Introduction to Data Science
STAT 2600: Spring 2021

Time & Location Mondays, Wednesdays, and Fridays 1:50-2:40 PM Hybrid - FLMG 155

Professor Dr. Osita Onyejekwe Osita.Onyejekwe@colorado.edu ECOT 226
Office hours: Thursday 1:30-3:30 PM **(or by appointment (may change)**
https://cuboulder.zoom.us/j/91806588648

TAs Constance Delannoy (Office Hours)
Recitation: Tuesdays: 1:50 – 2:40 PM
4:10 – 5:00 PM
OH: Tuesdays: 10:00 AM – NOON
https://cuboulder.zoom.us/j/93885854529
Email: Constance.Delannoy@Colorado.edu

Course Websites We will use Canvas for grading, readings, and assignments.

Readings Our textbook is *R for Data Science* by Wickham and Grolemund available at *r4ds.had.co.nz*.

Materials Bring to class a laptop or tablet computer capable of running R.

Course Description This course introduces importing, tidying, exploring, visualizing, summarizing, and modeling data and then communicating the results of these analyses to answer relevant questions and make decisions. Students will learn how to program in R using reproducible workflows.

Overall Learning Objective To develop technical and professional skills necessary to analyze data as a member of a team. This includes:

- Understanding fundamental statistical concepts
- Visualizing and exploring data
- Importing and tidying datasets
- Programming effectively in R
- Building basic statistical models
- Collaborating with teammates to discover and communicate interesting findings and recommendations based on data.
- Mastering reproducible statistical workflows.

In other words, to learn R to do interesting and useful things with data.

Assignments

Weekly assignments will take the form of a single Jupyter Notebook text file: namely, code snippets integrated with captions and other narrative. Except where otherwise noted, assignments are typically due on Wednesdays at Midnight MST on the dates indicated on Canvas. Homework usually is assigned on Fridays.

Your assignment score for the course will be calculated by averaging your **four (4) highest** homework scores. That is, your lowest homework score will not count toward your grade.

Each homework assignment will have **5 problems**, each of which may have several parts. Your score for each assignment will be assigned according to the scheme outlined in the rubric below.

Homework Rubric

Correctness: Each homework will have **5 problems**, which will often have multiple parts. Each of the **5 problems** will be worth **2 points**. Deductions will be made at the discretion of the TA.

Grade

Your grade in this course will be determined by a series of **5** homework assignments (**40%**), lab submission (**10%**), quizzes (**10%**) and a final project (**40%**).

Style: Coding style is very important. With the exception of Homework 1, you will receive a deduction of up to 1 point if you do not adhere to good coding style.

- No deduction if your homework is submitted with:
  - good, consistent coding style
  - appropriate use of variables
  - appropriate use of functions
  - good commenting
  - good choice of variable names
  - appropriate use of inline code chunks
- -0.5 if coding style is acceptable but fails on a couple of the criteria above.
- -1 if coding style is overall poor and fails to adhere to many of the above criteria.

Participation

Lab Activities: The Lab session is scheduled for Fridays. During the lab sessions, students will get hands-on practice with the week's material by working on assigned lab activities. I will be available over Zoom to introduce the activities and to answer any questions you may have. Tasks may include but are not limited to: running or modifying code from the lecture, pair coding, or completing short coding exercises. Note that even when this course shall be hybrid, I will still be teaching over zoom in the classroom and therefore you must still log into zoom in class. It will be run in a synchronous manner so as to keep the format consistent with students that are overseas.

All scheduled lectures will have an associated lab component. **Please note that you are required to have your zoom video turned on during lecture.** Your Lab participation score for the course will be calculated based on the number of labs that you submit, as indicated in the table below. There shall be a total of 13 labs this semester.

0-4	*5-7*	*9-10*	*11-13*
0	**5**	**7.5**	**10**

Quizzes

There will be **4 short quizzes** scheduled during the later weeks of class. Dates and times will be announced in advance. The purpose of these quizzes is to assess your understanding of various concepts that are central to the class. Your score on the quizzes will count for 10% of your final grade.

Final Project

The final project for the class will ask you to explore a broad policy question using a large publicly available dataset. This project is intended to provide students with the complete experience of going from a study question and a rich data set to a full statistical report. Students will be expected to (a) explore the data to identify important variables; (b) perform statistical analyses to address the policy question; (c) produce tabular and graphical summaries to support their findings; and (d) write a report describing their methodological approach, findings, and limitations thereof (Presented During Final Exam).

While students may work in small groups to decide on appropriate statistical methodology and graphical/tabular summaries, each student will be required to produce and submit their own code and final report.

