



PROFITING FROM THE DATA ECONOMY

Understanding the Roles of
Consumers, Innovators and Regulators
in a Data-Driven World

DAVID A. SCHWEIDEL

Praise for *Profiting from the Data Economy*

“David Schweidel’s book is a highly readable and comprehensive look at the impact that the Big Data phenomenon is having on both businesses and consumers. Managers can learn the implications of the availability of more data on how to reach consumers more efficiently and effectively, while taking consumer privacy and regulatory implications into consideration. This is an important and timely book.”

—**Russell Winer**, William H. Joyce Professor of Marketing,
Stern School of Business, New York University

“David’s book is a must-read for those wanting to turn big data into big value. First and foremost, this book recognizes that it is only through the combination of better data and statistical knowledge when coupled with business acumen that corporate and customer value can arise.”

—**Eric T. Bradlow**, Chairperson, Wharton Marketing Department, K.P. Chao Professor; Professor of Marketing, Statistics, and Education, Vice-Dean and Director, Wharton Doctoral Programs, Co-Director, Wharton Customer Analytics Initiative, The Wharton School of the University of Pennsylvania

“Every business leader, every data practitioner, every teacher and student of business should read this book. David Schweidel’s account of analytics in business is both accessible and thought-provoking. It is a pragmatic introduction to the way that new levels of insights and the realities of consumer sentiments in our data-driven world shape the future of business and the economy as a whole.”

—**Rasmus Wegener**, Partner, Bain & Company

“An extremely well-written and timely book on data economy. It is full of case studies and quotes. I particularly liked the discussion on the role of public policy in protecting consumers from cyber threats.”

—**Jagdish N. Sheth**, Charles H. Kellstadt Professor of Marketing,
Emory University

“In *Profiting from the Data Economy*, Schweidel brilliantly and accurately coins the term ‘digital exhaust’ to label the increasing amounts of granular data society’s engines generate. This book presents the innovators, regulators, and consumers who grapple with the opportunities and cautions this new resource creates, as well as the tools, techniques, players, projects, and legal issues surrounding Big Data’s ecosystem. As a practitioner of the art and science of Big Data, I promise this book will engage you. If you are already involved in Big Data, this will help you understand the landscape—and if you are not, it will convince you to start paying attention. An equally important read for consumers, business, public services, and the legal profession.”

—**Michael Knight**, CEO, CustomerAnalytics.com

“The increasing swath of consumer-level data that is now available to firms in many industries is enabling a culture of data-driven decision-making in these organizations. This can make them more efficient and profitable. Consumers should also realize that by sharing their information with firms, they can benefit from more personalized and curated products. They need to get over their knee-jerk reaction to privacy concerns from data sharing and realize that this give and take is how the world will increasingly operate in the coming years. Said simply, users have a lot more to gain than lose from information sharing with firms. This book lays out these important issues remarkably well and is an absolute must-read for practitioners and government officials alike.”

—**Anindya Ghose**, Professor of IT and Professor of Marketing,
and Co-Director of the Center for Business Analytics at
NYU Stern School of Business

“The data-driven world described in *Profiting from the Data Economy* identifies an expanding vista for consideration by innovators and management in the modern enterprise. In this book, David Schweidel addresses how all stakeholders can benefit from new data, advances in analytics, and new interpretations of results. This engaging narrative is rich in anecdotes, analogs, and lessons-learned, and is an important read for all managers, analysts, and students of analytics.”

—**Benn Konsynski**, George S. Craft Distinguished University Professor
of Information Systems, Goizueta Business School, Emory University

Profiting from the Data Economy

Understanding the Roles of Consumers,
Innovators, and Regulators
in a Data-Driven World

David A. Schweidel

Associate Professor of Marketing,
Co-Director of Emory Marketing Analytics Center,
Goizueta Business School, Emory University

Associate Publisher and Director of Marketing: Amy Neidlinger

Executive Editor: Jeanne Glasser Levine

Development Editor: Natasha Torres

Operations Specialist: Jodi Kemper

Cover Designer: Chuti Prasertsith

Managing Editor: Kristy Hart

Senior Project Editor: Lori Lyons

Copy Editor: Bart Reed

Proofreader: Apostrophe Editing Services

Indexer: Lisa Stumpf

Compositor: Nonie Ratcliff

Manufacturing Buyer: Dan Uhrig

© 2015 by David A. Schweidel

Pearson Education, Inc.

Upper Saddle River, New Jersey 07458

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Company and product names mentioned herein are the trademarks or registered trademarks of their respective owners.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

First Printing November 2014

ISBN-10: 0-13-381977-9

ISBN-13: 978-0-13-381977-9

Pearson Education LTD.

Pearson Education Australia PTY, Limited.

Pearson Education Singapore, Pte. Ltd.

Pearson Education Asia, Ltd.

Pearson Education Canada, Ltd.

Pearson Educación de Mexico, S.A. de C.V.

Pearson Education—Japan

Pearson Education Malaysia, Pte. Ltd.

Library of Congress Control Number: 2014947301

To my family.

This page intentionally left blank

Contents-at-a-Glance

	Foreword	xiv
	Preface	xviii
Chapter 1	Beyond Big Data	1
Chapter 2	Building Businesses	23
Chapter 3	Refining Practice	51
Chapter 4	Improving Public Service	77
Chapter 5	Today's Data Economy	97
Chapter 6	Cracks in the Foundation of the Data Economy	123
Chapter 7	Harbingers of Change	151
Chapter 8	In Need of Oversight?	173
Chapter 9	The Race for Resources.	203
Chapter 10	What's Next for the Data Economy?	235
	Afterword	257
	Index	259

Contents

	Foreword	xiv
	Preface	xviii
Chapter 1	Beyond Big Data	1
	Searching for the Next Generation of Quants	2
	From Big Data's Past to Its Future	5
	Characterizing Big Data	6
	Is Big Data a Strategy?	9
	Data Versus Insights	10
	Data and Value	12
	Value for Value	16
	Endnotes	20
Chapter 2	Building Businesses	23
	Back to Marketing Basics	23
	Putting Marketing Analytics to Use	27
	Internet-Based Businesses: Is Content or Context King?	32
	Social (Marketing) Networks	38
	Common Ground	44
	Discussion Questions: How Do We Reveal Ourselves Online?	45
	Endnotes	46
Chapter 3	Refining Practice	51
	Old Media? New Media? Just Media	52
	Better Data, Better Ad Targeting	57
	Old Media Meets New Media	59
	What's Your Life Worth?	64
	Timing's Everything	66
	You're Where?	69
	Discussion Questions: Reaching Today's Consumer	71
	Endnotes	72

Chapter 4	Improving Public Service	77
	Can Data Protect and Serve?	78
	Big Findings in Public Data	80
	Quality Trumps Quantity	83
	Compiling Data to Inform the Public	88
	Consumers and Providers of Data	90
	Discussion Questions: Data Science for Social Good	92
	Endnotes	93
Chapter 5	Today's Data Economy	97
	The Groundwork	97
	The Current Exchange	100
	The Foundation of the Data Economy:	
	Customer-Centric Marketing	108
	Customer-Centric Investments in Data	114
	Discussion Questions: The Collaborative Consumer	118
	Endnotes	119
Chapter 6	Cracks in the Foundation of the Data Economy	123
	Privacy in Customer Data	125
	Learning Who Your Customers Are	127
	Why Marketers Need to Engage in the Debate	130
	Transparent Practices, Informed Customers	135
	Sharing the Value of Data	140
	My Actions, My Data?	141
	Discussion Questions: The Hierarchy of	
	Personal Data	144
	Endnotes	146
Chapter 7	Harbingers of Change	151
	Demand-Based Pricing	151
	The Consumer Highway to Hell?	156
	Benefiting from Price Discrimination	160
	Consumers' Comfort with Leveraging the	
	Data Exhaust	163
	Discussion Questions: Valuing	
	Consumer Data	169
	Endnotes	170

Chapter 8	In Need of Oversight?	173
	Valuing Consumer Privacy	173
	Profiling by Association	176
	Data Sharing Free-for-All	180
	Consumer Data, But at What Cost?	185
	Data-Driven Discrimination	189
	Socially Acceptable Segmentation?	192
	Discussion Questions: Protecting Consumers Throughout the Data Value Chain	196
	Endnotes	197
Chapter 9	The Race for Resources.	203
	Want Consumer Data? Pay to Play	203
	Exchanging Products and Services for Consumer Data	205
	Data Acquisition Free-for-All	208
	Empowering and Informing Consumers	211
	Reshaping the Media Landscape	214
	Consumer Data as a Financial Asset	218
	Do We Need Regulators in the Data Economy?	220
	Education as Part of Data Regulation?	224
	Can Consumer Control Ensure Competition?	227
	Discussion Questions: Empowering Consumers to Regulate Access to Personal Data	228
	Endnotes	229
Chapter 10	What's Next for the Data Economy?	235
	Moving Beyond Double Jeopardy	235
	The Changing Face of Innovation	237
	Can Consumer Data Contribute to Competition?	239
	Smarter Practice, but How Far Is Too Far?	241
	The Cost of Data-Driven Innovation	244
	An Appropriate Role for Government?	246
	A Right to Digital Privacy?	249
	Endnotes	252
	Afterword	257
	Index	259

Acknowledgments

The ideas put forth here are the result of years of coffee breaks, lunches, random conversations, and more coffee breaks with collaborators and colleagues. I benefited from studying and working with the faculty at The Wharton School of the University of Pennsylvania. During my time at the University of Wisconsin-Madison, colleagues and advisory board members played a large role in refining the core ideas that underpin this work and in broadening the scope of my thinking. The same has held true of my colleagues at Emory University. Across the board, I have been fortunate to have found mentors who encouraged me to think big while not losing sight of what is practical and relevant.

