

Data Management Using R

Myo Minn Oo

2021-12-04

Contents

1	About	5
1.1	Usage	5
1.2	Render book	5
1.3	Preview book	6
2	Getting Started	7
2.1	Downloading and installing R	7
2.2	RStudio	7
2.3	R Packages	7
2.4	Workflows	7
2.5	Searching Help	7
2.6	Using this book	7
2.7	Disclaimer	8
3	Using existing data files	9
3.1	Working directories	9
3.2	Reading data files	9
3.3	Saving data files	11
4	Parts	13
5	Footnotes and citations	15
5.1	Footnotes	15
5.2	Citations	15

6	Blocks	17
6.1	Equations	17
6.2	Theorems and proofs	17
6.3	Callout blocks	17
7	Sharing your book	19
7.1	Publishing	19
7.2	404 pages	19
7.3	Metadata for sharing	19

Chapter 1

About

This is a *sample* book written in **Markdown**. You can use anything that Pandoc’s Markdown supports; for example, a math equation $a^2 + b^2 = c^2$.

1.1 Usage

Each **bookdown** chapter is an .Rmd file, and each .Rmd file can contain one (and only one) chapter. A chapter *must* start with a first-level heading: **# A good chapter**, and can contain one (and only one) first-level heading.

Use second-level and higher headings within chapters like: **## A short section** or **### An even shorter section**.

The `index.Rmd` file is required, and is also your first book chapter. It will be the homepage when you render the book.

1.2 Render book

You can render the HTML version of this example book without changing anything:

1. Find the **Build** pane in the RStudio IDE, and
2. Click on **Build Book**, then select your output format, or select “All formats” if you’d like to use multiple formats from the same book source files.

Or build the book from the R console:

```
bookdown::render_book()
```

To render this example to PDF as a `bookdown::pdf_book`, you'll need to install XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.org/tinytex/>.

1.3 Preview book

As you work, you may start a local server to live preview this HTML book. This preview will update as you edit the book when you save individual .Rmd files. You can start the server in a work session by using the RStudio add-in “Preview book”, or from the R console:

```
bookdown::serve_book()
```

Chapter 2

Getting Started

Introduction to R

2.1 Downloading and installing R

2.2 RStudio

download and install start screen console right or left workspace

2.3 R Packages

package masking essential packages for the book tidyverse mStats

2.4 Workflows

pipng

2.5 Searching Help

2.6 Using this book

All chapter sections start with a second-level (##) or higher heading followed by your section title, like the sections above and below here. You can have as many as you want within a chapter.

2.7 Disclaimer

First, this book is largely adapted from the book *Data Management Using Stata: A Practical Handbook*, Second Edition. I tried to replicates examples and exercises as close to the book as possible.

As Michael Mitchell recommended in his book, I also encourage you to type in the codes, see the results, and run the examples from this book for yourself. In this way, you will understand well how data management in R works.

The following codes create a `data` folder in current working directory and download the datasets from the book `dmus2`. We will discuss how to set current working directory in a minimal and effective way in a later section.

```
## create a directory under current working folder
if (!dir.exists("data")) {
  dir.create("data")
}

## set the URL to dmus2 datasets in zipped format
dmus2 <- "https://www.stata-press.com/data/dmus2/dmus2.zip"

## create a temporary file to store the zipped file
temp <- tempfile()

## download the dmus2 zipped file into temporary file
download.file(dmus2, temp)

## extract the dmus2 datasets into the data folder
unzip(temp, exdir = "data/", junkpaths = TRUE)
```

More online resources and recommended books for R

Chapter 3

Using existing data files

You have some data that you are eager to analyze using R. Before you can analyze the data, you must first read the data into R. This chapter describes how you can read common types of data files into R.

3.1 Working directories

3.2 Reading data files

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 3.1.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Don't miss Table 3.1.

```
knitr::kable(
  head(pressure, 10), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

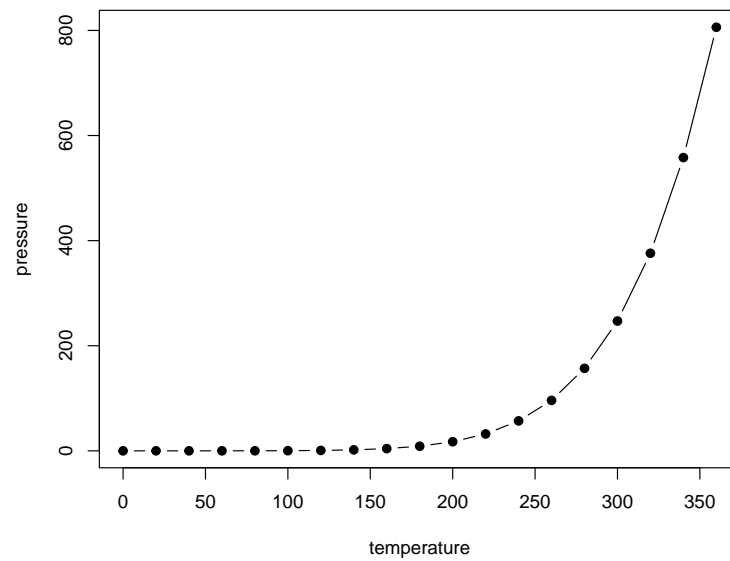


Figure 3.1: Here is a nice figure!

Table 3.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

3.2.1 Reading CSV spreadsheets

3.2.2 Reading Excel spreadsheets

3.2.3 Reading Stata datasets

3.2.4 Reading SPSS datasets

3.2.5 Reading other types

delimited types entering data directly

3.3 Saving data files

3.3.1 Saving CSV spreadsheets

3.3.2 Saving Excel spreadsheets

3.3.3 Saving Stata datasets

3.3.4 Saving SPSS datasets

3.3.5 Saving other types

delimited types entering data directly

Chapter 4

Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Chapter 5

Footnotes and citations

5.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

5.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2021] (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [Xie, 2015] (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 6

Blocks

6.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (6.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (6.1).

6.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 6.1.

Theorem 6.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

6.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Chapter 7

Sharing your book

7.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

7.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

7.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book—all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

<https://pkgs.rstudio.com/bookdown/reference/gitbook.html>

Or use:

```
?bookdown::gitbook
```

Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL <http://yihui.org/knitr/>. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2021. <https://github.com/rstudio/bookdown>.