

Syllabus

August 17, 2024

1 LFSC 507 / BCMB 424 Programming for Statistical and Graphical Analysis of Biological Data

Time: In-person: Monday and Wednesday, 10:20 am - 11:30 am (Tutorials). The online component is asynchronous.

Location: 306 Mossman

Instructor: Andrew Willems

Instructor e-mail: awillems@vols.utk.edu

Instructor: Tian Hong

Instructor e-mail: hongtian@utk.edu

Office hour: Friday 1pm-3pm

Location: 316 Mossman

Course description: This is a hands-on introductory programming course with the emphasis on analyzing and visualizing biological data. The course will cover basic programming concepts including core data types, functions, flow control, modules, classes and interactions with files. It will also introduce popular libraries for numerical computing, data analysis, machine learning and scientific plotting. Toy examples in areas of structural biology, cell imaging, transcriptomics, genomics and systems biology will be used. Students are encouraged to bring their own data analysis/modeling projects to the course and work on them as a part of the graded component. Upon completing the course, students will be able to use programming tools to perform data analysis, to make publication quality plots, and to learn advanced computational methods with the help of coding. Python will be used as the primary programming language of the course.

Materials and course website: All materials, including announcements, notes, instructions, code and data files, will be available at <https://bioinfo.tk>

Computing device: All assignments can be finished via remote access to the course server, so students may choose to use any UT or personal computers. Depending the selected projects, students may prefer to use their personal laptops. Instructions for software installation will be provided in the course.

Grading: Grade distribution: 20% Participation, 40% Homework assignment and 40% Project.

Attendance is optional for all in-person sessions, but students are expected to engage in discussions (in-person and online) and participate in course surveys. The mid-term and final surveys are each worth 2%.

Assignments are in the format of coding exercises due weekly. Completing these exercises is essential for understanding the subsequent lectures. No late assignments will be accepted.

A data analysis or modeling and simulation project needs to be proposed and finished by each student. The project can be a comprehensive pipeline to analyze complex biological datasets and obtain scientific conclusions. Alternatively, it can be a package to model and simulate a particular biological system. Students are encouraged to use their own datasets (e.g. collected for their own research interests or obtained from their current or previous labs) for the project. The project is evaluated based on a written research plan (10% of the project grade) and a final software package (BCMB 424: 90%, LFSC 507: 50%). For students taking LFSC 507, a written final report is required (40%). The report summarizes the scientific and/or technical contributions of the project to a research field.

The final grade for each student will be on an A-F scale:

A : 90 - 100
A-: 87-89
B+: 84-86
B : 80 - 83
B-: 76-79
C+: 71-75
C : 65 - 70
D : 50 - 64
F : 0 - 49

Office hours Please respect the office hour times given for the instructor. If these times are inconvenient, then make an appointment. We highly recommend that you keep up with assignments and practice problems and come to instructors with questions throughout the semester instead of right before assignments are due.

Academic integrity Honor Statement: *An essential feature of the University of Tennessee, Knoxville, is a commitment to maintaining an atmosphere of intellectual integrity and academic honesty. As a student of the university, I pledge that I will neither knowingly give nor receive any inappropriate assistance in academic work, thus affirming my own personal commitment to honor and integrity.*

Academic honesty practices: Much of this course will be done with the intention to openly help your classmates. Students are expected to take advantage of this group learning approach. However, any blatant duplication of code, figures, plots, etc. will be considered cheating and/or plagiarism and you will receive a zero on the assignment.

Students with Disabilities Policy Any student who feels he or she may need an accommodation based on the impact of a disability should contact the Office of Disability Services (ODS)

at 866-974-6087 in 2227 Dunford Hall to document their eligibility for services. ODS will work with students and faculty to coordinate reasonable accommodations for students with documented disabilities.

Tentative Schedule for in-person sessions

Week	Day	Lesson	Date	Topic	Note
1	M	1	8/19	Introduction	Project: Topic selection starts
1	W	2	8/21	Variables, strings and numbers	
2	M	3	8/26	Lists, tuples and sets	
2	W	4	8/28	Loops and conditional statement	
3	M	-	9/2	—	No classes
3	W	5	9/4	Working with files	
4	M	6	9/9	Functions 1	Project: Data analysis starts
4	W	7	9/11	Functions 2	
5	M	8	9/16	Working with libraries	
5	W	9	9/18	Dictionaries	
6	M	10	9/23	Statistical methods 1	
6	W	11	9/25	Statistical methods 2	
7	M	12	9/27	Statistical methods 3	
7	W	14	10/5	Regular expressions	
8	M	-	10/7	No classes	
8	W	15	10/9	NumPy 1	
9	M	16	10/14	NumPy 2	Project: Research plan due
9	W	17	10/16	Matplotlib 1	
10	M	18	10/21	Matplotlib 2	
10	W	19	10/23	Seaborn	
11	M	20	10/28	Data analysis 1	
11	W	21	10/30	Data analysis 2	
12	M	22	11/4	Data analysis 3	
12	W	23	11/6	Pandas	
13	M	24	11/11	Machine learning 1	
13	W	25	11/13	Machine learning 2	
14	M	26	11/18	Machine learning 3	
14	W	27	11/20	Machine learning 4	
15	M	-	11/25	Working on project	
15	W	-	11/27	No classes	
16	M	-	12/2	Working on project	Project: Report & code due

[]: