

## The R Book

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. 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 Turn your R code into packages that others can easily download and use. This practical book shows you how to bundle reusable R functions, sample data, and documentation together by applying author Hadley Wickham's package development philosophy. In the process, you'll work with devtools, roxygen, and testthat, a set of R packages that automate common development tasks. Devtools encapsulates best practices that Hadley has learned from years of working with this programming language. Ideal for developers, data scientists, and programmers with various backgrounds, this book starts you with the basics

and shows you how to improve your package writing over time. You'll learn to focus on what you want your package to do, rather than think about package structure. Learn about the most useful components of an R package, including vignettes and unit tests Automate anything you can, taking advantage of the years of development experience embodied in devtools Get tips on good style, such as organizing functions into files Streamline your development process with devtools Learn the best way to submit your package to the Comprehensive R Archive Network (CRAN) Learn from a well-respected member of the R community who created 30 R packages, including ggplot2, dplyr, and tidyr Forecasting is required in many situations. Stocking an inventory may require forecasts of demand months in advance. Telecommunication routing requires traffic forecasts a few minutes ahead. Whatever the circumstances or time horizons involved, forecasting is an important aid in effective and efficient planning. This textbook provides a comprehensive introduction to forecasting methods and presents enough information about each method for readers to use them sensibly.

R is the amazing, free, open-access software package for scientific graphs and calculations used by scientists worldwide. The R Student Companion is a student-oriented manual describing how to use R in high school and college science and mathematics courses. Written for beginners in scientific computation, the book assumes the reader has just some

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

R, an Open Source software, has become the de facto statistical computing

environment. It has an excellent collection of data manipulation and graphics capabilities. It is extensible and comes with a large number of packages that allow statistical analysis at all levels – from simple to advanced – and in numerous fields including Medicine, Genetics, Biology, Environmental Sciences, Geology, Social Sciences and much more. The software is maintained and developed by academicians and professionals and as such, is continuously evolving and up to date. *Statistics and Data with R* presents an accessible guide to data manipulations, statistical analysis and graphics using R. Assuming no previous knowledge of statistics or R, the book includes: A comprehensive introduction to the R language. An integrated approach to importing and preparing data for analysis, exploring and analyzing the data, and presenting results. Over 300 examples, including detailed explanations of the R scripts used throughout. Over 100 moderately large data sets from disciplines ranging from Biology, Ecology and Environmental Science to Medicine, Law, Military and Social Sciences. A parallel discussion of analyses with the normal density, proportions (binomial), counts (Poisson) and bootstrap methods. Two extensive indexes that include references to every R function (and its arguments and packages used in the book) and to every introduced concept.

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 *Advanced R* helps you understand how R works at a fundamental level. It is designed for R programmers who want to deepen their understanding of the language, and programmers experienced in other languages who want to understand what makes R different and special. This book will teach you the foundations of R; three fundamental programming paradigms (functional, object-oriented, and metaprogramming); and powerful techniques for debugging and optimising your code. By reading this book, you will learn: The difference between an object and its name, and why the distinction is important The important vector data structures,

how they fit together, and how you can pull them apart using subsetting The fine details of functions and environments The condition system, which powers messages, warnings, and errors The powerful functional programming paradigm, which can replace many for loops The three most important OO systems: S3, S4, and R6 The tidy eval toolkit for metaprogramming, which allows you to manipulate code and control evaluation Effective debugging techniques that you can deploy, regardless of how your code is run How to find and remove performance bottlenecks The second edition is a comprehensive update: New foundational chapters: "Names and values," "Control flow," and "Conditions" comprehensive coverage of object oriented programming with chapters on S3, S4, R6, and how to choose between them Much deeper coverage of metaprogramming, including the new tidy evaluation framework use of new package like rlang (<http://rlang.r-lib.org>), which provides a clean interface to low-level operations, and purrr (<http://purrr.tidyverse.org/>) for functional programming Use of color in code chunks and figures Hadley Wickham is Chief Scientist at RStudio, an Adjunct Professor at Stanford University and the University of Auckland, and a member of the R Foundation. He is the lead developer of the tidyverse, a collection of R packages, including ggplot2 and dplyr, designed to support data science. He is also the author of R for Data Science (with Garrett Grolemund), R Packages, and ggplot2: Elegant Graphics for Data Analysis.

