

## Package Apt R

Ubuntu Server is a complete, free server operating system that just works, with the extra Ubuntu polish, innovation, and simplicity that administrators love. Now, there's a definitive, authoritative guide to getting up and running quickly with the newest, most powerful versions of Ubuntu Server. Written by leading members of the Ubuntu community, *The Official Ubuntu Server Book, Third Edition*, covers all you need to know to make the most of Ubuntu Server, whether you're a beginner or a battle-hardened senior systems administrator. The authors cover Ubuntu Server from start to finish: installation, basic administration and monitoring, security, backup, troubleshooting, system rescue, and much more. They walk through deploying each of the most common server applications, from file and print services to state-of-the-art, cost-saving virtualization and cloud computing. In addition, you'll learn how to Make the most of Ubuntu Server's latest, most powerful technologies Discover easy, fast ways to perform key administration tasks Automate Ubuntu installs, no matter how many servers you're installing Quickly set up low-cost Web servers and e-mail Protect your server with Ubuntu's built-in and optional security tools Minimize downtime with fault tolerance and clustering Master proven, step-by-step server and network troubleshooting techniques Walk through rescuing an Ubuntu server that won't boot Deploy your own Ubuntu servers in the cloud

Make the most of R's dynamic capabilities and implement web applications with Shiny About This Book Present interactive data visualizations in R within the Shiny framework Construct web dashboards in a simple, intuitive, but fully flexible environment Apply your skills to create a real-world web application with this step-by-step guide Who This Book Is For If you are a data scientist who needs a platform to show your results to a broader audience in an attractive and visual way, or a web developer with no prior experience in R or Shiny, this is the book for you. What You Will Learn Comprehend many useful functions, such as lapply and apply, to process data in R Write and structure different files to create a basic dashboard Develop graphics in R using popular graphical libraries such as ggplot2 and GoogleVis Mount a dashboard on a Linux Server Integrate Shiny with non-R-native visualization, such as D3.js Design and build a web application In Detail R is nowadays one of the most used tools in data science. However, along with Shiny, it is also gaining territory in the web application world, due to its simplicity and flexibility. Shiny is a framework that enables the creation of interactive visualizations written entirely in R and can be displayed in almost any ordinary web browser. It is a package from RStudio, which is an IDE for R. From the fundamentals of R to the administration of multi-concurrent, fully customized web applications, this book explains how to achieve your desired web application in an easy and gradual way. You will start by learning about the fundamentals of R, and will move on to looking at simple and practical examples. These examples will enable you to grasp many useful tools that will assist you in solving the usual problems that can be faced when developing data visualizations. You will then walk through the integration of Shiny with R in general and view the different visualization possibilities out there. Finally, you will put your skills to the test and create your first web application! Style and approach This is a comprehensive, step-by-step guide that will allow you to learn and make full use of R and Shiny's capabilities in a gradual way, together with clear, applied examples.

Games provide mathematical models for interaction. Numerous tasks in computer science can be formulated in game-theoretic terms. This fresh and intuitive way of thinking through complex issues reveals underlying algorithmic questions and clarifies the relationships between different domains. This collection of lectures, by specialists in the field, provides an excellent introduction to various aspects of game theory relevant for applications in computer science that concern program design, synthesis, verification, testing and design of multi-agent or distributed systems. Originally devised for a Spring School organised by the GAMES Networking Programme in 2009, these lectures have since been revised and expanded, and range from tutorials concerning fundamental notions and methods to more advanced presentations of current research topics. This volume is a valuable guide to current research on game-based methods in computer science for undergraduate and graduate students. It will also interest researchers working in mathematical logic, computer science and game theory.

Over 100 recipes to get up and running with the modern Linux administration ecosystem Key Features Understand and implement the core system administration tasks in Linux Discover tools and techniques to troubleshoot your Linux system Maintain a healthy system with good security and backup practices Book Description Linux is one of the most widely used operating systems among system administrators, and even modern application and server development is heavily reliant on the Linux platform. The *Linux Administration Cookbook* is your go-to guide to get started on your Linux journey. It will help you understand what that strange little server is doing in the corner of your office, what the mysterious virtual machine languishing in Azure is crunching through, what that circuit-board-like thing is doing under your office TV, and why the LEDs on it are blinking rapidly. This book will get you started with administering Linux, giving you the knowledge and tools you need to troubleshoot day-to-day problems, ranging from a Raspberry Pi to a server in Azure, while giving you a good understanding of the fundamentals of how GNU/Linux works. Through the course of the book, you'll install and configure a system, while the author regales you with errors and anecdotes from his vast experience as a data center hardware engineer, systems administrator, and DevOps consultant. By the end of the book, you will have gained practical knowledge of Linux, which will serve as a bedrock for learning Linux administration and aid you in your Linux journey. What you will learn Install and manage a Linux server, both locally and in the cloud Understand how to perform administration across all Linux distros Work through evolving concepts such as IaaS versus PaaS, containers, and automation Explore security and configuration best practices Troubleshoot your system if something goes wrong Discover and mitigate hardware issues, such as faulty memory and failing drives Who this book is for If you are a system engineer or system administrator with basic experience of working with Linux, this book is for you.

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

R Markdown: The Definitive Guide is the first official book authored by the core R Markdown developers that provides a comprehensive and accurate reference to the R Markdown ecosystem. With R Markdown, you can easily create reproducible data analysis reports, presentations, dashboards, interactive applications, books, dissertations, websites, and journal articles, while enjoying the simplicity of Markdown and the great power of R and other languages. In this book, you will learn Basics: Syntax of Markdown and R code chunks, how to generate figures and tables, and how to use other computing languages Built-in output formats of R Markdown: PDF/HTML/Word/RTF/Markdown documents and ioslides/Slidy/Beamer/PowerPoint presentations Extensions and applications: Dashboards, Tufte handouts, xaringan/reveal.js presentations, websites, books, journal articles, and interactive tutorials Advanced topics: Parameterized reports, HTML widgets, document templates, custom output formats, and Shiny documents. Yihui Xie is a software engineer at RStudio. He has authored and co-authored several R packages, including knitr, rmarkdown, bookdown, blogdown, shiny, xaringan, and animation. He has published three other books, Dynamic Documents with R and knitr, bookdown: Authoring Books and Technical Documents with R Markdown, and blogdown: Creating Websites with R Markdown. J.J. Allaire is the founder of RStudio and the creator of the RStudio IDE. He is an author of several packages in the R Markdown ecosystem including rmarkdown, flexdashboard, learnr, and radix. Garrett Grolemund is the co-author of R for Data Science and author of Hands-On Programming with R. He wrote the lubridate R package and works for RStudio as an advocate who trains engineers to do data science with R and the Tidyverse.

A step-by-step solution-based guide to preparing building, training, and deploying high-quality machine learning models with Amazon SageMaker Key Features Perform ML experiments with built-in and custom algorithms in SageMaker Explore proven solutions when working with TensorFlow, PyTorch, Hugging Face Transformers, and scikit-learn Use the different features and capabilities of SageMaker to automate relevant ML processes Book Description Amazon SageMaker is a fully managed machine learning (ML) service that helps data scientists and ML practitioners manage ML experiments. In this book, you'll use the different capabilities and features of Amazon SageMaker to solve relevant data science and ML problems. This step-by-step guide features 80 proven recipes designed to give you the hands-on machine learning experience needed to contribute to real-world experiments and projects. You'll cover the algorithms and techniques that are commonly used when training and deploying NLP, time series forecasting, and computer vision models to solve ML problems. You'll explore various solutions for working with deep learning libraries and frameworks such as TensorFlow, PyTorch, and Hugging Face Transformers in Amazon SageMaker. You'll also learn how to use SageMaker Clarify, SageMaker Model Monitor, SageMaker Debugger, and SageMaker Experiments to debug, manage, and monitor multiple ML experiments and deployments. Moreover, you'll have a better understanding of how SageMaker Feature Store, Autopilot, and Pipelines can meet the specific needs of data science teams. By the end of this book, you'll be able to combine the different solutions you've learned as building blocks to solve real-world ML problems. What you will learn Train and deploy NLP, time series forecasting, and computer vision models to solve different business problems Push the limits of customization in SageMaker using custom container images Use AutoML capabilities with SageMaker Autopilot to create high-quality models Work with effective data analysis and preparation techniques Explore solutions for debugging and managing ML experiments and deployments Deal with bias detection and ML explainability requirements using SageMaker Clarify Automate intermediate and complex deployments and workflows using a variety of solutions Who this book is for This book is for developers, data scientists, and machine learning practitioners interested in using Amazon SageMaker to build, analyze, and deploy machine learning models with 80 step-by-step recipes. All you need is an AWS account to get things running. Prior knowledge of AWS, machine learning, and the Python programming language will help you to grasp the concepts covered in this book more effectively.

R is a wonderful thing, indeed: in recent years this free, open-source product has become a popular toolkit for statistical analysis and programming. Two of R's limitations -- that it is single-threaded and memory-bound -- become especially troublesome in the current era of large-scale data analysis. It's possible to break past these boundaries by putting R on the parallel path. Parallel R will describe how to give R parallel muscle. Coverage will include stalwarts such as snow and multicore, and also newer techniques such as Hadoop and Amazon's cloud computing platform.

Europe is getting closer. So are European social sciences. However, this is easier done in theory development and central research questions. When it comes to data the mutual understanding is far from perfect, due to a lack of knowledge about the data bases of the respective countries and the EU in general. This is particularly true when it comes to the regional level. This volume will help to improve the insight into the rich stock of European datasets which cover any kind of regional information. Many institutions ranging from statistical offices to more academic research centres and commercial enterprises report their offerings with special emphasis on the regional level (e.g. European Community Household Panel, European Social Survey, Labour Force Survey). Central categories such as NUTS and LAU are explained and discussed. In addition, typical examples of socio-economic cross-border and multi-level studies highlight the power of a regionalized European perspective. Furthermore, information about special tools for such type of analysis is included in the volume.

This introduction to both structured programming and object-oriented programming using the C++ language and the Turbo C++ compiler starts from the beginning, assuming no previous knowledge of any programming language. Covers topics such as getting acquainted with computers, programs and Turbo C++, writing simple C++ programs, and includes an introduction to OOP.

Hacking with Kali introduces you the most current distribution of the de facto standard tool for Linux pen testing. Starting with use of the Kali live CD and progressing through installation on hard drives, thumb drives and SD cards, author James Broad walks you through creating a custom version of the Kali live distribution. You'll learn how to configure networking components, storage devices and system services such as DHCP and web services. Once you're familiar with the basic components of the software, you'll learn how to use Kali through the phases of the penetration testing lifecycle; one major tool from each phase is explained. The book culminates with a chapter on reporting that will provide examples of documents used prior to, during and after the pen test. This guide will benefit information security professionals of all levels, hackers, systems administrators, network administrators, and beginning and intermediate professional pen testers, as well as students majoring in information security. Provides detailed explanations of the complete penetration testing lifecycle Complete linkage of the Kali information, resources

and distribution downloads Hands-on exercises reinforce topics

If you have always wanted to learn Linux but are still afraid to do so, this book is for you! A lot of people think of Linux as a sophisticated operating system that only hackers and geeks know how to use, and thus they abort their dream of learning Linux. Well, let me surprise you! Linux is simple and easy to learn, and this book is the ultimate proof! You may have stumbled across a variety of sources that all explain Linux in a complicated and dry manner. This book does exactly the opposite; it teaches you Linux in a delightful and friendly way so that you will never get bored, and you will always feel motivated to learn more. Learn Linux Quickly doesn't assume any prior Linux knowledge, which makes it a perfect fit for beginners. Nevertheless, intermediate and advanced Linux users will still find this book very useful as it goes through a wide range of topics. Learn Linux Quickly will teach you the following topics: · Installing Linux · Over 116 Linux Commands · User and Group Management · Linux Networking Fundamentals · Bash Scripting · Automate Boring Tasks with Cron Jobs · Create your Own Linux Commands · Linux Disk Partitioning and LVM · Finding Files on Linux · Understanding File Permissions · Linux Processes And much more! There is no time to waste here! Learn Linux Quickly and kick start your Linux career today!

This practical guide explains you to program and understand the power of Apache Cassandra 3.x. You will explore the integration and interaction of Cassandra components, and explore features such as the token allocation algorithm, CQL3, vnodes, lightweight transactions, and data modelling in detail.

From the Reviews "[This book] contains an excellent blend of both Shiny-specific topics ... and practical advice from software development that fits in nicely with Shiny apps. You will find many nuggets of wisdom sprinkled throughout these chapters...." Eric Nantz, Host of the R-Podcast and the Shiny Developer Series (from the Foreword) "[This] book is a gradual and pleasant invitation to the production-ready shiny apps world. It ...exposes a comprehensive and robust workflow powered by the {golem} package. [It] fills the not yet covered gap between shiny app development and deployment in such a thrilling way that it may be read in one sitting.... In the industry world, where processes robustness is a key toward productivity, this book will indubitably have a tremendous impact." David Granjon, Sr. Expert Data Science, Novartis Presented in full color, Engineering Production-Grade Shiny Apps helps people build production-grade shiny applications, by providing advice, tools, and a methodology to work on web applications with R. This book starts with an overview of the challenges which arise from any big web application project: organizing work, thinking about the user interface, the challenges of teamwork and the production environment. Then, it moves to a step-by-step methodology that goes from the idea to the end application. Each part of this process will cover in detail a series of tools and methods to use while building production-ready shiny applications. Finally, the book will end with a series of approaches and advice about optimizations for production. Features Focused on practical matters: This book does not cover Shiny concepts, but practical tools and methodologies to use for production. Based on experience: This book is a formalization of several years of experience building Shiny applications. Original content: This book presents new methodologies and tooling, not just a review of what already exists. Engineering Production-Grade Shiny Apps covers medium to advanced content about Shiny, so it will help people that are already familiar with building apps with Shiny, and who want to go one step further.

Predictive Soil Mapping (PSM) is based on applying statistical and/or machine learning techniques to fit models for the purpose of producing spatial and/or spatiotemporal predictions of soil variables i.e. maps of soil properties and classes at different resolutions. It is a multidisciplinary field combining statistics, data science, soil science, physical geography, remote sensing, geoinformation science and a number of other sciences. Predictive Soil Mapping with R is about understanding the main concepts behind soil mapping, mastering R packages that can be used to produce high quality soil maps, and about optimizing all processes involved so that also the production costs can be reduced. The online version of the book is available at: <https://envirometrix.github.io/PredictiveSoilMapping/> Pull requests and general comments are welcome. These materials are based on technical tutorials initially developed by the ISRIC's Global Soil Information Facilities (GSIF) development team over the period 2014-2017

