Introduction to R

Research Methods and Skills

12/10/2021

Who are we?

Dr Matt Craddock



Sarah Swift Building Room 2226 mcraddock@lincoln.ac.uk

Dr Tochukwu Onwuegbusi



Sarah Swift Building Room 2211 tonwuegbusi@lincoln.ac.uk

Replicability and reproducibility

Scientists Replicated 100 Psychology Studies, and Fewer Than Half Got the Same Results

The massive project shows that reproducibility proble top scientific journals

A question of trust: fixing the eplication crisis

he crisis of non-replications in experimental social psychology is crisis of trust. What's the solution?



RESEARCH ARTICLE



Estimating the reproducibility of psychological science

Open Science Collaboration*,†

+ See all authors and affiliations

Science 28 Aug 2015: Vol. 349. Issue 6251. aac4716 DOI: 10.1126/science.aac4716

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson & Marcus R. Munafò 🖾

Nature Reviews Neuroscience 14, 365–376 (2013) Download Citation 4



Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • https://doi.org/10.1371/journal.pmed.0020124



Scenario A

You've just started work in a psychology lab. You're asked to help analyse some old data. There is reaction time data from 50 participants. Each participant's data is stored in a separate text file.

- How do you combine the data from each participant together to be able to analyse the data?
- It turns out some of the participants only completed part of the experiment which ones, and what should you do with their data?
- What steps should you take to select and perform appropriate statistical analysis?

Scenario B

You've been asked to design, implement, and evaluate a new treatment regime across several psychiatric institutes. Several colleagues are skeptical that it can deliver the kind of improvements in outcomes indicated in a publication describing the method.

- How do you interpret the strength of the previously published evidence?
- How do you design a rigorous test of the treatment efficacy?
- How do you evaluate and report on the outcomes of your trial?

Research Methods and Skills





Course outline

Weeks 1-5

- Introduction to R
- Basic R programming
- Plotting with ggplot2
- Data import, selection and manipulation
- Describing and summarising your data

Weeks 11-13

• Qualitative methods

Weeks 5, 7-10

- Hypothesis testing and estimation
- *t*-tests and comparing two groups
- Correlation and linear regression
- One-way ANOVA
- Factorial ANOVA

Course outline

Weeks 1-5



Weeks 5, 7-10



Course outline

Weeks 1-5



Hadley Wickham & Garrett Grolemund

Weeks 5-10

DISCOVERING STATISTICS USING R



ANDY FIELD | JEREMY MILES | ZOË FIELD

(S)

How it's going to work

Each week you'll be provided with a set of pre-recorded videos introducing the core topics for that week.

Most weeks you will be provided with some exercises to work on.

In timetabled sessions we will spend the first hour going over the week's exercises. In the second hour, you can spend the time working together on the exercises, and we'll be available to help out.

If you get stuck **ask us**!

Introduction to R and RStudio

What is R?

R

R is a statistical, mathematical programming language

- Created in 1993
- Designed from the ground up to support many statistical tasks
- Covers all aspects of data analysis from import through to production of reports
- Free, open source
 - Can be downloaded from the R-project website
- Continually evolving
 - R has over 12,000 *packages* that add additional functions

But WHY?

What can you do with R?



Source: N. D. Phillips



Companies that use R for Analytics



Still not convinced?



Getting started

What is RStudio?



- An Integrated Development Environment (IDE)
- An interface for R that makes your life much, much easier
- Makes many things explicit that you would otherwise have to guess

| e e RStudio | |
|--|--|
| 🝳 🔹 🤮 🔹 🕞 🔚 🎒 🖉 Go to file/function 🔤 🛛 🔡 🔹 Addins 👻 | 送 Project: (None) 🔹 |
| <pre>Console -/ Console -/ C</pre> | Project: (Nonc) - Project: (Nonc) - Project: (Nonc) - Global Environment + Environment is empty Files Plots Packages Help Viewer Prover Folder Deciments SoftwareCarpentry r_gapminder Name Documents SoftwareCarpentry r_gapminder Name Size Modified r-novice-gapminder |
| | |

Getting started TODAY

- 1. Open up a web browser
- 2. Go to https://rstudio.cloud
- 3. Sign up! Use your REAL NAME, and your University of Lincoln email address.



