

GGG 771: Data Science for Genetics and Genomics

Syllabus

Catalog description

This graduate course is part of a first-year training program for Genetics & Genomics Scholars and provides a broad understanding how to apply principles of data science to large multi-faceted datasets that are central to modern-day genetics and genomics. The students will focus on the application of these principles in the analysis of genetics and genomic data. Students will develop basic skills for reproducible research, including project organization, version control and test-evaluate-diagnose development. While exploring the universe of genetic and genomics analysis packages, the students will focus on the R data-science platform. They will develop their skills in common genetics and genomics analyses, including RNA-seq differential expression and population genetics statistics. The final product for the course will be a collaborative, small groups project consisting of a defined analysis of a genetic or genomic dataset of their choice

Course description

Large multi-faceted datasets are central to modern-day genetics and genomics. In this course we will learn how to apply principles of data science to the analysis of genetics and genomic data. We will develop basic skills for reproducible research, including project organization, version control and test-evaluate-diagnose development. We will explore the universe of genetic and genomics analysis packages, with a focus on the R data-science platform. We will develop our skills in common genetics and genomics analyses, including RNA-seq differential expression and population genetics statistics.

Learning outcomes

- **Identify** key choices in a data science workflow, and **evaluate** how those choices affect analysis accuracy and reproducibility.
- **Demonstrate** basic proficiency with data science skills including use of the command line (bash), version control (git/Github), and R (Rstudio).
- **Identify** appropriate software packages for a genetics/genomics analysis, **execute** their installation and usage, and **articulate** problems with selected software.
- **Design** and **develop** a computational analysis of a genetics/genomics dataset.
- **Manage** and **negotiate** collaborative computational analyses.

Native Land Acknowledgement

I want to acknowledge that we are on the traditional territory of the Tuscarora, Coharie and Haliwa-Saponi here in Raleigh.

Inclusion Statement

We strive to create a learning environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ability, etc.). To help accomplish this:

1. If you have a name and/or a set of pronouns that differ from what appears on official University records, please let us know in person, via email, or on the personal information survey that you filled out in the fall semester.
2. If you feel like your performance in the class is impacted by experiences outside of the class, please let us know in person, via email, or in the personal information survey that you filled out before class. We want to be a resource for you. You can also make an anonymous statement that we can share with the class.

We are still learning about diverse perspectives and identities, if something was said in class that made you feel uncomfortable please let us know. Anonymous feedback is always welcome and should be emailed to the graduate student coordinator Jenni Wilson

(jenni_wilson@ncsu.edu) or the Director of the GG Scholars program Martha Burford Reiskind (mbreiski@ncsu.edu).

Course Structure

The course is composed of lectures, computational laboratories and computational homework assignments. A foundation of basic and key data science concepts will be used to explore several topics in genetics and genomics research. Each topic will be considered from the perspective of ensuring accuracy and reproducibility of the analysis, the ability to share and communicate the analysis with appropriate level of detail, and the mathematical/computational techniques used to analyze the data.

Students will be primarily evaluated on their contribution to class-wide visible resources. These will include a class-wide "wiki" on how to install and use bioinformatic software programs to do specific genetics/genomics tasks, and a project modifying published analyses workflows to test alternative genetics/genomics hypotheses. In class sessions will alternate between lecture and laboratory-based environments. In lab sessions students will work on defined problems in small groups, with the instructors as a resource.

Course Schedule

Class will meet two times a week for 1.5 hours: **TBD**.

Office hours **TBD**.

Class location **TBD** depending on university status will either be on Zoom or in a computer lab.

Course Materials

There is no required textbook purchase for this course. Readings and course notes will be provided in class. Suggested books (all freely available online): [R for Data Science](#), and [Modern Statistics for Modern Biology](#)

Course Requirements

Graduate standing is required; there are no prerequisites or co-requisites.

Instructors

Benjamin Callahan, bcallah@ncsu.edu, (919) 515-8536

Rafael Guerrero, rfquerre@ncsu.edu, (919) 515-5329

Grading

Grading components

Attendance: 20 points (10%)

Attendance to all class lectures and laboratory sessions is expected. Attendance at each session is worth 1 point and will be accounted for by a **minute paper** answering the three questions (1) in one sentence what did you learn today?, (2) what are you still confused about?, and (3) what would you like to know more about?. In cases where students are unable to attend the Zoom lecture, and listen to the recording asynchronously, they will turn in their **minute papers** online. Please see the Expectations & Policies section of the syllabus for more information.

Homework: 100 points (50%)

Homework will consist of data-science and analysis tasks related to the course materials. Homework will be submitted through GitHub. This will include 4 homework assignments, each worth 25 points.

Final Project: 80 points (40%)

Small groups will perform a final project consisting of a defined analysis of a genetic or genomic dataset of their choice. The points for the final project are broken down to 20 points for suitability of the analytic design to the research question, 20 points for the accurate execution of the analysis, 20 points for the level of reproducibility and comprehension of the analysis, and 20 points for the final presentation of the results.

