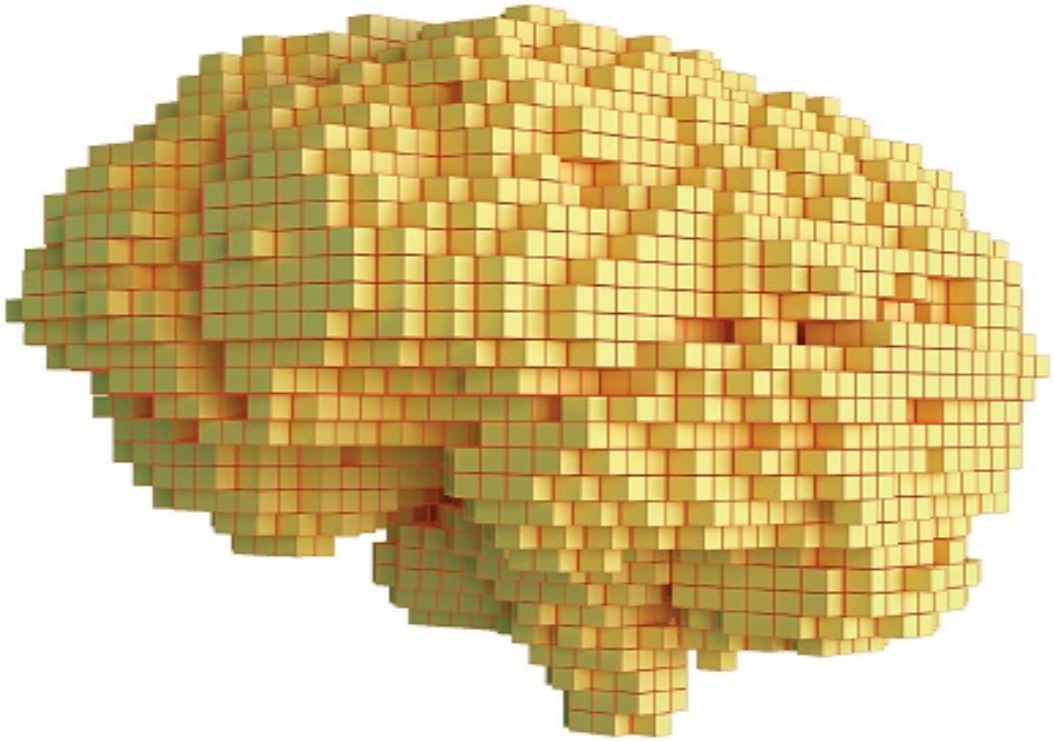
ALEX J. GUTMAN AND JORDAN GOLDMEIER



BECOMING A DATA HEAD

HOW TO THINK, SPEAK, AND UNDERSTAND DATA SCIENCE, STATISTICS, AND MACHINE LEARNING

PRAISE FOR *BECOMING A DATA HEAD*

Big Data, Data Science, Machine Learning, Artificial Intelligence, Neural Networks, Deep Learning … It can be buzzword bingo, but make no mistake, everything is becoming “datafied” and an understanding of data problems and the data science toolset is becoming a requirement for every business person. Alex and Jordan have put together a must read whether you are just starting your journey or already in the thick of it. They made this complex space simple by breaking down the “data process” into understandable patterns and using everyday examples and events over our history to make the concepts relatable.

—**Milen Mahadevan**, President of 84.51°

What I love about this book is its remarkable breadth of topics covered, while maintaining a healthy depth in the content presented for each topic. I believe in the pedagogical concept of “Talking the Walk,” which means being able to explain the hard stuff in terms that broad audiences can grasp. Too many data science books are either too specialized in taking you down the deep paths of mathematics and coding (“Walking the Walk”) or too shallow in over-hyping the content with a plethora of shallow buzzwords (“Talking the Talk”). You can take a great walk down the pathways of the data field in Alex and Jordan's without fear of falling off the path. The journey and destination are well worth the trip, and the talk.

—**Kirk Borne**, Data Scientist, Top Worldwide Influencer in Data Science

The most clear, concise, and practical characterization of working in corporate analytics that I've seen. If you want to

be a killer analyst and ask the right questions, this is for you.

—**Kristen Kehrer**, Data Moves Me, LLC, LinkedIn Top Voices in Data Science & Analytics

THE book that business and technology leaders need to read to fully understand the potential, power, AND limitations of data science.

