

Practical Analysis of Biological Data in R - BISC 444

Syllabus - 2024 Fall Semester

1. Basic Information

Course: Introduction to Bioinformatics, BISC 444, 2 credits
Textbook: "An Introduction to R" (<http://cran.r-project.org/doc/manuals/R-intro.pdf>)
Prerequisites: none
Place and time: Tuesdays and Thursdays: Two sections:
(either 9:00 am – 9:50 am or 10:00 am – 10:50 am)
Location: RRI 301

Faculty: Dr. Matthew Dean
Associate Professor, Molecular and Computational Biology
Office: 304A Ray I. Irani Building.
Telephone: 213-740-5513
Email: matthew.dean@usc.edu
Office Hours: Thursday 11:00 am – 11:50 am or by appointment

Teaching Assistant: Maeve Secor
Office:
Telephone:
Email: msecor@usc.edu
Office Hours:

2. Classroom policy

Students must bring laptops to class. Whether you use Windows, Mac, or other (Linux, Unix, etc.) does not matter, but laptops are critical because lectures include hands-on programming.

Any other electronic communication devices (phones, blackberries, and similar) must be turned off, and no instant messenger/chat type programs are allowed in class.

3. Course goals and learning objectives

The main goal of Introduction to Bioinformatics is to teach students how to use R and to gain knowledge of statistical approaches for the analysis of biological data. Students will learn how to use R, an open-source statistical programming environment that is widely used in biology. Our philosophy in this class is to learn R in a hands-on way, through tutorials and weekly homeworks that challenge the student to break down problems into manageable units. Students will apply their R skills to address a bioinformatic question of their own construction and present their results as a poster to the rest of the class.

Students, especially graduate students, are encouraged to bring their own data sets to analyze and to ask a question that is specific to their thesis. Students are encouraged to consult with the instructor to devise their own project.

There will be a slight emphasis on two broad topics. The first is genomics, including metagenomic data, genome sequencing, RNA sequencing, etc. (Please note however, that our class concerns the analysis of genomics data downstream of raw data processing. In other words, we do not cover upstream topics like “quality control” and “mapping reads”). The second is on clustering methods such as Principal Components Analysis and related methods that are useful in identifying and summarizing patterns in complex data.

In this class, bioinformatics refers to any computational approaches that are incorporated into the analysis of biological data. The ability to write code is a critical aspect of success, regardless of field of interest or type of data.

The only pre-requisite for this course is scientific curiosity. Students are not expected to know anything about bioinformatics, coding, or statistics. This class is not meant to teach advanced algorithmic design or statistics (such classes already exist at USC), though there are many themes that overlap with those fields. The emphasis in this course is on practical implementation, not on computational aesthetics.

Please note: This course is extremely challenging. Although it is meant for the beginner with no prior experience in coding, we cover a lot of material, including statistical approaches as well as programming in R.

4. Weekly topics

Week	Dates	Tuesday	Thursday
1	8/27-8/29	Basic R syntax	Basic R syntax (cont.)
2	9/3-9/5	"for" loops, "if" statements	Plotting in R
3	9/10-9/12	read / write data	Basic statistics in R
4	9/17-9/19	Basic statistics in R (cont.)	Statistical philosophy
5	9/24-9/26	Linear Models (lm, glm)	Nonlinear Models (nlms)
6	10/1-10/3	Manipulating strings	Class Project: stop codon usage
7	10/8-10/10	Midterm	Fall Break
8	10/15-10/17	Clustering data (PCA)	Clustering data (LDA)
9	10/22-10/24	Class Project: microbiome study	Randomization
10	10/29-10/31	Randomization (cont.)	Randomization (cont.)
11	11/5-11/7	Class Project: indel study	Writing/modifying functions
12	11/12-11/14	Bioconductor	Bioconductor (cont.)
13	11/19-11/21	Evolutionary statistics	Class Project: evolution of molting
14	11/26-11/28	Poster presentations	Thanksgiving
15	12/3-12/5	Poster presentations	Poster presentations

Final exam: Thursday, December 12, 11 am – 1pm (needs to be confirmed)

5. Professor

Dr. Matthew Dean
matthew.dean@usc.edu
213-740-5513
304A Ray R. Irani Building
1050 Childs Way
University of Southern California
Los Angeles, CA 90089

Dr. Dean maintains an active research program focused on evolutionary biology, genomics, and reproduction. Bioinformatics represents an integral part of his research.

6. Required material

- Textbook: "An Introduction to R" (<http://cran.r-project.org/doc/manuals/R-intro.pdf>)
- Additional online materials will be specified throughout the course
- Laptop computer (if you do not have one, we can provide one for you)

7. Assessment

Grades are based on four scores: 1) midterm exam grade, 2) final exam grade, 3) weekly homework assignments where students solve bioinformatic challenges by writing code, 4) final projects (documented code and poster). Please note: you will not have to pay for poster production, USC prints them for you.

Assessment Procedure	Percent
Midterm exam	25%
Final exam	25%
Weekly homeworks	25%
Final project (documented code and poster presentation)	25%

8.1. Criteria for grading: The final will be an open book test that consists of both written questions and answers as well as computer programming problems. Bioinformatics code will be graded according to proper annotation of code and ability to solve the problem of interest. The final presentation will be graded according to clarity of scientific hypothesis, appropriateness of data to address that hypothesis, ability of the student to effectively communicate their bioinformatic strategy, and on the substance of their conclusions.

Students who are not able to meet deadlines due to medical or other emergency must notify the instructor immediately. Any medical excuses must be accompanied by a letter from a doctor that explicitly states the student was unable to attend exam or

presentation.

8.2. *Course grade:* The course is not curved. Letter grades will follow a straight scale: 90% and above leading to A, 80-90% leading to B, etc. Pluses and minuses are assigned by dividing each range in corresponding halves (A, A-) or thirds (B+, B, B-, C+, ...).

8. Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Support Systems:

Student Health Counseling Services - (213) 740-7711 – 24/7 on call
engemannshc.usc.edu/counseling

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 – 24/7 on call
engemannshc.usc.edu/rsvp

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) | Title IX - (213) 740-5086
equity.usc.edu, titleix.usc.edu

Information about how to get help or help a survivor of harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations.

Bias Assessment Response and Support - (213) 740-2421
studentaffairs.usc.edu/bias-assessment-response-support

Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.

The Office of Disability Services and Programs - (213) 740-0776
dsp.usc.edu

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support and Advocacy - (213) 821-4710
studentaffairs.usc.edu/ssa

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu

Non-emergency assistance or information.