

# R

## R for Data Science

Learn how to use R to turn raw data into insight, knowledge, and understanding. 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, R for Data Science is designed to get you doing data science as quickly as possible. Authors Hadley Wickham and Garrett Grolemund guide you through the steps of importing, wrangling, exploring, and modeling your data and communicating the results. You'll get a complete, big-picture understanding of the data science cycle, along with basic tools you need to manage the details. Each section of the book is paired with exercises to help you practice what you've learned along the way. You'll learn how to: Wrangle—transform your datasets into a form convenient for analysis Program—learn powerful R tools for solving data problems with greater clarity and ease Explore—examine your data, generate hypotheses, and quickly test them Model—provide a low-dimensional summary that captures true "signals" in your dataset Communicate—learn R Markdown for integrating prose, code, and results

## Advanced R

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.

## Geocomputation with R

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/>.

## **Learning R**

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

## **An Introduction to Statistical Learning**

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, marketing, and 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, deep learning, survival analysis, multiple testing, and more. Color graphics and real-world examples are used to illustrate the methods presented. This book is targeted at statisticians and non-statisticians alike, who wish to use cutting-edge statistical learning techniques to analyze their data. Four of the authors co-wrote An Introduction to Statistical Learning, With Applications in R (ISLR), which has become a mainstay of undergraduate and graduate classrooms worldwide, as well as an important reference book for data scientists. One of the keys to its success was that each chapter contains a tutorial on implementing the analyses and methods presented in the R scientific computing environment. However, in recent years Python has become a popular language for data science, and there has been increasing demand for a Python-based alternative to ISLR. Hence, this book (ISLP) covers the same materials as ISLR but with labs implemented in Python. These labs will be useful both for Python novices, as well as experienced users.

## **Hands-On Machine Learning with R**

Hands-on Machine Learning with R provides a practical and applied approach to learning and developing intuition into today's most popular machine learning methods. This book serves as a practitioner's guide to the machine learning process and is meant to help the reader learn to apply the machine learning stack within R, which includes using various R packages such as glmnet, h2o, ranger, xgboost, keras, and others to effectively model and gain insight from their data. The book favors a hands-on approach, providing an intuitive understanding of machine learning concepts through concrete examples and just a little bit of theory. Throughout this book, the reader will be exposed to the entire machine learning process including feature engineering, resampling, hyperparameter tuning, model evaluation, and interpretation. The reader will be exposed to powerful algorithms such as regularized regression, random forests, gradient boosting machines, deep learning, generalized low rank models, and more! By favoring a hands-on approach and using real world data, the reader will gain an intuitive understanding of the architectures and engines that drive these algorithms and packages, understand when and how to tune the various hyperparameters, and be able to interpret model results. By the end of this book, the reader should have a firm grasp of R's machine learning

stack and be able to implement a systematic approach for producing high quality modeling results. Features:

- Offers a practical and applied introduction to the most popular machine learning methods.
- Topics covered include feature engineering, resampling, deep learning and more.
- Uses a hands-on approach and real world data.

## **R Packages**

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

## **The Art of R Programming**

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.

## **R Cookbook**

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

## Learn R

Learning a computer language like R can be either frustrating, fun, or boring. Having fun requires challenges that wake up the learner's curiosity but also provide an emotional reward on overcoming them. This book is designed so that it includes smaller and bigger challenges, in what I call playgrounds, in the hope that all readers will enjoy their path to R fluency. Fluency in the use of a language is a skill that is acquired through practice and exploration. Although rarely mentioned separately, fluency in a computer programming language involves both writing and reading. The parallels between natural and computer languages are many, but differences are also important. For students and professionals in the biological sciences, humanities, and many applied fields, recognizing the parallels between R and natural languages should help them feel at home with R. The approach I use is similar to that of a travel guide, encouraging exploration and describing the available alternatives and how to reach them. The intention is to guide the reader through the R landscape of 2020 and beyond. Features R as it is currently used Few prescriptive rules—mostly the author's preferences together with alternatives Explanation of the R grammar emphasizing the "R way of doing things" Tutoring for "programming in the small" using scripts The grammar of graphics and the grammar of data described as grammars Examples of data exchange between R and the foreign world using common file formats Coaching for becoming an independent R user, capable of both writing original code and solving future challenges What makes this book different from others: Tries to break the ice and help readers from all disciplines feel at home with R Does not make assumptions about what the reader will use R for Attempts to do only one thing well: guide readers into becoming fluent in the R language Pedro J. Aphalo is a PhD graduate from the University of Edinburgh, and is currently a lecturer at the University of Helsinki. A plant biologist and agriculture scientist with a passion for data, electronics, computers, and photography, in addition to plants, Dr. Aphalo has been a user of R for 25 years. He first organized an R course for MSc students 18 years ago, and is the author of 13 R packages currently in CRAN.

