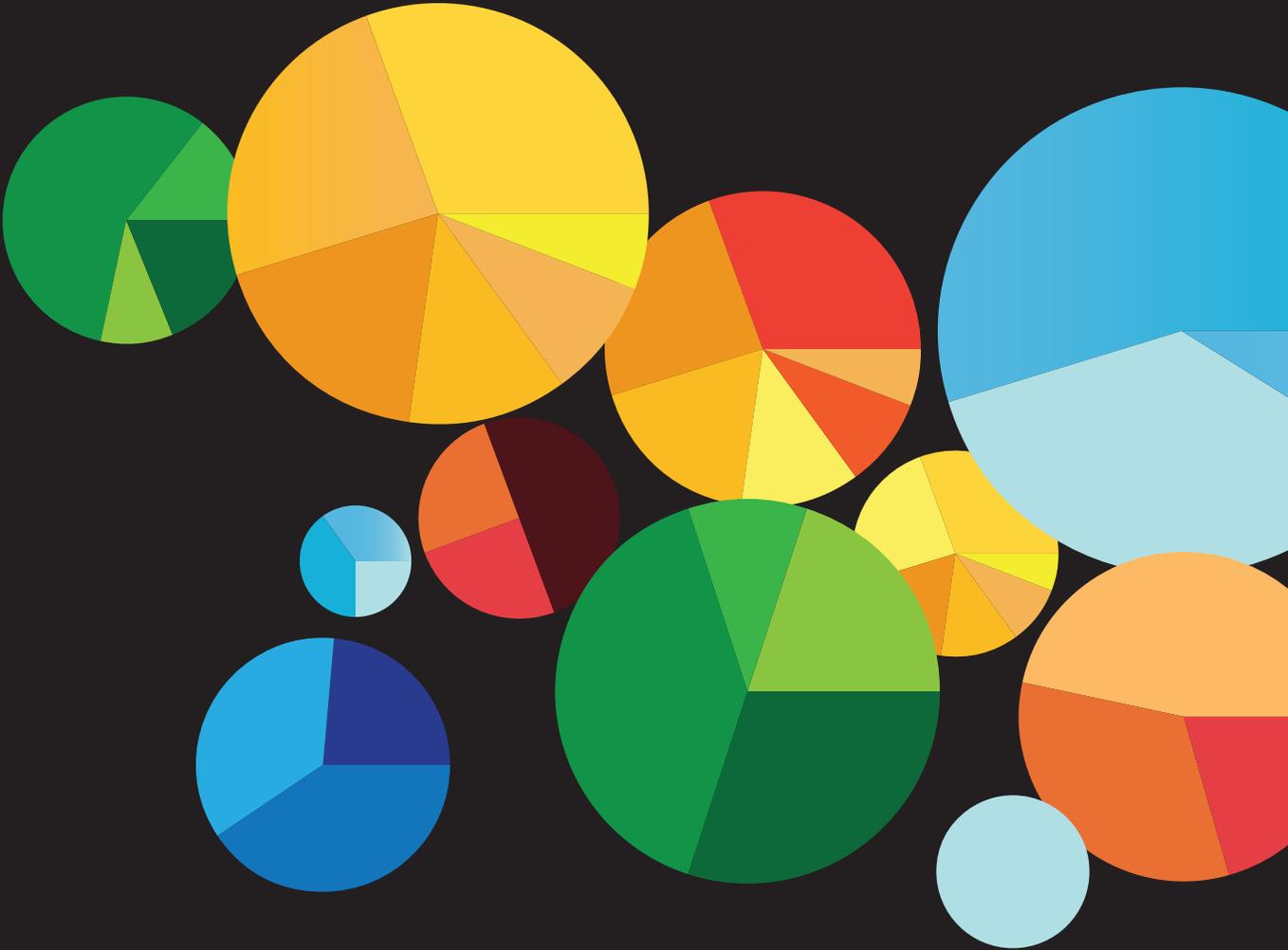


VISUALIZING DATA

A HARVARD BUSINESS REVIEW INSIGHT CENTER REPORT



Sponsored by



VISUALIZING DATA

The HBR Insight Center highlights emerging thinking around today's most important business ideas. In this Insight Center, we'll explore the power of using data visualization to drive business strategy. We'll talk about when (and when not) to use visualization, how to get started, how to know if you're getting a good return on your data visualization investment, and more.

- 1 **Editor's Welcome: Your Business Needs Insight Not Just Pretty Pictures**
by Scott Berinato
- 3 **The Power of Visualization's Aha Moments**
by Scott Berinato
- 5 **The Value of a Good Visual: Immediacy**
by Bill Franks
- 7 **The Question All Smart Visualizations Should Ask**
by Michael Schrage
- 8 **When Data Visualization Works and When It Doesn't**
by Jim Stikeleather
- 10 **We've Reached Peak Infographic, and We're No Smarter for It**
by Dylan C. Lathrop
- 12 **It's Time to Retire 'Crap Circles'**
by Gardiner Morse
- 14 **What Moleskine's Market Position Really Looks Like**
by Dan McGinn
- 15 **When Presenting Your Data, Get to the Point Fast**
by Nancy Duarte
- 17 **Ten Years of News Corp. Income Data in Less Than a Minute**
by Gretchen Gavett
- 18 **Data for All! How New Tools Democratize Visualization**
by Bill Franks
- 19 **Visualization as Process, Not Output**
by Jer Thorp
- 21 **Telling Stories with Visual Data: A Glimpse into the Future of Narrative**
by Jer Thorp
- 22 **When Creating Visualizations, Question Everything**
by Irene Ros and Adam Hyland
- 24 **Visualizing Hip Hop Lyrics as Cultural Indicator**
by Jeff Kehoe
- 25 **Tell Better Data Stories with Motion and Interactivity**
by Andrew DeVigal
- 27 **The Science of What We Do (and Don't) Know About Data Visualization**
by Robert Kosara
- 29 **The Chart Wars Have Begun**
by Jake Porway
- 30 **WEBINAR SUMMARY**
Get the Picture: Gaining Insight with Data Visualization
Featuring Jeremy Howard

EDITOR'S WELCOME: YOUR BUSINESS NEEDS INSIGHT NOT JUST PRETTY PICTURES

BY SCOTT BERINATO

In 2007, if you were a Starbucks shareholder and you opened your annual financial report, the first meaningful information you encountered was this:

REVENUE COMPONENTS

The following table shows the Company's revenue components for the fiscal year ended October 1, 2006:

REVENUES	% of Total Net Revenues	% of Specialty Revenues
Company-operated retail	85%	
Specialty:		
Licensing:		
Retail stores	7%	45%
Grocery and warehouse club	4%	24%
Branded products	<1%	2%
Total licensing	11%	71%
Foodservice and other:		
Foodservice	4%	27%
Other initiatives	<1%	2%
Total foodservice and other	4%	29%
Total specialty	15%	100%
Total net revenues	100%	

And if you read the “narrative” annual report from that year, the first thing you saw was this:

We all have a Starbucks we call our own.

For customers, it may be based on a cherished coffeehouse experience. A smiling barista who remembers how they prefer their morning latte. Or the welcoming vibe they feel each time they enter their neighborhood store.

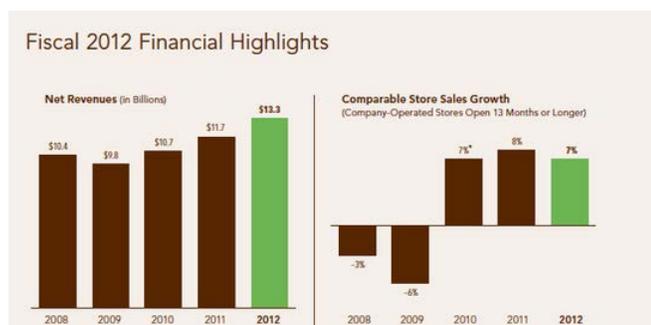
For coffee farmers, this sense of shared ownership may evolve as we work together to make their farms more productive and their communities more prosperous.

And for our employees, whom we call “partners,” it may represent something altogether different—a company with an enduring commitment to be caring, upfront and fair.

But no matter how you define your Starbucks, each unique meaning stems from the commitment that we've made to do business in a different way. A way that truly values individuals and their contributions. A way that balances good business practices with a passion to improve this world we share.

A better way.

Last year, if you were a Starbucks shareholder, you received only one annual report, and the very first page after the cover began with this:



The evolution of annual reports represents a broader trend in business communication: Data comes first, and it's increasingly visual.

Welcome to the Visualizing Data Insight Center, where, for the next month, we'll explore the broadening role data visualization plays in business communication. We believe this is not only an inevitable trend but also one you must embrace if you want to effectively communicate with all your stakeholders.

Data visualization is taking hold now because of two trends. The first: Big data is here, it must be analyzed, and one of the best ways to make sense of it is with visual representations. The second: The tools to create good data visualizations are being democratized, which has led to a growing community of programmers, designers, and statisticians who can apply their analytical and intuitive powers to creating meaningful visual stories.

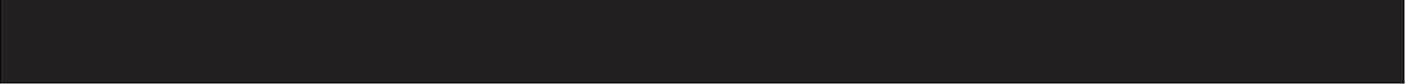
We've already declared the data scientist to be the sexiest job of the 21st century. Those scientists will need visualization experts the way writers need editors.

While we'll be sharing some stunning visuals over the next month, this Insight Center is not just focused on pretty pictures. Our aim is to get you acquainted with this burgeoning discipline, emphasizing how and when you can practically apply visualization in order to be a more effective communicator. We'll cover everything from the basic — how you can improve PowerPoint slides using visualization — to the advanced ways graphics can make sense of big data sets.

We'll also share our contributors' and editors' favorite visualizations, with explanations of why they find them so useful. And, of course, we'll share some all-time bad examples as cautionary tales.

We also want to know what your favorite — or least favorite — data visualizations are. You can email me or post links in the comments.

We're kicking off the Insight Center today with a good mix of introductory content. Designer and art director Dylan C. Lathrop explores the growth of data viz and the positives and negatives that have emerged during this period. I interviewed Amanda Cox, a graphics editor at *The New York Times*, to talk about how they create their stunning — and stunningly useful — visualizations. How do they decide what's worth doing? And how do they staff against it?



And for fun, we're running one of our most popular features on "crap circles," exploring the world of remarkably awful PowerPoint cycle graphics.

Thank you for joining us, and jump in with your ideas, insights, and visualizations as well.

THE POWER OF VISUALIZATION'S AHA MOMENTS

BY SCOTT BERINATO

Amanda Cox has been a graphics editor at The New York Times for eight years. Trained as a statistician, Cox develops visualizations across platforms, from simple print infographics to highly complex online interactive data tools. The Times is a visualization leader, but Cox believes the best is yet to come from this discipline, which she calls “both young and not young.” I spoke to Cox about The Times’ approach to visualization and the power of aha moments.

Do you think data visualization is entering a time when it’s becoming a core communication tool?

