

An Introduction to R

3rd - 5th December 2019

Alex Douglas

[@a.douglas@abdn.ac.uk](mailto:a.douglas@abdn.ac.uk)
[@Scedacity](https://twitter.com/Scedacity)

UNIVERSITY OF ABERDEEN




1

goals

- over the next three days.....
- introduce you to using R for data manipulation, graphics and statistics
- provide you with the practical skills necessary to explore and visualise your data
- provide you with the practical skills necessary to start your own data analysis
- BUT cannot teach you everything there is to know about R: that's down to you

2

course structure

- course split into 6 main components
 - introduction to R environment
 - basic R operations
 - dataframes
 - graphics and data exploration
 - basic statistics using R
 - R programming (optional)
- work through manual and associated exercises

3

what is R?

- environment for statistical computing, graphics and programming
- originally created by Ross Ihaka and Robert Gentleman (1996)
- currently maintained by international R-core development team
- very similar to S



4

R-Project and CRAN

- more information found at [www.r-project.com](http://www.r-project.org)
- download R from <http://cran.uk.r-project.org>

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and Mac OS. To download R, please choose your preferred CRAN mirror.

If you have questions about R or how to download and install the software, or any of the license terms etc., please read our answers to frequently asked questions before you send us email.

News

- R version 3.4.0 (You Biased Darkness) has been released on Friday 2017-04-21.
- R version 3.3.3 (Great裹 R) has been released on Monday 2017-03-06.
- RwiR 2017 July 4 - 7, in Brussels has opened registration and more at <http://useR2017.brussels/>
- Tomas Kalvoda has joined the R core team.
- The R Foundation welcome five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Torsten Kettner, and Satyabrata Narsimhan.
- The R Journal Volume 8/1 is available.
- The useR! 2017 conference will take place in Brussels, July 4 - 7, 2017.

R Foundation
 Nonprofit
 Board
 Members
 Donors
 Donate

5

why you shouldn't bother with R

relatively steep learning curve?
 it's command line driven
 you need to learn to 'speak' R



```
se.fnc <- function(x){ #start function for se
  std.x <- sd(x)           #calculate SD
  nos.x <- length(x)       #calculate number obs
  se.x <- std.x/(sqrt(nos.x))#calculate SE
  print(se.x)               # print SE
}                           #end of function
```

6

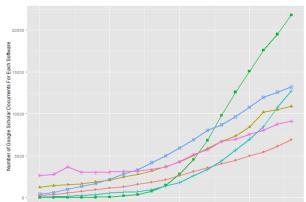
why you should learn 

it's free and platform independent

it's the software of choice for many students, academics, industries and charities worldwide

highly flexible and extensive

encourages you think about your analyses



7

why you should learn 

it allows you to keep an exact and reproducible record of your analyses*

excellent graphics facilities*

might get you that job/post-doc

R is a community not just a bit of software

8



9

why you should learn 

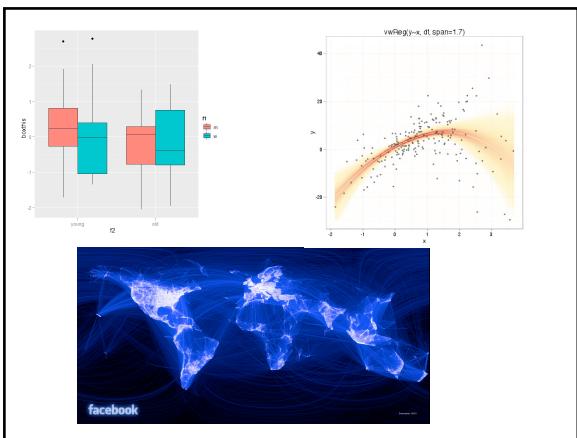
it allows you to keep an exact and reproducible record of your analyses

excellent graphics facilities*

might get you that job/post-doc

R is a community not just a bit of software

10



11

why you should learn 

it allows you to keep an exact and reproducible record of your analyses

excellent graphics facilities

might get you that job/post-doc

R is a community not just a bit of software

12

Using R - GUI

- fairly spartan GUI – very few menus or buttons

R is not menu driven

13

using R

- commands are typed (or sourced) into the console window at the > prompt

```
> 2+2
[1] 4
>
```

- R is object orientated. You can create variables and assign value(s) to them

```
> a <- 2+2
> a
[1] 4
>
```

assign variables a value using the 'gets' operator

display variable value by typing variable name

14

using R

- once created, operations can be performed on variables

```
> a <- 2+2
> b <- 3*2
> a+b
[1] 10
```

