they open up the way we think about data, reveal hidden patterns, and highlight connections between elements. Traditionally, it was researchers who designed visuals to make trends clear to an academic or lay audience. Now, because current web applications allow anyone with access to data to enter information and easily create a virtualization of it, students, informal learners, and the purely curious can now easily create visualizations that might reveal trends that are not obvious from the numbers alone. Such easy-to-access tools could simplify the interpretation of complex data sets and encourage cross-disciplinary interpretation. Whereas visualizations once were often too complex for quick assimilation, tools that create interactive visualizations provide users with some measure of control over how—and how quickly—the information is presented, making complex patterns easier to perceive and understand. Wordle and some of the tools at Many Eyes create visual montages of words, sentences, phrases, or paragraphs uploaded and processed so that the audience, in examining the end result, sees text in a new light.

What are the downsides?

Considering the rapid rise in information available and the power and ubiquity of web-based applications, there is a remarkable scarcity of compelling visualization examples that leverage these opportunities. While many tools are easy to use and turn out engaging graphics, there is a risk that a resulting image might not illustrate the most significant implications of the data. Similarly, these tools might encourage students to substitute gratuitous graphics for more meaningful content in their presentations. The low barrier of entry could make output from these tools vulnerable to over-interpretation or misinterpretation. Because individuals unaccustomed to interpreting statistics can plug data into one of these tools without a clear idea of its significance, the resulting visual information might shift focus to curious anomalies rather than capture a valid trend. As a result, the graphic representation could mislead those inexperienced in the use and interpretation of statistics.

Where is it going?

Data visualization has become more ubiquitous as familiarity with web tools has increased and as people have come to terms with the sheer volume of web-accessible data. Data overload from online public records, sensor data from weather stations accessible from any keyboard, and readily accessible educational and scholarly resources represents a growing challenge for academics and researchers, and so it makes sense to expect the data to drive the tools in the next few years. As general audiences become more discerning and comfortable with increased animation and complex interactivity, they will expect visualizations to map effectively to the data they are supposed to represent. Even as wall-size screens at public venues invite presenters to fill them with complex graphics, mobile devices will continue to require visual adaptation to their smaller viewing screens. Meanwhile, graphical representations, in whatever form they take, will be expected to clarify the narrative in an environment that combines increasingly sophisticated multimedia presentation with ever-increasing amounts and types of data.

What are the implications for teaching and learning?

For scholars, particularly those whose conclusions depend on interpretation of complex statistics, data visualization offers the promise of easier communication and a wider audience for their findings. As visualization tools become easier to implement, educators and students stand to reap considerable benefit. Educators can present their points in an increasingly engaging form, and students making presentations or delivering papers have the opportunity to graphically incorporate statistics to make points that their peers can easily understand. In addition to adding considerable value by making both meaning and data accessible, visualization tools allow end users to personalize or take ownership of data. Studying an engaging visual representation of information can help bring data to life and draw it into the reach of non-experts.