



Integrating programming into neuroscience courses

Ashley Juavinett, PhD

 @analog_ashley

UC San Diego Neurobiology



To get us started...

Please take the pre-workshop survey

<https://forms.gle/2HosnNnvGrtH3QWTA>

(Optional) Make a GitHub account:

<http://www.github.com>

(Optional) Request a Dandihub account:

<https://dandiarchive.org/>

All of the materials for today's workshop
live here: http://www.github.com/ajuavinett/FUN_2023



By the end of this workshop, you'll be able to:

- Identify multiple ways of integrating programming into your course
- Access and use internet-based computing platforms (e.g. Google Colab)
- Write & edit code in Jupyter (or similar) notebooks
- Access Neurodata Without Borders datasets for both teaching and research



Workshop plan

Time	Description
2:45 - 3:00 pm	Part I: Motivation & Framework
3:00 - 3:30 pm	Part II: The Tools (choosing a programming language, working with Jupyter/Colab notebooks)
3:30 - 3:40 pm	Break
3:40 - 4:20 pm	Part III: Cell Types Lesson
4:20 - 4:40 pm	Part IV: Neurodata Without Borders
4:40 - 4:50 pm	Part V: Wrap-up & promising practices for teaching coding



PART I

The Motivation



Neuroscience data is getting bigger

One $\sim 500 \text{ mm}^3$ mouse brain...

assuming $\sim 1,000$ connections for each neuron, the resulting connection matrix contains $\sim 10^{11}$ entries

= **a few hundred gigabytes**

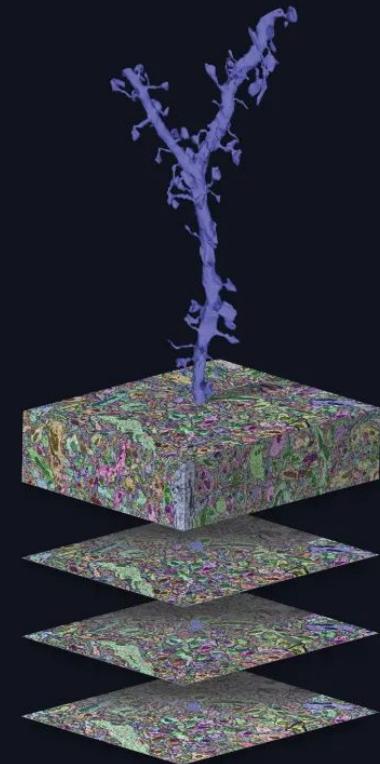
imaging at $5 \text{ nm} \times 5 \text{ nm} \times 40 \text{ nm}$ resolution

= **500 petabytes**

recording from each pixel for 20 min at 1000 Hz

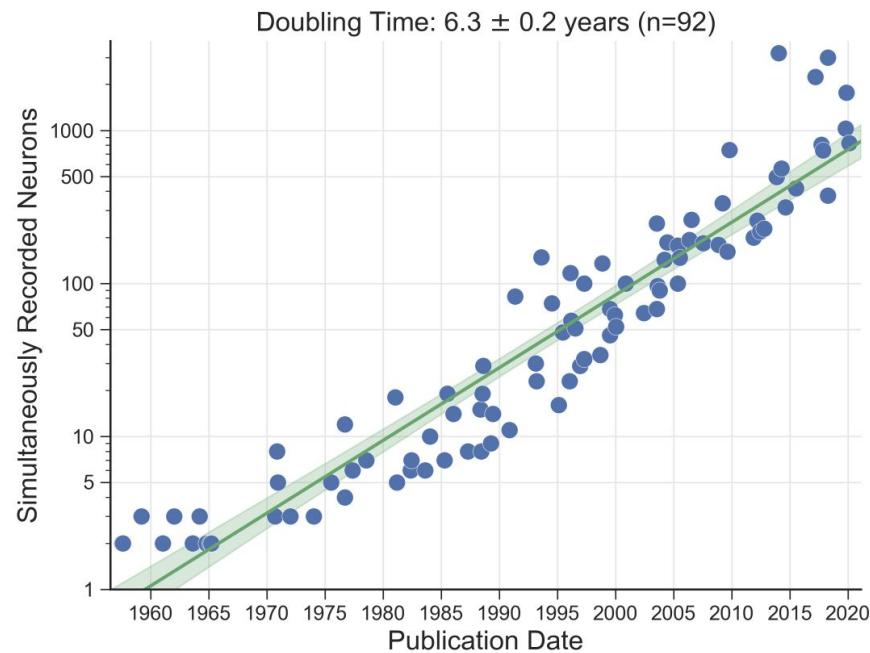
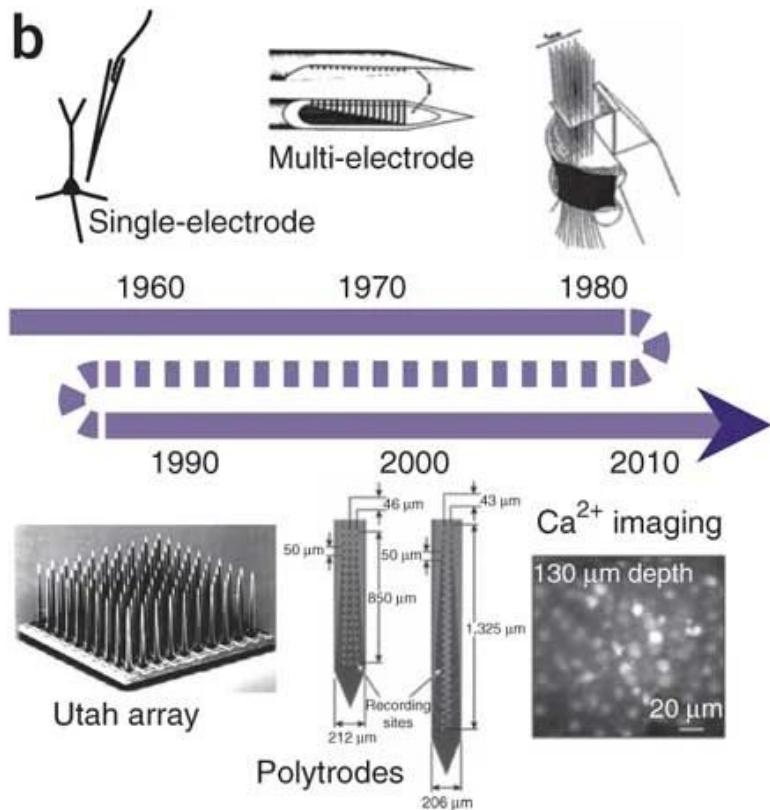
= **500 petabytes**

Estimates: Engert, 2014; Image: [EyeWire](#)



eyewire.org

The number of neurons we can record from is growing exponentially



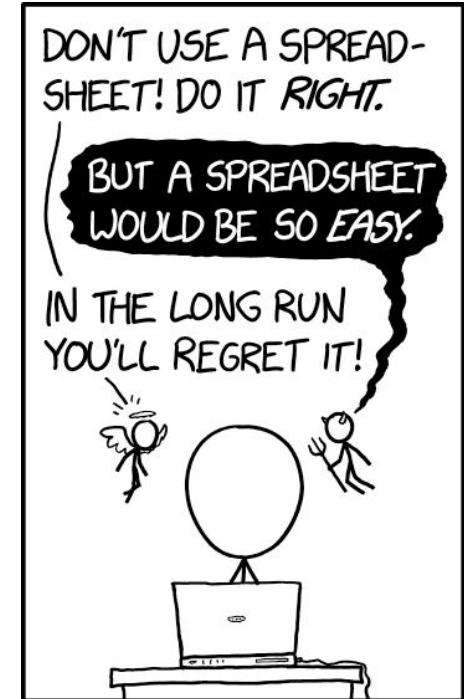
Stevenson & Kording, 2011;
<https://stevenson.lab.uconn.edu/scaling/>

Excel can only handle datasets with **“1 million rows, and 16,000 columns** — many datasets in biology are much larger than this!

