

PS9594A: Computational Text Analysis

Dr. Sebastián Vallejo Vera | Western University

2023-12-21

Contents

“Computational Text Analysis”	5
0.1 Software and Packages	5
0.2 Datasets	5
0.3 Acknowledgments	6
1 Hello bookdown	7
1.1 A section	7
2 Cross-references	9
2.1 Chapters and sub-chapters	9
2.2 Captioned figures and tables	9
3 Parts	13
4 Footnotes and citations	15
4.1 Footnotes	15
4.2 Citations	15
5 Blocks	17
5.1 Equations	17
5.2 Theorems and proofs	17
5.3 Callout blocks	17

6	Sharing your book	19
6.1	Publishing	19
6.2	404 pages	19
6.3	Metadata for sharing	19

“Computational Text Analysis”

This is the dedicated site for the course PS9594A: “Computational Text Analysis” at Western University, taught by Sebastián Vallejo Vera. In each week, you will find the code, exercises, and slides for corresponding topic.

Before you start, check the required software and packages below. Also, don’t forget to read the Syllabus and the Perusall with the readings for the course. This site will be corrected/updated throughout the semester.

0.1 Software and Packages

For the first part of this course (Weeks 1 - 5), we will be mainly using R. For the second part of the this course (Weeks 6 - 11), we will use a combination of R and Python. I will assume that you are familiar with R language, RStudio, and R packages. If you are not, please come to office hours and I can help you out¹. In R the main packages, we you will need to have the following packages installed:

- tidyverse (we will be piping)
- tidylog (helps keep track of what your are pipins)
- tidytext (great for working with text)
- quanteda (stands for “Quantitative Analysis of Textual Data”)

0.2 Datasets

Throughout the class, we will be using a number of sample datasets. Access to these datasets will be provided directly on the code. For your Final Essay, you can use one of the following datasets (or, even better, you can use your own):

¹It might even be faster to get one of your classmates or colleagues to help you out.

- data 1 + link
- data 2 + link
- data 3 + link

0.3 Acknowledgments

The organization of the first part of this course (Weeks 1 - 5) and the format of the assignments are borrowed from Christopher Barrie's excellent course on "Computational Text Analysis", a syllabus from the prolific Tiago Ventura, and Grimmer, Roberts, and Stewart's excellent book, "Text as data: A new framework for machine learning and the social sciences". The code used throughout the course is a patchwork of my own code, but my own code borrows heavily from the internet (but that's true for all code). I try my best to give credit to the original authors of the code (when and if possible).

Chapter 1

Hello bookdown

All chapters start with a first-level heading followed by your chapter title, like the line above. There should be only one first-level heading (#) per .Rmd file.

1.1 A section

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.

An unnumbered section

Chapters and sections are numbered by default. To un-number a heading, add a {.unnumbered} or the shorter {-} at the end of the heading, like in this section.

Chapter 2

Cross-references

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

2.1 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 2.
 - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

2.2 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 2.1.

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

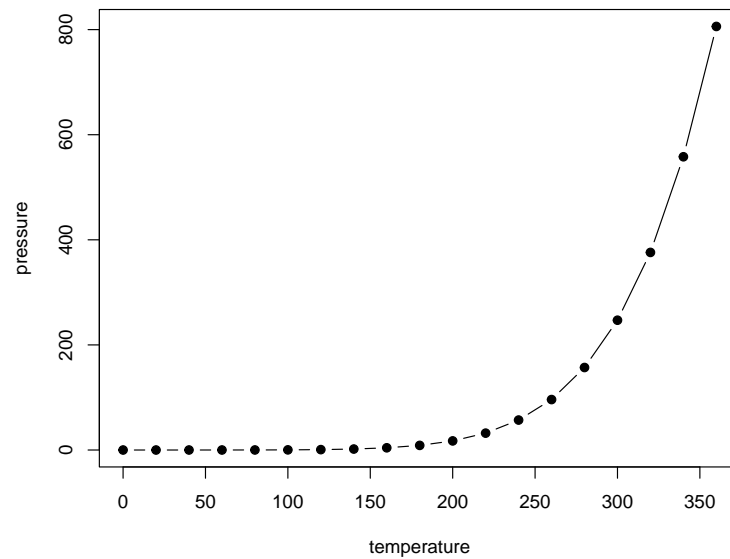


Figure 2.1: Here is a nice figure!

Don't miss Table 2.1.

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

Table 2.1: 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 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 [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 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.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2023. URL <https://CRAN.R-project.org/package=bookdown>. R package version 0.37.