My thanks to Jeanne Glasser Levine for encouraging me to pursue this project and shepherding it through the publication process. Over the course of working on this project, I conducted numerous interviews. The perspective gained from those conversations was invaluable. My thanks to Reggie Bradford, Drew Conway, Jim Davis, Mike Finnerty, Jim Guszcza, Seth Haberman, Claudio Marcus, George Pappachen, Justin Reilly, Doug Rollins, Damon Samuel, Indira Venkat, Rasmus Wegener, and Robert Welborn for their time.

This book would not have been possible without the love and support of my wife Jennifer. Working on this book while we tackled a few other projects probably wasn't ideal. Yet, you were a constant source of assurance throughout, not to mention being essential in keeping me on task.

About the Author

David A. Schweidel is Associate Professor of Marketing and Co-Director of Emory Marketing Analytics Center (EmoryMAC) at the Goizueta Business School of Emory University.

Schweidel received his B.A. in mathematics, M.A. in statistics, and Ph.D. in marketing from the University of Pennsylvania. Prior to joining Emory in 2012, he was on the faculty of the Wisconsin School of Business at the University of Wisconsin-Madison.

Schweidel is an expert in the areas of customer relationship management and marketing intelligence. His research focuses on the development and application of statistical models to understand customer behavior and inform managerial decisions. His research has appeared in leading business journals, including *Journal of Marketing*, *Journal of Marketing Research*, *Marketing Science*, and *Management Science*. His research has garnered numerous awards, including the Gaumnitz Junior Faculty Research Award from the Wisconsin School of Business and the Marketing Science Institute's Buzzell Award. He has been recognized as a leading scholar by the Marketing Science Institute's Young Scholar program and by *Poets and Quants*' "Top 40 Under 40." Based on his research, he has consulted for companies such as eBay and HP Labs.

Schweidel has previously taught courses in data analysis and customer relationship management. Currently, he teaches one of the first courses offered at a top business school in digital and social media strategy. His teaching has been recognized at the Wisconsin School of Business, where he received the Chipman Faculty Award for Excellence in Teaching (2011). He also received the junior faculty teaching award from the MBA program at Goizueta (2014). In addition to his work in the classroom, he has led tutorials at conferences, including

the INFORMS Business Analytics conference and the AMA Analytics with Purpose conference. He has also spoken at conferences such as the AMA Advanced Research Techniques forum, INFORMS Marketing Science conference, and the Marketing Science Institute's Marketing Analytics in a Data-Rich Environment conference.

Schweidel is the author of *Social Media Intelligence* (Cambridge University Press), in which he and his co-author discuss how organizations can leverage social media data to inform their marketing strategies.

Foreword: The Personalized and the Personal

Fifty years ago, the communication theorist Marshall McLuhan famously declared that “the medium is the message.” McLuhan’s idea was that the modern media—no less than the specifics of the content they convey—affect society in profound ways that merit closer scrutiny. In today’s world of Big Data, an analogous idea is at play in the marketing domain.

Many of the concepts central to David Schweidel’s timely and perceptive new book have been with us for a long time. Business analytics—the use of data analysis to drive better business decisions—dates back at least as far as the invention of credit scoring in the late 1950s. In marketing, the use of transactional and customer-level data dates back at least as far as the introduction of database marketing and lifetime value methods in the 1980s. So while such terms as “business analytics,” “data science,” and “Big Data” are new, the activities they denote are most certainly not.

Yet there are important ways in which the advent of Big Data represents something fundamentally new. One is sheer ubiquity. Moore’s Law tells us that the number of transistors that can be fit on a microchip grows exponentially, doubling approximately every two years. This has led to commensurately rapid growth in the availability of inexpensive computing power and data storage. The past decade has also witnessed a similarly rapid expansion of the quiver widely available of statistical and machine learning tools and methods. For example, the contributed packages to the popular, open-source statistical computing language R has grown exponentially for several years—a “Moore’s Law” for statistical learning algorithms. As a result, data scientists can visualize, explore, and analyze complex datasets with ever greater sophistication.

At the same time, vivid examples of analytics-driven innovation have captured the imaginations of business leaders across domains. In professional sports, the Oakland As used data analytics to tap into an inefficient market for talent and cheaply hire valuable players neglected by richer teams.¹ In politics, the 2012 Obama reelection campaign paired sophisticated database marketing techniques with insights from the behavioral sciences to identify and effectively prompt persuadable voters to cast their ballots for Obama.² In entertainment, Netflix is able to analyze rich databases of viewing patterns for relevant associations among titles, directors, and actors. One such analysis revealed a significant group of viewers who enjoy movies starring Kevin Spacey, films directed by David Fincher, as well as a certain politically themed BBC miniseries. Netflix's remake of *House of Cards* was thus built on a firm foundation of data.

In short, data analytics has become a prominent, permanent feature of a business landscape that Moore's Law has forever altered.

But the changes ushered in by the age of Big Data do not end with advances in tools, methods, and executive-level recognition of business analytics. Society as a whole is experiencing changes that are no less dramatic. Indeed, one of the signature issues of our time is the ever-increasing extent to which our everyday activities are being digitally mediated and recorded. Each time we text a friend, do an online search, stream a video, make a credit card purchase, pick up a book on an e-reader, map directions using a mobile app, buy groceries using a clubcard, "like" a piece of online content, or make a personal or professional network connection, we leave behind a digital trace—a transactional record in a database—recording the activity. The change isn't simply that organizations know more about us than they used to. It is that the degree to which they can learn about us is potentially boundless because so many personal aspects

¹ See *Moneyball* by Michael Lewis.

² See *The Victory Lab* by Sasha Issenberg

of our lives are now digitally mediated. Paraphrasing McLuhan, the digital medium of personal data collection is the message.

Personalized marketing promotions and collaborative filtering-generated book, movie, song, and news article recommendations are now part of everyday life. But they mark only the beginning of what is possible. The digital mediation of our daily activities can also bring forth unexpected innovations, mash-ups, and new business models. For example, a large supermarket chain is able to use its clubcard data to predict which of its customers is at high risk for diabetes.³ Insurance companies use personal credit scores (originally invented to underwrite mortgages) as behavioral proxies that turn out to be surprisingly predictive of who will crash their cars. The University of Cambridge psychometrics lab reported being able to predict both sexual and political orientation—as well as basic personality types—using nothing more than social media “like” data.⁴

While personal privacy concerns are never far from the surface of such applications, one can imagine socially desirable innovations as well. Returning to our examples, the previously mentioned supermarket chain makes its diabetes predictions in order to encourage at-risk customers to seek medical advice. Some auto insurers are now collecting telematics data that capture moment-by-moment details of their insured’s driving behavior. Such data can certainly be used in the traditional ways: to make ever-more refined promotion, underwriting, and pricing decisions. But one could also imagine pioneering insurers offering a new type of data product: personalized “report cards” detailing drivers’ specific strengths and weaknesses, designed to help student drivers learn, experienced drivers remain sharp, and elderly drivers safely stay behind the wheel longer. One can even imagine

³ See “Tesco: Clubcard data could be used to help shoppers make healthier eating choices”, *The Grocer*, May 23, 2013.

⁴ See “Private traits and attributes are predictable from digital records of human behavior” by Michal Kosinska, David Stillwell, and Thore Graepel, *Proceedings of the National Academy of Sciences*.

the sorts of analyses performed by the Cambridge psychometrics unit being used—on an opt-in basis—to make friendship or matchmaking recommendations.

