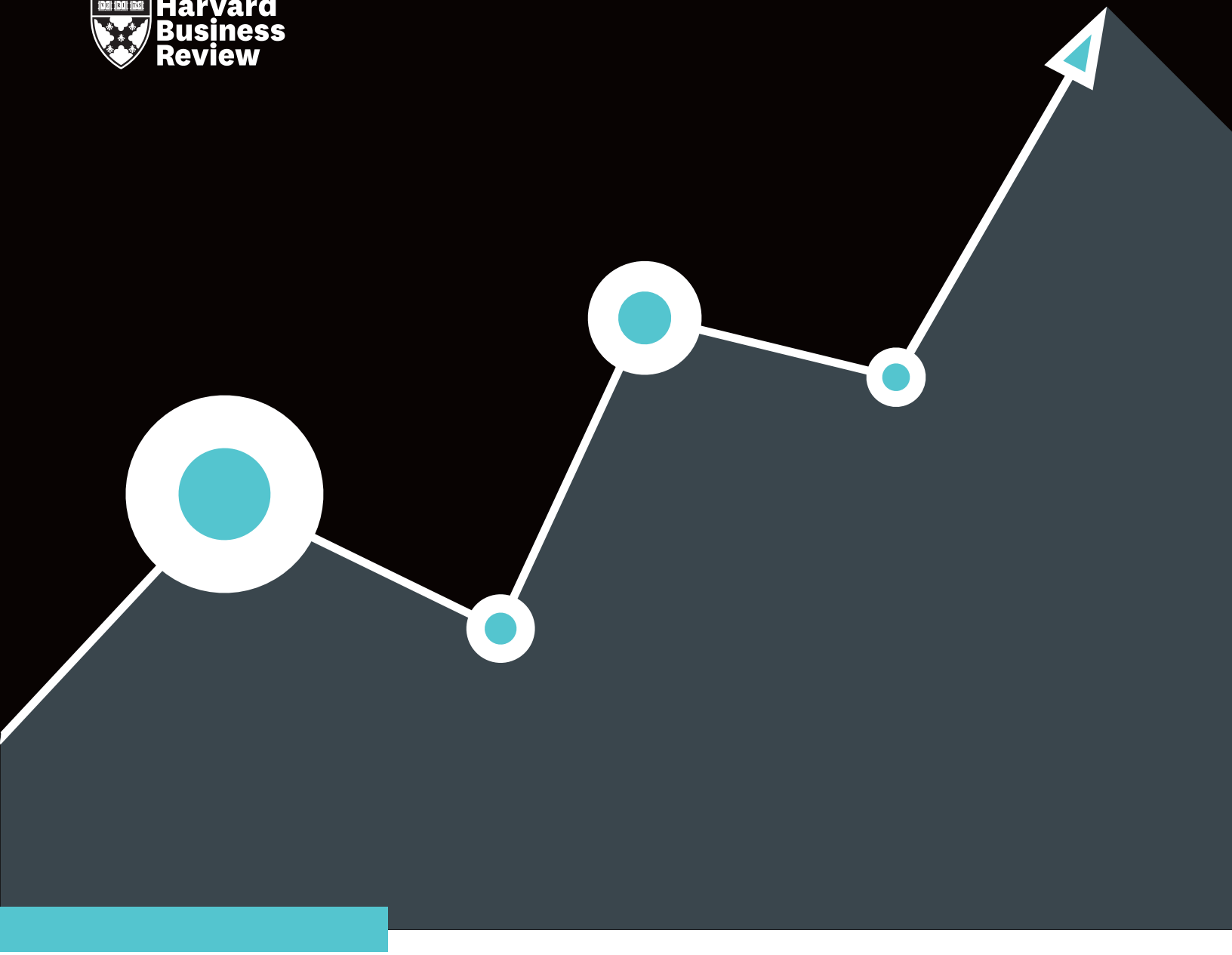
PERSUADING WITH DATA

A HARVARD BUSINESS REVIEW INSIGHT CENTER REPORT

Harvard Business Review



Sponsored by

sas

PERSUADING WITH DATA

The HBR Insight Center highlights emerging thinking around today's most important business ideas. In this Insight Center, we'll look at best practices for using graphics, data, and storytelling to make spectacular business cases, get organizational buy-in, train talent, and win customers. Topics will include selecting the most relevant and compelling metrics, emerging practices and common pitfalls of data visualization, and more.

HOW TO TELL A STORY WITH DATA

BY JIM STIKELEATHER

An excellent visualization, according to Edward Tufte, expresses "complex ideas communicated with clarity, precision and efficiency." I would add that an excellent visualization also tells a story through the graphical depiction of statistical information. As I discussed in an earlier post, visualization in its educational or confirmational role is really a dynamic form of persuasion. Few forms of communication are as persuasive as a compelling narrative. To this end, the visualization needs to tell a story to the audience. Storytelling helps the viewer gain insight from the data. (For a great example, how much do you think steroids have influenced baseball?)

So how does a visual designer tell a story with a visualization? The analysis has to find the story that the data supports. Traditional journalism does this all the time, and journalists have become very good at storytelling with visualization via infographics. In that vein, here are some journalistic strategies on telling a good story that apply to data visualizations as well.

1. Find the compelling narrative. Along with giving an account of the facts and establishing the connections between them, don't be boring. You are competing for the viewer's time and attention, so make sure the narrative has a hook, a momentum, or a captivating purpose. Finding the narrative structure will help you decide whether you actually have a story to tell. If you don't, then perhaps this visualization should support exploratory data analysis (EDA) rather than convey information. However, for the designer of an exploratory visualization it is still important to spark the viewer's imagination to encourage examining relationships among and facilitate interacting with the data—think gameification.

2. Think about your audience. What does the audience know about the topic? Is it meant for decision makers, general interested parties, or others? The visualization needs to be framed around the level of information the audience already has, correct and incorrect:

- Novice: first exposure to the subject, but doesn't want oversimplification
- Generalist: aware of the topic, but looking for an overview understanding and major themes
- Managerial: in-depth, actionable understanding of intricacies and interrelationships with access to detail
- Expert: more exploration and discovery and less storytelling with great detail
- Executive: only has time to glean the significance and conclusions of weighted probabilities

3. Be objective and offer balance. A visualization should be devoid of bias. Even if it is arguing to influence, it should be based upon what the data says–not what you want it to say. Tufte found numerous charts that misled viewers about the underlying data, and created a formula to quantify such a misleading graphic called the "Lie Factor." The Lie Factor is equivalent to the size of the effect shown in the graphic, divided by the size of the effect in the data. Sometimes it is unintentional—a number that is three times bigger than another will be perceived nine times bigger if represented in 3D. There are simple ways to encourage objectivity: labeling to avoid ambiguity, have graphic dimensions match data dimensions, using standardized units, and keeping design elements from compromising the data. Balance can come from alternative representations (multiple clusterings; confidence intervals instead of lines; changing timelines; alternative color palettes and assignments; variable scaling) of the data in the same visualization. Maintaining objectivity and balance is not a trivial effort and is easily unintentionally violated. Viewers and decision makers will eventually sniff out inconsistencies which in turn will cause the designer to lose trust and credibility, no matter how good the story.

4. Don't censor. Don't be selective about the data you include or exclude, unless you're confident you're giving your audience the best representation of what the data "says." This selectivity includes using discrete values when the data is continuous; how you deal with missing, outlier, and out of range values; arbitrary temporal ranges; capped values, volumes, ranges, and intervals. Viewers will eventually figure that out and lose trust in the visualization (and any others you might produce).

5. Finally, edit, edit, edit. Also, take care to really try to explain the data, not just decorate it. Don't fall into [the] "it looks cool" trap, when it might not be the best way to explain the data. As journalists and writers know, if you are spending more time editing and improving your visualization than creating it, you are probably doing something right.

HOW GE USES DATA VISUALIZATION TO TELL COMPLEX STORIES

BY GRETCHEN GAVETT

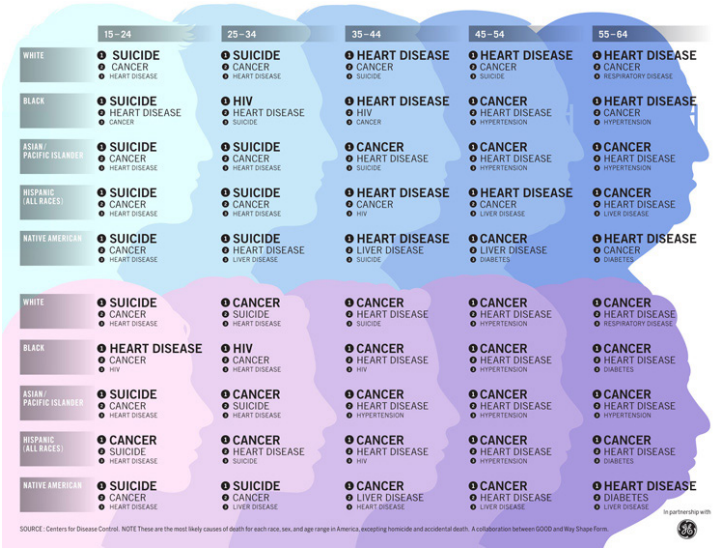
GE, perhaps more than any other major company, is dedicated to the use of data visualization as a key part of its marketing and communications efforts. Stemming from our Insight Center on visualizing data, I spoke with Linda Boff, GE's executive director of global brand marketing, about the benefits and challenges of this approach. An edited version of our conversation is below.

What's the history of data visualization at GE? How did your strategy around it develop?

GE specializes in complex challenges in solving the toughest problems in the world: infrastructure, renewable energy, affordable health care. Things you really have to get your mind around.

In trying to do that, the marketing communications brand group is always searching for compelling ways to bring these challenges to life. Five years ago or so, we started using data visualization.

One of our first was back in 2009, and [it] was about causes of death. We separated them in[to] male versus female, and via age spans. So if you're 24 to 36, or what have you, these are the three things you [are] most likely to die of. Now it seems so simple, but it was really compelling:



We got a tremendous response to it. The media loved it. Our different stakeholders—be it customers, employees—everybody thought, wow, what a great way to tell a story, and it was sort of born from that.

How do you think about using data visualization when it comes to different audiences and stakeholders, both within your company and outside of it?

As a large multinational company, we do have many audiences. And they range from employees and retirees to retail investors and thought leaders. Initially we thought about this—and I think to a large degree continue to—as a way to do external storytelling, but we have found that it works on so many different levels.

As a result, we have used data visualization in places as diverse as our annual reports or our annual report app, which is obviously geared toward investors. We've used it with thought leaders. When we released a white paper last fall on the industrial Internet, data visualization was a great way to tell that story.

It really works across different audiences. That's one of the things that's perhaps most exciting about it. ...

How do you staff for digging through all of that data, doing design work, and other digital elements?

The approach we took—and it's an approach we often take—is that a couple [of] people inside the brand marketing group spent a lot of time on it, but we also partner with the best of the best externally. And these are folks like Ben Fry, Lisa Strausfeld, Carlo Ratti at MIT, and Jer Thorp, then at *The New York Times*.

We didn't say, "OK, we're just going to work with the design studio Pentagram" or "We're just going to work with The New York Times," or what have you. And that was a fabulous approach because it gave us the eyes and the sensibilities of folks in a number of different areas.

Also, we're GE. We're involved in everything from transportation to health, curing people to [generating] energy to building things. We wanted a diversity of points of view on a diversity of subjects.

What projects have been the most successful for you? And how do you define success?

Because we have approached this largely as storytelling, we're always looking to experiment.

We also paid a lot of attention to the kinds of things that content publishers and marketers do for engagement, comments, news coverage. Over the years, we've had great pickup by people at publications, bloggers, all of whom are influential. That's meant a lot to us because it's a way for us to tell the GE story, and the amplification of that story is really, really important.

And it's also been a way for us to "double click" on certain things. Let me give you an example of what I'm talking about. We're an Olympic sponsor, and there are not that many Olympic sponsors. GE's in there, with Coca-Cola, with McDonald's, with their marketing machines, so to speak.

I was looking back at what some of what we did for the summer games last year in London. And data vis was a transformative way for us to talk about the data surrounding the games. We made this wonderful visualization that was 100 years of world records for the summer games. And you could sort of click into it obviously and see by country and time, etc.

So what I mean by "double click" is that it was another way for GE to talk about the importance of the Olympic games and give a bit of a perspective on them over time, but using a tool that a decade ago we never would have. It's enabled us to tell deeper, richer stories.

Another example, one that I really like, starts with the fact that GE generates about a quarter of the world's electricity. That's a lot. So we have a visualization right now that shows 713 turbines and the power generated over two weeks [see http://vimeo.com/36354487].

I can sit there and say, until the cows come home, "We generate a quarter of the world's electricity." But when you see it as a visualization, I think it's much more memorable.

What are some data visualization experiments or projects you're working on now?

We are working on one that I'm particularly excited about. Not long ago, we did what we called Flight Quest, an initiative we ran with Kaggle. We released some data from our customers [at airlines], as well as data from the National Airspace System on never-before-released flight times, arrival times, flight numbers, origins, arrival cities, all of these different elements.

We released all of it to the Kaggle community of data scientists globally and said: improve travel.

There were five winning algorithms that came up with a 40% improvement in flight arrival times. Eventually this could be software that could be incorporated into an airline's system to improve arrival times. But what we're working on now is a 3D interactive visualization of those winning algorithms. [Phase two of the project begins in June.]

What are your biggest challenges as you build new visualizations?

One of the biggest challenges today is that people expect data that is very real-time and current. And then the other piece of it is, how do you make it relevant? I think *The New York Times* has done a fantastic job [on these fronts]. I think *Wired* has done a nice job on this.

But at the same time, if the point is to simplify a story or make whatever the topic is a story well told, if it gets overly complicated it defeats the original purpose. So I think that's the line we all have to just watch out for a little bit. And we've learned this as we've gone, whether it's the topic or the how pleasing the interface on. Some things are just more inherently interesting than others.

