How to Wow with R:
A brief introduction to advanced topics

Luke McGuinness

Twitter: @mcguinlu

Department of Population Health Sciences,
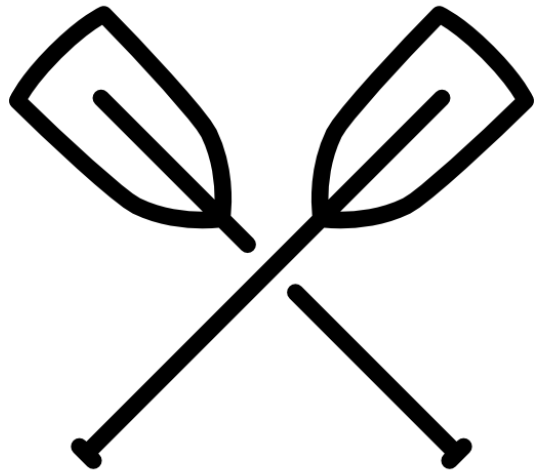### Bristol Medical School

2nd November, 2021

Intro to advanced topics

- Create **beautiful plots** with `ggplot2`

- **Literate programming** with `Rmarkdown`

- Build **interactive apps** with `shiny`

- **Share your work** as a package

Public Service Announcement

I pronounce the letter "R" oddly:

R
("oar")

means

R
("arr")

Pretty plots:
Intro to `ggplot2`

ggplot2

A package to create **highly customisable publication-ready plots**

Key elements:

- **aesthetic**: what you want to graph (e.g. x, y, colours, etc)

- **geom**: how you want to graph it (e.g. scatterplot, histogram)

- **options**: optional titles, themes, etc.

Elements are added together using **+**

```
ggplot(data = df) +
aes(x = status) +
geom_histogram(stat = "count") +
aes(fill = diet) +
labs(fill = "Diet status") +
labs(title = "Number of people by status",
     subtitle = "Colour indicates diet",
     x = "Status",
     y = "Number of people") +
theme(axis.title=element_text(colour="red"))+
theme(legend.position = "bottom") +
theme(panel.grid = element_blank()) +
theme(panel.background=element_rect(fill="bla
theme(title = element_text(size = 20))
```


Number of people by status
Colour indicates diet
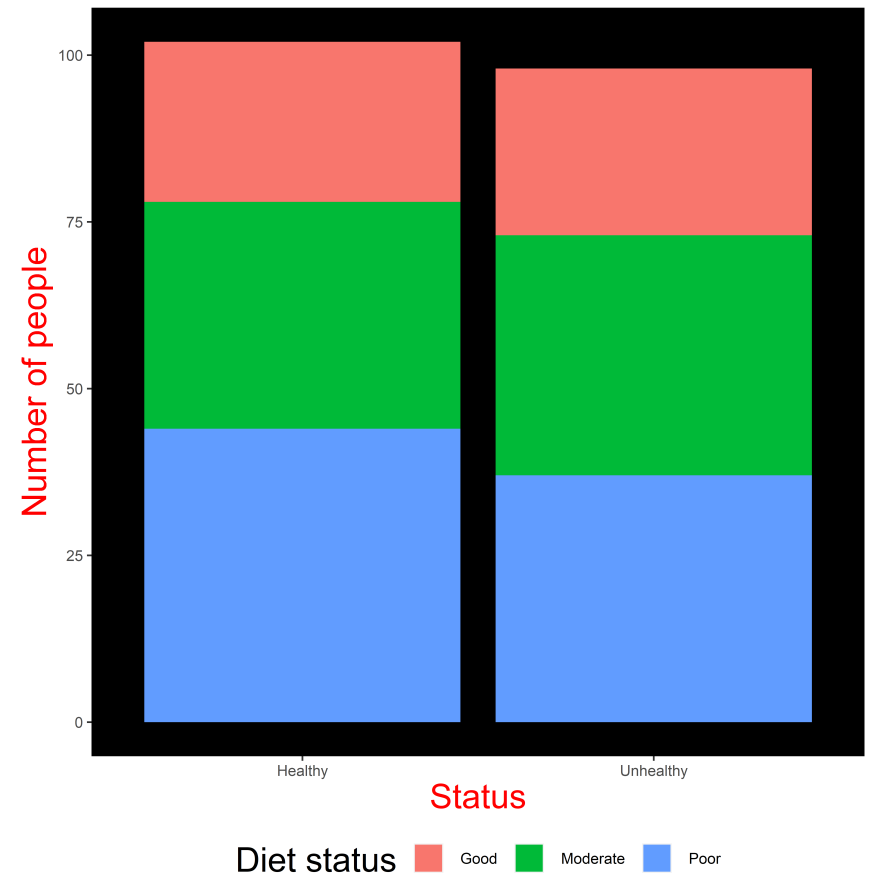
Virtually endless customisability