Such digital innovations must be evaluated and implemented using not only economic and technological frameworks, but regulatory, societal, and ethical ones as well. The MIT Media Lab professor Alex “Sandy” Pentland remarked that because “digital breadcrumbs” are about people, “there are enormous issues about privacy, data ownership, and data control. You can imagine using Big Data to make a world that is incredibly invasive, incredibly ‘Big Brother’... George Orwell was not nearly creative enough when he wrote *1984*.”⁵ Regardless of where one draws the line between the usefully personalized and the excessively personal, it is clear that not all technologically possible data innovations will be socially desirable ones. Companies, consumers, and regulators will need to work together to determine where the lines should be drawn. David’s book is a worthy entry in a much-needed new genre.

—**James Guszcza, Deloitte Consulting**

⁵ See the edge.org conversation with Alex (Sandy) Pentland: “Reinventing Society in the Wake of Big Data.”

Preface

We are constantly being watched. I don't intend this as hyperbole or as a means of criticizing national security efforts. Nor is this intended to be part of the description for a television series or movie. Rather, I say this as a statement of fact. From our website browsing to our shopping behavior, such routine activities generate volumes of data that are archived by both public and private entities. From a marketing perspective, the "data exhaust"—the data we generate through our activities—has the potential to be a valuable resource, a new source of insights into which organizations can tap to learn about consumers. But, with our activities monitored by retailers, advertisers, and social media platforms, a legitimate question arises as to what individuals should expect in exchange for unfettered access to their data exhaust. While the exchange occurs between the consumers providing their data and the organizations making use of it, recent reports from government agencies and proposed legislation suggest that efforts to capitalize on consumers' data exhaust may meet regulatory scrutiny.

These three parties—consumers, businesses, and regulators—will all be active players in the data economy. The building of the data economy is still in progress, and it changes on a regular basis. In some cases, these advancements are minor. In other cases, the landscape is irrevocably altered, leaving us to question where we go from here. In this book, I offer a look at the role of the three players in the data economy and how each—now and in the future—affects how data-driven insights will reshape marketing practice.

1

Beyond Big Data

While there has been a lot of discussion around the term “Big Data,” much of the discourse treats this as an abstract idea rather than a system in which we are all active participants. While the term has become ubiquitous, interest in the topic has not waned. A Google search for the term turns up approximately 13.7 million search results. A snapshot of Google Trends reveals the meteoric rise of queries for Big Data beginning in 2011 and increasing ever since then.¹ Searching archived Twitter messages using Topsy reveals more than three million tweets referencing Big Data and in excess of two million tweets mentioning #bigdata.

Some contend that this marks a dramatic shift in what businesses and organizations are capable of doing. Others deride or critique it. Author and Duke University professor Dan Ariely likened Big Data to teenage sex: “Everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.”² Whether you’re an ardent believer that the Big Data revolution is the best thing since sliced bread or you’re skeptical of the buzzword, there is no disputing the fact that more attention is being paid to the topic.

This attention isn’t just coming from corporate broom closets where statisticians are huddled over computers and poring over the output from complex analyses. Instead, data and analytics are garnering attention in the C-suite. In some companies, these topics fall

under the purview of the CIO, whereas they are part of the CMO's responsibilities for other companies. We've also seen the emergence of the chief data officer. Although you might expect to see this position at a financial institution or a company based in Silicon Valley, advertising juggernaut Ogilvy and Mather appointed its first global chief data officer in August 2013.³

The interest in capitalizing on the abundance of data extends beyond the boardroom to the public sphere. Microtargeting in political races was used as early as 2004. Local governments are also getting in on the act, with Philadelphia Mayor Michael Nutter naming the city's first chief data officer in 2012 and New York City Mayor Michael Bloomberg appointing the city's first chief analytics officer in 2013.

Searching for the Next Generation of Quants

Why are companies, campaigns, and governments focusing on individuals with a knack for data? Probably the same thinking that led Hal Varian, an emeritus economics professor from the University of California at Berkeley and chief economist for Google, to proclaim in 2009 that "the sexy job in the next ten years will be statisticians."⁴ If you're able to extract insights and act upon them, they can provide a strategic advantage. Across different types of organizations, statisticians can contribute immensely to improving operations, from increasing efficiency and cutting costs to increasing revenue. However, being well versed in statistics isn't enough. What many organizations are seeking is a data scientist with the holy trinity of skills: someone with expertise in a particular field, coupled with knowledge of sophisticated statistical tools, and the technical expertise to

develop and implement these algorithms on a large scale. This is often depicted as a Venn diagram (see Figure 1.1).

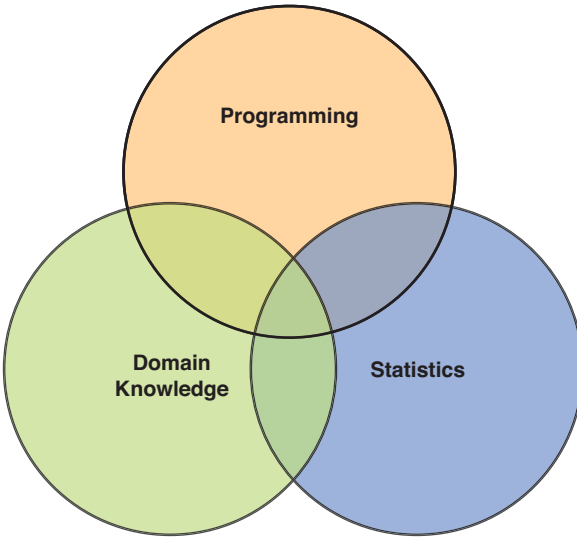


Figure 1.1 Data Science Venn diagram

Are they searching for mythical unicorns? Not necessarily. Both presidential campaigns in 2012 had chief data scientists. According to the vice president of Big Data products for IBM, a data scientist is “part analyst, part artist.”⁵ It’s not enough to crunch numbers in the background. The findings from advanced analytics are only as good as the way in which the insights are communicated to key decision makers. When we talk about using marketing analytics to guide strategy, we’re not just talking about a set of curve-fitting exercises. Instead, we’re talking about storytelling, informed by data, which has the potential to inform decision makers.

Are these three skills all essential, or can we get by with someone who is lacking one of them? With knowledge of statistics and the ability to code, a researcher can make data sing. However, although

this may be sufficient from the standpoint of conducting research, what practical decisions can they support? Without a sufficiently deep understanding of the domain in which they're operating, the impact of the insights on strategy will be limited. Meanwhile, someone who understands statistical models and knows the domain in which they operate is valuable from the standpoint of producing insights, but they are limited in their ability to convert those insights into a scalable solution. We similarly run into problems if individuals are fluent computer scientists but lack knowledge of the statistical models that are often used for evaluating business decisions.

Universities are making an effort to address the significant talent gap. Many have started to offer programs that tap into the interest in Big Data and data science. North Carolina State University, through its collaboration with SAS, launched the Institute for Advanced Analytics and offers an M.S. in analytics. At New York University, you'll find the Center for Data Science offering an M.S. in data science, as well as an M.S. in business analytics offered by NYU Stern School of Business. The marketing department at the Wharton School of the University of Pennsylvania houses the Wharton Customer Analytics Initiative, while the operations and information management department offers a track in business analytics. Northwestern University offers an M.S. in predictive analytics through the School of Continuing Studies, while Northwestern Engineering houses the M.S. in analytics. You'll also find an M.S. in analytics offered by the University of San Francisco and an M.S. in marketing analytics at the University of Maryland.

From this small sampling of the programs and the initiatives that have developed in higher education around analytics, you can start to see why it's difficult to prepare students for the roles that organizations are seeking to fill. There's a fundamental question about where the appropriate training for dealing with data takes place. The two logical schools in which programs focusing on Big Data would emerge are business schools, where the insights have the potential to

guide decisions, and computer science and engineering departments, where the technical tools may take center stage rather than strategic decisions.

From Big Data's Past to Its Future

The two sides of the Big Data coin may go back as far as the origin of the term. According to *The New York Times* story by Steve Lohr, author Erik Larson wrote a piece for *Harper's Magazine* in 1989 that was subsequently reprinted in *The Washington Post*.⁶ In it, Larson discusses the direct marketing industry and its practice of merging different facets of consumer data. Decades before data privacy and transparency came into the vernacular, Larson suggests that the “keepers of big data” will “track you for the rest of your consuming life—pitch you baby toys when you're pregnant, condos when you're fifty.”⁷

Interestingly enough, Lohr doesn't ascribe credit for the term “big data” to Larson. Instead, he argues that credit should go to John Mashey, chief scientist for Silicon Graphics in the 1990s. When queried about the use of the term, Mashey told Lohr that “I was using one label for a range of issues, and I wanted the simplest, shortest phrase to convey that the boundaries of computing keep advancing.” Lohr's logic in giving Mashey credit for coining the term was that it “should go to someone who was aware of the computing context.”

Regardless of to whom you feel credit is due, this exercise into the ancestry of one of today's most common phrases within technology and business circles offers a curious perspective. Larson exposes the potential benefits and risks associated with compiling disparate pieces of data. Like a jigsaw puzzle, each individual piece of data does not reveal much about the person who produced it. But assemble enough of these pieces, and the full picture begins to emerge. The image that Larson describes is one based on mortgage and tax records, consumer

surveys, and records from the Census Bureau. Compared to what we're capable of seeing today, what Larson envisioned may appear to look like a five-year-old's finger painting tacked onto a refrigerator, hanging next to a digital photograph taken with the latest digital SLR camera on the market today. The possibilities that he described pointed in the direction of the path that we eventually traveled. However, we have gone leaps and bounds beyond what many would have imagined in the past.

For that, credit is certainly due to technology. Dramatic increases in processing power have paved the way for sophisticated marketing analysis to be conducted on desktop computers instead of mainframes. Rather than speaking in terms of census tracts, we can now talk in terms of individual households or, better yet, members of the household. Retail purchase records from scanners in grocery stores, once heralded as a major advance for the marketing profession, pale in comparison to the amount of data produced by consumers visiting websites and making purchases online. None of what we are talking about today would be possible were it not for technological advances. At the same time, credit should be given to those who had the foresight and saw potential opportunities to put such granular data to use. These two forces, working in parallel, heralded the age of Big Data.

Characterizing Big Data

One of the most frequent ways in which Big Data is defined is in terms of the 3 V's: volume, variety, and velocity. As the term implies, anything falling under the umbrella of Big Data is large in size. Some have taken this to mean anything that is too voluminous to be stored on a desktop computer. Multiple students have expressed interest in studying Big Data, which I've generally taken to mean that they are interested in using software that can handle more rows than a single

spreadsheet in Microsoft Excel can accommodate. (Excel has a current limit of just more than one million rows—a vast improvement to the 65,536 rows of the previous version of the software.) In addition to the size of the data being stored, Big Data often comprises a variety of formats. On top of the quantifiable data (or “structured data”), Big Data encompasses “unstructured data” such as text comments, images, and multimedia file types.

Perhaps the most important of the 3 V’s is velocity. It is the velocity with which the data are collected and must be processed that can separate Big Data problems from those that simply involve large quantities of data. One illustration of problems that require data to be collected and processed rapidly is real-time marketing. Whether it is an auction to determine the order in which advertisements will appear alongside search results or targeted messages based on an individual’s mobile browsing history and current location, such problems require that data be processed rapidly.

Walt Disney World maintains an underground bunker called the Disney Operational Command Center to ensure that the theme park operates smoothly. Its staff may attempt to increase the speed with which visitors are moving through the queue of a ride if they find that it is too long, or dispatch costumed employees to entertain guests while they wait.⁸ If visitors are moving through rides more efficiently and having a more pleasant experience at the park, they’re apt to have more time during which they can visit stores and restaurants in the amusement parks, which provides Disney with additional revenue from its visitors. Phil Holmes, vice president of the Magic Kingdom, noted that “if we can increase the average number of shop or restaurant visits, that’s a huge win for us.”

While the majority of cases discussed in this book focus on marketing applications, perhaps one of the most important applications of Big Data today is security. Although Disney’s objective may be

to minimize wait times or maximize visitors' expenditures, imagine if we could apply similar approaches to law enforcement? Consider the FBI's development of its Next Generation Identification (NGI) system, which will explore the use of facial recognition tools.⁹ Such a system could be used after crimes have been committed, comparing the footage from security cameras to databases that have been compiled previously. If such technology could be deployed more rapidly, it has the potential to reduce the resources that need to be committed to pursuing offenders. For such tasks, time is of the essence. Faster processing of the available data can contribute not only to cost reductions but also to improvements in public safety.

Although the 3 V's are common to problems that fall under the auspices of Big Data, they ignore at least two other critical factors. First, there are issues as to the veracity of the data. No one questions that there is a lot of data available, but organizations trying to cut through the noise and identify the signal must ask themselves if they can trust the different data streams available. When decisions are being based on the results of data analysis, the findings are meaningless if the data on which they are based is biased in some way.

This has been one of the concerns raised by marketers about the potential use of social media data. Although such data are generally available, do the comments scraped off the Web reflect the thoughts of a brand's entire customer base? Neglecting to account for known biases in social media data could contribute to problems such as overestimating the importance of an issue to consumers or failing to capture shifts in brand sentiment.¹⁰

Second, and more important, is the value of the data. Many organizations talk about having a Big Data strategy. If they're referring to a plan to warehouse and access data relevant to their organization, there's nothing wrong with this statement. In fact, more organizations would probably benefit from having a well-thought-out strategy that integrates the IT function with the appropriate business processes.

The problem is, though, that they're often not referring to how data will be stored and made available to users. Instead, they're using the term "Big Data" as a crutch. Rather than thinking through what they are trying to achieve and gathering data that are appropriate to addressing those goals, they believe that they have a foolproof strategy: track "everything."

Is Big Data a Strategy?

Here's the problem with this so-called strategy. Suppose that reams of data are captured. All that we've done is kick the can down the road. Deep in the recesses of the warehouse we've built, in the multitude of haystacks we've collected, is the golden nugget that we've been looking for. The problem is that we now have to go through each of those haystacks. And, more likely than not, we'll find something. Whether we can do anything with what we find is another story. The focus on collecting more and more data has obscured what we should have been asking ourselves from the start: What actions are we going to take?

It's true that data are necessary to derive insights, and those insights inform the actions we take. However, thinking strategically requires that we work backward. Asking first what it is that we're trying to do, we can then identify the insights needed to inform such actions. Based on the insights we need, we can back into the data that are needed to yield such insights. If we start blindly by compiling data without considering where we're trying to end up, we run the risk that we've created more work for ourselves because we now have to sift through mountains of irrelevant data that have been captured in our dragnet.

The challenges faced by many organizations, from city governments to publically traded corporations, don't require Big Data.

Rather, these organizations should be taking a look at the key issues they're facing and considering how those issues can be investigated. It's not that more data are necessarily better. Sometimes more data are just more.

What the conversation should be focused on are the data that will lead to bigger insights. Sometimes this does in fact require more data or different types of data. In other cases, it requires rethinking the assumptions we currently hold and applying a different type of analysis. In specific circumstances, Big Data may be the raw fuel powering these insights. However, on its own, Big Data doesn't tell us what we should be doing. It doesn't tell political campaigns in which media markets to place advertisements. It doesn't tell retail stores to whom to send coupons. It doesn't tell city agencies how best to use their limited resources. Making such recommendations is where the artistry in data science comes into play. If Big Data is a natural resource, then the advanced statistical tools employed by analytic professionals and data scientists serve as the means of extracting and refining the raw material into something of value.

A colleague has suggested that Big Data is like "rocket science." There are actually people who are doing rocket science, but the common use of the phrase extends beyond the well-credentialed few. There are indeed organizations that are knee deep in Big Data, but the phrase has become a catchall for most things involving data. Like it or not, we are living in the age of Big Data.

Data Versus Insights

For all the interest in Big Data, this is not a book about Big Data. This book is about what can be done with Big Data. More accurately, it's about what can be done with data, both the good and the bad. When Larson first mentioned the "keepers of big data" back in 1989,

companies such as Google, Facebook, and Amazon.com didn't exist. And yet, marketing analytics and targeting were alive and well—perhaps not as efficiently as they are implemented today, but the ability to draw a portrait of a person based on multiple data sources was practiced by Claritas, Inc., which was eventually acquired by marketing research behemoth Nielsen.

The insights produced to serve marketers didn't hinge on the amount of data available to them. Instead, it was based on understanding how the available data could be put to use. Larson tees up ideas that are relevant even today, if not ahead of their time, such as permission-based marketing and consumers receiving compensation in exchange for their data. And yet, he determines that he isn't concerned because the organizations involved in direct marketing and the entities that provide organizations with access to consumer data “don't really know what they're doing—at least not yet.”

What makes the present day different from 1989? For one thing, organizations are much more versed in what can be accomplished with the right data. In addition to knowing what can be done with data, there's also more data available. The data exhaust—the digital trail produced by consumers through their everyday activities—is a potential goldmine. Although there are ample opportunities for organizations to leverage available sources of data, let's not assume that consumers are victims. Consumers have options at their disposal, and their choices have the potential to separate the winners from the losers.

Organizations of all stripes can be developed or refined based on the insights afforded by consumer data. As organizations have come to realize the tremendous potential that can be extracted from consumer data, the winners will be determined by those who have access to the data they need—that is, those who provide a compelling reason for consumers to share this data with them. Focusing on businesses, this is predicated on an exchange. Neither party has a gun to the head

of the other. Consumers and organizations are both willing participants in this data economy.

Data and Value

As with other economic systems, at the heart of the data economy is value. As we will discuss, organizations have made it clear that they see tremendous value in consumer data. Some have explicitly said as much. Others have revealed it through their actions. Depending on the organization we're discussing, there are a number of ways in which detailed data have been turned into value.

Collecting and storing data should be viewed as an investment. As with other investments, the question that organizations should be able to answer is, what is the payoff associated with compiling databases? Police forces, for example, have turned to predictive analytics. In addition to the hardware and software investment, there's a cost associated with hiring the analytic talent to conduct the necessary analyses. Here, the payoff can be viewed as increased public safety, as manifested through a reduction in crime.

Google demonstrated that certain search terms are correlated with flu activity.¹¹ From a public health perspective, such information could be useful in determining when it is most essential to ramp up efforts to encourage individuals to get vaccinated. Health insurance providers may take such efforts upon themselves, looking at the potential savings associated with reducing the number of hospitalizations. Employers may also promote vaccinations, hoping to curb the amount of worker productivity that is lost due to workers taking time off to recuperate. Viewing this problem from another perspective, pharmaceutical companies could identify the value of such data if it would enable them to make more efficient use of their marketing budgets. As you can see from this one example, the potential value

associated with a particular piece of data depends on the organization's goals.

Although these two illustrations demonstrate what can be gained by organizations turning to data that is generally available, consider briefly data that may not be available for public consumption: a consumer's purchasing habits. Consider simply the question of how strongly you prefer Coca-Cola to Pepsi. If Coca-Cola knew which consumers were only interested in its products, which consumers were only interested in Pepsi's product, and which consumers did not have a strong preference, it might change the way it approached marketing to each of these different consumers.

It might decide, for example, to spend just enough on marketing to loyal Coca-Cola consumers to encourage them to purchase more frequently. For these consumers, though, the company is not worried about them switching over to their competitor. For those consumers loyal to Pepsi, it may not make sense for Coca-Cola to exert any effort marketing toward these individuals. If their brand preferences are so strongly in favor of Pepsi over Coca-Cola, there's little that Coca-Cola would be able to do to sway them. For the consumers in the middle, perhaps that's where Coca-Cola's (and Pepsi's) marketing efforts have the potential to have the biggest impact.

Coca-Cola and Pepsi, as well as all other publically traded companies, are in the businesses of what's best for their shareholders, but we could also apply the same thinking to the presidential election. Across the country, there are states that are deep blue and there are states that are deep red. Although a candidate could pour money into advertising in those states where his party has not fared well historically, barring a huge shift in the demographics of the state, such advertising expenditures are not expected to yield much of a payoff. Instead, what we are left with is a deluge of advertising concentrated in battleground states, specifically in counties where the advertising is expected to yield the biggest impact.

As we'll discuss in more depth, not all data are equally valuable. Detailed data about the television programs viewed by a voter who lives in Wisconsin's Dane county are likely to be of less value compared to the same data about a voter who lives in Ohio's Hamilton county. Some pundits considered both Ohio and Wisconsin to be battleground states, so why the difference in the likely value of data from voters in these two counties? Dane county leans heavily to the left, as reflected by the 71.1% of the vote received by President Obama. In contrast, President Obama received only 51.8% of the votes coming out of Hamilton county. Political advertising can exert some sway on voters, but there's a limit to its effectiveness. Given the strong leaning of Dane county, not much could have been done there by either party to sway voters. Hamilton County, in contrast, was identified as one of the seven most important counties in the 2012 election by *The Washington Post*.¹² If campaigns knew the programs that different types of voters in Hamilton county were watching, such data could be used to ensure that advertising occurred in the programs viewed by the voters they were most interested in reaching. Regardless of context, the determining factor in how much data are worth to an organization is based on what the organization can do with the data and whether having the data can potentially further the organization's goals.

Although the ability to take actions that affect consumer behavior is necessary for data to be of value, a few other conditions will affect just how valuable an individual's particular data are to an organization. First, there needs to be a sufficient number of consumers who are "like you" with regard to your preferences and attitudes. If your outlook is so idiosyncratic that an organization can't identify other consumers who are similar to you, it's simply too inefficient for that organization to acquire your data and take actions tailored to you. There just isn't the scale for this to be viable. Fortunately, it turns out that consumers are not as different from one another as they might think. Or, more precisely, they are similar enough that they can be

grouped together into consumer segments, enabling organizations to pick and choose the segments on which they want to focus their efforts.

In addition to there being enough consumers like you, how much an organization is willing to spend on your data hinges on how valuable you (or the segment to which you belong) are to the organization. Frequent travelers are of interest to the airline and hotel industries because of the volume of business they generate, so they may provide these individuals with a separate telephone number for customer service, express check-in, or other perks. Casinos pay particular attention to their high rollers and provide them with a number of complimentary offerings to attract and retain their business because the amount that the casino stands to gain from the individual's gambling activities can be quite substantial. Undecided voters in swing states are targeted because they can decide which presidential candidate receives the state's electoral votes. In short, the segment has to matter to the organization.

With all these factors in place, ultimately the value of your data to organizations depends on how readily available such data are from other consumers. Although all consumers may be distinct from each other, the intent in forming a small number of market segments is to identify groups of consumers who are similar enough to each other and sufficiently different from other consumers. Claritas put this into practice with its Potential Rating Index for Zip Markets (more commonly referred to as PRIZM) segmentation scheme.

Members of the Executive Suites segment, for example, tend to place orders at barnesandnoble.com, play golf, and watch *Saturday Night Live*. This segment consists of upper-middle class singles and couples who typically work white collar jobs. In contrast, although the Bohemian Mix segment falls into the same age range, they are more inclined than the Executive Suites to live in cities and more likely to have children at home. Their media and lifestyle habits also

differ—they express an interest in foreign films, are more likely to rent rather than own their home, and are more likely to read *GQ*.

Organizations must determine which of the segments are of interest to them so that they may focus their efforts on those segments. Once that has been determined, though, it doesn't matter which particular individuals from the segment provide data to the organization. If other consumers like you are willing to share their data with organizations at no cost, then that's how much the organization should be willing to spend on data acquisition. However, if each individual in a segment has determined that there is a minimum the organization must offer for someone to be willing to share the data, then it may be in the organization's interests to invest in acquiring this data.

Value for Value

The ideas being put forth here are not new. The core ideas that we'll discuss are at least 20 years old. However, since these ideas were first introduced into the public sphere, the landscape has irrevocably changed. Disturbingly, some of the ways in which these changes have occurred may have gone unnoticed.

First, consider the digital marketing platforms available to organizations today. Although consumers may be most familiar with Google as a search engine or a "free" email provider, at the end of the day, Google is an advertising platform. Likewise, although we may think of Facebook and Twitter as social networking products, these too are advertising platforms. We don't pay a monthly fee or a one-time charge for the rights to use these tools. Rather, we pay every time we use the tools—each query entered into Google, each time we send an email message, each time we view a video on YouTube. With every Facebook post and each tweet we send, we're paying these companies.

Online platforms will give way to the Internet of Things and connected devices. The notion of the “quantified self” is only beginning to become concrete, with fitness products and other forms of wearable technology. In exchange for what we learn about ourselves based on the data we generate, perhaps in pursuit of our personal goals, we adopt technology and share our data with the developers of the technology. In doing so, we’re paying with our actions, with the data that we produce. And there’s absolutely nothing wrong with that, so long as we’re aware of what we’re doing.

Consumer awareness took on broader significance with the revelations offered by Edward Snowden’s disclosure of documents that detailed the surveillance activities undertaken by the National Security Administration (NSA). Many had talked in vague terms about “big brother” watching us. Set in the aftermath of 9/11, the CBS drama *Person of Interest* presents a machine that can identify those involved in crimes before the events take place, gathering data through traffic cameras and other devices. In 2008’s *The Dark Knight*, Batman builds a machine that turns every cell phone into a microphone, allowing him to build a citywide surveillance system to aid him in finding where the Joker is located. Like these fictitious examples, the NSA’s surveillance activities may have been well intentioned. However, in the wake of the information that has come to light, there is increased interest in organizations being transparent about the data they are collecting and how such data are being put to use.

The interest in data transparency permeates the marketing field, as well as the regulatory environment in which companies must operate. In 2012, the Federal Trade Commission (FTC) issued orders to a number of companies to disclose information about data collection and usage practices.¹³ In a March 2012 report, the FTC urged companies to improve transparency in their data collection efforts. In the same report, there is a call to increase consumers’ knowledge about the data practices in which companies engage.

Despite the growing interest and potentially increased scrutiny, how exactly will organizations' data collection and usage practices change from what they have been historically? And, more important, do consumers care? Whereas the answer to the first of these questions is something of a moving target, the second question is more important. After all, to what extent do consumers need protection against large organizations' data-hording activities if they are informed of such activities and are not fazed by it?

Forrester examined the second of these questions in a 2012 study.¹⁴ Distinguishing between behavioral data (such as your online browsing history or your purchase history at a retailer) and individual data (for example, your address or Social Security number), it found that all age groups were more concerned about how companies used individual data, with older consumers more wary than younger consumers. The study also found that exclusive deals attract a majority of young consumers to share their data with companies, but the appeal of such incentives do not attract older consumers nearly as much. Mirroring this, the study found that a larger share of older consumers do not complete an online transaction because of something they read in the company's privacy policy.

So what can be gleaned from these findings? It turns out that consumers are actually concerned with how companies are making use of their data. Although that might seem like a dark cloud hanging over marketing, it's important to note that consumers' concerns vary with the type of data being collected and across different consumer groups—and these data collection practices do appear to impact consumer expenditures.

As we'll discuss, one option for organizations is to collect the data about which consumers are less sensitive and hence more willing to share. However, the key is for the organization to engage consumers like they would approach another collaborator. The consumers have something that is of value to the organization. With access to such information, organizations may change their practices. Marketing

expenditures may be made with increased precision. Product lines may expand or contract. Consumers may see a more targeted message. New businesses may develop to meet the consumers' needs. All these practices may rely heavily on access to the right consumer data.

If we're willing to share data with companies, consumers have much to gain, as do the businesses that develop if they are successful. It is simply a matter of an exchange—consumers are more willing to share their data if it is clear what they are getting in return. That payment may be in the form of financial compensation or access to a product or service. The idea of an exchange, where value is traded for value, is straightforward enough. However, as recent events have revealed, a host of issues are related to conducting such a transaction.

At the same time, there are also issues focused on the protections that must be in place for consumers. Questions about the information that organizations should present to consumers and the measures that should be taken to ensure that consumers' data are secure need to be addressed. Just as there is the potential for businesses to grow larger than we would have imagined, fueled primarily by innovative thinking about how consumer data can be used, there is also potential for such data to be exploited and considerable harm levied against consumers. And because of this, there is a legitimate question as to what the role of the government should be. Is a watchdog agency akin to the Consumer Financial Protection Bureau needed to oversee such exchanges? Alternatively, should market forces dictate who will and will not have access to consumer data and how much it is worth?

The next few chapters look at success stories that have emerged in business, where a good part of the companies' successes have been thanks in no small part to the data provided by their consumers. We'll also see how existing businesses have refined their practices by leveraging more detailed information about consumers. In addition, we'll see the benefits that the public has accrued through the innovative work of government agencies using many of the same tools employed by businesses.

With these exemplars of what is possible as a backdrop, we'll cast a critical eye on the current exchange in which consumers and organizations participate. Although this model functions well in some regards, we put forth reasons that suggest that it may be in need of updating. There are some signs that these revisions are already underway. Based on these harbingers, we'll discuss the structure that may be needed to support the burgeoning data economy.

Taking into account the events that have unfolded in recent years—from Internet giants and social networking sites coming on the scene, to interest in data, infographics, and statistics going mainstream—this book is not an attempt to predict what course we'll chart in the near future. However, given how we have seen organizations use data and current activities, there are indicators that suggest a potential direction—one in which consumers, businesses, and the public all stand to benefit.

Endnotes

1. <http://www.google.com/trends/explore#q=%22big%20data%22>
2. <https://www.facebook.com/dan.ariely/posts/904383595868>
3. Kaye, Katie (2013), "Ogilvy Chief Data Officer Role May Be Sign of Things to Come," *Advertising Age*, August 13, accessed at <http://adage.com/article/datadriven-marketing/ogilvy-chief-data-officer-role-sign-things/243713/>
4. http://www.mckinsey.com/insights/innovation/hal_varian_on_how_the_web_challenges_managers
5. <http://www-01.ibm.com/software/data/infosphere/data-scientist/>
6. Lohr, Steve (2013), "The Origins of 'Big Data': An Etymological Detective Story," *The New York Times*, February 1, accessed at <http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/>

7. Larson, Erik (1989), “What Sort of Car-rt-sort Am I,” *Harper’s Magazine*, July, pp. 64–69.
8. Barnes, Brooks (2010), “Disney Tackles Major Theme Park Problem: Lines,” *The New York Times*, December 27, accessed at <http://www.nytimes.com/2010/12/28/business/media/28disney.html>
9. http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/ngi
10. Schweidel, David A. and Wendy W. Moe (2014), “Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice,” *Journal of Marketing Research*, 51 (4), 387–402.
11. <http://www.google.org/flutrends/about/how.html>
12. Blake, Aaron (2012), “The 7 most important counties in election 2012,” *The Washington Post*, November 6, accessed at <http://www.washingtonpost.com/blogs/the-fix/wp/2012/11/06/the-7-most-important-counties-in-election-2012/>
13. <http://www.ftc.gov/news-events/press-releases/2012/12/ftc-study-data-broker-industrys-collection-use-consumer-data>
14. Bernoff, Josh (2012), “Turns Out Consumers Really Do Care About the Data You’re Collecting,” *Advertising Age*, January 25, accessed at <http://adage.com/article/digitalnext/turns-consumers-care-data-collecting/232331/>

This page intentionally left blank

Index

A

- Abine.com, 248
- aboutthedata.com, 181
- accountability, television
 - advertising, 62
- ACLU, digital privacy, 251
- ACR (automatic content recognition), 60
- Axiom, 26-27, 180
 - Epsilon, 182
- Ad-free Outlook.com, 174
- advertisements
 - Google, 37-38
 - social networks, 39
- Advertising Age, 54, 169
- AdWords, Google, 215
- Affordable Care Act, 190
 - segmentation, 193-194
- AFP (Agence France-Presse), 142
- Aguilar, Luis, 246
- airline industry, pricing, 154
- algorithms, Facebook, 245
- alumni, donor relations, 113
- Amazon, 44
 - customer reviews, 33-34
 - data economy, 98
- Amazon Prime
 - loyalty programs, 34-35
 - personalization, 237
 - price increase, 35
- America Online (AOL), 175
- American Civil Liberties Union, 191
- analyzing public data, 80-83
- Annenberg Public Policy Center, 160
- Anti-Marketing, survey of consumers' preferences, 165
- Anti-Personalization, survey of consumers' preferences, 164
- Apple
 - iBeacon, 69
 - medical field, 241
- apps, 118
 - crowdsourcing, 90-92
 - television advertising, 60
- Arbitron, 56
 - pay-to-play model, 205
 - consumer data, 247
 - hotspots, 106-107
 - opt-in programs, 139
- Atlanta Police Department, 79
- audio-recognition feature,
 - Facebook, 217
- auto insurance, segmentation, 195
- auto insurance companies, driving data, 101-102
- automatic content recognition (ACR), 60
- Automatic Labs, Inc., 101
- Automatic Link device, fuel efficiency, 101-102

B

banks, 213
 Bayes, Reverend, 132
 Beacon, Facebook, 135-138
 Beckham, David, 217
 behavioral data, 18, 131
 behavioral tracking, 134
 behaviors that affect value,
 customers, 110
 beliefs and interests, marketing, 23-25
 Bernard, Julie, 125
 Berners-Lee, Sir Tim, 245
 Beso, 140
 Best Buy, 29-31
 Bezos, Jeff, 52
 personalized pricing, 159
 Big Data, 1
 characterizing, 6-9
 history of, 5-6
 as a strategy, 9-10
 Bloomberg, Michael, 2
 Blue CRUSH (Criminal Reduction
 Utilizing Statistical History), 79
 Boehner, House Speaker John, 161
 Bohemian Mix segment, 16
 Booker, Mayor Cory, 91
 Boston Marathon bombing, 85-86, 88
 brand-building strategies, television
 advertising, 62
 Brynjolfsson, Erik, 77, 91
 Bureau of Consumer Protection, 227

C

casinos, 15
 cell phones, 107
 privacy, 249-251
 Census Bureau, 27
 Center for Data Science, 4
 CFPB (Consumer Financial
 Protection Bureau), 224-225

Change.org, 207
 changing messages as people
 change, 66-69
 characterizing Big Data, 6-9
 Chip-and-PIN technology, 185
 churn, 128-129
 Citizens Connect app, 91
 civil rights, segmentation
 practices, 191
 Civil Rights Principles for the era of
 Big Data, 191
 Claritas, Inc., 11, 15
 clothing, demand, 151
 CLV (customer lifetime value),
 109-111
 Coca-Cola, 13
 collecting data, insurance industry,
 64-66
 college tuition, 194
 cost of, 161-162
 Comcast, 217
 NBCU+, 58
 compensation for data, 203-205, 220
 competition
 consumer data and, 239-241
 consumer protections, 227-228
 compiling data to inform the
 public, 88-90
 computerized decision making, 192
 concerns over data collection, 18-19
 consumer awareness, 17
 consumer data
 competition and, 239-241
 cost of, 185-189
 exchanging products and services
 for, 205-208
 as a financial asset, 218-220
 paying for, 203-205
 valuing, 169-170
 Consumer Financial Protections
 Bureau (CFPB), 225
 consumer liability, 187

Consumer Privacy Bill of Rights, 179-180

consumer protections, 221

 competition, 227-228

consumer segments, 15, 193

consumers

 empowering, 211-214

to regulate access to personal data, 228-229

 informing, 211-214

 mindshare, 62

 protecting, 196

 purchasing habits, 13

content analysis, 190

Coons, Senator Chris, 227

cost of

 college tuition, 161-162

 consumer data, 185-189

 data-driven innovation, 244-245

coupons, 111-112

CPG manufacturers, 169

credit card providers

 digital marketing, 29

 direct marketers, 28

credit scores, 41, 180

creditworthiness, 41-42, 190

crime, predicting, 78-80

CRM (customer relationship management), 23

crowdsourcing, 83-88

 apps, 90-92

Cukier, Kenneth, 82

current exchange, 100, 107

 Automatic Labs, Inc., 101-102

 Facebook, hotspots, 106

 Mint.com, 102-105

customer churn, 128-129

customer data, privacy, 125-127

customer lifetime value (CLV), 109, 110-111

customer preferences, in advance of hurricanes, 29

customer relationship management.

See CRM (customer relationship management)

customer reviews, 33-34

customer-centric marketing, 108-113

 investments, 114-118

customer-level transactional information, 128

customers

 behaviors that affect value, 110

 learning who they are, 127-130

 profiles, 145

Cycle Atlanta, 93

D

Dane county Wisconsin, 14

Dark Knight, 17

data

 compiling to inform the public, 88-90

 cost of, consumer data, 185-189

 versus insights, 10-12

 internal data, 27

 investments, 12

 sharing value of, 140

 value of, 9, 12-16

political advertising, 14

data acquisition, free-for-all model, 208-210

data aggregators, 213

data breaches, 124-126, 185

data brokers, 97

 data acquisition, free-for-all model, 208-210

data collection, concerns over, 18-19

data collection practices, privacy, 130-135

data economy, 97-100

 regulation, 220-224

data endowment, 168

data exhaust, consumers' comfort with leveraging, 163-169

Data Laissez-Faire, survey of consumers' preferences, 165

data literacy, 184

data ninjas, 78

data ownership, 141-144
 Facebook, 141
 Google, 142
 Twitter, 141

data restrictions, innovation and, 224

Data Science for Social Good - Atlanta, 93

data scientists, 3-5

data sharing, 160, 180-184

data transparency, 17

data-driven discrimination, 189-192

data-driven innovation, cost of, 244-245

Delete Me, Abine.com, 248

Delivery Agent, 64, 217

Dell, IdeaStorm, 87

demand, 151

demand-based pricing, 151-156

Democrats, advertisements, 57-58

demographics, 127-129
 personalized pricing, 159

Department of Veteran Affairs, 185

Desmond, Laura, 216

digital exhaust, 159

digital marketing, credit card providers, 29

digital marketing platforms, 16

digital privacy, 249-252

direct marketers, credit card providers, 28

Direct Marketing Association (DMA), 181

discounts, 111-112

Discover Bank, 211

Discover Financial Services, personalized pricing, 157

discrimination, data-driven, 189-192

Disney Operational Command Center, 7-8

dispersion, 84-85

DMA (Direct Marketing Association), 181

"Do Not Track" legislation, 131

donor relations, 113

Do-Not-Track Online Act of 2013, 131-134

Doritos, 87

double jeopardy, 235

driving data, 101-102

dynamic pricing, 154-155

E

educating consumers on what is being collected, 224-226

efficiency, 59

Elberse, Anita, 235

electric toothbrushes, 65

Electronic Fund Transfer Act, 187

empowering consumers, 211-214
 to regulate access to personal data, 228-229

Enoch, Glenn, 99

Epsilon, 182

Eric and Wendy Schmidt Data Science for Social Good fellowship, 93

ESPN, 99-100

European Union, right to be forgotten, 222-223

exchanging products and services for, consumer data, 205-208

Executive Suites segment, 15

Experian, 181

F

Facebook, 16, 45
 algorithms, 245
 audio-recognition feature, 217
 Beacon, 137-138
 data ownership, 141
 hotspots, 106
 innovation and, 238
 marketing, 39-40
 privacy controls, 176
 privacy policies, 135-136
 privacy settings, 207
 SecondSync, 59
 targeted advertising, 45
 transparency, 208, 246
 using credentials to log-in to other sites, 206-207
Facebook Beacon, 41
Facebook log-in tool, 211
FAFSA (Free Application for Federal Student Aid), 162
FCC, consumer protections, 247
Federal Trade Commission. See FTC (Federal Trade Commission)
Feinblatt, John, 89
finances, tracking, 102-105
financial assets, consumer data as, 218-220
financial service providers, 225
empowering consumers, 211
Financial Services and Markets Act 2000, 220
Financial Services Authority, 220
Firehose, 43
Fitbit, 241
Flowers, Michael, 81
Fourth Amendment, 227
free-for-all model, data acquisition, 208-210

FTC (Federal Trade Commission), 17, 227-228
 data brokers, recommendations, 210
 Facebook, privacy policies, 135
 Identity theft, 126
FuelBand, Nike, 241-242

G

geography, personalized pricing, 156-157
geo-targeting, 69-71
Getty Images, 142
Gillette, 68
Google, 16, 37-38, 239
 AdWords, 215
 data ownership, 142
 health, 242
 hotspots, 106
 location-aware advertising, 69
 marketing, 40
 privacy policies, 136
 Nest Labs, 137
 scanning emails, 175
 search terms, 12
Google Chrome, 215
Google Consumer Surveys, 204
Google Fiber, 247
Google Fit, 242
government, role in data collection, 19, 246-248
GPS monitoring, 226

H

H&M, 217
Hamilton county, Ohio, 14
Hammerbacher, Jeff, 78
Harvard, 162
Henn, Steve, 86
heterogeneity, 129-130

history of Big Data, 5-6
 Holmes, Phil, 7
 Home Depot, personalized pricing, 157
 homophily, 177
 hotels, collecting data, 51-52
 hotspots
 AT&T, 106-107
 Facebook, 106
 Google, 106
 Howe, Scott, 214, 248
Huffington Post, 207

I

IAB (Interactive Advertising Bureau), 133
 iBeacon, 69, 70
 IBM, 41
 Blue CRUSH (Criminal Reduction Utilizing Statistical History), 79
 IdeaStorm, Dell, 87
 identity theft, 125-126, 185
 individual data, 18
 informing consumers, 211-214
 InfoScout, 169
 innovation
 changing face of, 237-239
 data restrictions and, 224
 data-driven innovation, cost of, 244-245
 insights versus data, 10-12
 Instagram, 238
 Institute for Advanced Analytics, 4
 insurance industry
 collecting data, 64-66
 segmentation practices, 195
 interests and beliefs, marketing, 23-25
 internal data, 27
 Internet, 239
 internet-based businesses
 Amazon, 32-36
 Google, 37-38
 Pandora, 36-37

interpersonal influence, 43
 investments
 customer-centric marketing, 114-118
 data, 12
 It's All Good, survey of consumers' preferences, 165

J

Jankowski, Ben, 216
 Joly, Hubert, 30-31

K

Kalanick, Travis, 154
 Kantar Media, 56
 Kelly, Cameron, 221
 Kelman, Glenn, 78

L

larger brands versus small brands, 235
 Larson, Erik, 5
 Laurie, Hugh, 55
 lawsuits, Morel, Daniel, 142
 lead-generation companies, 28-29
 The Leadership Conference on Civil and Human Rights, 191
 liability, consumer liability, 187
 Liberty Mutual, 92
 life-event triggers, Experian, 181
 LinkedIn, 238
 Live Nation, 155
 location, personalized pricing, 156-157
 Location Privacy Protection, 226
 location-aware advertising, 69-71
 location-based data, 226
 Lohr, Steve, 5
 loyalty programs, Amazon Prime, 34-35

