

R In Action Data Analysis And Graphics With R

In this handy, practical book you will cover each concept concisely, with many illustrative examples. You'll be introduced to several R data science packages, with examples of how to use each of them. In this book, you'll learn about the following APIs and packages that deal specifically with data science applications: readr, dibble, forecasts, lubridate, stringr, tidyr, magrittr, dplyr, purrr, ggplot2, modelr, and more. After using this handy quick reference guide, you'll have the code, APIs, and insights to write data science-based applications in the R programming language. You'll also be able to carry out data analysis. What You Will Learn Import data with readr Work with categories using forcats, time and dates with lubridate, and strings with stringr Format data using tidyr and then transform that data using magrittr and dplyr Write functions with R for data science, data mining, and analytics-based applications Visualize data with ggplot2 and fit data to models using modelr Who This Book Is For Programmers new to R's data science, data mining, and analytics packages. Some prior coding experience with R in general is recommended.

Computational Genomics with R provides a starting point for beginners in genomic data analysis and also guides more advanced practitioners to sophisticated data analysis techniques in genomics. The book covers topics from R programming, to machine learning and statistics, to the latest genomic data analysis techniques. The text provides accessible information and explanations, always with the genomics context in the background. This also contains practical and well-documented examples in R so readers can analyze their data by simply reusing the code presented. As the field of computational genomics is interdisciplinary, it requires different starting points for people with different backgrounds. For example, a biologist might skip sections on basic genome biology and start with R programming, whereas a computer scientist might want to start with genome biology. After reading: You will have the basics of R and be able to dive right into specialized uses of R for computational genomics such as using Bioconductor packages. You will be familiar with statistics, supervised and unsupervised learning techniques that are important in data modeling, and exploratory analysis of high-dimensional data. You will understand genomic intervals and operations on them that are used for tasks such as aligned read counting and genomic feature annotation. You will know the basics of processing and quality checking high-throughput sequencing data. You will be able to do sequence analysis, such as calculating GC content for parts of a genome or finding transcription factor binding sites. You will know about visualization techniques used in genomics, such as heatmaps, meta-gene plots, and genomic track visualization. You will be familiar with analysis of different high-throughput sequencing data sets, such as RNA-seq, ChIP-seq, and BS-seq. You will know basic techniques for integrating and interpreting multi-omics datasets. Altuna Akalin is a group leader and head of the Bioinformatics and Omics Data Science Platform at the Berlin Institute of Medical Systems Biology, Max Delbrück Center, Berlin. He has been developing computational methods for analyzing and integrating large-scale genomics data sets since 2002. He has published an extensive body of work in this area. The framework for this book grew out of the yearly computational genomics courses he has been organizing and teaching since 2015.

This book addresses the difficulties experienced by wet lab researchers with the statistical analysis of molecular biology related data. The authors explain how to use R and Bioconductor for the analysis of experimental data in the field of molecular biology. The content is based upon two university courses for bioinformatics and experimental biology students (Biological Data Analysis with R and High-throughput Data Analysis with R). The material is divided into chapters based upon the experimental methods used in the laboratories. Key features include: • Broad appeal--the authors target their material to researchers in several levels, ensuring that the basics are always covered. • First book to explain how to use R and Bioconductor for the analysis of several types of experimental data in the field of molecular biology. • Focuses on R and Bioconductor, which are widely used for data analysis. One great benefit of R and Bioconductor is that there is a vast user community and very active discussion in place, in addition to the practice of sharing codes. Further, R is the platform for implementing new analysis approaches, therefore novel methods are available early for R users.

R for Data Analysis in easy steps is written using a proven easy-to-follow style for maximum appeal. It will be useful to anyone who wants to begin programming in R, with minimum fuss.

R is the world's most popular language for developing statistical software: Archaeologists use it to track the spread of ancient civilizations, drug companies use it to discover which medications are safe and effective, and actuaries use it to assess financial risks and keep economies running smoothly. The Art of R Programming takes you on a guided tour of software development with R, from basic types and data structures to advanced topics like closures, recursion, and anonymous functions. No statistical knowledge is required, and your programming skills can range from hobbyist to pro. Along the way, you'll learn about functional and object-oriented programming, running mathematical simulations, and rearranging complex data into simpler, more useful formats. You'll also learn to: –Create artful graphs to visualize complex data sets and functions –Write more efficient code using parallel R and vectorization –Interface R with C/C++ and Python for increased speed or functionality –Find new R packages for text analysis, image manipulation, and more –Squash annoying bugs with advanced debugging techniques Whether you're designing aircraft, forecasting the weather, or you just need to tame your data, The Art of R Programming is your guide to harnessing the power of statistical computing.

