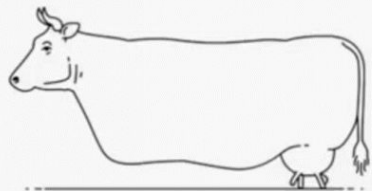


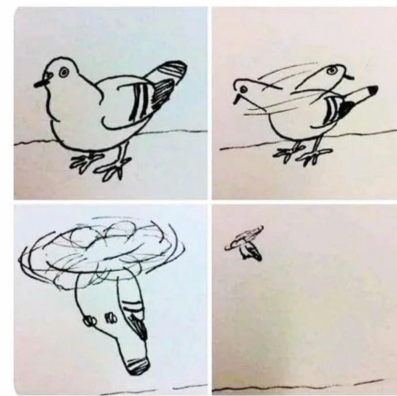
- If your code works fine don't touch it
+ my code:



COGS 108 Week 2

FA23, A04/A07

When your program
is a complete mess,
but it does its job



Plan for Today

- Basic introductions
- Tools we will be using in COGS 108 (anaconda, jupyter, git/github)
- Github Demo

Your TA

Jason Chen

- First-year graduate student in Data Science
- B.S. in Cognitive Science and Neurobiology @ UCSD
- OH: Tue/Thu 4-5 pm (over Zoom)
- Email: xic007@ucsd.edu

A Little Bit About Yourself...

slido



How are you doing?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

slido



What is your major?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

slido



What is your level of Python proficiency?

① Click **Present with Slido** or install our [Chrome extension](#) to activate this poll while presenting.

Section Philosophy

- Attendance is not mandatory (but encouraged)
- Reasons to attend the discussion sessions:
 - Demos to help you set up and start working
 - Hands-on experience and personalized guidance
 - Time to practice technical aspects of the course
 - Discussion is not supposed to be a monologue – Please ask questions

Programming Basics

- This course assumes basic programming knowledge... but not much
- Resources:
 - codecademy
 - 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/>
 - Stack Overflow
- AI helper
 - ChatGPT (warning: **DO NOT** rely on it too much as it can slow you down from being a good programmer!!! Please refer to the syllabus for information about the use of LLM)

Anaconda

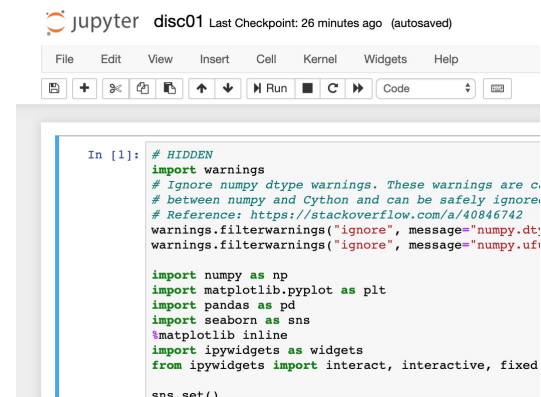
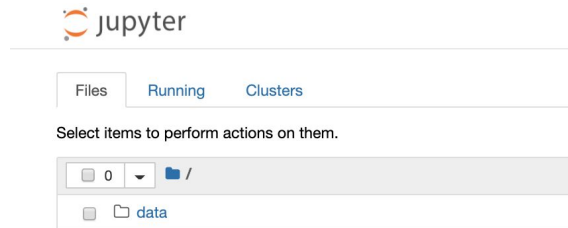
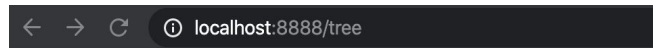
The Data Science Toolkit – contains Python and data science libraries (including jupyter notebooks)

- 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"

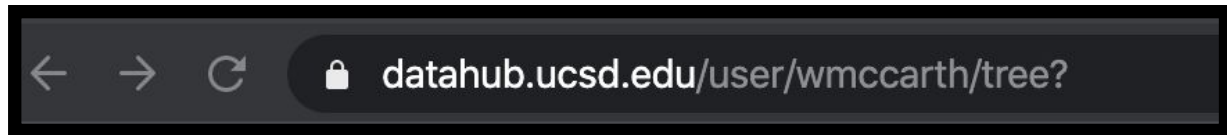
Jupyter



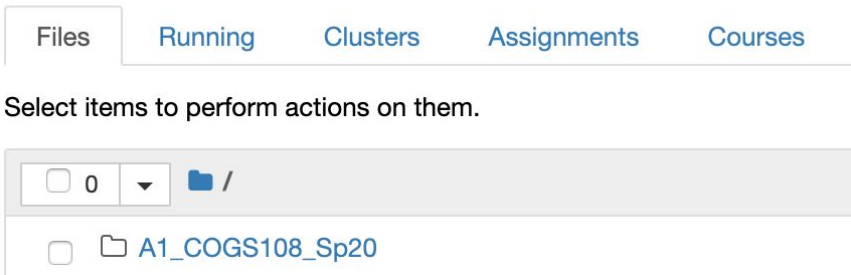
- Python code is run on a python interpreter
- Jupyter is a program that creates an interface for typing python code in a browser, that also runs that code in a python interpreter
- What does this mean?!
 - You can run python programs from a browser!