M

Macy's, 125
 Madrigal, Alexis, 86
 mailing lists, 27
 marketers, engaging in debate over data collection practices, 130-135
 marketing
 basics of, 23-27
 changing messages as people change, 66-69
 customer-centric marketing, 108-113
 on Facebook, 39-40
 Google, 40
 interests and beliefs, 23-25
 location-aware advertising, 69-71
 mobile advertising, 41
 one-on-one marketing, 155
 personal data, 160
 reaching today's consumer, 71
 on social networks, 38-44
 Twitter, 40
 marketing analytics, examples, 27-32
 Marshalls, 185
 Mashey, John, 5
 mass customization, 155
 mathematicians, 78
 Mayer-Schonberger, Viktor, 82
 McAfee, Andrew, 91
 McPhee, William, 235
 media, television industry, 52-57
 Memphis Police Department, Blue CRUSH (Criminal Reduction Utilizing Statistical History), 79
 Microsoft, Ad-free Outlook.com, 174
 microtargeting, 2, 32
 mindshare, 62
 Minority Report, 78
 Mint.com, 102-105
 mobile advertising, 41
 money, tracking, 102-105
 moods, Twitter, 99

Morel, Daniel, 142
 My Coke Rewards program, 207
 My Starbucks Idea, 87

N

NAACP, 191
 National Council of La Raza, 191
 National Science Foundation, 190
 National Security Administration (NSA), 17
 NBCU+, 58
 Neiman Marcus, 124
 Nest Labs, 243
 privacy policies, 137
 Netflix, 36, 44, 245
 data economy, 98
 New York City Fire Department, 81
 New York University, 4
 Next Generation Identification (NGI), 8
 Nielsen, 11
 pay-to-play model, 205
 viewer diaries, 55
 Nielsen Audio, 205
 Nielsen ratings, Twitter, 59
 Nike, FuelBand, 241-242
 No Harm in Data, survey of consumers' preferences, 165
 North Carolina State University, 4
 Northwestern University, 4
 NSA (National Security Administration), 17, 123, 179, 221
 Nutter, Michael, 2
 NYC Business Atlas, 89
 NYC Open Data, 81

O

Obama, President, 221
 Ocean Spray, 60
 Ocwen Financial Corporation, 183
 Office Depot, personalized pricing, 158

Office of Data Analytics, 82

NYC Business Atlas, 89

one-on-one marketing, 155

online platforms, 17

online profiles, 45-46

opt-in programs

AT&T, 139

privacy, 138

opt-out marketing programs, 138

Oxycodone, 82

P

Page, Larry, 223, 240

Pandora, 36-37

Pareto principle, 109

Pariser, Eli, 244

PathTracker, 70

Paul, Senator Rand, 227

paying, for consumer data, 203-205

pay-to-play model, 203-205

PCAST (President's Council of Advisors on Science and Technology), 184

People v. Weaver, 226

PeopleBrowsr, 43, 224

Pepsi, 13

perceptions of privacy, 123-125

personal data, 160

sharing, 173

personality surveys, 46

personalization, larger brands versus small brands, 235

personalization filters, 244

personalized pricing, 156-160

Rosetta Stone, 157

Staples, 156-157

personally identifiable information (PII), 130

pharmaceutical representatives, 112-113

PII (personally identifiable information), 130

Pillsbury, 60-62

Pinterest, 140, 238

Placemeter, 89

Podesta, John, 139, 179, 192

political advertising, 14, 57-59

political campaigns, Aristotle, 188

Pose, 140

predicting crime, 78-80

predictors, 116

preexisting conditions, 66

President's Council of Advisors on Science and Technology (PCAST), 184

Pretty Little Liars, 216

price discrimination, 160-162

price increase, Amazon Prime, 35

price points, 154

price skimming, 154

pricing

demand-based pricing, 151-156

dynamic pricing, 154-155

personalized pricing, 156-160

Prime, personalization, 237

print advertising, 241

privacy, 213

customer data, 125-127

data collection practices, 130-135

digital privacy, 249-252

Do-Not-Track Online Act of 2013, 131-133

opt-in programs, 138

opt-out marketing programs, 138

perceptions of, 123-125

profiling by association, 176-180

transparent practices, 135-139

value of, 173-176, 219

privacy externality, 179

privacy interdependence, 179

privacy policies, 18

Facebook, 135-136

Google, 136

Nest Labs, 137

privacy settings, Facebook, 207

profiles of consumers, 145

profiling by social media activity, 176-180

Progressive, 65

protecting, consumers, 196

protections for consumer, 221

public, compiling to inform the public, 88-90

public data, analyzing, 80-83

purchase patterns, 209

purchase process, 62

Q

quantified self, 17

quants, 78

R

reaching today's consumer, 71

recommendations

Amazon, 25

Netflix, 44

Reding, Viviane, 136

Reilly, Justin, 257-258

regulation

data economy, 220-224

educating consumers on what is being collected, 224-226

innovation and, 224

Reign, 216

relations based on customer data, 30

Republicans, advertisements, 57-58

restaurants, compiling data to inform the public, 90

return path data, 56

revealing ourselves online, 45-46

revenue streams, consumer data as financial assets, 218-220

RFID technology, 51

Rich, Jessica, 227

right to be forgotten, 222-223

right to digital privacy, 249-252

Rockefeller, Senator, 131, 180

role of government in data collection, 19

Rosetta Stone, personalized pricing, 157

S

Samsung, Simband, 241

Schmidt, Eric, 145

schools that offer M.S. degrees in analytics, 4

search terms, Google, 12

SEC (Securities and Exchange Commission), 222

second screen, 59

SecondSync, 59

Security Concerns, survey of consumers' preferences, 165

Security-Focused, survey of consumers' preferences, 164

segmentation practices, 192-196
civil rights, 191

set-top box, 56

sharing

data. *See* data sharing

personal data, 173

value of data, 140

Shazam, 60

Simband, Samsung, 241

SmartBand, Sony, 241

smarter practices, 241-243

smartphones, 107

location-aware advertising, 70

Snowden, Edward, 17, 123

social connections, 177

social influence, 43

social media, television advertising, 60

social media posts, 144

social networks

creditworthiness, 41-42

Facebook, 39-40

Google, 40

marketing, 38-44

Twitter, 40

social ratings, television, 216
 socialTV, 59
 Sony, 185
 SmartBand, 241
 Sotomayor, Justice Sonia, 226
 Staples, personalized pricing, 156-157
 Starbucks, My Starbucks Idea, 87
 State Farm, 65
 Stiegman, Patrick, 99
 STP (segmentation, targeting, positioning), 193
 strategies of Big Data, 9-10
 streaming media providers, 237
 Street Bump, 92
 subscribers, telecommunication providers, 217
 Surowiecki, James, 83
 survey of consumers' preferences, 163-169
 surveys, 204
 Google Consumer Surveys, 204
 Swisher, Kara, 52

T

Tableau, 80
 Target, 68-69, 124
 data breaches, 126, 185
 Target Red Card, 182-183
 targeted advertising
 Facebook, 45
 political advertising, 57-59
 T-commerce movement, 217
 TD Bank USA N.A., 183
 telecommunication providers, 217-218
 television, 216
 television advertising, 241
 accountability, 62
 apps, 60
 brand-building strategies, 62
 social media, 60

television industry, advertisements, 52-57
 terms of service, data ownership, 141-142
 terrorism, 190
 text analysis, 190
 TiVo, 218
 T.J. Maxx, 185
 tracking money, 102-105
 transparency, 193, 212
 Facebook, 208, 246
 transparent data collection policies, privacy, 135-139
 travelers, 15
 Tsarnaev, Dzhokhar, 85
 Tsarnaev, Tamerlan, 86
 Tumblr, 238
 TV of Tomorrow, 64
 tweets, social influence, 43
 Twitter, 16, 238
 data ownership, 141-142
 Firehose, 43
 marketing, 40
 moods, 99
 Nielsen ratings, 59
 television ratings, 216

U

Uber, 154
 United Kingdom, Financial Services Authority, 220
 United States, regulatory agencies, 222
 United States Supreme Court, right to digital privacy, 249-252
 University of Pennsylvania, 162
 University of Wisconsin-Madison, 161
 Upworthy, 245
 utility providers, 243

V

- value homophily, 177
- value of
 - data, 9, 12-16, 169-170
 - political advertising*, 14
 - privacy, 173-176, 219
- Varian, Hal, 2, 77
- velocity, 7
- Venn diagram, 3
- Verizon, 70, 217
 - consumer data, 247
- video games, demand, 152
- viewer diaries, 55
- Visible World, 64

W-X

- Walt Disney World, Disney
 - Operational Command Center, 7-8
- Washington Post, The*, 52
- Waze, 87, 91
- Weather Underground, 91
- website customization, 158
- Western Union, 239
- Wharton Customer Analytics
 - Initiative, 4
- Wheeler, Tom, 247
- Who Wants to Be a Millionaire?*, 83
- Wu, Tim, 239

[x+1], 159

Y

- Yelp, 90

Z

- Zuckerberg, Mark, 135