Learning how to code with open access neurobiology datasets

Ashley Juavinett, Ph.D.
Assistant Teaching Professor
UC San Diego, Division of Biological Sciences

Goals for this workshop

- Provide a general framework on teaching coding to neuroscience students
- Introduce Jupyter Notebooks as a tool for teaching
- Demonstrate the use of the Jupyter Notebooks

While we're warming up,
let's prepare for coding together.
https://bit.ly/ViABLEcoding

That link will send you to:
https://colab.research.google.com/github/ajuavinett/ViABLE_2021/

Additional materials are available: https://github.com/ajuavinett/ViABLE_2021

Where I'm coming from...

Postdoc: More MATLAB

Lafayette College (coding class optional — I didn't take it)

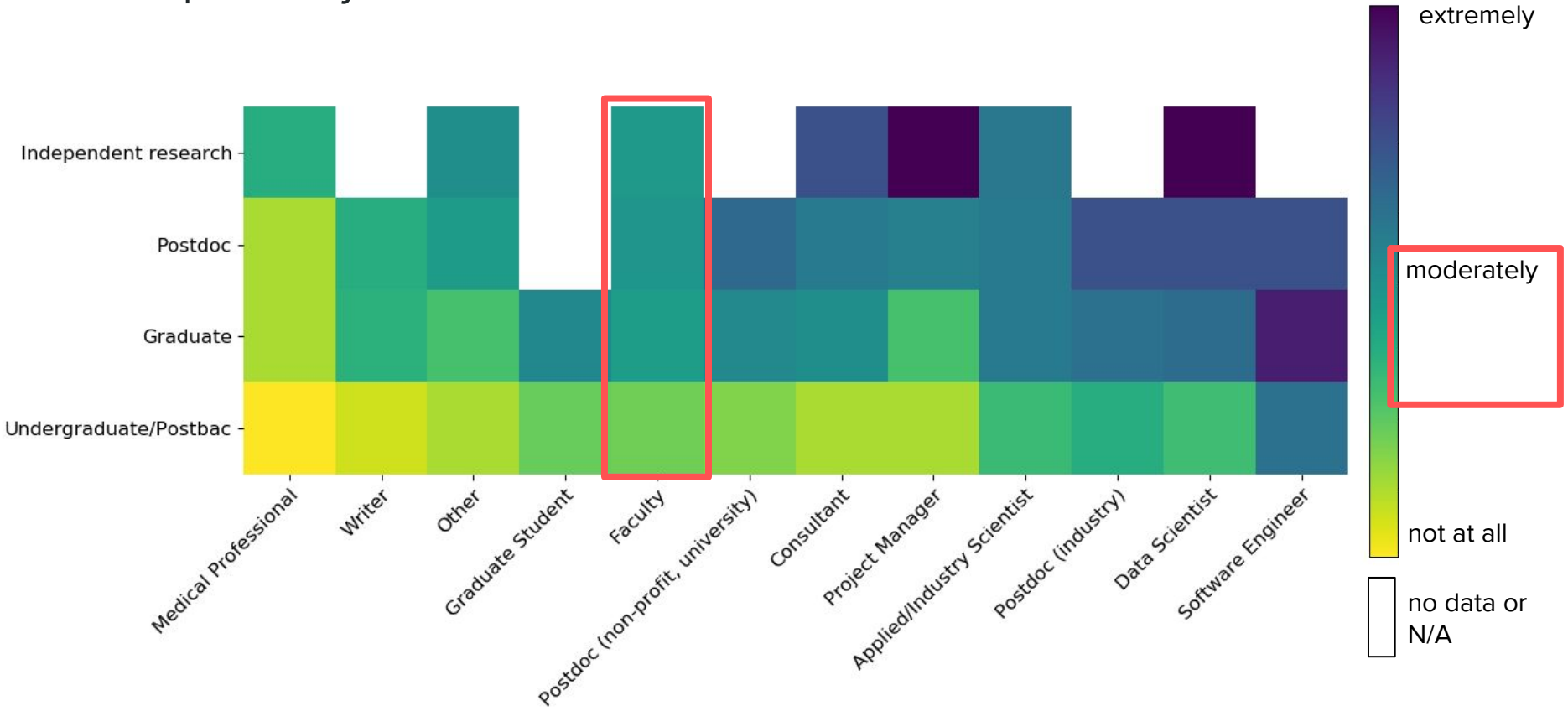
PhD: "learned how to code" in MATLAB, dabbled in Python

Faculty: When I learned Python 😬

I've *never* taken a formal coding class!

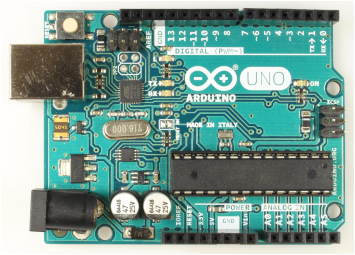
Even if you haven't, you can teach students how to code.

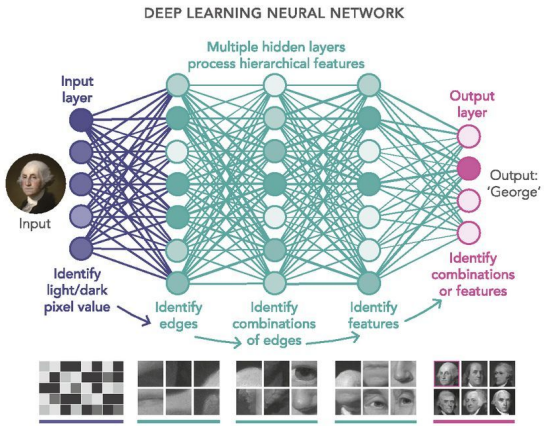
How comfortable did/do you feel working with code at this point in your career?



