

Python Computing for Data Science

Files for Today:

<https://goo.gl/dyKHxx> = this PDF

<https://goo.gl/N4r9xX> = Jupyter Notebook

Files for the Course:

git clone <https://github.com/profjsb/python-seminar.git>

(if you dont have *git*, please set it up later)

Signup (Piazza):

<https://piazza.com/berkeley/spring2018/ay250class13410>

Welcome to the *Python Computing for Data Science* Seminar

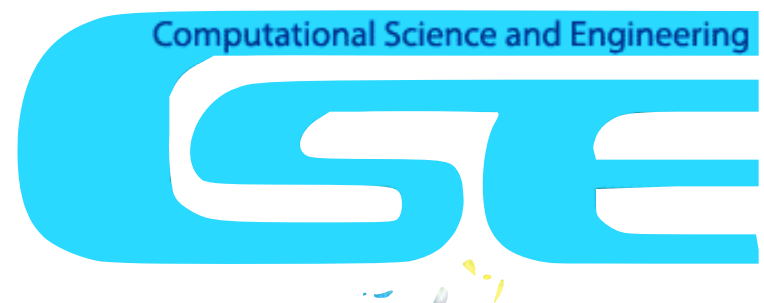
AY 250: Monday 2-5pm (Campbell Hall 131)

Instructor: Josh Bloom

GSI: Chelsea Harris



Instructor+GSI email:
ucbpythonclass+seminar@gmail.com

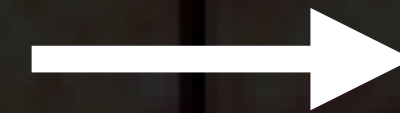


Award #0941742

Democratizing Trends in the Sciences

Data

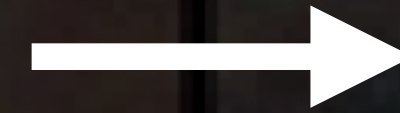
Decreasing cost to obtain,
move, store



open data,
more freely shared

Compute

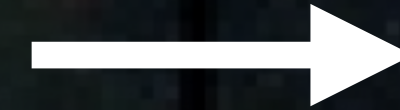
Decreasing cost,
increasing specialty



more accessible

**Technology/
Methodology**

Algorithmic innovation,
software tooling



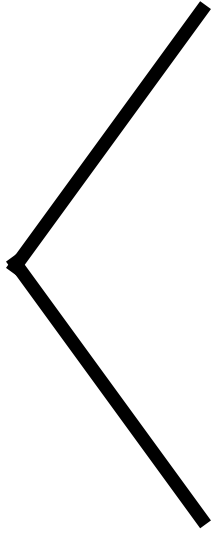
open source

If anyone can get Data, Compute, Tech... who wins?

domain expertise → She who asks the right questions

+

She who answers questions better & faster than others:

data science 

- ▶ computational access
- ▶ methodological inference (e.g., machine learning)
- ▶ better story telling, dissemination of results
- ▶ reproducibility (acceptance)

Motivation:

short version

FTW: leverage the Python ecosystem to do cutting-edge research

long version

- 1) get you using Python to do cutting-edge research in the physical, biological and/or social sciences
- 2) help you realize that Python is a viable framework to do just about any 21st century problem well (and costs zero). “Super Glue”
- 3) fold you into the Python community so you know how to navigate it yourself and so it potentially benefits from having you part of it

How we plan to do this:

- "formal" lectures on specialized topics each week by leading experts & local practitioners (Monday)
- "breakout work sessions" interspersed within the lectures
- homework assignments based on week's lecture
- final project

Prerequisites:

- working knowledge (or more) of the core Python language and/or Python BootCamp graduate
- installation of Python (3.6.X), scientific 3rd party packages (Anaconda distro), & git
- laptop for use in class and for homeworks
- tolerance for our terrible computer humor

<http://www.pythonbootcamp.info/preparation/software>

piazza for real-time/off-line interaction

homework updates, solutions, ...

<https://piazza.com/berkeley/spring2018/ay250class13410/home>

The screenshot shows the Piazza web interface for a course. The top navigation bar includes the Piazza logo, the course identifier 'AY 250', and tabs for 'Q & A', 'Course Page', and 'Manage Class'. A user profile for 'Joshua Bloom' is visible in the top right. Below the navigation bar, there are links for homework assignments: 'hw1', 'hw2', 'hw3', 'hw4', and 'logistics'. The main content area is divided into two columns. The left column contains a 'New Post' button, a search bar, and a list of pinned posts: 'Cloning the Git Repo for the ...', 'Getting up and running with ...', and 'First Day of Class Tomorrow'. The right column displays a specific note titled 'Cloning the Git Repo for the Class' by 'Josh'. The note contains instructions on how to clone a Git repository and pull updates. The note has 2 views and was posted 1 minute ago. Below the note, there is a section for 'followup discussions' and a button to 'Start a new followup discussion'.

piazza AY 250 Q & A Course Page Manage Class Joshua Bloom

hw1 hw2 hw3 hw4 logistics

Unread Updated Unresolved Following

New Post Search or add a post...

