SPE489 Computational Tools for Social Science

2+1 Hybrid Course 5:50-7:40 PM PT Thursday Evenings
Classroom: Stauffer 106
Fall 2021

Instructor. Tyler Reny tyler.reny@cgu.edu (more info here!)

Your TAs. Sean Merritt sean.merritt@cgu.edu, Abdullah Alswelh (Abdullah.Alswelh@cgu.edu) and Fawaz Alotaibi

Office Hours. TBD

Structure. 1 hour asynchronous + 2 hours synchronous/online hybrid.

Instructor Feedback and Communication. The best way to get in touch with me is via email tyler.reny@cgu.edu. I will do my best to respond to your messages within 1 business day.

Class Objectives and Student Learning Outcomes. The research aims of scholars in social sciences are diverse, as are approaches to answering research questions, but the field increasingly requires the ability to use *data* to bolster claims about how the social world works. This course will introduce you to the computational tools needed to collect, clean, and analyze data and provide you the foundation needed to further develop your skills as a programmer and social scientist. By the end of the course, students should be able to use R programming language to gather data from a variety of different sources, clean and shape the data to facilitate statistical analysis, and present the raw data and outcomes from any statistical analyses using effective visualizations.

Credits. 4

Class Requirements

1) *Attendance.* Students are expected to show up (in person or virtually), attend class, and participate.
2) *Laptop*. Bring a laptop with internet access to each class. If you plan to attend in person and do not have a working laptop with wireless internet capabilities, please contact me.
3) *Videos*. Watch videos before each in-person class. These videos are intended to give you the necessary context and scaffolding to understand the R code you'll be executing in class. The material I cover in the in-person lectures/labs will build on material in these videos.
4) *Problem sets*.
    a. Problem sets will be completed in R Markdown and turned in as a compiled HTML document
    b. I will not accept late problem sets (though there are always exceptions!)
    c. You are encouraged to discuss problems sets with other students in the class. However, each problem set should be produced by the individual student and the work written up independently. If you do work with others, please include the

names of other students you worked with on the problem set. I strongly suggest you use the problem sets as an opportunity to ensure that *you* are understanding the concepts that you are learning in this course as we go. This course builds on itself every week and if you fall behind and don't catch up, you will struggle exponentially more as the course progresses. As such, I suggest you try the problem sets on your own first, then meet up with or discuss where you've gotten stuck with other students second.

   d. Problem sets will be simply graded with a check plus (you tried hard and got most of the answer correct), check (you finished the problem set but you didn't try very hard or weren't very thorough), and check minus (you turned in an incomplete or did not turn in a problem set). I will always provide on Canvas an answer key that you should use to check your own answers and see how I would approach the different problems. Remember that there are many ways to get to the "right" answer in R! I will make every effort to grade and return problem sets within 1 week of their due dates.

5) *Final project*. Students should plan to submit either an a) research memo; b) research paper; or c) research poster at the end of the semester that uses the skills used in this class to answer a research question of interest to you. You will need to turn in the executable code used for the project that I should be able to fully replicate on my computer. Students are free to choose any topic they would like, though I encourage you to use this opportunity to look at data that could help you work toward a research paper, thesis, or dissertation! Take this opportunity to receive structured feedback on something that will help you through the program. Key progress dates are below.

Grading. Grades will be based on the problem sets (50%), participation (5%) and the final project (45%). Letter grades of A will indicate mastery of course material. Letter grades of B will indicate proficiency with course material. Letter grades of C will indicate gaps in proficiency with course material and will indicate that the student work was not at a level expected of the program.

Course Website. The course website is located on Canvas. This is where I will distribute course materials, including video, lecture slides, and problem sets, etc. All announcements will be made through Canvas. You are also encouraged to use the discussion board to discuss problems, questions, concerns, etc., with other students in the class. I will monitor this and chime in when needed. I encourage students to reply to each other's questions, and a student's respectful and constructive participation on Canvas will count toward his/her class participation grade.

Prerequisites. This course assumes no familiarity with programming. We will start from basics! That being said, if you have never been exposed to the R programming language, I would encourage you to work through some very basic coding exercises at one or more of the following locations:

- DataCamp Introduction to R course "Intro to Basics" lesson is free https://www.datacamp.com/courses/free-introduction-to-r
- CodeAcademy (some early lessons might be free) https://www.codecademy.com/learn/learn-r

- Work through the first few lessons here: https://stat545.com/r-basics.html

Getting Help

Rstudio Support: https://support.rstudio.com/hc/en-us
RStudio Webinars https://rstudio.com/resources/webinars/

Class Policies

The CGU institutional policies apply to each class offered at CGU. Students are encouraged to review the student handbook for their program as well as policy documentation in the Bulletin and on the Registrar's webpages: http://bulletin.cgu.edu/ and http://www.cgu.edu/registrar. The protocols defined by the CGU's Student Conduct Code must be upheld in all classes. For more information, please visit for CGU's Basic Code of Conduct (Links to an external site.).

