--- Chapter Two ---

# *What You Need To Know To Get Started With R*

What Is R?

R is an open source, and so freely available, analytical software package that is widely used by biologists for carrying out data processing and statistical analysis. In fact, it has become so widely used that knowing how to use R can now be considered a critical skill for almost all biologists. However, unlike most other widely used software packages, R does not have a true graphic user interface (GUI) that allows you to simply click on menus and/or buttons to carry out specific tasks. Instead, it is primarily command-driven. This means that you need to enter lines of code in order to get it to do what you want it to do. Many biologists find this rather daunting, and at first it can be, but you should not allow this to put you off using it. This is because R is an amazingly powerful tool that, if you learn to use it properly, can make your life as a biologist so much easier. In addition, once you start using it, you will find that R is not as difficult, or as complicated, to use as it may at first seem, and the effort you put into learning it will be repaid many times over by the benefits of being able to use it to process and analyse your data more quickly and more efficiently than you could ever hope to do without it.

Where Can I Get R From?

As an open source software package, R can be freely downloaded from the R Project website at *www.r-project.org*. This workbook assumes that you will be using version 3.6.1, 3.6.2 or 3.6.3 (the three most recent versions at the time of writing), but the instructions provided in it should work for almost any recent and future versions of R.

How Can I Use R?

There are two common ways that biologists use R. These are using the native R user interface, which will be referred to in this workbook as RGUI, and by accessing it using a third-party user interface. The RGUI option is the one that automatically appears when you open R, and it looks like this:
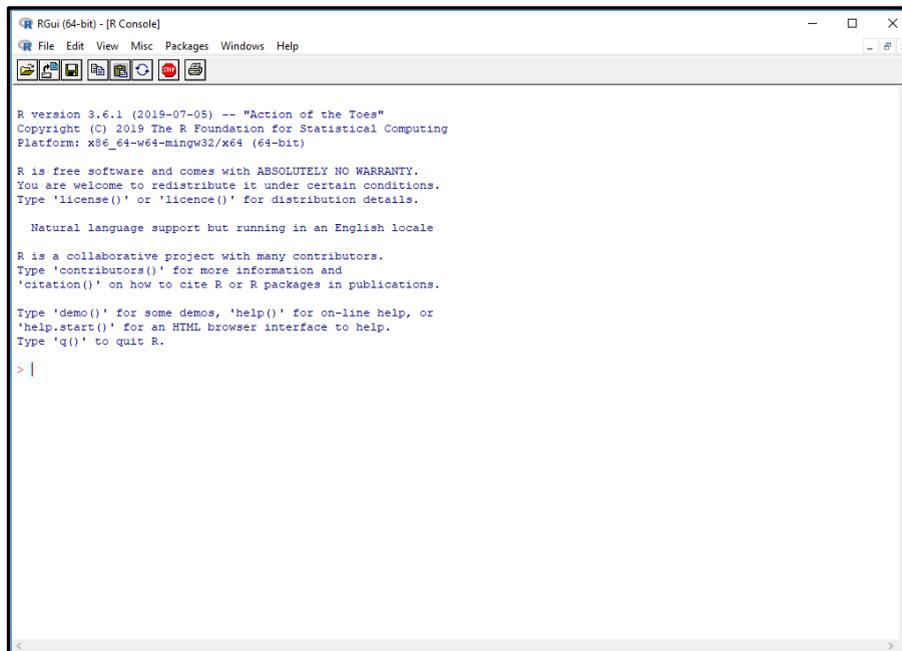
This window (which, when you first open it, contains text detailing of the version of R you are using and various other bits of information) is known as the R CONSOLE window. To use the RGUI, you simply type the R code you wish to run on the command line of this window (the line that starts with a > symbol) and press ENTER on your keyboard. To save your work, or to reload a previous work session, you can select the appropriate option from the menus along the top of the user interface (see below for more details). As and when you call them up, other windows, such as DATA VIEWER windows (which allow you to view data tables that you have within R), R GRAPHCS windows (which will display graphs and other graphics generated by the R commands you enter into the R CONSOLE window) and an EDITOR window, will appear in the main RGUI window alongside the R CONSOLE window.

Of the third-party user interfaces available for R, the one that is the most widely used by biologists is called RStudio. The main reason that many biologists prefer to use RStudio rather than the native RGUI is because it makes it easier to follow exactly what you are doing, to edit the commands you have used in the past to adapt them for use with new data sets and to archive them in a meaningful way so you have a record of exactly what you have done (see Appendix II). As a result, we recommend that you use RStudio to access R, rather than RGUI, both for completing the exercises in this workbook and for your own analyses. However, regardless of how you access R, you will use the same commands to complete the same tasks. Any points of difference between using these two interfaces for the exercises in this workbook are highlighted within the instructions provided. This means that, as far as this workbook is concerned, it does not matter which interface you choose to use.

