

## BRM5012 R practical

# What is R?

A programming language for data science and statistics.

Open source: free to use and build on.

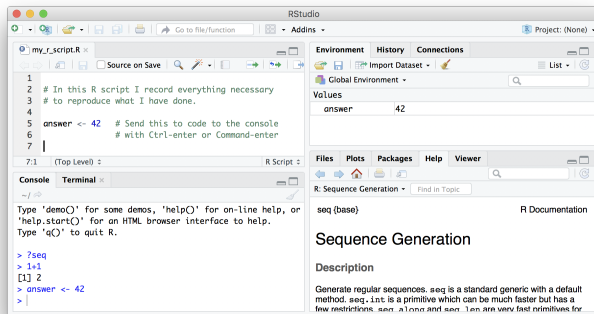
Many contributors and packages.

# R and RStudio

R is a programming language.

- ▶ Can be used directly from the “command line”.

RStudio is an environment for using R, either locally or over the web.



# Data science skills are generic

- ▶ Load data

- ▶ Tidy



- ▶ Explore, visualize

- ▶ Transform

- ▶ Summarize, model, statistical testing



- ▶ Report

See book “R for Data Science” <http://r4ds.had.co.nz/>

# R makes data science reproducible and automated

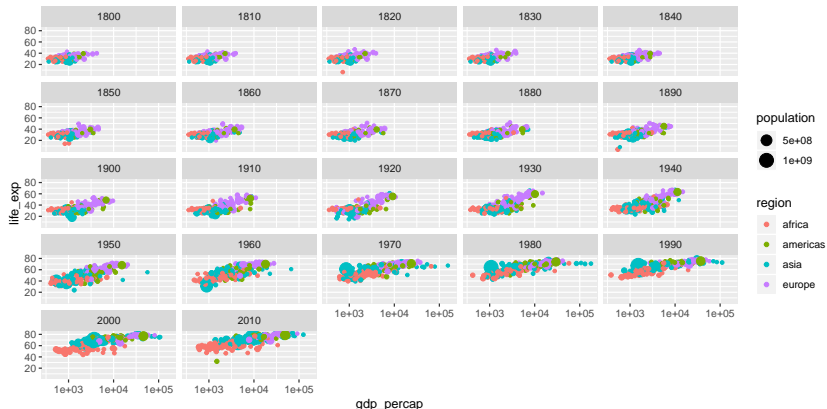
An “R script” is text that

- ▶ describes exactly what the computer should do.
- ▶ is a record of exactly what was done.

Can be turned into “functions” that become the building block of future analysis.

# R makes data science reproducible and automated

```
ggplot(gap_geo,  
  aes(x=gdp_percap, y=life_exp,  
    size=population, color=region)) +  
  geom_point() + facet_wrap(~ year) + scale_x_log10()
```



# Tidyverse

A modern collection of R packages that work well together, mostly written by Hadley Wickham.

Key packages:

- ▶ `dplyr` for manipulating tabular data
- ▶ `ggplot2` for visualization

# Bioconductor

Bioconductor is a repository of R packages for working with biological data.

- ▶ Special data types and file formats.
- ▶ Need to deal with large quantity of data.
- ▶  $p \gg n$ , high throughput experiments often produce detailed information about a small number of biological samples, requiring special statistical methods.



# Aims for these practicals

- ▶ Hands on experience using R.
- ▶ Enough about tidyverse for it to be immediately useful.
- ▶ Know enough to know the next question to ask, and where to look for the answer.

Not aiming to cover

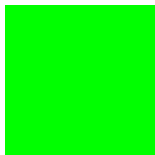
- ▶ Statistical modelling and testing (lm, etc).
- ▶ Bioconductor in depth.

Developing fluency in R will take further reading and practice.

# Workshop format

- ▶ Follow presenter in your own RStudio session, with variations.
- ▶ Short challenges to apply what you are learning, not assessed.
- ▶ Etherpad to share challenge solutions. Alternative place to ask questions.
- ▶ Multiple choice questions will be given out, and will be assessed. Hand in answers at the end of each day, also records your attendance.

## Sticky notes



All good  
Challenge completed



Something is broken  
It doesn't work  
Something doesn't make sense  
(or raise hand/call out)