

MARK GARDENER

The background features two large, stylized letters, 'R' and 'N', rendered in a dark grey color with a fine dotted pattern. The 'R' is positioned in the upper left, and the 'N' is in the lower right. The title text is overlaid on these letters.

The Essential **R** Reference

The
ESSENTIAL
REFERENCE **R**

The
ESSENTIAL
REFERENCE **R**

Mark Gardener



John Wiley & Sons, Inc.

The Essential R Reference

Published by
John Wiley & Sons, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Copyright © 2013 by Mark Gardener

Published by John Wiley & Sons, Inc., Indianapolis, Indiana
Published simultaneously in Canada

ISBN: 978-1-118-39141-9
ISBN: 978-1-118-39140-2 (ebk)
ISBN: 978-1-118-39138-9 (ebk)
ISBN: 978-1-118-39139-6 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

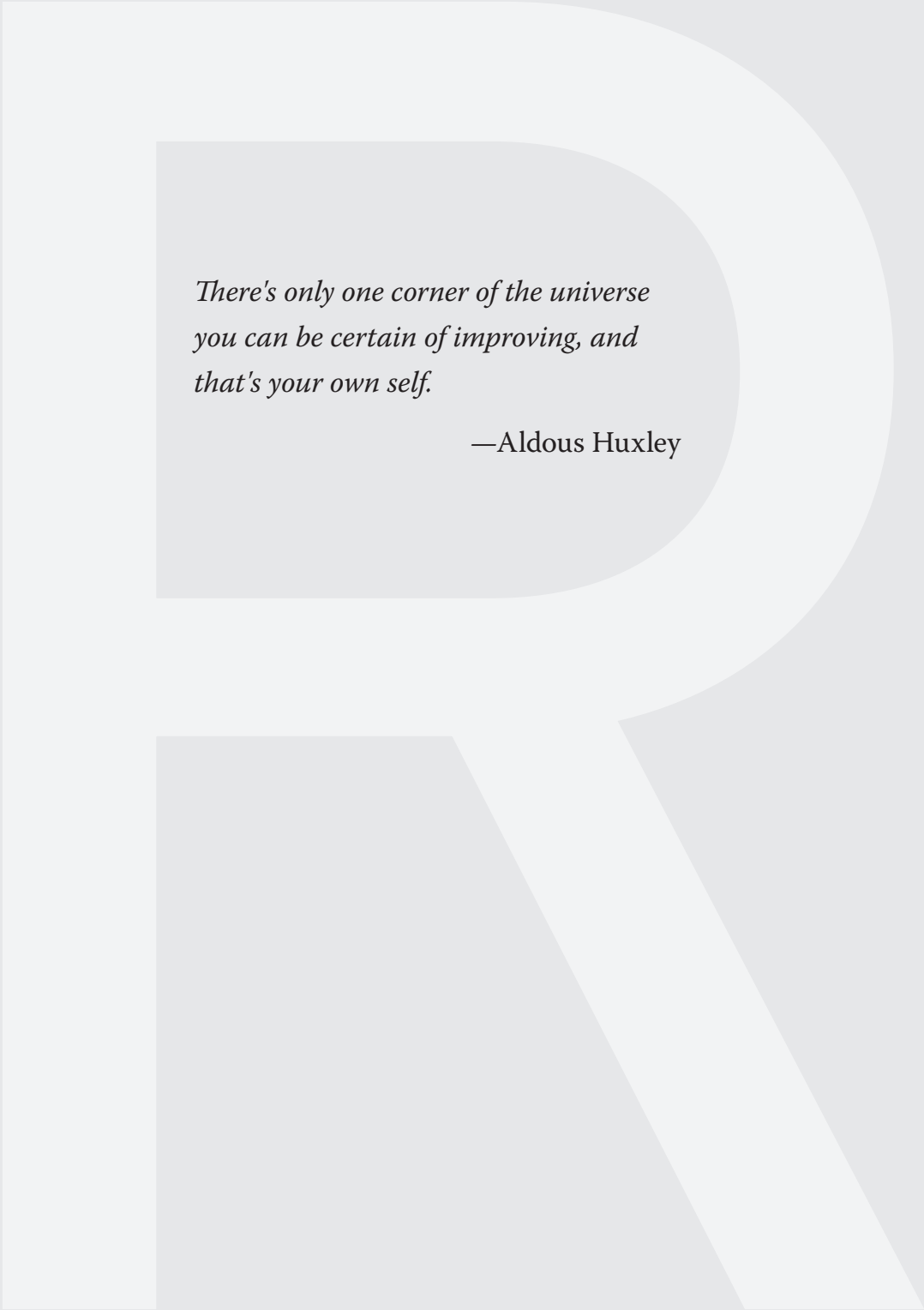
Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that Internet websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2012948918