Unlike the native RGUI, RStudio needs to be downloaded and installed separately. However, it is also freely available and you can download it from *www.rstudio.com*. Once you have installed RStudio, it will automatically connect to R (as long as you have already installed R on your computer) and it will be ready to use. When you first open RStudio, its user interface should look like this:
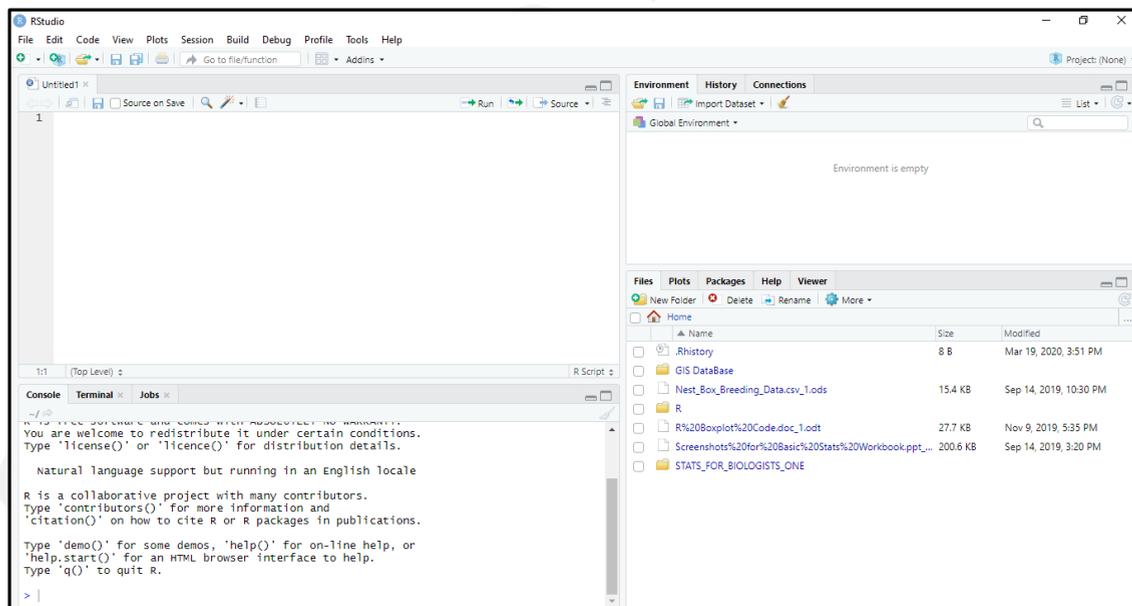
If your version of RStudio does not look like this, click on VIEW on the main menu bar, and select PANES> SHOW ALL PANES. This will make sure that all the required windows are set to display. If the upper left hand window is still not visible after you have

done this, click on FILE on the main menu bar and select NEW FILE> R SCRIPT. This should make the missing window appear.

The RStudio user interface is divided into four windows or panes. These are: 1. The SCRIPT EDITOR window (the upper left hand window); 2. The R CONSOLE window (the lower left hand window); 3. The ENVIRONMENT/HISTORY window (the upper right hand window); 4. A window in the lower right hand corner that includes tabs for FILES (which allows you to browse files on your computer – this means the list of files and folders that appears in this tab will be unique to your computer and will not match those in the image on page 8), PLOTS (which displays any graphs or other graphics created by commands you run in the R CONSOLE window), PACKAGES (which allows you to access additional packages – see below for information on what an R package is) and HELP (where you can get help if you get stuck with anything while using R).

As with the native RGUI, you can enter the R code you need to run a specific command directly in the R CONSOLE window in RStudio. However, the main advantage of using RStudio is that you can also use the SCRIPT EDITOR window to enter, edit, prepare, annotate and store any R commands you might wish to use before you run them in the R CONSOLE window. This makes it a much more flexible way to enter commands into R. In order to run a command from the SCRIPT EDITOR window, you need to transfer it to the R CONSOLE window. To do this, you simply select the appropriate command (or a block of commands) in the SCRIPT EDITOR window and click on the RUN button at the top of it.

What Special Terms Are Used When Describing Things In R?

For the most part, the terms that are used when completing data analysis tasks in R are the same as those you would use in any other statistical package. However, there are eight terms which are more specific to R that are useful to understand in order to be able to use it more easily. These are:

1. **Code:** In R, Code refers to the string of text and characters that you need to enter to make it do something, such as importing a data set, creating a graph or running a specific statistical test. A piece of Code is made up of a number of different components,