And you just have to experiment to figure out what works and what doesn't.

I think so. And I have no regret in experimentation because I think we wouldn't be where we are if we hadn't experimented along the way.

What advice would you have for other companies, be [they] big companies or small companies, about why they should take data vis seriously? And what lessons would you impart to them?

The power of a good story well told in any sort of medium cannot be overstated. Data vis has allowed us to do storytelling at its best. Experimentation is also key, getting in there, understanding a medium and a technique, and not being afraid to experiment with it and be open and collaborative. We have had data marathons with many universities where we've brought in students, given them a problem, and said, hey, let's work over the next couple of days to solve this.

This is an open space. This fact is incredibly important. Open experimentation is a great way to bring to life challenges through vivid storytelling.

DON'T READ INFOGRAPHICS WHEN YOU'RE FEELING ANXIOUS

BY NICHOLAS DIAKOPOULOS

Ever had to look at a data visualization while you were in a lousy mood? Chances are you were more likely to make an error in visual judgment than if you had been feeling more cheerful. New research suggests that putting users in an emotionally positive mind-set improves their accuracy in interpreting data visualizations. So even if your company publishes only the occasional graph, provoking a positive emotional response in the audience might just help you get your data across more accurately. Moreover, as research in this area grows more nuanced, companies producing visual analytic tools and products may be able to enhance their effectiveness by designing with emotion in mind.

It's already broadly accepted that emotion (or *affect* as it's called in psychology) can influence different cognitive processes like attention, memory, creativity, and problem solving. As design-thinking legend Don Norman writes in his book *Emotional Design*, "When you are in a state of negative affect, feeling anxious or endangered, neurotransmitters focus the brain processing," allowing for concentration on details. "Positive affect arouses curiosity, engages creativity, and makes the brain into an effective learning organism," allowing for a less-focused and broader view of the situation.

Recently published research from Ph.D. student Lane Harrison and collaborators extends this understanding of affect and brain processing to the realm of data visualization. Their results experimentally show that people who were positively primed—put in a good mood by reading a lighthearted news article—made less visual judgment errors across a range of different charts than people who were negatively primed. Moreover they found that it was the positive priming that was decreasing error rather than the negative priming increasing it.

What does this mean for companies designing data visualization tools and environments? If visual judgment and analytic accuracy [are] of the utmost importance, like in finance, intelligence, or health care, you might think about how to integrate positive priming into the user's experience. The broader idea is something that scholars Richard Thaler and Cass Sunstein, in their book *Nudge*, call "libertarian paternalism"—biasing experiences to correct for cognitive deficiencies in human reasoning.

So, if users are less error prone when they're happy, why not design that emotional response into their tools and environments?

If only it were that easy. The news articles used as the affective primes in Harrison's study only nudged the mood of 30% of participants. It's also unclear how long an affective prime really lasts, how "durable" it can be. In the study, the subjects looked at the visualization right after reading the positive or negative story, but in a real use-context people are exposed to all kinds of different things that might confound an attempt to bias their mood: nutty coworkers, startling news events, soothing music, frenetic social media, or dreamy personal ruminations, all of which can be hard to control.

Manipulating emotions may come with its own set of slippery slopes and gotchas too. Another study from 2011 used images of things like smiling babies and rollicking puppies to induce positive affect in a creativity task. But imagine a somber data visualization about a patient's health vitals surrounded by such an artificial attempt to manipulate the user into a positive, more accurate mood. The tension between the desired affect and the inherent tone of the content could be problematic or even backfire entirely.

Less intrusive ways to put users in a positive frame of mind include making the interface more aesthetically pleasing or changing the lighting or color in the user's environment. Environmental conditions that Don Norman cites as engendering a positive affect include sweet smells, soothing sounds, harmonious music, symmetry, and smooth objects. Such environmental changes might stand in less direct conflict with underlying content than an interface that is actively trying to manipulate mood.

It should be clear that there are challenges here: When, how, and how much might you want to provoke emotion? And should you design it into the tool, or into the environment? These questions will get sorted out as the research in this area improves, but what we do know is that whether the end users of your company's data visualizations are customers, the public, or a group of internal analysts, it might just be worthwhile to keep those users in a cheerful mood.

TO GO FROM BIG DATA TO BIG INSIGHT, START WITH A VISUAL

BY SINAN ARAL; VISUALIZATION BY NIKOLAOS HANSELMANN

Although data visualization has produced some of the most captivating artistic displays in recent memory, some of which have found their way into exhibits at the New York Museum of Modern Art and countless art installations around the world, business leaders are asking: is data visualization actionable?

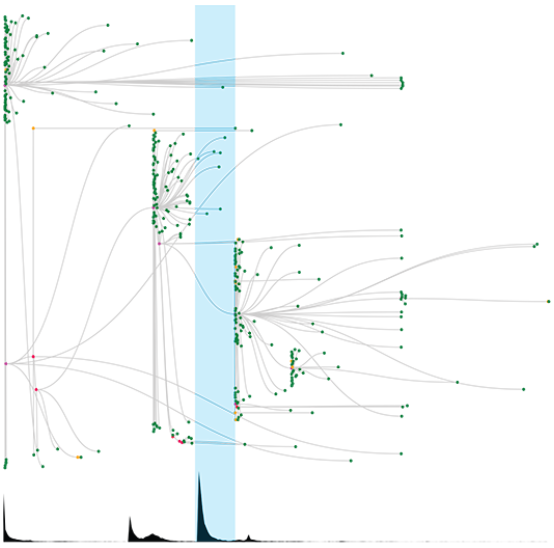
I think so. In my role as the Scholar-in-Residence at *The New York Times* R&D Lab, I am collaborating with one of the world's most advanced digital R&D teams to figure out how we can draw actionable insights from big data.

How big? Massive: We are documenting every tweet, retweet, and click on every shortened URL from Twitter and Facebook that points back to *New York Times* content, and then combining that with the browsing logs of what those users do when they land at the *Times*. This project is a relative of the widely noted Cascade project. Think of it as Cascade 2.0.

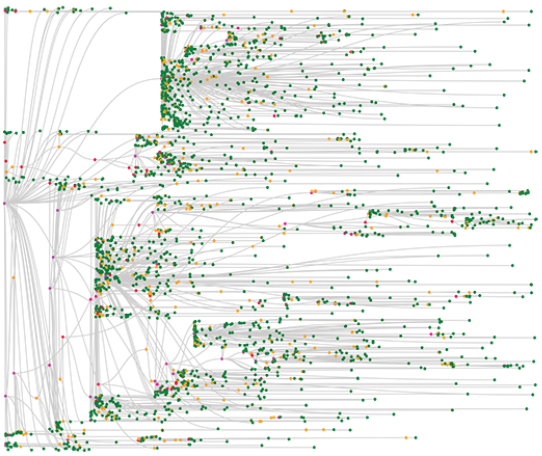
We're doing this to understand and predict when an online cascade or conversation will result in a tidal wave of content consumption on the Times, and also when it won't. More importantly we are interested in how the word-of-mouth conversation drives readership, subscriptions, and ad revenue; how the *Times* can improve their own participation in the conversation to drive engagement; how we can identify truly influential readers who themselves drive engagement; and how the Times can then engage these influential users in a way that complements the users' own needs and interests. Do it, and we can turn that statistical analysis, as you'll see below, into elegant, artistic real-time data streams.

Handling the streams, archiving the sessions, and storing and manipulating the information are in themselves herculean tasks. But the even bigger challenge is transforming beautiful, big data into actionable, meaningful, decision-relevant knowledge. We've found that visualization is one of the most important guideposts in this search for knowledge, essential to understanding where we should look and what we should look for in our statistical analysis.

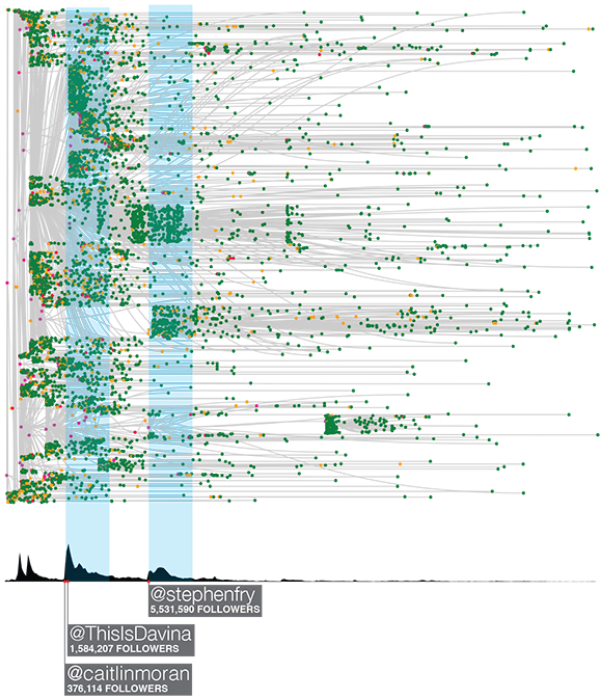
For example, here are three visualizations that have helped us gain knowledge. They show cascades of the tweets and retweets as lines and dots about three different *Times* articles over time, combined with the click-through volume on each article synced in time and displayed as a black graph under each cascade. Each panel tells a different story about engagement with the content.



For the first article, there is a sizable Twitter conversation and several large spikes in traffic. But the click-through volume seems independent of the Twitter conversation: The largest spike in traffic, highlighted in blue, occurs when there is very little Twitter activity. In this case, a prominent link on a blog or a news story that referred to the story, rather than the Twitter conversation itself, is probably driving the traffic.



On the second article, the Twitter conversation is intense. There are many, tweets and retweets of the article—yet the article itself gets very little traffic. People are talking about the article on Twitter, but not reading it. This sometimes happens when the main message of an article sparks a debate or a conversation that can happen without the content of the article being that important, for example, when a timely piece of news contains little analysis or editorial content, or when the conversation or debate gets away from the article and evolves [into] its own independent content.



In the third and final article, an intense Twitter conversation moves in lockstep with engagement. As people tweet and retweet the article, their followers are clicking through and engaging with the content itself. This tight relationship between the online conversation and the website traffic is most pronounced when the three "influencers" tagged in the figure inspire the two largest spikes in traffic over the engagement life cycle of the article.

With just these three data visualizations, we've gained understanding in important nuances about so-called virality. The relationship between online word-of-mouth conversations and engagement isn't as simple as something just "going viral." Different patterns emerge with different types of content.

Still, the visuals cannot tell the whole story. We see some clear correlations here, but complex conditional dependencies and temporal and network autocorrelation make it necessary to build more sophisticated causal statistical models that will generate true, reliable insights about word-of-mouth influence.

What these visuals do help with is getting us to know where to look and what questions to ask of the data. That is, we can't build the more complex models until we know the most suitable places for building them. These visuals give us some of that insight.

Cascade 2.0 will be built on sophisticated analytics, and it will require data visualization. Asking important questions and avoiding unnecessary ones [are] essential to moving forward effectively and efficiently with big data. Without visualization, we are much less efficient in getting to the questions whose answers teach us something. That's why visualizing data must be one of the most important tools for data scientists. It is our torch in a thick, dark forest.

TO TELL YOUR STORY, TAKE A PAGE FROM KURT VONNEGUT

BY ANDREA OVANS

In the 1989 movie *Dead Poet's Society,* Robin Williams, playing the iconoclastic English teacher John Keating, dismisses the notion that you can judge the perfection of a poem mathematically by plotting how artfully it employs meter, rhyme, and metaphor against how important the subject is. Rather than have his students think they could graph the relative merits of, say, a Shakespeare sonnet against a poem by Alan Ginsberg, he has them rip up their textbooks. Data can't tell us anything about stories, he's saying, as pages of *Understanding Poetry*, by Dr. J. Evans Pritchard, Ph.D., fly all over the room.

Businesspeople are often advised to turn their data into stories to make them more persuasive. And that is certainly good advice. But they are given precious few tools to help them do that. It turns out though, John Keating notwithstanding, that graphs can be remarkably useful in demonstrating the mechanics underpinning an effective story. One person who'd given this a lot of thought was novelist Kurt Vonnegut, a real-life literary iconoclast if there ever was one.

Harvard's Nieman Foundation for Journalism recently shed a spotlight on Vonnegut's story graphs in its publication Nieman Storyboard (a wonderful resource on the art of storytelling in itself). Vonnegut devoted his master's thesis at the University of Chicago to studying the shapes of stories. The thesis was rejected (apparently, Vonnegut's advisors were of the John Keating school of literary criticism). But his ideas are thriving online in various storytelling tutorials. Nieman offers up Vonnegut's original presentation, now on YouTube, in which he graphs some of the most basic story structures and explains how they work.

http://www.youtube.com/watch?v=oP3c1h8v2ZQ&feature=share&list=PL991B74289AE23E10

"There's no reason why the simple shapes of stories can't be fed into computers," Vonnegut begins. First up is one he calls *Man in a Hole*. "It needn't be a man, and he needn't fall into a hole," he adds, for the metaphorically challenged among us. "That's just an easy way to remember it."

In the tradition of J. Evans Pritchard, he starts by drawing the vertical Good Fortune/Ill Fortune (or G-I) axis, with "sickness and poverty" at the bottom and "wealth and boisterous good health" at the top. At the midpoint, he draws his x axis—B (for beginning) to E (for electricity). He's joking, of course, but he also wants to underscore the point that this is an exercise in relativity, since it's the shape of the curve that matters, not the specific data points.

Then he places his chalk on the y axis a bit above the midpoint ("Why start with a depressing person?" he quips), draws a sine wave dropping down toward the bottom and rising up again: Somebody gets into trouble and gets out of it. "People *love* that story," he says. "They *never* get sick of it!" (This is doubly obvious when you draw the business parallel by substituting a term like *business, strategy, revenue, IT, HR*, or whatever for the word *somebody*).

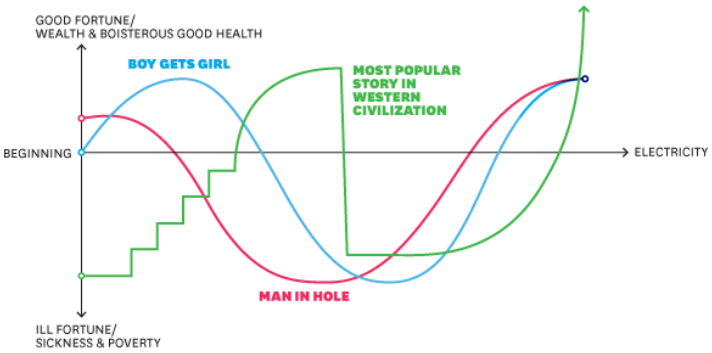
He goes on to graph *Boy Meets Girl,* starting right at the midpoint of the y axis—"an average person on an average day, not expecting anything." He draws a double sine wave rising up and then down and then up again. "Something wonderful happens, Oh hell. Got it back again." In business terms, the classic turnaround story (IBM comes to my mind here, and more than once).

The next one is more complicated, he warns. Despite what he's just said, he starts at the bottom and stays there—a wretched person, a little girl, no less, has lost her mother and her father has married again to a horrible woman. The curve hovers at the bottom. A fairy godmother arrives, bestowing much-needed resources (shoes, a dress, mascara). With each gift, the line goes up incrementally, like a bar graph. The girl goes to a dance. The clock strikes 12:00. The resources dry up. The line drops almost straight down, but not all the way back (she has those memories, and maybe some IP or a valuable customer base). The prince finds her, the shoe fits. Facebook buys your start-up, the curve shoots up as you achieve off-scale happiness.

It so happens, he says, that this Cinderella story is "the most popular story in our Western civilization. Every time it's retold somebody makes another million dollars. You're welcome to do it." Well, sure...

Here are all three stories, conveniently plotted on a single graph:

SIMPLE SHAPES OF STORIES
As told by Kurt Vonnegut.



SOURCE DAVID YANG, VISUAL.LY HBR.ORG

But watch the video (it's less than five minutes long), and two things become apparent. The first is certainly that so many successful business stories follow patterns embedded in Western civilization's most primal literary conventions. It's easy to see why marshaling data to tell one of these kinds of stories—rags turning into riches, mistakes rectified, challenges overcome, the right resources and the right contacts saving the day—would be so compelling. And there's probably an argument here for reading more fiction, to give John Keating his due.

The second is that Vonnegut's delivery matters as much as his ideas. His timing is perfect. His language is concrete and unexpected. He's showing you the simplicity that underlies apparent complexity—that's what data are so good at doing. But he's just as concerned with making sure you're paying attention—since no one is persuaded by something they don't remember.

THE QUICK AND DIRTY ON DATA VISUALIZATION

BY NANCY DUARTE

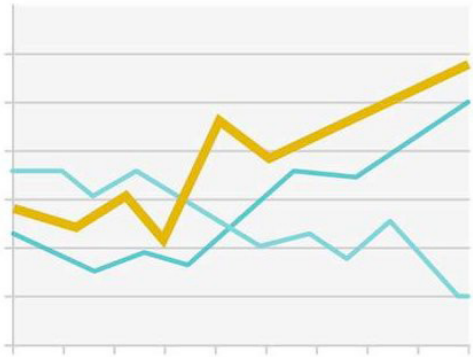
Displaying data can be a tricky proposition, because different rules apply in different contexts. A sales director presenting financial projections to a group of field reps wouldn't visualize her data the same way that a design consultant would in a written proposal to a potential client.

So how do you make the right choices for your situation? Before displaying your data, ask yourself these five questions:

1. Am I presenting or circulating my data?

Context plays a huge role in how best to render data. When delivering a presentation, show the conclusions you've drawn, not all the details that led you to those conclusions. Because your slides will be up for only a few seconds, your audience will need to process them quickly. People won't have time to chew on a lot of complex information, and they're not likely to run up to the wall for a closer look at the numbers. So, think in broad strokes when you're putting your charts together: What's the overall trend you're highlighting? What's the most striking comparison you're making? Those are the sorts of questions to answer with projected data.

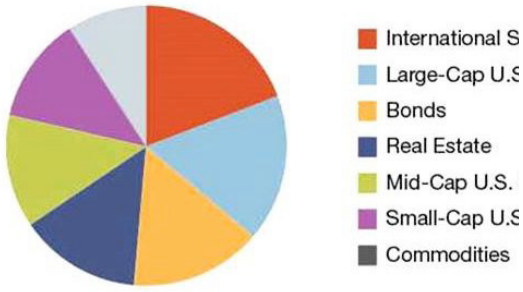
Scales, grid lines, tick marks, and such should provide context, but without competing with the data. Use a light neutral color, such as gray, for these elements so they'll recede to the background, and plot your data in a slightly stronger neutral color, such as blue or green. Then use a bright color to emphasize the point you're making, as in this example:



It's fine to display more detail in documents or in decks that you e-mail rather than present. Readers can study them at their own pace—examine the axes, the legends, the layers—and draw their own conclusions from your body of work. Still, you don't want to overwhelm them, especially since they won't have you there in person to explain what your main points are. Use white space, section heads, and a clear hierarchy of visual elements to help your readers navigate dense content and guide them to key pieces of data.

2. Am I using the right kind of chart or table?

When you choose how to visualize your data, you're deciding what type of relationship you want to emphasize. Take a look at this chart, which shows the breakdown of an investment portfolio:



Investment Portfolio Breakdown

In the pie, it's clear that this person holds a number of investments in different areas—but that's about all you see.

Here are the same data in a bar chart:

Investment Portfolio Breakdown

Now it's much easier to discern how much is invested in each category. If your focus is on comparing categories, the bar chart is the better choice. A pie chart would be more useful if you were trying to make the point that a single investment made up a significant portion of the portfolio.

3. What message am I trying to convey?

Whether you're presenting or circulating your charts, you need to highlight the most important items to ensure that your audience can follow your train of thought and focus on the right elements. For example, this chart is difficult to interpret because all the information is displayed with equal visual value:

Revenue Trends

	Q1	Q2	Q3	Q4	Total
Americas	-18%	7%	25%	2%	2%
Australia	47%	-7%	26%	15%	17%
China	15%	-5%	1%	7%	19%
Europe	57%	10%	-3%	7%	13%
India	57%	6%	-3%	8%	13%

Are we comparing regions? Quarters? Positive versus negative numbers? It's difficult to determine what matters most. By adding color, you can draw the eye to specific areas:

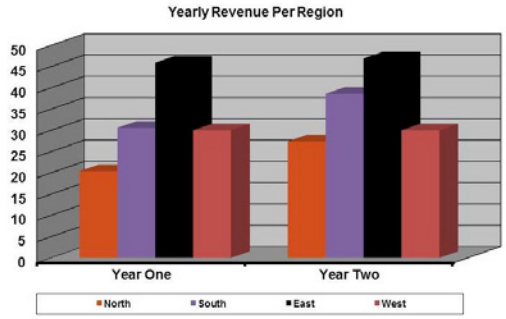
Revenue Trends

	Q1	Q2	Q3	Q4	Total
Americas	-18%	7%	25%	2%	2%
Australia	47%	-7%	26%	15%	17%
China	15%	-5%	1%	7%	19%
Europe	57%	10%	-3%	7%	13%
India	57%	6%	-3%	8%	13%

We now know that we should be focusing on when and in which regions revenue dropped.

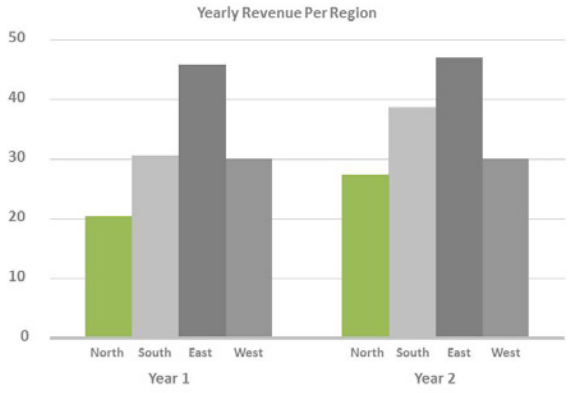
4. Do my visuals accurately reflect the numbers?

Using a lot of crazy colors, extra labels, and fancy effects won't captivate an audience. That kind of visual clutter dilutes the information and can even misrepresent it. Consider this chart:



Can you figure out the northern territory's revenue for year one? Is it 17? Or maybe 19? The way some programs create 3D charts would lead any rational person to think that the bar in question is well below 20. However, the data behind the chart actually says that bar represents 20.4 units. You can see that if you look at the chart in a very specific way, but it's difficult to tell which way that should be—even with plenty of time to scrutinize it.

It's much clearer if you simply flatten the chart:



5. Are my data memorable?

Even if you've rendered your data clearly and accurately, it's another challenge altogether to make the information stick. Consider using a meaningful visual metaphor to illustrate the scale of your numbers and cement the data in the minds of your audience members. A metaphor can also tie your insights to something that your audience already knows and cares about.

Author and activist Michael Pollan showed how much crude oil goes into making a McDonald's Double Quarter Pounder with Cheese through a striking visual demonstration: He placed glasses on a table and filled them with oil to represent the amount of oil consumed during each stage of the production process. At the end, he took a taste of the oil to drive home his point. (To add an element of humor, he later revealed his prop "oil" to be chocolate syrup.) Watch the video here: http://youtu.be/6As879M_kCs.

Pollan could have shown a chart, but this was more effective because he gave the audience a tangible visual—one that triggered a visceral response.

By answering these five questions as you're laying out your data, you'll visualize it in a way that helps people understand and engage with each point in your presentation, document, or deck. As a result, your audience will be more likely to adopt your overall message.

WHAT MAKES THE BEST INFOGRAPHICS SO CONVINCING

BY ANDREA OVANS

A great infographic is an instant revelation. It can compress time and space. *(Good gosh—Usain Bolt is that much faster than all the other 100-meter gold medalists who've ever competed?) It can illuminate patterns in massive amounts of data. (Sure, we're spending much more on health care and education than our grandparents did. But look how much less on housing.) It can make the abstract convincingly concrete. (Which player was ESPN's SportsCenter[']s most discussed during the 2012 football season? Tim Tebow—and by a colossal margin. Seriously?)*

These intriguing revelations come from a short trip around *The Best American Infographics 2013*. Spend serious time poring over graphs, pie charts, bar charts, flow charts, timelines, interactive diagrams, maps, cut-away diagrams, and narrative illustrations, as Gareth Cook did to compile the collection, and you'll come away with more than your share of these mind-bending moments—and a wide-ranging view of what infographics can do. A Pulitzer Prize–winning journalist, Cook is a regular contributor to NewYorker.com and *Scientific American Mind*.

The most compelling infographics, he says, mine relationships among overlooked variables to tell you something unexpected and get you thinking. *(Who knew it takes an annual income of $908,000 to break into the top 1% in Stamford, Connecticut, but only $609,000 in New York City—and just $558,000 in pricey San Francisco?)* The least effective confuse you (the food pyramid), overwhelm you with data (nutrition labels), or are just plain boring. I recently asked Cook to share his thoughts about what makes an infographic particularly persuasive.

What's special about the way infographics make their case?

Infographics have an emotional power because they can show you an idea—or a relationship, or how something works—very quickly. People respond to that. A persuasive infographic surprises the viewer. It moves them in some way and makes them want to keep looking at it or show it to other people.

Did you see commonalities in the ones you found most convincing?

First, I'd say, they all have a clear focus. The designer has gone in and removed all the extraneous details so you see just what you need to understand the message behind it. And yet the best ones also have a kind of openness—the person who's done it is transparent about what data they're using. That can be tricky because you need to give people a sense of all the data that's out there, and enough context, without overwhelming them. In the best cases, viewers feel that they are the ones stepping in and making the connection because they can see the bigger pattern naturally emerging from what you're showing them.

Can you give me an example?

Take a look at the first infographic of the collection. It's very simple. It starts with a question: "Which Birth Dates Are Most Common?" And what we see is a chart that shows every day of the year in various shades of a single color. The darker the color, the more babies were born on that day in the U.S.

It's effective because you can see all the data for the entire year, and yet the actual relationship emerges very strongly. You immediately see the dark band running through July, August, September, and into October. It's very clear that more people are being born then.

Once you've seen the main relationship, you can look at other things, as well, which is very satisfying. You can see, for example, on July 4th and 5th there's a sudden drop-off in people being born, presumably because it's around the holidays—you can see the same thing around the Christmas holidays. But then if you look over at February 14th there's a dark island where a lot of babies are being born. So you can see the main relationship, but then you can also do some exploring.

That's an important part of its persuasiveness: You want to show someone something, but you also want to give them a sense that they're free to move around and find their own relationships. When they do, they'll have confidence that you really are giving them the whole story.

In his introduction to the book, David Byrne talks about the power of infographics to let us see the invisible. He's thinking mainly of cutaway diagrams, as an explanatory tool, but I imagine that can be an effective tool of persuasion as well.

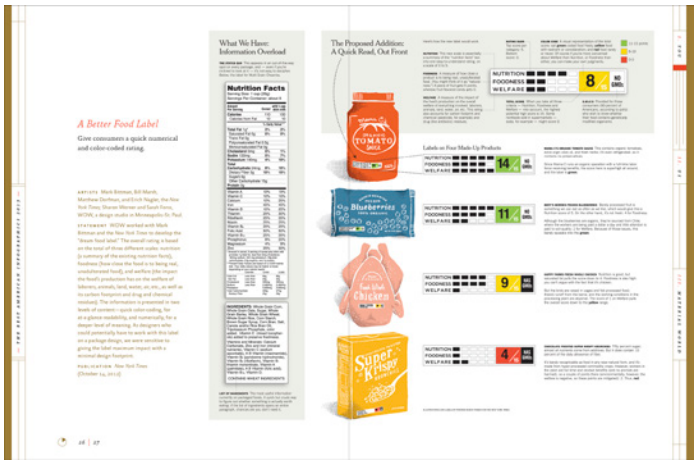
Sometimes people don't believe you because they can't relate to your argument or they can't understand it. Infographics can make an abstract subject concrete—let viewers put their hands around it. One of the 10 interactive infographics in the book does this especially well. It shows carbon emissions in New York City in real time, representing each ton of carbon dioxide as a giant blue sphere.