Datahub



- Datahub lets you interact with Jupyter remotely.
- 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



Check Your Work (in general on datahub)

- The tests built into the notebook are (very) minimal.
 - To write your own tests, add a cell with assert statements below your code.
- Make sure to click Validate before turning in your notebook!
 - This replicates what our autograder will do.
 - Hidden tests on the autograder aren't validated.
- Don't click submit after deadline! (Your assignment will be marked as a late submission)

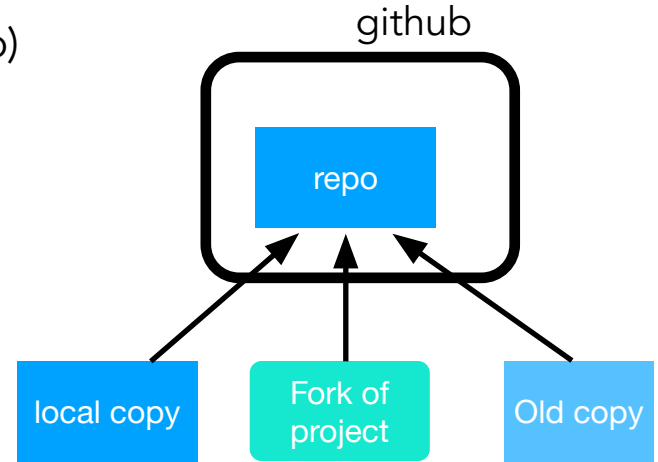
Git

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”

What is git + GitHub?

- Somewhere online to store a copy of a project (Github)
- Plus a tool to interact with this copy (Git)
 - Command line and desktop versions
- A way of keeping track of changes you make to this project
- Does everyone have a GitHub account?



Why use git + GitHub?

- Collaboration: Git allows you to work on code projects with other people. It's the preferred tool for many projects, like...
 - Python: <https://github.com/python/cpython>
 - Jupyter: <https://github.com/jupyter/>
 - COGS 108: <https://github.com/COGS108/>
- Backup
- Version control (*undo* on a large scale)
- Code reuse

Jupyter is NOT Good in Version Control

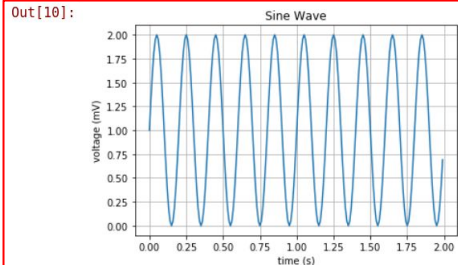
This is what the plot looks like in GitHub:

```
In [10]: import numpy as np
import matplotlib.pyplot as plt

# Data for plotting
t = np.arange(0.0, 2.0, 0.01)
s = 1 + np.sin((5 * 2) * np.pi * t)

# Note that using plt.subplots below is equivalent to using
# fig = plt.figure() and then ax = fig.add_subplot(111)
fig, ax = plt.subplots()
ax.plot(t, s)

ax.set(xlabel='time (s)', ylabel='voltage (mV)', title='Sine Wave')
ax.grid()
```



"outputs": [

{

"data": {

"image/png":