An essential library of basic commands you can copy and paste into R The powerful and open-source statistical programming language R is rapidly growing in popularity, but it requires that you type in commands at the keyboard rather than use a mouse, so you have to learn the language of R. But there is a shortcut, and that's where this unique book comes in. A companion book to Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics, this practical reference is a library of basic R commands that you can copy and paste into R to perform many types of statistical analyses. Whether you're in technology, science, medicine, business, or engineering, you can quickly turn to your topic in this handy book and find the commands you need. Comprehensive command reference for the R programming language and a companion book to Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics Combines elements of a dictionary, glossary, and thesaurus for the R language Provides easy accessibility to the commands you need, by topic, which you can cut and paste into R as needed Covers getting, saving, examining, and manipulating data; statistical test and math; and all the things you can do with graphs Also includes a collection of utilities that you'll find useful Simplify the complex statistical R programming language with The Essential R Reference. .

Geocomputation with R is for people who want to analyze, visualize and model geographic data with open source software. It is based on R, a statistical programming language that has powerful data processing, visualization, and geospatial capabilities. The book equips you with the knowledge and skills to tackle a wide range of issues manifested in geographic data, including those with scientific, societal, and environmental implications. This book will interest people from many backgrounds, especially Geographic Information Systems (GIS) users interested in applying their domain-specific knowledge in a powerful open source language for data science, and R users interested in extending their skills to handle spatial data. The book is divided into three parts: (I) Foundations, aimed at getting you up-to-speed with geographic data in R, (II) extensions, which covers advanced techniques, and (III) applications to real-world problems. The chapters cover progressively more advanced topics, with early chapters providing strong foundations on which the later chapters build. Part I describes the nature of spatial datasets in R and methods for manipulating them. It also covers geographic data import/export and transforming coordinate reference systems. Part II represents methods that build on these foundations. It covers advanced map making (including web mapping), "bridges" to GIS, sharing reproducible code, and how to do cross-validation in the presence of spatial autocorrelation. Part III applies the knowledge gained to tackle real-world problems, including



representing and modeling transport systems, finding optimal locations for stores or services, and ecological modeling. Exercises at the end of each chapter give you the skills needed to tackle a range of geospatial problems. Solutions for each chapter and supplementary materials providing extended examples are available at <https://geocompr.github.io/geocompkg/articles/>. Dr. Robin Lovelace is a University Academic Fellow at the University of Leeds, where he has taught R for geographic research over many years, with a focus on transport systems. Dr. Jakub Nowosad is an Assistant Professor in the Department of Geoinformation at the Adam Mickiewicz University in Poznan, where his focus is on the analysis of large datasets to understand environmental processes. Dr. Jannes Muenchow is a Postdoctoral Researcher in the GIScience Department at the University of Jena, where he develops and teaches a range of geographic methods, with a focus on ecological modeling, statistical geocomputing, and predictive mapping. All three are active developers and work on a number of R packages, including `stplanr`, `sabre`, and `RQGIS`.

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.

"Practical recipes for visualizing data"--Cover.

Despite the recent rapid growth in machine learning and predictive analytics, many of the statistical questions that are faced by researchers and practitioners still involve explaining why something is happening. Regression analysis is the best 'swiss army knife' we have for answering these kinds of questions. This book is a learning resource on inferential statistics and regression analysis. It teaches how to do a wide range of statistical analyses in both R and in Python, ranging from simple hypothesis testing to advanced multivariate modelling. Although it is primarily focused on examples related to the analysis of people and talent, the methods easily transfer to any discipline. The book hits a 'sweet spot' where there is just enough mathematical theory to support a strong understanding of the methods, but with a step-by-step guide and easily reproducible examples and code, so that the methods can be put into practice immediately. This makes the book accessible to a wide readership, from public and private sector analysts and practitioners to students and researchers. Key Features:

- 16 accompanying datasets across a wide range of contexts (e.g. academic, corporate, sports, marketing)
- Clear step-by-step instructions on executing the analyses.
- Clear guidance on how to interpret results.
- Primary instruction in R but added sections for Python coders.
- Discussion exercises and data exercises for each of the main chapters.
- Final chapter of practice material and datasets ideal for class homework or project work.

