# CDS 101 Syllabus (Spring 2022)
Introduction to Computational and Data Sciences

Section K02

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. For many students, CDS-101 may be their first formal computer language course. To accommodate students with different levels of coding experience, the course is presented as if all students are learning to code for the first time.

Your section of CDS 101 is taught by Professor Brian Colchao from the Department of Computational and Data Sciences.
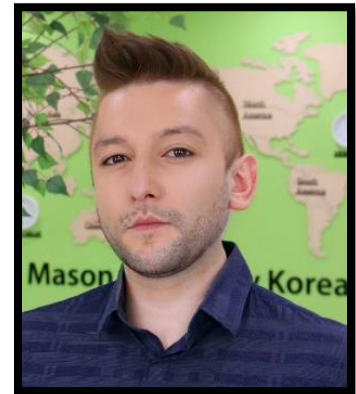
Your tutors for this course are Eun Won Kim and Yena Hong.
- They can be reached through Slack.
- Office hours are virtual, but can be arranged in-person upon request

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 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.

- **Classroom:** G301
- **Meeting times:** Thursdays 9:00AM – 11:40PM room G301
- **Office Hours:**
    - Hong: Monday 11AM – 12PM
    - Kim: Tuesday 12PM – 1PM
    - Colchao: Wednesday 4:00 – 5:00 PM
- **Credit hours:** 3.0 credit hours.
- **Prerequisites:** None, but a background in algebra is assumed.
- **Mason Core:**

    - Natural science requirement
    - Natural science + lab (when taken with CDS 102)

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. Please save these links as favorites in your internet browser for quick access:

1. R. Irizarry, *Introduction to Data Science: Data Analysis and Prediction Algorithms with R* (Chapman and Hall/CRC, 2019).
   - https://rafalab.github.io/dsbook/
2. C. Ismay and A.Y. Kim,, *Statistical Inference via Data Science: A ModernDive into R and the Tidyverse* (Chapman and Hall/CRC, 2019).
   - https://moderndive.com
3. 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/
4. 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 must also install the R programming language and RStudio on their own computers and will need to install the following applications 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
- Programming software: RStudio (https://www.rstudio.com)
- PDF export: LaTeX (https://www.latex-project.org)
  - TinyTex: https://bookdown.org/yihui/rmarkdown-cookbook/install-latex.html

Platforms

The course will be administered through the following online platforms:

- ==GitHub==: https://github.com
  - 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
- ==Slack==: https://www.slack.com
  - Slack is ==the primary communication medium==, replacing email (see the *Contact policy* below) while also serving as a discussion board.
- ==Blackboard==: https://mymasonportal.gmu.edu
  - Blackboard operates as the central ==resource for course materials==, homework instructions, and links to the lecture videos hosted on YouTube.
  - Blackboard is also used for ==module exercises, projects, assignments, and 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:

- ==Help will always be given in CDS 101 to those who ask for it.== If you are stuck, please reach out to your professor or the tutors on Slack. You are welcome to message us 24 hours a day, 7 days a week.
- Allow up to 24 hours for a response.
- ==Questions about an assignment should be asked before the due date. Questions asked after 5:00pm on a due date may not receive a response until after the time the assignment is due==.
- Messages received over ==weekends and holidays may receive delayed responses==.
- If your questions are complicated, I may ask you to join me in a Zoom meeting or in my office to work through it.
- 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 redirected to Slack.

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

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

1. What is happening when you run your code? What were you expecting?
   a. ==If there's an error message, write what it is. You must provide a screenshot.== Be sure to take a real screenshot, not a photograph of your screen using your phone.

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. During the COVID-19 pandemic this is more important than ever.

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

The above does not pertain to the projects and quizzes. Late submissions for the quizzes and 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.

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 contact Joanna Park (spark214@gmu.edu) at the Office of Disability Services (ODS) at +82-32-626-5071. All academic accommodations must be arranged through the ODS: http://ds.gmu.edu/ during the first 2 weeks of class. Please notify me after your accommodation has been approved by the Office of Disability Services.

Campus Closure or Class Cancelation/Adjustment Policy

If the campus closes, or if a class meeting needs to be canceled or adjusted due to weather or other concern, students should check Blackboard and Slack for updates on how to continue learning and for information about any changes to events or assignments.