I wish there were more examples in the high-end data viz world to back that up. I wish there were more examples where data viz actually mattered. The case studies for us to lean on are sparser than they should be. On the other hand, you can argue it’s a young field and people are doing all kinds of crazy interesting things, and that’s a good thing. There’s that classic idea that it’s useful for people to do crazy theoretical things even if they don’t know what they’re useful for. Like with origami. Folding patterns turn out to be really useful for many applications, including surgery. But it wasn’t clear when people were documenting origami patterns that they would eventually help save lives. I want applications that actually matter. Actionable results.

It seems like there’s more focus on trying to get data viz to go viral than to make it “matter.”

There’s a lot where not much action comes out of it. I don’t know if the ratio is different from the ratio of bad writing to good, or bad restaurant openings to good, but I think it’s an important idea to focus on. There’s a strand of the data viz world that argues that everything could be a bar chart. That’s possibly true but also possibly a world without joy.

Nicely designed posters with a few numbers on them aren’t really data viz.

What is the right skill set for building good data visualizations? Seems like the right person is a unicorn: someone who knows some design, some programming, and some statistics.

Here, anyway, I’ve heard us described as very liberal arts. Those skills you list are useful, but we start with curiosity and skepticism. Most people here don’t come from statistics, graphic design, or programming backgrounds. We have journalists, biologists, urban planners, mapmakers. The ability to ask good questions is really

what we start with. I come from a statistics background, and I’m finding statistics students’ portfolios are crazy weak compared to the computer science students, even though they’re playing with the same problems. I think it’s because comp sci students are encouraged to play, whereas with stats majors, it’s, “Here’s your rule book, now make things.” I don’t think that’s a good model for making better visualization.

But surely you have to have some of those three core skills.

In bigger projects, we put together teams where those skills are reflected, but it’s not like we all need to know how to program. I bring a statistical background. But I’m not a designer by any means.

That’s surprising that, as a graphics editor, you don’t consider design a strength.

When I first started, I thought design was ten minutes of “make it cute” at the end that I could talk someone into doing for me. Now I know that design thinking needs to be involved from conception. And after a while you see the math behind it. How do I minimize eye movement on this infographic? Something like that — now I know how to do that because we have design principles. Design and typography do matter. It’s about hierarchy of information and how people perceive information. Done properly, that cleanup work really matters. On the other hand, it’s easy to believe that it matters more than it does. If you make a fantastically interesting chart and some poor design decisions, the data will still come through. If you make a bad chart with a beautiful design, what have you done, really?

Is there resistance to data viz at *The Times*? Does anyone think it’s a lot of work for not much reward?

Not a ton of resistance here. Editors and reporters rightfully question why we might want to do a visualization. It’s not always immediately obvious to them the journalistic value, and sometimes they’re right. The parallel to me is that there’s room in a newspaper for quirky text stories too. Most everyone here would agree that the best way to tell some stories is through data. Some think this is so very rarely, some think this is so most of the time, but they would concede that telling the story with data is accepted.

What about the cost-benefit part of it? Data viz is still somewhat new, and sometimes it seems like it’s a lot of work to make a simple point.

We might be different than a non-news organization, but the cri-

THE VALUE OF A GOOD VISUAL: IMMEDIACY

BY BILL FRANKS

Our brains are meant to see in pictures. Grids and columns of data, while ubiquitous, make it very difficult to see trends or patterns. Additionally, a lot of the new data sources available today, such as genetic data or social network data, don't lend themselves to traditional spreadsheets and graphs. These data types require a different way of displaying them to allow us to see the underlying patterns and stories in the data.

I'd like to walk you through an exercise to illustrate how effective visualizations allow you to immediately comprehend a complex set of relationships. Consider a standard map, such as the map of the United States below. If I show you the map and ask you to describe how a few states are related to one another, you can immediately visualize and verbalize an answer. New York is up and to the right of Virginia, while Texas is down and to the left. There are several states in between the two, and Texas borders fewer states than Virginia does. Simple, right?

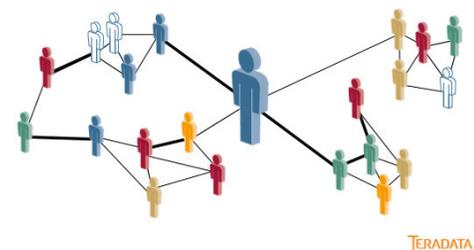


Next, I take away the map and give you a set of data that gives relative location, size, spatial border details, and other information about the states. Could you describe the same relationships in just a few seconds based on the data alone? I'll bet you'd throw your hands up in the air — just like I do. I have never been able to find a remotely reasonable way to explain the information immediately visible to me in a map without making use of a map.

You might not immediately grasp the importance of the above example because it seems self-evident that maps are important. I'd like to suggest to you that the reason for that is that you are aware that maps exist in the first place. Had you never seen a map, you'd be struggling to explain the information that a map conveys in some other way.

This is the key to the value proposition of data visualization. It could be that you are struggling to convey information without

being aware that there is a visual that can have the same type of impact as a map. Or there may be connections among all that data that you'd never make without a visualization. Until you see the visual for the first time, however, you won't appreciate the value it offers. Take a look at the social network graph below. It shows that there are several distinct groups of people who interact a lot among their own groups and don't really interact outside of their groups. However, there is a single person who has contact with each group and connects them. That person would be the critical one to reach if you want to influence those groups easily. See how easy it is to understand the connections between people and groups?



As the map example illustrates, data visualizations can make it easy to rapidly understand relationships, patterns, and stories that are contained within a complex data set. This is the reason they are so powerful. There is a reason for the saying "A picture is worth a thousand words." It also holds for data points.

One of the best features of modern visualization tools is that they permit interactivity with the underlying data. In other words, a visual isn't static. You can click on various parts of a visual to drill into different views of your data on the fly. While many business intelligence tools have enabled drill-down reports for years, they typically contain only common visuals and also typically constrain users to predetermined paths. Visualization tools today don't apply many limits on what users can do, which opens up a lot more options for analyzing data.

A few years ago, we put a popular visualization tool on my team's laptops. It was a huge hit. Over time, several members of my team stopped using traditional spreadsheet and presentation tools altogether in favor of the visualization tool. Even if all they need to show a client are some fairly standard bar and pie charts, the interactivity of the tool is a huge plus. When the chart is up on the

screen and a client asks a question that requires a different view of the data, it is easy to drill into that view on the fly. No more sending an email later in the day with another chart. The data in the charts can also be automatically updated with the latest data. That adds a lot of value on top of the visualizations themselves.

Don't underestimate how much an appropriate visual can help you get your point across. You have to see the power of high-impact visualizations in order to fully grasp what is possible. The good news is that modern visualization tools can help users at any skill level do a better job of analyzing, comprehending, and presenting information. Give it a shot.

FEATURED COMMENT FROM HBR.ORG

“No argument here. ... It's imperative to illustrate the data in a way that people can understand.”

— Phil Simon

THE QUESTION ALL SMART VISUALIZATIONS SHOULD ASK

BY MICHAEL SCHRAGE

“A picture is worth a thousand words” may be a lovely cliché, but it’s exactly the wrong way to view visualization. As admirable as the craft, message, and data-driven artistry of the Edward Tufte and Stephen Fewes may be, successful visualization is less about effectively conveying complex information than it is about creatively provoking human interaction.

Infographics should (quite literally) be seen more as interfaces to interpersonal engagement than aesthetically pleasing packages of numbers and analytics. The essential question smart “visualization” and “visualizers” should address is not “What’s the best and most accessible way of presenting the data?” but “What kinds of conversation and interaction should our visualization evoke?”

Visualizations work best when generating situational awareness and contexts that otherwise wouldn’t exist. Like the best maps and GPS, they simultaneously provide a sense of where you are and insight into where you might want to go next. Individual epiphany defers to interpersonal interaction: Is this where we really are? Given what we see here, do we really want to go there? Visualizations support and enhance teams and teamwork.

So I cannot overstate the power and importance of visualizations as portals. I will never forget an Excel-enabled experience almost 20 years ago when a presenter toggled from histograms to pie charts to cells to macros to raw data tables in a product performance review. Visualizations weren’t static summaries; they were digitally dynamic gateways into more detailed data (and the statistics manipulating them). The visualizations synthesized but not at the price and cost of denying deeper and more granular views.

The resulting conversations were, of course, immeasurably enriched by this approach. Questions and disagreements could be addressed at whatever layer — or layers — of visual representation was most fit for the purpose. The multiple visualizations were integrated, interoperable, and inspiring. They effectively facilitated a collaborative interaction that would otherwise not have been possible.

More recently, visualizations from the Consortium for Advanced Simulation of Light Water Reactors (CASL) had a huge impact on an innovation workshop because they facilitated a cross-functional design conversation that the organization’s existing visualization portfolio could not. Why? Because the firm’s simulations were dedicated to modeling known problems instead of inspiring collabora-

tive interaction. The CASL simulations effectively forced the organization to rethink how visualizations could be used as platforms to create common understanding across the enterprise rather than as high-resolution tools to support technical specialists.

I’ve (literally) seen this visualization influence recapitulated in hundreds of other design and innovation environments. Instead of pretty PowerPoints and charismatic keynotes engineered around the ideal of optimizing information presentation, visualization was treated as a compelling invitation to engage and interact with both the material and each other. In other words, visualization was less about presentation than it was about UX — user experience. Visualization was managed as a UX design challenge rather than as a way to put on a data-driven show.

Is this unduly harsh or cynical? No. Look at Gardiner Morse’s excellent post on “crap circles” or Dylan Lathrop’s post on peak infographics. The information/presentation design bias taints the overwhelming majority of visualization efforts. That’s both a pity and a problem. Until visualizers embrace the design imperative that their visualizations facilitate interaction as much as they convey information, visualizers are doomed to be high-resolution underachievers.

Yes, accessibility, understanding, and insight are the wonderful products of wonderful visualizations. But truly transformative visualizations invite people to touch, stroke, and go deeper into the data that underlie them. They engage. They encourage engagement. They give their users a new way to view each other and the data.

Is a picture worth a thousand words? Sure. Maybe even ten — or a hundred! — thousand. But you want to make sure they’re the right words. Don’t view visualization as a medium that substitutes pictures for words but rather as interfaces to human interactions that create new opportunities for new value creation.

FEATURED COMMENT FROM HBR.ORG

“Well said. Visualizations open new doors for many interpretations. And the success emerges when the data visualizations trigger meaningful discussions and lead to value creation and innovation.” — Anithasrnvsn

WHEN DATA VISUALIZATION WORKS — AND WHEN IT DOESN'T

BY JIM STIKELEATHER

I am uncomfortable with the growing emphasis on big data and its stylist, visualization. Don't get me wrong — I love infographic representations of large data sets. The value of representing information concisely and effectively dates back to Florence Nightingale, who developed a new type of pie chart to clearly show that more soldiers were dying from preventable illnesses than from their wounds. On the other hand, I see beautiful exercises in special effects that show off statistical and technical skills but do not clearly serve an informing purpose. That's what makes me squirm.

Ultimately, data visualization is about communicating an idea that will drive action. Understanding the criteria for information to provide valuable insights and the reasoning behind constructing data visualizations will help you do that with efficiency and impact.

For information to provide valuable insights, it must be interpretable, relevant, and novel. With so much unstructured data today, it is critical that the data being analyzed generate interpretable information. Collecting lots of data without the associated metadata — such as what it is and where, when, how, and by whom it was collected — reduces the opportunity to play with, interpret, and gain insights from the data. It must also be relevant to the persons who are looking to gain insights and to the purpose for which the information is being examined. Finally, it must be original or shed new light on an area. If the information fails any one of these criteria, then no visualization can make it valuable. That means that only a tiny slice of the data we can bring to life visually will actually be worth the effort.

