

Class Project

Your final project for this class is to create a program that does something new and interesting, potentially in your own field of study, although that is not a strict requirement.

This is 15% of your final grade.

The size of your project should be roughly equivalent to TWO labs. This means a program which inputs data from one small file and produces a single type of graph output -- is not large enough. You want to look for problems that have enough data which your program must do multiple operations on, perhaps a problem that has multiple avenues for analysis. Related, merging two types of data from two sources into one useful output/analysis is a great idea. You aren't limited to data, though, and the following are suggestions:

Suggested Topic Areas

Data Analysis: This will probably be the most popular project area. Find an interesting dataset and write a program that helps to analyze what is in the data. You'll want to look for large datasets or datasets that allow for multiple avenues of analysis. Here are just a few websites with open datasets:

- [US Govt Datasets](#) (some cool stuff in here)
- [Kaggle datasets](#) (a bunch of junk, but also some good nuggets hidden in here)
- [Decent List of Dataset links](#)
- [Full text of novels](#)
- [Lots of weather data](#)
- Tons of tweets (ask Prof Chambers for access)
- GOOGLE SEARCH - searching for your topic name with "CSV" or "dataset" might find what you're looking for!

Build an "App": Lots of websites have APIs where you can call them from code. Go back to our ChatBot Lab 3 to remind yourself we queried Wikipedia! This type of project would be a friendly user program that queries websites for the user without requiring a web browser. Many websites have APIs, like Wikipedia, Reddit, and Yahoo Finance. You could make a program read and post to Reddit, for instance.

Machine Learning: You could build a machine learning program to make predictions from data. The upcoming Author ID lab introduces one such machine learner. The main characteristic here is that you want to find some labeled data for your topic of interest. For instance, data about home sales have known sales prices. Speeches made on the Senate floor have known political parties by who made the speech. Stocks have closing prices (up or down). Your project would mainly focus on inputting your data, getting it into a learnable format, and run the machine learner on it.

Image Manipulation: We've had a couple labs with images (Lab 2, 4, and event 7). Image manipulation can quickly become complex and interesting. For instance, perhaps you liked Lab 7 with the GPS tracker but didn't get to the drawing stage. You could find a different GPS dataset online, learn to parse it, and do some drawing on the map to analyze it.

Simulation or Games: Your project doesn't have to involve data. Perhaps you are studying a field like Physics and want to simulate some natural phenomena. Or perhaps you want to make a video game! We haven't done any labs in this area, so this topic would be a good exploratory one with sufficient challenge.

Research Project: Some of you have upcoming research ideas. A valid project is doing some preliminary programming for a chunk of the research. Finding a result is not required in this case, but the program must be of sufficient complexity and usefulness to the future research.

Last Year's Topics: Varsity Sport Stats Tracker, MechE material strength simulator, Eye Pupil Detector, Simulate Canons, Zillow Data Analysis, Buying v Renting Real Home Data, Stock Price Prediction, Baking Recipe Suggestions, Book Word Usage.

Pick a Topic

Due March 25 (5% penalty each day late)

Write one substantial paragraph describing the topic and your general approach to programming a solution for it. You must describe at a minimum:

1. Expected input data and type of output.
2. Clear description of major steps to accomplish in your program.
3. Clear description of why this is larger than a lab and worthy of a project.

Write Your Program

Due May 2 @ 2359 (10% penalty each day late)

The following is required upon turnin of your project:

1. All of your Python code, thoroughly documented with code comments.
2. A **readme.txt** file that clearly shows how to run your code and specifies what data files (if any) are needed. You must include the python command that successfully runs your program, and include examples of user input (if any) that is required once the program is running.
3. All input data files (if any)
4. Output products -- such as generated graphs, plots, or images. Some projects will not have these, others may have many.

Presentation to the Class

Due May 3

Show the class! Our final lab time will be devoted to presentations. You will have ~10 minutes each to show your code (yes you'll show the code), explain any data used, and show the program running. You may use slides and visuals when appropriate. Some project topics might be more code-based, and others data analysis which relies on output analysis.

What to turn in

1. **Code comments:** did you comment your code? Every function should have comments at the top. Each main 'paragraph' of your program should have a short comment explaining what that portion of your code does.
2. **readme.txt:** Include your name, a title for your project, and very specific instructions on how to run your program. You should give me the exact python3 command to run, and then explain any data files that it might need access to. Does your program use a special library? Tell me so I know what to install.
3. **All required files:** if your program uses data files, those should be uploaded with your program files.
4. **Upload:** visit [the submit website](#) and submit to the 'FinalProject' assignment. Upload all required files.

If I cannot run your program based on your submission, your grade will be penalized.