—**Jennifer L. L. Morgan**, PhD, Analytical Chemist at Procter and Gamble

You've heard it before: “We need to be doing more machine learning. Why aren't we doing more sophisticated data science work?” Data science isn't the magic unicorn that will solve all of your company's problems. *Data Head* brings this idea to life by highlighting when data science is (and isn't) the right approach and the common pitfalls to watch out for, explaining it all in a way that a data novice can understand. This book will be my new “pocket reference” when communicating complicated concepts to non-technically trained leaders.

—**Sandy Steiger**, Director, Center for Analytics and Data Science at Miami University

Individuals and organizations want to be data driven. They say they are data driven. *Becoming a Data Head* shows them how to actually become data driven, without the assumption of a statistics or data background. This book is for anyone, or any organization, asking how to bring a data mindset to the whole company, not just those trained in the space.

—**Eric Weber**, Head of Experimentation & Metrics Research, Yelp

What is keeping data science from reaching its true potential? It is not slow algorithms, lack of data, lack of computing power, or even lack of data scientists. *Becoming a Data Head* tackles the biggest impediment to data

science success—the communication gap between the data scientist and the executive. Gutman and Goldmeier provide creative explanations of data science techniques and how they are used with clear everyday relatable examples. Managers and executives, and anyone wanting to better understand data science will learn a lot from this book. Likewise, data scientists who find it challenging to explain what they are doing will also find great value in *Becoming a Data Head*.

—**Jeffrey D. Camm**, PhD, Center for Analytics Impact, Wake Forest University

Becoming a Data Head raises the level of education and knowledge in an industry desperate for clarity in thinking. A must read for those working with and within the growing field of data science and analytics.

—**Dr. Stephen Chambal**, VP for Corporate Growth at Perduco (DoD Analytics Company)

Gutman and Goldmeier filter through much of the noise to break down complex data and statistical concepts we hear today into basic examples and analogies that stick. *Becoming a Data Head* has enabled me to translate my team's data needs into more tangible business requirements that make sense for our organization. A great read if you want to communicate your data more effectively to drive your business and data science team forward!

—**Justin Maurer**, Engineering and Data Science Manager at Google

As an aerospace engineer with nearly 15 years experience, *Becoming a Data Head* made me aware of not only what I personally want to learn about data science, but also what I need to know professionally to operate in a data-rich environment. This book further discusses how to filter through often overused terms like artificial intelligence. This is a book for every mid-level program manager

learning how to navigate the inevitable future of data science.

—**Josh Keener**, Aerospace Engineer and Program Manager

A must read for an in-depth understanding of data science for senior executives.

—**Cade Saie**, PhD, Chief Data Officer

Gutman and Goldmeier offer practical advice for asking the right questions, challenging assumptions, and avoiding common pitfalls. They strike a nice balance between thoroughly explaining concepts of data science while not getting lost in the weeds. This book is a useful addition to the toolbox of any analyst, data scientist, manager, executive, or anyone else who wants to become more comfortable with data science.

—**Jeff Bialac**, Senior Supply Chain Analyst at Kroger

Gutman and Goldmeier have written a book that is as useful for applied statisticians and data scientists as it is for business leaders and technical professionals. In demystifying these complex statistical topics, they have also created a common language that bridges the longstanding communication divide that has — until now — separated data work from business value.

—**Kathleen Maley**, Chief Analytics Officer at datazuum

Becoming a Data Head

How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning

ALEX J. GUTMAN
JORDAN GOLDMEIER

WILEY

Published simultaneously in Canada

does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that Internet websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (877) 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 booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2021934226

ISBN: 978-1-119-74174-9
ISBN: 978-1-119-74176-3 (ebk)
ISBN: 978-1-119-74171-8 (ebk)

For my children Allie, William, and Ellen.

Allie was three when she discovered dad was a “doctor.” Puzzled, she looked at me and said, “But, you don't help people… .”
In that spirit, I also dedicate this book to you, the reader.

I hope this helps you.
—Alex

For Stephen and Melissa
—Jordan

About the Authors

Alex J. Gutman is a data scientist, corporate trainer, Fulbright Specialist grant recipient, and Accredited Professional Statistician® who enjoys teaching a wide variety of data science topics to technical and non-technical audiences. He earned his Ph.D. in applied math from the Air Force Institute of Technology where he currently serves as an adjunct professor.