Once we've narrowed the universe of data down to those that satisfy these three requirements, we must also understand the legitimate reasons to construct data visualizations and recognize what factors affect the quality of data visualizations. There are three broad reasons for visualizing data:

- **Confirmation:** If we already have a set of assumptions about how the system we are interested in — for example, a market, customers, or competitors — operates, visualizations can help us check those assumptions. They can also enable us to observe whether the underlying system has deviated from the model we had and assess the risk of the actions we are about to undertake based upon those assumptions. You see this approach in some enterprise dashboards.

- **Education:** There are two forms of education that visualization offers. One is simply reporting: Here is how we measure the underlying system of interest, and here are the values of those measures in some comparative form — for instance, over time or against other systems or models. The other is to develop intuition on and new insights into the behavior of a known system as it evolves and changes over time, so that humans can get an experiential feel of the system in an extremely compressed time frame. You often see this model in the gamification of training and development.
- **Exploration:** When we have large sets of data about a system we are interested in and the goal is to provide optimal human-machine interactions to that data to tease out relationships, processes, models, etc., we can use visualization to help build a model to allow us to predict and better manage the system. The practice of using visual discovery in lieu of statistics is called exploratory data analysis (EDA), and too few businesses make use of it.

Assuming the visualization creator has gotten it all right — a well-defined purpose, the necessary and sufficient amount of data and metadata to make the visualization interpretable, enabling relevant and original insights for the business — what gives us confidence that these insights are now worthy of action? Our ability to understand and, to a degree, control three areas of risk can define the visualization's resulting value to the business:

- **Data quality:** The quality of the underlying data is crucial to the value of visualization. How complete and reliable is it? As with all analytical processes, putting garbage in means getting garbage out.
- **Context:** The point of visualization is to make large amounts of data approachable so we can apply our evolutionarily honed pattern detection computer — i.e., our brain — to draw insights from it. To do so, we need to access all the potential relationships of the data elements. This context is the source of insight. To leave out any contextual information or metadata (or more appropriately, “metacontent”) is to risk hampering our understanding.
- **Biases:** The creator of the visualization may influence the semantics of the visualization and the syntax of the elements of the visualization via color choices, positioning, and visual

tricks (such as unnecessary 3D, or 2D when 3D is more informative) — any of which can challenge the interpretation of the data. This also creates the risk of prespecifying discoverable features and results via the embedded algorithms used by the creator (something EDA is intended to overcome). These, in turn, can significantly influence how viewers understand the visualization and what insight they will gather from it.

Ignoring these requirements and risks can undermine the visualization's purpose and confuse rather than enlighten.

WE'VE REACHED PEAK INFOGRAPHIC, AND WE'RE NO SMARTER FOR IT

BY DYLAN C. LATHROP

If I were to chart the evolution of my attitude toward infographics over time, it would start with a soaring arc, then dip and rise, and finally drop into a steady flat line. My personal interest in data visualizations has waned, but demand is higher than ever — and I've devoted much of my career to propping up that demand. For three years I've distilled data into graphs and illustrations that I'm proud to call my own. But I've also scratched my head at more slickly designed, low-info infographics than I care to count.

Infographics may seem like the design trend of the moment, but they have a long history. In 1920s Europe, Otto Neurath and Gerd Arntz launched The Isotype Project, pioneering the use of simple visual methods for conveying critical data. Neurath, a social scientist, wanted to create an “international picture language” that could explain important political and social issues to people who couldn't read. With designer Arntz he created 4,000 pictograms to communicate information necessary to ordinary lives. Many of their original icons still grace our bus stops and government buildings.

Neurath and Arntz were designing for an audience that couldn't read. Today, many people just don't want to. There's never been more data at our fingertips, but most of us have trouble making sense of that glut of information unless it's shaped into cohesive nuggets. Enter the modern infographic, which has moved away from the elegant simplicity of the Isotype icons in favor of communicating entire data sets in one smartly designed package.

Our data is growing more complicated just as readers are getting less patient. Even the best illustration can't bridge the comprehension gap. Sometimes good design even enables mental shortcuts. Glance at an infographic and you can feel like you've processed massive amounts of information. Share, tweet, or like the infographic online and your friends can see just how much you've learned without even looking themselves. Designers (and the outlets that commission them) have figured out that they can emphasize pretty and clicky over useful and interesting, which means Neurath and Arntz's work has taken on an exploitative new dimension.

Ever peered deep into the belly of an infographic and wondered, “Is there anything connecting these numbers other than this illustration?” You're just beginning to see the Matrix, Neo. The model of simplicity executed by Neurath and Arntz has mutated to become more about design than data, creating a glut of attractive imagery

with little reliable or even compelling information. Isotype worked because it collected relevant civic data and distributed it using functional design. Today, the goal is to make infographics viral, regardless of the stories they tell. This practice has broadened the definition of what we refer to as an “infographic.”

I've been part of the problem. Once, I was asked to create an infographic using some weak data on marijuana. I pushed through a design that amounted to three weed-joke illustrations, no charts, and almost no numbers. Forgive me, Otto, for I have sinned. May this joint light the way to my redemption.

It's time we acknowledge the shortcomings of infographics as much as we celebrate their upsides. That's harder to do now that infographics have exploded as a cottage industry within design. The boom began in 2006 with the launch of *GOOD* magazine, my former employer, which emphasized infographics from the start and was named a National Magazine Award finalist for its work in 2008. *Fast Company* and *Wired* entered the game. Today, entire design studios are dedicated to creating these illustrative data dumps.

Since leaving my job earlier this year, I've had the opportunity to create infographics on a freelance basis for several publications. When I speak to potential clients, I often realize that they're more interested in churning out the form than in telling a story. An idea without supporting data, a list, a business plan, a resume — none of these are infographics, no matter what they're labeled. And even if the data exists, just because you can create an infographic doesn't mean you should.

Worse yet, private businesses are increasingly controlling the form. Companies looking to illustrate their capabilities and showcase their potential collaborate with a “media producer” to present the facts as the company sees them. There's nothing wrong with a business leveraging visual vocabulary to tell its story. But potential customers tend to view an infographic less skeptically than they do an ad.

These visual interlopers haven't undercut the form entirely. There are still honest-to-goodness data visualizers out there, like Bloomberg *BusinessWeek's* Jennifer Daniel and Facebook's Nicholas Felton, and other talented artists pushing infographics in new directions, like *The New York Times* contributor Andrew Kuo. But more and more infographics are starting with questions like “Will this blow up on the Internet?” or demands such as “I want an info-

graphic! About what? Whatever!” And these so-called information graphics threaten to undermine even the most shining examples.

Infographics can evolve by transcending cold data breakdown and combining data visualization with more human narratives. Some publications have begun to present well-designed information in tandem with deeply reported pieces online, and the future it represents is thrilling. I’m not ready for an infographic about the death of infographics, but I’m sure someone somewhere has already assigned that piece and is just waiting for us all to click.

This article first appeared in TOMORROW magazine and was reprinted with the author’s permission.

IT'S TIME TO RETIRE 'CRAP CIRCLES'

BY GARDINER MORSE

Every time I encounter a crap circle my heart sinks. I first wrote about these contemptible “information” graphics in HBR in 2005, and since then they’ve only seemed to multiply. You know what these are — you may have even used them — though you may not have had a name for them. I aim to change that. These pernicious circles-and-arrows diagrams infest PowerPoint and other business presentations, purporting to clarify an idea while actually obscuring it.



As I wrote back then, when you find yourself about to drop a crap circle into your slide deck, stop. And the next time a presenter trots out a circle to make a point, call him on it.

Here’s the original article. I urge you to forward it to violators, and submit examples in the comments below of the worst (or best?) crap circles you’ve encountered.

The most dubious business plan can look solid — even smart — if it’s cast as a virtuous circle: “See, we invest our profits in innovation to create delightful products that customers buy — which generate profits that we invest in innovation!” Who could argue with that? Indeed, the merit of self-reinforcing systems seems so obvious that businesspeople instinctively describe their strategies as cyclical activities that magically fuel themselves. Meanwhile, audiences demand snappy-looking, easy-to-digest graphics that, almost by definition, strip away nuance. It’s no surprise, then, that business communications are lousy with circle-and-arrow diagrams that range from the dumb to the deceptive.

Though you’ve seen a million of these, you’ve probably never thought much about them. That’s because, like optical illusions, they play on your expectations and trick you into seeing something that isn’t there: If one arrow leads to the next, then *of course* the

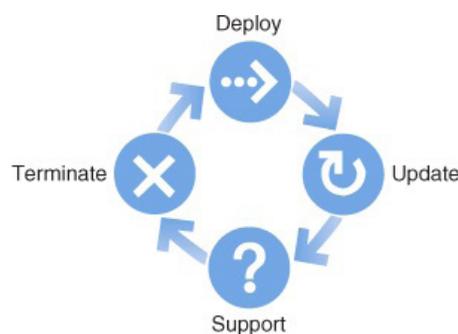
steps follow. But once you start examining these ubiquitous diagrams, you’ll be amazed by what you don’t see.

Consider these examples:

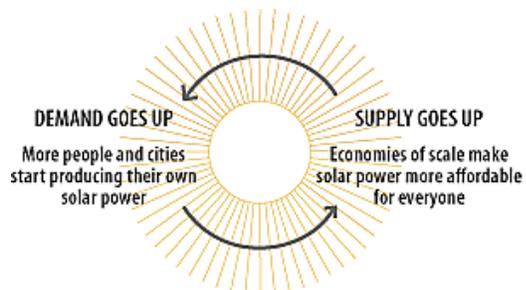
The circle below, from a global accounting firm’s website, is used to illustrate the company’s consulting services for owner-managed businesses. It shows the business life cycle “Maturity Phase” leading, inexplicably, into the “Conception/Start-Up Phase.” This company’s clients should ask whether they really want to be guided in circles.



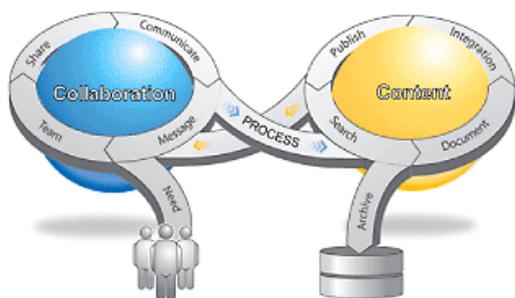
With the next design, a Boston-based software company helpfully illustrates the stages of its application management life cycle. Through some trick of causality, termination leads to deployment. This may be a good model from a consultancy’s standpoint — when a client’s projects end, they start again — but if you’re paying the tab, you probably want the project to actually end when it’s terminated.



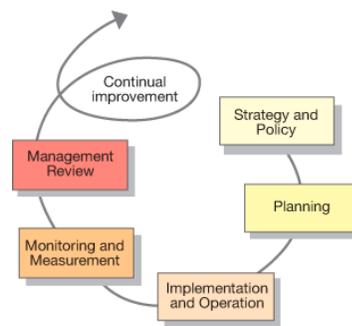
The friendly-looking sunburst that follows, captured from the website of a solar energy advocacy group, shows how to create an unlimited market for your product. Here, as the supply of solar energy increases, so does the demand — in an apparently endless cycle. If these folks are right, we’re all in the wrong business.