You can automate analyses in Excel, but this is quite limited.

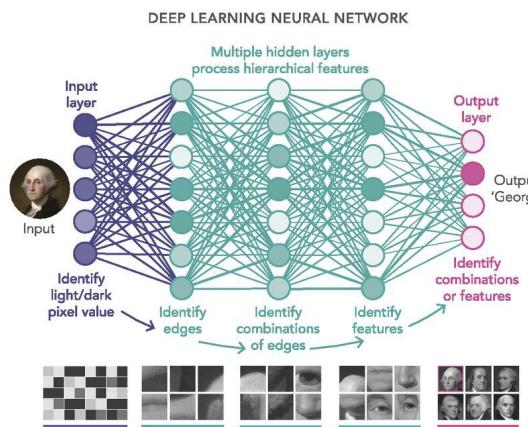
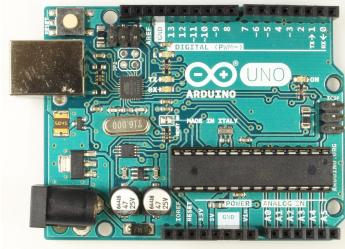
There are also specialized biological data analysis software programs, but often these are limited in how much they can be customized.

Code is *infinitely* customizable.



<https://xkcd.com/2180/>

Why do neurobiology students need to learn how to code?



wrangling large datasets

custom data acquisition

statistics, data analysis & visualization

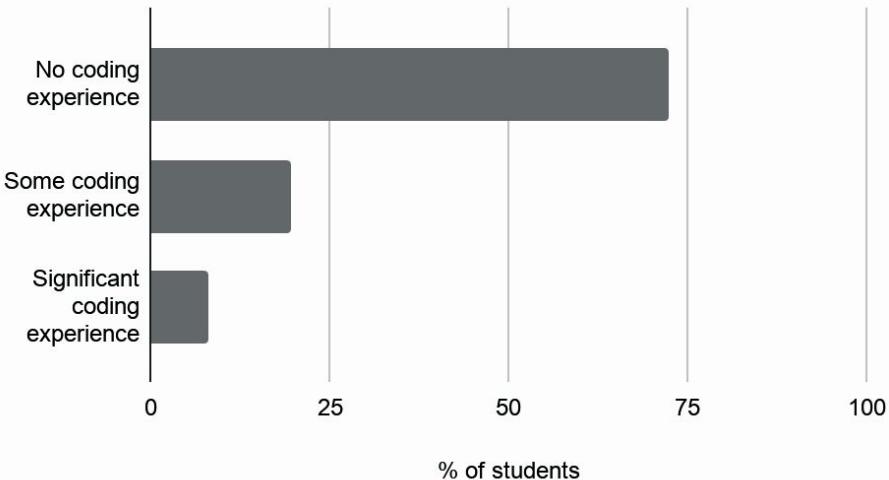
computational modeling



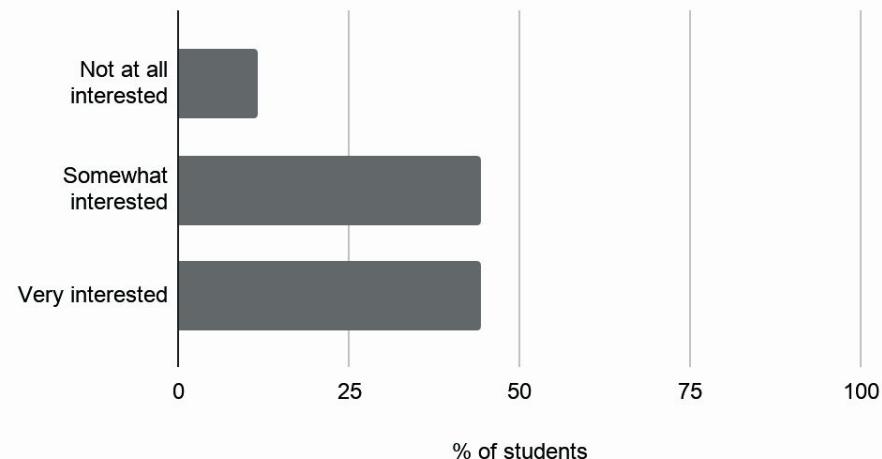
**The next generation of
neuroscientists needs new
skills to adapt to this
changing landscape of data.**

Mismatch between experience and interest

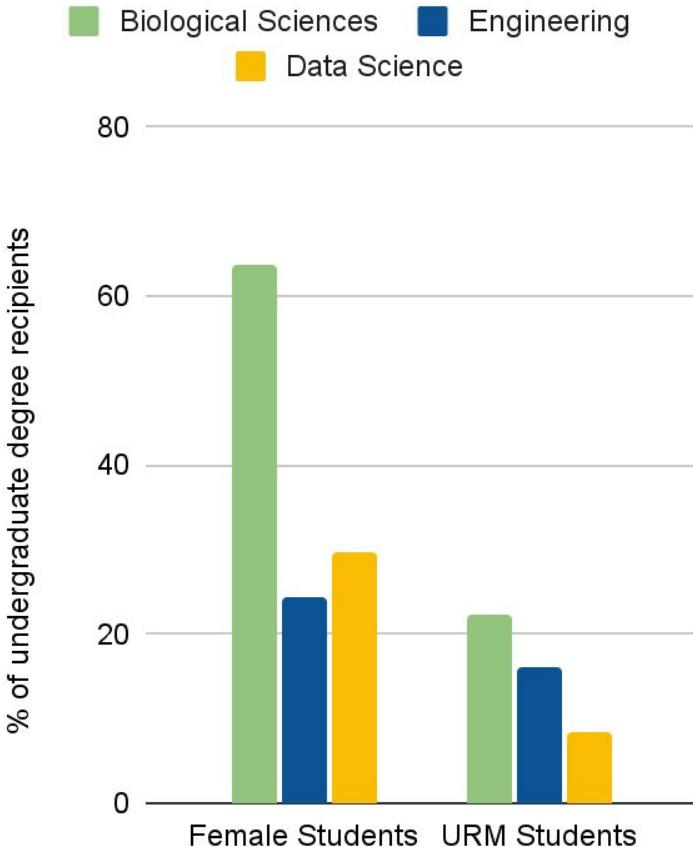
How much coding experience do you have?



How interested are you in learning how to code?



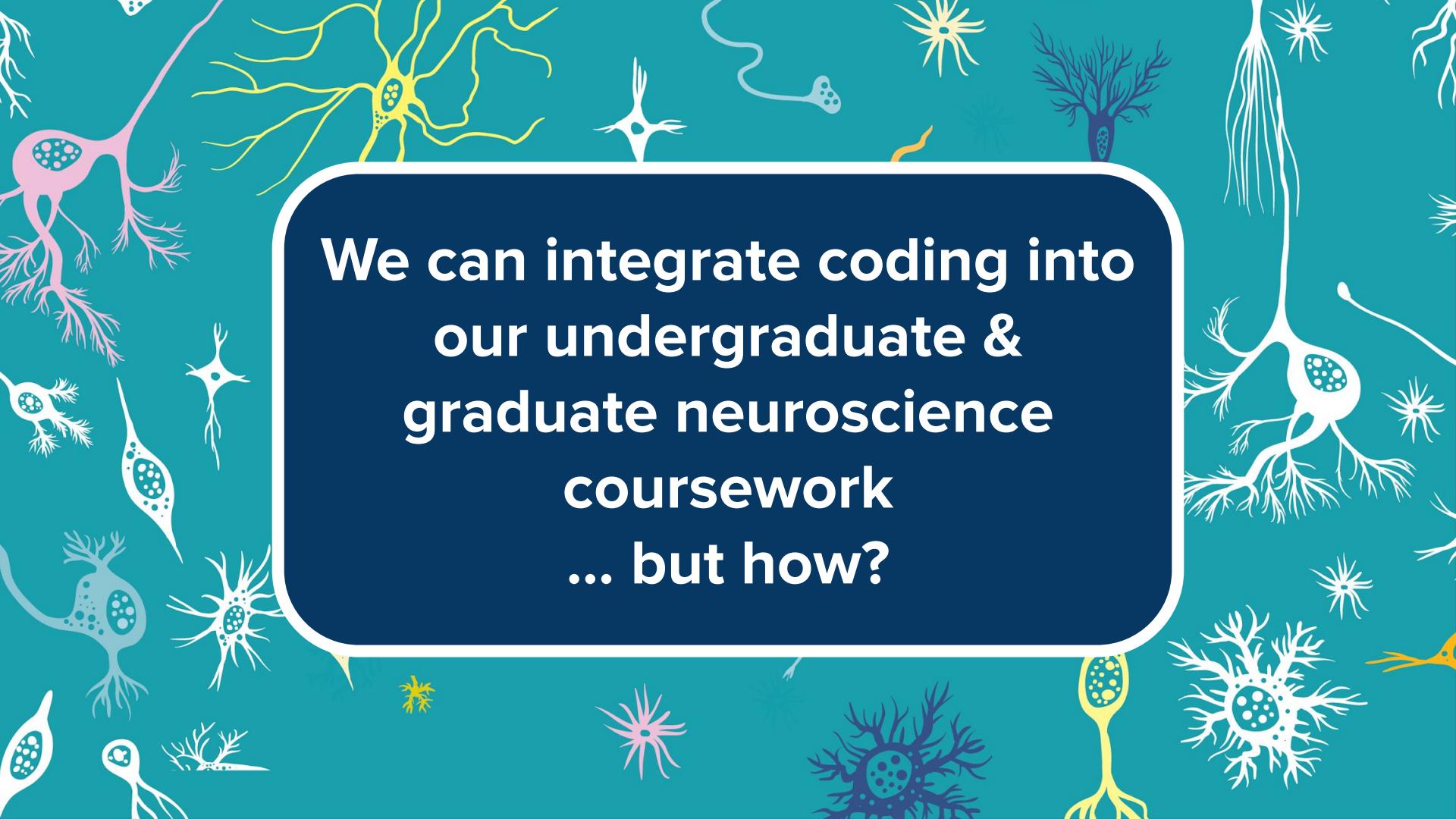
4 sections of neurobiology lab; 138 students total



This is also a matter of equity.

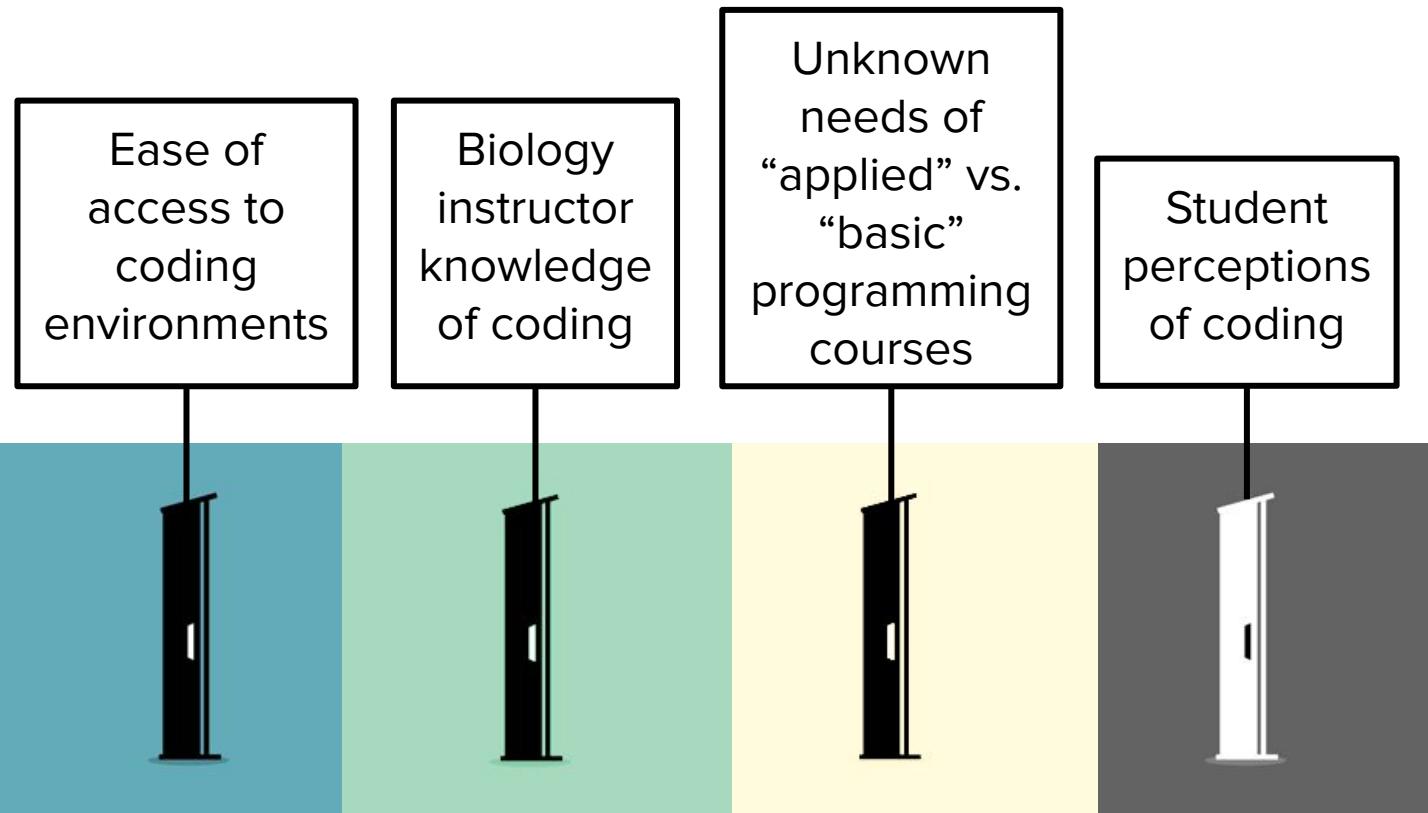
More and more jobs require programming experience, and the jobs that do pay more.

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



We can integrate coding into
our undergraduate &
graduate neuroscience
coursework
... but how?

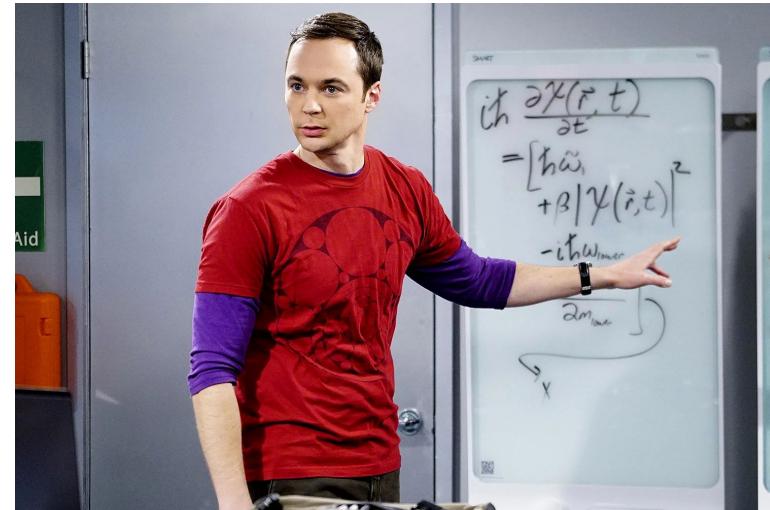
Barriers to bringing coding into neuroscience courses



Juavinett, *Neuron*,
“[The next generation
of neuroscientists
needs to learn how to
code, and we need
new ways to teach
them](#)” 2022

Student perceptions of coding & computer science (CS)

- **singularly focused** - CS requires an obsession with CS, necessarily at the exclusion of having other interests or meeting personal needs.
- **asocial** - computer scientists do not have social skills and CS requires working in isolation.
- **competitive** - CS courses (and thus potentially the field as a whole) are competitive and CS work is done individually as opposed to in collaboration with others.
- **male** - success in CS requires one to identify as male, or men are innately more talented in CS than women.



Lewis et al., “I Don’t Code All Day”: Fitting in Computer Science When the Stereotypes Don’t Fit (2016)

See also Cheryan et al., 2009, 2011, & 2013; Cvencek et al., 2015; Ensmenger, 2012; Goodenow, 1993; Lee 1998; Margolis & Fisher, 2003; Margolis et al., 2010; Steele, 1997

Neurobiology students start 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

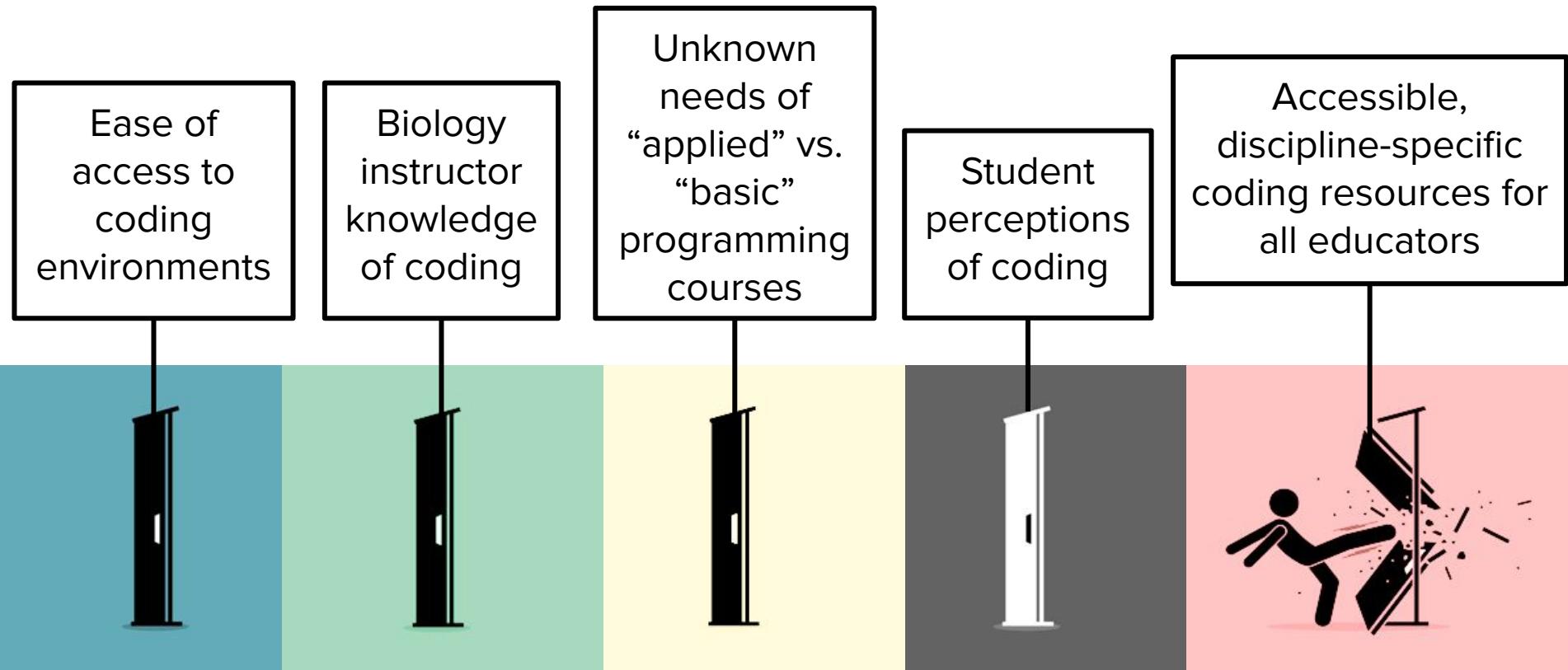
I've heard very awful things about computer science courses but it seems interesting especially if the connection to neurobiology is there!

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?*”

Barriers to bringing coding into neuroscience courses



Three possible ways to include coding into the curriculum

Exposure to coding in coursework

- Limited time to devote to coding
- Students see utility and *edit* code but do not *write* their own code
- Using code may even be optional



Less independent coding
Lower instructor proficiency

Integration of coding in coursework

- 1-2 hours dedicated to contextualized, introductory programming lessons
- Students edit code and *begin* to write their own
- Lessons exist in standalone modules
- Students can ask their own questions of data

More independent coding
Higher instructor proficiency



Files Running Clusters Assignments

Select items to perform actions on them.

<input type="checkbox"/> 0	<input type="button" value="▼"/>	<input type="button" value="📁"/> / ComputerLabs
		<input type="button" value="📁"/> ..
<input type="checkbox"/>	<input type="button" value="📝"/>	Lab1_Introduction to Jupyter Notebooks.ipynb
<input type="checkbox"/>	<input type="button" value="📝"/>	Lab3_CompareCellFeatures.ipynb
<input type="checkbox"/>	<input type="button" value="📝"/>	Lab4_Statistics.ipynb
<input type="checkbox"/>	<input type="button" value="📝"/>	Lab5_testReactionTimes.ipynb
<input type="checkbox"/>	<input type="button" value="📝"/>	Lab6_compareCreLines.ipynb
<input type="checkbox"/>	<input type="button" value="📝"/>	Lab6_plotAPfeatures.ipynb

Notebooks for a neurobiology lab class

Choose cell features to compare — *data from the Allen Institute*

Demonstrate statistical concepts and implement two-sample statistics

Compare electrophysiological properties between cell types — *data from the Allen Institute*

