A Minimal Book Example

John Doe

2023-05-18

Contents

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

Preface

This book is dedicated to my wife Balqisu.

To all the young scientists who would benefit from this book and colleagues who through every small or big way helped me write this, I say a big thank you. **?**

The motivation for writing this book comes from the dire and urgent need for many, especially those in developing countries to have and use a freely available yet sophisticate statistical software. In regions where buying a statistical software is often beyond the means of most persons R comes as a great relief. There is a huge gap between theoretical and practical knowledge of using statistical applications by many scientists. **R**, an open source statistical software offers a unique but important opportunity to bridge this gap.

The focus of this book is to introduce data analysis and R to persons with very little knowledge in both. This step by step introduction to data analysis and R is deliberately organised to have very limited text but lots of practical examples. Readers are advised to replicate the examples as they read along, at least to have a feel of how these work in R.

The book comes with a set of data in many file formats. Some are fictitious while a few are real data modified for public use. The datasets are freely available online at https://dl.dropboxusercontent.com/u/1673032/Intro_R_STI_Data_14_10_2014.rar.

As this is essentially an introductory text, it is set up to sparingly use packages1 other than the R base packages that comes with the default setup. Occasionally however other packages are mentioned especially when it comes to the graphics where the "lattice" and "ggplot2" packages are used. These are kept to the minimum as new readers may find it confusing when they are constantly being asked to download and install one package after the other.

As a general rule for this book, functions and their options are printed in

monospaced font with opening and closing brackets. Packages are written in monospaced without brackets. Console output are written in monospaced font and with a grey background. "Urls" and internet links are written in hypertext format for easy access to these sites once a reader is connected to the internet. Specific datasets are written in courier new font with a file extension such as .csv, .txt etc.

Chapter 3

Introduction

3.1 The statistical data analyst

Statistical data analysis is more than just using a computer software to generate a set of results. It involves the basic understanding of the data type being analysed, the best way to analyse and then present such data to make sense to the population at large. Thus the data analyst :

1. Must understand the genesis (study methodology) involved in obtaining the data in the first place. Conclusions from the same data may be different depending on the study methodology used and hypothesis being tested. It is very prudent therefore that the statistical data analyst be involved in the data collection process right from the beginning.
2. Should be able to point out errors in the data collection process in the very early stages to avoid wasting valuable resources on data that may not answer the question at hand.
3. Provide valuable advice on the best method of analysing the data at hand.
4. Perform the analysis in a scientifically sound manner applying the most current and statistically appropriate principles.
5. Be able to present the result of the analysis in a manner that makes it easy for literally all persons without statistical and analytical expertise to completely understand with the least effort. This requires the statistical data analyst to be in a position to explain the analysis in the ”layman’s language”.
6. Finally the data analyst must know his limit. There are often instances where the analyst should seek ”professional” help even though he/she may feel he/she is on the right path. It never hurts to seek a second opinion from your peers.

It therefore goes without saying from the prior discussion that the data analyst must have a firm understanding of statistical and research methodology.

3.2 Statistical software

Some years back, statistical analysis was one of the most tedious processes done mainly by dedicated statisticians. With the advent of computers and statistical software it has become rather handy with many advantages but some disadvantages as well.

The main advantages are:

1. The tremendous speed with which large data is processed and results obtained.
2. The accuracy of the statistical calculations performed. Computers do not make mistakes but one has to beware of rounding in some software. Some software can perform calculations to specific number of decimal places. Therefore when one is confronted with a figure such as 1.00377655432 the software may work with 1.0037765, leaving out the last four digits. Calculations using this truncated value is likely to have different result from the non-truncated figure, thus affecting the accuracy ofthe result.
3. Many modern statistical software have the ability to read data from varied sources and formats. This makes it easy to transfer data from one software to another without having to re-enter the data collected into the second computer or software.This transferability has enabled the use of other digital equipment such as smart phones, personal digital assistants and tablets for data collection. Data collected in this manner is said to be ready for cleaning and analysis, bypassing the data entry stage.
4. Plotting graphs is one of the most important uses of modern day computerised data analysis. Statistical software tend to make this rather tedious process almost hassle free and accords one the ability to redo the plot from scratch at the click of a button.

Despite all these advantages many disadvantages are also inherent in the use of computers and statistical software. Some include:

1. Many persons with very little or no statistical knowledge can manipulate data and come up with conclusions that often tend to be very spurious. The cliche "Garbage In Garbage Out" could not apply better than in this situation.
2. Many commonly used software tend to be very reliable and accurate. With the large number of often user written statistical software freely available online one needs to be cautious of the output generated. Some of these could actually be wrongly written codes producing faulty results.
3. Unfortunately the most used, reliable and accurate statistical software tend to be expensive as well. This not withstanding there are few, such as R that combines free and open source with versatility and reliability. This forms the basis for my choice of R for this book.

3.3 Obtaining and installing R

R is a free software programming language and environment for data manipulation, calculation and graphical display. It can run on Windows, MacOS X and Unix systems. It has great applications in many academic fields including mathematics, economics and epidemiology. This capability has been enhanced by the many packages written by individuals over the years. R has a great advantage of being able to handle many datasets simultaneously. However this functionality comes at a cost which would be discussed in the subsequent chapters. R also has great graphics functionality but requires practice to harness full use of this.

Several advanced statistical and mathematical functionality such as regression and survival analysis are also implemented in R.

R and its many packages are obtainable free from http://cran.r-project.org/. The most current version as at writing this book is R-3.2.1-win32.exe. The Windows version is installable on both the 32 and 64 bit operating systems. Download the base file from https://cran.r-project.org/bin/windows/base/, save on your computer and install (preferably as an administrator) by double clicking the executable file and following the on-screen instructions.

3.4 Running R

After installation, two shortcuts similar to the figure below are placed on the desktop (windows version). On a Windows 8 64 bit machine both shortcuts for the 32 and 64 bit versions are placed on the desktop. Double click the desktop icon to launch R. A welcome screen as shown below is displayed in the console launched.

The $>$ at the end of the output above is the command prompt and indicates R is ready to be used. The R Console before you at this stage if running for the first time will show a wide grey screen with a smaller white screen. The white screen is the referred to as the console. We can type straight and obtain requisite output on the console but I will strongly discourage this habit.

3.5 R as a calculator

The applicability of R in data analysis starts from its use as an overgrown calculator. In this section we learn this simple task. Calculations in R can be achieved simply by inputting and executing using the appropriate mathematical operational symbols. These are

$+$ addition
$*$ multiplication
$-$ substraction

/ division
∧ raised to power or exponent

A couple of examples

```
7+8
[1] 15

(12-6)/7
[1] 0.8571429
```

3.6 Name assignment in R

All objects in R can be assigned names, values or attributes. This assignment is achieved by using "->" or "<-". The example below assigns the values 5 to Kwame and 10 to Ama, their respective ages.

```
Kwame <- 5
Ama <- 10
```

Calling these names yield the values (ages) assigned. We can do that by typing the names in the console.

```
Kwame
[1] 5
Ama
[1] 10
```

To add the ages of Kwame and Ama we can just do

```
Kwame + Ama
#> [1] 15
```

Chapter 4

Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

4.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
    - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world` = `# Hello world {#hello-world}`.
    - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.
2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 4.
    - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

4.2 Captioned figures and tables

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 4.1.

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

Don't miss Table 4.1.

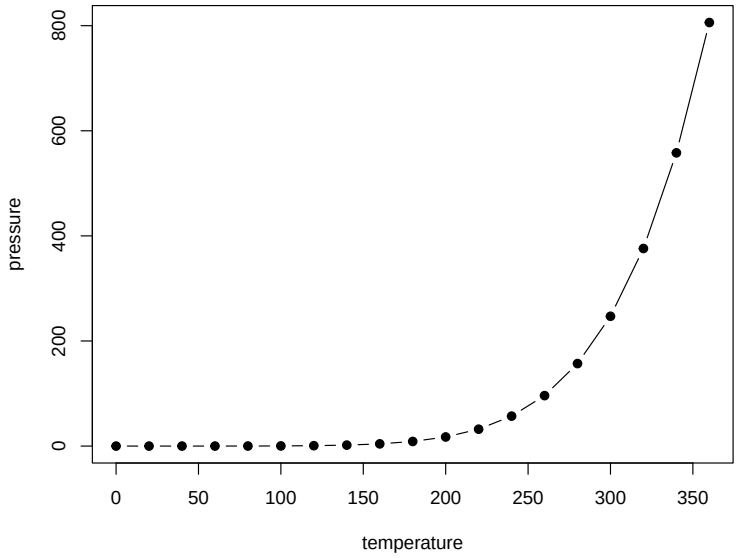

Figure 4.1: Here is a nice figure!

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

Table 4.1: 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

Chapter 5

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 6

Footnotes and citations

6.1 Footnotes

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

6.2 Citations

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

For example, we are using the **bookdown** package (Xie, 2023) (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 `bs4_book` theme makes footnotes appear inline when you click on them. In this example book, we added `csl: chicago-fullnote-bibliography.csl` to the `index.Rmd` YAML, and include the `.csl` file. To download a new style, we recommend: https://www.zotero.org/styles/

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

[1] This is a footnote.

Chapter 7

Blocks

7.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{7.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (7.1).

7.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 7.1.

Theorem 7.1. *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.

7.3 Callout blocks

The `bs4_book` theme also includes special callout blocks, like this `.rmdnote`.

You can use **markdown** inside a block.

```
head(beaver1, n = 5)
#>   day time  temp activ
#> 1 346  840 36.33     0
#> 2 346  850 36.34     0
#> 3 346  900 36.35     0
#> 4 346  910 36.42     0
#> 5 346  920 36.55     0
```

It is up to the user to define the appearance of these blocks for LaTeX output.

You may also use: `.rmdcaution`, `.rmdimportant`, `.rmdtip`, or `.rmdwarning` as the block name.

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 8

Sharing your book

8.1 Publishing

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

8.2 404 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.

8.3 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 `bs4_book` provides enhanced metadata for social sharing, so that each chapter shared will have a unique description, auto-generated based on the content.

Specify your book's source repository on GitHub as the `repo` in the `_output.yml` file, which allows users to view each chapter's source file or suggest an edit. Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/bs4_book.html

Or use:

```
?bookdown::bs4_book
```

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2023). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.34.