**Cell 6351 - Intro to Reproducible Data Analysis Using R**

**Summer 2025**

**1-2 Credit Hours**

**Course Location: TBD**

**Time: Tuesday, Thursday 9:00 AM – 10:20 AM**

**Course Date Range: 7/1/2025-7/28/2025**

*This syllabus is intended as a guide for this course. Dates, assignments, and evaluation are subject to revision by the instructor and will be announced in advance when possible.*

**COURSE FACULTY**

Course Director: Lindsay Hayes Teaching Assistant: Eleana Cabello

Title: Assistant Professor E-mail: Eleana-Cabello@ouhsc.edu

E-mail: Lindsay-Hayes@ouhsc.edu

Office: BMSB 103

Office Hours: Mondays 1-2pm

Telephone: (405) 271-2377 x 33708

**WELCOME FROM THE COURSE DIRECTOR / TEACHING PHILOSOPHY**

Current biomedical research regularly generates large datasets that require increasing computational skills to process and analyze including in genomics, behavior, imaging, and molecular biology. As a result, scientists today require fundamental computational and programming skillsets to efficiently perform data analyses, generate complex models for testing, develop reproducible data analyses, and effectively communicate results through clear and meaningful graphics. However, these skillsets are often challenging due to a lack of structured training in data science.

***R for Data Analysis*** is a hands-on course that will introduce the R statistical software including data structures, wrangling data, plotting data, basic statistics, data modeling, and literate coding practices. There is no textbook for the course. All course materials will be provided on Canvas. While no formal background in statistics or programming is necessary, a general knowledge of basic statistics, experimental design, and computer file storage systems will be helpful. All classes will involve active learning/coding activities so a laptop with power cord is required for all sessions.

**COURSE DESCRIPTION**

***Reproducible Data Analysis Using R*** prepares students to explore large datasets, generate publication-quality graphics, perform reproducible and literate data analysis pipelines, and communicate analysis results. The goal is to teach novice programmers to write modular and descriptive code using R to address relevant scientific questions.

**PREREQUISITE(S)**

Permission of instructor

**CLASS TYPE**

Lecture/Discussion

**COMPUTER REQUIREMENTS**

A laptop is required for the course. The laptop requirements set forth by the College of Medicine can be found [here](https://itsupport.ou.edu/TDClient/30/Unified/KB/ArticleDet?ID=2333). Specifically, laptops must be encrypted in accordance with University policy. It is the student’s responsibility to ensure that one’s laptop is connected to the OUHSC wireless network.

**In addition, students need to have R and R Studio installed on their computers before the first class session. Detailed information on this can be found on the course Canvas site.**

**COURSE OBJECTIVES**

* Students will be able to **import** and **summarize** **data** in R using reproducible and literate programming principles.
* Students will be able to **analyze and evaluate data** manipulations and outcomes.
* Students will be able to **seek help** for roadblocks, **debug** problems, and **communicate** **results** to a broad audience.
* Students will be able to take appropriate steps to **evaluate data quality** and choose appropriate analyzes.

**LEARNING OUTCOMES**

The goal of this course is to teach novice programmers to write modular code and impart best practices for using R for data analysis. R is commonly used across many scientific disciplines and will be useful to a wide array of students across campus. The course will focus on teaching the fundamentals of the programming language R including data analysis and plotting but will not teach statistical analyses.

**CLASS PREPARATION**

Before each class, there will be a set of pre-lecture activities and/or readings to complete. It is important these activities be done before class so that the students can benefit the most from the in-class activities. Students are responsible for completing all assigned readings and materials, whether discussed in class or not (including any written or verbal updates, all lecture material, case studies, independent study, other information provided, etc.). Students are responsible for processing and analyzing data outside of the instruction time, so the student and the data are ready for in-class activities.

**PROFESSIONALISM**

Students and faculty will treat each other with respect by 1) showing up on time, 2) being prepared to participate in class, and 3) demonstrating collegiality and civility in all discussions.

