

# Analytics


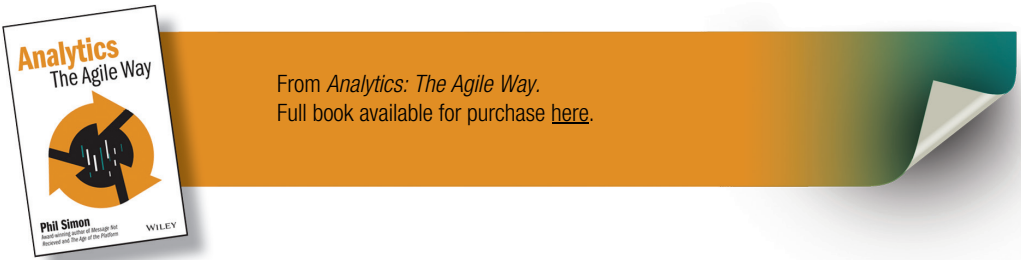
## The Agile Way



**Phil Simon**

Award-winning author of *Message Not Recieved* and *The Age of the Platform*

**WILEY**



From *Analytics: The Agile Way*.  
Full book available for purchase [here](#).

# Contents

**Preface: The Power of Dynamic Data    xvii**

List of Figures and Tables xxvii

## Introduction: It Didn't Used to Be This Way 1

A Little History Lesson 2Analytics and the Need for Speed 5Book Scope, Approach, and Style 9Intended Audience 12Plan of Attack 13Next 14Notes 14

## Part One Background and Trends ..... 17

Chapter 1 Signs of the Times: Why Data and Analytics Are Dominating Our World 19The *Moneyball* Effect 20Digitization and the Great Unbundling 22Amazon Web Services and Cloud Computing 24Not Your Father's Data Storage 26Moore's Law 28The Smartphone Revolution 28The Democratization of Data 29The Primacy of Privacy 29The Internet of Things 31The Rise of the Data-Savvy Employee 31The Burgeoning Importance of Data Analytics 32Data-Related Challenges 40Companies Left Behind 41The Growth of Analytics Programs 42Next 43Notes 43

Chapter 2 The Fundamentals of Contemporary Data: A Primer  
on What It Is, Why It Matters, and How to Get It 45

Types of Data 46

Getting the Data 52

Data in Motion 61

Next 63

Notes 63

Chapter 3 The Fundamentals of Analytics: Peeling Back the  
Onion 65

Defining Analytics 66

Types of Analytics 69

Streaming Data Revisited 72

A Final Word on Analytics 74

Next 75

Notes 75

**Part Two Agile Methods and Analytics.....77**

Chapter 4 A Better Way to Work: The Benefits and Core Values  
of Agile Development 79

The Case against Traditional Analytics Projects 80

Proving the Superiority of Agile Methods 82

The Case for Guidelines over Rules 84

Next 88

Notes 88

Chapter 5 Introducing Scrum: Looking at One of Today's Most  
Popular Agile Methods 89

A Very Brief History 90

Scrum Teams 91

User Stories 94

Backlogs 97

Sprints and Meetings 98

Releases 101

Estimation Techniques 102

Other Scrum Artifacts, Tools, and Concepts 109

Next 112

Chapter 6 A Framework for Agile Analytics: A Simple Model for  
Gathering Insights 113

Perform Business Discovery 115

Perform Data Discovery 117

Prepare the Data	118
Model the Data	120
Score and Deploy	127
Evaluate and Improve	128
Next	130
Notes	130

## **Part Three Analytics in Action ..... 131**

### **Chapter 7 University Tutoring Center: An In-Depth Case Study on Agile Analytics 133**

The UTC and Project Background	134
Project Goals and Kickoff	136
Iteration One	139
Iteration Two	140
Iteration Three	145
Iteration Four	146
Results	147
Lessons	148
Next	148

### **Chapter 8 People Analytics at Google/Alphabet: Not Your Father's HR Department 149**

The Value of Business Experiments	150
PiLab's Adventures in Analytics	151
A Better Approach to Hiring	153
Staffing	156
The Value of Perks	158
Results and Lessons	162
Next	162
Notes	163

### **Chapter 9 The Anti-Google: Beneke Pharmaceuticals 165**

Project Background	166
Business and Data Discovery	167
The Friction Begins	168
Astonishing Results	169
Developing Options	171
The Grand Finale	172
Results and Lessons	173
Next	174

Chapter 10 Ice Station Zebra Medical: How Agile Methods  
Solved a Messy Health-Care Data Problem 175

Paying Nurses 176  
Enter the Consultant 178  
User Stories 179  
Agile: The Better Way 182  
Results 183  
Lessons 183  
Next 184

Chapter 11 Racial Profiling at Nextdoor: Using Data to Build a  
Better App and Combat a PR Disaster 185

Unintended but Familiar Consequences 187  
Evaluating the Problem 188  
Results and Lessons 193  
Next 195  
Notes 195

**Part Four Making the Most Out of Agile Analytics ..... 197**

Chapter 12 The Benefits of Agile Analytics: The Upsides of  
Small Batches 199

Life at IAC 200  
Life at RDC 203  
Comparing the Two 206  
Next 206

Chapter 13 No Free Lunch: The Impediments to—and  
Limitations of—Agile Analytics 209

People Issues 210  
Data Issues 212  
The Limitations of Agile Analytics 216  
Next 219

Chapter 14 The Importance of Designing for Data: Lessons from  
the Upstarts 221

The Genes of Music 222  
The Tension between Data and Design 226  
Next 229  
Notes 229

**Part Five Conclusions and Next Steps .....231****Chapter 15 What Now?: A Look Forward 233**

A Tale of Two Retailers 234

The Blurry Futures of Data, Analytics, and Related Issues 239

Final Thoughts and Next Steps 242

Notes 243

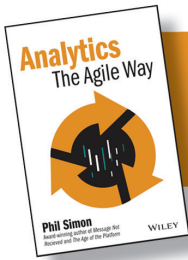
**Afterword 245****Acknowledgments 247****Selected Bibliography 249**

Books 249

Articles and Essays 251

**About the Author 253****Index 255**

From *Analytics: The Agile Way* by Phil Simon. Copyright © 2017, SAS Institute Inc., Cary, North Carolina, USA.  
ALL RIGHTS RESERVED.



From *Analytics: The Agile Way*.  
Full book available for purchase [here](#).

# 1

## CHAPTER

# Signs of the Times

## *Why Data and Analytics Are Dominating Our World*

*Technology is eating the world.*

—Marc Andreessen, August 20, 2011

On August 20, 2011, the ex-Netscape founder and current rock-star venture capitalist uttered these five words—perhaps the most telling and quoted words of the Internet age. In a nutshell, technology has spawned powerful new companies and industries and decimated others. It has led to revolutions, unprecedented wealth, and new social mores and change that many institutions and individuals are barely beginning to process.

