

Visualizations Make Big Data Meaningful

New techniques are designed to translate “invisible numbers” into visible images.

UNTIL RECENTLY, a good spreadsheet or perhaps a pie chart might have been sufficient to get a firm grip on a dataset. However, making sense of big data requires more, and with our increasing inundation with data comes new and creative opportunities to build unique interfaces.

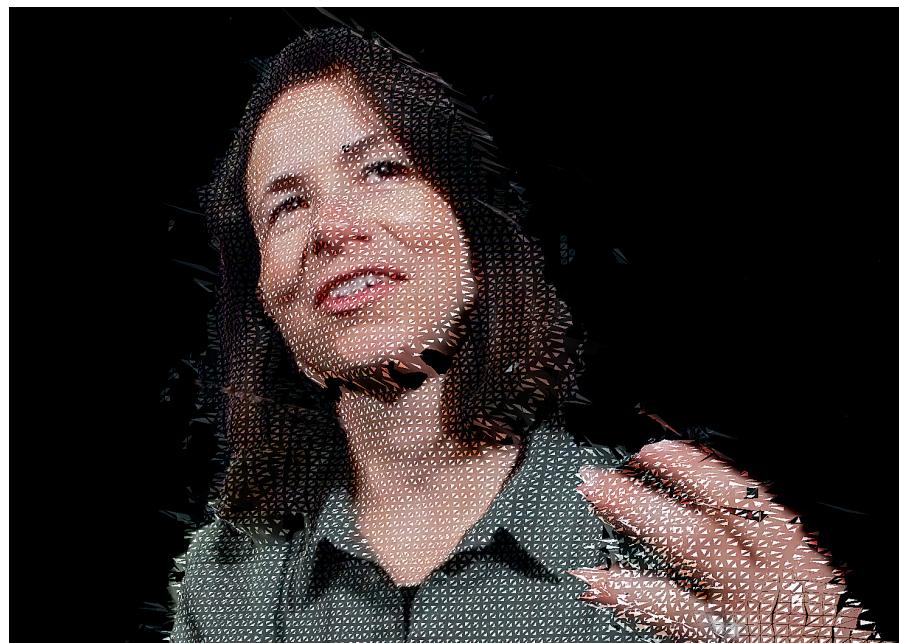
That is why, powered by new technologies such as touchscreens and giant LCD monitors, so-called “data artists” are experimenting with ways to make huge amounts of data comprehensible and accessible to a broader range of people, and then displaying their data visualizations inside venues like the corporate facilities of companies like Microsoft and Google.

Whether people are familiar with the term “data visualization” or not, they are certainly aware of some of the more popular implementations of it. For instance, the Facebook Timeline is a visualization tool designed by data artist Nicholas Felton as a way of manipulating and organizing the information in Facebook’s database.

“Those who create data visualization are providing the beautiful, easy-to-use interfaces that lie on top of massive amounts of number crunching from a growing array of sources,” observes Justin Langseth, CEO of Washington, D.C.-based startup Zoomdata.

Indeed, like all art, data visualizations come in many shapes and flavors—from the inventive software-based work of Jer Thorp to the big-screen animated films from Pixar Animation Studios; from the business solutions of Zoomdata to the documentary “Clouds” by James George on the emerging field of data visualization (filmed using Microsoft Kinect’s Depth camera paired with a digital SLR video), which premiered recently at the Sundance Film Festival.

In fact, George’s work in video and



An image from “Clouds,” a “computational documentary” by James George in collaboration with Jonathan Minard, which depicts people discussing creative use of code.

photography epitomizes the relationship between art and computer science. Besides being an independent artist, he is also the first artist-in-residence at Microsoft Research’s Studio 99 Gallery, and is a lecturer on computational processes in video art at New York University’s Interactive Telecommunications Program.