With the advancement of statistical methodology inextricably linked to the use of computers, new methodological ideas must be translated into usable code and then numerically evaluated relative to competing procedures. In response to this, *Statistical Computing in C++ and R* concentrates on the writing of code rather than the development and study of numerical algorithms per se. The book discusses code development in C++ and R and the use of these symbiotic languages in unison. It emphasizes that each offers distinct features that, when used in tandem, can take code writing beyond what can be obtained from either language alone. The text begins with some basics of object-oriented languages, followed by a "boot-camp" on the use of C++ and R. The authors then discuss code development for the solution of specific computational problems that are relevant to statistics including optimization, numerical linear algebra, and random number generation. Later chapters introduce abstract data structures (ADTs) and parallel computing concepts. The appendices cover R and UNIX Shell programming. Features Includes numerous student exercises ranging from elementary to challenging Integrates both C++ and R for the solution of statistical computing problems Uses

C++ code in R and R functions in C++ programs Provides downloadable programs, available from the authors' website The translation of a mathematical problem into its computational analog (or analogs) is a skill that must be learned, like any other, by actively solving relevant problems. The text reveals the basic principles of algorithmic thinking essential to the modern statistician as well as the fundamental skill of communicating with a computer through the use of the computer languages C++ and R. The book lays the foundation for original code development in a research environment.

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](http://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](http://manning.com/kabacoff2)

"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"--

An Essential Reference for Intermediate and Advanced R Programmers Advanced R presents useful tools and techniques for attacking many types of R programming problems, helping you avoid mistakes and dead ends. With more than ten years of experience programming in R, the author illustrates the elegance, beauty, and flexibility at the heart of R. The book develops the necessary skills to produce quality code that can be used in a variety of circumstances. You will learn: The fundamentals of R, including standard data types and functions Functional programming as a useful

framework for solving wide classes of problems The positives and negatives of metaprogramming How to write fast, memory-efficient code This book not only helps current R users become R programmers but also shows existing programmers what's special about R. Intermediate R programmers can dive deeper into R and learn new strategies for solving diverse problems while programmers from other languages can learn the details of R and understand why R works the way it does.

Computer software is an essential tool for many statistical modelling and data analysis techniques, aiding in the implementation of large data sets in order to obtain useful results. R is one of the most powerful and flexible statistical software packages available, and enables the user to apply a wide variety of statistical methods ranging from simple regression to generalized linear modelling. *Statistics: An Introduction using R* is a clear and concise introductory textbook to statistical analysis using this powerful and free software, and follows on from the success of the author's previous best-selling title *Statistical Computing*. \* Features step-by-step instructions that assume no mathematics, statistics or programming background, helping the non-statistician to fully understand the methodology. \* Uses a series of realistic examples, developing step-wise from the simplest cases, with the emphasis on checking the assumptions (e.g. constancy of variance and normality of errors) and the adequacy of the model chosen to fit the data. \* The emphasis throughout is on estimation of effect sizes and confidence intervals, rather than on hypothesis testing. \* Covers the full range of statistical techniques likely to be need to analyse the data from research projects, including elementary material like t-tests and chi-squared tests, intermediate methods like regression and analysis of variance, and more advanced techniques like generalized linear modelling. \* Includes numerous worked examples and exercises within each chapter. \* Accompanied by a website featuring worked examples, data sets, exercises and solutions: <http://www.imperial.ac.uk/bio/research/crawley/statistics> *Statistics: An Introduction using R* is the first text to offer such a concise introduction to a broad array of statistical methods, at a level that is elementary enough to appeal to a broad range of disciplines. It is primarily aimed at undergraduate students in medicine, engineering, economics and biology - but will also appeal to postgraduates who have not previously covered this area, or wish to switch to using R.

R is rapidly becoming the standard software for statistical analyses, graphical presentation of data, and programming in the natural, physical, social, and engineering sciences. *Getting Started with R* is now the go-to introductory guide for biologists wanting to learn how to use R in their research. It teaches readers how to import, explore, graph, and analyse data, while keeping them focused on their ultimate goals: clearly communicating their data in oral presentations, posters, papers, and reports. It provides a consistent workflow for using R that is simple, efficient, reliable, and reproducible. This second edition has been updated and expanded while retaining the concise and engaging nature of its predecessor, offering an accessible and fun introduction to the packages *dplyr* and *ggplot2* for data manipulation and graphing. It expands the set of basic statistics considered in the first edition to include new examples of a simple regression, a one-way and a two-way ANOVA. Finally, it introduces a new chapter on the generalised linear model. *Getting Started with R* is suitable for undergraduates, graduate students, professional researchers, and practitioners in the biological sciences.

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.

