R For Data Science Journal Of Statistical Software

Statistical analysis is common in the social sciences, and among the more popular programs is R. This book provides a foundation for undergraduate and graduate students in the social sciences on how to use R to manage, visualize, and analyze data. The focus is on how to address substantive questions with data analysis and replicate published findings. Using R for Data Analysis in Social Sciences adopts a minimalist approach and covers only the most important functions and skills in R to conduct reproducible research. It emphasizes the practical needs of students using R by showing how to import, inspect, and manage data, understand the logic of statistical inference, visualize data and findings via histograms, boxplots, scatterplots, and diagnostic plots, and analyze data using one-sample t-test, difference-of-means test, covariance, correlation, ordinary least squares (OLS) regression, and model assumption diagnostics. It also demonstrates how to replicate the findings in published journal articles and diagnose model assumption violations. Because the book integrates R programming, the logic and steps of statistical inference, and the process of empirical social scientific research in a highly accessible and structured fashion, it is appropriate for any introductory course on R, data analysis, and empirical social-scientific research.
Instead of presenting the standard theoretical treatments that underlie the various numerical methods used by scientists and engineers, Using R for Numerical Analysis in Science and Engineering shows how to use R and its add-on packages to obtain numerical solutions to the complex mathematical problems commonly faced by scientists and engineers. This practical guide to the capabilities of R demonstrates Monte Carlo, stochastic, deterministic, and other numerical methods through an abundance of worked examples and code, covering the solution of systems of linear algebraic equations and nonlinear equations as well as ordinary differential equations and partial differential equations. It not only shows how to use R's powerful graphic tools to construct the types of plots most useful in scientific and engineering work, but also: Explains how to statistically analyze and fit data to linear and nonlinear models Explores numerical differentiation, integration, and optimization Describes how to find eigenvalues and eigenfunctions Discusses interpolation and curve fitting Considers the analysis of time series Using R for Numerical Analysis in Science and Engineering provides a solid introduction to the most useful numerical methods for scientific and engineering data analysis using R.
Do you want to learn R? This book is built on the premise that anyone with a bit of free time and a healthy curiosity can learn to use R in their studies or at work. The authors focus on using R to do useful things like writing reports, creating data and graphs, accessing datasets collected by others, preparing data, and conducting simple data analysis. In this book you'll learn how to: install R and RStudio®, and set up an RStudio® project and folders; write an essay with graphs based on simple real-world data using R Markdown; create variables from everyday numeric information and visualize data through five types of charts—bar plot, histogram, pie chart, scatter plot, and time series line plot—to identify patterns in the data; write and run R programs, and prepare your data following the tidyverse approach; import external datasets into R, install R data packages, and carry out initial data validity checks; conduct exploratory data analysis through three exercises involving data on voting outcomes, natural resource consumption, and gross domestic product (GDP) via data visualization, correlation coefficient, and simple regression; and write a research paper on the impact of GDP per capita on life expectancy using R Markdown. Student-friendly language and examples (such as binge-watched shows on Netflix, and the top 5 songs on Spotify), cumulative learning, and practice exercises make this a must-have guide for a variety of courses where data are used and reports need to be written. Code and

datasets used to carry out the examples in the book are available on an accompanying website.

An Introduction to Data Science by Jeffrey S. Saltz and Jeffrey M. Stanton is an easy-to-read, gentle introduction for people with a wide range of backgrounds into the world of data science. Needing no prior coding experience or a deep understanding of statistics, this book uses the R programming language and RStudio® platform to make data science welcoming and accessible for all learners. After introducing the basics of data science, the book builds on each previous concept to explain R programming from the ground up. Readers will learn essential skills in data science through demonstrations of how to use data to construct models, predict outcomes, and visualize data.

Key features: Unique in its combination of serving as an introduction to spatial statistics and to modeling agricultural and ecological data using R Provides exercises in each chapter to facilitate the book's use as a course textbook or for self-study Adds new material on generalized additive models, point pattern analysis, and new methods of Bayesian analysis of spatial data. Includes a completely revised chapter on the analysis of spatiotemporal data featuring recently introduced software and methods Updates its coverage of R software including newly introduced packages Spatial Data Analysis in Ecology and Agriculture Using R, 2nd Edition provides practical instruction on the use of the R programming language to analyze spatial data arising from research in ecology, agriculture, and environmental science. Readers have praised the book's practical coverage of spatial statistics, real-world examples, and user-friendly approach in presenting and explaining R code, aspects maintained in this update. Using data sets from cultivated and uncultivated ecosystems, the book guides the reader through the analysis of each data set, including setting research objectives, designing the sampling plan, data quality control, exploratory and confirmatory data analysis, and drawing scientific conclusions. Additional material to accompany the book, on both analyzing satellite data and on multivariate analysis, can be accessed at https://www.plantsciences.ucdavis.edu/plant/additionaltopics.htm.