And this one, from a Canadian enterprise content management company, is notable for its sleight of hand. Circles rotating in opposite directions (in which, among other oddities, “publish” gives rise to “search”) link through arrows whose origins and destinations, on close inspection, are obscure or completely hidden. Maybe the intent of this diagram is to make prospects too dizzy to ask questions.



Kudos, though, to the author of the disarmingly honest graphic below, from a U.S. safety engineers group — a refreshing bit of out-of-the-circle thinking. He seems to have had an epiphany as he created the diagram, realizing that the development of safety processes doesn’t always chase its tail — that “management review” needn’t slavishly feed into “strategy and policy” in the service of “continual improvement.”



By fighting the impulse to think in circles, he’s set an example for everyone who has uncritically accepted or, worse, actually constructed a crap circle — and that’s most of us. The next time you find yourself preparing a circle for a presentation, ask yourself whether the process you’re describing really works the way you say it does. And the next time a presenter trots out a circle to make a point, find the bogus links and put him on the spot. We could all benefit from a little more linear thinking.

WHAT MOLESKINE'S MARKET POSITION REALLY LOOKS LIKE

BY DAN MCGINN

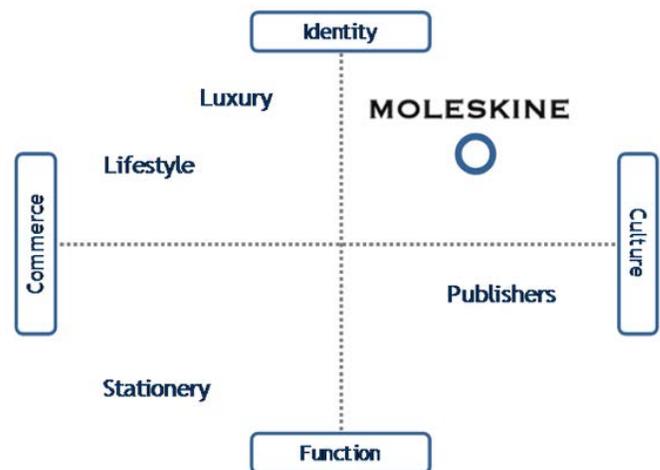
The editors managing the Visualizing Data Insight Center have asked us to submit a visualization that we find compelling or effective. Mine won't win any design awards, but I was so struck by it that I passed it along.

A *Quartz* article on the IPO of hipster journal company Moleskine included this chart that comes from the company's prospectus.

Now, if I had read a paragraph of text Moleskine marketers had written that said "Moleskine is not a stationery company — in fact, it's the opposite of a stationery company," I likely would have rolled my eyes and glossed over the rest of the positioning.

This simple 2x2 chart, though, conveys that idea better than any text ever could, without any marketing jargon, and in a way that leaves an impression with me. There is some craft here as well, despite the artless design. For example, the two axes could be positioned any other way, with Identity on the right, bottom, or left, but by positioning the axes as they have, Moleskine ends up by itself in the top right — the classic position for success on charts.

I'm not saying I'm going to run out and buy shares, but this example struck me immediately when I saw it.



WHEN PRESENTING YOUR DATA, GET TO THE POINT FAST

BY NANCY DUARTE

Projecting your data on slides puts you at an immediate disadvantage: When you're giving a presentation, people can't pull the numbers in for a closer look or take as much time to examine them as they can with a report or white paper. That's why you need to direct their attention. What do you want people to get from your data? What's the message you want them to take away?

Data slides aren't really about the data. They're about the *meaning* of the data. And it's up to you to make that meaning clear before you click away. Otherwise the audience won't process — let alone buy — your argument.

Take this table, for instance:

European Sales

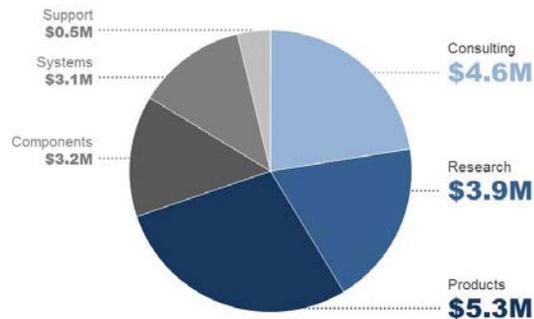
GEO and Desc	Consulting	Research	Products	Components	Systems	Support	TOTAL
Grand Total	4,634,068.91	3,883,621.28	5,843,307.44	2,890,064.81	2,531,168.87	831,178.79	20,395,627.67
UK	1,812,064.34	1,659,130.10	1,677,140.18	1,734,038.89	1,593,727.52	287,646.49	8,763,747.52
Germany	385,195.50	145,927.68	737,681.62	520,211.67	129	70,969.04	1,860,114.51
France	53,743.00	360,629.35	296,678.27	57,801.30	261,631.24	41,861.04	1,072,342.20
Italy	9,730.00	62,204.26	287,966.62	28,900.65	0	37,208.19	426,009.72
Spain	30,246.48	935,909.79	160,940.14	86,701.94	0	29,158.87	1,242,957.22
Portugal	0.00	124,531.50	115,279.71	31,790.71	0.00	11,678.77	283,280.69
Switzerland	1,194,630.10	209,724.09	302,458.78	31,790.71	278,337.21	53,392.28	2,070,333.17
Netherlands	317,484.50	85,999.93	320,678.71	34,680.78	0.00	44,457.28	803,301.20
Western Europe Other	119,394.74	284,601.35	1,281,434.33	28,900.65	397,085.90	199,308.40	2,310,725.37
Russia	94,053.01	6,532.05	211,934.18	14,450.32	0.00	8,548.41	335,517.97
Poland	325,191.36	8,431.18	237,751.17	21,675.49	0.00	7,490.30	600,539.50
Eastern Europe Other	292,335.88	62,453.86	213,365.73	18,785.42	258.00	39,459.72	626,658.61

It's confusing — especially if you project it for five seconds and then move on. And even if you leave it up for five minutes while you talk, anyone who's struggling to derive meaning from it won't be paying much attention to what you have to say. They'll be too busy squinting from their seats, trying to navigate all those heavy grid lines that give every single cell equal weight. It's not at all clear where the eye should go. Your audience won't know what direction to read — horizontally or vertically — or what conclusions to draw. Though the grand total line is emphasized, is that really the main point you want to convey?

Now let's look at the data presented more simply. Say you've identified three business units with potential for sustained growth in Europe. By eliminating the dense matrix and connecting only key

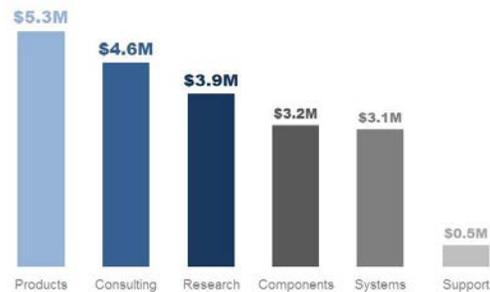
numbers to a pie with leader lines, you remove clutter that distracts from your message. And notice the clear hierarchy of information: You can highlight important pieces of the pie by rendering them in color and their corresponding annotations in large, blue type. Other sections recede to the background, where they belong, with their neutral shades and small, gray labels.

Total Europe Sales



But pie charts can be tricky for an audience to process when segments are similar in size — it's hard to distinguish between them at a glance. If you're running into that problem, consider displaying the same data in a linear way. In this bar chart, for example, you draw attention to the poorest-performing unit, a point that got lost in the pie:

Total Europe Sales



These few tricks will help audiences see what you *want* them to see in your data. By focusing their attention on the message behind the numbers, not on the numbers themselves, you can create presentations that resonate with them and compel them to act.

WFEATURED COMMENT FROM HBR.ORG

“The recommendations ... are well worth considering and no doubt add value to the presentation. Any information that the presenter consider of significant importance should be provided in another form as significant takeaways to the discussion. Great post.”

— Tim

TEN YEARS OF NEWS CORP. INCOME DATA IN LESS THAN A MINUTE

BY GRETCHEN GAVETT

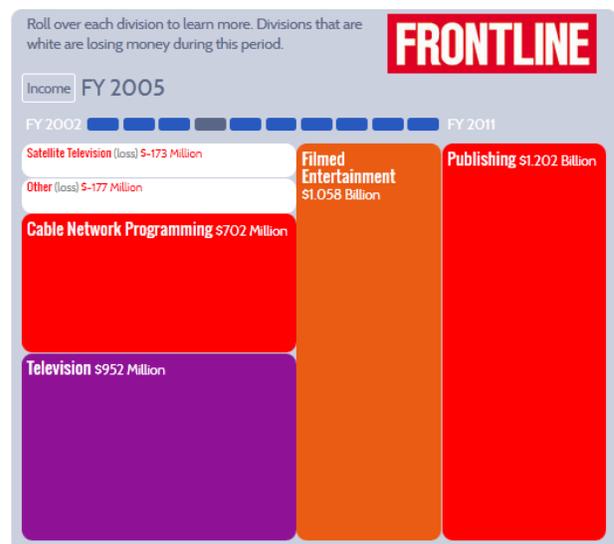
We're posting some of our favorite visualizations as part of this month's Insight Center on the topic. My colleague Dan McGinn shared his, on Moleskine, last week. Here's mine:

At my previous job with the PBS series *Frontline*, my colleague Sam Bailey and I wanted to answer a couple of questions about Rupert Murdoch's media empire: How, exactly, does News Corp. make its money? And has this changed over time? The answers, we were hoping, might help our readers better understand how Murdoch's beloved, hack-riddled broadsheets fit into the rest of his organization.

So Sam pulled together some pretty dry-looking data from a decade of News Corp. annual reports to create this animated treemap. It tells the story of the company's financial priorities and evolution better than a few hundred words or pages of tables ever could (click on the rectangles between FY 2002 and 2011 for speedier viewing):

It's visual evidence that, indeed, the massive organization had become "a sports and entertainment company with a newspaper problem." Whether or not this is a good business strategy wasn't necessarily up to us to decide; however, it laid the pure economics of it out in an easy-to-grasp story. What's more, the simple animated transitions helped bring to life year-over-year fluctuations of the various groups' performance in a way that static snapshots of the data could not.

Building this was, for us, a journalistic endeavor, but this doesn't mean you can't use a treemap to analyze your own financial or digital trends over time. Not only can it call attention to what you might otherwise have missed, but it's also a powerful narrative device for presenting information to others.



Source: News Corporation annual reports, 2002 - 2011. Chart © 2012 WGBH Educational Foundation

DATA FOR ALL! HOW NEW TOOLS DEMOCRATIZE VISUALIZATION

BY BILL FRANKS

When I talk to organizations about how they are using data visualization tools, I am often struck by the fact that they use these tools mostly to generate charts and graphs that really aren't all that different from what they could have created with standard business intelligence or desktop tools. People get very excited about their output nonetheless. At first this surprised me, but then I realized what was going on.

I would like to suggest that data visualization tools such as Tableau and Spotfire, to name just a couple, offer two great value propositions — that are often intertwined into a single value proposition. Understand the two propositions and you'll create more value for your organization.