| R Studio Cloud 🛛 | Your Workspace | Projects Info | | MC M | /latt Craddock |
|---|----------------|---------------|-----------------|--------------------------|----------------|
| Spaces | Your Projects | | New Project 🗸 🗸 | Options | × |
| PSY9219M; Research Methods New Space | | No Projects | | Search Projects | |
| Learn | | | | Sort Projects By name | |
| Guide Primers DataComp Courses | | | | By date created | |
| Cheat Sheets | | | | | |
| Feedback and Questions | | | | | |
| Terms and Conditions Image: System Status | | | | | |
| ↓ → → | | | | | 22 / 56 |





| R RStudio Cloud × + | | | | | - 0 | × |
|---|-------------------------------------|------------------------------------|--------------------------|----------------------|------------------------------------|------------------|
| ← → C https://rstudio.cloud/project/84901 | | | * 🕈 🗋 | a 🕫 💁 | ¢ 🛆 | 🌍 : |
| Your Workspace / Untitled Project & Click to name your project | | | | ¢ | MC Matt Cr | addock |
| File Edit Code View Plots Session Build Debug Profile | Tools Help | | _ | | | |
| 💽 - 🚰 - 📄 🦳 Co to file/function 🛛 🔡 - Addins - | 🔟 Install Packages | - | | | R 3.5.0 🗸 | |
| Console Terminal × Jobs × | Check for Packag | e Updates | ory Connections | | | |
| /cloud/project/ ≈ | Version Control | • | Dataset 🗸 🔏 | | E List | • C • |
| R version 3.5.0 (2018-04-23) "Joy in Playing" Copyright (C) 2018 The R Foundation for Statistical Computi Platform: x86_64-pc-linux-gnu (64-bit) | Shell Terminal Addins | • | nt - | | Q, | |
| R is free software and comes with ABSOLUTELY NO WARRANTY. You are welcome to redistribute it under certain conditions Type 'license()' or 'licence()' for distribution details. | Keyboard Shortcu Modify Keyboard | its Help Shift+Alt+K Shortcuts | Environment | t is empty | | |
| R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications | Project Options Global Options | | ages Help Viewer | | | |
| Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. | | New Folder Cloud > proj Name | Upload 🥸 Delete 📑 ect | Rename 🌼 M Size M | Nore • Nodified | © *** |
| <pre>> </pre> | | L P Rhistory Roject.Rp | roj | 0 B S 205 B S | ep 24, 2018, 11 ep 24, 2018, 12 | :53 AM :29 PM |



How it works!





How to use R

- The R Console
 - REPL: Read/Evaluate/Print/Loop
 - Type stuff in, it tries to do it

When you see the > symbol -

>

... R is waiting for your input.

> 5 [1] 5



Try using R like a calculator!

Basic arithmetic operators

| Symbol | Operation |
|--------|----------------|
| + | addition |
| - | subtraction |
| * | multiplication |
| / | division |
| Λ | exponentiation |
| %% | modulo |

Warming up

You can break up long maths expressions over multiple lines:

```
2 + 4 + 5 +
5 + 6 + 7 + 8 +
10
```

[1] 47

Note that when you do that, the ">" symbol changes to a "+"

| > 5 + | F | | |
|-------|----|--|--|
| + 5 | | | |
| [1] 1 | LO | | |

Remember!

> means R is waiting for input.

>

+ means R is waiting for you to finish your command.

+

Either finish your command, or press the **Esc** key to cancel it.

Text input

R can also accept text strings as input.

```
"hello world!"
[1] "hello world!"
```

You need to use quotation marks ("") to tell R that this is text:

hello world!

```
## Error: <text>:1:7: unexpected symbol
## 1: hello world
## ^
```

Otherwise, you'll receive an error like the one above.

Why is there an error?

In R, you can assign values to an **object** for subsequent use. **Objects** have names that are written as text.

The assignment operator is the two-character symbol:

You assign values to objects by putting the <- sign between the name of the object and the value you want to give it:

example <- 5 example

[1] 5

<-

Note that R does not immediately provide output when you assign the output to an object.