Up-to-Date Guidance from One of the Foremost Members of the R Core Team Written by John M. Chambers, the leading developer of the original S software, *Extending R* covers key concepts and techniques in R to support analysis and research projects. It presents the core ideas of R, provides programming guidance for projects of all scales, and introduces new, valuable techniques that extend R. The book first describes the fundamental characteristics and background of R, giving readers a foundation for the remainder of the text. It next discusses topics relevant to programming with R, including the apparatus that supports extensions. The book then extends R's data structures through object-oriented programming, which is the key technique for coping with complexity. The book also incorporates a new structure for interfaces applicable to a variety of languages. A reflection of what R is today, this guide explains how to design and organize extensions to R by correctly using objects, functions, and interfaces. It enables current and future users to add their own contributions and packages to R. A 2017 Choice Outstanding Academic Title

This book aims to provide a broad introduction to the R statistical environment in the context of applied regression analysis, which is typically studied by social scientists and others in a second course in applied statistics.

The contents of *The R Software* are presented so as to be both comprehensive and easy for the reader to use. Besides its application as a self-learning text, this book can support lectures on R at any level from beginner to advanced. This book can serve as a textbook on R for beginners as well as more advanced users, working on Windows, MacOs or Linux OSes. The first part of the book deals with the heart of the R language and its fundamental concepts, including data organization, import and export, various manipulations, documentation, plots, programming and maintenance. The last chapter in this part deals with oriented object programming as well as interfacing R with C/C++ or Fortran, and contains a section on debugging techniques. This is followed by the second part of the book, which provides detailed explanations on how to perform many standard statistical analyses, mainly in the Biostatistics field. Topics from mathematical and statistical settings that are included are matrix operations, integration, optimization, descriptive statistics, simulations, confidence intervals and hypothesis testing, simple and multiple linear regression, and analysis of variance. Each statistical chapter in the second part relies on one or more real biomedical data sets, kindly made available by the Bordeaux School of Public Health (Institut de Santé Publique, d'Épidémiologie et de Développement - ISPED) and described at the beginning of the book. Each chapter



ends with an assessment section: memorandum of most important terms, followed by a section of theoretical exercises (to be done on paper), which can be used as questions for a test. Moreover, worksheets enable the reader to check his new abilities in R. Solutions to all exercises and worksheets are included in this book.

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.

Statistical Inference via Data Science: A ModernDive into R and the Tidyverse provides a pathway for learning about statistical inference using data science tools widely used in industry, academia, and government. It introduces the tidyverse suite of R packages, including the `ggplot2` package for data visualization, and the `dplyr` package for data wrangling. After equipping readers with just enough of these data science tools to perform effective exploratory data analyses, the book covers traditional introductory statistics topics like confidence intervals, hypothesis testing, and multiple regression modeling, while focusing on visualization throughout. Features:

- ? Assumes minimal prerequisites, notably, no prior calculus nor coding experience
- ? Motivates theory using real-world data, including all domestic flights leaving New York City in 2013, the Gapminder project, and the data journalism website, `FiveThirtyEight.com`
- ? Centers on simulation-based approaches to statistical inference rather than mathematical formulas
- ? Uses the `infer` package for "tidy" and transparent statistical inference to construct confidence intervals and conduct hypothesis tests via the bootstrap and permutation methods
- ? Provides all code and output embedded directly in the text; also available in the online version at `moderndive.com`

This book is intended for individuals who would like to simultaneously start developing their

data science toolbox and start learning about the inferential and modeling tools used in much of modern-day research. The book can be used in methods and data science courses and first courses in statistics, at both the undergraduate and graduate levels.

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.

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.

Praise for previous editions: "Gandrud has written a great outline of how a fully reproducible research project should look from start to finish, with brief explanations of each tool that he uses along the way... Advanced undergraduate

students in mathematics, statistics, and similar fields as well as students just beginning their graduate studies would benefit the most from reading this book. Many more experienced R users or second-year graduate students might find themselves thinking, 'I wish I'd read this book at the start of my studies, when I was first learning R!'...This book could be used as the main text for a class on reproducible research ..." (The American Statistician) Reproducible Research with R and R Studio, Third Edition brings together the skills and tools needed for doing and presenting computational research. Using straightforward examples, the book takes you through an entire reproducible research workflow. This practical workflow enables you to gather and analyze data as well as dynamically present results in print and on the web. Supplementary materials and example are available on the author's website. New to the Third Edition Updated package recommendations, examples, URLs, and removed technologies no longer in regular use. More advanced R Markdown (and less LaTeX) in discussions of markup languages and examples. Stronger focus on reproducible working directory tools. Updated discussion of cloud storage services and persistent reproducible material citation. Added discussion of Jupyter notebooks and reproducible practices in industry. Examples of data manipulation with Tidyverse tibbles (in addition to standard data frames) and `pivot_longer()` and `pivot_wider()` functions for pivoting data. Features Incorporates the most important advances that have been developed since the editions were published Describes a complete reproducible research workflow, from data gathering to the presentation of results Shows how to automatically generate tables and figures using R Includes instructions on formatting a presentation document via markup languages Discusses cloud storage and versioning services, particularly Github Explains how to use Unix-like shell programs for working with large research projects