I am in the privileged position to have lived through all of this; consequently, I can wax poetic about things to which students probably cannot relate. (In a few of my books, I have done just that.) Yes, I remember getting my first e-mail as a sophomore at Carnegie Mellon in 1991. (I was blown away.) I too once thought that entering your credit card information into a computer was downright weird. I recall telephone booths, answering machines, flip phones, primitive web browsers, search results that weren't remotely accurate, when Napster was a thing, and nascent social networks such as Friendster that went down more often than they stayed up.

This is not a book about technology per se; it is a book about one of the most important consequences of ubiquitous technology: the explosion of data and the practice of analytics. Make no mistake: These are direct descendants of our tech-centric times. Absent the arrival of the World Wide Web, the smartphone explosion, cheap data storage, and the digitization of books, newspapers, songs, photos, and more, analytics and data wouldn't be nearly as critical as they are today. This chapter looks at those trends.

## THE *MONEYBALL* EFFECT

Billy Beane attained fame in baseball and analytics circles long before Brad Pitt portrayed him in the 2011 film *Moneyball*. In fact, Beane was making quite the name for himself even prior to Michael Lewis's 2004 book of the same name.

It's no overstatement to claim that as general manager of the small-market Oakland A's, Beane changed the game of baseball forever. Big-market powerhouses such as the New York Yankees, Boston Red Sox, and Los Angeles Dodgers could effectively print their own money. Not Beane. He had to compete with a relatively paltry annual budget of roughly \$60 million. That meant that he couldn't even dream of chasing other teams' pricey free agents. In fact, he had to let many of his own stars walk.\* Case in point: Beane had no

---

\*One of my favorite quotes from the movie: "The problem we're trying to solve is that there are rich teams and there are poor teams. Then there's fifty feet of crap, and then there's us."



shot of re-signing all-star first baseman Jason Giambi in 2001. The slugger and later admitted steroid user upped with the Yankees for nearly \$120 million over seven years. Beane couldn't justify spending nearly 30 percent of his budget on a single player—no matter how prolific.

Instead, Beane proved that necessity is the mother of invention. He famously plucked players off other teams' scrap heaps, especially if they possessed odd skills. Player X can't hit home runs? No problem. Can he frustrate opposing pitchers by being "a tough out"? Can he just get on base? Beane drafted players who "just didn't look" like effective baseball players. His unorthodox methods angered many longtime Oakland scouts, men who had spent their careers watching players and developing a supposed eye for talent, not staring at spreadsheets. (The case study in Chapter 9 will have much more to say about resistance to analytics.)

You probably know how this story turns out. Would Michael Lewis write a book about you that turns into a movie starring Brad Pitt if you failed miserably? Pretty soon, even big-market teams such as the Yankees and Red Sox began hiring their own analytics experts, and later, *teams* of experts. The *Moneyball* movement spread beyond baseball to all other major sports. In fact, analytics are starting to move from the back room to the field. In their 2016 book *The Only Rule Is It Has to Work: Our Wild Experiment Building a New Kind of Baseball Team*, Ben Lindbergh and Sam Miller wrote:

The Denver Broncos of the National Football League announce[d] that the team's director of analytics, Mitch Tanney, will break the front-office fourth wall in their upcoming games, speaking on a headset to head coach Gary Kubiak to offer his input on which plays the probabilities and percentages support. As managers increasingly come from cohorts that tend to be more perceptive to sabermetrics, it seems inevitable that something similar will happen in baseball.

Today, you'll find examples of *Moneyball* in areas and industries without any ties to sports at all. Case in point: the U.S. justice system.

In 2007, Anne Milgram became the attorney general of New Jersey. In her words, she “wanted to moneyball criminal justice.”\* Yes, *moneyball* is now a verb.

## DIGITIZATION AND THE GREAT UNBUNDLING

Not *that* long ago, it was impossible to buy an individual song in a digital format. One could either hunt for the seven-inch 45 rpm record or just buy the CD—often at the exorbitant price of \$15.99. The same held true for individual articles in newspapers and magazines, unless you wanted to read one at a newsstand.

Put differently, the Internet has created plenty of new industries, but disrupted or “unbundled” plenty of others. Writing for the *Harvard Business Review* in 2014, Justin Fox noted:

Much of the business story of the digital age so far has been about taking products and institutions apart—unbundling them. Music CDs were unbundled into MP3s that were sold (and illicitly downloaded) individually. Newspapers have been unbundled by blogs and classified ad sites. Now, digital-education upstarts are trying to unbundle the university.<sup>1</sup>

And this trend has only intensified. In April 2015, HBO became the first cable network to offer its service independent of cable TV providers.<sup>†</sup> Dubbed HBO NOW, the service, for a flat monthly fee of \$14.99, allows consumers to watch award-winning shows such as *Veep*, *Curb Your Enthusiasm*, and *Last Week Tonight* on their Apple TVs. No longer do customers need to buy pricey and wasteful cable packages from Verizon, Comcast, DIRECTV, and Cox Communications. Many industry analysts saw the move as an inevitable attempt to curtail piracy and lure tech-savvy cord-cutters and “cord-nevers” who consider the

---

\* See Milgram’s TED talk at <http://bit.ly/2ndikqg>.

<sup>†</sup> HBO announced the service at a much-ballyhooed Apple event in San Francisco at the Yerba Buena Center on March 9, 2015.

idea of paying \$120 per month for a full-fledged cable TV subscription laughable. To goose subscriptions, HBO offered a free one-month trial to boot.

In several ways, this was a harbinger of things to come. HBO soon made its new service available on other “over-the-top” (OTT) devices such as Chromecast, Roku, and iOS and Android phones. It didn’t take long for other cable TV channels to follow HBO’s lead. Showtime Anytime launched on July 8, 2015, for \$10.99 per month. In February 2017, AT&T fell in line with a similar offering for Starz and Starz Encore for \$8 per month.

Now fans can legally binge-watch *Game of Thrones* at a fraction of what it would have cost two years ago.\* Still, there’s a more important benefit here. Think about the data implications of this move. Prior to launching HBO Now, the company did not maintain direct relationships with its customers. It needed to rely on intermediaries—in this case, cable companies—and pay for the privilege. Under the old model, HBO needed to compensate the Verizons of the world. Even worse, HBO could not directly and easily ascertain which shows its viewers were watching, when, and on which device. The cable networks can now do this—at least for their customers using apps via OTT devices. In other words, they can now do what Netflix has successfully done for years.

## STEVE JOBS, THE *NEW YORK TIMES*, AND THE DATA WARS

Apple advocates customer privacy and often takes shots at Google/Alphabet for its “privacy-challenged” business model. As the following story demonstrates, though, it’s folly to claim that Apple doesn’t “get” the import of data.