The assignment operator

Think of <- as meaning "is now". i.e.

example <- 5

can be read as

The object "example" is now 5

Working with objects

Once an object is assigned, the name that you gave it *stands in* for the *value* that you assigned to it, and can be used as if it were that value:

| example | | |
|---------------------------|--|--|
| ## [1] 5 | | |
| example + 10 | | |
| ## [1] 15 | | |
| example + 13 - 1 * 2 %% 4 | | |
| ## [1] 16 | | |



```
\times
RStudio Cloud
                            ×
              https://rstudio.cloud/project/84901
\leftarrow
                                                                                                     ☆
                                                                                                                                               :
                                                                                                                                     Matt Craddock
 Your Workspace / Untitled Project .
 File
      Edit Code View Plots Session
🗘 - 🔿 -
            🕞 🗐 📥 🛛 🍌 Go to file/function
                                                                                                                                   R 3.5.0 -
                                                                                Environment History Connections
Console Terminal × Jobs ×
                                                                           Ð
                                                                                                                                            -\Box
/cloud/project/ 🖈
                                                                                🕣 🔚 🖙 Import Dataset 🗸 💰
                                                                                                                                     📃 List 🗸 🛛 📿 🗸
                                                                                🛑 Global Environment 🗸
                                                                                                                                 Q
R version 3.5.0 (2018-04-23) -- "Joy in Playing"
                                                                                Values
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86 64-pc-linux-gnu (64-bit)
                                                                                  example
                                                                                                  10
                                                                                  hi there
                                                                                                  "hi there!"
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
                                                                                Files Plots Packages Help Viewer
'citation()' on how to cite R or R packages in publications.
                                                                                                                                            💁 New Folder 🛛 🔾 Upload 🛛 Delete 📑 Rename
                                                                                                                           🚳 More 🗸
Type 'demo()' for some demos, 'help()' for on-line help, or
                                                                                                                                            R ...
                                                                                   Cloud > project
'help.start()' for an HTML browser interface to help.
                                                                                       ▲ Name
                                                                                                                  Size
                                                                                                                             Modified
Type 'q()' to quit R.
                                                                                    £
>
                                                                                    .Rhistory
                                                                                                                             Sep 24, 2018, 11:53 AM
                                                                                                                  0 B
> example <- 10
> hi there <- "hi there!"
                                                                                    project.Rproj
                                                                                                                  205 B
                                                                                                                             Sep 24, 2018, 1:06 PM
> hi there
[1] "hi there!"
> example
[1] 10
> Í
```

| R RStudio Cloud | × + | - 🗆 X |
|--|---|--|
| \leftrightarrow \rightarrow C $$ https://rstudie | o.cloud/project/84901 | 🖈 🔹 🗈 🛯 🖉 💁 🦓 🤤 : |
| R Studio Cloud 🛛 🛛 | Your Workspace / Untitled Project & Click to name your project | MC Matt Craddock |
| Spaces | File Edit Code View Plots Session Build Debug Profile | Tools Help |
| 📥 Your Workspace | • • | R 3.5.0 • |
| BSY9219M; Research Methods and | Console Terminal × Jobs × | Environment History Connections |
| + New Space | R version 3.5.0 (2018-04-23) "Joy in Playing" Copyright (C) 2018 The R Foundation for Statistical Computi g Platform: x86_64-pc-linux-gnu (64-bit) | example <- 10 hi_there <- "hi there!" in hi_there example |
| Learn | R is free software and comes with ABSOLUTELY NO WARRANTY. | |
| 🧭 Guide | You are welcome to redistribute it under certain conditions Type 'license()' or 'licence()' for distribution details. | 5. |
| U Primers | R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications | Files Plots Packages Help Viewer |
| DataCamp Courses | Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. | e Modified |
| | Type q() to quit R. | B Sep 24, 2 |
| Feedback and Questions | > example <- 10 | Image: Sep 24, 2 |
| Info | <pre>> hi_there [1] "hi there!" > example</pre> | |
| Terms and Conditions | [1] 10 > | |
| 네 System Status | | |
| | | |
| | | |
| | | |
| | | |



Try a few things out!

- 1. Assign some values to objects using the assignment operator (<-)
- 2. Try using arithmetic operations (e.g. *, /, %%) on those objects
- 3. Try using arithmetic operations to combine multiple numerical objects
- 4. Try using arithmetic operations on text

Combining multiple things

Sometimes you want to allocate more than one value to an object. You can use the **c()** function to do this.

c(8, 5, 10)
[1] 8 5 10
example <- c(8, 5, 10)
example <10
[1] 8 5 10
c("hello", "how", "are", "you")</pre>

[1] "hello" "how" "are" "you"

IMPORTANT: BRACKETS () AFTER A WORD MEAN THAT THIS IS A FUNCTION

Vectors

The function **c()** is creating **vectors**.

Vectors are simply a one-dimensional collection of things that all have the same *type* (we will cover data types next week!).

Note that mixing, for example, text and numbers, will yield a *character* vector.

c(5, "five", 2) ## [1] "5" "five" "2"



Functions are commands that operate on **objects**.

For example, to calculate the *mean* of several numbers, you can use the function **mean()**. The output of functions can also be assigned to **objects** using <-.

```
mean(c(8, 5, 10))
## [1] 7.666667
example <- c(8, 5, 10)
mean(example)
## [1] 7.666667
example_mean <- mean(example)
example_mean
## [1] 7.666667</pre>
```

Try it out!

- 1. Use **c()** to create a vector of numbers.
- 2. Use **c()** to create a vector of strings.
- 3. Calculate the **mean()** of a vector of numbers.
- 4. Try guessing some other simple statistics (e.g. other types of *average*) that you can use.

Getting help

If you don't know how to use a function, R has built-in help!

There are several ways you can access it:





Packages are the key to R's versatility. Over 12000 are currently available from the **Comprehensive R Archive Network** - CRAN. The **install.packages()** function can be used to install packages.

Let's install the "cowsay" package. **cowsay** is an extraordinarily useful package, as you'll see.

One way to install the package is using the console:

```
install.packages("cowsay")
```

Once it's installed, use the library() function to load the package!

library(cowsay)

But **another** way to install is using the GUI!

| RStudio Cloud × + | - 🗆 X | | | | |
|--|--|--|--|--|--|
| ← → C | 🕶 🚖 🕈 🗋 🗟 📓 💁 🌍 E | | | | |
| Your Workspace / Untitled Project * Click to name your project | Matt Craddock | | | | |
| File Edit Code View Plots Session Build Debug Profile Tools Help | | | | | |
| • ← ← ← ← ← ← ← ← ← ← ← ← ← ← ← ← ← ← ← | R 3.5.0 🗸 | | | | |
| Console Terminal × Jobs × | Environment History Connections | | | | |
| /cloud/project/ A | 💣 📊 📑 To Console 📑 To Source 🕴 🖌 🔍 | | | | |
| R version 3.5.0 (2018-04-23) "Joy in Playing" Copyright (C) 2018 The R Foundation for Statistical Computing Platform: x86_64-pc-linux-gnu (64-bit) R is free software and comes with ABSOLUTELY NO WARRANTY. You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details. R is a collaborative project with many contributors. | example <- 10 hi_there <- "hi there!" hi_there example ?mean | | | | |
| Type 'contributors()' for more information and | Files Plots Packages Help Viewer | | | | |
| Citation() on now to cite it of it packages in publications. | 🗿 Install 🜘 Update 📑 Packrat 🔍 🔍 | | | | |
| Type 'demo()' for some demos, 'help()' for on-line help, or 'help start()' for an HTML browser interface to help. | Name Description Version | | | | |
| Type 'q()' to quit R. | System Library | | | | |
| | boot Bootstrap Functions (Originally by Angelo 1.3-20 Canty for S) | | | | |
| > example <- 10 > hi there <- "hi there!" | classFunctions for Classification7.3-14 | | | | |
| > hi_there [1] "bi there!" | Cluster "Finding Groups in Data": Cluster Analysis 2.0.7-1 () Extended Rousseeuw et al. | | | | |
| > example | Code Analysis Tools for R 0.2-15 | | | | |
| | compiler The R Compiler Package 3.5.0 🐵 🗠 | | | | |
| > rmean > | ✓ datasets The R Datasets Package 3.5.0 ⊕ | | | | |
| | foreign Read Data Stored by 'Minitab', 'S', 'SAS', 0.8-70 (2010) 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', | | | | |
| | graphics The R Graphics Package 3.5.0 · 3.5.0 | | | | |
| | ✓ grDevices The R Graphics Devices and Support for 3.5.0 ⊕ ⊙ Colours and Fonts | | | | |
| | grid The Grid Graphics Package 3.5.0 🖮 🛇 | | | | |
| | KernSmoothFunctions for Kernel Smoothing Supporting Wand & Jones (1995)2.23-15 | | | | |
| | Lattice Trellis Graphics for R 0.20-35 | | | | |
| | MASS Support Functions and Datasets for 7.3-49 Image of the second secon | | | | |

