Learning goals:

- Understand Jupyter
- Understand Datahub
- Final Project team building and ideas
- Q & A

Introductions

COGS 108 Fall 2020
Atman Patel
Discussion 1

a2patel@eng.ucsd.edu
OH: Tue 11:30am -12:30pm

Discussion slides and materials adapted from Sam Lau (TA: WI20)

Welcome to COGS 108!

- **Atman Patel**
- **2nd year MS student in ECE** – Specialization in Machine Learning and Data Science
- **Research:** Deep Learning, applications in Computer Vision
- **Interests:** Ukulele (started learning), Soccer, table tennis

Section Philosophy

- Attendance is not required
- Reasons to never miss the discussion sessions:
  - Demos to help you set up and start working
  - Hands-on experience and personalized guidance
  - Discussion is not supposed to be a monologue – Please ask questions
  - Condensed information >> searching the infinite www

Zoom format

- **Lecture -> Demo -> Q/A.**
  Please mute your microphone, but feel free to unmute and ask questions!
- **If you’re more comfortable with text ->** Keep asking questions on chat -> TI or I will address them.
- **We will take up assignment specific questions towards the end of Discussion or during OH.**

Project

- From a group of 4-5 students (if you opt for group project option)
- Feel free to talk to others right now! Post a message on common chat with your interests, region, skills etc.
- Use piazza -> pinned post
- Start working towards the project as soon as possible

Programming

- **This course assumes basic programming knowledge**
  - But not much
- **Resources:**
  - codeacademy
  - Start Here: https://github.com/COGS108/Tutorials/blob/master/01-Python.ipynb
  - Python in detail: https://jakevdp.github.io/PythonDataScienceHandbook/
  - Pandas: https://www.dataschool.io/python-pandas-tips-and-tricks/
  - Git: https://guides.github.com/activities/hello-world/
- **Cheatsheets**
  - Google: 'python cheatsheet', 'pandas cheatsheet', 'git cheatsheet' (find one that's good for you)

Anaconda

The Data Science Toolkit – contains Python and data science libraries (including jupyter notebooks)

- **Download:** https://www.anaconda.com/products/individual
- **Installation:** https://docs.anaconda.com/anaconda/install/
- **Verify installation:** https://docs.anaconda.com/anaconda/install/verify-install/
- **Make sure anaconda is added to the system path:** For mac: export PATH="/usr/local/anaconda3/bin:$PATH"

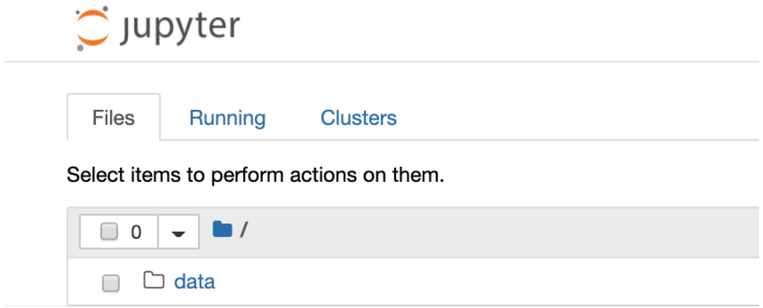
Git

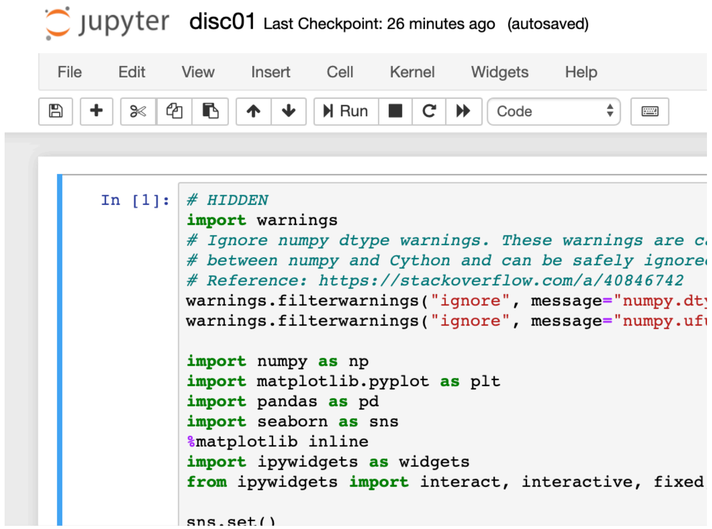
Version control system!

- **Go to** https://git-scm.com/downloads
- **Choose your Operating System** (Windows/OS X/Linux)
- **Follow the steps specific to your OS**
- **Verify installation:** In terminal type “git --version”

Jupyter

- Python code is run on a python interpreter
- Jupyter is a program that creates an interface for typing python code in a browser, that also runs that code in a python interpreter
- What does this mean?!
  - **Jupyter is a way of running python programs from a browser (like chrome) (hooray!)**





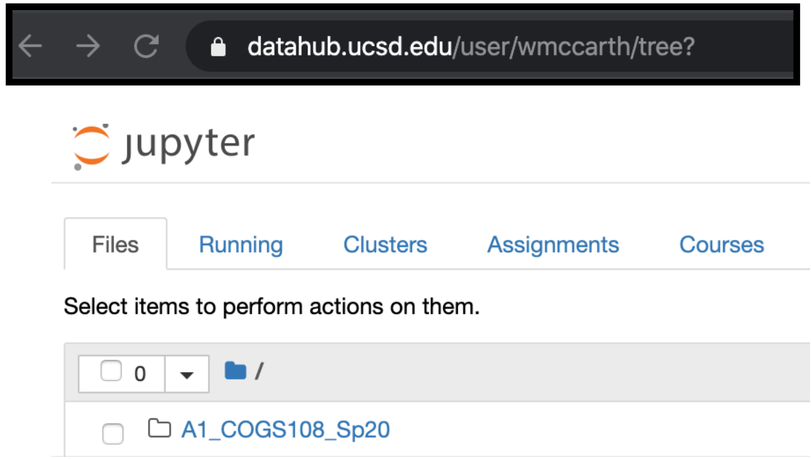
Jupyter Demo

An example of what you can do with Jupyter

For today's demo (includes both code and data):
https://github.com/COGS108/Section-Fa20

But before we start, we need to learn about datahub!!

datahub



- Jupyter runs python code in a browser.
  - But Jupyter is itself just a program that's running on a computer somewhere.
- **datahub** lets you interact with Jupyter that's running somewhere else.
- What does this mean?!
  - **You don't need to worry about installing Jupyter**
  - **You can use datahub to create and run python programs (online)**
  - **You can use this interface to fetch and submit assignments**

Oakland License Plates

- **Get data and jupyter notebook –**
  1. Open datahub
  2. Create a new folder called “Discussions”
  3. Open Terminal
  4. Go to Discussions folder
  5. Type “Git clone **https://github.com/COGS108/Section-Fa20”**
  6. Close Terminal
- **Open Jupyter notebook called disc01.ipynb** (located in Discussions/Section_Fa20/disc01/)

Working on your assignments

- **Log into datahub.ucsd.edu**
- Go to **Assignments tab**
- **‘fetch’ assignments** you have access to -> Submit after completion

Resources

\Most of the course related information:
https://github.com/COGS108/Overview
(It'll be worth bookmarking this page. You'll use it a lot)

For a long list of interesting datasets:
https://tinyletter.com/data-is-plural

All Discussion Materials:
https://github.com/COGS108/Section-Fa20

(Page above also has links to today's demo and extra practice with Python.)

Next week: A1 help, git walkthrough