(Continued)

---

\*I am intentionally oversimplifying things here. For years, one could purchase individual seasons of popular shows via Apple’s iTunes, Amazon’s Prime Video, and other services. I gladly pony up \$20 for seasons of AMC’s *Better Call Saul*. In point of fact, I’d pay \$20 per episode.

(Continued)

On April 3, 2010, Apple CEO Steve Jobs, to much hubbub, announced the launch of his company's latest game-changing product: the iPad. Consumers flocked to the device, although the product ultimately has not approached the iPhone's lofty sales figures. (As of November 2016, Apple had sold nearly 340 million iPads<sup>2</sup>, making untold billions in the process.)

This was the first tablet that really mattered, and Jobs knew that he held tremendous power over potential partners. Not long after the launch, major media companies scurried to develop apps. In particularly cantankerous negotiations with the *New York Times*, Jobs insisted that Apple “own” the relationship with the customer and collect a fee for the privilege. That is, customers would subscribe to the *Times*'s new app either through Apple's app store or not at all.

Apple was flush with cash. It did not need a few dollars from the *Times*'s subscribers. This wasn't about money; it was about control and data. Jobs understood the cardinal importance of customer data, so much so that he was willing to anger an iconic newspaper and its loyal subscribers.

### **? TIP**

In business there have always been two ways to make money: to bundle and to unbundle. Right now we're living in an age of unbundling. In the process, data and analytics provide companies with an unprecedented opportunity to understand their customers, employees, products, and users.

---

## **AMAZON WEB SERVICES AND CLOUD COMPUTING**

Up until 2006, most enterprises found it time-consuming and expensive to deploy new applications. Hardware and infrastructure vendors such as IBM largely charged their clients substantial installation and maintenance fees. Getting up and running was no small endeavor.

In 2006, one company started to change that.

Amazon launched a tiny experiment named Amazon Web Services (AWS). To run its core business, Amazon required powerful data

centers, servers, and infrastructure, but even the e-commerce juggernaut didn't use all of that capacity. Why let all of that excess compute power vanish into the ether? Why not rent its powerful infrastructure to companies as a service? It was a revolutionary and extremely profitable idea that the tech establishment didn't recognize until Amazon had built a sizable—perhaps even insurmountable—lead.

The AWS model did two critical things. First, it allowed companies of all sizes to *virtually* run their own applications, significantly reducing deployment times in the process. Second, AWS did not charge its clients flat fees. Rather, costs varied in direct proportion to usage. In other words, its pricing model was flexible or elastic. This appealed to many CIOs whose IT budgets plunged in the wake of the 2008 financial crisis. Organizations could now seriously consider whether they wanted to exit the IT business and “stick to their knitting.” To this end, in 2008, Nicholas Carr penned the influential book *The Big Switch: Rewiring the World, from Edison to Google*. A century ago, most large companies produced their own electricity. Was a similar computing switch inevitable?

To call AWS *successful* is the acme of understatement. Amazon invented a multibillion-dollar product category. Its offerings currently go by clunky monikers such as Platform as a Service and Infrastructure as a Service. Although Amazon has spawned prominent imitators in the form of Microsoft, Google, and IBM, it continues to hold the pole position by a substantial margin. What's more, growth will continue for the foreseeable future. As Charles Babcock wrote on *InformationWeek*, AWS

continues to grow at a 50% per year pace. It's headed for \$13 billion in revenues in 2017. In less than five years, it will be another \$100 billion Amazon business if the current pace holds up.<sup>3</sup>

Lest you think that AWS appeals only to tiny start-ups searching for a business model, consider that the Central Intelligence Agency, Netflix, Airbnb, Zillow, and many other prominent organizations rely on Amazon. In the rare event of an AWS outage (such as the one that happened on March 28, 2017), people learn about it very quickly.

**? TIP**

More powerful, reliable, and affordable infrastructure means more business and consumer applications. This results in more data and—you guessed it—more analytics.

---

## NOT YOUR FATHER’S DATA STORAGE

Up until fairly recently, data storage was both very expensive and very limited. Today, neither of those is true, and a brief history lesson will show just how far and fast we have come with those 1s and 0s.

### How? Hadoop and the Growth of NoSQL

For a long time, organizations needed to store all of their data either in mainframes or in relational databases. (For decades, data storage was almost always on-premise.) These options haven’t gone away. To be sure, many enterprises today continue to run smoothly with “back ends” such as Microsoft SQL Server, Oracle, and IBM’s DB2. It’s absurd to think, though, that these options represent the only or even the best way to store all types of enterprise data. Amazon has introduced a seismic shift in data storage and application deployment.

For starters, countless organizations have put their data “in the cloud.”\* Amazon Redshift and Microsoft Azure Data Lake are just two of the many cloud-based alternatives to storing data. Beyond *where* the data is stored, *how* it is stored may not look anything like it did in 1998. Distributed file systems such as Hadoop and popular NoSQL<sup>†</sup> databases such as Apache’s Cassandra can yield significant advantages over their traditional and relational counterparts: Fault tolerance and parallel processing<sup>‡</sup> are at the top of the list.

Brass tacks: Today, organizations face no shortage of flexible, powerful, and—as we’ll see next—extremely affordable options for storing their data.

---

\*I’m not a fan of this phrase because there is no one “cloud.” For more on this, see <http://bit.ly/2ILRZTh>.

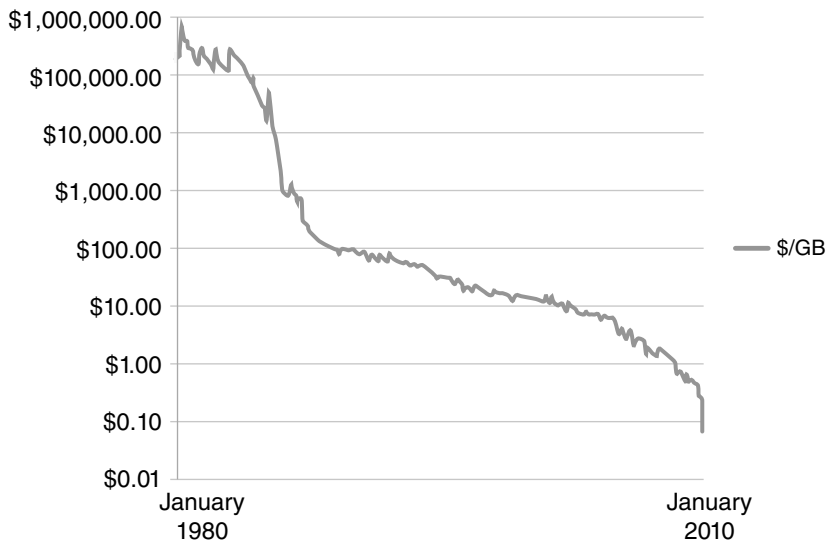
<sup>†</sup> Short for “not only SQL,” not “no SQL.”