This document is designed to be a resource for those Linux users wishing to seek clarification on Linux/UNIX/POSIX related terms and jargon. At approximately 24000 definitions and two thousand pages it is one of the largest Linux related dictionaries currently available. Due to the rapid rate at which new terms are being created it has been decided that this will be an active project. We welcome input into the content of this document. At this moment in time half yearly updates are being envisaged. Please note that if you wish to find a 'Computer Dictionary' then see the 'Computer Dictionary Project' at <http://computerdictionary.tsf.org.za/> Searchable databases exist at locations such as: <http://www.swpearl.com/eng/scripts/dictionary/> (SWP) Sun Wah-PearL Linux Training and Development Centre is a centre of the Hong Kong Polytechnic University, established in 2000. Presently SWP is delivering professional grade Linux and related Open Source Software (OSS) technology training and consultant service in Hong Kong. SWP has an ambitious aim to promote the use of Linux and related Open Source Software (OSS) and Standards. The vendor independent positioning of SWP has been very well perceived by the market. Throughout the last couple of years, SWP becomes the Top Leading OSS training and service provider in Hong Kong. <http://www.geona.com/dictionary?b=> Geona, operated by Gold Vision Communications, is a new powerful search engine and internet directory, delivering quick and relevant results on almost any topic or subject you can imagine. The term "Geona" is an Italian and Hebrew name, meaning wisdom, exaltation, pride or majesty. We use our own database of spidered web sites and the Open Directory database, the same database which powers the core directory services for the Web's largest and most popular search engines and portals. Geona is spidering all domains listed in the non-adult part of the Open Directory and millions of additional sites of general interest to maintain a fulltext index of highly relevant web sites. <http://www.linuxdig.com/documents/dictionary.php> LINUXDIG.COM, "Yours News and Resource Site", LinuxDig.com was started in May 2001 as a hobby site with the original intention of getting the RFC's online and becoming an Open Source software link/download site. But since that time the site has evolved to become a RFC distribution site, linux news site and a locally written technology news site (with bad grammer :) with focus on Linux while also containing articles about anything and everything we find interesting in the computer world. LinuxDig.Com contains about 20,000 documents and this number is growing everyday! <http://linux.about.com/library/glossary/blglossary.htm> Each month more than 20 million people visit About.com.

Whether it be home repair and decorating ideas, recipes, movie trailers, or car buying tips, our Guides offer practical advice and solutions for every day life. Wherever you land on the new About.com, you'll find other content that is relevant to your interests. If you're looking for "How To" advice on planning to re-finish your deck, we'll also show you the tools you need to get the

job done. If you've been to About before, we'll show you the latest updates, so you don't see the same thing twice. No matter where you are on About.com, or how you got here, you'll always find content that is relevant to your needs. Should you wish to possess your own localised searcheable version please make use of the available "dict", <http://www.dict.org/> version at the Linux Documentation Project home page, <http://www.tldp.org/> The author has decided to leave it up to readers to determine how to install and run it on their specific systems. An alternative form of the dictionary is available at: <http://elibrary.fultus.com/covers/technical/linux/guides/Linux-Dictionary/cover.html> Fultus Corporation helps writers and companies to publish, promote, market, and sell books and eBooks. Fultus combines traditional self-publishing practices with modern technology to produce paperback and hardcover print-on-demand (POD) books and electronic books (eBooks). Fultus publishes works (fiction, non-fiction, science fiction, mystery, ...) by both published and unpublished authors. We enable you to self-publish easily and cost-effectively, creating your book as a print-ready paperback or hardcover POD book or as an electronic book (eBook) in multiple eBook's formats. You retain all rights to your work. We provide distribution to bookstores worldwide. And all at a fraction of the cost of traditional publishing. We also offer corporate publishing solutions that enable businesses to produce and deliver manuals and documentation more efficiently and economically. Our use of electronic delivery and print-on-demand technologies reduces printed inventory and saves time. Please inform the author as to whether you would like to create a database or an alternative form of the dictionary so that he can include you in this list. Also note that the author considers breaches of copyright to be extremely serious. He will pursue all claims to the fullest extent of the law.

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 tidy

This book is designed as an Ubuntu 20.04 LTS Server administration and reference source, covering the Ubuntu servers and their support applications. Server tools are covered as well as the underlying configuration files and system implementations. The emphasis is on what administrators will need to know to perform key server support and management tasks. Coverage of the systemd service management system is integrated into the book. Topics covered include software management, systemd service management, systemd-networkd and Netplan network configuration, AppArmor security, OpenSSH, the Chrony time server, and Ubuntu cloud services. Key servers are examined, including Web, FTP, CUPS printing, NFS, and Samba Windows shares. Network support servers and applications covered include the Squid proxy server, the Domain Name System (BIND) server, DHCP, distributed network file systems, IPtables firewalls, and cloud computing.