Regardless of grading basis, students must receive a score of at least 50% on the final project to pass the class.

Course Grading

Your final course grade will be calculated according to the following breakdown:

Assignments	**40%**
Participation	**10%**
Quizzes	**10%**
Final Project	**40%**

Late Submission

Homework is to be submitted by **11:59 PM MST on Wednesdays** on the due date indicated, unless an alternate due date is announced. Late homework will not be accepted for credit.

Note that your lowest homework score will not count toward your grade, so you can miss one homework without it counting toward your course grade.

Collaboration/ Plagiarism

You are encouraged to discuss homework problems with your fellow students. However, the work you submit must be your own. You must acknowledge in your submission any help received on your assignments. That is, you must include a comment in your homework submission that clearly states the name of the student, book, or online reference from which you received assistance.

Submissions that fail to properly acknowledge help from other students or non-class sources will receive no credit. Copied work will receive no credit. Any and all violations will be reported to the College of Arts and Science.

All students are expected to comply with the CU Boulder student honor code policy. This policy can be found online at: https://www.colorado.edu/policies/student-honor-code-policy.

The course collaboration policy allows you to discuss the problems with other students, but requires that you complete the work on your own. Every line of text and line of code that you submit must be written by you personally. You may not refer to another student's code, or a "common set of code" while writing your own code. You may, of course, copy/modify lines of code that you saw in lecture or lab.

It's paramount to make sure that the assistance you receive consists of general advice that does not cross the boundary into using code or answers written by someone else. It is fine to discuss ideas and strategies, but you should be careful to write your programs on your own.

You must not share actual program code with other students. In particular, you should not ask anyone to give you a copy of their code or, conversely, give your code to another student who asks you for it; nor should you post your solutions on the web, in public repositories, or any other publicly accessible place. [You may not work out a full communal solution on a whiteboard/blackboard/paper and then transcribe the communal code for your submission.] Similarly, you should not discuss your algorithmic strategies to such an extent that you and your collaborators end up turning in [essentially] the same code. Discuss ideas together, but do the coding on your own.

Modifying code or other artifacts does not make it your own. In many cases, students take deliberate measures -- rewriting comments, changing variable names, and so forth -- to disguise the fact that their work is copied from someone else. It is still not your work. Despite such cosmetic changes, similarities between student solutions are easy to detect. Programming style is highly idiosyncratic, and the chance that two submissions would be the same except for changes of the sort made easy by a text editor is vanishingly small. In addition to solutions from previous years or from other students, you may come across helpful code on the Internet or from other sources outside the class. Modifying it does not make it yours.

I allow exceptions in certain obvious instances. For example, you might be assigned to work with a project team. In that case, developing a solution as a team is expected. The instructor might also give you starter code, or permit use of local libraries. Anything which the instructor

explicitly gives you doesn't normally need to be cited. Likewise, help you receive from course staff doesn't need to be cited.

If you have any questions about any of the course policies, please don't hesitate to ask.

Policies

Classroom Behavior

Both students and faculty are responsible for maintaining an appropriate learning environment in all instructional settings, whether in person, remote or online. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. For more information, see the policies on classroom behavior and the Student Code of Conduct.

Requirements for COVID-19

As a matter of public health and safety due to the pandemic, all members of the CU Boulder community and all visitors to campus must follow university, department and building requirements, and public health orders in place to reduce the risk of spreading infectious disease. Required safety measures at CU Boulder relevant to the classroom setting include:

- maintain 6-foot distancing when possible,
- wear a face covering in public indoor spaces and outdoors while on campus consistent with state and county health orders,
- clean local work area,
- practice hand hygiene,
- follow public health orders, and
- if sick and you live off campus, do not come onto campus (unless instructed by a CU Healthcare professional), or if you live on-campus, please alert CU Boulder Medical Services.

Students who fail to adhere to these requirements will be asked to leave class, and students who do not leave class when asked or who refuse to comply with these requirements will be referred to Student Conduct and Conflict Resolution. For more information, see the policies on COVID-19 Health and Safety and classroom behavior and the Student Code of Conduct. If you require accommodation because a disability prevents you from fulfilling these safety measures, please see the "Accommodation for Disabilities" statement on this syllabus.

All students who are new to campus must complete the COVID-19 Student Health and Expectations Course. Before coming to campus each day, all students are required to complete the Buff Pass. {***In this class, you may be reminded of the responsibility to complete the Buff Pass and given time during class to complete it.***}

Students who have tested positive for COVID-19, have symptoms of COVID-19, or have had close contact with someone who has tested positive for or had symptoms of COVID-19 must stay home. In this class, if you are sick or quarantined. Email me and let me know that you are ill. **Because of FERPA student privacy laws, do not state to me the nature of your illness. Just let me know that you are ill. Do not send me a "doctor's notes" for classes missed.**

Accommodation for Disabilities

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website. Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition, see Temporary Medical Conditions on the Disability Services website.