?The objective of this text is to introduce RStudio to practitioners and students and enable them to use R in their everyday work. It is not a statistical textbook, the purpose is to transmit the joy of analyzing data with RStudio.

Practitioners and students learn how RStudio can be installed and used, they learn to import data, write scripts and save working results. Furthermore, they learn to employ descriptive statistics and create graphics with RStudio. Additionally, it is shown how RStudio can be used to test hypotheses, run an analysis of variance and regressions. To deepen the learned content, tasks are included with the solutions provided at the end of the textbook. This textbook has been

recommended and developed for university courses in Germany, Austria and Switzerland.

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples

Learn how to write object-oriented programs in R and how to construct classes and class hierarchies in the three object-oriented systems available in R. This book gives an introduction to object-oriented programming in the R programming language and shows you how to use and apply R in an object-oriented manner. You will then be able to use this powerful programming style in your own statistical programming projects to write flexible and extendable software. After reading Advanced Object-Oriented Programming in R, you'll come away with a practical project that you can reuse in your own analytics coding endeavors. You'll then be able to visualize your data as objects that have state and then manipulate those objects with polymorphic or generic methods. Your projects will benefit from the high degree of flexibility provided by polymorphism, where the choice of concrete method to execute depends on the type of data being manipulated. What You'll Learn Define and use classes and generic functions using R Work with the R class hierarchies Benefit from implementation reuse Handle operator overloading Apply the S4 and R6 classes Who This Book Is For Experienced programmers and for those with at least some prior experience with R programming language. /div

This invaluable addition to any data scientist's library shows you how to apply the R programming language and useful statistical techniques to everyday business situations as well as how to effectively present results to audiences of all levels. To answer the ever-increasing demand for machine learning and analysis, this new edition boasts additional R tools, modeling techniques, and more. Practical Data Science with R, Second Edition takes a practice-oriented approach to explaining basic principles in the ever-expanding field of data science. You'll jump right to real-world use cases as you apply the R programming language and statistical analysis techniques to carefully explained examples based in marketing, business intelligence, and decision support. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

Longitudinal data analysis for biomedical and behavioral sciences This innovative book sets forth and describes methods for the analysis of longitudinal data, emphasizing applications to problems in the biomedical and behavioral sciences. Reflecting the growing importance and use of longitudinal data across many areas of research, the text is designed to help users of statistics better analyze and understand this type of data. Much of the material from the book grew out of a course taught by Dr. Hedeker on longitudinal data analysis. The material is, therefore, thoroughly classroom tested and includes a number of features designed to help readers better understand and apply the material. Statistical procedures featured within the text include: * Repeated measures analysis of variance * Multivariate analysis of variance for repeated measures * Random-effects regression models (RRM) * Covariance-pattern models * Generalized-estimating equations (GEE) models * Generalizations of RRM and GEE for categorical outcomes Practical in their approach, the authors emphasize the applications of the methods, using real-world examples for illustration. Some syntax examples are provided, although the authors do not generally focus on software in this book. Several datasets and computer syntax examples are posted on this title's companion Web site. The authors intend to keep the syntax examples current as new versions of the software programs emerge. This text is designed for both undergraduate and graduate courses in longitudinal data analysis. Instructors can take advantage of overheads and additional course materials available online for adopters. Applied statisticians in biomedicine and the social sciences can also use the book as a convenient reference.

See How Graphics Reveal Information Graphical Data Analysis with R shows you what information you can gain from graphical displays. The book focuses on why you draw graphics to display data and which graphics to draw (and uses R to do so). All the datasets are available in R or one of its packages and the R code is available at rosuda.org/GDA. Graphical data analysis is useful for data cleaning, exploring data structure, detecting outliers and unusual groups, identifying trends and clusters, spotting local patterns, evaluating modelling output, and presenting results. This book guides you in choosing graphics and understanding what information you can glean from them. It can be used as a primary text in a graphical data analysis course or as a supplement in a statistics course. Colour graphics are used throughout.

