COMS W3132: ~~Advanced~~ Intermediate Computing in Python

Columbia University
Spring 2025

Dr. Daniel Bauer, original slides by Jan Janak

Teaching Team

- Dr. Daniel Bauer <bauer@cs.columbia.edu>
  (Senior Lecturer in Computer Science)
  https://www.cs.columbia.edu/~bauer
  Office: 704 CEPSR

- Assistants:

Darien Moment	dem2187@columbia.edu	Wed 4:00-5:30pm
Sophie Tsanang-Tigoumo	sgt2125@columbia.edu	Thu 10:30am-12pm
Kavika Krishnan	kk3526@columbia.edu	TBA
Julien Remy	jr4404@columbia.edu	Tue 5:00-6:30pm

Columbia Curriculum for CS Non-majors

Humanities

Engineering

COMS W1002

- Python Basics
- Art
  Economics
  Linguistics
  Social Sciences

ENGI E1006

- Python Basics
- Data Processing
  Machine Learning
  Visualization
  Networking

COMS W2132 (NEW!)

- Python Software Engineering
- Data Structures and Algorithms
- Context-Specific Individual Project

What is this Course About?

- “Intermediate course in computing for CS non-majors”
- A follow-up course to COMS W1002 and ENGI E1006
  - You asked for it, we listened…
  - Fill a gap in Columbia’s Python curriculum
- Introduction to Python software engineering
- Essential data structures and algorithms in Python
  - Practical alternative to core CS Java data structure classes (COMS 3134 / 3137)
- 4-6-week individual project to deepen your Python skills
  - Learn to design, develop, and publish a Python program

This Course is NOT About

- Python language basics.
- Rigorous theory of data structures and algorithms.
- Passive learning (you will need to write programs and design / work on your own project)

Learning Objectives

By the end of this course, you should

1. understand essential data structures (linked lists, stacks, queues, trees, graphs) and their space / time requirements. Be able to decide when to use which data structure.
2. be able to implement the data structures and corresponding algorithms in Python.
3. apply data structures and algorithms in a non-trivial project.

Disclaimer

- This is a new course.
  - Based on requests from COMS W1002 and ENGI E1006 students
- Please expect adjustments, changes, and course corrections
- Feedback is greatly appreciated

Appeal to CS non-majors: Please help us find a model for W2132 that is useful to you!

Programming / Python Prerequisites

Concepts you should know from COMS W1002 or ENGI E1006

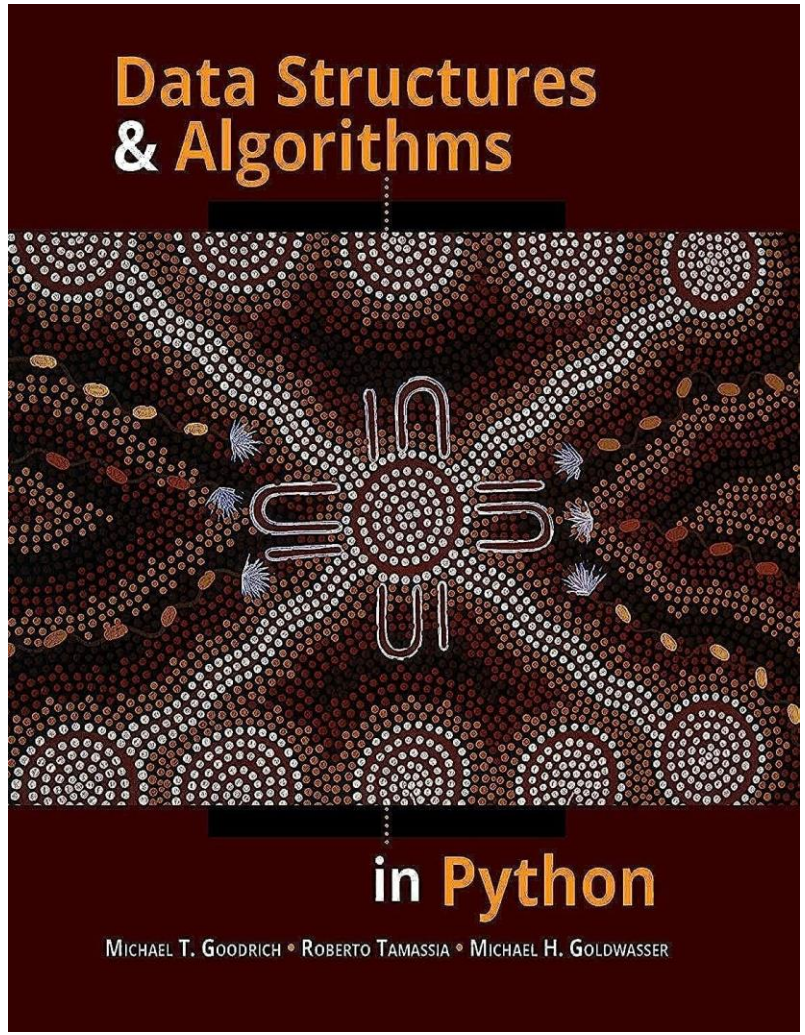
1) Names, variables, and expressions
2) Control structures (**if** and **switch** statements)
3) Iteration structures (**for** and **while** loops)
4) Functions
5) Classes and objects (basic object-oriented programming)

