QTM 151 - Introduction to Statistical Computing II

Danilo Freire

Summer 2025

Contents

Course Description

Welcome to QTM 151! This course introduces students to data analysis and statistical computing using Python and SQL. It is ideal for those with little or no programming experience who want to develop skills for data-driven decision making.

Over the semester, we will cover version control for collaborative coding, Jupyter Notebooks for reproducible research, Python programming basics, data wrangling and merging in SQL, data visualisation, and introductions to linear modelling and time series analysis.

Students will work with real-world datasets and problems, gaining practical experience in using these tools to extract insights from data. The course aims to develop both technical skills and critical thinking needed for complex data challenges. By the end, students will be ready for advanced study in quantitative methods and data science.

Learning Objectives

By the end of this course, students will be able to:

1. Perform basic operations and write functions in Python.

2. Conduct data wrangling and manipulate data using Python libraries such as Pandas.

3. Merge and manage databases using SQL.

4. Create visualisations to effectively communicate data insights.

5. Implement linear models and understand the principles of time series analysis.

6. Use Jupyter Notebooks for reproducible research.

7. Develop problem-solving skills relevant to data analysis and statistical computing.

Prerequisites

There are no prerequisites for this course. All students are welcome to join, regardless of their prior experience with programming or data analysis. Please feel free to reach out if you have any questions about the course content or your readiness to take the class.

Materials

This course is designed to be self-contained, providing all the necessary resources and materials to succeed in mastering the core concepts. However, students are encouraged to explore the following suggested books and online courses to deepen their understanding of Python and SQL.

Suggested Books

- Python for Data Analysis by Wes McKinney
- Elements of Data Science by Allen Downey
- Automate the Boring Stuff with Python by Al Sweigart
- Python for Everybody by Charles Severance
- SQL for Data Scientists by Renee M. P. Teate

Online Courses

- Coursera: Python for Everybody Specialisation
- edX: Python Basics for Data Science
- Codecademy: Learn Python
- DataCamp: Introduction to SQL
- Coursera: SQL for Data Science

Additional Resources

- Python Documentation
- Pandas Documentation
- PostgreSQL Documentation
- SQLBolt
- DataLemur for SQL interview practice
- Github Guides

Course Information

We will meet every Monday and Wednesday from 16:00 to 16:50 in the Anthropology Building, 303. It is important that you read the materials before class. All information about the course is available on the course's GitHub repository at `https://github.com/danilofreire/qtm151`. While I will try to adhere to the course schedule as much as possible, I also want to adapt to your learning pace and style. The syllabus and course plan may change in the semester. Again, please check the course repository regularly to check for updates. I will also announce any changes in class and via email.

Software

We will mainly use Python in this course. Python is a free, versatile, and powerful programming language that is widely used in data science, machine learning, and scientific computing. I recommend using the Anaconda distribution as it comes with many necessary Python libraries for data analysis, such as Pandas, NumPy, and Jupyter.

You can write your Python code in any text editor, but I recommend VS Code with the Python extension. Pycharm is also well-regarded by developers. If you are feeling adventurous, you can also use Neovim with the coc-pyright plugin. That is, if you can exit the editor. :)

We will use PostgreSQL for database management. You can download PostgreSQL from the official website. Please also install pgAdmin and the VS Code extension for PostgreSQL to interact with the database.

We will also use Jupyter Notebooks in class. Jupyter itself comes pre-installed with Anaconda, but please install the Jupyter extension for VS Code as well. We will have a hands-on session to learn how to use Jupyter effectively.

To help you get started, I have prepared a series of tutorials on how to install Anaconda, Jupyter, PostgreSQL, VS Code, and open a free educational account on GitHub. Please follow these tutorials as soon as possible to ensure that you have all the necessary tools for the course.

Office Hours

I am very flexible with office hours, but it is easier to contact me via email. Feel free to send me a message any time at danilo.freire@emory.edu, and I will likely reply within a few hours. You can also book an appointment with me on Calendly, or just email me your availability. If you prefer, you can meet me in the afternoon at my office. I am in the Department of Quantitative Theory and Methods almost every weekday. My office address is in the Psychology and Interdisciplinary Sciences Building, 36 Eagle Row, 5th Floor, room XXXX. If possible, please email me before coming to ensure that no two students book the same time slot.

Academic Integrity

Upon every individual who is a part of Emory University falls the responsibility for maintaining in the life of Emory a standard of unimpeachable honour in all academic work. The Honour Code of Emory College is based on the fundamental assumption that every loyal person of the University not only will conduct his or her own life according to the dictates of the highest honor, but will also refuse to tolerate in others action which would sully the good name of the institution. Academic misconduct is an offense generally defined as any action or inaction which is offensive to the integrity and honesty of the members of the academic community. Any suspected case of academic misconduct will be referred to the Emory Honour Council.

Artificial Intelligence

Students have to submit ten problem sets and complete five in-class quizzes. You are allowed to use AI to assist with your assignments. I recommend using GitHub Copilot to generate code snippets, as it is free for students and provides good suggestions and explanations. Claude, ChatGPT, and Perplexity AI are also good tools. I am available to provide support and assistance with these tools during office hours or by appointment. However, please note that any errors or omissions resulting from the use of AI tools are your responsibility. Do not rely solely on AI to complete your assignments; you must

always double-check your work. Remember to cite all sources used in your problem sets and projects, including AI tools. Please include a note at the end of any document indicating that AI was used in its development.