Credit Hour

Credit hour refers to the units or credits earned by a student for the successful completion of a course at CGU. These are the units recorded on transcripts and the units that are counted toward degree requirements. For CGU courses, a single unit or credit is determined by 10.5 hours of instructional activity per semester. Instructional activity includes direct instructor contact in a physical or virtual classroom as well as asynchronous instructional content for online or hybrid courses. See the full policy here.

Academic Integrity

The work you do in this class must be your own. Information on CGU's Policy and Procedures for Violations of Standards of Academic Integrity can be found at: https://cgu.policystat.com/policy/3903703/latest/. In addition, the Claremont Colleges Library has a number of resources on academic honesty and integrity, including the following online tutorial: https://library.claremont.edu/exploring-academic-integrity/

Standards of Appropriate Behavior for Online Courses

Web based programs define the area in view of the camera as part of the classroom. Students must therefore present themselves and their surroundings as much as possible as though they are in a physical classroom. Attending class remotely, you can easily forget that your home environment is fully visible and can be distracting to you and your classmates. Here are some guidelines to help you to be more intentional and effective in how you attend class online:

Appropriate Dress: Even though you are at home, dress in casual professional attire for class so that you present yourself appropriately to your instructor and peers on camera. Remember, there is no back row in a video meeting. What you wear is amplified and speaks loudly!

Location: Attend class in a location that allows for your full attention and participation. Aim to be on-camera in an environment free of auditory or visual distractions, and that allows for open,

focused participation in class discussion. While home situations differ and some interruptions might be unavoidable, aim for the following:

- Be mindful of your background. Your camera captures you and anything in your background, so find a background that is appropriate to be viewed by your classmates. You can always set up a virtual background in Zoom that will hide the real background in your location.

- Stay on camera – leaving the camera frequently or for extended periods of time disengages you and others from class work.

- Make prior arrangements with family members to give you dedicated uninterrupted space and time for the duration of each class meeting. Avoid public locations, especially locations that are noisy and distracting or where you cannot freely speak to engage in discussions.

- Stay focused. Avoid the following to the extent you are able: Interacting with persons not part of the class, behaving in an overly inattentive manner, multi-tasking with non-class related work.

We understand that you may have challenges with caregiving or finding a distraction-free space for the duration of each class meeting. Please speak with me about this so that we can work out ways to optimize your engagement and reduce interruption and distractions.

Accommodations

Accommodations for Students with Different Abilities. CGU is committed to creating courses that are inclusive and accessible. If you would like to request academic accommodations due to temporary or permanent disability, contact the CGU Dean of Students and Coordinator for Student Disability Services at DisabilityServices@cgu.edu or (909) 607-9448. Reasonable accommodations are considered after you have conferred with the Office of Disability Services (ODS) and presented the required documentation of your disability to the ODS. Planning is essential, so please communicate to the ODS as soon as possible.

Religious Accommodations. Students who expect to miss classes or assignments as a consequence of their religious observance shall be provided with a reasonable alternative opportunity to complete such academic responsibilities.

Mental Health and Well Being. Graduate school is a context where mental health struggles can arise or be exacerbated. If you ever find yourself struggling, please ask for help. If you wish to seek out campus resources, here is some basic information: services.claremont.edu/mcaps/

Monsour Counseling and Psychological Services (MCAPS) is committed to promoting psychological wellness for all students at The Claremont Colleges. Professional and well-trained psychologists, psychiatrists, and post-doctoral and intern therapists offer support for a range of psychological issues in a confidential and safe environment.

Phone (909) 621-8202
After hours emergency (909) 607-2000
Tranquada Student Services Center, 1st floor
757 College Way
Claremont, CA 91711

Title IX. One of my responsibilities as an instructor is to help create a safe learning environment. I am a mandatory reporter. Thus, if I learn of any potential violation of CGU's gender-based misconduct policy (e.g., rape, sexual assault, dating violence, domestic violence, or stalking) by any means, I am required to notify the CGU Title IX Coordinator at Deanof.Students@cgu.edu or (909) 607-9448. Students can request confidentiality from the institution, which I will communicate to the Title IX Coordinator. If students want to speak with someone confidentially, the following resources are available on and off campus: EmPOWER Center (909) 607-2689, Monsour Counseling and Psychological Services (909) 621-8202, and The Chaplains of The Claremont Colleges (909) 621-8685. Speaking with a confidential resource does not preclude students from making a formal report to the Title IX Coordinator if and when they are ready. Confidential resources can walk students through all of their reporting options. They can also provide students with information and assistance in accessing academic, medical, and other support services they may need.

