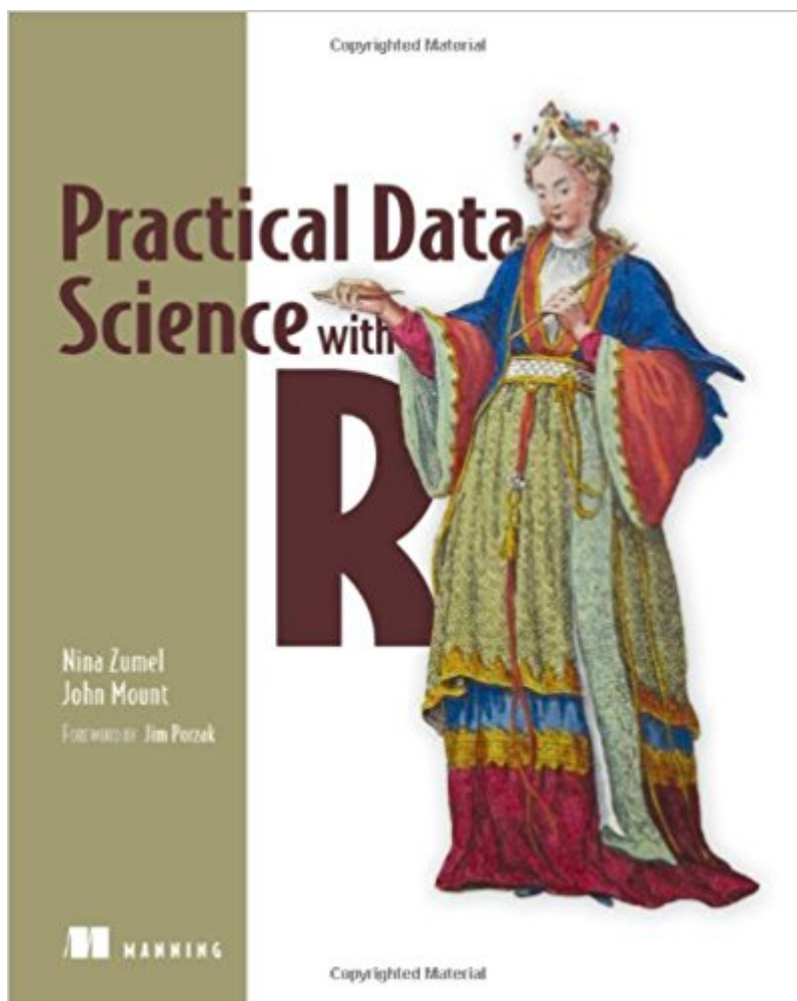


The book was found

# Practical Data Science With R



## Synopsis

Summary Practical Data Science with R lives up to its name. It explains basic principles without the theoretical mumbo-jumbo and jumps right to the real use cases you'll face as you collect, curate, and analyze the data crucial to the success of your business. You'll apply the R programming language and statistical analysis techniques to carefully explained examples based in marketing, business intelligence, and decision support. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

About the Book Business analysts and developers are increasingly collecting, curating, analyzing, and reporting on crucial business data. The R language and its associated tools provide a straightforward way to tackle day-to-day data science tasks without a lot of academic theory or advanced mathematics. Practical Data Science with R shows you how to apply the R programming language and useful statistical techniques to everyday business situations. Using examples from marketing, business intelligence, and decision support, it shows you how to design experiments (such as A/B tests), build predictive models, and present results to audiences of all levels. This book is accessible to readers without a background in data science. Some familiarity with basic statistics, R, or another scripting language is assumed.

What's Inside Data science for the business professional Statistical analysis using the R language Project lifecycle, from planning to delivery Numerous instantly familiar use cases Keys to effective data presentations

About the Authors Nina Zumel and John Mount are cofounders of a San Francisco-based data science consulting firm. Both hold PhDs from Carnegie Mellon and blog on statistics, probability, and computer science at win-vector.com.

