Adv Topics II: From Data to Science Communication in R (01:185:412)

Rutgers University

Dr. Michelle Hurst

Fall 2024

Syllabus last updated on 2024-08-13.

Logistics

In-Person Lectures on Mondays and Wednesdays, 2:00pm - 3:20pm in LSH B111 Livingston Campus.

The course is on Canvas, which is where announcements, slides, and other resources will be regularly posted. We will also be on Piazza, where you can post questions and get help from both the professor and other students.

Communication with the Professor

Contact Information:

michelle.hurst@rutgers.edu

Psychology Annex A103, Busch Campus

Piazza: We will use Piazza, where students can post and answer each other's questions. This is the best place to go for everything from logistics to specific questions about lecture, as your classmates can also answer questions and they may be faster to respond than I am! I will also monitor the conversations there to acknowledge others' responses or provide responses directly.

Student Drop-In Hours: Any student can stop by my office (see above) or zoom room (link below) with questions or just to say hi during my student drop-in hours, scheduled on: Thursdays from 3:30pm to 5pm. If this time doesn't work for you, you can set up an appointment by email.

Zoom link: https://rutgers.zoom.us/j/97734172928?pwd=yaRky1pBM5M0Z3j1e9W6pFp2f1p57m.1

Email Guidelines: You can also email me. I typically respond to emails within 24-48 hours, not including the weekend. Please include "D2SC with R:" in the subject line. For example, "D2SC with R: question about assignment 3"

Description

This course is an introduction to using R, the open-source programming environment, for data organization, wrangling, visualization, and communication. We will build these skills in the context of research questions and datasets often found in the interdisciplinary field of Cognitive Science. Throughout the course, we will

highlight the relations between research questions, data structure, and decisions about data visualization and communication.

This course does not assume any prior programming knowledge or knowledge of R and has only one prerequisite: 01:185:201.

Why should a Cognitive Science student learn about data? Cognitive Science is an interdisciplinary field, bringing together research from Psychology, Computer Science, Philosophy, Linguistics, and others, with the shared goal of understanding the human mind and brain. In many cases (though not all!), Cognitive Science involves working with *data*: performance metrics of how quickly and accurately humans and other animals can perform a visual search task, characterizing children's errors when learning to count, brain regions involved in spatial navigation, and so on. Learning to appropriately and effectively work with data is an important skill.

Course Materials and Technology

Classes will be interactive lectures. My expectation is that you attend class and will have a laptop computer to use throughout the lecture. If you are unlikely to have a computer to use, let me know and we'll find a solution.

If you do not have the appropriate technology for financial reasons, please email the Dean of Students at deanofstudents@echo.rutgers.edu for assistance. If you are facing other financial hardships, please visit the Office of Financial Aid: https://financialaid.rutgers.edu/

All software and readings will be freely available to you.

Primary Readings

Most of our reading will come from:

R for Data Science, 1st Edition

Hadley Wickham & Garrett Grolemund. O'Reilly Media. 978-1491910399

It is freely available here: https://r4ds.had.co.nz/index.html

In the schedule, readings marked as coming from "R4DS" are from this book.

Other readings and practice assignments will be posted on canvas throughout the course, also from freely available sources - and these will be posted with a direct link.

Software and Tools

All software is freely available, with some additional free upgrades because of your educational affiliation.

R: https://www.r-project.org/

RStudio: https://posit.co/download/rstudio-desktop/

Github (https://github.com/), with a free educational account (which you can request here: https://education.github.com/)

If you're not familiar with git, github, or terminal, then you'll also need an IDE for working with github, such as Github Desktop (https://desktop.github.com/)

Piazza: https://piazza.com/rutgers/fall2024/cogsci412

Canvas: https://rutgers.instructure.com/courses/286665

Learning Goals

The learning goals of this course are meant to align with the learning goals of the Cognitive Science undergraduate major, which you can learn more about here: https://ruccs.rutgers.edu/academics/undergraduate/learning-goals

After fully participating in this course you will:

. . . build skills: students will learn how to use the programming environment R, including tidyverse and R Markdown, to wrangle, visualize, and communicate data.

