

too BIG to IGNORE

THE BUSINESS CASE FOR BIG DATA



PHIL SIMON

Award-Winning Author of
THE AGE OF THE PLATFORM

WILEY

Contents

[Cover](#)

[Praise](#)

[Wiley & SAS Business Series](#)

[Title Page](#)

[Copyright](#)

[Other Books by Phil Simon](#)

[Epigraph](#)

[List of Tables and Figures](#)

[Acknowledgments](#)

[Preface](#)

[NOTES](#)

[Introduction: This Ain't Your Father's Data](#)

[BETTER CAR INSURANCE THROUGH DATA](#)

[POTHOLE AND GENERAL ROAD HAZARDS](#)

[RECRUITING AND RETENTION](#)

[HOW BIG IS BIG? THE SIZE OF BIG DATA](#)

[Why Now? Explaining the Big Data Revolution](#)

[CENTRAL THESIS OF BOOK](#)

[PLAN OF ATTACK](#)

[WHO SHOULD READ THIS BOOK?](#)

[SUMMARY](#)

Chapter 1: Data 101 and the Data Deluge

THE BEGINNINGS: STRUCTURED DATA

STRUCTURE THIS! WEB 2.0 AND THE ARRIVAL OF BIG DATA

THE COMPOSITION OF DATA: THEN AND NOW

THE CURRENT STATE OF THE DATA UNION

THE ENTERPRISE AND THE BRAVE NEW BIG DATA WORLD

SUMMARY

NOTES

Chapter 2: Demystifying Big Data

CHARACTERISTICS OF BIG DATA

THE ANTI-DEFINITION: WHAT BIG DATA IS NOT

SUMMARY

NOTES

Chapter 3: The Elements of Persuasion: Big Data Techniques

THE BIG OVERVIEW

STATISTICAL TECHNIQUES AND METHODS

DATA VISUALIZATION

AUTOMATION

SEMANTICS

BIG DATA AND THE GANG OF FOUR

PREDICTIVE ANALYTICS

LIMITATIONS OF BIG DATA

SUMMARY

NOTES

Chapter 4: Big Data Solutions

PROJECTS, APPLICATIONS, AND PLATFORMS

OTHER DATA STORAGE SOLUTIONS

WEBSITES, START-UPS, AND WEB SERVICES

HARDWARE CONSIDERATIONS

THE ART AND SCIENCE OF PREDICTIVE ANALYTICS

SUMMARY

NOTES

Chapter 5: Case Studies: The Big Rewards of Big Data

QUANTCAST: A SMALL BIG DATA COMPANY

EXPLORYS: THE HUMAN CASE FOR BIG DATA

NASA: HOW CONTESTS, GAMIFICATION, AND OPEN

INNOVATION ENABLE BIG DATA

SUMMARY

NOTES

Chapter 6: Taking the Big Plunge

BEFORE STARTING

STARTING THE JOURNEY

AVOIDING THE BIG PITFALLS

SUMMARY

NOTES

Chapter 7: Big Data: Big Issues and Big Problems

PRIVACY: BIG DATA = BIG BROTHER?

BIG SECURITY CONCERNS

BIG, PRAGMATIC ISSUES

SUMMARY

NOTES

Chapter 8: Looking Forward: The Future of Big Data

PREDICTING PREGNANCY

BIG DATA IS HERE TO STAY

[BIG DATA WILL EVOLVE](#)

[PROJECTS AND MOVEMENTS](#)

[BIG DATA WILL ONLY GET BIGGER...AND SMARTER](#)

[THE INTERNET OF THINGS: THE MOVE FROM ACTIVE TO
PASSIVE DATA GENERATION](#)

[BIG DATA: NO LONGER A BIG LUXURY](#)

[STASIS IS NOT AN OPTION](#)

[SUMMARY](#)

[NOTES](#)

[Final Thoughts](#)

[SPREADING THE BIG DATA GOSPEL](#)

[NOTES](#)

[Selected Bibliography](#)

[About the Author](#)

[Index](#)

“Today Big data affects everybody and will continue to do so for the foreseeable future. In *Too Big to Ignore*, Phil Simon makes the topic accessible and relatable. This important book shows people how to put Big Data to work for their organizations.”

—William McKnight, President, McKnight Consulting Group

“Simon has an uncanny ability to connect business cases with complex technical principles, and most importantly, clearly explain how everything comes together. In this book, Simon demystifies Big Data. Simon’s vision helps the rest of us understand how this evolving and pervasive subject affects businesses today.”

—Dalton Cervo, co-author of *Master Data Management in Practice—Achieving True Customer MD* and president of Data Gap Consulting

“From Twitter feeds to photo streams to RFID pings, the Big Data universe is rapidly expanding, providing unprecedented opportunities to understand the present and peer into the future. Tapping its potential while avoiding its pitfalls doesn’t take magic; it takes a map. In *Too Big to Ignore*, Phil Simon offers businesses a comprehensive, clear-eyed, and enjoyable guide to the next data frontier.”