This guide for practicing statisticians, data scientists, and R users and programmers will teach the essentials of preprocessing: data leveraging the R programming language to easily and quickly turn noisy data into usable pieces of information. Data wrangling, which is also commonly referred to as data munging, transformation, manipulation, janitor work, etc., can be a painstakingly laborious process. Roughly 80% of data analysis is spent on cleaning and preparing data; however, being a prerequisite to the rest of the data analysis workflow (visualization, analysis, reporting), it is essential that one become fluent and efficient in data wrangling techniques. This book will guide the user through the data wrangling process via a step-by-step tutorial approach and provide a solid foundation for working with data in R. The author's goal is to teach the user how to easily wrangle data in order to spend more time on understanding the content of the data. By the end of the book, the user will have learned: How to work with different types of data such as numerics, characters, regular expressions, factors, and dates The difference between different data structures and how to create, add additional components to, and subset each data structure How to acquire and parse data from locations previously inaccessible How to develop functions and use loop control structures to reduce code redundancy How to use pipe operators to simplify code and make it more readable How to reshape the layout of data and manipulate, summarize, and join data sets

Summary R in Action, Second Edition presents both the R language and the examples that make it so useful for business developers. Focusing on practical solutions, the book offers a crash course in statistics and covers elegant methods for dealing with messy and incomplete data that are difficult to analyze using traditional methods. You'll also master R's extensive graphical capabilities for exploring and presenting data visually. And this expanded second edition includes new chapters on time series analysis, cluster analysis, and classification methodologies, including decision trees, random forests, and support vector machines. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Business pros and researchers thrive on data, and R speaks the language of data analysis. R is a powerful programming language for statistical computing. Unlike general-purpose tools, R provides thousands of modules for solving just about any data-crunching or presentation challenge you're likely to face. R runs on all

important platforms and is used by thousands of major corporations and institutions worldwide. About the Book R in Action, Second Edition teaches you how to use the R language by presenting examples relevant to scientific, technical, and business developers. Focusing on practical solutions, the book offers a crash course in statistics, including elegant methods for dealing with messy and incomplete data. You'll also master R's extensive graphical capabilities for exploring and presenting data visually. And this expanded second edition includes new chapters on forecasting, data mining, and dynamic report writing. What's Inside Complete R language tutorial Using R to manage, analyze, and visualize data Techniques for debugging programs and creating packages OOP in R Over 160 graphs About the Author Dr. Rob Kabacoff is a seasoned researcher and teacher who specializes in data analysis. He also maintains the popular Quick-R website at statmethods.net. Table of Contents PART 1 GETTING STARTED Introduction to R Creating a dataset Getting started with graphs Basic data management Advanced data management PART 2 BASIC METHODS Basic graphs Basic statistics PART 3 INTERMEDIATE METHODS Regression Analysis of variance Power analysis Intermediate graphs Resampling statistics and bootstrapping PART 4 ADVANCED METHODS Generalized linear models Principal components and factor analysis Time series Cluster analysis Classification Advanced methods for missing data PART 5 EXPANDING YOUR SKILLS Advanced graphics with ggplot2 Advanced programming Creating a package Creating dynamic reports Advanced graphics with the lattice package available online only from manning.com/kabacoff2

Improve Your Analytical Skills Incorporating the latest R packages as well as new case studies and applications, Using R and RStudio for Data Management, Statistical Analysis, and Graphics, Second Edition covers the aspects of R most often used by statistical analysts. New users of R will find the book's simple approach easy to understand while more

This introductory statistics textbook conveys the essential concepts and tools needed to develop and nurture statistical thinking. It presents descriptive, inductive and explorative statistical methods and guides the reader through the process of quantitative data analysis. In the experimental sciences and interdisciplinary research, data analysis has become an integral part of any scientific study. Issues such as judging the credibility of data, analyzing the data, evaluating the reliability of the obtained results and finally drawing the correct and appropriate conclusions from the results are vital. The text is primarily intended for undergraduate students in disciplines like business administration, the social sciences, medicine, politics, macroeconomics, etc. It features a wealth of examples, exercises and solutions with computer code in the statistical programming language R as well as supplementary material that will enable the reader to quickly adapt all methods to their own applications.

From a review of the first edition: "Modern Data Science with R... is rich with examples and is guided by a strong narrative voice. What's more, it presents an organizing framework that makes a convincing argument that data science is a course distinct from applied statistics" (The American Statistician). Modern Data Science with R is a comprehensive data science textbook for undergraduates that incorporates statistical and computational thinking to solve real-world data problems. Rather than focus exclusively on case studies or programming syntax, this book illustrates how statistical programming in the state-of-the-art R/RStudio computing environment can be leveraged to extract meaningful information from a variety of data in the service of addressing compelling questions. The second edition is updated to reflect the growing influence of the tidyverse set of packages. All code in the book has been revised and styled to be more readable and easier to understand. New functionality from packages like `sf`, `purrr`, `tidymodels`, and `tidytext` is now integrated into the text. All chapters have been revised, and several have been split, re-organized, or re-imagined to meet the shifting landscape of best practice.