PINNED

- Instr **Cloning the Git Repo for the ...**
- Instr **Getting up and running with ...**
- Instr **First Day of Class Tomorrow**

TODAY

Welcome to Piazza!

Note History:

note stop following **2 views**

Cloning the Git Repo for the Class

You should do this once (it will take awhile) while on a good network:

```
git clone https://github.com/profjsb/python-seminar.git
```

Then, at the start of each class:

```
git pull
```

(If you don't have git, see the [bottom of this link](#)).

Josh


logistics

edit • good note | 0 1 minute ago by Joshua Bloom

followup discussions for lingering questions and comments

Start a new followup discussion

github is the main data portal for us...

PUBLIC  profjsb / python-seminar

[Pull Request](#) [Unwatch](#) [★ Star](#) 1 [Fork](#) 0

[Code](#) [Network](#) [Pull Requests](#) 0 [Issues](#) 0 [Wiki](#) [Graphs](#) [Admin](#)


Python Seminar Course at UC Berkeley (AY 250) — [Read more](#)

[Clone in Mac](#) [ZIP](#) [HTTP](#) [SSH](#) [Git Read-Only](#) [Read+Write access](#)






[branch: master](#) [Files](#) [Commits](#) [Branches](#) 2 [Tags](#) [Downloads](#)

🕒 Latest commit to the **master** branch

setting up basic structure

 **profjsb** authored 6 minutes ago [commit ad44019458](#)

python-seminar /

name	age	message	history
 Breakouts	6 minutes ago	setting up basic structure [profjsb]	
 DataFiles_and_Notebooks	6 minutes ago	setting up basic structure [profjsb]	
 Lectures	6 minutes ago	setting up basic structure [profjsb]	
 LICENSE	6 minutes ago	setting up basic structure [profjsb]	
 README.md	4 days ago	fixed classroom number [profjsb]	

Scientific Research Computing with Python

Files for Today:

<https://goo.gl/dyKHxx> = this PDF

<https://goo.gl/N4r9xX> = Jupyter Notebook

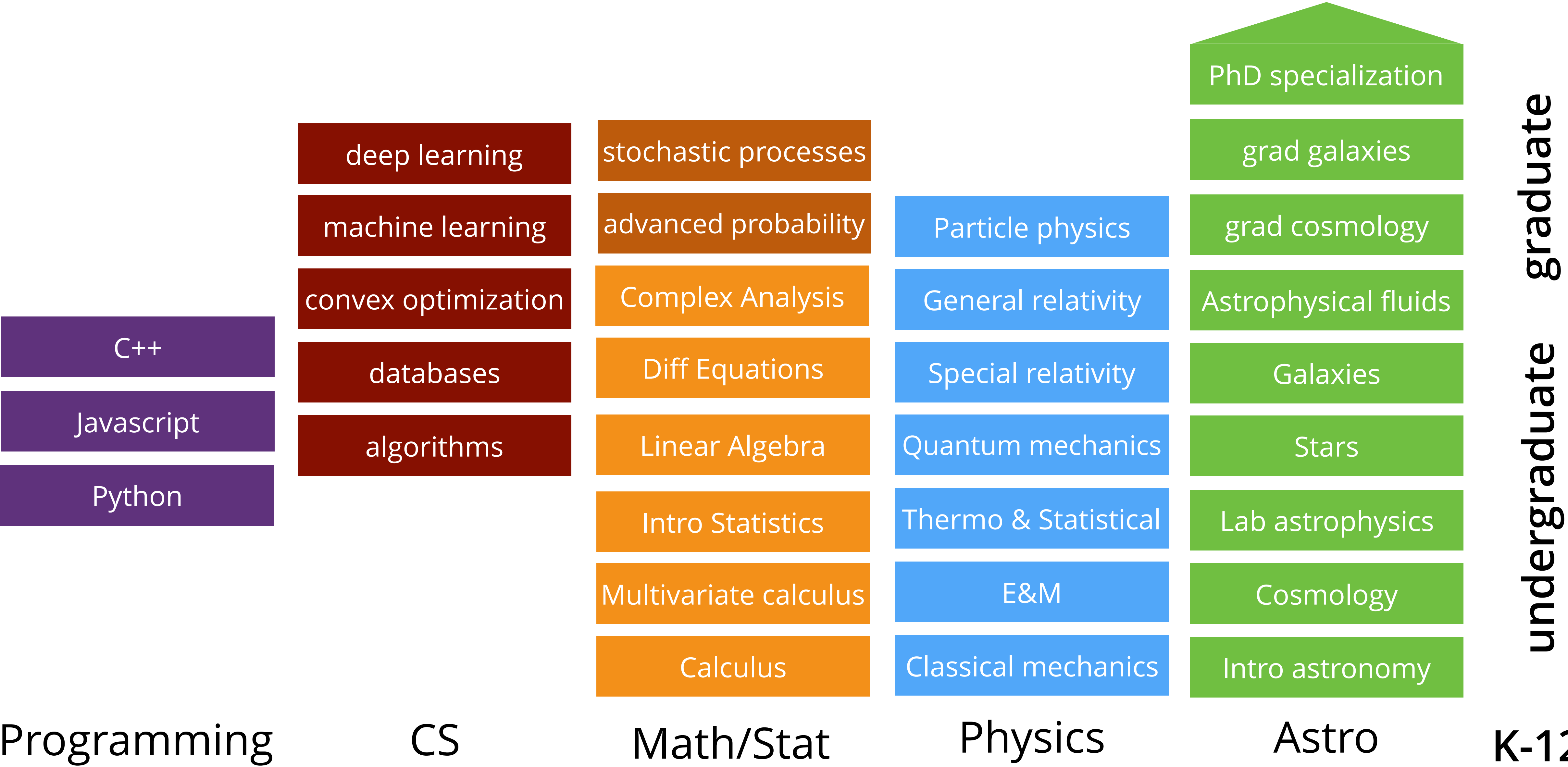
/DataFiles_and_Notebooks/00_AdvancedPythonConcepts = follow-along files

