BIPN 162 | Neural Data Science

SUMMER 2020

Instructor
Ashley Juavinett, PhD
ajuavine@ucsd.edu

Course GitHub: http://github.com/BIPN162

(Virtual) Office hours
Thursdays, 10-11 am, or by appointment
https://ucsd.zoom.us/my/ajuavinett

Instructional Assistant
Jennifer Grundman, jagrundm@ucsd.edu
Office hours (Fridays, 10-11 am):
https://ucsd.zoom.us/j/5022465808

Class Schedule
Lectures: Mon/Wed, 11 am - 1:50 pm
Discussion: Mon/Wed, 10-10:50 am
(See links on Canvas)

Course Description

Project-based course in which students will use computational notebooks to perform exploratory data analyses and to test hypotheses in large neuroscience datasets, including the differences between unique neuron types, leveraging text mining of the neuroscience literature, and human neuroimaging analyses. *Prerequisites*: BIPN100 & MATH11 (or comparable courses)

How class will work remotely

Given that four hours (Discussion + Lecture) is an unreasonably long time to sit in front of a computer, Discussion sections will end promptly at 10:50 am, and we will break our Lecture sessions into two chunks. Most days, I will livestream (and record) a lecture and introduction to the day at 11 am, followed by time for you to code with your group (and get lunch). We'll regroup at 1 pm on that same day to recap, or discuss a new topic.

Lecture sessions will largely be **interactive** — in other words, you'll be encouraged to talk with classmates and code while I'm coding. However, I recognize that everyone may not be able to tune in. So, lectures are **not mandatory** and will be recorded and shared. **You are *not* required to attend class, but you are encouraged to if you can**. Similarly, you're not required to turn on your video, but you're encouraged to if you feel comfortable. You may complete the coding activities with your group on your own time, but I encourage you to complete them during our synchronous sessions when the teaching staff is online and available.

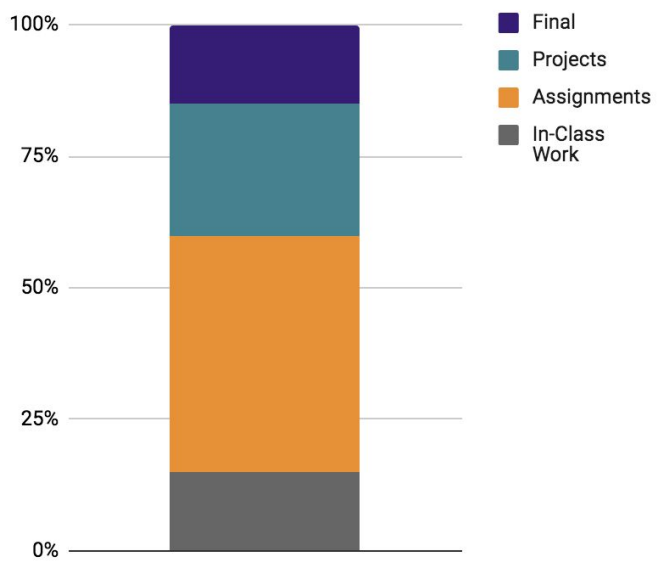
We'll be using Slack as the primary way to communicate and to hopefully build a bit of community, even in this online setting. Alternatively, you're welcome to use email with questions.

If you need additional help getting online, please check out this resource from the Teaching+Learning Commons: https://digitallearning.ucsd.edu/learners/learning-remote.html. If you're still having issues, please feel free to reach out to the teaching staff.

Course learning objectives

- Write and edit Python code, particularly in Jupyter Notebooks
- Develop hypotheses specific to big data environments in neuroscience
- Design a big data experiment and excavate data from open sources
- Integrate data from multiple datasets to answer a biological question

Grading



- **In-class work (15%)**: We will complete several activities in class for credit. If you miss these, you'll need to contact the IA and make them up before Friday at 5 pm.

- **Assignments (45%)**: Weekly take-home coding assignments will support your progression through the course topics and will directly relate to the larger class projects. Assignments will be submitted through the DataHub (http://datahub.ucsd.edu) and graded automatically using a tool called NBGrader.
    - Unless noted otherwise, assignments are due **Monday at 5 pm** and are worth 5-10% each.
    - These assignments should be completed individually and should take you about 1-2 hours. Discussion sections will greatly help you complete these assignments.

- **Final Project (25%)** Includes the project proposal, code, and deliverables. Your final project should integrate **two** different datasets to address a question about brain function. For example, you could investigate specific cell types in the brain, combining information across different data sets of your choosing (e.g., gene expression, visual responses, connectivity, and/or intrinsic activity.) Or, you could choose one brain region in humans and integrate datasets of your choosing (e.g., gene expression, Neurosynth, LISC, Human Brain Project) to address the function of that brain region and identify possible links between genes, circuits, and behavior. We will discuss possibilities for your project as we move through the course. You can see projects from WI20 here.

- **Final exam (15%)** The final in this class will largely test you on the biology content we cover, as well as some of the coding concepts that we have encountered throughout the quarter. The final is an open notes "take home" exam, and will be due on August 1st at 5 pm.

Additional notes about grading:

- We will be using Canvas (http://canvas.ucsd.edu)as well as the DataHub to manage grades and assignments.
- **Late policy**: Assignments and projects will lose -10% for each day they are late.
- **Grading Scheme:** Final scores will be converted to letter grades, where A=100-90%, B=89-80%, C=79-70%, D=69-60%, and F=59-0%. For positive and minus grades, A+ = 97-100, A = 93-96.99, A- = 90-92.99, B+ = 87-89.99, B = 83-86.99, and so on.