## Meta-Analysis with R

This book provides a comprehensive introduction to performing meta-analysis using the statistical software R. It is intended for quantitative researchers and students in the medical and social sciences who wish to learn how to perform meta-analysis with R. As such, the book introduces the key concepts and models used in meta-analysis. It also includes chapters on the following advanced topics: publication bias and small study effects; missing data; multivariate meta-analysis, network meta-analysis; and meta-analysis of diagnostic studies.

## Deep Learning with R

Summary Deep Learning with R introduces the world of deep learning using the powerful Keras library and its R language interface. The book builds your understanding of deep learning through intuitive explanations and practical examples. Continue your journey into the world of deep learning with Deep Learning with R in Motion, a practical, hands-on video course available exclusively at Manning.com ([www.manning.com/livevideo/deep-learning-with-r-in-motion](http://www.manning.com/livevideo/deep-learning-with-r-in-motion)). Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Machine learning has made remarkable progress in recent years. Deep-learning systems now enable previously impossible smart applications, revolutionizing image recognition and natural-language processing, and identifying complex patterns in data. The Keras deep-learning library provides data scientists and developers working in R a state-of-the-art toolset for tackling deep-learning tasks. About the Book Deep Learning with R introduces the world of deep learning using the powerful Keras library and its R language interface. Initially written for Python as Deep Learning with Python by Keras creator and Google AI researcher François Chollet and adapted for R by RStudio founder J. J. Allaire, this book builds your understanding of deep learning through intuitive explanations and practical examples. You'll practice your new skills with R-based applications in computer vision, natural-language processing, and generative models. What's Inside Deep learning from first principles Setting up your own deep-learning environment Image classification and generation Deep learning for text and sequences About the Reader You'll need intermediate R programming

skills. No previous experience with machine learning or deep learning is assumed. About the Authors François Chollet is a deep-learning researcher at Google and the author of the Keras library. J.J. Allaire is the founder of RStudio and the author of the R interfaces to TensorFlow and Keras. Table of Contents PART 1 - FUNDAMENTALS OF DEEP LEARNING What is deep learning? Before we begin: the mathematical building blocks of neural networks Getting started with neural networks Fundamentals of machine learning PART 2 - DEEP LEARNING IN PRACTICE Deep learning for computer vision Deep learning for text and sequences Advanced deep-learning best practices Generative deep learning Conclusions

## **The R Software**

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.

## **Learning Statistics with R**

There are many excellent R resources for visualization, data science, and package development. Hundreds of scattered vignettes, web pages, and forums explain how to use R in particular domains. But little has been written on how to simply make R work effectively—until now. This hands-on book teaches novices and experienced R users how to write efficient R code. Drawing on years of experience teaching R courses, authors Colin Gillespie and Robin Lovelace provide practical advice on a range of topics—from optimizing the set-up of RStudio to leveraging C++—that make this book a useful addition to any R user's bookshelf. Academics, business users, and programmers from a wide range of backgrounds stand to benefit from the guidance in Efficient R Programming. Get advice for setting up an R programming environment Explore general programming concepts and R coding techniques Understand the ingredients of an efficient R workflow Learn how to efficiently read and write data in R Dive into data carpentry—the vital skill for cleaning raw data Optimize your code with profiling, standard tricks, and other methods Determine your hardware capabilities for handling R computation Maximize the benefits of collaborative R programming Accelerate your transition from R hacker to R programmer

## **Efficient R Programming**

This book introduces readers to the fundamentals of creating presentation graphics using R, based on 111 detailed and complete scripts. It shows how bar and column charts, population pyramids, Lorenz curves, box plots, scatter plots, time series, radial polygons, Gantt charts, heat maps, bump charts, mosaic and balloon charts, and a series of different thematic map types can be created using R's Base Graphics System. Every example uses real data and includes step-by-step explanations of the figures and their programming. This second edition contains additional examples for cartograms, chord-diagrams and networks, and interactive

visualizations with Javascript. The open source software R is an established standard and a powerful tool for various visualizing applications, integrating nearly all technologies relevant for data visualization. The basic software, enhanced by more than 14000 extension packs currently freely available, is intensively used by organizations including Google, Facebook and the CIA. The book serves as a comprehensive reference guide to a broad variety of applications in various fields. This book is intended for all kinds of R users, ranging from experts, for whom especially the example codes are particularly useful, to beginners, who will find the finished graphics most helpful in learning what R can actually deliver.