. . . build knowledge: students will learn about common data structures in Cognitive Science, and how they relate to research questions.

Overall, by the end of the course, students will have the skills and knowledge necessary to move Cognitive Science projects from a research question to data communication using the R programming environment. Specifically, all students will learn principles and tools for data management, wrangling, visualization, and communication, with a focus on the open source programming language R and version control using GitHub. This will result in a fully reproducible data communication project. For students doing a thesis project, this project will be an initial draft of your thesis.

Grading

Grades will be assigned using the standard scale:

- 90 - 100 = A
- 85 - 89.99 = B+
- 80 - 84.99 = B
- 75 - 79.99 = C+
- 70 - 74.99 = C
- 60 - 69.99 = D
- < 60 = F

Assessments

Your final grade will be calculated based on the following components:

- Weekly Assignments = 54
- In Class Mini Activities (ICMAs) = 12
- Piazza Engagement = 6
- Final Project = 28

Weekly Assignments (54%)

What are they? There will be 11 weekly assignments, which are small problem sets specifically geared toward practicing the content from that week. They will involve working with real datasets in R. The exact number of problems and size of problems will vary depending on the content of that week. You can work with your classmates, ask questions on piazza, or use search engines to help solve the problem. However, at submission time, the code and supporting information you include must be *your work*. You cannot use ChatGPT (or similar) to generate code/responses (see Academic Integrity section).

How are they graded? The lowest TWO assignments will be dropped, and the remaining NINE assignments will each be worth 6% of your overall grade. Each assignment might have a different number of points, depending on the specific problems. All assignments will be graded based on the accuracy, implementation,

and readability of your code and on your "plain English" response to the prompt (for some problems). A full rubric is provided on Canvas, and deviations from it for specific assignments will be provided in the assignment instructions.

Late Policies: You can use SIX (6) late days TOTAL on Weekly Assignments. For example, you could submit one assignment six days late, three assignments each two days late, etc. Any late days beyond the six provided above will be a 10% deduction per day. Assignments will NOT be accepted more than 6-days late (i.e., they will not be accepted after Thursday at midnight the week after they are due).

Why are they assigned? The best way to learn a programming language (or really, any language!) is to practice. The weekly assignments are meant to give you substantial structured practice using the skills and knowledge we learn in class. It also provides a way for me to assess your learning, both to assign grades and so that I can quickly adapt my teaching in class if needed.

In Class Mini Activities (12%)

What are they? In each lecture, there will occasionally be a "mini problem" that we work on together and I expect you to follow along. You will do these activities in one continuous **R Notebook File**, which you will **push** back to github at the end of each class with "ICMA [date]" in the commit message (more on all of this the first week of class!).

How are they graded? These are "graded" based on doing/trying them (i.e., submitting the R file to github) - so being wrong or incomplete is OK. If you miss a class, forget your laptop, or for any other reason aren't able to actually do and/or submit the activities *during class*, you must still submit the in class activities before the beginning of the next class. These late activities will still be given full credit. If you are missing class for an extended time and you also cannot keep up with the ICMAs (e.g., serious illness), let me know and we can work together to set a date by which you need to catch up.

Every class will have an ICMA. You only need to submit 24 of them (out of 28) to get full credit, which means you can skip 4 of them entirely. Any additional missing assignment will be a deduction of 0.5 points each.

Why are they assigned? Learning the skills and knowledge for this class will be easier and faster with practice. The ICMAs are designed to maintain engagement during class and provide hands on practice with every lecture. Even though you can skip 4 of them, I highly recommend you plan on doing all of them!

Piazza Engagement (6%)

What is this? We will use Piazza for Q&A about course material, including assignments and other assessments. You can register here: https://piazza.com/rutgers/fall2024/cogsci412. In Piazza, you can post a question, comment, or response to someone else.