Jan 22	Advanced Python Language Concepts (decorators, OrderedDict, Generators, Iterables, Context Managers)	- GIT - scipy §2.1	Josh
Jan 29	Pandas, Scipy, & Numpy	- scipy §§ 1.3, 1.5, 2.2 - numpy - skim chap 4/5 of McKinney	Josh
Feb 5	Data vizualization (Matplotlib, Bokeh, Altair, Plotly, mayavi)	- Skim Tufte's Vizualization book - colormap talk (Scipy 2015)	Josh
Feb 12	Interacting with the world (requests, email, IoT/pyserial)	None	Josh
Feb 19	Holiday (no class)		
Feb 26	Parallelism (asyncio, dask, IPython cluster)	- [ipyparallel docs] (http://ipyparallel.readthedocs.io/en/latest/intro.html)	Josh
Mar 5	Database interaction (sqlite, postgres, SQLAlchemy, peewee), Large datasets (xarray, HDF5)	None	Josh
Mar 12	Machine Learning I (sklearn, NLP)	None	Josh
Mar 19	Machine Learning II (keras [tensorflow])	None	Josh
Mar 26	Spring Break		
Apr 2	Image processing (OpenCV, skimage)	None	Stefan van der Walt
Apr 9	Web frameworks & RESTful APIs, Flask	None	Josh
Apr 16	Bayesian programming & Symbolic math	Probabalistic Programming eBook install: pip install pymc3	TBD
Apr 23	Speeding it up (Numba, Cython, wrapping legacy code)	TBD	Josh
Apr 30/Onward	final project work		

Course Schedule

<https://github.com/profjsb/python-seminar>

Challenge of Data-Driven Domain Education Stack



Concepts/Practices in this Course

- Jupyter & JupyterLab
- using git & github
- Docker
- Data science workflows
- reproducible research
- application building
- debugging
- testing

"Data science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured, which is a continuation of some of the data analysis fields such as statistics, data mining, and predictive analytics..."

-wikipedia

Workflow for a *typical* week

Friday:

email from next week's instructor w/ special installation instructions, reading/tutorials

Monday:

2:00 cd python-seminar; git pull

2:10 - 3pm Intro topics Lecture

3 - 3:30pm Breakout coding

3:30-4:50pm Detailed topics lecture (+stretch)

4:50-5:00pm Work on homework

Thursday?:

TBD Supervised help with homework [place TBD]

Monday Morning:

Homework project due

Course Grade

- 10% participation in lectures/breakouts
- 60% Problem Sets
There will be 10 assignments. Do at least 6.
We will keep your best 5.
- 30% Final Project, due May 10 (no final exam)

Final Project

a) Build a substantial framework for doing something in your own research, based on at least two topics from different weeks. Something you will use for a long time...

e.g., image analysis package, hardware control software, a webservice that does some crunching under the hood, provide a parallelization of some algorithm or code you use, etc.

- or -

b) Contribute code/functionality to an open-source Python project (Jupyter, scipy, Cython, numpy, matplotlib, etc.)