Elements of Data Science for the Physical and Life Sciences
SCTC 1013
Fall 2022

Section 003
Tuesday/Thursday 3:30 – 4:50 pm SERC 108A

Instructor: Jonathan M. Smith
Office: SERC 435
Email: jonathan.m.smith@temple.edu
Zoom Personal Room:
https://temple.zoom.us/my/jonathan.m.smith
Office Hours: Tuesday 10-12:30 and by arrangement.
Textbook: Free online digital textbook: Inferential Thinking 2nd Edition by Ani Adhikari, John DeNero, David Wagner.
Online Course Materials: https://sites.temple.edu/eds8/
Course Delivery: In-class meetings
Course Anticipated Costs: Free digital text and online materials. Necessary but typical computer equipment.
Meeting Times and Location: Tuesday/Thursday 3:30-4:30 PM SERC 108A
Virtual Laboratory Office Hours: (See Google sheet): https://bit.ly/3T5nB1X
Course Assistants:

Evan Gerlach	tun92624@temple.edu
Harrison Fedor	tun23723@temple.edu
Liem Mackay	tun32942@temple.edu
Edward Pike	tuj97761@temple.edu

Overview

Data is all around us and the amount of data in the world is constantly growing. Data analysis, simulation, and modeling is central to all fields of science. At the same time data fluency informs critical thinking and strengthens problem solving skills. Data science leverages computer programming and statistical concepts to develop formalized tools suitable to tackle tasks ranging from experimental data analysis to bioinformatics and to informing societal issues. This course aims to provide a foundation in computer programming approaches using the Python language to quickly begin work with real-world data sets drawn from the sciences and beyond. This hands-on approach will lay the foundation for domain or major specific applications with increased sophistication.

Course Goals

- Develop mathematical problem-solving skills
- Enhance critical thinking through data analysis and computational modeling
- Build foundational data fluency readied for major specific applications
- Understand algorithmic logic
- Acquire and process real-world data
- Perform data manipulation operations using basic Python coding constructs
- Visualize data including 2D relationships, distributions, geospatial, and 3D.

- Numerical simulation and distributions
- Draw inferences from data by applying statistical concepts
- Use machine learning concepts such as regression and classification to perform prediction tasks

Course Modules

- **Programming and data visualization**: Programming concepts, Jupyter Notebooks and Python, data types, functions, and visualization
- **Statistical inference**: Basics of probability, sampling and simulation, hypothesis testing, A/B testing, and confidence intervals
- **Prediction and machine learning**: Sample means, correlation, regression, and classifiers

Course Structure

Data science is a hands-on activity. Optimized learning will occur by manipulating and analyzing data using web based Jupyter notebooks in parallel with classroom teaching and activities. Videos may be utilized to introduce topics prior to the week's lecture. Students will be expected to view these videos and come prepared to class with questions about the topic. During class students will be introduced to class via lectures and lab work. Students will work on course labs during class times with the guidance of instructors and course assistants. Lab hours will also be held to provide help on the labs outside of class meeting times. To get credit for work on labs, you will need to download your lab submission after completing it and submit it on Canvas.

Technology Statement

A computer with a working browser and reliable internet connection, is required. Recommended internet speed is 8mbps download & 5mbps upload. You may test your connection at https://www.speedtest.net. If you do not have the required educational technology, you may submit a request to procure university resources (limited availability) using the Student Emergency Aid Fund. The University will endeavor to meet needs, such as with a long-term loan of a refurbished computer/laptop, or subsidized internet access.
A webcam, microphone and earphones will also be necessary for participating in laboratory coding sessions.

Assessment and Grading

Component	
Laboratory Data Assignments	60%
Module Projects	15%
Quizzes	15%
Participation	10%

Laboratory Data Assignments

Students will complete labs online to get hands on experience with data, visualization, and coding. Assignments should include comments, #, and

markdown to annotate work much like showing work on a problem in other subjects. It is best to break up a complicated code block into two or more blocks. Completed assignments will be uploaded on Canvas in both .html and .ipynb format (downloaded from Jupyter notebook). Classmates can provide helpful discussion will working through data and coding challenges but submitted work should be your own. A late penalty of 5% will be applied for every 12 hours beyond the due date. Late labs submitted more than 48 hours after the due date will receive no credit. Non-submission of a lab will count as zero, and your lowest lab assignment score in the course will be dropped. Extensions will be granted on a case-by-case basis. Students need to request extensions prior to the due date.