How are they graded? These are graded based on completion. Responses to others will be worth 2pts and posting new questions or comments will be worth 1pt. You can earn up to 6pts - though I recommend you engage beyond that as well, to support your and each others' learning! Comments that do not provide additional meaningful information (e.g., saying "me too", "I don't know") will not receive points. Questions that can be answerable from the syllabus or canvas (e.g., when is assignment 3 due) will also not receive points. You can also post anonymously, but only identifiable posts will count toward the grade.

Why is this graded? To encourage asking and answering questions to/from your peers. It also provides a consistent way for me to address questions, so that all students can see the answers and you can benefit from your fellow classmates asking questions. Questions/comments can be specific to course material (e.g., about the lecture information, ICMAs, assignment problems) or about R and related course content more generally.

Final Project (28%)

What is this? Throughout the course, we will be building up skills with the goal of creating a fully reproducible scientific research project. The research project must include coming up with a research question, answering that question using data, and communicating the results of that research question through visuals and writing in a reproducible R Markdown document. The actual form of the research project can differ depending on your goals, including a traditional written paper (recommended for thesis students to be the first draft of your thesis project), a scientific poster, an infographic, or a shinyapp. The research question must fit within the broad discipline of Cognitive Science, but is otherwise up to you. You can use primary data (e.g., your thesis, from a lab you're working in) or secondary data (i.e., an openly available dataset found online). Regardless of the format, all projects will be graded using the same rubric.

Additionally, there are two other components you must complete:

1. You must submit a written proposal and schedule a 15 minute meeting with Prof Hurst to discuss your planned project. The written proposal should include a paragraph or two describing: your research question, data source, and planned format. During the meeting we will discuss your plan, including any potential road blocks or contingency plans. If we identified any necessary changes or next steps during the meeting, then you must resubmit your written proposal outlining your updated plans within 2 weeks of our meeting. If you decide later to deviate from this proposal, you must talk to Prof Hurst. If you do not, you risk your project being deemed inappropriate for the course.
2. You must do a 5-minute mini presentation in class that tells us your research question, data source/method, and one plot/graph illustrating something about your data. These presentations will take place over the last three days of class.

How is this graded? Your final project (20%) will be graded based on: the effectiveness and readability of your code, your use of the R Markdown document structure, your data visualization and communication choices, and how you describe your research question, motivation, and interpretation of the results. A full rubric is posted on Canvas. The final project is due on our scheduled exam date (instead of a formal exam), which is TBD by Rutgers. This cannot be submitted late except for extreme circumstances, as late submissions may delay your final grades.

The project proposal (due between Monday Oct 14 - Monday Nov 4, at least one hour before our scheduled meeting time) will be graded based on completion only (2%). Importantly, this is a check on the appropriateness and feasibility of your project plan. If you do not complete it, you risk having your final project be deemed inappropriate and receiving a 0. If submitted between Nov 4 and Nov 11, the maximum you will receive is 1pt. If submitted after Nov 11, you'll receive 0 points.

The presentation (6%) must be submitted to Canvas by Tuesday Dec 3 and you will present to the class between Dec 4-11th. This is a short presentation giving a high-level overview of your project in three slides. We will go over the precise structure of what's required in class. It will be graded based on clarity of your materials, overall structure, and the code used to generate the slides/content. If you submit your presentation on Canvas between Dec 4 - 11, you'll receive a 0.5 pt deduction per day. Submitting after Dec 11 will result in a 0.

Course Schedule

A full course schedule is provided at the end of the syllabus. The course will generally be organized into sections that will cover various aspects of the R programming environment. Some classes will provide more information and other classes will be about applying that information in real data settings, while including additional content about specific use cases or variations. These classes are noted on the schedule as "data exercises" and will use a real dataset from Cognitive Science, with a leaning toward a specific interdisciplinary subfield: Psychology, Computer Science, or Linguistics. Note that these classes will still provide substantial *new* information and are not "just" work sessions. The difference is that the new information will be fully

embedded in a real-life Cognitive Science research problem, so that we can see the progression from beginning to end.