Letter Grades

The course uses Standard NCSU Letter Grading:

$97 \leq \mathbf{A+} \leq 100$	$93 \leq \mathbf{A} < 97$	$90 \leq \mathbf{A-} < 93$
$87 \leq \mathbf{B+} < 90$	$83 \leq \mathbf{B} < 87$	$80 \leq \mathbf{B-} < 83$
$77 \leq \mathbf{C+} < 80$	$73 \leq \mathbf{C} < 77$	$70 \leq \mathbf{C-} < 73$
$67 \leq \mathbf{D+} < 70$	$63 \leq \mathbf{D} < 67$	$60 \leq \mathbf{D-} < 63$
$0 \leq \mathbf{F} < 60$		

Detailed Schedule (Tentative, Spring 2020)

Week	Subject
0	Pre-class survey of student computational backgrounds.
1	File systems. The command line. Installation.
2	Intro to R.
3	Version control. Git/GitHub.
4	Data project organization. Tabular data/spreadsheets. Homework 1 (version control) due.
5	Documentation and reproducible analysis.
6	Additional topics in base R. Homework 2 (documentation) due.
7	Visualization in R and ggplot2.
8	RNA-seq data and analysis. Homework 3 (data manipulation and visualization) due.
9	RNA-seq data and analysis.
10	Clustering and dimension reduction. Homework 4 (sequence alignment) due. Final project topics determined.

11	Clustering and dimension reduction.
12	Population genetics.
13	Population genetics.
14	Final projects: results presentations

Expectations & Policies

Attendance

Please read the Attendance Regulations (REG 02.20.03) found at <https://policies.ncsu.edu/regulation/reg-02-20-03>.

- You are required to be in class and on time. Part of your grade comes from your participation during class; therefore you will lose participation points for frequent absences. All students are given one free pass for an unexcused absence. After missing one class, students will lose 1 point per unexcused absence.
- If you have to be absent on the day of a presentation or handing in an assignment you have to provide appropriate documentation for your absence.
- You will not be able to makeup participation activities we conduct during class. Only activities that become take-home assignments can be made up, if you have appropriate documentation for your absence.
- If you discontinue class attendance without following proper procedures for dropping or withdrawing, you will receive a grade of F in the course and your last date of attendance will be documented in your final grade.

Assignments

- You are responsible for obtaining lecture material from the course web site and for readings as listed on the schedule of topics in this syllabus.

Integrity

- All course work submitted for a grade in both lecture and field part must be your own. University standards of academic integrity forbid either giving or receiving unauthorized help on graded work. Violations of University standards will be prosecuted. You will need to sign the academic integrity statement on each written assignment.

Please read the **Code of Student Conduct** (POL 11.35.01) found at <http://policies.ncsu.edu/policy/pol-11-35-01> and go to the **Office of Student Conduct** at: <http://studentconduct.ncsu.edu/>

- Be a team player and be considerate of others in class by following simple rules of politeness.
- Turn off completely all electronic devices (iPods, cell phones, laptops) during lecture, unless you are using them to read lecture notes, or take notes.
- This course is for you—you will get as much out of it as you're willing to put in.

Email Etiquette

Make sure to include your full name in the body of all emails you send to us. In the subject heading, **please write: GGI_DataScience**. Please only write GGI_DataScience and nothing else. If you are attaching a document, **include your name in the document and use your name and course number for the file name**. Also, include your full name in the attached document. If you're replying to an email, please include any previous exchanges in the email reply. All emails should begin with a salutation and close with your name. If you have not received a reply, make sure you have the correct subject.

Statement on Disabilities: If you have a disability that will affect your performance in this course, reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with the Disability Resource Office at Holmes Hall, Suite 304, 2751 Cates Avenue, Campus Box 7509, 919-515-7653. For more information on NC State's policy on working with students with disabilities, please see the **[Academic Accommodations for Students with Disabilities Regulation \(REG02.20.01\)](#)**. Also, please be sure to discuss any issues with the course coordinator.

Non-discrimination Policy: NC State University provides equality of opportunity in education and employment for all students and employees. Accordingly, NC State affirms its commitment to maintain a work environment for all employees and an academic environment for all students that is free from all forms of discrimination. Discrimination based on race, color, religion, creed, sex, national origin, age, disability, veteran status, or sexual orientation is a violation of state and federal law and/or NC State University policy and will not be tolerated. Harassment of any person (either in the form of quid pro quo or creation of a hostile environment) based on race, color, religion, creed, sex, national origin, age, disability, veteran status, or sexual orientation also is a violation of state and federal law and/or NC State University policy and will not be tolerated. Retaliation against any person who complains about discrimination is also prohibited. NC State's policies and regulations covering discrimination, harassment, and retaliation (POL 04.25.05) may be accessed at <http://policies.ncsu.edu/policy/pol-04-25-05> and <http://oied.ncsu.edu/oied/policies.php>. Any person who feels that he or she has been the subject of prohibited discrimination, harassment, or retaliation should contact the Office for Equal Opportunity (OEO) at 919-515-9617.