Do you teach or demonstrate programming in your courses?

Why do neurobiology students need to learn how to code?





custom data acquisition (controlling hardware, image acquisition, etc)

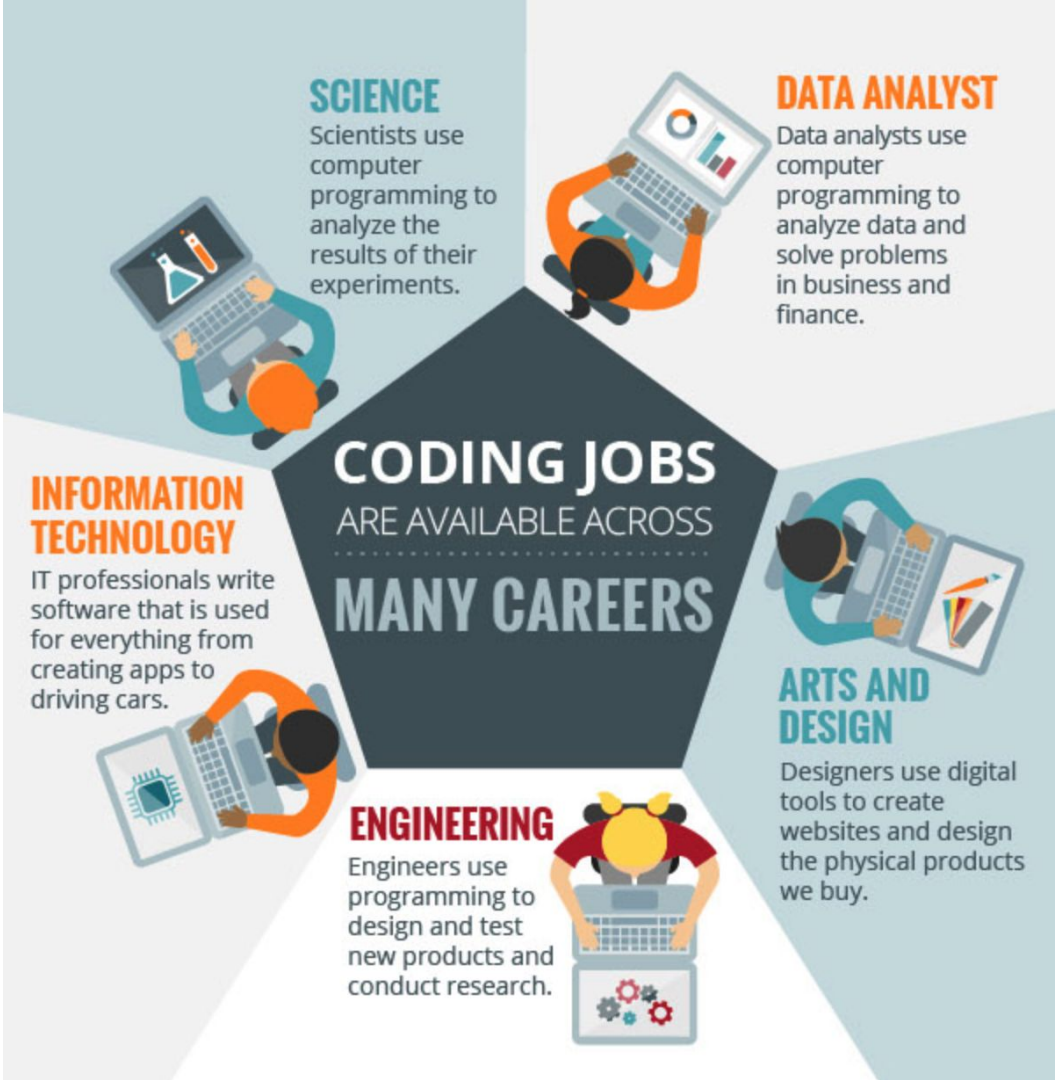
statistics, data analysis & visualization

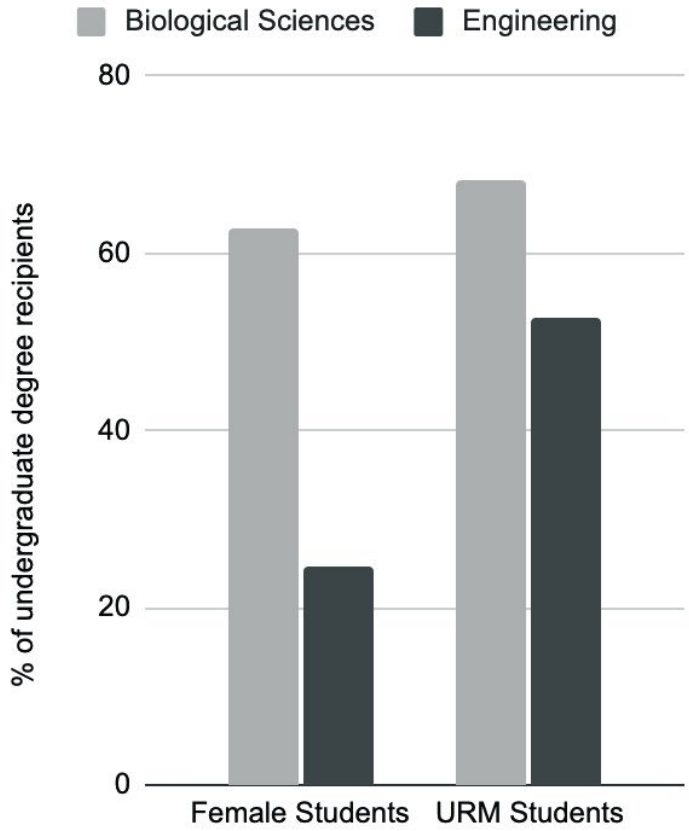
computational modeling

Why do neuroscience students need to learn how to code?

