

CMEE Masters: Computing Coursework Assessment

Note that:

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

In the weekly feedback/assessments, please compare with the solution 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: Abigail Baines

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! You will likely find this useful: <https://github.com/github/gitignore>

You had a readme file with a list of the weeks' content, and then within each week, a readme with a detailed directory/file tree for each week, presumably using the nifty `tree` command. OK, but you may not always want/need to do it in this detail. For example, no need to list files in Sandbox or Data. After all, the Code files would presumably contain sufficient info about what sorts of data are expected as input.

Of equal or more importance is a description of what the overall project structure is and what the language and dependencies requirements are, which would be equally useful for a new user trying to understand and run your workflow. As you become a seasoned programmer, you will learn to make the readme file descriptions more informative yet succinct.

Your Git repo size when I checked week 7 was a modest 8.00 MB — 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, Results Sandbox, Code

Found 11 code files: ConcatenateTwoFiles.sh, UnixPract1.txt, CountLines.sh, variables.sh, CompileLaTeX.sh, csvtospace.sh, FirstExample.tex, MyExampleScript.sh, FirstBiblio.bib, tabtocsv.sh, boilerplate.sh

No unwanted files, great!

UnixPract1.txt was fine - almost identical to my solutions, interesting! Each solution could have been described in a comment, breaking the description down into the key components of the unix command.

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: 100

WEEK 2

Found the Code, Sandbox, Data, Results directories

Found 18 code files: lc2.py, boilerplate.py, basic.csv.py, dictionary.py, debugme.py, scope.py, tuple.py, basic_io.py, lc1.py, oaks.py, My_best.txt, loops.py, using_name.py, cfexercises.py, align_seqs.py, sysargv.py, control_flow.py, test_control_flow.py

Found the following extra file: DNA.csv; -0.5pt

lc1.py, lc2.py, dictionary.py, tuple.py were all fine. The first three could have an better formatted output – Compare with the solution on the repo; -3pts (1 each).

align_seqs.py was nicely done. You could have written it as a self-sufficient script that could also take external inputs. Compare with the solution.

All other scripts were fine; some missing docstrings.

Points for this week: 90

WEEK 3

Found directories Practicals, Code, Data, Results

Found 27 code files: browse.R, PP_Regress.R, apply1.R, var_geoms.R, Pub_ready_ggplot.R, sample.R, run_get_TreeHeight.sh, get_TreeHeight.py, boilerplate.R, case_study3.R, Multi-f-plots.R, TreeHeight.R, PP_Lattice.R, next.R, Bin_width.R, Ricker.R, Vectorize1.R, tapply.R, basic_io.R, Try.R, apply2.R, get_TreeHeight.R, TAutoCorr.R, case_study1.R, Vectorize2.R, TAutoCorr.tex, case_study2.R

The Results directory was missing and remained missing; most of the errors in this week were initially because of this. A blanket -10 pts for this.

Found the following extra files: MyResults.Rout, Makefile, TAutoCorr1.pdf; -1.5 pts

The “extra” Rplots.pdf file is a R quirk — I am not actually deleting points for it, as it only happens when a R script is run externally from unix using **RScript**, and there is no easy solution for this.

Vectorize1.R was fine.

Vectorize2.R was fine, but only printed one of the two speed checks to terminal — -1.5pts .

PP_Regress.R was OK, but could have been more compact — have a look at my solution.

TAutoCorr.R was OK, but messy – compare with the solution for a different approach — does your solution capture every pair of successive years? The report: Nicely done – nice, succinct job. You could have plotted the histogram of the permuted correlation coefficients as well. Some more interpretation would have been nice.

Points for this week: 87 pts

WEEKS 4, 5 & 6

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

WEEK 7

Found directories Code, Data, and Results

Found a README

Found the Code, Sandbox, Data, Results directories

Found 29 code files: running_LV.sh, TestR.py, regexs.py, LV3.py, testing.py, timeitme.py, blackbirds.py, First_RNB.ipynb, TestR.R, profileme.py, LV4.py, fmr.R, First_IPYNB.ipynb, using_os.py, nets.py, LV1.py, sql.py, LV5.py, sql2.py, LotkaVoltera.ipynb, DrawFW.py, Pandas.ipynb, LV3.ipynb, Nets.R, Models.ipynb, run_fmr.R.py, r6_sympy.ipynb, LV2.py, Testing_pandas.ipynb

Found the following extra files ('gitignore these): Pandas-checkpoint.ipynb, LotkaVoltera-checkpoint.ipynb, Testing_pandas-checkpoint.ipynb, Models-checkpoint.ipynb, LV3-checkpoint.ipynb, First_IPYNB-checkpoint.ipynb, First_RNB-checkpoint.ipynb; -3.5pts

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

LV2.py gave an error: -5pts

Nets.py needed the pandas package. Good job, but note that the goal was to avoid using additional packages unless absolutely necessary ;).

blackbirds.py was fine.

You did ALL the LV* scripts, Great! But only some had profiling. Also, running_LV.sh gave an error. +6 Extra credit points.

Otehr scripts file. Some docstrings missing.

Points for this week: 92.5 pts

Overall Assessment

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

Very few errors.

You also went just that extra mile in some cases. You clearly like coding, especially in python!

You joined the course late, and had to really get going. And you did. Well done.

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 L^AT_EX and git, you did very well! In particular, you seem to have become quite comfortable with both Python and R – that’s great!

It was a tough set of weeks, but I hope it gave you an inkling of why and how you would/could use Python, R, UNIX, etc together or as and when required.

Provisional Grade

A*	87
A	
B	
C	
F	

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 Appendix of the SilBioComp document.

We will discuss where 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.

Signed: Samraat Pawar

January 31, 2018