R Markdown: The Definitive Guide is the first official book authored by the core R Markdown developers that provides a comprehensive and accurate reference to the R Markdown ecosystem. With R Markdown, you can easily create reproducible data analysis reports, presentations, dashboards, interactive applications, books, dissertations, websites, and journal articles, while enjoying the simplicity of Markdown and the great power of R and other languages. In this book, you will learn Basics: Syntax of Markdown and R code chunks, how to generate figures and tables, and how to use other computing languages Built-in output formats of R Markdown: PDF/HTML/Word/RTF/Markdown documents and ioslides/Slidy/Beamer/PowerPoint presentations Extensions and applications: Dashboards, Tufte handouts, xaringan/reveal.js presentations, websites, books, journal articles, and interactive tutorials Advanced topics: Parameterized reports, HTML widgets, document templates, custom output formats, and Shiny documents. Yihui Xie is a software engineer at RStudio. He has authored and co-authored several R packages, including knitr, rmarkdown, bookdown, blogdown, shiny, xaringan, and animation. He has published three other books, Dynamic

Documents with R and knitr, bookdown: Authoring Books and Technical Documents with R Markdown, and blogdown: Creating Websites with R Markdown. J.J. Allaire is the founder of RStudio and the creator of the RStudio IDE. He is an author of several packages in the R Markdown ecosystem including rmarkdown, flexdashboard, learnr, and radix. Garrett Golemund is the co-author of R for Data Science and author of Hands-On Programming with R. He wrote the lubridate R package and works for RStudio as an advocate who trains engineers to do data science with R and the Tidyverse.

Learn how to program by diving into the R language, and then use your newfound skills to solve practical data science problems. With this book, you'll learn how to load data, assemble and disassemble data objects, navigate R's environment system, write your own functions, and use all of R's programming tools. RStudio Master Instructor Garrett Golemund not only teaches you how to program, but also shows you how to get more from R than just visualizing and modeling data. You'll gain valuable programming skills and support your work as a data scientist at the same time. Work hands-on with three practical data analysis projects based on casino games Store, retrieve, and change data values in your computer's memory Write programs and simulations that outperform those written by typical R users Use R programming tools such as if else statements, for loops, and S3 classes Learn how to write lightning-fast vectorized R code Take advantage of R's package system and debugging tools Practice and apply R programming concepts as you learn them

The R Book John Wiley & Sons

An Introduction to Statistical Learning provides an accessible overview of the field of statistical learning, an essential toolset for making sense of the vast and complex data sets that have emerged in fields ranging from biology to finance to marketing to astrophysics in the past twenty years. This book presents some of the most important modeling and prediction techniques, along with relevant applications. Topics include linear regression, classification, resampling methods, shrinkage approaches, tree-based methods, support vector machines, clustering, and more. Color graphics and real-world examples are used to illustrate the methods presented. Since the goal of this textbook is to facilitate the use of these statistical learning techniques by practitioners in science, industry, and other fields, each chapter contains a tutorial on implementing the analyses and methods presented in R, an extremely popular open source statistical software platform. Two of the authors co-wrote The Elements of Statistical Learning (Hastie, Tibshirani and Friedman, 2nd edition 2009), a popular reference book for statistics and machine learning researchers. An Introduction to Statistical Learning covers many of the same topics, but at a level accessible to a much broader audience. This book is targeted at statisticians and non-statisticians alike who wish to use cutting-edge statistical learning techniques to analyze their data. The text assumes only a previous course in linear regression and no knowledge of matrix algebra.