The ability to interpret and act on the massive amounts of information locked in web and enterprise systems is critical to success in the modern business economy. R, a free software environment for statistical computing and graphics, is a comprehensive package that empowers developers and analysts to capture, process, and respond intelligently to statistical information. R in Action is the first book to present both the R system and the use cases that make it such a compelling package for business developers. The book begins by introducing the R language, and then moves on to various examples illustrating R's features. Coverage includes data mining methodologies, approaches to messy data, R's extensive graphical environment, useful add-on modules, and how to interface R with other software platforms and data management systems.

Summary Gnuplot in Action, Second Edition is a major revision of this popular and authoritative guide for developers, engineers, and scientists who want to learn and use gnuplot effectively. Fully updated for gnuplot version 5, the book includes four pages of color illustrations and four bonus appendixes available in the eBook. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Gnuplot is an open-source graphics program that helps you analyze, interpret, and present numerical data. Available for Unix, Mac, and Windows, it is well-maintained, mature, and totally free. About the Book Gnuplot in Action, Second Edition is a major revision of this authoritative guide for developers, engineers, and scientists. The book starts with a tutorial introduction, followed by a systematic overview of gnuplot's core features and full coverage of gnuplot's advanced capabilities. Experienced readers will appreciate the discussion of gnuplot 5's features, including new plot types, improved text and color handling, and support for interactive, web-based display formats. The book concludes with chapters on graphical effects and general techniques for understanding data with graphs. It includes four pages of color illustrations. 3D graphics, false-color plots, heatmaps, and multivariate visualizations are covered in chapter-length appendixes available in the eBook. What's Inside Creating different types of graphs in detail Animations, scripting, batch operations Extensive discussion of terminals Updated to cover gnuplot version 5 About the Reader No prior experience with gnuplot is required. This book concentrates on practical applications of gnuplot relevant to users of all levels. About the Author Philipp K. Janert, PhD, is a programmer and scientist. He is the author of several books on data analysis and applied math and has been a gnuplot power user and developer for over 20 years. Table of Contents PART 1 GETTING STARTED Prelude: understanding data with gnuplot Tutorial: essential gnuplot The heart of the matter: the plot command PART 2 CREATING GRAPHS Managing data sets and files Practical matters: strings, loops, and history A catalog of styles Decorations: labels, arrows, and explanations All about axes PART 3 MASTERING TECHNICALITIES Color, style, and appearance Terminals and output formats Automation, scripting, and animation Beyond the defaults: workflow and styles PART 4 UNDERSTANDING DATA Basic techniques of graphical analysis Topics in graphical analysis Coda: understanding data with graphs

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience"--

Learn how to perform data analysis with the R language and software environment, even if you have little or no programming experience. With the tutorials in this hands-on guide, you'll learn how to use the essential R tools you need to know to analyze data, including data types and programming concepts. The second half of Learning R shows you real data analysis in action by covering everything from importing data to publishing your results. Each chapter in the book includes a quiz on what you've learned, and concludes with exercises, most of which involve writing R code. Write a simple R program, and discover what the language can do Use data types such as vectors, arrays, lists, data frames, and strings Execute code conditionally or repeatedly with branches and loops Apply R add-on packages, and package your own work for others Learn how to clean data you import

from a variety of sources Understand data through visualization and summary statistics Use statistical models to pass quantitative judgments about data and make predictions Learn what to do when things go wrong while writing data analysis code Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and STAN examines the Bayesian and frequentist methods of conducting data analyses. The book provides the theoretical background in an easy-to-understand approach, encouraging readers to examine the processes that generated their data. Including discussions of model selection, model checking, and multi-model inference, the book also uses effect plots that allow a natural interpretation of data. Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and STAN introduces Bayesian software, using R for the simple modes, and flexible Bayesian software (BUGS and Stan) for the more complicated ones. Guiding the reader from easy toward more complex (real) data analyses in a step-by-step manner, the book presents problems and solutions—including all R codes—that are most often applicable to other data and questions, making it an invaluable resource for analyzing a variety of data types. Introduces Bayesian data analysis, allowing users to obtain uncertainty measurements easily for any derived parameter of interest Written in a step-by-step approach that allows for eased understanding by non-statisticians Includes a companion website containing R-code to help users conduct Bayesian data analyses on their own data All example data as well as additional functions are provided in the R-package blmeco