An R Companion to Applied Regression is a broad introduction to the R statistical computing environment in the context of applied regression analysis. John Fox and Sanford Weisberg provide a step-by-step guide to using the free statistical software R, an emphasis on integrating statistical computing in R with the practice of data analysis, coverage of generalized linear models, and substantial web-based support materials. The Third Edition has been reorganized and includes a new chapter on mixed-effects models, new and updated data sets, and a de-emphasis on statistical programming, while retaining a general introduction to basic R programming. The authors have substantially updated both the car and effects packages for R for this edition, introducing additional capabilities and making the software more consistent and easier to use. They also advocate an everyday data-analysis workflow that encourages reproducible research. To this end, they provide coverage of RStudio, an interactive development environment for R that allows readers to organize and document their work in a simple and intuitive fashion, and then easily share their results with others. Also included is coverage of R Markdown, showing how to create documents that mix R commands with explanatory text.

This comprehensive and richly illustrated volume provides up-to-date material on Singular Spectrum Analysis (SSA). SSA is a well-known methodology for the analysis and forecasting of time series. Since quite recently, SSA is also being used to analyze digital images and other objects that are not necessarily of planar or rectangular form and may contain gaps. SSA is multi-purpose and naturally combines both model-free and parametric techniques, which makes it a very special and attractive methodology for solving a wide range of problems arising in diverse areas, most notably those associated with time series and digital images. An effective, comfortable and accessible implementation of SSA is provided by the R-package Rssa, which is available from CRAN and reviewed in this book. Written by prominent statisticians who have extensive experience with SSA, the book (a) presents the up-to-date SSA methodology, including multidimensional extensions, in language accessible to a large circle of users, (b) combines different versions of SSA into a single tool, (c) shows the diverse tasks that SSA can be used for, (d) formally describes the main SSA methods and algorithms, and (e) provides tutorials on the Rssa package and the use of SSA. The book offers a valuable resource for a very wide readership, including professional statisticians, specialists in signal and image processing, as well as specialists in numerous applied disciplines interested in using statistical methods for time series analysis, forecasting, signal and image processing. The book is written on a level accessible to a broad audience and includes a wealth of examples; hence it can also be used as a textbook for undergraduate and postgraduate courses on time series analysis and signal processing.

Machine learning has finally come of age. With H2O software, you can perform machine learning and data analysis using a simple open source framework that's easy to use, has a wide range of OS and language support, and scales for big data. This hands-on guide teaches you how to use H2O with only minimal math and theory behind the learning algorithms. If you're familiar with R or Python, know a bit of statistics, and have some experience manipulating data, author Darren Cook will take you through H2O basics and help you conduct machine-learning experiments on different sample data sets. You'll explore several modern machine-learning techniques such as deep learning, random forests, unsupervised learning, and ensemble learning. Learn how to import, manipulate, and export data with H2O Explore key machine-learning concepts, such as cross-validation and validation data sets Work with three diverse data sets, including a regression, a multinomial classification, and a binomial classification Use H2O to analyze each sample data set with four supervised machine-learning algorithms Understand how cluster analysis and other unsupervised machine-learning algorithms work

This book is designed as an Ubuntu 21.04 Server administration and reference source, covering the Ubuntu servers and their support applications. Server tools are covered as well as the underlying configuration files and system implementations. The emphasis is on what administrators will need to know to perform key server support and management tasks. Coverage of the systemd service management system is integrated into the book. Topics covered include software management, systemd service management, AppArmor security, OpenSSH, the Chrony time server, and Ubuntu cloud services. Key servers are examined, including Web, FTP, CUPS printing, NFS, and Samba Windows shares. Network support servers and applications covered include the Squid proxy server, the Domain Name System (BIND) server, DHCP, distributed network file systems, IPtables firewalls, and cloud computing.

This hands-on guide demonstrates how the flexibility of the command line can help you become a more efficient and productive data scientist. You'll learn how to combine small, yet powerful, command-line tools to quickly obtain, scrub, explore, and model your data. To get you started—whether you're on Windows, OS X, or Linux—author Jeroen Janssens introduces the Data Science Toolbox, an easy-to-install virtual environment packed with over 80 command-line tools. Discover why the command line is an agile, scalable, and extensible technology. Even if you're already comfortable processing data with, say, Python or R, you'll greatly improve your data science workflow by also leveraging the power of the command line. Obtain data from websites, APIs, databases, and spreadsheets Perform scrub operations on plain text, CSV, HTML/XML, and JSON Explore data, compute descriptive statistics, and create visualizations Manage your data science workflow using Drake Create reusable tools from one-liners and existing Python or R code Parallelize and distribute data-intensive pipelines using GNU Parallel Model data with dimensionality reduction, clustering, regression, and classification algorithms

This book is intended for developers who have worked with the Raspberry Pi and who want to learn how to make the most of the Raspbian operating system and their Raspberry Pi. Whether you are a beginner to the Raspberry Pi or a seasoned expert, this book will make you familiar with the Raspbian operating system and teach you how to get your Raspberry Pi up and running.