Review in Lab this Friday. See topic checklist on Courseworks:
https://courseworks2.columbia.edu/courses/214113/pages/python-basics-checklist

Recommended Textbook

- Excellent textbook for CS non-majors !
- The right level of rigor
- Lectures will mostly follow the textbook
- Recommended but not required
- Columbia Library ebook:

  https://clio.columbia.edu/catalog/17854614



Goodrich, Tamassia, Goldwasser, "*Data Structures & Algorithms in Python*", 1st edition, Wiley, 2013

CourseWorks

- Announcements
- Syllabus
- Important Times & Dates
- Contact Information
- Course Materials
- Homework
- Gradebook

Getting Help

Computer science and programming can be difficult for people at all levels.
Getting stuck is normal!
Ask for help early and often!

- Ed Discussions (online forum)
- TA and Instructor office hours
- Friday Labs!

Lectures and Labs

- Lectures: Mon/Wed 2:40-3:55pm, 451 CSB
- Labs (Mandatory and possibly more useful): Fri 1:10-2:25pm, 313 Fayerweather
- Both in-person.
- Attendance mandatory and graded. If you need to miss a session, please let the instructor know.

Attendance Poll - January 22, 2025

PollEv.com/danielbauer757

- Login with your Columbia UNI.
- Make sure account is associated with your real name.
- “Vote” for the word of the day.

Grading

3 individual homework assignments	24% (8% each)
Midterm Exam (in class, 65 minutes)	10%
Final Exam (120 minutes)	10%
Attendance & Participation	6%
Project	**50%**

Individual Homework Assignments

- Mostly first half of the semester
- Will cover data structures and algorithms, programming and theory (written answers)
- Must be completed individually
- Two-week windows to complete
- Submit through GitHub Classrooms (to be discussed)

Academic Honesty Policy

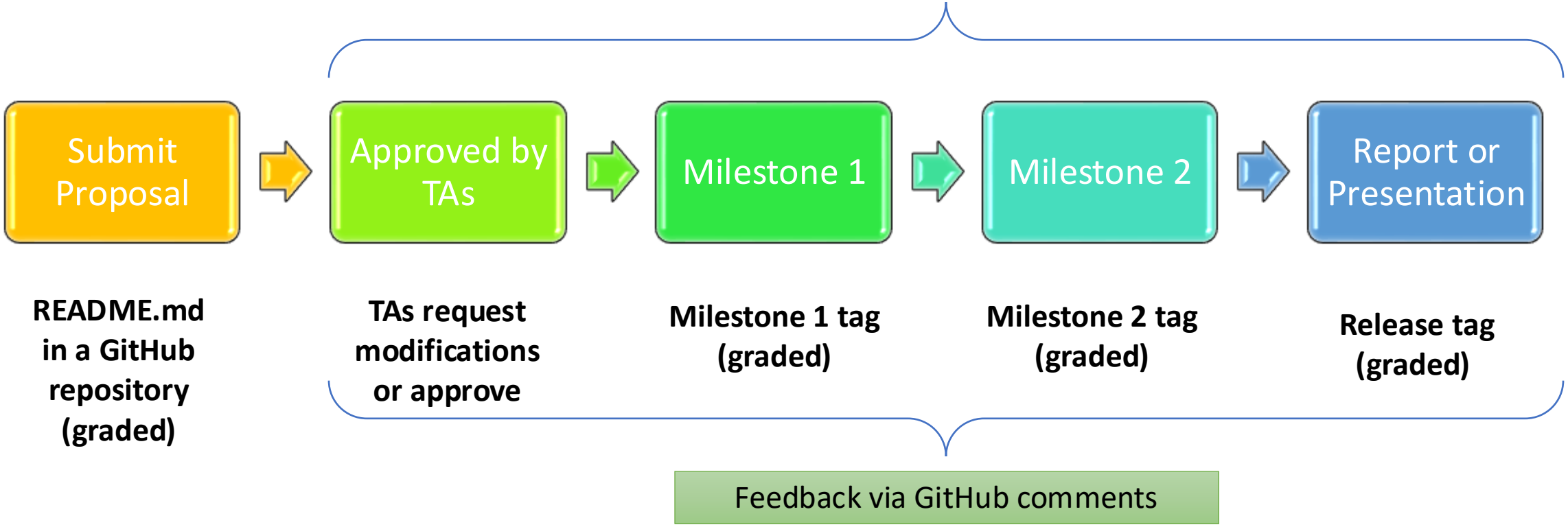
- Part 1, collaboration policy:
  - Individual homework assignments and project.
  - You are encouraged to **discuss** classroom material, problems, and homework with other students, the TAs, and the instructor.
  - You can work out the theory together on the whiteboard or paper.
  - **However, you must do all final writing and programming by yourself without further interaction.**
  - Do not share your solutions with others.
  - Do not copy anyone else's work without attribution.

