Reproducible Research (Biol 4800/7800)

Location:

Time:

Instructor: Dr. Tad Dallas (tadallas@lsu.edu)

Office: A343 Life Sciences

Office hours: T and Th from 2:30 - 4:20pm

Course Overview:

Scientific knowledge builds upon existing scientific knowledge. Much of this knowledge has a firm basis in empirical or simulated data and resulting analyses of those data. However, much of these analyses are no longer able to confirmed due to a degradation of the analytical tools used (e.g., fortran punchcards). This is not ideal, as we are forced to take the findings of previous researchers at face value. Recognizing that programming languages and data storage formats are not permanent, it is the goal of scientists to not only make their code and data freely available, but to ensure that analyses can be run regardless of operating system, elapsed time, or analytic dependencies. This course will provide you the tools to ensure reproducible research.

In doing so, we will learn some algorithmic basics, the R programming language, version control, and a bit of the unix command line and associated tools. Get stoked.

Course Goals:

Over the course, it is expected that students gain

- a knowledge of the issues surrounding reproducibility and openness of scientific research
- an ability to design and implement reproducible scientific workflows
- experience with algorithmic thinking
- experience using the unix command line, R, and other computational tools
- an appreciation and/or deep hatred for reproducible research

Syllabus Subject to Change:

Changes to the syllabus may be made during the semester. The most up-to-date and current syllabus will always be available on the course website (Moodle).

Grading

There will be a total of 500 points, consisting of 6 assignments, some attendance/participation points, and a final project. The final project will be performed in small groups for undergraduate students (4800), and independently for graduate students (7800).

Assignments:

There will be 6 assignments throughout the semester. *Each assignment will be worth 50 points.* The assignments will be submitted through Github, a version control platform.

Attendance:

Much of the material presented will not be available if you aren't in class to hear it. Given the pace of the course, and the computational and programming hurdles you will almost certainly encounter, you should probably come to class. Attendance and participation will be worth 50 points, so there's some incentive as well.

Project:

The project . . .

The project is worth 150 points

Academic honesty

Louisiana State University adopted the Commitment to Community in 1995 to set forth guidelines for student behavior both inside and outside of the classroom. The Commitment to Community charges students to maintain high standards of academic and personal integrity. All students are expected to read and be familiar with the LSU Code of Student Conduct and Commitment to Community, found online at www.lsu.edu/saa. It is your responsibility as a student at LSU to know and understand the academic standards for our community.

Students who are suspected of violating the Code of Conduct will be referred to the office of Student Advocacy & Accountability. For undergraduate students, a first academic violation could result in a zero grade on the assignment or failing the class and disciplinary probation until graduation. For a second academic violation, the result could be suspension from LSU. For graduate students, suspension is the appropriate outcome for the first offense.

Further information is provided on the LSU website

Special needs statement

My goal is to help you learn. Students who have any difficulty (either permanent or temporary) that might affect their ability to perform in class can contact me privately, or reach out to the LSU Disability

Services staff.

More information on registering a disability is available at LSU Disability Services, located at 124 Johnston Hall. Contact the Center by telephone at 225-578-5919 or via email at `disability@lsu.edu`.

Late Assignments:

Assignments are expected to be turned in on time. I will not accept late assignments. This is motivation to commit partial work, as I will judge the work based on the last commit made before the deadline.

Schedule

Week	Topic	Assignment
1	Background	–
2	Intro to tools (git, Github, unix command line)	`gitIntro`
3	Analytical pipeline structure (make, Travis)	–
4	R basics (including Rmd)	`rIntro`
5	R data manipulation	`rDataManip`
6	R functions, unit testing, and line profiling	`rTesting`
7	R, APIs, and data visualization	`rAndAPI`
8	Dealing with spatial data	`rSpatial`
9	Dependency hell and containerization	–
10	Parallel computing	–
11	Copyright garbage	–
12	Final project presentations	–

Final projects

Info.