GT 4801-C Syllabus

Open Data with R, 1 credit hour

Instructor Information

Instructor	Email	Office Hours & Location
Jay Forrest	jay.forrest@library.gatech.edu	By appointment: Jay's Calendly

General Information

Description

Open Data with R is an introduction to data analysis with the R statistical programming language using the principles of Open Science as a framework. This class will cover importing, transforming, visualizing, modeling, and communicating data with R software in an open manner.

Pre- or Co-Requisites

CS 1301 or CS 1315 or CS 1371 with a minimum grade of C.

Course Goals and Learning Outcomes

The goal of the course is to familiarize students with data analysis using R. Upon successful completion of Open Data with R, you will have a foundation to:

1. Advance comprehension about Open Science principles including: data sharing, open data tools, and open reproducible research.
2. Identify how to input data from a variety of formats into R (readr)
3. Describe how and why to transform data in R (dplyr, tidyr, stringr)
4. Explain how to visualize data with R (ggplot2)
5. Be able to conduct Exploratory Data Analysis in R
6. Be able to create basic statistical models in R (, lm, rpart)
7. Be able to communicate and share data (RMarkdown language)

Course Requirements & Grading

The course is evaluated on a letter grade scale (see below). Points are awarded based on attendance, participation in the online discussion forums, weekly project checkpoints, and a group/individual project:

Assignment	Date	Weight (Percentage, points, etc)
Attendance and Participation	Weekly	25 points
Discussion Posts	Weekly	20 points
Project Checkpoints	Weeks 1-4	20 points
Overall Project		
Presentation, Poster Presentation and/or Written Report	Week 5	30 points
Constructive Feedback		5 points

Extra Credit Opportunities

A student may earn up to 10 points of extra credit by attending 1-2 (5 pts each) Library data workshops or IDEAS (http://ideas.gatech.edu/ (Links to an external site.)) presentations. To receive points students must write a 1-page review of the event and show a working code example to the instructors.

Description of Graded Components

Attendance and participation: Up to 5 points per week actively attended.

Online Discussion: Each week a discussion question will be posted related to the ethical or open use of data. 5pts will be earned for each discussion posts, 3pts for your own post and and 2pts for responding to at least one post of a peer student. Discussion posts are due by Sunday, 11:30PM EST.

PROJECT

The project can be done individually or as a group of up to 3 students. For group work, each group must clearly define the individual contributions of each member of the project team. Please email me if you wish to work as a team.

Weekly project checkpoints will put into practice the skills you learn from the weekly workshops and help you progress toward the final project. You can earn 5 points for each checkpoint due at the end of weeks 1-4. Weekly project checkpoints will be due Sunday at 11:30PM EST, and instructors will provide feedback by the following Tuesday.

1. Week 1: 1-2 page description of a selected data set, including what the dataset describes, a brief data dictionary, and a preliminary research question.
2. Week 2: Submit an R script of transformations on the data, and exploratory data analysis, and an annotated bibliography of 3-5 references related to your data/research question focusing on methodology used.
3. Week 3: 1-2 page description of the model you plan to use, including which R packages or verbs you will use. Cite at least 1 reference work that uses or explains your model/package.
4. Week 4: Submit an R script file for your data visualization and data model.

Overall project reports: For the overall project you will present an analysis of your data, prepare a written report, **and/or** create a poster presentation. The overall project builds directly from the weekly project checkpoints.

Option A: Project Presentation

1. Each project team will provide a 10-15 minute recorded presentation on their data, research question, steps taken in R, and their results in visual form. Each team member should contribute and speak during the presentation.
2. Required Elements:
    1. Title (1 pt)
    2. Description of the Data Set (5 pts)
    3. Research Question (5 pts)
    4. What you did in R (8 pts) (e.g. tell us the model, the package, and the verbs you used)
    5. What were your results (9 pts) (e.g. the output and visualization of the output of the model)
    6. Bibliography (2 pt)

