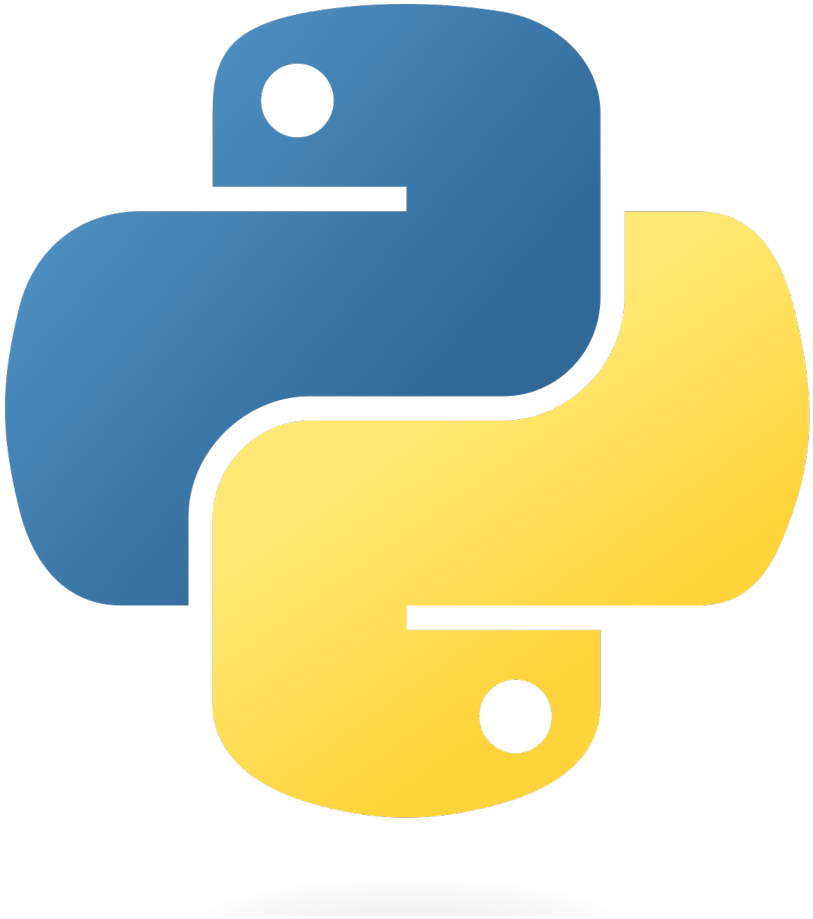
Python for Scientific Data Analysis

Note: **SYLLABUS IS SUBJECT TO CHANGE!!!**



Instructor: Thayne Currie
Department: Physics and Astronomy
Modality: In person, Mondays, Wednesdays, 1pm-2:15pm
Office Hours: Monday – 2:45-3:30pm; Thursday – 11:30am - 12:15pm
Email: thayne.currie@utsa.edu
Office Location: AET 3.388
Course Github Page: github.com/thaynecurrie/phys7943_fall2023

Course Syllabus Fall 2023

Table of Contents

Course Syllabus – Fall 2023

Public Health Considerations

The health and safety of our campus community is a shared responsibility of all Roadrunners. It is important to note that none of us can guarantee a COVID-19 free environment. We all must, however, follow the guidelines outlined in the UTSA Public Health Task Force Report ("Report") and any other applicable policies as may be communicated by the University from time to time. This will include regulating behaviors outlined in the Report that mitigate risk, including:

- Encouraging the use of face coverings,
- Self-monitoring for symptoms before coming to campus,
- Getting tested for COVID-19 if showing symptoms or after a close contact with a COVID-19 positive individual (if you are not already fully vaccinated and are not symptomatic),
- Following proper hygiene practices, including frequent hand sanitization, using cleansing wipes to disinfect surfaces, and minimizing the use of shared devices, tools and equipment,
- Avoid congregating (i.e. bottlenecking) near the entrances and exits before and after class – keeping your distance to reduce possible transmission from symptomatic or asymptomatic individuals.
- Communicating any COVID-19 related health concern to your supervisor or professor, and
- Submitting a self-report to report your positive test results or exposure (if not fully vaccinated and are also symptomatic for COVID-19 infection).

