BIPN 162 / BGGN 240 | Neural Data Science

WINTER 2020

Instructor
Ashley Juavinett, PhD
ajuavine@ucsd.edu

Office hours
York 4070C, Mondays 1-3 pm

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

Instructional Assistant
Zarek Siegel
zssiegel@ucsd.edu

Class Schedule
Lectures & Discussion in TATA 2501
Lectures: T/Th, 2-3:20 pm
Discussion: Friday, 11-11:50 am

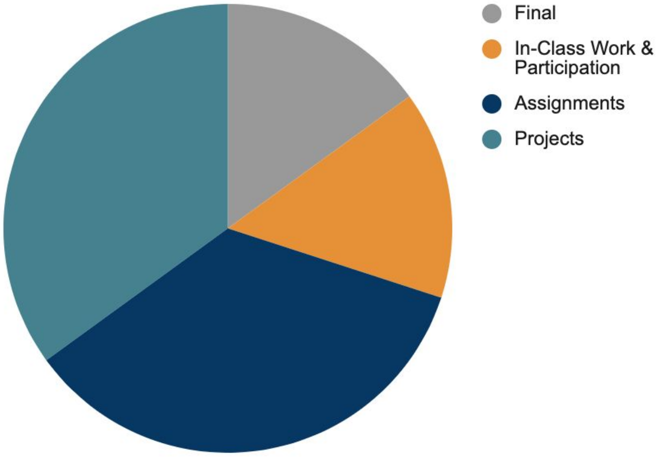
Course Description: Project-based course in which students will use computational notebooks to perform exploratory data analyses and to test hypotheses in large neuroscience datasets, including the differences between unique neuron types, leveraging text mining of the neuroscience literature, and human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

Students will be able to:

- 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 & participation (15%)**:
    - ○ Attendance (10%): If you attend >75% of coding/ discussion sections, you'll receive all 10 points. Alternatively, you may choose to allocate these participation points to the final exam. To do so, inform me before Week 10.



  - In-class work (5%): 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 (35%)**: 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.
  - All assignments are due **Monday at 5 pm** and are worth 2.5-7.5% each.
  - These assignments should be completed individually and should take you about 1-2 hours. Discussion sections will greatly help you complete these assignments.

- **Projects (35%)** 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 type(s)** 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 (20%, groups of 2-3): Your final project should 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. Alternatively, you can significantly expand on the cell types project. We will discuss possibilities for your project as we move through the course.

- **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. We will discuss the contents & format of the final as it approaches.

Additional notes about grading:

- We will be using Canvas (http://canvas.ucsd.edu) 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.

- 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, **Whirlwind Tour of Python**
- VanderPlas, **Python Data Science Handbook** (available free online or in print)
- 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

(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.	
Jan 7	**Introduction to Data Science** Tools; Jupyter Notebooks	Read Mark Humphries, *The Spike*, "A Neural Data Science: How & Why" [link]
Jan 9	**Programming Fundamentals I** Python syntax, expressions, & variables	*Before Class*: Listen to lecture & Complete "Programming in Python" and "Variables and Data Types" on DataQuest. Note: *No discussion section this week.*
Week 2	Focuses on fundamental coding skills in Python.	
Jan 14	**Programming Fundamentals II** Data structures: lists, tuples, and dictionaries	*Due Monday*: **A0** Computer Setup & the Entry Survey
Jan 16	**Programming Fundamentals III** Booleans, conditionals, and loops	*Before class*: Complete DataQuest "Lists & For Loops" tutorial.
Week 3	This week we'll start working with gene expression data from Allen's RNAseq database. We'll extend what you've learned in weeks 1 & 2 to begin working with objects and NumPy arrays.	
Jan 21	**Introduction to the Allen RNAseq database** Brain organization, RNA sequencing & gene expression	*Due Monday at 5pm*: **A1** Programming Fundamentals I & II
Jan 23	**Object-oriented programming** Objects, classes, and review weeks 1 & 2	*Before class*: Complete DataQuest "Object-Oriented Python" tutorial.
Week 4	This week we'll introduce a few packages that are useful for scientific computing and data science, as well as cover the science behind the Allen Cell Types Atlas.	
Jan 28	**NumPy & Pandas**	*Due Monday at 5pm*: **A2** Conditionally Expressed *Before class*: Complete DataQuest "Introduction to NumPy", "Boolean

		Indexing with NumPy", & "Introduction to Pandas" tutorials.
Jan 30	**Introduction to the Cell Types Atlas** Genetic engineering, patch clamp electrophysiology, intrinsic physiology	
Week 5	This week, we'll dive into the Allen Brain Cell Types Atlas and talk about ways we can visualize the data and run statistics.	
Feb 4	**Visualizing data** Best practices for data visualization & implementation using Matplotlib	*Due Monday at 5pm*:: **A3** LISC & OOP *Before class*: Complete DataQuest "Exploratory Data Analysis: Line Charts" tutorial.
Feb 6	**Statistics** An introduction to simple statistics in Python	
Week 6	We'll work in another Allen database, the Brain Observatory, to see how different cell types perceive different aspects of the visual world.	
Feb 11	**Brain Observatory Dataset** Two-photon calcium imaging & the visual system **Details for Cell Types Project**	*Due Monday at 5pm*: **A4** Mouse vs Human Cell Types
Feb 13	**Image processing in Python**	
Week 7	**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.	
Feb 18	**Functional neuroimaging in humans & the Neurosynth database**	*Due Monday at 5pm*: **Cell Types Project Proposal**
Feb 20	**Case study:** Richard Gao's analysis of marmoset data across data types	
Week 8	During the next two weeks, we'll talk about uses of programming in neuroscience that aren't directly related to data science, but can be helpful for processing your data.	
Feb 25	**Clustering & dimensionality reduction**	*Due Monday at 5pm*: **A5** Neurosynth

	PCA, spike sorting, & other ways in which dimensionality reduction is used in neuroscience.	
Feb 27	**Signal processing & brainwaves** EEG, ECog Data, Fourier transforms **Details for Cell Types Project**	*Due Wednesday at 5 pm:* **Cell Types Project**

Week 9

Additional Python applications

Mar 3	**Single-cell correlations & correlations between brain regions**	*Due Monday at 5pm*: **A6** Dimensionality reduction & signal processing
Mar 5	**Behavioral data in the Visual Coding Neuropixels dataset**	

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.

Mar 10	**Next steps in neural data science & computational approaches to big data**	*Due Monday at 5pm*: **A7** Brain correlations
Mar 12	**Final Project Round Table** Share your final project with your classmates and visitors	*Due Wednesday at 5 pm:* **Final Projects**

Final Exam: March 19th @ 3 pm