

R4Stats

Colin Madland

2022-06-26

Contents

About	5
Usage	5
Render book	5
Preview book	6
1 Getting Started with R	7
1.1 Problems	7
2 Univariate data	11
Levels of measurement	11
2.1 Data Vectors	12
3 Parts	15
4 Footnotes and citations	17
4.1 Footnotes	17
4.2 Citations	17
5 Blocks	19
5.1 Equations	19
5.2 Theorems and proofs	19
5.3 Callout blocks	19

6	Sharing your book	21
6.1	Publishing	21
6.2	404 pages	21
6.3	Metadata for sharing	21

About

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

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.

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/>.

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 1

Getting Started with R

1.1 Problems

```
1+2*(3+4)
```

```
## [1] 15
```

```
(4^3)+(3^(2+1))
```

```
## [1] 91
```

```
sqrt(4+3)*(2+1)
```

```
## [1] 7.937254
```

```
(1+(2*3^4))/(5/6)-7
```

```
## [1] 188.6
```

```
(0.25 - 0.2) / (0.2 * (1 - 0.2)/100)^(1/2)
```

```
## [1] 1.25
```

```
x <- 2
y <- 3
z <- 4
w <- 5

x*y*z*w
```

```
## [1] 120
```

```
rivers
```

```
## [1] 735 320 325 392 524 450 1459 135 465 600 330 336 280 315 870
## [16] 906 202 329 290 1000 600 505 1450 840 1243 890 350 407 286 280
## [31] 525 720 390 250 327 230 265 850 210 630 260 230 360 730 600
## [46] 306 390 420 291 710 340 217 281 352 259 250 470 680 570 350
## [61] 300 560 900 625 332 2348 1171 3710 2315 2533 780 280 410 460 260
## [76] 255 431 350 760 618 338 981 1306 500 696 605 250 411 1054 735
## [91] 233 435 490 310 460 383 375 1270 545 445 1885 380 300 380 377
## [106] 425 276 210 800 420 350 360 538 1100 1205 314 237 610 360 540
## [121] 1038 424 310 300 444 301 268 620 215 652 900 525 246 360 529
## [136] 500 720 270 430 671 1770
```

```
Orange
```

```
## Tree age circumference
## 1 1 118 30
## 2 1 484 58
## 3 1 664 87
## 4 1 1004 115
## 5 1 1231 120
## 6 1 1372 142
## 7 1 1582 145
## 8 2 118 33
## 9 2 484 69
## 10 2 664 111
## 11 2 1004 156
## 12 2 1231 172
## 13 2 1372 203
## 14 2 1582 203
## 15 3 118 30
## 16 3 484 51
## 17 3 664 75
## 18 3 1004 108
```


## 19	3	1231	115
## 20	3	1372	139
## 21	3	1582	140
## 22	4	118	32
## 23	4	484	62
## 24	4	664	112
## 25	4	1004	167
## 26	4	1231	179
## 27	4	1372	209
## 28	4	1582	214
## 29	5	118	30
## 30	5	484	49
## 31	5	664	81
## 32	5	1004	125
## 33	5	1231	142
## 34	5	1372	174
## 35	5	1582	177

```
mean(Orange$age)
```

```
## [1] 922.1429
```

```
max(Orange$circumference)
```

```
## [1] 214
```


Chapter 2

Univariate data

Levels of measurement

The view in most textbooks is from Stanley Smith Stevens (1964) ::: {.definition} **Nominal**

Such data is qualitative or descriptive, but not numeric. An example might be the name of a person or the town they are from, or the number on a bib a runner wears in a race. :::

Definition 2.1. Ordinal

Ordinal data is data with some order, so that we can sort the data from largest to smallest. An example might be the place a runner takes in a race.

Definition 2.2. Interval

Interval data is ordinal data where the difference between two values has some interpretation. The clock time a person finishes might be an example. If we know runner A finishes at noon and runner B at 1PM then we know that runner B took longer. Since we haven't specified when they started, we don't know what percent longer though.

Definition 2.3. Ratio

Ratio data has a meaningful 0. If we record not the time of finishing, but the time since starting, then 0 has a meaning and we can take a ratio of the total time for runner A and B to compare the two.

However, working with data on a computer is different, requiring different categories...

Definition 2.4. Factor

When we look at many variables, some may simply record categories used to group the data. In R we will use *factors* to store these variables. An example might be the browser a user has used to view a website, as gleaned from a log.

Definition 2.5. character

Some categorical data are factors, but others are really just identifiers, and are not used for grouping. An example might be a user's IP address. Difference can be thought of as distinguishing between *categorizing* a case or *characterizing* a case. While both factor and categorical data are *nominal*, we keep the distinction as we will interact with the data differently.

Definition 2.6. discrete

Discrete data comes from measurements where there are essentially only distinct and separate possible values that can be counted. For example, the number of visits a person makes to a website will always be integer data, as will other counting data.

Definition 2.7. continuous

Data which could conceivably come from a continuum of variables. The recording of time in milliseconds of a visit to a website might be such data. A useful distinction is that for discrete data we expect that cases will share values, whereas for continuous data this will be impossible, or at least very unlikely. We can also turn continuous data into discrete data by truncating (record the minute instead of the millisecond) or by binning. Rather than draw distinctions between ordinal, interval, and ratio, it is more important for statistical theory - in finding a model for the recorded data - to know if the data is discrete or continuous.

Definition 2.8. time and date

Though we just saw that time and date can be considered continuous or discrete, for computers there are often separate ways to handle date and time data. Issues that complicate matters are leap days and time zones, but also scale (some people want millisecond data)

Definition 2.9. hierarchical

while much data is several measurements for several cases and fits nicely onto a rectangular spreadsheet, data on networks does not fit this

2.1 Data Vectors

Suppose the number of whale beachings in Texas during the 1990s was

74 122 235 111 292 111 211 133 156 79

We can combine these into a data set through

```
whale <- c(74, 122, 235, 111, 292, 111, 211, 133, 156, 79)
```

The `whale` object is a *data vector*.

the size of the data set is retrieved with the `length` function

```
length(whale)
```

```
## [1] 10
```

```
sum(whale)
```

```
## [1] 1524
```

Average can be found with combining the two...

```
sum(whale)/length(whale)
```

```
## [1] 152.4
```

or

```
mean(whale)
```

```
## [1] 152.4
```

2.1.1 Vectorization

```
whale - mean(whale)
```

```
## [1] -78.4 -30.4 82.6 -41.4 139.6 -41.4 58.6 -19.4 3.6 -73.4
```

```
whale^2 / length(whale)
```

```
## [1] 547.6 1488.4 5522.5 1232.1 8526.4 1232.1 4452.1 1768.9 2433.6 624.1
```

```
sqrt(whale)
```

```
## [1] 8.602325 11.045361 15.329710 10.535654 17.088007 10.535654 14.525839  
## [8] 11.532563 12.489996 8.888194
```


Chapter 3

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 4

Footnotes and citations

4.1 Footnotes

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

4.2 Citations

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

For example, we are using the **bookdown** package [?] (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 5

Blocks

5.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (5.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (5.1).

5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 5.1.

Theorem 5.1. *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>.

5.3 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 6

Sharing your book

6.1 Publishing

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

6.2 404 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.

6.3 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.