CDS 102 Syllabus (Spring 2021)

Introduction to Computational and Data Sciences Lab

Section 102/2P3

Description

In this lab, students will develop basic skills for obtaining, cleaning, transforming, and visualizing real-world datasets using the R programming language and the RStudio integrated development environment. Statistical methods for analyzing, interpreting, and predicting dataset trends are then introduced and approached from a computational point of view using randomization and simulation. Emphasis is placed on documenting one's scientific work using RStudio in conjunction with GitHub to fulfill the principles of reproducible research. Connections are highlighted between statistical inference and the scientific method and how this is related to both the scientific method's power and its limitations. The labs complement and allow further practice for the topics covered in CDS 101.

- **Classroom:** On-Campus, INN 222
- **Meeting times:** Thursdays, 4:30pm-5:45pm
- **University holidays:** N/A (No spring break this semester
- **Credit hours:** 1.0 credit hour
- **Prerequisites:** None, but a background in algebra is assumed
- **Mason Core:** Natural science + lab when taken with CDS 101

Instructor – Mr. Matthew "Jordan" Miller

- **Office**: Zoom -- https://gmu.zoom.us/j/7677876729 *By Appointment*
- **Phone:** (540) 324-8751
- **Email:** mmille64@gmu.edu
- **Online Office hours:** TBA, or by appointment
- **Teaching Assistant:** N/A

Objectives

By the end of the course, students will be able to:

- Use GitHub for collaborating on a reproducible research document
- Obtain, clean, transform, and visualize a dataset using the R programming language
- Interpret, and predict dataset trends using statistical inference and models

Materials

Software

During the course we will use RStudio Server available at https://rstudio.cos.gmu.edu, which provides a complete computing environment that is accessible using any computer with a modern web browser (Firefox and Chrome). Students are welcome to install RStudio on their own computers and will need to install the following applications in order to match what is available on RStudio Server:

- Programming language: R (https://www.r-project.org)
  - Windows: https://cran.cnr.berkeley.edu/bin/windows/base/
  - Mac: https://cran.cnr.berkeley.edu/bin/macosx/
- Version control: Git (https://git-scm.com)
  - Windows: https://git-scm.com/download/win
  - Mac: https://git-scm.com/download/mac
- PDF export: LaTeX (https://www.latex-project.org)
  - Windows: https://mirrors.rit.edu/CTAN/systems/win32/miktex/setup/windows-x64/basic-miktex-2.9.6753-x64.exe
  - Mac: http://tug.org/cgi-bin/mactex-download/MacTeX.pkg
- Programming software: RStudio (https://www.rstudio.com)
  - Windows: https://s3.amazonaws.com/rstudio-ide-build/desktop/windows/RStudio-1.2.830.exe
  - Mac: https://s3.amazonaws.com/rstudio-ide-build/desktop/macos/RStudio-1.2.830.dmg

Technical support will only be provided for RStudio Server.

Platforms

The course will be administered through the following online platforms:

- Course website (https://labs.cds101.com)
- GitHub
- Slack
- Blackboard

The course website operates as the central repository for the lab instructions. Slack is the primary communication medium, replacing email (see the *Contact policy* below) and serving as a discussion board. GitHub is used for storing your classroom files, distributing and collecting homework assignments, handing out example code, and for project collaborations. Blackboard will be used for posting grades.

Policies

Contact policy

All correspondence is to be done using the private, invite-only Slack workspace for the course, email, or text. Direct messages on Slack are the most preferred method of contact. Your Slack username *must* be registered and associated with your Mason @masonlive.gmu.edu email address (yes, the @masonlive one).

My ground rules for direct messages are as follows:

- I check and respond to messages throughout the day.
- Allow up to approximately 24 hours for a response during normal hours.
- Just because I view a message does not mean I will respond right away.
- Please allow extra lag in response time during weekends and school holidays.
- If your questions are involved enough, I will ask you to talk to me in person or (via zoom).

Tech support: R, RStudio, GitHub, and your computer

Post all technical issues or error messages for R, RStudio, GitHub, and your computer in the designated Slack channel
#r-rstudio-github-help: https://masoncds101.slack.com/messages/CASUNTFNX.
This is so that other students can either help out or see how to resolve what is likely a common problem. If it becomes clear that the error or issue is highly specific, then discussion can be moved to Direct Message or handled via a remote desktop sharing session.

When posting about an issue, here are some basic questions to answer that will help with troubleshooting:

1. What did you expect to happen when you ran your code?
2. What is actually happening when you run your code?
    a. If there's an error message, tell us what it is. A screenshot works, provided you a) don't crop the image as that can remove useful information by accident, and b) take a real screenshot, not a photograph of your screen using your phone.
3. Is there any other context we should know? For example, if a file won't load, did you check that you are in the correct project or that the file actually exists? Did your issue appear only after you worked on a different assignment? Did you recently install a package not used in class?