Beyond research, there are more and more jobs for software engineers, and they pay well

(see report by Burning Glass)





https://ir.ucsd.edu/undergrad/stats-data/ug-degree.html

This is also a matter of equity.

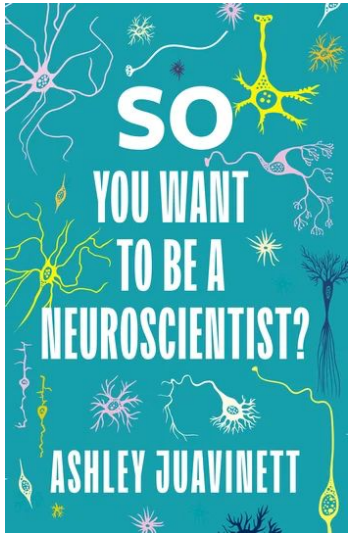
More and more jobs require programming experience, and the jobs that do pay more.

(see report by Burning Glass: https://www.burning-glass.com/research-project/coding-skills/)

3) the ability to integrate disparate information and think critically;
4) quantitative skills and familiarity with use of statistical applications and scientific programming/coding;
5) facility in the application of scientific inquiry, with analytic and research skills;

"The New Blueprints: Undergraduate Neuroscience Education in the Twenty-First Century" (Wierterlak et al., *JUNE*, 2018)

(Grisham et al., *Frontiers in Neuroinformatics*, 2016)



See also Akil et al., "Neuroscience Training for the 21st Century," *Neuron* (2016)

Summarized in *So you want to be a neuroscientist?* (Columbia University Press)

Proposed Training to Meet Challenges of Large-Scale Data in Neuroscience

William Grisham[1], Barbara Lom[2], Linda Lanyon[3] and Raddy L. Ramos[4]*

[1] Department of Psychology, University of California, Los Angeles, CA, USA, [2] Biology Department, Davidson College, Davidson, NC, USA, [3] International Neuroinformatics Coordinating Facility, Stockholm, Sweden, [4] Department of Biomedical Sciences, College of Osteopathic Medicine, New York Institute of Technology, Old Westbury, NY, USA

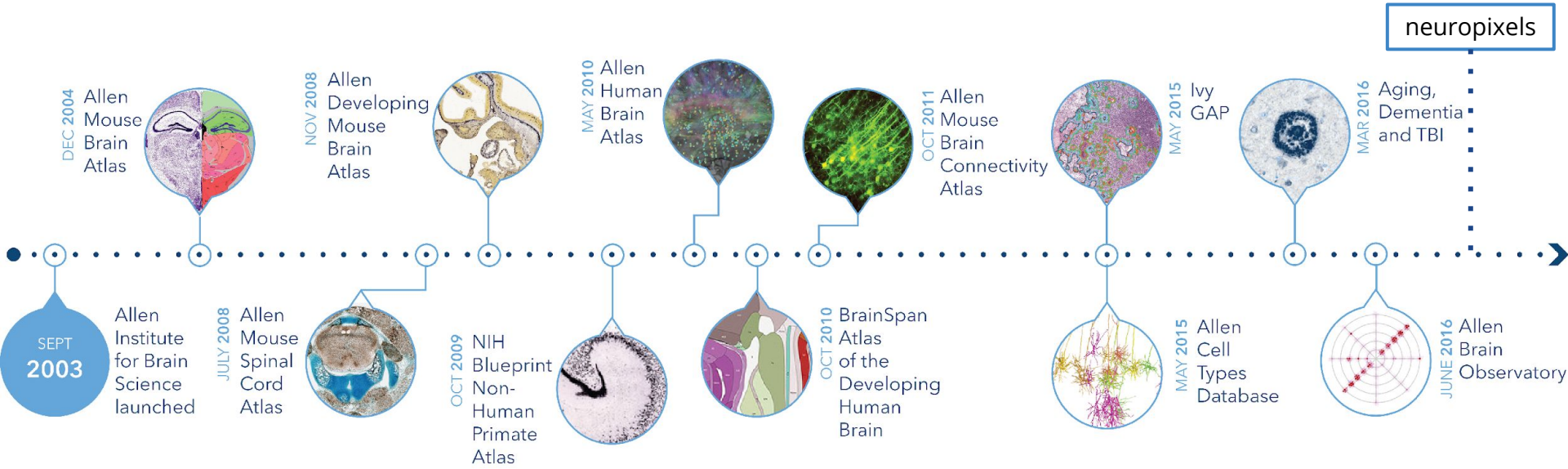
The scale of data being produced in neuroscience at present and in the future creates new and unheralded challenges, outstripping conventional ways of handling, considering, and analyzing data. As neuroinformatics enters into this big data era, a need for a highly trained and perhaps unique workforce is emerging. To determine

Yet, few neuroscience programs require coding, and only some offer it as an elective

- In a study of undergraduate neuroscience majors, only **six** of 118 institutions required coding, and **nine** included coding as an elective (Pinard-Welyczko et al., *JUNE*, 2017)
  - In a smaller study of 22 programs, 28% require programming as a coursework, 50% offer it as an elective (SfN, 2016)
- 15% of neuroscience PhD programs require programming, and 55% of including programming as an elective (SfN, 2016)