Course JupyterBook:
bipn145.github.io

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')
```

Three possible ways to include coding into the curriculum

Exposure to coding in coursework

- Limited time to devote to coding
- Students see utility and *edit* code but do not *write* their own code
- Using code may even be optional

Contextualized (discipline-based) coding classes

- An entire class dedicated to teaching coding in the context of neuroscience
- Students design & write their own code
- May include novel research projects



Less independent coding
Lower instructor proficiency

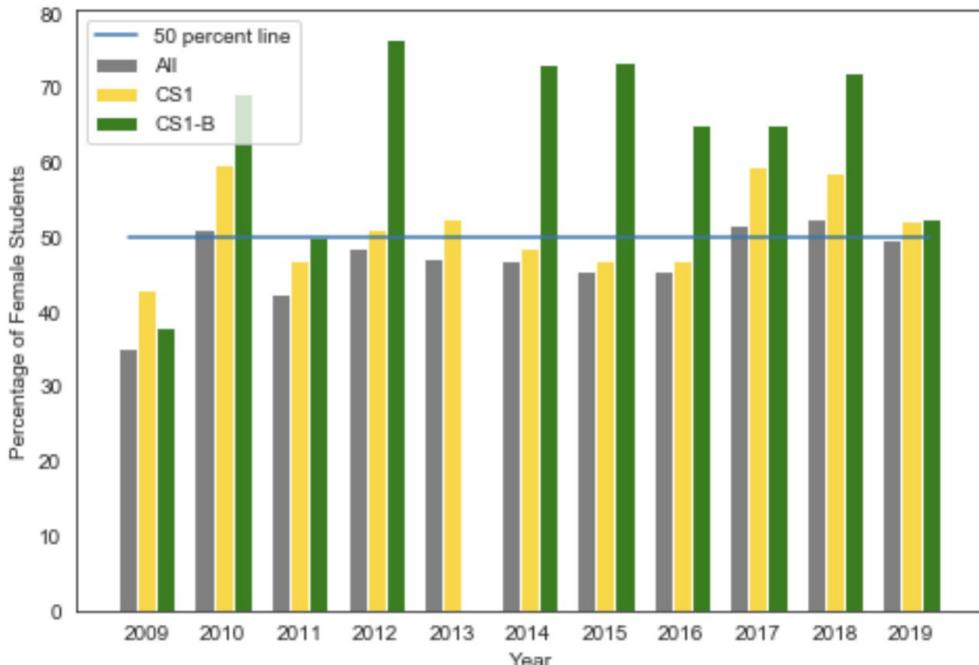
Integration of coding in coursework

- 1-2 hours dedicated to contextualized, introductory programming lessons
- Students edit code and *begin* to write their own
- Lessons exist in standalone modules
- Students can ask their own questions of data

More independent coding
Higher instructor proficiency

Contextualized coding invites in more female students

Figure 2: Female Students: % in CS1 and CS1-B, per year



learn how
to code
while
looking at
> genes
> cells
> brains

Introduction to Python for
Biologists (BILD 62)
Spring 2022, T/Th@2pm

Dodds et al., S/GCSE 2021
“A Biology-based CS1: Results and
Reflections, Ten Years in”
(results from this study less clear for
first-generation or URM students)



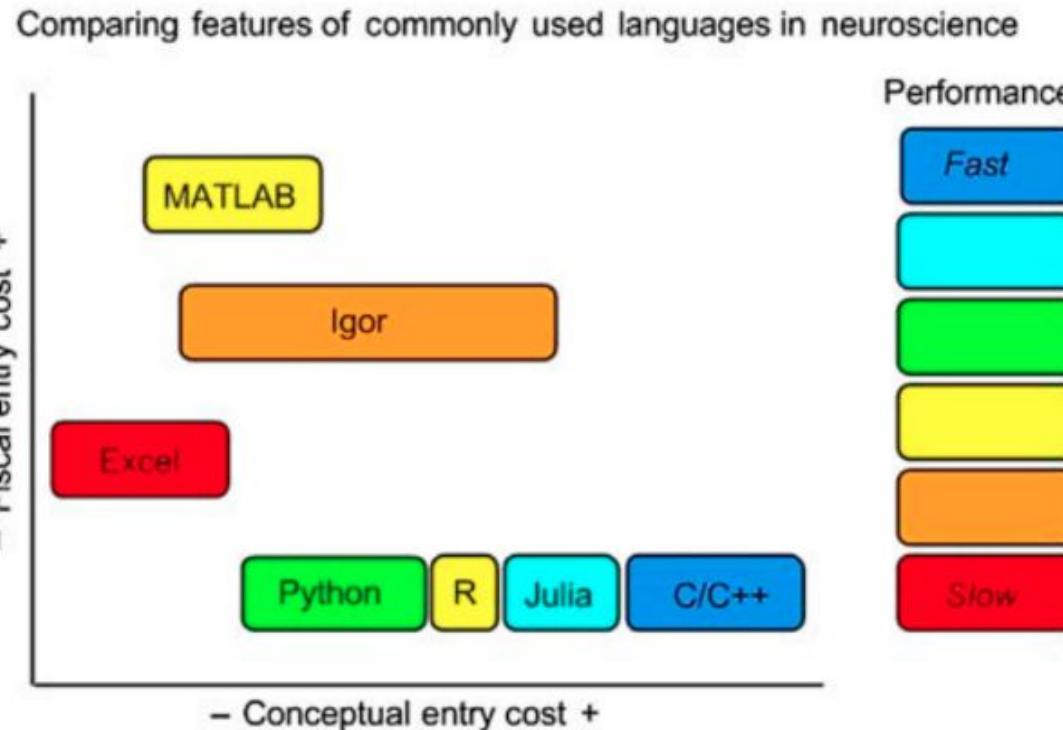
PART II

The Tools



Considerations for choosing a programming language

- Fiscal & conceptual entry (see right)
- Usage in neuroscience
- Ease of use
- *My recommendation:*
Python or R



From Wallisch (2017)

DATA SCIENCE / MACHINE LEARNING PLATFORM



Information Technology Services - Educational Technology Services

Help ▾ FAQ



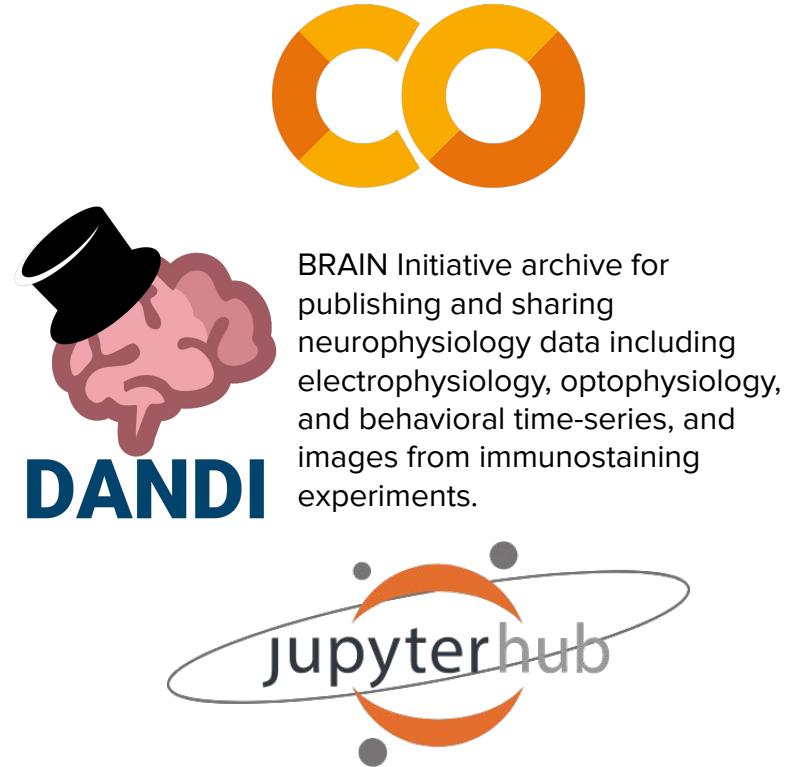
UCSD has a data science online platform designed for exactly this, and Google Colab works well for educators at institutions without a Jupyter Hub

Considerations when choosing tools & a teaching approach

- How much do you want students to write code *de novo* versus editing provided code?
 - Carpentries Model: live coding on a blank page
- Do you want students to be dependent on (sometimes expensive) school or course-provided software?
- Do students have their own laptops (or do they have iPads)?
- Do students have access to a computer with a decent internet connection?

Tools for today

- **Parts II & III:** Google Colab, a cloud-based computing *notebook*
 - Very similar to Jupyter notebooks, with small differences.
 - Conveniently saves & launches from Google Drive
- **Parts IV:** Dandihub (optional), cloud-based computational environment that includes Jupyter Notebooks & pre-configured docker container





Let's open up our first
notebook from
[http://www.github.com/
ajuavinett/FUN_2023](http://www.github.com/ajuavinett/FUN_2023)



PART III

Cell Types Lesson

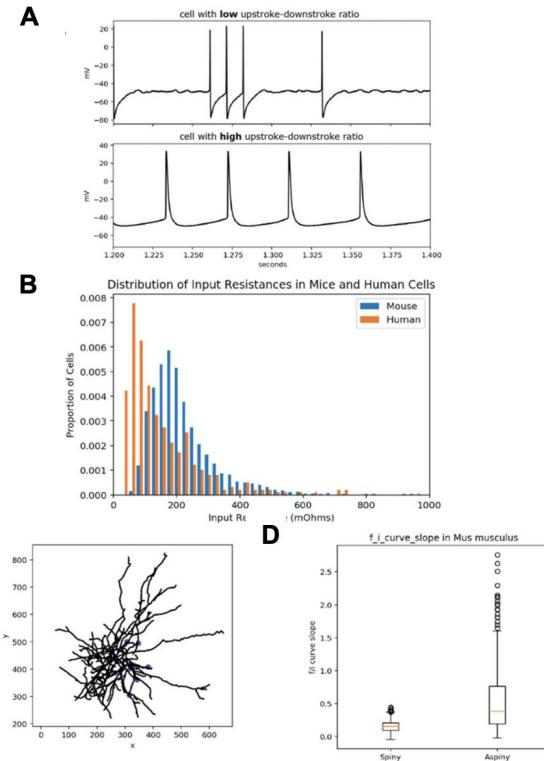


Cell Types Lesson Learning Objectives