The 2nd edition of R for Marketing Research and Analytics continues to be the best place to learn R for marketing research. This book is a complete introduction to the power of R for marketing research practitioners. The text describes statistical models from a conceptual point of view with a minimal amount of mathematics, presuming only an introductory knowledge of statistics. Hands-on chapters accelerate the learning curve by asking readers to interact with R from the beginning. Core topics include the R language, basic statistics, linear modeling, and data visualization, which is presented throughout as an integral part of analysis. Later chapters cover more advanced topics yet are intended to be approachable for all analysts. These sections examine logistic regression, customer segmentation, hierarchical linear modeling, market basket analysis, structural equation modeling, and conjoint analysis in R. The text uniquely presents Bayesian models with a minimally complex approach, demonstrating and explaining Bayesian methods alongside traditional analyses for analysis of variance, linear models, and metric and choice-based conjoint analysis. With its emphasis on data visualization, model assessment, and development of statistical intuition, this book provides guidance for any analyst looking to develop or improve skills in R for marketing applications. The 2nd edition increases the book's utility for students and instructors with the inclusion of exercises and classroom slides. At the same time, it retains all of the features that make it a vital resource for practitioners: non-mathematical exposition, examples modeled on real world marketing problems, intuitive guidance on research methods, and immediately applicable code.

Probability and Statistics for Data Science: Math + R + Data covers "math stat"—distributions, expected value, estimation etc.—but takes the phrase "Data Science" in the title quite seriously: * Real datasets are used extensively. * All data analysis is supported by R coding. * Includes many Data Science applications, such as PCA, mixture distributions, random graph models, Hidden Markov models, linear and logistic regression, and neural networks. * Leads the student to think critically about the "how" and "why" of statistics, and to "see the big

picture." * Not "theorem/proof"-oriented, but concepts and models are stated in a mathematically precise manner. Prerequisites are calculus, some matrix algebra, and some experience in programming. Norman Matloff is a professor of computer science at the University of California, Davis, and was formerly a statistics professor there. He is on the editorial boards of the Journal of Statistical Software and The R Journal. His book Statistical Regression and Classification: From Linear Models to Machine Learning was the recipient of the Ziegel Award for the best book reviewed in Technometrics in 2017. He is a recipient of his university's Distinguished Teaching Award.

An invaluable step-by-step, pedagogically engaging guide to data management in R for social science researchers. This book shows students how to recode and document data, as well as how to combine data from different sources, or import from statistical packages other than R.

This book is built on the premise that anyone can learn to use the R software. The authors emphasize using R to do useful things like writing papers and reports, creating data and graphs, accessing datasets collected by others, preparing data, and conducting simple data analysis. After a first chapter on installing the software and project setup, the second chapter shows how to write an essay using R Markdown, rewarding readers with an immediate tangible result, and taking the fear out of working with a new software. After walking readers step by step through data creation, visualization, preparation, access, and exploration across the next four chapters, the book ends on a high note of writing an empirical research paper in R. Student-friendly language and examples (e.g. binge-watched shows on Netflix, top 5 songs on Spotify), cumulative learning and repetition across chapters, and practice exercises make this a must-have guide for a variety of courses where data is used and reports need to be written (including, but not limited to intro statistics and research methods). Code and datasets used to carry out the examples in the book are available on an accompanying website.

The world is becoming increasingly complex, with larger quantities of data available to be analyzed. It so happens that much of these "big data" that are available are spatio-temporal in nature, meaning that they can be indexed by their spatial locations and time stamps. Spatio-Temporal Statistics with R provides an accessible introduction to statistical analysis of spatio-temporal data, with hands-on applications of the statistical methods using R Labs found at the end of each chapter. The book: Gives a step-by-step approach to analyzing spatio-temporal data, starting with visualization, then statistical modelling, with an emphasis on hierarchical statistical models and basis function expansions, and finishing with model evaluation Provides a gradual entry to the methodological aspects of spatio-temporal statistics Provides broad coverage of using R as well as "R Tips" throughout. Features detailed examples and applications in end-of-chapter Labs Features "Technical Notes" throughout to provide additional technical detail where relevant Supplemented by a website featuring the associated R package, data, reviews, errata, a discussion forum, and more The book fills a void in the literature and available software, providing a bridge for students and researchers alike who wish to learn the basics of spatio-temporal statistics. It is written in an informal style and functions as a down-to-earth introduction to the subject. Any reader familiar with calculus-based probability and statistics, and who is comfortable with basic matrix-algebra representations of statistical models, would find this book easy to follow. The goal is to give as many people as possible the tools and confidence to analyze spatio-temporal data.

