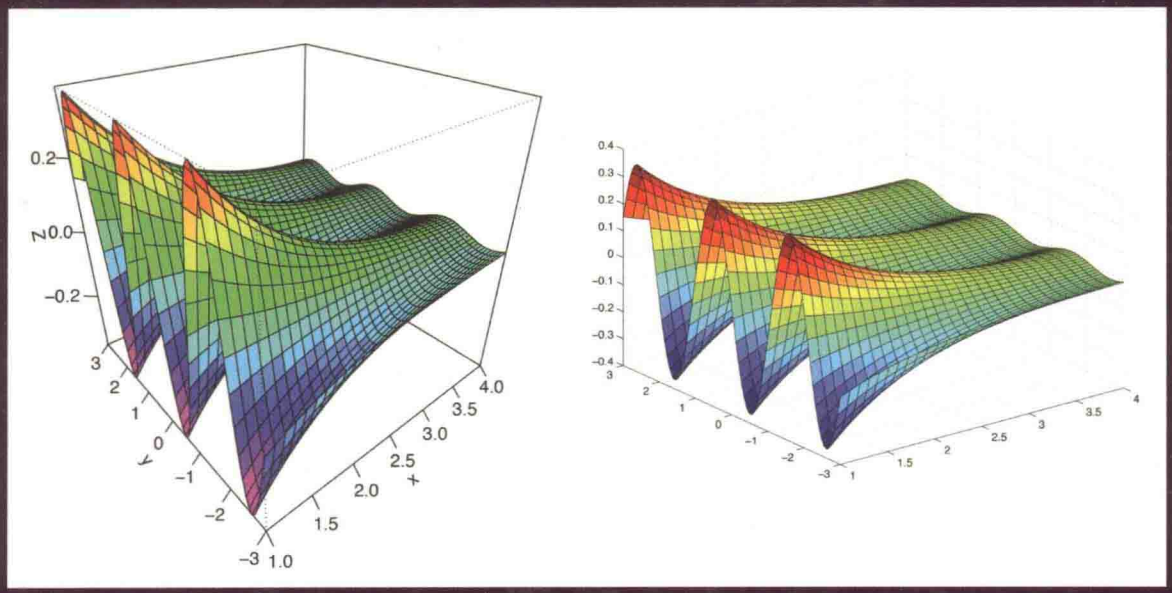
The R Series

R and MATLAB®



David E. Hiebeler

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

R and MATLAB®

David E. Hiebeler
University of Maine
Orono, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20150406

International Standard Book Number-13: 978-1-4665-6838-9 (Hardback)

Library of Congress Cataloging-in-Publication Data

Hiebeler, David E.
R and MATLAB / David E. Hiebeler.
pages cm. -- (Chapman & Hall/CRC the R series ; 30)
"A CRC title."
Includes bibliographical references and index.
ISBN 978-1-4665-6838-9 (alk. paper)
1. Multivariate analysis--Data processing. 2. R (Computer program language) 3. MATLAB. I. Title.

QA278.H54 2016
510.285'53--dc23 2015006954

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

R and MATLAB®

Chapman & Hall/CRC
The R Series

Series Editors

John M. Chambers
Department of Statistics
Stanford University
Stanford, California, USA

Torsten Hothorn
Division of Biostatistics
University of Zurich
Switzerland

Duncan Temple Lang
Department of Statistics
University of California, Davis
Davis, California, USA

Hadley Wickham
RStudio
Boston, Massachusetts, USA

Aims and Scope

This book series reflects the recent rapid growth in the development and application of R, the programming language and software environment for statistical computing and graphics. R is now widely used in academic research, education, and industry. It is constantly growing, with new versions of the core software released regularly and more than 6,000 packages available. It is difficult for the documentation to keep pace with the expansion of the software, and this vital book series provides a forum for the publication of books covering many aspects of the development and application of R.

The scope of the series is wide, covering three main threads:

- Applications of R to specific disciplines such as biology, epidemiology, genetics, engineering, finance, and the social sciences.
- Using R for the study of topics of statistical methodology, such as linear and mixed modeling, time series, Bayesian methods, and missing data.
- The development of R, including programming, building packages, and graphics.

The books will appeal to programmers and developers of R software, as well as applied statisticians and data analysts in many fields. The books will feature detailed worked examples and R code fully integrated into the text, ensuring their usefulness to researchers, practitioners and students.

