Reproducibility

Practical Session

R and Stats Workshop 23rd and 24th November 2016

Susan Jarvis

Steps to reproducibility

- *Script where possible/ Script everything* - All analysis will be conducted in **R** including data preparation

- *Document your code* - We will both comment the code we write and demonstrate how to incorporate code and text within an **RMarkdown** document

- *Document your workflow* - We will use an **RStudio Project**, which will group all the files required for a particular analysis

- *Use version control* - The Project we create will be version controlled with **git** and could later be published on **Github**

Dataset

Elemental concentrations in fish from lakes in Northwest England

This resource is made available under the terms of the Open Government Licence

Barnett, C.L.; Wells, C.; Thacker, S.; Guyatt, H.J.; Fletcher, J.M.; Lawlor, A.J.; Winfield, I.J.; Beresford, N.A. (2015)
https://doi.org/10.5285/ed90df1b-462c-46bb-afbd-59794fb03f6b

Data comprise concentrations of elements in ashed fish sampled from lakes in the English Lake District in 2012 and 2013. Fish were collected from three lakes (Windermere, Bassenthwaite Lake and Derwent Water) by the Centre for Ecology & Hydrology (CEH) Lake Ecosystems group. Fish species collected were Roach (Rutilus rutilus), Perch (Perca fluviatilis), Ruffe (Gymnocephalus cernuus), Brown trout (Salmo trutta), Pike (Esox lucius) and Vendace (Coregonus albula). All samples were ashed prior to analysis by ICPMS or ICPOES to determine elemental concentrations.

Publication date: 2015-03-04

Where/When

Study area

Temporal extent
2012-01-02 to ...

Online Resources

Supporting information

Get the data

⬇ Download the data

📄 Supporting documentation

Format of the dataset: text/csv

Use of this resource is subject to these restrictions

This resource is made available under the terms of the Open Government Licence

You must cite: Barnett, C.L.; Wells, C.; Thacker, S.; Guyatt, H.J.; Fletcher, J.M.; Lawlor, A.J.; Winfield, I.J.; Beresford, N.A. (2015). Elemental concentrations in fish from lakes in Northwest England. NERC Environmental Information Data Centre. https://doi.org/10.5285/ed90df1b-462c-46bb-afbd-59794fb03f6b

BibTeX | RIS

Registration is required to access this data

Provide feedback / Report a problem

RStudio Projects

- *Document your workflow* - We will use an **RStudio Project**, which will group all the files required for a particular analysis

- Can be set up later on using an existing folder

- Essential for using git in RStudio

Script everything (in R)

You might be tempted to use things other than R:

- Exploratory data analysis

- Bringing datasets together (SQL)

- Spatial analysis

- Publication quality plots

Reproducible scripting and plotting

- Make your script easy to run in one go (like a workflow) - i.e. don't organise so you need to run something at the bottom of the script first

- Don't overwrite object names! Can lead to a lot of confusion/wasted effort

- Write plots directly to file (i.e. don't create in RStudio and then save) - more control over appearance and easier to reproduce

Writing up

- Often a cause of errors - copying and pasting or typing by hand results from one program to word processing

- Also hard to track which version of results and figures are in each document - wastes time double checking and re-running

- Difficult to get around completely when using Word - there are better alternatives e.g. LaTeX but these have a steep learning curve

- Markdown is a good middle ground, especially for writing up as you go (probably not suitable for a thesis)

Version control

- Probably the hardest to get to grips with, but potentially the most useful in the long run
- Basic idea is that version control program 'tracks' changes you make to files (code or data) so that you can
  - See how your code has changed over time
  - Revert back to a previous version if your new version doesn't work
  - Easily collaborate with others as all people have access to the most up-to-date version
  - Can track changes by different people and merge them together
- Key is that a record is made not just of the files but of all the changes you make