This textbook provides future data analysts with the tools, methods, and skills needed to answer data-focused, real-life questions; to carry out data analysis; and to visualize and interpret results to support better decisions in business, economics, and public policy. Data wrangling and exploration, regression analysis, machine learning, and causal analysis are comprehensively covered, as well as when, why, and how the methods work, and how they relate to each other. As the most effective way to communicate data analysis, running case studies play a

central role in this textbook. Each case starts with an industry-relevant question and answers it by using real-world data and applying the tools and methods covered in the textbook. Learning is then consolidated by 360 practice questions and 120 data exercises. Extensive online resources, including raw and cleaned data and codes for all analysis in Stata, R, and Python, can be found at www.gabors-data-analysis.com.

Gain sharp insights into your data and solve real-world data science problems with R—from data munging to modeling and visualization About This Book Handle your data with precision and care for optimal business intelligence Restructure and transform your data to inform decision-making Packed with practical advice and tips to help you get to grips with data mining Who This Book Is For If you are a data scientist or R developer who wants to explore and optimize your use of R's advanced features and tools, this is the book for you. A basic knowledge of R is required, along with an understanding of database logic. What You Will Learn Connect to and load data from R's range of powerful databases Successfully fetch and parse structured and unstructured data Transform and restructure your data with efficient R packages Define and build complex statistical models with glm Develop and train machine learning algorithms Visualize social networks and graph data Deploy supervised and unsupervised classification algorithms Discover how to visualize spatial data with R In Detail R is an essential language for sharp and successful data analysis. Its numerous features and ease of use make it a powerful way of mining, managing, and interpreting large sets of data. In a world where understanding big data has become key, by mastering R you will be able to deal with your data effectively and efficiently. This book will give you the guidance you need to build and develop your knowledge and expertise. Bridging the gap between theory and practice, this book will help you to understand and use data for a competitive advantage. Beginning with taking you through essential data mining and management tasks such as munging, fetching, cleaning, and restructuring, the book then explores different model designs and the core components of effective analysis. You will then discover how to optimize your use of machine learning algorithms for classification and recommendation systems beside the traditional and more recent statistical methods. Style and approach Covering the essential tasks and skills within data science, Mastering Data Analysis provides you with solutions to the challenges of data science. Each section gives you a theoretical overview before demonstrating how to put the theory to work with real-world use cases and hands-on examples.

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.

This textbook explains SQL within the context of data science and introduces the different parts of SQL as they are needed for the tasks usually carried out during data analysis. Using the framework of the data life cycle, it focuses on the steps that are very often given the short shift in traditional textbooks, like data loading, cleaning and pre-processing. The book is organized as follows. Chapter 1 describes the data life cycle, i.e. the sequence of stages from data acquisition to archiving, that data goes through as it is prepared and then actually analyzed, together with the different activities that take place at each stage. Chapter 2 gets into databases proper, explaining how relational databases organize data. Non-traditional data, like XML and text, are also covered. Chapter 3 introduces SQL queries, but unlike traditional textbooks, queries and their parts are described around typical data analysis tasks like data exploration, cleaning and transformation. Chapter 4 introduces some basic techniques for data analysis and shows how SQL can be used for some simple analyses without too much complication. Chapter 5 introduces additional SQL constructs that are important in a variety of situations and thus completes the coverage of SQL queries. Lastly, chapter 6 briefly explains how to use SQL from within R and from within Python programs. It focuses on how these languages can interact with a database, and how what has been learned about SQL can be leveraged to make life easier when using R or Python. All chapters contain a lot of examples and exercises on the way, and readers are encouraged to install the two open-source database systems (MySQL and Postgres) that are used throughout the book in order to practice and work on the exercises, because simply reading the book is much less useful than actually using it. This book is for anyone interested in data science and/or databases. It just demands a bit of computer fluency, but no specific background on databases or data analysis. All concepts are introduced intuitively and with a minimum of specialized jargon. After going through this book, readers should be able to profitably learn more about data mining, machine learning, and database management from more advanced textbooks and courses.

Using data from one season of NBA games, Basketball Data Science: With Applications in R is the perfect book for anyone interested in learning and applying data analytics in basketball. Whether assessing the spatial performance of an MBA player's shots or doing an analysis of the impact of high pressure game situations on the probability of scoring, this book discusses a variety of case studies and hands-on examples using a custom R package. The codes are supplied so readers can reproduce the analyses themselves or create their own. Assuming a basic statistical knowledge, Basketball Data Science with R is suitable for students, technicians, coaches, data analysts and applied researchers. Features: · One of the first books to provide statistical and data mining methods for the growing field of analytics in basketball. · Presents tools for modelling graphs and figures to visualize the data. · Includes real world case studies and examples, such as estimations of scoring probability using the Golden State Warriors as a test case. · Provides the source code and data so readers can do their own analyses on NBA teams and players.