Jordan Goldmeier is an internationally recognized analytics professional and data visualization expert, author, and speaker. A former chief operations officer at Excel.TV, he has spent years in the data training trenches. He is the author of *Advanced Excel Essentials* and *Dashboards for Excel*. His work has been cited by and quoted in the Associated Press, Bloomberg BusinessWeek, and American Express OPEN Forum. He is currently an Excel MVP Award holder, an achievement he's held for six years, allowing him to provide feedback and direction to Microsoft product teams. He once used Excel to save the Air Force $60 million. He is also a volunteer Emergency Medical Technician.

About the Technical Editors

William A. Brenneman is a Research Fellow and the Global Statistics Discipline Leader at Procter & Gamble in the Data and Modeling Sciences Department and an Adjunct Professor of Practice at Georgia Tech in the Stewart School of Industrial and Systems Engineering. Since joining P&G, he has worked on a wide range of projects that deal with statistics applications in his areas of expertise: design and analysis of experiments, robust parameter design, reliability engineering, statistical process control, computer experiments, machine learning, and statistical thinking. He was also instrumental in the development of an in-house statistics curriculum. He received a Ph.D. in Statistics from the University of Michigan, an MS in Mathematics from the University of Iowa, and a BA in Mathematics and Secondary Education from Tabor College. William is a Fellow in both the American Statistical Association (ASA) and the American Society for Quality (ASQ). He has served as ASQ Statistics Division Chair, ASA Quality and Productivity Section Chair, and as Associate Editor for *Technometrics*. William also has seven years of experience as an educator at the high school and college level.

Jennifer Stirrup is the Founder and CEO of Data Relish, a UK-based AI and Business Intelligence leadership boutique consultancy delivering data strategy and business-focused solutions. Jen is a recognized leading authority in AI and Business Intelligence Leadership, a Fortune 100 global speaker, and has been named as one of the Top 50 Global Data Visionaries, one of the Top Data Scientists to follow on Twitter, and one of the most influential Top 50 Women in Technology worldwide.

Jen has clients in 24 countries on 5 continents, and she holds postgraduate degrees in AI and Cognitive Science. Jen has authored books on data and artificial intelligence and has been featured on CBS Interactive and the BBC as well as other well-known podcasts, such as Digital Disrupted, Run As Radio, and her own Make Your Data Work webinar series.

Jen has also given keynotes for colleges and universities, as well as donated her expertise to charities and non-profits as a Non-Executive Director. All of Jen's keynotes are based on her 20+ years of global experience, dedication, and hard work.

Acknowledgments

I've noticed a trend in acknowledgment sections—the author's spouse is often mentioned at the end. I suppose it's a saving-the-best-for-last gesture, but I promised my wife if I ever wrote a book, I'd mention her first to make it perfectly clear whose contributions mattered most to me. So, to my wife Erin, thank you for your love, encouragement, and smile. As I write this, you are taking our three young children on a bike ride, giving me time to write one final page. (I assure all readers this act is a representative sample of our lives this past year.)

I'd also like to thank my parents, Ed and Nancy, for being the best cheerleaders in whatever I do and for showing me what being a good parent looks like, and to my siblings Ryan, Ross, and Erin for their support.

This book is the culmination of many discussions with friends and colleagues, ranging from whether I should attempt to write a book about data literacy to potential topics that should appear in it. Thank you especially to Altynbek Ismailov, Andy Neumeier, Bradley Boehmke, Brandon Greenwell, Brent Russell, Cade Saie, Caleb Goodreau, Carl Parson, Daniel Uppenkamp, Douglas Clarke, Greg Anderson, Jason Freels, Joel Chaney, Joseph Keller, Justin Maurer, Nathan Swigart, Phil Hartke, Samuel Reed, Shawn Schneider, Stephen Ferro, and Zachary Allen.

I'm also indebted to the hundreds of engineers, business professionals, and data scientists I've interacted with, personally or online, who've taught me how to be a better data scientist and communicator. And to my “students” (colleagues) who have given candid feedback about the courses I've taught, I heard you and I thank you.