\* An indispensable resource for Fedora users who must now work without customer support from Red Hat, Inc., covering critical troubleshooting techniques for networks, internal servers, and external servers \* Chris Negus is a well-known Linux authority and also the author of the top-selling Red Hat Linux Bible (0-7645-4333-4); Thomas Weeks is a trainer and administrator who manages hundreds of Red Hat Linux systems \* Covers all of the most common Fedora problem areas: firewalls, DNS servers, print servers, Samba, NFS, Web servers, FTP servers, e-mail servers, modems, adding hardware, and hardware certification \* Features easy-to-use flowcharts that guide administrators step by step through common Fedora troubleshooting scenarios \* A companion Web site offers troubleshooting updates to keep pace with the frequent Fedora Core releases as well as a forum for exchanging troubleshooting tips

Rcpp is the glue that binds the power and versatility of R with the speed and efficiency of C++. With Rcpp, the transfer of data between R and C++ is nearly seamless, and high-performance statistical computing is finally accessible to most R users. Rcpp should be part of every statistician's toolbox. -- Michael Braun, MIT Sloan School of Management "Seamless R and C++ integration with Rcpp" is simply a wonderful book. For anyone who uses C/C++ and R, it is an indispensable resource. The writing is outstanding. A huge bonus is the section on applications. This section covers the matrix packages Armadillo and Eigen and the GNU Scientific Library as well as RInside which enables you to use R inside C++. These applications are what most of us need to know to really do scientific programming with R and C++. I love this book. -- Robert McCulloch, University of Chicago Booth School of Business Rcpp is now considered an essential package for anybody doing serious computational research using R. Dirk's book is an excellent companion and takes the reader from a gentle introduction to more advanced applications via numerous examples and efficiency enhancing gems. The book is packed with all you might have ever wanted to know about Rcpp, its cousins (RcppArmadillo, RcppEigen .etc.), modules, package development and sugar. Overall, this book is a must-have on your shelf. -- Sanjog Misra, UCLA Anderson School of Management The Rcpp package represents a major leap forward for scientific computations with R. With very few lines of C++ code, one has R's data structures readily at hand for further computations in C++. Hence, high-level numerical programming can be made in C++ almost as easily as in R, but often with a substantial speed gain. Dirk is a crucial person in these developments, and his book takes the reader from the first fragile steps on to using the full Rcpp machinery. A very recommended book! -- Søren Højsgaard, Department of Mathematical Sciences, Aalborg University, Denmark "Seamless R and C ++ Integration with Rcpp" provides the first comprehensive introduction to Rcpp. Rcpp has become the most widely-used language extension for R, and is deployed by over one-hundred different CRAN and BioConductor packages. Rcpp permits users to pass scalars, vectors, matrices, list or entire R objects back and forth between R and C++ with ease. This brings the depth of the R analysis

framework together with the power, speed, and efficiency of C++. Dirk Eddelbuettel has been a contributor to CRAN for over a decade and maintains around twenty packages. He is the Debian/Ubuntu maintainer for R and other quantitative software, edits the CRAN Task Views for Finance and High-Performance Computing, is a co-founder of the annual R/Finance conference, and an editor of the Journal of Statistical Software. He holds a Ph.D. in Mathematical Economics from EHESS (Paris), and works in Chicago as a Senior Quantitative Analyst.

Advance your skills in efficient data analysis and data processing using the powerful tools of Scala, Spark, and Hadoop About This Book This is a primer on functional-programming-style techniques to help you efficiently process and analyze all of your data Get acquainted with the best and newest tools available such as Scala, Spark, Parquet and MLlib for machine learning Learn the best practices to incorporate new Big Data machine learning in your data-driven enterprise to gain future scalability and maintainability Who This Book Is For Mastering Scala Machine Learning is intended for enthusiasts who want to plunge into the new pool of emerging techniques for machine learning. Some familiarity with standard statistical techniques is required. What You Will Learn Sharpen your functional programming skills in Scala using REPL Apply standard and advanced machine learning techniques using Scala Get acquainted with Big Data technologies and grasp why we need a functional approach to Big Data Discover new data structures, algorithms, approaches, and habits that will allow you to work effectively with large amounts of data Understand the principles of supervised and unsupervised learning in machine learning Work with unstructured data and serialize it using Kryo, Protobuf, Avro, and AvroParquet Construct reliable and robust data pipelines and manage data in a data-driven enterprise Implement scalable model monitoring and alerts with Scala In Detail Since the advent of object-oriented programming, new technologies related to Big Data are constantly popping up on the market. One such technology is Scala, which is considered to be a successor to Java in the area of Big Data by many, like Java was to C/C++ in the area of distributed programming. This book aims to take your knowledge to next level and help you impart that knowledge to build advanced applications such as social media mining, intelligent news portals, and more. After a quick refresher on functional programming concepts using REPL, you will see some practical examples of setting up the development environment and tinkering with data. We will then explore working with Spark and MLlib using k-means and decision trees. Most of the data that we produce today is unstructured and raw, and you will learn to tackle this type of data with advanced topics such as regression, classification, integration, and working with graph algorithms. Finally, you will discover at how to use Scala to perform complex concept analysis, to monitor model performance, and to build a model repository. By the end of this book, you will have gained expertise in performing Scala machine learning and will be able to build complex machine learning projects using Scala. Style and approach This hands-on guide dives straight into implementing Scala for machine learning without delving much into mathematical proofs or validations. There are ample code examples and tricks that will help you sail through using the standard techniques and libraries. This book provides practical examples from the field on how to correctly tackle data analysis problems, particularly for modern Big Data datasets.