Policies

Attendance and Participation

I expect you to attend class, in accordance with Rutgers' in-person policies. Participation with class activities (via ICMAs, described above) will count toward your grade. However, I have already built in ways to get full credit for missed classes, without needing to provide an explanation (see details in the In Class Mini Activity (12%) section above). If you are sick, please stay home. I do not need to know why you miss a class, but letting me know when you will be absent is helpful and appreciated. When possible, you can report absences using the University self-reporting absence system: https://sims.rutgers.edu/ssra/

What if I get really behind or things go wrong? If things happen in your life that make it difficult to submit course assessments within the guidelines above, let's talk and work together to make a plan. You can also use the resources detailed at the end of the syllabus to support your academic and personal well-being at Rutgers.

What if I add the course after the semester starts? Talk to me as soon as possible and no later than Friday Sept 13 so that we can arrange for you to catch up, without penalty. If you do not reach out to me by Sept 13th, your missed assignments will incur a late penalty.

Disability Accommodation

Rutgers University is committed to the creation of an inclusive and safe learning environment for all students, and welcomes students with disabilities into all the University's educational programs. The Office of Disability Services (ODS) is responsible for the determination of appropriate accommodations for students who encounter barriers due to disability. Once a student has completed the ODS process (registration, initial appointment, and submitted documentation) and reasonable accommodations are determined to be necessary and appropriate, a Letter of Accommodation (LOA) can be requested and will be sent to the you (the student) and me (the instructor). This should be done as early in the semester as possible as accommodations are not retroactive. Once I receive the LOA, we will have a discussion about how the accommodations will be implemented. More information can be found at www.ods.rutgers.edu. You can contact ODS at (848)445-6800 or via email at dsoffice@echo.rutgers.edu.

Civility and Community

Learning a new (programming) language can be hard, and at times alternates between frustrating and exciting. I want students to embrace this challenge in a supportive and thoughtful community. To do that, I expect that all students will be courteous, supportive, and thoughtful to other students - both in person and online. Responses and other communication among each other (e.g., in class, on Piazza):

- should not be demeaning or rude
- should not negatively refer to a person's ability to succeed in the course (you also should be kind to yourself about this!!)
- should not be racist, sexist, ableist, or hateful in any other way
- should give each other the benefit of the doubt
- should be kind
- should celebrate small and big coding wins!

Academic Integrity

Rutgers University takes academic dishonesty very seriously. By enrolling in this course, you assume responsibility for familiarizing yourself with the Academic Integrity Policy and the possible penalties (including suspension and expulsion) for violating the policy. As per the policy, all suspected violations will be reported to the Office of Student Conduct. Academic dishonesty includes (but is not limited to):

- Cheating
- Plagiarism
- Aiding others in committing a violation or allowing others to use your work
- Failure to cite sources correctly
- Fabrication
- Using another person's ideas or words without attribution, including re-using a previous assignment
- Unauthorized collaboration
- Sabotaging another student's work

If you are ever in doubt, consult with me (your instructor)

Note that for the purposes of this class, you *can and should* use search engines and trusted websites (e.g., stackoverflow) to help work through your assignments, work together in small groups, or ask questions on Piazza when stuck. However, all work submitted must be your own. I recommend you start working on a problem first, and turn to the internet or a friend when you get stuck.

You **cannot use ChatGPT**, or similar systems, to generate code for you. **Why not?** My goal, and I assume your goal, is that you *learn* in this course. Although ChatGPT, and similar, *might* get you a correct answer very quickly, you are less likely to learn when a solution is generated for you. Also - it may not generate the right answer! When you already know the programming language, ChatGPT might speed up your coding. But it is not usually a good tool for learning.