## **Data Visualisation with R**

Data science has taken the world by storm. Every field of study and area of business has been affected as people increasingly realize the value of the incredible quantities of data being generated. But to extract value from those data, one needs to be trained in the proper data science skills. The R programming language has become the de facto programming language for data science. Its flexibility, power, sophistication, and expressiveness have made it an invaluable tool for data scientists around the world. This book is about the fundamentals of R programming. You will get started with the basics of the language, learn how to manipulate datasets, how to write functions, and how to debug and optimize code. With the fundamentals provided in this book, you will have a solid foundation on which to build your data science toolbox.

## **R Programming for Data Science**

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.

## **Introductory Statistics with R**

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.

## **Advanced R, Second Edition**

This Brief provides a roadmap for the R language and programming environment with signposts to further resources and documentation.

## **A Tiny Handbook of R**

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 Grolemund 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

## **Hands-On Programming with R**

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

## **Introduction to Data Science**

Much of the data available today is unstructured and text-heavy, making it challenging for analysts to apply their usual data wrangling and visualization tools. With this practical book, you'll explore text-mining techniques with tidytext, a package that authors Julia Silge and David Robinson developed using the tidy principles behind R packages like ggraph and dplyr. You'll learn how tidytext and other tidy tools in R can make text analysis easier and more effective. The authors demonstrate how treating text as data frames enables you to manipulate, summarize, and visualize characteristics of text. You'll also learn how to integrate

natural language processing (NLP) into effective workflows. Practical code examples and data explorations will help you generate real insights from literature, news, and social media. Learn how to apply the tidy text format to NLP Use sentiment analysis to mine the emotional content of text Identify a document's most important terms with frequency measurements Explore relationships and connections between words with the ggraph and widyr packages Convert back and forth between R's tidy and non-tidy text formats Use topic modeling to classify document collections into natural groups Examine case studies that compare Twitter archives, dig into NASA metadata, and analyze thousands of Usenet messages

## **Text Mining with R**

This book offers an application-oriented guide to random forests: a statistical learning method extensively used in many fields of application, thanks to its excellent predictive performance, but also to its flexibility, which places few restrictions on the nature of the data used. Indeed, random forests can be adapted to both supervised classification problems and regression problems. In addition, they allow us to consider qualitative and quantitative explanatory variables together, without pre-processing. Moreover, they can be used to process standard data for which the number of observations is higher than the number of variables, while also performing very well in the high dimensional case, where the number of variables is quite large in comparison to the number of observations. Consequently, they are now among the preferred methods in the toolbox of statisticians and data scientists. The book is primarily intended for students in academic fields such as statistical education, but also for practitioners in statistics and machine learning. A scientific undergraduate degree is quite sufficient to take full advantage of the concepts, methods, and tools discussed. In terms of computer science skills, little background knowledge is required, though an introduction to the R language is recommended. Random forests are part of the family of tree-based methods; accordingly, after an introductory chapter, Chapter 2 presents CART trees. The next three chapters are devoted to random forests. They focus on their presentation (Chapter 3), on the variable importance tool (Chapter 4), and on the variable selection problem (Chapter 5), respectively. After discussing the concepts and methods, we illustrate their implementation on a running example. Then, various complements are provided before examining additional examples. Throughout the book, each result is given together with the code (in R) that can be used to reproduce it. Thus, the book offers readers essential information and concepts, together with examples and the software tools needed to analyse data using random forests.

## **Random Forests with R**

Presenting a practitioner's guide to capabilities and best practices of quality control systems using the R programming language, this volume emphasizes accessibility and ease-of-use through detailed explanations of R code as well as standard statistical methodologies. In the interest of reaching the widest possible audience of quality-control professionals and statisticians, examples throughout are structured to simplify complex equations and data structures, and to demonstrate their applications to quality control processes, such as ISO standards. The volume balances its treatment of key aspects of quality control, statistics, and programming in R, making the text accessible to beginners and expert quality control professionals alike. Several appendices serve as useful references for ISO standards and common tasks performed while applying quality control with R.

## **Quality Control with R**

Gain the R programming language fundamentals for doing the applied statistics useful for data exploration and analysis in data science and data mining. This book covers topics ranging from R syntax basics, descriptive statistics, and data visualizations to inferential statistics and regressions. After learning R's syntax, you will work through data visualizations such as histograms and boxplot charting, descriptive statistics, and inferential statistics such as t-test, chi-square test, ANOVA, non-parametric test, and linear regressions. Learn R for Applied Statistics is a timely skills-migration book that equips you with the R programming fundamentals and introduces you to applied statistics for data explorations. What You Will



Learn Discover R, statistics, data science, data mining, and big data Master the fundamentals of R programming, including variables and arithmetic, vectors, lists, data frames, conditional statements, loops, and functions Work with descriptive statistics Create data visualizations, including bar charts, line charts, scatter plots, boxplots, histograms, and scatterplots Use inferential statistics including t-tests, chi-square tests, ANOVA, non-parametric tests, linear regressions, and multiple linear regressions Who This Book Is For Those who are interested in data science, in particular data exploration using applied statistics, and the use of R programming for data visualizations.

## **Learn R for Applied Statistics**

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions

## **Mastering Spark with R**

Master the Shiny web framework—and take your R skills to a whole new level. By letting you move beyond static reports, Shiny helps you create fully interactive web apps for data analyses. Users will be able to jump between datasets, explore different subsets or facets of the data, run models with parameter values of their choosing, customize visualizations, and much more. Hadley Wickham from RStudio shows data scientists, data analysts, statisticians, and scientific researchers with no knowledge of HTML, CSS, or JavaScript how to create rich web apps from R. This in-depth guide provides a learning path that you can follow with confidence, as you go from a Shiny beginner to an expert developer who can write large, complex apps that are maintainable and performant. Get started: Discover how the major pieces of a Shiny app fit together Put Shiny in action: Explore Shiny functionality with a focus on code samples, example apps, and useful techniques Master reactivity: Go deep into the theory and practice of reactive programming and examine reactive graph components Apply best practices: Examine useful techniques for making your Shiny apps work well in production

## **Mastering Shiny**

Although there are currently a wide variety of software packages suitable for the modern statistician, R has the triple advantage of being comprehensive, widespread, and free. Published in 2008, the second edition of *Statistiques avec R* enjoyed great success as an R guidebook in the French-speaking world. Translated and updated, *R for Statistics* in

## **R for Statistics**

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, Second Edition*, 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, manipulation, and visualization; 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. After all this you'll make your code reproducible with LaTeX, RMarkdown, and Shiny. 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 Explore R, RStudio, and R packages Use R for math: variable types, vectors, calling functions, and more Exploit data structures, including data.frames, matrices, and lists Read many different types of data Create attractive, intuitive statistical graphics Write user-defined functions Control program flow with if, ifelse, and complex checks Improve program efficiency with group manipulations Combine and reshape multiple datasets Manipulate strings using R's facilities and regular expressions Create normal, binomial, and Poisson probability distributions Build linear, generalized linear, and nonlinear models Program basic statistics: mean, standard deviation, and t-tests Train machine learning models Assess the quality of models and variable selection Prevent overfitting and perform variable selection, using the Elastic Net and Bayesian methods Analyze univariate and multivariate time series data Group data via K-means and hierarchical clustering Prepare reports, slideshows, and web pages with knitr Display interactive data with RMarkdown and htmlwidgets Implement dashboards with Shiny Build reusable R packages with devtools and Rcpp Register your product at [informit.com/register](http://informit.com/register) for convenient access to downloads, updates, and corrections as they become available.

## **R for Everyone**

The richly illustrated Interactive Web-Based Data Visualization with R, plotly, and shiny focuses on the process of programming interactive web graphics for multidimensional data analysis. It is written for the data analyst who wants to leverage the capabilities of interactive web graphics without having to learn web programming. Through many R code examples, you will learn how to tap the extensive functionality of these tools to enhance the presentation and exploration of data. By mastering these concepts and tools, you will impress your colleagues with your ability to quickly generate more informative, engaging, and reproducible interactive graphics using free and open source software that you can share over email, export to pdf, and more. Key Features: Convert static ggplot2 graphics to an interactive web-based form Link, animate, and arrange multiple plots in standalone HTML from R Embed, modify, and respond to plotly graphics in a shiny app Learn best practices for visualizing continuous, discrete, and multivariate data Learn numerous ways to visualize geo-spatial data This book makes heavy use of plotly for graphical rendering, but you will also learn about other R packages that support different phases of a data science workflow, such as tidyr, dplyr, and tidyverse. Along the way, you will gain insight into best practices for visualization of high-dimensional data, statistical graphics, and graphical perception. The printed book is complemented by an interactive website where readers can view movies demonstrating the examples and interact with graphics.