Health and Participation in Class

We are most concerned about your health and the health of your classmates and instructors.

- If you test positive for COVID-19, or are told by a healthcare provider that you are presumed positive for the virus, please work with your instructor on health accommodations and follow other university guidelines, including self-reporting ([Coronavirus Self Reporting](#)): Self-reporting is not only to help provide support to you, but also to assist in contact tracing for containing the spread of the virus.
- If you feel unwell, even if you have not been knowingly exposed to COVID-19, please do not come to class.
- If you are in quarantine, have been notified that you may have been exposed to COVID-19, or have a personal or family situation related to COVID-19 that prevents you from

attending this course in person (or synchronously), please connect with your instructor to discuss the situation and make alternative plans, as necessary.

- If you need to make a request for an academic consideration related to COVID-19, such as a discussion about possible options for remote learning, please talk with your instructor for the appropriate process to make a COVID-19 request (a university-level form can be found [here](#)).

Health and Well-Being Resources

These are difficult times, and academic and personal stress are natural results. Everyone is encouraged to [take care of themselves](#) and their peers. If you need additional support, there are many resources on campus to help you:

- Counseling Center ([NCSU Counseling Center](#))
- Health Center ([Health Services | Student](#))
- If the personal behavior of a classmate concerns or worries you, either for the classmate's well-being or yours, we encourage you to report this behavior to the NC State CARES team: ([Share a Concern](#)).
- If you or someone you know are experiencing food, housing or financial insecurity, please see the Pack Essentials Program ([Pack Essentials](#)).

Community Standards related to COVID-19

We are all responsible for protecting ourselves and our community. Please see the [community standards](#) (released on 7/28/2020) and Rule 04.21.01 regarding Personal Safety Requirements Related to COVID-19 [RUL 04.21.01 – Personal Safety Requirements Related to COVID-19 – Policies, Regulations & Rules](#).

Course Expectations Related to COVID-19:

- **Face Coverings:** All members of the NC State academic community are required to follow all university guidelines for personal safety with face coverings, physical distancing, and sanitation. Face coverings are required in class and in all NC State buildings. Face coverings should be worn to cover the nose and mouth and be close fitting to the face with minimal gaps on the sides. In addition, students are responsible for keeping their course/work area clean. Please follow the cleaning guidelines described by the university.
- **Course Attendance:** NC State attendance policies can be found at: [REG 02.20.03 – Attendance Regulations – Policies, Regulations & Rules](#) . Please refer to the course's attendance, absence, and deadline policies for additional details. If you are quarantined or otherwise need to miss class because you have been advised that you may have been exposed to COVID-19, you should not be penalized regarding attendance or class participation. However, you will be expected to develop a plan to keep up with your coursework during any such absences. If you become ill with COVID-19, you should follow the steps outlined in the health and participation section above. COVID 19-related absences will be considered excused; documentation need only involve communication with your instructor.

- **Course Meeting Schedule:** Your course might not have a traditional meeting schedule in Spring 2021. Be sure to pay attention to any updates to the course schedule as the information in this syllabus may have changed. Please discuss any questions you have with the instructor.
- **Classroom Seating:** To support efficient, effective contact tracing, please sit in the same seat when possible and take note of who is sitting around you; instructors may also assign seats for this purpose.
- **Technology Requirements:** This course may require particular technologies to complete coursework. Be sure to review the syllabus for these expectations, and see the [syllabus technology requirements](#) for your course. If you need access to additional technological support, please contact the Libraries' Technology Lending Service: ([Technology Lending](#)).

Course Delivery Changes Related to COVID-19

Please be aware that the situation regarding COVID-19 is frequently changing, and the delivery mode of this course may need to change accordingly, including from in-person to online. Regardless of the delivery method, we will strive to provide a high-quality learning experience.

Grading/Scheduling Changing Options Related to COVID-19

If the delivery mode has a negative impact on your academic performance in this course, the university has provided tools to potentially reduce the impact:

- **Enhanced S/U Grading Option:** [Enhanced Satisfactory/ Unsatisfactory Grading Option](#)
- **Late Drop:** [Enhanced Late Drop Option](#)

In some cases, another option may be to request an incomplete in the course. Before using any of these tools, discuss the options with your instructor and your academic advisor. Be aware that if you use the enhanced S/U, you will still need to complete the course and receive at least a C- to pass the course.

Other Important Resources

- **Keep Learning:** [Keep Learning](#)
- **Protect the Pack FAQs:** [Frequently Asked Questions | Protect the Pack](#)
- **NC State Protect the Pack Resources for Students:** [Resources for Students | Protect the Pack](#)
- **NC State Keep Learning, tips for students opting to take courses remotely:** [Keep Learning Tips for Remote Learning](#)
- **Introduction to Zoom for students:** <https://youtu.be/5LbPzzPbYEW>
- **Learning with Moodle, a student's guide to using Moodle:** <https://moodle-projects.wolfware.ncsu.edu/course/view.php?id=226>
- **NC State Libraries** [Technology Lending Program](#)