<sup>‡</sup> For more on this, see my book *Too Big to Ignore*.

## How Much? Kryder's Law

In 2004, Google shocked the world by launching Gmail, a free e-mail service with one gigabyte (GB) of storage. This was a big deal. Today, Gmail offers 10 gigabytes of storage space for free accounts. Microsoft's free Outlook offers virtually unlimited storage.

All of this costs consumers zero. Ditto for basic accounts for Dropbox, Box.net, GoogleDrive, and Microsoft's OneDrive. None of these free services would exist if storage costs had not plummeted. As Figure 1.1 shows, they certainly have.



**Figure 1.1** Data-Storage Costs over Time

Source: Data from Matthew Komorowski (see [www.mkomo.com/cost-per-gigabyte](http://www.mkomo.com/cost-per-gigabyte))  
Figure from Phil Simon.

This phenomenon is known as Kryder's law. Not that long ago, data storage used to be terribly expensive for both enterprises and individuals. This has not been the case for a long time. Data storage is now orders of magnitude cheaper than it was in past decades.

### **? TIP**

Cheaper storage and more powerful alternatives mean that organizations can store more data, and, by extension, develop more powerful and sophisticated analytics.

## MOORE'S LAW

In 1965, Gordon Moore predicted that processing power for computers would double every 12 months.\* At the same time, the cost of that technology would drop by half over the same time.

Although it cannot continue forever, Moore's law has held true for more than 50 years. It is the reason that if you own a contemporary smartphone, you hold in your hand more computing power than the Apollo 11 spacecraft that put the first man on the moon.

## THE SMARTPHONE REVOLUTION

As a kid, I was scared of getting lost. I can remember losing track of my parents—or maybe it was vice versa—at a county fair when I was about 10 years old. I can remember someone announcing over the loudspeaker, “Will Philip Simon please come to the information booth?” Of course, I heard the announcement; it wasn't as if I was walking around listening to my iPhone or iPod.

I also remember my early trips on airplanes, some of which wound up delayed on the tarmac. When the pilot or flight attendant would announce that we weren't going anywhere for 15 minutes, the passengers and I would audibly moan.

Both of these scenarios are less likely to happen these days. Smartphones allow parents to call their children wherever and for whatever reason. Garden-variety travel delays don't seem to bother us as much as they once did. (Being physically removed from a plane, however, is another matter.) Most of us are oblivious to the outside world, dialed into our electronic devices. If we're listening to music, then we might not even *hear* those announcements in the first place.

The growth of smartphones—and to a lesser but nevertheless increasing extent, wearable technology—is absolutely staggering. In February 2017, Cisco reported that global data traffic from mobile devices “grew 63 percent in 2016” and reached “7.2 exabytes per month at the end of 2016, up from 4.4 exabytes<sup>†</sup> per month at the end of 2015.”<sup>‡</sup> Both

---

\* For different reasons not worth discussing here, people often quote this as 12, 18, and 24 months.

<sup>†</sup> One exabyte is equivalent to 1 billion gigabytes and 1,000 petabytes.

<sup>‡</sup> For the whole report, see <http://bit.ly/1W26UQo>.



actively and passively, we are generating more data than ever with no end in sight. To wit, an increasing amount of this data is *contextual*. That is, our smartphones can easily determine what we're doing when and where we're doing it (if we let them, of course). There's a reason that my iPad "recommends" that I open the YouTube app when I go to Lifetime Fitness and step on the treadmill. (I like to watch concert footage when I run. The news today just upsets me.)

## THE DEMOCRATIZATION OF DATA

Let's go back in time to the 1984 Hackers Conference. The American writer and entrepreneur Stewart Brand said the following to legendary Apple cofounder Steve Wozniak:

On the one hand information wants to be expensive, because it's so valuable. The right information in the right place just changes your life. On the other hand, information wants to be free because the cost of getting it out is getting lower and lower all the time. So you have these two fighting against each other.<sup>4</sup>

Brand was nothing short of clairvoyant. The trends described so far in this chapter reflect the fundamental conflict he referenced more than 30 years ago.

To be sure, not all information is free today. Many media sites have for years experimented with paywalls designed to protect their content from free riders. More recently, they have declared war on ad blockers. Music and movie piracy abound despite their industries' attempts to thwart them.

The tension between free and paid isn't going anywhere. For now, suffice it to say that "data" is more available, or, if you like, more *democratic* than ever. It's also much larger. After all, it's called *Big Data* for a reason. (Chapter 2 delves deeper into this subject.)

## THE PRIMACY OF PRIVACY

In June 2015, Apple CEO Tim Cook spoke at the Electronic Privacy Information Center (EPIC) Champions of Freedom event in Washington, D.C. After being honored for excellent leadership in the area of

privacy, Cook excoriated the companies that he felt were lagging in this regard:

Like many of you, we at Apple reject the idea that our customers should have to make tradeoffs between privacy and security. We can and we must provide both in equal measure. We believe that people have a fundamental right to privacy. The American people demand it, the constitution demands it, morality demands it.

Some of the most prominent and successful companies have built their businesses by lulling their customers into complacency about their personal information.<sup>5</sup>

Without naming names, no one could accuse Cook of being particularly subtle here. He was slamming Facebook and Google for what he believes are their cavalier stances on user privacy. Irrespective of Cook's unknown private opinions on the matter, his public stance certainly jibes with Apple's product philosophy. Apple doesn't need to monetize user data. Its high profit margins allow it to market "privacy as a feature," something that Microsoft has also adopted in recent years.<sup>6</sup>

Many consumers also report feeling this way. (It remains to be seen whether or not they vote with their wallets, though.) The public seems to be growing increasingly skeptical of what takes place online—and with good reason. From a 2014 Harris Interactive and TRUSTe study:

. . . 84 percent of consumers are less likely to click on an online ad. Three-fourths are less likely to enable location tracking. In addition, a full 89 percent won't do business with a company that doesn't do a good enough job protecting them online. And 76 percent are likely to check websites and apps for a privacy certification seal.\*

Companies that routinely ignore privacy concerns often face swift retribution and incur the wrath of their customers and users. Uber is arguably the poster child today, but it is hardly alone.

---

\* Read the whole study here: <http://bit.ly/2nEkrUb>.

## THE INTERNET OF THINGS

My friend Heather recently gave me a Nixplay Wi-Fi digital picture frame as a gift at my housewarming party. (Ironically, I had been thinking about getting one for a few months.) Within minutes, I connected it to my home network, installed the app on my smartphone and iPad, and started adding photos. It didn't take long before many of my favorite pictures appeared for my guests to see.

