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

Computational Science and Engineering

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

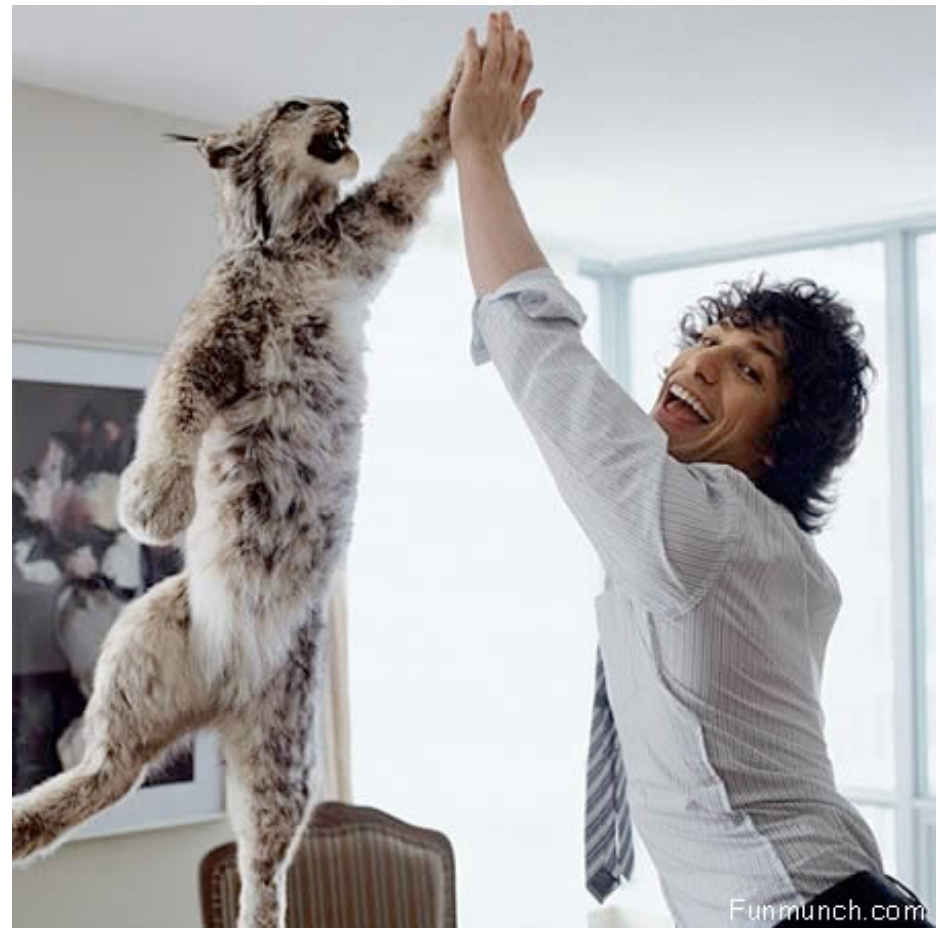
How we plan to do this:

- "formal" lectures on specialized topics each week by leading experts & local practitioners (Friday)

