How to Wow with R

Luke McGuinness (+ Matt Lee)

Twitter: @mcguinlu

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

21st October, 2020

A brief intro to advanced topics

- Create beautiful plots with `ggplot2`

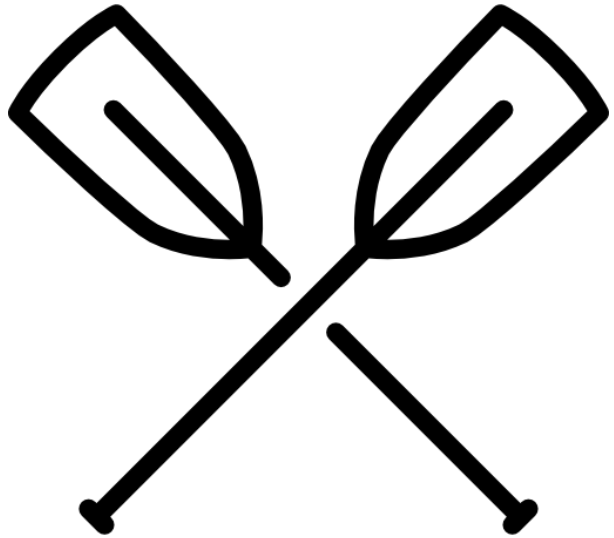
- Build interactive apps with `shiny`

- Literate programming with `Rmarkdown`

- 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

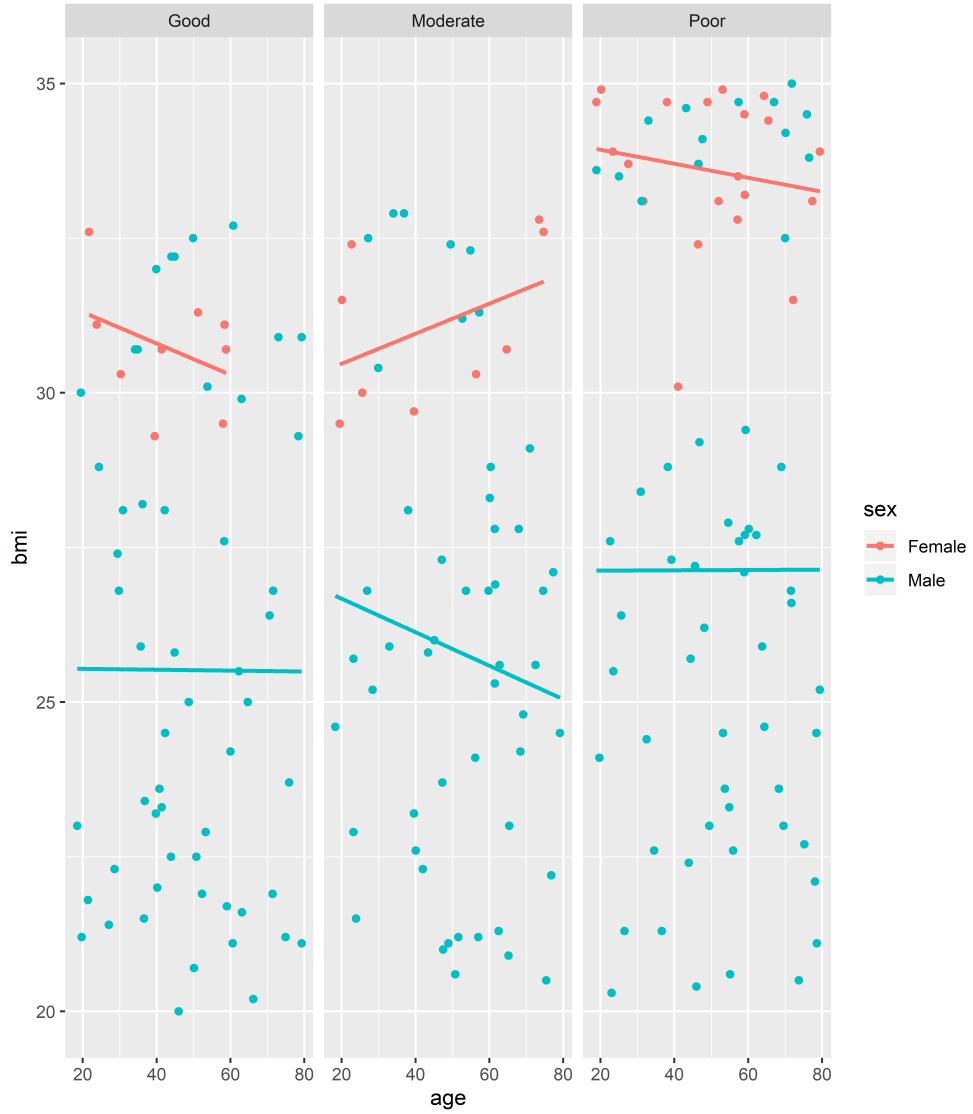
A package to create highly customisable publication-ready plots