We're still in the early innings, but items such as these offer glimpses into a very connected future—specifically, the much-hyped Internet of Things. Make no mistake: It is coming, and not just for techies like me. As I wrote in a recent article for SAS, “Even traditionally conservative sectors such as farming stand to reap enormous rewards and savings in the form of greater crop yields, more efficient use of water, and the like.”\*

Cisco Systems has predicted that by 2020, 50 billion devices will connect to the Internet.<sup>†</sup> Even if that estimate is off by a factor of three, it's still an enormous number with obvious implications for data and analytics—to say nothing about security and privacy.

## THE RISE OF THE DATA-SAVVY EMPLOYEE

Historically, a garden-variety marketing analyst didn't have to be a mathematical wunderkind. Today, though, that same entry-level analyst more likely than not knows her way around a spreadsheet. What's more, she's probably proficient with Google Analytics. The trend is unmistakable: Not only are employees increasingly tolerating working with data, but they are beginning to *demand* it.

Kathy Marshall serves as the director of recruitment quality and client engagement at Decision Toolbox, a recruitment products and services company. In her words:

Data-savvy workers think differently, ask questions, challenge the establishment, and demand improvements. If a data-driven employee isn't challenged, can't affect

---

\* Read the article at <http://bit.ly/2mt8OB8>.

† Read the white paper at <http://bit.ly/1LgfMSb>.

change, or isn't able to access the information they need to drive results, [they] will quickly move on to other opportunities.<sup>7</sup>

But what if you just don't "do numbers"? You're in trouble. It's getting harder for numerically challenged employees to find hiding places within organizations.

## THE BURGEONING IMPORTANCE OF DATA ANALYTICS

On many levels, the trends described in this chapter have been nothing short of transformative. With regard to this book, they have individually, collectively, and *exponentially* increased the importance of data and analytics. These twins are more important than ever, but don't take my word for it.

Let's look at what happened on July 29, 2016.

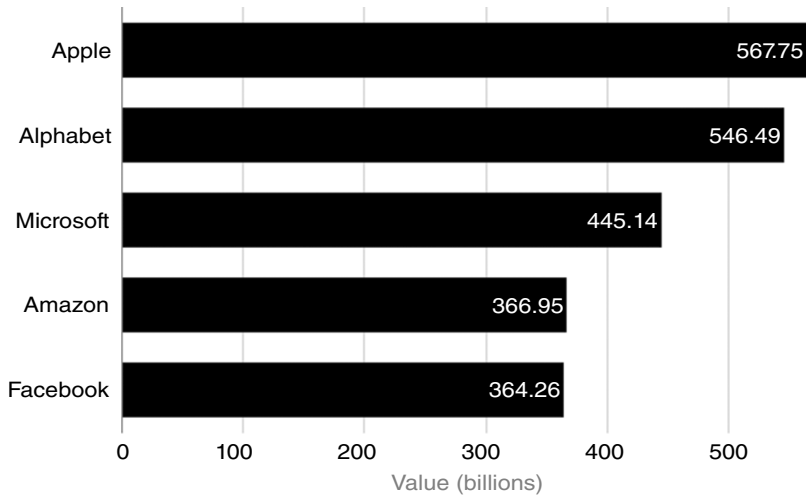
### A Watershed Moment

Lost in all of the hubbub of the most contentious presidential campaign of our lifetime, something remarkable and probably unprecedented happened on that summer day. For perhaps the first time in history, each of the five most valuable companies in the world (as measured by market capitalization) belonged to the technology sector. At least for a short while, Exxon, General Electric, and Johnson & Johnson took backseats to their tech-savvy brethren. Writing for *Slate*, Will Oremus noted:

To be clear, this leaderboard in itself has no direct impact on, well, anything, other than public perception. No one gets a prize for ranking at the top of it, nor a penalty for dropping down the list. Still, it feels like a cultural moment to see what are sometimes called the "big five" U.S. tech companies surpass all others in market value.<sup>8</sup>

Figure 1.2 presents these companies' market capitalizations.

Fundamental differences certainly exist among these companies beyond a few hundred billion dollars in market cap. For instance, as mentioned earlier in this chapter, Amazon has aggressively expanded



**Figure 1.2** The World's Most Valuable Companies by Market Cap as of July 29, 2016, at 10:50 a.m. ET

*Source: Data from Google Finance. Figure from Phil Simon.*

its focus from its relatively humble beginnings. It's downright wrong—and has been for years—to think of Amazon as an online bookstore. Its cloud-computing division, AWS, generates more than \$15 billion per year in revenue as well as the lion's share of the company's profits.<sup>9</sup>

For its part, Facebook has stayed largely true to its initial vision of making the world more open and connected. To this end, Mark Zuckerberg has made his fair share of high-profile acquisitions. By adding WhatsApp, Instagram, and Oculus, Facebook has cemented its position as the world's most popular and valuable social network. Unlike Amazon, Facebook has not made significant inroads into non-adjacent areas such as enterprise sales. (Facebook's recent efforts to hawk its new collaboration tool Workplace don't appear to be bearing much fruit.)

## Common Ground

The cardinal similarity among these behemoths are data, and, by extension, analytics. Each company possesses incredibly large and valuable troves of customer and user information, although their strategies diverge considerably. Make no mistake: Amazon, Apple, and Microsoft

may differ from Facebook and Alphabet/Google in how they treat privacy, but all five of these market leaders understand the import of data. Case in point: On June 13, 2016, Microsoft ponied up a mind-boggling \$26.2 billion for professional social network LinkedIn. Many industry insiders consider the move to be nothing more than a large data grab.

## CREATING CULTURES OF ANALYTICS

In September 2015, I gave the keynote speech at the JMP Discovery Summit in San Diego. The title of my talk was “How to Create a Culture of Analytics.”\* Over the course of an hour, I described how Netflix, Amazon, Facebook, and Google have institutionalized analytics in their organizations. To paraphrase the title of the book by ex-IBM CEO Lou Gerstner, it ain’t easy, but some elephants can indeed dance.

And to these victors go significant spoils. Sure, these companies have succeeded for other reasons, including their competitors’ missteps and dumb luck. Nevertheless, Wall Street likes what it has seen. Netflix, Amazon, Facebook, and Google have consistently outperformed their peers. On a personal level, I have found that as a general rule, organizations that embrace analytics and experimentation are able to move more nimbly compared to bureaucratic, data-agnostic organizations. They are less likely to suffer the fates of Blockbuster, Kodak, and other erstwhile powerhouses.

## The Data Business Is Alive and Well and Flourishing

There’s a reason that Facebook’s stock is flying so high. Unlike Apple, a company decidedly not in the data business, Facebook has figured out how to monetize petabytes of user data. As we have seen, Wall Street has taken notice but, as always, isn’t satisfied. The next logical step for Facebook is to open its vast data trove to those hungering for it the most. I’m talking here about the social network’s *real* customers: its advertisers.<sup>†</sup> And Mark Zuckerberg is doing exactly that.

\* Watch the talk at <http://bit.ly/2moA5G8>.