In turn, faculty members or supervisors will submit a COVID Case Referral to alert the COVID Response Team about positive COVID-19 cases for operational action. Failure to abide by these guidelines and requirements may result in disciplinary action in accordance with the Student Code of Conduct or applicable employment policies and procedures. Violations should be reported to the Office of Institutional Compliance via the UTSA Hotline for appropriate action.

For Face-to-Face Classroom and Other Academic Sessions: Face coverings/masks are recommended in indoor public and common spaces, especially for those individuals who are high risk and/or not vaccinated. **Due to my personal circumstances, I will strongly request students to wear masks during any in-person interactions (e.g. in-person office hours).** For the latest information, please review the Roadrunner Roadmap.

COURSE DESCRIPTION

Phys 4953/7903. Data Analysis Methods in Physics and Astronomy: Python for Scientific Data Analysis. (3-0) 3 Credit Hours. This course focuses on data analysis methods useful for physics and astronomy research in the Python programming language. Course coverage is split between learning pure methods and widely-used Python packages.

Credit hours: 3

Prerequisites/co-requisites: Prior instructor approval and basic familiarity with the fundamentals of coding.

Course Objectives: At the end of the course students will:

1. Understand the basics of Python relevant to scientific research, including core syntax and data structures.
2. Understand core Scientific Python functions in the NumPy and SciPy packages, including array indexing and slicing, conditional statements, numerical linear algebra, and statistical tests
3. Master reading in and plotting/displaying scientific data with Matplotlib
4. Understand basic functionality of the Pandas package for sorting/selecting data
5. Understand key functionality of the AstroPy and AstroQuery packages for professional astrophysics research
6. Understand key modeling techniques as coded in Python; understand the basics of concepts like Markov Chain Monte Carlo simulations and principal component analysis
7. Better write original Python code for scientific research incorporating concepts described in 1-6.

Time permitting, my goal also is to cover the following topics:
7. Machine Learning in Physics and Astronomy
8. Rigorous Hypothesis Testing
9. Code Optimization and Parallel Processing

Course Format: This course will be taught in person.
In-Person Requirements:

1. a laptop computer**;**
2. Python-3.7 or later installed on your laptop under the Anaconda Python distribution with the following packages installed: NumPy, SciPy, Matplotlib, Pandas, Seaborn, AstroPy, Astroquery, and other packages as announced. **DO NOT PROCRASTINATE WITH THIS INSTALLATION!**
3. Zoom installed on your laptop. My recommendation: create a separate Anaconda Python environment for this class to avoid clashes with your normal codes.

COMMUNICATION PLAN

1. **Email** - I respond best to email – thayne.currie@utsa.edu -- and will make every attempt to send you a reply within 24 hours (usually much faster).

If more than 24 hours has elapsed without a response, I likely never received your email: please resend. University policy requires me to communicate with you from your utsa email address.

2. **Office Hours** – Office Hours are Monday – 2:45-3:30pm; Thursday – 11:30am - 12:15pm

INSTRUCTOR BIO

Thayne Currie: I am happy to be your instructor. I earned my PhD from UCLA, doing most my thesis work at the Harvard-Smithsonian Center for Astrophysics. Prior to joining UTSA in August 2022, I was an astrophysicist at NASA Ames Research Center. I have a separate affiliated position with the Subaru Telescope on Maunakea in Hawai`i. My research focuses on studying extrasolar planets by direct imaging using Subaru and other optical/infrared telescopes on the Earth and in space.

Like many (all?) of you, Python is a second language for me: I started coding in Fortran (!!!) and for over a decade used IDL as my primary language before switching to Python. However, I hope that this fact will be an advantage for the class: I still remember what it was like to learn Python, what made sense, what didn't, and how existing literature could be improved.

COURSE MATERIALS

Required Textbook

There are no required textbooks for this course. However, my notes will draw heavily from a few sources. All of these are available on Amazon for reasonable prices: the third one has a free online version on Github. I would strongly suggest that you purchase at least one of these books:

1) *Python for Data Analysis*, 3rd edition (2022)– Wes McKinney [a good general overview]

2) *Scientific Computing with Python*, 2nd edition (2021) – Claus Fuhrer [can be rigorous]

3) *Python Data Science Handbook*, (2017) – Jake van der Plas [excellent Pandas discussion]

If we get to Machine Learning – which I hope we will – I will also likely be using a very new text (not yet officially published) called:

4) *Machine Learning in Physics & Astronomy*, by Viviana Acquaviva

I am also evaluating other textbooks on statistics and linear algebra in Python. I may also recommend some of these.