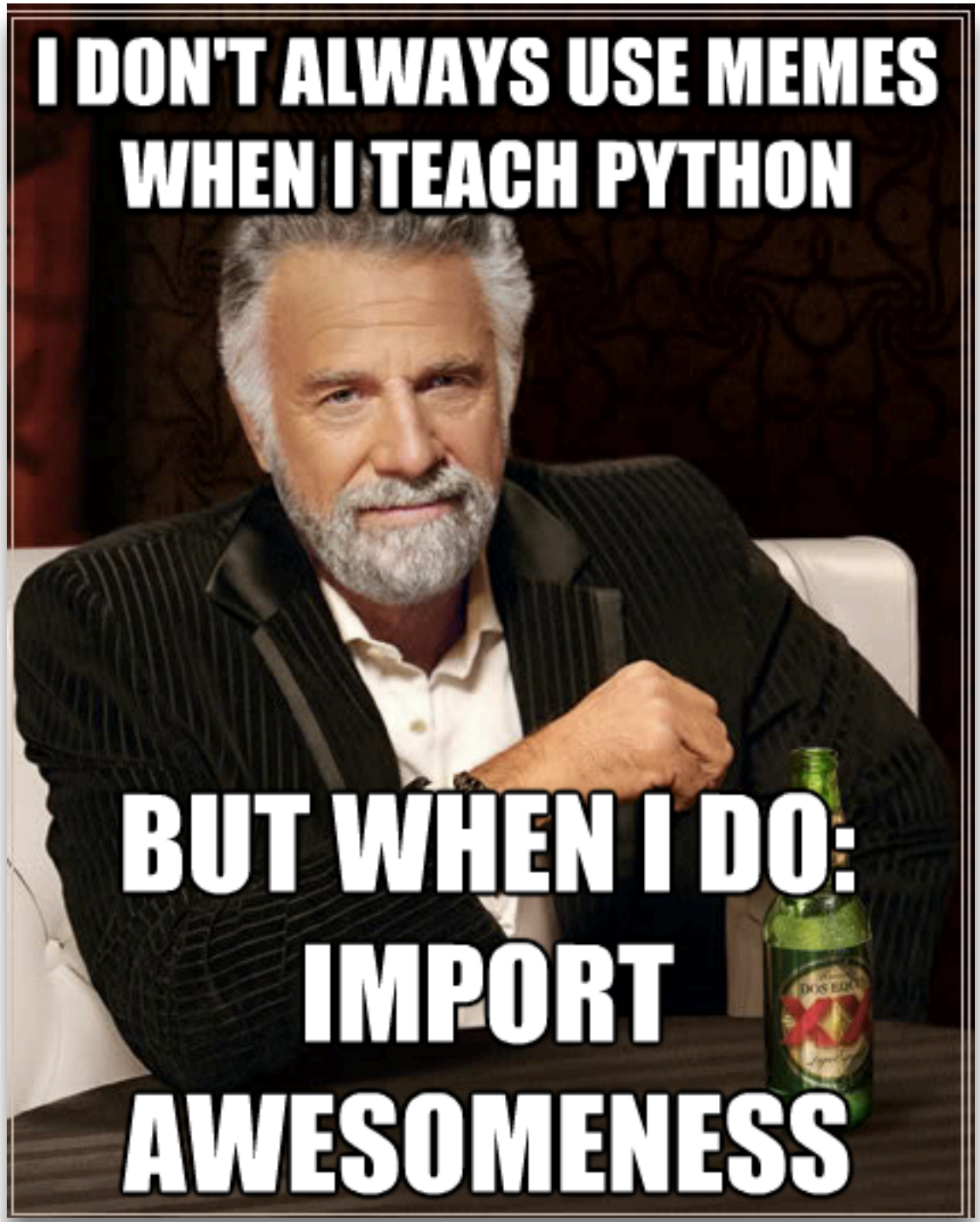
- "breakout work sessions" interspersed within the lectures

