

# Excellent scientific practice

reproducible, open, reusable  
data and analysis

- Adopting a workflow that will make your results reproducible will ultimately make your life easier; if a problem (or question) arises somewhere down the line, it will be much easier to correct (or explain).
- To be efficient, you will need to use scripts, but you will see it is not so difficult!

# 1. Data organisation

- 1 project = 1 directory
  - get the data in the most-raw form possible and document its provenance.
  - Separate raw data from derived data
  - Write ReadMe files/ get a master file explaining your organisation and describing the experiments
- Choose file names carefully
  - Avoid using "final" in a file name.
  - File names should be explanatory, they can be long
  - Avoid spaces, use \_
  - use read.txt files to explain your naming, if not in the master file

# Data analysis

- Document every step of your analysis (every copy-paste, every modification of the data, every click and every variable value)
- At best, only use scripts : it will be quicker to document and fast to redo whatever the modification of the analysis is.

# Use script

- To download data
- To convert data type
- To change anything in the data, anything!
- To subset, filter, or re-arrange the data
- To analyse the data and produce graphs

# Specifics

- organise your code text the same way you organise your data
- Use relative path in the code.
- Random number: save your seed (save the random number generated)

# Organise your scripts, step 1

- 1. Use comment a lot: one line of comment for one line of script: what is this doing? why are you doing it?
- 2. Make short codes doing only one thing, call these codes into a master code.

# Organise your scripts, step 2

- 1. transform comments into visible outputs using knitr: comments are normal text, while your code is viewed differently.
- 2. transform some small codes into functions
- 3. pack the functions together into packages

# Version control

- Git for yourself
- Git and github online: backup, work at different places, work with different people.

# Publish everything

- Licence your code (not CC!)
- Use a script if possible :)
- From Github directly in figshare...