We can integrate coding into neuroscience courses *(especially remote laboratory courses)*

Neuroscience datasets are increasingly being shared online



NEURODATA WITHOUT BORDERS

DANDI
dandiarchive.org

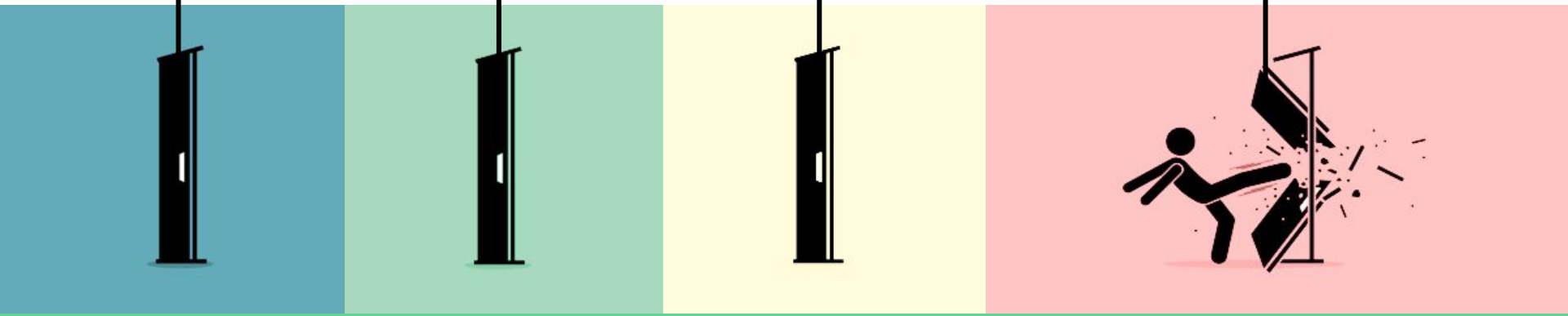
Barriers to bringing coding into neuroscience courses

Ease of access to coding environments

Biology instructor knowledge of coding

Unknown needs of “applied” vs. “basic” programming courses

Accessible, discipline-specific coding resources for all educators



Students also come in with preconceived notions about coding

I am intimidated because I have some buddies who are computer science majors and share their stresses to me sometimes

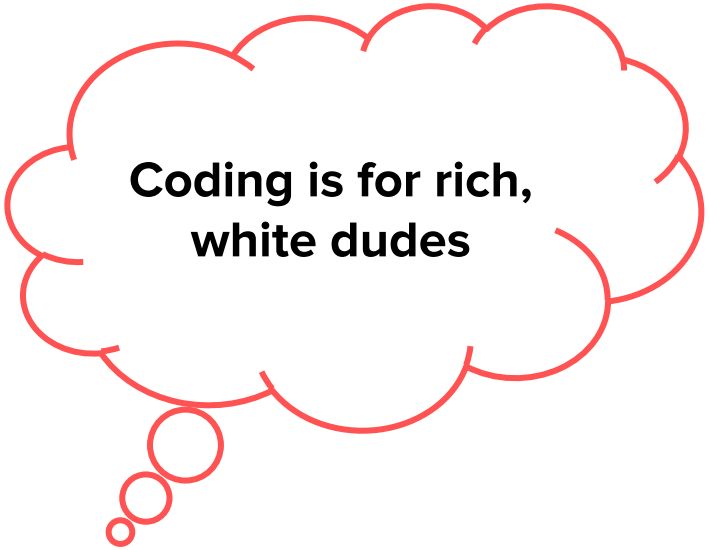
I've heard very awful things about computer science courses but it seems interesting especially if the connection to neurobiology is there!

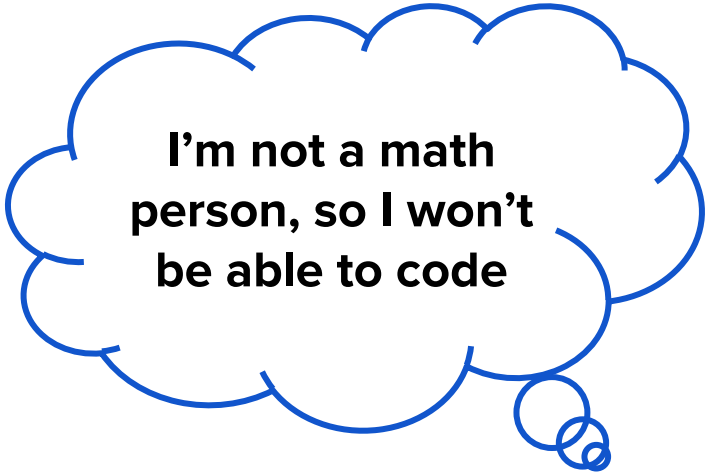
Not very comfortable since I'm not very much a computer person, also **numbers stress me out**

I'm curious about it. I don't know if I'll be a good coder or not.