I'm fortunate to have many academic and professional mentors who've given me numerous opportunities to find my voice and confidence as a statistician, data scientist, and trainer. Thank you to Jeffery Weir, John Tudorovic, K. T. Arasu, Raymond Hill, Rob Baker, Scott Crawford, Stephen Chambal, Tony White, and William Brenneman (who kindly served as a technical editor on this book). It's impossible not to become wiser hanging around a group like that.

Thanks to the team at Wiley: Jim Minatel for believing in the project and giving us a chance, Pete Gaughan and John Sleeva for guiding us through the process, and the production staff at Wiley for meticulously combing through our chapters. And to our technical editors, William Brenneman and Jen Stirrup, we appreciate your suggestions and expertise. The book is better because of you.

Special thanks to my coauthor Jordan Goldmeier, for one obvious reason (the book in your hands) and one not so obvious. Early in my career, I complained to Jordan that people didn't share my interest in statistics and statistical thinking. He said if I'm bothered by it, then it's my obligation to change it. I've been working to fulfill that obligation ever since.

Finally, I'd like to thank my wife Erin one final time (because you've got to save the best for last).

—Alex

I would like to acknowledge the many people who brought this book together.

First, and foremost, I would like to acknowledge my coauthor-in-crime, Alex Gutman. For years, we discussed writing a book together. When the moment was right, we pulled the trigger. I couldn't have asked for a better coauthor.

Thanks to the wonderful folks at Wiley who helped put this together, including acquisition editor Jim Minatel, and project editor John Sleeva. Also, I would like to acknowledge our technical editors, William Brenneman and Jen Stirrup for your hard work reviewing the book. We took your comments to heart.

Last but not least, thank you to my partner, Katie Gray, who always believed in this project—and me.

—Jordan

Table of Contents

- Cover
- Title Page
- Copyright
- Dedication
- About the Authors
- About the Technical Editors
- Acknowledgments
- Foreword
  - NOTE
- Introduction
  - THE DATA SCIENCE INDUSTRIAL COMPLEX
  - WHY WE CARE
  - DATA IN THE WORKPLACE
  - YOU CAN UNDERSTAND THE BIG PICTURE
  - WHO THIS BOOK IS WRITTEN FOR
  - WHY WE WROTE THIS BOOK
  - WHAT YOU'LL LEARN
  - HOW THIS BOOK IS ORGANIZED
  - ONE LAST THING BEFORE WE BEGIN
  - NOTES
- PART I: Thinking Like a Data Head
  - CHAPTER 1: What Is the Problem?
    - QUESTIONS A DATA HEAD SHOULD ASK
    - UNDERSTANDING WHY DATA PROJECTS FAIL
    - WORKING ON PROBLEMS THAT MATTER

CHAPTER SUMMARY
NOTES
CHAPTER 2: What Is Data?
DATA VS. INFORMATION
DATA TYPES
HOW DATA IS COLLECTED AND STRUCTURED
BASIC SUMMARY STATISTICS
CHAPTER SUMMARY
NOTES
CHAPTER 3: Prepare to Think Statistically
ASK QUESTIONS
THERE IS VARIATION IN ALL THINGS
PROBABILITIES AND STATISTICS
CHAPTER SUMMARY
NOTES
PART II: Speaking Like a Data Head
CHAPTER 4: Argue with the Data
WHAT WOULD YOU DO?
TELL ME THE DATA ORIGIN STORY
IS THE DATA REPRESENTATIVE?
WHAT DATA AM I NOT SEEING?
ARGUE WITH DATA OF ALL SIZES
CHAPTER SUMMARY
NOTES
CHAPTER 5: Explore the Data
EXPLORATORY DATA ANALYSIS AND YOU
EMBRACING THE EXPLORATORY MINDSET

CAN THE DATA ANSWER THE QUESTION?
DID YOU DISCOVER ANY RELATIONSHIPS?
DID YOU FIND NEW OPPORTUNITIES IN THE DATA?
CHAPTER SUMMARY
NOTES
CHAPTER 6: Examine the Probabilities
TAKE A GUESS
THE RULES OF THE GAME
PROBABILITY THOUGHT EXERCISE
BE CAREFUL ASSUMING INDEPENDENCE
ALL PROBABILITIES ARE CONDITIONAL
ENSURE THE PROBABILITIES HAVE MEANING
CHAPTER SUMMARY
NOTES
CHAPTER 7: Challenge the Statistics
QUICK LESSONS ON INFERENCE
THE PROCESS OF STATISTICAL INFERENCE
THE QUESTIONS YOU SHOULD ASK TO CHALLENGE THE STATISTICS
CHAPTER SUMMARY
NOTES
PART III: Understanding the Data Scientist's Toolbox
CHAPTER 8: Search for Hidden Groups
UNSUPERVISED LEARNING
DIMENSIONALITY REDUCTION
PRINCIPAL COMPONENT ANALYSIS
CLUSTERING