"iVBORw0KGgoAAAANSUHEUgAAAYwAAAEWCAYAAAB1xKBvAAAABHNCSVQICAgIfAhkIAAA
AAIwSFIZAAAEgAACxIB0t1+/AAAAADlORVh0U29mdHdhcmUAAbWF0cGxvdGxpYiB2ZXJzaW9uID
luMi4yLCBodHRwOi8vbWF0cGxvdGxpYi5vcmlvLmUwY29mdHdhcmUAAbWF0cGxvdGxpYiB2ZXJzaW9uID
EgABHeQEkVSXGGRFLembFNSPn7Wyy45i5UXh5ZjvcSy4xcr78WK5bwkzvKSEllOqaVxZKcOJL
N+Fhc0dxJEVxAAgQBAiClDbDP0tPT+80fVdXdmOnl1q17ezBm/T6f+QDdXVXnVtU996z3HFFKE
SNGjBgxYvRDYrkHECNGjBgvVgZigREjRowYmbQQC4wYMWLEiKGFwGDEiBEjRgwtXAljRowY
MWJoIRYymWLEiBFDC7HAiBEDEJG/JiKPL/c4YsQ4nxELjBgfGojIXSLyoojMiMgZEXIBRH4lQCn
1B0qp+x3QfExE/q+2z5tERHX5boNt+jFi2EQsMGJ8KCAiE8CfAr8BrAE2Ab8CIB2Tfha4t+3zPcC7H
b57Tyk15XgsMWJEQiwYynxY8BEApdS3lVJ1pdSCUupxpdRbAClyRRF5PjjY1/i/JCLvichZEftEZ
G23/8PEdnl//aYiFzWhe6zwJ0iEvDa3cC/A7Ys+u5Z/7qrReRPRSkf+0/FZGL/d8+LyLb2i8ul8RkYf9
/2dF5F+LyEEROS4i3xCR4YjPLUaMJmKBEEPDgj1AXUT+k4h8VkrWa5zzY8APATcCfwX4NiCI/O
/APwL+AnAB8Bzw7S7X+AGQ9a8BnjXxBLB30XfP+v9PAL8PXAzcCiwAv+..."

Jupyter is NOT Good in Version Control

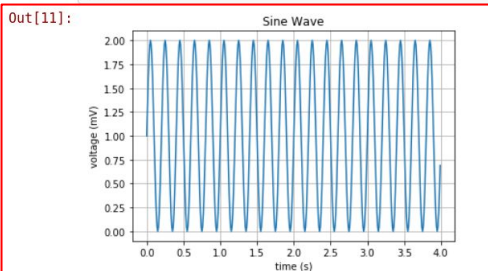
When we make a little change:

```
In [11]: import numpy as np
import matplotlib.pyplot as plt

# Data for plotting
t = np.arange(0.0, 4.0, 0.01)
s = 1 + np.sin((5 * 2)* np.pi * t)

# Note that using plt.subplots below is equivalent to using
# fig = plt.figure() and then ax = fig.add_subplot(111)
fig, ax = plt.subplots()
ax.plot(t, s)

ax.set(xlabel='time (s)', ylabel='voltage (mV)', title='Sine Wave')
ax.grid()
```



The difference in plot will be marked as:

```
"outputs": [
```

{

```
"data": {
```

"image/png":

"IVBORw0KGgoAAAANSUheUgAAAYwAAAEWCAYAAAB1xKBvAAAABHNCSVQICAgIfAhkIAAA
AAIwSFZlAAAEgAACxIB0t1+/AAAADl0RVh0U29mdHdhcmUAbWF0cGxvdGxpYiB2ZXJzaW9uID
luMi4yLzBodHRwOi8vbWF0cGxvdGxpYiY5cmcmcvcnp/UCwAAIAIBJREFUeJzsvXmcHND13/s9vc4+2
EgABHeQEKV SXGGRFLemBfNSPn7Wyy45i5UXh5ZjvcSy4xcr78WK5bkwzvKSeIIQqaVxZKcOJL
N+FHcd0xJEVxAAgQBAiClbDP0tPT+80fVdXdmOn1q17ezBm/76f+QDdXVXNvU996z3HFFKE
SNGjBgxYvRDYrkHECNGjBgxVgZigREjRowYmBQQC4wYMWLEiKGFWFGDEiBEjRgwtXAljRowY
MWJoIRYYMWLEiBFDC7HAiBEDEJG/JiKPL/c4YsQ4nxELjBgfGojI XSLyoojMiMgZEXIBRH4IQcn
1B0qp+x3QfExE/q+2z5tERHX5bnOt+jfIEQsMGJ8KCAiE8CfAr8BrAE2Ab8CIB2Tfha4t+3zPcC7H
b57Tgk15XsMWJEQiwweYnxY8BEADp3S1VJ1pdSCUupxpdRbAClyRRF5Pjj1/i/JCLvichZEfkZ
6273/8PEDnl/aYifQhweWzJ0iEvaD3cA/AY7s+u5Z/7qrREPRReSkf+0RrF228u1l8RkEY9
/2dF5F+LyEEROS4i3xCR4YjPLUaMjmkBEePDgj1AXUT+k4h8VKrW5szZY8APATcCfwX4NICl/O
/APwL+AnAB8Bzw7S7X+AGQ9a8BnjXxBLB30XfP+v9PAL8PXAzcCiwAv+..."

