BIPN 162 / BGGN 240

Neural Data Science

SPRING 2024

T/Th 12:30 - 2 pm
TATA 3201

Course GitHub:
http://github.com/BIPN162/BIPN162_SP24

	Instructor	**Instructional Assistant**
	Ashley Juavinett, PhD	Jeffrey Liu
	ajuavine@ucsd.edu	jcliu@ucsd.edu
Office hours:	Mondays @ 3 pm	Wednesdays @ 2 pm

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 (SP24)*: BILD62 (or other Python course), BIPN140 & MATH11 (or comparable courses).

Students will be able to:

- Develop hypotheses specific to big data environments in neuroscience
- Design a neural data science experiment and excavate data from open sources
- Integrate data from multiple datasets to answer a biological question
- Describe the fundamentals of statistical machine learning tools used in neuroscience
- Dissect data analysis sections of computational / data-heavy neuroscience papers
- Interpret results from common methods in neural data science

Class Schedule & Attendance Policies

Although we have a scheduled discussion section, we will *not* be using it. Instead, I will be assigning you outside material to either read or watch. We will discuss this material in class, and you will be responsible for knowing it for assignments and quizzes. We will be using our time in lecture for active discussion and exploration of course content. Lecture attendance is not strictly mandatory. However, as you soon will see, our lecture sessions will not be unidirectional — these will be active learning sessions where we co-create our learning. These lecture sessions will be podcasted, but you will be expected to make up any in-class activities that you missed.

Grading

- **Quizzes** (3 in-class, see syllabus for tentative dates; **25%**)
- **In-class work (10%)**: We will complete several activities in class for credit. If you miss one, you'll need to contact our IA to make it up.
- **Assignments (25%)**: Weekly take-home coding assignments or Canvas quizzes will support your progression through the course topics and will directly relate to the larger class projects. Coding assignments will be submitted through the DataHub (http://datahub.ucsd.edu) and graded automatically using a tool called NBGrader.
- **Projects (40%)** Includes the project proposal, code, and deliverables.
    - Cell Types Project (15%, groups of 2-3): The first project will ask you to investigate specific **cell types** in the brain, combining information across at least **two** different data sets of your choosing (e.g., gene expression, visual responses, connectivity, and/or intrinsic activity.)
    - Final Project (25%, groups of 2-3): There are two options for this project.

      Option 1: Integrate **three** different datasets to address a question about brain function. For example, 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.

      Option 2: Choose a paper that uses a new data analysis technique that has been published no more than 10 years ago (most should have code and data publicly available). Implement this method using a new data set (either your own or from publicly available sources).

Additional notes about grading:

- We will be using Canvas (http://canvas.ucsd.edu) as well as the DataHub (http://datahub.ucsd.edu) to manage grades and assignments.
- **Late policy**: Assignments 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.
- If you're enrolled in the graduate version of the course (BGGN 240), you'll have an additional small assignment to complete. We'll discuss once the quarter begins.

Course Resources

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

- VanderPlas, **Python Data Science Handbook** (available free online or in print)
- Wallisch, **Neural Data Science**
- Adhikari & DeNero, **The Foundations of Data Science**

If you'd like a more thorough Python review, you're 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 (including AI assistants) 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.

We'll be relying a lot on other people's code as we learn. Here are some guidelines as to how you should use other code in the process of writing your own, as well as how you can talk to your classmates about the code we're working with in class:

Do explain the thought process behind your code.
Do share the general steps you took to solve a problem.
Do describe your code to others, either verbally or in writing.
Do use examples on the internet to inform your code.

Do not screenshot someone else's code.
Do not directly share your code with others, either in text or image format.

Do not directly copy 5+ lines of code from examples on the internet or chatGPT.
Do not share the values of variables that are explicitly asked for in the validation of the question.

We will be discussing how we can best use generative AI tools such as chatGPT to help us learn how to code. You're encouraged to use these tools to help you learn, but *not* to complete assignments. One of our goals in this class is for you to learn how to read and inspect code for data analysis, including code written by an AI assistant. I reserve the right to ask you to explain any code you have written on an assignment or on the final project, and in person quizzes will not permit the use of an AI assistant.

Syllabus

(*very* tentative and subject to change!)

Date	Topic	Assignments
Week 1	What is neural data science? 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.	*Take Home Readings*: 1. Mark Humphries, "A Neural Data Science: How & Why" 2. Paninski & Cunningham 2018
April 2	**Introduction to Neural Data Science** Tools; Jupyter Notebooks	
April 4	**Python Fundamentals** From variables to data structures, and a reminder of object-oriented programming ***Note***: In-Class Python Activity, due ~~Friday~~ **Monday** at 5 pm	*Optional before class*: Complete "Programming in Python," "Variables and Data Types," and "Lists & For Loops" tutorials on DataQuest.
Week 2	Focuses on data science skills in Python. We'll also start working with gene expression data from the Allen Brain Institute.	*Take Home Reading*: Mosher et al. 2020
April 9	**Python for Data Science** Learning with AI assistants; Using scientific Python (NumPy, Pandas); Introduction to gene expression methods	*Optional before class*: Complete DataQuest "Introduction to NumPy", "Boolean Indexing with NumPy", & "Introduction to Pandas" tutorials. *Due Wednesday @ 5 pm*: **a0** Computer Setup

April 11	**Introduction to the Cell Types Atlas** Genetic engineering, patch clamp electrophysiology, intrinsic physiology *Discussion of Week 1 & 2 readings*	
Week 3	Exploratory data analysis	*Take Home Reading*: Richiardi et al. 2015
April 18	**Exploratory data visualization** Matplotlib and working with 2D, 3D, and more in Python; best practices for data visualization	*Optional*: Complete DataQuest "Exploratory Data Analysis: Line Charts" tutorial. *Due Wednesday at 5pm*: **a1** Conditionally Expressed
April 16	**Gene Expression in the Brain** Brain organization, RNA sequencing & gene expression; Patch-Seq ***Note***: In-Class Concept Maps, due Monday at 5 pm	
Week 4		
April 23	**Statistical Underpinnings** Describing a distribution, correlations, the utility of SciPy *Note*: In-Class Quiz (Python for Data Science) **Details for Cell Types Project**	*Due Wednesday at 5pm*: **a2** Mouse vs Human Cell Types
April 25	**Vectors & Matrices** Linear algebra in Python, describing a correlation in terms of linear algebra	*Before class*: Review NMA Tutorials 1 & 2
Week 5		
April 30	**Data formats** Working with Neurodata Without Borders physiology data	*Due Wednesday at 5pm*: **a3** Correlations
May 2	**Signal processing & brainwaves** Time series, EEG, ECog Data, Fourier transforms	*Before class*: Review NMA Tutorial 5

Week 6 Linear Models

May 7	**Building models**: Can we predict behavior? Is everything a linear regression? **Details for Final Project**	*Due Wednesday at 5pm*: **Cell Types Project**
May 9	**Implementing GLMs with fMRI data** *(Guest speaker Yixin Yuan, Serences & Aoi labs)*	

Week 7 Dimensionality Reduction

May 14	**A Geometric View of Data** Building an intuition for dimensionality *Note*: In-Class Quiz (Statistics for Data Science)	*Due Wednesday at 5 pm*: Final Project Proposal
May 16	**Clustering & dimensionality reduction** PCA, spike sorting, & other ways in which dimensionality reduction is used in neuroscience.	

Week 8

Parameterizing heterogeneous datasets & leveraging big data. 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.

May 21	Behavioral data in the Visual Coding Neuropixels dataset	*Due Wednesday at 5pm*: **a6** Dimensionality reduction & signal processing
May 23	Hierarchical models	

Week 9 Leveraging big data

May 28	A return to time series with hidden Markov models	
May 30	Causal inference and real world data	

Week 10	**Looking forward & final projects**. In this final week, we'll talk about additional neuroscience-related applications for coding and data science, and you'll share your final projects with the class.	
June 4	Final project check-ins *Note*: In-Class Quiz (Advanced Models)	*Due Wednesday at 5pm*: **a7**
June 6	Next steps in neural data science & computational approaches to big data	

Final Project Round Table: