Introduction to ‘R’ statistical software for statistical analysis

George Savva, (QIB)

# Preface

This course will introduce R statistical software, some basics of how it works, guides to performing common operations, where to go for further support and some tips for good practice.

The aim is to become familiar with the R/RStudio environment and some common functions so that you can learn the specific functions that you need on your own or with further training. Specifically this course will bring you up to the level you need to attend the externally run statistics courses provided at by NBI.

The data you need to complete the training exercises is in the accompanying file (Introstat.xlsx). All of the commands for the worked examples and the exercises is in the file TutorialScript.R. The commands for the worked examples are also typed out here.

# Learning objectives

#Specific tasks:

##Day 1:

* Making an Rstudio project
* Loading packages
* Loading a dataset into R from an Excel file
* Exploring your data; calculating descriptive statistics
* Simple hypothesis tests

##Day 2:

* Estimating and diagnosing a regression model

##Day 3:

* Making graphics using ‘base’ graphics and the ggplot2 package

How to use R \* R and RStudio \* RStudio ‘projects’ and a good workflow \* The command line \* Using scripts for reproducibility \* Getting help \* Using packages

How R works \* Objects and functions \* Data frames and vectors \* Types of data: numerics, factors, and strings \* Representing and handling ‘missing’ values \* The importance of ‘tidy’ data

# Introduction

## What are R and RStudio?

R is an open source statistics package, initially developed during the 1990s, and that has now become the world’s most widely used and comprehensive statistical software. R calls itself a ‘programming language *and* environment for statistic computing’.

That is, ‘R’ refers both to the software itself ‘and’ the programming language that you use to interact with it.

RStudio is a integrated development environment (IDE) for R that makes working R much easier. Most R users use RStudio and I recommend using RStudio for new users.

The great strength of R is in its contributed packages, these are community written add-ons that provide functions to perform almost any statistical, programming, or data-related task. We will introduce several packages for data management, analysis and graphing during this course.

# Exercise 1

## Getting R and RStudio

Download and install the latest version of R from <https://cran.r-project.org/>

Then download and install RStudio from <https://www.rstudio.com/>

Start RStudio. It will detect your installation of R, and you should see a screen like this:

On the left is the console window, where you type commands and see output. The windows on the right hold various useful tabs, here the top pane is showing the data I happen to currently have loaded (yours will be empty) and part of my filesystem at the bottom. These right-hand windows can also show graphs, help files, and your command history.

# Exercise

First task: check RStudio is working. Click in the console window and type:

1+2

Press return on your keyboard. You should see:

## [1] 3

This is the basic way in which R works. We enter commands at the command prompt (or via a script), and we get the output in the console window.

To chain together multiple commands, and to keep our work safe we will write scripts, which are just sequences of commands that are executed in order.

# Exercise

Try a few other mathematical functions at the R console.

# Using projects

Before we go any further, we are going to start an RStudio ‘project’ to organise our work. Using projects helps us to keep all of the data and analysis for a particular piece of work in the same place.

Click on ‘New -> New project’ in the toolbar. Click ‘Start a new project in a brand new working directory’. Then click ‘new project’ on the next screen.

Now you can choose where to create the new directory for your R project, and what to call it. Make a project called ‘Rtraining’ or something like that, somewhere in your personal filestore.

Now when you return to the main RStudio window you are working within your project. Notice that the working directory has switched to the new directory that you created.

Important: Keep the R scripts and the data associated with this training course in that directory, so that you can access them easily.