In 2010, as we're told in the introduction, New York City added the equivalent of 54 million metric tons of carbon dioxide to the atmosphere. That's two tons every second. As you watch, the giant spheres emerge from the ground and start to float upward, two every second. You can see how much they build up over time. By the end of a day, the pile has reached the top of the Empire State Building. It's amazing; you get a visceral sense for how much pollution that is.

Can you give me an example of an infographic that's good at boiling down a mass of big data?

One is the Better Food Label, which Mark Bittman and a team of designers at *The New York Times* came up with. Look at the food label on your breakfast cereal in the morning and you see this overwhelming amount of data—vitamin A, vitamin C, calcium, all these percentages, two columns, with and without milk. It's hard to make sense of it all. Imagine someone at the grocery store trying to decide between two products: Which is going to be better for me and my family? It's just too hard to get the answer.

So they came up with a chart designed to address just a few basic questions that someone might want to know when trying to decide how good this food is. How healthy is it nutritionally? How free is it from possible contaminants? How safely was it produced, environmentally? And when you look at their label, you can take in all of that information in two or three seconds.



This is something infographics are naturally designed to do—give you the gist of a really big data set. I think this is one of the reasons why we're seeing infographics used in so many different realms right now.

I know you're talking to a number of business groups while working on next year's collection of infographics. What are some of the ways forward-thinking businesses are beginning to use them?

Certainly, businesspeople are working with designers to develop infographics that present ideas. But more broadly, they['re] working with them to help solve problems. People adept at creating visual solutions bring a different basic set of questions to bear. In considering a data set, they may say "Oh, we can look at this unusual variable and see how that changes over time." Or they may come up with a new way to explain something to a customer who just can't seem to understand your current pitch.

I was not at all anticipating this when I set out to do this collection, but I've definitely heard from readers who use this as a source book. When they have a problem they flip through it and may notice something that gives them an idea they wouldn't have thought of before.

Many of the infographics in this collection are pretty funny. If you are in the serious business of trying to persuade people of something, do you see a role for humor?

I think it's often the case that when people are designing something to persuade, they forget the importance of whimsy. Humor opens people up and makes them more willing to hear messages they might otherwise reject out of hand. When you're working really hard on designing something or making something clear, it's very easy to lose that sense of fun yourself, and the work shows it. You want your audience to sense that at a certain level you are enjoying this. A lot of the pieces in this collection just make me smile.

THE RIGHT COLORS MAKE DATA EASIER TO READ

BY SHARON LIN AND JEFFREY HEER

What is the color of money? Of love? Of the ocean? In the United States, most people respond that money is green, love is red, and the ocean is blue. Many concepts evoke related colors—whether due to physical appearance, common metaphors, or cultural conventions. When colors are paired with the concepts that evoke them, we call these "semantically resonant color choices."

Artists and designers regularly use semantically resonant colors in their work. And in the research we conducted with Julie Fortuna, Chinmay Kulkarni, and Maureen Stone, we found they can be remarkably important to data visualization.

Consider the charts of (fictional) fruit sales below.

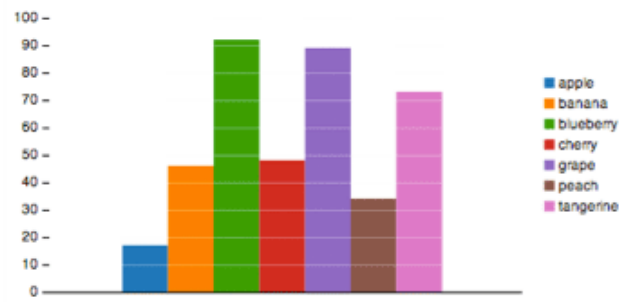
The only difference between the charts is the color assignment. The left-hand chart uses colors from a default palette. The right-hand chart has been assigned semantically resonant colors. (In this case, the assignment was computed automatically using an algorithm that analyzes the colors in relevant images retrieved from Google Image Search using queries for each data category name.)

Now, try answering some questions about the data in each of these charts. Which fruit had higher sales: blueberries or tangerines? How about peaches versus apples? Which chart do you find easier to read?

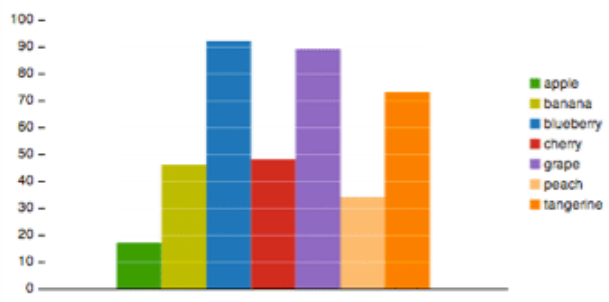
If you answered the chart on the right, you're not alone. To determine the impact of semantically resonant colors on graph analysis, we ran experiments to measure how quickly people can complete data-comparison tasks on bar charts using either default colors or semantically resonant colors. On average, people took a full second less to complete a single comparison task when they were looking at semantically resonant colors (whether chosen by our algorithm or by an expert designer). That may not sound like a lot but it's about 10% of the total task time. These time savings can add up, particularly for data analysts making untold numbers of such comparisons throughout their workday.

What's going on here? We see a number of ways in which semantically resonant colors could be helping improve graph-reading performance. First, semantically resonant colors can enable you to take advantage of familiar existing relationships, thus requiring you to use less conscious thought and speeding recall. Non-resonant colors, on the other hand, can cause semantic interference: the colors and concepts interfere with each other (as anyone familiar with the famous Stroop test from psychology knows—the one in which you're asked to name the text colors of color names printed in conflicting colors: green, red, and so on). Second, because your recall of the concept-color relationship is improved when looking at semantically resonant data, you may not need to repeatedly look at the legend to remember which column is which, and so can focus more on the data itself.

To make effective visualization color choices, you need to take a number of factors into consideration. To name just two: All the colors need to be suitably different from one another, for instance, so that readers can tell them apart—what's called "discriminability." You also need to consider what the colors look like to the color



Default color assignment



Semantically resonant color assignment

blind—roughly 8% of the U.S. male population! Could the colors be distinguished from one another if they were reprinted in black and white?

One easy way to assign semantically resonant colors is to use colors from an existing color palette that has been carefully designed for visualization applications (ColorBrewer offers some options) but assign the colors to data values in a way that best matches concept color associations. This is the basis of our own algorithm, which acquires images for each concept and then analyzes them to learn concept color associations. However, keep in mind that color associations may vary across cultures. For example, in the United States and many Western cultures, luck is often associated with green (four-leaf clovers), while red can be considered a color of danger. However, in China, luck is traditionally symbolized with the color red.

There are a few other factors to consider when using semantically resonant colors:

Type of data: So far, we have only discussed data that represent discrete categories. Other data may be numerical or rank-ordered ("poor," "fair," "good" for example). In these cases, a diverging or sequential color scheme may be preferred, in which a single color becomes darker or lighter depending on the relative order of the values.

Similar color associations: Some concepts map to very similar colors. For example, "magazine" and "newspaper" might both map to gray. We could consider assigning two different shades of gray to both concepts, but then it may be more difficult to remember which shade of gray maps to which one in the visualization. In this case, we might prefer less-resonant colors that ensure discriminability.

Concept-color association strength: Some concepts are simply more colorable than others. For example, people generally agree on the colors of asset categories such as "gold," "silver," "cash." However, what is the color of "social security," "national defense," or "income security"? Overall, we found that using semantically resonant colors for categories that were more colorable unsurprisingly tends to provide greater performance improvements.

Semantically resonant colors can reinforce perception of a wide range of data categories. We believe similar gains would likely be seen for other forms of visualizations like maps, scatterplots, and line charts. So when designing visualizations for presentation or analysis, consider color choice and ask yourself how well the colors resonate with the underlying data.

DATA GOES BEST WITH A GOOD STORY (AND VICE VERSA)

BY WALTER FRICK

Storytelling with data is having its moment in the journalism world. After leaving *The New York Times* last year, Nate Silver has launched his new data journalism venture with ESPN, former Washington Post blogger and columnist Ezra Klein is expanding his chart-filled approach beyond public policy at Vox.com, and both the Times and the Post are starting their own data-driven sites to replace their departed stars.

While none of these ventures rely solely on charts and data to serve readers, the Post's is explicitly betting that data on its own isn't enough. Its forthcoming policy-focused website is premised on the notion that data and narrative go best together. The lesson for anyone looking to communicate or persuade using data is that the addition of human drama goes a long way.

I recently spoke with the Post's Jim Tankersley, the economic correspondent leading the project, about what he sees as the benefits of pairing narrative and data to get your point across. An edited version of our conversation follows.

I wanted to start with this idea of "In One Chart" journalism to get your sense of the ways in which this has succeeded, what's missing from it, and if it is fair to see this as a starting point for what you're building.

I love "In One Chart" because I sometimes think that way, and I know a lot of people who do. Where I think it's been really successful is in drawing people into something they might not have read a 700-word article about but can understand the thrust of quite quickly. And then very often they go on to read that article.

But I do think it's just a jumping-off point. We will do cool data visualizations on our site. But even in the context of those visualizations, we want to make them do a better job of telling stories. We're going to be very interested in the sort of data plots that really help you understand change over time—a narrative thread that you can represent visually. If we can give [readers] themes and narratives and plots, that helps them understand really complicated concepts. And I think you can do that visually. People already do it, and they do it really well, and we want to build on those kind[s] of successes.

There's a sense that the narrative, the story, is a human thing that can draw someone in. Embedded in how we talk about that is the idea that the data is the medicine. You're going to draw someone in with a story so that they're consuming the stuff they didn't want to consume—the data. One of the things that this "chart journalism" has done is make the chart the part that jumps off the page, what draws people in. How do you think about which is the lure and which is the substance?

Data has gone from a thing that we try to bury to now the shiny thing that makes you want to swallow it. I do still think there are things that you can't do in charts, that you have to do in exposition, that by themselves are not that sexy if you just made a story around the numbers.

Data and visualizations and storytelling—human drama—when combined help readers understand concepts. I can get a lot from one great visual chart, but I can get more context from several paragraphs just really laying out the conclusions of data. And I can get a lot out of a story about someone like me doing something that is interesting to me. But I think that when you bring them all together, what you get is this experience where people have a process of discovery for themselves, which I think is important for learning. If you are just looking at a chart, you can try to figure out what it's telling you and there's a little bit of discovery as [you] decipher it. If that chart occurs in the natural flow of narrative or storytelling, then I think it's almost a clue popping up as you're trying to solve this mystery. I think that when they all work together, they're more powerful than any of them alone.

I can imagine a case where you find an interesting story and you find data to buttress it and to better tell that story. Or the reverse: you find a really interesting dataset, but you think you need a story or a narrative to really draw people in. There's potentially a little bit of a tension here. As a reporter, there can be an impulse to go out and say, "I want the best possible story I can find." The most extreme, interesting, compelling, sometimes bizarre story. And often that can be in tension with a dataset. How do you find stories that are doing service to data and vice versa, rather than having them be in conflict?

It's quite often that you find stories that lead you then to [ask], "Hey, is there an interesting data point that can help flesh this out?" Any really great story will be a window into something bigger. And if you look hard enough you can figure out where that window leads you. I think it's easier—and it's going to be by far the dominant type that we're going to be doing—that the story works the reverse way. That you start with the question, you start with the data, you start

with "What are we trying to figure out here?" Once you come to a good conclusion based on your own analysis of the data, then you go out and find the story.

A lot of times reporters get caught up because they want every person whom they feature to be perfectly emblematic of a central tendency in a dataset. But the truth is, it's really hard to find someone who represents that entire thing. So what I like to think about is, "Who is someone who gives us the best entry point here, who represents something about this?" And you're honest with your readers about the things that they are not representative of, but you focus the camera most on the parts where they are.

I don't think there've been a lot of times where I've read a story and thought, "Wow, I really wish this guy more perfectly represented the dataset underlying what the writer's trying to tell me." But there have been plenty of times where I've read a story and thought, "This person's not interesting." And being interesting is a really important way to get people to read and digest what you're trying to help them learn.

How are you thinking about trust with the reader, especially since they're trusting you to handle data properly?

We have an extra obligation to try and earn that trust. People are naturally skeptical of what they read in the newspaper or on the internet, and they should be. We want to build this relationship with readers where they feel like we are treating numbers honestly. We owe it to our readers to be constantly keeping their skepticism in mind, and [we] try to be very, very careful with how we use numbers so that they don't feel like we've misled them in any way.

How much of this kind of storytelling is made possible by access to data that wasn't available before?

It's absolutely easier. It also ups the expectation for what a good story is going to be. I think the easiest place to see this is in sports reporting. There's always been a basic set of statistics available to everybody, but now if you're trying to make the case for why a guy's contract is fair and all you're talking about is RBIs, you get laughed out of the room, even by a casual newspaper reader. I think that's good.

To what extent does this phenomenon go beyond journalism—the way that politicians communicate with citizens, or the way that businesses communicate to their investors or consumers, etc.? Are we going to see data as storytelling increasing in other areas?

Think about what really successful trial lawyers do, what really successful CEOs do with their boards and shareholders: they tell stories, they use numbers, and they make arguments. I don't think it's unique to journalism, but I hope that we can do it in a really beneficial way to help readers.