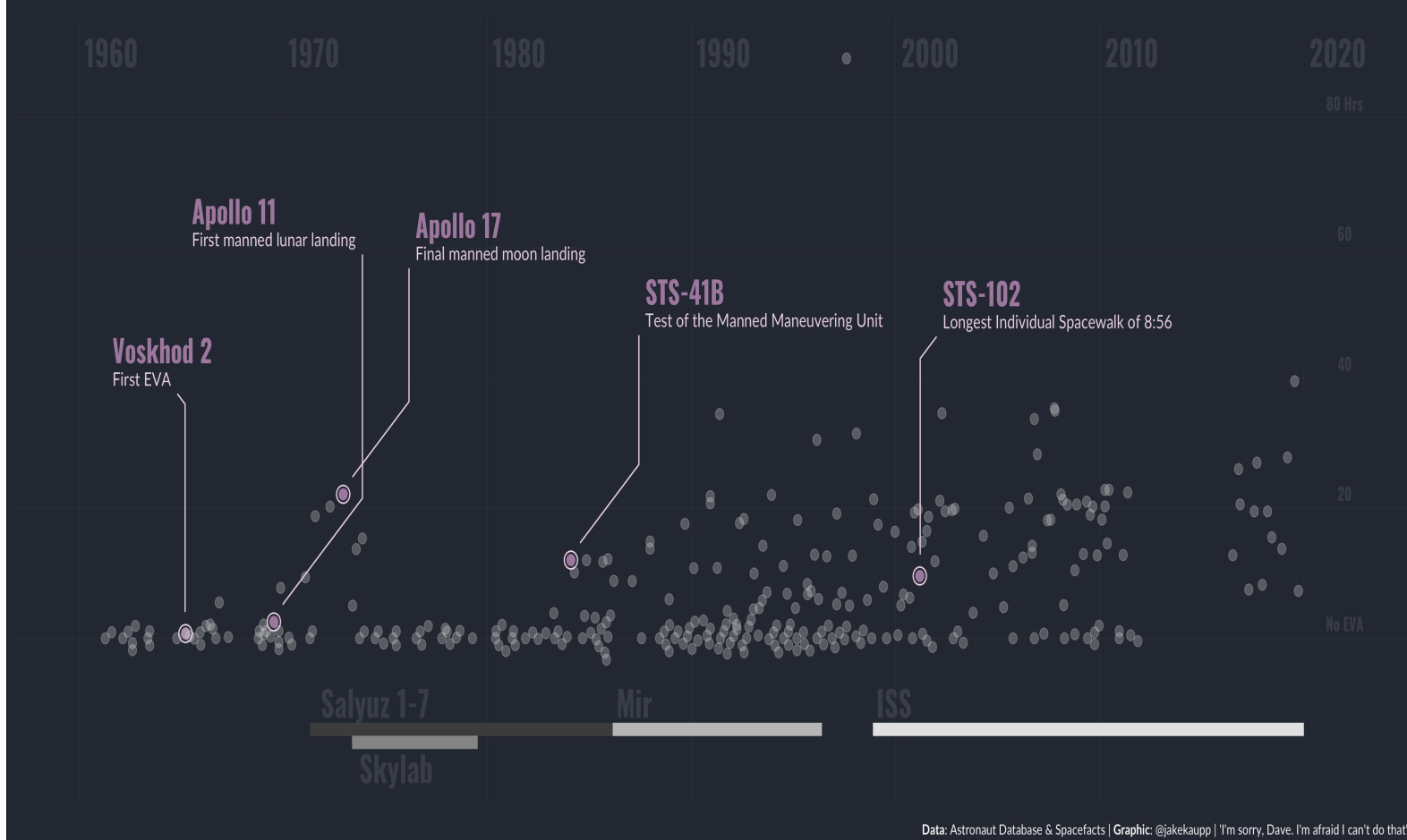Plots on previous slides are **only a taster**