**ATTENDANCE**

Attendance is required for all class sessions. All anticipated absences should be requested in writing at least 24 hours in advance. A limited number of reasons will justify excused absences. These include, but are not limited to, an illness, a personal emergency, a family emergency, or presentation or representation in a leadership capacity at a professional meeting. Students may be required to provide documentation of the reason for the absence and may be required to complete supplementary assignments to make up for missed activities, but the course director and/or instructors are not required to provide make-up opportunities for missed discussions. Each unexcused absence will trigger a 10 point reduction in attendance points (see evaluation and grading below). If a session cannot be in person due to illness or weather, the session may be given by zoom. In such cases, every effort will be made to communicate any changes and to record the session for those that cannot attend synchronously.

*Unexcused lateness of more than 10 minutes will be recorded as an absence.*

*Egregious lack of engagement (e.g., work unrelated to the course, messaging/texting, or using social media for more than 10 minutes of class) will be counted as an absence.*

**PARTICIPATION**

Students are required to process and analyze data provided during the course. Students are required to write their own scripts and analyze data independently. Students will be required to submit a final project processing their own data or publicly available data using literate programming concepts.

**CONDUCT OF THE COURSE**

**Instruction Mode**

**Traditional: Face-To-Face Contact:** Instructors interact with students in the same physical space 75% or more of the time.

**Required Texts and Materials**

Required readings will be posted online before the relevant course sessions.

**Assignments (if applicable)**Homework assignments and final projects will be completed independently by students.

**Evaluation Methods**In addition to attendance and active participation in all classes, students are expected to complete three homework assignments and a final presentation. Homework reports will use data and analysis pipelines provided in class. The final presentation will use data collected by the student from their research lab (with mentor permission) or a publicly available dataset.

**Grading Rubrics**

**Rubric for homework reports**

|  |  |
| --- | --- |
|  | **Did the student successfully complete the analysis and script? Was the correct output generated?** |
| 5 pts (Good) | Student understood the assignment, completed the script, and generated the correct output. |
| 3 pts (Fair) | The student attempted the assignment but was unable to complete the script and generated partially correct output. |
| 1 pt (Poor) | The student did not understand the assignment, turned in scripts that were missing the majority of the components and failed to generate correct outputs. |

**Rubric for the final project**

|  |  |  |
| --- | --- | --- |
|  | Did the final project have a clearly stated analysis goal? | Did the final project successfully achieve that goal? |
| 5 pts (Good) | Student understood the assignment, picked a task that could be completed by data science, and clearly communicated the problem and goals. | The student was able to write a script to address their problem and present the problem, logic and organization of the script, and achieved the stated goals. |
| 3 pts (Fair) | The student attempted the assignment, identified a problem, and was able to partially communicate the goals of the project. | The student was able to communicate the problem, attempted a script to achieve the goal, but was only partially able to achieve the stated goals. |
| 1 pt (Poor) | The student did not understand the assignment, was not able to communicate a problem or identify achievable goals. | The student poorly communicated the problem, did not find a logical solution, and was not able to achieve the stated goals. |

**Rubric For Attendance**

|  |  |
| --- | --- |
| -10 points | Each unexcused absence (defined as per in the attendance section above) |
|  | Up to a maximum of 40 points. |

**Student Evaluation of the Course (cannot be included in course grade)**

We will allocate 15 minutes in the last course session for students to perform the standard course evaluation, along with comments for the instructor.

**Use of large language models (LLMs) such as ChatGPT and Copilot.** The instructor supports the use of LLM, such as ChatGPT and Copilot to help with troubleshooting and to expand the student’s grasp of course materials. However, the instructor mostly values a deep student understanding and accuracy of course materials. Below are guidelines for using LLMs for a statistical programming course.

1. Use LLM to guide conceptual understanding

Encourage students to use LLMs as a **study aid** rather than a solution engine. They can ask for clarifications on concepts, request examples, or generate explanations of statistical programming topics. However, they should be required to cross-check and verify answers using lecture materials or external reliable sources.

