



---

# Journal of Statistical Software

January 2009, Volume 29, Book Review 4.

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

---

Reviewer: Byron Ellis  
AdBrite, Inc.

---

## Software for Data Analysis

John M. Chambers  
Springer-Verlag, New York, 2008.  
ISBN 978-0-387-75935-7. 498 pp. USD 79.95.  
<http://www.springer.com/978-0-387-75935-7>

---

## Introduction

Despite claims to the contrary *Software for Data Analysis* provides a comprehensive introduction to R and the basic concepts of statistical computing. In conjunction with lecture notes it could be a good text for use in a statistical computing course. However, a class that focuses on R as a tool for statistical analysis would probably fare better with another text.

## Contents

The text itself is laid out in a traditional fashion. The first three chapters provide R basics from interactive to the introduction of functions as units of packaging and then to the collection of functions into a script. Chapter 3 provides an introduction to R's often overlooked debugging system and should provide new material for all but the most experienced R users.

Chapters 4 and 5 complete the introduction to R's hierarchy of organization. Chapter 4 introduces the mechanics of R's package system and the highest level of code organization. Chapter 5 does the same for the organization of data through the introduction of R's object system, environments along with a discussion of closures as a unit of organization.

From the discussion of organizational mechanics the book then introduces R's computational facilities. Chapter 6, in particular, introduces the notion of the vector-based computational model that will likely be unfamiliar to both new programmers and experienced programmers coming from procedural languages such as C or Java.

After building up the computational model, Chapter 7 and 8 then focus on using R for specific tasks. In many ways these chapters are most like extended examples of the techniques and ideas presented in the first six chapters. Chapter 7 is a discussion of R's visualization models including both so-called base graphics as well as the **lattice** package. The discussion of the graphical system is brief, but there are several specific texts for both base graphics and the

**lattice** package. Chapter 8 focuses on computations with text in R and introduces the notion of R existing in a larger system by interacting with Perl to aid in the text processing task rather than attempting the entire thing in R.

Chapters 9 and 10 introduce “new” classes (aka S4 classes) and their method dispatch system. For developers familiar with other object-oriented systems these two chapters are the most important since they describe a system that might be unfamiliar to users of most “mainstream” languages. These two chapters are the most useful to developers who are accustomed to working with functions in scripts looking to move to a system that is more tightly integrated with the types of data being manipulated by these functions.

Chapters 11 and 12 provide an in-depth discussion of “foreign function” interfaces in R. Chapter 11 discusses the traditional R mechanisms for interfacing with C, Fortran and C++ code. Chapter 12 expands on this notion to include interfaces via other mechanisms, including databases and spreadsheets as a mechanism for communication.

Finally, Chapter 13 digs into the R interpreter itself. This chapter includes some discussion of the garbage collector and the R evaluator. It also discusses the internals of some of R’s more uncommon syntax such as replacement functions.

## Conclusion

In general, this is the sort of book that has a little bit of something for everyone except for perhaps the data analyst who will find little discussion of R’s statistical analysis packages. For the new developer there is comprehensive coverage of R’s syntax and basic building blocks. For the advanced user there is a nice discussion of the deeper aspects of the language, a fair amount of wisdom and best practices and probably a trick or two to be learned. However, this book is best targeted toward the intermediate R user who is looking to move from using scripts to the development of dissemination of packages with a complete system of organizing computations and data.

## Reviewer:

Byron Ellis  
AdBrite, Inc.  
San Francisco, CA, United States of America  
E-mail: [byron.ellis@gmail.com](mailto:byron.ellis@gmail.com)  
URL: <http://byronellis.name/>