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: Petra Guy

Overall Project workflow

Found all the expected weekly directories in your parent directory.

You had a .gitignore throughout, but could have included many more pattern exclusions; You will likely find this useful: `https://github.com/github/gitignore`

You had a readme file for the overall project with nothing really useful in it. Within each week, you had a readme with a detailed file list, except in week 7. These decriptions were a bit anecdotal, almost lijke a diary. That's fine, but 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 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 reasonable 62.50 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 the way lost parts of your git history — but I won't check these possibilities!

WEEK 1

Found directories Data, Results Sandbox, Code

Found 9 code files: ConcatenateTwoFiles.sh, CompileLatex.sh, UnixPrac1.sh, CountLines.sh, variables.sh, csvtospace.sh, MyExampleScript.sh, tabtocsv.sh, boilerplate.sh

You put FirstExample.tex in `Data`, not sure why. Not a biggie though. -1.5pt

Found the following extra files: FirstExample.pdf; -0.5pt

UnixPrac1.txt was commented out. Each solution could have been described in a comment, breaking the description down into the key components of the unix command. Compare with the solutions. -5pts

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: 93

WEEK 2

Found the Code, Sandbox, Data, Results directories

Found 21 code files: lc2.py, boilerplate.py, basic_csv.py, test_oaks.py, dictionary.py, LV1.py, debugme.py, scope.py, tuple.py, profilename.py, basic_io.py, lc1.py, oaks.py, loops.py, using_name.py, cfexercises.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, test_oaks.py, basic_io.py gave errors because of a missing file. -7.5pts

lc1.py, lc2.py, dictionary.py, tuple.py were all fine. They could have given more nicely-formatted output – compare with the solution on the repo.

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

You did align_seqs_fasta.py — so +2.5 extra credit points.

All other scripts were fine except one with missing docstring(s).

Points for this week: 94

WEEK 3

Found directories Practicals, Code, Data, Results

Found 26 code files: browse.R, PP_Regress.R, Vectorize2.py, apply1.R, CompileLatex.sh, run_get_TreeHeight.sh, get_TreeHeight.py, Control.R, boilerplate.R, TreeHeight.R, PP_Lattice.R, next.R, Sample.R, Vectorize1.R, break.R, basic_io.R, Vectorize1.py, try.R, apply2.R, get_TreeHeight.R, TAutoCorr.R, Vectorize2.R, DataWrangTidy.R, DataWrang.R, PP_Regress.py, GPDD.R

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.

You used an absolute path in CompileLatex.sh which gave an error; -2pts

Control.R, apply2.R, gave a syntax error: -10pts

Vectorize1.R was fine.

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

PP_Regress.R solution was far too lengthy — have a look at my solution(s).

TAutoCorr.R: Looks like you tried to calculate autocorrelation using the `pracma` package. In general, the goal, in the initial weeks in particular, was to avoid using additional packages in R or Python. Compare with the solution for a different, simpler approach. Nevertheless, you clearly spent some thought and time in the analyses, which I appreciate. I was not able to re-generate ay of your plots though, so that's a fundamental problem. -2pts

The report was fine. The interpretation was good, but I was unable to see the plots because of the ...

You had some real problems with mapping extra credit, looks like – +1pt for trying!

Plus you did PP_Regress.py on your own volition, which is not a bad thing! – +1 pt.

You also did Vectorize*.py. Great. Vectorize2.py did not really give a meaningful output to screen though – 4 extra credit points added on for both.

+2.5 extra credit points for DataWrangTidy

Points for this week: 94.5

WEEKS 4, 5 & 6

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

WEEK 7

Found directories Code, Data, and Results

.ipynb_checkpoints could have been .gitignore-d

Found a README, but was really brief

Results directory was missing, hence many of the errors named below.

Found 19 code files: using_os.py, Nets.py, regexs.py, timeitme.py, LV4.py, DrawFW.py, TestR.R, LV1_2.sh, LV3.py, Nets.R, LV1.py, fmr.R, TestR.py, blackbirds.py, run_fmr_R.py, profileme.py, LV5.py, LV2.py, script_to_run_scripts.py

LV1_2.sh (instructions for naming not followed!) gave an error: -5pts

A number of other scripts gave errors because Results directory was missing: -10pts overall

LV3.py gave a IndexError: -5pts

using_os.py gave an indentation error: -5pts Compare with the solution. The code could have provided some meaningful output to screen. -1pt

Nets.py needed the pandas package. Good job, but note that the goal was to avoid using additional packages unless absolutely necessary ;). Same comment about some other scripts where you used pandas.

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

You attempted all the LV* scripts with profiling. very good. There were errors though. Also look at the solution. +5 extra credit points.

Otehr scripts fine, but some docstrings missing.

Points for this week: 78 pts

Overall Assessment

You did an OK job overall, including attempting almost all extra credit Qs. I was happy to see your efforts to understand as many details of the programming languages and programming as possible.

Too many errors, especially some silly ones at the end in week 7. But you went just that extra mile in many cases, especially when python was involved. You apparently like coding in Python (more than R)!

Some of the scripts could have been more compact/short – see the solutions for how you could have found (potentially better) solutions/implementations.

Overall, You delivered on many 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 have taken to Python – 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*	
A	
B	66
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