Biol 801/Evrn 420/720, Homework for unit testing unit

deadline 1 March 31, midnight; deadline 2 April 9, midnight

Introduction to what this assignment is all about

This assignment serves two purposes: to let you practice unit testing, and as your proposal for your final project. Your grade will be based on the unit testing and coding, and the quality of your presentation of your ideas and preliminary analyses for the final project; you will additionally be given feedback on the suitability of the proposed final project idea. So you won't be graded on whether we think your final project idea is a good one, but you will be graded on whether you present it well and write good code to carry out preliminary analyses to support it. As usual, your grade will be based on serious, effort-full work on all parts of the assignment.

What you must turn in by deadline 2

By deadline 2, you must:

- You must share a git repository with Terry and Dan and send 2 pdfs by email to us
- In the repo must be a file called `Proposal.Rmd` and another called `Tests.Rmd` which, when knit, generate the two pdfs you email to us. `Proposal.Rmd` generates your proposal, including all computations that produce preliminary results for the proposal as well as well-written text and figures/tables describing your preliminary results and your plan for the final project. `Tests.Rmd` carries out your unit tests.
- In the repo must also be at least three R functions in their own scripts (.R files) that are `source`'d by both `Test.Rmd` (where they are tested) and `Proposal.Rmd` (where they are used to carry out your analyses). These should be functions representing real steps or components in your analysis.

I will read your two pdfs. I will check you are using git effectively. I will knit your two documents, making sure that works. So, unlike prior assignments, you need to pay attention to whether the knit will succeed on my machine, and will complete in reasonable time. Perhaps the best way to do this is, when you think you are done, clone your git repo onto another part of your hard drive and try knitting the "fresh" copy there, and see if it works - that's essentially what I will be doing, so it's a good way to test whether it will work when I do it.

What you must turn in by deadline 1

By deadline 1, you must:

- You must share a git repository with Terry and Dan (this is the same repo as item 1 in the previous section, so actually you will only share it with us once, for deadline 1, and then when deadline 2 rolls around all the additional commits you will have been making will already be available to us when we do a pull)

- In the repo must be at least three R functions in their own scripts (.R files), each with the *complete specification* of a function in it. This should include the first line of the function (where the function name and arguments are listed), the header comments in which all the arguments and outputs are described, and the return statement at the end. The header comments should be written such that there is only one interpretation of what the function requires for inputs, gives as outputs, and does. You need real precision of language here. Please do NOT return complete code, as that defeats the purpose of this exercise, which is about spec'ing your functions and then writing unit tests at that stage, before writing the functions themselves. You may include comments in the body of your function that designate the purpose of blocks of code that accomplish the constituent parts of what the function does (i.e., pseudocode).
- In the repo must also be a complete version of `Tests.Rmd`. Yes, this means you have to write your tests before you actually write your functions. They won't run, but they will stand ready for when the functions are written.
- The repo must have a branch tag called `halfway` that I can use to recover the state of the repo at deadline 1. This is because you will likely move on and make commits to your code in pursuit of deadline 2 before I have a chance to look at everyone's code for deadline 1, so I need to be able to use git to checkout the version of the code that was current at deadline 1.

Yes, the tasks due for deadline 1 are meant to be a half-way point along the road to completing what is due at deadline 2. If you take the unit testing and code design lessons we have learned seriously and execute them well, you should find that your final version of `Tests.Rmd` and your function specs are the same or very similar in your final repo as they are on deadline 1.

OK, I am writing a proposal for a final project, what makes a good final project?

A final project should be a mini research project, executed from start to finish using the tools and techniques we have learned (e.g., it'll be an R markdown controlling both the write-up and the code, all in a git repo, demonstrating modularity, unit testing, and a variety of the other techniques and practices we have learned). You make up your own final project and find the data for it (or you can do a theory project with no data if you want). Hopefully your final project will read like an interesting, and possibly even a novel short research paper. Now is when you can start thinking about your final project, and what you will do and what data you will use, since starting that thinking is one of the things this assignment is about. If you are having trouble thinking of a final project idea, or you have some ideas but are not sure they are suitable, please chat with one of us any time. Final projects do not have to be ambitious scientifically - there is not really enough time for that, and what you will be trying to demonstrate is what you have learned in this couse, not how well you can do research. (Though those overlap, to some extent, of course, since we hope this course improves your ability to do research!) So scientfically simple final projects can be good. BUT, the final project does have to be complex enough to make a good canvas for demonstrating all the skills you have gained in this course. And if you can make your final project serve your own research goals, that's a great synergy (perhaps it later becomes your first these chapter, or senior thesis or something).

Other comments

This is a fairly open-ended assignment. You may have noticed the assignments have been getting less prescribed and more open-ended as the course progresses, which is consistent with the idea of helping you in your actual research or other analytic/computational projects, which are open ended. In any open ended assignment, including this one, the goal is not to accomplish certain pre-defined tasks which have right and wrong answers, but instead to demonstrate the extent of your knowledge and skills, impressing the grader. In other words, as in research, there is no one correct path, and we don't have in mind an ideal result, but some projects will be very well done and impressive and others may be . . . less so (though I hope they will all be good!).