One of George’s experiments in computational photography began with a story he heard on the radio about a surveillance system that was to be installed to combat crime in the New York City subway system. Cameras were to be mounted in all subway stations and, instead of having the monitors watched by people, AI algorithms would be used to identify threats.

“The system failed miserably because there were too many passengers, too much data being generated,” George recalls. “So the system was switched off, but the cameras are still

there today, attached to computers that aren’t turned on.”

George and his team had been experimenting with the Microsoft Kinect camera, which sees the world in three dimensions, transforming it into a dataset that can be manipulated in post-production.

“We took the camera into the subway system and began capturing image data with the intention of re-creating the way the security system might have seen the world if it had worked,” he explains. “We called it ‘DepthEditorDebug,’ after the name of the app we used to make it, and what it did was illustrate that, as in any art form, there is a level of interpretation depending on what you leave in, what you leave out, or what you fix—known as ‘cleaning the data.’ So data visualization is a very expressive medium in the same way that painting can be expressive. The role of the artists is to apply their insight to the data and then choose

what to show and how to show it."

A co-founder of The Office For Creative Research, an R&D group for hire, Jer Thorp says his team is hired by corporations and museums to approach novel problems concerning data and to engage with that data in ways that probably have not been done before. His two biggest clients are currently Microsoft and the Musum of Modern Art (MOMA) in New York City (where he and his colleagues collectively serve as artists in residence).

"Regardless of whom we work for, everyone is being overwhelmed by data, and the promise of simplification becomes really attractive," Thorp says. "Being able to see a single graphic that represents a complicated thing makes peoples' cognitive load a little easier."

The ability to lighten that load, he says, comes from three advances in technology: the fact that storage of data has become almost free; the cloud and GPU-driven calculations that have opened computing to the masses; and the availability of software tools like Apache Hadoop that have simplified the processing of large-scale datasets.

Take, for example, the project Thorp completed last year for the Vancouver

"Being able to see a single graphic that represents a complicated thing makes people's cognitive load a little easier."

Art Gallery called "Grand Hotel." Starting with a database of about 700,000 hotels—not every hotel in the world, but pretty close—he and his team wondered what would be the most engaging way to visualize all that information; the hotels' ratings, their locations, photos, and so on.

"We could have plotted points on a map or built graphs around their room rates," he says, "but instead, we imagined what it would look like if characters from famous literature were to make their travels today;

where would they stay, what would they think of the hotels. It was a way to display the data we had, but in an entertaining fashion."

They started with characters from the book *Lolita* and had them travel down the eastern coast of the U.S., stopping at rundown hotels as they do in the novel, with information about those hotels projected on a 16-by-30-foot screen.

"Then we got more abstract and did the same with Ulysses and his army, plotting all the points where, theoretically, he would have stayed at hotels," Thorp says. Then they did the same with *Lawrence Of Arabia*, Jack Kerouac's *On The Road*, and other traveling characters of literature.

"The data visualization is in some ways a data narrative," says Thorp. "It explains, but it also relates. It intrigues. It compels. It allows us to unpack that dataset in a way which I feel is a lot more human than it might have been if it were just, say, a chart or spreadsheet."

Humanizing the use of leading-edge technology is also the task of Tony DeRose, who heads up the Research Group at Pixar, the animation

International

Panel Mulls Outlook for European CS Research

A special open session during a recent ACM Europe Council meeting at the Royal Society in London included a panel discussion moderated by former ACM president Dame Wendy Hall on "The Future of Computer Science Research in Europe."

The panel, made up of representatives of government, industry, and academia, addressed weaknesses in Horizon 2020, a program in which the European Union has pledged to invest nearly 80 billion euros (about \$110 billion) through 2020 to ensure Europe's global competitiveness in research and innovation. As noted in the Information Society Technologies Advisory Group (ISTAG) Working Group on Software Technologies report "Software Technologies: The Missing Key Enabling Technology," the Horizon 2020 plan identified numerous "key enabling technologies" but failed

to include software development as one of them.