- homework assignments based on week's lecture

- final project




Bring me
a shrubbery!!
ICANHASCHEEZBURGER.COM


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

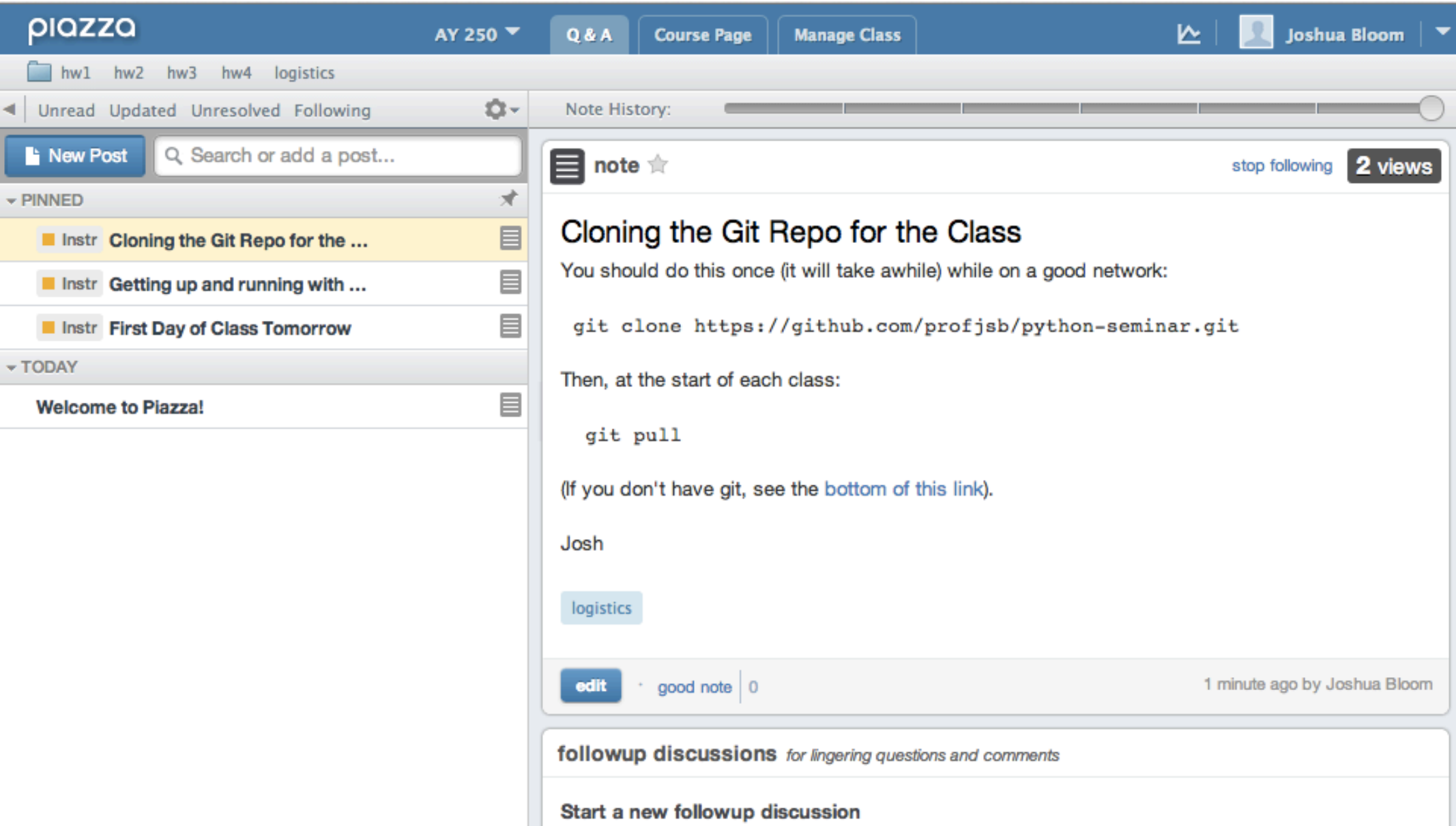
- laptop for use in class and for homeworks

