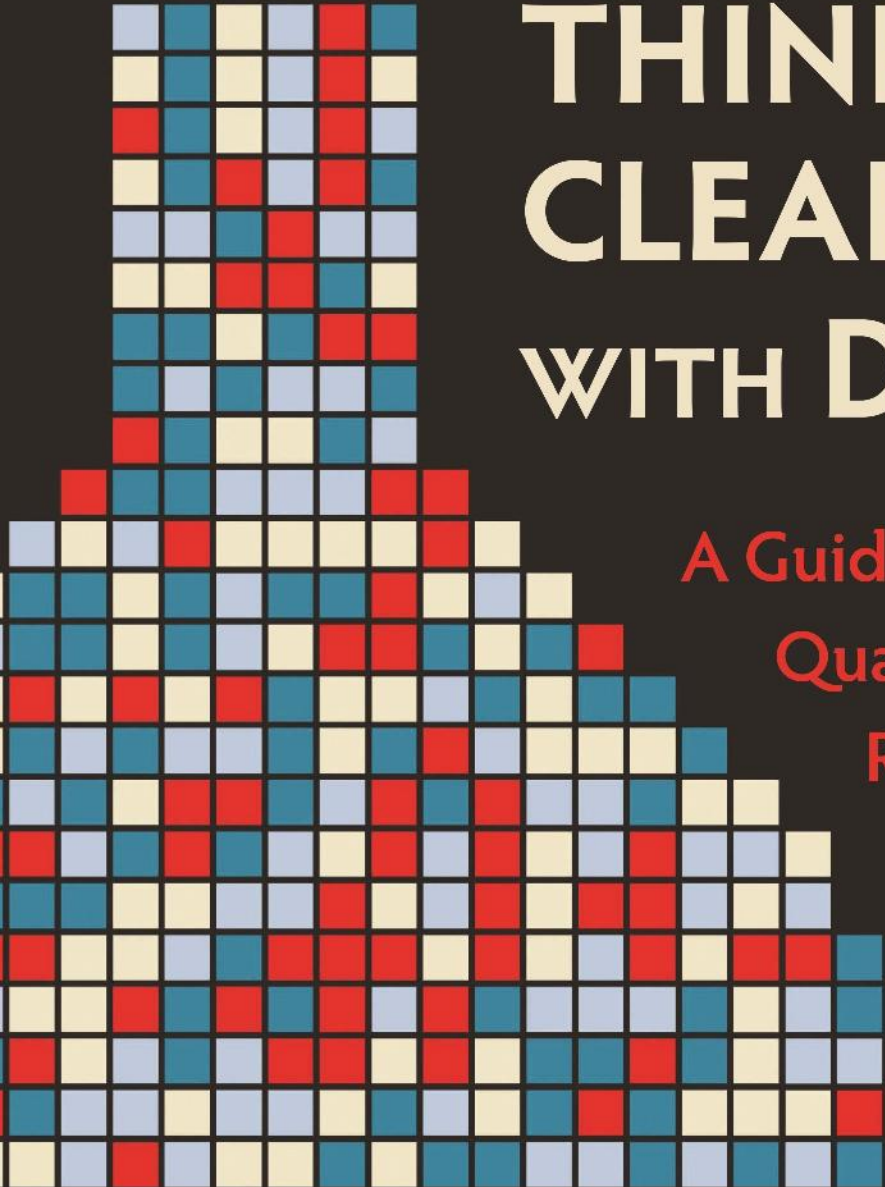




THINKING CLEARLY WITH DATA



A Guide to
Quantitative
Reasoning
and Analysis



Ethan
Bueno de Mesquita

Anthony
Fowler

Thinking Clearly with Data

Thinking Clearly with Data

A Guide to Quantitative Reasoning
and Analysis

ETHAN BUENO DE MESQUITA
ANTHONY FOWLER

PRINCETON UNIVERSITY PRESS
Princeton and Oxford

Copyright © 2021 by Princeton University Press

Princeton University Press is committed to the protection of copyright and the intellectual property our authors entrust to us. Copyright promotes the progress and integrity of knowledge. Thank you for supporting free speech and the global exchange of ideas by purchasing an authorized edition of this book. If you wish to reproduce or distribute any part of it in any form, please obtain permission.

Requests for permission to reproduce material from this work should be sent to permissions@press.princeton.edu

Published by Princeton University Press
41 William Street, Princeton, New Jersey 08540
6 Oxford Street, Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

All Rights Reserved

Library of Congress Cataloging-in-Publication Data

Names: Bueno de Mesquita, Ethan, 1974– author. | Fowler, Anthony, author.