Integrates the theory and applications of statistics using R A Course in Statistics with R has been written to bridge the gap between theory and applications and explain how mathematical expressions are converted into R programs. The book has been primarily designed as a useful companion for a Masters student during each semester of the course, but will also help applied statisticians in revisiting the underpinnings of the subject. With this dual goal in mind, the book begins with R basics and quickly covers visualization and exploratory analysis. Probability and statistical inference, inclusive of classical, nonparametric, and Bayesian schools, is developed with definitions, motivations, mathematical expression and R programs in a way which will help the reader to understand the mathematical development as well as R implementation. Linear regression models, experimental designs, multivariate analysis, and categorical data analysis are treated in a way which makes effective use of visualization techniques and the related statistical techniques underlying them through practical applications, and hence helps the reader to achieve a clear understanding of the associated statistical models. Key features: Integrates R basics with statistical concepts Provides graphical presentations inclusive of mathematical expressions Aids understanding of limit theorems of probability with and without the simulation approach Presents detailed algorithmic development of statistical models from scratch Includes practical applications with over 50 data sets

AN UP-TO-DATE, COMPREHENSIVE TREATMENT OF A CLASSIC TEXT ON MISSING DATA IN STATISTICS The topic of missing data has gained considerable attention in recent decades. This new edition by two acknowledged experts on the subject offers an up-to-date account of practical methodology for handling missing data problems. Blending theory and application, authors Roderick Little and Donald Rubin review historical approaches to the subject and describe simple methods for multivariate analysis with missing values. They then provide a coherent theory for analysis of problems based on likelihoods derived from statistical models for the data and the missing data mechanism, and then they apply the theory to a wide range of important missing data problems. Statistical Analysis with Missing Data, Third Edition starts by introducing readers to the subject and approaches toward solving it. It looks at the patterns and mechanisms that create the missing data, as well as a taxonomy of missing data. It then goes on to examine missing data in experiments, before discussing complete-case and available-case analysis, including weighting methods. The new edition expands its coverage to include recent work on topics such as nonresponse in sample surveys, causal inference, diagnostic methods, and sensitivity analysis, among a host of other topics. An updated "classic" written by renowned authorities on the subject Features over 150 exercises (including many new ones) Covers recent work on important methods like multiple imputation, robust alternatives to weighting, and Bayesian methods Revises previous topics based on past student feedback and class experience Contains an updated and expanded bibliography Statistical Analysis with Missing Data, Third Edition is an ideal textbook for upper undergraduate and/or beginning graduate level students of the subject. It is also an excellent source of information for applied statisticians and practitioners in government and industry.

An Up-to-Date, All-in-One Resource for Using SAS and R to Perform Frequent Tasks The first edition of this popular guide provided a path between SAS and R using an easy-to-understand, dictionary-like approach. Retaining the same accessible format, SAS and R: Data Management, Statistical Analysis, and Graphics, Second Edition explains how to easily perform an analytical task in both SAS and R, without having to navigate through the extensive, idiosyncratic, and sometimes unwieldy software documentation. The book covers many common tasks, such as data management, descriptive summaries, inferential procedures, regression analysis, and graphics, along with more complex applications. New to the Second Edition This edition now covers RStudio, a powerful and easy-to-use interface for R. It incorporates a number of additional topics, including using application program interfaces (APIs), accessing data through database management systems, using reproducible analysis tools, and statistical analysis with Markov chain Monte Carlo (MCMC) methods and finite mixture models. It also includes extended examples of simulations and many new examples. Enables Easy Mobility between the Two Systems Through the extensive indexing and cross-referencing, users can directly find and implement the material they need. SAS users can look up tasks in the SAS index and then find the associated R code while R users can benefit from the R index in a similar manner. Numerous example analyses demonstrate the code in action and facilitate further exploration. The datasets and code are available for download on the book's website.

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 users through the practical and powerful tools that the R system 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 2003 and has new material on survival analysis, random coefficient models, and the handling of high-dimensional data.

Statistical Computation for Programmers, Scientists, Quants, Excel Users, and Other Professionals Using the open

source R language, you can build powerful statistical models to answer many of your most challenging questions. R has traditionally been difficult for non-statisticians to learn, and most R books assume far too much knowledge to be of help. R for Everyone is the solution. Drawing on his unsurpassed experience teaching new users, professional data scientist Jared P. Lander has written the perfect tutorial for anyone new to statistical programming and modeling. Organized to make learning easy and intuitive, this guide focuses on the 20 percent of R functionality you'll need to accomplish 80 percent of modern data tasks. Lander's self-contained chapters start with the absolute basics, offering extensive hands-on practice and sample code. You'll download and install R; navigate and use the R environment; master basic program control, data import, and manipulation; and walk through several essential tests. Then, building on this foundation, you'll construct several complete models, both linear and nonlinear, and use some data mining techniques. By the time you're done, you won't just know how to write R programs, you'll be ready to tackle the statistical problems you care about most.

COVERAGE INCLUDES • Exploring R, RStudio, and R packages • Using R for math: variable types, vectors, calling functions, and more • Exploiting data structures, including data.frames, matrices, and lists • Creating attractive, intuitive statistical graphics • Writing user-defined functions • Controlling program flow with if, ifelse, and complex checks • Improving program efficiency with group manipulations • Combining and reshaping multiple datasets • Manipulating strings using R's facilities and regular expressions • Creating normal, binomial, and Poisson probability distributions • Programming basic statistics: mean, standard deviation, and t-tests • Building linear, generalized linear, and nonlinear models • Assessing the quality of models and variable selection • Preventing overfitting, using the Elastic Net and Bayesian methods • Analyzing univariate and multivariate time series data • Grouping data via K-means and hierarchical clustering • Preparing reports, slideshows, and web pages with knitr • Building reusable R packages with devtools and Rcpp • Getting involved with the R global community

With more than 200 practical recipes, this book helps you perform data analysis with R quickly and efficiently. The R language provides everything you need to do statistical work, but its structure can be difficult to master. This collection of concise, task-oriented recipes makes you productive with R immediately, with solutions ranging from basic tasks to input and output, general statistics, graphics, and linear regression. Each recipe addresses a specific problem, with a discussion that explains the solution and offers insight into how it works. If you're a beginner, R Cookbook will help get you started. If you're an experienced data programmer, it will jog your memory and expand your horizons. You'll get the job done faster and learn more about R in the process. Create vectors, handle variables, and perform other basic functions Input and output data Tackle data structures such as matrices, lists, factors, and data frames Work with probability, probability distributions, and random variables Calculate statistics and confidence intervals, and perform statistical tests Create a variety of graphic displays Build statistical models with linear regressions and analysis of variance (ANOVA) Explore advanced statistical techniques, such as finding clusters in your data "Wonderfully readable, R Cookbook serves not only as a solutions manual of sorts, but as a truly enjoyable way to explore the R language—one practical example at a time."—Jeffrey Ryan, software consultant and R package author

R is a powerful language for statistical computing and graphics that can handle virtually any data-crunching task. It runs on all important platforms and provides thousands of useful specialized modules and utilities. This makes R a great way to get meaningful information from mountains of raw data. R in Action, Second Edition is a language tutorial focused on practical problems. Written by a research methodologist, it takes a direct and modular approach to quickly give readers the information they need to produce useful results. Focusing on realistic data analyses and a comprehensive integration of graphics, it follows the steps that real data analysts use to acquire their data, get it into shape, analyze it, and produce meaningful results that they can provide to clients. Purchase of the print book comes with an offer of a free PDF eBook from Manning. Also available is all code from the book.

This thoroughly updated new edition of the bestselling text trains students to critically read a research article from start to finish. Containing twenty-five engaging articles of both ideal and flawed research, it discusses how to assess them in terms of soundness of the design and appropriateness of the statistical analyses.

Utilize R to uncover hidden patterns in your Big Data About This Book Perform computational analyses on Big Data to generate meaningful results Get a practical knowledge of R programming language while working on Big Data platforms like Hadoop, Spark, H2O and SQL/NoSQL databases, Explore fast, streaming, and scalable data analysis with the most cutting-edge technologies in the market Who This Book Is For This book is intended for Data Analysts, Scientists, Data Engineers, Statisticians, Researchers, who want to integrate R with their current or future Big Data workflows. It is assumed that readers have some experience in data analysis and understanding of data management and algorithmic processing of large quantities of data, however they may lack specific skills related to R. What You Will Learn Learn about current state of Big Data processing using R programming language and its powerful statistical capabilities Deploy Big Data analytics platforms with selected Big Data tools supported by R in a cost-effective and time-saving manner Apply the R language to real-world Big Data problems on a multi-node Hadoop cluster, e.g. electricity consumption across various socio-demographic indicators and bike share scheme usage Explore the compatibility of R with Hadoop, Spark, SQL and NoSQL databases, and H2O platform In Detail Big Data analytics is the process of examining large and complex data sets that often exceed the computational capabilities. R is a leading programming language of data science, consisting of powerful functions to tackle all problems related to Big Data processing. The book will begin with a brief introduction to the Big Data world and its current industry standards. With introduction to the R language and presenting its development, structure, applications in real world, and its shortcomings. Book will progress towards revision of major R functions for data management and transformations. Readers will be introduced to Cloud based Big Data solutions (e.g. Amazon EC2 instances and Amazon RDS, Microsoft Azure and its HDInsight clusters) and also provide guidance on R connectivity with relational and non-relational databases such as MongoDB and HBase etc. It will

