Introduction to Data Science: Data Analysis and Prediction Algorithms with R introduces concepts and skills that can help you tackle real-world data analysis challenges. It covers concepts from probability, statistical inference, linear regression, and machine learning. It also helps you develop skills such as R programming, data wrangling, data visualization, predictive modeling, and big data. This book is a textbook for a first course in data science. No previous knowledge of R is necessary, although some experience with programming may be helpful. The book is divided into six parts: R, data visualization, statistics with R, data wrangling, machine learning, and productivity tools. Each part has several chapters meant to be presented as one lecture. The authors use motivating case studies that realistically mimic a data scientist’s experience. He starts by asking specific questions and answers these through data analysis so concepts are learned as a means to answering the questions. Examples of the case studies included are: US murder rates by state, self-reported student heights, trends in world health and economics, the impact of vaccines on infectious disease rates, the financial crisis of 2007-2008, election forecasting, building a baseball team, image processing of handwritten digits, and movie recommendation systems. The statistical concepts used to answer the case study questions are only briefly introduced, so complementing with a probability and statistics textbook is highly recommended for in-depth understanding of these concepts. If you read and understand the chapters and complete the exercises, you will be prepared to learn more advanced concepts and skills needed to become an expert.

This new edition of a standard reference on the use of statistical methods for categorical data has increased dramatically, particularly for applications in the biomedical and social sciences. An Introduction to Categorical Data Analysis, Third Edition summarizes these methods and shows readers how to use them using software. Readers will find a unified generalized linear models approach that connects logistic regression and loglinear models for discrete data with normal regression for continuous data. Adding to the value in the new edition is: illustrations of the use of R software to perform all the analyses in the book; a new chapter on alternative methods for categorical data, including smoothing and regularization methods (such as the lasso); classification methods such as linear discriminant analysis and classification trees; and classification + New sections in many chapters introducing the Bayesian approach for the methods of that chapter; > Than 70 analyses of data sets to illustrate application of the methods; and > About 200 exercises, many containing other data sets; > An appendix showing how to use SAS, Stata, and SPSS, and an appendix with short solutions to most exercises. The authors present analyses in an applied, nontechnical style, this book illustrates the methods using a wide variety of real data, including medical clinical trials, environmental questions, drug use by teenagers, horseshoe crab mating, basketball shooting, correlates of happiness, and much more. An Introduction to Categorical Data Analysis, Third Edition is an invaluable tool for statisticians and biostatisticians as well as methodologists, epidemiologists, medical researchers, and other scientists, educators, and students in health, medicine, and the behavioral and social sciences. Now in its third edition, this classic book is widely considered the leading text on Bayesian methods, lauded for its accessible, practical approach to analyzing data and solving research problems. Bayesian Data Analysis, Third Edition continues to take an applied approach to analysis using up-to-date Bayesian methods. The authors— all leaders in the statistics community—introduce basic concepts from a data-analytic perspective before presenting advanced methods. Throughout the text, numerous worked examples drawn from real applications and research emphasize the use of Bayesian inference in practice. New to the Third Edition Four new chapters on nonparametric modeling Coverage of weighted informative priors and boundary-avoiding priors Updated discussion of cross-validation and predictive information criteria Improved convergence monitoring and effective sample size calculations for iterative simulation Presentations of Hamiltonian Monte Carlo, variational Bayes, and expectation propagation New and revised software for doing Bayesian analyses Updated discussion of cross-validation and predictive information criteria Improved convergence monitoring and effective sample size calculation. These expansions of the book have been updated to provide a more cohesive discussion of Bayesian processing with examples and applications enabling the comprehension of alternative approaches to solving estimation/detection problems. The second edition of Bayesian Signal Processing features: “Classical” Kalman filtering for linear, linearized, and non-linear systems; “modern” uncentred and ensemble Kalman filters; and the “next-generation” Bayesian particle filters Sequential Bayesian detection techniques incorporating model-based schemes for a variety of real-world problems Practical Bayesian processor designs including comprehensive methods of performance analysis ranging from simple sanity testing and ensemble techniques to sophisticated information metrics New case studies on adaptive particle filtering and sequential Bayesian detection are covered detailing more Bayesian approaches to applied problem solving MATLAB® and R notes at the end of each chapter help readers solve complex problems using readily available software commands and point out other software packages available Problem sets included to test readers’ knowledge and help them put their new skills into practice Bayesian Signal Processing, Second Edition is an invaluable tool for statisticians and biostatisticians as well as methodologists, epidemiologists, medical researchers, and other scientists, educators, and students in health, medicine, and the behavioral and social sciences. This book outlines Bayesian statistical analysis in great detail, from the development of a model to the process of making statistical inference. The key feature of this book is that it covers models that are most commonly used in social science research — including the linear regression model, generalized linear models, hierarchical models, and multivariate regression models — and it thoroughly develops each real-data example in painstaking detail.

Using a practical, hands-on approach, this book will teach anyone how to carry out Bayesian analyses and interpret the results. Taken literally, the title “All of Statistics” is an exaggeration. But in spirit, the title is apt, as the book does cover a much broader range of topics than a typical introductory book on mathematical statistics. This book is for people who want to learn probability and statistics quickly. It is suitable for graduate or advanced undergraduate students in computer science, mathematics, statistics, and related disciplines. The book includes modern topics such as nonparametric curve estimation, bootstrapping, and classification, and topics that are usually relegated to follow-up courses. The reader is presumed to know calculus and a little linear algebra. No previous knowledge of probability and statistics is required. Statistics, data mining, and machine learning are all concerned with collecting and analysing data.

If you are interested in a comprehensive approach to Bayesian data analysis, this book is an excellent choice. It covers a wide range of topics in Bayesian statistics, including hierarchical modeling, smoothing, and predictive inference. The authors use real-world data sets to illustrate the methods, and the book contains many exercises and examples to help readers understand the concepts. Overall, this is a well-written and accessible book that is highly recommended for anyone interested in Bayesian data analysis.
such as chi-square and t-test. Proceeds to examine more advance methods, from regression and analysis of variance, through to generalized linear models, generalized mixed models, time series, spatial statistics, multivariate statistics and much more. The R Book is aimed at undergraduates, postgraduates and professional researchers, as well as anyone else who wishes to learn how to use R. The book is designed to help you get the most out of R, whether you are a complete novice or an experienced analyst.

The R Book will provide a comprehensive introduction to the R system and its most widely used packages. The book is structured into two parts: the first part introduces the R language, while the second part covers advanced topics. Each chapter contains practical examples and exercises, and the book concludes with a collection of useful scripts to help you get started with R.

The R Book provides a self-contained and comprehensive introduction to R. It is particularly useful for beginners, as well as for experienced users who want to brush up on their skills. The book is designed to be accessible, with a focus on practical examples and exercises. It is ideal for students, researchers, and professionals who want to learn how to use R effectively.

