

Data-Based Research with Python: Course Syllabus

001.2.4065 (3 CPT)

Lecturer: Victor Yashunsky

Course Overview

Modern research depends on reliable data collection, processing, and analysis. This course is designed for graduate students with little or no programming experience who need to integrate data analysis into their work. The emphasis is on hands-on practice and problem-solving, with Python as the primary tool. By the end, you will be able to automate workflows, handle diverse datasets, and build computational solutions using Python's scientific libraries.

Each week blends short lectures with guided labs and assignments grounded in real research applications.

Learning Objectives

By the end of the course, you will be able to:

- Write, debug, and adapt Python code for scientific research.
- Analyze and manipulate data using NumPy and Pandas.
- Produce publication-quality visualizations with Matplotlib and Seaborn.
- Process and extract information from images using Scikit-image.
- Apply basic machine learning techniques with Scikit-learn.
- Automate workflows and manage code with the command line and Git.

Grading Scheme

Total: 100%

- **Assignments – 40%**
Coding home tasks.
- **Participation – 25%**
Attendance and engagement.
- **Final Project – 35%**
Proposal (5%); Code & Analysis (10%); Presentation (10%)

Two/Three-Weekly Modules

Module 1: Python Fundamentals—core programming skills and the Python language

- Getting Started: IDEs, Jupyter Notebooks, and interactive coding.
 - Data Types: Integers, floats, strings, and booleans.
 - Data Structures: Lists, tuples, dictionaries, and sets.
 - Control Flow: If/else, for/while loops.
 - Functions: Writing and reusing functions effectively.
-

Module 2: Data Analysis with Pandas and NumPy—tools for scientific data handling

- Working with structured datasets with **Pandas**: importing, querying, filtering, reshaping, and organizing information.
 - Using arrays for efficient computation: indexing, slicing, and applying mathematical operations with NumPy.
 - Case Studies: To be announced (drawn from real research contexts).
-

Module 3: Visualization and Image Processing—from graphs to images

- Designing and customizing figures with statistical visualizations suitable for publication with Matplotlib.
 - Handling, transforming, and quantifying images with OpenCV and Scikit-image.
 - Case Studies: To be announced (based on research applications).
-

Module 4: Machine Learning Basics using Scikit-learn—an introduction to computational modelling

- Core workflows for machine learning with Scikit-learn.
 - Understanding the distinction between supervised and unsupervised learning.
 - Algorithms: Decision Trees, K-Nearest Neighbors (classification), K-Means (clustering).
 - Evaluation: Accuracy, precision, recall, and limitations of metrics.
-

Module 5: Research Workflows and Final Project—putting it all together

- Command Line: Running Python scripts and managing files.
 - Version Control: Introduction to Git and GitHub for reproducibility and collaboration.
 - Final Project: Independent or group-based research project applying course skills, presented at the end of the course.
-