

Basic Statistical Computing and Data Science
Using R

Contents

1	Basic Installation	5
1.1	R and RStudio	5
1.2	LateX	6
1.3	Orientation	6
1.4	Libraries in R	6
1.5	Markdown	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

R is a programming language and open-source software environment used for statistical computing, data analysis, and visualization. It was created by Ross Ihaka and Robert Gentleman at the University of Auckland, in the mid-1990s. R is widely used in both academia and industry for a variety of data analysis tasks, from simple data cleaning and exploration to advanced statistical modeling and machine learning.

The main strengths of R is its flexibility and ease of use. R has a rich set of

built-in functions and packages, but users can also write their own functions and packages, making it possible to customize and extend the language to suit their needs. R has a large and active community of users and developers, who contribute new packages and functionality to the language on a regular basis. This means that R is constantly evolving, and new tools and techniques are being added all the time. It also means that if you are having trouble with R, finding the answer is often only a Google search away.

R is particularly well-suited for statistical analysis and data visualization. It has a wide range of built-in statistical functions for data analysis, including basic descriptive statistics, hypothesis testing, regression analysis, and time series analysis. R also has powerful data visualization tools, with a variety of built-in plotting functions and packages for creating charts, graphs, and interactive visualizations. These visualization tools allow users to explore data and communicate their findings in a clear and effective way.

Another key feature of R is its ability to handle large datasets. R can easily handle datasets with millions of observations and hundreds of variables, making it a popular choice for big data analysis. R also has tools for working with messy or incomplete data, including functions for data cleaning, imputation, and transformation.

This book is a compilation of videos that will help you learn R. We will walk you through everything from installing R and RStudio to installing packages to basic statistics and some deeper data science algorithms. We will spend time discussing how to use R for statistical calculations and what the output that R generates means. Hopefully, when you are finished with this book, you will have a deeper understanding of statistics and data science and how these techniques can be applied to real life.

Chapter 1

Basic Installation

1.1 R and RStudio

R has become a popular tool for machine learning and artificial intelligence applications. It has a wide range of packages for machine learning, including tools for clustering, classification, and regression analysis. R is also used for deep learning and neural network analysis.

RStudio is an integrated development environment (IDE) for R, designed to make it easier to write and run R code. It was created by the RStudio team, led by JJ Allaire, and first released in 2011. RStudio has become a popular tool for R users, particularly those who are new to the language.

One of the key features of RStudio is its user-friendly interface. The IDE has a clean and intuitive layout, with multiple panes that allow users to view and edit their code, interact with the console, and view plots and other output. RStudio also has a variety of built-in tools and features that make it easier to write and debug R code, including syntax highlighting, code completion, and integrated debugging tools.

RStudio also has a variety of tools for working with projects and packages. Users can create and manage projects, which are collections of R code, data, and other files that are organized in a single directory. RStudio also has a package manager, which allows users to install and manage R packages, and create their own packages.

Together, R and RStudio form a powerful toolset for data analysis and statistical modeling. Let's begin by getting R and RStudio downloaded.

How to download R and RStudio

1.2 LaTeX

LaTeX is a document preparation system that allows users to typeset and format high-quality documents, particularly those with mathematical equations, formulas, and symbols.

RStudio integrates with LaTeX by allowing users to create and compile LaTeX documents directly from within the RStudio environment. This means that users can seamlessly integrate their R code and output into their LaTeX documents, making it easy to create reproducible research reports, manuscripts, and presentations.

How to download LaTeX

1.3 Orientation

In this tutorial, we will learn how to open an RScript. We will learn about the numeric data type in RStudio and how to perform operations using the numeric data type; such as addition, subtraction, multiplication, division, square roots, exponentiation, and logarithms.

R Orientation Part 1

In this tutorial, we will learn about the integer data type in RStudio. We will learn how to perform operations with the integer data type and the logical data type. We will also learn about the character data type.

R Orientation Part 2

1.4 Libraries in R

In R, a library is a collection of pre-written R functions that extend the functionality of the base R system. Libraries can be loaded into R to provide additional capabilities for data manipulation, visualization, statistical analysis, and more.

There are many libraries available in R, and they can be installed from the Comprehensive R Archive Network (CRAN), GitHub, or other sources. To use a library in R, it must first be installed and then loaded into the R session using the `library()` function. Once a library is loaded, its functions and other objects become available for use in the R session.

Libraries in R can be incredibly useful for performing specific tasks that may not be included in the base R system. They can also save time and effort by providing pre-written code for common operations, allowing users to focus on their specific analysis or research questions rather than writing code from scratch. Finally, because R is an open-source language, users can contribute their own libraries and share them with the wider R community.

This tutorial will give you a brief introduction to libraries in R.

R Libraries

1.5 Markdown

R Markdown is a markup language that enables the creation of dynamic and reproducible documents, reports, and presentations that integrate text, code, and data. R Markdown is based on the Markdown syntax and is designed to facilitate the creation of documents that combine R code, data analysis, and narrative text.

With R Markdown, users can create documents in a variety of formats, including HTML, PDF, Word, and PowerPoint, among others.

R Markdown documents are highly customizable, allowing users to control the appearance of their documents using a combination of Markdown syntax and customizable options. R Markdown also provides tools for including tables, figures, and interactive visualizations in documents.

R Markdown is fully integrated with RStudio, making it easy to create, preview, and publish R Markdown documents from within the RStudio environment. This integration streamlines the workflow for creating and sharing reproducible research reports, making R Markdown a valuable tool for researchers, data scientists, and other professionals who need to communicate their findings in a clear and reproducible way.

R Markdown

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)
```

Don't miss Table 2.1.



Figure 2.1: Here is a nice figure!

```
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 0.33.