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 = age) +
aes(y = bmi) +
geom_point() +
aes(colour = sex) +
geom_smooth(method = "lm") +
facet_grid(diet ~ .) +
facet_grid(. ~ diet)
```



Virtually endless customisability

Plots on previous slides are only a taster

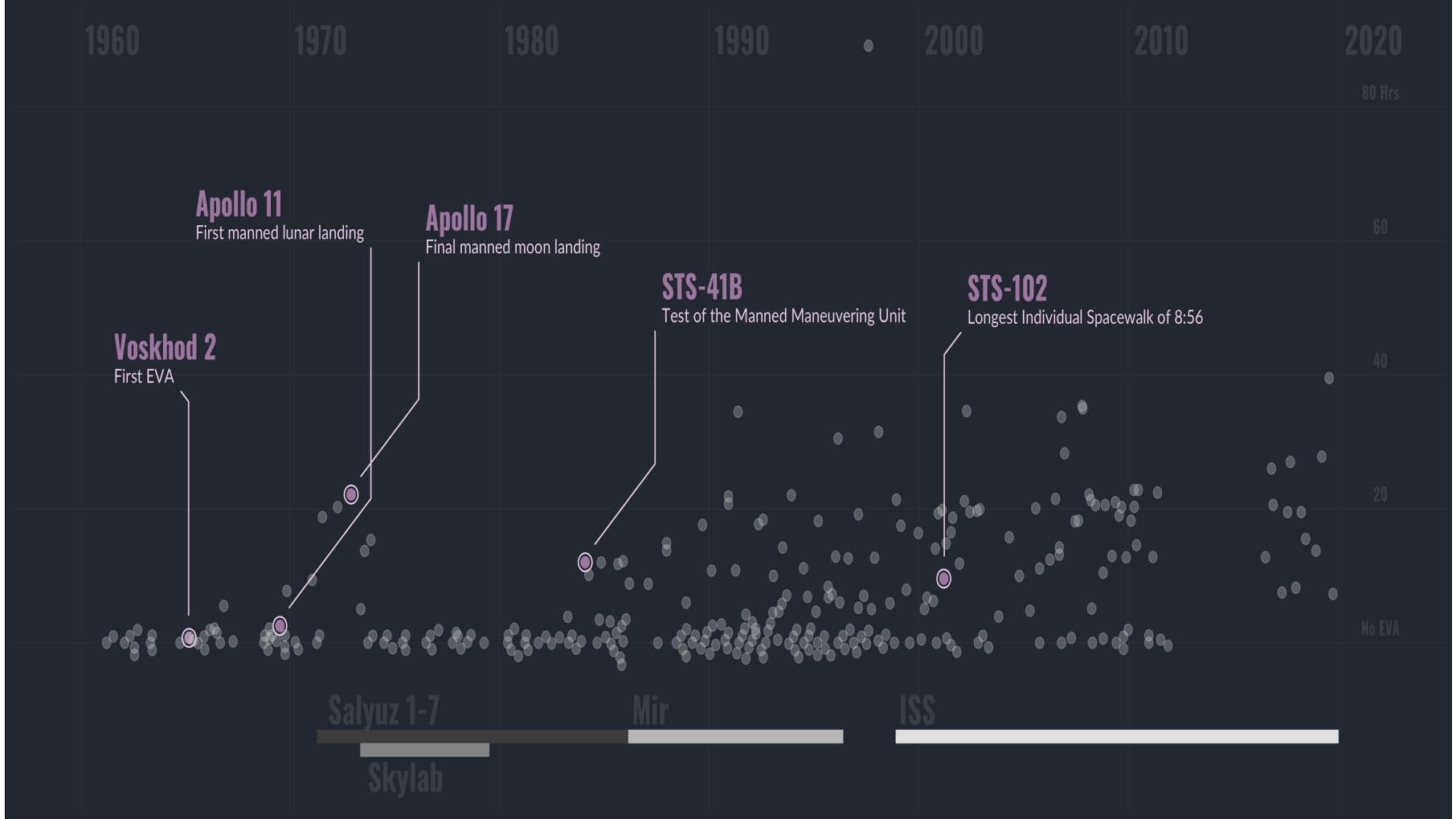
Can make any plot you can think up

Key resources, both open-source:

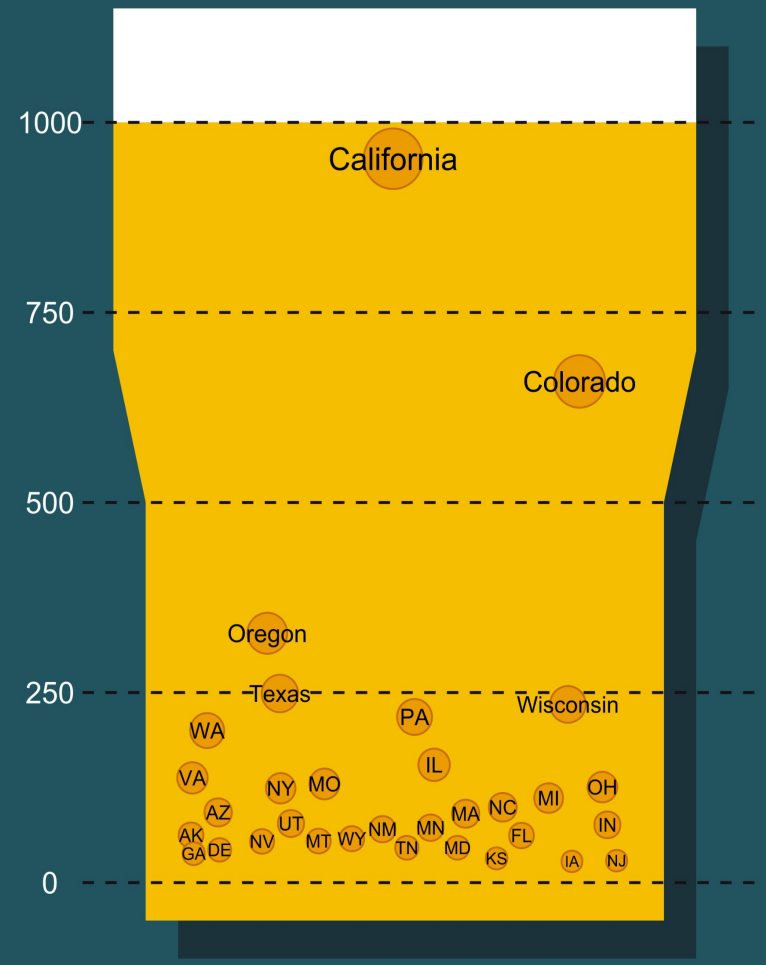
- `ggplot2` documentionation and cheatsheet - great starting point.
- "Data visualisation" by Kieran Healy - great introductory text on the science of data visualization, with examples in R.
- Tidy Tuesday - weekly community-run visualization exercise

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.



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

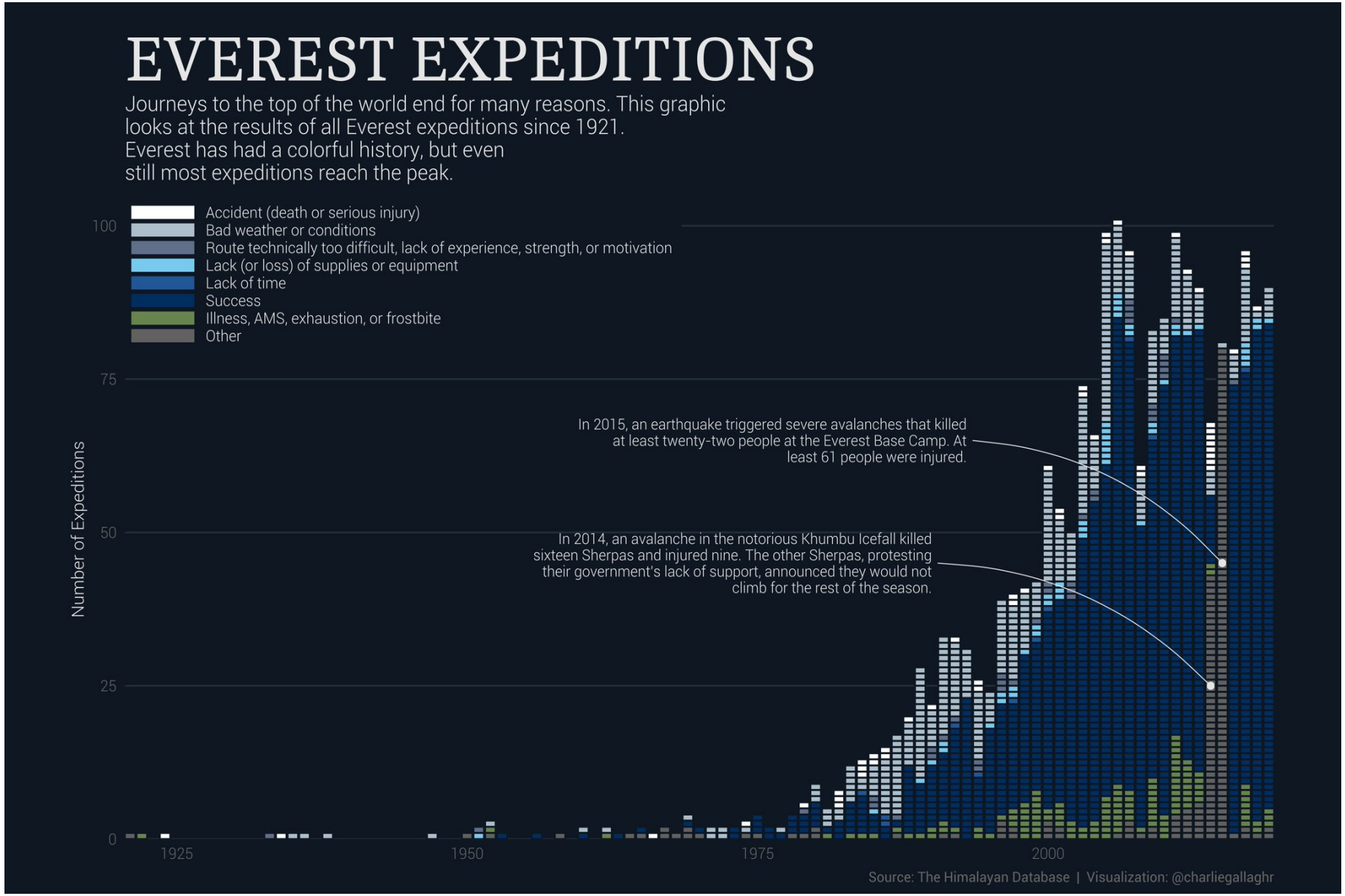


Great American Beer Festival

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
Number of Expeditions
100
75
50
25
0
1925
1950
1975
2000
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.
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:
R Markdown

R Markdown

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 `R Markdown` 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 reproduce these verbatim not, using the `echo` argument

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)` ``

R Markdown Demo

Why bother?

Avoids errors in copying numerical results into a manuscript

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