## **Interactive Web-Based Data Visualization with R, plotly, and shiny**

Networks have permeated everyday life through everyday realities like the Internet, social networks, and viral marketing. As such, network analysis is an important growth area in the quantitative sciences, with roots in social network analysis going back to the 1930s and graph theory going back centuries. Measurement and analysis are integral components of network research. As a result, statistical methods play a critical role in network analysis. This book is the first of its kind in network research. It can be used as a stand-alone resource in which multiple R packages are used to illustrate how to conduct a wide range of network analyses, from basic manipulation and visualization, to summary and characterization, to modeling of network data. The central package is igraph, which provides extensive capabilities for studying network graphs in R. This text builds on Eric D. Kolaczyk's book Statistical Analysis of Network Data (Springer, 2009).

## **Statistical Analysis of Network Data with R**

An essential guide to the trouble spots and oddities of R. In spite of the quirks exposed here, R is the best computing environment for most data analysis tasks. R is free, open-source, and has thousands of contributed packages. It is used in such diverse fields as ecology, finance, genomics and music. If you are using spreadsheets to understand data, switch to R. You will have safer -- and ultimately, more convenient -- computations.

## **The R Inferno**

R is a language and environment for data analysis and graphics. It may be considered an implementation of S, an award-winning language initially developed at Bell Laboratories since the late 1970s. The R project was initiated by Robert Gentleman and Ross Ihaka at the University of Auckland, New Zealand, in the early 1990s, and has been developed by an international team since mid-1997. Historically, econometricians have favored other computing environments, some of which have fallen by the wayside, and also a variety of packages with canned routines. We believe that R has great potential in econometrics, both for research and for teaching. There are at least three reasons for this: (1) R is mostly platform independent and runs on Microsoft Windows, the Mac family of operating systems, and various flavors of Unix/Linux, and also on some more exotic platforms. (2) R is free software that can be downloaded and installed at no cost from a family of mirror sites around the globe, the Comprehensive R Archive Network (CRAN); hence students can easily install it on their own machines. (3) R is open-source software, so that the full source code is available and can be inspected to understand what it really does, learn from it, and modify and extend it. We also like to think that platform independence and the open-source philosophy make R an ideal environment for reproducible econometric research.

## **Applied Econometrics with R**

Applied Survival Analysis Using R covers the main principles of survival analysis, gives examples of how it is applied, and teaches how to put those principles to use to analyze data using R as a vehicle. Survival data, where the primary outcome is time to a specific event, arise in many areas of biomedical research, including clinical trials, epidemiological studies, and studies of animals. Many survival methods are extensions of techniques used in linear regression and categorical data, while other aspects of this field are unique to survival data. This text employs numerous actual examples to illustrate survival curve estimation, comparison of survivals of different groups, proper accounting for censoring and truncation, model variable selection, and residual analysis. Because explaining survival analysis requires more advanced mathematics than many other statistical topics, this book is organized with basic concepts and most frequently used procedures covered in earlier chapters, with more advanced topics near the end and in the appendices. A background in basic linear regression and categorical data analysis, as well as a basic knowledge of calculus and the R system, will help the reader to fully appreciate the information presented. Examples are simple and straightforward while still illustrating key points, shedding light on the application of survival analysis in a way that is useful for graduate students, researchers, and practitioners in biostatistics.

## **Applied Survival Analysis Using R**

R by Example is an example-based introduction to the statistical computing environment that does not assume any previous familiarity with R or other software packages. R functions are presented in the context of interesting applications with real data. The purpose of this book is to illustrate a range of statistical and probability computations using R for people who are learning, teaching, or using statistics. Specifically, this book is written for users who have covered at least the equivalent of (or are currently studying) undergraduate level calculus-based courses in statistics. These users are learning or applying exploratory and inferential methods for analyzing data and this book is intended to be a useful resource for learning how to

implement these procedures in R.