Jupyter is NOT Good in Version Control

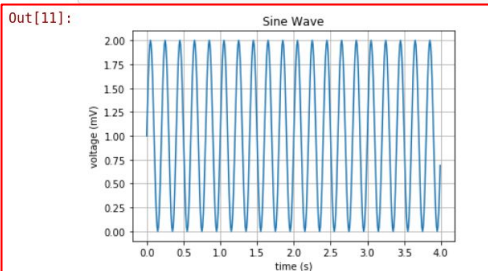
When we make a little change:

```
In [11]: import numpy as np
import matplotlib.pyplot as plt

# Data for plotting
t = np.arange(0.0, 4.0, 0.01)
s = 1 + np.sin((5 * 2)* np.pi * t)

# Note that using plt.subplots below is equivalent to using
# fig = plt.figure() and then ax = fig.add_subplot(111)
fig, ax = plt.subplots()
ax.plot(t, s)

ax.set(xlabel='time (s)', ylabel='voltage (mV)', title='Sine Wave')
ax.grid()
```



The difference in plot will be marked as:

```
"outputs": [
```

{

```
"data": {
```

"image/png":

"IVBORw0KGgoAAAAANSuHEuGAAAYwAAAEWCAYAAAB1xKBvAAAABHNCSVQICAgIfAhkIAAA
 AAlwSFZlAAAEgAACxiB0t1+/AAAADI0RVh0U29mdHdhcmUAbWF0cGxvdGxpYiB2ZXJzaW9uID
 luMi4yLXBodHRwOi8vbWF0cGxvdGxpYiY5cmcmcvcnp/UCwAAIAABJREFUeJzsvXmcHnd13/s9vc4+2
 EgABHeQEKV SXGGRFLemBfNSPn7Wyy45i5UXh5JzvcSy4xcr78WK5b5wkzvKSeIIQqaVxZKcOJL
 N+FHcd0xJEVxAAgQBAiCldBDP0tPT+80fvdXdmOn1q17ezBm/T6f+QDdXVXnVtU996z3HFFKE
 SNGjBgxYvRDYrkHECNGjBgxvGZigREjRowYMbQQC4wYMWLEiKGFwGDEiBEjRgwtXAljRowY
 MWJoIRYYMWLEiBFDC7HAiBEDEJG/JiKPL/c4YsQ4nxELjBgfGojI XSLyoojMiMgZEXIBRH4IQcn
 1B0qp+x3QfExE/q+2z5tERHX5bnT+jfIEQsMGJ8KCAiE8CfAr8BrAE2Ab8CIB2Tfha4t+3zPcC7H
 b57Yk13XGSMWJEQiwWYnxY8BEADp3SIVJ1pdSCUupxpdRBAClyRRF55j1/1/JCLvichZEfkTZ
 g278/PEDnll+yiYFQhwe6zWJ0iEvdA3cC/A7Ys+u5CZ/7qrREPRReSkf+0/FZgl+2d+LylL2i8ull8RkZY9
 /2dF5F+LyEEROS4i3cR4YjPLUAmJmKBEePDgj1AXUT+k4h8vKRWa5ZZY8APATcCfwX4NICl/O
 /APwL+AnAB8Bzw7S7X+AGQ9a8BnjXxBLB30XfP+v9PAL8PXAzcCiwAv+..."

Jupyter is NOT Good in Version Control

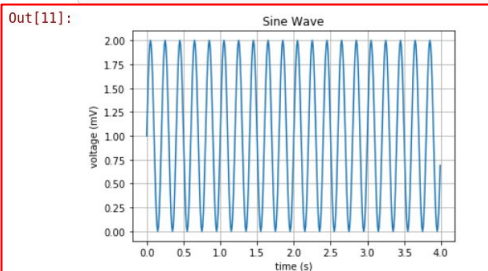
When we make a little change:

```
In [11]: import numpy as np
import matplotlib.pyplot as plt

# Data for plotting
t = np.arange(0.0, 4.0, 0.01)
s = 1 + np.sin((5 * 2)* np.pi * t)

# Note that using plt.subplots below is equivalent to using
# fig = plt.figure() and then ax = fig.add_subplot(111)
fig, ax = plt.subplots()
ax.plot(t, s)

ax.set(xlabel='time (s)', ylabel='voltage (mV)', title='Sine Wave')
ax.grid()
```



The difference in plot will be marked as:

```
"outputs": [
```

{

```
"data": {
```

"image/png":

