CDS 101 Syllabus (Fall 2020)

Introduction to Computational and Data Sciences

Section 002

The One-Page Summary

What:

This course will teach you how to use computers to analyze data. Data analysis using computers is now a core part of every scientific field, as well as many other areas of human endeavor. You will learn how to program in the R programming language, and become proficient in using R to manipulate tables of data, create graphs, and conduct basic statistical analyses.

When & Where:

Your section of CDS 101 will meet Tuesdays and Thursdays from 3:00-4:15pm online via Blackboard Collaborate.

Who:

Your section of CDS 101 is taught by Dr. Dominic White, an Assistant Professor in the Department of Computational and Data Sciences. Your section also has two STARS (undergraduate TAs): Collin Adams and Zachary Stept.

How:

Each week we will cover a different module. As part of each module:

- You will have a set of preparatory materials (a combination of videos, interactive tutorials, and readings). You must do these before coming to the first class of the week.
- There will be a short Blackboard quiz on those preparatory materials (this will be due by the start of the first class of the week).
- In class, we will work on actual programming assignments. There will be a new assignment each week.
- In addition, there will be a Midterm Project half way through the semester, as well as a Final Project at the end of the semester.

There are two policies that are particularly important to remember:

- *The "Hogwarts" Policy*: Help will always be given in CDS 101 to those who ask for it. If you are stuck, please reach out to your instructor or a STAR, and we will guide you through the material.
- *Academic Honesty*: Plagiarism, copying, cheating, sharing answers, etc. are strictly prohibited. You will be caught and reported to the University, which can result in failing the course or expulsion.

The Fine Print

Description

During this course, students will develop basic skills for loading, 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. An emphasis is placed on documenting one's scientific work using the R Markdown format 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. These tools will also be used to critically examine statistical claims reported in mass media, demonstrating how scientific literacy and a basic knowledge in statistics are indispensable tools to making sense of our modern world. Special topics like machine learning and dashboards may also be covered as time allows.

- **Classroom:** Online via Blackboard Collaborate
- **Meeting times:** Tuesday and Thursday 3:00-4:15pm
- **University holidays:** Labor Day (Mon Sep 7), Fall Break (Mon Oct 12), Thanksgiving (Weds Nov 25 - Sun Nov 29)
- **Credit hours:** 3.0 credit hours
- **Prerequisites:** None, but a background in algebra is assumed.
- **Mason Core:** Natural science + lab (when taken with CDS 102)

Instructor: Dr. Dominic White	
Office	242 Research Hall
Slack	@Dr White
Email	dwhite34@gmu.edu
Office hours (online via BB Collaborate)	Thursdays 1-2pm

CDS STARS (Student Teaching Assistants and Researchers)		
Name	Collin Adams	Zachary Stept
Slack	@Collin	@Zach
Email	cadams31@masonlive.gmu.edu	zstept@masonlive.gmu.edu
Office hours (online via BB Collaborate)	TBD	TBD

Objectives

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

- Obtain, clean, transform, and visualize a dataset using the R programming language
- Interpret, and predict dataset trends using statistical inference and models
- Document their work using R Markdown, a reproducible research format
- Manage files and source code using GitHub

Materials

Textbooks

This course utilizes three textbooks that are freely available online under Creative Commons licenses:

R. Irizarry, *Introduction to Data Science: Data Analysis and Prediction Algorithms with R* (Chapman and Hall/CRC, 2019). `https://rafalab.github.io/dsbook/`

H. Wickham and G. Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (O'Reilly Media, Sebastopol, CA, 2017). `http://r4ds.had.co.nz/`

D.M. Diez, C.D. Barr, and M. Çetinkaya-Rundel, *Introductory Statistics with Randomization and Simulation, 1st ed.* (OpenIntro, United States, 2014). `https://www.openintro.org/stat/index.php?stat_book=isrs`

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-2.9.7086-x64.exe
    - Mac: http://tug.org/cgi-bin/mactex-download/MacTeX.pkg
- Programming software: RStudio (https://www.rstudio.com)
    - Windows: https://download1.rstudio.org/desktop/windows/RStudio-1.2.1335.exe
    - Mac: https://download1.rstudio.org/desktop/macos/RStudio-1.2.1335.dmg

Technical support will only be provided for RStudio Server.

Platforms

The course will be administered through the following online platforms:

- GitHub: https://github.com
- Slack: Link on Blackboard
- Blackboard: https://mymasonportal.gmu.edu

The course website operates as the central repository for course materials, copies of the lecture slides and handouts, homework instructions, and links to the lecture videos hosted on YouTube. Slack is the primary communication medium, replacing email (see the *Contact policy* below) while also serving as a discussion board. GitHub is used for connecting your class files to RStudio Server, tracking changes, distributing starter files for homework assignments and certain module exercises, and for project collaborations. Blackboard is used for module exercises, part of the midterm exam, for submitting assignments, and returning grades.

Policies

Contact policy

All correspondence is to be done using the private, invite-only Slack workspace for the course. **Direct messages on Slack are to be used for contacting me instead of emails**. My ground rules for direct messages are as follows:

- I check and respond to messages during normal university hours (i.e. weekdays 9am-5pm).
- Allow up to 24 hours for a response during normal hours.
- Questions about homework problems or troubleshooting a technical issue should be posted to the relevant Slack channels.