Quizzes

Periodic quizzes will be taken to check your understanding of the material presented in the lectures and lab. Quizzes will be scheduled in advance, and students can see the course timeline and weekly module activities on Canvas to know what material will be tested. You will be allowed to have one page of notes to be used on the quizzes. No make-ups shall be granted for missed quizzes. Non-submission of a quiz will count as zero, and your lowest two quiz scores will be dropped.

Participation

Active participation is necessary and will be assessed through both attendance and engagement. Periodic worksheets may be used as part of the participation grade. In addition, students will be required to attend three lab hours a semester, one per module.

Leaving the Course

The last day to add/drop the course is Tuesday, September 6, 2022. The last day to withdraw from the course is Monday, November 28, 2022. These are strict deadlines. Incompletes will only be given in unusual circumstances.

Academic Freedom

Freedom to teach and freedom to learn are inseparable facets of academic freedom. The University has adopted a policy on Student and Faculty Academic Rights and Responsibilities (Policy 03.70.02) that can be accessed at http://policies.temple.edu

Academic Honesty

Please review Temple's page on academic honesty and other student responsibilities in the undergraduate bulletin: https://bulletin.temple.edu/undergraduate/about-temple-university/student-responsibilities/. Cheating on any graded component will have serious consequences: a score of zero will be given on any work where you are detected cheating. **The penalty remains the same irrespective of whether the graded work was copied from or by the student**.

Accommodation

Any student who has a need for accommodations based on the impact of a documented disability or medical condition should contact Disability Resources and Services (DRS) in 100 Ritter Annex (drs@temple.edu; 215-204-1280) to request accommodations and learn more about the resources available to you. If you have a DRS accommodation letter to share with me, or you would like to discuss your accommodations, please contact me as soon as practical. I will work with you and with DRS to coordinate reasonable accommodations for all students with documented disabilities. All discussions related to your accommodations will be confidential.

Attendance and Your Health

To achieve course learning goals, students must attend and participate in classes, according to your instructors' requirements. However, if you feel unwell or if you are under quarantine or in isolation because you have been exposed to the COVID virus or tested positive for it, you should not come to campus or attend in-person classes or activities. It is the student's responsibility to contact their instructors to create a plan for participation and engagement in the course as soon as they can do so, and to make a plan to complete all assignments in a timely fashion, when illness delays their completion.

Topics

Week	Topic	Assessments
M1	**Programming and data visualization**	
1	Course introduction & Python Bootcamp Week 1	Lab 01 (*Introduction):* Practice with Jupyter Notebooks
2	Python Bootcamp Week 2 (Data Types, Tables, Functions)	Lab 02 (*Data Types*): Acceleration of Gravity on the Moon
3	Understanding Data with Table Methods	Lab 03 (*Data Tables*): Environmental Protection Index and Sustainability
4	Visualizing Data	Lab 04 (*Functions and Visualization*): COVID data
5	Summary: Module 1	Project 1: Programming & Data Visualization: Olympics mini-project
M2	**Statistical inference**	
6	Iteration and Probability	Lab 05 (*Probability & Randomization*): Groundhog Weather Prognostication
7	Hypothesis Testing (Sampling, Simulation)	Lab 06 (*Sampling & Simulation*): DNA GC Content and Wordle

8	A/B Testing & Causality	Lab 07 (*Inference):*Global Climate Change
9	Summary: Module 2	Project 2: *Statistical Inference:* Mount Saint Helens and biodiversity recovery
M3	**Prediction and machine learning**	
10	Normal Distribution (Mean, Standard Deviation)	Lab 08 (*Variance of Sample Means and Correlation*): Old Faithful Eruption Data
11	Regression & Regression Inference	Lab 09 (*Regression*): Estimation of the Age of the Universe
12	Classifiers and Classification	Lab 10 (*Machine Learning, Classification, & Chemistry*): Flower species classification and Molecular Acidity Prediction
13	Summary: Module 3	Project 3: Prediction and Machine Learning
14	Project Presentations	Project 3 Presentations