—Chris Berdik, author of *Mind over Mind: The Surprising Power of Expectations*

“Business leaders are drowning in data, and the deluge has only just begun. In *Too Big to Ignore*, Simon delves into the world of Big Data, and makes the business case for capturing, structuring, analyzing, and visualizing the immense amount of information accessible to businesses. This book gives your organization the edge it needs to turn data into intelligence, and intelligence into action.”

—Paul Roetzer, Founder & CEO, PR 20/20; author of *The Marketing Agency Blueprint*

“Phil Simon’s *Too Big to Ignore* clearly demonstrates the increasing role and value of Big Data. His illustrative case studies and engaging style will dispel any doubts executives may have about how Big Data is driving success in today’s economy.”

—Adrian C. Ott, award-winning author of *The 24-Hour Customer*

Wiley & SAS Business Series

The Wiley & SAS Business Series presents books that help senior-level managers with their critical management decisions.

Titles in the Wiley and SAS Business Series include:

Activity-Based Management for Financial Institutions: Driving Bottom-Line Results by Bren Bahnum

Big Data Analytics: Turning Big Data into Big Money by Frank Ohlhorst

Branded! How Retailers Engage Consumers with Social Media and Mobility by Bernie Brenner and Lori Schafer

Business Analytics for Customer Intelligence by Gert Laursen

Business Analytics for Managers: Taking Business Intelligence Beyond Reporting by Gert Laursen and Jesper Thorlund

The Business Forecasting Deal: Exposing Bad Practices and Providing Practical Solutions by Michael Gilliland

Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy by Olivia Parr Rud

CIO Best Practices: Enabling Strategic Value with Information Technology, Second Edition by Joe Stenzel

Connecting Organizational Silos: Taking Knowledge Flow Management to the Next Level with Social Media by Frank Leistner

Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors by Clark R. Abrahams and Mingyuan Zhang

Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring by Naem Siddiqi

The Data Asset: How Smart Companies Govern Their Data for Business Success by Tony Fisher

Demand-Driven Forecasting: A Structured Approach to Forecasting by Charles Chase

The Executive's Guide to Enterprise Social Media Strategy: How Social Networks Are Radically Transforming Your Business by David Thomas and Mike Barlow

Executive's Guide to Solvency II by David Buckham, Jason Wahl, and Stuart Rose

Fair Lending Compliance: Intelligence and Implications for Credit Risk Management by Clark R. Abrahams and Mingyuan Zhang

Foreign Currency Financial Reporting from Euros to Yen to Yuan: A Guide to Fundamental Concepts and Practical Applications by Robert Rowan

Human Capital Analytics: How to Harness the Potential of Your Organization's Greatest Asset by Gene Pease, Boyce Byerly, and Jac Fitz-enz

Information Revolution: Using the Information Evolution Model to Grow Your Business by Jim Davis, Gloria J. Miller, and Allan Russell

Manufacturing Best Practices: Optimizing Productivity and Product Quality by Bobby Hull

Marketing Automation: Practical Steps to More Effective Direct Marketing by Jeff LeSueur

Mastering Organizational Knowledge Flow: How to Make Knowledge Sharing Work by Frank Leistner

The New Know: Innovation Powered by Analytics by Thornton May

Performance Management: Integrating Strategy Execution, Methodologies, Risk, and Analytics by Gary Cokins

Retail Analytics: The Secret Weapon by Emmett Cox

Social Network Analysis in Telecommunications by Carlos Andre Reis Pinheiro

Statistical Thinking: Improving Business Performance, Second Edition by Roger W. Hoerl and Ronald D. Snee

Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics by Bill Franks

The Value of Business Analytics: Identifying the Path to Profitability by Evan Stubbs

Visual Six Sigma: Making Data Analysis Lean by Ian Cox, Marie A. Gaudard, Philip J. Ramsey, Mia L. Stephens, and Leo Wright

Win with Advanced Business Analytics: Creating Business Value from Your Data by Jean Paul Isson and Jesse Harriott

For more information on any of the above titles, please visit www.wiley.com.

Too Big to Ignore

The Business Case for Big Data

Phil Simon

WILEY

Cover image: © Baris Simsek/iStockphoto

Cover design: John Wiley & Sons, Inc.

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the Web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best effort in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

ISBN 9781118638170 (Hardcover)

ISBN 9781118642108 (ebk)

ISBN 9781118641682 (ebk)

ISBN 9781118641866 (ebk)

Other Books by Phil Simon

Why New Systems Fail: An Insider's Guide to Successful IT Projects

The Next Wave of Technologies: Opportunities in Chaos

The New Small: How a New Breed of Small Businesses Is Harnessing the Power of Emerging Technologies

The Age of the Platform: How Amazon, Apple, Facebook, and Google Have Redefined Business

101 Lightbulb Moments in Data Management: Tales from the Data Roundtable (Editor)

The fact that we can now begin to actually look at the dynamics of social interactions and how they play out, and are not just limited to reasoning about averages like market indices is for me simply astonishing. To be able to see the details of variations in the market and the beginnings of political revolutions, to predict them, and even control them, is definitely a case of Promethean fire. Big Data can be used for good or bad, but either way it brings us to interesting times. We're going to reinvent what it means to have a human society.

—Sandy Pentland, Professor, MIT

Knowledge is good.

