O'REILLY®

Data Visualization

A New Language for Storytelling

Mike Barlow

SAN JOSE

LONDON

NEW YORK

SINGAPORE

Strata+ Hadoop WORLD

Make Data Work

strataconf.com

Presented by O'Reilly and Cloudera, Strata + Hadoop World is where cutting-edge data science and new business fundamentals intersect—and merge.

- Learn business applications of data technologies
- Develop new skills through trainings and in-depth tutorials
- Connect with an international community of thousands who work with data

Job # 15420

Data Visualization

A New Language for Storytelling

Mike Barlow

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo O'REILLY®

Data Visualization

by Mike Barlow

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://my.safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editor: Mike Loukides

October 2014: First Edition

Revision History for the First Edition:

2014-10-14: First release

2015-03-06: Second release

The cover image is a visualization of New York hospitals using *https://mapsdata.co.uk*.

ISBN: 978-1-491-94503-2

[LSI]

Table of Contents

Data Visualization: A New Language for Storytelling

What is good visualization? It is a representation of data that helps you see what you otherwise would have been blind to if you looked only at the naked source. It enables you to see trends, patterns and outliers that tell you about yourself and what surrounds you.

—Nathan Yau, *Data Points* (Wiley)

An Emerging Universal Medium

When was the last time you saw a business presentation that did not include at least one slide with a bar graph or a pie chart? Data visualizations have become so ubiquitous that we no longer find them remarkable.

And yet they *are* remarkable. Consider this observation from the Second Edition of *The Visual Display of Quantitative Information* (Graphics Pr) by Edward R. Tufte:

> The use of abstract, non-representational pictures to show numbers is a surprisingly recent invention, perhaps because of the diversity of skills required—the visual-artistic, empirical-statistical, and mathematical. It was not until 1750–1800 that statistical graphics—length and area to show quantity, time-series, scatterplots, and multivariate displays—were invented, long after such triumphs of mathematical ingenuity as logarithms, Cartesian coordinates, the calculus, and the basics of probability theory.

It seems counterintuitive to believe that a phenomenon can be remarkable and commonplace at the same time. But there are plenty of examples: birdsong, beautiful sunsets, pizza, sex—to name a few.

Some argue that data graphics have already become a sort of lingua franca, a common global language crossing boundaries of culture and politics. Nathan Yau sees data visualization "as a medium rather than a specific tool." Good data visualizations are more than just endpoints of analytic processes; they are platforms for telling stories, conveying knowledge, eliciting emotions, and sparking curiosity.

At their most basic level, visualizations enable us to compare numbers (or sets of numbers) quickly. Visualizations rely on our innate human ability to discern patterns rapidly and convert them into usable information. Our early ancestors needed pattern-recognition skills to keep them safe from camouflaged predators.

Data visualizations appeal to similar circuits in our brains. The major difference between us and our ancestors is situational. They were looking for signs of predators or prey; we're trying to figure out where to invest the money in our retirement accounts.

"When you're dealing with more than two numbers, it's much easier to compare them if they're shown in a chart than if they're shown in a tabular format," says Francois Ajenstat, director of product management at Tableau. "Maybe it's better to ask, 'When is a visualization *not* the right approach?' When you're looking at an invoice, for example, you just want to see the numbers. But when you're looking at rows and columns of data, then visualization is actually the beginning of the analytics process."

Noah Iliinsky works at IBM's Center for Advanced Visualization. An evangelist for data visualization, he advocates a rigorously disciplined approach.

"There are four rules that I've come up with, and I think they're pretty sound," he told an audience at the O'Reilly Strata Conference + Hadoop World in New York City in October 2013. "The first is purpose: why are you doing this visualization? The second is content: what are you trying to visualize? The third is structure: how are you going to visualize it? how do we best reveal the most important data and relationships? The fourth is formatting: how will it look and feel? How will it be consumed? Formatting is the icing on the cake!"

Even if your purpose is clear and your data is sound, your choice of structure is critical.[1] For example, if you're trying to highlight relationships among data points, use scatterplots, matrix charts, or network diagrams. For showing parts of a whole, use pie charts or treemaps.

If your goal is comparing a set of values, use bar charts, block histograms, or bubble charts. When you're tracking data that rises and falls over time, use line graphs or stack graphs. When you're analyzing text, use word trees or tag clouds.