Chris L. Hankin of Imperial College, who is on the European Commission Directorate General for Communications Networks, Content & Technology (DG Connect) Advisory Forum (CAF), addressed the Digital Agenda for Europe, the EU's strategy to help digital technologies deliver sustainable economic growth to "reboot Europe's economy." As the first of seven flagship initiatives under Europe 2020, the EU's strategy to deliver sustainable and inclusive growth, Hankin said the Digital Agenda aimed to double ICT (information and communication technologies) R&D spending from 5.5 billion euros to 11 billion euros by 2020.

He added, "It's really the responsibility of groups like ACM Europe that that kind of investment is properly directed and achieves the kind of impact that the Digital

Agenda for Europe foresaw."

Robert Madelin, EC DG Connect's director general, said, "I think we're going in the right direction, but the 'right direction' has to be more explicitly open to cooperation upstream with different interests, and it has to be more engaged with ordinary people and issues."

Paola Inverardi, first female rector of the University of L'Aquila, Italy, and Italy's representative to the EC for ICT, says Horizon 2020 "to some extent reduces ICT and computer science to an ancillary focus."

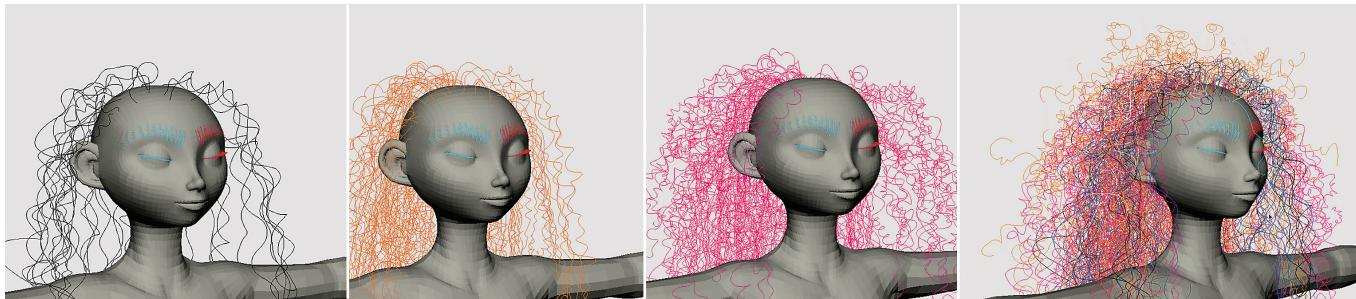
Last year's High-Level Vision 2030 report by industry association ARTEMIS projected jobs in software and services for Europe will reach about 8.9 million, Inverardi said, compared to less than a million working on semiconductors and other hardware. "We have a mass of people working on the soft part of

ICT, and I think that we need to take care of their future."

Microsoft Research Connections vice president Tony Hey stressed the importance of teaching computer science to "stimulate young people into the potential of science." However, he added, "it seems absolutely incredible if you don't have a serious computer science research agenda in Horizon 2020. That's where your future is going to be."

Andreas Reuter, managing director of the Heidelberg Institute for Theoretical Studies, said, "I think funding agencies should not primarily try to fund directions of research ... they should rather address the environmental conditions, political, legal, economical infrastructure that will foster leading-edge research and will support the take-up of research results by industry and society."

—Lawrence M. Fisher



The challenges of visualizing lifelike hair in the animated fantasy *Brave*.

studio known for its big-screen features. “Our group tries to chip away at the big, hard, technical challenges coming down the pike,” he says, “especially the ones where the probability of failure is high.”

One such challenge that sounds simple was creating lifelike hair on Merida, the main character in *Brave*, the 2012 animated fantasy about a princess in the Scottish Highlands. To create her mass of long, bouncy red hair that had to fall on her shoulders in a believable fashion, DeRose’s team needed to create a mathematical model to describe how hair should move, and then devise a simulation system that would use that model to create the on-screen motion.

