# QTM 531: COMPUTING II

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

## COURSE OBJECTIVES

This class is the second sequence of the two computing courses connecting to QTM 530 Computing I. Assuming that students know how to explore and manipulate data, and do the basic programming, this course will focus on gaining building blocks for programming related to data analysis and machine learning. In addition, the class will introduce practical concepts relevant for reproducible research with big data. By the end of the course, students are expected to (1) fluently reshape data into the most convenient form for analysis, (2) know how to implement methods related to data analysis, (3) know how to implement algorithms in machine learning, (4) know how to implement statistical methods and machine learning algorithms using cloud, (5) know how to make a research reproducible by understanding from the development and source control to the deployment. Students would primarily write code in Jupyter/IPython notebooks. Most of the computing exercises will be based on Python.

## PREREQUISITES

- QTM 530 Computing I

## LECTURES

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/qtm531spring2024

## REQUIRED TEXTBOOK

- [IML] Introduction to Machine Learning with Python, by Andreas C. Müller and Sarah Guido, O'Reilly
- [PD] Python for Data Analysis, by Wes McKinney, O'Reilly
- [IPBS] Introduction to Python Programming for Business and Social Science Applications, by Frederick Kaefer and Paul Kaefer

## OPTIONAL MATERIAL

[FPP] Foundations of Python Programming --------
https://runestone.academy/runestone/books/published/fopp/index.html
[PE] Python for Everybody, by Charles R. Severance, http://do1.dr-chuck.com/pythonlearn/EN_us/pythonlearn.pdf [GLM] statsmodels.org – GLM notebook -------

--- FUNCTION REFERENCE
https://www.statsmodels.org/stable/examples/notebooks/generated/glm.html
[HDS] How to think like a Data Scientist,
(https://runestone.academy/runestone/books/published/httlads/index.html
[BB] Bash for Beginners [BB], by Machtelt Garrels,
https://www.tldp.org/LDP/Bash-Beginners-Guide/Bash-Beginners-Guide.pdf
[IL] Introduction to Linux [IL], by Machtelt Garrels, https://www.tldp.org/LDP/intro-linux/html/
[ACP] AWS Cloud Practitioner Essentials [ACP], 2nd Ed,
https://www.aws.training/Details/Curriculum?id=27076&scr=path-cp

CLASS REQUIREMENTS
Grades will be based on
• homework assignments (45%)
• mid-semester project (20%)
• final project (30%)
• class participation and in-class exercises (5%)

HOMEWORK
The homework assignment consists of 7 computer-based problem sets. Any assignment submitted after the due date/time will be considered for half points. To accommodate unexpected circumstances, your lowest homework grade will be automatically dropped at the end of the semester. Working together on the homework assignments is encouraged, but you must write your own answers. It is highly recommended that you make your solo effort on all the problems before consulting others.

HONOR CODE
All students enrolled at Emory are expected to abide by the Emory College Honor Code. Any type of academic misconduct is not allowed which includes 1) receiving or giving information about the content or conduct of an examination knowing that the release of such information is not allowed and 2) plagiarizing, whether intentionally or unintentionally, in any assignment. For the activities that are considered to be academically dishonest, refer to the Honor Code:

http://catalog.college.emory.edu/academic/policies-regulations/honor-code.html.

DISABILITY ACCOMMODATIONS
If you are seeking classroom accommodations or academic adjustments under the Americans with Disabilities Act, you are required to register with Office of Accessibility Services (OAS), http://accessibility.emory.edu/. Once registration is finalized, students must request accommodation needs to be communicated or facilitated. Students are expected to give two

weeks' notice of the need for accommodations for any class activities including the exams. For more information, please see http://accessibility.emory.edu/students/new-to-oas/registering.html. Please make sure to contact me with the relevant letter at the beginning of the semester.

COURSE SCHEDULE

| Week | Date | Topic |
|---|---|---|
| Module 1: Programming Essentials in Python | | |
| Week 1 | 17-Jan | Introduction to version control |
| Week 2 | 22-Jan | Pandas, Mathematical Operations, and Lists. |
| | 24-Jan | Boolean Types, Ifelse, Data Subsetting |
| Week 3 | 29-Jan | Recoding, and aggregating variables |
| | 31-Jan | Merging, Codebooks, and Chaining |
| Module 2: Process Automation | | |
| Week 4 | 5-Feb | Scraping 1: HTML, JSON, Dictionaries |
| | 6-Feb | Assignment 1 due (10 p.m.) |
| | 7-Feb | Scraping 2: Retrieving and Processing |
| Week 5 | 12-Feb | Flow Control / Loops |
| | 13-Feb | Assignment 2 due (10 p.m.) |
| | 14-Feb | Parallelization: Multiprocessing and Multithreading |
| Week 6 | 19-Feb | Time Series, Pivoting, and Panel Data |
| | 21-Feb | Regular Expressions and Text Wrangling |
| | 22-Feb | Assignment 3 due (10 p.m.) |
| Module 3: Model Deployment | | |
| Week 7 | 26-Feb | Data analysis with OLS estimator |
| | 28-Feb | Data analysis with randomized experiment |
| | 29-Feb | Assignment 4 due (10 p.m.) |
| Week 8 | 4-Mar | Logit model |
| | 6-Mar | Poisson model |
| | 7-Mar | Midsemester Project Due |

| | | |
|---|---|---|
| Week 9 | 11-Mar | Spring Break (no classes, no office hours) |
| | 13-Mar | Spring Break (no classes, no office hours) |
| Week 10 | 18-Mar | Example 1: Decision Tree |
| | 19-Mar | Assignment 5 due (10 p.m.) |
| | 20-Mar | Classification |
| Week 11 | 25-Mar | Model Evaluation |
| | 27-Mar | Example 2: Neural Networks |
| Week 12 | 1-Apr | Example 3: K-Means clustering Part I |
| | 2-Apr | Assignment 6 due (10 p.m.) |
| | 3-Apr | Visualizing model outputs |
| Module 4: Databases, Development and Production | | |
| Week 13 | 8-Apr | Basic Query Language |
| | 9-Apr | Assignment 7 due (10 p.m.) |
| | 10-Apr | Relational SQL |
| Week 14 | 15-Apr | Managing virtual environments: Conda, Dockers and containers |
| | 17-Apr | Scripting and Operating System |
| Week 15 | 22-Apr | Source Code and APIs |
| | 24-Apr | Deployment |
| | 25-Apr | Assignment 7 due (10 p.m.) |
| Week 16 | 29-Apr | Parallel Computing and GNU Computing |
| | 1-May | Final Project Due |