Properties and Best Uses of Visual Encodings

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	**yes**	**infinite**	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	**optional** (alphabetical or numbered)	**infinite**	Good	Good	Good	Good
	length	**yes**	**many**	Good	Good		
	size, area	**yes**	**many**	Good	Good		
	angle	**yes**	medium/few	Good	Good		
	pattern density	**yes**	few	Good	Good		
	weight, boldness	**yes**	few		Good		
	saturation, brightness	**yes**	few		Good		
	color	no	few (< 20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	**infinite**			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	**yes**	few		Good		

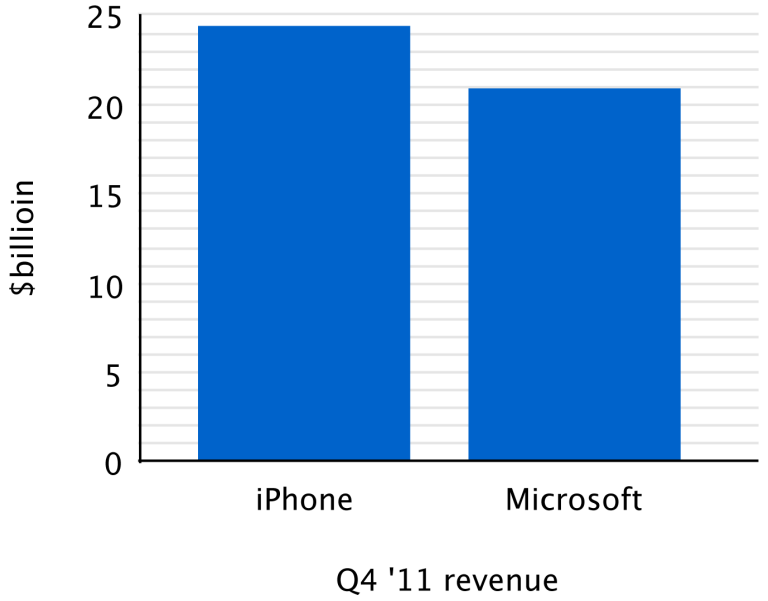
Noah Iliinsky • ComplexDiagrams.com/properties • 2012-06

Your choice of "visual encoding"—all of the possible formatting options available—is also crucial. Picking the wrong structure or the wrong format can obscure your data or create misleading impressions.

In many instances, the simplest structures and formats can be the most powerful. The value of stark simplicity is illustrated by the following

1. Commenting on an earlier draft of this paper, Jeffrey Heer, professor of Computer Science at the University of Washington, writes, "The systematic study of visual encoding traces back to French cartographer and designer Jacques Bertin; his seminal book, *The Semiology of Graphic*, is a powerhouse! Visual encoding was made the object of experimental study by statistician William S. Cleveland and colleagues, who published papers on the topic back in the 1980s. These are true giants in the field of data visualization, on par with Edward Tufte."

bar chart, which compares iPhone sales with total Microsoft revenues over a three-month period in 2011.



Ideally, says Iliinsky, following the rules enables you to create a visualization that "tells the story" of the underlying data set. "We have incredible software in our brain and incredible hardware in our optical system that make us extremely good at pattern recognition and pattern matching," he says. "We're also good at spotting where the pattern is broken, where there are gaps and outliers." Good data visualizations bring patterns, trends, gaps, and outliers to the surface, making them visible to our eyes and accessible to our brains.

"Visualizations give us access to huge amounts of data, very rapidly," says Iliinsky. "Visualizations play to the skills that are wired into our brains. Those are skills we don't have to learn—we already have them, free of charge."

Making Points, Deflating Arguments

Author and researcher Richard Florida has built a successful career on data analysis and data visualization. Florida, a professor at New York University and the University of Toronto, is the author of three bestsellers, *The Rise of the Creative Class* (Basic Books), *Cities and the*

Creative Class (Routledge), and *The Flight of the Creative Class* (Harper Business). He is also a senior editor at *The Atlantic*.

"Data visualizations, especially maps, have been extremely helpful in my writing at *Atlantic Cities*. They've helped me provide readers with a visual understanding of complex issues, specifically when looking at questions of geography," says Florida. "For example, in order to help visualize the significant class and workforce divide in our cities, I have used a series of maps to illustrate the point. The maps have been useful in identifying patterns and understanding economic development trends. If you were to look at the body of my work at *Atlantic Cities*, you'll see that maps…are a central piece of my work."

Data visualizations can also deflate an argument. A good example of this is "512 Paths to the White House," a comprehensive interactive graphic that showed the inevitability of President Obama's reelection. The graphic was created for *The New York Times* by Mike Bostock and Shan Carter and was published at a point in the campaign when many journalists, politicians, and pollsters were describing the race as a largely even match.

"Shan and I felt like TV anchors spent a lot of time talking about hypotheticals prior to election night," says Bostock. Although there were at least 512 possible scenarios, the anchors "could only discuss one scenario at a time," recalls Bostock. As a result, viewers "had very little understanding of how likely this particular scenario was, and what the overall probabilities were."