1. Prohibited direct copy-paste for assignments

Students are **prohibited from copy and pasting** LLM-generated responses directly into their assignments. If using LLM, students must 1) engage with the answers critically, 2) reformulate or rewrite solutions in their own words, incorporating personal insights, and 3) include citations when ChatGPT is used as an informational resource.

1. Verification Requirement

The instructor has concern that sometimes LLMs return **inaccurate or nonsensical answers**. Therefore, students should validate any solution or explanation provided by LLMs to ensure the results make sense. For example, cross-reference with textbooks or peer-reviewed materials, explain the solution based on principles taught in class, or acknowledge discrepancies provided by LLMs and provide insight into why the incorrect advice was provided.

1. Encourage debugging skills

The best use of LLMs is to assist in debugging code and explaining errors. However, students should **document their process** when seeking help. For example, provide the original code, show how LLMs suggested changes, and analyze or explain why the solution worked or not.

1. Transparency in LLM usage

Students should be **transparent** about using LLMs in assignments, indicating where and how they used it similar to citing references. If students turn to extensive use of LLMs for assignments, they could be required to provide extensive explanations for solutions and a reflection on the limitations of LLMs.

**EVALUATION AND GRADING**

Course grading will be S/U. A total of more than 75 points will be required for a grade of “S”.

**Grading Scale**

|  |  |
| --- | --- |
| **Activity** | **Points/ % of Final Grade** |
| Homework 1 | 15 |
| Homework 2 | 15 |
| Homework 3 | 15 |
| Final Presentation | 35 |
| Attendance/Participation | 20 |
| **Total Points** | 100 |

**COURSE OUTLINE/SCHEDULE**

Dates, topics, and assignments are subject to change. You will be notified if a change occurs. Each session will be 1.5 hours.

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Class** | **Topic** | **Homework** |
| 7/01 | 01 | Introduction to R/RStudio |  |
| 7/03 | 02 | Navigating Data Types in R | Homework 1 (due 7/25) |
| 7/08 | 03 | Data Wrangling with tidyverse |  |
| 7/10 | 04 | Data Visualization with ggplot2 | Homework 2 (due 7/25) |
| 7/14 | 05 | Dimension Reduction with tidymodels | Homework 3 (due 7/25) |
| 7/16 | 06 | Machine Learning with tidymodels |  |
| 7/18 | 07 | Building a Portfolio Using GitHub |  |
| 7/22 | 08 | In-Class work on final with instructor Q&A | Final project proposal due |
| 7/24 | 09 | Final Project Presentations | 5 - 10 min presentation |
| 7/28 | 10 | Final Project Presentations | 5 - 10 min presentation  (with time for student course evaluation) |

**Required Policy Statements**

**This syllabus is intended as a guide for this course. Dates, assignments, and evaluation are subject to revision by the instructor. Any such revisions will be announced in advance.**

**Copyright.** This syllabus and all related course material are protected under US Copyright Law and may not be further disseminated in any form or format without the prior explicit written consent of the faculty member. Failure to comply with this provision may subject the student to disciplinary action and/or state or federal action.

**Student Professional Behavior in an Academic Program.** Ethical and professional behaviors are considered a core competency in an academic program and thus are key factors in a student’s good academic standing. Upon acceptance of an offer of admission, the student commits to comply with all professional conduct regulations established by the University, respective college, and program. The complete Student Professional Behavior in an Academic Program policy is at [www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf](http://www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf).

**Academic Misconduct Code.** The Academic Misconduct Code describes academic misconduct as any acts intended to improperly affect the evaluation of a student’s academic performance or achievement. Academic Misconduct includes but is not limited to cheating, plagiarism, fabrication, fraud, destruction, bribery or intimidation, assisting others in any act proscribed by the Code, or attempting to engage in such acts. The policy and procedures related to academic misconduct are detailed in the Academic Misconduct Code found in Appendix C of the Faculty Handbook at [www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf](http://www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf).