Student Support and Mental Wellness

- Student Success Essentials: https://success.rutgers.edu
- Student Support Services: https://www.rutgers.edu/academics/student-support

- The Learning Centers: https://rlc.rutgers.edu/
- Rutgers Libraries: https://www.libraries.rutgers.edu/
- Bias Incident Reporting: https://studentaffairs.rutgers.edu/bias-incident-reporting
- Dean of Students – Student Support Office: https://success.rutgers.edu/resource/dean-students-student-support-office
- Office of Veteran and Military Programs and Services: https://veterans.rutgers.edu
- Student Health Services: http://health.rutgers.edu/
- Counseling, Alcohol and Other Drug Assistance Program & Psychiatric Services (CAPS): http://health.rutgers.edu/medical-counseling-services/counseling/
- UWill: free immediate access to teletherapy; you can choose a therapist based on your preferences including issue, gender, language, ethnicity. http://health.rutgers.edu/uwill/
- Office for Violence Prevention and Victim Assistance: www.vpva.rutgers.edu/
- Office of Disability Services: https://ods.rutgers.edu/

- Basic Needs Assistance (food, housing, and other essentials): https://ruoffcampus.rutgers.edu/basic-needs
- Rutgers Student Food Pantry: https://ruoffcampus.rutgers.edu/food-pantry

Full Schedule (subject to change!)

week	date	Topic	Readings	Assessments
1	Wed Sep 04	intro & installs		
2	Mon Sep 09	git, github, R ecosystem	https://intro2r.com/chap1.html	
2	Wed Sep 11	R basics	https://intro2r.com/objects-in-r.html https://intro2r.com/help.html https://style.tidyverse.org/index.html (Chpt 1, 2, 3)	WA#1 due Fri
3	Mon Sep 16	readr and data types	R4DS Chp 12 (Tidy data, up to 12.3), 18 (Pipes)	
3	Wed Sep 18	dplyr	R4DS Chp 5 (Data transformation)	WA#2 due Fri
4	Mon Sep 23	tidyr	R4DS Chp 12 (Tidy data, 12.3 +)	
4	Wed Sep 25	ggplot2 & intro to data visualization	R4DS Chp 3 (Data visualization)	WA#3 due Fri
5	Mon Sep 30	data exercise 1: Decision Making	https://intro2r.com/data-types.html	
5	Wed Oct 02	data exercise 1: Decision Making	R4DS Chp 16 (Strings)	WA#4 due Fri
6	Mon Oct 07	back to basics: data structures, missings, logic	https://intro2r.com/data-structures.html https://intro2r.com/data-types.html	
6	Wed Oct 09	combining data	R4DS Chp 13 (Relational Data)	WA#5 due Fri
7	Mon Oct 14	iteration	R4DS Chp 21 (Iteration)	Proposal Window Opens
7	Wed Oct 16	iteration		WA#6 due Fri
8	Mon Oct 21	data exercise 2: Behavioral Neuroscience	TBD	
8	Wed Oct 23	data exercise 2: Behavioral Neuroscience		WA#7 due Fri
9	Mon Oct 28	Communicating Research	TBD	
9	Wed Oct 30	R Markdown	R4DS Chp 27	WA#8 due Fri
10	Mon Nov 04	R Markdown	R4DS Chp 29	Proposal Window Closes
10	Wed Nov 06	APA papers & papaja	TBD	WA#9 due Fri
11	Mon Nov 11	data exercise 3: Linguistics		
11	Wed Nov 13	data exercise 3: Linguistics		WA#10 due Fri
12	Mon Nov 18	data exercise 3: Linguistics		
12	Wed Nov 20	Infographics, Dashboards, & Shiny Apps	TBD	WA#11 due Fri
13	Mon Nov 25	Infographics, Dashboards, & Shiny Apps	TBD	
13	Wed Nov 27	NO CLASS (adjusted class days)		
14	Mon Dec 02	catch up + missed topics	TBD	Presentation File Due (Dec 3)
14	Wed Dec 04	Project Presentations		
15	Mon Dec 09	Project Presentations		
15	Wed Dec 11	Project Presentations		
TBD				Final Project Due