QTM 151 Introduction to Statistical Computing II, Spring 2023

Instructor: Dr. Alejandro Sánchez Becerra
Email: alejandro.sanchez.becerra@emory.edu
Office: Psychology and Interdisciplinary Sciences (PAIS), 571

Communications:

Each module includes a discussion board where you can post questions that may be relevant to the entire class. You are encouraged to post questions related to the class here.

If your question is individual in nature, please use email to contact me outside of class time and office hours. I will try to respond to all written communication within 24 hours. Please try to limit the use of email and ask questions during office hours instead.

Overview

This course is an introduction to the Python programming language and SQL for students without prior programming experience. The purpose of this course is to prepare students for upper-level electives in data analysis related courses. It will cover the programming basics of Python which include understanding data types, controlling flow using loops and conditional statements, and writing functions. In addition to these basics, this course will put emphasis on skills that are relevant for data analysis which include 1) data manipulation such as merging, appending, and reshaping using SQL and Python, 2) making various plots for descriptive analysis using Python, and 3) using cloud-based services such as version control and cloud computing.

Learning objectives

1. Understand the version control workflow using GitHub
2. Understand Python data types such as int, float, string, list, and dictionary, NumPy array and Pandas data frame
3. Control flow with if statements and loops (for loop, while loop)
4. Write custom functions
5. Design and write programs to solve problems
6. Data manipulation such as merging, appending, and reshaping using SQL and Python pandas
7. Plot various graphs for descriptive analysis using matplotlib and seaborn

Materials

The class will be entirely based off of lectures provided by the instructor for each class and stored in the following Github repository:

https://github.com/alejandrosanchezbecerra/qtm151spring2023

Optional Materials

The following readings are not required, but are suggested for students interested in further readings:

- Elements of Data Science: https://allendowney.github.io/ElementsOfDataScience/README.html
- Think Python: https://greenteapress.com/thinkpython/html/index.html
- SQL: https://runestone.academy/ns/books/published/ac1/sql/toctree.html

Course Schedule

The following table outlines the proposed schedule for the semester. While I intend to follow this as closely as possible, this is subject to change. Check Canvas regularly for updates to deadlines and any changes to the assigned material. All homework due dates are in blue.

Week	Date	Title	Lecture	Percentage
Module 1: Introduction to GitHub, Python and Jupyter Notebooks				
Week 1	11-Jan	Introduction to Computing Environment	0	
Week 2	16-Jan	Martin Luther King Holiday (no classes)	1	
	17-Jan	Assignment 1 due (10 p.m.)		5%
	18-Jan	Introduction to Jupyter Notebooks	2	
Module 2: Python data types and controlling flows				
Week 3	23-Jan	Variables and Lists	3	
	24-Jan	Assignment 2 due (10 p.m.)		5%
	25-Jan	Mathematical Operations, Arrays, and Random Numbers	4	
Week 4	30-Jan	Boolean variables and if/else statements	5	
	31-Jan	Assignment 3 due (10 p.m.)		5%
	1-Feb	Loops – while, for, break, continue	6	
Week 5	6-Feb	Applications 1: Simulation Studies	7	
	7-Feb	Assignment 4 due (10 p.m.)		5%
	8-Feb	Applications 2: Operations over multiple datasets	8	Quiz 1: 6%
Module 3: Writing and Running Functions				
Week 6	13-Feb	User defined functions, lambda, apply	9	
	14-Feb	Assignment 5 due (10 p.m.)		5%
	15-Feb	Local/global variables	10	
Week 7	20-Feb	Reading, Subsetting, Factor variables	11	
	21-Feb	Assignment 6 due (10 p.m.)		5%
	22-Feb	Application 1: Linear Regression in Python	12	
	24-Feb	Partial Withdrawal Deadline		
Week 8	27-Feb	Application 2: Out-of-sample Model prediction	13	Quiz 2: 6%
		Mid-semester Student Survey		0.50%

