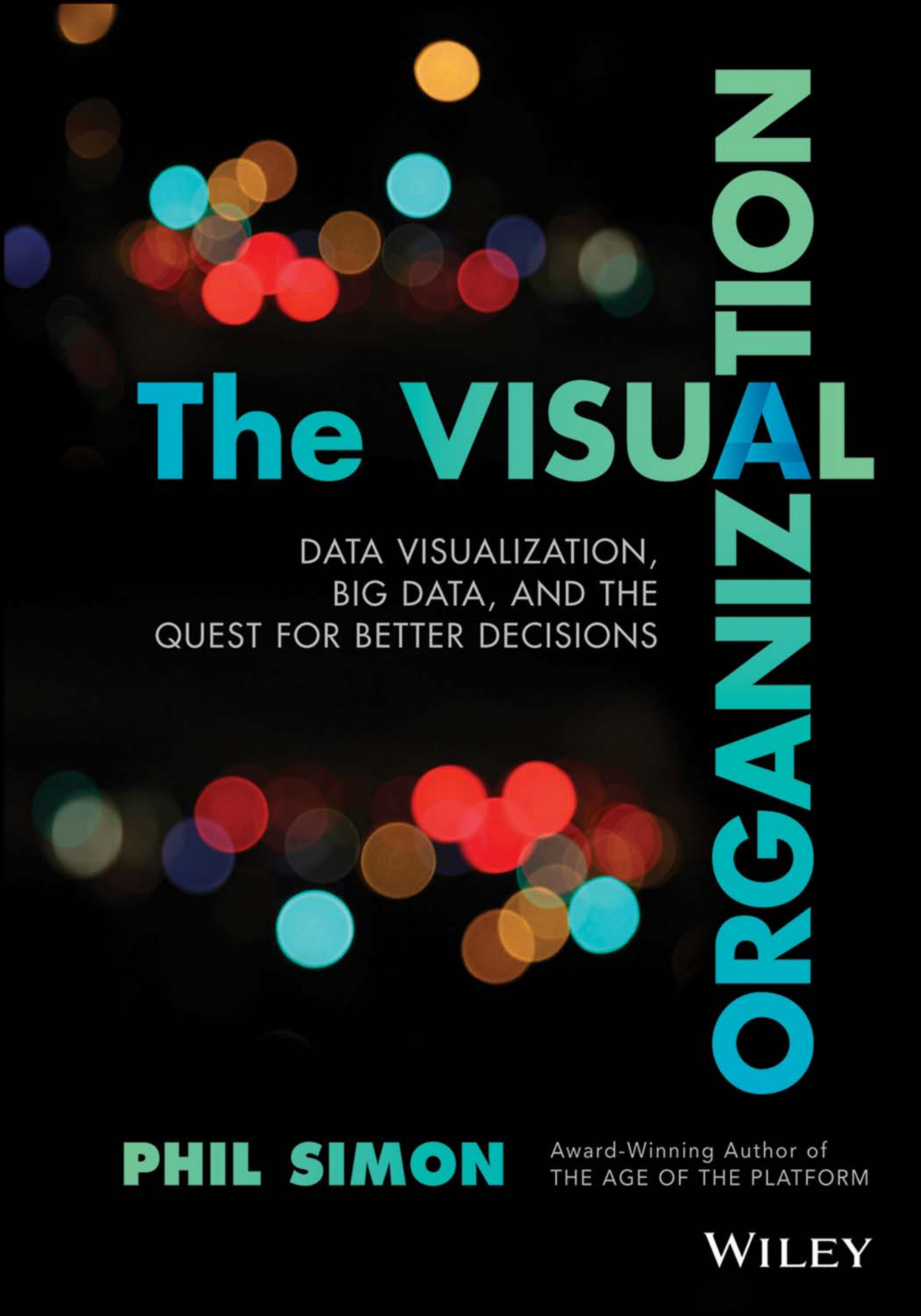
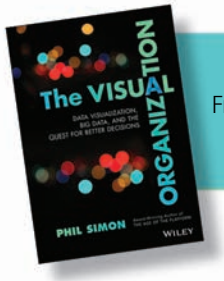
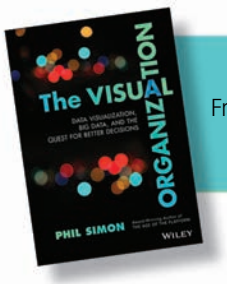

The VISUAL ORGANIZATION
DATA VISUALIZATION, BIG DATA, AND THE QUEST FOR BETTER DECISIONS
PHIL SIMON
Award-Winning Author of THE AGE OF THE PLATFORM
WILEY



Contents



CHAPTER 1

The Ascent of the Visual Organization

Where is the knowledge we have lost in information?
—T. S. Eliot

Why are so many organizations starting to embrace data visualization? What are the trends driving this movement? In other words, why are organizations becoming more *visual*?

Let me be crystal clear: data visualization is by no means a recent advent. Cavemen drew primitive paintings as a means of communication. "We have been arranging data into tables (columns and rows) at least since the second century C.E. However, the idea of representing quantitative information graphically didn't arise until the seventeenth century."* So writes Stephen Few in his paper "Data Visualization for Human Perception."

In 1644, Dutch astronomer and cartographer Michael Florent van Langren created the first known graph of statistical data. Van Langren displayed a wide range of estimates of the distance in longitude between Toledo, Spain, and Rome, Italy. A century and a half later, Scottish engineer and political economist William Playfair invented staples like the line graph, bar chart, pie chart, and circle graph.†

Van Langren, Playfair, and others discovered what we now take for granted: compared to looking at individual records in a spreadsheet or database table, it's easier to understand data and observe trends with simple graphs and charts.

.

(The neurological reasons behind this are beyond the scope of this book. Suffice it to say here that the human brain can more quickly and easily make sense of certain types of information when they are represented in a visual format.)

This chapter explores some of the social, technological, data, and business trends driving the visual organization. We will see that employees and organizations are willingly representing—or, in some cases, being forced to represent—their data in more visual ways.

Let's start with the elephant in the room.

THE RISE OF BIG DATA

We are without question living in an era of Big Data, and whether most people or organizations realize this is immaterial. As such, compared to even five years ago, today there is a greater need to visualize data. The reason is simple: there's just so much more of it. The infographic in Figure 1.1 represents some of the many statistics cited about the enormity of Big Data. And the amount of available data keeps exploding. Just look at how much content people generate in one minute on the Internet, as shown in Figure 1.2.

Figures 1.1 and 1.2 manifest that Big Data is, well, big—and this means many things. For one, new tools are needed to help people and organizations make sense of this. In *Too Big to Ignore*, I discussed at length how relational databases could not begin to store—much less analyze—petabytes of unstructured data. Yes, data storage and retrieval are important, but organizations ultimately should seek to use this information to make better business decisions.

OPEN DATA

Over the past few years, we've begun to hear more about another game-changing movement: open data. (Perhaps the seminal moment occurred when Sir Tim Berners-Lee gave a 2010 TED talk on the subject.*) Put simply, open data represents "the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents, or other mechanisms of control."[1]

While critical, the arrival of Big Data is far from the only data-related trend to take root over the past decade. The arrival of Big Data is one of the key factors explaining the rise of the Visual Organization.