Makes updating a paper with new data really straightforward

Things you can produce using `R Markdown`

- Academic papers
- Theses
- Supervisory reports (great if format is similar each time)
- Slides (this slide deck was made using `R Markdown`!)

Resources

R Markdown tutorial from RStudio

R Markdown: 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

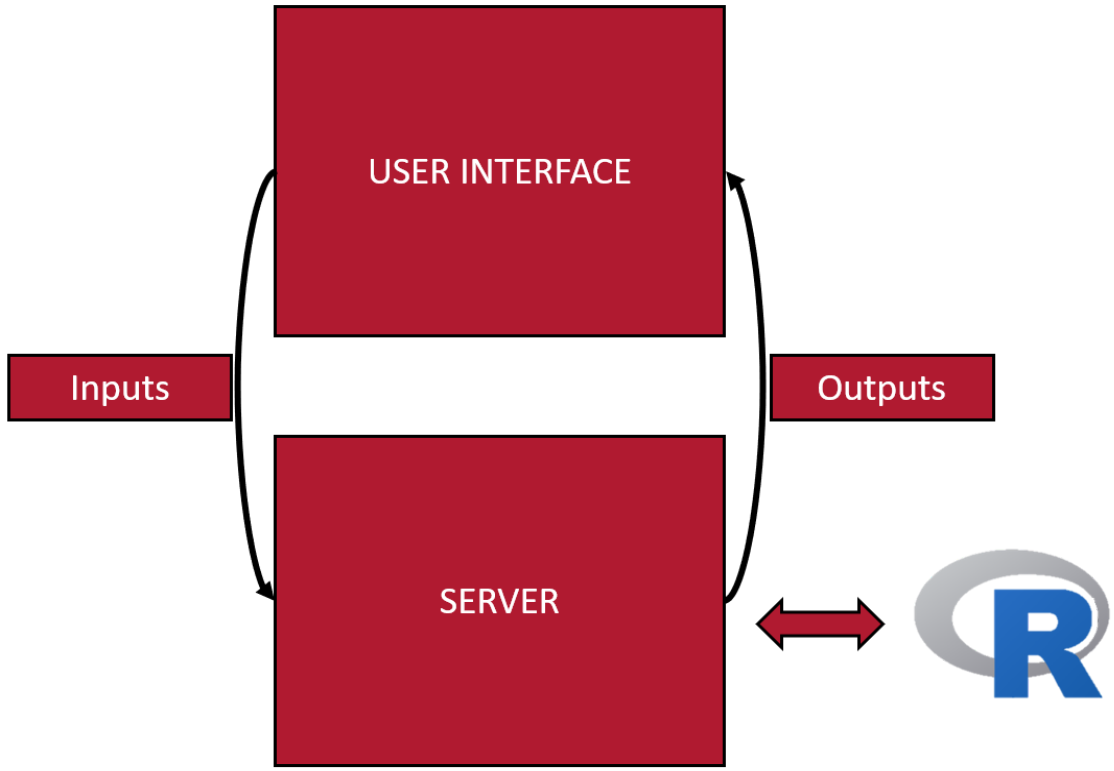
Structure of a `shiny` app

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

- **User interface**
- **Server**
- **Call to the `shinyApp` function**

How it works

Essentially using a webpage rather than command line to instruct R



Why bother?

- R has a high barrier to entry for new users