Table of Contents

PART 1 INTRODUCTION TO DATA SCIENCE The data science process Loading data into R Exploring data Managing data

PART 2 MODELING METHODS Choosing and evaluating models Memorization methods Linear and logistic regression Unsupervised methods Exploring advanced methods

PART 3 DELIVERING RESULTS Documentation and deployment Producing effective presentations

## Book Information

Paperback: 389 pages

Publisher: Manning; 1st edition (April 13, 2014)

Language: English

ISBN-10: 1617291560

ISBN-13: 978-1617291562

Product Dimensions: 7.3 x 1 x 9.1 inches

Shipping Weight: 1.6 pounds (View shipping rates and policies)

Average Customer Review: 4.3 out of 5 stars 33 customer reviews

Best Sellers Rank: #74,194 in Books (See Top 100 in Books) #12 in Books > Computers & Technology > Programming > Software Design, Testing & Engineering > Tools #21 in Books > Computers & Technology > Programming > Software Design, Testing & Engineering > Testing #27 in Books > Textbooks > Computer Science > Artificial Intelligence

## Customer Reviews

Nina Zumel co-founded Win-Vector, a data science consulting firm in San Francisco. She holds a PH.D. in robotics from Carnegie Mellon and was a content developer for EMC's Data Science and Big Data Analytics Training Course. Nina also contributes to the Win-Vector Blog, which covers topics in statistics, probability, computer science, mathematics and optimization. John Mount co-founded Win-Vector, a data science consulting firm in San Francisco. He has a Ph.D. in computer science from Carnegie Mellon and over 15 years of applied experience in biotech research, online advertising, price optimization and finance. He contributes to the Win-Vector Blog, which covers topics in statistics, probability, computer science, mathematics and optimization.

This is the most practical R book on enterprise approach to data analytics. Rarely any book that can spare several chapters on preparing data, which in fact build the foundation of a good modeling. However, I would prefer the data cleansing and the big data algorithm on data mining algorithm be expanded further. As a reader, I need additional exploration with other materials rather than only this book.

Practical Data Science is a fun interesting book.. There are parts that lost me.... like watching a ball under the three shells kind of thing... but-- the 70% of the book that I DID get is remarkable. I learned many things that I will put to use. Well worth the price of the book. For example-- I loved the lookup vectors to change values.. very interesting graphs. This book does not waste your time. Hopefully soon I can grasp everything. Good job. I recommend for all R users. I love that the examples are about real business situations and not plant life.

Love this book. Any one interested in data science should get this hands-on experiential learning book.

An awesome book that's helped me on occasion with my coursework and my job.

The book has a lot of good examples and gets you jump started on being productive.

This book is what I was looking for for my new job as a Credit Risk Data Modeler (basically data science applied to credit problems). It's coverage is broad, but deep and applied enough that I have been able to apply its contents in my day-to-day work. I look forward to a second edition which will hopefully rectify the following: Earlier in the book it seemed the authors took great pains to explain in layman's terms the various statistical elements of the topic they were covering. They provided very clear and meaningful explanations which made a lot of sense of complex topics. But later in the book it seemed that that approach largely went out the window and they started using more technical boiler plate to describe the various statistical tests and procedures. Rather than perhaps give the technical boilerplate (as you'd see it in a textbook) and then elaborate on it with a more human-centric explanation, they would just leave it at the nearly impervious technical description and then proceed to explain how to conduct the test/procedure/etc in R. But without understanding of what you're trying to accomplish and why, it's hard to write the code to actually do it. Keep in mind that I'm relatively well prepared for this book too, having had as much stats and econometrics as I could fit into my four-year degree. If I found some sections of the book too technical to understand then it seems likely that the book would benefit from some additional explanation and discussion in those later sections. Also, I have a good deal of "boots on the ground" experience with this book in my attempts to apply it in my daily work. I've found that it is useful, but could be more useful if there was more discussion of various practical problems. For instance, much of my work is focused on producing a predictive model of likelihood of charge-off. I.e., if we approve and fund this application, how likely is it to perform or charge-off. The book shares some high-level approaches to finding problems in data (using plots and summaries), fixing those problems using various techniques, selecting variables, and how to conduct the statistical modeling (logistic in my case). But it fails to really tie those areas together beyond the high-level. For instance, what are the assumptions of a logistic regression? How do you resolve issues in your data to ensure that you meet those assumptions and can perform a valid logistic regression? How do you really select variables when you're faced with at least 20 possibilities (and potentially many many more if you count interaction terms, unfixed variables, and variables which have been fixed in different ways)? I suppose, for what it was, that it is "mission accomplished." I'd just like to see a lot more. Perhaps there's need for a second volume? Perhaps "Advanced Practical Data Science with R?" Either this book could have a second edition with a lot more content covering finding data problems, resolving