Think of open data as the liberation of valuable information that fosters innovation, transparency, citizen participation, policy measurement, and better, more efficient government. Examples of open or public datasets include music metadata site MusicBrainz and geolocation site OpenStreetMap. But it

* To watch the talk, go to http://tinyurl.com/tim-open-data.

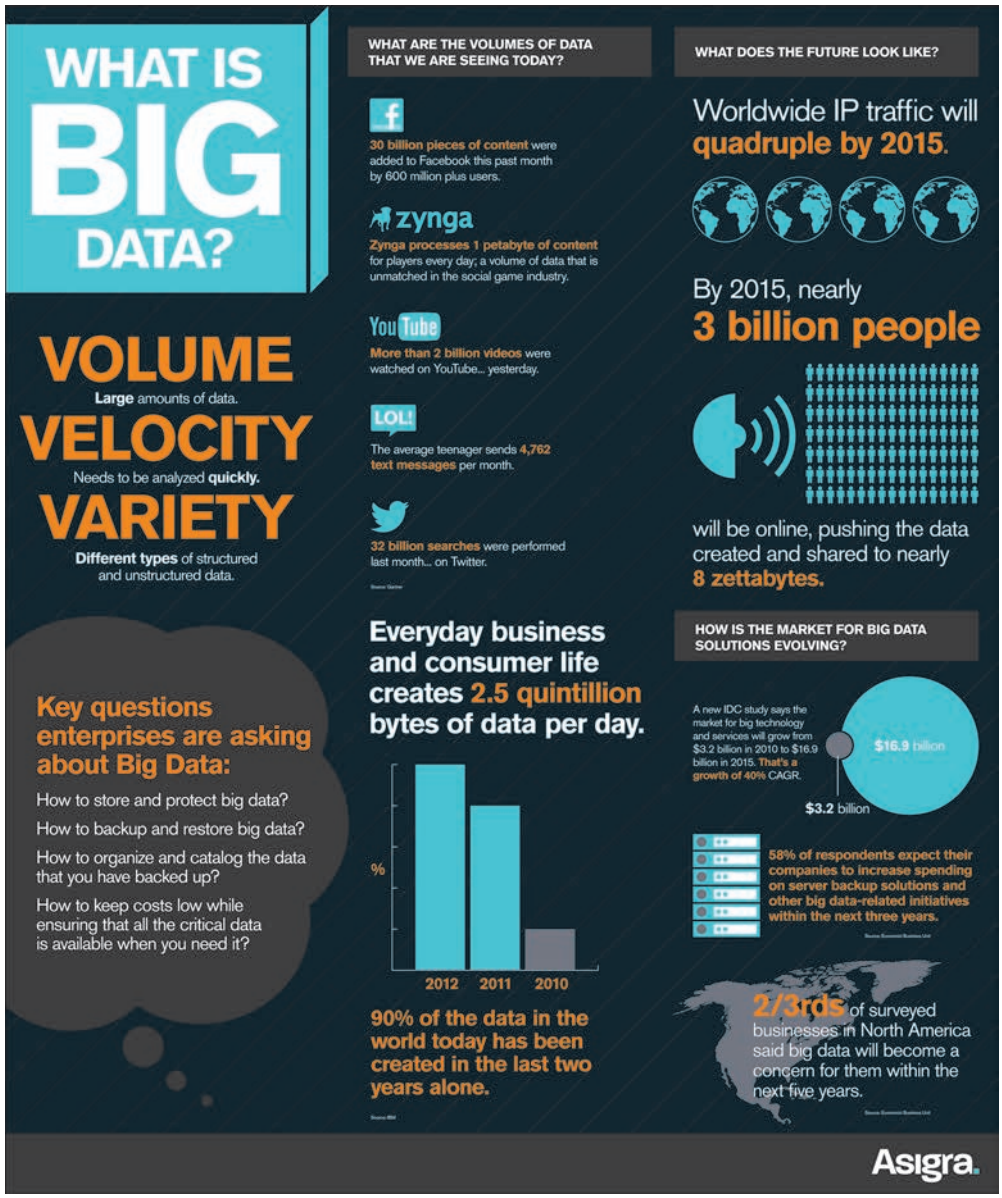


Figure 1.1 What Is Big Data?
Source: Asigra

doesn't stop there. Anyone today can access a wide trove of scientific, economic, health, census, and government data. Data sources and types are being released every day.* And, as Chapter 2 will show, there's no dearth of powerful and user-friendly tools designed specifically to visualize all this data.†

† For some of them, see http://opendata-tools.org/en/visualization.