Master the art of building analytical models using R About This Book Load, wrangle, and analyze your data using the world's most powerful statistical programming language Build and customize publication-quality visualizations of powerful and stunning R graphs Develop key skills and techniques with R to create and customize data mining algorithms Use R to optimize your trading strategy and build up your own risk management system Discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R Who This Book Is For This course is for data scientist or quantitative analyst who are looking at learning R and take advantage of its powerful analytical design framework. It's a seamless journey in becoming a full-stack R developer. What You Will Learn Describe and visualize the behavior of data and relationships between data Gain a thorough understanding of statistical reasoning and sampling Handle missing data gracefully using multiple imputation Create diverse types of bar charts using the default R functions Familiarize yourself with algorithms

written in R for spatial data mining, text mining, and so on Understand relationships between market factors and their impact on your portfolio Harness the power of R to build machine learning algorithms with real-world data science applications Learn specialized machine learning techniques for text mining, big data, and more In Detail The R learning path created for you has five connected modules, which are a mini-course in their own right. As you complete each one, you'll have gained key skills and be ready for the material in the next module! This course begins by looking at the Data Analysis with R module. This will help you navigate the R environment. You'll gain a thorough understanding of statistical reasoning and sampling. Finally, you'll be able to put best practices into effect to make your job easier and facilitate reproducibility. The second place to explore is R Graphs, which will help you leverage powerful default R graphics and utilize advanced graphics systems such as lattice and ggplot2, the grammar of graphics. You'll learn how to produce, customize, and publish advanced visualizations using this popular and powerful framework. With the third module, Learning Data Mining with R, you will learn how to manipulate data with R using code snippets and be introduced to mining frequent patterns, association, and correlations while working with R programs. The Mastering R for Quantitative Finance module pragmatically introduces both the quantitative finance concepts and their modeling in R, enabling you to build a tailor-made trading system on your own. By the end of the module, you will be well-versed with various financial techniques using R and will be able to place good bets while making financial decisions. Finally, we'll look at the Machine Learning with R module. With this module, you'll discover all the analytical tools you need to gain insights from complex data and learn how to choose the correct algorithm for your specific needs. You'll also learn to apply machine learning methods to deal with common tasks, including classification, prediction, forecasting, and so on. Style and approach Learn data analysis, data visualization techniques, data mining, and machine learning all using R and also learn to build models in quantitative finance using this powerful language.

This book is a refreshingly practical yet theoretically sound roadmap to leveraging data analytics and data science. The vast amount of data generated about us and our world is useless without plans and strategies that are designed to cope with its size and complexity, and which enable organizations to leverage the information to create value in marketing. Creating Value with Data Analytics in Marketing provides a nuanced view of big data developments and data science, arguing that big data is not a revolution but an evolution of the increasing availability of data that has been observed in recent times. Building on the authors' extensive academic and practical knowledge, this book aims to provide managers and analysts with strategic directions and practical analytical solutions on how to create value from existing and new big data. The second edition of this bestselling text has been fully updated in line with developments in the field and includes a selection of new, international cases and examples, exercises, techniques and methodologies. Tying data and analytics to specific goals and processes for implementation makes this essential reading for advanced undergraduate and postgraduate students and specialists of data analytics, marketing research, marketing management and customer relationship management. Online resources include chapter-by-chapter lecture slides and data sets and corresponding R code for selected chapters.

Practical Data Science for Information Professionals provides an accessible introduction to a potentially complex field, providing readers with an overview of data science and a framework for its application. It provides detailed examples and analysis on real data sets to explore the basics of the subject in three principle areas: clustering and social network analysis; predictions and

forecasts; and text analysis and mining. As well as highlighting a wealth of user-friendly data science tools, the book also includes some example code in two of the most popular programming languages (R and Python) to demonstrate the ease with which the information professional can move beyond the graphical user interface and achieve significant analysis with just a few lines of code. After reading, readers will understand: · the growing importance of data science · the role of the information professional in data science · some of the most important tools and methods that information professionals can use. Bringing together the growing importance of data science and the increasing role of information professionals in the management and use of data, Practical Data Science for Information Professionals will provide a practical introduction to the topic specifically designed for the information community. It will appeal to librarians and information professionals all around the world, from large academic libraries to small research libraries. By focusing on the application of open source software, it aims to reduce barriers for readers to use the lessons learned within.

"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"--

This is the first three chapters of a textbook for data scientists who want to improve how they work with, analyze, and extract information from data. The focus of the textbook is how to appropriately apply statistical methods, both simple and sophisticated, to 21st century data and problems. This book contains the first three chapters: Introduction -- Data Science and Statistics, Descriptive Statistics, and Data Visualization -- as well as the book front matter. Subsequent chapters will be published in 3- to 5-chapter sets as they become available.The textbook is intended for current and future data scientists, and for anyone interested in deriving information from data. It requires some mathematical sophistication on the part of the reader, as well as comfort using computers and statistical software.Data science is a new field that has arisen to exploit the proliferation of data in the modern world. Mathematical statistics dates back to the mid-18th century, where the field began as the systematic collection of population and economic data by nations. The modern practice of statistics – which includes the collection, summarization, and analysis of data – dates to the early 20th century. Today statistical methods are widely used by governments, businesses and other organizations, as well as by all scientific disciplines.It has been said that a data scientist must have a better grasp of statistics than the average computer scientist and a better grasp of programming than the average statistician. This book will give data scientists a firm foundation in statistics.

