Scientific Data Analysis in Python

Meeting Times: Tuesdays & Fridays (85 minutes per session; specific meeting time TBD)
Dates: May 15 – June 24, 2025
Instructor: Brian Hill
Course Duration: 6 ½ weeks, 2 sessions per week (Total: 12 to 13 sessions)

Course Description:

This course introduces core Python programming and scientific computing skills using *Astronomical Python* (by Imad Pasha) and transitions into fundamental data science concepts with *Data Science from Scratch* (by Joel Grus). The first half of the course covers Python essentials, visualization, numerical computing, and scientific packages used in astronomy. The second half introduces data science principles, covering statistics, machine learning, and real-world data applications.

Required Materials:

- Imad Pasha, *Astronomical Python* (Sections 1–2, Chapters 1–8)
- Joel Grus, *Data Science from Scratch*, 2nd Edition

Daily Schedule:

Part 1: Scientific Python (Pasha)

Week 1 (May 16):

- **Before May 15:** Self study chapter 1 through chapter 3
- **Ch. 4:** Introduction to Python
    - Variables, data types, indexing, slicing
    - Control flow, loops, basic debugging

Week 2 (May 20 & 23):

- **Ch. 5:** Visualization with Matplotlib
- **Ch. 6:** Numerical Computing with NumPy

Week 3 (May 27 & 30):

- **Ch. 7:** Scientific Computing with SciPy
- **Ch. 8:** Astropy and Astronomical Packages

Part 2: Data Science Foundations (Grus)

Week 4 (June 3 & 6):

- **Ch. 1–2:** What is Data Science? Python Review
- **Ch. 3:** Visualizing Data

Week 5 (June 10 & 13):

- **Ch. 4–6:** Linear Algebra, Statistics, Probability

Week 6 (June 17 & 18):

- **Ch. 7–8:** Hypothesis Testing, Gradient Descent

Week 7 (June 24):

- **Ch. 9–10:** Getting Data, Working with Data
- (Optional bonus topic of Ch. 11–12: Machine Learning intro)

THE PRECEDING DAILY SCHEDULE IS SUBJECT TO ADJUSTMENT.

THE FINAL/ACTUAL SCHEDULE WILL BE RECORDED HERE:

https://brianhill.github.io/scientific-data-analysis

Grading

- Per-Session Coding Exercises: 5% each, 65% in total
- Final project: mini data analysis using tools from both books: 20%
- Participation: 15%