“That subtle motion is something an animator isn’t going to want to go in and control every little detail by hand,” DeRose explains.

One of the interesting applied math/computer science questions was what gives real hair its volume. It turns out that approximately 100,000 hairs rest on each other, colliding with the hairs around them as a head moves. “The potential number of collisions is enormous—about 100,000 squared potential collisions,” DeRose says. “So we needed to develop algorithms that would quickly determine which subset of hairs to check for collisions. If we had to do all of that by hand, we’d still be waiting for the images to complete themselves.”

What has enabled such data manipulation, says DeRose, is the computing power “that has gone through the roof in the last 10 years. GPUs being developed today are opening up new levels of complexity that we’ll be able to bring to the screen. As John Lasseter, our chief creative officer, has said, ‘the artistic vision inspires new technical development...and new technical development

inspires new art.’ And so we get this nice, positive feedback cycle between the two disciplines, technology and art.”

Meanwhile, in the world of commerce, companies like Zoomdata are using data artists to design interfaces strictly for business purposes. “The business world is pretty adept at creating pie charts and bar charts,” says Zoomdata CEO Langseth, “but it thirsts for more interesting, more intuitive interactive visualizations of data. That’s the space we’re in: coming up with novel, dynamic ways of looking at the data, as opposed to staring at tables and columns.”

For instance, one Zoomdata “visualization” is a cascading chord chart that displays connections from one place to another, perhaps the flow of traffic on a Web site from one area of the site to another.

“The most competitive retailers, for example, are using advanced analytics to track customers and to target advertisements very specifically to the information they have on those customers,” says Langseth. “And that involves new sources of data and new ways of looking at and exploring that data.”

“Since Apple reinvented the phone, many Silicon Valley startups have three co-founders: a technologist, a business person, and an artist who has recently become critical to almost everything people do with computers,” he says. “And data is no exception. It’s not something that’s received much artistic attention over the years, but now companies like ours are starting to focus on that.”

Indeed, Google is sponsoring a global competition to find an up-and-coming software developer who pushes the boundaries of art using data and code, which Google calls “DevArt.” The winner of the competition will receive a \$41,000 prize, as well as Google Developer assistance

and curating, and production support from the Barbican, Europe’s largest multi-arts venue in London, to help transform their concept into a digital art installation.

“I sometimes hear phrases like ‘artists create, developers code,’ but nothing could be further from the truth,” says Google developer advocate Paul Kinlan in a recent blog. “We are all a creative bunch with a passion for exploring and creating amazing works that push the boundaries of what we believe is possible with modern computing technology. Sometimes, we just need some inspiration and an outlet.”

Further Reading

“DepthEditorDebug,” a video and photographs posted February, 2011 by James George and Alexander Porter, at <http://jamesgeorge.org/works/deptheditordebug.html>

“Clouds,” a video documentary posted January, 2014 by James George and Jonathan Minard, at <http://jamesgeorge.org/works/clouds.html>

“Jer Thorp: The Weight Of Data,” a video posted February, 2012 by TEDx, at <http://www.youtube.com/watch?v=Q9wcvFkWpsM>

“Math In The Movies,” an article posted by the Mathematical Association of America, at <http://www.maa.org/meetings/calendar-events/math-in-the-movies>

“Pixar And Math – Disney Pixar’s *Brave* – Wonder Moss,” a video posted November 2012 by Inabottle, at <http://www.youtube.com/watch?v=EnaA9ZRPXiE>

“Data As Paint, And The Rise Of The Data Artist,” a blog posted March, 2013 by Justin Langseth, at <http://datametaphors.wordpress.com/>

“DevArt: Your Code Belongs In An Art Gallery,” a blog posted February, 2014 by Paul Kinlan, at <http://googledevelopers.blogspot.com/2014/02/devart-your-code-belongs-in-art-gallery.html>