<sup>†</sup> If you use Facebook, then you are the product. Never confuse *users* with *customers*.

In 2013, Facebook purchased Microsoft's Atlas ad serving and tracking business for an estimated price of \$100 million.<sup>26</sup> Relaunched by Facebook the following year, Atlas helps marketers and publishers at a high level:

- Load online ads into web pages and even individual apps.
- Measure which people their online ads are reaching.

For the past few years, Facebook limited the availability of some of its most promising ad products. Not anymore. Scott Shapiro currently serves as Facebook's product marketing director for measurement. "Previously, these tools were only available in Atlas for the largest enterprise marketers," he said in March 2017. "We're expanding the access of functionality that only existed in Atlas to tens of thousands of marketers who use Facebook's ad tools."<sup>11</sup>

Facebook clearly possesses something that scores of established companies want: incredibly detailed user demographic and psychographic data. Want to find female millennials in Manhattan who dig the HBO show *Girls*? No problem. Looking for male fans of the English prog-rock band Marillion who reside in London? Check. With nearly two billion users (many of them legitimate and routinely engaged), Facebook has advertisers frothing at the mouth. Atlas isn't exactly a hard sell.

## ANTICIPATORY COMMERCE

For decades, Amazon has made it easy for millions of customers to quickly find what they want and place orders with a click of the mouse. Want to subscribe to toothpaste or even beef jerky? No problem. But what about fulfilling an order *before* you actually place it? That thought has occurred to Amazon as well.

In January 2014, word broke that Jeff Bezos's company had filed a patent for anticipatory commerce. As Greg Bensinger of the *Wall Street Journal* reported: "Amazon says it may box and ship products it expects customers in a specific area will want—based on previous orders and other factors. According to the patent, the packages could wait at the shippers' hubs or on trucks until an order arrives. If implemented well, this strategy has the potential to take predictive analytics to the next level, allowing the data-savvy company to greatly expand its base of loyal customers."<sup>12</sup>

## Not Just the Big Five

Amazon, Facebook, Google, Microsoft, and—to a lesser extent—Apple are hardly alone in reaping outsize rewards from their early analytics and data-related efforts. Consider Netflix, a company that I discuss at length in *The Visual Organization*. I won't repeat myself too much here, but suffice it to say that Netflix knows a great deal about what its nearly 90 million subscribers are watching, when, and on which device. To put it mildly, detailed information on viewing habits informs the company's decisions on producing expensive original content—and leasing existing shows and movies. Netflix doesn't green-light shows willy-nilly. At roughly \$5 billion in 2016 alone, content-acquisition costs represent one of Netflix's largest annual expenses.<sup>13</sup>

### *Uber: The Economist's Dream*

Forget publicly traded companies for a moment. Uber, the world's richest *private* company at the time of this writing, sports a valuation of roughly \$70 billion. Travis Kalanick's outfit knows far more about its customers and "driver-partners"\* than most people realize. In fact, Uber's vaunted algorithm and app allow it to do things about which most CEOs can only dream.

If you have studied economics, you most likely learned about a concept called *consumer surplus*. Put simply, this represents the difference between the price that consumers are willing and able to pay for a good or service and the price that they actually pay (read: the market price). Let's say that you are really hankering for some blueberries. You go to your local grocery store and head to the fruit section. You would pay \$5 for a carton of them. Fortunately, blueberries are on sale for \$1. In this example, that \$4 difference represents the consumer surplus.

Keeping with this example, it's practically impossible for your neighborhood Albertsons or Whole Foods to capture that \$4. There's just no way for a store employee to effectively gauge how much each customer is willing to pay. That \$1 price reflects a number of factors.

---

\* For legal reasons, the company created this clunky term. By refusing to classify its drivers as *employees*, Uber saves billions in employee taxes and benefits. To be fair, many other "on-demand" companies do the same thing.



Perhaps there is an excess of blueberries and the store is “blowing them out” in lieu of tossing them out. Blueberries are perishable goods, after all.

Of course, Uber doesn’t face this problem. Via data and its app, the company is able to offer dynamic and surge pricing. (Some would use the much less benign term *price gouging*.) In theory, surge pricing can match drivers and customers in real time. Remember that Uber’s “driver-partners” are free to drive if and when they like. If Uber can’t compel people to drive, then how can it meet consumer demand?

The answer lies in economics. Uber relies on market incentives to lure drivers out of their homes and onto the streets when its algorithm detects a dearth of them in a city. Uber automatically raises driver rates when customer demand rises. Put differently, nothing at Uber is fixed: not the per-mile rate, not the number of drivers, not the schedule. The same trip that cost you \$15 last week might cost you twice that much today. It all hinges on supply and demand. For instance, consider New Year’s Eve, typically a massive mismatch between supply of and demand for drivers. Over the course of Uber’s history, many customers have taken to social media to vent about exorbitant rates on this day. After the ball dropped in Manhattan’s Times Square in 2014, some inebriated riders were shocked to find that surge pricing resulted in fares six times higher than normal.<sup>14</sup>

Brass tacks: Uber is arguably able to capture more of the consumer surplus than any company in the history of modern-day capitalism. It may, in fact, be an economist’s dream.\* That’s a far cry, though, from saying that it’s all sunshine and lollypops for the ride-sharing juggernaut. It most certainly is not.

In 2014, the company faced a public maelstrom when word leaked that it was furtively employing “God View” tracking capability on unsuspecting customers. With a few clicks of a mouse, nosy employees could easily stalk celebrities, journalists, and rival ride-sharing exes.† The company ultimately settled a New York lawsuit.

---

\* For more on this, see the *Freakonomics* podcast “Why Uber Is an Economist’s Dream” at <http://tinyurl.com/freakuber>.

† To read the court documents, see <http://tinyurl.com/uber-courttx>.

Things have continued to go downhill for Uber. In a remarkable three-week period in March 2017, the company faced a string of PR nightmares, including:

- An ex-employee’s viral blog post about the company’s rampant harassment and discrimination.\*
- A lawsuit from Alphabet (Google’s parent company) over stolen intellectual property.
- A viral video in which Kalanick berated an Uber driver over fares.†
- A spate of executive turnover.
- A crash of a driverless car in Arizona and the subsequent suspension of its testing.
- Revelations that the company furtively developed technology designed to identify and circumvent government officials’ efforts to detect if the service was operating illegally—aka Greyball.‡

April of 2017 didn’t bode much better for Uber. Stories surfaced on its hyper-aggressive business practices, including illegally spying on Lyft drivers and tracking users’ behavior after they had deleted the Uber app. (This practice, known as “fingerprinting,” violates Apple’s terms of service.)