those problems intelligently (for instance, resolving missing data is basically left as "either drop the effected records" or "use the mean as a replacement or the missing value," but there are alternative methods which may be more suitable), what data problems will cause issues in OLS regression, logistic regression, and machine learning; And how to practically select variables and a model. I feel like the book gave me some tools to apply (like a small box of tools you might purchase from a hardware store), but left a lot out. So now I'm in deep water trying to figure out why my logistic regression isn't predictive enough and what I can do about it. Is it the data and how I fixed variables? Is it the variables I've selected? Should I have used automated variable selection techniques? Or just manually tried different variables? How does an experienced practitioner approach these problems? I know they iterate: explore data, clean data, select variables, select model, test model, look at data, change data, change variables, etc... but practically speaking what does it look like? In the book they offer a hand-coded basic variable selection script, and mention that one could also use stepwise variable selection. In the real world I'm reasonably sure that this is not actually done--mostly because their selection script does about as well as stepwise at selecting appropriate variables. There are many other better ways of selecting variables, I've discovered, and I wish that they'd discussed some of those ways (pros and cons), and shown how to conduct them in a meaningful fashion. Same thing with building a model. In my case, I have a whole bunch of variables, limited data (about 2000 records, with the desired outcome only occurring in 120 of those), and the automated tools (various R packages I've discovered and applied) either take a long time to run and/or yield poor results. But if not automated tools then what? Manually add variables and ANOVA test the difference between the first and second model? I'd just like more...more discussion and elaboration and examples of how practical data science is conducted. This book seems like it does a fantastic job as an introduction to the topic, but you'll quickly find that you'll be in deep water without a clue how to swim--as in my case. You'll be left to your own devices, and find yourself wishing, as I do, that there was more in the book (or another book) that I could study after this one which would help take me from beginner data scientist to intermediate. Overall, I'm very glad I bought and read the book.

Good Read

Great very detail book! I am an analytics professional looking to starting using R programming to leverage my data the best way I can.

[Download to continue reading...](#)

Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data Book 1) Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics (Big Data Analytics, Data Science, Business Intelligence Book 6) Big Data For Business: Your Comprehensive Guide to Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Data Analytics and Python Programming: 2 Bundle Manuscript: Beginners Guide to Learn Data Analytics, Predictive Analytics and Data Science with Python Programming Analytics: Business Intelligence, Algorithms and Statistical Analysis (Predictive Analytics, Data Visualization, Data Analytics, Business Analytics, Decision Analysis, Big Data, Statistical Analysis) Data Analytics For Beginners: Your Ultimate Guide To Learn and Master Data Analysis. Get Your Business Intelligence Right - Accelerate Growth and Close More Sales (Data Analytics Book Series) Discovering Knowledge in Data: An Introduction to Data Mining (Wiley Series on Methods and Applications in Data Mining) Practical Data Management for Risk Data Aggregation and BCBS 239 Compliance Second Edition Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) Principles of Data Wrangling: Practical Techniques for Data Preparation Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton Series in Modern Observational Astronomy) Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (Morgan Kaufmann Series in Data Management Systems) Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems) Analytics: Data Science, Data Analysis and Predictive Analytics for Business Data Smart: Using Data Science to Transform Information into Insight R for Data Science: Import, Tidy, Transform, Visualize, and Model Data

[Contact Us](#)

[DMCA](#)

[Privacy](#)