K-MEANS CLUSTERING
CHAPTER SUMMARY
NOTES
CHAPTER 9: Understand the Regression Model
SUPERVISED LEARNING
LINEAR REGRESSION: WHAT IT DOES
LINEAR REGRESSION: WHAT IT GIVES YOU
LINEAR REGRESSION: WHAT CONFUSION IT CAUSES
OTHER REGRESSION MODELS
CHAPTER SUMMARY
NOTES
CHAPTER 10: Understand the Classification Model
INTRODUCTION TO CLASSIFICATION
LOGISTIC REGRESSION
DECISION TREES
ENSEMBLE METHODS
WATCH OUT FOR PITFALLS
MISUNDERSTANDING ACCURACY
CHAPTER SUMMARY
NOTES
CHAPTER 11: Understand Text Analytics
EXPECTATIONS OF TEXT ANALYTICS
HOW TEXT BECOMES NUMBERS
TOPIC MODELING
TEXT CLASSIFICATION
PRACTICAL CONSIDERATIONS WHEN WORKING WITH TEXT
CHAPTER SUMMARY

NOTES
CHAPTER 12: Conceptualize Deep Learning
NEURAL NETWORKS
APPLICATIONS OF DEEP LEARNING
DEEP LEARNING IN PRACTICE
ARTIFICIAL INTELLIGENCE AND YOU
CHAPTER SUMMARY
NOTES
PART IV: Ensuring Success
CHAPTER 13: Watch Out for Pitfalls
BIASES AND WEIRD PHENOMENA IN DATA
THE BIG LIST OF PITFALLS
CHAPTER SUMMARY
NOTES
CHAPTER 14: Know the People and Personalities
SEVEN SCENES OF COMMUNICATION BREAKDOWNS
DATA PERSONALITIES
CHAPTER SUMMARY
NOTES
CHAPTER 15: What's Next?
Index
End User License Agreement

List of Tables

Chapter 2
TABLE 2.1 Example Dataset on Advertisement Spending and Revenue

Chapter 3

TABLE 3.1 Probability Dentists Agree to an Advertising Claim

TABLE 3.2 Possible Combinations of 4 out of 5 Dentists Agreeing

Chapter 6

TABLE 6.1 Probabilities Scenarios with Associated Notation

TABLE 6.2 Cumulative Probability of a Die Roll Less than 7

Chapter 7

TABLE 7.1 Questions, Null Hypotheses (H_0), and Alternative Hypotheses (H_a)

TABLE 7.2 False Positive vs. False Negative Decision Errors

Chapter 8

TABLE 8.1 Which of These Two Athletes are “Closest” to Each Other?

TABLE 8.2 Clustering Algorithms Get Confused If Your Data Isn't Scaled.

TABLE 8.3 Summarizing Unsupervised Learning and the Supervision Required

Chapter 9

TABLE 9.1 Applications of Supervised Learning

TABLE 9.2 Multiple Linear Regression Model Fit to Housing Data. All correspon...

TABLE 9.3 Sample Housing Data

Chapter 10

TABLE 10.1 Simple Dataset for Logistic Regression: Using GPA to Predict Inter...

TABLE 10.2 Snapshot of the Intern Dataset from HR. The majors are CS = Comput...

TABLE 10.3 Confusion Matrix for Predictions from a Classification Model with ...

TABLE 10.4 Confusion Matrix for Predictions from a Classification Model with ...

Chapter 11

TABLE 11.1 Converting Text to Numbers as a Bag of Words . The numbers represe...

TABLE 11.2 Extending the Bag-of-Words Table with Bigrams. The resulting docum...

TABLE 11.3 Representing Words as Vectors with Word Embeddings

TABLE 11.4 A Basic Spam Classifier Example

Chapter 13

TABLE 13.1 Success Rates of Surgical Techniques to Remove Kidney Stones

TABLE 13.2 Simpson's Paradox Lurking in the Success Rates of Surgical Techniq...