- this is very powerful and flexible (as you will see)
- much of the functionality of R is enhanced by using variables called functions

15

using R

- functions contain a set of instructions that allow you to perform a specific task(s)
- you can use functions that are inbuilt in R (or R packages)

```
> numbers <- c(2,3,4,5,6)
> numbers
[1] 2 3 4 5 6
> mean(numbers)
[1] 4
> var(numbers)
[1] 2.5
```

concatenate

calculate the mean of variable numbers

calculate the variance of variable numbers

16

using R

- or write your own functions
- function to calculate standard error

```
> se.fnc <- function(x){ #start function for se
  std.x <- sd(x)           #calculate SD
  nos.x <- length(x)        #calculate number obs
  se.x <- std.x/(sqrt(nos.x))#calculate SE
  print(se.x)
} #end of function
```

- using your new function

```
> se.fnc(numbers)
[1] 0.7071068
```

notice the use of comments

17

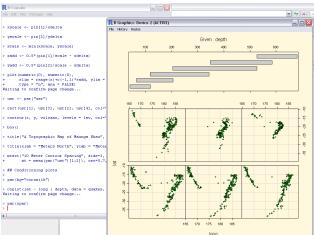
using R - syntax

- R is case sensitive ‘A’ is not the same as ‘a’
- commands are generally separated by a new line, occasionally you might see a semicolon ;
- anything that follows the hash symbol (#) will be ignored by R. Use this to comment your code
- a series of commands can be grouped using braces { }
- don’t worry too much about spaces
- previous commands can be recalled using ↑ and ↓ keys

18

using R

- some functions open other devices
 - i.e. many graphic functions open a plotting device
 - plots can be copied and pasted into word
 - or saved as a pdf, jpg etc
 - you can open more than one device



19

getting help in R

- R has extensive help facilities
- from within R the main method of getting help is to use the `help()` function

```
> help("plot")
or
?plot
to search for help

> help.search("plot")
or
??plot
```



20

getting help in R

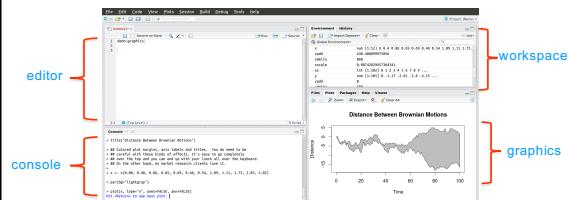
- wealth of information on web
 - R-project web site
 - Rhelp mailing list with searchable archives
 - free pdfs from R-project website
 - www.Rseek.org
 - UoA Rusergroup mailing list

<https://mlisthost.abdn.ac.uk/mailman/listinfo/rusergroup>

21

using an IDE: RStudio

- Using an Integrated Development Environment (IDE)
- RStudio (<http://www.rstudio.com/>)



22

keep scripting

- always use a script editor or IDE to write your R code rather than typing directly into console
- the more you use R the more complicated your code will become
- allows you to keep a permanent and reproducible record of your analyses

```
se.fnc <- function(x) {
  std.x <- sd(x)
  nos.x <- length(x)
  se.x <- std.x/(sqrt(nos.x))
  print(se.x)
}
```

23

course survival guide (and beyond)

- keep a careful record of your code, analysis & plots
 - Rstudio, MS word, R markdown(?) etc
- annotate your R code with plenty of comments (#)
- remember R has an extensive help facility
- ask plenty of questions
- start using R to explore and analyse your own data as soon as possible

24

where to start?

- work through sections of manual
- complete exercises for each section
- datafiles and exercises can be found at:

<https://alexd106.github.io>



Welcome to QUAData! This repository is for the first course. Over the next 7 weeks we'll be introduced to using R as an instrument for statistical computing. It is best suited for 3 hours per week, but just like any new skill, the initial learning curve can be a little steep and you will need to use it frequently otherwise it's easier to forget.

All the material is freely available online so you can use as much as possible and have based it on 15 years experience teaching (and learning) R. It is not intended as an introductory statistics course, although I will be using some basic statistics to highlight some of the features of R. I hope that by the end of the course you will be more comfortable using R and that you will be able to use your experience in using R, I have included a number of practice [exercises](#) for you to work through during the course and encourage you complete those in your own time; you certainly won't learn how to use R by watching other people do it.

FAQ