- tolerance for our terrible computer humor

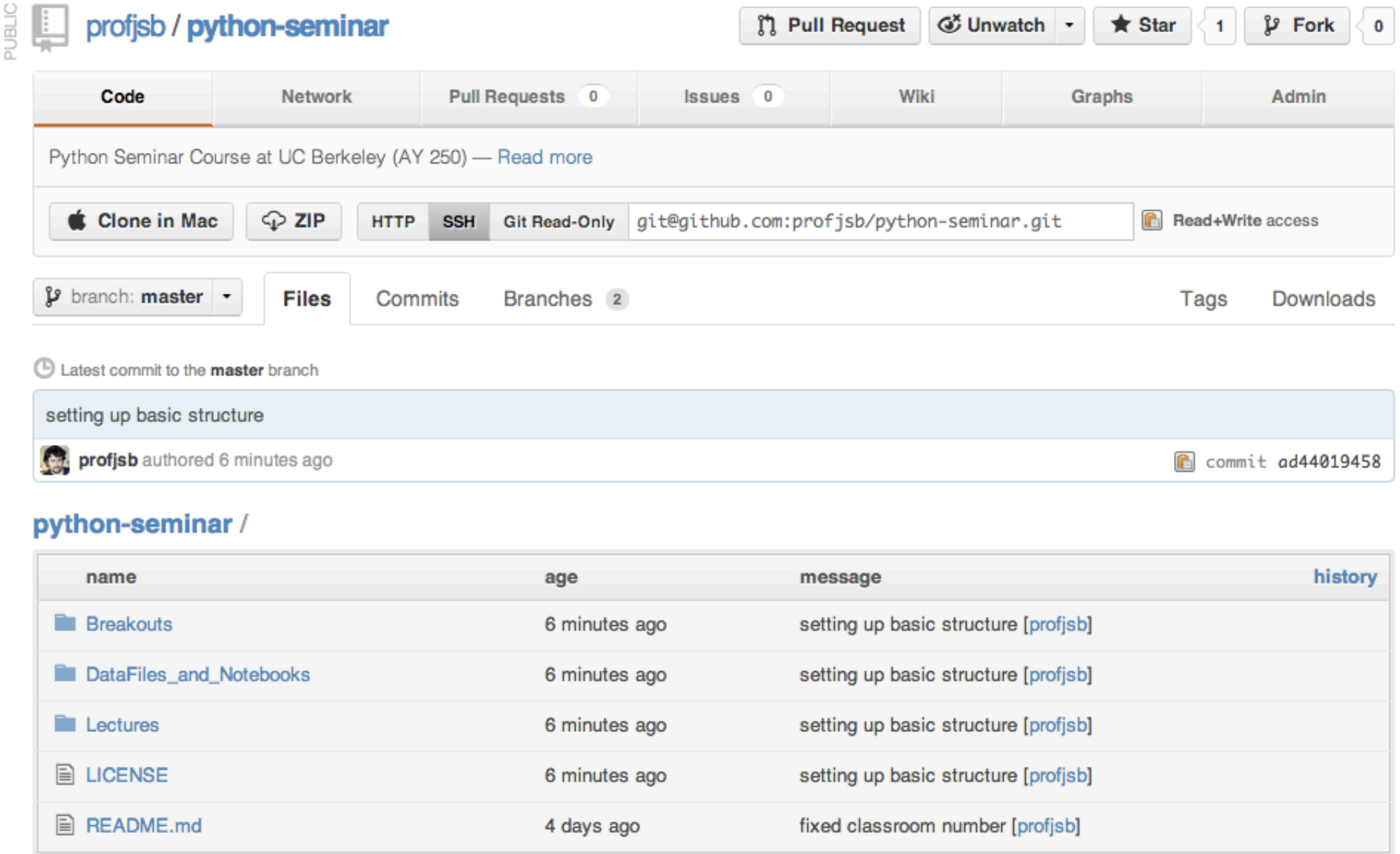
http://www.pythonbootcamp.info/preparation/software

piazza for real-time/off-line interaction homework updates, solutions, ...

piazza.com/berkeley/fall2016/ay250class13410/home



`github` is the main data portal for us...



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

Concepts/Practices in this Course

- Jupyter & JuypterLab
- 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	`cd python-seminar; git pull`
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 Dec12
(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 (Juypter,scipy, Cython,numpy, matplotlib, etc.)