



Journal of Statistical Software

April 2011, Volume 40, Book Review 3.

<http://www.jstatsoft.org/>

Reviewer: James P. Howard, II
University of Maryland University College

R Cookbook

Paul Teetor

O'Reilly, Sebastopol, CA, 2011.

ISBN 978-0-569-80915-7. xviii + 413 pp. USD 39.99 (P).

<http://oreilly.com/catalog/9780596809164>

R Cookbook by Paul Teetor is O'Reilly's second book on R following [Adler \(2009\)](#) and like [Adler \(2009\)](#), *R Cookbook* is neither an introductory text on R nor a text targeting specific field. Instead, *R Cookbook* provides a collection of more than 200 simple methods ranging from simple data transformations, through graphing, to advanced time series analysis. Readers unfamiliar with the underlying statistical methods are unlikely to find value in this book. This collection is aimed at the R professional who is well-versed in statistics and data management, but does not necessarily know (or remember) how to perform a specific task, such as generating a random sample, in R.

Teetor has satisfied this goal. With many books available applying R to specific domains, such as environmental statistics or econometrics, *R Cookbook* provides a valuable desktop reference for R users that goes beyond the programming language into the statistics. Organized around topical themes and providing specific targeted guidance on accomplishing narrow tasks, *R Cookbook* aims to provide the reader with targeted methods addressing specific problems that the reader can already articulate. This follows in the tradition of other books in the O'Reilly Cookbook series, e.g., [Carlson and Richardson \(2006\)](#).

The book's organization aims to provide a desk reference for R. The book is structured as a discrete collection of recipes the user can follow to solve a stated problem and the author has focused on the abstract principles behind the recipe to provide a simple and usually elegant solution. Functions within R are not described in detail nor are the statistical methods behind the function.

The book's overall structure divides the recipes into 14 chapters collecting and focusing recipes by broad topic. Examples of chapters include "Data Transformations" and "Time Series Analysis" but the *R Cookbook* is richer than a list of chapters can produce. The book spans data management, core statistical topics such as sample analysis, linear regression, and statistical and data visualization, but also touches on matrix algebra and a full chapter called, "Useful Tricks," a hodge podge of recipes that do not fit elsewhere, and R's internal matters, such as the library search path. Also included in the first chapter are recipes for installing and starting R, as well as getting help for R.

A typical recipe is structured around three sections. The first, a problem statement, briefly describes the problem the reader might have in plain language by placing the reader (directly addressed as “you” throughout the text) in the driver’s seat and approaching the problem in the imperative. The problem is not usually a grand statement of a statistical procedure nor a complete analysis, but a simple step, such as Recipe 5.29 which offers to help the reader “combine the contents of two data frames” into one.

The problem is appropriately followed by a section listing the solution. The solution lists which function or sequence of functions addresses the problem. Each includes a one or two sentence explanation of the function and the key function arguments. There is no detail and the solution is not a substitute for the `help()` information for the given function. In many cases, cross referencing with the online help in R assists the reader in understanding the solution provided.

The interesting piece of each recipe is the associated discussion section. The discussion provides insight into how to apply the solution to the problem. Within the discussion, Teetor will also expound on the best applications and usage for the solution. Finally, the discussion section provides examples with data and graphs where appropriate.

A significant strength of the book lay in the end of the recipes. Many have a section called “See Also” that provides a quick reference to other recipes within the book that complement, support, or may be related to the recipe in question. An example is recipe 4.10, “Reading Tabular or CSV Data from the Web” that lists two important recipes (“Reading from CSV Files” and “Reading Files with a Complex Structure”). Other recipes point the reader directly to specific functions within R. With this, the reader does not necessarily need to know the exact method they are looking for as the cross references can walk a reader to the best solution for their problem.

Another strength of this book is Teetor’s grace and wit in describing the technical details of a complex solution. An example is in Recipe 9.8, “Testing the Mean of a Sample (*t* Test)” that calls the `t.test` function a “many-splendored thing” before suggesting related recipes. The conversational style adopted by the author cuts through the formality common among the academic treatments of R. This gives the book a readable aspect the average desk reference is lacking, and makes *R Cookbook* suitable for browsing when looking for new and interesting things to do with R.

There are two significant drawbacks to this book. The first is inherent in the format of a printed book. The numerous cross references and need for a simple method to search the book for the right recipe suggests it would be better served as a website, perhaps allowing user contributions. While the printed book offers some assurance of editorial oversight, the flexibility of a wiki or blog with multiple indexing methods (that is, tagging) would have vastly increased the simple value of this knowledge. This should not be interpreted to mean the cross referencing, indexing, and organization of the book are insufficient for its purpose. It is only a note that similar projects may wish to provide room to grow. A related issue is that the first chapter, on installing R and similar issues, is excessive when the information is readily available via the R website.

The second drawback is the simplicity of the contrived examples. Using real world data would have contravened the goal of an application agnostic text, but the ability to tie a method to a familiar outcome cannot be overstated. Unlike some other books on R, *R Cookbook* does not have an associated package that includes the data (or any supporting functions) that

back up the text. This keeps the reader from exploring the data-intensive examples from an established base line and hampers some of the experimental experience of learning.

These drawbacks do not significantly detract from the immediate purpose of the book. *R Cookbook*'s structure is intelligent without overwhelming the reader or attempting an encyclopedic approach to the material. The recipes are broken down into simple and appropriately sized pieces allowing cross references to link more complex examples. The drawbacks only limit the text to the pages between the covers.

As a desk reference, *R Cookbook* is not a replacement for other texts, such as Venables and Ripley (2002) or domain specific texts for social science, biostatistics, or other fields. However, *R Cookbook* is an excellent companion for other references on the market. This book is a useful reference for experienced R users and but it may also provide guidance to new R users through the adoption process.

References

Adler J (2009). *R in a Nutshell*. O'Reilly, Sebastopol.

Carlson L, Richardson L (2006). *Ruby Cookbook*. O'Reilly, Sebastopol.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer-Verlag, New York.

Reviewer:

James P. Howard, II

Mathematics and Statistics

University of Maryland University College

E-mail: jh@jameshoward.us or jphoward@faculty.umuc.edu

Twitter: [@howardjp](https://twitter.com/howardjp)

URL: <http://www.jameshoward.us/>