

Rationale for making your analysis reproducible

When delivering results to your client, it is important to include code and any other information necessary to reproduce your results. This is true even if your client is not very technically inclined. There are several reasons for this:

- Cleaning up the analysis scripts in preparation for sending them to the client can help you ensure that the code you wrote is clear and correct.
- If the client has some skill with the software you are using, he can modify the code you give him if he just wants a small tweak to the analysis or to a plot.
- If the client wants you to make a small tweak to the analysis or to a plot, it will be easy for you to do later. While you are in the middle of a project, it's hard to imagine that you will forget everything you've done. But if you need to make a modification a couple of months later, you will be happy to have an easily-reproducible record of everything taht you did.
- If the client wants to build on your analysis later, either with you or with another consultant, your work will be well documented and easy to adapt for a new project.

Some notes on structuring data/code for reproducible analysis

There are different ways to go about this, but a good simple way is to include three components: a README file, an analysis script or a directory containing the analysis scripts, and a data file or a folder containing all the data files.

1. README file (usually called `README.txt` or `README.md`).
 - If data is not included, this should include a description of how to get the data and where it should be stored/how it should be named.
 - You should include which packages are required to run your script.
 - Useful: `sessionInfo()` in R will tell you all the packages that were loaded, their version, and the version of R.
 - Should have a step-by-step description of what needs to be done to recreate your analysis (e.g.: step 1: Download data to a certain directory, step 2: make sure packages are installed, step 3: run Rmd).
 - Any information about running R scripts or Rmd files should specify the directory they should be run from.
2. Either a file with the analysis scripts, or a source folder containing the analysis scripts. The scripts should either refer to the data used with relative paths, or the README file should tell the user which paths need to be changed for the script to run on their computer.
3. Either a data file or a data folder with the data.