

Syllabus

(#course-overview)

Welcome to BIOL4800 section 3, *Introduction to Data Science*. We are living in a rapidly changing world, with the current pandemic as an immediate example among many other complex issues (e.g., climate warming, biodiversity loss). Meanwhile, increasing large volumes of data have become available during the past decades because of advances in technology, measurement approaches, establishment of citizen science programs, etc. This increasing wealth of data provides opportunities to solve complex problems that were unsolvable previously, fueling the rise of the new field of data science. Data science is interdisciplinary and usually combines elements of computer programming, statistics, graphic design, communication, and domain knowledge. A data scientist normally needs to identify relevant questions, gather appropriate data from different sources, clean, organize, and manage datasets, extract useful information to generate answers to the questions, and communicate results with others.

This course will provide an introduction to data science by covering data science tools (e.g., command line, Git, etc.), basic to intermediate R programming, data management, data manipulation, data visualization, basic statistical models, and reproducible workflows to generate reports. This course will use R as the main programming language. However, the concepts and skills learned in this course will be easily transferable to other programming languages.

(#learning-objectives)

Over the course, students should be able to:

- Use **Git** and **GitHub** to create repositories, manage projects, track changes of files, recover old versions of files, push/pull changes to/from remote repositories, discuss through issues, and submit pull requests
- Be able to use **R** to get, clean, explore, visualize data, and conduct simple statistic analyses
- Organize the above processes into a *project* or an R package
- Use unit tests in **R** to assure code functionality
- Use **R Markdown** to combine text, code, and results into reports, slides, or even websites (i.e., literature programming
([https://en.wikipedia.org/wiki/Literate_programming#:~:text=Literate%20programming%20is%20a%20programming,source%20code%20can%20be%20generated.\)\)](https://en.wikipedia.org/wiki/Literate_programming#:~:text=Literate%20programming%20is%20a%20programming,source%20code%20can%20be%20generated.))))
- Googling

(#instructor)

Daijiang Li (<https://dlilab.com>)

Assistant Profesoor

Pronouns: He/Him/His

Department of Biological Sciences

Center for Computation & Technology

LSU

Email address: dli30@lsu.edu (<mailto:dli30@lsu.edu>)

(#meeting-times--locations)۱۵

Tu/Th 1:30-2:50 PM

Class location: 225 Tureaud Hall

(<https://www.google.com/maps/place/Tureaud+Hall/@30.4094801,-91.1791112,15z/data=!4m5!3m4!1s0x0:0x92922201c8e77288!8m2!3d30.4094801!4d-91.1791112>)

Zoom link: <https://lsu.zoom.us/j/96140507560>

(<https://lsu.zoom.us/j/96140507560>)

Office room: LSB125

Office hours: Tuesday 12-1pm in LSB 125 or by appointment

(#pre-requisites)

No pre-requisites; ideally, you should have taken the BIOL 4800 FOUND COMP BIOLOGIST course.


(#textbooks-andor-readings)gs

There is no required text for the course. Online readings will be provided for each learning topic.

(#syllabus-subject-to-change)

This syllabus represents my current plans and objectives. As the semester progresses, those plans may need to change to enhance the classes learning opportunity. Such changes, communicated clearly, are not unusual (especially during a pandemic) and should be expected. Syllabus and grades will be available on Moodle. But the most up-to-date and current syllabus will always be available on the course web page (/).

(#technology)


This course will have lots of hands-on computing exercises during and after lectures. Therefore, students are strongly recommended to bring their own laptop to class. For students without access to a personal laptop, it is possible to check out one  from the library (<https://lib.lsu.edu/services/borrowing/gear>) (<https://lib.lsu.edu/services/borrowing/gear>)).

All softwares used in this course are free and cross platforms (macOS, Linux, or Windows operating systems). Main softwares that we will use are: Git (Git) (for version control), R (<https://www.r-project.org>) and use RStudio (<https://www.rstudio.com/products/rstudio/download/>) as an integrated development environment (IDE).