- Relate the diverse intrinsic electrophysiological features of neurons to their structure and function
- Compare electrophysiological characteristics of neurons in humans and mice
- Practice using Jupyter Notebooks to run and edit Python code
- Develop a sense of belonging and self-efficacy in coding and neuroscience research

More information & resources:

<https://sites.google.com/ucsd.edu/neuroedu/cell-types>



Juavinett, “Learning to code while analyzing an open access dataset,” JUNE, 2020



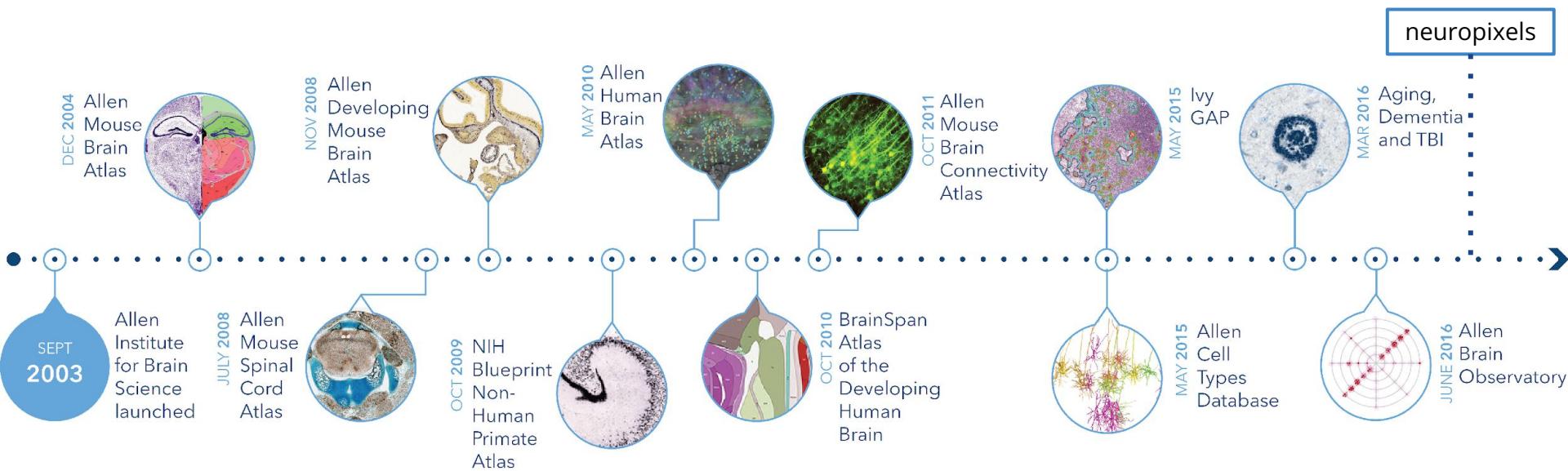
PART IV

Teaching with Open Datasets

(e.g. Neurodata Without Borders)



Neuroscience datasets are increasingly being shared online



NEURODATA
WITHOUT BORDERS



DANDI
dandiararchive.org

Neuroscience Datasets star file refresh

File Edit View Insert Format Data Tools Add-ons Help Last edit was seconds ago

1 Database Name

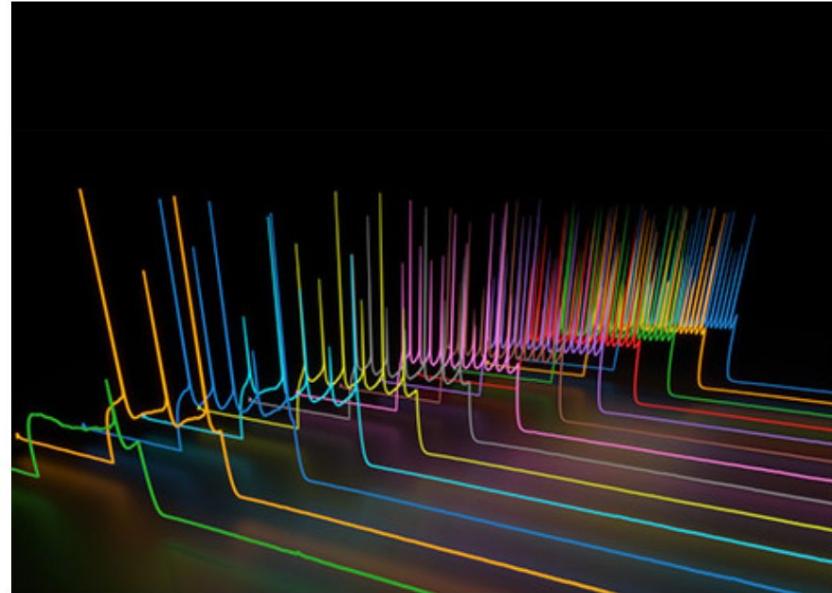
A	B	C	D	E	F	G	
1	Database Name	URL	Data Summary	Student Objective	Accessing API/SDK	Using Pandas	Text Mining
2	NeuroSynth	https://neurosynth.org/	Summarizes fMRI data from many studies	Perform meta analyses of fMRI data			X
3	LISC	https://lisc-tools.github.io/	Text analysis of papers	Search publications for terms to identify interesting intersections.			X
4	Allen Cell Types Atlas	http://celltypes.brain-map.org/	Whole-cell electrophysiology in genetically-identified cell types in mice and humans	Compare electrophysiology metrics for different cell types in mice and humans.	X	X	
5	Allen Brain Observatory	http://observatory.brain-map.org/	In vivo 2p imaging in genetically-identified cell types in mice	Compare visual responses of cells recorded via two-photon imaging and analyze correlations.	X	X	
6	Allen Neuropixels	https://allensdk.readthedocs.io/en/stable/api/allen.brain_map.allen_neuropixels.html	Extracellular recording & behavior in mice performing a task	Compare visual & behavioral responses of cells recorded with high density extracellular recording arrays. Also includes LFP data.	X	X	
7	Allen RNAseq	https://portal.brain-map.org/analysis/rnaseq	RNAseq in mice & humans	Compare gene expression in mice and humans in different cells & brain regions.		X	
8	Allen Connectivity	http://connectivity.brain-map.org/					
9	Allen Synaptic Physiology	https://portal.brain-map.org/analysis/synaptic_physiology					
10	Neuropixels Spike Sorting	http://repository.cshl.edu/index.html?category=spike%20sorting	Lots of spikes	PCA analysis to isolate units in extracellular recording			

Open Neuroscience Datasets <https://bit.ly/openneurodatasets>

Neurodata Without Borders

Neurodata Without Borders (NWB) is a data standard for neurophysiology, providing neuroscientists with a common standard to share, archive, use, and build analysis tools for neurophysiology data. NWB is designed to store a variety of neurophysiology data, including data from intracellular and extracellular electrophysiology experiments, data from optical physiology experiments, and tracking and stimulus data.

