

# *Python Computing for Data Science*

## **Files for Today:**

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

<http://goo.gl/XFiZYI> = 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):**

<http://piazza.com/berkeley/fall2016/ay250class13410>

# Welcome to the *Python Computing for Data Science* Seminar

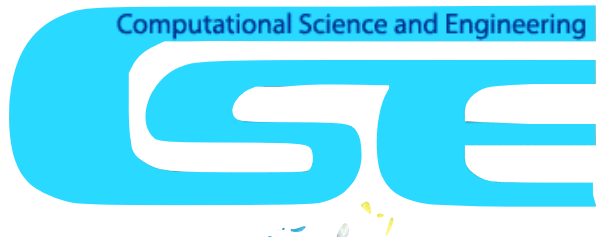
AY 250: Friday 1-4pm (Campbell Hall 131)

*Instructor:* Josh Bloom

*GSIs:* Goutam Murlidhar, Hadrien Renold



Instructor+GSI email:  
[ucbpythonclass+seminar@gmail.com](mailto:ucbpythonclass+seminar@gmail.com)



Award #0941742

## Motivation:

### short version

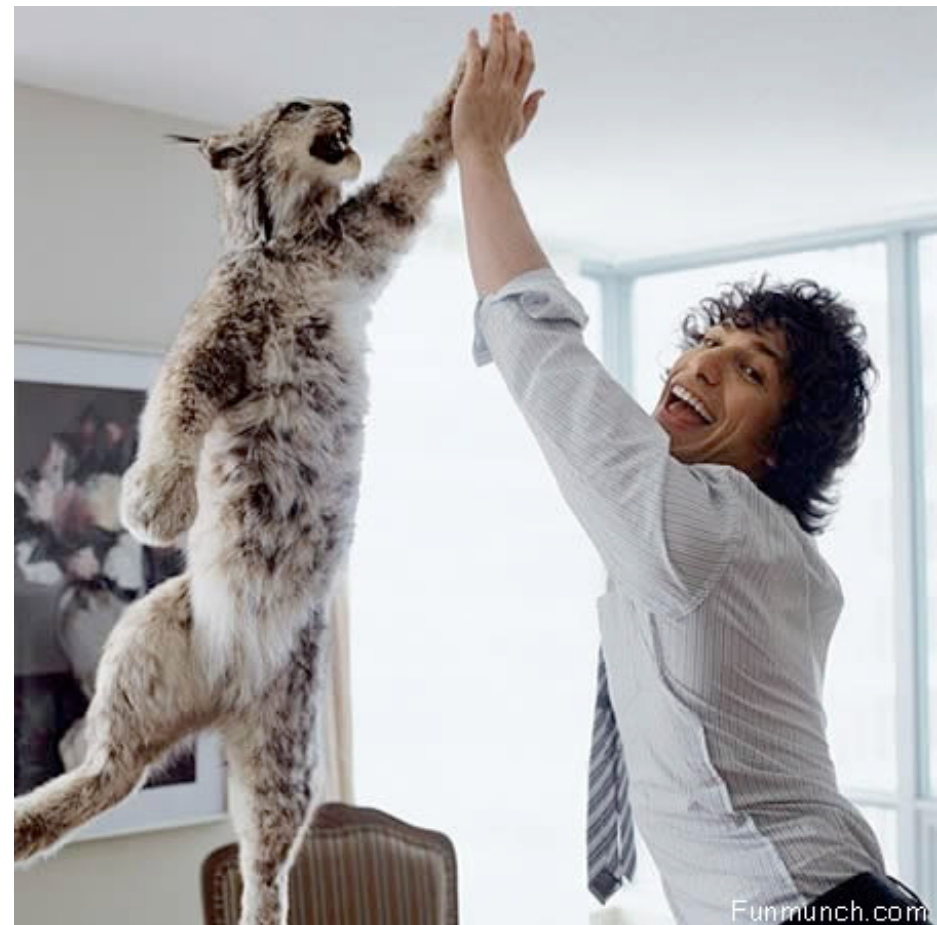
get you using Python to do cutting-edge research

### long version

- 1) get you using Python to do cutting-edge research,
- 2) helping 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 it 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 (Friday)
- "breakout work sessions" interspersed within the lectures
- homework assignments based on week's lecture
- final project



**Bring me**

**a shrubbery!!**



**I DON'T ALWAYS USE MEMES  
WHEN I TEACH PYTHON**

**BUT WHEN I DO:  
IMPORT  
AWESOMENESS**



# Prerequisites:

- working knowledge (or more) of the core Python language
  - and/or -

Python BootCamp graduate

- installation of Python (3.5.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, ...

[piazza.com/berkeley/fall2016/ay250class13410/home](https://piazza.com/berkeley/fall2016/ay250class13410/home)

**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

git@github.com:profjsb/python-seminar.git

 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:

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

<http://goo.gl/XFiZYI> = Jupyter Notebook

/DataFiles\_and\_Notebooks/00\_AdvancedPythonConcepts = follow-along files

# Course Schedule

Date	Content	Reading	Leader
Aug 26	<b>Advanced Python Language Concepts</b> (decorators, OrderedDict, Generators, Iterables, Context Managers)	- <a href="#">GIT</a> - <a href="#">scipy §2.1</a>	Josh
Sep 2	<b>Pandas, Scipy, &amp; Numpy</b>	- <a href="#">scipy §§ 1.3, 1.5, 2.2</a> - <a href="#">numpy</a>	Josh
Sep 9	<b>Data vizualization</b> (Matplotlib, Bokeh, Altair, Plotly, mayavi)	TBD	Josh
Sep 16	<b>Interacting with the world</b> (requests, email, IoT/pyserial)	TBD	Josh
Sep 23	<b>Parallelism</b> (asyncio, dask, IPython cluster)	TBD	Josh
Sep 30	<b>Database interaction</b> (sqlite, postgres, SQLAlchemy), <b>Large datasets</b> (xarray, HDF5)	TBD	Josh
Oct 7	<b>Image processing</b> (OpenCV, skimage)	TBD	<i>Stéfan van der Walt</i>
Oct 14	<b>Machine Learning I</b> (sklearn, NLP)	TBD	Josh
Oct 21	<b>Machine Learning II</b> (keras [tensorflow])	TBD	Josh
Oct 28	<b>Bayesian programming &amp; Symbolic math</b>	TBD	<i>Brett</i>
Nov 4	<b>Web frameworks &amp; RESTful APIs, Flask</b>	TBD	Josh
Nov 11	holiday		
Nov 18	<b>Computational Frameworks</b> (Docker, AWS, Azure, AWS-Lambda)	TBD	Josh
Nov 25	holiday		
Dec 2	<b>Speeding it up</b> (Numba, Cython, wrapping legacy code)	TBD	Josh

# 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

Thursday:

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

Friday:

1:00	<code>cd python-seminar; git pull</code>
1:10 - 2pm	Intro topics Lecture
2 - 2:30pm	Breakout coding
2:30-3:50pm	Detailed topics lecture (+stretch)
3:50-4:00pm	Work on homework

Monday:

TBD Supervised help with homework [place TBD]

Friday Morning:

Homework project due

# Course Grade

10% participation in lectures/breakouts

60% Homeworks

there will be 11 assignments & we will drop your lowest score

30% Final Project, due Dec 12  
(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 a major open-source Python project (Jupyter, scipy, Cython, numpy, matplotlib, etc.)