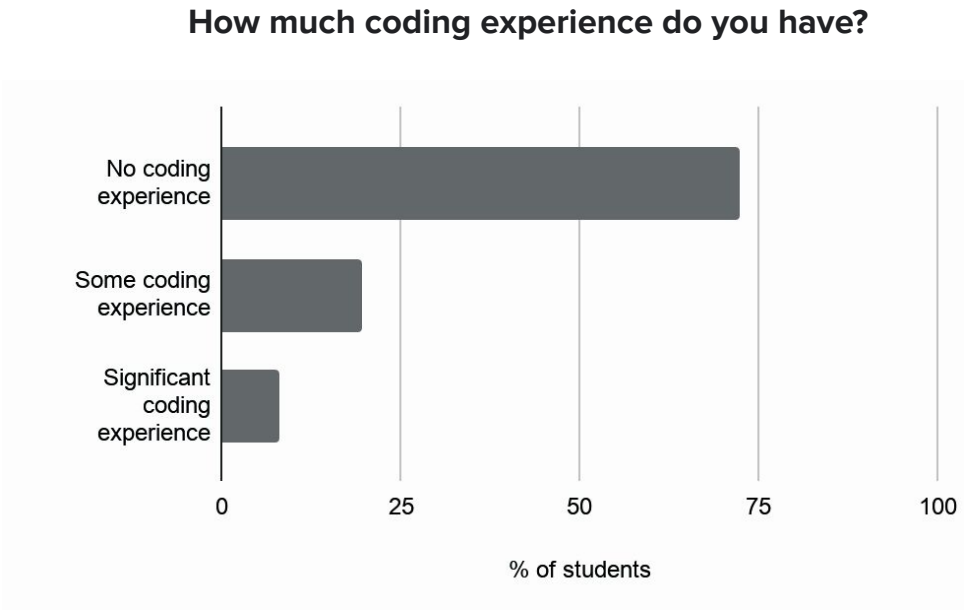
Responses to the question *"How do you feel about learning how to code?"*

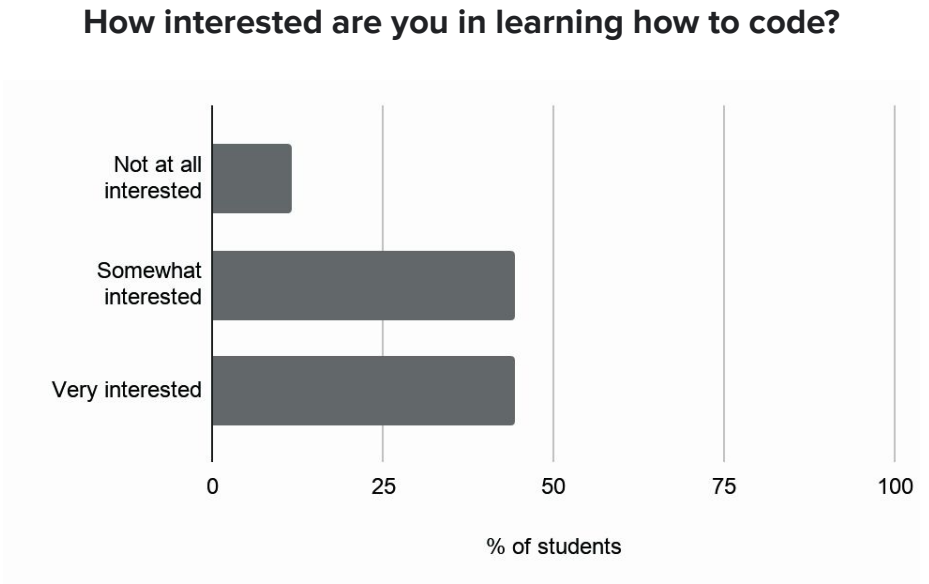
Unspoken beliefs about coding





Mismatch between experience and interest





4 sections of neurobiology lab; 138 students total

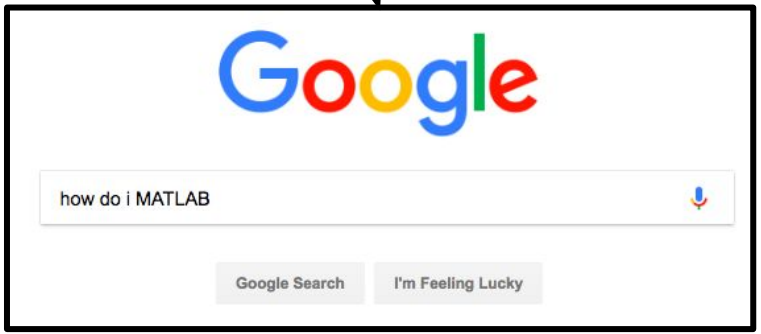
A general framework for teaching coding to neuroscience students

- There is no **programming gene**
- Frame coding as a **skill** — *it's like learning a language*
- Encourage **growth mindsets** around coding
  - Students with growth mindsets perform better over time (even in organic chemistry)
  - Beliefs about whether computational abilities are fixed or malleable impact: sense of belonging, how we respond to difficulties, and ultimately, achievement

More thoughts on this: Medium | There is No Such Thing as a Computational Person

A general framework for teaching coding to neuroscience students

- Live code in front of students
- Encourage **pair programming**
- Normalize errors and **searching for answers**