further expand to include Big Data tools such as Apache Hadoop ecosystem, HDFS and MapReduce frameworks. Also other R compatible tools such as Apache Spark, its machine learning library Spark MLlib, as well as H2O. Style and approach This book will serve as a practical guide to tackling Big Data problems using R programming language and its statistical environment. Each section of the book will present you with concise and easy-to-follow steps on how to process, transform and analyse large data sets.

This book provides an elementary-level introduction to R, targeting both non-statistician scientists in various fields and students of statistics. The main mode of presentation is via code examples with liberal commenting of the code and the output, from the computational as well as the statistical viewpoint. Brief sections introduce the statistical methods before they are used. A supplementary R package can be downloaded and contains the data sets. All examples are directly runnable and all graphics in the text are generated from the examples. The statistical methodology covered includes statistical standard distributions, one- and two-sample tests with continuous data, regression analysis, one-and two-way analysis of variance, regression analysis, analysis of tabular data, and sample size calculations. In addition, the last four chapters contain introductions to multiple linear regression analysis, linear models in general, logistic regression, and survival analysis.

The majority of data sets collected by researchers in all disciplines are multivariate, meaning that several measurements, observations, or recordings are taken on each of the units in the data set. These units might be human subjects, archaeological artifacts, countries, or a vast variety of other things. In a few cases, it may be sensible to isolate each variable and study it separately, but in most instances all the variables need to be examined simultaneously in order to fully grasp the structure and key features of the data. For this purpose, one or another method of multivariate analysis might be helpful, and it is with such methods that this book is largely concerned. Multivariate analysis includes methods both for describing and exploring such data and for making formal inferences about them. The aim of all the techniques is, in general sense, to display or extract the signal in the data in the presence of noise and to find out what the data show us in the midst of their apparent chaos. An Introduction to Applied Multivariate Analysis with R explores the correct application of these methods so as to extract as much information as possible from the data at hand, particularly as some type of graphical representation, via the R software. Throughout the book, the authors give many examples of R code used to apply the multivariate techniques to multivariate data.

The high-level language of R is recognized as one of the most powerful and flexible statistical software environments, and is rapidly becoming the standard setting for quantitative analysis, statistics and graphics. R provides free access to unrivalled coverage and cutting-edge applications, enabling the user to apply numerous statistical methods ranging from simple regression to time series or multivariate analysis. Building on the success of the author's bestselling Statistics: An Introduction using R, The R Book is packed with worked examples, providing an all inclusive guide to R, ideal for novice and more accomplished users alike. The book assumes no background in statistics or computing and introduces the advantages of the R environment, detailing its applications in a wide range of disciplines. Provides the first comprehensive reference manual for the R language, including practical guidance and full coverage of the graphics facilities. Introduces all the statistical models covered by R, beginning with simple classical tests 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 professionals in science, engineering and medicine. It is also ideal for students and professionals in statistics, economics, geography and the social sciences.

This textbook offers an easy-to-follow, practical guide to modern data analysis using the programming language R. The chapters cover topics such as the fundamentals of programming in R, data collection and preprocessing, including web scraping, data visualization, and statistical methods, including multivariate analysis, and feature exercises at the end of each section. The text requires only basic statistics skills, as it strikes a balance between statistical and mathematical understanding and implementation in R, with a special emphasis on reproducible examples and real-world applications. This textbook is primarily intended for undergraduate students of mathematics, statistics, physics, economics, finance and business who are pursuing a career in data analytics. It will be equally valuable for master students of data science and industry professionals who want to conduct data analyses.

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 programming in R, including how to write data frames, create functions, and use variables, statements, and loops –Statistical concepts like exploratory data analysis, probabilities, hypothesis 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 statistics and the depth of R's functionality. Make The Book of R your doorway into the growing world of data analysis.

Newcomers to R are often intimidated by the command-line interface, the vast number of functions and packages, or the processes of importing data and performing a simple statistical analysis. The R Primer provides a collection of concise

examples and solutions to R problems frequently encountered by new users of this statistical software. This new edition adds coverage of R Studio and reproducible research.