Special Needs and Accessibility Services

I am committed to providing necessary accommodations to ensure all students have an equal opportunity to succeed in this course. Students with medical or health conditions that may impact their academic performance should visit the Department of Accessibility Services (DAS) to determine eligibility for appropriate accommodations. Those who receive accommodations should provide me with an Accommodation Letter from DAS at the beginning of the semester or as soon as the accommodation is granted. Please note that DAS accommodations, such as extra time or quiet spaces, will apply only to quizzes, not assignments. This is because assignments are released in advance, allowing students to work at their own pace. Athletes and students with other commitments should also inform me of any scheduling conflicts at the beginning of the semester. I will do my best to accommodate these students, but I cannot guarantee that all requests will be granted. If you have any questions or concerns, please contact me.

English Language Learners

Emory University welcomes students from around the country and the world, and the unique perspectives international and multilingual students bring enrich the campus community. To empower multilingual learners, an array of support is available including language and culture workshops and individual appointments. For more information about English Language Learning support at Emory, please contact the ELLP Specialists at `https://writingcenter.emory.edu`. No student will be penalised for their command of the English language.

Assignments and Grading Policy

Problem Sets (50%). There will be five problem sets throughout the course. These assignments are designed to reinforce concepts covered in lectures and readings, and to provide hands-on practice with statistical programming. Problem sets will include a mix of theoretical questions and practical applications. They will be assigned regularly and must be completed individually. You may discuss your work with other colleagues as long as you do not copy entire sentences, just changing a few words. If you worked with other students, please write down their names on your assignment. Please also acknowledge any sources you used in your work, including textbooks, articles, and AI resources. *Any assignment submitted after the due date/time will automatically be graded for half points.* To accommodate unexpected circumstances, your lowest assignment grade will be automatically dropped at the end of the semester. *The same applies to in-class quizzes.* Please submit all assignments in PDF format via Canvas until midnight on the due date.

    Class Quizzes (30%). Students will also take three in-class quizzes throughout the semester. These quizzes will be based on the lectures from the previous weeks. They will be designed to test your understanding of the material and your ability to apply the concepts to new problems. Quizzes will be open-book and open-notes. You are **not** allowed to use AI tools. They are individual assessments, and students are not allowed to discuss the questions with their colleagues in class.

Grading Scale

Each student's final grade will be based on the following after rounding up to the nearest point:

Grade	A	A-	B+	B	B-	C	D	F
Range	91%–100%	86%–90%	81%–85%	76%–80%	71%–75%	66%–70%	60%–65%	<60%

Course Outline and Suggested Readings

The lecture notes cover all the necessary material for the course, and the weekly suggested readings are recommended for those who want to deepen their understanding of the course topics. As mentioned above, the course outline is subject to change, and I will update the syllabus if needed. Please remember to check the course GitHub repository regularly.

- Tuesday, May 13: Lecture 01: Welcome to QTM 151 - Introduction and Course Overview. Please make sure to install the necessary software for the course by following the Course Tutorials: How to Install Anaconda, Jupyter, PostgreSQL, VSCode, and Open a Free Educational Account on GitHub.

- Wednesday, May 14: Lecture 02: Jupyter Notebooks, Packages, Variables, and Lists. Lecture Slides, Lecture Notebook.

- Thursday, May 15: Lecture 03: Maths Operations, Arrays, and Boolean Logic. Lecture Slides, Lecture Notebook

  - Kahoot Quiz. The quiz is optional and will not be graded. We will do it in class! It is a fun way to review the material covered in the lecture.

- Friday, May 16: Lecture 04: Quiz 01 - Based on Lectures 01-03. The repository contains the quiz questions and will be made online on the day of the quiz.

  - Assignment 01 due.

- Monday, May 19: Lecture 05: Conditional Statements. Lecture Slides, Lecture Notebook.

  - Kahoot Quiz.

- Tuesday, May 20: Lecture 06: While and For Loops. Lecture Slides, Lecture Notebook.

  - Kahoot Quiz.

- Wednesday, May 21: Lecture 07: Functions. Lecture Slides, Lecture Notebook.

- – Kahoot Quiz.

- – Assignment 02 due.

- Thursday, May 22: Lecture 08: Data Wrangling with Pandas. Lecture Slides, Lecture Notebook.

  - – Kahoot Quiz.

- Friday, May 23: Lecture 09: Quiz 02 - Based on Lectures 05-08. As with Quiz 01, the repository will be available on the day of the quiz.

  - – Assignment 03 due.

- Monday, May 26: Lecture 10: Subsetting and Filtering Data. Lecture Slides, Lecture Notebook.

  - – Kahoot Quiz.

- Tuesday, May 27: Lecture 11: Introduction to SQL. Lecture Slides, Lecture Notebook.

  - – Kahoot Quiz.

- Wednesday, May 28: Lecture 12: SQL in Python. Lecture Slides, Lecture Notebook.

  - – Kahoot Quiz.
  - – Assignment 04 due.

- Thursday, May 29: Lecture 13: Merge Tables in SQL. Lecture Slides, Lecture Notebook.

  - – Kahoot Quiz.

- Friday, May 30: Lecture 14: Quiz 03 - Based on Lectures 10-13.

  - – Assignment 05 due.

Each lecture folder contains an HTML file or a Jupyter notebook (`.ipynb`) with code examples and explanations, along with any additional resources or datasets used in the lecture.