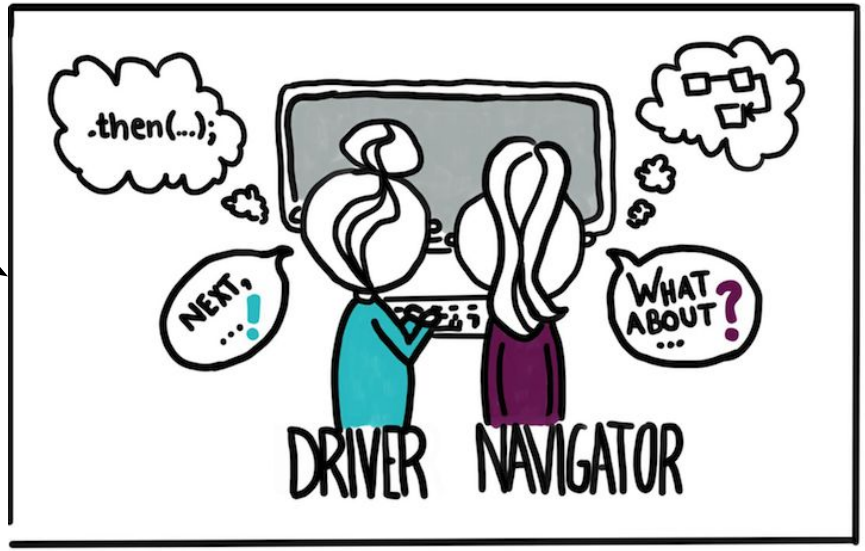


Image Source

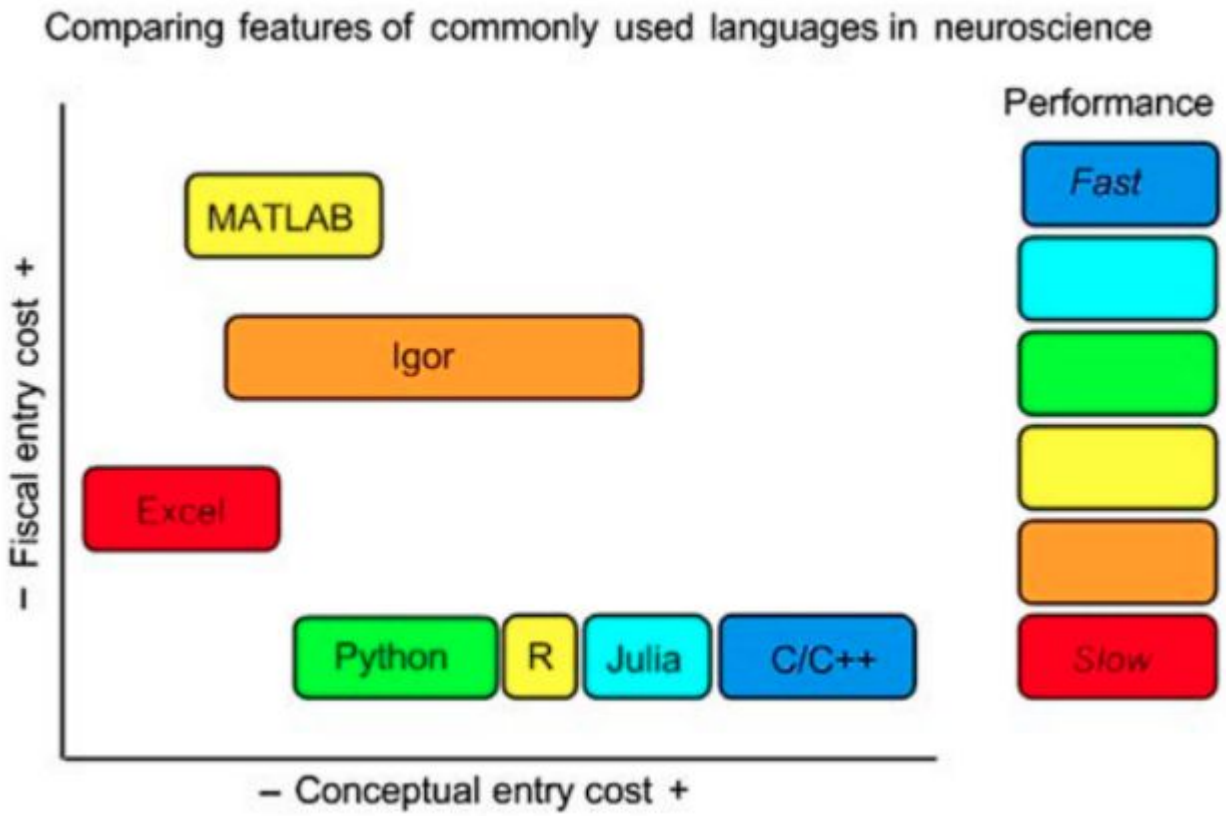
More thoughts on this:

Goals for this workshop

- Provide a general framework on teaching coding to neuroscience students
- **Introduce Jupyter Notebooks as a tool for teaching**
- Demonstrate the use of the Jupyter Notebooks

Considerations for choosing a programming language

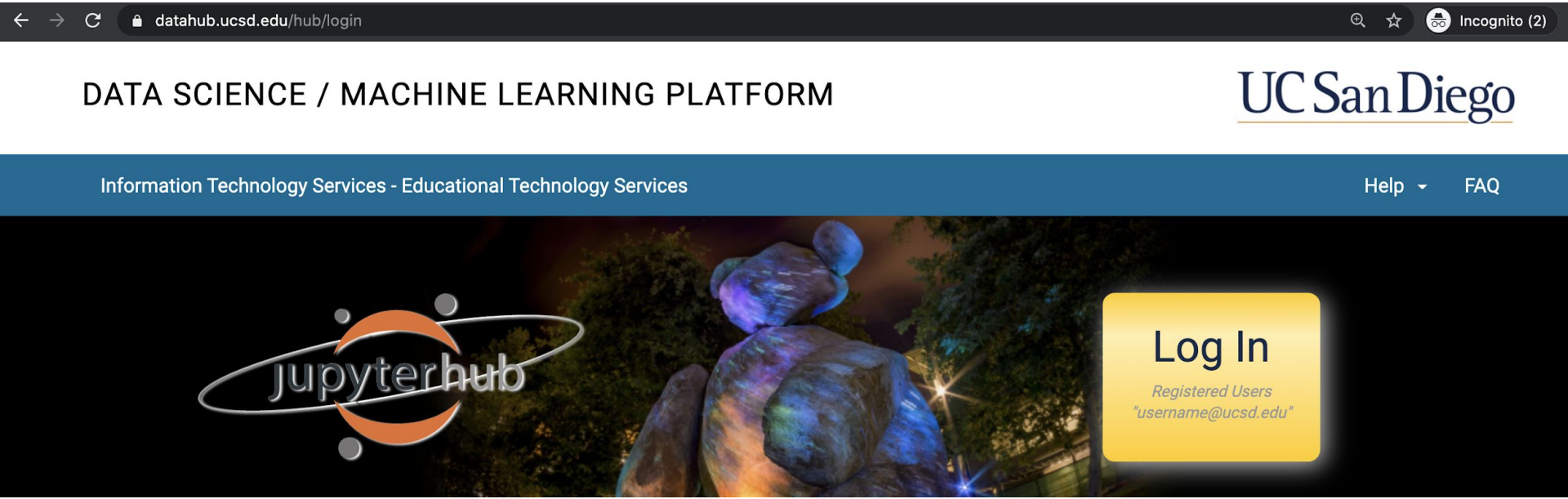
- Fiscal & conceptual entry
- Usage in neuroscience
  - Python, MATLAB, and R



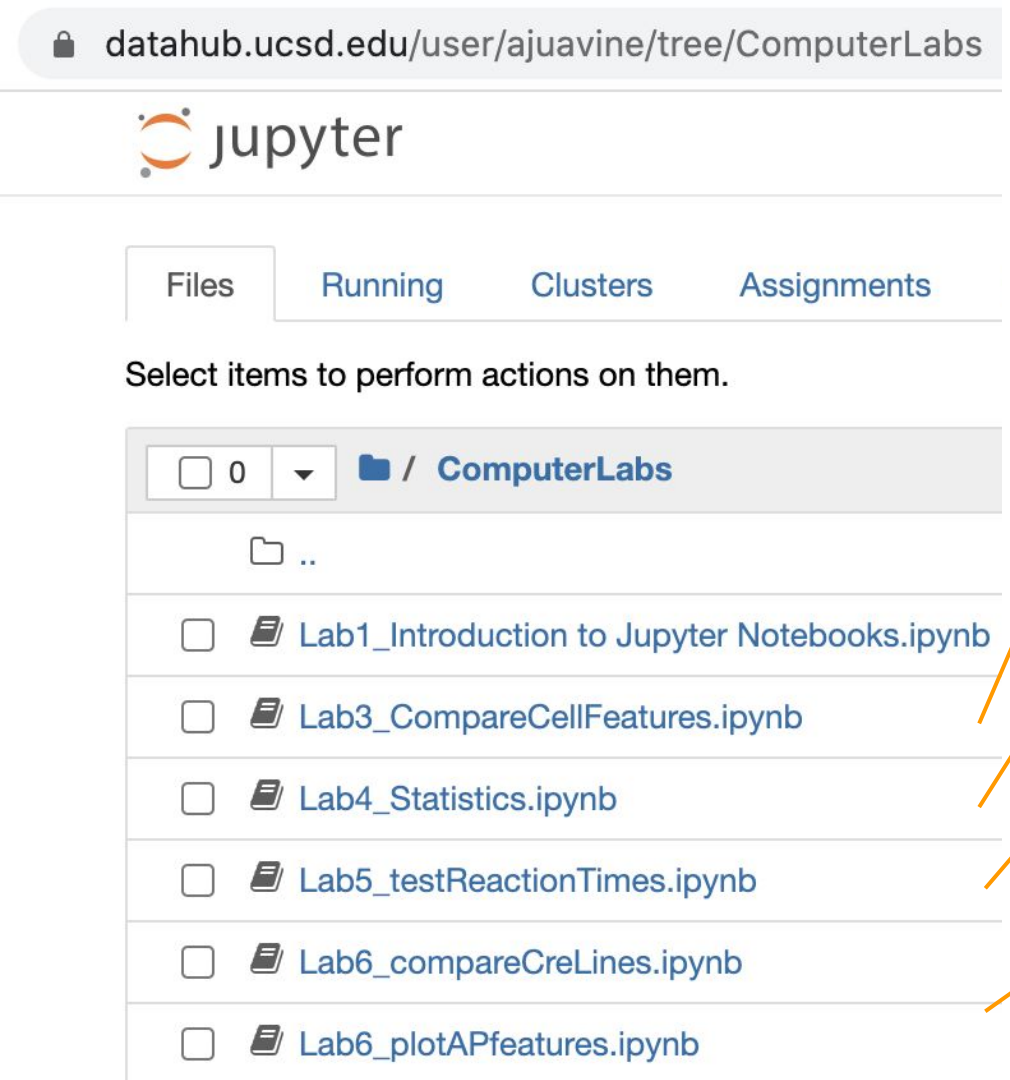
From Wallisch (2017)

There are several ways to implement Jupyter Notebooks:

- Run notebooks via Binder or Google Colaboratory, as we'll do today
- Set up a JupyterHub for your course or university
- Students download the Anaconda Distribution on their computer
  - Includes Jupyter Notebook, many useful packages and more



UCSD has a data science online platform designed for exactly this, and Google Colab works well for educators at institutions without a Jupyter Hub



Choose cell features to compare (e.g., capacitance and surface area) — *data from the Allen Institute*

Demonstrate statistical concepts and implement simple two-sample statistics

Run auditory & reaction time tests & test for differences

Compare electrophysiological properties (e.g., spike width, tau, resistance) in different Cre-lines — *data from the Allen Institute*