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.



*There's only one corner of the universe
you can be certain of improving, and
that's your own self.*

—Aldous Huxley

ABOUT THE AUTHOR



Mark Gardener (<http://www.gardenersown.co.uk>) is an ecologist, lecturer, and writer working in the UK. He has a passion for the natural world and for learning new things. Originally he worked in optics, but returned to education in 1996 and eventually gained his doctorate in ecology and evolutionary biology. This work involved a lot of data analysis and he became interested in R as a tool to help in research. He is currently self-employed and runs courses in ecology, data analysis, and R for a variety of organizations. Mark lives in rural Devon with his wife Christine (a biochemist) and still enjoys the natural world and learning new things.

ABOUT THE TECHNICAL EDITOR

Richard Rowe started his professional life as a physicist, but switched fields to earn a PhD in insect behavior. He has taught data analysis courses, mainly to biologists, at Canterbury, then James Cook University, since 1982. He has worked with R since 1997 when a friend forced a very early copy onto him. The R system has exponentially improved over the past decade, and in light of the fact that Richard's individual capacity is more linear, he retired in 2011 but keeps his hand in data-analysis consultancies and master-class workshops regularly (the best way to learn is to teach). Based on life, his belief is that ecological and behavioral data is usually the dirtiest and most ill-behaved, and hence the most fun to explore for pattern. His other hobby is dragonfly biology.

CREDITS

Executive Editor

Carol Long

Project Editor

Victoria Swider

Technical Editor

Richard Rowe

Production Editor

Kathleen Wisor

Copy Editor

Kim Cofer

Editorial Manager

Mary Beth Wakefield

Freelancer Editorial Manager

Rosemarie Graham

Associate Director of Marketing

David Mayhew

Marketing Manager

Ashley Zurcher

Business Manager

Amy Knies

Production Manager

Tim Tate

**Vice President and Executive Group
Publisher**

Richard Swadley

Vice President and Executive Publisher

Neil Edde

Associate Publisher

Jim Minatel

Project Coordinator, Cover

Katie Crocker

Composer

Jeff Lytle, Happenstance Type-O-Rama

Proofreader

James Saturnio, Word One

Indexer

Jack Lewis

Cover Designer

Ryan Sneed

ACKNOWLEDGMENTS

First of all my thanks go out to the R project team and the many authors and programmers who work tirelessly to make this a peerless program. I would also like to thank my wife, Christine, who has had to put up with me during this entire process, and in many senses became an R-widow! Thanks to Wiley, for helping this book become a reality, especially Carol Long and Victoria Swider. I couldn't have done it without you. Thanks also to Richard Rowe, the technical reviewer, who first brought my attention to R and its compelling (and rather addictive) power. Last but not least, thanks to the R community in general. I learned to use R largely by trial and error and using the vast wealth of knowledge that is in this community. I hope that this book is a worthwhile addition to the R knowledge base and that it will prove useful to all users of R.