—Motto of fictitious Faber College, *Animal House*

List of Tables and Figures

[Figure P.1 Michael Lewis and Billy Beane with Katty Kay at IBM Information on Demand 2011](#)

[Table I.1 Big Data Improves Recruiting and Retention](#)

[Figure I.1 The Internet in One Minute](#)

[Figure I.2 The Drop in Data Storage Costs](#)

[Figure I.3 The Technology Adoption Life Cycle \(TALC\)](#)

[Table 1.1 Simple Example of Structured Customer Master Data](#)

[Table 1.2 Simple Example of Transactional Sales Data](#)

[Figure 1.1 Entity Relationship Diagram \(ERD\)](#)

[Figure 1.2 Flickr Search Options](#)

[Figure 1.3 The Ratio of Structured to Unstructured Data](#)

[Figure 1.4 The Organizational Data Management Pyramid](#)

[Figure 2.1 Google Trends for Big Data](#)

[Figure 2.2 The Deep Web](#)

[Table 3.1 Sample Regression Analyses](#)

[Table 3.2 Simple CapitalOne A/B Test Example with Hypothetical Data](#)

[Figure 3.1 Reis's Book Cover Experiment Data](#)

[Figure 3.2 Tableau Interactive Data Visualization on How We Eat](#)

[Figure 3.3 RFID Tag](#)

[Figure 3.4 Google Autocomplete](#)

[Table 4.1 The Four General Types of NoSQL Databases](#)

[Table 4.2 Google Big Data Tools](#)

[Table 4.3 Is Big Data Worth It? Hardware Considerations](#)

[Figure 5.1 Quantcast Quantified Dashboard](#)

[Table 6.1 Big Data Short- and Long-Term Goals](#)

[Figure 8.1 Retail Awareness of Big Data](#)

Preface

Errors using inadequate data are much less than those using no data at all.

—Charles Babbage

It's about 7:30 a.m. on October 26, 2011, and I'm driving on The Strip in Las Vegas, Nevada. No, I'm not about to play craps or see Celine Dion. (While very talented, she's just not my particular brand of vodka.) I'm going for a more professional reason. Starting sometime in mid-2011, I started hearing more and more about something called *Big Data*. On that October morning, I was invited to IBM Information on Demand (IOD) conference. It was high time that I learned more about this new phenomenon, and there's only so much you can do in front of a computer.

Beyond my insatiable quest for knowledge on all matters technology, truth be told, I went to IOD for a bunch of other reasons. First, it was convenient: The Strip is a mere fifteen minutes from my home. Second, the price was right: I was able to snake my way in for free. It turns out that, since I write for a few high-profile sites, some people think of me as a member of the media. (Funny how I never would have expected that ten years ago, but far be it from me to look a gift horse in the mouth.) Third, it was a good networking opportunity and my fourth book, *The Age of the Platform*, had just been published. I am familiar enough with the book business to know that authors have to get out there if they want to generate a buzz and move copies. These were all valid reasons to hop in my car, but for me there was an extra treat. I had the opportunity to meet and listen firsthand to the conference's two keynote speakers: Michael Lewis (one of my favorite writers) and a man by the name of Billy Beane.

For his part, Lewis wasn't at IOD to promote his latest opus like I was. On the contrary, he was there to speak about his 2003 book *Moneyball: The Art of Winning an Unfair Game*. The book had been enjoying a huge commercial resurgence as of late, thanks in no small part to the recent film of the same name starring some guy named Brad Pitt. I hadn't read *Moneyball* in some years, but I remember breezing through it. Lewis's writing style is nothing if not engaging. (He even made subprime mortgages and synthetic collateralized debt obligations [CDOs] interesting in *The Big Short*.)

I've always been a bit of a stats geek, and *Moneyball* instantly hit a nerve with me. It told the story of Beane, the general manager (GM) of the budget-challenged Oakland A's. Despite his team's financial limitations, he consistently won more games than most other mid-market teams—and even franchises like the New York Yankees that effectively printed their own money. The obvious question was how? Beane bucked convention and routinely ignored the advice of long-time baseball scouts, often earning their derision in the process. Instead, Beane predicated his management style on a rather obscure, statistics-laden field called *sabermetrics*. He signed free agents who he believed were undervalued by other teams. That is, he sought to exploit market inefficiencies.

One of Beane's favorite bargains: a relatively cheap player with a high on-base percentage (OBP). In a nutshell, Beane's simple and irrefutable logic could be summarized as follows: players more likely to get on base are more likely to score runs. By extension, higher-scoring teams tend to win more games than their lower-scoring counterparts. But Beane didn't stop there. He was also partial to players (again, only at the right price) who didn't swing at the first pitch. Beane liked hitters who consistently made opposing pitchers work deep into the count. These patient batters were more likely to make opposing pitches tired—and then give *everyone* on the A's better pitches to hit. (Again, more

runs would result, as would more wins.)

Figure P.1 Michael Lewis and Billy Beane with Katty Kay at IBM Information on Demand 2011¹

Source: Todd Watson



Back then, evaluating players based on unorthodox stats like these was considered heresy in traditional baseball circles. And that resistance was not just among baseball outsiders. In the late 1990s and early 2000s, a conflict within the A's organization was growing between Beane and his most visible employee: manager Art Howe. A former infielder with three teams over twelve years, Howe for one wasn't on board with Beane's unconventional program, to put it mildly. As Lewis tells us in *Moneyball*, Howe was nothing if not old school. He certainly didn't need some newfangled, stats-obsessed GM telling him the X's and O's of baseball.

Oakland's internal conflict couldn't persist; a GM and manager have to be on the same page in any sport, and baseball is no exception. Rather than fire Howe outright (with the A's eating his \$1.5 million salary), Beane got creative, as he is wont to do. He cajoled the New York Mets into taking him off their hands, not that the Mets needed much convincing. The team soon signed its new leader to a then-bawdy four-year, \$9.4 million contract. After all, Howe had won a more-than-respectable 50 percent of his games with the small-market A's and *he just looks managerial*. The man has a great jaw. Imagine what Howe could do for a team with a big bankroll like the Mets?

Howe's tenure with the Mets was ignominious. The team won only 42 percent of its games over Howe's watch. After two seasons, the Mets realized what Beane knew long ago: Howe and his managerial jaw were much better in theory than in practice. In September 2004, the Mets parted ways with their manager.

While Beane may have been the first GM to embrace sabermetrics, he soon had company. His success bred many disciples in the baseball world and beyond. Count among them Theo Epstein, currently the President of Baseball Operations for the Chicago Cubs. In his previous role as GM of the Boston Red Sox, Epstein even hired Bill James, the godfather of sabermetrics. And it worked. Epstein won two World Series for the Sox, breaking the franchise's 86-year drought. Houston Rockets's GM Daryl Morey is bringing *Moneyball* concepts to the NBA. As a November 2012 *Sports Illustrated*

article points out, the MBA grad takes a radically different approach to player acquisition and development compared to his peers.²

And then there's the curious case of Kevin Kelley, the head football coach at the Pulaski Academy, a high school in Little Rock, Arkansas. Kelley isn't your average coach. The man "stopped punting in 2005 after reading an academic study on the statistical consequences of going for the first down versus handing possession to the other team."³ Coach Kelley simply refuses to punt. Ever. Even if it's four and 20 from his own ten-yard line. But it gets even better. Ever the contrarian, after Pulaski scores, Kelley has his kicker routinely try on-side kicks to try to get the ball right back. In one game, Kelley's team scored twenty-nine points before the opponent even touched the football!⁴ The results? The Bruins have won multiple state championships using their coach's unconventional style.

So why were Lewis and Beane the keynote speakers at IOD, a corporate information technology (IT) conference? Because, as *Moneyball* demonstrates so compellingly, today new sources of data are being used across many different fields in very unconventional and innovative ways to produce astounding results—and a swath of people, industries, and established organizations are finally starting to realize it.

This book explains why Big Data is a big deal. For example, residents in Boston, Massachusetts, are automatically reporting potholes and road hazards via their smartphones. Progressive Insurance tracks real-time customer driving patterns and uses that information to offer rates truly commensurate with individual safety. HR departments are using new sources of information to make better hiring decisions. Google accurately predicts local flu outbreaks based on thousands of user search queries. Amazon provides remarkably insightful, relevant, and timely product recommendations to its hundreds of millions of customers. Quantcast lets companies target precise audiences and key demographics throughout the Web. NASA runs contests via gamification site TopCoder, awarding prizes to those with the most innovative and cost-effective solutions to its problems. Explorys offers penetrating and previously unknown insights into health care behavior.

How do these organizations and municipalities do it? Technology is certainly a big part, but in each case the answer lies deeper than that. Individuals at these organizations have realized that they don't have to be statistician Nate Silver to reap massive benefits from today's new and emerging types of data. And each of these organizations has embraced Big Data, allowing them to make astute and otherwise impossible observations, actions, and predictions.

It's time to start thinking big.

This book is about an unassailably important trend: Big Data, the massive amounts, new types, and multifaceted sources of information streaming at us faster than ever. Never before have we seen data with the volume, velocity, and variety of today. Big Data is no temporary blip of a fad. In fact, it's only going to intensify in the coming years, and its ramifications for the future of business are impossible to overstate.

Put differently, Big Data is becoming too big to ignore. And that sentence, in a nutshell, summarizes this book.

Phil Simons
Henderson, NV
March 2011

NOTES

1. Watson, Todd, "Information on Demand 2011: A Data-Driven Conversation with Michael Lewis & Billy Beane," October 26, 2011, <http://turbotodd.wordpress.com/2011/10/26/information-on-demand-2011-a-data-driven-conversation-with-michael-lewis-billy-beane/>, retrieved December 11, 2012.

2. Ballard, Chris, "Lin's Jumper, GM Morey's Hidden Talents, More Notes from Houston," November 30, 2012, http://sportsillustrated.cnn.com/2012/writers/chris_ballard/11/30/houston-rockets-jeremy-lin-james-harden-daryl-morey/index.html, retrieved December 11, 2012.

3. Easterbrook, Gregg, "New Annual Feature! State of High School Nation," November 15, 2007, <http://sports.espn.go.com/espn/page2/story?page=easterbrook/071113>, retrieved December 11, 2012.

4. Wertheim, Jon, "Down 29-0 Before Touching the Ball," September 15, 2012, <http://sportsillustrated.cnn.com/2011/writers/scorecasting/09/15/kelley.pulaski/index.html>, retrieved December 11, 2012.

5 For those of you not familiar with the term, *OBP* represents the true measure of how often a batter reaches base. It includes hits, walks, and times hit by a pitch. Beane also sought out those with high on-base plus slugging percentages. OPS equals the sum of a player's OBP and slugging percentage (total bases divided by at bats).

Acknowledgments

Kudos to the Wiley team of Tim Burgard, Shelly Sessoms, Karen Gill, Johnna VanHoose Dinse, Chris Gage, and Stacey Rivera for making this book possible so quickly. You all were a “big” help.

I am grateful to smart cookies Charlie Lougheed, Jim McKeown, Jason Crusan, Jag Duggal, Jim Kelly, Clinton Bonner, William McKnight, Scott Kahler, and Seth Grimes for their time and expertise. Talking to these folks made research fun. A tip of the hat to Hope Nicora, Andy Havens, Adrian Ott, Brad Feld, Chris Berdik, Terri Griffith, Jim Harris, Dalton Cervo, Jill Dyché, Todd Hamilton, Tom Fisher, Ellen French, Dick and Bonnie Denby, Kristen Eckstein, Bob Charette, Andrew Botwin, Thomas and Keri Sandell, Clair Byrd, Jay and Heather Etchings, Karlena Kuder, Luke “Heisenberg” Fletcher, Michael, Penelope, and Chloe DeAngelo, Shawn Graham, Chad Roberts, Sarah Terry, Jeff Lee, Marc Cenicola, Brenda Blakely, Colin Hickey, Bruce Webster, Alan Berkson, Michael West, John Spatola, Marc Paoella, Angela Bowman, and Brian and Heather Morgan and their three adorable kids.

Next up are the usual suspects: my longtime Carnegie Mellon friends Scott Berkun, David Sandberg, Michael Viola, Joe Mirza, and Chris McGee.

My heroes from Rush (Geddy, Alex, and Neil), Dream Theater (Jordan, John, John, Mike, and James), Marillion (h, Steve, Ian, Mark, and Pete), and Porcupine Tree (Steven, Colin, Gavin, John, and Richard) have given me many years of creative inspiration through their music. Keep on keepin’ on!

Vince Gillian, Aaron Paul, Bryan Cranston, Dean Norris, Anna Gunn, Betsy Brandy, RJ Mitte, and the rest of the cast and team of *Breaking Bad* make me want to do great work.

Next up: my parents. I’m not here without you.

Introduction: This Ain't Your Father's Data

Throughout history, in one field after another, science has made huge progress in precisely the areas where we can measure things—and lagged where we can't.

—Samuel Arbesman

Car insurance isn't a terribly sexy or dynamic business. For decades, it has essentially remained unchanged. Nor is it an egalitarian enterprise: while a pauper and a millionaire pay the same price for a stamp (\$0.45 in the United States as of this writing), the car insurance world works differently. Some people just pay higher rates than others, and those rates have at least initially very little to do with whether one is a "safe" driver, whatever that means. Historically, many if not most car insurance policies were written based on very few independent variables: age, gender, zip code, previous speeding tickets and traffic violations, documented accidents, and type of car. As I found out more than twenty years ago, a newly licensed, seventeen-year-old guy in New Jersey who drives a sports car has to pay a boatload in car insurance for the privilege—even if he rarely drives above the speed limit, always obeys traffic signals, and has nary an accident on his record. Like just about every kid my age, I wasn't happy about my rates. After all, I was an "above average" driver, or at least I liked to think so. Why should I have to pay such exorbitant fees?

Of course, we all can't be above average; it's statically impossible. Truth be told, I'm sure that back then I occasionally didn't come to a complete stop at every red light. While I've never been arrested for DUI, to this day I don't always obey the speed limit. (Shhh . . . don't tell anyone.) When I'm driving faster than the law says I should, I sometimes think of the famous George Stigler picture of Milton Friedman taken in the mid-twentieth-century. Friedman was paying a speeding ticket with a paradoxically, a big smile on this face. Why such joy? Because Friedman was an economist and, like such, he was rational to a fault. In his view of the world, the time that he regularly saved by exceeding the speed limit was worth more to him than the risk and fine of getting caught. To people like Friedman and me, speeding is only a simple expected value calculation: Friedman sped because the rewards outweighed the risks. When a cop pulled him over, he was glad to pay the fine. But I digress.

So why do most car insurance companies base their quotes and rates on relatively simple variables? The answer isn't complicated, especially when you consider the age of these companies. Allstate opened its doors in 1931. GEICO was founded in 1936, and the Progressive Casualty Insurance Company set up shop only one year later. Think about it: seventy-five years ago, those primitive models represented the best that car insurance companies could do. While each has no doubt tweaked its models since then, old habits die hard, as we saw with Art Howe and Billy Beane in the Preface. For real change to happen, somebody needs to upset the applecart. In this way, car insurance is like baseball.

BETTER CAR INSURANCE THROUGH DATA

The similarities between the ostensibly unrelated fields of baseball and car insurance don't end there. Much like the baseball revolution pioneered by Billy Beane, car insurance today is undergoing a fundamental transformation. Just ask Joseph Tucci. As the CEO at data storage behemoth EMC Corporation, he knows a thing or thirty about data. On October 3, 2012, Tucci spoke with Cor

Johnson of Bloomberg Television at an Intel Capital event in Huntington Beach, California. Tucci talked about the state of technology, specifically the impact of Big Data and cloud computing on his company—and others.¹ At one point during the interview, Tucci talked about advances in GPS mapping, mobile technologies, and telemetry, the net result of which is revolutionizing many businesses, including car insurance. No longer are rates based upon a small, primitive set of independent variables. Car insurance companies can now get much more granular in their pricing. Advances in technology are letting them answer previously unknown questions like these:

- Which drivers routinely exceed the speed limit and run red lights?
- Which drivers routinely drive dangerously slow?
- Which drivers are becoming less safe—even if they have received no tickets or citations? That is, who used to generally obey traffic signals but don't anymore?
- Which drivers send text messages while driving? (This is a big no-no. In fact, texting while driving [TWD] is actually considerably more dangerous than DUI.² As of this writing, fourteen states have banned it.)
- Who's driving in a safer manner than six months ago?
- Does a man with two cars (a sports car and a station wagon) drive each differently?
- Which drivers and cars swerve at night? (This could be a manifestation of drunk driving.)
- Which drivers checked into a bar using FourSquare or Facebook and drove their own cars home (as opposed to taking a cab or riding with a designated driver)?

Thanks to these new and improved technologies and the data they generate, insurers are effectively retiring their decades-old, five-variable underwriting models. In their place, they are implementing more contemporary, accurate, dynamic, and data-driven pricing models. For instance, in 2011 Progressive rolled out Snapshot, its Pay As You Drive (PAYD) program.³ PAYD allows customers to voluntarily install a tracking device in their cars that transmits data to Progressive and possibly qualifies them for rate discounts. From the company's site:

How often you make hard brakes, how many miles you drive each day, and how often you drive between midnight and 4 a.m. can all impact your potential savings. You'll get a Snapshot device in the mail. Just plug it into your car and drive like you normally do. You can go online to see your latest driving details and projected discount.

Is Progressive the only, well, progressive insurance company? Not at all. Others are recognizing the power of new technologies and Big Data. As Liane Yvkoff writes on CNET, "State Farm subscribers self-report mileage and GMAC uses OnStar vehicle diagnostics reports. Allstate's Drive Wise goes one step further and uses a similar device to track mileage, braking, and speeds over 80 mph, but only in Illinois."⁴

So what does this mean to the average driver? Consider two fictional people, both of whom hold car insurance policies with Progressive and opt in to PAYD:

- Steve, a twenty-one-year-old New Jersey resident who drives a 2012, tricked-out, cherry red Corvette
- Betty, a forty-nine-year-old grandmother in Lincoln, Nebraska, who drives a used Volvo station wagon

All else being equal, which driver pays the higher car insurance premium? In 1994, the answer was obvious: Steve. In the near future, however, the answer will be much less certain: *it will depend on the data*. That is, vastly different driver profiles and demographic information will mean less and less to car insurance companies. Traditional levers like those will be increasingly supplemented with data c

drivers' individual patterns. What if Steve's flashy Corvette belies the fact that he always obeys traffic signals, yields to pedestrians, and never speeds? He is the embodiment of safety. Conversely, despite her stereotypical profile, Betty drives like a maniac while texting like a teenager.

In this new world, what happens at rate renewal time for each driver? Based upon the preceding information, Progressive happily discounts Steve's previous insurance by 60 percent but triples Betty's renewal rate. In each case, the new rate reflects new—and far superior—data that Progressive has collected on each driver.

Surprised by his good fortune, Steve happily renews with Progressive, but Betty is irate. She calls the company's 1-800 number and lets loose. When the Progressive rep stands her ground, Betty decides to take her business elsewhere. Unfortunately for Betty, she is in for a rude awakening. Allstate, GEICO, and other insurance companies have access to the same information as Progressive. All companies strongly suspect that Betty is actually a high-risk driver; her age and Volvo only tell part of her story—and not the most relevant part. As such, Allstate and GEICO quote her a policy similar to Progressive's.

Now, Betty isn't happy about having to pay more for her car insurance. However, Betty *should* pay more than safer drivers like Steve. In other words, simple, five-variable pricing models no longer represent the best that car insurance companies can do. They now possess the data to make better business decisions.

Big Data is changing car insurance and, as we'll see throughout this book, other industries as well. The revolution is just getting started.

POTHOLES AND GENERAL ROAD HAZARDS

Let's stay on the road for a minute and discuss the fascinating world of potholes. Yes, potholes. Historically, state and municipal governments have had a pretty tough time identifying these pesky devils. Responsible agencies and departments would often scour the roads in search of potholes and general road hazards, a truly reactive practice. Alternatively, they would rely upon annoyed citizens to call them in, typically offering fairly generic locations like "on Main Street, not too far from the corner of Eleven." In other words, there was no good automatic way to report potholes to the proper authorities. As a result, many hazards remained unreported for significant periods of time, no doubt causing considerable damage and earning the ire of many a taxpayer. Many people agree with the quote from acerbic comedian Dennis Miller, "The states can't pave [expletive deleted] roads."

Why has the public sector handled potholes and road hazards this way? For the same reason that car insurance companies relied upon very few basic variables when quoting insurance rates to their customers: in each case, it was the best that they could do at the time.

At some point in the past few years, Thomas M. Menino (Boston's longest-serving mayor) realized that it was no longer 1950. Perhaps he was hobnobbing with some techies from MIT at dinner one night. Whatever his motivation, he decided that there just had to be a better, more cost-effective way to maintain and fix the city's roads. Maybe smartphones could help the city take a more proactive approach to road maintenance. To that end, in July 2012, the Mayor's Office of New Urban Mechanics launched a new project called Street Bump, an app that

allows drivers to automatically report the road hazards to the city as soon as they hear the unfortunate "thud," with their smartphones doing all the work.

The app's developers say their work has already sparked interest from other cities in the U.S.

Europe, Africa and elsewhere that are imagining other ways to harness the technology.

Before they even start their trip, drivers using Street Bump fire up the app, then set their smartphones either on the dashboard or in a cup holder. The app takes care of the rest, using the phone's accelerometer—a motion-detector—to sense when a bump is hit. GPS records the location and the phone transmits it to a remote server hosted by Amazon Inc.'s Web services division.⁵

But that's not the end of the story. It turned out that the first version of the app reported far too many false positives (i.e., phantom potholes). This finding no doubt gave ammunition to the many naysayers who believe that technology will never be able to do what people can and that things are just as fine as they are, thank you. Street Bump 1.0 "collected lots of data but couldn't differentiate between potholes and other bumps."⁶ After all, your smartphone or cell phone isn't inert; it moves in the car naturally because the car is moving. And what about the scores of people whose phones "moved" because they check their messages at a stoplight?

To their credit, Menino and his motley crew weren't entirely discouraged by this initial setback. In their gut, they knew that they were on to something. The idea and potential of the Street Bump app were worth pursuing and refining, even if the first version was a bit lacking. Plus, they have plenty of examples from which to learn. It's not like the iPad, iPod, and iPhone haven't evolved over time.

Enter InnoCentive Inc., a Massachusetts-based firm that specializes in open innovation and crowdsourcing. (We'll return to these concepts in Chapters 4 and 5.) The City of Boston contracted InnoCentive to improve Street Bump and reduce the number of false positives. The company accepted the challenge and essentially turned it into a contest, a process sometimes called *gamification*. InnoCentive offered a network of 400,000 experts a share of \$25,000 in prize money donated by Liberty Mutual.

Almost immediately, the ideas to improve Street Bump poured in from unexpected places. Ultimately, the best suggestions came from

- A group of hackers in Somerville, Massachusetts, that promotes community education and research
- The head of the mathematics department at Grand Valley State University in Allendale, Michigan
- An anonymous software engineer

The result: Street Bump 2.0 is hardly perfect, but it represents a colossal improvement over its predecessor. As of this writing, the Street Bump website reports that 115,333 bumps have been detected. What's more, it's a quantum leap over the manual, antiquated method of reporting potholes, no doubt still being used by countless public works departments throughout the country and the world. And future versions of Street Bump will only get better. Specifically, they may include earthquake detection capability and different uses for police departments.

Street Bump is not the only example of an organization embracing Big Data, new technologies, and, arguably most important, an entirely new mind-set. With the app, the City of Boston was acting less like a government agency and more like, well, a progressive business. It was downright refreshing to see.

Crowdsourcing roadside maintenance isn't just cool. Increasingly, projects like Street Bump are resulting in substantial savings. And the public sector isn't alone here. As we've already seen with examples like Major League Baseball (MLB) and car insurance, Big Data is transforming many industries and functions within organizations. Chapter 5 will provide three in-depth case studies of organizations leading the Big Data revolution.

RECRUITING AND RETENTION

In many organizations, Human Resources (HR) remains the redheaded stepchild. Typically seen as the organization's police department, HR rarely commands the internal respect that most SVPs and Chief People Officers believe it does. I've seen companies place poor performers in HR because they couldn't cut it in other departments. However, I've never seen the reverse occur (e.g., "Steve was horrible in HR, so we put him in Finance."). For all of their claims about being "strategic partners," many HR departments spend the majority of their time on administrative matters like processing new hire paperwork and open enrollment. While rarely called *Personnel* anymore (except on *Mad Men*), many HR departments are anachronistic: they operate now in much the same way as they did for decades ago.

My own theory about the current, sad state of HR is as follows: As a general rule, HR folks tend not to make decisions based upon data. In this way, HR is unique. Employees rely almost exclusively on their gut instincts and corporate policy. What if employees in other departments routinely made important decisions sans relevant information? Absent data, the folks in marketing, sales, product, and finance wouldn't command a great deal of respect either. W. Edwards Deming once said, "In God we trust, all others must bring data." Someone forgot to tell this to the folks in HR, and the entire function suffers as a result.

