

# Learning R

Peter K Dunn



# Contents

<b>I</b>	<b>Introduction to R tutorials</b>	<b>1</b>
<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Why learn R? . . . . .	3
1.2	About R . . . . .	4
1.2.1	What is R? . . . . .	4
1.2.2	What are the differences between S-Plus and R? . . . . .	5
1.3	Basic R commands . . . . .	5
<b>2</b>	<b>Basic arithmetic in R</b>	<b>7</b>
<b>3</b>	<b>Getting help in R</b>	<b>11</b>
3.1	Help in R itself . . . . .	11
3.2	Help outside R . . . . .	11
3.3	Using R functions . . . . .	12
3.3.1	An initial look . . . . .	12
3.3.2	Another look . . . . .	14
<b>4</b>	<b>Loading data</b>	<b>17</b>
4.1	Introduction . . . . .	17
4.2	Entering data directly . . . . .	17
4.3	Loading text data files . . . . .	18
4.4	More on loading data . . . . .	19
4.5	Data frames . . . . .	21
4.6	Examining data . . . . .	22
4.6.1	Initial examination of the data frame . . . . .	22
4.6.2	Initial examination of the variables . . . . .	23
4.6.3	Initial plotting of the data . . . . .	24
<b>5</b>	<b>Matrices in R</b>	<b>29</b>

<b>6</b>	<b>Plotting and graphics in R</b>	<b>31</b>
6.1	Plotting data . . . . .	31
6.2	Plotting functions . . . . .	33
<b>7</b>	<b>Simple analyses</b>	<b>35</b>
7.1	Formula notation . . . . .	35
7.2	Function output . . . . .	37
<b>8</b>	<b>How to run your code in R</b>	<b>41</b>
<b>9</b>	<b>More on plotting</b>	<b>43</b>
9.1	Introduction . . . . .	43
9.2	Adding labels . . . . .	44
9.3	Changing the plotting characters . . . . .	45
9.4	Changing colours . . . . .	45
9.5	Changing the format of the axes . . . . .	45
9.6	Selecting parts of a data frame . . . . .	46
9.7	Adding a legend . . . . .	47
9.8	Saving figures . . . . .	48
9.9	Other options . . . . .	49

# Part I

## Introduction to R tutorials



# Chapter 1

## Introduction

### 1.1 Why learn R?

- R is available for most operating systems (including Windows, Mac, and most flavours of Linux including BSD)<sup>1</sup>.
- R is open source.
- Because R is open source, the code is freely available to anyone. This has the advantages that:
  - Anyone can examine the code to find (and fix) errors;
  - Unlike some statistical (and mathematical) packages, it is possible to know *exactly* what the software does by looking at the code.

This means the software has been examined by many experts to ensure quality of the statistical procedures.

- Because R is free, it is easy and cheap to ensure you have the most up-to-date versions.
- R is very powerful.
- R is extendable, with numerous add-on packages available.
- Because R is open source and used by many active and renowned scientists and statisticians, R is often the software of choice for implementing new ideas.

---

<sup>1</sup>For example, SPSS is not available for Macs, and only for some flavours of Linux, and not for BSD

- Many research institutions, universities and commercial companies use R.
- R is actively supported through a variety of means (see Section 3).
- R is actively supported by many well-known and accomplished statisticians and scientists.
- R is programmable: if R can't do a particular task, you can *program* R to do it.
- R produces publication quality graphics.
- In many areas, R is (or is becoming) the standard environment for analysis, notably bioinformatics (using the BioConductor package for R).
- R is an object-oriented language.

## 1.2 About R

### 1.2.1 What is R?

R is a statistics package, or a statistics environment. It is a command-line package, in a similar way that MATLAB is command-line driven, unlike many popular statistics packages.

R is based on the S language, a very high level language and an environment for data analysis and graphics. In 1998, the Association for Computing Machinery (ACM) presented its Software System Award to John M. Chambers, the principal designer of S, for

- the S system, which has forever altered the way people analyse, visualise, and manipulate data. . .
- S is an elegant, widely accepted, and enduring software system, with conceptual integrity, thanks to the insight, taste, and effort of John Chambers.

The two main expression of the S language are:

- S-Plus, a commercial (and often expensive) implementation of S; and
- R, an free, Open Source implementation of S.

### 1.2.2 What are the differences between S-Plus and R?

Since both S-Plus and R are implementations of S, they are very similar. There are some differences, however:

- S-Plus is commercial; R is free and Open Source.
- There are a few differences in the language also.
- S-Plus has a fancier GUI.
- S-Plus often has fancier graphics capabilities.
- R has some useful functions not found in S-Plus (such as adding Greek letters to plots). Likewise, S-Plus has some functionality not found in R.
- S-Plus has many (additional cost) add-ons that extend its functionality beyond R, or are free in R.
- R is available “out of the box” on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux). It also compiles and runs on Windows and MacOS. In contrast, S-Plus is available for Windows and some flavours of Linux, but not MacOS.

The functionality of R is growing rapidly, and for most purposes they can be considered together apart from the interface. Low-level users might never notice the differences.

## 1.3 Basic R commands

- Once R has been started, you can quit by typing `q()`.
- R is command-line based (like, say, MATLAB), not point-and-click (like, say, SPSS or Word).
- In the command window, a `>` character at the far left (the ‘command prompt’) indicates R is awaiting your instructions.