Published Titles

Stated Preference Methods Using R, *Hideo Aizaki, Tomoaki Nakatani, and Kazuo Sato*

Using R for Numerical Analysis in Science and Engineering, *Victor A. Bloomfield*

Event History Analysis with R, *Göran Broström*

Computational Actuarial Science with R, *Arthur Charpentier*

Statistical Computing in C++ and R, *Randall L. Eubank and Ana Kupresanin*

Reproducible Research with R and RStudio, Second Edition, *Christopher Gandrud*

R and MATLAB® *David E. Hiebeler*

Nonparametric Statistical Methods Using R, *John Kloke and Joseph McKean*

Displaying Time Series, Spatial, and Space-Time Data with R, *Oscar Perpiñán Lamigueiro*

Programming Graphical User Interfaces with R, *Michael F. Lawrence and John Verzani*

Analyzing Sensory Data with R, *Sébastien Lê and Theirry Worch*

Parallel Computing for Data Science: With Examples in R, C++ and CUDA, *Norman Matloff*

Analyzing Baseball Data with R, *Max Marchi and Jim Albert*

Growth Curve Analysis and Visualization Using R, *Daniel Mirman*

R Graphics, Second Edition, *Paul Murrell*

Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving, *Deborah Nolan and Duncan Temple Lang*

Multiple Factor Analysis by Example Using R, *Jérôme Pagès*

Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R, *Daniel S. Putler and Robert E. Krider*

Implementing Reproducible Research, *Victoria Stodden, Friedrich Leisch, and Roger D. Peng*

Graphical Data Analysis with R, *Antony Unwin*

Using R for Introductory Statistics, Second Edition, *John Verzani*

Advanced R, *Hadley Wickham*

Dynamic Documents with R and knitr, Second Edition, *Yihui Xie*

To my parents,
for encouraging me as I got started
with that first Apple][+.

Preface

Who this book is for

I have interacted with many people over the past several years after making my "MATLAB/R Reference" available on my Web site. Based on those conversations, there is a large population of people out there who have used MATLAB®[1] for some period of time, but who now find themselves working with biologists, statisticians, or some other professionals who speak R rather than MATLAB. There is a complementary group of people who use R, who now find themselves trying to work and share code with colleagues who use MATLAB. The intended reader is someone who already knows one package, and for whatever reasons, now needs or wants to learn the other.

This book grew out of the MATLAB/R Reference document mentioned above. That document is a concise reference summary of many of the key topics presented in this book, and continues to be available on my Web site, at `http://www.math.umaine.edu/~hiebeler`.

My own experience learning R

I fall into the first category of users described above. I used MATLAB for many years, primarily for prototyping my research simulations (which I would then rewrite in C for faster performance) and for data visualization and graphics. I also regularly taught a course on "Modeling and Simulation" in which I used MATLAB as the software platform. The course covers various biological models, including stochastic spatial models. I found more and more biologists signing up for the class over the years, and some of them started asking if they could use R because they had already started learning to use it in their statistics classes. I was somewhat tired of learning new programming languages/environments, and was not particularly eager to learn another one.[2] But I reluctantly decided to look into R. At first I was very frustrated with some of the differences that struck me right away — for example, when you edit a file defining a new R function, you cannot just type its name to call the function, but instead must "**source**" it first. Typing a matrix into R is certainly more tedious than doing so in MATLAB.

However, over time, I came to appreciate the power and flexibility of many features of R. The fact that it is available for free on Windows, Mac OS-X, and Linux certainly has

[1] MATLAB® and Simulink® are registered trademarks of The MathWorks, Inc. For product information, please contact: The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098 USA; Tel: 508-647-7000; Fax: 508-647-7001; E-mail: `info@mathworks.com`; Web: `www.mathworks.com`.

[2] The main programming languages I had used over the years at that point were Applesoft BASIC, Apple][+ machine language, Pascal, Modula-2, Forth, C, Lisp, APL, SNOBOL, PostScript, Java, Perl, MATLAB, and a bit of Python. Plus various libraries/environments and programming-like things such as SunView, X11, csh scripting, LaTeX, and HTML. That list is certainly smaller than that of many computer scientists, but long enough for me.

made it easier for the many students who work on research with me or take my courses to obtain a copy for their own computers.

There are still some times when I miss the more concise/simple way of expressing things in MATLAB (in particular, how much more convenient it is to type in a matrix), but I have now come to enjoy the time I spend working in R.

Formatting conventions and terminology

R and MATLAB commands, i.e., things you could actually type at the command prompt, are generally formatted using a `non-bold typewriter font`, for example `x = sqrt(7)`. A **bold font** is used when referring to the name of a function, file, variable, or keyword, for example the **sqrt** function, the variable **foo**, or **for** loops.

R and MATLAB are referred to in the book as "platforms," rather than the perhaps more natural "software packages," because the word "package" is a loaded word with specific meaning in R.

In some parts of the book, R and MATLAB code are placed side by side. This is done for brief commands and concepts which do not need much explanation. In other parts, R material is presented, followed by MATLAB material, with some differences between the platforms emphasized.

Commands vs. GUI

There are two main approaches to working with software platforms like MATLAB and R: doing things primarily with commands, and doing things primarily via menus through a graphical user interface (GUI). I will admit that I am old-school, and started using computers before they *had* graphical user-interfaces. Personally, I can type commands much more quickly than I can click my way to equivalent commands via menus, so I prefer the command-line approach over the GUI approach.

As one justification for my preference, using commands to achieve goals is usually a bit more portable and easier to share with others than using menus. The placements of various items in menus is sometimes different on different operating systems such as Mac OS-X vs. Microsoft Windows, so if you are trying to explain to someone how to accomplish something, and they have a different operating system than you (or even simply a different version of the software), you may have some trouble. Admittedly, some commands may also differ between operating systems or between versions of the software. But for the most part, if you write an R or MATLAB script on one operating system, it will work identically on others, and can be usefully shared with other people more easily than a description of which menu items to select and which buttons to push. The configuration and details of user-interfaces also tend to change more frequently than the commands in packages like MATLAB and R; if you tell someone how to do something using commands, it is likely to work on a wider variety of versions of the software. I therefore focus more on using commands than on clicking on menu items in this book, for those tasks where both approaches can be used.

What this book is not

This book is not intended to be a comprehensive introduction or overview of either the R or MATLAB platforms. I should warn you up front that this book will not teach you what many may consider the "best" way to do things in R or MATLAB, according to the deeper philosophy of either platform. My goal, based on my experience, is to try and show how to do things in either platform which is most similar to the ways they are done in the other, to make the transition from one platform to the other as quick and painless as possible for you. If you want to delve into the deeper ways of thinking behind using either platform, this reference can help get you started, but you should follow up with some of the other plentiful resources available.

This book is also not intended to steer you toward one platform or the other. Quite often, the choice of software platform to use depends on your context, i.e., your employer and colleagues. If you truly have a wide-open choice, it is difficult to say which platform is best for you. My personal feeling is that many things are *easier* to do in MATLAB, but can be done more *flexibly* in R, at the expense of being more complicated. That is just a general impression, and there are of course many counterexamples. Another thing to note is that The MathWorks, Inc. (developer of MATLAB) has been quite aggressive recently about improving the performance of various types of MATLAB code; many of my MATLAB computations run many times faster in newer versions of MATLAB than they used to. If you are at a large company or academic institution, you likely have access to a site license to use MATLAB. If you are looking for routines to perform specialized statistical tests, R has a vast and quickly growing library of packages to fit that need. R is of course available to download free of charge, although MATLAB can be evaluated in a free trial, and there is very affordable pricing for students or for personal use.

Acknowledgments

Thanks to my editor, John Kimmel, for being far too patient with me as I worked on this book.

Bill Halteman helped me learn R, first by going through my MATLAB simulation lab exercises and determining it would not be too difficult to implement them all in R. He then patiently answered my many questions about R (and pretty much everything else under the sun) for the following several years. Many people have also sent me suggestions and corrections for my MATLAB/R Reference that this book was originally based on. Those contributors include Juan David Ospina Arango, Berry Boessenkool, Robert Bryce, Thomas Clerc, Alan Cobo-Lewis, Richard Cotton, Stephen Eglen, Andreas Handel, Niels Richard Hansen, Luke Hartigan, Roger Jeurissen, David Khabie-Zeitoune, Seungyeon Kim, Michael Kiparsky, Isaac Michaud, Andy Moody, Ben Morin, Lee Pang, Manas A. Pathak, Rachel Rier, Rune Schjellerup Philosof, Rachel Rier, William Simpson, David Winsemius, Corey Yanofsky, and Jian Ye.

Finally, like most authors, I must thank my family and apologize to them for all of the times I chose to spend with drafts of this book rather than with them.

About the Author

David E. Hiebeler is an Associate Professor in the Department of Mathematics & Statistics at the University of Maine. He was previously a Visiting Lecturer at Cornell University. He has degrees from Rensselaer Polytechnic Institute, Harvard University, and a Ph.D. in applied mathematics from Cornell University.

He previously worked at the Center for Nonlinear Studies, the Theoretical Division, and the Advanced Computing Lab at Los Alamos National Laboratory, the Santa Fe Institute, and Thinking Machines Corporation. He began programming at age 12 on an Apple][+ computer. His research involves mathematical and computational stochastic spatial models in population ecology and epidemiology. He also dabbles in iOS (iPhone/iPad) development, primarily for K–12 outreach.