DATA DOESN'T SPEAK FOR ITSELF

BY THOMAS C. REDMAN

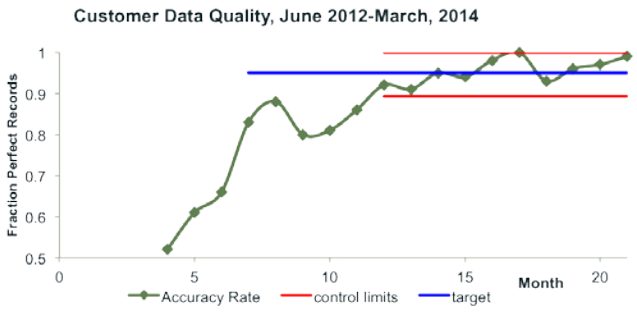
I trained as a statistician and first joined Bell Labs in the network performance group. A year or two after I started, it was time for my first big presentation at AT&T Headquarters. I completed my prep well in advance and rehearsed carefully. Then I was off to the big meeting.

It could not have gone worse. The only impressions I left were bad ones. Young hothead that I was, I blamed everyone but myself, including the audience: "The average manager up here can't even understand a pie chart!"

An established veteran of many such presentations looked me square in the eye and said, "Of course not, Tom. It's your job to make it so they don't have to."

That was my first lesson in data presentation. As a data presenter, you face a tall order in getting others to comprehend and believe data. You have to think through your audience's background and present data in ways that advance their understanding. The best way to do so is to make your plots and the accompanying explanations easy to understand. As Edward Tufte advises, label the axes, don't distort the data, and keep chart-junk to a minimum.

The plot below is a typical result of a well-conceived and executed data quality program. But it features too many unfamiliar terms such as "accuracy rate" and "fraction perfect records." Without additional explanation, the audience may find themselves lost.



Start by explaining how to interpret the chart at its most basic level: "Here is a time-series plot of the results of our data quality program. I know most of you are familiar with such plots, but let's make sure we're all on the same page here. As you can see, we focused on the quality of customer data. The x-axis is time, and here I'm showing one point every month. The y-axis is the fraction of data records that were created perfectly each month. That's how we're measuring accuracy. It is a tall standard, and I'll have more to say about that in a minute." Then, explain to your audience how to read the data presented within the chart: "The green line displays our actual results. The blue line shows the target we set for ourselves, and the red lines are control limits. These are a bit technical. I'll explain later. Now before I dig in, are there any questions about how to read the chart?"

Note that you've told your audience where you'll be expanding, but you're focusing on the basics of reading the chart first. This lets them fully comprehend the visual, so they can then put their full attention toward listening to your explanation of the data to come.

Now tell the story of the data in a powerful, animated fashion. In this case, there is much to tell, including how and why the program started; the joys and challenges surrounding the documentation of customer requirements; measurements against those requirements, including the logic of the choice of metric on the y-axis; improvement projects; and how you established control—essentially the implications of those red lines. Point out the impact of each on the plot as you proceed.

Different audiences will have different needs, and you should tell the story in the simplest and most direct way you can for each. For example, a technical community may wish to understand the details in your choice of metric and the software used to draw the plots. A senior leader may wish to understand the significance of the story for extending data quality across the organization. While the main story will be the same for each, the emphases should be very different.

Be aware that many people are skeptical about analytics, big data, data mining, and statistics (perhaps recalling the famous observation, "There are three kinds of lies. Lies, damned lies, and statistics."). Whether this skepticism is justified or not, it does enormous damage by slowing, or even stopping, the penetration of good ideas into organizations. As a data presenter, you have a sacred trust to build support for data among your audience. You must:

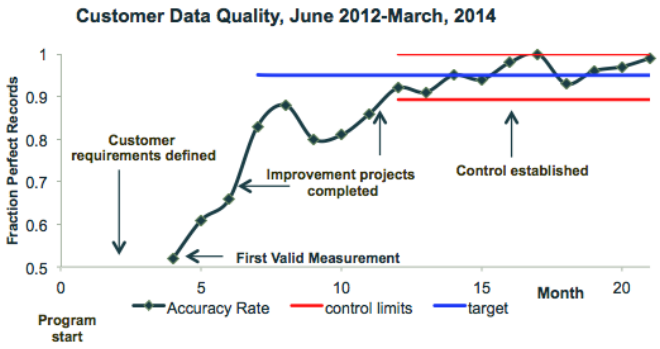
1. Present the facts in the most straightforward, accurate way you can. This is especially true when the results aren't favorable. Further, if your results are counter to established wisdom, simply state that this is the case.

2. Present a comprehensive picture. Leaving out a key fact is the worst kind of lie.
3. Provide proper context, including where the data originate and what you've done to ensure they are of high quality. (If you've done little, you must explicitly state, "The data are of unknown quality. This could impact results.")
4. Summarize your analysis, including shortcomings and alternative explanations for the results you see.

It is fine (and often appropriate) to state your opinion, but you must clearly separate your opinion from the facts. Even the best analysis goes only so far; then intuition takes over. Make the dividing line clear.

Now take your concern for the audience a step further. Successful oral presentations live on as people pass on PowerPoint decks or links to them. People reading a slide deck alone will not have the benefit of your oral explanations, so you must think of their needs as well. As I heard in my early days at Bell Labs, "People spend an average of 15 seconds looking at a chart. Don't make them spend 13 of those seconds figuring out how to read the chart. Build in explanations wherever possible. Even better, make the graphic tell the story."

With this in mind, take two steps. First, provide your explanation of how to read the chart in the notes page of your PowerPoint or slide deck program. Second, annotate the graphic, as below. While annotations do not replace a well-told story, they do give the reader some inkling of what's involved.



To most audiences, an ounce of insight is worth a ton of analysis. Thus, one outstanding graphic that cuts to the heart of the issue at hand and guides next steps is worth more than hundreds of mediocre ones. Seek that graphic. Presented this way, data are power.

None of what I've proposed here is particularly difficult in practice, once you have an important insight or result to share. Leaders—even skeptics—hunger for ways to improve their departments and companies. Your job as a data presenter is to tap into and satisfy that hunger in the simplest, most transparent way that you can.

VISUALIZING ZERO: HOW TO SHOW SOMETHING WITH NOTHING

BY ANDY KIRK

In its most revealing form, data visualization makes the "invisible" visible. It enables people to move beyond just looking at data towards actually seeing the shapes and magnitudes of its physical properties to inform and enlighten.

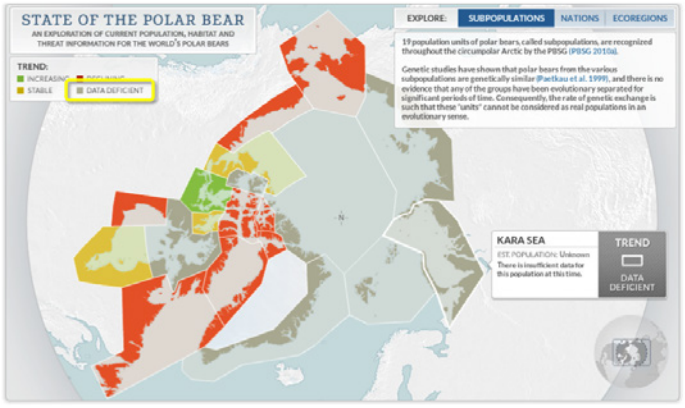
Discernibility is a prominent guiding decision: making the size of values as readable as possible, the distinction between categories as identifiable as possible, and the nature of relationships between entities as evident as possible.

Yet, what if there is no size? What if there are no amounts for a category? What if no relationships exist?

Welcome to the design of nothing, a delicate and often neglected aspect of data visualization concerned with showing the absence of data, representing zero, and utilizing the property of emptiness. How do we make these slippery attributes of nothingness visible?

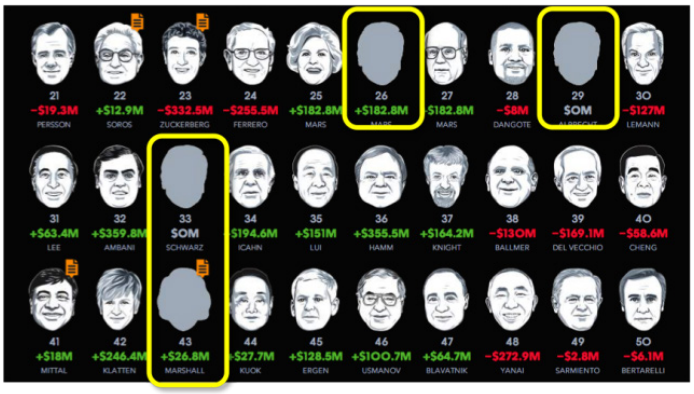
Showing the absence of data: Though analysts and designers naturally seek to work with data that is complete, missing data can be just as revealing as the data itself. Here are two projects that make the "nothing" of missing data into something.

Firstly, in the "State of the Polar Bear," by Periscopic, we see an attempt to convey the change in the population and habitat of the [p]olar bear around the Arctic.



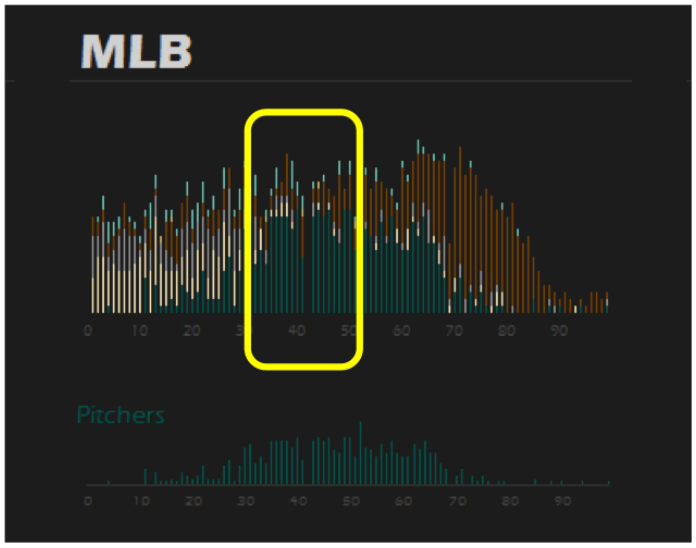
Aside from the sadly apparent areas of population decline, a key observation from this project is the "data deficiency" status for a large section of the displayed region, mostly part of Russia. As one of the project's authors, Kim Rees, explains, these areas without any data were not excluded, just grayed out: "It was a political statement to Russia to release the data they have about polar bears." The absence of data does not hinder the project, it adds a whole new dimension to it.

In the "Billionaires" visualization, by Bloomberg Visual Data, we see a number of generic blank faces included among the illustrations of the top 200 billionaires.



Those blanked out faces represent the reclusive and elusive, the ones for whom only a college yearbook photograph exists in the public domain. There's something compelling about those who manage to preserve such anonymity, and the inclusion of their absence adds intrigue to the visualization.

Representing zero: When it comes to the challenge of representing zeros, we are not talking about the absence of data but rather the absence of amount. In the below snippet from a graphic titled "The Uniform Distribution," created by Dark Horse Analytics, we see the number of MLB players who wear each different numbered jersey.



What is striking about this chart is the complete absence of players who wear the number 42. This isn't a gap in the data but simply illustrates the significance of the retiring of the #42 shirt in 1997 to honor Jackie Robinson. The retirement of Mariano Rivera in 2013, the last player allowed to wear #42, secured the zero value in the chart above.

Another example of portraying zero comes from the famous dot-point map created by John Snow to plot the deaths from the outbreak of cholera in London in 1854. One of the most striking findings from the map is the lack of deaths registered at the brewery, despite its proximity to the proposed source of the disease.

On investigation it was discovered that the brewery workers drank the beer they were making, rather than the water originating from the pump on Broad Street.

In both [of] these examples, zero needs to be given a home, as it is a crucial element of the story. The absence of an amount in each case is evident through contrast with non-zero values, though to identify these as zeros and not gaps in data does require some domain knowledge for interpretation.

Utilizing emptiness: The final aspect of designing nothing concerns the deliberate deployment of blank space. Through utilizing emptiness we can create quite striking displays of data and help increase readability.

In the example below, taken from the front page of the *Independent* newspaper, the large area of empty space on the right hand side provides context for the interpretation of the few countries that voted "No" for an immediate cease-fire in the Middle East.

The relative scales of each side are driven home by the generous use of white space, made all the more eye-grabbing by the fact that it is so seldom used on a newspaper's front page.

The design of nothing may sound like an oxymoron but it is anything but. Though visualizations most often focus on significant quantities and relationships, their absence is sometimes equally as interesting. For visualization design, there is always something in nothing.

10 KINDS OF STORIES TO TELL WITH DATA

BY TOM DAVENPORT

For almost a decade I have heard that good quantitative analysts can "tell a story with data." Narrative is—along with visual analytics—an important way to communicate analytical results to non-analytical people. Very few people would question the value of such stories, but just knowing that they work is not much help to anyone trying to master the art of analytical storytelling. What's needed is a framework for understanding the different kinds of stories that data and analytics can tell. If you don't know what kind of story you want to tell, you probably won't tell a good one.

This insight came in an interview a couple of years ago with Joe Megibow, a leading analytics practitioner who was head of web analytics at Expedia and is now Senior Vice President of Omnichannel E-Commerce at American Eagle. We simultaneously realized that there are several different types of analytical stories, and that it might be useful to create a typology of them. I later created what I think is the first typology of analytical stories in my book (with Jinho Kim) *Keeping Up with the Quants* and, since its publication last year, I've refined the typology further. Practically speaking, there are four key dimensions that determine the type of story you can tell with data and analytics:

Time: Analytical stories can be about the past, present, or future. The most common type of analytical story is about the past—it's a *reporting* story using descriptive analytics to tell what happened last week, month, quarter, or year. By the way, most visual analytics stories are also of this type. They're not the most valuable form of story, but it's undeniably useful to know what happened.