Title: Thinking clearly with data : a guide to quantitative reasoning and analysis / Ethan Bueno de Mesquita and Anthony Fowler.

Description: 1st. edition. | Princeton : Princeton University Press, 2021. | Includes index.

Identifiers: LCCN 2021011897 (print) | LCCN 2021011898 (ebook) | ISBN 9780691214368 (hardback) | ISBN 9780691214351 (paperback) | ISBN 9780691215013 (epub)

Subjects: LCSH: Sociology–Statistical methods. | Sociology–Methodology.

Classification: LCC HM535 .B84 2021 (print) | LCC HM535 (ebook) | DDC 302.01/5195–dc23

LC record available at <https://lcn.loc.gov/2021011897>

LC ebook record available at <https://lcn.loc.gov/2021011898>

British Library Cataloging-in-Publication Data is available

Editorial: Bridget Flannery-McCoy, Alena Chekanov

Jacket/Cover Design: Wanda España

Production: Erin Suydam

Publicity: Kate Hensley

Copyeditor: Elizabeth J. Asbornio

Cover image: Igor Kisselev / Alamy Stock Vector

This book has been composed in Minion Pro

Printed on acid-free paper. ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Ethan: for Abe and Hannah

Anthony: for Gloria

Short Contents

	<i>Preface</i>	xvii
CHAPTER 1	Thinking Clearly in a Data-Driven Age	1
PART I	ESTABLISHING A COMMON LANGUAGE	11
CHAPTER 2	Correlation: What Is It and What Is It Good For?	13
CHAPTER 3	Causation: What Is It and What Is It Good For?	37
PART II	DOES A RELATIONSHIP EXIST?	53
CHAPTER 4	Correlation Requires Variation	55
CHAPTER 5	Regression for Describing and Forecasting	74
CHAPTER 6	Samples, Uncertainty, and Statistical Inference	94
CHAPTER 7	Over-Comparing, Under-Reporting	113
CHAPTER 8	Reversion to the Mean	138
PART III	IS THE RELATIONSHIP CAUSAL?	157
CHAPTER 9	Why Correlation Doesn't Imply Causation	159
CHAPTER 10	Controlling for Confounders	193
CHAPTER 11	Randomized Experiments	218
CHAPTER 12	Regression Discontinuity Designs	243

CHAPTER 13	Difference-in-Differences Designs	266
CHAPTER 14	Assessing Mechanisms	290
PART IV	FROM INFORMATION TO DECISIONS	303
CHAPTER 15	Turn Statistics into Substance	305
CHAPTER 16	Measure Your Mission	336
CHAPTER 17	On the Limits of Quantification	357
	<i>Index</i>	371

Contents

	<i>Preface</i>	xvii
	Organization	xviii
	Who Is This Book For?	xix
	Acknowledgments	xx
CHAPTER 1	Thinking Clearly in a Data-Driven Age	1
	What You'll Learn	1
	Introduction	1
	Cautionary Tales	2
	Abe's hasty diagnosis	2
	Civil resistance	3
	Broken-windows policing	5
	Thinking and Data Are Complements, Not Substitutes	7
	Readings and References	9
PART I	ESTABLISHING A COMMON LANGUAGE	11
CHAPTER 2	Correlation: What Is It and What Is It Good For?	13
	What You'll Learn	13
	Introduction	13
	What Is a Correlation?	13
	Fact or correlation?	17
	What Is a Correlation Good For?	19
	Description	19
	Forecasting	20
	Causal inference	23
	Measuring Correlations	24
	Mean, variance, and standard deviation	24
	Covariance	27
	Correlation coefficient	28
	Slope of the regression line	29
	Populations and samples	29
	Straight Talk about Linearity	30
	Wrapping Up	33
	Key Terms	33