Most of the data that companies collect is related to customer behaviors, such as clicks on a website or purchases in a supermarket. But data science algorithms and predictive analytics tools aren't that specific, so customer data is treated the same way as, for example, astronomical or genomic data. This practical guide introduces powerful methods for behavioral data analysis that you're probably not aware of. Advanced experimental design will help you get the most out of your A/B tests, while causal diagrams will allow you to tease out causality from correlation even when you can't run experiments. Written in an accessible style for data scientists, business analysts, and behavioral scientists, this practical book provides complete examples and exercises in R

and Python to help you gain more insight from your immediately. Understand the specifics of behavioral data Explore the differences between measurement and prediction Learn how to clean and prepare behavioral data Design and analyze experiments to drive optimal business decisions Use behavioral data to understand and measure cause and effect Segment customers in a transparent and insightful way

The only how-to guide offering a unified, systemic approach to acquiring, cleaning, and managing data in R Every experienced practitioner knows that preparing data for modeling is a painstaking, time-consuming process. Adding to the difficulty is that most modelers learn the steps involved in cleaning and managing data piecemeal, often on the fly, or they develop their own ad hoc methods. This book helps simplify their task by providing a unified, systematic approach to acquiring, modeling, manipulating, cleaning, and maintaining data in R. Starting with the very basics, data scientists Samuel E. Buttrey and Lyn R. Whitaker walk readers through the entire process. From what data looks like and what it should look like, they progress through all the steps involved in getting data ready for modeling. They describe best practices for acquiring data from numerous sources; explore key issues in data handling, including text/regular expressions, big data, parallel processing, merging, matching, and checking for duplicates; and outline highly efficient and reliable techniques for documenting data and recordkeeping, including audit trails, getting data back out of R, and more. The only single-source guide to R data and its preparation, it describes best practices for acquiring, manipulating, cleaning, and maintaining data Begins with the basics and walks readers through all the steps necessary to get data ready for the modeling process Provides expert guidance on how to document the processes described so that they are reproducible Written by seasoned professionals, it provides both introductory and advanced techniques Features case studies with supporting data and R code, hosted on a companion website A Data Scientist's Guide to Acquiring, Cleaning and Managing Data in R is a valuable working resource/bench manual for practitioners who collect and analyze data, lab scientists and research associates of all levels of experience, and graduate-level data mining students.

Handbook of Probabilistic Models carefully examines the application of advanced probabilistic models in conventional engineering fields. In this comprehensive handbook, practitioners, researchers and scientists will find detailed explanations of technical concepts, applications of the proposed methods, and the respective scientific approaches needed to solve the problem. This book provides an interdisciplinary approach that creates advanced probabilistic models for engineering fields, ranging from conventional fields of mechanical engineering and civil engineering, to electronics, electrical, earth sciences, climate, agriculture, water resource, mathematical sciences and computer sciences. Specific topics covered include minimax probability machine regression, stochastic finite element method, relevance vector machine, logistic regression, Monte Carlo simulations, random matrix, Gaussian process regression, Kalman filter, stochastic optimization, maximum likelihood, Bayesian inference, Bayesian update, kriging, copula-statistical models, and more. Explains the application of advanced probabilistic models encompassing multidisciplinary research Applies probabilistic modeling to emerging areas in engineering Provides an interdisciplinary approach to probabilistic models and their applications, thus solving a wide range of practical problems

Statistics for Linguists: An Introduction Using R is the first statistics textbook on linear models for linguistics. The book covers simple uses of linear models through generalized models to more advanced approaches, maintaining its focus on conceptual issues and avoiding excessive mathematical details. It contains many applied examples using the R statistical programming environment. Written in an accessible tone and style, this text is the ideal main resource for graduate and advanced undergraduate students of Linguistics statistics courses as well as those in other fields, including Psychology, Cognitive Science, and Data Science.