Stories about the present are most likely to involve some form of *survey*—an analysis of what people or objects are currently up to. It may actually involve survey research—asking people what they think about something. In some cases survey analysis involves a statistical model of what factors drive others. We might call those *explanatory survey* stories. In general there are lots of minor variations on the survey story. In the book we talk about social science surveys, surveys of cable TV viewers, and surveys of bombers in World War II.

Stories about the future are *predictions*; they use, of course, predictive analytics. They take data from the past (it's hard to get data from the future!) to create a statistical model, which is then used to predict the future. Quants create prediction stories all the time—about what customers are likely to buy, about how likely it is for an event to happen, about future economic conditions. These types of prediction stories always involve assumptions (notably that the future will be like the past in some key respects) and probability. The good news is that we can specify the likelihood that the story will be true; wouldn't that have been nice for fairy tales?

Focus: Are you trying to tell a *what* story, a *why* story, or a *how to address the issue* story? (I am thankful to several Procter & Gamble executives for pointing this out.) *What* stories are like reporting stories—they simply tell what happened. *Why* stories go into the underlying factors that caused the outcome. *How to address the issue stories* explore various ways to improve the situation identified in the what and the why stories. A really complete story may have all of these focus elements. P&G has made considerable progress in getting agreement on the *what* story quickly, and then spending more time and energy on the *why* and *how* topics. Key to doing that is having all parties involved in telling the story working off the same data.

Depth: There is also a depth dimension to analytical stories. When I spoke with Joe Megibow, then at Expedia, about this, he said that many of their stories were "CSI" projects—relatively small, ad hoc investigations to find out why something suboptimal was happening. One of his favorite examples involved discovering why some Irish customers were dropping online transactions when they got to the postal code input form. It turned out that some rural Irish locations don't have postal codes. Just like on CSI—story solved in a short time.

The alternative I call "*Eureka*" stories, which involve long, analytically driven searches for a solution to a complex problem. When you solve it, you want to yell, "Eureka!" My primary example in the book involved discovering the right way to refer and price potential buyers to real estate agents at Zillow. The project was core to the company's business model and was worth a long story; after a few false starts and the use of several different analytical methods, they got it right. These types of stories are typically long, important, and expensive, so getting stakeholder buy-in is critical if you plan to reach the end of them.

Methods: Finally, there are different types of stories based on the analytical method used. Are you trying to tell, for example, a *correlation* story—in which the relationships among variables rose or fell at the same time—or a *causation* story, in which you'll argue that

one variable caused the other? In most cases, doing some sort of controlled experiment is really the only way to establish causation. People—particularly those in the media—tell bad stories all the time because they confuse causation with correlation.

These ten kinds of stories are not mutually exclusive. There are certainly other method-based stories, and probably other important dimensions as well. But knowing that there are at least ten ways to tell analytical stories is much more useful than knowing only that you should tell one. There are other important aspects of analytical storytelling as well, such as that the story told to businesspeople should generally begin with the result and recommended outcome. You can save the details of how you got there analytically for the footnote of your report or presentation. And terms like R2, coefficient, logistic, and heteroskedasticity should not be appearing in your public story at all!

THE CASE FOR THE 5-SECOND INTERACTIVE

BY SCOTT BERINATO

Watching a "cold read" of a data visualization is revelatory. Try this: Hand some friends a printout of an infographic and ask what they think. Then, watch them. People's eyes will dart, their jaws will clench. Sometimes they will move the paper around, as if meaning will emerge at just the right angle. Loose bits of internal dialogue sneak out: *OK, what does that bar—right, got it.* For some, defiance takes over if meaning eludes them: *What am I looking at? Why is it green? I don't get it.* One colleague of mine grabs the paper—possessing it seems to be part of making sense of it—and then takes a deep breath, seemingly preparing for a mental fight.

I have no doubt about the value of data visualization. We don't need to belabor the fact that visual information rules and that data-viz literacy will be as fundamental a business skill in the future as spreadsheet literacy is today. But this problem with the cold read bothers me. Even reasonably simple visuals seem to spark confusion, and that's a problem if the goal is to improve understanding and be persuasive with data. I've begun actively trying to build graphics that reclaim that mental energy lost to parsing a graphic for what it could be better spent on: actually analyzing data and forming ideas based on it.

This was the first prototype for beating the cold read problem: http://blogs.hbr.org/2014/05/the-case-for-the-5-second-interactive/.

You might be asking, *That's it?* The entire experience requires three clicks and about five seconds. It generates a stacked bar chart. All I've done is deconstruct a chart into three simpler ones, then allowed the user to put it back together at his or her own pace. Nothing's particularly clever here; the idea behind it is simple: If cold reads of visual information are difficult because users have to make decisions about where to focus, [and] you remove as many of those decisions as possible, users will require less cognitive work to comprehend the information. I'm betting the five seconds it takes to get through the graphic more than makeup for the amount of work the user would have done parsing a static image that presented all the data at once.

Of course, it's a dilettante's hypothesis. I'm no scientist, but I did ask the folks at the Harvard Vision Sciences Lab to weigh in on the idea. The team there said that studies that looked specifically at how our brains process infographics as a whole were limited; most of the research is far more (mind-numbingly more) specific than that—but some themes in the research do seem to support an approach that simplifies visuals, and removes choices for our brains.

Good enough for me. And my method, while not scientific, wasn't haphazard either. I followed these four basic principles when building this five-second interactive:

- *Eliminate data whenever possible.* This is borrowed from Braess' Paradox. The actual theory applies to how adding route options in traffic systems reduces overall performance. My simple version is: The more data points you present, the harder you make it for users to decide where to focus and how to proceed.
- *Explain the data as simply and explicitly as possible.* That's just good editing.
- *Use animation to inform, not decorate.* Movement on a screen should help the user make sense of the information; otherwise, minimize or eliminate it.
- *Tell a story.* I was inspired here by John McPhee, a master storyteller who often employs what I'd call "micronarratives"—a few sentences within an essay that tell a story, often to explain something complex. One such McPhee micronarrative, for example, explained how a log burns; another explained why there's gold in mountains. Point is, humans like narrative; it helps us engage and deal with complexity, even if the narrative is as small as the one above.

Here's how those principles were applied in the above chart on Football v. Rugby: The opening scene eliminates more than half of all of the information contained in the source dataset. Data are explained and labeled unambiguously: Ask a question (How long are games?), get an answer (60 minutes and 80 minutes).

Next we add five new pieces of information, but crucially we also *remove three*—the previous subhead and labels. This lets the user (OK, forces the user to) focus on the most important information: the new data. The animation to expand the new bars over the originals isn't a design flourish; it's functional. It explains the relationship between data points. The second bar (how much action occurs) literally fills up part of the first one; it's a subset. State three repeats this same melody, and the narrative concludes with the full reveal of all of the data. By that point, hopefully, no parsing is necessary. We already get it, and we can spend our time thinking about the meaning rather than trying to figure out the meaning.

We've tried this approach at HBR here, and here. I hope it can become a format that's as natural a component of digital content as static infographics are to print. But I don't think it's right for every dataviz. Sometimes, complexity is the point, and some would argue there's value, even joy, in spending time with complex visuals and discovering as we go. Sometimes we don't know the answers the data are providing, so we want to spend time coming up with multiple possible interpretations rather than being told what the narrative is. Datasets that allow for user-chosen variables to define the visualization probably won't work in this model, either.

But for the countless good examples of "here's the point" data visualizations that stream through Twitter and get social love on the Internet every day, it could be a valuable tool that increases the effectiveness of the visualization by beating the cold read.

HOW DATA VISUALIZATION ANSWERED ONE OF RETAIL'S MOST VEXING QUESTIONS

BY GRETCHEN GAVETT

Sometimes it's relatively easy to know what your customers are doing. In e-commerce, advances in tracking and analytics have made it possible for retailers to understand what individual customers are doing before they make a purchase, and to gather and analyze hundreds and thousands of data points to identify trends.

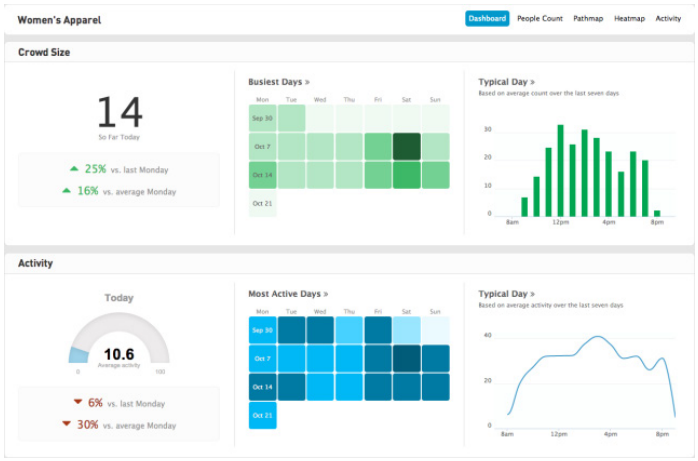
Brick-and-mortar stores haven't had the same advantage.

"Retailers are all using scanner data to track what happened at the point of *sale*," says Sam Hui, an associate professor of marketing at NYU's Stern School of Business. "But they have no idea what's really happening at a point-of-purchase *decision*."

This is changing with the emergence of location analytics. Take Alex and Ani, which designs and retails jewelry, and Belk, a department store chain. Both have signed on with Prism Skylabs, a software company, to map in-store customer behavior.

By using a store's existing security cameras, or installing new ones, Prism (no relation to the NSA program) is able to track the movement of a store's customers and identify patterns. "We're not really looking at any individual; we're looking at what a group of people over a period of time do," says senior vice president of managed services Cliff Crosbie. "That's the really big thing: identifying what a volume of people do over a period of time, and how you read that information."

In many ways, Prism is capturing the simplest aspects of shopping, aspects ecommerce websites now take for granted. "Retailers want to know what parts of their store are busy, and where customers particularly shop. So, if there's a promotion on, when do people stop there and what do they do?" Prism can also track what happens on individual days, or over time, using a dashboard like this (rather than reams of Excel spreadsheets):



These simple, color-coded data visualizations allow retailers to turn a store floor into an analytics narrative. (A new version also takes weather into account.)

Prism can also convey information on customer movements as a heat map. Consider this example from Alex and Ani during a pilot program during last year's holiday season, which tracked customer movement on the floor over a three-week period. The redder the location, the more frequently it was trafficked:

For chief technology officer Joe Lezon, the results were both helpful and surprising. "We now know that there was a certain area in our store people went to more often," he told me. "We also realized that 98% of the people turned right when they first entered the store." Lezon, along with Alex and Ani's head of merchandising and head of sales operations, used the data to inform product placement.

In one instance, a slower-moving product was moved to a more trafficked location, resulting in an uptick in sales. And when the location of [a] store's more popular items were shifted, Lezon and his team were able to watch the process by which customers were able to locate them.

Both Lezon and Greg Yin, Belk's vice president of innovation, told me that the heat mapping is particularly valuable when it comes to maximizing the value of staffing—making sure customers have a salesperson to assist them, easing the burden of the busiest times on sales associates.

Yin also says collecting and visualizing this data ha[ve] helped his company test out in-store assumptions quickly. "I don't think we're in a place in the industry right now where we can invest 12, 18 months in a long [research] project because the technology will have changed by then," he explained. "It's not about building out big, long-term solutions. It's about building a foundation in our stores and online so we can move as our customer moves."

And while there are some privacy concerns, Prism, unlike other kinds of online tracking, promises a level of anonymity.

"We've had cameras in stores for years," Yin reminds me. "But the nice thing about Prism is that it's anonymizing. There's no personal data being reflected because it's all aggregated." At the same time, he recognizes that "when we're talking about location-based marketing, we're really talking about personalization."

And when it comes to personalization, there has to be a give-and-take between the customer and the store; "research shows that many customers are willing to opt in to these kinds of things as long as there's some kind of [benefit] in exchange."

He notes, however, that the kind of bartering with personal information that's resulted in so many successful recommendation algorithms, for example, doesn't necessarily translate to the in-store experience. "We have to understand that the online customer is different than the in-store customer, and that the expectations might be different," he says. "When you get into facial recognition and trying to assess the demographics of a customer coming into your store, then you're getting into a little bit more of a gray area."

And when it comes to just physically walking into a store, there's no real way for a customer to opt out of becoming a data point that, presumably, might make the shopping experience better in the future. A 2013 Pew study found that 64% of American adults cleared their cookies and browser history to become less visible online; even Prism's Cliff Crosbie notes that more people are switching off their Wi-Fi in stores. While Prism's technology removes the actual images of customers—something Crosbie says is "the right thing" to do—being tracked is still a hidden part of the shopping experience.

This is all the more important considering the fact that companies are just starting to experiment with how location analytics can both improve a shopper's experience and boost [companies'] sales. "We can correlate a slight uplift in the sales for slower moving products," says Lezon. "But in general, this is a tough metric."

"We could definitely see, after changing a display, the traffic really picking up there," Yin explained. "The next obvious piece is to really be able to triangulate some sales against that."

Those sales are what's most important to Yin. "I don't come in every morning and say, 'How am I going to innovate today?' That doesn't really exist," he explains. "The question is, 'How do we drive business? How do we provide a great customer experience? How do we best equip our associates?'"

"This is a really interesting time in retail," he continues. "All of these technologies are starting to come together—whether it's mobile, whether it's social, whether it's analyzing a lot of data—and they're coming together to meet the customer. At the end of the day, understanding customer behavior in stores and being able to take actions on it [are] problem[s] we're trying to solve."

For Joe Lezon, the ultimate goal is the coupling of data based on a customer's online and in-store experience.

"My ideal situation, to be honest, is: Gretchen, you walk into my store," he says to me. "I know who you are. I know why you're there: Your daughter's birthday is next week and you want to buy her a gift. At the same time, I know what you've purchased in the past so I can actually help direct you to the right products.

"How do you merge all the data together to get a full 360-degree view of the customer? That's where all this is going."

Imagine what that visualization might look like.

DECISIONS DON'T START WITH DATA

BY NICK MORGAN

I recently worked with an executive keen to persuade his colleagues that their company should drop a long-time vendor in favor of a new one. He knew that members of the executive team opposed the idea (in part because of their well-established relationships with the vendor) but he didn't want to confront them directly, so he put together a PowerPoint presentation full of stats and charts showing the cost savings that might be achieved by the change.

He hoped the data would speak for itself.

But it didn't.

The team stopped listening about a third of the way through the presentation. Why? It was good data. The executive was right. But, even in business meetings, numbers don't ever speak for themselves.

To influence human decision making, you have to get to the place where decisions are really made—in the unconscious mind, where emotions rule, and data is mostly absent. Yes, even the most savvy executives begin to make choices this way. They get an intent, or a desire, or a want in their unconscious minds, then decide to pursue it and act on that decision. Only after that do they become consciously aware of what they've decided and start to justify it with rational argument. In fact, recent research from Carnegie Mellon University indicates that our unconscious minds actually make better decisions when left alone to deal with complex issues.

Data is helpful as supporting material, of course. But, because it spurs thinking in the conscious mind, it must be used with care. Effective persuasion starts not with numbers, but with stories that have emotional power because that's the best way to tap into unconscious decision making. We decide to invest in a new company or business line not because the financial model shows it will succeed but because we're drawn to the story told by the people pitching it. We buy goods and services because we believe the stories marketers build around them: "A diamond is forever" (De Beers), "Real Beauty" (Dove), "Think different" (Apple), "Just do it" (Nike). We take jobs not only for the pay and benefits but also for the self-advancement story we're told, and tell ourselves, about working at the new place.

Sometimes we describe this as having a good "gut feeling." What that really means is that we've already unconsciously decided to go forward, based on desire, and our conscious mind is seeking some rationale for that otherwise invisible decision.

I advised the executive to scrap his PowerPoint and tell a story about the opportunities for future growth with the new vendor, reframing and trumping the loyalty story the opposition camp was going to tell. And so, in his next attempt, rather than just presenting data, he told his colleagues that they should all be striving toward a new vision for the company, no longer held back by a tether to the past. He began with an alluring description of the future state—improved margins; a cooler, higher-tech product line; and excited customers—then asked his audience to move forward with him to reach that goal. It was a quest story, and it worked.

Good stories—with a few key facts woven in—are what attach emotions to your argument, prompt people into unconscious decision making, and ultimately move them to action.

THAT *MAD MEN* COMPUTER, EXPLAINED BY HBR IN 1969

BY ANDREA OVANS

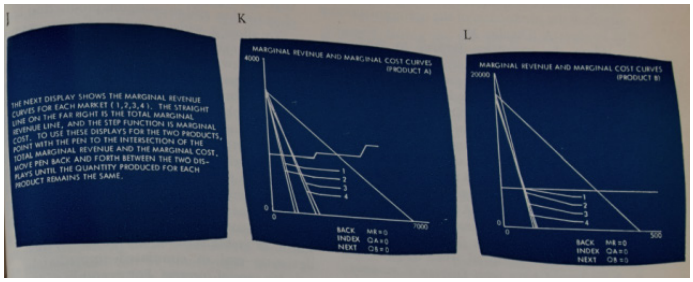
It's 1969 on this season's *Mad Men*, and a glass-enclosed climate-controlled room is being built to house Sterling Cooper's first computer—a soon-to-be-iconic IBM System/360—in the space where the copywriters used to meet.

That same year, in an article entitled "Computer Graphics for Decision Making," IBM engineer Irvin Miller introduced HBR's readers to a potent new computing technology that was part of the 360—the interactive graphical display terminal.

Punch cards and tapes were being replaced by virtual data displays on glass-screened teletypes, but those devices still displayed mainly text. Now the convergence of long-standing cathode-ray-tube and light-pen hardware with software that would accept English language commands was about to create a revolution in data analysis.

Previously, if executives had wanted to investigate, say, the relationship of plant capacity to the cost of production, marginal costs to quantity produced, or marginal revenues to quantities sold, they'd have to fill out a requisition, wait for a data analyst to run a query through the machine using some computer language like Fortran, and then generate a written report. That could take months.

But interactive graphics offered the possibility of providing realistic answers quickly and directly. As Miller explains: "With such a console in his office, an executive can call for the curves that he needs on the screen; then, by touching the screen with the light pen, he can order the computer to calculate new values and redraw the graphs, which it does almost instantaneously."

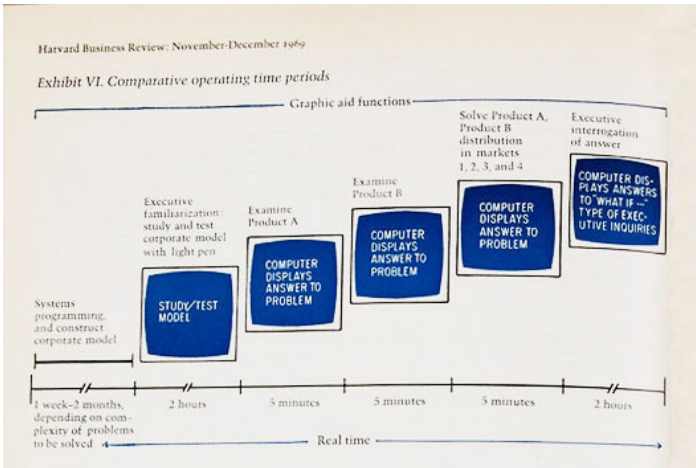


To read Miller's tutorial is to return to some first principles that may still be worth bearing in mind, even in today's world of vastly greater amounts of data and computing power (the largest mainframe Miller refers to has a capacity of two megabytes). The first is his almost offhand initial stipulation that the factors affecting a business that a computer can process are quantitative.

The second is his explanation (or, for us, reminder) of what the computer does when it delivers up the graphs: "To solve business problems requiring executive decisions, one must define the total problem and then assign a mathematical equation to each aspect of the problem. A composite of all equations yields a mathematical model representing the problem confronting the executive." Miller suggests, as an example, that a system programmed with data on quantities produced and sold, plant capacity, marginal cost, marginal revenues, total cost, total revenue, price, price for renting, and price for selling could enable businesspeople to make informed decisions about whether to hold inventory; expand plant production; rent, buy, or borrow; increase production; and examine the effects of anomalies on demand or the effects of constraints.

Even in this simple example it's easy to see how hard it is to "define the total problem"—how, for instance, decisions might be skewed by the absence of, say, information on interest rates (which in 1969 were on the threshold of skyrocketing to epic proportions) or of any data on competitors, or on substitutes (a concept Michael Porter wouldn't introduce until 1979).

Miller is hardly oblivious to the dangers (the term "garbage in; garbage out" had been coined in 1963); and in [response] to the question of why an executive should rely on the differential calculus and linear programming that underpinthe models (interestingly, Miller assumes senior business executives haven't [taken] calculus), he replies that the point of the equations is only to *"anticipate and verify intuitive guesses which are expected to be forthcoming from the businessman"* [italics original]. In other words, the mathematics are essentially meant to serve as an amplification of the executive's judgment, not as a substitute.



Intuition-support is, in fact, the point for Miller. For him, the real benefit of the new technology isn't just the ability to perform what-if analyses on current data, as powerful as that is, but that executives could do it in the privacy of their own offices, which would afford them the time for the private reflection from which intuition springs. "The executive needs a quiet method whereby he alone can anticipate, develop, and test the consequences of following various of his intuitive hunches before publicly committing himself to a course of action," Miller says, before he even begins to explain how the technology works.

In this it's enlightening to revisit Miller's estimates of how much time the entire process was supposed to take: a few weeks to construct the model, five minutes to conduct each what-if scenario—and then two full hours for the executive to consider the implications of the answers. In this, HBR's first examination of data visualization, it is in those two hours of solitary quiet time that the real value of interactive computing lies.

WHO'S AFRAID OF DATA-DRIVEN MANAGEMENT?

BY JEFF BLADT AND BOB FILBIN

From a management perspective, making decisions based on data is a clear win. Yet it's often difficult to adopt a data-informed culture. In every organization, there are teams and employees who embrace this transition, and those who undermine it. To convert your biggest data skeptics, the first step is to understand the psychology of their resistance.

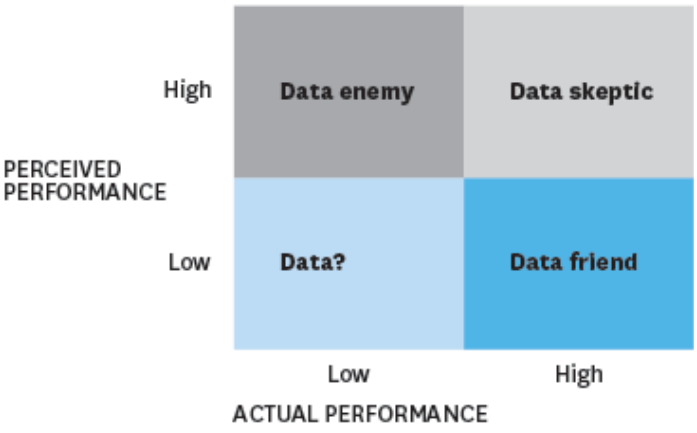
A data insight without a subsequent action is like a key without someone to turn it: worthless. A good data scientist can identify which coworkers will use insights from data to open new doors for the business, and which will continue to rely on intuition. This is because employees who act on data will do so for two main reasons: to improve their perceived or actual performance. From our perspective, there are four types of employees in any organization:

(1) Highly regarded, high performing
(2) Highly regarded, low performing
(3) Lowly regarded, high performing
(4) Lowly regarded, low performing

And their willingness to embrace data looks something like this:

FOUR TYPES OF EMPLOYEES AND THEIR WILLINGNESS TO EMBRACE DATA

Star performers are most resistant to data, while unsung workhorses are most likely to embrace it.



SOURCE DOSOMETHING.ORG HBR.ORG

Now intuitively, you would think that the first group (high-high, your overachieving all-stars) would be the easy converts to a data-informed culture; of course, they'll want the best tools and analysis at their disposal. But in our experience, the high-highs are the most likely to be data skeptics. Quantifying their domain and performance offers little upside. They are already perceived as doing quality work; adding hard numbers can, at best, affirm this narrative, and at worst submarine the good thing they have going. There is a reasonable fear that the outputs used to measure their performance will not fully capture the true value of their contributions. Skepticism is especially strong in any workplace where attribution is difficult (think marketing and media).

But, this group can be convinced: involve them early, give them a voice in creating the new metrics that will underpin the data-informed culture, and give them opportunities to push back. These efforts can make the data culture feel like their creation, not something that was forced upon them.

Your main challenge lies next down the list: the high-lows. These are your data antagonistics. Coworkers love them, but deep down they always fear they will be found out. Their ideas are occasionally fantastic, but too often they are just shooting in the dark. When things go right, they are never exactly certain why (their instincts are just that good), and when things go wrong, they instinctively turn to ass-covering mode. Quantifying their work (on someone else's terms, no less) only has downside. Swinging for the fences every at-bat is great, until the manager and fans learn to calculate (and value) on-base percentage. Then, 30 home runs with a .150 OBP is no longer getting the job done.

There's not a lot that can be done for this group. The malleable ones will eventually come around, but those stuck in their heuristicways will undermine and cavil the creeping in of a data-informed culture.

After this group, you have the low-highs. This group will be your biggest champion. Too long have they toiled on the lower reaches of the totem pole. Giving these overachieving, underappreciated employees the information and framework to make their work comparable—to allow their true value to be understood—provides only upside. Give this group early wins by focusing on tying their outputs to organizational success. They will love you for it, and they will help promote your cause. And senior management will be impressed.

This brings us to the last group, the low-lows. They aren't going to fight data culture. Or embrace it. They'll simply turn their heads 10 degrees and think: data? In general, low-lows either swim with the current, which means they'll come around when coming around is the safe thing to do, or against the current, meaning they won't be around long enough to matter.

Data-informed decision making, and the culture change inherent therein, doesn't happen in a vacuum. Asking what do the data say before acting is a disruptive action, displacing prior norms. There will be employees like the low-highs who welcome this kind of change, and those like the high-lows who subvert it. Understanding the psychology underlying these behaviors is the necessary first step toward pushing past intuition and silencing the data skeptics.

AN INTRODUCTION TO DATA-DRIVEN DECISIONS FOR MANAGERS WHO DON'T LIKE MATH

BY WALTER FRICK

Not a week goes by without us publishing something here at HBR about the value of data in business. Big data, small data, internal, external, experimental, observational—everywhere we look, information is being captured, quantified, and used to make business decisions.

Not everyone needs to become a quant. But it is worth brushing up on the basics of quantitative analysis, so as to understand and improve the use of data in your business. We've created a reading list of the best HBR articles on the subject to get you started.

Why data matters

Companies are vacuuming up data to make better decisions about everything from product development and advertising to hiring. In their 2012 feature on big data, Andrew McAfee and Erik Brynjolfsson describe the opportunity and report that "companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors" even after accounting for several confounding factors.

This shouldn't come as a surprise, argues McAfee in a pair of recent posts. Data and algorithms have a tendency to outperform human intuition in a wide variety of circumstances.

Picking the right metrics

"There is a difference between numbers and numbers that matter," write Jeff Bladt and Bob Filbin in a post from last year. One of the most important steps in beginning to make decisions with data is to pick the right metrics. Good metrics "are consistent, cheap, and quick to collect." But most importantly, they must capture something your business cares about.

The difference between analytics and experiments

Data can come from all manner of sources, including customer surveys, business intelligence software, and third-party research. One of the most important distinctions to make is between analytics and experiments. The former provides data on what is happening in a business; the latter actively tests out different approaches with different consumer or employee segments and measures the difference in response. For more on what analytics can be used for, read Thomas Davenport's 2013 HBR article "Analytics 3.0." For more on running successful experiments, try these two articles.