The **NWB team** consists of neuroscientists and software developers who recognize that adoption of a unified data format is an important step toward breaking down the barriers to data sharing in neuroscience.



(Courtesy of the Allen Institute for Brain Science)

See also Rübel et al. (2022):
<https://elifesciences.org/articles/78362>

d.

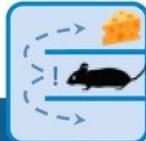
Experimental Subject

- Species
- Genotype
- Age and weight
- Custom subject-specific fields



Experimental Design

- Stimuli
- Environment
- Trial structure
- Epochs
- Perturbations



Data Acquisition

- Device settings
- Filtering parameters
- Sampling rate
- Recording/imaging area
- Emerging technologies



Behavior

- Position and speed
- Choice and response time
- Video and audio
- Eye tracking
- Experiment-specific measures



Neuronal Activity

- Extracellular electrophysiology
- Intracellular electrophysiology
- Optophysiology
- Pre-processed data
- New data modalities





Teaching and Learning with NWB Datasets

Teaching and Learning with NWB Datasets

For many, many years in neuroscience, the landscape of file formats has been the wild west. Everyone either uses a format either dictated by their lab or by the equipment they're using. This poses a huge challenge to data sharing: if you want to collaborate with the lab next door and share data with them, one of the first questions is, "what format is your data in?" While static images taken from a microscope and standard biology lab procedures like western blots are fairly straightforward to share, physiology and behavior data types are much more diverse.

That's where [Neurodata Without Borders](#) (NWB) comes in. NWB is a consortium of researchers who are aiming to *standardize neuroscience data on an international scale*. The use of this standard dataset format will make sharing and contributing to open source projects easier, ultimately accelerating discovery.

NWB has pioneered a data format called [Neurodata Without Borders: Neurophysiology \(NWB:N\)](#). NWB:N is a standard to share, store, and build analysis tools for neuroscience data. It is built on [HDF5 \(hierarchical data format\)](#), which was designed for large data storage and multi-language accessibility. In other words, it's good at doing exactly what we need it to do.

Standardized data for education

A standardized and accessible data format like NWB also opens possibilities for more easily integrating these cutting-edge datasets into the classroom. **This online textbook aims to provide interactive and self-**



In progress: online lessons using NWB datasets at nwb4edu.github.io

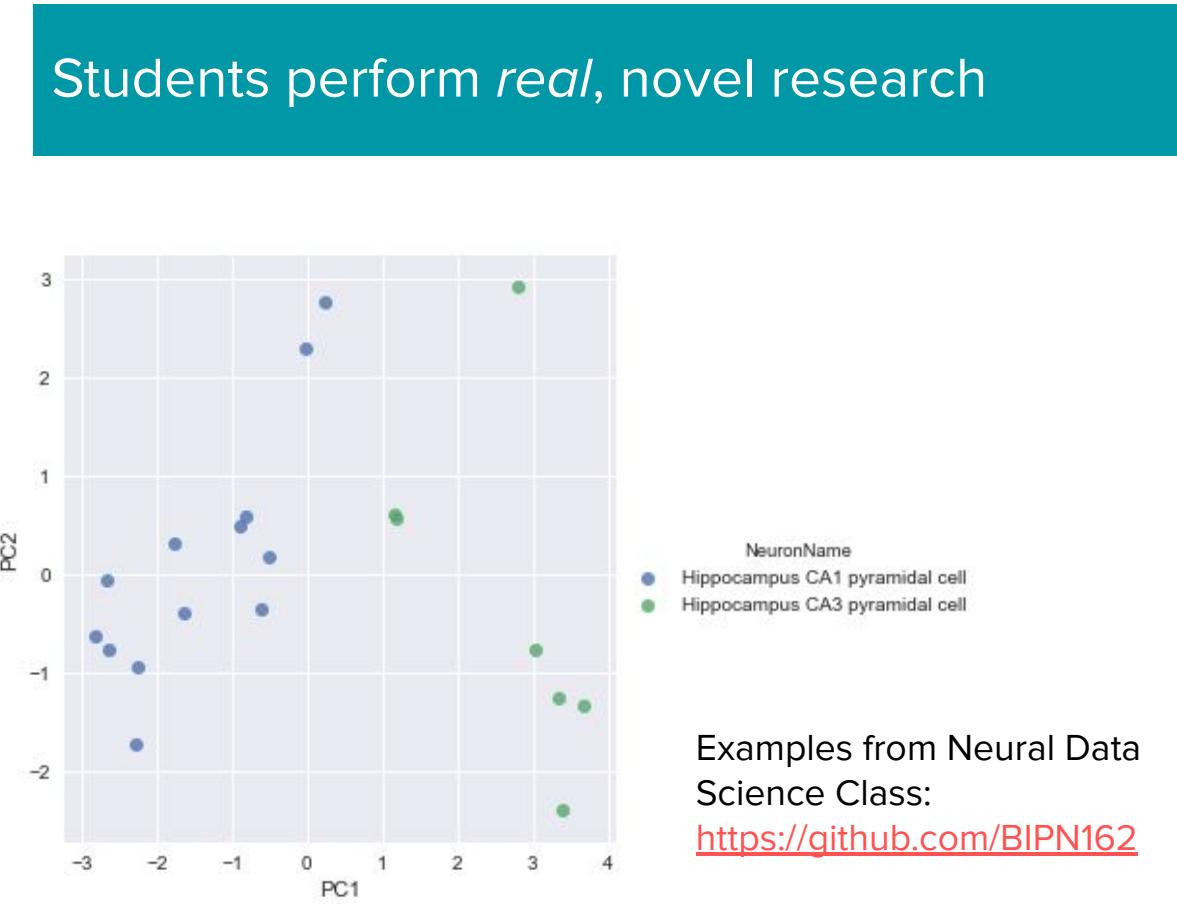
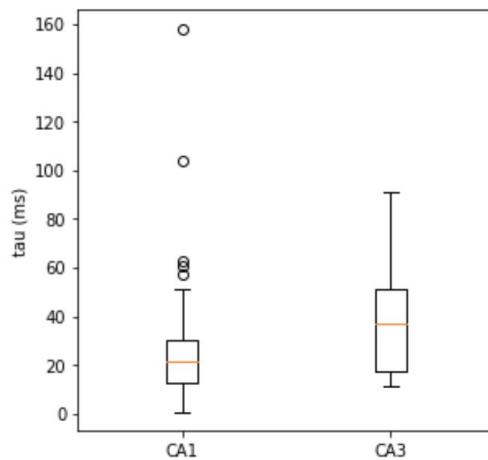
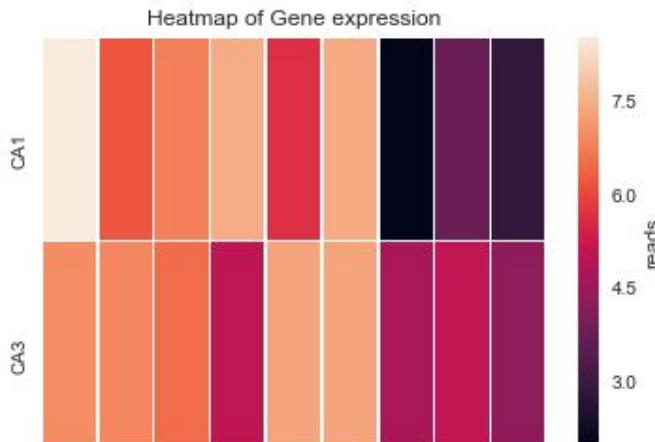


Contents

- Standardized data for education
- Using this textbook
- Content outline
- Support
- Additional resources
- License

Looking for
paid test
users!







For the final demo, we'll work out of the
Dandihub: hub.dandiarchive.org