		Module 3: Data Manipulation with Pandas		
	1-Mar	Creating and replacing variables	14	
Week 9	6-Mar	Spring Break (no classes, no office hours)		
	8-Mar	Spring Break (no classes, no office hours)		
Week 10	13-Mar	Aggregating Data	15	
	14-Mar	Assignment 7 due (10 p.m.)		*5%*
	15-Mar	Chaining and Lambda Functions	16	*Quiz 3: 6%*
Week 11	20-Mar	Merging Data	17	
		Instructions for final project		
	21-Mar	Assignment 8 due (10 p.m.)		*5%*
		Module 4: Data Manipulation with SQL		
	22-Mar	Introduction to SQL notebooks	18	
Week 12	27-Mar	Uploading and Merging Tables in SQL	19	
	22-Mar	Assignment 9 due (10 p.m.)		*5%*
	29-Mar	Import SQL Data into Python	20	*Quiz 4: 6%*
		Module 5: Time and Text Data		
Week 13	3-Apr	Manipulating Text Data	21	
	4-Aprl	Assignment 10 due (10 p.m.)		
	5-Apr	Time and Panel Data	22	
Week 14	10-Apr	Pivot Table	23	
	12-Apr	Advanced Plots 1	24	*Quiz 5: 6%*
		Module 6. Miscellaneous Topics		
Week 15	17-Apr	Advanced Plots 2	25	
	19-Apr	Advanced Plots 3	26	
Week 16	24-Apr	Review		
	25-Apr	Final Project Due		*20%*

Course Assignments

Your grade comprises homework (50%), class quizzes (30%), and the final project (20%). The assignments correspond to the structure outlined in the schedule above.

Course Assignments Description

Assignments

Working together on the homework assignments is encouraged, but you must write your own solutions. It is highly recommended that you make your solo effort on all the problems before consulting others. Each assignment has its due date indicated in this Syllabus and on Canvas. Any assignment submitted after the due date/time will not be given any points. To accommodate unexpected circumstances, your lowest homework grade will be automatically dropped at the end of the semester. **The same applies to in class quizzes.**

Final Project

The final term project is group-based, in teams of 3-4 students. The details of the project will be discussed in class.

Participation (extra credit) + 0.5%

Your participation will consist of a midsemester in-class survey. This will include teaching feedback and knowledge assessment questions that I will use for up to 0.5% extra credit.

Grading Scale

Your final grade will be assigned based on this grading scale: Grades will generally follow the standard scale, but the professor may curve up:

A	A-	B+	B	B-	C+	C	C-	D	F
93+	87-92	83-87	80-82	75-79	70-74	65-69	60-64	55-59	0-54

Throughout the semester, I will keep the Canvas gradebook up to date so that you have an ongoing understanding of where your grades fall in this scale.

Policy

General: Students are expected to adhere to the Emory College Honor Code as well as its Conduct Code, see http://college.emory.edu/home/academic/policy/conduct_code.html. Specifically, the honor code is in effect throughout the semester. By taking this course, you affirm that it is a violation of the code to cheat on assignments, end-of-chapter assessments, to plagiarize, to deviate from the teacher's instructions about collaboration on work that is submitted for grades, to give false information to a faculty member, and to undertake any other form of academic misconduct. You also affirm that if you witness others violating the code you have a duty to report them to the honor council.

Special circumstances: Students requiring any type of special classroom/testing accommodation for a disability, religious belief, scheduling conflict, or other impairment that might affect his or her successful completion of this course must personally present the requested remedy or other adjustment in written form (signed and dated) to the instructor, i.e., supporting memorandum of accommodation from the Office of Disabilities Services, http://www.ods.emory.edu/. Requests for accommodations must be received and authorized by the instructor in written form no less than two weeks in advance of need. No accommodation should be assumed unless so authorized. In the event of needs identified later in the course, or for which an adjustment cannot be made on a timely basis, a grade of "I" Incomplete for the course, will be given to accommodate the unanticipated request.

Attendance:

- There is no separate grade for attendance. However, attendance will be monitored through quizzes which will take place during class.

- The OUE student self-service absence form must be submitted if you would like to have your absence excused due to the reasons recognized by the university (See the top of the form). Be sure to complete this form in advance, and you must include all classes you are currently enrolled at the time of completing the form so that all your instructors will be notified all at once. If you wish to explain your situation further, you may also email your instructor in advance, and if necessary, submit official documentations.
- Check our Canvas page regularly! All course materials including lecture materials, quizzes, assignments, due dates, and important announcements will be posted on Canvas throughout the semester.

In/Out of classroom conduct: Students are expected to adhere to the Emory University Code of Conduct, see http://conduct.emory.edu/policies/code/index.html