Grading

Breakdown

Category	Weight
Setup (FERPA)	1%
Quizzes:	9%
Weekly Assignments:	45%
Inference Project:	10%
Midterm project:	15%
Final project:	20%

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

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

- You will have a combination of videos, interactive tutorials, and readings.
- There will be a short Blackboard quiz on those materials every week.
- There will be an assignment each week.
- Instead of in-class exams, there will be a Midterm, Inference, and Final Project.

Interactive Tutorials and Readings

Weekly tutorials are scheduled during the semester and students are expected to complete these on their own. These activities are an important part of the course and are prerequisites for the assignments that we will work on in class. While there is not a grade category for the interactive tutorials and readings, information presented in the tutorials will be on the quizzes.

Students are encouraged to ask questions in Slack:

- Do…
    - Ask questions that are on-topic and about the reading's subject matter
    - Your question should also explain which part of a concept you do not understand and why, never say you don't understand everything
    - Look up words you do not understand
    - Look up information on the internet to help you better write your question

Quizzes

Weekly quizzes are designed to check your understanding of the course material. These will be a regular part of the course and collectively factor into the Quiz grade category.

Assignments

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

- <mark>Your R Markdown file must successfully knit to PDF</mark> in a clean RStudio environment to be eligible for full credit.
- Screen captures of RMarkdown files that do not knit will receive a grade of 0%.
- Full sentences with proper grammar and punctuation are to be used throughout the submission.
    - I understand that English is a second language for the majority of students, <mark>so I may ask you to rewrite your question to make it clearer so I better understand</mark> your problem

<mark>Grades for the individual assignments will be primarily based on the correctness of your answers to each question</mark>, as well as document formatting and visualization quality. For example, do not submit a PDF containing figure with unreadable overlapping labels, code blocks that run off the edge of the page, <mark>or are 50–100 pages in length</mark> because of printing out long tables.

<mark>General questions about the assignments should be posted in the Slack channel for that week.</mark>

Midterm, Inference & Final projects

<mark>Students will complete three projects</mark> where you will perform an exploratory data analysis and statistical analysis on a dataset. More detailed information about the project is to come later in the semester.

Conduct

Academic integrity

<mark>Plagiarism, copying, cheating, sharing answers, etc. are strictly prohibited</mark>. You will be caught and reported to the University, which can result in failing the course or expulsion. <mark>All course materials and class sessions are considered Intellectual Property</mark> and therefore copyrighted by the instructor. Sharing of any recorded, copied, or other class materials without the professor's written permission is considered stealing and is STRICTLY prohibited.

Students are permitted to ask questions about the assignments on Slack and discuss assignments in private communications, however <mark>it is important to make sure that you write your assignments by yourself and in your own words</mark>, meaning that students are not permitted to collaborate on write-ups for assignments and projects. <mark>Do not duplicate another person's material or ideas and represent them as your own. It is not a sign of respect and both students will be punished.</mark> "Individual assignment" is the default classification for all assignments, exams, and projects in the course. Content that comes from a resource or another student should be properly cited.

ANY MATERIAL THAT IS TAKEN 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. <mark>This report goes to both Mason Korea and the Fairfax campus</mark>. The minimum sanctions I will recommend are:

- <mark>Minor infraction</mark> (e.g. improper citation, copying from another student's assignment):

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

Decorum/discourse

<mark>Students are expected to be civil in their conduct and respectful of their fellow classmates and the professor for the duration of the course on all discussion platforms.</mark> Students are expected to follow proper grammar and punctuation in their posted messages and to refrain from using internet slang, abbreviations, and sarcasm.

<mark>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</mark>. Penalties for decorum violations cannot be negotiated or appealed.

Mason diversity statement

<mark>George Mason University promotes a living and learning environment for outstanding growth and productivity among its students, faculty and staff.</mark> An emphasis upon diversity and inclusion throughout the campus community is essential to achieve these goals<mark>. Diversity is broadly defined to include such characteristics as, but not limited to, race, ethnicity, gender, gender identity, religion, age, disability, and sexual orientation</mark>. 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.

Disclaimer

<mark>The professor reserves the right to modify this syllabus at any time</mark> during the course to improve the learning experience and classroom environment. <mark>The professor reserves the right to adjust grades as he/she deems appropriate.</mark> The pacing of the course and the list of covered topics may also be adjusted in response to student progress.