The first value proposition is the obvious one, which is enabling users to create better visuals that bring their data and their analysis to life. This is the value proposition that most people focus on and that gets the most attention. It is also the primary reason organizations invest in visualization tools.

The second value proposition, which is often either overlooked or vastly under-credited, is that visualization tools democratize big data by giving users wide flexibility to analyze data within a self-service business intelligence environment. Visualization tools allow users to explore, summarize, and visualize data in the way they see fit as opposed to the way someone else saw fit to allow them. By having the flexibility to join different data sources as desired, view patterns on the fly, and iterate, users can discover important insights and trends more easily and more rapidly.

Users may be able to access massive data sources in traditional environments, but they can do so only via predefined paths. On the other hand, common desktop tools such as PowerPoint or Excel that enable charting and graphing either require data extracts, which must be small, or more complex configurations than many are comfortable with. They are too complex and the visuals they generate aren't very robust or interactive.

While many users of the new visualization tools spend most of their time generating basic output, they get really excited about their newfound freedom to navigate the data and view it from any angle desired. While the graphics generated may be simple, users are much more confident that these graphics contain the right content.

The implication is that many organizations may not be getting the full benefit of their big data and visualization investments. But it'd be a mistake to make those tools available only to those users with advanced data skills. Using the tools should help even non-numerate users gain greater comfort with the data (one hopes, at least), and along with that comes growing the ability to draw increasingly sophisticated insights. And that's when the big data investments really start to pay off.

FEATURED COMMENT FROM HBR.ORG

[“The post you have written on data visualization is quite compelling. Visualizations help people see things that were not obvious to them before.”](#) — Will Waugh

VISUALIZATION AS PROCESS, NOT OUTPUT

BY JER THORP

“Please make me a visualization.”

I get a lot of emails that say this or some variation of it. They tend to make me think of other requests that could be made in the same form, like:

“Please make me a roast beef sandwich.”

“Please make me a scale model of the Eiffel Tower.”

Roast beef sandwiches and scale models of the Eiffel Tower, in these sentences, are common nouns. Visualization, on the other hand, is a verbal noun. The word “visualization” encapsulates a process. And it’s really that process, not the thing that results, that’s the essential part. A much more exciting email — one that, sadly, I receive much more rarely — uses the verb. Something along the lines of:

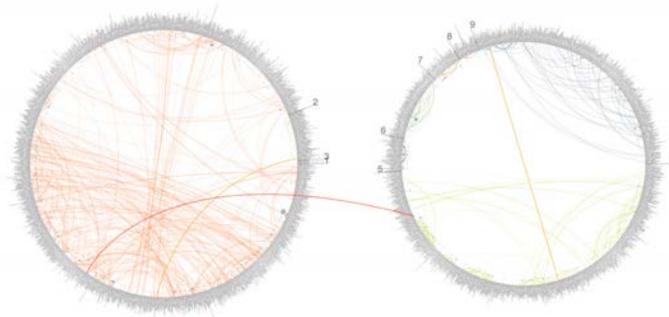
“Please visualize our data.”

The nice thing about this sentence is that it may result in many things. When I set out to make a roast beef sandwich, I almost always end up with ... a roast beef sandwich. If we set out to visualize instead of make a visualization, we can end up with any number of outcomes. In fact, many of those outcomes may not even be visualizations but rather solutions, new ideas, and better questions. Any good visualization process is iterative. And if we allow ourselves to think more about the value of the branching points of that process than about a single result, we leave ourselves open to many more possibilities. A verb-based approach to visualization also lets us think of it as a tool that can be used in many different projects, not only those whose results involve charts and graphs or sticks and balls.

In 2009 I was asked by Jake Barton to design an algorithm that would place the names of nearly 3,000 victims in specific places on the 9/11 memorial in Manhattan so that certain names could be near each other, per the wishes of their next of kin. This was a novel and difficult challenge, given the number of possible combinations and the complexity of the personalization required. When I started, I wasn’t even sure that it could be done.

The first step, then, was to get some idea of the scale and peculiarities of this particular problem. The data that I was sent was in a spreadsheet. Here’s something that I’ve learned about spreadsheets: No amount of staring at one is going to teach you about the data. In order to get a sense of the data and therefore the problem,

I created a visualization. This simple graphic, hand-rolled in a tool called Processing, shows the pools of the memorial as circles, with each name arranged on the edge of a ring. The lines between those names are the requested adjacencies — names that should, as per the wishes of family members, be placed together:



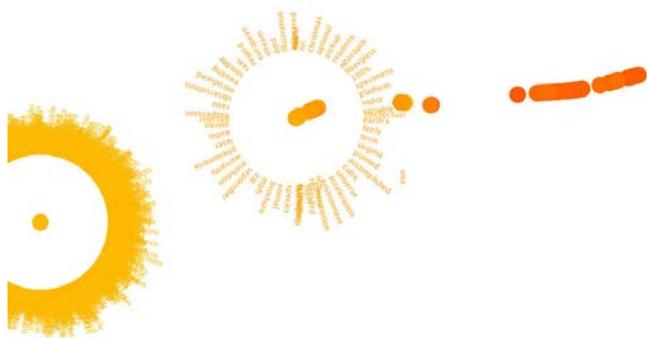
A visualization of the victims’ names and requested adjacencies for the 9/11 memorial, fall 2009.

I could have read the number of names and the number of requested connections from the spreadsheet. However, the key part of the problem ended up being in the physical distribution of those connections, which showed up only after sketching. This quick visualization also showed me that the connections were not evenly balanced between the two pools; indeed, they were heavily concentrated in one of the two pools. By building a bespoke visualization quickly, I put the data into a visual form that fit its structure specifically and got to the core of the problem.

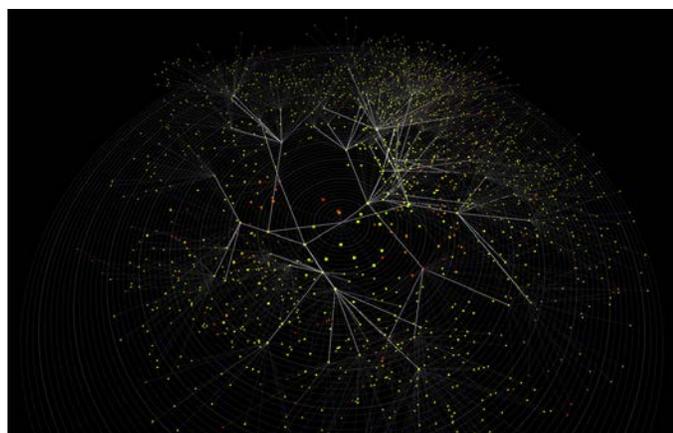
Getting this insight into the character of the data quickly changed my sense that developing an algorithm was impossible. I could now see that it seemed possible.

I think of these small visualization steps as “sketch points.” I don’t have to put too much thought into their aesthetics because they aren’t built for public consumption. I make my sketch points in some kind of expressive medium (like Processing) as opposed to a quicker but more constrained tool (like Excel or Tableau) in order to tailor them to the specifics of the data as closely as possible. In this fashion these stops along the way become low-investment testing grounds for new ideas and unusual approaches.

Here are a few sketch points from recent projects, each of which represented a turningpoint in my thinking:



A sketch point from a 2011 visualization of 138 years of *Popular Science*.



A sketch point from the development of Cascade at *The New York Times* R&D lab, 2009.



A sketch point from the development of a new visualization of ad placement networks, 2013.

None of these is meant for public consumption. You're looking at my efforts to work out a problem, see what I'm up against, and find in the sketches potential ways forward. By thinking about visualization as a process instead of an outcome, we arm ourselves with an incredibly powerful thinking tool. By splitting this process into small, bespoke sketch points, we can engage with the character of our data more specifically and access a broader and more varied solution space. Data visualization becomes much more than just the end of a sentence.

FEATURED COMMENT FROM HBR.ORG

“Thank you so much for sharing and encouraging right-brain thinking in the ‘process’ of presenting data.”

— Angela222

TELLING STORIES WITH VISUAL DATA: A GLIMPSE INTO THE FUTURE OF NARRATIVE

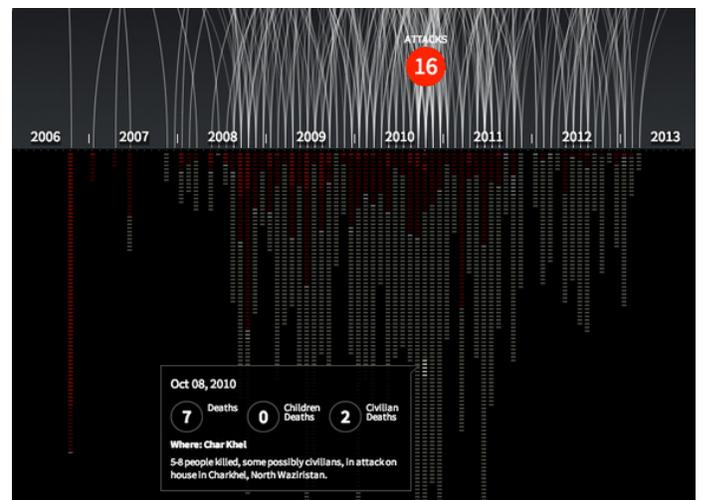
BY JER THORP

Editor's note: We've asked contributors to the Visualizing Data Insight Center to show us some of their favorite examples of data viz, with short explanations of what makes those visualizations so effective. Today, Jer Thorp shares one of his favorites.

Below is a screen grab of a masterful interactive data visualization. This narrative-driven piece by Pitch Interactive manages the extra-tricky task of balancing heavy subject matter with a clear story and compelling visuals. It's a glimpse into the future of data-driven storytelling. Perhaps the most interesting thing about the piece is that it wasn't commissioned by a media organization. It was built by Pitch as a way to explore and understand this complex topic. Bravo.

The full data visualization is here — it's worth watching and scrolling over for a more in-depth view.

View more data visualization examples and best practices in our monthlong series on data visualization.



WHEN CREATING VISUALIZATIONS, QUESTION EVERYTHING

BY IRENE ROS AND ADAM HYLAND

As organizations expand their digital infrastructures, the amount of data being collected is growing at an ever-increasing pace. This deluge presents a new set of challenges, and the most crucial of these — making sense of it all — depends on data visualization. Digital tools like Many Eyes and Tableau software have empowered companies and the public to create visualizations using built-in templates, but they have also spurred a desire for more control over visual method, layout, style, and branding.

The Open Web, a platform for building web-scale software using open technologies, is changing the landscape of data visualization. Open source libraries such as Raphael JavaScript and D3 JavaScript democratize data visualization, making the process and code transparent. Additionally, the communities supporting these tools are vibrant.

With these new opportunities, however, comes a need to unravel the production pipeline from messy and complex data sources to the polished results we are all used to. It's not as easy as it sounds. Deconstructing what actually has to take place in order for data visualization to happen is something we've been refining here at Bocoup.

Here's our formula:

Planning for data visualization starts with a conversation. When our clients initially state their goals — “We want to create data visualization of X” — we tend to ask “Why?” repeatedly, in a user-centric way. By asking questions, we identify the audience and our goals as creators, which, in turn, informs the rest of the process. Answering these questions is not just important to moving forward — it often transforms the project entirely.

