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: Yuheng Wang

Overall Project workflow

Found all the expected weekly directories in your parent directory.

You had a .gitignore throughout, with meaningful exclusions specific to certain languages/weeks – great. I am glad you used gitignore templates that you found online. However, at times the exclusions were not relevant (e.g., LaTeX beamer).

You had a readme file with a list of the weeks' content, and then within each week, a readme for each week. The readmes were excellent - great job. Also check out this resource: `https://github.com/jehna/readme-best-practices`

Your Git repo size when I checked week 7 was about 99.61 MB — a OK 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, Results Sandbox, Code

Found 12 code files: ConcatenateTwoFiles.sh, Firstbiblio.bib, CountLines.sh, variables.sh, CompileLaTeX.sh, tiff2png.sh, csvtospace.sh, FirstExample.tex, MyExampleScript.sh, UnixPrac1.txt, tabtocsv.sh, boilerplate.sh

Found a FirstExample.pdf: -0.5pt

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

csvtospace.sh was fine, but one addition you could have made to the script was to throw an error (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. Similar comment for ConcatenateTwoFiles.sh (running without two input files will not work), tabtocsv.sh and CompileLaTeX.sh, CountLines.sh, Variables.sh. But it's OK. No points deleted for this.

Points for this week: 99.5

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

Found no extra files; great!

basic_csv.py gave an error because of a missing data file: -5 pts

lc1.py, lc2.py, dictionary.py, tuple.py were all fine. They could have given an better formatted output – Compare with the solutions on the repo; -1 pts each.

align_seqs.py was nicely done. Glad you wrote it as a self-sufficient script that could also take external inputs. Also compare with the solution.

You did align_seqs_fasta.py and align_seqs_better.py — so +5 extra credit points.

All other scripts were fine.

Points for this week: 100

WEEK 3

Found directories Code, Data, Results

Found 29 code files: browse.R, PP_Regress.R, Vectorize2.py, apply1.R, sample.R, run_get_TreeHeight.sh, get_TreeHeight.py, compile_TAutoCorr.sh, boilerplate.R, TreeHeight.R, PP_Lattice.R, PlotLin.R, next.R, Girko.R, Vectorize1.R, time_compare.sh, break.R, basic_io.R, GDPP_map.R, Vectorize1.py, try.R, apply2.R, get_TreeHeight.R, TAutoCorr.R, Vectorize2.R, DataWrangTidy.R, preallocate.R, TAutoCorr.tex, control.R

Found a populated Results directory – should be empty other than, perhaps, a readme. -2pts

Vectorize1.R was fine.

Vectorize2.R —check the solution ,as it takes longer than it should. -3pts

PP_Regress.R: good — also have a look at my solution.

basic_io.R gave an error: -5pts

TAutoCorr.R gave an error (problem with ddata file)) – -5pts.

But the solution was correct. Also compare with the solution for a slightly different approach.

The report: OK. You could have plotted the correlation, and histogram of the permuted correlation coefficients as well. Some more interpretation would have been nice!

You did the Mapping and PP_Regress_loc extra credit correctly - +5 pts.

Points for this week: 90 pts

WEEKS 4, 5 & 6

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

WEEK 7

Found directories Code, Data, and Results

Found a README

Found 16 code files: TestR.py, regexs.py, profileme2.py, blackbirds.py, Nets.py, TestR.R, profileme.py, fmr.R, using_os.py, LV1.py, DrawFW.py, Nets.R, run_fmr_R.py, run_LV.py, LV2.py, timeitime.py

MyFirstJupyterNb.ipynb was missing: -10 pts

Found a populated Results directory – should be empty other than, perhaps, a readme. -2pts

using_os.py worked, but compare with the solution. The script could have provided some meaningful output to screen.

blackbirds.py was fine. Do look at the solution as well.

You did the two LV* scripts with profiling. Good. Also look at the solution, which is simpler, and covers the other LV challenges as well.

You also did the extra-credit nets.py: +2.5 pts

Points for this week: 90.5 pts

Overall Assessment

You did an excellent job overall, including many extra credit Qs.

Very few errors. You also went just that extra mile in many cases. You clearly like coding!

Overall, You delivered on most fronts, and if 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 well! In particular, you seem to be well on your way to becoming comfortable with both Python and R – that's great!

I was impressed by your efforts to understand as many details of the programming languages and programming as possible.

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

Provisional Mark: 85

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