GT 4801/GT8801 Syllabus

Open Data with R, A, 1 credit hour

Instructor Information

Instructor	Email	Office Hours & Location
Jay Forrest	jay.forrest@library.gatech.edu	TBD, Library
Instructor	**Email**	**Office Hours & Location**
Ameet Doshi	ameet.doshi@library.gatech.edu	TBD, Library

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

There are no co- or pre-requisites for the class.

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. Understand how to input data from a variety of formats into R (readr)
3. Understand why and how to transform data in R (dplyr, tidyr, stringr)
4. Understand 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 (modelr, lm, rpart)
7. Be able to communicate and share data (RMarkdown language, GitHub)

Course Requirements & Grading

The course is evaluated using pass/fail grading. 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	25 points
Project Checkpoints	Weeks 1-4	20 points
Overall Project Project Presentation Presentation Feedback Written Project	Week 5	 10 points 5 points 15 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/) 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. Students will provide their own response with at least 1 citation (3 points/week) and respond to the answer of at least two other students (1 pts each). Discussion posts are due by Friday, 11:30PM, and responses to other students due by Sunday, 11:30PM.

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.

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, 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 during the final week of class and prepare a written report. The overall project builds directly from the weekly project checkpoints.

1. Project Presentation
    a. During the final week of class, each project team will provide a 10-15 minute 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. Project Document
    a. Final report in IEEE Conference Format 6-10 pages https://www.ieee.org/conferences/publishing/templates.html
    b. Project Document will have the following sections:
        i. Abstract
        ii. Introduction (1-2 pages): Research Problem, Purpose of the Study, Motivation, Audience, Contribution, Goal and Paper Organization

iii. Related Works (1-2 pages):
iv. Techniques used in the study and Proposed Methods (1-2 pages)
v. Results (1-2 pages)
vi. Conclusions (1+ paragraphs)
vii. Future Work (1+ paragraphs)
viii. References

Grading Scale

Open Data with R is offered Pass/Fail: A student will pass Open Data with R if their point total at the end of the program is greater than or equal to 70 points, **AND** they accumulate points in all four of the categories listed above (Attendance, Discussion, Project Checkpoints, Final Project).

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 available through the GT Library at https://learning.oreilly.com/library/view/r-for-data/9781491910382/

Additional Materials/Resources

Students are encouraged to download R and R Studio to their own computers. Both software packages are Open Source and work with Mac/Linux/Windows. Optionally, a student can use the cloud version of R rstudio.cloud

Course Website and Other Classroom Management Tools

Canvas Link: TBD

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/ or http://www.catalog.gatech.edu/rules/18/.

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/, 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 the instructors know in advance, so that we can work out an alternative access to the workshop materials. Attendance is required for the last week of class.

Collaboration & Group Work

Students should use any resource 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.

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. Weekly project checkpoints will receive a 20% penalty for each day submitted late. Late submission for the discussion posts is not permitted, nor can an extension be granted for the written and presentation components of the project.

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/ 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 step out of the room. The exception is during the student project presentations where mobile devices should not be used.

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 Schedule

Date	Topic	Deliverables
Week 1	Intro to Open Science, Open Data and Intro to R	Attendance, Discussion Post, Project checkpoint 1
Week 2	Data Transformation and Exploratory Data Analysis	Attendance, Discussion Post, Project checkpoint 2
Week 3	Data Modeling and R Programming	Attendance, Discussion Post, Project checkpoint 3
Week 4	Data Visualization and RMarkdown	Attendance, discussion post, project checkpoint 4
Week 5	Communicating Data for Reproducibility	Project Presentation and Written Report