The interactive visualization created by Bostock and Carter enabled readers to consider all 512 paths and assess for themselves the likelihood of a Romney victory. "So we lost that edge-of-your-seat dramatic television experience, but we gained a better understanding of what was happening," says Bostock.

Exploratory Versus Explanatory Visualization

It seems fair to say that data visualization is essentially a form of storytelling. But in the same way that you wouldn't necessarily share a Stephen King story with a group of toddlers or tell children's bedtime stories to middle-aged adults at a cocktail party, different audiences need different types of data visualization.

"Exploratory graphics are something you make for yourself, while expository (or explanatory) graphics are something you make for oth-

ers," says Bostock. "The primary goal of exploratory visualization is speed—to find insights quickly—and preferably in a comprehensive, unbiased way."

Anne Milley, a senior director of analytics at SAS, sees visualization as "the key to achieving efficiency and effectiveness in the value creation process." Visualization, from her perspective, unlocks the real value of data. "In the discovery phase of analysis, visualization maximizes use of the analyst's visual bandwidth and frees up working memory, of which we have so little. As the analyst, you are both information producer and consumer. As you visually explore the data, what you see *informs* your next step," says Milley.

Because an analyst typically looks at *many* graphs during the exploratory phase, "those graphs should be quick and easy to create," says Milley. "Visual data exploration lets you stay in flow and keep yourself focused on solving the problem at hand. And it also helps you see if there's something interesting in the data that you might have missed when it was in tabular form."

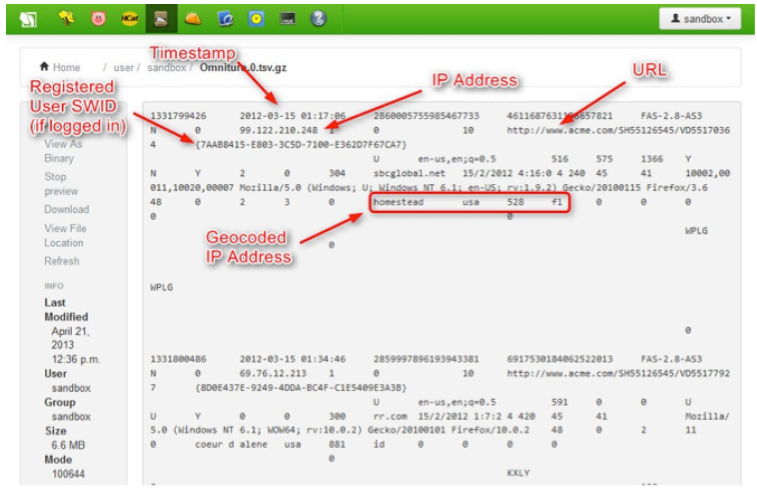
The process for creating explanatory visualizations is generally slower, "because you have to externalize the context you gained exploring, which means annotations and views intended to reveal those specific insights. Think of exploratory graphics as reading and expository graphics as teaching," says Bostock.

Rachel Binx is a cofounder of Meshu, a company that converts personal travel data into custom jewelry. In a previous role at San Francisco-based Stamen Design, she worked with clients such as MTV, Facebook, and the MoMA. "Exploratory visualization is most often done by and for the people closest to the data," says Binx. "So you can get away with making obtuse, unclear, or hard-to-use visualizations, because the 'audience' usually already understands the data, and is invested in the exploration."

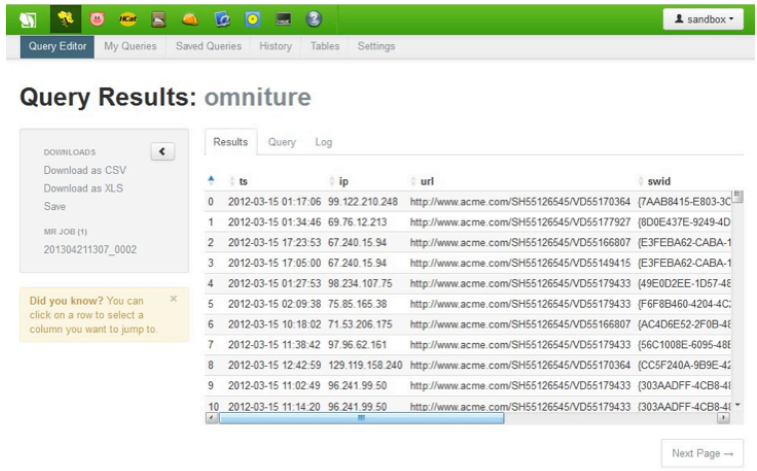
Exploratory visualization can also be used to test insights with small audiences before the data is "ready for prime time." Ofer Mendelevitch, director of data sciences at Horntonworks, uses healthcare data as an example. "Let's say you've got data about patients and medications. As a data scientist, you can run a model. But you might not have the expertise to know if the data is good or bad. So it makes sense to create a simple chart, just something with an X and Y axis, and show the chart to a subject matter expert. The expert should be able to tell you if

something looks strange. Maybe you have the wrong algorithm, or maybe your data is skewed."