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 Julia is a well-constructed programming language with fast execution speed, eliminating the classic problem of performing analysis in one language and translating it for performance into a second. This book will help you develop and enhance your programming skills in Julia to solve real-world automation challenges. This book starts off with a refresher on installing and running Julia on different platforms. Next, you will compare the different ways of working with Julia and explore Julia's key features in-depth by looking at design and build. You will see how data works using simple statistics and analytics, and discover Julia's speed, its real strength, which makes it particularly useful in highly intensive computing tasks and observe how Julia can cooperate with external processes in order to enhance graphics and data visualization. Finally, you will look into meta-programming and learn how it adds great power to the language and establish networking and distributed computing with Julia.

The statistics profession is at a unique point in history. The need for valid statistical tools is greater than ever; data sets are massive, often measuring hundreds of thousands of measurements for a single subject. The field is ready to move towards clear objective benchmarks under which tools can be evaluated. Targeted learning allows (1) the full generalization and utilization of cross-validation as an estimator selection tool so that the subjective choices made by humans are now made by the machine, and (2) targeting

the fitting of the probability distribution of the data toward the target parameter representing the scientific question of interest. This book is aimed at both statisticians and applied researchers interested in causal inference and general effect estimation for observational and experimental data. Part I is an accessible introduction to super learning and the targeted maximum likelihood estimator, including related concepts necessary to understand and apply these methods. Parts II-IX handle complex data structures and topics applied researchers will immediately recognize from their own research, including time-to-event outcomes, direct and indirect effects, positivity violations, case-control studies, censored data, longitudinal data, and genomic studies.

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.

Geocomputation with R is for people who want to analyze, visualize and model geographic data with open source software. It is based on R, a statistical programming language that has powerful data processing, visualization, and geospatial capabilities. The book equips you with the knowledge and skills to tackle a wide range of issues manifested in geographic data, including those with scientific, societal, and environmental implications. This book will interest people from many backgrounds, especially Geographic Information Systems (GIS) users interested in applying their domain-specific knowledge in a powerful open source language for data science, and R users interested in extending their skills to handle spatial data. The book is divided into three parts: (I) Foundations, aimed at getting you up-to-speed with geographic data in R, (II) extensions, which covers advanced techniques, and (III) applications to real-world problems. The chapters cover progressively more advanced topics, with early chapters providing strong foundations on which the later chapters build. Part I describes the nature of spatial datasets in R and methods for manipulating them. It also covers geographic data import/export and transforming coordinate reference systems. Part II represents methods that build on these foundations. It covers advanced map making (including web mapping), "bridges" to GIS, sharing reproducible code, and how to do cross-validation in the presence of spatial autocorrelation. Part III applies the knowledge gained to tackle real-world problems, including representing and modeling transport systems, finding optimal locations for stores or services, and ecological modeling. Exercises at the end of each chapter give you the skills needed to tackle a range of geospatial problems. Solutions for each chapter and supplementary materials providing extended examples are available at <https://geocompr.github.io/geocompkg/articles/>. Dr. Robin Lovelace is a University Academic Fellow at the University of Leeds, where he has taught R for geographic research over many years, with a focus on transport systems. Dr. Jakub Nowosad is an Assistant Professor in the Department of Geoinformation at the Adam Mickiewicz University in Poznan, where his focus is on the analysis of large datasets to understand environmental processes. Dr. Jannes Muenchow is a Postdoctoral Researcher in the GIScience Department at the University of Jena, where he develops and teaches a range of geographic methods, with a focus on ecological modeling, statistical geocomputing, and predictive mapping. All three are active developers and work on a number of R packages, including stplanr, sabre, and RQGIS.

Geocomputation with RCRC Press

This book provides a general introduction to the R Commander graphical user interface (GUI) to R for readers who are unfamiliar with R. It is suitable for use as a supplementary text in a basic or intermediate-level statistics course. It is not intended to replace a basic or other statistics text but rather to complement it, although it does promote sound statistical practice in the examples. The book should also be useful to individual casual or occasional users of R for whom the standard command-line interface is an obstacle. [tinyurl.com/RcmdrBook](http://tinyurl.com/RcmdrBook) The site includes data files used in the book and an errata list. <http://socserv.mcmaster.ca/jfox/Books/RCommander/Writing-Rcmdr-Plugins.pdf> Writing R Commander Plug-in Packages