	Exercises	34
	Readings and References	36
CHAPTER 3	Causation: What Is It and What Is It Good For?	37
	What You'll Learn	37
	Introduction	37
	What Is Causation?	38
	Potential Outcomes and Counterfactuals	39
	What Is Causation Good For?	40
	The Fundamental Problem of Causal Inference	41
	Conceptual Issues	42
	What is the cause?	42
	Causality and counterexamples	44
	Causality and the law	47
	Can causality run backward in time?	47
	Does causality require a physical connection?	48
	Causation need not imply correlation	49
	Wrapping Up	49
	Key Terms	50
	Exercises	50
	Readings and References	52
PART II	DOES A RELATIONSHIP EXIST?	53
CHAPTER 4	Correlation Requires Variation	55
	What You'll Learn	55
	Introduction	55
	Selecting on the Dependent Variable	56
	The 10,000-hour rule	57
	Corrupting the youth	59
	High school dropouts	62
	Suicide attacks	63
	The World Is Organized to Make Us Select on the Dependent Variable	64
	Doctors mostly see sick people	65
	Post-mortems	65
	The <i>Challenger</i> disaster	67
	The financial crisis of 2008	69
	Life advice	69
	Wrapping Up	70
	Key Term	70
	Exercises	70
	Readings and References	72
CHAPTER 5	Regression for Describing and Forecasting	74
	What You'll Learn	74
	Introduction	74
	Regression Basics	74
	Linear Regression, Non-Linear Data	79

	The Problem of Overfitting	87
	Forecasting presidential elections	87
	How Regression Is Presented	89
	A Brief Intellectual History of Regression	89
	Wrapping Up	91
	Key Terms	91
	Exercises	92
	Readings and References	93
CHAPTER 6	Samples, Uncertainty, and Statistical Inference	94
	What You'll Learn	94
	Introduction	94
	Estimation	94
	Why Do Estimates Differ from Estimands?	96
	Bias	96
	Noise	97
	What Makes for a Good Estimator?	98
	Quantifying Precision	99
	Standard errors	99
	Small samples and extreme observations	101
	Confidence intervals	102
	Statistical Inference and Hypothesis Testing	103
	Hypothesis testing	103
	Statistical significance	104
	Statistical Inference about Relationships	105
	What If We Have Data for the Whole Population?	106
	Substantive versus Statistical Significance	107
	Social media and voting	107
	The Second Reform Act	108
	Wrapping Up	109
	Key Terms	109
	Exercises	110
	Readings and References	111
CHAPTER 7	Over-Comparing, Under-Reporting	113
	What You'll Learn	113
	Introduction	113
	Can an octopus be a soccer expert?	113
	Publication Bias	118
	p -hacking	119
	p -screening	120
	Are Most Scientific "Facts" False?	122
	ESP	122
	Get out the vote	123
	p -hacking forensics	124
	Potential Solutions	126
	Reduce the significance threshold	126
	Adjust p -values for multiple testing	127
	Don't obsess over statistical significance	127

	Pre-registration	127
	Requiring pre-registration in drug trials	128
	Replication	128
	Football and elections	129
	Test important and plausible hypotheses	130
	The power pose	131
	Beyond Science	131
	Superstars	132
	Wrapping Up	134
	Key Terms	134
	Exercises	134
	Readings and References	136
CHAPTER 8	Reversion to the Mean	138
	What You'll Learn	138
	Introduction	138
	Does the truth wear off?	138
	Francis Galton and Regression to Mediocrity	139
	Reversion to the Mean Is Not a Gravitational Force	142
	Seeking Help	145
	Does knee surgery work?	146
	Reversion to the Mean, the Placebo Effect, and Cosmic Habituation	147
	The placebo effect	147
	Cosmic habituation explained	148
	Cosmic habituation and genetics	150
	Beliefs Don't Revert to the Mean	150
	Wrapping Up	152
	Key Words	152
	Exercises	152
	Readings and References	155
PART III	IS THE RELATIONSHIP CAUSAL?	157
CHAPTER 9	Why Correlation Doesn't Imply Causation	159
	What You'll Learn	159
	Introduction	159
	Charter schools	160
	Thinking Clearly about Potential Outcomes	163
	Sources of Bias	168
	Confounders	168
	Reverse causality	169
	The 10,000-hour rule, revisited	170
	Diet soda	173
	How Different Are Confounders and Reverse Causality?	174
	Campaign spending	174
	Signing the Bias	176
	Contraception and HIV	179
	Mechanisms versus Confounders	181
	Thinking Clearly about Bias and Noise	183