Illness and emergencies

It is a student's responsibility to inform me about illnesses or personal/family emergencies that will interfere with attendance or submitting work on time. This must be done as soon as possible. In case of illness, you may be asked to provide a doctor's note before being granted an assignment extension or exemption.

I understand that certain emotional or physical situations can impact a student's willingness to communicate what is going on and that it can take a day or two to inform me about a personal emergency or severe illness. At the same time, all students are expected to exercise personal responsibility. It is not acceptable to wait to tell me about the impacts of a personal illness or emergency until you're about to fail the course due to not attending class or missing lab submission deadlines.

Attendance policy

Students are expected to attend every lab, and links to begin work will only be distributed to students that show up in person. Students must get in contact with the instructor to obtain the access link if they are absent, and the due date for the lab remains the same regardless of when you get the link. Unexcused absences, frequent tardiness, or leaving class early will reduce your participation grade.

Students are responsible for informing me about upcoming absences due to religious holidays, scheduled varsity sports trips, or a school-sponsored activity. Any make-up work is to be completed within the time-frame I set forth. Exemptions may be granted at my discretion.

Students With Four or More Unexcused Absences Will Receive An Automatic Zero For The Course

Late work policy

Unless otherwise noted, labs are to be submitted by the given due date. The following penalties apply for labs (please note that weekends count as days):

- First day late, by 11:59pm: -10%
- Second day late, by 11:59pm: -20%
- Third day late, by 11:59pm: -30%
- Fourth day late or later: no credit

Extensions or exemptions may be granted at my discretion.

Students With Four or More Labs That Are Submitted 4+ Days late or Not Submitted At All Will Receive An Automatic Zero For The Course, No Exceptions.

Regrading appeals policy

Regrade appeals need to be in writing, printed out, and hand delivered to me (or emailed as a PDF) **within 48 hours of receiving back an assignment** (not including weekends). Appeals via Slack or email will not be accepted, no exceptions. Appeals are only to be used for correct answers being marked as incorrect, misapplication of the grading rubric, or incorrectly tallied points. Submissions need to clearly state what you want regraded and to justify the request by citing evidence[1]. The number of points a question, exercise, or rubric category is worth or that were deducted for an incorrect answer or mistake cannot be appealed and are not up for debate or negotiation.

Extra credit and grading curves policy

Individual requests for extra credit or a grading curve will not be granted, no exceptions. Any opportunities to earn extra points will be offered to the entire class. Grading curves are handled on a per-lab basis and are applied to all students equally.

Accommodations policy

Students with disabilities who need academic accommodations, please see me and contact the Office of Disability Services (ODS) at (703) 993-2474. All academic accommodations must be arranged through the ODS: http://ods.gmu.edu/.

Grading

Breakdown

Category	Weight
Lab Reports:	90%
Participation:	10%

Schema

Based on the final total score, your final grade will be determined as follows: A+ [97-100], A [93-96], A- [90-92], B+ [87-89], B [83-86], B- [80-82], C+ [77-79], C [73-76], C- [70-72], D [65-69], F [<65].

[1] Acceptable evidence includes class notes from CDS 101 (provide date of class), a reading passage (provide full citation), or another valid source (textbooks, official publications, etc).

Expectations

Participation

Students are expected to come to class prepared and ready to participate in class activities and exercises, which at times may be completed in pairs or in groups. Students are expected to give their full effort while completing an in-class exercise and to complete it within a reasonable timeframe. A combination of speed of completion and quality of work will be factored into your participation grade. A student's number of absences during the semester will also factor into his/her participation grade.

Lab Reports

At the beginning of each lab you will be provided with a set of instructions and a link to obtain your lab repository on GitHub. The lab reports will be completed and submitted using the starter files provided to you in the lab repository. To fill out the lab reports, you will need to read and follow the instructions. The labs are similar to interactive online tutorials, where you will read some instructions, see an example, and then complete an exercise to demonstrate that you understand a concept. The exercises consist of many kinds of activities, such as creating visualizations, manipulating datasets, performing statistical analysis, and building models. Many of the labs then conclude with a set of questions where you will perform additional analysis and interpret the meaning of your results. When you have completed writing up your report in the RMarkdown file format, you will submit the final result on GitHub using a *Push Request*.

You will be provided with lab report templates each class, and a complete lab report should have the following:

- Your name
- A title
- The date you are submitting the lab report
- A sub header for each exercise and each question, with your answers and work going in the space in between the headers
- Your responses will document your procedure for completing each exercise and question. This involves explaining each code block that you write. Be precise. Write how you would write a tutorial for a new user. Do not assume that the reader will fill in any missing steps.
- The finished lab report should have proper grammar and spelling.
- Follow the rule that anytime you need to write a response or explain a code block, you should write in **complete sentences**. If you are interpreting something like a visualization or a table, it is likely that you will need to write more than one sentence.