Course Resources

There is no official textbook for this course. Instead, we'll be relying on several online resources:

- VanderPlas, **Whirlwind Tour of Python**
- VanderPlas, **Python Data Science Handbook** (available free online or in print)
- Donaghue, Ellis, and Voytek, **Data Science in Practice**
- Wallisch, **Neural Data Science**
- Adhikari & DeNero, **The Foundations of Data Science**
- Software Carpentry, **Plotting and Programming in Python**

You're also strongly encouraged to sign up for DataQuest. They have many free tutorials in their Data Scientist Path that echo the coding skills we'll be learning in this class. Corresponding tutorials are denoted **in blue** on the syllabus, and you're encouraged to complete them before lecture.

Course Philosophy

A note on our course's environment

We'll be working together to create an equitable and inclusive environment of mutual respect, in which we all feel comfortable to share our moments of confusion, ask questions, and challenge our understanding. Everyone should be able to succeed in this course. If you do not feel that is the case please let me know.

Course accommodations

If you need accommodations for this course due to a disability, please contact the Office for Students with Disabilities (osd@ucsd.edu) for an Authorization for Accommodation letter. Please speak with me in the first week of class if you intend to apply for accommodations. For more information, visit http://disabilities.ucsd.edu.

This course, and the work it entails, is for *you*

You won't benefit if others do your work. If you're unclear about what constitutes cheating in this course, please ask. Cases of academic dishonesty or cheating will be first handled by me, and then by the Academic Integrity Office. If you become aware of cheating in this class, you can anonymously report it: https://academicintegrity.ucsd.edu/.

Syllabus

(this is a living document and is subject to change!)

Date	Topic	Before class
Week 1	What is neural data science? What kind of questions can you ask of your data? To set the foundation for this course, we'll introduce the approaches and tools that are commonly used to analyze big data sets in neuroscience, and learn some fundamental Python skills.	
Jun 29	*No discussion section on Monday of this week* **Introduction to Data Science and our Course** Why Python? Jupyter Notebooks **Programming Fundamentals I** Python syntax, expressions, & variables	*Before Class*: Read Mark Humphries, *The Spike*, "A Neural Data Science: How & Why" [link]
July 1	**Programming Fundamentals II** Data structures: lists, tuples, and dictionaries *ExpressionsVariablesStrings Worksheet* **Programming Fundamentals III** Booleans, conditionals, and loops	*Before Class*: Complete "Programming in Python" and "Variables and Data Types" on DataQuest. *Due Friday at 5 pm*: **A0** Computer Setup & the Entry Survey
Week 2	This week we'll start working with gene expression data from the Allen Brain Institute. We'll extend what you've learned in Week 1 to start thinking about object-oriented programming.	
July 6	**Programming Fundamentals III** Booleans, conditionals, and loops *(continued)* **Introduction to the Allen RNAseq database** Brain organization, RNA sequencing & gene expression	*Before class*: Complete DataQuest "Lists & For Loops" tutorial. *Due Monday at 5pm*: **A1** Programming Fundamentals
July 8	**Object-oriented programming** Objects, classes, and review weeks 1 & 2 *Discussion: Genetic Geography of the Brain* **Introduction to Final Projects**	~~*Before class*: Complete DataQuest "Object-Oriented Python" tutorial.~~ (Unfortunately this tutorial is no longer free)

Week 3	This week we'll introduce a few packages that are useful for scientific computing and data science, as well as dive into the Allen Brain Cell Types Atlas. We'll also talk about ways we can visualize the data and run statistics.	
July 13	**NumPy & Pandas** *Designing Problems In-Class Assignment* **Introduction to the Cell Types Atlas** Genetic engineering, patch clamp electrophysiology, intrinsic physiology	*Before class*: Complete DataQuest "Introduction to NumPy", "Boolean Indexing with NumPy", & "Introduction to Pandas" tutorials. **A2** *due Monday at 5pm*
July 15	**Visualizing data** Best practices for data visualization & implementation using Matplotlib ~~**Brain Observatory Dataset**~~ ~~Two-photon calcium imaging & the visual system~~	*Before class*: Complete DataQuest "Line Plots" tutorial.
Week 4	We'll work in another Allen database, the Brain Observatory, to see how different cell types perceive different aspects of the visual world.	
July 20	**Statistics** An introduction to simple statistics in Python **Brain Observatory Dataset** Two-photon calcium imaging & the visual system	*Due Monday at 5pm*: **Final Project Proposal**
July 22	**Clustering & dimensionality reduction** PCA, spike sorting, & other ways in which dimensionality reduction is used in neuroscience. *Discussion: Final Project Tasks* ~~**Signal processing & brainwaves**~~ ~~EEG, ECog Data, Fourier transforms~~	**A3** *Due Wednesday at 5pm*

Week 5	**Parameterizing heterogeneous datasets.** Modern neuroscience incorporates various types of data, both physiological and behavioral. This week, we'll address how we integrate diverse types of physiological & behavioral data to address an experimental question.	
July 27	**Single-cell correlations & correlations between brain regions** *Correlations In-Class Assignment*	**A4** *Due ~~Monday~~ Wednesday at 5pm*
July 29	**Behavioral data in the Visual Coding Neuropixels dataset**	
July 31	**Project Showcase (During "Final")** Share your final project with your classmates	**Final Project due Friday at 11:30 am**

Take Home Final due on August 1st @ 5 pm