Preferred Student Names and Pronouns

CU Boulder recognizes that students' legal information doesn't always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructors' class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name.

Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code (honor@colorado.edu); 303-492-5550). Students found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found at the Honor Code Office website.

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

The University of Colorado Boulder (CU Boulder) is committed to fostering an inclusive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct (harassment, exploitation, and assault), intimate partner violence (dating or domestic violence), stalking, or protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, anonymous reporting, and the campus resources can be found on the OIEC website.

Please know that faculty and graduate instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, dating and domestic violence, stalking, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to reasonably and fairly deal with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, please send me e-mail or visit me in office hours to notify me of such a situation at least two weeks in advance of the event.

See the campus policy regarding religious observances for full details.

Final Examination Policy

Purpose

This policy enforces a uniform final examination schedule in order to facilitate fairness and help maximize the educational experience for students.

Policy

It is the policy of the University of Colorado Boulder to adhere to the final examination schedule as published by the Office of the Registrar each semester. Unless notified otherwise in writing during the first week of classes, students enrolled in undergraduate courses should assume that an examination will be given. While it may be appropriate not to give a final in some undergraduate courses, such as independent studies, laboratory courses, seminars, project-based courses and colloquia, final examinations are integral parts of the instructional program and should be given in all other undergraduate courses. Graduating seniors are not exempt from final

examinations. Students enrolled in graduate courses should consult with their instructors on whether final examinations will be administered.

Procedures

1. The final examination in a course should be given as scheduled by the Office of the Registrar and not at other times, even if the instructor and all students in a course agree to such a change. An instructor may allow individual students to take the final examination at an earlier or later time if the instructor is satisfied that an exception is based on good and sufficient reasons and if such an exception is unlikely to materially advantage or disadvantage the interests of other students in the course.

2. Students should consult with their instructors and course syllabi for final examination information for courses that are scheduled at non-standard times and are not included in the published final examination schedule.

3. Final examinations in summer courses and in courses offered during special (variable length) sessions in fall and spring semesters are typically administered on the last regular meeting day of the course. No Common Exams for multiple sections of a single course may be scheduled for special session courses.

4. The week of classes preceding the scheduled final examination period should be used primarily for continued instruction and may include the introduction of new material. For courses in sessions of ten weeks or longer, no examinations may be given during the week of classes preceding the start of the campus's final examination period; however, assignments listed in the syllabus such as papers, lab practicums, presentations, portfolios and projects may be due during that week.

5. Class sessions or graded assignments of any kind, including papers, lab practicums, presentations, portfolios and projects, may not take place or be due on a day designated in the academic calendar as a Reading Day.

6. When students have three or more final examinations on the same day, they are entitled to arrange an alternative examination time for the last exam or exams scheduled on that day. When students have two final examinations scheduled to meet at the same time, they are entitled to arrange an alternative examination time for the course that meets later in the week during the term or, if the two courses meet on the same day during the term, the course that meets later in the day. Students must make arrangements with the instructor of the affected course(s) by the standard

deadline to drop a course in that term and are expected to provide supporting written information of these situations to qualify for exceptions.

7. The submission deadline for grades each semester or special session is 96 hours after the conclusion of the final examination, excluding designated university holidays.

8. The Provost (or designee), in consultation with the Senior Vice Chancellor (or designee), may either reschedule or cancel final examinations in response to inclement weather or other emergencies that result in a campus closure during the final examination period. The administration's determination of whether to reschedule or to cancel final examinations is based on the number of exam periods affected by the closure and the timing of the closure.

    - Rescheduled exams: A final examination may be rescheduled within the final examination period. Students who do not participate in a rescheduled final examination are not guaranteed any make-up examination or alternative assignment, and in such cases students will be assigned course grades based on tests, assignments and other graded work completed up to the end of the term.
    - Canceled exams: If a final examination is canceled, course grades will be assigned based on tests, assignments and other graded work completed up to the end of the term. Faculty may also offer a make-up examination and/or allow for an alternative assignment to be submitted after the final examination period. However, such opportunities must be made available to all students enrolled in the course, all final work must be received and graded no later than the end of the first week of the subsequent term (inclusive of summer session), and faculty must then submit change of record information for students whose course grades change due to their work on make-up examinations or alternative assignments.