Course Syllabus – CBB 410

Course Information

Course Number: CBB 410 FA24
Course Name: Introduction to Data Science
Term: FA 2024
Start Date: 09/04/2024
End Date: 12/06/2024
Credits: 3.0

Meeting Days / Times

Mondays and Wednesdays, 11:30am-1:00pm PT / 2:30-4:00pm ET
(See Calendar in Canvas for the most up-to-date schedule.)

Aside from the first week, Mondays will be in person lectures and Wednesdays will be hands-on virtual classes focused on coding.

Location

CA: Graduate Office Dining Room (Hazen Theory Building)
(Exception: Wed. September 4 - Grad Office Seminar Room)
FL: C304

Course Managers

Role	Last Name	First Name	Email Address
Course Director	Bagsic	Samantha	bagsic.samantha@scrippshealth.org
TA	Kogachi	Leticia	lkogachi@scripps.edu
TA	Ng	Karenna	kang@scripps.edu

Course Description

This course will provide students with training on general data analysis techniques. It expands upon the computational boot camps that have previously been offered to matriculating students. Students will learn how data is organized and common approaches to presenting data, coding, and other tools to aid them in their own analyses. The course structure will include didactic lectures on select topics intermixed with hands-on applications of these principles.
Topics covered:

- Data organization, visualization, and basic statistics
- Pseudo-coding, introduction to R, and data manipulation using R
- Building models and creating journal-quality figures
- Introduction to advanced methodologies

Program Learning Outcomes

By the end of the program, students will have accomplished these objectives:
PLO1: Original Research – graduate students are expected to develop the skills critical for generating high-quality research output. This would include absorbing, recalling, and contextualizing scientific knowledge, evaluating scientific information and data, creating testable hypotheses and investigating hypotheses, mastering scientific tools and techniques, displaying ethical behavior, and receiving and giving feedback.
PLO2: Communication – graduate students are expected to demonstrate the oral, written, and media skills to effectively communicate the impact of a study or a body of work to the greater scientific community and to the public at large using a number of methods.
PLO3: Critical Thinking – graduate students are expected to develop a self-directed process to analyze information, form opinions or judgments, and use this process to improve the quality of their scientific thoughts, navigate problems, and make informed decisions.
PLO4: Intellectual Curiosity – graduate students are expected to acquire the capacity to build their intellectual curiosity and demonstrate problem solving approaches that serve their professional growth and ability to impact a field.
PLO5: Career and Professional Development – graduate students are expected to develop a variety of transferable skillsets throughout their graduate experience, including management and leadership, inclusiveness, resilience, scientific rigor, collaboration, accountability, time management, teamwork, networking, and career planning.

For a detailed description of each outcome and specific success indicators, please refer to this web page: https://education.scripps.edu/graduate/doctoral-program/.

Course Learning Outcomes

Upon completion of this course students will be able to:
CLO1: Transform and join multiple datasets in preparation for analyses
CLO2: Write efficient code and functions in R
CLO3: Create publication-quality figures and tables to describe datasets
CLO4: Build and communicate simple statistical models

Background Preparation (Prerequisites)

This course does not assume any prior coding/programming or statistics knowledge and is designed for beginners in data science. No prerequisites are required.

Course Materials

This course will follow the text and modules from *R for Data Science* by Garrett Grolemund & Hadley Wickham. This text is available for free online but may be purchased in hard copy if desired.

Welcome | R for Data Science (had.co.nz)

Class Format

Aside from the first week, which will be in person on Wednesday September 6 for the course intro, Mondays will be in-person lectures focused on introducing each topic, and Wednesdays will be virtual hands-on sessions where we will run code to illustrate the week's topic together via zoom. The final week of the course will be in person both days for final presentations.

Expectations and Logistics

The tentative topics and schedule of the course is described in the table below. There will be two homework assignments, a midterm project, and a final project. Final projects will be presented to the class during the final week of the term.

Week:	Day:	Date:	Topic	Location
Week 1	Wednesday	9/4/2024	Course Introduction: R, R Studio, and R Markdown	In Person
Week 2	Monday	9/9/2024	Data Visualization	In Person
	Wednesday	9/11/2024	Using ggplot2 to visualize data	Zoom
Week 3	Monday	9/16/2024	Data Transformation	In Person
	Wednesday	9/18/2024	Using dplyr to transform datasets	Zoom
Week 4	Monday	9/23/2024	Exploratory data analysis basics	In Person
	Wednesday	9/25/2024	Combining data visualization and transformation tools	Zoom
Week 5	Monday	9/30/2024	Data structures and imports	In Person
	Wednesday	10/2/2024	Using tibble and readr to efficiently work with data sets	Zoom
Week 6	Monday	10/7/2024	Tidy data	In Person
	Wednesday	10/9/2024	Using tidyr to organize datasets	Zoom
Week 7	Monday	10/14/2024	Relational data	In Person
	Wednesday	10/16/2024	Integrating relational datasets	Zoom
Week 8	Monday	10/21/2024	Non-numeric data types	In Person
	Wednesday	10/23/2024	Manipulating strings, factors, dates, and times	Zoom
Week 9	Monday	10/28/2024	Writing clear code	In Person
	Wednesday	10/30/2024	Using pipes as a programming tool	Zoom
Week 10	Monday	11/4/2024	Writing and using functions	In Person
	Wednesday	11/6/2024	Using functions and conditional expressions	Zoom
Week 11	Monday	11/11/2024	Vectors & iteration	In Person

	Wednesday	11/13/2024	Working with loops	Zoom
Week 12	Monday	11/18/2024	Building models	In Person
	Wednesday	11/20/2024	Predictive modelling	Zoom
Week 13	Monday	11/25/2024	Working with data models	In Person
	Wednesday	11/27/2024	Communicating your findings	Zoom
Week 14	Monday	12/2/2024	Final Presentations	In Person
	Wednesday	12/4/2024	Final Presentations	In Person

Attendance Statement

Students are expected to attend all classes. Students who are unable to attend class must seek permission for an excused absence from the course director or teaching assistant. Unapproved absences or late attendance for three or more classes may result in a lower grade or an "incomplete" for the course. If a student has to miss a class, he or she should arrange to get notes from a fellow student and is strongly encouraged to meet with the teaching assistant to obtain the missed material.