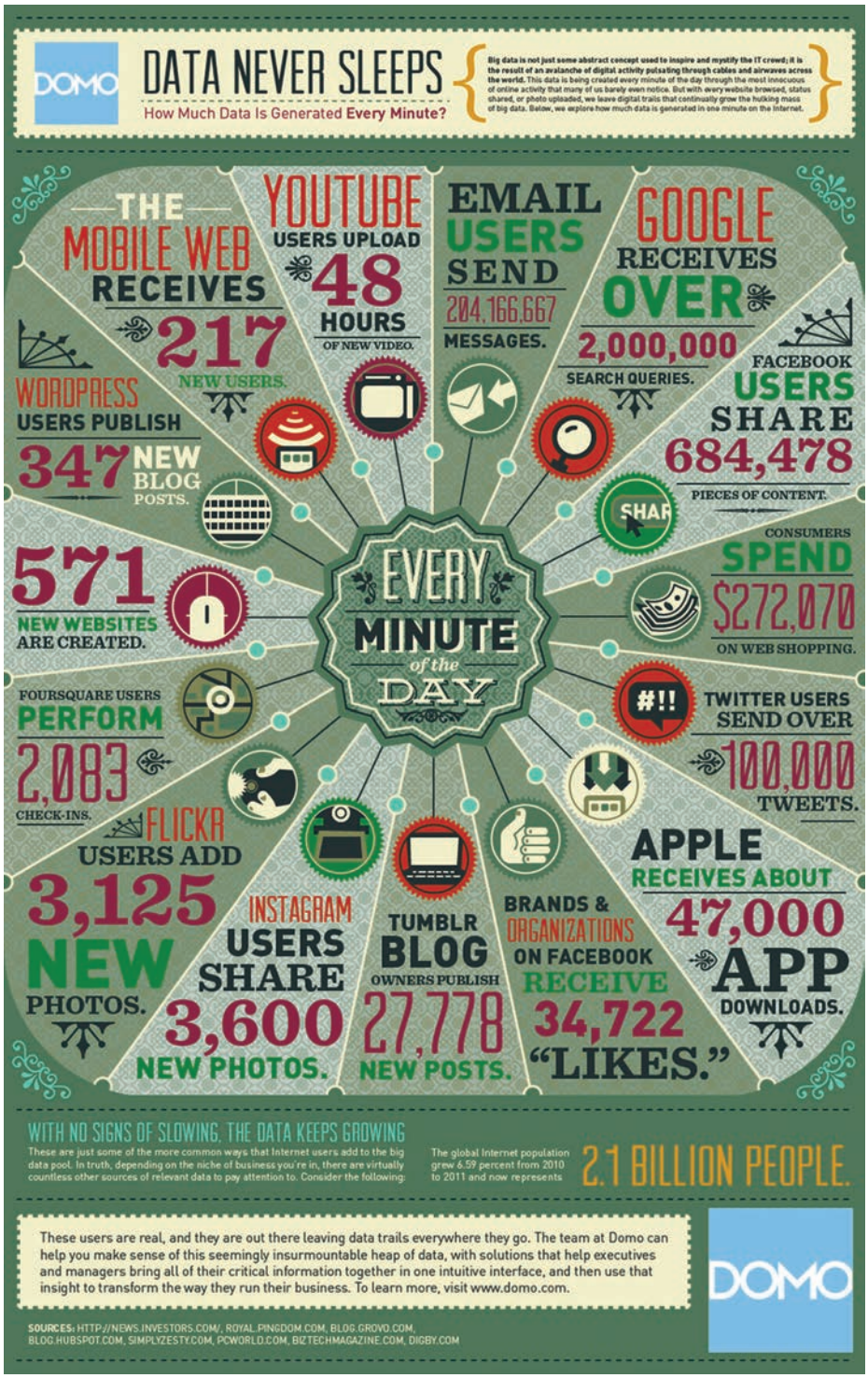


Figure 1.2 The Internet in One Minute
Source: Image courtesy of Domo; www.domo.com

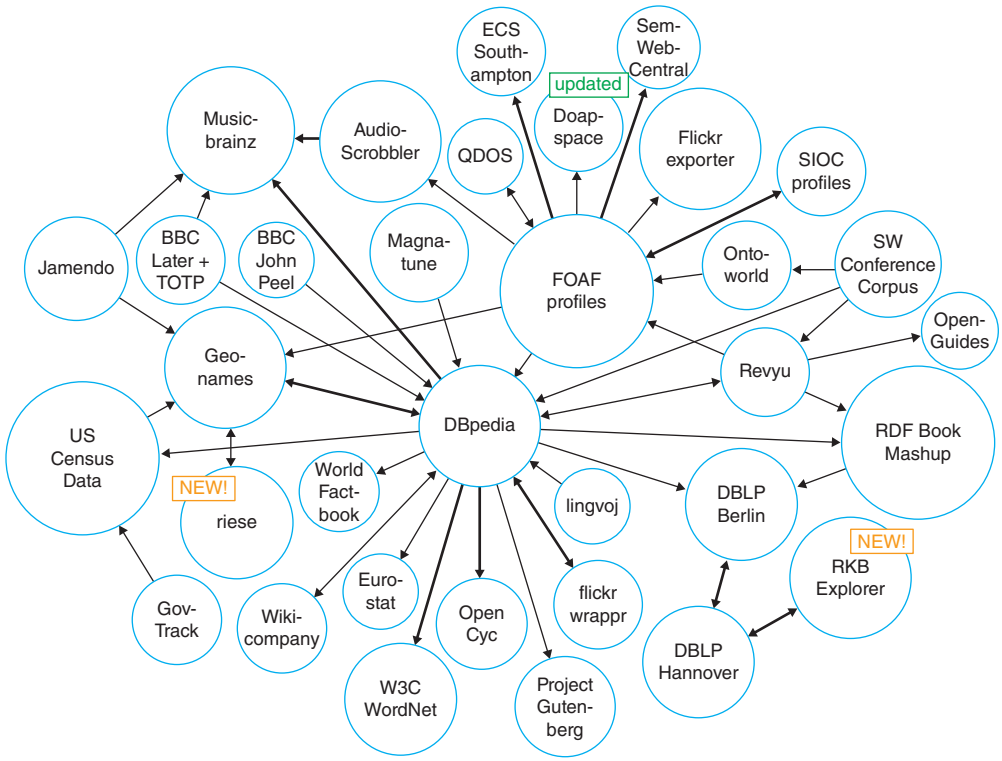


Figure 1.3 Examples of Mainstream Open Datasets as of 2008
Source: Richard Cyganiak, licensed under Creative Commons

Figure 1.3 represents a mere fraction of the open datasets currently available to anyone with an Internet connection and a desire to explore.*

Of course, the benefits of open data are not absolute. Unfortunately, and not surprisingly, many people use open data for malevolent purposes. For instance, consider Jigsaw, a marketplace that pays people to hand over other people's contact information. (I won't dignify Jigsaw with a link here.) As of this writing, anyone can download this type of data on more than 7 million professionals. Beyond annoying phone calls from marketers and incessant spam, it's not hard to imagine terrorist organizations accessing open data for nefarious purposes. Still, the pros of open data far exceed their cons.

THE BURGEONING DATA ECOSYSTEM

In the Introduction, I discussed how anyone could easily visualize their Facebook, Twitter, and LinkedIn data. I used Vizify to create an interesting visual profile of my professional life, but Vizify and its ilk are really just the tip of the

* For a larger visual of what's out there, see http://tinyurl.com/open-data-book.

iceberg. Through open APIs, scores of third parties can currently access that data and do some simply amazing things. For instance, MIT's Immersion Project lets Gmail users instantly visualize their e-mail connections.*

As far as I know, Google has no legal obligation to keep any of its APIs open. Nor is it compelled to regularly lease or license its user data or metadata to an entity. (Government edicts to turn over user data like the 2013 PRISM affair are, of course, another matter.) The company chooses to make this information available. If you're curious about what Google permits itself to do, check out its end-user license agreement (EULA).†

Perhaps Google will incorporate some of the Immersion Project's features or technology into Gmail or a new product or service. And maybe Google will do the same with other third-party efforts. Progressive companies are keeping tabs on how developers, partners, projects, and start-ups are using their products and services—*and the data behind them*. This is increasingly becoming the norm. As I wrote in *The Age of the Platform*, Amazon, Apple, Facebook, Google, Salesforce.com, Twitter, and other prominent tech companies recognize the significance of ecosystems and platforms, especially with respect to user data.

THE NEW WEB: VISUAL, SEMANTIC, AND API-DRIVEN

Since its inception, and particularly over the past eight years, the Web has changed in many ways. Perhaps most significantly to laypeople, it has become much more visual. Behind the scenes, techies like me know that major front-end changes cannot take place sans fundamental technological, architectural, and structural shifts, many of which are data driven.

The Arrival of the Visual Web

Uploading photos to the Web today is nearly seamless and instant. Most of you remember, though, that it used to be anything but. I'm old enough to remember the days of text-heavy websites and dial-up Internet service providers (ISPs) like Prodigy and AOL. Back in the late 1990s, most people connected to the Internet via glacially slow dial-up modems, present company included. Back then I could hardly contain my excitement when I connected at 56 kilobits per second. Of course, those days are long gone, although evidently AOL still counts nearly three million dial-up customers to this day.[2] (No, I couldn't believe it either.)