Over 100 hands-on recipes to effectively solve real-world data problems using the most popular R packages and techniques About This Book Gain insight into how data scientists collect, process, analyze, and visualize data using some of the most popular R packages Understand how to apply useful data analysis techniques in R for real-world applications An easy-to-follow guide to make the life of data scientist easier with the problems faced while performing data analysis Who This Book Is For This book is for those who are already familiar with the basic operation of R, but want to learn how to efficiently and effectively analyze real-world data problems using practical R packages. What You Will Learn Get to know the functional characteristics of R language Extract, transform, and load data from heterogeneous sources Understand how easily R can confront probability and statistics problems Get simple R instructions to quickly organize and manipulate large datasets Create professional data visualizations and interactive reports Predict user purchase behavior by adopting a classification approach Implement data mining techniques to discover items that are frequently purchased together Group similar text documents by using various clustering methods In Detail This cookbook offers a range of data analysis samples in simple and straightforward R code, providing step-by-step resources and time-saving methods to help you solve data problems efficiently. The first section deals with how to create R functions to avoid the unnecessary duplication of code. You will learn how to prepare, process, and perform sophisticated ETL for heterogeneous data sources with R packages. An example of data manipulation is provided, illustrating how to use the "dplyr" and "data.table" packages to efficiently process larger data structures. We also focus on "ggplot2" and show you how to create advanced figures for data exploration. In addition, you will learn how to build an interactive report using the "ggvis" package. Later chapters offer insight into time series analysis on financial data, while there is detailed information on the hot topic of machine learning, including data classification, regression, clustering, association rule mining, and dimension reduction. By the end of this book, you will understand how to resolve issues and will be able to comfortably offer solutions to problems encountered while performing data analysis. Style and approach This easy-to-follow guide is full of hands-on examples of data analysis with R. Each topic is fully explained beginning with the core concept, followed by step-by-step practical examples, and concluding with detailed explanations of each concept used.

Harness the full power of the behavioral data in your company by learning tools specifically designed for behavioral data analysis. Common data science algorithms and predictive analytics tools treat customer behavioral data, such as clicks on a website or purchases in a supermarket, the same as any other data. Instead, this practical guide introduces powerful methods specifically tailored for behavioral data analysis. Advanced experimental design helps you get the most out of your A/B tests, while causal diagrams allow you to tease out the causes of behaviors even when you can't run experiments. Written in an accessible style for data scientists, business analysts, and behavioral scientists, thispractical book provides complete examples and exercises in R and Python to help you gain more insight from your data--immediately. Understand the specifics of behavioral data Explore the differences between measurement and prediction Learn how to

clean and prepare behavioral data Design and analyze experiments to drive optimal business decisions Use behavioral data to understand and measure cause and effect Segment customers in a transparent and insightful way

In this age of information, the manipulation, analysis, and interpretation of data have become a fundamental part of professional life; nowhere more so than in the delivery of healthcare. From the understanding of disease and the development of new treatments, to the diagnosis and management of individual patients, the use of data and technology is now an integral part of the business of healthcare. Those working in healthcare interact daily with data, often without realising it. The conversion of this avalanche of information to useful knowledge is essential for high-quality patient care. R for Health Data Science includes everything a healthcare professional needs to go from R novice to R guru. By the end of this book, you will be taking a sophisticated approach to health data science with beautiful visualisations, elegant tables, and nuanced analyses. Features Provides an introduction to the fundamentals of R for healthcare professionals Highlights the most popular statistical approaches to health data science Written to be as accessible as possible with minimal mathematics Emphasises the importance of truly understanding the underlying data through the use of plots Includes numerous examples that can be adapted for your own data Helps you create publishable documents and collaborate across teams With this book, you are in safe hands – Prof. Harrison is a clinician and Dr. Pius is a data scientist, bringing 25 years' combined experience of using R at the coal face. This content has been taught to hundreds of individuals from a variety of backgrounds, from rank beginners to experts moving to R from other platforms.

"One of the few books that provide an accessible introduction to quantitative data analysis with R. A particular strength of the text is the focus on 'real world' examples which help students to understand why they are learning these methods." - Dr Roxanne Connelly, University of York Relevant, engaging, and packed with student-focused learning features, this book provides the step-by-step introduction to quantitative research and data every student needs. Gradually introducing applied statistics and R, it uses examples from across the social sciences to show you how to apply abstract statistical and methodological principles to your own work. At a student-friendly pace, it enables you to: - Understand and use quantitative data to answer questions - Approach surrounding ethical issues - Collect quantitative data - Manage, write about, and share the data effectively Supported by incredible digital resources with online tutorials, videos, datasets, and multiple choice questions, this book gives you not only the tools you need to understand statistics, quantitative data, and R software, but also the chance to practice and apply what you have learned. Brian J. Fogarty is a Lecturer in Quantitative Social Science on the Glasgow Q-Step Programme in the School of Social and Political Sciences at the University of Glasgow.

