CMEE Masters: Computing Coursework Assessment

Note that:

All script/code errors and other info mentioned below are in the weekly log files

In the weekly feedback/assessments, please compare with the solutions whenever needed to see why I might have taken off points for a particular exercise/script or code file. We can then discuss these in your 1:1 post-assessment feedback session.

Assignment Objectives: To work on a series of computing/programming exercises and problems in a coherent, modular, reproducible workflow under version control.

Student's Name: Hannah O'Sullivan

Overall Project workflow

Found all the expected weekly directories in your parent directory.

You had a .gitignore throughout, with meaningful exclusions specific to certain weeks – great. But there were many more exclusions possible; check this out: `https://www.gitignore.io`.

You had an overall readme file, and then one within each week. The weekly Readmes were succinct and clear, including a description of what the overall project structure. You could have also included language versions and dependencies requirements. Check out this resource: `https://github.com/jehna/readme-best-practices`. As you become a seasoned programmer, you will learn to make the readme file descriptions more informative yet succinct. *And you are welcome, Rey!*

Your Git repo size when I checked week 7 was about 2.01 MB — a small size, suggesting you did not keep unnecessary binary files under VC, and that you did not commit excessively. It could also mean that you did not commit enough, and/or somehow along the the way lost parts of your git history — but I won't check these possibilities!

WEEK 1

Found directories Data, Sandbox, Code

Found 12 code files: CompileLatex.sh, ConcateTwoFiles.sh, CountLines.sh, variables.sh, tiff2png.sh, csvtospace.sh, UnixPrac.txt, FirstExample.tex, MyExampleScript.sh, FirstBiblio.bib, tabtocsv.sh, boilerplate.sh

UnixPrac1.txt was fine. Each solution was described in a comment, great. You could have broken the description down into the key components of the unix command, but that's OK. Compare with the solutions, especially the last one.

csvtospace.sh and other shell scripts were fine. You could have made them throw an error (with a message) if no input csv file(s) were provided. In general, it is a good idea to add some input checks and return a meaningful message with error for utility files like this, especially in case somebody else uses it. But it's OK. No points deleted for this.

Points for this week: 100

WEEK 2

Found the Code, Sandbox, Data, Results directories

Found 21 code files: lc2.py, boilerplate.py, basic_csv.py, cfexercises2.py, align_seqs_better.py, dictionary.py, debugme.py, scope.py, cfexercises1.py, tuple.py, basic_io.py, lc1.py, oaks_debugme.py, oaks.py, loops.py, using_name.py, align_seqs.py, sysargv.py, align_seqs_fasta.py, control_flow.py, test_control_flow.py

You had a gitignore in the results directory in multiple weeks; this is not necessary, as all file/directory patterns in subdirectories can be excluded from a single .gitignore in the parent directory of the repository. See `https://git-scm.com/docs/gitignore` and `https://labs.consol.de/development/git/2017/02/22/gitignore.html` : -2.5pts

scope.py gave an error: -5pts

lc1.py, lc2.py, dictionary.py, tuple.py were all nicely done. They could have produced better-formatted output – Compare with the solutions on the repo; -1 pt each.

align_seqs.py was nicely done. You could have written it as a self-sufficient script that could also take external inputs optionally (though I did not ask for it specifically). Compare with the solution.

You did align_seqs_fasta.py and align_seqs_better.py correctly: +5 extra credit pts.

All other scripts were fine.

Points for this week: 97.5

WEEK 3

Found directories Practicals, Code, Data, Results

Found 32 code files: browse.R, PP_Regress.R, Vectorize2.py, apply1.R, sample.R, TimeVecs.sh, run_get_TreeHeight.sh, get_TreeHeight.py, CompileLaTeX.sh, boilerplate.R, TAutoCorrLaTeX.tex, gpdd_map.R, TreeHeight.R, PP_Lattice.R, next.R, Girko.R, Vectorize1.R, break.R, plotLin.R, basic_io.R, Vectorize1.py, try.R, apply2.R, get_TreeHeight.R, TAutoCorr.R, Vectorize2.R, DataWrangTidy.R, preallocate.R, PP_Regress_loc.R, DataWrang.R, MyBars.R, control.R

Found TAutoCorrLaTeX.pdf in Results directory: ideally this directory should be empty other than, perhaps, a readme. Also, same comment as before about .gitignore

DataWrang.R gave a path error: -2 pts

Vectorize1.R was fine.

Vectorize2.R was fine, except for a bunch of commented out lines. Also, compare with the solution.

PP_Regress.R: fine. Also have a look at my solution.

TAutoCorr.R was OK, but you should have shuffled the entire time series instead of the two equal sets separately – compare with the solution; does your solution permute every pair of successive years? -3pts

The report: Nicely done, but you could have kept the writeup in a completely separate directory. Glad you plotted the histogram of the permuted correlation coefficients as well. You could also have plotted the correlation pattern itself. Some more interpretation of the results would have been nice. -2pts

You did the Mapping (good description of the data and potential biases) and pp_regress_loc extra credits – +5 pts.

Points for this week: 98 pts

WEEKS 4, 5 & 6

Not assessed, but happy you kept everything organized as much as possible.

WEEK 7

Found directories Code, Data, and Results

Found a README

Found 16 code files: TestR.py, regexs.py, LV3.py, profileme2.py, timeitme.py, blackbirds.py, TestR.R, profileme.py, LV4.py, fmr.R, using_os.py, LV1.py, DrawFW.py, runLV.py, run_fmr_R.py, LV2.py

MyFirstJupyterNb.ipynb missing: -10 pts

regexs.py and gave an error: -5pts

blackbirds.py gave a path error: -2pts

blackbirds.py was fine otherwise as far as I can tell, compare with solution anyway.

The two LV* scripts with profiling were fine. You also did the LV3-4 extra credits: +5pts . Also look at the solution for run_LV, which is shorter.

using_os.py is fine. The scrip provided some more meaningful output to screen – good. Compare with the solution.

Docstrings missing: -7pts

Points for this week: 81

Overall Assessment

You did a very good job overall, including most extra credit Qs.

Neat, clean project organization and code, but a few silly errors. I was impressed by your efforts to understand as many details of the programming languages and coding as possible. You clearly like coding!

Overall, as this is the first time you have done programming in a heady mix of UNIX, Python, & R with a sprinkling of LaTeXand git, you did very, very well! In particular, you seem to on your way to becoming equally comfortable with Python and R.

It was a tough set of weeks, but I believe your hard work in them has given you a great start towards further training, a quantitative masters dissertation, and ultimately a career in quantitative biology!

Provisional Mark: 81

The overall assessment will typically have significantly lesser marks than a simple weighted average of each week's points because the overall assessment is based on not just the "Computing Coursework Assessment Criteria", but also the the "Marking Criteria for Exams, Essays and Coursework". Both sets of marking criteria are in the Assessment Appendix of the online TheMulQuaBio notes and git repository.

We will discuss where you gained or lost marks, and what you could have improved further in your 1:1 post-assessment feedback session. To the extent possible, please come with questions about specific scripts based upon the feedback you have received. This may require you to compare your code with the solution code in many cases.

Signed: Samraat Pawar

January 18, 2019