Think about Pinterest for a moment. As of this writing, the company sports a staggering valuation of $3.5 billion without a discernible business model—or

† A EULA establishes the user's or the purchaser's right to use the software.

at least a publicly disclosed one beyond promoted pins. As J.J. Colao writes on Forbes.com, "So how does one earn such a rich valuation without the operating history to back it up? According to Jeremy Levine, a partner at Bessemer Venture Partners who sits on Pinterest's board, the answer is simple: 'People love it.'"[3] (In case you're wondering, I'll come clean about Pinterest. It's not a daily habit, but I occasionally play around with it.*)

It's no understatement to say that we are infatuated with photos, and plenty of tech bellwethers have been paying attention to this burgeoning trend. On April 9, 2012, Facebook purchased Instagram, the hugely popular photo-sharing app. The price? A staggering $1 billion in cash and stock. At the time, Instagram sported more than 30 million users, but no proper *revenue*, let alone profits. Think Zuckerberg lost his mind? It's doubtful, as other titans like Google were reportedly in the mix for the app.

Thirteen months later, Yahoo CEO Marissa Mayer announced the much-needed overhaul of her company's Flickr app. Mayer wrote on the company's Tumblr blog, "We hope you'll agree that we have made huge strides to make Flickr awesome again, and we want to know what you think and how to further improve!"[4] And hip news-oriented apps like Zite and Flipboard are heavy on the visuals.

Facts like these underscore how much we love looking at photos, taking our own, and tagging our friends. Teenagers and *People* aficionados are hardly alone here. Forget dancing cats, college kids partaking in, er, "extracurricular" activities, and the curious case of Anthony Weiner. For a long time now, many popular business sites have included and even featured visuals to accompany their articles. It's fair to say that those without photos look a bit dated. Many *Wall Street Journal* articles include infographics. Many blog posts these days begin with featured images. Pure text stories seem so 1996, and these sites are responding to user demand. Readers today *expect* articles and blog posts to include graphics. Ones that do often benefit from increased page views and, at a bare minimum, allow readers to quickly scan an article and take something away from it.

Linked Data and a More Semantic Web

It's not just that data has become bigger and more open. As the Web matures and data silos continue to break down, data becomes increasingly interconnected. As recently as 2006, only a tiny percentage of data on the Web was linked to other data.† Yes, there were oodles of information online, but tying one dataset to another was often entirely manual, not to mention extremely challenging.

NOTE

Some degree of overlap exists among the terms *linked data* and *open data* (discussed earlier in this chapter). That is, some open data is linked and arguably most linked data is open, depending on your definition of the term. Despite their increasing intersection, the two terms should *not* be considered synonyms. As Richard MacManus writes on ReadWriteWeb, open data "commonly describes data that has been uploaded to the Web and is accessible to all, but isn't necessarily 'linked' to other data sets. [It] is available to the public, but it doesn't link to other data sources on the Web."[5]

Today we are nowhere near connecting all data. Many datasets cannot be easily and immediately linked to each other, and that day may never come. Still, major strides have been made to this end over the past eight years. The Web is becoming more semantic (read: more meaningful) in front of our very eyes. (David Siegel's book *Pull: The Power of the Semantic Web to Transform Your Business* covers this subject in more detail.)

The term *linked data* describes the practice of exposing, sharing, and connecting pieces of data, information, and knowledge on the semantic Web. Both humans and machines benefit when previously unconnected data is connected. This is typically done via Web technologies such as uniform resource identifiers* and the Resource Description Framework.†

A bevy of organizations—both large and small—is making the Web smarter and more semantic by the day. For instance, consider import.io, a U.K.-based start-up that seeks to turn webpages into tables of structured data. As Derrick Harris of GigaOM writes, the "service lets users train what [CEO Andrew] Fogg calls a 'data browser' to learn what they're looking for and create tables and even an application programming interface out of that data. The user dictates what attributes will comprise the rows and columns on the table, highlights them, and import.io's technology fills in the rest."[6]

The Relative Ease of Accessing Data

Yes, there is more data than ever, and many organizations struggle trying to make heads or tails out of it. Fortunately, however, all hope is not lost. The data-management tools available to organizations of all sizes have never been more powerful.

Prior to the Internet, most large organizations moved their data among their different systems, databases, and data warehouses through a process

* In computing, a uniform resource identifier is a string of characters used to identify a name or a Web resource. It should not to be confused with its two subclasses: uniform resource locator and uniform resource name.

† The Resource Description Framework is a family of World Wide Web Consortium specifications originally designed as a metadata data model.

known as *extract, transform, and load*, or *ETL*. Database administrators and other techies would write scripts or stored procedures to automate this process as much as possible. Batch processes would often run in the wee hours of the morning. At its core, ETL extracts data from System A, transforms or converts that data into a format friendly to System B, and then loads the data into System B. Countless companies to this day rely upon ETL to power all sorts of different applications. ETL will continue to exist in a decade, and probably much longer than that.

Now, ETL has had remarkable staying power in the corporate IT landscape. Today it is far from dead, but the game has changed. ETL is certainly not the only way to access data or to move data from Point A to Point B. And ETL is often not even the best method for doing so. These days, many mature organizations are gradually supplanting ETL with APIs. And most start-ups attempt to use APIs from the get-go for a number of reasons. Data accessed via APIs is optimized for consumption and access as opposed to storage.

In many instances, compared to ETL, APIs are just better suited for handling large amounts of data. In the words of Anant Jhingran, VP of products at enterprise software vendor Apigee:

> The mobile and apps economy means that the interaction with customers happens in a broader context than ever before. Customers and partners interact with enterprises via a myriad of apps and services. Unlike traditional systems, these new apps, their interaction patterns, and the data that they generate all change very rapidly. In many cases, the enterprise does not "control" the data. As such, traditional ETL does not and will not cut it.[7]

Jhingran is absolutely right about the power of—and need for—APIs. No, they are not elixirs, but they allow organizations to improve a number of core business functions these days. First, they provide access to data in faster and more contemporary ways than ETL usually does. Second, they allow organizations to (more) quickly identify data quality issues. Third, open APIs tend to promote a generally more open mind-set, one based upon innovation, problem solving, and collaboration. APIs benefit not only companies but their *ecosystems*—that is, their customers, users, and developers.

In the Twitter and Vizify examples in the Introduction, I showed how real-time data and open APIs let me visualize data without manual effort. In the process, I discovered a few things about my tweeting habits. Part III will provide additional examples of API-enabled data visualizations.

Greater Efficiency via Clouds and Data Centers

I don't want to spend too much time on it here, but it would be remiss not to mention a key driver of this new, more efficient Web: cloud computing. It is no

understatement to say that it is causing a tectonic shift in many organizations and industries.

By way of background, the history of IT can be broken down into three eras:

1. The Mainframe Era
2. The Client-Server Era
3. The Mobile-Cloud Era

Moving from one era to another doesn't happen overnight. While the trend is irrefutable, the mainframe is still essential for many mature organizations and their operations. They're called *laggards* for a reason. Over the foreseeable future, however, more organizations will get out of the IT business. Case in point: the propulsive success of Amazon Web Services, by some estimates a nearly $4 billion business *by itself*.[8] (Amazon frustrates many analysts by refusing to break out its numbers.) Put simply, more and more organizations are realizing that they can't "do" IT as reliably and inexpensively as Amazon, Rackspace, VMware, Microsoft Azure, and others. This is why clunky terms like *infrastructure as a service* and *platform as a service* have entered the business vernacular.