THIS COURSE ON CANVAS

There should be a Canvas page for this course. As this is not a large class like AST 1013 and Canvas is new (and I am still scarred from my experience with Blackboard), I will try to rely on Canvas as little as possible. In some instances, using Canvas will be unavoidable.

ASSIGNMENTS AND ASSESSMENTS

Readings and Lectures – Reading assignments are optional – as the lectures will cover this material – but they will likely be beneficial. I will try to provide lecture notes in PDF form before each class. Lectures may include files, separate Python scripts/modules, etc. I will also try to provide these ahead of time.

My plan right now is to have class sessions consist of two blocks of lectures separated by a period where you have time to work on a subset of problems.
Your grade will be based on the following:

- **Homework Problem Sets** – The problems will be cover material discussed in lectures. Except for Week 1 and possibly one more week, every week will have problem sets. These problems sets will be due the following Monday before class.

  A completed assignment will consist of i) source code showing how you solved each problem, ii) the result (including any figures, output data, or printed statements), and iii) responses to questions (where applicable).

  It is possible – though not guaranteed – that I will reserve some problems for graduate students only.

  During each class, I will set aside some time (~15 minutes) to work on a subset of homework problems. I will be available to help in case you get stuck on coding issues.

- **Class Project –** You will devise and work on a focused Python coding project relevant to your own interests that goes beyond material covered in this class. This could be original code you write yourself or analysis of data with a sophisticated pipeline/analysis package not covered in the class. The subject matter will be approved by me. You are allowed to work in groups of up to 3 for this assignment, provided that each member directly contributes an equal amount to the project. (continued below …)

  Due at the end of the course, you will submit a write-up of your project: describing its motivation and coding skills tested and instructions for executing code to replicate your results. If you write original code, this code should be submitted; if you use existing code, you will be required to provide a copy/link to the source code. Whether you write original code or utilize existing code, you will be expected to explain what the code is doing at each step

Working in Groups is Acceptable

You are encouraged to work in groups on problem set assignments and for your class

project PROVIDED THAT the work you submit is your own. E.g. you write down your own answers to problem sets yourself: do not copy and paste others' work.

GRADING INFORMATION

Final grades will be based on earning the following percentage breakdown.

Activity	Percentage
Problem Sets	80%
Class Project	20%

Grade Distribution and Letter Grade

Percentage	Grade
97-100%	A+
93-96%	A
90-92%	A-
87-89%	B+
83-86%	B
80-82%	B-
77-79%	C+
73-76%	C
70-72%	C-
67-69%	D+
63-66%	D
60-62%	D-
<60%	F

ESSENTIAL STUDENT INFORMATION

- **Important:** Bookmark and visit the Common Syllabus Information webpage to find important and valuable resources about counseling services, transitory/minor medical issues, supplemental instruction, tutoring services, academic success coaching, sexual harassment and sexual misconduct, campus safety and emergency preparedness, inclusivity statement, and the Roadrunner Creed.
- For technical requirements, support, and academic resources, visit the Student Support Gateway, where you can find all your tech and academic support resources in one place.

- Follow Online Learning Netiquette standards for your online communication activities. Please be mindful of the communication tools available in your course and use them for learning purposes. Class discussions take place in a respectful and safe environment, whether online or in person. UTSA encourages everyone to openly share their ideas and opinions without penalty or judgment, but learning should always be based on facts and research. It is possible to disagree without being disagreeable.
- UTSA provides reasonable accommodations to students via the Student Disability Services. For more details on eligibility, policies, and requirements, please visit www.utsa.edu/disability or call (210) 458-4157.
- **UTSA Wellbeing Resources:** your wellbeing is a priority for us. UTSA is proud to partner with Wellness 360 and MySSP to provide students with access to quality health and mental health care. Visit the UTSA Students Wellbeing Resources to explore the services available.