Ask the right questions of data

Though statistical analysis will be left to quantitative analysts, managers have a critical role to play in the beginning and end of the process, framing the question and analyzing the results. In the 2013 article "Keep Up with Your Quants," Thomas Davenport lists six questions that managers should ask to push back on their analysts' conclusions:

1. What was the source of your data?
2. How well do the sample data represent the population?
3. Does your data distribution include outliers? How did they affect the results?
4. What assumptions are behind your analysis? Might certain conditions render your assumptions and your model invalid?
5. Why did you decide on that particular analytical approach? What alternatives did you consider?
6. How likely is it that the independent variables are actually causing the changes in the dependent variable? Might other analyses establish causality more clearly?

The article offers a primer on how to frame data questions as well. For a shorter walk-through on how to think like a data scientist, try this post on applying very basic statistical reasoning to the everyday example of meetings.

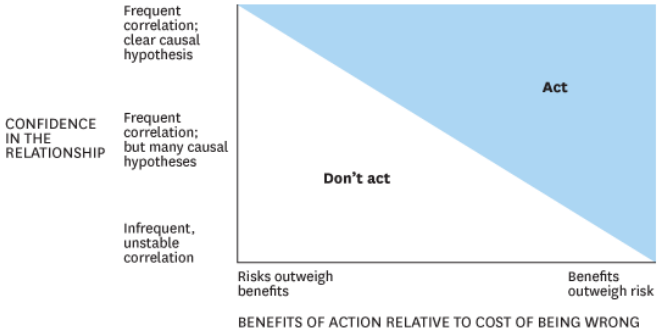
Correlation vs. cause-and-effect

The phrase "correlation is not causation" is commonplace, but figuring out just what it implies in the business context isn't so easy. When is it reasonable to act on the basis of a correlation discovered in a company's data?

In this post, Thomas Redman examines causal reasoning in the context of his own diet, to give a sense of how cause-and-effect works.

And BCG's David Ritter offers a framework for deciding when correlation is enough to act on here:

WHEN TO ACT ON A CORRELATION IN YOUR DATA
How confident are you in the relationship? And do the benefits of action outweigh the risks?



SOURCE DAVID RITTER, BCG HBR.ORG

The more frequent the correlation, and the lower the risk of being wrong, the more it makes sense to act based on that correlation.

Know the basics of data visualization

Rule #1: No more crap circles. To decide how to best display your data, ask these five questions. Make sure to browse some of the best infographics of all time. And before you present your data to the board, consult this series on persuading with data. (Don't forget to tell a good story.)

Learn statistics

A couple of years ago, Davenport declared in HBR that data scientists have the sexiest job of the 21st century. His advice to the rest of us? If you don't have a passing understanding of introductory statistics, it might be worth a refresher.

That doesn't have to mean going back to school, as Nate Silver advises in an interview with HBR. "The best training is almost always going to be hands-on training," he says. "Getting your hands dirty with the data set is, I think, far and away better than spending too much time doing reading and so forth."

WEBINAR SUMMARY

MAKING BIG DATA ACTIONABLE: HOW DATA VISUALIZATION AND OTHER TOOLS CHANGE THE GAME

FEATURING BILL FRANKS

Contributors
Bill Franks, Chief Analytics Officer, Teradata

Angelia Herrin (Moderator), Editor, Research and Special Projects, *Harvard Business Review*

Overview

With many organizations still using very basic tools to visualize their data, they are missing a huge opportunity. To make big data actionable and profitable, firms must find new and innovative ways to fully leverage their data and change their business through analytics.

One option is to adopt and use powerful visualization tools that are available today. Through visualization, organizations can find and communicate new insights more easily. The key to making these new insights actionable is to make analytics operational. This requires embedding analytics into business processes in an automated way and generating analytics on the fly, whenever decisions need to be made.

Context

Bill Franks discussed how to make big data more actionable by using compelling data visualization tools and techniques.

Key Learnings

Big data is changing the way companies do business.

Although big data can be defined in many ways, results are more meaningful than definitions. Some believe the "big data bubble" might be ready to burst, but Bill Franks has a different perspective. It's true that many people have unrealistic expectations about acquiring and using big data, and that hype will be deflated over time. However, the value of big data is real. Mr. Franks made the following observations about the ways in which big data is transforming business:

- **Big data can enhance the customer experience.** Disney recently introduced "magic wristbands" at its parks. These RFID bracelets provide insight into visitors' movements around the theme parks. That data is used to "create magic" and improve the customer experience.
- **Big data is blurring industry classifications.** In the past, Nike was known as a sportswear company. That has changed with the Nike FuelBand. This high-tech device expanded Nike's business into software, hardware, and storage of downloaded data. The skill sets that Nike used in the past are no longer sufficient, and new skills are needed to succeed. This dynamic is not exclusive to Nike but is playing out in companies across industries.
- **Companies are monetizing data and analytics.** Data that companies collect may be used to generate new revenue streams. For example, telecommunications companies have insight into the speed and volume of highway traffic, thanks to information collected from customers' mobile phones. This data can be repurposed into map applications. Companies must bear in mind, however, that customer privacy is paramount when sharing data.
- **Organizations are looking at new ways to make analytics operational.** Analytics can be embedded into processes in new ways. For example, the International Air Transport Association is considering airport tunnels that incorporate security detectors. This process would be less intrusive for passengers than current security practices. However, it would require real-time collection and analysis of data.

Over time, the process for creating powerful data visualizations has evolved and become more accessible.

Powerful data visualizations have always been possible. Before the advent of computers, however, they were difficult to create and limited for special circumstances. With computers, visualizations became more widespread.

Although early computer-generated graphics were very basic, people used text in ingenious ways to create simple but effective charts. By

"There are many definitions of big data, but results count more than definitions. The most important thing is finding data that matters to the business."—Bill Franks

the early 2000s, everyone could create nice visuals with common desktop tools. One drawback of these, however, was static data.

Today, customized visualizations that are tied to a specific type of analysis or concept are becoming more common. Examples include social network and affinity graphs. Over the past year, infographics have also become very popular. These tell a complex story by combining multiple visuals with short text excerpts. The next frontier of visualization is immersive intelligence that offers three-dimensional, interactive exploration of data.

Visualizations increase the impact of data and aid in discovery of new insights.

Visualizations enable people to see stories that words alone can't tell. For instance, a double helix illustration dramatically enhances the way individuals interpret information about DNA.

While visualizations can be powerful, Mr. Franks made two important comments:

1. **Visualization tools alone don't uncover insights—people are an essential element.** Visualization tools aid with discovery of new information but are not a silver bullet. People are needed to find, communicate, implement, and operationalize the insights that come from big data. Only people can ask the right questions and assemble the right visuals.
2. **Visuals shouldn't distract.** Visualization is all about impact. It's not about getting fancy for the sake of fanciness. In fact, overly ornate charts can reduce the impact of data. Customization options should be used in charts only if they add value (figure 1).

Visualization is about impact, not about getting fancy for the sake of it.

Today's visualization tools support decision making through interactive exploration of data.

Modern visualization technologies offer several distinct but related value propositions. They allow teams to create better visuals that bring data and analysis to life. In addition, they support forward-looking decision making rather than simply report on what happened in the past. Since visualization tools provide a flexible, self-service environment, they also democratize data within organizations.

Mr. Franks characterized modern visualization tools as "a spreadsheet plus a presentation package on steroids." Because they connect live to data sources, these technologies are interactive. With desktop tools, teams can collaborate and share graphic packages.

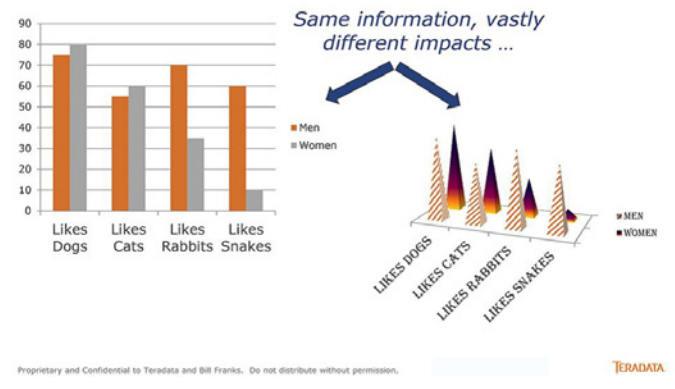


Figure 1

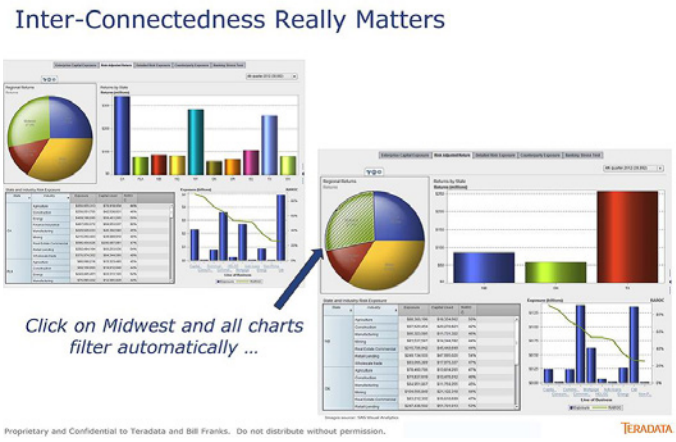


Figure 2

Today's visualization tools are also interconnected. Several charts can be linked together, which facilitates rapid exploration of data. It is possible to click on one chart and then immediately see several other charts with related data. This is compelling and makes presentations more effective.

It is hard to understand the impact of the interconnectedness of today's visualization tools without experiencing it.

Operational analytics are automated and actionable.

Until recently, most analytics were "artisanal"; they were typically generated by hand and weren't the product of a consistent process. Today, however, analytics are experiencing an "industrial revolution," and the result is operational analytics.

Operational analytics are embedded in business processes in an automated way. They are prescriptive and generated in "decision time"; that is, they are available whenever decisions need to be made. This enables scalability at a level never seen before. The power of big data lies in the new information that is provided to analytic processes, such as cross-channel data to support ecommerce analysis or web browsing behavior to support consumer behavior analysis.

It is important to note that "decision time" operational analytics are not the same as batch analytics, which are applied operationally.

"Visualization is not a magic bullet but a useful tool to help people find and communicate new insight." —Bill Franks

An example of the latter is a website that is updated nightly with new customer recommendations. In contrast, an example of decision time operational analytics is a website that is optimized in real time based on customer clicks and preferences.

To make big data actionable, IT must remove the barriers between employees and data.

Organizations derive the greatest value from big data when employees are free to explore and conduct their own analyses. Creating this type of environment often means that the IT team must change from a serving model to one of enabling.

Mr. Franks described the differences between the traditional IT model and an enabling model:

- **The traditional IT model.** In this case, the IT team controls access to data and analysis tools. Mr. Franks compared this approach to traditional frozen yogurt shops, where servers prepare orders and customers pay per cup and per topping.
- **The discovery model.** To move from serving to enabling, IT teams must remove the barriers between customers and data. Once this occurs, people can generate their own insights. Mr. Franks compared the discovery model to new frozen yogurt shops where customers select and dispense their own yogurt and toppings, and pay per ounce.

Important Point

Who owns analytics? Some organizations have two distinct roles related to analytics: the chief analytics officer and the chief data officer. The CAO reports to the business, while the CDO reports to the chief information officer. The two positions are closely linked.

Biographies

Bill Franks, Chief Analytics Officer, Teradata

Bill Franks is chief analytics officer for Teradata, where he provides insight on trends in the analytics and big data space and helps clients understand how Teradata and its analytic partners can support their efforts. In addition, Franks is a faculty member of the International Institute for Analytics and the author of the book *Taming the Big Data Tidal Wave.* His second book, *The Analytics Revolution,* is coming in fall 2014. He is also a sought-after speaker and frequent blogger. His work has spanned a variety of industries, with clients ranging in size from Fortune 100 companies to small nonprofit organizations. You can learn more at http://www.bill-franks.com.

Angelia Herrin, Editor for Research and Special Projects, *Harvard Business Review*

Angelia Herrin is editor for research and special projects at *Harvard Business Review.* At *Harvard Business Review*, Herrin oversaw the relaunch of the management newsletter line and established the conference and virtual seminar division for *Harvard Business Review.* More recently, she created a new series to deliver customized programs and products to organizations and associations.

Prior to coming to *Harvard Business Review,* Herrin was the vice president for content at womenConnect.com, a website focused on women business owners and executives.

Herrin's journalism experience spans twenty years, primarily with Knight Ridder newspapers and USA Today. At Knight Ridder, she covered Congress as well as the 1988 presidential elections. At USA Today, she worked as Washington editor, heading the 1996 election coverage. She won the John S. Knight Fellowship in Professional Journalism at Stanford University in 1989–90.

LEARN MORE

DATA VISUALIZATION: WHAT IS IT AND WHY IS IT IMPORTANT?
→ sas.com/en_us/insights/big-data/data-visualization.html

FOR MORE ABOUT SAS DATA VISUALIZATION AND ANALYTIC CAPABILITIES
→ sas.com/visualanalytics

TRY SAS® VISUAL ANALYTICS: GET DEMO ACCESS TO EXPLORE A SAMPLE DATA SET AND BUILD A REPORT
→ sas.com/en_us/offers/13q2/va-try-before-you-buy/register.html

ABOUT SAS

The leader in business analytics and the largest independent vendor in the business intelligence market, SAS specializes in deriving value from data of any size. SAS® Visual Analytics is a BI platform combining high-performance in-memory technology with easy-to-use analytics, data visualization and reporting. It enables you to spot patterns from massive data volumes, identify opportunities for further analysis and share results via the Web or mobile devices. SAS gives customers at more than 70,000 sites around the world THE POWER TO KNOW®.

A HARVARD BUSINESS REVIEW INSIGHT CENTER REPORT

Harvard Business Review