Students of business history will realize that we've seen this movie before. Remarkably, a century ago, many enterprises generated their own electricity. One by one, each eventually realized the silliness of its own efforts and turned to the utility companies for power. Nicholas Carr makes this point in his 2009 book *The Big Switch: Rewiring the World, from Edison to Google*. Cloud computing is here to stay, although there's anything but consensus after that.* For instance, VMware CEO Pat Gelsinger believes that it will be "decades" before the public cloud is ready to support all enterprise IT needs.[9]

Brass tacks: the Web has become much more visual, efficient, and data-friendly.

BETTER DATA TOOLS

The explosion of Big Data and Open Data did not take place overnight. Scores of companies and people saw this coming. Chief among them are some established software vendors and relatively new players like Tableau, Cloudera, and HortonWorks. These vendors have known for a while that organizations will soon need new tools to handle the Data Deluge. And that's exactly what they provide.

Over the past 15 years, we have seen marked improvement in existing business intelligence solutions and statistical packages. Enterprise-grade applications from MicroStrategy, Microsoft, SAS, SPSS, Cognos, and others have upped their games considerably.† Let me be clear: these products can without question do

* Even the definition of *cloud computing* is far from unanimous. Throw in the different types of clouds (read: public, semi-public, and private), and brouhahas in tech circles can result.

† IBM acquired both SPSS and Cognos, although each brand remains.

NOTE

Chapter 2 describes these new, more robust applications and services in much more detail.

more than they could in 1998. However, focusing exclusively on the evolution of mature products does not tell the full story. To completely understand the massive wave of innovation we've seen, we have to look beyond traditional BI tools. The aforementioned rise of cloud computing, SaaS, open data, APIs, SDKs, and mobility have collectively ushered in an era of rapid deployment and minimal or even zero hardware requirements. New, powerful, and user-friendly data-visualization tools have arrived. Collectively, they allow Visual Organizations to present information in innovative and exciting ways. Tableau is probably the most visible, but it is just one of the solutions introduced over the past decade.

Today, organizations of all sizes have at their disposal a wider variety of powerful, flexible, and affordable dataviz tools than ever. They include free Web services for start-ups to established enterprise solutions.

Equipped with these tools, services, and marketplaces, employees are telling fascinating stories via their data, compelling people to act, and making better business decisions. And, thanks to these tools, employees need not be proper techies or programmers to instantly visualize different types and sources of data. As you'll see in this book, equipped with the right tools, laypersons are easily interacting with and sharing data. Visual Organizations are discovering hidden and emerging trends. They are identifying opportunities and risks buried in large swaths of data. And they are doing this often without a great deal of involvement from their IT departments.

VISUALIZING BIG DATA: THE PRACTITIONER'S PERSPECTIVE

IT operations folks have visualized data for decades. For instance, employees in network ops centers normally use multiple screens to monitor what's taking place. Typically of great concern are the statuses of different systems, networks, and pieces of hardware. Record-level data was rolled into summaries, and a simple red or green status would present data in an easily digestible format.

This has changed dramatically over the past few years. We have seen a transformation of sorts. Tools like Hadoop allow for the easy and inexpensive collection of vastly more data than even a decade ago. Organizations can now maintain, access, and analyze petabytes of raw data. Next-generation dataviz tools can interpret this raw data on the fly for *ad hoc* analyses. It's now easy to call forth thousands of data points on demand for any given area into a simple webpage, spot anomalies, and diagnose operational issues *before* they turn red.[10]

Scott Kahler works as a senior field engineer at Pivotal, a company that enables the creation of Big Data software applications.

NOTE

Visual Organizations deploy and use superior dataviz tools and, as we'll see later in this book, create new ones as necessary.

GREATER ORGANIZATIONAL TRANSPARENCY

At the first Hackers' Conference in 1984, American writer Stewart Brand famously said, "Information wants to be free." That may have been true two or three decades ago, but few companies were particularly keen about transparency and sharing information. Even today in arguably most workplaces, visibility into the enterprise is exclusively confined to top-level executives via private meetings, e-mails, standard reports, financial statements, dashboards, and key performance indicators (KPIs). By and large, the default has been sharing only on a need-to-know basis.

To be sure, information hoarding is alive and well in Corporate America. There's no paucity of hierarchical, conservative, and top-down organizations without a desire to open up their kimonos to the rank and file. However, long gone are the days in which the idea of sharing data with employees, partners, shareholders, customers, governments, users, and citizens is, well, weird. These days it's much more common to find senior executives and company founders who believe that transparency confers significant benefits. Oscar Berg, a digital strategist and consultant for the Avega Group, lists three advantages of greater transparency:

1. Improve the quality of enterprise data
2. Avoid unnecessary risk taking
3. Enable organizational sharing and collaboration[11]

An increasing number of progressive organizations recognize that the benefits of transparency far outweigh their costs. They embrace a new default modus operandi of sharing information, not hoarding it. It's not hard to envision in the near future collaborative and completely transparent enterprises that give their employees—and maybe even their partners and customers—360-degree views of what's going on.

Even for organizations that resist a more open workplace, better tools and access to information are collectively having disruptive and democratizing effects, regardless of executive imprimatur. Now, I am not advocating the actions of PRISM leaker Edward Snowden. The former technical contractor-turned-whistleblower at Booz Allen Hamilton provided *The Guardian* with highly classified NSA documents. This, in turn, led to revelations about U.S. surveillance on cell phone and Internet communications. My only point is that today the forces advancing freedom of information are stronger than ever. Generally speaking, keeping data private today is easier said than done.

THE COPYCAT ECONOMY: MONKEY SEE, MONKEY DO

When a successful public company launches a new product, service, or feature, its competition typically notices. This has always been the case. For instance, Pepsi launched Patio Diet Cola in 1963, later renaming it Diet Pepsi. Coca-Cola countered by releasing Diet Coke in 1982. Pharmaceutical companies pay attention to one another as well. Merck launched the anti-cholesterol drug Zocor in January 1992. Four years later, the FDA approved Pfizer's Lipitor, a drug that ultimately became the most successful in U.S. history.

Depending on things like patents, intellectual property, and government regulations, launching a physical me-too product could take years. Mimicking a *digital* product or feature can often be done in days or weeks, especially if a company isn't too concerned with patent trolls.

In *The Age of the Platform*, I examined Amazon, Apple, Facebook, and Google—aka *the Gang of Four*. These companies' products and services have become ubiquitous. Each pays close attention to what the others are doing, and they are not exactly shy about "borrowing" features from one another. This copycat mentality goes way beyond the Gang of Four. It extends to Twitter, Yahoo, Microsoft, and other tech behemoths. For instance, look at what happened after the initial, largely fleeting success of Groupon. Soon after its enormous success, Amazon, Facebook, and Google quickly added their own daily deal equivalents. Also, as mentioned in the Introduction, Facebook introduced Twitter-like features in June 2013, like video sharing on Instagram, verified accounts, and hashtags.* Facebook's 1.2 billion users didn't have to do a thing to access these new features; they just automatically appeared.