This introduction to R for students of psychology and health sciences aims to fast-track the reader through some of the most difficult aspects of learning to do data analysis and statistics. It demonstrates the benefits for reproducibility and reliability of using a programming language over commercial software packages such as SPSS. The early chapters build at a gentle pace, to give the reader confidence in moving from a point-and-click software environment, to the more robust and reliable world of statistical coding. This is a thoroughly modern and up-to-date approach using RStudio and the tidyverse. A range of R packages relevant to psychological research are discussed in detail. A great deal of research in the health sciences concerns questionnaire data, which may require recoding, aggregation and transformation before quantitative techniques and statistical analysis can be applied. R offers many useful and transparent functions to process data and check psychometric properties. These are illustrated in detail, along with a wide range of tools R affords for data visualisation. Many introductory statistics books for the health sciences rely on toy examples - in contrast, this book benefits from utilising open datasets from published psychological studies, to both motivate and demonstrate the transition from data manipulation and analysis to published report. R Markdown is becoming the

preferred method for communicating in the open science community. This book also covers the detail of how to integrate the use of R Markdown documents into the research workflow and how to use these in preparing manuscripts for publication, adhering to the latest APA style guidelines.

This book builds on the previous book by the same authors, An Introduction to R for Spatial Analysis and Mapping, to consider Spatial Data (ie the location attached to data), issues of inference, linking Big Data, Geography / GIS / Mapping and Spatial Analytics. A 'learning by doing' text book, it covers important theoretical issues and helps to develop practical skills in the reader for addressing these.

Data Science in Education Using R is the go-to reference for learning data science in the education field. The book answers questions like: What does a data scientist in education do? How do I get started learning R, the popular open-source statistical programming language? And what does a data analysis project in education look like? If you're just getting started with R in an education job, this is the book you'll want with you. This book gets you started with R by teaching the building blocks of programming that you'll use many times in your career. The book takes a "learn by doing" approach and offers eight analysis walkthroughs that show you a data analysis from start to finish, complete with code for you to practice with. The book finishes with how to get involved in the data science community and how to integrate data science in your education job. This book will be an essential resource for education professionals and researchers looking to increase their data analysis skills as part of their professional and academic development.

Computational statistics and statistical computing are two areas that employ computational, graphical, and numerical approaches to solve statistical problems, making the versatile R language an ideal computing environment for these fields. This second edition continues to encompass the traditional core material of computational statistics, with an

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.

R for Political Data Science: A Practical Guide is a handbook for political scientists new to R who want to learn the most useful and common ways to interpret and analyze political data. It was written by political scientists, thinking about the many real-world problems faced in their work. The book has 16 chapters and is organized in three sections. The first, on the use of R, is for those users who are learning R or are migrating from another software. The second section, on econometric models, covers OLS, binary and survival models, panel data, and causal inference. The third section is a data science toolbox of some the most useful tools in the discipline: data imputation, fuzzy merge of large datasets, web mining, quantitative text analysis, network analysis, mapping, spatial cluster analysis, and principal component analysis. Key features: Each chapter has the most up-to-date and simple option available for each task, assuming minimal prerequisites and no previous experience in R Makes extensive use of the Tidyverse, the group of packages that has revolutionized the use of R Provides a step-by-step guide that you can replicate using your own data Includes exercises in every chapter for course use or self-study Focuses on practical-based approaches to statistical inference rather than mathematical formulae Supplemented by an R package, including all data As the title suggests, this book is highly applied in nature, and is designed as a toolbox for the reader. It can be used in methods and data science courses, at both the undergraduate and graduate levels. It will be equally useful for a university student pursuing a PhD, political consultants, or a public official, all of whom need to transform their datasets into substantive and easily interpretable conclusions.

R for Cloud Computing looks at some of the tasks performed by business analysts on the desktop (PC era) and helps the user navigate the wealth of information in R and its 4000 packages as well as transition the same analytics using the cloud. With this information the reader can select both cloud vendors and the sometimes confusing cloud ecosystem as well as the R packages that can help process the analytical tasks with minimum effort, cost and maximum usefulness and customization. The use of Graphical User Interfaces (GUI) and Step by Step screenshot tutorials is emphasized in this book to lessen the famous learning curve in learning R and some of the needless confusion created in cloud computing that hinders its widespread adoption. This will help you kick-start analytics on the cloud including chapters on both cloud computing, R, common tasks performed in analytics including the current focus and scrutiny of Big Data Analytics, setting up and navigating cloud providers. Readers are exposed to a breadth of cloud computing choices and analytics topics without being buried in needless depth. The included references and links allow the reader to pursue business analytics on the cloud easily. It is aimed at practical analytics and is easy to transition from existing analytical set up to the cloud on an open source system based primarily on R. This book is aimed at industry practitioners with basic programming skills and students who want to enter analytics as a profession. Note the scope of the book is neither statistical theory nor graduate level research for statistics, but rather it is for business analytics practitioners. It will also help researchers and academics but at a practical rather than conceptual level. The R statistical software is the fastest growing

analytics platform in the world, and is established in both academia and corporations for robustness, reliability and accuracy. The cloud computing paradigm is firmly established as the next generation of computing from microprocessors to desktop PCs to cloud.