Over insightful 90 recipes to get lightning-fast analytics with Apache Spark About This Book Use Apache Spark for data processing with these hands-on recipes Implement end-to-end, large-scale data analysis better than ever before Work with powerful libraries such as MLLib, SciPy, NumPy, and Pandas to gain insights from your data Who This Book Is For This book is for novice and intermediate level data science professionals and data analysts who want to solve data science problems with a distributed computing framework. Basic experience with data science implementation tasks is expected. Data science professionals looking to skill up and gain an edge in the field will find this book helpful. What You Will Learn Explore the topics of data mining, text mining, Natural Language Processing, information retrieval, and machine learning. Solve real-world analytical problems with large data sets. Address data science challenges with analytical tools on a distributed system like Spark (apt for iterative algorithms), which offers in-memory processing and more flexibility for data analysis at scale. Get hands-on experience with algorithms like Classification, regression, and recommendation on real datasets using Spark MLLib package. Learn about numerical and scientific computing using NumPy and SciPy on Spark. Use Predictive Model Markup Language (PMML) in Spark for statistical data mining models. In Detail Spark has emerged as the most promising big data analytics engine for data science professionals. The true power and value of Apache Spark lies in its ability to execute data science tasks with speed and accuracy. Spark's selling point is that it combines ETL, batch analytics, real-time stream analysis, machine learning, graph processing, and visualizations. It lets you tackle the complexities that come with raw unstructured data sets with ease. This guide will get you comfortable and confident performing data science tasks with Spark. You will learn about implementations including distributed deep learning, numerical computing, and scalable machine learning. You will be shown effective solutions to problematic concepts in data science using Spark's data science libraries such as MLLib, Pandas, NumPy,

SciPy, and more. These simple and efficient recipes will show you how to implement algorithms and optimize your work. Style and approach This book contains a comprehensive range of recipes designed to help you learn the fundamentals and tackle the difficulties of data science. This book outlines practical steps to produce powerful insights into Big Data through a recipe-based approach.

Make life at the office easier for server administrators by helping them build resilient Ubuntu server systems About This Book Tackle the issues you come across in keeping your Ubuntu server up and running Build server machines and troubleshoot cloud computing related issues using Open Stack Discover tips and best practices to be followed for minimum maintenance of Ubuntu Server 3 Who This Book Is For This book is for a vast audience of Linux system administrators who primarily work on Debian-based systems and spend long hours trying fix issues with the enterprise server. Ubuntu is already one of the most popular OSes and this book targets the most common issues that most administrators have to deal with. With the right tools and definite solutions, you will be able to keep your Ubuntu servers in the pink of health. What You Will Learn Deploy packages and their dependencies with repositories Set up your own DNS and network for Ubuntu Server Authenticate and validate users and their access to various systems and services Maintain, monitor, and optimize your server resources and avoid tremendous load Get to know about processes, assigning and changing priorities, and running processes in background Optimize your shell with tools and provide users with an improved shell experience Set up separate environments for various services and run them safely in isolation Understand, build, and deploy OpenStack on your Ubuntu Server In Detail Ubuntu is becoming one of the favorite Linux flavors for many enterprises and is being adopted to a large extent. It supports a wide variety of common network systems and the use of standard Internet services including file serving, e-mail, Web, DNS, and database management. A large scale use and implementation of Ubuntu on servers has given rise to a vast army of Linux administrators who battle it out day in and day out to make sure the systems are in the right frame of operation and pre-empt any untoward incidents that may result in catastrophes for the businesses using it. Despite all these efforts, glitches and bugs occur that affect Ubuntu server's network, memory, application, and hardware and also generate cloud computing related issues using OpenStack. This book will help you end to end. Right from setting up your new Ubuntu Server to learning the best practices to host OpenStack without any hassles. You will be able to control the priority of jobs, restrict or allow access users to certain services, deploy packages, tackle issues related to server effectively, and reduce downtime. Also, you will learn to set up OpenStack, and manage and monitor its services while tuning the machine with best practices. You will also get to know about Virtualization to make services serve users better. Chapter by chapter, you will learn to add new features and functionalities and make your Ubuntu server a full-fledged, production-ready system. Style and approach This book contains topic-by-topic discussion in an easy-to-understand language with loads of examples to help you take care of Ubuntu Server. Plenty of screenshots will guide you through a step-by-step approach.

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.

Bioinformatics derives knowledge from computer analysis of biological data. In particular, genomic and transcriptomic datasets are processed, analysed and, whenever possible, associated with experimental results from various sources, to draw structural, organizational, and functional information relevant to biology. Research in bioinformatics includes method development for storage, retrieval, and analysis of the data. Bioinformatics in Aquaculture provides the most up to date reviews of next generation sequencing technologies, their applications in aquaculture, and principles and methodologies for the analysis of genomic and transcriptomic large datasets using bioinformatic methods, algorithm, and databases. The book is unique in providing guidance for the best software packages suitable for various analysis, providing detailed examples of using bioinformatic software and command lines in the context of real world experiments. This book is a vital tool for all those working in genomics, molecular biology, biochemistry and genetics related to aquaculture, and computational and biological sciences.

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.

[Copyright: 4293042c021bd72e72564336da5fc09c](https://www.copyright.com/0-90-4293042c021bd72e72564336da5fc09c/)