Books and Arts

Big data: Distilling meaning from data

Felice Frankel[1] & Rosalind Reid[2]

Buried in vast streams of data are clues to new science. But we may need to craft new lenses to see them, explain Felice Frankel and Rosalind Reid.

It is a breathtaking time in science as masses of data pour in, promising new insights. But how can we find meaning in these terabytes? To search successfully for new science in large datasets, we must find unexpected patterns and interpret evidence in ways that frame new questions and suggest further explorations. Old habits of representing data can fail to meet these challenges, preventing us from reaching beyond the familiar questions and answers.

To extract new meaning from the sea of data, scientists have begun to embrace the tools of visualization. Yet few appreciate that visual representation is also a form of communication. A rich body of communication expertise holds the potential to greatly improve these tools. We propose that graphic artists, communicators and visualization scientists should be brought into conversation with theorists and experimenters before all the data have been gathered. If we design experiments in ways that offer varied opportunities for representing and communicating data, techniques for extracting new understanding can be made available.

D. ARMENDARIZ

Discussing visual communication before designing experiments may reveal new science.

Visual representation is familiar in data-intensive fields. Years before a detector is built for a facility such as the Large Hadron Collider near Geneva, for example, physicists will have pored over simulations. They examine how important events will 'look' in the displays that reveal and communicate what is going on inside the machine. Such discussions tend to take place within the visual conventions of a field. But perhaps conversations might be broadened to consider alternative representations of the same data. These might suggest other approaches to

collecting, organizing and querying data that will maximize the transparency of experimental results and thus aid intuition, discovery and communication.

Unfortunately, visualization experts and communicators are often consulted only after data are organized and stored, in the hope that they will create effective computer displays, slides and figures for publication. Meanwhile, they may be developing their tools in isolation, kept at arm's length by scientists who are busy getting their experiments done. Opportunities for useful dialogue are thus squandered.

When scientists, graphic artists, writers, animators and other designers come together to discuss problems in the visual representation of science, such as at the Image and Meaning workshops run by Harvard University (**http://www.imageandmeaning.org**), it becomes clear that representations repeatedly fail to communicate understanding or address obvious questions about the underlying data. A three-dimensional volume rendering may give no hint of important uncertainties or data gaps; solid surfaces or sharp edges may suggest data where they do not exist. A graphic artist might propose ways to reveal gaps or deviations from expectation early in an experiment, guiding subsequent data collection or highlighting new avenues of enquiry. When we asked Harvard University chemist George Whitesides to change the geometry of a self-assembled mono layer with clearly delineated hydrophobic and hydrophilic areas to create an image for submission to a journal, he found himself redesigning the experiment, and unexpected science emerged.

Student workshops and exercises, such as those run by the US National Science Foundation's Picturing to Learn project (**http://www.picturingtolearn.org**), teach us that attempting to visually communicate scientific data and concepts opens a path to understanding. When science and design students collaborate, their drive to understand one another's ideas pushes them to create new ways of seeing science. Investment in visual communication training for young scientists will pay off handsomely for any data-intensive discipline.

The ingrained habits of highly trained scientists make them rarely as adventurous as these young minds. We think we are on the path to insight when shading reveals contours in 3D renderings, or when bursts of red appear on heat maps, for example. But the algorithms used to produce the graphics may create illusions or embed assumptions. The human visual system creates in the brain an apparent understanding of what a picture represents, not necessarily a picture of the underlying science. Unless we know all the steps from hypothesis to understanding — by conversing with theorists, experimentalists, instrument and software developers, visualization scientists, graphic artists and cognitive psychologists — we cannot be sure whether a display is accurate or misleading.

The greatest opportunity and risk lie in that last step in the path: understanding. Whether verbal or visual, any language that is garbled and inconsistent fails to do its job. Let's talk. Let's all talk.

See Editorial, page 1.

1. Felice Frankel is senior research fellow in the faculty of arts and sciences at Harvard University, Cambridge, Massachusetts 02138, USA. With G. M. Whitesides, she is co-author of *On the Surface of Things: Images of the Extraordinary in Science*. Email: **felice_frankel@harvard.edu**
2. Rosalind Reid is executive director of the Initiative in Innovative Computing at Harvard University and former Editor of *American Scientist*.