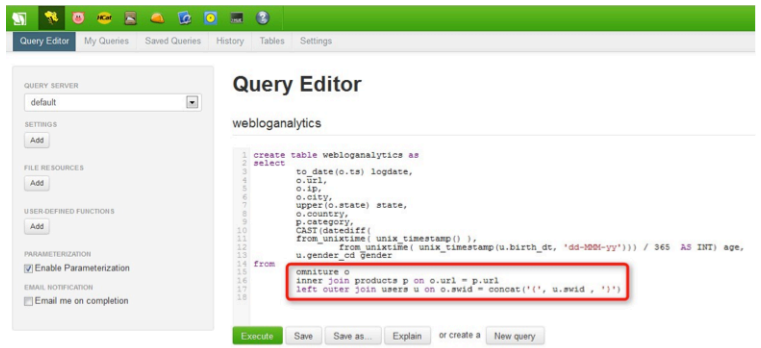
Following are six images from a Hortonworks tutorial on visualized clickstream data. In this example, the weblog data is combined with CRM data to visualize customer behavior. The following image shows raw data received from the Hortonworks website.



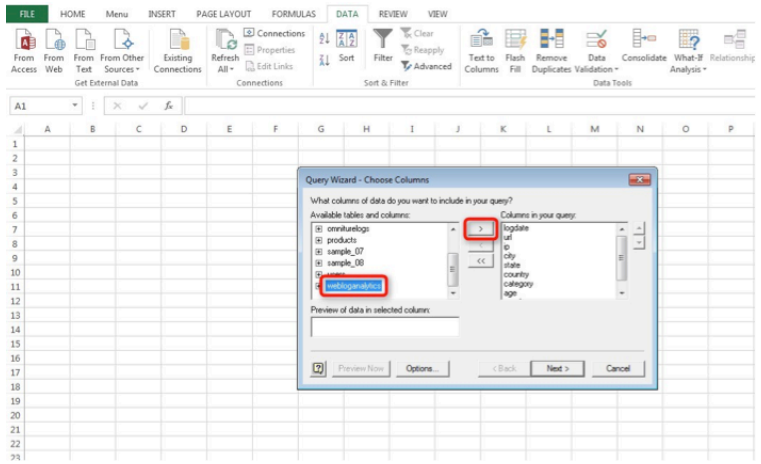
The following image shows data brought into HDFS and placed in a table.



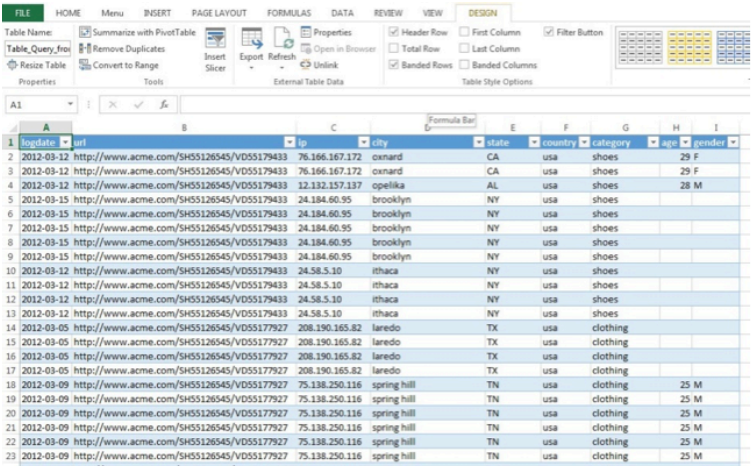
Next, the data is processed using Hive.



After being processed in Hive, the data is brought into Excel.

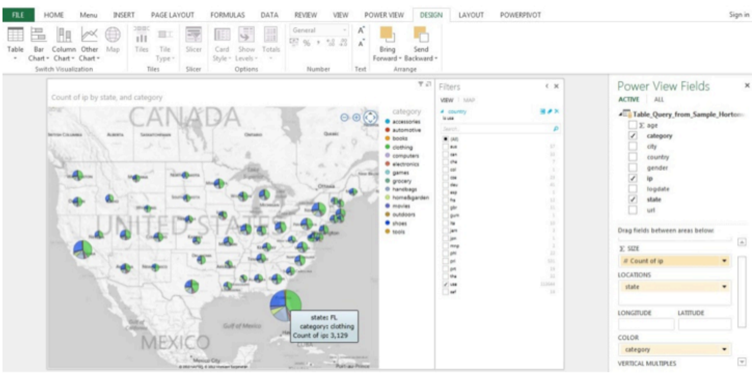


Now all the data from Hadoop is in Excel.