Social networks aren't the only ones capable of rapidly rolling out new product features and updates. These days, software vendors are increasingly using the Web to immediately deliver new functionality to their customers. Companies like Salesforce.com are worth billions in large part due to the popularity of SaaS. As a result, it's never been easier for vendors to quickly deploy new tools and features. If Tableau's latest release or product contains a popular new feature, other vendors are often able to swiftly ape it—and get it out to their user bases. Unlike the 1990s, many software vendors today no longer have to wait for the next major release of the product, hoping that their clients upgrade to that version and use the new feature(s). The result: the bar is raised for everyone. Chapter 2 will cover data-visualization tools in much more depth.

DATA JOURNALISM AND THE NATE SILVER EFFECT

Elon Musk is many things: a billionaire, a brilliant and bold entrepreneur, the inspiration for the *Iron Man* movies, and a reported egomaniac. Among the companies (yes, plural) he has founded and currently runs is Tesla Motors.

* Facebook also borrowed trending topics in January of 2014.

Tesla is an electric-car outfit that aims to revolutionize the auto industry. Its Model S sedan is inarguably stunning but, at its current price, well beyond the reach of Joe Sixpack. Less certain, though, are Musk's grandiose performance claims about his company's chic electric vehicle.

New York Times journalist John Broder decided to find out for himself. In early 2013, he took an overnight test-drive up Interstate 95 along the U.S. eastern seaboard, precisely tracking his driving data in the process.*

On February 8, 2013, the *Times* published Broder's largely unflattering review of the Model S. In short, the reporter was not impressed. Chief among Broder's qualms was the "range anxiety" he experienced while driving. Broder claimed that the fully charged Model S doesn't go nearly as far as Musk and Tesla claim it does. The reporter worried that he would run out of juice before he made it to the nearest charging station. In Broder's words, "If this is Tesla's vision of long-distance travel in America's future . . . and the solution to what the company calls the 'road trip problem,' it needs some work."

A negative review published in the *New York Times* has legs; this wasn't a teenager's Tumblr account. Musk quickly went on the offensive, attempting to prove that Broder got it wrong. Musk's smoking gun: the data—sort of. In a piece for the Tow Center for Digital Journalism (an institute within Columbia University's Graduate School of Journalism), Taylor Owen wrote that, "Tesla didn't release the data from the review. Tesla released [its] *interpretation* of the data from the review."[12] [Emphasis mine.]

Musk appeared on a number of television shows to plead his case and to question Broder's ability to follow simple instructions. Broder retaliated in a separate *Times* piece. The story blew over after a few days, but its impact has been anything but ephemeral. If this was hardly the first kerfuffle between a journalist and a public figure, then what was special about this one? In short, it was the cardinal role that data played in the dispute. Both Musk and Broder tried to *prove* their positions by using data.

Broder is one of an increasing cadre of high-profile reporters taking a more data-oriented approach to journalism these days. *Bloomberg Businessweek* formally refers to some of its staff as *data journalists*. *New York Times* Op-Ed columnist David Brooks has written extensively about the impact, benefits, and limitations of Big Data. But if there's a poster boy for contemporary data journalism, he goes by the name of Nate Silver.

In 2009, *Time* named the thirty-something statistician, pundit, and blogger as one of the most influential people on the planet. A picture of the wunderkind is presented in Figure 1.4.

From 2010 to July 2013, the *New York Times* licensed and hosted his blog FiveThirtyEight, and the results were nothing short of staggering. For instance,

* Read the entire review here: http://tinyurl.com/broder-tesla.

Figure 1.4 Nate Silver Speaking at SXSWi in 2009
Source: Randy Stewart, Seattle, WA, USA[13]

on the Monday before the 2012 U.S. presidential election, more than 20 percent of all visitors to the *Times* website read some of Silver's musings. Predictably (pun intended), his 2012 book *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't* quickly became a bestseller.

In his writing, Silver frequently uses data, statistical models, charts, and graphs on a veritable bouillabaisse of topics. Beyond politics, Silver opines about subjects like these:

- Blogging: "The Economics of Blogging and *The Huffington Post*"
- Hockey: "Why Can't Canada Win the Stanley Cup?"
- Baseball: "Money on the Bench"
- Basketball: "Heat's Clutch Stats Meet Match in Spurs' Strategy"

Although the subjects change, the general methodology does not. Silver's readers observe firsthand how he uses data to support his hypotheses so convincingly, although he has his detractors. Many FiveThirtyEight fans read Silver's data-driven articles while commuting or at work. When making arguments to their own bosses and colleagues, it's likely that Silver's thought process and articles persuade them to use data and dataviz as well.

In early 2013, Silver spoke to an audience at Washington University about what Max Rivlin-Nadler of *Gawker* described as the "statistical pitfalls

of accruing such a large following." After the presidential election, Silver had become so popular that he was starting to exert considerable influence over the democratic process. For some time, Silver wondered if he should do the unthinkable: cease blogging, at least about politics and elections. "I hope people don't take the forecasts too seriously," Silver said in February 2013. "You'd rather have an experiment where you record it off from the actual voters, in a sense, but we'll see. If it gets really weird in 2014, in 2016, then maybe I'll stop doing it. I don't want to influence the democratic process in a negative way."[14]

The possibility of Silver leaving the *Times* became a reality on July 19, 2013. Silver announced that he was taking a position with ESPN in August of that year. (Undue influence was unlikely the sole factor in Silver's decision; he probably attracted a colossal contract, and his love of sports is well documented.) In a statement released by ESPN, Silver said, "This is a dream job for me. I'm excited to expand FiveThirtyEight's data-driven approach into new areas, while also reuniting with my love of sports. I'm thrilled that we're going to be able to create jobs for a great team of journalists, writers, and analysts. And I think that I've found the perfect place to do it."[15]

In a way, however, Silver's departure changes nothing. No doubt that the popularity and data-driven style of his writing will continue to influence many current and future journalists throughout the world.

DIGITAL MAN

For a bunch of reasons covered in this chapter, we have become increasingly more comfortable with—and reliant upon—data and technology in our daily lives. It seems that we are almost always tethered to devices of one kind or another. This section explains the arrival of the digital man and how it has led to the Visual Organization. To summarize, as citizens, we have become more tech savvy, and not just at home. We take this newfound knowledge into the workplace. If our current employer isn't providing us with the tools we need to do our jobs, many of us will just bring our own.

The Data Deluge is transforming many areas of our lives, including journalism. To be sure, there will still be disputed stories that ultimately hinge upon "he said, she said." More and more, however, data will be able to tell more of the story.

The Arrival of the Visual Citizen

Although precise statistics are hard to come by, social networks and blogging platforms have exploded. There's no government agency that releases official or validated government statistics. For instance, Google claims that more than 400 million people use Google Plus, but it's important to take that claim with more than a grain of salt. Numbers like these are bogus for many reasons.

First, consumer-facing companies face a strong incentive to exaggerate their reported users. Next, it's not difficult for people, groups, and enterprises to create multiple accounts on any network. For instance, I created and actively manage four separate Twitter handles, each with a different purpose:

1. `@philsimon`: my main handle*
2. `@motionpub`: the handle for my small publishing company
3. `@thenewsmall`: the handle for my third book
4. `@newsmallapp`: the handle for my app, based upon the third book

At least I'm a human being, though. Fake handles are rampant. Fortunately, services like ManageFlitter allow me to detect Twitter handles likely run by spambots. With a few clicks, I can remove them *en masse*.

Even if we ignore those considerations, we're still not out of the woods yet. The question of what represents an "active user" is open to wide interpretation. The term is fuzzy; there's no universally accepted way to define it, much less monitor it for accuracy. Are you considered active if you create an account? If you log in every week? Every month? If Google automatically creates a Plus account when users sign up for Gmail, are they active users even if they never set up circles or "+1" anything?

I don't have answers to these questions and, for our purposes, exactitude doesn't matter. Social networks are huge, and hundreds of millions of us legitimately spend a great deal of time on them. In the process, we generate and consume oodles of data. As we do this, we become more comfortable with it. Increasingly these networks are presenting their data in visual, interactive formats. Mark Zuckerberg, LinkedIn CEO Reed Hoffman, Twitter boss Dick Costolo, and others know that we don't want to stare at raw data in spreadsheets any more than we have to. We do enough of that at work. We prefer to view data in more visual ways. This is why these sites allow us to easily see the page views, impressions, or "engagement" of our status updates, posts, videos, and photos. This type of data makes many of us more likely to buy Facebook or Twitter ads, sponsor a story, and promote a tweet.

In the Introduction, I described how Twitter allows users to easily represent their tweets via interactive and visual means. On October 29, 2013, Twitter went even further. The company announced that it was making its timeline decidedly more visual via an update that "will insert previews of images and Vines directly into tweets on the web and in Twitter's iOS and Android apps. To see the entire image or Vine, just tap on it."[16]

It would be folly, however, to claim that Twitter and its ecosystem collectively hold a monopoly on data visualization for social networks. Nothing could

* True story: I inadvertently deleted `@philsimon` a few years ago and started a `#savephilsimon` campaign under `@philsimon2`. It worked. The Twitter powers that be gave me `@philsimon back`.

be further from the truth. Facebook's famous social graph provides a visual means for both users and the company to see who is connected to whom. Zuckerberg understands that contemporary dataviz requires the deployment of *structurally different* technologies. (Behind the scenes, the social graph utilizes a graph database, not a relational designed for processing backend operations.* This difference is anything but semantic. Graph databases assume that the relationships are as important as the records.[17])

For its part, LinkedIn launched a major redesign of its member profiles in June 2013.† The goals were twofold: to make profiles simpler and decidedly more visual. Premium members can access more advanced analytics, many of which are visualized. Rather than show my own profile again, let's look at part of the new profile of Mark Kelly, the talented keyboardist for the English progressive rock band Marillion. It is shown in Figure 1.5.

A look at Figure 1.5 reveals the obvious: Kelly is endorsed most frequently for his expertise in the music industry. Aside from his duties in Marillion, Kelly serves as head of the Featured Artists Coalition, a group that campaigns for the protection of performers' and musicians' rights. As of this writing, 18 people have endorsed Kelly for this skill, including yours truly. Clicking on any image on the right takes the user to that endorser's profile.

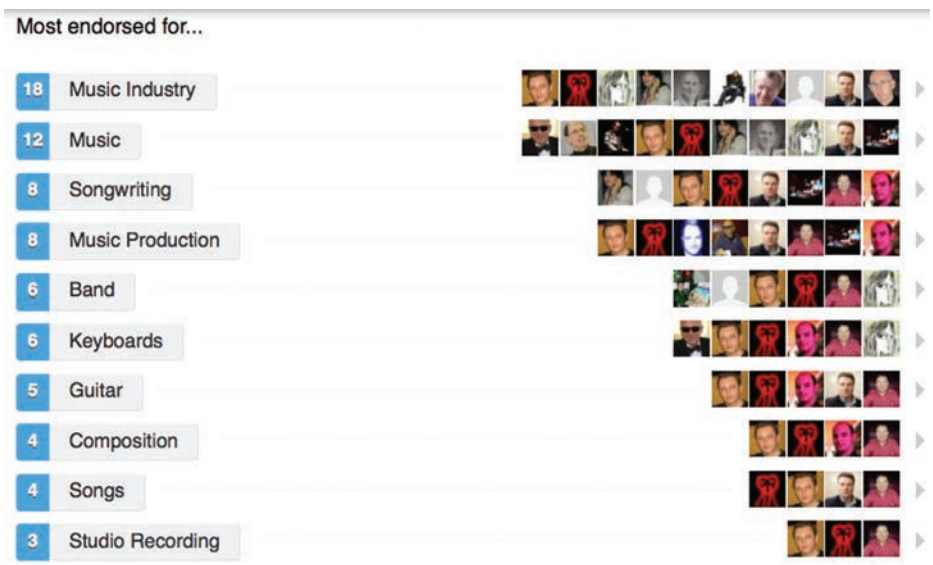


Figure 1.5 LinkedIn Endorsements of Marillion Keyboardist Mark Kelly
Source: LinkedIn

* For more on this, see http://tinyurl.com/fb-graph2.

† In November of 2013, Netflix relaunched its home page, making it decidedly more visual.

LinkedIn's recent redesign reflects a much larger Web trend. Today, most prominent social networks provide users with access to free, powerful, and increasingly visual analytics and data. Examples include Pinterest, Google, and Facebook.

Today, laypeople are looking at and working with data more than ever. More and more business decisions now require data. To make sense of it, data needs to become visual. Dataviz is becoming the norm. The LinkedIn redesign was a case in point. Without effective dataviz, how can we cope with the Data Deluge?

Mobility

It would be remiss here to ignore the enormous impact that mobility has had on our data-consuming and -generation habits. To be sure, the rise of smartphones, apps, and near-constant communication has driven an increase in both the supply of—and demand for—data. Over the past five years, we have seen the explosion of tablets and other touch-based devices. While not the only show in town, the iPad reigns supreme, with more than 100 million units sold as of this writing.[18]

The full political, economic, and cultural impact of mobility is way beyond the scope of this book. For now, suffice it to say that, more than ever, mobility has made data more pervasive, visual, and even touchable.

The Visual Employee: A More Tech- and Data-Savvy Workforce

In his 2008 book *Grown Up Digital: How the Net Generation Is Changing Your World*, Don Tapscott discusses how today's young people are using technology in fascinating and unprecedented ways. Yes, there are slackers within such a large group; it's not as if they all spend their time passively watching television, texting, and eating Cheetos. Rather, "Net Geners" are constantly active. Tethered to their smartphones, they are almost always viewing, creating, and distributing data in one form or another via Twitter, Facebook, Snapchat, Vine, YouTube, Instagram, and a host of other apps and sites. They aren't just constantly consuming information; they are actively generating lots of it. As a group, Millennials are extremely proficient with gadgets.