- Questions about an assignment should be asked before the due date. Questions asked after 5:00pm on the due date will be answered the following morning.[1]
- Just because I view a message does not mean I will respond right away.
- I generally don't respond to messages over weekends and school holidays.
- If your questions are involved enough, I will ask you to schedule an appointment with me.
- On email: Emails sent during the first week of classes will be responded to, but I will respond to you using Slack. Emails sent to me after the first week will be ignored.[2]

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`. 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 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 and ability to communicate what is going on and that it can take a few days 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 missing multiple submission

---

[1] Exceptions to this rule are determined on a case-by-case basis. For example, I try to respond when there are unexpected technical glitches.

[2] If there are special circumstances requiring that we communicate via email, it is your responsibility to inform me about it as soon as possible.

deadlines.

Late work policy

Unless otherwise noted, assignments are to be submitted by 11:59pm on the due date. The following penalties apply for most assignments (please note that weekends count as days):

- First day late, by 11:59pm: -15%
- Second day late, by 11:59pm: -30%
- Third day late or later: no credit

The above does not pertain to the quizzes, which must be completed by the designated date and time to receive credit. Late submissions for the midterm and final projects will not be accepted.

Students are responsible for informing me about any religious holidays, scheduled varsity sports trips, or other school-sponsored activities that will interfere with submitting an assignment on time. Extensions are to be completed within the time-frame I set forth. Exemptions may be granted at my discretion.

Regrading appeals policy

Regrade appeals need to be written and formatted as a formal letter and submitted to me as a PDF file within 48 hours of receiving back an assignment (not including weekends). Appeals sent in plain text 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 justify the request by citing evidence[3]. 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-assignment 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/`.

---

[3] Acceptable evidence includes class notes (provide class module number), a reading passage (provide full citation), or another valid source (textbooks, official publications, etc).

Grading

Breakdown

Category	Weight
FERPA form	1%
Quizzes:	9%
Weekly Assignments:	45%
Midterm project:	15%
Final project:	30%

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].

Expectations

Quizzes

Weekly quizzes for checking your understanding and to reinforce what you've learned will be a regular part of the course and collectively factor into the Quizzes grade category.

Readings

Reading assignments are regularly scheduled during the semester and students are expected to complete all readings on their own. These readings are an important part of the course and the lecture videos should not be seen as a replacement for them. While there is not a grade category for the course readings, students are responsible for understanding the content and students that skip them are under-prepared for the later parts of the course.

Students are encouraged to ask questions about and discuss the readings in Slack. To maintain a high quality discussion, here's a list of things you should do before submitting your question in the Slack channel:

- Do…
    - use the search feature to check if your question has already been asked, and if so, contribute to that ongoing discussion thread instead of asking the question again
    - ask questions that are on-topic and about the reading's subject matter instead of asking about tangential commentary made by the reading's author or extraneous details in an example
    - explain, as part of your question, which part of a concept or method you do not understand and why, instead of just saying you don't understand it
    - look up words you do not understand and information you can quickly do a web search for to help you better phrase and contextualize your question

Assignments

The assignments are to be completed using the R Markdown format introduced during the first couple of weeks of the course.

- Your R Markdown file must successfully knit to PDF in a clean RStudio environment to be eligible for full credit.
- Full sentences with proper grammar and punctuation are to be used throughout the submission.
- You must explain what you are doing with each chunk of code and interpret the meaning of what you calculate so that a person that is not familiar with the problem could follow your logic.

Grades for the individual assignments will be primarily based on the correctness of your answers to each question, as well as document formatting, visualization quality, writing quality, and code style. For example, do not engage in unorthodox practices such as writing sentences in a section header, placing code outside of code blocks, or submitting a PDF containing figures with unreadable overlapping labels, code blocks that run off the edge of the page, and is 50–100 pages in length because you keep printing out long tables. Also, unless specifically requested in the instructions, screenshots should never be included as part of your submission.

General questions about the assignments should be posted in your section's Slack channel. Take care to ensure that your submissions are in your own words and code. See the academic integrity section for specifics.

Midterm & Final projects

Students will complete two projects where you will perform an exploratory data analysis on a dataset. More detailed information about the project is to come later in the semester.

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."[4]

Students are permitted to ask questions about the assignments on Slack and discuss assignments in private communications, however it is important to make sure that you write your assignments by yourself and in your own words, meaning that students are not permitted to collaborate on write-ups for assignments and projects. In the same vein, do not duplicate or paraphrase another person's material or ideas and represent them as your own. "Individual assignment" is the default classification for all assignments, exams, and projects in the course; any exceptions to the rule will be noted in the instructions. Content that comes from a resource or another student should be properly cited.

---

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

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:

- For individual assignments, you cannot reuse anything from another student's work (past or present), including but not limited to R Markdown documents, code, plain text explanations, etc.
- You are not permitted to consult or use solution sets for any of the assignments, activities, and projects for the course.
- You are not permitted to ask other students from this or previous semesters for copies of their assignments or projects, even to use for reference.

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. The minimum sanctions I will recommend are

- Minor infraction (e.g. improper citation):
    - First offense (university-wide): 0% on the assignment.
    - Second offense: F for the course.
    - Third offense: expulsion from the university.
- Major infraction (e.g. cheating in an exam, or copying another student's assignment/project):
    - First offense: F for the course.
    - Second offense: expulsion from the university.

Decorum/discourse

Students are expected to be civil in their conduct and respectful of their fellow classmates and the instructor for the duration of the course on all discussion platforms. Students are expected to follow proper grammar and punctuation in their posted messages 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 disruptive or repeat violations. Penalties for decorum violations cannot be negotiated or appealed.

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 pacing of the course and the list of covered topics may also be adjusted 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.

Schedule

A provisional calendar for the course modules will be made available on Blackboard. Timings of different modules may be changed subject to our progress throughout the semester.