**QTM 530: COMPUTING I**

**FALL 2024**

**Instructor:** [**Dr. Alejandro**](https://www.juanestrada.info/) **Sánchez Becerra**

Email: [alejandro.sanchez.becerra@emory.edu](mailto: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.

**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/qtm530fall2024>

**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 (60%)

* 6 homeworks, worth 10% each

• self-directed projects 30%)

* Three written deliverables, worth 10% each

• in-class workshop (10%)

**LATE ASSIGNMENT POLICY**

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.

**GRADING**

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 | 83-86 | 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.

**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** | 29-Aug | 0 | Introduction to version control |
| **Week 2** | 3-Sep | 1 | Pandas, Mathematical Operations, Variables, and Lists. |
|  | 5-Sep | 2 | Boolean Types, Ifelse, Data Subsetting |
| **Week 3** | 9-Sep |  | Assignment 1 due (10 p.m.) |
|  | 10-Sep | 3 | Recoding, and aggregating variables |
|  | 12-Sep | 4 | Merging, Codebooks, and Chaining |
| **Module 2: Process Automation** | | | |
| **Week 4** | 17-Sep | 5 | Scraping 1: HTML, JSON, Dictionaries |
|  | 18-Sep |  | Assignment 2 due (10 p.m.) |
|  | 19-Sep | 6 | Scraping 2: Retrieving and Processing |
| **Week 5** | 24-Sep | 7 | Flow Control / Loops |
|  | 26-Sep | 8 | Functions and Parallel Computing |
| **Week 6** | 1-Oct | 9 | Time Series, Pivoting, and Panel Data |
|  | 2-Oct |  | Assignment 3 due (10 p.m.) |
|  | 3-Oct | 10 | Regular Expressions and Text Wrangling |
| **Module 3: Model Deployment** | | | |
| **Week 7** | 8-Oct | 11 | Data analysis with OLS estimator |
|  | 9-Oct |  | Assignment 4 due (10 p.m.) |
|  | 10-Oct | 12 | Data analysis with randomized experiment |
| **Week 8** | 14-Oct |  | **FALL BREAK (No Class, No Office Hours)** |
|  | 16-Oct | 13 | Logit model |
|  | 17-Oct |  | Project 1 due (10 p.m.) |
| **Week 9** | 22-Oct | 14 | Example 1: Regression Trees and Random Forests |
|  | 24-Oct | 15 | Cross-validation |
| **Week 10** | 29-Oct | 16 | Classification Metrics |
|  | 30-Oct |  | Project 2 due (10 p.m.) |
|  | 31-Oct | 17 | Example 2: Neural Networks |
| **Week 11** | 5-Nov | 18 | Example 3: PCA and K-means clustering |
|  | 7-Nov | 19 | Machine Learning Pipelines |
|  | 8-Nov |  | Assignment 5 due (10 p.m.) |
| **Week 12** | 12-Nov | 20 | Text Analysis **(CLASS ON ZOOM)** |
|  | 14-Nov | 21 | Advanced Github **(SELF STUDY – No Class, No Office Hours)** |
| **Module 4: Databases and Production Environment** | | | |
| **Week 13** | 18-Nov | 22 | Introduction to SQL |
|  | 19-Nov |  | Assignment 6 due (10 p.m.) |
|  | 20-Nov | 23 | Running SQL from Python |
| **Week 14** | 26-Nov | 24 | Python Classes + Deployment **(CLASS ON ZOOM)** |
|  | 28-Nov |  | **THANSKGIVING BREAK (No Class, No Office Hours)** |
| **Week 15** | 3-Dec | 25 | Managing virtual environments: Conda, Dockers and containers |
|  | 4-Dec |  | Project 3 due (10 p.m.) |
|  | 5-Dec | 25 | In-class workshop |
| **Week 16** | 10-Dec | 26 | In-class workshop |