	Wrapping Up	186
	Key Terms	187
	Exercises	187
	Readings and References	191
CHAPTER 10	Controlling for Confounders	193
	What You'll Learn	193
	Introduction	193
	Party whipping in Congress	193
	A note on heterogeneous treatment effects	197
	The Anatomy of a Regression	198
	How Does Regression Control?	201
	Controlling and Causation	209
	Is social media bad for you?	210
	Reading a Regression Table	211
	Controlling for Confounders versus Mechanisms	213
	There Is No Magic	214
	Wrapping Up	215
	Key Terms	215
	Exercises	216
	Readings and References	217
CHAPTER 11	Randomized Experiments	218
	What You'll Learn	218
	Introduction	218
	Breastfeeding	219
	Randomization and Causal Inference	221
	Estimation and Inference in Experiments	224
	Standard errors	224
	Hypothesis testing	225
	Problems That Can Arise with Experiments	225
	Noncompliance and instrumental variables	226
	Chance imbalance	232
	Lack of statistical power	234
	Attrition	235
	Interference	236
	Natural Experiments	237
	Military service and future earnings	238
	Wrapping Up	239
	Key Terms	239
	Exercises	240
	Readings and References	242
CHAPTER 12	Regression Discontinuity Designs	243
	What You'll Learn	243
	Introduction	243
	How to Implement an RD Design	247
	Are extremists or moderates more electable?	249
	Continuity at the Threshold	251
	Does continuity hold in election RD designs?	255

	Noncompliance and the Fuzzy RD	256
	Bombing in Vietnam	257
	Motivation and Success	261
	Wrapping Up	262
	Key Terms	262
	Exercises	262
	Readings and References	264
CHAPTER 13	Difference-in-Differences Designs	266
	What You'll Learn	266
	Introduction	266
	Parallel Trends	267
	Two Units and Two Periods	269
	Unemployment and the minimum wage	269
	N Units and Two Periods	272
	Is watching TV bad for kids?	273
	N Units and N Periods	275
	Contraception and the gender-wage gap	276
	Useful Diagnostics	278
	Do newspaper endorsements affect voting decisions?	278
	Is obesity contagious?	279
	Difference-in-Differences as Gut Check	282
	The democratic peace	282
	Wrapping Up	285
	Key Terms	285
	Exercises	286
	Readings and References	288
CHAPTER 14	Assessing Mechanisms	290
	What You'll Learn	290
	Introduction	290
	Causal Mediation Analysis	291
	Intermediate Outcomes	292
	Cognitive behavioral therapy and at-risk youths in Liberia	293
	Independent Theoretical Predictions	294
	Do voters discriminate against women?	294
	Testing Mechanisms by Design	295
	Social pressure and voting	295
	Disentangling Mechanisms	296
	Commodity price shocks and violent conflict	296
	Wrapping Up	298
	Key Terms	299
	Exercises	299
	Readings and References	300
PART IV	FROM INFORMATION TO DECISIONS	303
CHAPTER 15	Turn Statistics into Substance	305
	What You'll Learn	305

Introduction	305
What's the Right Scale?	305
Miles-per-gallon versus gallons-per-mile	306
Percent versus percentage point	309
Visual Presentations of Data	309
Policy preferences and the Southern realignment	311
Some rules of thumb for data visualization	314
From Statistics to Beliefs: Bayes' Rule	314
Bayes' rule	317
Information, beliefs, priors, and posteriors	318
Abe's celiac revisited	319
Finding terrorists in an airport	322
Bayes' rule and quantitative analysis	325
Expected Costs and Benefits	328
Screening frequently or accurately	329
Wrapping Up	331
Key Words	331
Exercises	332
Readings and References	334
 CHAPTER 16	
Measure Your Mission	336
What You'll Learn	336
Introduction	336
Measuring the Wrong Outcome or Treatment	337
Partial measures	337
Metal detectors in airports	337
Intermediate outcomes	339
Blood pressure and heart attacks	340
Ill-defined missions	341
Climate change and economic productivity	342
Do You Have the Right Sample?	343
External validity	343
Malnutrition in India and Bangladesh	343
Selected samples	344
College admissions	345
Why can't major league pitchers hit?	345
Strategic Adaptation and Changing Relationships	349
The duty on lights and windows	349
The shift in baseball	350
The war on drugs	351
Wrapping Up	353
Key Words	353
Exercises	353
Readings and References	355
 CHAPTER 17	
On the Limits of Quantification	357
What You'll Learn	357
Introduction	357
Decisions When Evidence Is Limited	358
Cost-benefit analysis and environmental regulation	358

Floss your teeth and wear a mask	359
Floss your teeth	359
Wear a mask	360
Quantification and Values	361
How quantitative tools sneak in values	361
Algorithms and racial bias in health care	361
How quantification shapes our values	363
Think Clearly and Help Others Do So Too	367
Exercises	367
Readings and References	368
 <i>Index</i>	 371