## **R by Example**

Master the programming language of choice among statisticians and data analysts worldwide Coming to grips with R can be tough, even for seasoned statisticians and data analysts. Enter *R For Dummies*, the quick, easy way to master all the R you'll ever need. Requiring no prior programming experience and packed with practical examples, easy, step-by-step exercises, and sample code, this extremely accessible guide is the ideal introduction to R for complete beginners. It also covers many concepts that intermediate-level programmers will find extremely useful. Master your R ABCs ? get up to speed in no time with the basics, from installing and configuring R to writing simple scripts and performing simultaneous calculations on many variables Put data in its place ? get to know your way around lists, data frames, and other R data structures while learning to interact with other programs, such as Microsoft Excel Make data dance to your tune ? learn how to reshape and manipulate data, merge data sets, split and combine data, perform calculations on vectors and arrays, and much more Visualize it ? learn to use R's powerful data visualization features to create beautiful and informative graphical presentations of your data Get statistical ? find out how to do simple statistical analysis, summarize your variables, and conduct classic statistical tests, such as t-tests Expand and customize R ? get the lowdown on how to find, install, and make the most of add-on packages created by the global R community for a wide variety of purposes Open the book and find: Help downloading, installing, and configuring R Tips for getting data in and out of R Ways to use data frames and lists to organize data How to manipulate and process data Advice on fitting regression models and ANOVA Helpful hints for working with graphics How to code in R What R mailing lists and forums can do for you

## **R For Dummies**

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.

## **R in Action**

*R Visualizations: Derive Meaning from Data* focuses on one of the two major topics of data analytics: data visualization, a.k.a., computer graphics. In the book, major R systems for visualization are discussed, organized by topic and not by system. Anyone doing data analysis will be shown how to use R to generate any of the basic visualizations with the R visualization systems. Further, this book introduces the author's *lessR* system, which always can accomplish a visualization with less coding than the use of other systems, sometimes dramatically so, and also provides accompanying statistical analyses. *Key Features* Presents thorough coverage of the leading R visualization system, *ggplot2*. Gives specific guidance on using base R graphics to attain visualizations of the same quality as those provided by *ggplot2*. Shows how to create a wide range of data visualizations: distributions of categorical and continuous variables, many types of scatterplots including with a third variable, time series, and maps. Inclusion of the various approaches to R graphics organized by topic instead of by system. Presents the recent work on interactive visualization in R. David W. Gerbing received his PhD from Michigan State University in 1979 in quantitative analysis, and currently is a professor of quantitative analysis in the School of Business at Portland State University. He has published extensively in the social and behavioral sciences with a focus on quantitative methods. His *lessR* package has been in development since 2009.

## R Visualizations

The objective of this book is to address the needs of students in using software for data science. The pace and content of this book is set in such a way so that it gives user friendly approach to students and other learners. The key features of this book are more than 60 packages have been discussed with relevant examples. The author has also incorporated the data analysis using R-Commander which is Graphic User Interface (GUI) of R programming. This book is a primer for upcoming fields such as Data Science and Machine Learning.

## Data Analysis Using R (Low Priced Edition): A Primer for Data Scientist

<https://works.spiderworks.co.in/~15026955/lfavourk/zpreventd/rinjurex/industrial+facilities+solutions.pdf>  
<https://works.spiderworks.co.in/@27602500/garisepe/echargew/uheadl/2014+jeep+wrangler+owners+manual.pdf>  
<https://works.spiderworks.co.in/@82229244/carisex/vpreventd/punitey/satchwell+room+thermostat+user+manual.pdf>  
<https://works.spiderworks.co.in/^58385420/wcarvez/gsmashc/frescueq/power+up+your+mind+learn+faster+work+smarter.pdf>  
<https://works.spiderworks.co.in/-62771064/wawardih/preventl/ypackb/braddocks+defeat+the+battle+of+the+monongahela+and+the+road+to+revolution.pdf>  
<https://works.spiderworks.co.in/@94862096/fcarvep/jsparew/tgetd/textbook+of+operative+urology+1e.pdf>  
<https://works.spiderworks.co.in/=31153522/cembodys/xpoura/kinjurei/elna+3003+manual+instruction.pdf>  
<https://works.spiderworks.co.in/~15532439/vembarkk/reditt/lguaranteeh/mini+projects+using+ic+555+earley.pdf>  
[https://works.spiderworks.co.in/\\_46247648/jtacklek/hchargeu/ninjurew/austin+a55+manual.pdf](https://works.spiderworks.co.in/_46247648/jtacklek/hchargeu/ninjurew/austin+a55+manual.pdf)  
<https://works.spiderworks.co.in/^40428514/xtackleh/kpreventj/ypackd/iutam+symposium+on+combustion+in+superconducting+materials.pdf>