My Inclusivity Statement

The University of Texas at San Antonio, a Hispanic Serving Institution situated in a global city that has been a crossroads of peoples and cultures for centuries, values diversity and inclusion in all aspects of university life. As an institution expressly founded to advance the education of Mexican Americans and other underserved communities, our university is committed to ending generations of discrimination and inequity. UTSA, a premier public research university, fosters academic excellence through a community of dialogue, discovery, and innovation that embraces the uniqueness of each voice.

I also strongly support academic freedom and viewpoint diversity. I believe that a university thrives more when it provides a welcoming environment for students with different backgrounds, life experiences, political views, and religious beliefs or lack thereof to learn, debate, and express themselves without fear of ostracization or need for self-censorship. I commit to treating students fairly. Finally, science is for everyone, period.

COURSE MANAGEMENT AND POLICIES

Student Code of Conduct and Scholastic Dishonesty/Copyright and Fair Use The Student Code of Conduct is Section B of the Appendices in the Student Information Bulletin. Scholastic Dishonesty is listed in the Student Code of Conduct (Sec. B of the Appendices) under Sec. 203. It is important to understand the issue of intellectual property rights. You may not use the images or thoughts of others for profit or gain without their written permission. The UTSA library has a Copyright Laws and Public Performance Rights (PPR) page.

Clarification of the above for this class: For this class, the following will also be considered as examples of academic dishonesty: using AI to write code, copy and pasting code from fellow classmates.

Students with Disabilities
The University of Texas at San Antonio, in compliance with the Americans with Disabilities Act and Section 504 of the Rehabilitation Act, provides "reasonable accommodations" to students with disabilities. Only those students who have officially registered with Student Disability Services and requested accommodations for this course will be eligible for disability accommodations. Instructors at UTSA must be provided official notification of accommodation through Student Disability Services. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.utsa.edu/disability or by calling Student Disability Services at (210) 458-4157. Accommodations are not retroactive.

Family Educational Rights and Privacy Act (FERPA)
FERPA grants students the right to control certain disclosures of their educational records. For a full explanation of your rights and to grant access to FERPA educational records, go to Student Catalog Annual FERPA Letter and One Stop Enrollment – FERPA Proxy Access. Without your consent or authorization of proxy access, UTSA may release Directory Information, such as but not limited to your name, email, phone, place of birth, and photograph, unless you have opted out of the release of Directory Information. To opt out, go to Restrict Directory Information Form. ***Mandatory Reporting of Sexual Misconduct and Reporting of Health and Safety Information***: If a student discloses an incident of sexual misconduct to any UTSA employee (other than to a designated confidential employee such as mental health counselor or PEACE advocate, a UTSA police officer using a pseudonym form or at a public awareness event), that information is not confidential, and the UTSA employee must report all known information to the UTSA Office of Equal Opportunity Services. Employees may also report any concerns about the health and safety of students or others to other school officials and/or law enforcement. For a complete list of exceptions to FERPA, please see Student Catalog Annual FERPA Letter and HOP 5.01.

Video and audio recording
As the instructor of this course, I may record meetings and lessons. You are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Recordings may not be published, reproduced, or shared with those not in the class. If the instructor or a UTSA office plans any other uses for the recordings, consent of the students identifiable in the recordings is required before such use unless an exception is allowed by law. For more information on your privacy and class recordings, review Student Privacy (FERPA) in Virtual Classrooms and Other Educational Recordings and the Guide to Secure Video Conferencing Tools.

Firearm Exclusion Zone Notice
I invoke my right to designate my office as an exclusion zone. Pursuant to HOP 9.48, Carrying of Concealed Handguns on Campus, my private office, AET 3.344, is a designated exclusion zone. As set out in Section 30.06, Penal Code (trespass by

license holder with a concealed handgun), a person licensed to carry a Concealed Handgun under Subchapter H, Chapter 411 Government Code (handgun licensing law), may not enter this property/office with a concealed handgun. De conformidad con HOP 9.48, Llevar Armas de Fuego Encubiertas en el Campus, mi oficina privada, AET 3.344, es una zona designada de exclusin. Conforme a la seccin 30.06 del cdigo penal (trespasar portando armas de fuego) personas con licencia bajo del sub-capitulo H, capitulo 411, codigo de gobierno (ley de portar armas), no deben entrar a esta propiedad portando un arma de fuego.