```
git clone https://github.com/ajuavinett/FUN_2023
```



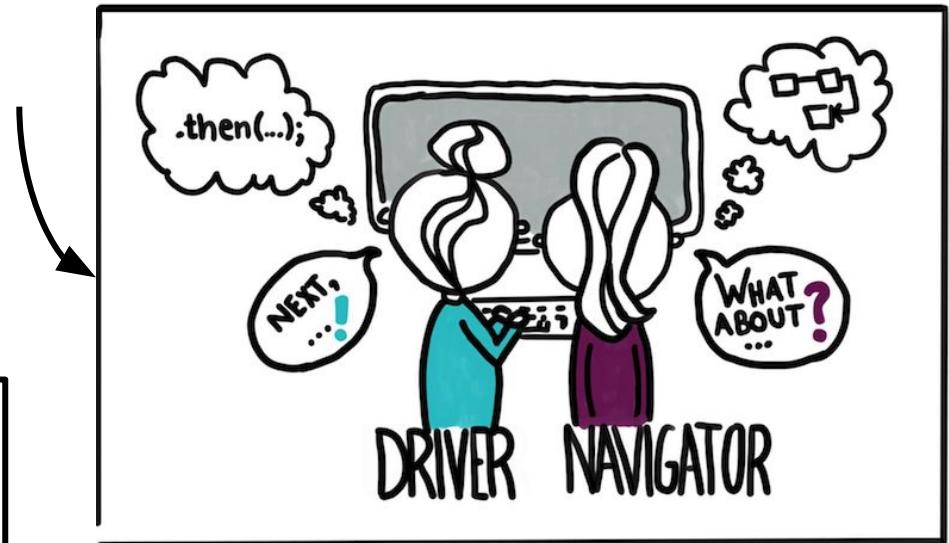
PART V

Wrap-Up & Promising Practices



A general framework for teaching coding to neuroscience students

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



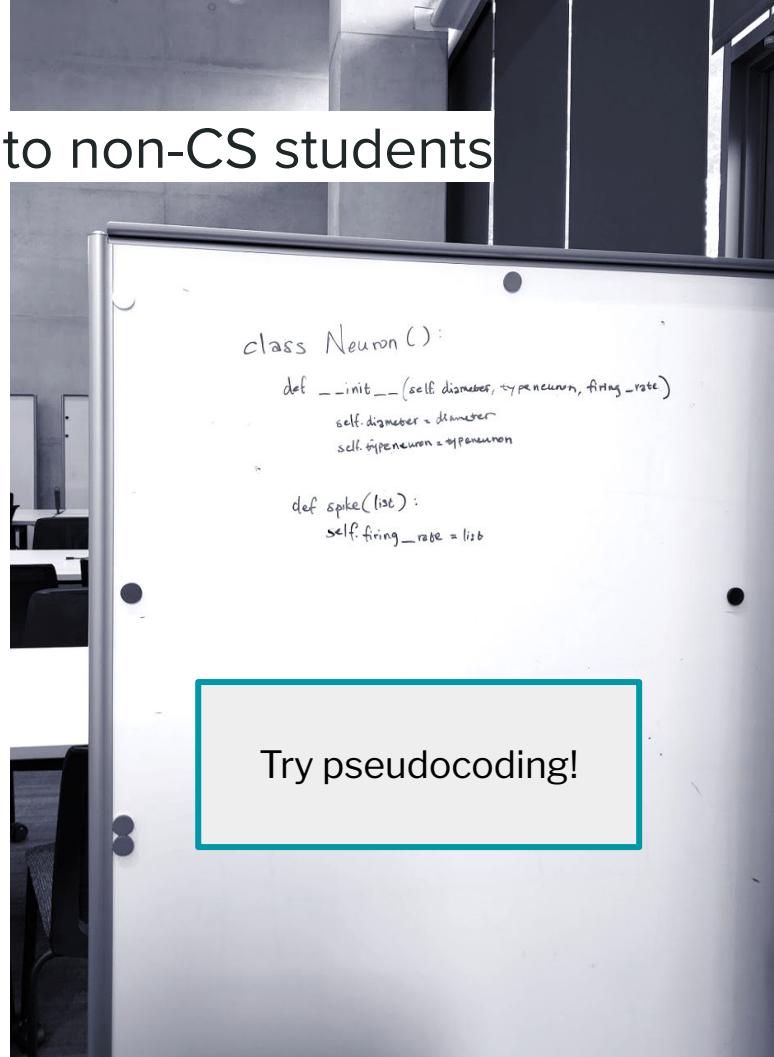
[Image Source](#)

More thoughts on this: [Ten quick tips for teaching programming](#)

A general framework for teaching coding to non-CS students

- Frame coding as a **skill** — *it's like learning a language*
- Encourage **growth mindsets** around coding
 - Students with growth mindsets perform better over time
 - 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](#)



Oh yeah, what about chatGPT (or other LLMs)?

**You (and your students) can and
should use it to learn!**

How you integrate it into your class
depends on your learning objectives!



Extensions of what you've seen here

- Addition of transcriptomic data (ongoing work)
- Recent publication by Ho et al. (2021)
- Your work?



The Journal of Undergraduate Neuroscience Education (JUNE), Fall 2021, 20(1):A100-A110

ARTICLE

Pandemic Teaching: Using the Allen Cell Types Database for Final Semester Projects in an Undergraduate Neurophysiology Lab Course

Yi-Yun Ho¹, Andrea Roeser¹, Gwenda Law², and Bruce R. Johnson¹

¹Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853; ²Department of Biomedical Engineering, Cornell University, Ithaca, NY 14853.

We designed a final semester research project that allowed students to apply the electrophysiological concepts they learned in a lab course to propose and answer experimental questions without access to laboratory equipment. We created the activity based on lesson plans from Ashley Juavinett and the Allen Institute for Brain Science (AIBS) Allen SDK online examples. An interactive graphic interface

final semester project allowed students to ask real-world medical and scientific questions from “start to end”. Through this project, students developed skills to navigate an extensive online database and gained experience with coding-based data analysis. They chose neuronal populations from human and mouse brains to compare passive properties and neuronal excitability

Thank you.

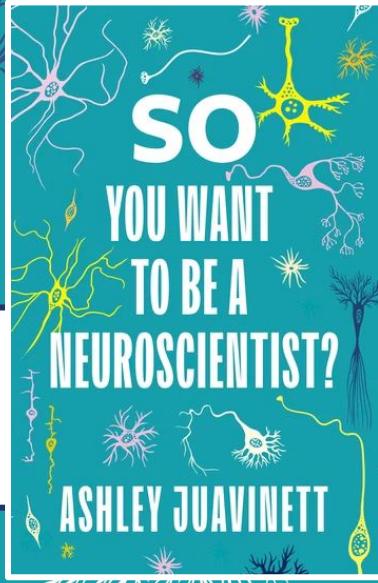
THE  KAVLI FOUNDATION

Support for online neural
data science textbook

Victor Magdaleno-Garcia,
mentee & research intern



@analog_ashley



ALLEN INSTITUTE *for*
BRAIN SCIENCE

Support for using & sharing datasets

Kaitlyn Casimo, Saskia De Vries, Tom Chartrand

Austin Zuckerman

Current research on
computing attitudes in
biology students





Post Workshop Survey

<https://forms.gle/65oaDctnPVccnUMGA>