Notebooks are designed to expose students to coding, with straightforward tasks to complete

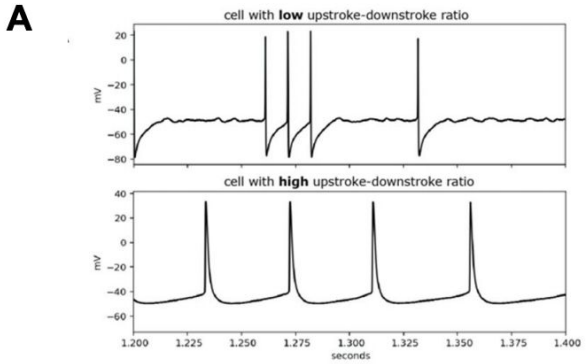
> **Task**: Choose two different Cre lines to compare, and assign them to the variables below by replacing the `...`. The value of your variable needs to be a **string** -- in other words, it should have quotes around it. The cell will print how many cells are in your datasets. If you have less than 10 cells, consider choosing a different Cre-line.

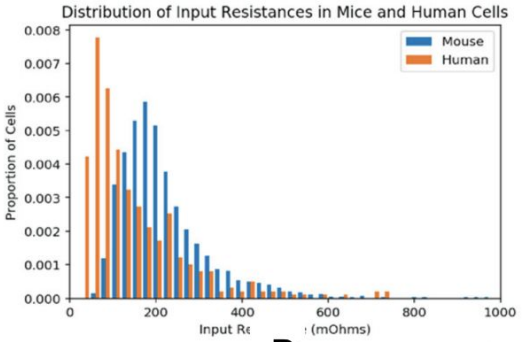
> You can find information on the different cre-lines that are available in this glossary or on the Allen Institute's website.

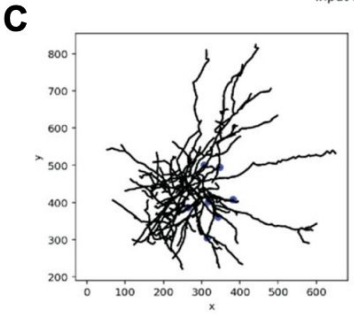
> **Note**: Be sure that you are using the *entire* name of the Cre line -- that means *everything* within the single quotes above.

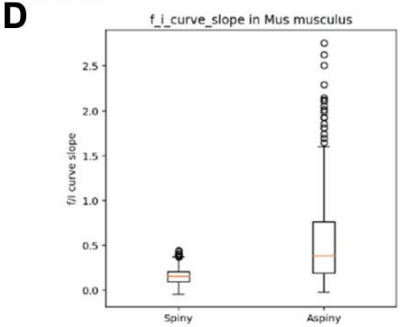
In [ ]:
```
cre_line_1 = ...
cre_line_2 = ...

cre_line_1_df = mouse_ephys_df[mouse_ephys_df['transgenic_line']==cre_line_1]
cre_line_2_df = mouse_ephys_df[mouse_ephys_df['transgenic_line']==cre_line_2]

print(cre_line_1 + ' has ' + str(len(cre_line_1_df)) + ' cells')
print(cre_line_2 + ' has ' + str(len(cre_line_2_df)) + ' cells')
```

Students can plot raw electrophysiology traces and choose their own features to compare across Cre-lines, species, or spiny/aspiny cell types









Juavinett, “Learning to code while analyzing an open access dataset,” *Journal of Undergraduate Neuroscience Education*, 2020

Juavinett, "Learning to code while analyzing an open access dataset," *Journal of Undergraduate Neuroscience Education*, 2020

Goals for this workshop

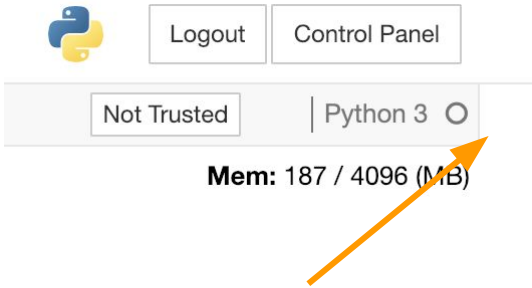
- Provide a general framework on teaching coding to neuroscience students
- Introduce Jupyter Notebooks as a tool for teaching
- **Demonstrate the use of the Jupyter Notebooks for:**
  - Statistics
  - Signal Processing
  - Neural Data Science

About Jupyter Notebooks

- Jupyter is a loose acronym for Julia, Python, and R 😏
- Run in a web browser!
- Shows plots directly in the notebook as you work your way through, performing analyses in real-time
- **Cell**: the main organizational structure of the notebook
  - Use **Shift+Enter** to run a cell (or press Run)
  - You can run cells out of order, and move cells around!
  - Cells can be **code** or **markdown** (descriptive text or images)

Using Jupyter Notebooks (continued)

- Processing-intensive cells can take longer to run, but your code may also get stuck in a cell.
  - Interrupt a stuck cell using Kernel > Interrupt
- **If you change anything in the cell, you need to re-run it.**
- For help:
  - Help > User Interface tour
  - Help > Keyboard Shortcuts



You can tell if the kernel is busy by whether or not the circle next to Python 3 (upper right corner) is filled or not. (filled = busy)

or if you have a stop sign in Colab

Let’s code!

https://bit.ly/ViABLEcoding

That link will send you to:
https://colab.research.google.com/github/ajuavinett/ViABLE_2021/

Additional materials are available: https://github.com/ajuavinett/ViABLE_2021