**Academic Appeals.** The Academic Appeals policy outlines the procedure a student must follow to request a hearing for appeals related to evaluation in a course, thesis or dissertation defense, or general or comprehensive exam. It also outlines the appeal process for a suspension or dismissal or under the Student Professional Behavior in an Academic Program Policy and for appeals of decisions resulting in dismissal, expulsion, or suspension from a program or of being required to repeat a semester of year. The sole basis for an academic appeal is an alleged prejudiced or capricious academic evaluation or decision. Policy and procedure details are in Appendix C of the Faculty Handbook at [www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf](http://www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf).

**Accommodation on the Basis of Disability.** The University of Oklahoma is committed to the goal of achieving equal educational opportunity and full participation for students with disabilities. Accommodations on the basis of disability are available by contacting the Accessibility and Disability Resource Center (ADRC) by email at [adrc@ou.edu](mailto:adrc@ou.edu) or by calling (405) 325-3852 or Voice (405) 217-3494 (VP). Information on policies and registration with the Accessibility and Disability Resource Center may be found on the ADRC website at: [www.ou.edu/adrc](http://www.ou.edu/adrc). Students requesting accommodations related to work in a course must contact the ADRC as soon as possible; accommodations are not made retroactively.

**Sexual Misconduct.** For issues regarding gender-based discrimination, sexual harassment, sexual misconduct, stalking, or intimate partner violence, the University offers a variety of resources, including Advocates-On-Call 24/7, counseling services, mutual “No Contact orders,” scheduling adjustments, and disciplinary sanctions against the perpetrator. Information is available from the Institutional Equity Office at (405) 325-2215 (8AM-5PM) or the OU Advocates at (405) 615-0013 (24/7).

**Adjustment for Pregnancy/Childbirth Related Issues.** Students needing modifications or adjustments to course requirements because of documented pregnancy-related or childbirth-related issues should contact the college’s Assistant/Associate Dean for Student Affairs (or academic advisor) or the Accessibility and Disability Resources Center ([www.ou.edu/adrc](file:///C:\Users\lklimkow\AppData\Local\Microsoft\Windows\INetCache\Content.Outlook\W59K6WNN\www.ou.edu\adrc)) as soon as possible to discuss. Generally, modifications will be made where medically necessary and similar in scope to accommodations based on temporary disability. See [www.ou.edu/content/eoo/faqs/pregnancy-faqs.html](http://www.ou.edu/content/eoo/faqs/pregnancy-faqs.html) for answers to commonly asked questions.

**Course Drop/University Withdrawal.** The student is responsible to submit required University paperwork before the deadlines to drop or withdraw from a course, shown in the Academic Calendar at <http://admissions.ouhsc.edu/AcademicCalendar.aspx>. Missed homework and examination grades will be entered as a grade of zero if a student fails to formally drop the course or withdraw from the University.

**Laptop/Device Encryption and Anti-Virus Software.** In advance of examinations, students must check that their laptop or PC includes up-to-date encryption software and the necessary programs for securing the device. Students who obtain new or replacement devices at any time can request access to the Student Virtual Desktop at <http://ouhsc.edu/sde>. OU IT will grant student access within 24 hours of request submission. Students can log in at <https://mydesk.ou.edu>. All students should continue to encrypt their devices with Windows 10 and MacOS encryption tools and install anti-virus software. Instructions and recommendations are linked at: [Windows 10 Encryption](https://itsupport.ou.edu/TDClient/35/Norman/KB/ArticleDet?ID=72&SIDs=161), [MacOS Encryption](https://support.apple.com/en-us/HT204837), and [Anti-Virus Software](https://it.ouhsc.edu/services/desktopmgmnt/antivirussoftware.asp).

**Absences**. If you will be absent from a course activity for any reason, it is your responsibility to notify the instructor as specified by the course syllabus.

**COVID-19 Related Academic Policy Addendum (Attached).**

**Responsible Conduct of Research.** Students, as members of the University community, have the responsibility to ensure the integrity and ethical standards of any research activity with which they are associated directly or of which they have sufficient knowledge to determine its appropriateness. Students are governed by the Policy on Ethics in Research (Faculty Handbook Section 3.25) at [www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf](http://www.ouhsc.edu/provost/documents/FacultyHandbookOUHSC.pdf).