But don't think for a minute that Millennials are the only ones interacting with data and technology on a near-constant basis. Yes, legitimate differences among generations exist. (Don't they always?) But the consumerization of IT has ushered in an entirely new and tech-centric era. We are *all* becoming more tech- and data-savvy, not to mention fidgety. According to a 2012 Pew survey of 2,254 people, 52 percent of *all* cell phone owners said they had used their mobile devices to do a variety of things while watching TV.[19]

NOTE

As a group, consumers are becoming much more familiar with—and skilled at—using, interpreting, and representing data. Tech-savvy citizens take these skills and this data-oriented mind-set with them to work. They don't leave their brains at the door. This is causing increasing friction in organizations tied to "data-free" ways of doing things.

Navigating Our Data-Driven World

Knowing that better dataviz tools exist only gets us so far. For any organization to be as successful as possible, all of its employees need to step up. Truly understanding today's data streams requires more than just purchasing, downloading, and creating dataviz tools. Employees must actually use them.

Fortunately, there's never been greater access to user-friendly and powerful dataviz applications. The past ten years have brought about a much more democratic technology ethos into the workplace. Many employees no longer take as a given that they have to use only programs endorsed by IT. Sure, many organizations still cling to restrictive policies about the applications that employees *officially* can and can't use while on the clock. This hostility to "nonsanctioned" technologies, however, is starting to wane. Over the past decade, we've seen the rise of the Freemium model, BYOD (bring your own device), Web-based services, and open-source software. The success of workplace social networks like Yammer (acquired by Microsoft for $1.2 billion in June of 2012*) underscores a critical trend: in many organizations, the adoption of new technologies is becoming much more organic and bottom up, especially compared to the mid-1990s.

As mentioned earlier in this chapter, employees today are increasingly tech savvy. If they are dissatisfied with their employer's current applications and systems, they can and often will look elsewhere for superior alternatives. This is true with respect to many technologies, and dataviz is no exception to this rule. What's more, in most instances, there's not a great deal that IT can realistically do about employees "flying under the radar."

> We are all becoming more comfortable with data. Data visualization is no longer just something we have to do at work. Increasingly, we want to do it as consumers and as citizens. Put simply, visualizing helps us understand what's going on in our lives—and how to solve problems.

A simple Google search on "best free data-visualization tools" may confirm what skeptical employees have long suspected: that their employers are a little behind the times and better options are available. This "use whatever tools are needed" mind-set is particularly pronounced at small businesses and start-ups.

* At the time, Microsoft already sold SharePoint, a workplace social network of sorts. Unlike Yammer, SharePoint needed to be deployed in a top-down manner. In this sense, it was the antithesis of Yammer.

NEXT

Chapter 2 looks at the specific dataviz applications, services, and tools that Visual Organizations are using. We'll see that the new boss isn't the same as the old boss.

NOTES

1. Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. (2007). "DBpedia: A Nucleus for a Web of Open Data." *The Semantic Web*. Lecture Notes in Computer Science 4825. p. 722.
2. Aguilar, Mario, "3 Million Suckers Still Pay for AOL Dial-Up," Gizmodo, July 27, 2012, http://gizmodo.com/5929710/3-million-suckers-still-pay-for-aol-dial+up, Retrieved August 27, 2013.
3. Colao, J.J., "Why Is Pinterest a $2.5 Billion Company? An Early Investor Explains," *Forbes*, May 8, 2013, http://www.forbes.com/sites/jjcolao/2013/05/08/why-is-pinterest-a-2-5-billion-company-an-early-investor-explains, Retrieved August 27, 2013.
4. Mayer, Marissa, "Your World, in Full Resolution," Yahoo's Tumblr, May 20, 2013, http://yahoo.tumblr.com/post/50934634700/your-world-in-full-resolution, Retrieved August 27, 2013.
5. MacManus, Richard, "It's All Semantics: Open Data, Linked Data & The Semantic Web," readwrite.com, March 31, 2010, http://readwrite.com/2010/03/31/open_data_linked_data_semantic_web, Retrieved August 26, 2013.
6. Harris, Derrick, "Import.io Wants to Help Turn Web Pages into Data—Fast," Gigaom, August 29, 2013, http://gigaom.com/2013/08/29/import-io-wants-to-help-turn-web-pages-into-data-fast, Retrieved August 29, 2013.
7. Jhingran, Anant, "From ETL to API—A Changed Landscape for Enterprise Data Integration," Apigee blog, October 10, 2012, https://blog.apigee.com/detail/from_etl_to_api_a_changed_landscape_for_enterprise_data_integration, Retrieved, June 27, 2013.
8. Dignan, Larry, "Amazon's AWS: $3.8 Billion Revenue in 2013, Says Analyst," ZDNet, January 7, 2013, http://www.zdnet.com/amazons-aws-3-8-billion-revenue-in-2013-says-analyst-7000009461, Retrieved August 28, 2013.
9. Hiner, Jason, "We've Entered the Third Generation of IT, Says Vmware," TechRepublic, August 27, 2013, http://www.techrepublic.com/blog/tech-sanity-check/weve-entered-the-third-generation-of-it-says-vmware, Retrieved August 28, 2013.
10. Personal conversation with Kahler, September 25, 2013.
11. Berg, Oscar, "3 Reasons Why Organizations Need to Increase Transparency," CMS Wire, July 5, 2011, http://www.cmswire.com/cms/

enterprise-collaboration/3-reasons-why-organizations-need-to-increase-transparency-011886.php, Retrieved July 20, 2013.

12. Owen, Taylor, "What the Tesla Affair Tells Us About Data Journalism," Tow Center blog, February 21, 2013, http://www.towcenter.org/blog/what-the-tesla-affair-tells-us-about-data-journalism, Retrieved September 15, 2013.
13. Nate Silver—SXSWi 2009. [CC-BY-SA-2.0 (http://creativecommons.org/licenses/by-sa/2.0)], via Wikimedia Commons http://commons.wikimedia.org/wiki/File:Nate_Silver_2009.png.
14. Rivlin-Nadler, Max, "Nate Silver Might Stop Blogging if He Starts to 'Influence the Democratic Process,'" Gawker, February 16, 2013, http://www.gawker.com/477097844, Retrieved June 21, 2013.
15. "Nate Silver Makes Move to ESPN," July 22, 2013. http://espn.go.com/espn/story/_/id/9499752/nate-silver-joins-espn-multi-faceted-role, Retrieved July 22, 2013.
16. Newton, Casey, "Twitter Timeline Becomes More Visual with Previews of Images and Vines," The Verge, October 29, 2013, http://www.theverge.com/2013/10/29/4848184/twitter-timeline-becomes-more-visual-with-previews-of-images-and-vines, Retrieved October 29, 2013.
17. Neo4j, "Social Networks in the Database: Using a Graph Database," Neo4j blog, September 15, 2009, blog.neo4j.org/2009_09_01_archive.html, Retrieved September 6, 2013.
18. Statistic Brain, "Apple Computer Company Statistics," September 22, 2012, http://www.statisticbrain.com/apple-computer-company-statistics, Retrieved June 12, 2013.
19. Reardon, Marguerite, "Trend watch: We're using our cell phones while watching TV," CNET, July 17, 2012, http://news.cnet.com/8301-1035_3-57473899-94/trend-watch-were-using-our-cell-phones-while-watching-tv/.