



# Closing Remarks: Final Comments and Encouragement

So, you really do know how to use R. You have acquired skills that allow you to import data, summarize and explore it, and plot it in a wide range of formats. You have acquired skills as well that allow you to use R to do basic statistics. Throughout, we have ensured that you understand the value of writing down and saving the instructions you want R to carry out in a script. As a result, you end up with a *permanent, repeatable, annotated, shareable archive of your analysis*. Your entire analysis, from transferring your data from your notebook to making figures and performing analyses, is all in one, secure, repeatable, annotated place.

We also have introduced you to a very fundamental rule for data analysis. *Never* start an analysis with statistical analysis. *Always* start an analysis with a picture. Why? If you have done a replicated experiment, conducted a well-organized, stratified sampling programme, or generated data using a model, it is highly likely that some theoretical relationship underpinned your research effort. You have in your head an *expected* pattern in your data. Make the picture that should tell you the answer—make

the axes correspond to the theory. If you can do this, *and* see the pattern you *expected*, you are in great shape. You will know the answer!

Throughout the book, we emphasized a workflow centred on graphing and statistical analysis. Here we expand this workflow to 17(!) steps. The additional steps here include such things as snacks and organization.

1. On your computer, create a folder for your project.
2. Enter your data into a spreadsheet (e.g. Excel), with observations in rows and variables in columns. Save it to your new folder.
3. Print a hard copy of the spreadsheet and check it against the original data sheets. Correct any errors in the spreadsheet.
4. Save the spreadsheet in comma-separated values (.csv) format.
5. Protect the spreadsheet and .csv file by making them read-only files.
6. Take a coffee, tea, or other beverage break.
7. Start up RStudio, open a new script file, and save it to the folder you created in step 1. Write your instructions for R in this script file, and save it regularly (really!). This script file and your data file are all that you need to keep safe (but you need to keep them really safe!).
8. Always put comments in your script.
9. Import your data into R. Check that the numbers of rows and variables are correct. Check that variable types are correct. Check that the number of levels of categorical variables is correct. Check that the range and distribution of numeric variables are sensible. Check for any missing values and make sure you know where they are. Check how many data points per treatment combination. Check everything!
10. Don't forget to put comments in your script.
11. Have a snack break.
12. Explore your data, primarily by using R's awesome plotting functions. *dplyr* and *ggplot2* are particularly useful at this stage. Guess the answer that your statistical tests should give you. Make some

other graphs to further investigate your guesses. Make some more figures, just to be sure.

13. Use a statistical test to check if your guess is right or wrong. If it's right, rejoice and publish. If it's wrong, figure out why.
14. Don't forget to put comments in your script.
15. Communicate your findings to your colleagues, in an email, report, manuscript, poster, web page, etc.
16. Go back and organize the files in your folder and the script you've written. Leave it all in a state that you'll be happy to come back to in six months' time—that's a long time from now, so take your time doing this.
17. Retire to your preferred socializing establishment. Try not to bore your non-R friends by talking about R all night.

If you stick to this basic workflow you will have a very solid and efficient foundation for using R. Yes, there are many ways you could go about doing what we described above. However, why not just use one way and keep your R-life simple? Make all of your scripts and analyses cover each of these steps. Make sure you annotate your script so that each of these sections is clear. If you do all of this, you will be an efficient and happy R user. And you, your friends and colleagues, and editors will appreciate all of the effort when you need to revisit, share, or publish your analysis.

Happy R-ing.