	logdate	url	ip	city	state	country	category	age	gender
2	2012-03-12	http://www.acme.com/SH55126545/VD55179433	76.166.167.172	oxnard	CA	usa	shoes	29	F
3	2012-03-12	http://www.acme.com/SH55126545/VD55179433	76.166.167.172	oxnard	CA	usa	shoes	29	F
4	2012-03-12	http://www.acme.com/SH55126545/VD55179433	12.132.157.137	opelika	AL	usa	shoes	28	M
5	2012-03-15	http://www.acme.com/SH55126545/VD55179433	24.184.60.95	brooklyn	NY	usa	shoes		
6	2012-03-15	http://www.acme.com/SH55126545/VD55179433	24.184.60.95	brooklyn	NY	usa	shoes		
7	2012-03-15	http://www.acme.com/SH55126545/VD55179433	24.184.60.95	brooklyn	NY	usa	shoes		
8	2012-03-15	http://www.acme.com/SH55126545/VD55179433	24.184.60.95	brooklyn	NY	usa	shoes		
9	2012-03-15	http://www.acme.com/SH55126545/VD55179433	24.184.60.95	brooklyn	NY	usa	shoes		
10	2012-03-12	http://www.acme.com/SH55126545/VD55179433	24.58.5.10	ithaca	NY	usa	shoes		
11	2012-03-12	http://www.acme.com/SH55126545/VD55179433	24.58.5.10	ithaca	NY	usa	shoes		
12	2012-03-12	http://www.acme.com/SH55126545/VD55179433	24.58.5.10	ithaca	NY	usa	shoes		
13	2012-03-12	http://www.acme.com/SH55126545/VD55179433	24.58.5.10	ithaca	NY	usa	shoes		
14	2012-03-05	http://www.acme.com/SH55126545/VD55177927	208.190.165.82	laredo	TX	usa	clothing		
15	2012-03-05	http://www.acme.com/SH55126545/VD55177927	208.190.165.82	laredo	TX	usa	clothing		
16	2012-03-05	http://www.acme.com/SH55126545/VD55177927	208.190.165.82	laredo	TX	usa	clothing		
17	2012-03-05	http://www.acme.com/SH55126545/VD55177927	208.190.165.82	laredo	TX	usa	clothing		
18	2012-03-09	http://www.acme.com/SH55126545/VD55177927	75.138.250.116	spring hill	TN	usa	clothing	25	M
19	2012-03-09	http://www.acme.com/SH55126545/VD55177927	75.138.250.116	spring hill	TN	usa	clothing	25	M
20	2012-03-09	http://www.acme.com/SH55126545/VD55177927	75.138.250.116	spring hill	TN	usa	clothing	25	M
21	2012-03-09	http://www.acme.com/SH55126545/VD55177927	75.138.250.116	spring hill	TN	usa	clothing	25	M
22	2012-03-09	http://www.acme.com/SH55126545/VD55177927	75.138.250.116	spring hill	TN	usa	clothing	25	M
23	2012-03-09	http://www.acme.com/SH55126545/VD55177927	75.138.250.116	spring hill	TN	usa	clothing	25	M

The final step is creating a visualization in Excel.



In this example, the visualization is likely to be used for exploratory purposes. But it could also be used as an explanatory visualization in a presentation for internal users or partners.

It's important to distinguish between exploratory and explanatory visualization because each represents a different use case, according to Scott Murray, a code artist and an assistant professor of design at the University of San Francisco, where he teaches data visualization and interaction design. "Exploratory visualization is helpful when you have a new data set, but don't yet know what story it's trying to tell you. So you need to explore the data, visually, to get a sense of any interesting patterns and trends. This usually involves either an interactive visualization (so you can quickly compose different views of the data) or using a tool that quickly generates and outputs multiple views on

its own. For example, R and Tableau are used heavily for this exploratory process, because they can quickly generate tons of different views into the data," says Murray. "Once you know the story, you enter the explanatory phase, in which you're designing a more limited, yet refined view optimized for communicating that story to someone else. Usually, that 'someone else' isn't familiar with the data, and doesn't have a sense of the larger context," he continues. "A good explanatory visualization will provide that context and highlight the portions of the data that you feel are most meaningful."

Best of Both Worlds?

Not every data visualization falls neatly into the "exploratory" or "explanatory" category. Some appear to reside happily in both worlds.

The MasterCard Mobile Payments Readiness Index presents more than 50 different kinds of data from multiple sources in a coherent, interactive, and highly appealing graphic. The index was produced by MasterCard's Global Insights team. The visualization was led by Adam Bell, a Vice President/Business Leader for Global Insights and MasterCard Advisors.