The next thing to remember is that it's common to expend considerable effort transforming the data into machine-consumable forms, cleaning out any rogue data points, and reformatting it into shapes (like tables or JSON formats) that lend themselves to computation. This is a process that anyone will tell you takes longer than expected, so don't get too frustrated. It's worth the effort.

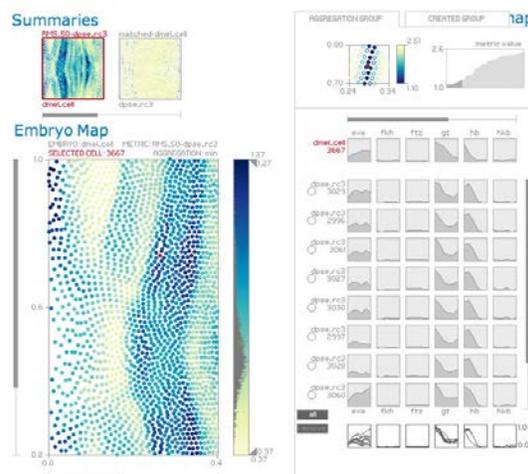
Next, data discovery focuses on understanding the parameters of the data, because the data alone can't offer information on the underlying relationships. What outliers exist in the data? Are they meaningful or simply adding noise? For example, when exploring subprime lending, economists from Stanford uncovered what they thought was an error in their data: a large, unaccountable surge in loans in the early months of the year. Upon further investigation, they discovered the anomalous rise was due to the effect of the

Earned Income Tax Credit, dramatically changing the direction of their research.

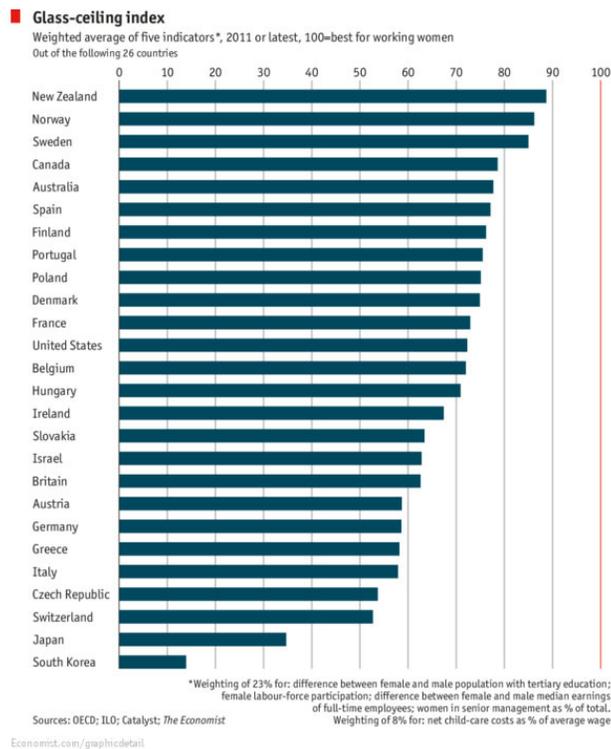
When the data's forms, shapes, and curves become second nature, the really fun part begins: asking it more questions. It is tempting to start a data visualization project with an idea for a striking visual artifact illustrating a conclusion and then work backward. Much like our initial conversation to discuss goals, clients will often say, “We know that X is true, and we want a data visualization showing that.” But is it really true? Even if it is, is that really the right fact to communicate about the data?

More often than not, challenging some of these assumptions can have a profound impact — not just from the perspective of the immediate project but also for the organization as a whole. **The data, then, exists for two purposes: verifying what we already know and exposing us to what we don't.** Having the time and space to explore both of these goals in a phase of analysis is crucial to preparing the data to be visualized. When we truly find the data to support what we're trying to communicate, the visual forms will emerge naturally.

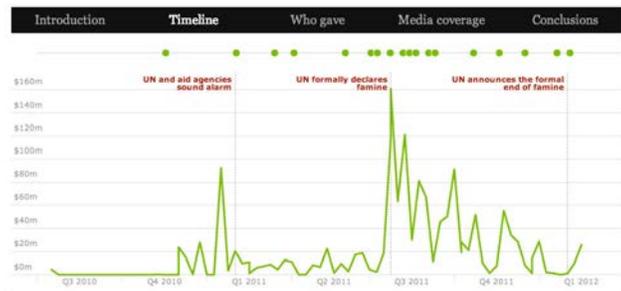
Now — yes now — we start actually visualizing the data. It's time to think about the audience. What are its characteristics? What decisions are readers or constituents expected to make based on the visualization? For highly engaged audiences with their own questions, an exploratory tool is best, like this one from the University of Utah:



For a busy audience, a high-level, static graphic may be most appropriate, such as this one from *The Economist*:



If the data is complex, an interactive step-through interface focusing on a single narrative — like this feature from *The Guardian* — may best serve your readers:



Many of these approaches can be combined in the right circumstances.

When the delivery format is understood and the constraints of the audience and their expectations clear, it's time to play around with visual designs. **Experimentation is crucial.** Not all visual designs are created equal; depending on the relationships in the data, some may be more appropriate than others. Targeting a certain format and screen size introduces a different set of constraints and possibilities. The best solution is often the simplest one, and rapid iteration and prototyping are the only ways to ensure that we are representing the data and narrative accurately.

In practice, this pipeline feeds itself. As we understand our data, it brings up new and better questions for us to answer. Budgeting time for this exploration is necessary in order to create a valuable resource that will help you communicate with your audience. The greatest mistake we see with data visualization is taking the easy way out — oversimplification of the data (such as averaging it out) can hide intricacies and important patterns, but visualizing all the data often results in “hairballs,” or distributions with no patterns.

In the end, the data visualization workflow requires you to be surprised. When working with data, you can expect to be wrong, to fail fast, and to fail often. More than any other engineering practice, data visualization requires an iterative approach to account for the changing nature of your findings as you work.

This approach is strongly supported by modern data visualization practices on the Open Web. The web offers us tools for rapid prototyping, instant support and feedback, development practices that can evolve fast and grow with the changing ecosystem, and powerful tools for interactive software development.

Working on the Open Web does not eliminate the complexity of the underlying problems we're trying to solve, but it does offer users and new developers the ability to read, tinker with, and share code because of the collaborative nature of the ecosystem. **You're never alone.** Regardless of whether you're a startup or an established organization, the Open Web can prove invaluable in making sense of your data and presenting it well.

FEATURED COMMENT FROM HBR.ORG

“Always asking questions is such a vital part of getting to a worthwhile place.” — Chicago Style SEO

VISUALIZING HIP HOP LYRICS AS CULTURAL INDICATOR

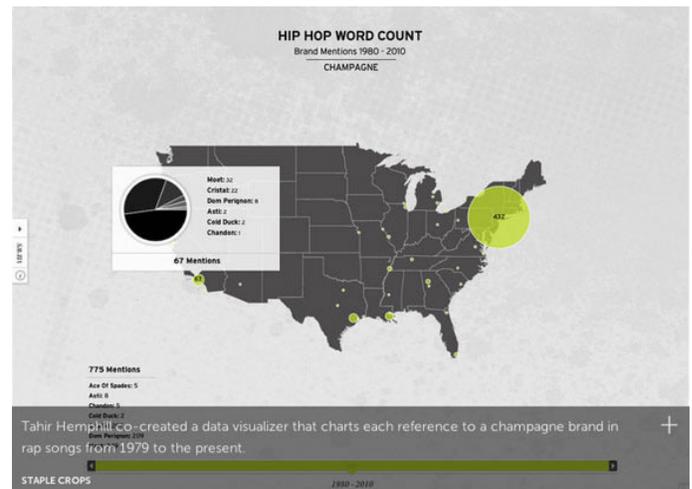
BY JEFF KEHOE

Editor's note: We've asked contributors to the Visualizing Data Insight Center to show us some of their favorite examples of data viz, with short explanations of what makes those visualizations so effective. Today, HBR senior editor Jeff Kehoe shares one of his favorites.

People often think of data visualization as an alternative to representation in words. But there are creative, effective visualizations that have been developed that vividly illustrate how — and how often — words themselves are used by people in various contexts. Note the recent rise in popularity of word cluster tools.

These kinds of tools can go even deeper to highlight interesting and powerful cultural indicators. I'm a hip hop fan, so this particular data viz caught my attention last week. It charts brand references to Champagne in rap songs from 1979 to the present (see the full visualization at Fast Company):

The idea of tracking changing cultural references among artists, or the changing vocab styles among groups or individual artists, turns me on. As more and more of the world's textual knowledge recorded in books and other printed material gets digitized, we'll be able to visualize and explore more about ourselves. The possibilities are pretty mind-boggling.



TELL BETTER DATA STORIES WITH MOTION AND INTERACTIVITY

BY ANDREW DEVIGAL

When it comes to making sense of vast amounts of complicated data, time really is on your side. It's a simple concept, one that everyone understands: An action starts and then eventually stops. The distance between those two points conveys information — information about *then*, about *now*, and about the differences between the two.

If you apply that simple yet elegant measuring stick to an overwhelming glut of information, you have the beginnings of a powerful data visualization that can simplify the complex, identify trends, and shape your audience's comprehension of the story you want to tell.

However, when time is the canvas for your data, you'll need one or both of these techniques: motion and interactivity.

Hans Rosling, who gained popular fame in his 2006 TED Talk on stats that reshape your worldview, uses the power of motion in the software that runs his Gapminder trend-finding operation.

http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

As he told *Discover*, “[W]hen you show time as graphic movement, as animation, people suddenly understand.” BBC's *The Joy of Stats* series added a bit of special effects to a condensed version of the lecture in 2010: “Hans Rosling's 200 Countries, 200 Years, 4 Minutes.”

http://www.youtube.com/watch?feature=player_embedded&v=jbkSRLYSojo

These videos exemplify the effectiveness of movement in explaining the complexity in health, poverty, and economic trends. And by pacing through the years, the audience is able to consume the data as a visual narrative one frame at a time.

There are also boundless opportunities to illustrate data from a series of events over time in tandem with geographic overlays. Take “1945-1998” by Isao Hashimoto, which traces the growing number of nuclear explosions around the world by marking each blast with a visual and audio blip. In Hashimoto's visualization, 1962 was a dramatic year, with over 175 nuclear explosions. In contrast, the video's activity during the previous year carries implied meaning — fewer blips, fewer nuclear blasts.

<http://blip.tv/ctbtomedia/1945-1998-by-isao-hashimoto-japan-2003-1671472>

Visualizations that span seconds instead of decades can also sift through an avalanche of information. One of my favorite interactives from *The New York Times* is “Fractions of a Second: An Olympic Musical” by Amanda Cox, who was interviewed for this series. As with the previous example, silence is golden. Listening to the margins between skiers' finish times provides extremely effective insights.

Blending time and motion can also help the audience feel emotionally closer to the story. Take the admittedly low-info — yet amusing and thirst-inducing — graphic “The American Beer Revival” by Nate Whitson. Nate used some simple (read *approachable*) visual techniques to tell the story of U.S. breweries in the 100 years before Prohibition.

<http://vimeo.com/hilittleguy/americanbeerrevival>

The examples mentioned so far have mainly been published in linear form, with the ability to hit “Play” and/or “Pause” to step through the information. Adding a layer of interactivity allows your audience to filter through volumes of data and creates a second axis to compare information across time. Hans Rosling's Trendalyzer software, which he uses in Gapminder's analysis of “Wealth & Health of Nations,” empowers the user to highlight geographic regions and even isolate specific countries.