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

Introduces professionals and scientists to statistics and machine learning using the programming language R Written by and for practitioners, this book provides an overall introduction to R, focusing on tools and methods commonly used in data science, and placing emphasis on practice and business use. It covers a wide range of topics in a single volume, including big data, databases, statistical machine learning, data wrangling, data visualization, and the reporting of results. The topics covered are all important for someone with a science/math background that is looking to quickly learn several practical technologies to enter or transition to the growing field of data science. The Big R-Book for Professionals: From Data Science to Learning Machines and Reporting with R includes nine parts, starting with an introduction to the subject and followed by an overview of R and elements of statistics. The third part revolves around data, while the fourth focuses on data wrangling. Part 5 teaches readers about exploring data. In Part 6 we learn to build models, Part 7 introduces the reader to the reality in companies, Part 8 covers reports and interactive applications and finally Part 9 introduces the reader to big data and performance computing. It also includes some helpful appendices. Provides a practical guide for non-experts with a focus on business users Contains a unique combination of topics including an introduction to R, machine learning, mathematical models, data wrangling, and reporting Uses a practical tone and integrates multiple topics in a coherent framework Demystifies the hype around machine learning and AI by enabling readers to understand the provided models and program them in R Shows readers how to visualize results in static and interactive reports Supplementary materials includes PDF slides based on the book's content, as well as all the extracted R-code and is available to everyone on a Wiley Book Companion Site The Big R-Book is an excellent guide for science technology, engineering, or mathematics students who wish to make a successful transition from the academic world to the professional. It will also appeal to all young data scientists, quantitative analysts, and analytics professionals, as well as those who make mathematical models.

This book provides a general introduction to the R Commander graphical user interface (GUI) to R for readers who are unfamiliar with R. It is suitable for use as a supplementary text in a basic or intermediate-level statistics course. It is not intended to replace a basic or other

statistics text but rather to complement it, although it does promote sound statistical practice in the examples. The book should also be useful to individual casual or occasional users of R for whom the standard command-line interface is an obstacle. [tinyurl.com/RcmdrBook](http://tinyurl.com/RcmdrBook) The site includes data files used in the book and an errata list.

<http://socserv.mcmaster.ca/jfox/Books/RCommander/Writing-Rcmdr-Plugins.pdf> Writing R Commander Plug-in Packages

R is now the most widely used statistical software in academic science and it is rapidly expanding into other fields such as finance. R is almost limitlessly flexible and powerful, hence its appeal, but can be very difficult for the novice user. There are no easy pull-down menus, error messages are often cryptic and simple tasks like importing your data or exporting a graph can be difficult and frustrating. Introductory R is written for the novice user who knows a little about statistics but who hasn't yet got to grips with the ways of R. This new edition is completely revised and greatly expanded with new chapters on the basics of descriptive statistics and statistical testing, considerably more information on statistics and six new chapters on programming in R. Topics covered include: A walkthrough of the basics of R's command line interface Data structures including vectors, matrices and data frames R functions and how to use them Expanding your analysis and plotting capacities with add-in R packages A set of simple rules to follow to make sure you import your data properly An introduction to the script editor and advice on workflow A detailed introduction to drawing publication-standard graphs in R How to understand the help files and how to deal with some of the most common errors that you might encounter. Basic descriptive statistics The theory behind statistical testing and how to interpret the output of statistical tests Thorough coverage of the basics of data analysis in R with chapters on using chi-squared tests, t-tests, correlation analysis, regression, ANOVA and general linear models What the assumptions behind the analyses mean and how to test them using diagnostic plots Explanations of the summary tables produced for statistical analyses such as regression and ANOVA Writing your own functions in R Using table operations to manipulate matrices and data frames Using conditional statements and loops in R programmes. Writing longer R programmes. The techniques of statistical analysis in R are illustrated by a series of chapters where experimental and survey data are analysed. There is a strong emphasis on using real data from real scientific research, with all the problems and uncertainty that implies, rather than well-behaved made-up data that give ideal and easy to analyse results.

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

Statistical Rethinking: A Bayesian Course with Examples in R and Stan builds readers' knowledge of and confidence in statistical modeling. Reflecting the need for even minor programming in today's model-based statistics, the book pushes readers to perform step-by-step calculations that are usually automated. This unique computational approach ensures that readers understand enough of the details to make reasonable choices and interpretations in

their own modeling work. The text presents generalized linear multilevel models from a Bayesian perspective, relying on a simple logical interpretation of Bayesian probability and maximum entropy. It covers from the basics of regression to multilevel models. The author also discusses measurement error, missing data, and Gaussian process models for spatial and network autocorrelation. By using complete R code examples throughout, 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 prepares 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.

[Copyright: 47f504c58e60dcfb31d9416dab70882f](#)