### *Airbnb: Better (If Possibly Illegal) Living through Data*

Next on the decacorn<sup>s</sup> list is Airbnb, the disruptive and possibly illegal online marketplace and hospitality service. Its current private valuation is \$30 billion. Much like at Uber, CEO Brian Chesky and his cofounders understand exactly what is happening on his company’s “platform” at all times. Chesky et al. built Airbnb with data in mind from early on. As a result, Airbnb can easily answer questions such as:

- How many proper homes are available in Montreal right now? What’s the average price per home, and how does that compare with available apartments?
- What was the average length of stay of Airbnb guests in New York last December? How did that change from the previous December?

---

\* See <http://bit.ly/2kX7hju>.

† See <http://bit.ly/2mgAs2y>.

‡ See <http://nyti.ms/2nY3J54>.

<sup>s</sup>If *unicorns* are start-ups worth \$1 billion, then *decacorns* are worth \$10 billion.

- What is the trend in Paris? Are prices increasing or decreasing, and by how much?
- How many individual rooms are available in San Francisco? What about entire homes and apartments?
- How many hosts are violating local housing ordinances? (This may include listing multiple properties or listing properties for lengths of time outlawed by local statutes.)

Airbnb possesses an astonishing quantity and level of information on its customers and hosts—not that Chesky is keen on releasing that data to regulatory agencies, hotel lobbyists, affordable-housing activists, and others curious about whether the company is flouting the law. Under the guise of user privacy, Airbnb has routinely resisted calls to release its data. In fact, independent websites such as Inside Airbnb ([insideairbnb.com](http://insideairbnb.com)) exist that seek to “add data to the debate.” For more on this, see Brad Stone’s excellent 2017 book *The Upstarts: How Uber, Airbnb, and the Killer Companies of the New Silicon Valley Are Changing the World*.

But it doesn’t stop there. Airbnb aims to maximize customer recommendations and repeat customers—in this case, bookings. Like many companies, Airbnb uses Net Promoter Score (NPS), a customer-loyalty measure that dates back to 2003. At a high level, NPS asks one simple question: “How likely are you to recommend Airbnb?” In industry parlance, this is called *likelihood to recommend*. The company’s data scientists methodically analyze and tweak the following variables:

- Overall review score and responses to review subcategories on a scale from 1 to 5
- Guest acquisition channel (organic or marketing campaigns)
- Trip destination
- Guest origin
- Previous bookings from the guest on Airbnb
- Trip length
- Number of guests
- Price per night
- Month of checkout (to account for seasonality)

- Room type (entire home, private room, shared room)
- Other listings owned by the host<sup>15</sup>

It may seem that Netflix, Uber, and Airbnb have built insurmountable leads via their extensive use of data and technology. For several reasons, though, that type of thinking is misplaced and shortsighted.

For starters, disruption happens faster than ever. Case in point: Up until the iPhone launched in 2007, the BlackBerry constituted nearly 40 percent of the cellphone market. On February 15, 2017, research firm Gartner reported that the BlackBerry's market share had plunged to 0.0 percent,<sup>16</sup> not that it has been remotely relevant for nearly a decade. Add to that the fact that *dominance* and *monopoly* are not synonyms. Hulu, YouTube, and Amazon are hot on Netflix's trail. Lyft is still a viable national ride-share alternative in the United States that continues to raise substantial funds.<sup>17</sup> (Some cities that have banned Uber and Lyft have endorsed smaller, local players that choose to abide by their rules. Ride Austin in Austin, Texas, is just one example.) Given Uber's burgeoning legal and PR troubles, success is anything but assured. And if you think that Airbnb is the only short-term rental marketplace around, you're mistaken.

## DATA-RELATED CHALLENGES

This is not to say that today's powerful tech companies haven't struggled with data-related matters. Nothing could be further from the truth.

Facebook and Twitter received a great deal of valid criticism during and after the 2016 U.S. presidential election. Each social network failed to prevent or at least contain the spread of fake news. In one oft-cited example dubbed "Pizzagate," a North Carolina man "self-investigated" an apocryphal and politically charged story falsely claiming that Hillary Clinton was running a child-sex ring out of a pizzeria in Washington, D.C. As Joshua Gillin wrote on PolitiFact:

Edgar Maddison Welch, a 28-year-old man from Salisbury, N.C., walked into Comet Ping Pong in the capital around 3 p.m. on Dec. 4. Police said he pointed his gun at a worker, who fled, and then Welch started firing the rifle inside the restaurant.<sup>18</sup>

While extreme, stories such as these on Facebook weren't uncommon during a most contentious election. In fact, the social network's algorithm creates these types of "filter bubbles," to borrow a phrase from Eli Pariser's remarkably insightful and prescient 2012 book. In March 2017, Facebook finally responded by adding a "mark as disputed" option to stories in its NewsFeed.\* Many critics argued that the response was too little, too late. No argument here.

For its part, Amazon has filed several lawsuits against Fiverr, "a global online marketplace offering tasks and services, beginning at a cost of \$5 per job performed." Amazon first took Fiverr to court in April 2015. Six months later, Amazon and its lawyers were back at it. The obvious question is: Why? As Sarah Perez wrote for TechCrunch:

The defendants in the new case, listed as "John Does," each used Fiverr.com to sell fake positive or 5-star Amazon reviews. In some cases, they even offered "verified" reviews, meaning those where they buy the product—provided they're compensated for that, of course. Other times, they also tell the purchaser to just provide the product review and they'll post it.<sup>19</sup>

Those faux reviews aren't just inconvenient. If left unchecked, they represent a significant threat to Amazon's business because customer reviews are downright essential today. In December 2015, search engine optimization outfit BrightLocal found that 88 percent of consumers trust online reviews as much as personal recommendations. Nearly 40 percent "regularly" read online product reviews and only 12 percent did not.<sup>†</sup>

## COMPANIES LEFT BEHIND

Of course, not every company and industry has benefited from the technology and data revolutions. To borrow a line from *Breaking Bad*, I want to tread lightly here. I am certainly not looking to make enemies.

---

\*See *60 Minutes'* fascinating segment on how fake news spreads at <http://cbsn.ws/2o75MB0>.

<sup>†</sup>Read the whole study at <http://bit.ly/2dDB2o8>.

It's fair to say, though, that the publishing industry as a whole arrived late to the data and analytics party. Compared to Amazon, traditional publishers and large bookstores such as Barnes & Noble have struggled. (Borders filed for Chapter 11 bankruptcy in February 2011.) In large part, these firms haven't maintained direct relationships with their customers. They often don't know who buys their books. At least they are not alone.

Many brick-and-mortar retailers also have suffered from their inability to tailor offerings to potential customers who walk through their doors. This is why there's so much industry excitement around location-based technologies such as Apple's iBeacon. (Think of the infamous commercials in *Minority Report*.) Equipped with real-time data, Target, Home Depot, Macy's, and other big-box retailers may represent more than mere showrooms for Amazon, a company that largely avoids the considerable expenses associated with maintaining physical stores.