"IVBORw0KGgoAAANSUheUgAAAYwAAAEWCAYAAAB1xKBvAAAABHNCSVQICAgIfAhkIAAA
AAIwSFZfAAAEgAACxiB0t1+/AAAADI0RVh0U29mdHdhcmUAbWF0cGxvdGxpYiB2ZXJzaW9uID
luMi4yLXBodHRwOi8vbWF0cGxvdGxpYi5vcmcvhp/UCwAAIAJBREFUeJzsvXmcHNd13/s9vc4+2
EgABHeQEKV SXGGRFLembFNSPn7Wyy45i5UXh5ZjvcSy4xcr78WK5bkwzvKSeIIQqaVxZKcOJL
N+FHcd0xJEVxAAgQBAiCIdbDP0tPT+80fvdXdmOnl1q17ezBm/T6f+QDdXVXnVtU996z3HFFKE
SNGjBgxYvRDYrkHECNGjBgxVgZigREjRowYMbQQC4wYMWLEiKGFwGDEiBEjRgwtXAljRowY
MWJolRYYMWLEiBFDC7HAiBEDEJG/JiKPL/c4YsQ4nxELjBgfGojI XSLyoojMiMgZEXIBRH4IQcn
1B0qp+x3QfExE/q+2z5tERHX5boNt+jfIEQsMGJ8KCAiE8cFAr8BrAE2Ab8CIB2Tfha4t+3zPcC7H
b57Yk15XsgMWJEQiwYnXy8BEADpcS3iVj1pdSCUupxpdRACLvRRF5PjY1/i/JCLvichZEftkEZ
G23/8PEdnl//aYiFzWhe6wJ0iEVDa3C4i4Ys+uSZ/7qrReRPRSkf+0/FZGL/d8+LyLb2i8u1l8RkyY9
/2dF5P+LyEEROS4S3xCR4YjPLUAmJmKBEePDgj1AXUT+k4h8VKrWf+5ZZzLd8APATcCfwX4NICl/O
/APwL+AnAB8Bzw7SX+AGQ9a8BnjXxBLB30XP+v9PAL8PXAZcCiwAv+..."

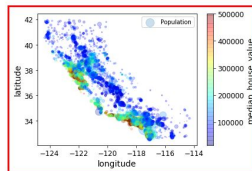
When you have a lot of plots, your changes in git will be thousands-pages long... And it is impossible to tell what the differences are...

How to Avoid This Issue?

- Clear your outputs before pushing back your notebook to GitHub. (You will only be able to see changes in your CODE)
- Use other GitHub Apps, like ReviewNB.
- This app will VISUALIZE the plots, instead of showing those nonsense words (Yay!)

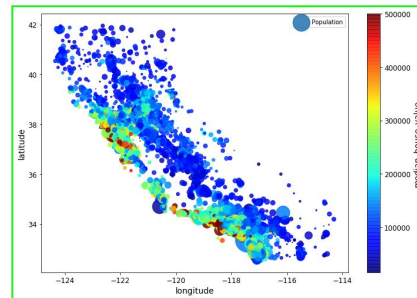
```
1 housing.plot(kind="scatter", x="longitude", y="latitude", alpha=0.2,  
2             s=housing['population']/100, label='Population', figsize=(8,8),  
3             c="median_house_value", cmap=plt.get_cmap('jet'), colorbar=True,  
4             sharex=False)  
5 plt.legend()  
6 save_fig('housing_prices_scatterplot')
```

Saving figure housing_prices_scatterplot



```
1 housing.plot(kind="scatter", x="longitude", y="latitude", alpha=0.8,  
2             s=housing['population']/100, label='Population', figsize=(8,8),  
3             c="median_house_value", cmap=plt.get_cmap('jet'), colorbar=True,  
4             sharex=False)  
5 plt.legend()  
6 save_fig('housing_prices_scatterplot')
```

Saving figure housing_prices_scatterplot



Source: <https://nextjournal.com/schmudde/how-to-version-control-jupyter>

Demo of GitHub Repo and Commands

- Cloning a repo on your local machine
- Working on your local repo (making changes to files)
- Stage, commit and push these changes to your Github repo
- Commands you should know:
 - Git clone -This will download the latest version of the repo to your local PC
 - Git status (not really needed – but really helpful)
 - Git add - This adds the changes in the staging area
 - Git commit -m "" - This will save a “Snapshot” of your most recent changes
 - Git push - This will upload your local changes to your GitHub Repo
 - Git pull - This will update your local repo to the latest version which is on GitHub

A Few Reminders

- Lecture Quiz Q1 Due **TODAY 11:59 PM**
- Practice Assignment Due **Wednesday 11:59 PM**
- Pre-course Survey Due **Wednesday 11:59 PM**
- Discussion Lab D1 Due **Friday 11:59 PM**