Overview

Most subdisciplines in modern biology involve the analysis of large amounts of data, inference based on probabilistic models, or both. This course builds on students’ previous experience in computational biology and extends it into the context of applied, computational analysis of biological data. In particular, the analyses will be performed for biologist clients, who will provide real data from on-going projects that require computational statistical analysis. The clients will lay out the scientific context and motivation, and be the subdisciplinary expert in biology. Teams of students will work together, with consultation and direction from the instructor, to answer the scientific questions and perform the analyses for the client. The research products will be delivered, evaluated and refined at intermediate stages, through written and in-person presentations with the client. Likewise, final research products will be delivered to the client and will utilize best practices in report generation and reproducible research. The delivery of analyses to clients is a valuable experience for data science careers in academics, government or business.

Computational biology can open a world of possibilities for research and employment. The computational and analysis skills we will practice are increasingly imperative for modern biologists (e.g., see the mention of computation, statistics and quantitative analysis in many job advertisements). Without them the scope and ambition for biological research are unnecessarily constrained. In addition, the problems and computational approaches we will study are good application domains for students from outside biology who are interested in computational and data science.

We will have a few main goals in this class: 1) to gain expertise and become proficient in the application of probability, mathematics and computational methods in biology, 2) to understand philosophies and conceptual frameworks underlying analytical methods, and 3) to develop skills in teamwork and reporting of scientific analysis.

Assessment and class structure

Assessment of your work in this course will be done regularly by you. You will do independent work (e.g., reading, exercises, problem solving) and determine through interactions with the class what you did correctly, where you need further study or have questions, and so forth. It will be your responsibility to gauge your competency and understanding of these exercises. I will help motivate you by offering challenging and engaging course material, but your success will depend heavily on your personal motivation.

Your performance will be evaluated formally based on three writing assignments (initial description of analysis and questions, interim report on findings, and final report of analysis). Assessment will be of the group’s written work, as well as of individual knowledge
and contributions to the project through instructor interviews with each student. The interim and final reports will be accompanied by computer code and instructions by which the analysis could be repeated, which will also be graded. Additionally, your interactions and presentations to the client will be evaluated through peer and instructor assessments. Grading will be on a standard scale (i.e., 90–100=A, 80–89=B, etc.).

\[
\begin{align*}
\text{Writing assignments (group)} & \quad 20 + 60 + 100 = 180 \\
\text{Instructor interviews} & \quad 25 + 35 = 60 \\
\text{Client presentations (group)} & \quad 2 \times 15 = 30 \\
\text{Client and team interactions (individual)} & \quad 30 \\
\end{align*}
\]

Early in the semester, biologists will visit the class to describe a set of scientific questions, data to address the questions, and the challenge of analyzing this set of data. Teams of students will work together to address the questions and implement an analysis. The first deliverable product of analysis will be a write-up of the scope of science and analysis that synthesizes what has been presented by the scientific client (due in week 5 of the semester). This assignment will be assessed for its clarity, the extent to which the scientific questions and analysis tasks are clearly defined, and questions that have been identified by the analysis team.

The teams will then work toward performing analyses and obtaining answers to the scientific questions. In week 11 of the semester, teams will deliver a written interim report of findings, make a presentation to the client, and gather feedback on the analysis. Again, the written work of the group (written report, along with code and instructions for repeating the analysis) and its in-person presentation will be assessed. Additionally, individual students will meet with the instructor for an assessment interview, to answer questions about the interim report to the client and the analysis problem.

A final report of the analysis will be delivered and presented in the final week of the semester, with assessment of group and individual work as for the interim report.

Throughout the semester students will make short presentations on analyses to other students and to the clients. These will be assessed for accuracy and effectiveness. Students will work in teams on different projects, so class time will include instances in which we break into groups for separate discussions of projects. But the class will typically work as a whole, so that students will be exposed to all projects and analyses that the teams are pursuing, and there will be occasional exchanges of students among groups for sharing of knowledge and skills.

Note-taking and questions
You will want to take notes on readings, make outlines of important concepts, and make notes on computer code and concepts. In class meetings I will present lectures, answer questions, offer further explanations of material and facilitate discussions. The course’s WyoCourses website and Github repository are additional resources for asking questions and discussing topics outside of class, and we will use them in various ways to interact outside of class.
Computer use and etiquette

We are using a room full of computers for this class. Occasionally, a student has difficulty focusing on the class material and not being distracted by the many distractions that a networked computer offers. These distractions are typically detrimental to the attention of the individual user, but also to the students who sit nearby. Therefore I ask that you find a solution to manage and minimize these distractions for yourself. For example, I ask that you do not consult the web while I or anyone else is presenting material to the class, because it will interfere with your ability to follow the presenter. Likewise, there is no reason to use personal email or other messaging software during class. Of course, legitimate uses of these resources during class exist, particularly during those times when I ask you to use the computers. These include transferring a copy of a file to yourself or another student via email, consulting the web for computational science resources related to class work, taking notes, and trying out code as I present it on screen. Please use your best judgment and minimize distractions to yourself and others.

Additional Items

- The schedule of topics, assignments, and all other details in this syllabus are subject to change with fair warning, including announcements in class or via university email addresses and the course website.
- Any student who has a disability and is in need of classroom accommodations please contact the instructor and the University Disability Support Services.
- Students whose religious activities conflict with the class schedule should contact the instructor at the beginning of the semester to make alternative arrangements.
- Cheating and other forms of academic dishonesty are listed in University Regulation 802, Revision 2. If you are found to be engaged in academic misconduct, at a minimum you will receive no credit for that exam or assignment. Repeat or serious offenders can expect more serious consequences.

Course schedule

1. Weeks 1–4
   Meetings with scientific clients
   Software tools and best practices for reproducible research
   Tools for collaborative coding and writing
   Introduction to UNIX and cluster computing
   Development of initial document with scope and nature of analysis

2. Weeks 5–10
   Learn statistical and computational tools for analysis for projects
   Perform analyses and share knowledge among students
   Development and delivery of interim report on analysis

3. Weeks 11–15
   Learn statistical and computational tools for analysis for projects
   Perform analyses and share knowledge among students
   Development and delivery of final report on analysis