— MARK GARDENER

CONTENTS

| | |
|---|------------|
| <i>Introduction</i> | xv |
| Theme 1: Data | 1 |
| Data Types | 3 |
| Types of Data | 3 |
| Altering Data Types | 16 |
| Testing Data Types | 18 |
| Creating Data | 22 |
| Creating Data from the Keyboard | 22 |
| Creating Data from the Clipboard | 29 |
| Adding to Existing Data | 29 |
| Importing Data | 39 |
| Importing Data from Text Files | 39 |
| Importing Data from Data Files | 46 |
| Saving Data | 49 |
| Saving Data as a Text File to Disk | 50 |
| Saving Data as a Data File to Disk | 59 |
| Viewing Data | 61 |
| Listing Data | 61 |
| Data Object Properties | 74 |
| Selecting and Sampling Data | 107 |
| Sorting and Rearranging Data | 117 |
| Summarizing Data | 121 |
| Summary Statistics | 121 |
| Summary Tables | 136 |
| Distribution of Data | 146 |
| Density Functions | 148 |
| Probability Functions | 152 |
| Quantile Functions | 158 |
| Random Numbers | 161 |
| Theme 2: Math and Statistics | 167 |
| Mathematical Operations | 169 |
| Math | 169 |
| Logic | 188 |
| Complex Numbers | 194 |
| Trigonometry | 202 |
| Hyperbolic Functions | 206 |
| Matrix Math | 206 |
| Summary Statistics | 223 |
| Simple Summary Stats | 223 |
| Tests of Distribution | 235 |
| Differences Tests | 240 |
| Parametric Tests | 241 |
| Non-parametric | 252 |

| | |
|--|------------|
| Correlations and Associations | 267 |
| Correlation | 267 |
| Association and Goodness of Fit | 275 |
| Analysis of Variance and Linear Modeling | 283 |
| ANOVA | 284 |
| Linear Modeling | 300 |
| Miscellaneous Methods | 317 |
| Clustering | 318 |
| Ordination | 324 |
| Time Series | 331 |
| Non-linear Modeling and Optimization | 333 |
| Theme 3: Graphics | 347 |
| Making Graphs | 348 |
| Types of Graphs | 349 |
| Saving Graphs | 389 |
| Adding to Graphs | 398 |
| Adding Data | 398 |
| Adding Lines | 404 |
| Adding Shapes | 416 |
| Adding Text | 422 |
| Adding Legends | 432 |
| Graphical Parameters | 436 |
| Using the par Command | 437 |
| Altering Color | 439 |
| Altering Axis Parameters | 446 |
| Altering Text Parameters | 453 |
| Altering Line (and Box) Parameters | 456 |
| Altering Plot Margins | 459 |
| Altering the Graph Window | 462 |
| Theme 4: Utilities | 475 |
| Install | 476 |
| Installing R | 477 |
| Installing Packages | 477 |
| Using R | 480 |
| Using the Program | 480 |
| Additional Packages | 486 |
| Programming | 490 |
| Managing Functions | 491 |
| Saving and Running Scripts | 498 |
| Conditional Control | 502 |
| Returning Results | 507 |
| Error Trapping | 525 |
| Constants | 527 |
| <i>Index</i> | <i>529</i> |

INTRODUCTION

R is rapidly becoming the *de facto* standard among professionals, and is used in every conceivable discipline from science and medicine to business and engineering. R is more than just a computer program; it is a statistical programming environment and language. R is free and open source and is, therefore, available to everyone with a computer.

R is a language with its own vocabulary and grammar. To make R work for you, you communicate with the computer using the language of R and tell it what to do. You accomplish this by typing commands directly into the program. This means that you need to know some of the words of the language and how to put them together to make a “sentence” that R understands. This book aims to help with this task by providing a “dictionary” of words that R understands.

The help system built into R is extensive, but it is arranged by command name; this makes it hard to use unless you know some command names to start with. That’s where this book comes in handy; the command names (the vocabulary of R) are arranged by topic, so you can look up the kind of task that you require and find the correct R command for your needs.

I like to think of this book as a cross between a dictionary, a thesaurus, and a glossary, with a fair sprinkling of practical examples. Even though some may consider me an “R expert” at this point, I am still learning and still forgetting! I often have to refer to notes to remind me how to carry out a task in R. That is why I wrote this book—to help novice users learn more easily, and to provide more experienced users with a reference work they can delve into time and time again. I also learned a great deal more about R by writing about it, and I hope that you will find it an essential companion in your day-to-day conversations with R.

Who This Book Is For

This book is for anyone who needs to analyze any data, whatever their discipline or line of work. Whether you are in science, business, medicine, or engineering, you will have data to analyze and results to present. R is powerful and flexible and completely cross-platform. This means you can share data and results with anyone. R is backed by a huge project team, so being free does not mean being inferior!

Whether you are a student or an experienced programmer, this book is meant to be an essential reference. If you are completely new to R, this book will enable you to learn more quickly by providing an easy-to-use “dictionary.” You may also consider reading my previous book,

Beginning R: The Statistical Programming Language, which provides a different learning environment by taking you from simple tasks to more complex ones in a linear fashion.

If you are already familiar with R, this book will help as a useful reference work that you can call upon time and time again. It is easy to forget the name of a command or the exact syntax of the command. In addition to jogging your memory, the examples in the book will help put the commands into context.

What This Book Covers

Each command listed in this book has an explanation of what the command does and how to use it—the “grammar,” if you will. Related commands are also listed, as in a thesaurus, so if the word you are looking at is not quite what you need, you are likely to see the correct one nearby.

I can’t pretend that this reference book covers every command in the R language, but it covers a lot (more than 400). I’ve also not covered some of the more obscure parameters (formally called “arguments” in R) for some of the commands. I called this book “Essential” because I believe it covers the essentials. I also hope that you will find it essential in your day-to-day use of R.

One of the weaknesses of the R help system is that some of the examples are hard to follow, so each command listed in this book is accompanied by various examples. These show you the command “in action” and hopefully help you to gain a better understanding of how the command works. The examples are written in R code and set out as if you had typed them into R yourself. And unlike the built-in help system in R, you get to see the results, too!

How This Book Is Structured

This book is not a conventional textbook; it is intended as a reference work that you can delve into at any point.

This book is organized in a topic-led, logical manner so that you can look for the kind of task that you want to carry out in R and find the command you need to carry out that task as easily as possible, even if you do not know the name of the command. The book is split into four grand themes:

- Theme 1: “Data”
- Theme 2: “Math and Statistics”
- Theme 3: “Graphics”
- Theme 4: “Utilities”

These are hopefully self-explanatory, with the exception perhaps of “Utilities”; this covers the commands that did not fit easily into one of the other themes, particularly those relating to the programming side of R.

You can use the table of contents to find your way to the topic that matches the task you want to undertake. If the command you need is not where you first look, there is a good chance that the command you did find will have a link to the appropriate topic or command (some commands have entries on more than one topic).

The index is also a helpful tool because it contains an alphabetical list of all the commands, so you can always find a specific command by its name there.

The following is a brief description of each of the four main themes:

Theme 1: “Data”—This theme is concerned with aspects of dealing with data. In particular:

- **Data types**—Different kinds of data and converting one kind of data into another kind.
- **Creating data**—Commands for making data items from the keyboard.
- **Importing data**—Getting data from sources on disk.
- **Saving data**—How to save your work.
- **Viewing data**—Seeing what data you have in R.
- **Summarizing data**—Ways of summarizing data objects. Some of these commands also appear in Theme 2, “Math and Statistics.”
- **Distribution of data**—Looking at different data distributions and the commands associated with them, including random numbers.

Theme 2: “Math and Statistics”—This theme covers the commands that deal with math and statistical routines:

- **Mathematical operations**—Various kinds of math, including complex numbers, matrix math, and trigonometry.
- **Summary statistics**—Summarizing data; some of these commands are also in Theme 1, “Data.”
- **Differences tests**—Statistical tests for differences in samples.
- **Correlations and associations**—Including covariance and goodness of fit tests.
- **Analysis of variance and linear modeling**—Many of the commands associated with ANOVA and linear modeling can be pressed into service for other analyses.
- **Miscellaneous Tests**—Non-linear modeling, cluster analysis, time series, and ordination.

Theme 3: “Graphics”—This theme covers the graphical aspects of the R language:

- **Making graphs**—How to create a wide variety of basic graphs.
- **Adding to graphs**—How to add various components to graphs, such as titles, additional points, and shapes.
- **Graphical parameters**—How to embellish and alter the appearance of graphs, including how to create multiple graphs in one window.

Theme 4: “Utilities”—This theme covers topics that do not fit easily into the other themes:

- **Installing R**—Notes on installing R and additional packages of R commands.
- **Using R**—Accessing the help system, history of previously typed commands, managing packages, and more.
- **Programming**—Commands that are used mostly in the production of custom functions and scripts. You can think of these as the “tools” of the programming language.

Each of the topics is also split into subtopics to help you navigate your way to the command(s) you need. Each command has an entry that is split into the following sections:

- **Command Name**—Name of the command and a brief description of what it does.
- **Common Usage**—Illustrates how the command looks with commonly used options. Use this section as a memory-jogger; if you need fine details you can look in the “Command Parameters” section.
- **Related Commands**—A list of related commands along with the page numbers or a link to their entries so you can easily cross-reference.
- **Command Parameters**—Details of commonly used parameters for the command along with an explanation of what they do.
- **Examples**—Examples of the command in action. The section is set out in code style as if you had typed the commands from the keyboard yourself. You also see the resulting output that R produces (including graphical output).

Some commands are relevant to more than one theme or section; those commands either have a cross-reference and/or have an entry in each applicable place.

What You Need to Use This Book

R is cross-platform technology and so whatever computer you use, you should be able to run the program. R is a huge, open-source project and is changing all the time. However, the basic commands have altered little, and you should find this book relevant for whatever version you are using. I wrote this book using Mac R version 2.12.1, Windows R version 2.14.2, and Linux R version 2.14.1.

Having said that, if your version of R is older than about 2009, I recommend getting a newer version.

Conventions

To help you get the most from the text and keep track of what’s happening, we’ve used a number of conventions throughout the book.

R CODE

The commands you need to type into R and the output you get from R are shown in a monospace font. Each example that shows lines that are typed by the user begins with the > symbol, which mimics the R cursor like so:

```
> help()
```

Lines that begin with something other than the > symbol represent the output from R (but look out for typed lines that are long and spread over more than one line). In the following example the first line was typed by the user and the second line is the result:

```
> data1  
[1] 3 5 7 5 3 2 6 8 5 6 9
```

ANNOTATIONS

The hash symbol (#) is used as an annotation character in R (see the following example). Anything that follows is ignored by R until it encounters a new line character. The examples used throughout this book contain plenty of annotations to help guide you through the complexities and facilitate your understanding of the code lines.

```
## Some lines begin with hash symbols; that entire line is ignored by R.  
## This allows you to see the commands in action with blow by blow notes.  
> help(help) # This line has an annotation after the command
```

OPERATIONAL ASSIGNMENT

R uses two forms of “assignment.” The original form (the form preferred by many programmers) uses a kind of arrow like so: `<-`. This is used to indicate an assignment that runs from right to left. For example:

```
> x <- 23
```

This assigns the value 23 to a variable named `x`. An alternative form of assignment is mathematical type of assignment, the equals sign (`=`):

```
> x = 23
```

In most cases the two are equivalent and which you use is entirely up to you. Most of the help examples found in R and on the Internet use the arrow (`<-`). Throughout this book I have tended to use the `=` operator (because that is what I am used to), unless `<-` is the only way to make the command work.

COMMAND PARAMETERS

Most R commands accept various parameters; you can think of them as additional instructions that make the command work in various ways. Some parameters have default values that are used if you do not explicitly indicate an alternative. These parameters are also “order specific.” This means that you can specify the value you want the parameter to take without naming it as long as the values are in the correct order. An example should clarify this; the `rnorm` command generates random numbers from the normal distribution. The full command looks like this:

```
rnorm(n, mean = 0, sd = 1)
```

You supply `n`, the number of random values you want; `mean`, the mean of the values; and `sd`, the standard deviation. Both the `mean` and `sd` parameters have defaults, which are used if you do not specify them explicitly. You can run this command by typing any of the following:

```
> rnorm(n = 10, mean = 0, sd = 1)  
> rnorm(10, 0, 1)  
> rnorm(10)
```

These all produce the same result: ten values drawn randomly from a normally distributed set of values with a mean of zero and a standard deviation of one. The first line shows the full version of the command. The second line shows values for all the parameters, but unnamed. The third line shows only one value; this will be taken as `n`, with the other parameters having their default values.

This is useful for programming and using R because it means you can avoid a lot of typing. However, if you are trying to learn R it can be confusing because you might not remember what all the parameters are.

Some commands will also accept the name of the parameters in abbreviated form; others will not. In this book I have tried to use the full version of commands in the examples; I hope that this will help clarify matters.

CROSS-REFERENCES

You can find many cross-references in this book in addition to the commands listed in the “Related Commands” section of each command’s entry. These cross-references look like this:



The magnifying glass icon indicates a cross reference.

Cross references are used in the following instances:

- Relevant commands in the same section or a different section.
- Relevant sections in the same theme or in a different theme.
- An instance in which the command in question appears in another theme or section.
- An instance in which the command in question has related information in another theme.

Data Downloads

If you come across a command that has an example you would like to try on your own, you can follow along by manually typing the example into your own version of R. Some of these examples use sample data that is available for download at <http://www.wiley.com/go/EssentialRReference>. You will find all examples that require the data are accompanied by a download icon and note indicating the name of the file so you know it’s available for download and can easily locate it in the download file. The download notes look like this:



The download icon indicates an example that uses data you need to download.

Once at the site, simply locate the book’s title and click the Download Code link on the book’s detail page to obtain all the example data for the book.

There will only be one file to download and it is called `Essential.RData`. This one file contains the example data sets you need for the whole book; it contains very few because I have tried to make all data fairly simple and short so that you can type it directly. Once you have the file on your computer you can load it into R by one of several methods:

- For Windows or Mac you can drag the `Essential.RData` file icon onto the R program icon; this opens R if it is not already running and loads the data. If R is already open, the data is appended to anything you already have in R; otherwise, only the data in the file is loaded.

-
- If you have Windows or Macintosh you can also load the file using menu commands or use a command typed into R:
 - For Windows use File ➤ Load Workspace, or type the following command in R:

```
> load(file.choose())
```
 - For Mac use Workspace ➤ Load Workspace File, or type the following command in R (same as in Windows):

```
> load(file.choose())
```
 - If you have Linux, you can use the `load()` command but you must specify the filename (in quotes) exactly. For example:

```
> load("Essential.RData")
```

The `Essential.RData` file must be in your default working directory and if it is not, you must specify the location as part of the filename.

sample content of The Essential R Reference

- [How to Grow Up: A Memoir pdf, azw \(kindle\), epub](#)
 - **[download Personal Finance](#)**
 - [Bride of the Castle \(Castle Perilous, Book 8\) pdf](#)
 - [click Siege 13: Stories pdf](#)
 - **[click The Inimitable Jeeves \(Jeeves, Book 2\)](#)**
 - [read The Poacher's Son online](#)
-
- <http://thermco.pl/library/Personal-Finance--An-Encyclopedia-of-Modern-Money-Management.pdf>
 - <http://patrickvincitore.com/?ebooks/Personal-Finance.pdf>
 - <http://junkrobots.com/ebooks/Bride-of-the-Castle--Castle-Perilous--Book-8-.pdf>
 - <http://junkrobots.com/ebooks/La-Tete-d-un-Homme--Inspector-Maigret--Book-5-.pdf>
 - <http://pittiger.com/lib/The-Inimitable-Jeeves--Jeeves--Book-2-.pdf>
 - <http://jaythebody.com/freebooks/The-Poacher-s-Son.pdf>