Interactive graphics encourage people to lean forward and participate in the storytelling. By adding layers of information and the mechanics to view the data in varying perspectives, you're essentially allowing your audience to fill in and add their own narratives — their second stories, if you will.

This purposeful interaction reminds me of a story Internet artist Jonathan Harris tells about his fourth-grade teacher, Baz. In the story he describes the difference between the “wow” when you lean back stunned and the “wow” when you lean in to participate. The latter moments linger in your memory. Similarly, interactive graphics are an invitation to participate and an opportunity to have the information stick with you.

Tread lightly, however, when adding interactivity to your chart. You may very well be adding design and interface hurdles that you'll need to compensate for, increasing the complexity of comprehending your data.

This leads me to my three takeaway points when presenting data visualization using motion and time:

Pace through the data. By layering information over time, you isolate specific data sets to minimize the odds that you overwhelm your audience. Comparing numbers over time can also help identify patterns and highlight trends. Animating these patterns helps bridge the comprehension gap between two sets of data.

Minimize the number of canvases. By keeping to one canvas throughout the video, your viewer will be able to understand the full frame of reference. You can zoom in and pan across the overall chart to highlight or tease out nuggets of information. For example, *The New York Times* classic “One Race, Every Medalist Ever” relies solely on camera moves throughout the entire video. The objects and data remain static. With intentional angles and views, this single canvas afforded multiple perspectives to visualize the data and tell the historical story of the event.

Add motion purposefully. Do not add animation for animation’s sake; make sure that you’re using it to convey meaning. Avoid using motion as a type and motion study, as in many college animation assignments. Use motion to show growth, demonstrate a shift over time, or emphasize a piece of data.

These points will give your visualized data more impact only if they work in concert with the main goals: Simplify the complex and ensure that the data provide insight.

THE SCIENCE OF WHAT WE DO (AND DON'T) KNOW ABOUT DATA VISUALIZATION

BY ROBERT KOSARA

Visualization is easy, right? After all, it's just some colorful shapes and a few text labels. But things are more complex than they seem, largely due to the ways we see and digest charts, graphs, and other data-driven images. While scientifically backed studies do exist, there are actually many things we don't know about how and why visualization works. To help you make better decisions when visualizing your data, here's a brief tour of the research.

The Early Years of Understanding Data

While the early days of visualization go back over 200 years, actual research to understand how it works really only started in the 1960s. Jacques Bertin's *Sémiologie Graphique* (*Semiology of Graphics*), published in 1969, was the first systematic treatment of the different ways graphical representations encode data. Bertin coined many terms of the trade, such as the *mark*, which is the basic unit of every visualization, like a bar, line, or circle sector. He also defined a number of *retinal variables*, which are the visual properties we use to express the data. These include color, size, location, etc.

In the early 1980s, Bertin's work was picked up by researchers in statistical graphics and the nascent field of visualization (which didn't quite have its name yet). William Cleveland and Robert McGill performed experiments to find out which of Bertin's retinal variables were best suited for particular types of data, while Jock Mackinlay built a system that put Bertin's and Cleveland and McGill's work to use in order to create visualizations from data.

Thanks to Cleveland and McGill we know that our perception is the most precise when it comes to understanding the location of a mark, followed closely by our ability to perceive length. We're even less adept at perceiving area and orientation, and our ability to distinguish colors is even worse. We can see tiny differences in direction between lines that are almost but not exactly parallel, but we have a hard time quantifying an angle to say what percentage it represents in a pie chart. We can tell fewer than a dozen colors apart when their hues are very distinct and can precisely compare shades of colors next to each other, but move them apart and surround them with very different ones, and it all goes out the window.

This may all seem interesting, but its practical uses are not obvious. To turn the theory into practice, Mackinlay built a system that

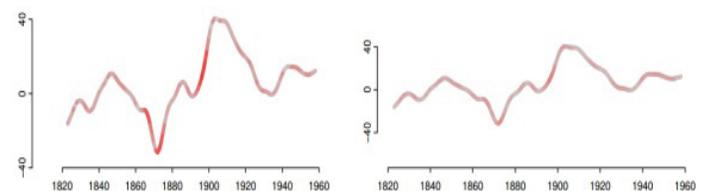
assigned data fields to visual variables automatically in a way that optimized readability. Most visualization tools today still don't offer that kind of intelligence, though Tableau's Show Me feature is built on a very similar idea.

More Knowledge, More Questions

A lot has happened since the 1980s, but we seem to be at a bit of a standstill when it comes to understanding the basics. There are many open questions today, and we also realize the gaps and problems with some of the work performed.

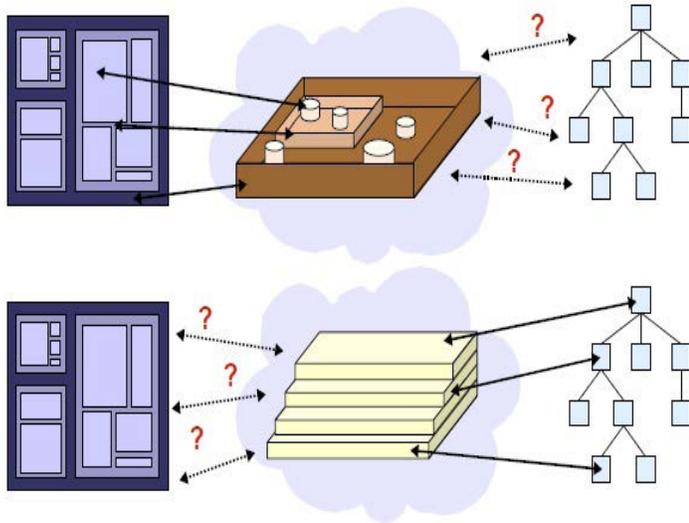
As a case in point, Cleveland promoted an idea that he called banking to 45 degrees. The idea is simple: In a line chart, the average slope should be 45 degrees. That makes intuitive sense, because very steep charts tend to look overly dramatic and very flat ones make it hard to see any change in the data at all. Cleveland's recommendation was based on research on how well we are able to compare the slopes of lines. He found that the highest accuracy was achieved when the lines being compared had an average of 45 degrees inclination.

But as it turns out, that is not the entire truth. There were some limitations in Cleveland's study that made 45 degrees look like the best option, but it seems that shallower angles are actually better. This was shown in a research paper that Justin Talbot, John Gerth, and Pat Hanrahan presented in October 2012 at the annual VisWeek conference. The left line graph below is closer to 45 degrees on average, but the right one, while shallower, has fewer areas that produce large errors (which are indicated by the dark red color).



There is more. My former student Caroline Ziemkiewicz and I found that there is a potential interaction between the visual metaphor used to show data and the linguistic metaphor used to ask a

question. We found this when looking at visualizations of trees, or hierarchies. The two most popular visualization techniques for this type of data, treemaps and node-link diagrams, differ in the way they show the hierarchy. Node-link diagrams use levels (or “above-ness”), while treemaps use nesting. A question asked using a levels metaphor (“Which of the nodes below node D ...”) is easier to answer using the node-link diagram, which uses a compatible metaphor, than is one asked using containment (“Which of the directories inside directory D ...”), which works better with a treemap. The different metaphors are illustrated below, with treemaps on the left and node-link diagrams on the right.



We only scratched the surface on this. There are many other metaphors that are used in visualization, whether obvious or not. Barbara Tversky and Jeff Zacks found in the early 2000s that lines imply transitions whereas bars imply individual values. The seemingly simple choice between a bar graph and a line chart has implications on how we perceive the data.

Bizarrely, so does gravity. In our work on metaphors, Ziemkiewicz and I found that people interpreted round shapes as unstable because, they said, they might roll away. But for shapes to roll, there must be a force that causes the movement. After studying this effect some more, we found that the points in a scatter plot attract each other and that they are seemingly pulled down by gravity. We remember points not where they are in the plot but rather shift them toward clusters in our memory and let them drift slightly downward.

Findings and distinctions in visualization can be subtle, but they can have a profound impact on how well we can read the information and how we interpret it. There is much more to be learned about how visualization works and how best we can use it to represent, analyze, and communicate data.

THE CHART WARS HAVE BEGUN

BY JAKE PORWAY

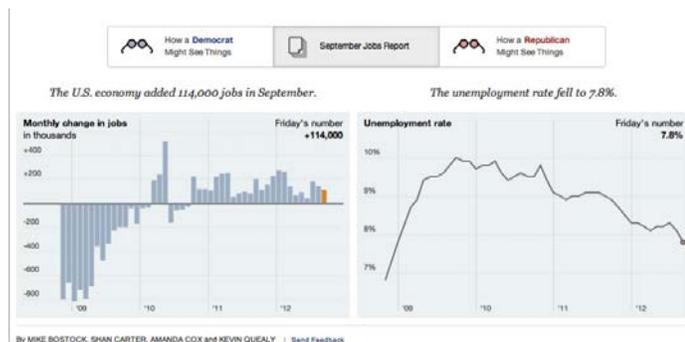
A few years ago, data scientist Alex Lundry gave a fantastic presentation describing the ways data visualization is being used by political parties to push their own agendas. He showed a Republican visualization of the House Democrats' health plan — an infographic full of sinuous pipes and literal red tape, smattered with ugly unreadable fonts and unwelcoming 8-bit color palettes — next to a Democratic visualization of the same plan, which instead looked like an Easter basket, a perfectly designed and welcoming bundle of pastel circles.

<http://blip.tv/ignitedc/alex-lundry-chart-wars-the-political-power-of-data-visualization-3021845>

The difference between the two visualizations, which present the same information, was striking. It's clear that the spin we're accustomed to hearing from politicians is now something we're going to be *seeing* from politicians as well.

The most troubling part of all this is that “we the people” rarely have the skills to see how data is being twisted into each of these visualizations. We tend to treat data as “truth,” as if it is immutable and only has one perspective to present. If someone uses data in a visualization, we are inclined to believe it. This myopia is not unlike imagining the red velvet cake we see in front of us to be the only thing that could have been created from the eggs and milk we mixed together to make it. We don't see in the finished product the many transformations and manipulations of the data that were involved, along with their inherent social, political, and technological biases.

This is why I love this *New York Times* interactive, which *Times* graphics editor Amanda Cox discussed on this site a few weeks ago:



Source: *The New York Times*.

The Times begins with the “raw” data,* from which you can compare the Republican interpretation (red-colored glasses) and Democrat interpretation (blue-colored glasses). As the window slides to either set of glasses, we're greeted by that party's familiar talking points, accompanied by the interpretation of the data that supports that view. The beauty of this visualization is not in either the Democrat or the Republican end products but rather in the concise way it draws our attention to the process of visualizing data to suit our own ends.

Taken in isolation, either final visualization gives us an answer. Taken together, the opposing visualizations force us to ask questions.

Beyond raising awareness about political bias in data visualization, this piece employs a technique that many visualizations can benefit from: comparison. For example, while I can see how well I kept to my own budget by visualizing my monthly expenses, I see a different picture when I visualize my expenses against those of others in my demographic. Similarly, visualizing my spending by type vs. by time of day gives me entirely different views of the same data. The power in each of these examples comes from seeing the data from many perspectives that altogether form a more informed view than any one individually.

