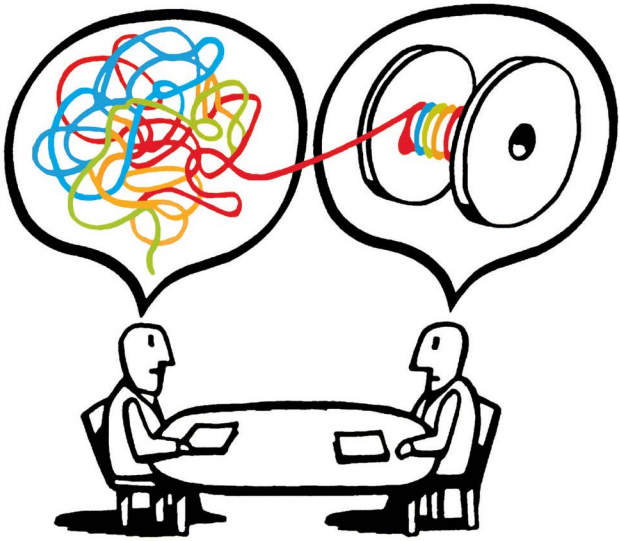
O'REILLY®

Thinking with Data

How to Turn Information into Insights



Max Shron

Praise for *Thinking with Data*

"*Thinking with Data* gets to the essence of the process, and guides data scientists in answering that most important question—what's the problem we're really trying to solve?"

— *Hilary Mason*
Data Scientist in Residence at Accel Partners; co-founder of the DataGotham Conference

"*Thinking with Data* does a wonderful job of reminding data scientists to look past technical issues and to focus on making an impact on the broad business objectives of their employers and clients. It's a useful supplement to a data science curriculum that is largely focused on the technical machinery of statistics and computer science."

— *John Myles White*
Scientist at Facebook; author of *Machine Learning for Hackers* and *Bandit Algorithms for Website Optimization*

"This is a great piece of work. It will be required reading for my team."

— *Nick Kolegraff*
Director of Data Science at Rackspace

"Shron's *Thinking with Data* is a nice mix of academic traditions, from design to philosophy, that rescues data from mathematics and the regime of pure calculation. … These are lessons that should be included in any data science course!"

— *Mark Hansen*
Director of David and Helen Gurley Brown Institute for Media Innovation; Graduate School of Journalism at Columbia University

This page intentionally left blank

Thinking with Data

Max Shron

O'REILLY® Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

THINKING WITH DATA

by Max Shron

Copyright © 2014 Max Shron. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://my.safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editors: Mike Loukides and Ann Spencer
Production Editor: Kristen Brown
Copyeditor: O'Reilly Production Services
Proofreader: Kim Cofer

Cover Designer: Karen Montgomery
Interior Designer: David Futato
Illustrator: Rebecca Demarest

February 2014: First Edition

Revision History for the First Edition:

2014-01-16: First release

See *http://oreilly.com/catalog/errata.csp?isbn=9781449362935* for release details.

Nutshell Handbook, the Nutshell Handbook logo, and the O'Reilly logo are registered trademarks of O'Reilly Media, Inc. *Thinking with Data* and related trade dress are trademarks of O'Reilly Media, Inc.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps.

While every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

ISBN: 978-1-449-36293-5
[LSI]

Contents

Preface | vii

1 | Scoping: Why Before How 1

2 | What Next? 17

3 | Arguments 31

4 | Patterns of Reasoning 43

5 | Causality 57

6 | Putting It All Together 67

A | Further Reading 77