Programming for Data Science in R

Efstathios (Stathis) D. Gennatas, MBBS AICSM PhD

2020-12-05

Contents

PDSR

Welcome to PDSR!
This book is aimed as an introductory- to intermediate-level R programming learning resource.
It is the online book for UCSF Biostat 213/4, currently being updated regularly.

EDG
UCSF[1] LCM[2],
San Francisco, CA, November 2020

[1]https://www.ucsf.edu/
[2]https://lambdamd.org/

Chapter 1

Introduction

Throughout this book you will see boxes with R code followed by its output, if any. The code (or input) is decorated with a teal border on the left to separate it from its output, like in the following example:

```
x <- rnorm(200)
x[1:20]
```

```
 [1] -0.82847083 -1.58936059  0.26891099  0.16006956  0.97151651  1.06612773
 [7]  2.12868481 -0.82863315  0.20806149  0.79409205 -1.27790280 -0.70165927
[13]  0.43853741 -1.36409009  1.38606876  1.81914918  0.05142488  0.21419057
[19] -0.84332025  0.01965687
```

Notice that R adds numbers in brackets in the beginning of each row. This happens when R prints the contents of a vector. The number is the integer index of the first element in that row. Therefore, the first one is always `[1]` and the number of the subsequent rows depends on how many elements fit in each line. If the output is a single element, it will still have `[1]` in front of it.

Also notice that if we enclose the assignment operation of a variable in parentheses, this prints the resulting value of the variable. Therefore, this:

```
(y <- 4)
```

```
[1] 4
```

is equivalent to:

```
y <- 4
y
```

```
[1] 4
```

Note that if you mouse over the input code box, a clickable "Copy to clipboard" appears on the top right of the box allowing you to copy paste into an R session