Your Physical Health. I am also committed to ensuring the health and safety of the CGU community. Information on CGU's COVID Fall Semester protocol can be found here https://info.cgu.edu/emergency/fall-2021-semester/ and I suspect it will be updated as we progress through the Fall.

Campus security. Campus security can be reached 24 hours/day at (909) 607-2000. Please download the CGU Safety Resource Card to your phone's contacts.

Tech Issues. The Office of Information Technology has a helpdesk to support you with CGU wireless access and email issues. They also have good documentation you can use to learn to connect and use online resources. **Website**: https://mycampus.cgu.edu/web/it

Center for Writing and Rhetoric (CWR). CGU has a graduate studies-focused Center for Writing and Rhetoric that works with you no matter where you are in the writing process. The CWR is not just for remediation of your writing, but for all writers to provide partnership and consultation to improve your academic work at all levels. The CWR can work with you in planning, outlining, drafting, and final review of documents and presentations for class work, conferences, and publications. **Website**: https://mycampus.cgu.edu/web/writing-rhetoric

Library. The Claremont Colleges Library has a wealth of resources, including subject specialist librarians, to support your academic work. Use the library for class work and research to access and use data-bases for articles, books, and data sets, to understand how to conduct effective searches and evaluate sources, use digital tools, and much more. The library offers workshops and 1-1 consultations with students as well. **Website:** https://library.claremont.edu/

Development of Course. Finally, this course draws heavily on the work of many others who have developed similar courses before me. Special thanks to Erin Hartmann and Jesse Acevedo for sharing materials. Substantial portions of this course are also modeled on courses by Jenny Bryan, Rachel Bernhard, Jae Yeon Kim, and many others.

Schedule. This is a preliminary schedule of the course. It is very likely that some of this will shift if we progress more quickly or slowly through the material than initially anticipated! I will update this document on Canvas and will alert you of changes when they are made.

Date	Before Class Video	In Class	Assignments Due
Class 1 - Sept 2	Intro, syllabus, expectations, installing R and R Studio, Canvas	Basics of R, getting help	1. Watch asynchronous video 2. Read syllabus 3. Download and install R and R Studio 4. Download LaTeX distribution 5. Fill out survey 6. Bring laptop to class! *References* Concepts: Wickham (Chpt 1); R and R Studio Installation Bryan *This will be updated each week depending on course progress*
Class 2 - Sept 9	Thinking through Tidy Data, data formats, intro to Dplyr packages and the Tidyverse, reproducibility, rules for troubleshooting	R Markdown lesson for PS1, packages and libraries Reading/Writing Data, Intro to dplyr 1	1. Watch asynchronous video 2. R Practice Problem Set *Readings* 1.Wickham; 2. Tidy Data (1.1-1.5); 3. Ioannidis; 4. Review R Markdown materials here *References* Stat545; Wickham (Chpt 27); Wright et al (Chpt 1 & 2); Soltoff; Soltoff 2; Soltoff 3
Class 3 - Sept 16	Obstacles to learning; Tips and Tricks for Data Exploration	dplyr 2: Exploring and Wrangling Data (long vs wide); intro to graphics	1. Watch asynchronous video 2. Submit idea for final project as word doc *References*

			Wickham (Chpt 7); Wright et al (Chpt 3)
Class 4 - Sept 23	Data formats 1	Numeric, Factors, Character Strings	1. Watch asynchronous video 2. PS1 due *References* Wickham (Chpt 14-15); Wright et al (Chpt 3.5, 3.7)
Class 5 - Sept 30 (Tyler away at conference)		**No Class. Virtual office hours. Potential lab with TA.**	
Class 6 - Oct 7	Data formats 2	Dates and times, Merging data (manipulating character strings)	1. Watch asynchronous video 2. Check in with Tyler progress on final project *References* Wickham (Chpt 16); Wright et al (Chpt 3.6); Wright et al (Chpt 2.10.4)
Class 7 - Oct 14	Repetitive Tasks in R 1	Loops (Bootstrapping)	1. Watch asynchronous video 2. PS2 due *References* Wickham (Chpt 21); Wright et al (Chpt 3.9); Soltoff
Class 8 - Oct 21	Repetitive Tasks in R 2	Functions	1. Watch asynchronous video 2. Check in with Tyler progress on final project (number 2) *References* Wickham (Chpt 21); Wright et al (Chpt 3.9); Soltoff
Class 9 - Oct 28	Why simulate data? Intro to regression in R	Simulation and basic regression analysis in R	1. Watch asynchronous video 2. PS3 due