## THE GROWTH OF ANALYTICS PROGRAMS

Do you still doubt that analytics matter more than ever? I hope that this chapter has convinced you. In case you're still on the fence, consider the following.

Higher education needs to evolve to stay relevant, especially in an era of massive open online courses, staggering levels of college debt, and Thiel Fellows.\* More U.S. colleges and universities than ever are offering formal programs in analytics and data science.<sup>20</sup> I am fortunate enough to teach at one of the best ones. I would be mystified if this trend abates anytime soon. Beyond accredited academic programs, there are plenty of informal ways to increase your knowledge in the field. Coursera, Udemy, and other online learning companies have followed suit with individual courses and programs, many of which cost nothing.

---

\* Controversial billionaire Peter Thiel gives 20 smart young cookies \$100,000 each year to start their own companies and forgo college.

---

## CHAPTER REVIEW AND DISCUSSION QUESTIONS

- What do you think of Steve Jobs's stance regarding the *New York Times*?
  - What types of things could Apple do with this type of customer information? What types of apps could it recommend?
  - What types of data does Uber capture? How can it analyze that data in ways that traditional taxi and transportation companies cannot?
    - What types of experiments can Uber run?
    - How else would you use this data?
    - How could you change the Uber app to collect even more information?
  - Now imagine that you are Airbnb CEO Brian Chesky. What kinds of questions could you ask and answer of your company's data?
  - How would you use that information?
  - Could bookstores and traditional publishers have embraced new technologies and data sources more quickly? What specifically could each have done?
  - Do you think that any of these moves ultimately would have made a difference, or are disruption and marginalization inevitable?
- 

## NEXT

This chapter establishes today's tech-, data-, and analytics-heavy business context. Now it's time to take a step back. Just what do *data* and *analytics* mean, anyway?

The next chapter answers this key question.

## NOTES

1. Justin Fox, "How to Succeed in Business by Bundling—and Unbundling," *Harvard Business Review*, June 24, 2014, <http://tinyurl.com/h8ld3eb>.
2. Daniel Nations, "How Many iPads Have Been Sold?," March 15, 2017, <http://tinyurl.com/z93lpml>.
3. Charles Babcock, "Amazon: The Self-Fueling, Perpetual Motion Machine," *InformationWeek*, January 9, 2017, <http://ubm.io/2j4NkJF>.
4. Roger Clarke, "Information Wants to Be Free . . .," August 28, 2001, <http://bit.ly/12V0o8m>.

5. Matthew Panzarino, "Apple's Tim Cook Delivers Blistering Speech on Encryption, Privacy," *TechCrunch*, June 2, 2015, <http://tcrn.ch/2nSK2YI>.
6. Peter Sayer, "Apple, Microsoft Wield Privacy as Marketing Tool," *Computerworld*, September 29, 2015, <http://tinyurl.com/z7ckvyc>.
7. Lisa Morgan, "12 Ways to Cultivate a Data-Savvy Workforce," *InformationWeek*, July 7, 2016, <http://ubm.io/29pbjX>.
8. Will Oremus, "Tech Companies Are Dominating the Stock Market as Never Before," *Moneybox*, July 29, 2016, <http://tinyurl.com/gkpsdtk>.
9. Alexei Oreskovic, "Amazon Isn't Just Growing Revenue Anymore—It's Growing Profits," *Business Insider*, April 28, 2016, <http://tinyurl.com/z3d42j2>.
10. Peter Kafka, "Sold! (Finally) Facebook Takes Atlas from Microsoft, So It Can Get Serious about Ads," *All Things*, February 28, 2013, <http://tinyurl.com/d6htpbm>.
11. Anthony Ha, "Facebook Is Making Its Cross-Device Atlas Data Available to More Advertisers," *TechCrunch*, March 7, 2017, <https://techcrunch.com/2017/03/07/facebook-advanced-measurement>.
12. Greg Bensinger, "Amazon Wants to Ship Your Package Before You Buy It," *Wall Street Journal*, January 17, 2014, <http://on.wsj.com/1h7X1m8>.
13. Jessica Rawden, "The Insane Amount of Money Netflix Will Spend on Content in 2016," *CinemaBlend*, March 1, 2016, <http://tinyurl.com/h9jdytt>.
14. Cristina Alesci and Kate Trafecante, "Uber Prices Surge on New Year's Eve," CNN, January 2, 2015, <http://tinyurl.com/cnnuber6>.
15. Lisa Qian, "How Well Does NPS Predict Rebooking?," Airbnb Engineering & Data Science, December 10, 2016, <http://nerds.airbnb.com/nps-rebooking>.
16. Dan Frommer, "Chart of the Day: The BlackBerry's Fall to 0.0 Percent Market Share," *Recode*, February 26, 2017, <http://tinyurl.com/jcbxfwu>.
17. Marco della Cava, "Lyft Looks to Raise \$500M as Uber Stumbles," *USA Today*, March 2, 2017, <http://usat.ly/2mDwLWM>.
18. Joshua Gillin, "How Pizzagate Went from Fake News to a Real Problem for a D.C. Business," *Politifact*, December 5, 2016, <http://bit.ly/2gZE862>.
19. Sarah Perez, "Amazon Files Suit against Individuals Offering Fake Product Reviews on Fiverr.com," *TechCrunch*, October 16, 2015, <http://tcrn.ch/2IHlOnR>.
20. J. Maureen Henderson, "Are Degrees in Big Data a Fad or a Fast Track to Career Success?," *Forbes*, July 30, 2013, <http://tinyurl.com/hhrqc7d>.



---

## About the Author

**Phil Simon** is a sought-after keynote speaker and recognized technology authority. He has written seven other business books, two of which have won awards: *The Age of the Platform* and *Message Not Received*. In the fall of 2016, he joined the faculty at Arizona State University's W. P. Carey School of Business. He teaches courses on system design, enterprise analytics, and business intelligence.

When not speaking, writing, and teaching, Simon advises organizations on analytics, strategy, emerging trends, and management. His contributions have been featured by the *New York Times*, *Wired*, *Harvard Business Review*, CNN, NBC, CNBC, *Inc.* magazine, *BusinessWeek*, the *Huffington Post*, and many other sites. He holds degrees from Carnegie Mellon and Cornell University.

Simon's website is [www.philsimon.com](http://www.philsimon.com). Find him on Twitter at @philsimon.

# Ready to take your SAS® and JMP® skills up a notch?



Be among the first to know about new books,  
special events, and exclusive discounts.

**[support.sas.com/newbooks](https://support.sas.com/newbooks)**

Share your expertise. Write a book with SAS.

**[support.sas.com/publish](https://support.sas.com/publish)**

 [sas.com/books](https://sas.com/books)  
for additional books and resources.

  
THE POWER TO KNOW.

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. © 2017 SAS Institute Inc. All rights reserved. M1588358 US.0217