BIOL 567
Biological data analysis
Fall 2017

Instructor: Scott Sherrill-Mix
Course location: 123 Someplace
Course time: Wed 1-3pm
Course website: `http://somewhere.edu/BIOL657/`

Office: 345 Somewhere
Office hours: By appointment or Wed 3-5pm
Email: shescott@upenn.edu
Phone: (234) 567-7890

Course description

This course aims to prepare students for data analysis in biological research. The primary objective is to enable students to incorporate bioinformatic, statistical and scripting tools into their research. The course is intended for students with little familiarity with programming but a desire to learn. A basic knowledge of cell biology is suggested. Class time will be split between lectures, demonstrations and in-class exercises. I think hands on application is the best way to learn these topics so there will be weekly assignments outside class intended to take 1-3 hours to complete and a final project.

Equipment

Each student will need a laptop (with internet connection and the ability to install free open-source software) for in class activities. Please try to avoid websites/media that are not class related while in the classroom.
A powerful laptop is *not* required. If you do not have access to a laptop, please email me.

Book

No textbook is required for this course. Reading assignments will be detailed in the weekly homeworks and sources of additional information will be given at the end of each lecture.

Objectives

The goal of the course is for students to achieve a comfortable familiarity with:

- Command line and text editors
- Source control
- Scripting and batch processing
- Reproducible analysis and plotting
- Common bioinformatic algorithms and databases

I hope this will enable students to use these techniques in their own research and continue learning after leaving the class.

Grading

Weekly assignments (40%)

There will be weekly assignments in applied data analysis. Some components of these may be used during in-class activities.

Final project (40%)

A final research project with written documentation and code (30%) along with an in class presentation (10%).

Final exam (10%)

This will be a short exam similar to a programming test one might encounter at a job interview.

Class participation (10%)

Participation entails class attendance, participation during in-class activities and mutual respect for other students and teachers.

Late grading

Assignments are due at midnight on the due date. Late assignments with be accepted up to one week after the due date with a 30% grade deduction. Each student will get one free late assignment. Final project write-ups will be accepted up to 3 days late with a cumulative 10% deduction per day.

Plagiarism policy

Students are allowed to work together and discuss assignments but each student is responsible that they independently complete the assignments and completely understand what is turned in. Direct reuse of another person's code in not acceptable in the weekly assignments. The point of the assignments is the journey not the destination. In the final project, code can be reused with correct licensing and documentation (*all included code written by anyone other than the student must be explicitly documented even if not required by the software license*).

Schedule

Week	Topics	Assignments due
1	Introduction to ssh and command line Introduction to text editors Getting help	—
2	Bash scripting Introduction to python Programming basics (variables, data types)	Assignment 1
3	Creating functions Benefits of .txt Source control	Assignment 2
4	Introduction to R Plotting	Assignment 3
5	Flow control and logic Command line tools	Assignment 4
6	Data structures overview Lists and dictionaries	Assignment 5
7	Best practices Reuseable code Reproducible research	Final project description
8	Python modules scipy, multiprocessing, subprocess, biopython	Assignment 6
9	String algorithms Levenshtein distance	Assignment 7
10	BLAST Bio databases	Assignment 8
11	Statistics in R	Assignment 9
12	Final test	—
13	Student presentations	Final report due