

An Introduction to 

(for robust and reproducible research practices)

Alex Douglas

 a.douglas@abdn.ac.uk
 @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
- download R from <http://cran.uk.r-project.org>



The R Project for Statistical Computing

Getting Started

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- R version 3.4.3 (Now Shiny Darkness) has been released on Friday 2017-04-21.
- R version 3.3.3 (Another Cance) has been released on Monday 2017-03-06.
- useR! 2017 July 4 - 7 in (Brussels) has opened registration and more at <http://useR2017.brussels/>
- Tomas Kalibera has joined the R core team.
- The R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Tomas Kalibera, and Balasubramanian Narasimhan.
- The R Journal Volume 9(1) is available.
- The useR! 2017 conference will take place in Brussels, July 4 - 7, 2017.

5

why you shouldn't bother with

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 R

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 R

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

SIGNIFICANCE The reproducibility crisis in science: A statistical counterattack

Science Estimating the reproducibility of psychological science

BBC NEWS Most scientists can't replicate studies by their peers'

PLOS MEDICINE Why Most Published Research Findings Are False

NATURE IS THERE A REPRODUCIBILITY CRISIS?

1,576 researchers surveyed

9

Why?

structural pressures

fraud -> QRPs

journal policy

closed science

10

why you should learn R

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

11

12

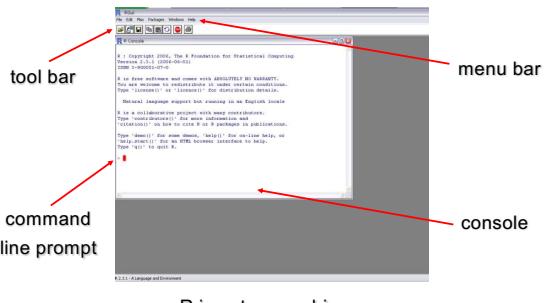
why you should learn R

- 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

13

Using R - GUI

- fairly spartan GUI – very few menus or buttons



R is not menu driven

14

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

15

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

16

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

17

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)               #print SE
}                                #end of function
```

- using your new function

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

notice the use of comments

18

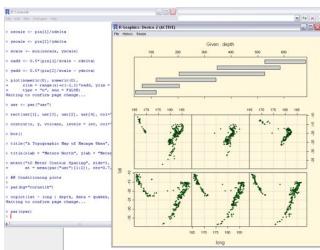
using R - syntax

- 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

19

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



20

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
```



21

getting help in R

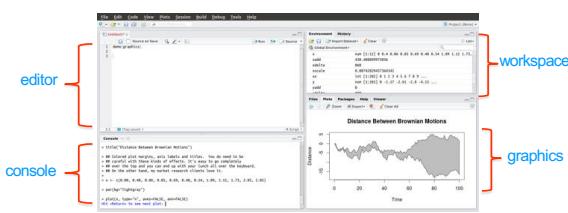
- wealth of information on web
- R-project web site
- Rhelp mailing list with searchable archives
- free pdfs from R-project website
- www.Rseek.org
- UoA Rusergroup mailing list

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

22

using an IDE: RStudio

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



23

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)
}
```

24

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

25

where to start?

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

<https://alexd106.github.io/intro2R>



An Introduction to R

Course Introduction

Welcome to the Introduction to R training course. Over the next 3 days we hope to introduce you to using R, an interactive environment for statistical computing. It is itself is not difficult to learn, but just like any new language the initial learning curve can be a little steep and you will need to practice.

A few notes about this course. I have tried to simplify the content of this course 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 have included a number of practical exercises for you to work through during the course and encourage you to complete these in your own time - you certainly won't need to wait for the rest of the class to finish.

26