- Turning your script into an app and hosting it online means even those without any experience of R can explore and benefit from your work
- From a self-promotion point of view - the more people who can use it, the better known/cited it will be!

Demo of LSHTM COVID Tracker App

Resources

`shiny` gallery - collection of example apps, presented beside the code that powers them

`shiny` cheatsheet

`shiny contest` - annual competition with prizes for the most impressive/most imaginative app. Previous winners are good for learning new techniques.

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 - benefits your own work!

Where to get/share packages?

CRAN

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

Bioconductor

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

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

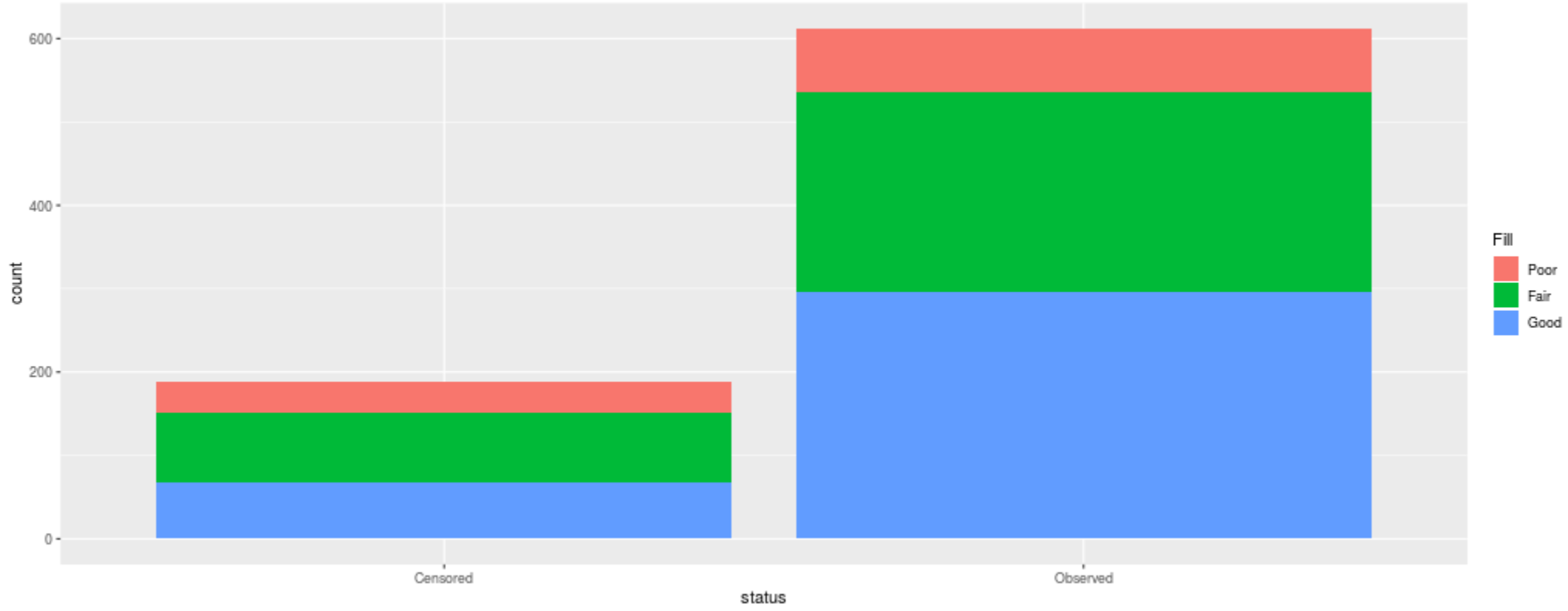
Final points:

- Can do an awful lot with R, once you have a good grasp of the basics
- There is an R package for everything (even making memes!)
- Best way to learn is to find a problem you want to solve and try and use R to do so.
- Don't be afraid to Google!
- Ugly working code >> perfect broken code

Demo of a basic `shiny` app

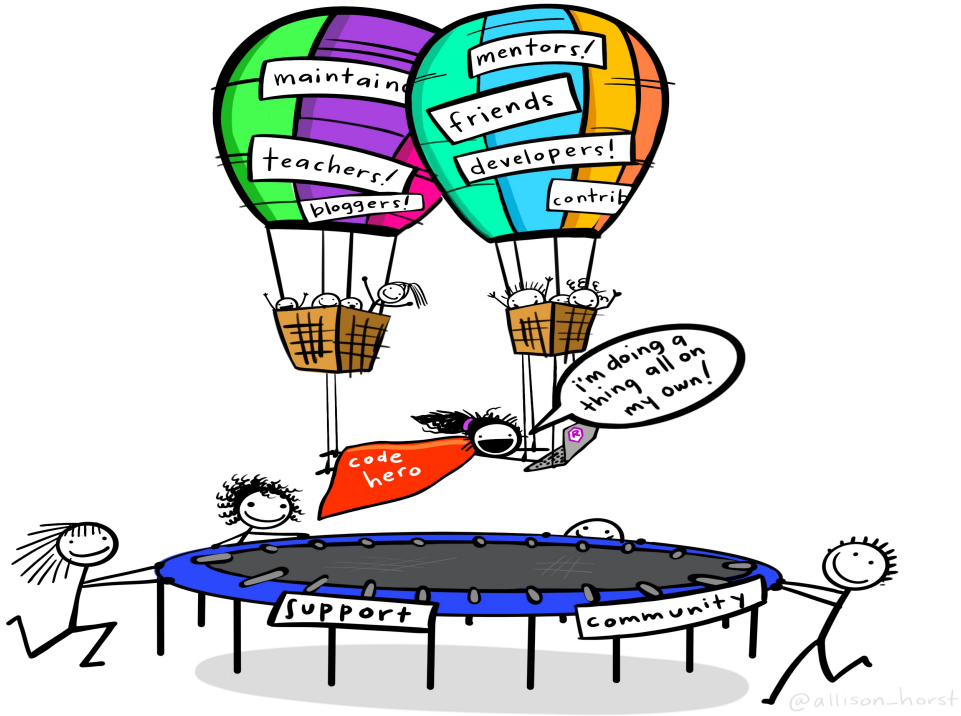
Variable to fill by:

Health



R Community

- MRC IEU code-sharing channel on Slack
- `#rstats` hashtag on Twitter and "R" tagged posts on stackoverflow



Artwork by @allison_horst

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/teaching-shiny-slides.html