It began as an effort to leverage data about mobile payments from 34 markets in various parts of the world. Developing a better understanding of the data was important to MasterCard, which is a key player in the emerging mobile payments sector.

"We were trying to determine which markets were ready to adopt mobile payment methods," says Bell. "Knowing where to focus your efforts is important for us and important for the industry."

The project had two distinct phases. The first was collecting, gathering, organizing, and analyzing data from multiple markets across the globe. The second phase was creating a presentation layer that was easy to understand and also conveyed the complexity of the underlying data. "We needed to formulate a strong hypothesis, follow it up with evidence, write a story around the evidence, and then bring it to life visually," Bell explains. "The visualization really showcases the data, so the design and execution are absolutely critical."

Figure 1. MasterCard Mobile Payments Readiness Index

The index was used internally as a tool for exploratory analysis and released externally as a dual-purpose tool for both exploration and explanation. Bell says the company will produce and publish similar visualizations in the future.

Complex visualizations such as the MasterCard Mobile Payments Readiness Index can require teams of data scientists, software engineers, marketers, and business leaders. Most organizations do not have the resources necessary to create visualizations at that level. But if the demand for Hollywood-grade data visualizations rises, enterprising vendors will doubtlessly step into the breach with affordable solutions.

Challenges, Perils, and Pitfalls

In addition to offering a variety of technical challenges, data visualization presents a mixed bag of moral and ethical dilemmas. "First, it is far too easy to take data out of context and represent it in a dishonest way," says Scott Murray. "Second, humans are far too trusting of visual images! If a chart looks halfway respectable, people will interpret it as unquestionable 'fact.' People who work with data know the reality is much messier and more open to interpretation. Where did the original data values come from? Who recorded them, and how? Is the source trustworthy? What is the source's motivation and intent in sharing this data? What is the intent of the visual designer? What is the designer

trying to persuade me to do or think?" Murray suggests that society needs "a cultural push toward data literacy that encourages citizens to be critical of data-images and to question their sources."

There is also the danger of data visualizations becoming a form of decoration. "There's always a temptation to produce something superficially beautiful or impressive but lacking insight," says Mike Bostock.

Jerzy Wieczorek, a Ph.D. candidate in the Department of Statistics at Carnegie Mellon University, spent several years at the US Census Bureau as a mathematical statistician. He warns against "using the wrong tool for the job" simply because it seems the easiest choice. For example, if your dataset includes the names of countries, it might be tempting to plot the data on an international map. "But a bar chart might be the better choice, since it's easier to compare the heights of two bars than the color-intensities of two filled-in areas on a map," says Wieczorek.

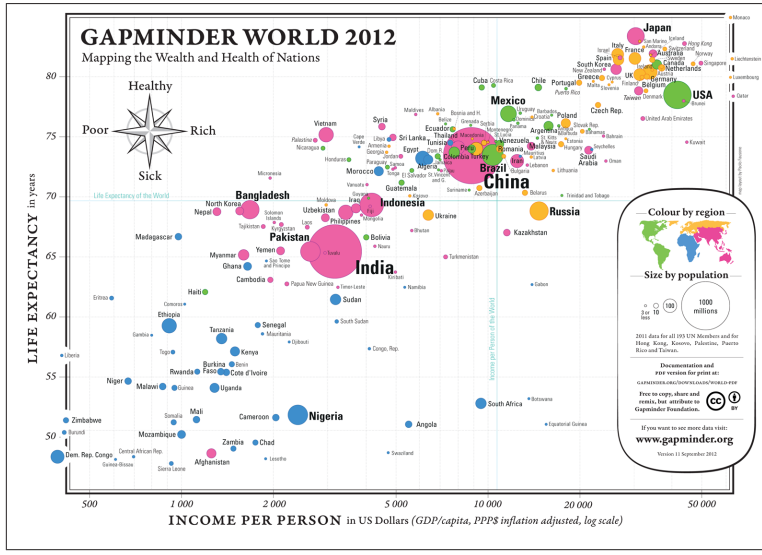


Figure 2. Image courtesy of the Gapminder Foundation.

For example, Gapminder's visualization showing relationships between income and life expectancy is essentially a scatterplot. As Wieczorek notes, "This scatterplot is much richer and more informative than a shaded map (which would have been a patchwork of hard-to-

compare colors). Here, you see immediately which countries are above or below average on either variable."

Figure 3. "Small multiples" facilitate comparison between and within groups. These back-to-back bar charts efficiently relate 714 numbers (17 degree fields, by 2 genders, across 21 years) about bachelor's degrees earned in the US. Image courtesy of Jerzy Wieczorek.

