Chapter 1

Introducing R

R is a free language that can easily be downloaded from the internet. Regardless of whether you work in industry, academics, or government, or if you just use the language for personal use, you will always have access to the R language. So why use R? There are at least six answers to this question. First, R is freeware. It can be used on a desktop or a laptop computer. Second, R is capable of performing numerical calculations on scalars, vectors and matrices, and can be used as a high-powered calculator. Third, R has built-in functions for performing probability and statistical calculations, and R can run simulation experiments using these functions. Fourth, R has extensive graphics capabilities so that it can produce production-level graphics. Fifth, R can be used as a programming language. R allows the user to execute several types of loops for iteration and also supports conditional execution of code. Sixth, hundreds of contributors have written R code contained in "packages," which continue to extend the language in the same way that apps have transformed handheld devices.

The R language, originally known as S, was developed in 1976 at Bell Labs by John Chambers and his colleagues. S-Plus is the current commercial version of S. In 1993, the free, open source R language was first developed by Ross Ihaka and Robert Gentleman (notice the common first initial) from the University of Auckland. R and S have a very large overlap of capabilities. If you know one, you essentially know the other. The differences are minor. One advantage to R over S, however, is the number of packages that have been written in R that are capable of extending the base language. These packages can be quite useful in specific niche applications.

The orientation for R is as follows. R is a vector-based language that uses vectors as its primary data structure. R has a command-line orientation rather than a Graphical User Interface (GUI) menu orientation. Although this might seem a bit antiquated at first, the size and capability of the language force this orientation. The focus here will be on the syntax, that is, the rules for issuing R commands. There are plenty of other R books on the market that are much more encyclopedic in nature or cover specific applications. This book is designed to be a quick introduction to the language for R novices. More advanced books on R will be much easier to read once you master the basics presented here.

You will notice that R tends to favor short, abbreviated variable and function names, which is helpful on your fingers in terms of saving keystrokes. It is also a heritage from the C language and the Unix environment that was present at Bell Labs in the 1970s.

The R language can be located by doing a web search on the letter R in a browser. Alternatively, you can use the website `http://www.r-project.org`. Once you get to the website you should pick a mirror site that is near your location and install the binary version. You will need to choose an appropriate platform: Windows, Mac, or Linux. For your first installation, it is probably best