| RStudio Cloud × + | | | | | |
|--|---|--|---|------------|------|
| ← → C | | | or 🖈 🕈 🗋 🔊 🛛 🖕 약 | Δ 🤇 | : |
| ■ Your Workspace / Untitled Project * Click to ne | ame your project | | ¢ (M | Matt Crade | dock |
| File Edit Code View Plots Session Build I | Debug Profile Tools Help | | | | |
| 🝳 📲 📲 🔚 📄 🛛 🧼 Go to file/function | - Addins - | | | R 3.5.0 • | |
| Console Terminal × Jobs × | Ð | Environment Hist | ory Connections | | |
| /cloud/project/ ⇔ | | 🕣 🔒 📑 To Con | sole 📑 To Source 🧕 🍯 | | |
| R version 3.5.0 (2018-04-23) "Joy in Playing Copyright (C) 2018 The R Foundation for Statist Platform: x86_64-pc-linux-gnu (64-bit) R is free software and comes with ABSOLUTELY NG You are welcome to redistribute it under certa Type 'license()' or 'licence()' for distributi | g" tical Computing D WARRANTY. in_conditions Install Packages | example <- 10 hi_there <- "hi hi_there example ?mean | there!" | | |
| R is a collaborative project with many contrib Type 'contributors()' for more information and 'citation()' on how to cite R on R packages in | Install from: (? Repository (CRAN, RSPM) | Configuring Repositorie | s es Help Viewer | | |
| | Packages (separate multiple with space or comma): | | | | IC |
| Type 'demo()' for some demos, 'help()' for on- 'help.start()' for an HTML browser interface t | cowsay | | Description | Version | |
| Type 'q()' to quit R. | Cowsay | | | | |
| > | /home/rstudio-user/R/x86_64-pc-linux-gr | nu-library/3.5 [Default] 🔻 | Bootstrap Functions (Originally by Angelo Canty for S) | 1.3-20 | |
| > example <- 10 > hi there <- "hi there!" | 🕜 Install dependencies | | Functions for Classification | 7.3-14 | |
| <pre>> hi_there [1] "bi there!"</pre> | Install dependencies | | "Finding Groups in Data": Cluster Analysi Extended Rousseeuw et al. | 2.0.7-1 | • 8 |
| > example | | nstall Cancel | Code Analysis Tools for R | 0.2-15 | |
| [1] 10 | | | The R Compiler Package | 3.5.0 | |
| > intedit | | ✓ datasets | The R Datasets Package | 3.5.0 | |
| | | foreign | Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', | 0.8-70 | |
| | | graphics | The R Graphics Package | 3.5.0 | |
| | | grDevices | The R Graphics Devices and Support for Colours and Fonts | 3.5.0 | |
| | | 🗌 grid | The Grid Graphics Package | 3.5.0 | |
| | | KernSmooth | Functions for Kernel Smoothing Supporting Wand & Jones (1995) | 2.23-15 | |
| | | lattice | Trellis Graphics for R | 0.20-35 | |
| | | MASS | Support Functions and Datasets for Venables and Ripley's MASS | 7.3-49 | |

Try out the cowsay package

cowsay adds a function called say(). Load the function in as follows, and look at the help for say().

Warning: package 'cowsay' was built under R version 4.1.1

library(cowsay)
?say()

Remember that help appears in the bottom right window!

Look at Usage and Arguments

Usage is how to use the function.

Arguments are what the functions expect and understand.



say(what = "Feed me, human.", by = "cat")



Try out the say() function

- 1. Try a few different animals by changing the **by** argument
- 2. Change what the animals say by changing the **what** argument.
- 3. Assign the output to an object using the <- operator.
- 4. Print out the value that you assigned to the object.

Additional resources



Hadley Wickham & Garrett Grolemund





(S)

There are copies of both these books in the library.

R for Data Science is available freely online at http://r4ds.had.co.nz/

An additional recommendation...

Hands-on Programming with R

Basic R programming book, also available for free online

https://rstudio-education.github.io/hopr/index.html



WRITE YOUR OWN FUNCTIONS AND SIMULATIONS

Garrett Grolemund Foreword by Hadley Wickham

This week's goals

- 1. Download R and RStudio! You'll find links and instructions on Blackboard.
- 2. Read through Chapter 1 of R for Data Science
- 3. Try out some of the introductory things mentioned in the slides, just to get a feel for using RStudio!