

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: Shiyun Liu

Overall Project workflow

Found all the expected weekly directories in your parent directory.

Your Git repo size when I checked week 7 was about 84.25 MB — a good 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 way lost parts of your git history — but I won't check these possibilities!

You had a .gitignore throughout, with meaningful exclusions specific to certain weeks – great. You could have had more patterns; 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 is. You could have been explicit about the language requirements and package dependencies. Also 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 even more informative yet succinct. No need to list files in sandbox, for example.

WEEK 1

Found directories Data, Sandbox, Code

Found 5 code files: CompileLaTeX.sh, csvtospace.sh, FirstExample.tex, FirstBiblio.bib, UnixPrac1.txt

ConcatenateTwoFiles.sh, CountLines.sh, variables.sh, tiff2png.sh, tabtocsv.sh, boilerplate.sh
missing: -50pts

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

csvtospace.sh was fine. You could have made it give (with a message) if no input csv file was 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.

Points for this week: 50

WEEK 2

Found the Code, Sandbox, Data, Results directories

Found 20 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, loops.py, using_name.py, align_seqs.py, sysargv.py, align_seqs_fasta.py, control_flow.py, test_control_flow.py

oaks.py missing: -10pts

dictionary.py, oaks_debugme.py, align_seqs_fasta.py gave errors: -15pts

lc1.py, lc2.py, dictionary.py (apart from the error), tuple.py were all fine. They could have produced slightly better-formatted output – Compare with the solutions on the repo.

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 (except for the error) and align_seqs_better.py correctly — so +5 extra credit pts.

All other scripts were fine – good job.

Points for this week: 80

WEEK 3

Found directories Practicals, Code, Data, Results

Found 28 code files: PP_Regress.R, apply1.R, sample.R, Vectorize.R, run.get_TreeHeight.sh, get_TreeHeight.py, CompileLaTeX.sh, boilerplate.R, TreeHeight.R, PP_Lattice.R, next.R, Girko.R, Vectorize1.R, break.R, plotLin.R, basic_io.R, correlation_result.tex, try.R, apply2.R, get_TreeHeight.R, TAutoCorr.R, browser.R, Vectorize2.R, DataWrangTidy.R, preallocate.R, DataWrang.R, map.R, control.R

Found a populated Results directory; ideally, this directory should be empty other than, perhaps, a Readme.

Girko.R, plotLin.R, TAutoCorr.R (silly typo!), map.R gave errors: -20 pts

You did the Vectorize* scripts, but did not compare them. They were not quite correct either — compare with the solutions.

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

You did the Mapping (good description of the data and potential biases) extra credit, though it had an error (don't put a install.package command in scripts) – +1 pt .

Points for this week: 80 pts

WEEKS 4, 5 & 6

Not assessed, but happy you kept things somewhat organized.

WEEK 7

Found directories Code, Data, and Results

Found a README

Found 17 code files: TestR.py, LV3.py, profileme2.py, timeitme.py, blackbirds.py, regex.py, TestR.R, profileme.py, LV4.py, fmr.R, MyFirstJupyterNb.ipynb, using_os.py, LV1.py, DrawFW.py, run_fmr_R.py, run_LV.py, LV2.py

MyFirstJupyterNb.ipynb missing: -10 pts

LV3.py, regex.py, LV4.py, fmr.R gave errors: -20 pts

The LV* scripts with profiling were OK, but with some small mistakes — compare with the solutions.

You did the LV3-4 extra credits, but with errors: +2pts for effort.

using_os.py is OK, but compare with the solution. The script could have provided some more meaningful output to screen. -2pts

Blackbirds.py was fine, but the output could have been better formatted (why repeat “Kingdom”, “Phylum” etc in every row?).

Points for this week: 70

Overall Assessment

You did an OK overall, including some extra credit Qs.

Reasonably clean project organization and code, but too many errors, most of which could have been easily fixed.

You clearly found these weeks very challenging, but I was impressed by your persistence. However, as this is the first time you have done programming in a heady mix of UNIX, Python, & R with a sprinkling of L^AT_EX and git, you did very well, keep going!

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: 58

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