You may need to finish writing up the lab report outside of class. See each lab report for the due date. Typically, the due date is the beginning of the next lab.

Final exam

There is no final exam for the labs.

Conduct

Academic integrity

"Student members of the George Mason University community pledge not to cheat, plagiarize, steal, or lie in matters related to academic work."[2]

Your lab report is to be written by yourself and in your own words. This is the rule even when you are working in groups of two or three. The groups are for discussing ideas and how to implement those ideas using R. Do not duplicate or paraphrase another person's material or ideas and represent them as your own. Content that comes from a resource or another student should be properly cited.

A note on sharing or reusing code found on other GitHub repos or on websites like *Wikipedia* or *Stack Overflow*. I am aware that there are solution sets, sample snippets of code, etc. that can be of use while working on your assignments, projects and exercises during the course. It's common knowledge that researchers in both industry and academia will use search engines while writing code. Being able to search for existing solutions so that you don't "reinvent the wheel" is a useful skill. Therefore, unless I specify otherwise, you are permitted to use these resources **as long as you provide a citation**.

Exceptions to this rule are:

- You and your lab partners cannot submit identical lab reports, although you can discuss your code. You and your lab partners cannot submit identical lab reports or a collaborative lab document where all group members contributed to different parts *unless* explicitly stated by the lab instructions.
- You are not permitted to share code snippets with another student that is not your partner.
- You are not permitted to reference lab reports completed by students from previous semesters.
- You are not permitted to look for, consult, or use solutions to labs that are similar to our labs

Any Material That Is Taken In Whole Or In Part From Another Source And Not Properly Cited Will Be Treated As A Violation Of MASON's Academic Honor Code.

Other violations of Mason's Honor Code will be treated similarly. Suspected violations will be reported to the Office of Academic Integrity. Please see the Honor Code page for details.

[2] Office for Academic Integrity. *2017-2018 Honor Code and Honor System.* Web. 27 Aug. 2017.

Decorum/discourse

Students are expected to be civil in their classroom conduct and respectful of their fellow classmates and the instructor for the duration of the course. Examples of expected behavior include, but are not limited to:

- Showing up to class on time
- Not interrupting your classmates or the instructor
- Silencing your cell phone
- Refraining from texting/messaging
- Refraining from using devices for anything other than coursework[3]
- Removing ear-buds/headphones and sunglasses when class begins

The expectations of civil and respectful behavior still apply for all online discussions. Students are expected to follow proper grammar and punctuation in online posts and to refrain from using internet slang, abbreviations, and sarcasm.

I will address violations of classroom decorum on a case-by-case basis and reserve the right to enact grade-based penalties for highly disruptive or repeat violations. Penalties for decorum violations cannot be negotiated or appealed.

[3] The term "devices" is meant to be broad and includes classroom computers, laptops, cell phones, tablets and e-readers, smart watches, etc. Exceptions can be made in cases of family or personal emergencies, please speak to me before class.

Mason diversity statement

George Mason University promotes a living and learning environment for outstanding growth and productivity among its students, faculty and staff. An emphasis upon diversity and inclusion throughout the campus community is essential to achieve these goals. Diversity is broadly defined to include such characteristics as, but not limited to, race, ethnicity, gender, religion, age, disability, and sexual orientation. Diversity also entails different viewpoints, philosophies, and perspectives. Attention to these aspects of diversity will help promote a culture of inclusion and belonging, and an environment where diverse opinions, backgrounds and practices have the opportunity to be voiced, heard and respected.

Support services

The Math Tutoring Center is in 344 Johnson Center; http://math.gmu.edu/tutor-center.php. The Math Department also maintains a list of persons that have identified themselves as math tutors: http://math.gmu.edu/tutor-list.php

Mason's Writing Center is in A114 Robinson Hall; (703) 993-1200; http://writingcenter.gmu.edu/

George Mason provides Counseling and Psychological Services (CAPS) for students. Contact them at (703) 993-2380 or http://caps.gmu.edu/.

Disclaimer

The instructor reserves the right to modify this syllabus at any time during the course to improve the learning experience and classroom environment. The digital version of the syllabus on the course website will be updated to reflect the changes. The pacing of the course and the list of covered topics may also be altered in response to student progress.

The course objectives reflect what a student is expected to understand by the end of the course after putting in the necessary time and effort both inside and outside the classroom and completing all assignments. These outcomes are not a guarantee, and students will get more out of the course the more they put into it. Any acquired skills and knowledge can fade over time if not reviewed or practiced after the course concludes.

Lab schedule is available on https://labs.cds101.com