Introduction

Key learning objectives

a) Build fundamental skills in using R
b) Understand types of data
c) Gain experience working with data

The material developed in these R Sessions provides you with a basic foundation for using a software package called, R. You will use R throughout your training in data analysis in this course, and beyond. R is a basic tool that you will use throughout the remainder of your Biomedical Sciences education and it is vital that you learn how to use it now.

Learning R could help you in a future career. Many jobs are opening up now (try Googling "Data Science") that require employees to know how to use R. In short, if you think that you might be interested any form of career in science (e.g., working for the NHS to analyze data), you'll likely need to know how to use R. These R session are a great place to start, and can help you in the long run.

We recommend that you complete one R session per week, starting in the first week of semester 1. Each session should take between 45 min and 1 hour to complete.

1.What are R and RStudio?

As you begin this course, you will join a very large community of scientists who use 'R' to analyse data. So, you might be asking, "What is R, and why are we using it?".

R is many things at once. We will mostly use R for to conduct powerful **statistical analyses**. But, R is also a **programming language**, and can also be used to make amazing **graphics** to present your data and analyses.

Importantly, the source code for R is openly available to users, i.e., R is 'open source'. This means that users also continually develop new tools, so that R can do many things that standard statistical packages cannot. Additionally, R does not cost anything to use, so it is always available to you (some data analysis programs cost sizeable sums).

RStudio is a friendly interface for R; i.e., think of R as a car's engine and chassis – it is what makes a car 'go' ,but it is not very comfortable to use on its own. Think of RStudio, on the other hand, as the 'body' of a car (with comfortable seats and cupholders), which makes R more pleasurable and easier to use. We will use the terms R and RStudio loosely (interchangeably), always implying that we will be using R's functionality but assuming that you are using R through RStudio.

So, why use R? In short, it is easily available and is very widely used to analyze

data, with many tools at hand. But, using R requires a fair amount of knowledge. This may seem daunting at first, but you will quickly get a grasp of R; we will provide you much help and opportunity for practice during this course. As you'll see below, other opportunities for help are also available.

But, why not just use Excel? Good question. Excel works well to organize data, ***and we recommend it for that purpose***. But, it has very few sophisticated tools to analyze data; in other words, it is inadequate for your needs later in your Biomedical Sciences education. In fact, you will find very few (if any) professional scientists using Excel for statistical analyses, whereas R is an "industry standard" for statistical analyses. In the very short term, you may find little benefit in using R compared to Excel, but over the longer term (within your university degree) the reasons for using R will become obvious.

Make your analyses reproducible by using scripts

To use R, you will be typing commands. While you might expect that 'clicking' commands would be a more current approach, this view is misled. Typing commands and storing them in files termed 'scripts', has at least one major advantage over 'clicking': it allows you to have a record of the commands you used for an analysis. Often, you will want to re-visit an analysis and change it or add to it; if you have the commands available, you can repeat an entire analysis exactly as you did it previously in a matter of seconds, instead of potentially hours by 'clicking'. As well, documenting your commands allows you to check the steps you took in an analysis (did you make a mistake?). With practice, which this course will provide, you will become comfortable with writing scripts, and we expect you will see their benefits.

How did R come to be?

The development of R followed the creation of a highly similar language called S, which was designed by a group of researchers led by John Chambers at the Bell Laboratories in New Jersey. Bell Laboratories also developed Unix, an important computer operating system, and the C programming language, which have both remained essential computer tools since their invention in the 1970's. S was developed in the 1980's and 1990's, and was owned by AT&T. It is now currently available as 'S PLUS'.

Robert Gentleman and Ross Ihaka began to build R, modeled from the S language, in the early 1990s. Now, R is widely used for scientific purposes.

S and R differ in an important way. S is a closed system, developed by software engineers who were paid for their work, and the underlying computer code is a company secret. This means that the inner workings of S, like most software that you must pay to use, remain unknown to the user.

2. Sessions Aims:

Session 1: Learn how to interact with R
Session 2: Learn how to store data and change your data
Session 3: Learn how to use data in "dataframes"
Session 4: Learn basic methods for plotting histograms

3. What help is available?

a) Session 1 outlines help that is available that is specific to R. See Session 1, point 3.
b) If you are a BMS2 student, this course has a discussion forum on LEARN, where students can help one another, and ask course instructors questions pertaining
c) Students at the University of Edinburgh can always e-mail Dr. Crispin Jordan for help: crispin.jordan@ed.ac.uk
d) University of Edinburgh students can attend Crispin's Office hours, available weekly, for help with this (and any data analysis) material. Venue and timing will be announced on LEARN.

4. How to use these sessions?

We recommend the following:
a) you complete only one session at a time; don't try to do too many at once.
b) you read and work through all of the main text for each session, and that you complete two questions from the Exercises. In most cases we provide more than two questions to provide additional opportunities for practice.
c) you work with a friend, and participate in the LEARN online forum.

5. How to use RStudio to your own computer

You have two ways of using Rstudio:

a) you can use RStudio via the University of Edinburgh computers.

b) You can install RStudio on your own computer.

Both (a) and (b) are good options for using R/RStudio. However, we encourage students to use R/RStudio downloaded and installed onto you own computer (option (b)).

How do you install R and RStudio on your computer?

In order to use R via RStudio, you must install both R as well as RStudio (these are separate programs). Occasionally, students contact Crispin and explain that they cannot get R to run. Almost always, this is because the student installed RStudio but not R (you need both); following our analogy, above, this situation is like trying to drive a car without an engine.

Therefore, you should first install R. Here is a link to download and install R:

https://www.r-project.org

If you are unfamiliar with installing programs on your computer, the simplest advice we can provide to perform a search in either Youtube or at the site, "rseek.org" for guidance (you will find a lot of help here). In both cases, you might perform a search such as, "How to install R on a Mac", (or PC, or Linux, etc).

Once you have installed R, you should download and install RStudio from here:

https://www.rstudio.com/products/rstudio/#Desktop

Once again, we recommend using Youtube or rseek.org to search for help if you need it. Students at the University of Edinburgh can also always ask Crispin for help (crispin.jordan@ed.ac.uk).

6. Format of session

Each R-session contains didactic material as well as instructions for you to type. The instructions are presented in (un-numbered) boxes.

Useful tips for using R are presented in italics, and additional information is presented in numbered boxes.