I hope that, as we move into a world where people will increasingly be exposed to shiny infographics and visually stunning data interactives, more pieces remind us of the process and motives behind them. Maybe as data visualization becomes more democratized we'll learn this lesson through doing, or perhaps *The New York Times* and others will still have to remind us. Either way, I hope others will help bring to light what's going on behind the scenes so that we can take the task of visualizing data on with our eyes open.

*Let's leave aside the biases and assumptions in the data itself (e.g., How do we define “unemployment”?).

GET THE PICTURE: GAINING INSIGHT WITH DATA VISUALIZATION

FEATURING JEREMY HOWARD

Contributors

Jeremy Howard, President and Chief Scientist, Kaggle, Inc.

Insights on visualization from the HBR Blog Network

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

Overview

In today's data-rich environment, where the quantity and speed of data are growing exponentially, it is more critical than ever to be able to accurately interpret and quickly act on incoming information. But humans are poorly equipped to efficiently find meaningful patterns in overwhelmingly large databases and numerical tables; pictures—born from data visualization—work much better. Great visualization provides efficiency at understanding data patterns, helps users derive more insights, and aligns people around a shared view of a situation. Visualization allows users to confirm what they know and exposes an organization to what it doesn't know.

Tools can help even non-statisticians visualize data relationships that lead to faster insights and better decisions. Users, though, should be careful not to misrepresent data or to assume they already know the data's meaning.

Context

In a prescient HBR blog post, former Harvard Business School professor John Sviokla discussed why data visualization is so important and what its benefits are.

In a recent HBR blog post, Jim Stikeleather, an executive strategist on innovation at Dell Services, described the key elements necessary to make data visualization work.

In this *Harvard Business Review* webinar, data scientist Jeremy Howard, who has won numerous data visualization competitions, discussed and demonstrated how simple tools can turn information into insightful pictures.

Key Learnings

Visualization helps people cope with massive amounts of data.

Graphic depictions have always been a key part of how humans communicate and share information, going back to the days of cave painting. In the Cabinet War Rooms used by Churchill during World War II, walls were covered with maps and color-coded pieces of string. The ability to visualize critical information allowed leaders to be able to quickly and easily understand the situation—which remains true today.

Since WWII, the amount of data in our “digital universe” has grown exponentially. The volume, mode, quality, speed, and granularity of data have all changed. But the importance of visualization has not. In this ocean of information, the need for visualization to help understand the data is already great and is growing.

“The model you really need is to be showing people pictures, because pictures are what the human brain understands best.” —Jeremy Howard

Data visualization provides multiple benefits, particularly the ability to glean new insights.

Sviokla sees three primary benefits of superior graphic visualization:

- **Efficiency.** Great visualizations are efficient in that they let people look at vast quantities of data quickly.
- **Alignment.** A great visualization helps create a shared view of a situation. Users are able to verify and share what they know, and people can become aligned on needed actions.
- **Insight.** Perhaps most important, the ability to visualize data can help an analyst or group achieve more insight into a problem and discover a new or even greater understanding.

“I believe that we will naturally migrate toward superior visualizations to cope with this information ocean.” —John Sviokla

For example, a property and casualty insurer was able to combine visual information from Google Earth with data showing flood plain information (Figure 1). This helped the company better assess risk and enabled its salespeople to communicate to customers why they might have higher premiums.



FIGURE 1
Insights from data visualization.

Visualization of data can also help show salespeople the best opportunities in their territories and can help managers understand how well those opportunities are being realized and how to best allocate resources. Visualization of data allows users to verify what they know and exposes users to what they don't know.

“In addition to arranging the information to create shared understanding, visualization gives us the ability to combine data in order to create new insight—quickly and clearly.” —John Sviokla

In his blog, Stikeleather provides his take on the reasons to construct data visualizations.

He sees them as:

- **Confirmation.** If we already have assumptions about how a system operates, visualizations can help us check those assumptions. They can enable us to observe whether an underlying system has deviated from the model we had and assess the risk of actions we are about to undertake. This approach is used in some enterprise dashboards.
- **Education.** Visualization offers two forms of education:
 - **Reporting.** This is how we measure the performance of an underlying system, often in comparison to other systems or models.
 - **New insights.** Visualization can develop intuition and new insights.

- **Exploration.** Visualization can be used to help build a new model that allows us to predict and better manage a system. The practice of using visual discovery in lieu of statistics is called exploratory data analysis, and too few businesses currently make use of it.

“Ultimately, data visualization is about communicating an idea that will drive action.” —Jim Stikeleather

For visualization to have value, the data used must be interpretable, relevant, and novel.

In Stikeleather's blog, he emphasizes his support for visualization, but he stresses that visualization must serve an informing purpose and must be based on meaningful data.

Stikeleather's criteria for the type of data needed to produce meaningful insights through visualization are that data must be:

- **Interpretable.** With so much unstructured data today, it is critical that the data being analyzed generate interpretable information. Collecting lots of data without the associated metadata—such as what is it and where when, how, and by whom it was collected—reduces the opportunity to play with, interpret, and gain insights from it.
- **Relevant.** The data must be relevant to the persons who are looking to gain insights and to the purpose for which the information is being examined.
- **Novel.** For a visualization to be meaningful, the data used must be original or shed new light on an area.

If the information fails any of these criteria, then even the greatest special effects can't make a visualization valuable. That means that only a tiny slice of the data we can bring to life visually will actually be worth the effort.

Visualizations can mislead as well as inform.

Although they are powerful, data visualizations can also distort, hide, or otherwise mislead depending on how the underlying data is selected and displayed. For example, a chart of U.S. incomes over the past 28 years can show healthy growth or complete stagnation, depending on whether the focus is on the mean income (growth) or income growth by percentile (which shows that most income growth has come from just the top 1%). During the recent Venezuelan election, the victor's voting chart was manipulated to look like a landslide because the Y axis was not anchored to a zero point; a newspaper's version was much more statistically accurate.

Another issue is that frequently users of data have preconceived notions about their conclusion and then look for data to support that conclusion. This is in contrast to first looking at the data and letting the analysis of that data drive the conclusion.

“In most of the visualizations I see, people first decide on the result they’re looking for or the stories that they’re creating. And then they create a chart to show it. You would be much better off first creating the chart, finding out what the data says, and then creating your story.”

—Jeremy Howard

One advantage of using third-party data scientists is that they look first at data without even understanding the business and then present this data to domain experts to help interpret what it means. Fresh eyes of outsiders can often find relationships in the data that daily users overlook.

In addition to being misleading, there are other risks related to visualization that Stikeleather laid out.

- **Data quality.** The quality of the underlying data is crucial to the value of visualization. How complete and reliable is it? As with all analytical processes, putting garbage in means getting garbage out.
- **Context.** The point of visualization is to make large amounts of data approachable so we can detect patterns and draw insights from it. To do so, we need to be able to access all the potential relationships of the data elements. This context is the source of insight. To leave out any contextual information or metadata is to risk hampering our understanding.
- **Biases.** The creator of a visualization may influence the semantics and syntax of the elements of the visualization via color choices, positioning, and visual tricks—any of which can challenge the interpretation of the data. This can significantly influence how viewers understand the visualization and what insight they will gather from it.

Ignoring these risks can undermine the visualization’s purpose and confuse rather than enlighten.

Several tools help non-statisticians get started in extracting insights from data.

While there is a growing number of data visualization tools available, a simple way to begin getting experienced with data visualization can be found in Microsoft Excel. Excel offers three useful functions that can assist in revealing data relationships:

1. **Sparklines.** These are tiny charts embedded in single Excel cells. They turn data into shapes that make it easier to spot consistent patterns as well as outliers.
2. **Conditional formatting.** Adding conditional formatting to a cell allows the application of different formatting options, such as color or font style, to a cell based on the data in the cell(s). For example, a cell could be colored red when its value is between two values. By assigning conditional formatting to different data ranges, an analyst can quickly map information frequency patterns that would be tedious to extract from a purely numerical table.
3. **Pivot tables.** This tool sorts, counts, and sums spreadsheet data and then creates a second table to display the summarized data. Done iteratively, it is a quick way to get a sense of what the data looks like.

Although data sets can appear overwhelmingly large, it is almost never necessary to analyze the full set. Excel users should learn how to randomly sample rows, perhaps 10%, and then analyze just that selection.

“When you turn things from tables of numbers to pictures, you get a lot more insight.” —Jeremy Howard

Conclusion

An excellent visualization, according to Edward Tufte, expresses “complex ideas communicated with clarity, precision, and efficiency.” Clearly, excellent data visualization also tells a story through the graphical depiction of statistical information. When you are creating a visualization in an educational or confirmational role, it 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. It’s the story that helps the viewer gain insight from the data.

Additional Information

Play with the data. The data sets used by Howard in this webinar are available for users to play with. They can be accessed at www.jphoward.wordpress.com.

Sparklines by Tufte. For more information about Sparklines, read an excerpt from Tufte’s book *Beautiful Evidence*.

Color blindness. When using Excel’s conditional formatting, opt for blue and red cells rather than blue and green, which are much more difficult for color-blind viewers to distinguish.

Higher resolution. Improved visualization is possible today because of improved computing power and also because of the higher-resolution screens and projectors that now exist to display data.

To learn more about SAS data visualization and analytic capabilities, visit sas.com/visualanalytics.

Speaker Biographies

**Jeremy Howard,
President and Chief Scientist, Kaggle, Inc.**

Jeremy Howard is president and chief scientist at Kaggle. Previously, he founded FastMail (sold to Opera Software) and Optimal Decisions (sold to ChoicePoint—now called LexisNexis Risk Solutions). Prior to that he worked in management consulting at McKinsey & Company and at A.T. Kearney.

Howard received his BA in philosophy from the University of Melbourne.

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

Angelia Herrin is editor for research and special projects at *Harvard Business Review*. She oversaw the relaunch of the management newsletter line and established the conference and virtual seminar division for *Harvard Business Review*. More recently, Herrin 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 20 years, primarily with Knight-Ridder newspapers and *USA Today*. At Knight-Ridder, she covered Congress as well as the 1988 presidential election. 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.

The information contained in this summary reflects BullsEye Resources, Inc.'s, subjective condensed summarization of the applicable conference session. There may be material errors, omissions, or inaccuracies in the reporting of the substance of the session. In no way does BullsEye Resources or Harvard Business Review assume any responsibility for any information provided or any decisions made based upon the information provided in this document.

Created for *Harvard Business Review* by BullsEye Resources.

www.bullseyeresources.com

© 2012 Harvard Business School Publishing.

LEARN MORE

DATA VISUALIZATION: WHAT IS IT AND WHY IS IT IMPORTANT?

→ sas.com/data-visualization/overview.html

FOR MORE ABOUT SAS DATA VISUALIZATION AND ANALYTIC CAPABILITIES

→ sas.com/visualanalytics

TRY SAS® VISUAL ANALYTICS: BROWSE INTERACTIVE REPORTS OR GET DEMO ACCESS TO EXPLORE A SAMPLE DATA SET AND BUILD A REPORT

→ sas.com/software/visual-analytics/demos/all-demos.html

A HARVARD BUSINESS REVIEW INSIGHT CENTER REPORT



SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration. Other brand and product names are trademarks of their respective companies.
106560_S107916.0613

hbr.org