Jeffrey Heer is a cofounder and CXO (Chief Experience Officer) at Trifacta. He is also a professor of Computer Science at the University of Washington, where he leads the Interactive Data Lab. Previously, he led the Stanford Visualization Group. Members of the group created a number of popular tools, including D3.js (Data-Driven Documents)[2] and Data Wrangler. Heer says that finding the right data and making sure that it is properly structured for visualization is critical. "It is surprisingly easy to overlook the immense amount of work that goes into preparing data for analysis," he says. "Visualization alone is

2. Mike Bostock created the initial prototype of D3 while on leave of absence from Stanford. According to Bostock, Jason Davies subsequently became D3's coauthor and maintainer and contributed significant parts of D3's functionality, such as the geographic projection pipeline.

not a magic bullet. For example, owning a hammer does not make you a master carpenter—though it certainly helps you be better at it. Similarly, effective visualization arises through the combination of tools and skilled use. Some amount of automation (e.g., automatic presentation methods) can help, but a thoughtful (and skeptical) analyst with driving questions is essential."

Worth a Thousand Words?

"Visualization is an extremely powerful tool. We all know a picture is worth a thousand words—it's all true," says Ofer Mendelevitch of Hortonworks. "When you've got a good visualization, people get it right away and you get a conversation going. You get feedback. It accelerates productivity. It's far better than talking on the phone or sending an email. You instantly convey the same idea to many minds."

But there's a flipside. "If you bring the wrong visualization, then a lot of bad things can happen." Finding the right ways of conveying information to an audience can be as important as the information itself, says Mendelevitch. "You need to find the best combination of techniques to create visualizations that inspire a wonderful flow of productivity." Resist the urge to show off, he says. "A lot of times as a data scientist, you tend to think, 'Wow, this is cool stuff!' Cool is okay, but you really have to think it through. Who is your audience? What is their level of attention? What are you trying to convey? Which tools will do the best job of conveying the message?"

A Range of Techniques for Visualizing Data

Hadley Wickham is Chief Scientist at RStudio and an adjunct assistant professor at Rice University. He builds tools (both computational and cognitive) that make data preparation, visualization, and analysis easier. His contributions to R include over 30 R packages for data analysis (ggplot2, plyr, reshape), making frustrating parts of R easier to use (lubridate for dates, stringr for strings, httr for accessing web APIs), and streamlining the R package development process (roxygen2, testthat, devtools, profr, staticdocs).

"One of the hardest parts of visualization is getting the data in the right form. A lot of analysts spend 80 percent of their time just getting the data ready to analyze," says Wickham. "When I was a student, I thought

it would be wonderful to get hired by a company with a great database where all the data was beautiful, clean, and correct."

The reality, says Wickham, is that most data is messy and disorganized. "So your first step is usually tidying up the data. When you've got it cleaned up and in the right form, then you can begin applying your visualization tools," he says. You are much more likely to uncover hidden nuggets of useful insight in visualizations that are built on a foundation of clean data, he says.

"The goal is asking a precise question that can be answered with an algorithm. Visualizations help you refine your question. When you can answer your question with a number or with a handful of numbers, then you have a model." He continues, "Some people like to model first and then visualize. I like to visualize first, and then model."

Wickham says that he generally works in R "because R currently has the best tools." He begins by tidying up the data with reshape 2. Then he puts in the data into ggplot2, his tool for visualization. "The goal of ggplot2 is to declare how your data should map to things in the visualization, and then ggplot2 goes away and generates that plot," he explains. Last, he feeds the data into plyr, which is a tool (or set of tools) for splitting big data into manageable pieces. "Plyr makes it easier to express common data manipulation operations (e.g., selecting variables, selecting rows, rearranging rows, adding new variables and summarizing data from multiple values to a single value)."

At the other end of the visualization spectrum is Microsoft, which made a strategic decision four years ago to "double down" on self-service visualization tools. Rather than reinvent the wheel, Microsoft decided to embed the newer tools within Excel. "Our users were telling us they didn't want to learn new tools, and so the obvious answer was integrating self-service visualization capabilities into Excel," says Herain Oberoi, a director of product management at Microsoft. "We made a conscious decision to funnel innovation and new capabilities, including R&D efforts from our Microsoft Research Group, into Excel. There was a huge effort behind this. You can see it not just in terms of the next version of Excel, but also in Power Map, which is a 3D geospatial visualization tool that lets you literally 'fly through' the data."