Chapter 14

TABLE 14.1 Seven Scenes of Communication Breakdown

List of Illustrations

Chapter 1

FIGURE 1.1 Sentiment analysis trends

Chapter 3

FIGURE 3.1 Weekly Customer Survey Results: Percent of Positive Reviews. The ...

FIGURE 3.2 Reprint of *American Scientist* figure

Chapter 4

FIGURE 4.1 Plot of test drives with critical component failures as a functio...

FIGURE 4.2 Plots of flights with incidents of O-ring thermal distress as a f...

FIGURE 4.3 Plots of flights with incidents of O-ring thermal distress as a f...

FIGURE 4.4 Plot of test drives with and without critical component failures ...

Chapter 5

FIGURE 5.1 A histogram showing the shape of sales price

FIGURE 5.2 Using box plots to compare sales prices at different quality rank...

FIGURE 5.3 A bar chart showing the counts by types of electrical installatio...

FIGURE 5.4 A line chart showing the number of houses sold in different month...

FIGURE 5.5 A scatter plot showing square footage and sales price

FIGURE 5.6 Square footage and sales price have a correlation of 0.62, which ...

FIGURE 5.7 Two datasets with a correlation of 0.8

FIGURE 5.8 Datasaurus: Data is free to download and explore.[9] Like Anscombe’...

Chapter 6

FIGURE 6.1 Venn diagram showing the probability of two events happening toge...

FIGURE 6.2 Tree diagram for scanning computers for a virus at a large compan...

Chapter 8

FIGURE 8.1 Sorting cars based on different composite features. Notice how th...

FIGURE 8.2 Principal component analysis groups and condenses the *columns* of ...

FIGURE 8.3 PCA finds optimal weights that are used to create composite featu...

FIGURE 8.4 The PCA algorithm creates a new dataset, the same size as the ori...

FIGURE 8.5 Clustering is a technique that groups *rows* of a dataset together....

FIGURE 8.6 The company's 200 locations, before clustering

FIGURE 8.7 *k*-means in action on retail locations

Chapter 9

FIGURE 9.1 Basic paradigm of supervised learning: mapping inputs to outputs...

FIGURE 9.2 Many lines would fit this data reasonably well, but which line is...

FIGURE 9.3 Least squares regression is finding the line through the data tha...

FIGURE 9.4 Two competing models. The model on the left generalizes well, whi...

FIGURE 9.5 In this plot, you can see how the model does not do well predicti...

Chapter 10

FIGURE 10.1 Fitting different logistic regression models to the data. The mo...

FIGURE 10.2 Applying the logistic regression model to make predictions at GP...

FIGURE 10.3 Simple decision tree applied to the HR intern dataset

FIGURE 10.4 A random forest is a “forest” of several decision trees, usually...

Chapter 11

FIGURE 11.1 A word cloud for the text in this chapter

FIGURE 11.2 Processing text down to a bag of words

FIGURE 11.3 Clustering documents and terms together with topic modeling. Can...

Chapter 12

FIGURE 12.1 The simplest neural network possible. The four inputs are proces...

FIGURE 12.2 A neural network with a hidden layer. The middle layer is “hidde...

FIGURE 12.3 A deep neural network with two hidden layers

FIGURE 12.4 Theoretical performance curves of traditional regression and cla...

FIGURE 12.5 How a grayscale image “looks” to a computer, and how that data w...

FIGURE 12.6 Color images are represented as 3D matrices for the pixel values...

FIGURE 12.7 Convolution is like a series of magnifying glasses, detecting di...

FIGURE 12.8 A simple representation of a recurrent neural network

FIGURE 12.9 Deep learning is a subfield of machine learning, which is a subf...

Foreword

Becoming a Data Head is well-timed for the current state of data and analytics within organizations. Let's quickly review some recent history. A few leading companies have made effective use of data and analytics to guide their decisions and actions for several decades, starting in the 1970s. But most ignored this important resource, or left it hiding in back rooms with little visibility or importance.

But in the early to mid-2000s this situation began to change, and companies began to get excited about the potential for data and analytics to transform their business situations. By the early 2010s, the excitement began to shift toward “big data,” which originally came from Internet companies but began to pop up across sophisticated economies. To deal with the increased volume and complexity of data, the “data scientist” role arose with companies—again, first in Silicon Valley, but then everywhere.