			References Wickham (Chpts 22-25); Wright et al (Chpt 5)
Class 10 - Nov 4	Fundamentals of Effective Data Visualization	Basic R Graphics	1. Watch asynchronous video 2. PS4 due *Readings* Healy (Chpt 1); Bryan *References* Wickham (Chpt 28); Wright et al (Chpt 4); The rest of Healy's Book; Soltoff
Class 11 - Nov 11	Data visualization 2	Advanced R Graphics	1. Watch asynchronous video *References* Colors in R Bryan; The rest of Healy's Book; Wright et al (Chpt 4)
Class 12 - Nov 18	Collecting Data from the Internet	Web Scraping: APIs + Parsing websites	1. Watch asynchronous video 2. PS5 due *References* Wright et al (Chpt 2.11); Soltoff; Soltoff 2
No Class Nov 25			
Class 13 - Dec 2	Fundamentals of Mapping	Spatial Data in R with sf package	1. Watch asynchronous video 2. Check in with Tyler progress on final project (number 3) Readings Healy (Chpt 7)
Class 14 - Dec 9	Fundamentals of Text as Data	Text as Data	1. Watch asynchronous video 2. PS6 due

			Reading Silge and Robinson (Chpt 1) *References* Silge and Robinson (rest of book); Wright et al (Chpt 3.8); Soltoff 1; Soltoff 2
Class 15 - Dec 16	TBD	TBD: Machine Learning or other topics of interest	1. Watch asynchronous video *References* Hill; Soltoff
December 18			*Final Project Due*

Diversity and Inclusion. There are a variety of resources available both at CGU as well as online. Below is just a sampling of groups

- R LGBTQ Twitter: Affinity group for LGBTQ people in the R community – Twitter often promotes events, panels and talks by and for LGBTQ R users.
- Gayta Science Twitter: Alliance that uses data science techniques to give LGBTQ+ experiences a voice – Twitter will often share data-driven work concerning the LGBTQ+ community.
- RLadies Community Slack: A global programming meetup for non-binary, trans, and female R users.
- R Ladies LA: Local LA chapter of R Ladies – hosts events, talks, and projects in the LA area.
- People of Color Code Meetup: A meetup for POC software developers – has events where POC developers can work on personal projects, collaborate, and learn.
- R Forwards: A task force set up by the R Foundation to address the under-representation of under-represented groups in the R community – collects representation data in the R community, produces workshops and teaching materials
- R Community Diversity and Inclusion Working Group: Working group set up by the R Consortium to encourage and support diversity and inclusion across a variety of events and platforms in the R community

Additional Reference Materials

R Books

R for Data Science https://r4ds.had.co.nz/
Advanced R http://adv-r.had.co.nz/
Data Visualization https://socviz.co/index.html#preface
Data Visualization https://clauswilke.com/dataviz/
Text Mining https://www.tidytextmining.com/
Data Analysis and Prediction in R https://rafalab.github.io/dsbook/
Modern Data Science with R https://mdsr-book.github.io/mdsr2e/index.html

Cheatsheets

Multiple R topics: https://rstudio.com/resources/cheatsheets/
Packages: https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages

Style Guides

Rstudio Style guide: http://adv-r.had.co.nz/Style.html
Google's R Style Guide: https://google.github.io/styleguide/Rguide.html
Tidyverse style guide: http://style.tidyverse.org/

Fundamentals of programming

Don’t Repeat Yourself
http://web.archive.org/web/20131204221336/http://programmer.97things.oreilly.com/wiki/index.php/Don%27t_Repeat_Yourself

Open Science

Nosek and friends “How Open Science Helps Researchers Succeed”
https://elifesciences.org/articles/16800.pdf
Nosek et al Promoting an Open Research Culture
https://science.sciencemag.org/content/sci/348/6242/1422.full.pdf?casa_token=762GJ4WpZVUAAAAA:BNiPtqK6xD9D3AZ5dTVBOyZwibkSO3oKWtAt6HNxbmVF1c3wOy1T1MlB7IIX_1lGGiBMfrdPHIKLxvs
https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf

Other Courses

https://github.com/ribernhard/PS239T
https://cfss.uchicago.edu/
https://qss.princeton.press/