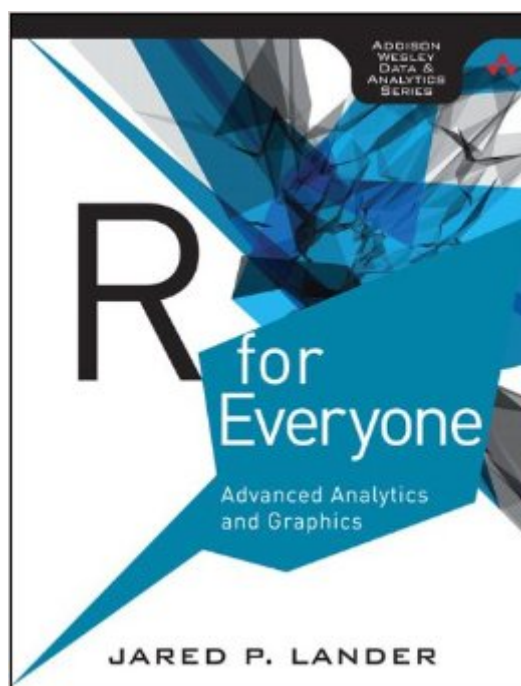# R For Everyone: Advanced Analytics And Graphics (Addison-Wesley Data & Analytics Series)

Synopsis

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

   Using the open source R language, you can build powerful statistical models to answer many of your most challenging questions. R has traditionally been difficult for non-statisticians to learn, and most R books assume far too much knowledge to be of help. R for Everyone is the solution. Drawing on his unsurpassed experience teaching new users, professional data scientist Jared P. Lander has written the perfect tutorial for anyone new to statistical programming and modeling. Organized to make learning easy and intuitive, this guide focuses on the 20 percent of R functionality youâ ™ll need to accomplish 80 percent of modern data tasks. Landerâ ™s self-contained chapters start with the absolute basics, offering extensive hands-on practice and sample code. Youâ ™ll download and install R; navigate and use the R environment; master basic program control, data import, and manipulation; and walk through several essential tests. Then, building on this foundation, youâ ™ll construct several complete models, both linear and nonlinear, and use some data mining techniques. By the time youâ ™re done, you wonâ ™t just know how to write R programs, youâ ™ll be ready to tackle the statistical problems you care about most. Â COVERAGE INCLUDES â ¢ Exploring R, RStudio, and R packages â ¢ Using R for math: variable types, vectors, calling functions, and more â ¢ Exploiting data structures, including data.frames, matrices, and lists â ¢ Creating attractive, intuitive statistical graphics â ¢ Writing user-defined functions â ¢ Controlling program flow with if, ifelse, and complex checks â ¢ Improving program efficiency with group manipulations â ¢ Combining and reshaping multiple datasets â ¢ Manipulating strings using Râ ™s facilities and regular expressions â ¢ Creating normal, binomial, and Poisson probability distributions â ¢ Programming basic statistics: mean, standard deviation, and t-tests â ¢ Building linear, generalized linear, and nonlinear models â ¢ Assessing the quality of models and variable selection â ¢ Preventing overfitting, using the Elastic Net and Bayesian methods â ¢ Analyzing univariate and multivariate time series data â ¢ Grouping data via K-means and hierarchical clustering â ¢ Preparing reports, slideshows, and web pages with knitr â ¢ Building reusable R packages with devtools and Rcpp â ¢ Getting involved with the R global community Â

Book Information

File Size: 70877 KB

Print Length: 464 pages

Simultaneous Device Usage: Up to 5 simultaneous devices, per publisher limits

Publisher: Addison-Wesley Professional; 1 edition (December 20, 2013)

Customer Reviews

To get the negatives out of the way: it's unfortunate that, having invested in appealing graphic design - the book looks just so much nicer than the spartan O'Reilly titles - Addison Wesley have not provided the author with similarly solid editorial support, resulting in a book that definitely feels rushed. There are typos, cosmetic blemishes (one regular annoyance is a table that's too wide for a page - one could fit it using a smaller font, but instead the table ends up split across twice as many rows), a couple of statistical blunders (on pp. 172 and 263), things that could have been left out, things that should have been included (oddly, the chapter on joins never mentions outer joins, and, in fact, does not explain what a join is) - and, finally, time and again, things that should have been explained better. I do not feel that "R for Everyone" is the best available introduction to R, and continue to endorse Robert Kabacoff's high-quality "R in Action" in that capacity.Where "R for Everyone" differs from "R in Action" - and, coming to the positives, where it wins out - is in intermediate-R territory. One important example is coverage of "ggplot2". Whereas "R in Action" discusses the "old school" R graphics, "R for Everyone" goes with "ggplot2", becoming the second popular book (after Winston Chang's "R Graphics Cookbook") to discuss the package - and although its explanation of "ggplot2" syntax is sketchy, the samples found throughout the book do build into a useful "ggplot2" gallery that actually brought me over the fence. "plyr" package, an important data-manipulation aid, is another example, and another "R in Action" no-show. So is "data.table".

In spite of the good reviews, the reality is very different. This is a bad R book with many shortcomings that range from bad explanations of the basics (functions, data types) to no explanation at all when showing the use of R in data analysis.The book has a nice layout in color, but this is misleading. Long R output without proper formatting for a textbook is always displayed because the author wrote the book directly in the code as he himself states and printed it out as it is. And it feels like. Most of the text looks just like comments in a program code. The treatment of functions is very poor (they are also very rarely used in the book) and the explanation of the different R data types lacks depth and is misguided. Silly examples are used to show the basics as in printing the author's name. The later chapters get even worse, literally damaging all the more interesting parts, where the book leaves the very basics and moves on to data handling and then to advanced data analytics in R.The part of the book that deals with data analytics is sincerely a bit of a tragedy. Rushed text with no clear or sometimes whatsoever explanations of what is actually being done, with just little text and lots of code output and charts taking most of the space. Ironically the book that is "for everyone" makes hard for "everyone" to understand anything that uses statistics, about 60% of the book!.It is harder even for those trained on statistics or related "hard" sciences.For example, In chapter 22, right in the beginning the author uses a value for the predicted number of clusters in the data under analysis. This value is taken out of the blue and only later it is shown how this value can be found using two methods.

Download to continue reading...

R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data & Analytics Series) R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data and Analytics) Data Just Right: Introduction to Large-Scale Data & Analytics (Addison-Wesley Data and Analytics) Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem (Addison-Wesley Data & Analytics Series) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data Science, ... (Hacking Freedom and Data Driven Book 2) Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem (Addison-Wesley Data & Analytics) Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics Series) Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2 (Addison-Wesley Data & Analytics) Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference (Addison-Wesley Data & Analytics) Advanced Programming in the UNIX Environment

(Addison-Wesley Professional Computing Series) Advanced Programming in the UNIX(R) Environment (Addison-Wesley Professional Computing Series) Win32 Programming (Addison-Wesley Advanced Windows Series)(2 Vol set) Big Data For Beginners: Understanding SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More! Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault The Design and Implementation of the 4.4 BSD Operating System (Addison-Wesley UNIX and Open Systems Series) First Principles of Discrete Systems and Digital Signal Processing (Addison-Wesley Series in Electrical Engineering) Principles of Compiler Design (Addison-Wesley series in computer science and information processing) Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation (Adobe Reader) (Addison-Wesley Signature Series (Fowler)) Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions (Addison-Wesley Signature Series (Fowler)) Circuits, Interconnections, and Packaging for Vlsi (Addison-Wesley VLSI systems series)