03-701: Data Analysis for Biological Sciences

Fall 2024

**Instructor:**  En Cai

Office: MI 612

Email: [encai@andrew.cmu.edu](mailto:encai@andrew.cmu.edu)

**Class time:** Mondays and Fridays, 11:00-12:20 pm, DH1212

**TA:** Suzi Kim (suzik@andrew.cmu.edu), Elianna Lai (eliannal@andrew.cmu.edu)

**Course objectives:** In the biological sciences, researchers often encounter diverse data types, ranging from numerical datasets to complex images. To draw meaningful conclusions from these data, it is essential to have the right analytical tools and the skills to apply them effectively. This course is designed to equip biology students with practical data analysis skills, specifically tailored for those with little or no coding experience. This course will cover:

1. **Introduction to Python:** Learn the fundamentals of Python programming, focusing on applications relevant to biological data analysis.
2. **Basics of Image Analysis:** Gain an understanding of the principles of image analysis, including key techniques and tools used in the field.
3. **Image Analysis with Python and MATLAB:** Learn the basics of MATLAB and compare its features and capabilities with Python. Students will practice coding techniques taught in this course, applying them to image analysis using both Python and MATLAB.

**Coursework:** Students will be graded on three components:

1. Participation (25%) – Students are required to participate in the class which will account for 25% of their final grade.
2. Homework (50%) – Students are required to complete weekly homework and turn in their homework on time. This will account for 50% of their final grade.
3. Final project (25%) -- Students are required to complete a final project at the end of the course. This will account for 25% of their final grade.

Policies

**Absences:** Students are required to attend class, and this accounts for 25% of their final grade. One excused absence is provided by default. Additional absences can be excused for extenuating circumstances. Absences for religious observances must be submitted by email to the instructor during the first two weeks of the course.

**Academic integrity:** Study groups can be helpful in understanding course material, but students are responsible for preparing their own assignments independently. All material turned in must be done independently unless explicitly indicated on the assignment handout. The University Policy on Academic Integrity can be found here: <http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html>.

**Homework**: Homework will be assigned every week and will be submitted online. One late homework is allowed (you can submit one assignment up to 7 days late). Additional late submissions will incur the following grade penalties:

Submitted within 48 hours of the deadline - 10% reduction in the homework assignment grade.

Submitted 2-7 days after the deadline - 50% reduction in the homework assignment grade.

Submitted more than 7 days after the deadline will receive a zero grade.

**Course Calendar (subject to revision)**

The course will be organized into two modules. The two modules and planned class sessions are as follows:

**Module 1: Get Started with Python**

Aug 26. Lecture 1: Course introduction

Aug 30. Lecture 2: Python Basics – Variables, Data Structures

Sep 2. No Class: Labor Day

Sep 6. Lecture 3: Python Basics- Built-In Functions, Loops

Sep 9. Lecture 4: Python Basics- Working with Files, Functions

Sep 13. Lecture 5: Python - Packages for Scientific Computing

Sep 16. Lecture 6: Python - Numpy and Pandas

Sep 20. Lecture 7: Python - Data Visualization

**Module 2: Image Analysis for Biological Sciences**

Sep 23. Lecture 8: Image Analysis: Concepts and Tools

Sep 27. Lecture 9: Image Analysis: Examples in ImageJ

Sep 30. Lecture 10: Image Analysis with Python

Oct 4. Lecture 11: Matlab Basics and Applications in Image Analysis

Oct 7. Lecture 12: Course Wrap-up

Oct 11. Lecture 13: Final Project Due

**Diversity and Inclusion:** At Carnegie Mellon University, students, faculty, staff, and researchers come from diverse backgrounds. Students from diverse backgrounds and perspectives should be able to benefit from this course. In order to maintain our diversity, we must not discriminate against others because of their personal identities. In the event that you suspect there has been a bias based on an individual's identity, you can contact me or the Center for Student Diversity and Inclusion at csdi@andrew.cmu.edu, (412) 268-2150.

**Textbook and Reading:**

1. An Introduction to Statistics with Python, by Thomas Haslwanter.
2. Computing Skills for Biologists, by Stefano Allesina and Madlen Wilmes.