The R Book is an excellent resource for anyone who wants to learn how to use R. It is well-structured, clearly written, and contains many practical examples and exercises. It is suitable for beginners as well as experienced users who want to refresh their skills. The book is highly recommended for anyone who wants to learn how to use R effectively.
Core Statistics is a compact starter course on the theory, models, and computational tools needed to make informed use of powerful statistical methods. The Book of R is a comprehensive, beginner-friendly guide to R, the world’s most popular programming language for statistical analysis. Even if you have no programming experience and little more than a grounding in the basics of mathematics, you’ll find everything you need to begin using R effectively for statistical analysis. You’ll start with the basics, like how to handle data and write simple programs, before moving on to more advanced topics, like producing statistical summaries of your data and performing statistical tests and modeling. You’ll even learn how to create impressive data visualizations with R’s basic graphics tools and contributed packages, like ggplot2 and ggvis, as well as interactive 3D visualizations using the rgl package. Dozens of hands-on exercises (with downloadable solutions) take you from theory to practice, as you learn: - The fundamentals of program- ming in R, including how to write data frames, create functions, and use variables, statements, and loops. - Statistical concepts like exploratory data analysis, probability distributions, tests and regression modeling, and how to execute them in R. - How to access R’s thousands of functions, libraries, and data sets. - How to draw valid and useful conclusions from your data. - How to create publication-quality graphics of your results. Combining detailed explanations with real-world examples and exercises, this book will provide you with a solid understanding of both the theory and the depth of R’s functionality. Make The Book of R your doorway into the growing world of data analysis. An introduction to the Bayesian approach to statistical inference that demonstrates its superiority to orthodox frequentist statistical analysis. This book offers an introduction to the Bayesian approach to statistical inference, with a focus on nonparametric and distribution-free methods. It covers not only well-developed methods for doing Bayesian statistics but also novel tools that enable Bayesian statistical analyses for cases that previously did not have a full Bayesian solution. The book’s premise is that there are fundamental problems with orthodox frequentist statistical analyses that dis- tort the scientific process. Side-by-side comparisons of Bayesian and frequentist methods illustrate the mismatch between the needs of experimental scientists in making inferences from data and the properties of the standard tools of classical statistics. The book first covers elementary probability theory, the binomial model, and models for comparing different experimental conditions or groups. It then turns its focus to distribution-free statistics that are based on having ranked data, examining data from experimental studies and rank-based correlational methods. Each chapter includes exercises that help readers achieve a more complete understanding of the material. The book devotes considerable attention not only to the linkage of statistics to practices in experimental science but also to the theoretical foundations of statistics. Frequentist statistical practices often violate their own theoretical premises. The beauty of Bayesian statistics, readers will learn, is that it is an internally coherent system of scientific inference that can be proved from probability theory. This is the first text in a generation to re-examine the purpose of the mathematical statistics course. The book’s approach interweaves traditional topics with data analysis and reflects the use of the computer with close ties to the practice of statistics. The author stresses analysis of data, examines real problems with real data, and motivates the theory. The book’s descriptive statistics, graphical displays, and realistic applications stand in strong contrast to traditional texts that are set in abstract settings. Important Notice: Media content referenced within the product description or the product may not be available in the ebook version. Join the revolution ignited by the ground-breaking R system! Starting with an introduction to R, covering standard regression methods, then presenting more advanced topics, this book guides readers through the practical and powerful tools that R provides. The emphasis is on hands-on analysis, graphical display and interpretation of data. The many worked examples, taken from real-world research, are accompanied by commentary on what is done and why. A website provides computer code and data sets, allowing readers to reproduce all analyses. Updates and solutions to selected exercises are also available. Assuming only basic statistical knowledge, the book is ideal for research scientists, final-year undergraduate or graduate level students of applied statistics, and practising statisticians. It is both for learning and for reference. This revised edition reflects changes in R since 2013 and has new material on survival analysis, random coefficient models, and the handling of high-dimensional data. This integrated introduction to fundamentals, computation, and software is your key to understanding and using advanced Bayesian methods. Unlike many other books that emphasize the Bayesian framework, this textbook distinguishes between the philosophy and the computational methods of using Python: Tutorial guide that will take you through the journey of Bayesian analysis with the help of sample problems and practice exercises; Learn how and when to use Bayesian analysis in your applications with this guide. Who This Book Is For Students, researchers and data scientists who wish to learn about Bayesian analysis in R. The book is appropriate for R users with some programming experience. The book is accessible to advanced students and professionals who need to draw quantitative conclusions from experimental data, although most of the examples are taken from natural and social sciences, it prepares them for more advanced or specialized statistical modeling. This book provides a practical foundation for performing statistical inference. Designed for both PhD students and seasoned professionals in the natural and social sciences, it provides them for more advanced or specialized statistical modeling. Web Resource The book is accompanied by an R package (rethinking) that is available on the author’s website and GitHub. The two core functions (map and map2stan) of this package allow a variety of statistical models to be constructed from standard model formulas. This book is a guide to learning Bayesian statistics and probability through unusual and illustrative examples. Probability and statistics are increasingly important in a huge range of professions. But many people use data in ways they don’t even understand, meaning they aren’t getting the most from it. Bayesian Statistics the Fun Way will change that. This book will give you a complete understanding of Bayesian statistics through simple explanations and un-boring examples. Find out the probability of UFOs landing in your garden; how likely Han Solo is to survive a flight through an asteroid shower, how to win an argument about conspiracy theories, and whether a burglary really was a burglary, to name a few examples. By using these off-the- beaten-track examples, the author actually makes learning statistics fun. And you’ll learn real skills, like how to do - How to measure your own level of uncertainty in a conclusion or belief - Calculate Bayes theorem and understand what it’s useful for - Find the posterior, likelihood, and prior to check the accuracy of your calculations - Calculate distributions to see the range of your data - Compare hypotheses and draw reliable conclusions from them Next time you find yourself with a sheet of survey results and no idea what to do with them, turn to Bayesian Statistics the Fun Way to get the most value from your data.