Option B: Project Report

1. Final report in IEEE Conference Format 6-10 pages https://www.ieee.org/conferences/publishing/templates.html
2. Project Document will have the following sections:
    1. Abstract (1 pt)
    2. Introduction (1-2 paragraphs): Research Problem, Purpose of the Study, Motivation, Audience, Contribution, Goal and Paper Organization (8 pts)
    3. Related Works (2-3 paragraphs) (2 pts)
    4. Techniques used in the study and Proposed Methods (1-2 pages) (7 pts)
    5. Results (1-2 pages) (8 pts)
    6. Conclusions (1+ paragraphs) (1 pt)
    7. Future Work (1+ paragraphs) (1 pt)
    8. References (2 pts)

Option C: Poster Presentation

1. Create a poster presentation using the UROP template: http://www.urop.gatech.edu/sites/default/files/documents/sample.postertemplates09.pdf
2. Required Elements:
    1. Title (1 pt)
    2. Description of the Data Set (5 pts)
    3. Research Question (5 pts)
    4. What you did in R (8 pts) (e.g. tell us the model, the package, and the verbs you used)
    5. What were your results (9 pts) (e.g. the output and visualization of the output of the model)
    6. Bibliography (2 pt)

Constructive Feedback

Students will be asked to critique two final presentations/papers prepared by their colleagues and provide constructive feedback. You will be graded up to 5 pts for this feedback, and your feedback will be shared with your colleague.

Grading Scale

Open Data with R is offered as a Letter Grade:

- A: 90-100
- B: 80-90
- C: 70-80
- D: 60-70
- F: < 60

Course Materials

Course Reference

This is a "no-cost materials" course and all reading material will be available open access or through the library e-book collection. The course will draw from *R for Data Science* by Hadley Wickham and Garrett Grolemund, O'Reilly, 2016. The textbook is freely available online: https://r4ds.had.co.nz/ Links to an external site. (Links to an external site.)

Additional Materials/Resources

We will use Posit Cloud in this course: https://posit.cloudLinks to an external site. A free account is sufficient for use in both class sessions and for working on your course project.

Campus Resources for Students

The Library has many resources for assisting with data analysis and with R. See the Library Research Guide for more details: http://libguides.gatech.edu/RStudio

Course Expectations & Guidelines

Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on Georgia Tech's Academic Honor Code, please visit http://www.catalog.gatech.edu/policies/honor-code/ (Links to an external site.) or http://www.catalog.gatech.edu/rules/18/ (Links to an external site.).

Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404)894-2563 or http://disabilityservices.gatech.edu/ (Links to an external site.), as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Attendance and/or Participation

Attendance and participation are graded and recommended. Attendance includes attending the R studio workshops scheduled during class time and participating in the discussion forum. Participation includes executing example code in R and posting discussion responses to other's work. If you cannot attend a class, please let me know, for attendance credit you will need to watch the recording and send screenshots of your running R code.

Collaboration & Group Work

Students should use any resource to help them better understand R and/or data analysis and may work collaboratively. For the overall project, each team member must develop R code for at least a portion of the project. Finally, if you feel that you need to ask someone to develop R code in its entirety for an assignment or project, please reach out to the instructors so that we can help you learn how to use R better.

Extensions, Late Assignments, & Re-Scheduled/Missed Exams

This is a five-week class, and each phase builds from the prior week, so it is not recommended to fall behind. All assignments and discussion posts will be open on the first day of class. Canvas will indicate a recommended due date, but with one exception all assignments are open until Sunday 9/22/2024 at 11:30 PM EDT and all work will be excepted without penalty by that due date. The exception is that you must turn in a draft of your course project to the Peer Review Discussion Board by Tuesday, September 17th at 11:30PM EDT.

Student-Faculty Expectations Agreement

At Georgia Tech we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. See http://www.catalog.gatech.edu/rules/22/ (Links to an external site.) for an articulation of some basic expectation that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build

the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

Student Use of Mobile Devices in the Classroom

Most of the class is hands-on, you will spend class time developing R code. Mobile phones should be on mute during the class, and if you need to have a conversation, please mute your audio/video.