

Practical Analysis of Biological Data in R - BISC 444

Syllabus - 2018 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
Class materials: Available through Dropbox
Place and time: Mondays: 10:00 am - 12:15 pm
Location: To be announced (but probably RRI 421)

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 10:00 am – 11:00 am or by appointment

Teaching Assistant: Sara Keeble
Office: RRI 302
Telephone:
Email: keeble@usc.edu
Office Hours: Tuesdays 3-4 pm

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. The class is divided into two main parts. In the first part, we 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. In the second part of the course, students will apply their R skills to address a bioinformatic question of their own construction. 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 without their own data will be given important bioinformatic questions by the instructor.

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. This class is not meant to teach advanced algorithmic design or statistics (such classes already exist in our department), 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, we cover a lot of material, including statistical approaches as well as programming in R.

4. Course plan and weekly readings

To maximize the benefit of attending class, you must read the selected pages listed below before coming to class.

Week	Date	Topic
1	August 20	Intro to R usage
2	August 27	Base R functions
3	September 3	Labor Day
3	September 10	Reading in and analyzing large datasets in R
4	September 17	Plotting options
5	September 24	Commonly used statistical tests
6	October 1	Linear models
7	October 8	Time series and smoothing
8	October 15	Midterm exam
9	October 22	Randomization: Permutation, Bootstrapping, Jackknifing, and Basic Simulations
10	October 29	Revisiting lm() but with randomization
11	November 5	Writing your own functions
12	November 12	Final presentations
13	November 19	Final presentations
14	November 26	Final Presentations
	December 10	Final exam (8-10am)

Weeks 1-11: These initial weeks will be spent learning R from the ground up, as well as basic statistical approaches, in a hands-on way. After 11 weeks, students will be fluent in R. We will then apply our newly gained knowledge to address a specific scientific question. Students (especially graduate students) are encouraged to bring their own data and their specific question to class for this purpose. Otherwise, genomics level problems will be assigned to them.

Weeks 12-14: The last three weeks of the course will be dedicated to student presentations, where students go from hypothesis, to data analysis, to conclusions using computational approaches.

6. 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 these endeavors.

7. 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)

8. 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; 10-pp, double-spaced report; and 20- to 30-minute presentation).

Assessment Procedure	Percent
Midterm exam	25%
Final exam	25%
Weekly homeworks	25%
Final project	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.

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+, ...).

9. Statement on Academic Conduct and Support Systems

All USC students are responsible for reading and following the Student Conduct Code, which appears in the SCampus and at <https://scampus.usc.edu/university-student-conduct-code/>. This policy does not apply to discussion or exchange of ideas. On the contrary, such interactions represent an important way to clear programming hurdles.

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 Section 11, Behavior Violating University Standards <https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct/>.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* <http://equity.usc.edu/> or to the *Department of Public Safety* <http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us>. This is important for the safety whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* <http://www.usc.edu/student-affairs/cwm/> provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems

A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* <http://dornsife.usc.edu/ali>, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* <http://emergency.usc.edu/> will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.

10. Resources

10.1. Web page: A class website will be setup on Backboard containing information about the course: syllabus, laboratory handouts, grades, miscellaneous information about weekly class

activities, and an email directory of all people in the class. Use it as much as you find it useful. The web page can be accessed through the main stem <https://Blackboard.usc.edu>.

10.2 Office Hours: Office hours will be held weekly. Time and location for my office hours are at the beginning of the syllabus. Those of the unofficial teaching assistant will be decided with you in class. Both of us are available by email to help you as much as you need.

During weeks 1-11, every student will meet with me at least once outside of class so that progress on projects can be assessed and any obstacles encountered solved.