However, just as firms were beginning to adjust to big data, the emphasis shifted again—around about 2015 to 2018 in many firms—to a renewed focus on artificial intelligence. Collecting, storing, and analyzing big data gave way to machine learning, natural language processing, and automation.

Embedded within these rapid shifts in focus were a series of assumptions about data and analytics within organizations. I am happy to say that *Becoming a Data Head* violates many of them, and it's about time. As many who work with or closely observe these trends are beginning to admit, we have headed in some unproductive directions based on these assumptions. For the rest of this

foreword, then, I'll describe five interrelated assumptions and how the ideas in this book justifiably run counter to them.

Assumption 1: Analytics, big data, and AI are wholly different phenomena.

It is assumed by many onlookers that “traditional” analytics, big data, and AI are separate and different phenomena. *Becoming a Data Head*, however, correctly adopts the view that they are highly interrelated. All of them involve statistical thinking. Traditional analytics approaches like regression analysis are used in all three, as are data visualization techniques. Predictive analytics is basically the same thing as supervised machine learning. And most techniques for data analysis work on any size of dataset. In short, a good Data Head can work effectively across all three, and spending a lot of time focusing on the differences among them isn't terribly productive.

Assumption 2: Data scientists are the only people who can play in this sandbox.

We have lionized data scientists and have often made the assumption that they are the only people who can work effectively with data and analytics. However, there is a nascent but important move toward the democratization of these ideas; increasing numbers of organizations are empowering “citizen data scientists.” Automated machine learning tools make it easier to create models that do an excellent job of predicting. There is still a need, of course, for professional data scientists to develop new algorithms and check the work of the citizens who do complex analysis. But organizations that democratize analytics and data science—putting their “amateur” Data Heads to work—

can greatly increase their overall use of these important capabilities.

Assumption 3: Data scientists are “unicorns” who have all the skills needed for these activities.

We have assumed that data scientists—those trained in and focused upon the development and coding of models—are also able to perform all the other tasks that are required for full implementation of those models. In other words, we think they are “unicorns” who can do it all. But such unicorns don't exist at all, or exist only in small numbers. Data Heads who not only understand the rudiments of data science, but also know the business, can manage projects effectively, and are excellent at building business relationships will be extremely valuable in data science projects. They can be productive members of data science teams and increase the likelihood that data science projects will lead to business value.

Assumption 4: You need to have a really high quantitative IQ and lots of training to succeed with data and analytics.

A related assumption is that in order to do data science work, a person has to be very well trained in the field and that a Data Head requires a head that is very good with numbers. Both quantitative training and aptitude certainly help, but *Becoming a Data Head* argues—and I agree—that a motivated learner can master enough of data and analytics to be quite useful on data science projects. This is in part because the general principles of statistical analysis are by no means rocket science, and also because “being useful” on data science projects doesn't require an extremely high level of data and analytics mastery. Working with professional data scientists or automated AI programs only requires the

ability and the curiosity to ask good questions, to make connections between business issues and quantitative results, and to look out for dubious assumptions.

Assumption 5: If you didn't study mostly quantitative fields in college or graduate school, it's too late for you to learn what you need to work with data and analytics.

This assumption is supported by survey data; in a 2019 survey report from Splunk of about 1300 global executives, virtually every respondent (98%) agreed that data skills are important to the jobs of tomorrow.[1] 81% of the executives agree that data skills are required to become a senior leader in their companies, and 85% agree that data skills will become more valuable in their firms. Nonetheless, 67% say they are not comfortable accessing or using data themselves, 73% feel that data skills are harder to learn than other business skills, and 53% believe they are too old to learn data skills. This “data defeatism” is damaging to individuals and organizations, and neither the authors of this book nor I believe it is warranted. Peruse the pages following this foreword, and you will see that no rocket science is involved!

So forget these false assumptions, and turn yourself into a Data Head. You'll become a more valuable employee and make your organization more successful. This is the way the world is going, so it's time to get with the program and learn more about data and analytics. I think you will find the process—and the reading of *Becoming a Data Head*—more rewarding and more pleasant than you may imagine.

Thomas H. Davenport
Distinguished Professor, Babson College
Visiting Professor, Oxford Saïd Business School