There is an explosion of interest in Bayesian statistics, primarily because recently created computational methods have finally made Bayesian analysis tractable and accessible to a wide audience. Doing Bayesian Data Analysis, A Tutorial Introduction with R and BUGS, is for first year graduate students or advanced undergraduates and provides an accessible approach, as all mathematics is explained intuitively and with concrete examples. It assumes only algebra and 'rusty' calculus. Unlike other textbooks, this book begins with the basics, including essential concepts of probability and random sampling. The book gradually climbs all the way to advanced hierarchical modeling methods for realistic data. The text provides complete examples with the R programming language and BUGS software (both freeware), and begins with basic programming examples, working up gradually to complete programs for complex analyses and presentation graphics. These templates can be easily adapted for a large variety of students and their own research needs. The textbook bridges the students from their undergraduate training into modern Bayesian methods. Accessible, including the basics of essential concepts of probability and random sampling Examples with R programming language and BUGS software Comprehensive coverage of all scenarios addressed by non-bayesian textbooks- t-tests, analysis of variance (ANOVA) and comparisons in ANOVA, multiple regression, and chi-square (contingency table analysis). Coverage of experiment planning R and BUGS computer programming code on website Exercises have explicit purposes and guidelines for accomplishment

R in Action, Third Edition teaches you to use the R language, including the popular tidyverse packages, through hands-on examples relevant to scientific, technical, and business developers. Built specifically for statistical computing and graphics, the R language, along with its amazing collection of libraries and tools, is one of the most powerful tools you can use to tackle data analysis for business, research, and other data-intensive domains. This revised and expanded third edition of R in Action covers the new tidyverse approach to data analysis and R's state-of-the-art graphing capabilities with the ggplot2 package. R in Action, Third Edition teaches you to use the R language, including the popular tidyverse packages, through hands-on examples relevant to scientific, technical, and business developers. Focusing on practical solutions to real-world data challenges, R expert Rob Kabacoff takes you on a crash course in statistics, from dealing with messy and incomplete data to creating stunning visualisations. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

Perform data analysis with R quickly and efficiently with more than 275 practical recipes in this expanded second edition. The R language provides everything you need to do statistical work, but its structure can be difficult to master. These task-oriented recipes make you productive with R immediately. Solutions range from basic tasks to input and output, general statistics, graphics, and linear regression. Each recipe addresses a specific problem and includes a discussion that explains the solution and provides insight into how it works. If you're a beginner, R Cookbook will help get you started. If you're an intermediate user, this book will jog your memory and expand your horizons. You'll get the job done faster and learn more about R in the process. Create vectors, handle variables, and perform basic functions Simplify data input and output Tackle data structures such as matrices, lists, factors, and data frames Work with probability, probability distributions, and random variables Calculate statistics and confidence intervals and perform statistical tests Create a variety of graphic displays Build statistical models with linear regressions and analysis of variance (ANOVA) Explore advanced statistical techniques, such as finding clusters in your data

Summary Beyond Spreadsheets with R shows you how to take raw data and transform it for use in computations, tables, graphs, and more. You'll build on simple programming techniques like loops and conditionals to create your own custom functions. You'll come away with a toolkit of strategies for analyzing and visualizing data of all sorts using R and RStudio. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Spreadsheets are powerful tools for many tasks, but if you need to interpret, interrogate, and present data, they can feel like the wrong tools for the task. That's when R programming is the way to go. The R programming language provides a comfortable environment to properly handle all types of data. And within the open source RStudio development suite, you have at your fingertips easy-to-use ways to simplify complex manipulations and create reproducible processes for analysis and reporting. About the Book With Beyond Spreadsheets with R you'll learn how to go from raw data to meaningful insights using R and RStudio. Each carefully crafted chapter covers a unique way to wrangle data, from understanding individual values to interacting with complex collections of data, including data you scrape from the web. You'll build on simple programming techniques like loops and conditionals to create your own custom functions. You'll come away with a toolkit of strategies for analyzing and visualizing data of all sorts. What's inside How to start programming with R and RStudio Understanding and implementing important R structures and operators Installing and working with R packages Tidying, refining, and plotting your data About the Reader If you're comfortable writing formulas in Excel, you're ready for this book. About the Author Dr Jonathan Carroll is a data science consultant providing R programming services. He holds a PhD in theoretical physics. Table of Contents Introducing data and the R language Getting to know R data types Making new data values Understanding the tools you'll use: Functions Combining data values Selecting data values Doing things with lots of data Doing things conditionally: Control structures Visualizing data: Plotting Doing more with your data with extensions

[Copyright: 4031b710f8d79a0c04360a2f6f41c197](#)