

# 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:** Xiaosheng Luo

## Overall Project workflow

Found all the expected weekly directories in your parent directory.

Your Git repo size when I checked week 7 was about 64.57 MB — a reasonable 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 will likely find this useful as well: <https://www.gitignore.io>.

You had a very nice 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 and what the language and dependencies requirements are – great job. 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.

## WEEK 1

Found directories Data, Sandbox, Code

Found 11 code files: ConcatenateTwoFiles.sh, CountLines.sh, variables.sh, CompileLaTeX.sh, csvtospace.sh, FirstExample.tex, MyExampleScript.sh, FirstBiblio.bib, UnixPrac1.txt, tabtocsv.sh, boilerplate.sh

tiff2png.sh missing: -10pts

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 was fine. You could have made it give 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 other shell scripts.

Points for this week: 100

## WEEK 2

Found the Code, Sandbox, Data, Results directories

Found 20 code files: lc2.py, boilerplate.py, basic\_csv.py, cfexercises2.py, scopy.py, align\_seqs\_better.py, dictionary.py, debugme.py, cfexercises1.py, tuple.py, basic\_io.py, lc1.py, oaks\_debugme.py, oaks.py, loops.py, align\_seqs.py, sysargv.py, align\_seqs\_fasta.py, control\_flow.py, test\_control\_flow.py

scope.py, using\_name.py missing: -20 pts

lc1.py, lc2.py, dictionary.py, tuple.py were all fine. They could have produced better-formatted output (lc2.py produced none) – Compare with the solutions on the repo. -0.5pt 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 — so +5 extra credit pts.

All other scripts were fine – great job.

Points for this week: 83

## WEEK 3

Found directories Practicals, Code, Data, Results

Found 31 code files: browse.R, PP\_Regress.R, Vectorize2.py, apply1.R, sample.R, Vectorize\_Compare.sh, run\_get\_TreeHeight.sh, get\_TreeHeight.py, Mapping.R, boilerplate.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, TAutoCorr.tex, MyBars.R, control.R

MyBars.R, Girko.R had errors: -10 pts

Vectorize1.R was fine.

Vectorize2.R was fine, nice job — compare with the solution.

PP\_Regress.R: fine. Also have a look at my solution.

TAutoCorr.R was fine; compare with the solution as well.

The report: Nicely done. Glad you plotted the histogram of the permuted correlation coefficients as well. You could also have plotted the correlation pattern itself. You could have kept the writeup in a completely separate directory

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

Points for this week: 95 pts

## **WEEKS 4, 5 & 6**

Not assessed.

## WEEK 7

Found directories Code, Data, and Results

Found a README

Found 16 code files: timeitem.py, TestR.py, LV3.py, profileme2.py, blackbirds.py, Nets.py, profileme.py, LV4.py, fmr.R, using\_os.py, LV1.py, DrawFW.py, Nets.R, run\_fmr\_R.py, run\_LV.py, LV2.py

regexs.py and MyFirstJupyterNb.ipynb missing: -20 pts

The two LV\* scripts with profiling were fine. You also did the LV3-4 extra credits: +5pts . Also look at the solution.

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

Nets.py: Good job!: +2.5pts

blackbirds.py was fine.

Points for this week: 14.5

## Overall Assessment

You did an excellent job overall, including all extra credit Qs!

Neat, clean project organization and code, and practically no 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 L<sup>A</sup>T<sub>E</sub>X and git, you did very, very well! In particular, you seem to be well on your way to becoming comfortable with both Python and R – that’s great!

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:** 90

*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