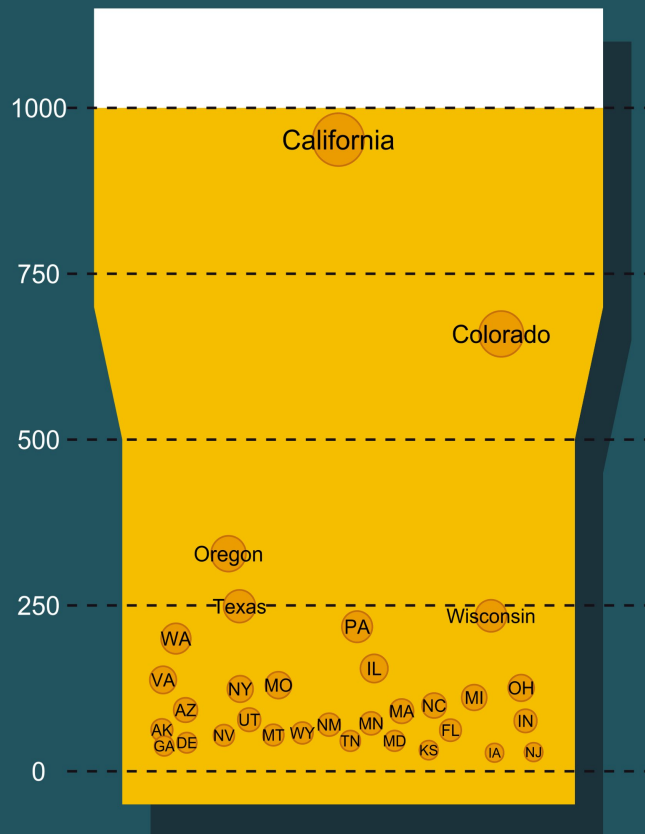
Can make **any plot you can think up**