NOTE: **The syllabus is subject to change at the instructor's discretion**. Any changes/corrections to the course materials, assignment dates, or other updates will be communicated to the students ahead of time

PROVISIONAL COURSE SCHEDULE (HIGHLY SUBJECT TO CHANGE!!!!!)

Below is a rough outline of the course. I have the details defined for the first 2-3 weeks. The rest will be filled in later.

Week 1: August 21-23

M - Setting up Anaconda, Course Overview; Crash course in Python (printing, reading/writing/prompting, functions)

Notes: Basic, Pt. 1-3
Reading: Ch2 - McKinney, Ch2 - Fuhrer
Homework: Python interests survey

W - Basic Python: If-Then Statements; For-Loops; Data Structures – Lists, Tuples, Arrays, and Dictionaries; Slicing

Notes: Basic, Pt. 4 and 5; Data Structures: 1-3
Reading: Ch3 - McKinney, Ch3 - Fuhrer
Homework: none

Week 2: August 28-30

M - Data Structures - A Brief Review of Tuples, Lists, and Arrays; Dictionaries; Slicing; Comprehensions; Sequence Functions, and Lambda Functions; NumPy - Numpy Arrays (briefly)

Notes: Data Structures: 2-3; NumPy: 1

Reading: Ch 3, Ch 4.1- McKinney; Ch 3; 4.1, 4.3 - Fuhrer
Homework: Data Structures 2-3, NumPy 1

W - Numpy - Numpy Arrays, Array Arithmetic and Universal Functions Slicing & Reshaping, More Array Operations, and Broadcasting

Notes: NumPy: 1 (last part), 2-4
Reading: Ch 4.1, 4.3, Appendix A2, A3 - McKinney; Ch 4.6, 4.7, 4.8 - Fuhrer
Homework: NumPy 2-4

Week 3: September 6

W – NumPy/Broadcasting; Numpy/SciPy Basic Linear Algebra, Solving Linear Equations

Notes: NumPy: 5
Reading: Ch 4.6 - McKinney; Ch 4.1.3, 4.2, 4.9 - Fuhrer
Homework: NumPy 5

Week 4: September 11-13

M – Numpy/Advanced linear algebra: singular value decomposition
W – Numpy/Advanced linear algebra: eigendecomposition, PCA

Notes: NumPy: 6, 7
Reading: Ch 4.6 - McKinney; Ch 4.2, 4.9 - Fuhrer
Homework: NumPy: 6-7

Week 5: September 18-20

M – Numpy/Advanced linear algebra: PCA
W - Numpy/Advanced linear algebra: PCA; SciPy/optimization: root finding

Notes: NumPy: 8,9; Intro SciPy 1
Reading: van der Plas/Ch 5-09
Homework: NumPy 8,9

Week 6: September 25-27

M – SciPy/optimization: root finding and minimization; SciPy/interpolation
W – SciPy/signal processing and statistics [signal processing, statistics]

Notes: Intro SciPy 1,2,3

Reading:
Homework: SciPy 1,2,3

Week 7: October 2-4

M – Matplotlib

W – Matplotlib

Notes: Matplotlib 1,2
Reading: vander Plas 4.00-4.10; McKinney, Ch 9; Fuhrer, Ch 6
Homework: Matplotlib 1,2

Week 8: October 9-11

M – Matplotlib

W – Matplotlib

Notes: Matplotlib 3,4
Reading: vander Plas 4.00-4.10; McKinney, Ch 9; Fuhrer, Ch 6
Homework: Matplotlib 3,4

Week 9: October 16-18

M - Pandas

W - Pandas, AstroPy

Week 10: October 23-25

M – AstroPy [over Zoom]

W – AstroPy, Classes [over Zoom]

Week 11: October 30-November 1

M - AstroPy

W - AstroPy

Week 12: November 6-8

M – Classes

W – Statistical Tests

Week 13: November 13-15

M – Classes

W – Statistical Tests

Week 14: November 20

M – Machine Learning

Week 15: November 27-29

M – Machine Learning

W – Machine Learning?

Week 16: December 4-6

TBD