

# 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:** Jacob Griffiths

## Overall Project workflow

Found all the expected weekly directories in your parent directory.

You had a .gitignore, with meaningful exclusions specific to certain weeks – good. Hard-coding the paths to specific files is not ideal though. You will likely find this useful: <https://www.gitignore.io>. Also see <https://git-scm.com/docs/gitignore> and <https://labs.consol.de/development/git/2017/02/22/gitignore.html>. In my notes I had also shown how to exclude files above certain sizes.

You had an overall readme file (with nothing useful), and then a readme for 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 – good job. However, you had more readmes including in code as well, with directory tree etc. 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. -2.5 pts

Your Git repo size when I checked week 7 was about 4.24 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!

## WEEK 1

Found directories Data, Sandbox, Code

Found 23 code files: installing.sh, ConcatenateTwoFiles.sh, UNIX\_commands.txt, README.txt, CountLines.sh, variables.sh, CompileLaTeX.sh, tiff2png.sh, dir\_structure.txt, csvtospace.sh, FirstExample.tex, MyExampleScript.sh, FirstBiblio.bib, UnixPrac1.txt, tabtocsv.sh, finding\_files.sh, UNIX\_shortcuts.txt, grep.sh, bash\_command\_challenge.sh, redir\_pipes.sh, boilerplate.sh, wild-cards.sh, tr.sh

That's a lot more scripts than I was expecting! Looks like you kept a really detailed record – good.

UnixPrac1.txt was fine. But compare with the solutions, especially the last one. Each solution was described in a comment. You also explained the key components of the unix command, great!

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. Similar comment for other shell scripts.

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

align\_seqs\_better.py and basic\_io.py gave an error: -10pts

lc1.py, lc2.py, dictionary.py, tuple.py were all nicely done. They gave some well-formatted output – great! Also 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 and align\_seqs\_better.py correctly (though latter had a minor syntax error) — so +5 extra credit pts.

All other scripts were fine – great job.

Points for this week: 95

## WEEK 3

Found directories Practicals, Code, Data, Results

Found 33 code files: browse.R, README.txt, PP\_Regress.R, Vectorize2.py, apply1.R, basic\_plot.R, sample.R, run\_get\_TreeHeight.sh, get\_TreeHeight.py, boilerplate.R, TreeHeight.R, PP\_Lattice.R, Vectorize\_speed.sh, next.R, Vectorize1.R, SQLinR.R, break.R, plotLin.R, basic\_io.R, Vectorize1.py, try.R, apply2.R, get\_TreeHeight.R, TAutoCorr.R, Vectorize2.R, DataWrangTidy.R, pre\_allocate.R, DataWrang.R, TAutoCorr.tex, map.R, MyBars.R, control.R, poundhill\_example.R

Found the following extra files: TAutoCorr.aux, TAutoCorr.log, TAutoCorr.synctex.gz, TAutoCorr.fls, TAutoCorr.fdb\_latexmk, .DataWrang.R.swp: -0.5 pt per file

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 OK, but compare with the solution.

The report: Zero interpretation! You could also have plotted the correlation pattern itself. You could also have kept the writeup in a completely separate directory. -5pts

You did the Vectorize and Mapping (good account of the data and potential biases) extra credit – +7.5 pts.

Points for this week: 96.5 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 18 code files: README.txt, TestR.py, regexs.py, LV3.py, profileme2.py, timeitme.py, blackbirds.py, TestR.R, profileme.py, LV1LV2LV3LV4\_run.sh, LV4.py, fmr.R, using\_os.py, LV1.py, DrawFW.py, Num\_computing.py, run\_fmr\_R.py, LV2.py

MyFirstJupyterNb.ipynb missing: -10 pts

timeitme.py, TestR.R, and Num\_computing.py gave an error: -15pts

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

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, including nicely-formatted output .

Points for this week: 78

## Overall Assessment

You did an GOOD job overall, including most extra credit Qs

Reasonably neat project organization and code. Some 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 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!

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: 83**

*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