Annotated Table of Contents

Preface and Acknowledgments *page* xiii

Introduction 1

Section I Thinking Like a Data Scientist 11

1 How the Rule of 72 Can Provide Guidance to Advance Your Wealth, Your Career, and Your Gas Mileage 14
Exponential growth is something human intuition cannot comprehend. In this chapter we illustrate this with several examples drawn from history and current experience. Then we introduce a simple rule of thumb, often used to help financial planners tame the cognitive load of exponential growth, and show how it can be used more widely to help explain a broad range of other issues. The Rule of 72 illustrates the power of having such "rules" in your toolbox for use as the need arises.

2 Piano Virtuosos and the Four-Minute Mile 19
The frequency of truly extreme observations and the size of the sample of observations being considered are inexorably related. Over the last century the number of musical virtuosos has ballooned to include copious numbers of high school students who can perform pieces that would have daunted all but the most talented artists in the past. In this chapter we find that a simple mathematical model explains this result as well as why a runner breaking the four-minute barrier in the mile has ceased to be newsworthy.

3 Happiness and Causal Inference 22
Here we are introduced to Rubin's Model for Causal Inference, which directs us to focus on measuring the effect of a cause rather than chasing a chimera by trying to find the cause of an effect. This reorientation leads us naturally to the random assignment-controlled experiment as a key companion to the scientific method. The power of this approach is illustrated by laying out how we can untie the Gordian knot that entangles happiness and performance. It also provides a powerful light that can be used to illuminate the dark corners of baseless claims.

4 Causal Inference and Death 29
The path toward estimating the size of causal effects does not always run smoothly in the real world because of the ubiquitous nuisance of missing data. In this chapter we examine the very practical situation in which unexpected events occur that unbalance carefully constructed experiments. We use a medical example in which some patients die inconveniently mid-experiment, and we must try to estimate a causal effect despite this disturbance. Once again Rubin's Model guides us to a solution, which is unexpectedly both subtle and obvious, once you get used to it.

5 Using Experiments to Answer Four Vexing Questions 43
Public education is a rich field for the application of rigorous methods for the making of causal inferences. Instead, we find that truthiness manifests itself widely within the discussions surrounding public education, the efficacy of which is often measured with tests. Thus it is not surprising that many topics associated with testing arise in which the heat of argument on both sides of the question overwhelms facts. We examine four questions that either have already been decided in courts (but not decisively) or are on the way to court as this chapter was being written.

6 Causal Inferences from Observational Studies: Fracking, Injection Wells, Earthquakes, and Oklahoma 61
It is not always practical to perform an experiment, and we must make do with an observational study. Over the past six years the

number of serious earthquakes in Oklahoma (magnitude 3.0 or more) has increased from less than two a year to almost two a day. In this chapter we explore how we can use an observational study to estimate the size of the causal effect of fracking and the disposal of wastewater through its high-pressure injection into the earth on seismicity. The evidence for such a connection is overwhelming despite denials from state officials and representatives of the petroleum industry.

7 Life Follows Art: Gaming the Missing Data Algorithm 72
A compelling argument can be made that the biggest problem faced by data scientists is what to do about observations that are missing (missing data). In this chapter we learn of what initially seem like completely sensible approaches for dealing with the inevitable missing data, yet they were being exploited to improperly game the system. It also illustrates what may be the most effective way to deal with such shenanigans.

Section II Communicating Like a Data Scientist 79

8 On the Crucial Role of Empathy in the Design of Communications: Genetic Testing as an Example 82
Graphical display is perhaps the most important tool data science possesses that allows the data to communicate their meaning to the data scientist. They are also unsurpassed in thence allowing the scientist to communicate to everyone else as well. By far, the most crucial attitude that anyone wishing to communicate effectively can have is a strong sense of empathy. In this chapter we discuss two different communications and show how the lessons learned from Princeton University's acceptance letter could be efficaciously applied in communicating the results of a genetic test for mutations that increase a woman's likelihood of cancer.

9 Improving Data Displays: The Media's and Ours 91
In the transactions between scientists and the general public, where influence flows in both directions, we see how advances in graphical display pioneered in the scientific literature were adopted by the mass media, and how advances in the media have been unfortunately slow to catch on among scientists.