International Federation of Classification Societies The International Federation of Classification Societies (IFCS) is an agency for the dissemination of technical and scientific information concerning classification and multivariate data analysis in the broad sense and in as wide a range of applications as possible; founded in 1985 in Cambridge (UK) by the following Scientific Societies and Groups: - British Classification Society - BCS - Classification Society of North America - CSNA - Gesellschaft fUr Klassification - GfKI - Japanese Classification Society - JCS - Classification Group ofItalian Statistical Society - CGSIS - Societe Francophone de Classification - SFC Now the IFCS includes also the following Societies: - Dutch-Belgian Classification Society - VOC - Polish Classification Section - SKAD - Portuguese Classification Association - CLAD - Group at Large - Korean Classification Society - KCS IFCS-98, the Sixth Conference of the International Federation of Classification Societies, was held in Rome, from July 21 to 24, 1998. Five preceding conferences were held in Aachen (Germany), Charlottesville (USA), Edinburgh (UK), Paris (France), Kobe (Japan).

Data Science for Business and Decision Making covers both statistics and operations research while most competing textbooks focus on one or the other. As a result, the book more clearly defines the principles of business analytics for those who want to apply quantitative methods in their work. Its emphasis reflects the importance of regression, optimization and simulation for practitioners of business analytics. Each chapter uses a didactic format that is followed by exercises and answers. Freely-accessible datasets enable students and professionals to work with Excel, Stata Statistical Software®, and IBM SPSS Statistics Software®. Combines statistics and operations research modeling to teach the principles of business analytics Written for students who want to apply statistics, optimization and multivariate modeling to gain competitive advantages in business Shows how powerful software packages, such as SPSS and Stata, can create graphical and numerical outputs

Introduction to R for Social Scientists: A Tidy Programming Approach introduces the Tidy approach to programming in R for social science research to help quantitative researchers develop a modern technical toolbox. The Tidy approach is built around consistent syntax, common grammar, and stacked code, which contribute to clear, efficient programming. The authors include hundreds of lines of code to demonstrate a suite of techniques for developing and debugging an efficient social science research workflow. To deepen the dedication to teaching Tidy best practices for conducting social science research in R, the authors include numerous examples using real world data including the American National Election Study and the World Indicators Data. While no prior experience in R is assumed, readers are expected to be acquainted with common social science research designs and terminology. Whether used as a reference manual or read from cover to cover, readers will be equipped with a deeper understanding of R and the Tidyverse, as well as a framework for how best to leverage these powerful tools to write tidy, efficient code for solving problems. To this end, the authors provide many suggestions for additional readings and tools to build on the

concepts covered. They use all covered techniques in their own work as scholars and practitioners.

Harness actionable insights from your data with computational statistics and simulations using R About This Book Learn five different simulation techniques (Monte Carlo, Discrete Event Simulation, System Dynamics, Agent-Based Modeling, and Resampling) in-depth using real-world case studies A unique book that teaches you the essential and fundamental concepts in statistical modeling and simulation Who This Book Is For This book is for users who are familiar with computational methods. If you want to learn about the advanced features of R, including the computer-intense Monte-Carlo methods as well as computational tools for statistical simulation, then this book is for you. Good knowledge of R programming is assumed/required. What You Will Learn The book aims to explore advanced R features to simulate data to extract insights from your data. Get to know the advanced features of R including high-performance computing and advanced data manipulation See random number simulation used to simulate distributions, data sets, and populations Simulate close-to-reality populations as the basis for agent-based micro-, model- and design-based simulations Applications to design statistical solutions with R for solving scientific and real world problems Comprehensive coverage of several R statistical packages like boot, simPop, VIM, data.table, dplyr, parallel, StatDA, simecol, simecolModels, deSolve and many more. In Detail Data Science with R aims to teach you how to begin performing data science tasks by taking advantage of Rs powerful ecosystem of packages. R being the most widely used programming language when used with data science can be a powerful combination to solve complexities involved with varied data sets in the real world. The book will provide a computational and methodological framework for statistical simulation to the users. Through this book, you will get in grips with the software environment R. After getting to know the background of popular methods in the area of computational statistics, you will see some applications in R to better understand the methods as well as gaining experience of working with real-world data and real-world problems. This book helps uncover the large-scale patterns in complex systems where interdependencies and variation are critical. An effective simulation is driven by data generating processes that accurately reflect real physical populations. You will learn how to plan and structure a simulation project to aid in the decision-making process as well as the presentation of results. By the end of this book, you reader will get in touch with the software environment R. After getting background on popular methods in the area, you will see applications in R to better understand the methods as well as to gain experience when working on real-world data and real-world problems. Style and approach This book takes a practical, hands-on approach to explain the statistical computing methods, gives advice on the usage of these methods, and provides computational tools to help you solve common problems in statistical simulation and computer-intense methods.