Scientific and Professional Ethics

The work you do in this course must be your own. Feel free to build on, react to, criticize, and analyze the ideas of others but, when you do, make it known whose ideas you are working with. You must explicitly acknowledge when your work builds on someone else's ideas, including ideas of classmates, professors, and authors you read. If you ever have questions about drawing the line between others' work and your own, ask the course professor who will give you clear guidance. Exams must be completed independently. Any collaboration on answers to exams, unless expressly permitted, may result in an automatic failing grade and possible expulsion from the Graduate Program.

Technology Requirements and Support

For issues related to Canvas, please contact the Graduate Office by email at: gradprgm@scripps.edu or by phone at: 858-784-8469.

Course Grading

Grading is in accordance with the academic policies of the Skaggs Graduate School. The breakdown of grading is as follows:

- Homework (2 assignments): 25% each
- Midterm Project: 25%
- Final Project/Presentation: 25%

Letter Grade	Percent	GPA	Description
A	93-100	4.00	Outstanding achievement. Student performance demonstrates full command of the course subject matter and evinces a high level of originality and/or creativity that far surpasses course expectations.
A-	90-92	3.67	Excellent achievement. Student performance demonstrates thorough knowledge of the course subject matter and exceeds course expectations by completing all requirements in a superior manner.
B+	87-89	3.33	Very good work. Student performance demonstrates above-average comprehension of the course subject matter and exceeds course expectations on all tasks as defined in the course syllabus. There is notable insight and originality.
B	83-86	3.00	Satisfactory work. Student performance meets designated course expectations and demonstrates understanding of the course subject matter at an acceptable level.
B-	80-82	2.67	Marginal work. Student performance demonstrates incomplete understanding of course subject matter. There is limited perception and originality.
C+	77-79	2.33	Unsatisfactory work. Student performance demonstrates incomplete and inadequate understanding of course subject matter. There is severely limited or no perception or originality. Course will not count toward degree.
C	73-76	2.00	Unsatisfactory work. Student performance demonstrates incomplete and inadequate understanding of course subject matter. There is severely limited or no perception or originality. Course will not count toward degree.
P	73-100	0.00	Satisfactory work. Student performance demonstrated complete and adequate understanding of course subject matter. Course will count toward degree.
F	0-72	0.00	Unacceptable work/Failure. Student performance is unacceptably low level of knowledge and understanding of course subject matter. Course will not count toward degree. Student may continue in program only with permission of the Dean.
I		0.00	Incomplete is assigned when work is of passing quality but is incomplete for a pre-approved reason. Once an incomplete grade is assigned, it remains on student's permanent record until a grade is awarded.
W		0.00	Withdrew from the course with Dean's permission beyond the second week of the term.

- All courses will be recorded and maintained in the student's permanent academic record; only courses that apply towards the degree will appear on the academic transcript. Non-credit or audited courses will not appear on the transcript.
- 4 core courses taken for a letter grade (pass = B- or higher for a core course)
- 2 elective courses taken pass/fail (pass = A, B, C for an elective)

Because students are encouraged to take electives outside their area of expertise, a "C" letter grade is passing.

Course Summary:

Date	Details
Mon Sep 2, 2024	Labor Day (No Class)
Wed Sep 4, 2024	Course Introduction: R, R Studio, and R Markdown - SEMINAR ROOM
Fri Sep 6, 2024	Grad Student Symposium
Mon Sep 9, 2024	Data Visualization
Wed Sep 11, 2024	Using ggplot2 to visualize data
Mon Sep 16, 2024	Data Transformation
Wed Sep 18, 2024	Using dplyr to transform datasets
Mon Sep 23, 2024	Exploratory data analysis basics
Wed Sep 25, 2024	Combining data visualization and transformation tools
	Homework 1
Mon Sep 30, 2024	Data structures and imports
Wed Oct 2, 2024	Using tibble and readr to efficiently work with data sets
Mon Oct 7, 2024	Tidy data
Wed Oct 9, 2024	Using tidyr to organize datasets
Mon Oct 14, 2024	Relational data
Wed Oct 16, 2024	Integrating relational datasets
Mon Oct 21, 2024	Non-numeric data types
Wed Oct 23, 2024	Manipulating strings, factors, dates, and times
	Midterm Project
Mon Oct 28, 2024	Writing clear code
Wed Oct 30, 2024	Using pipes as a programming tool
Mon Nov 4, 2024	Writing and using functions
Wed Nov 6, 2024	Using functions and conditional expressions
Mon Nov 11, 2024	Vectors & iteration
	Homework 2
Wed Nov 13, 2024	Working with loops
Mon Nov 18, 2024	Building models
Wed Nov 20, 2024	Predictive modelling
Mon Nov 25, 2024	Working with data models
Wed Nov 27, 2024	Communicating your findings
Thu Nov 28, 2024	Thanksgiving Holiday (No Class)
Fri Nov 29, 2024	Thanksgiving Holiday (No Class)
Mon Dec 2, 2024	Final Presentations
	Final Project
	Slide Upload for Final Presentation
Wed Dec 4, 2024	Final Presentations