



Journal of Statistical Software

October 2018, Volume 87, Book Review 1.

doi: 10.18637/jss.v087.b01

Reviewer: Pedro M. Valero Mora
Universitat de València

bookdown: Authoring Books and Technical Documents with R Markdown

Yihui Xie
Chapman & Hall/CRC, Boca Raton, 2017.
ISBN 9781138700109. 138 pp. USD 17.99 (P).
<https://bookdown.org/yihui/bookdown/>

Are you surprised to find a review of a book on software for formatting books in the Journal of Statistical Software? Well, do not be: there are strong arguments that make the **bookdown** package worth to be known by people who analyze data and then write books, reports, web sites, or e-books about their results. So, the **bookdown** package extends the literate programming facilities of **knitr** and the text-formatting capabilities of R Markdown to longer documents such as books or multipage sites by including automatic reference figuring, separated files for chapters or html pages, and easy export to \LaTeX book documents. After using it for a few months, I can say it is a very good option, if not the best, for books or academical writing. However, although the **bookdown** package works very well once you master it, there is a downside: in order to take advantage of its capabilities you have to get acquainted also with other packages, programming languages, formats, syntax, commands, environments, nomenclature, and else. As in the Bach quote: “It’s easy to play any musical instrument: all you have to do is hit the right keys at the right time and the instrument will play itself”. The **bookdown** package will pay off very quickly if you have invested a fair amount of time in the past learning the technologies that it integrates, but if you have not, then the reward will take longer and you may even wonder if it is worth of the effort. And, if you are not willing to acquire the ABCs of these technologies, then you might probably be better off sticking to alternatives such as, cough, the word processor that we all love to hate.

The first thing you have to know to take advantage of **bookdown** is, of course, R. If you are not familiar with it, and not interested in learning it, I think you can skip **bookdown** safely. But if you are not enticed by the currently most exciting statistical system it can be only because you are not interested in data analysis at all, and, in that case, why are you reading book reviews published in the Journal of Statistical Software?

The second thing you need to know is **knitr**, one of the implementations in R of literate programming, a programming paradigm in which natural language explains the logic of code that is interspersed with it, with the particularity that the code can be extracted automatically and executed as a normal program later on. This way, programs and documentation are not separated and programmers can write them following the logic of their thoughts. Literate

programming's first realization in R took place with the package **Sweave**, but **knitr** has been gaining traction among the R community since its inception a few years ago. **knitr** is a better option than **Sweave** nowadays because it uses R Markdown – more on this later – instead of \LaTeX , and also because it exports to .docx so that collaboration with people living in the Word-only world is possible. With **knitr**, the narrative part of text, the code from different programming languages such as R, Python or SQL, and the output – figures, tables or text – coexist in the same document. No more of keeping different windows for input and output of data analysis, and then another one – usually part of a word processor or similar – for the narrative part. No more of having a multitude of files residing on your hard disk with graphs, figures, tables, and text. No more of forgetting to update a result after performing a new analysis or correcting an error. (In fact, I can not help wondering why not all the statistical systems work in this way).

The third thing you need to know is R Markdown, a markup language that, as its name implies, implements the Markdown markup language in R. Other popular markup languages are HTML or \LaTeX . Markup languages use formatting codes embedded within the text to give indications about the final aspect that the document should have. These documents are then processed by programs such as in HTML viewers, or \LaTeX compilers in order to get the desired result.

Markup languages have several advantages as, for example, that the general aspect of the document can be manipulated easily – just change the template and your text can look formal, modern, suited for paper printing, for e-books, and so forth. However, it also has disadvantages such as that the formatting codes can interfere visually with the text and make reading somewhat more difficult than in WYSIWYG word processors, and that introducing the codes can be sometimes really cumbersome, as with the often long and laborious tags needed for formatting documents in HTML. However, here is where Markdown comes to the rescue: I can attest that their codes are both non-intrusive visually and very manageable. Besides, it is a breeze to learn the basics (I am writing this review with Markdown and it took me about ten minutes to kick off).

However, plain-vanilla Markdown cannot rival with another well-known markup language, \LaTeX , for static, paper-oriented, documents. Markdown was originally created for HTML output and its formatting capabilities are not really up to level of the ones in \LaTeX , but still another piece of software, **pandoc**, adapts the basic syntax of Markdown to multiple formats, permitting to transform documents coded in Markdown to \LaTeX , HTML and many more. **pandoc** is a command-line program that you can run on its own, but also can be called from within other programs. Indeed, with R Markdown, you do not need to interact directly with **pandoc** or \LaTeX and you only have to specify PDF output in yet another markup language called YAML (which stands for Yet Another Markup Language, of course).

So, the fourth thing you need to know to use **bookdown** is a bit of **pandoc**/YAML. Again, not really very complicated but I found that the information about their options was not easy to find.

Finally, if the word \LaTeX makes you think of terrible ordeals rather than of a useful software for formatting text, you'd be better off learning a little bit about its terminology, structure, and functionality before you set out with **bookdown**. You will not need to write anything directly in \LaTeX as far as you are content with **bookdown**'s output by default, but if you are somewhat finicky, and we all are sooner or later, then that is the way to go. Similarly, HTML

terminology is used profusely in the **bookdown** book so if you are completely illiterate on it, you may have trouble understanding some explanations.

Now that I have provided an overview of the different languages and technologies that the **bookdown** package integrates, it is time for providing an evaluation of the **bookdown** book itself. Good writers know their audience. Equally, smart readers are better off choosing well their books. My impression is that, if you are familiarized with the languages or programs mentioned previously, or at least with a few of them, this book is for you, but otherwise you may have a hard time following its explanations. By way of example, although the author indicates that the two first chapters are introductory and the other four discuss more advanced options, in Chapter 1, Section 1.4 the explanation of the difference between the strategies “Merge and Knit” versus “Knit and Merge” requires relatively advanced knowledge of how R works, or in Chapter 2, Section 2.1.3, it is declared that “You can also use math environments inside $\$$, $\$ \$$, $\$ \$ \$$, e.g.,” but what is a math environment, its purpose, or the dollar signs function is not mentioned, so it is assumed you know some \LaTeX . These are only two examples but the book has many of them. With respect to the other chapters, the author warns that Chapter 3 and 4 “may look very technical if you are not familiar with HTML/CSS and \LaTeX ”, Chapter 5 “the technical details are not important unless you do not use the **RStudio IDE**”, and finally “you may feel overwhelmed by the commands presented in Chapter 6”. So, in summary, beginners beware.

Of course, if you are well acquainted with all the languages or programs mentioned before, cutting all the unnecessary explanations and going directly to the advanced stuff may be perfect for you. From this perspective, the book gives advice about the fine points that only a top-expert could explain such as under-the-hood tricks, future developments, or buggy/unfinished parts. Indeed, that this book exists is really a nice thing, but I do not expect that many people will read it as an ordinary book as I did, i.e., starting at the first page and progressing page by page until the last one: most of the time people will only check it for solutions for specific problems, or, even more likely as the book is free online, they will stumble on it when doing internet searches. In fact, browsing the book online is probably the most convenient way of using this book, as the rather austere explanations are counterweighted with numerous links to other sources giving more complete accounts of the points presented.

This review is mainly on a book that refers to an R package, but I think that a valuation about the software is also fitting, and that is that the **bookdown** package is really splendid. This impression comes from my own experience using it for a small project: as I had already decided that a review of this book would require trying the software in practice, but I was not writing a book at that moment, I simply tried to reorganize the class notes for my introductory statistics course using **bookdown**. This involved reusing some material that I did not have the time for transforming, adding some links to datasets, and then producing an HTML version. A \LaTeX /PDF version of the notes was not a priority at this moment but knowing that I could make it in the future was clearly a plus. As a kick off, I looked at the instructions in page 3 of the book, and then, I basically followed my intuition about what I could change here or there, and where I had to write my stuff. Occasionally, I would have a glance to the book in search of some option I remembered to be somewhere or I would just search the web for a missing piece of information. The result? The instrument started to play itself mostly very well. Not that I pressed the keys at the right time all the time, but I managed a set of HTML pages that looked reasonably elegant, worked fine, and my students seemed to find easy to use.

In conclusion, **bookdown** is a very useful package that combines capabilities of other packages and programs already of much interest for statisticians. Although the book that explains it is not the best entry point for non-initiated, it can be a useful source to search for solutions when the default options do not satisfy you. All in all, if you work with data and text, you really could do a lot worse than using **bookdown** for making your web sites, books or reports.

Reviewer:

Pedro M. Valero Mora

Universitat de València

Departamento de Metodología de las Ciencias del Comportamiento

Valencia, Spain 46022

E-mail: valerop@uv.es

URL: <http://www.uv.es/valerop/>