10 Inside Out Plots 109

One of the greatest design challenges faced in visually displaying high-dimensional data (data involving more than just two variables) is the limitations of a two-dimensional plotting medium (a piece of paper or a computer screen). In this chapter we illustrate how to use inside out plots to reveal many of the secrets that might be contained in such data sets. The example we chose compares the performances of six baseball all-stars on eight variables.

11 A Century and a Half of Moral Statistics: Plotting Evidence to Affect Social Policy 122

Any data set that mixes geographic variables with other measures (like election outcomes by state or population per census tract) cries out for a map. Maps are the oldest of graphical displays with surviving examples from Nilotic surveyors in ancient Egypt and even older Chinese maps that long predate the Xia dynasty. A map is the obvious visual metaphor, using the two dimensions of the plotting plane to represent the geographic information. Only much later were other, nongeographic variables added on top of the geographic background. In this chapter we are introduced to Joseph Fletcher, a nineteenth-century British lawyer and statistician, who plotted such moral statistics as the prevalence of ignorance, bastardy, crime, and improvident marriages, on maps of England and Wales. The discussion of his works ranges broadly over what Fletcher did, why, and how his goal of increased social justice might have been aided through the use of more modern display methods.

Section III Applying the Tools of Data Science to Education 139

Public education touches everyone. We all pay for it through our local property taxes and almost all of us have partaken of it ourselves as well as through our children. Yet it is hard to think of any other area of such a broad-based activity that is so riddled with misconceptions borne of truthiness. In this section we will examine five different areas in which public opinion has been shaped by claims made from anecdotes and not from

evidence. Each chapter describes a claim and then presents widely available evidence that clearly refutes it. This section is meant as a consilience in which the methods introduced and illustrated in sections I and II are used to reinforce an attitude of skepticism while providing an evidence-based approach to assessing the likelihood of the claims being credible.

12 Waiting for Achilles 143
The educational system of the United States is regularly excoriated for the poor performance of its students and the disappointingly robust differences between the scores of black and white students. In this chapter we use evidence to clarify both of these issues and, in so doing, discover that the situation is not anywhere near as bleak as truthiness-driven critics would have us believe.

13 How Much Is Tenure Worth? 146
Critics of public education often lay a considerable portion of the blame for its shortcomings at the door of teacher tenure. In this chapter we trace the origins of tenure and provide evidence that eliminating it may be far more expensive and less effective than its critics suggest.

14 Detecting Cheating Badly: If It Could Have Been, It Must Have Been 152
Whenever tests have important consequences there is always the possibility of cheating. To limit cheating, student performances are scrutinized and severe punishments are sometime meted out. In this chapter we describe two instances in which the fervor of the investigation outstripped the evidence supporting the claim of a security infraction.

15 When Nothing Is Not Zero: A True Saga of Missing Data, Adequate Yearly Progress, and a Memphis Charter School 161
Increasingly often, the efficacy of schools is determined by the test scores earned by their students. In this chapter we learn of a charter school in Memphis that was placed on probation because the average scores of its students were too low. Unfortunately this apparent deficiency was not due to the school but rather to the city's treatment of missing data.

16 Musing about Changes in the SAT: Is the College Board Getting Rid of the Bulldog? 167
Modern college entrance exams in the United States have existed for about ninety years, and in that time, changes in the exams, their scoring, and their use have been made steadily. In this chapter we use evidence and statistical thinking to focus our discussion of three changes to the SAT that have been recently announced by the College Board. Two of them are likely to have almost no effect, but the third is a major change. I hypothesize why these particular changes were selected and conclude that the College Board might have been guided by the strategy developed by Dartmouth's former president John Kemeny when he led the school to coeducation in the 1970s.

17 For Want of a Nail: Why Worthless Subscores May Be Seriously Impeding the Progress of Western Civilization 175
In the 2010 U.S. Census, it cost about forty dollars for each person counted. This seems like an extravagant amount because changes in the population of the United States can be accurately estimated by noting that it increases by about one person every thirteen seconds. Yet the price is sensible because of the many small area estimates that Census data provide. In this chapter we examine the costs of testing in this same light and conclude that the opportunity costs of too long tests are likely extensive enough that they may be impeding progress in a serious way.

Section IV Conclusion: ~~Don't~~ Try This at Home 187

Bibliography 195
Sources 203
Index 205