Introduction to Statistical Thought.

Michael LAVINE. Published at *http://www.stat.duke.edu/~michael/book. html*, 2006. ISBN: none. 328 pp. $0.00.

Traditional curricula in statistics, at both the graduate and the advanced undergraduate levels, divide coursework into theory and application. Over time, theory courses have incorporated more applied material. In this book, suitable for a first-year graduate or upper-level undergraduate theory course, the author blends theory and application. This is done quite successfully, and the links between the two areas are much stronger than in other theory books that attempt the same blend. As an added bonus, the book introduces the reader to computation with R.

The author's goals are laid out in the Preface: to convey "statistical thinking," placing a strong emphasis on the ideas behind the procedures that we use and somewhat less emphasis on the technical details that underlie them; to emphasize the likelihood function as a unifying concept for data analysis, as opposed to a plethora of specialized criteria; and to integrate the use of software into the learning process. The author hits his mark on all accounts with a lively, engaging book that, remarkably, is available as a free download.

The book is organized into seven chapters: "Probability," "Modes of Inference," "Regression," "More Probability," "Special Distributions," "More Models," and "Mathematical Statistics." Although the titles of the chapters give a feel for the contents, much more material is covered. The first chapter, "Probability," for example, covers the usual material on the basics of probability, random variables, summaries of distributions, and selected families of distributions. It also contains extended examples that will hook the student and that demonstrate, in substantial fashion, *why* we are interested in the various summaries of distributions.

The book focuses on likelihood as the main driver of inference. It does so in a sophisticated fashion. Estimation is presented through likelihood, first treating maximum likelihood estimation, following with interval estimation based on the likelihood, then introducing Bayesian methods, and closing with hypothesis testing. The justification of the normal likelihood as an approximation to an actual likelihood is particularly nice. There is scant discussion of unbiasedness, no discussion of the method of moments, and no development of uniformly minimum variance unbiased estimators.

The book integrates computing with theory and application. In contains a wealth of R code. The code is well documented and explained. It progresses from introductory, fairly standard code—the first bit of code takes the reader from no knowledge of R through a simulation of the game of craps—to code tailored to look at features of particular datasets. The amount of detail and level of explanation of the code progresses at an appropriate pace. The reader who works through the examples will, on finishing the book, have acquired a sound working knowledge of R.

The quality of discussion on applications is a strength of this book. Chapter 3, Regression, provides strong motivation for the questions that we ask in regression analysis in the context of specific, lengthily described examples. The focus is on data exploration and modeling, as opposed to the all-too-common cursory numerical treatment in most theory-oriented books. The examples are accompanied by development of normal theory and generalized linear models. Along the way, the accompanying R code lets the reader explore the techniques. Formal inference is deferred until later chapters.

The book is well written. Experienced statisticians will find themselves reading along as if it were a novel. The clarity of the writing and judicious choice of examples let the ideas shine forth. Most theorems contain results that capture the main ideas while avoiding extensions that require additional technical details and that can obscure the essence of the results. Commentary in the text tells the reader how to extend the results.

The reader should be warned that there are a few gaps in the book's current version. Some sections, particularly toward the end, are sketches rather than full presentations. With the evolutionary nature of an on-line book, look for these sections to be filled out in the future.

I would recommend this book to diverse audiences. For first-year graduate students, I recommend downloading a copy to read in parallel with their theory course. The extra attention given to explanation of thoughts behind the techniques, along with the space devoted to more complete examples, should help students make the transition from undergraduate-level to graduate-level work. For instructors of theory courses, this book provides thoughtful explanations of many core concepts. It also shows how applications can be successfully incorporated into the course. A nice set of exercises is included at the end of each chapter. Finally, with his characteristic innovation, the author has produced an intriguing book that is available for free. Anyone with the vaguest interest in statistical thought material should at least take a look.

Steven N. MACEACHERN
Ohio State University

The Grammar of Graphics.

Leland WILKINSON. New York: Springer, 2005. ISBN 0-387-24544-8. xviii + 690 pp. $79.95.

I found the first edition of this book (Wilkinson 1999) quite remarkable. So did others. In fact, previous reviewers heralded it as "destined to become a land mark in statistical graphics" (Wegman 2000, p. 1009), "a worthy descendent of Bertin's *Sémiologie*, of Tukey's *EDA*" (Wainer 2001, p. 308), "a new ambitious step for the development and application of statistical graphics" (Polasek 2002, p. 562), and "a *tour de force* of the highest order" (Wainer 2001, p. 307). One reviewer suggested that "all geography and map libraries should add this book to their collections; the serious scholar of quantitative data graphics will place this book on the same shelf with those by Edward Tufte, and volumes by Cleveland, Bertin, Monmonier, [and] MacEachren" (Kirby 2001, pp. 55–56). This indeed is a tough act to follow. So what about the second edition?

The second edition continues the intellectual momentum of the author's original work. It is 282 pages longer (almost 70%), with 6 new chapters, for a total of 20. There are other notable revisions, including more examples and approximately 16 more pages (80%) of references. The depth and breadth of insight, information, and discussion, as well as graph types, continues to be remarkable. The text remains dense. It is not an easy task for a beginner to grasp all of the insight and details on the first pass, but sufficient sign posts are provided along the way for guidance. A more experienced reader will move through the text more deliberately and find delight in the details. Despite the potential for intimiding the reader by its encyclopedic coverage, this edition maintains the author's conversational tone, and his sense of humor expands into the new materials.

The second edition is divided into an introductory chapter (which is similar to that in the first edition), followed by two major parts. Part 1, "Syntax," focuses on the structure needed to create charts. It contains most of the material from the first edition, with some revisions and rearrangements. It comprises 11 chapters: "How to Make a Pie," "Data," "Variables," "Algebra," "Scales," "Statistics," "Geometry," "Coordinates," "Aesthetics," "Facets," and "Guides."

Part 2, "Semantics," focuses on content rather than structure. It examines the field of visualization, especially cognition and perception, with a strong emphasis on effective communication. It contains the 6 new chapters: "Space," "Time," "Uncertainty," "Analysis," "Control," and "Automation," as well as two chapters similar to those in the first edition, "Reader" and "Coda" (titled "Semantics" in first edition). Chapter 13, "Space," explores the role that spaces play in organizing graphs, specifically mapping of the underlying mathematical space into the display space. Chapter 14, "Time," discusses three models that predominate the measurement of time (deterministic, stochastic, and chaotic), the psychology of time, and the three types of time graphs (static, dynamic, and real-time). Chapter 15, "Uncertainty," reviews graphics that guide, qualify, or soften our judgments of uncertain data. Chapter 16, "Analysis," presents several analytical methods that depend on graphics for their usefulness, specifically variance analysis, shape analysis, sequence analysis, and pattern analysis. Chapter 17, "Control," describes two aspects of user control: methods for building charts and interactive methods for exploring data. Chapter 18, "Automation," describes two complementary methods for automating production graphs: graphics production language (GPL) and visualization markup language (VIZml).

The first edition was "highly recommended for anyone—statistician, engineer, or social scientist—with an interest in statistics" (Wegman 2000, p. 1009) and was written for "scientists working in the area of computer science, statistics, geography, and related disciplines." (Polasek 2002, p. 562). In addition, the author successfully targeted both undergraduate and graduate students. I prefer to extend Wilkinson's (1999, p. x) simple suggestion from the first edition. I find the second edition to be invaluable to anyone with an interest in communicating through any of a wide variety of either business or scientific graphics.

In summary, the second edition is an impressive expansion beyond a quite remarkable first edition. The text remains dense and even more encyclopedic,