Academic Honesty Policy Cont'd

- Part 2, AI and online resources policy:
  - You must attribute **any** code taken from an online source (full URL, AI prompt, …).
  - Only permitted to copy trivial, short code snippets ("how do I open/read from a file again?")
  - Do not use AI to solve full exercises or large portions of your project.

  - This course focuses on fundamentals – we recommend you use generative AI as little as possible.

Project: Workflow on GitHub



Similar to typical open-source software development workflow on GitHub

Project Proposal

- Short README.md addressing the following:

    1) What are you making?
    2) Why is it important, relevant or interesting?
    3) Data structures, tools, libraries, or services you intend to use
    4) Proposed milestones
    5) Optional: How do you want to publish your work?

We will provide an example when the time comes

How to Pick a Project?

- Try to think of a problem that is fun or important to you
- Perhaps there is a tool or a script that you have always wanted?
- Maybe you have an idea for a web service or website?
- Maybe you have a chore that could be automated?
- If you have a hobby, can you use Python to get better at it?
- There are no bad ideas (only undeveloped). If it is something you want, others probably want it too.

Pick a Project Area (examples)

- Computational linguistics / NLP
- Computational biology
- Social sciences
- AI and machine learning
- Systems, networking, Internet
- Data analysis and databases
- Computing in the arts

Pick a Project Area (continued)

- **Computational linguistics / NLP**
  - understand or respond to text (Llama, GPT), information extraction, …
- Computational biology
- Social sciences
- AI and machine learning
- Systems, networking, Internet
- Data analysis and databases
- Computing in the arts

Pick a Project Area (continued)

- Computational linguistics / NLP
- **Computational biology**
  - Simulations and modeling of biological systems
- Social sciences
- AI and machine learning
- Systems, networking, Internet
- Data analysis and databases
- Computing in the arts

Pick a Project Area (continued)

- Computational linguistics / NLP
- Computational biology
- **Social sciences**
  - Economics, political science (analyze datasets), education, environmental studies, law
- AI and machine learning
- Systems, networking, Internet
- Data analysis and databases
- Computing in the arts

Pick a Project Area (continued)

- Computational linguistics / NLP
- Computational biology
- Social sciences
- **AI and machine learning**
  - General AI or machine learning to present/analyze datasets, games, Peter Norvig's Pytudes
- Systems, networking, Internet
- Data analysis and databases
- Computing in the arts

Pick a Project Area (continued)

- Computational linguistics / NLP
- Computational biology
- Social sciences
- AI and machine learning
- **Systems, networking, Internet (talk to Jan)**
  - Internet services, websites, APIs, network analysis
- Data analysis and databases
- Computing in the arts

Pick a Project Area (continued)

- Computational linguistics / NLP
- Computational biology
- Social sciences
- AI and machine learning
- Systems, networking, Internet
- **Data analysis and databases**
  - Analyze public datasets, Google BigQuery, and Datalab
- Computing in the arts

Pick a Project Area (continued)

- Computational linguistics / NLP
- Computational biology
- Social sciences
- AI and machine learning
- Systems, networking, Internet
- Data analysis and databases
- **Computing in the arts**
  - Video / Audio demo, algorithmic art, games,
    Embedded systems (MicroPython), Internet of Things, Raspberry Pi

Projects by Application Type

1) Command line tool
2) GUI program (TkIntern, PyGame, PyQT, Web UI)
3) Internet service (API) or website
4) Jupyter Notebook
5) Embedded program (Raspberry Pi)

Some Pointers for Inspiration

- Peter Norvig's Jupyter Notebooks: https://github.com/norvig/pytudes
- Data science: https://www.dataquest.io/path/data-analyst/
- Machine learning: https://scikit-learn.org/stable/index.html
- Building websites with Django: https://www.tangowithdjango.com/
- Making games with Pygame: https://www.pygame.org/wiki/tutorials
- Learning robotics using Python (book): https://www.amazon.com/dp/B00YEVZ6UK
- Programming the Raspberry Pi (book): https://www.adafruit.com/product/1089

We will also offer a library of more concrete ideas later this semester