I wrote a book on botched IT projects and system implementations, many of which involved HR and payroll applications. Years of consulting on these types of engagements have convinced me that most employees in HR just don't think like employees in other departments. Most HR people don't seek out data in making business decisions or even use the data available to them. In fact, far too many HR folks actively try to *avoid* data at all costs. (I've seen HR directors manipulate data to justify the decision to recruit at Ivy League schools, despite the fact that trying to hire Harvard and Yale alumni didn't make the slightest bit of financial sense.) And it's this lack of data—and, in that vein, a data mind-set—that has long undermined HR as a function. As we'll see throughout this book, however, ignoring data (big or small) doesn't make it go away. Pretending that it doesn't exist doesn't make it so. In fact, Big Data can be extremely useful, even for HR.

As the *Wall Street Journal* recently reported,⁷ progressive and data-oriented HR departments are turning to Big Data to solve a long-vexing problem: how to hire better employees and retain them. It turns out that traditional personality tests, interviews, and other HR standbys aren't terribly good at predicting which employees are worth hiring—and which are not. Companies like Evolv "utiliz[e] Big Data predictive analytics and machine learning to optimize the performance of global hour workforces. The solution identifies improvement areas, then systematically implements changes to core operational business processes, driving increased employee retention, productivity, and engagement. Evolv delivers millions of dollars in operational savings on average for each client, and guarantees its impact on operating profitability."⁸

Millions in savings? Aren't these just lofty claims from a start-up eager to cash in on the Big Data buzz? Actually, no. Consider some of the specific results generated by Evolv's software, as shown in [Table I.1](#).

Table I.1 Big Data Improves Recruiting and Retention

Employee Problem	Big Data Solution
Compensation	Caesars casino found that increasing pay within certain limits had no impact on turnover.
Attrition	Xerox found that experience was overrated for call-center positions. What's more, overly inquisitive employees

tended to leave soon after receiving training.

Sick Time — Richfield Management tests applicants for opinions on drugs and alcohol. The company found that those who partake in “extracurricular” activities are more prone to get into accidents.

The lesson here is that Big Data can significantly impact each area of a business: its benefits can touch every department within an organization. Put differently, Big Data is too big to ignore.

HOW BIG IS BIG? THE SIZE OF BIG DATA

How big is the Big Data market? IT research firm Gartner believes that Big Data will create \$2 billion in worldwide spending in 2012, a number that will rise “to \$34 billion in 2013. Most of the spending will involve upgrading ‘traditional solutions’ to handle the flood of data entering organizations from a variety of sources, including clickstream traffic, social networks, sensors, and customer interactions; the firm believes that a mere \$4.3 billion in sales will come from ‘new Big Data functionality.’”⁹ For its part, consulting firm Deloitte expects massive Big Data growth, although precisely “estimating the market size is challenging.”¹⁰

High-level projections from top-tier consulting firms are all fine and dandy, but most people won't be able to get their arms around abstract numbers like these. The question remains: just how big is Big Data? You might as well ask, “How big is the Internet?”³⁰ We can't precisely answer these questions; we can only guess. What's more, Big Data got bigger in the time that it took me to write that sentence. The general answer is that Big Data is really big—and getting bigger all the time. Just look at these 2011 statistics on videos from website monitoring company Pingdom:

- **1 trillion:** The number of video playbacks on YouTube
- **140:** The number of YouTube video playbacks per person on Earth
- **48:** Hours video uploaded to YouTube every minute
- **82.5:** Percentage of the U.S. Internet audience that viewed video online
- **201.4 billion:** Number of videos viewed online per month (October 2011)
- **88.3 billion:** Videos viewed per month on Google sites, including YouTube (October 2011)
- **43:** Percentage share of all worldwide video views delivered by Google sites, including YouTube

If those numbers seem abstract, look at the infographic in [Figure I.1](#) to see what happens on the Internet every minute of every day.

[Figure I.1](#) The Internet in One Minute

Source: Image courtesy of Domo; www.domo.com

sample content of Too Big to Ignore: The Business Case for Big Data (Wiley and SAS Business Series)

- [read online Analysis I](#)
- [The Cambridge Companion to Medieval French Literature pdf, azw \(kindle\)](#)
- [The Sickness Unto Death: A Christian Psychological Exposition For Upbuilding And Awakening \(Kierkegaard's Writings, Volume 19\) pdf](#)
- [Wealth Secrets of the One Percent: A Modern Manual to Getting Marvelously, Obscenely Rich for free](#)
- [click Grant: Savior of the Union](#)

- <http://unpluggedtv.com/lib/Electrical-Engineering--Principles-and-Applications-5th---Solutions.pdf>
- <http://yachtwebsitedemo.com/books/The-Cambridge-Companion-to-Medieval-French-Literature.pdf>
- <http://thewun.org/?library/The-Beatles-Way--Fab-Wisdom-for-Everyday-Life.pdf>
- <http://metromekanik.com/ebooks/Derby-Dugan-s-Depression-Funnies--Funny-Papers-Trilogy--Book-2-.pdf>
- <http://poulterandmac.com/?books/Grant--Savior-of-the-Union.pdf>