QTM 350: Data Science Computing

Summary of Topics

Along with a primer on linux, scripting, Python and elementary programming concepts, the course introduces foundational knowledge of cloud computing. In the final project, students use Amazon Web Services for data science and machine learning applications.

Relationship to other courses offered by the Institute for Quantitative Theory and Methods (QTM)

This course is intended as a follow up to QTM 220, exposing students to popular languages for data analysis that they haven't used yet, important programming and computing concepts useful when working with computers, and emphasizing good workflows for reproducible research.

Course objectives

Each student will:

- Become comfortable working from the command line and using the command line to automate workflow (e.g. write a script for data cleaning a dataset)
- Learn basic syntax and concepts for popular languages (e.g. Python) and packages (e.g. Numpy, Pandas) used in data science.
- Learn foundational knowledge of cloud computing
- Creatively implement what they have been learning in a final project using cloud resources.

General

- Class hours and location: online (enrolled students will receive a Zoom meeting invite by email)
- Prerequisites: QTM 220
- Schedule

Instructor

- Jeremy Jacobson.
- Office hours: schedule online.

References

- Elements of Data Science

Elements of Data Science is a textbook by Allen Downey written in the form of Juptyer notebooks. It provides an introduction to data science in Python for people with no programming experience.

- AWS Academy Cloud Foundations (login required)

AWS experts develop and maintain the Cloud Foundations reference material specifically for AWS Academy courses such as this to ensure it keeps pace with cloud innovation and to prepare students for real-world, industry challenges.

Grading

- 3 Homework assignments: 30%
- Weekly take-home quizzes: 20%
- Project proposal: 10%
- Final project: 30%
- Final exam: 10%

Project Proposal

- Your group will submit a 5-minute video presentation describing your proposed final project and a Juypyter notebook presenting preliminary results.

Final Project

- Your group will publish the final project in the form of a blog post hosted on AWS that presents the primary results and (2) a Jupyter notebook and github repository that contains the details.
- You will work in groups for the homework and project.
- Your work is governed by the Emory Honor Code. Honor code violations (e.g., copies from any source including your colleagues and internet sites without reference and delineation of your contributions) will be referred to the Emory Honor Council.

Final Exam

The final exam is cumulative. Questions from the assigned quizzes are pooled and a selection of these, or similar, make up the Final Exam. It is administered on Canvas using Respondus.

Accomodations

All information shared with me regarding your identity, disability and accommodations is confidential. Please contact me as soon as possible so that I can provide the approved accommodations outlined on the accommodation notification letter. I am happy to consult with ADSR in providing additional accommodations.