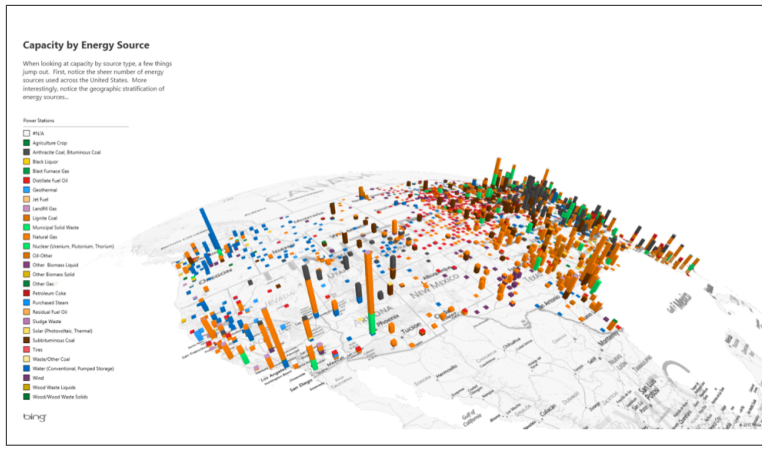


Figure 4. 3D visualization overlays bar charts and a map to show capacity by source across the United States.

Oberoi has no argument with Hadley Wickham's observation that most of the hard work of data visualization involves cleaning and tidying up messy datasets. "Visualization is the tip of the iceberg," says Oberoi. "You have to do a ton of work behind the scenes." The main difference between their approaches is that Microsoft, characteristically, is happy to provide a black box for users without Jedi-level data-wrangling skills.

"The reason we're offering those capabilities is because our customers are telling us that they want to do their own data visualizations," says Oberoi. "When you get into nonrelational data sources like Hadoop, it becomes even more important to have tools for formatting, massaging, and cleaning data so you can do visualizations against it. It's the old story of garbage in, garbage out. You can have a great visualization, but it can be completely misleading if the data was bad or dirty. That's not a sexy part of data visualization, but it's a big deal and it's very important."

Going Mainstream?

As data visualization becomes less of an esoteric art and more of a do-it-yourself process, it's likely that more of us will begin incorporating homemade charts and graphics into our daily communications. It certainly seems as though data visualization has already become an essential part of nonfiction storytelling.

Will data visualization transform modern culture, perhaps the same way that mathematics transformed science during the Renaissance? Or will data visualization become increasingly baroque and exotic? Nathan Yau compares data visualization to the written word. Like Iliinsky, he believes there are rules for data visualization, but the rules "aren't dictated by design or statistics." He writes, "Rather they are governed by human perception, and they ensure accuracy when readers interpret encoded data."

There is, of course, a danger inherent in visualizing data. When data is transformed into a visual image, it is no longer data—it becomes something else. Visualized data is not an abstraction; it has form, dimension, and various qualities. Visualizing data is like breathing life into dust—it's an act of creation. In the wrong hands, data visualizations can become tools of propaganda.

In his brilliant book, *Thinking, Fast and Slow* (Farrar, Straus and Giroux), Nobel laureate Daniel Kahneman describes two kinds of thought processes: System 1, which relies on intuition and produces snap judgments, and System 2, which relies on deliberative analysis and produces what most of us would refer to as "well-reasoned" judgment. Data visualizations, it seems, are perfect for System 1 scenarios, since they allow us to see the "big picture" instantly and draw conclusions very rapidly.

But if data science is about careful analysis—precisely the kind of slow rumination that characterizes System 2 thinking—isn't the whole idea of visualizing data something of a cheat? Raw data demands our full attention, while visualized data requires only a passing glance. Is that a good thing, or a bad thing? It probably depends on where you sit. If you're a data scientist, raw data is like raw vegetables—you might not love them, but you know that you need them. If you're a more casual consumer of data, a good visualization is like a take-out gourmet meal—it satisfies your appetite and doesn't require hours of your time to prepare.

About the Author

Mike Barlow is an award-winning journalist, author and communications strategy consultant. Since launching his own firm, Cumulus Partners, he has represented major organizations in numerous industries.

Mike is coauthor of *The Executive's Guide to Enterprise Social Media Strategy* (Wiley, 2011) and *Partnering with the CIO: The Future of IT Sales Seen Through the Eyes of Key Decision Makers* (Wiley, 2007). He is also the writer of many articles, reports, and white papers on marketing strategy, marketing automation, customer intelligence, business performance management, collaborative social networking, cloud computing, and big data analytics.

Over the course of a long career, Mike was a reporter and editor at several respected suburban daily newspapers, including *The Journal News* and the *Stamford Advocate*. His feature stories and columns appeared regularly in *The Los Angeles Times*, *Chicago Tribune*, *Miami Herald* , *Newsday*, and other major US dailies.

Mike is a graduate of Hamilton College. He is a licensed private pilot, an avid reader, and an enthusiastic ice hockey fan. Mike lives in Fairfield, Connecticut, with his wife and two children.