(#course-communications)

Important notes about updates to the syllabus, clarifications of assignments, or changes to the schedule will be communicated via LSU email. Please be sure to check your account regularly. If you have an email that you would like to send to the entire class, please communicate with the course instructor before doing so. Short questions should be addressed to me via email *only* after searching online and did not find an answer. I will try to respond to your email within 48 hours. More complicated questions should be addressed to me in person either after class or during office hours or appointment scheduled. In addition, we will use the **Issues** feature of **GitHub** to post questions and help each other.

(#classroom-conduct)

I will strive to create a learning environment for students that supports a diversity of thoughts, perspectives, and experiences, and honors your identities (including race, gender, class, sexuality, religion, nationality, ability, etc.). I welcome and appreciate  any constructive criticisms, suggestions, ideas, comments, and any other feedbacks for the course.

I encourage all of us to use welcoming and inclusive language, show respect toward others, and acknowledge our differences. As a student, you should also strive to respect and honor the diversity of your classmates. If something was said/done in class (by anyone) that made you feel uncomfortable, please talk to me about it.

Examples of behavior that contributes to a positive environment for our community include:

- Demonstrating empathy and kindness toward other people
- Being respectful of differing opinions, viewpoints, and experiences
- Giving and gracefully accepting constructive feedback
- Accepting responsibility and apologizing to those affected by our mistakes, and learning from the experience
- Focusing on what is best not just for us as individuals, but for the overall community

Examples of unacceptable behavior include:

- The use of sexualized language or imagery, and sexual attention or advances of any kind
- Trolling, insulting or derogatory comments, and personal or political attacks
- Public or private harassment
- Publishing others' private information, such as a physical or email address, without their explicit permission
- Other conduct which could reasonably be considered inappropriate in a professional setting

Note: If you believe you have been a victim of an alleged violation of the Code of Student Conduct (<https://www.lsu.edu/saa/students/codeofconduct.php>), or you are aware of such an alleged violation, you have the right to report it to the University (<https://www.lsu.edu/saa/students/codeofconduct.php>).

(#expectations)

LSU's general policy states that for each credit hour, you (the student) should plan to spend at least two hours working on course related activities outside of class. Since this course is for three credit hours, you should expect to spend a minimum of six hours outside of class each week working on assignments for this course.

(#evaluation)

Students will be evaluated on their knowledge of course content and ability to communicate their results via individual homework assignments (60%), their attendance and active participation in class (15%), and a final project (25%).

Homework (60%)

There will be 7 homework assignments, with 10 points for each assignment. All assignments should be finished as *Rmarkdown* files and I need to be able to reproduce all results by running it on my own computer. I expect that each student will complete the assignment independently (discussing about some questions is fine). The assignments are due by next Thursday before class. *No late assignment will be accepted*, for any reason. So don't wait until the last hour to do an assignment. Plan ahead and pace yourself. Instead of late assignment, I will drop your lowest score and use the other 6 scores for your final grade. *At least 5 assignments must be turned in to achieve a passing grade.*

Participation (15%)

A key component of data science is communicating your results with others. It is thus critical to attend class and participate in discussing of concepts, results, etc. Another important reason to come to class is that it is easier for us to overcome the computational and programming burdles that you likely will encounter. Attendance and participation will be worth 15 points.

Final project (25%)

The final project will demonstrate your knowledge of the course content and beyond. It will be a full report in an *R Markdown* format, mixed text with code, results, figures, tables, citations, etc. The final report should be in high quality and publishable with some additional effort later. The final project should be hosted on Github, organized as we discussed in class and should be fully reproducible. This means that I can clone your repository and re-run your files on my computer and get the same results. **The final project deadline is December 3rd.** More information about it can be found here ([/](#)).

Grading

LSU has a plus-minus grading scale. There will be no curve.

grade	percent
A+	> 97
A	93-97
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D	65-69
F	< 64

(#academic-integrity)

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 <https://www.lsu.edu/saa/> (<https://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 (<https://www.lsu.edu/online/faculty/academicintegrity.php>)

(#special-needs-statement)

Our 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 us privately, or reach out to the LSU Disability Services staff.

More information on registering a disability is available at LSU Disability Services (<https://www.lsu.edu/disability/>), located at 124 Johnston Hall. Contact the Center by telephone at 225-578-5919 or via email at disability@lsu.edu (<mailto:disability@lsu.edu>).