Key resources, all open-source:

- `ggplot2` documentionation and cheatsheet - great starting point.

- "Data visualisation" by Kieran Healy - great introductory text on the science of data visualization, with examples in R.

- Tidy Tuesday - weekly community-run visualization exercise

Since We Started Making Space Stations, Everybody Just Wanted To Go Outside

Illustrated below are the total EVA hours per American and Russian Space Missions since 1960. Overlaid at the bottom are the lifespans of the 10 space stations constructed by Russia or America. The annotations highlight notable missions of interest.

1960   1970   1980   1990   2000   2010   2020

80 Hrs

60

Apollo 11
First manned lunar landing

Apollo 17
Final manned moon landing

40

STS-41B
Test of the Manned Maneuvering Unit

STS-102
Longest Individual Spacewalk of 8:56

Voskhod 2
First EVA

20

No EVA

Salyuz 1-7   Mir   ISS

Skylab

Data: Astronaut Database & Spacefacts | Graphic: @jakekaupp | 'I'm sorry, Dave. I'm afraid I can't do that'

Great American Beer Festival

The Professional Judge Panel awards gold, silver or bronze medals that are recognized around the world as symbols of brewing excellence. These awards are among the most coveted in the industry and heralded by the winning brewers in their national advertising.

In this beer, each US State with greater than 25 medals since 1987 is represented by a bubble. The higher and larger the bubble in the pint the greater the number of medals recieved. California is way in front, with 962 medals!

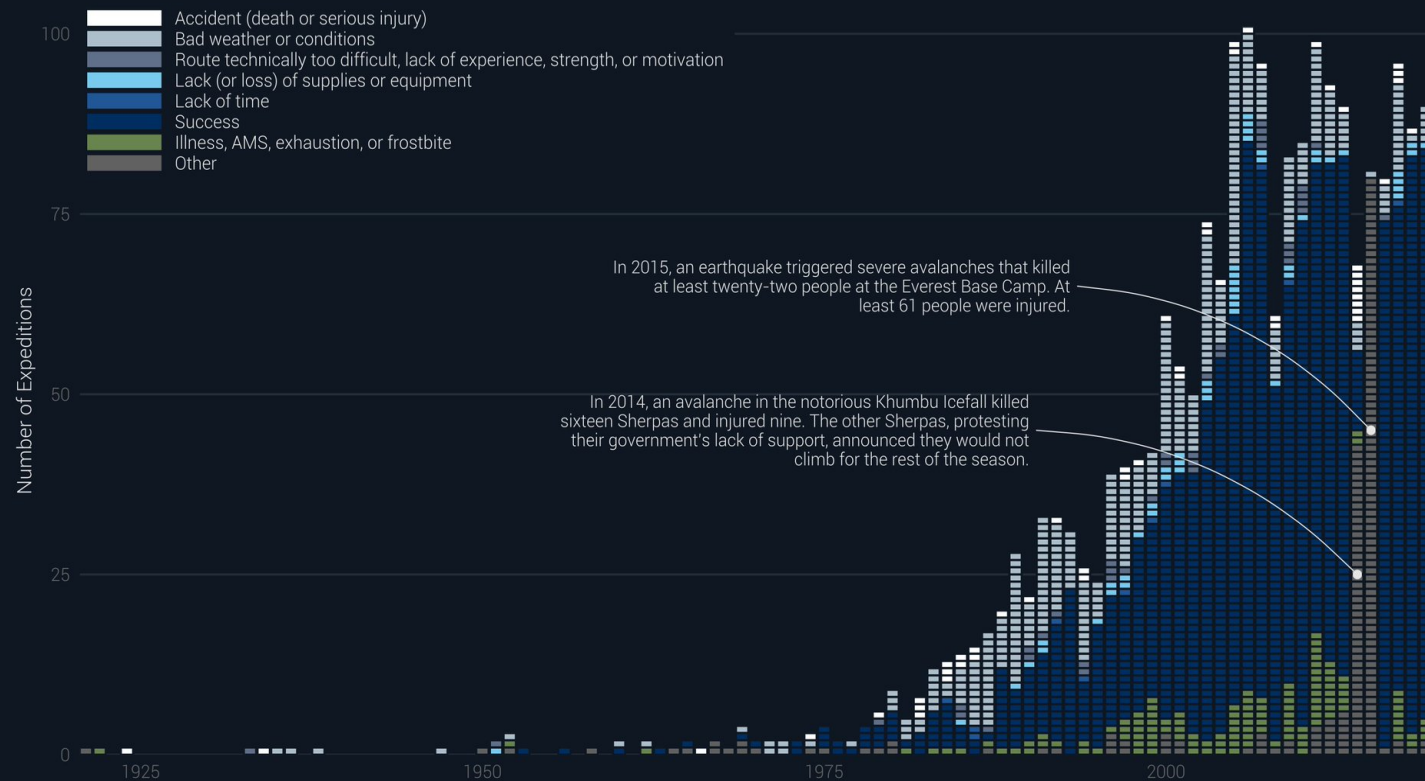Data from greatamericanbeerfestival.com/the-competition/winners
Visualisation by Jack Davison (Twitter @JDavison_)
Code found at github.com/jack-davison

EVEREST EXPEDITIONS

Journeys to the top of the world end for many reasons. This graphic
looks at the results of all Everest expeditions since 1921.
Everest has had a colorful history, but even
still most expeditions reach the peak.

Accident (death or serious injury)
Bad weather or conditions
Route technically too difficult, lack of experience, strength, or motivation
Lack (or loss) of supplies or equipment
Lack of time
Success
Illness, AMS, exhaustion, or frostbite
Other

100

75

Number of Expeditions

50

In 2015, an earthquake triggered severe avalanches that killed
at least twenty-two people at the Everest Base Camp. At
least 61 people were injured.

In 2014, an avalanche in the notorious Khumbu Icefall killed
sixteen Sherpas and injured nine. The other Sherpas, protesting
their government's lack of support, announced they would not
climb for the rest of the season.

25

0

1925   1950   1975   2000

Source: The Himalayan Database | Visualization: @charliegallaghr

Recap

Key concept is that you are **adding elements together**:

- first the data

- then the elements of the data you want to display (`aesthetics`)

- then the way you want to display them (`geoms`)

- then extra options (colour/themes/titles)

Literate programming: RMarkdown

RMarkdown

A single file (with `.Rmd` extension) that incorporates text, code and output (results and figures)

Can be converted to multiple formats:

- Word

- PDF

- HTML (web-page)

Components of a RMarkdown file

A (optional) YAML header surrounded by three hyphens (---)

- Defines metadata, such as the title, author and date
- Defines the output format (Word, PDF, HTML, etc.)

Code chunks: R code chunks surrounded by backticks (`)

- This is where the core computation happens
- Can choose to show the underlying code or just the results

Text with inline code

- Great for presenting results as part of a sentence
- E.g. The mean of the age variable was `r mean(age)`

RMarkdown Demo

Why bother?

Improves **reproducibility** and **transparency**, as the origin of every single result in the paper can be examined

Makes a lot of tasks more **straightforward**:

- Updating a manuscript with new **data/analyses**

- **Citing** other work

- Controlling **figure placement**

Things you can produce using RMarkdown

- Academic papers (example)

- Theses (`bristolthesis`)

- Supervisory reports (great if format is similar each time)

- Slides (this slide deck was made using `RMarkdown`!)

Resources

RMarkdown tutorial from RStudio

RMarkdown: The Definitive Guide by Yihui Xie

Making R accessible:
Intro to shiny

What is shiny?

shiny is an R package that allows users to **build interactive web applications** ("apps") straight from R.

Shiny apps are web-pages that
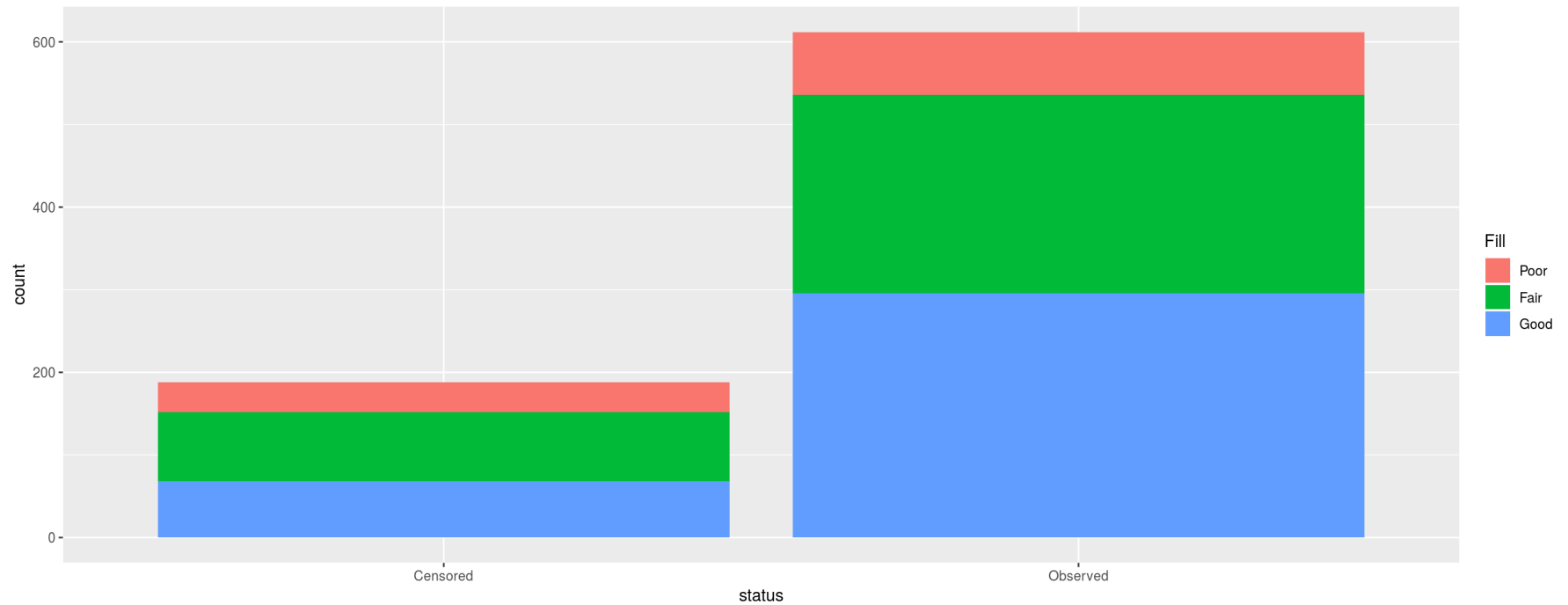users can interact with to:

- Explore data

- Perform analyses

- Create plots
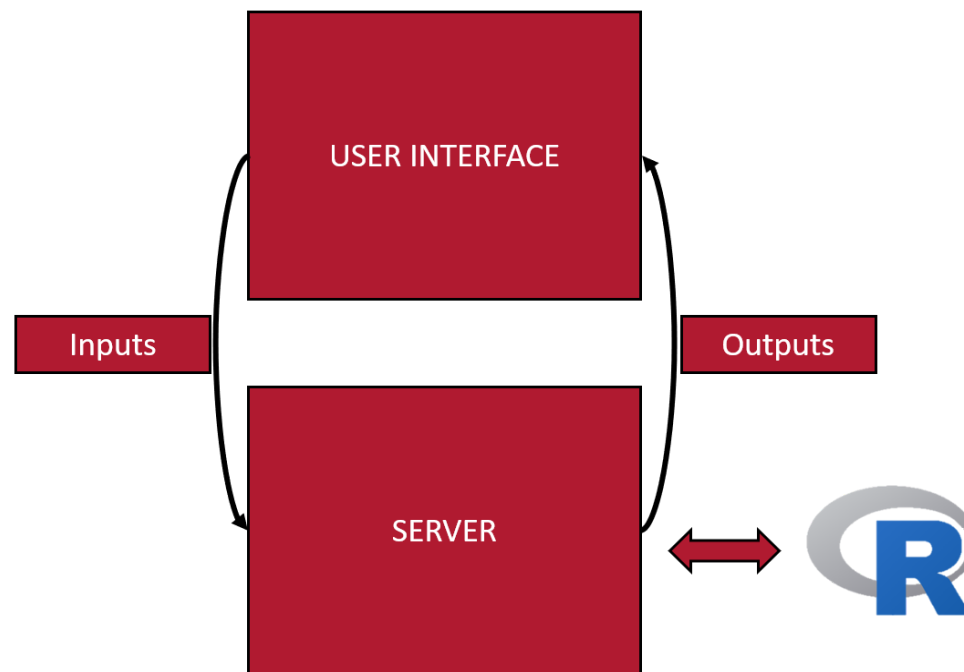
Demo of a basic `shiny` app

Variable to fill by:

Health ▾

Structure of a `shiny` app

A `shiny` app is a special type of **R** file that has three component:

- User interface

- Server

- Call to the `shinyApp`

How it works

Essentially using a **webpage** rather than **console** to instruct R

Why bother?

- R has a **high barrier to entry** for new users

- Apps allow those without any experience of R to **explore** and **benefit** from your work

- The more people who can use it, the **better known/cited** it will be!

Demo

Resources

shiny gallery - collection of example apps, with the underlying R code

shiny cheatsheet - 2-page summary of functions

shiny contest - annual competition with prizes for the most impressive/most imaginative app

Bristol short course on data visualisation and web applications has a half-day dedicated to building shiny apps

Packaging your code

Why package?

> "I wish I'd left this code across scattered .R files instead of combining it into a package" said no one ever #rstats http://t.co/udeNH4T67H
>
> — David Robinson (@drob) June 19, 2015

Benefits:

- **Reliable** and **universal** way to share code/data

- Makes it possible for others to **cite your work**

- Well-developed **testing framework**

Where to get/share packages?

CRAN

- Official repository
- Strict submission process to guarantee quality
- `install.packages("packagename")`

Bioconductor

- Topic specific repository, with a focus on bioinformatics
- Strict submission requirements, similar to CRAN
- `BiocManager::install("packagename")`

GitHub

- Popular for open source projects
- Wild west of packages
- `devtools::install_github("username/packagename")`

Resources

R Packages book by Hadley Wickham

Writing an R package from scratch by Hilary Parker

Wrapping up

Common question: Why should I use R rather than STATA/SPSS/etc?

R is **open-source** and **free** to use

Better **interface**

More **resources**

**Future-proofing** for a career outside academia

Wrapping up

Final points:

- Can do a **lot with R**, once you have a good grasp of the basics

- There is an **R package** for pretty much everything you may want to do.

- Best way to learn is to **find a problem** and try to **use R to solve it**.

R Community

- MRC IEU code-sharing channel on Slack

- `#rstats` hashtag on Twitter and "R" tagged posts on stackoverflow
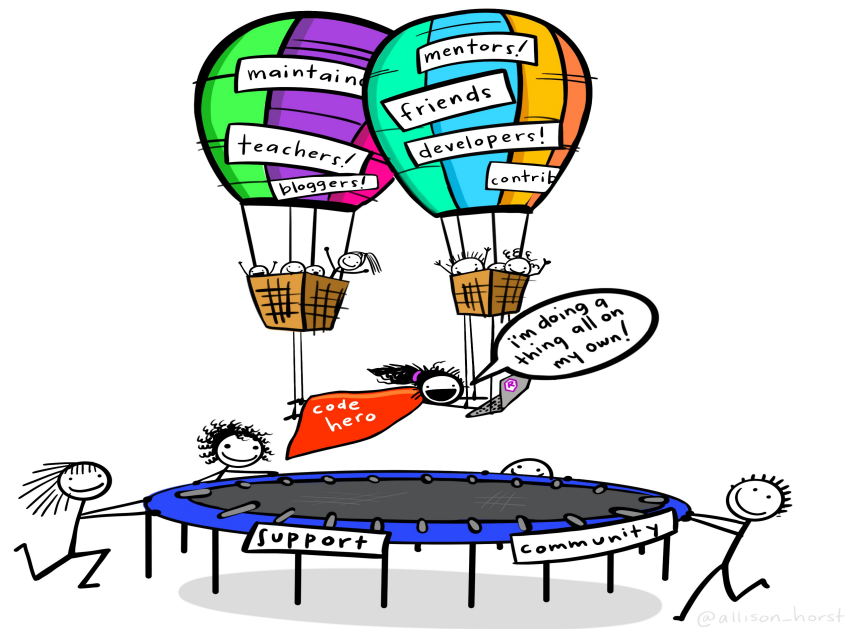
R Community

Please do get in touch if I can be of help:

Luke McGuinness

- Email: luke.mcguinness@bristol.ac.uk
- Twitter: @mcguinlu
- GitHub: @mcguinlu

Slides:

https://mcguinlu.github.io/slides/intro-to-r-phd/index.html