COGS 138 | Neural Data Science

Meeting: Winter 2022, TuTh 3:30p-4:50p (all times Pacific)
Instructor: Professor Bradley Voytek
Teaching Assistants (TAs): Eena Kosik
Location: RWAC 0103

Course GitHub: https://github.com/NeuralDataScience
Course Slack: https://neuraldatascience.slack.com

One-on-one Q&As:

Date & Time	Location	Instructional Staff	email
†Mondays 2:00p-3:00p	†CSB 169	Prof.: Bradley Voytek	bvoytek@ucsd.edu
Wednesdays 2:00p-3:00p	https://ucsd.zoom.us/j/99833389799	TA: Eena Kosik	ekosik@ucsd.edu

†And by appointment

Section

Section	Day	Time	Location	Staff
A01 (65228)	W	3:00-3:50p	https://ucsd.zoom.us/j/95827498984 CSB 004 (after week 2)	Eena

---

COURSE OBJECTIVES

Learn how to:

- think from a "data first" perspective: what data *would* you need to answer your scientific questions of interest?
- develop hypotheses specific to big data environments in neuroscience.
- work with many different neuroscience data types that might include data on behavior, brain structure and connectivity, single-unit spiking, field potential, gene expression, and even text-mining of the peer-reviewed neuroscientific literature.
- read and analyze data stored in standard formats (*e.g.*, Neurodata Without Borders and Brain Imaging Data Structure).
- integrate multiple heterogeneous datasets in scientifically meaningful ways.
- choose statistical model(s) informed by the underlying data.
- design a big data experiment and integrate data from multiple open data sources.
- consider alternative hypotheses and assess for spurious correlations and results.

COURSE MATERIALS

There is no official textbook for this course. Instead we'll rely on free online resources:

- **Whirlwind Tour of Python**: https://github.com/jakevdp/WhirlwindTourOfPython/
- **Python Data Science Handbook**: https://jakevdp.github.io/PythonDataScienceHandbook/
  - This book is available free online or in print.

GRADING & ATTENDANCE

Grading:

	% of Total Grade
(5) Assignments	50
(4) Readings	20
(1) Final Project	30

Letter Grade	From	To
A+	97.00	100.00
A	93.00	96.99
A-	90.00	92.99
B+	87.00	89.99
B	83.00	86.99
B-	80.00	82.99
C+	77.00	79.99
C	73.00	76.99
C-	70.00	72.99
D+	67.00	69.99
D	63.00	66.99
D-	60.00	62.99
F	0.00	59.99

Notes:

- Final exam date: No final exam, only final project deadline.

PANDEMIC & FIRST IN-PERSON OFFERING CAVEAT

All of this is subject to change if we find it's not working. This is the first-ever in-person offering of this class, and so it's a bit of an experiment in progress. Especially because this kind of course doesn't exist anywhere else! So please bear with me and be willing to speak up and give feedback.

Lecture Attendance

Our goal is to make lecture and discussion section worth your while to attend. However, I will not be taking attendance. But this is a project-based course, where the final project will be a comprehensive notebook demonstrating the work you and your group do. The more time you spend in class and discussions working on this, the better your project will be. The students who took the first pandemic-virtual iteration of this course last year have gone on to great things!

Grades

Grades for assignments will be released on Canvas approximately a week after the submission date. *It is your responsibility to ensure your assignments are submitted on time and to check your grades and get in touch if any are missing or if you think there is a problem.*

Assignments

There are five assignments. Assignments will focus on applying the concepts covered in lecture and ensuring you're on the right track for your final project.

Assignments will be posted on the Thursdays of the week before they're due, and you have until *Friday at 11:59 PM* of the following week (so a little over 8 days) to complete each assignment. Assignments will be released and submitted on datahub. Assignments are submitted individually. You will receive feedback along with a grade within a week from the due date. *Feedback from A2 should be incorporated into your final project.*

Late assignments earn fractional credit (75% within one week late; no late assignments accepted after one week).

Readings

There will be four weeks where readings are assigned. Just like the Assignments, Readings will be posted on the Thursdays of the week before they're due, and you have until *Friday at 11:59 PM* of the following week (so a little over 8 days) to complete the reading quiz assignment *on Canvas*. You are not timed. **You *must* click submit** to submit your reading quizzes. Your **most recent submission** will be graded—**you only get three attempts** (with unlimited attempts it would be possible to figure out all the answers simply through trial-and-error). If you fail to finish and submit your quiz answers before the deadline, it will not be graded. Your lowest reading quiz score will be dropped.

No late credit will be given if Reading quiz assignments are submitted after the due date.

Assignment Regrades

We will work hard to grade everyone fairly and return assignments quickly. But we know you also work hard and want you to receive the grade you've earned. Occasionally, grading mistakes do happen, and it's important to us to correct them.

If you think there is a mistake in your grade for an assignment, submit a regrade request to Prof. Voytek within 72 hours of receipt of the grade. This request should include evidence of why you think your answer was correct (*i.e.*, a specific reference to something said in lecture) and should point to the specific part of the assignment for us to reconsider.

Discussion Sections

Discussion sections will be used to review content from lectures, discuss readings, help with coding and the Final Project, and guide your assignments.

COURSE TOPICS & ASSIGNMENTS

Neuroscience is a rapidly changing field that is increasingly moving towards ever larger and more diverse datasets that are analyzed using increasingly sophisticated computational statistical methods. There is a strong need for neuroscientists who can think deeply about problems that incorporate information from a wide array of domains including psychology and behavior, cognitive science, genomics, pharmacology and chemistry, biophysics, statistics, and AI/ML. With its focus on combining many large, multidimensional, heterogeneous datasets to answer questions and solve problems, data science provides a framework for achieving this goal.

Determining what data one needs, and how to effectively combine datasets, is a creative process. For example, a neural data scientist might be tasked with combining:

> 1) *demographic information* and 2) *multiple cognitive and behavioral measures*, from people from whom we might collect;
> 3) *biometric data*, 4) *motion capture data* to understand motor control, and 5) *eye-tracking* to study attention, along with;
> 6) *structural connectomic* and 7) *functional brain imaging* data collected using methods with different spatial and temporal resolution (such as fMRI and EEG), and then place those results into context relative to;
> 8) *average human brain gene expression patterns* and 9) *the existing knowledge embedded within the peer-reviewed neuroscience literature* (>3,000,000 papers).

These types of data are very different: continuous and ordinal, time-series, video and images, directed graphs, spatial, high-dimensional categorical / nominal, and unstructured natural language. What is the appropriate way to aggregate and synthesize these data? What are the benefits and caveats for, say, aggregating spatially versus temporally? Being able to conceptualize how to carry out this integration is necessary before leveraging any technical skills will even be useful.

Course logistics

Neural Data Science will be a lecture and project-based course. It will consist of three hours of lecture (two 1.5 hour sessions per week); one that is designated for lecture, and one designated for in-class project work and discussion. Homework will consist of problem sets designed to support students with their progression through the larger class projects. Students will be evaluated on their assignments, readings, and their large Final Project. In lieu of a final exam, there is just the Final Project deadline.

OTHER GOOD STUFF

Class Conduct

In all interactions in this class, you are expected to be respectful. This includes following the UC San Diego principles of community: https://ucsd.edu/about/principles.html

This class will be a welcoming, inclusive, and harassment-free experience for everyone, regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, ethnicity, religion (or lack thereof), political beliefs/leanings, or technology choices.

At all times, you should be considerate and respectful. Always refrain from demeaning, discriminatory, or harassing behavior and speech. Last of all, take care of each other.

If you have a concern, please speak with Dr. Voytek, your TAs, or IAs. If you are uncomfortable doing so, that's ok! The OPHD (Office for the Prevention of Sexual Harassment and Discrimination) and CARE (confidential advocacy and education office for sexual violence and gender-based violence) are wonderful resources on campus.

OPHD: https://blink.ucsd.edu/HR/policies/sexual/OPHD.html
CARE: https://care.ucsd.edu/

Academic Integrity

Don't cheat.

You are encouraged to (and at times will have to) work together and help one another. However, you are personally responsible for the work you submit. For group work, it's your responsibility to ensure that you understand everything your group has submitted, and to make sure that the correct file has been uploaded, that the uploaded file is uncorrupted, and that it renders correctly. Projects may include ideas and code from other sources—but these other sources must be documented with clear attribution. Please review academic integrity policies here: https://academicintegrity.ucsd.edu/

We anticipate you all doing well in this course; however, if you are feeling lost or overwhelmed, that's ok! Should that occur, we recommend: (1) attending discussions and leveraging the time there, (2) attending weekly one-on-one hours with Dr. Voytek and the course TAs and IAs and/or, (3) browsing Slack.

Cheating and plagiarism have been and will be strongly penalized. If, for whatever reason, something else prohibits you from being able to turn in an assignment on time, immediately contact me by emailing your assignment (bvoytek@ucsd.edu), or else it will be graded as late.

Disability Access

Students requesting accommodations due to a disability must provide a current Authorization for Accommodation (AFA) letter. These letters are issued by the Office for Students with Disabilities (OSD), which is located in *University Center 202* behind Center Hall. Please make arrangements to contact Dr. Voytek privately to arrange accommodations.

Contacting the OSD can help you further:
858.534.4382 (phone)
osd@ucsd.edu (email)
http://disabilities.ucsd.edu

How to Get Your Question(s) Answered and/or Provide Feedback

It's *great* that we have so many ways to communicate, but it can get tricky to figure out who to contact or where your question belongs or when to expect a response. These guidelines are to help you get your question answered as quickly as possible *and* to ensure that we're able to get to everyone's questions.

That said, to ensure that we're respecting their time, TAs and IAs have been instructed they're only obligated to answer questions between normal working hours (M-F 9am-5pm). However, I *know* that's not when you may be doing your work. So, please feel free to post whenever is best for you while knowing that if you post late at night or on a weekend, you may not get a response until the next weekday. As such, do your best not to wait until the last minute to ask a question.

Finally…

If you have…

- **Questions about course content**: These are awesome! We want everyone to see them and have their questions answered too… so post these to Slack!
- **Questions about course logistics**: First, check the syllabus. If you can't find the answer, ask a classmate. If still unsure, post on Slack.
- **Questions about a grade**: If for an assignment, submit a regrade request or question as an email or a private message on Slack to your TAs.
- **Something super cool to share related to class**: Feel free to email Dr. Voytek (bvoytek@ucsd.edu), share it on Slack, or come to office hours.
- **Something you want to talk about in-depth**: Come to office hours or schedule a time to meet by email (bvoytek@ucsd.edu).

Schedule **(subject to change!)**

Week 1 | Introduction to Data Science

To set the foundation for this course, we'll introduce tools that are commonly used to analyze large neuroscientific data sets including Python and Jupyter. The first two weeks will serve as a crash course to bring students up to speed on the basics of using these tools to analyze data.

- What is data science, and how does it intersect with neuroscience?
- Asking the right types of questions of your data
- Exploratory data analysis in Jupyter

Week 2 | Neuroanatomy and physiology

- Getting you all caught up on the neuroscience side of neural data science
- **DUE: Assignments 0 & 1**

Week 3 | Tools for Data Science

- NWB and BIDS Data I/O
- GitHub and version control
- **DUE: Reading Quiz 1**

Week 4 | Data cleaning and visualization

- Outlier detection
- Choosing your plot types
- **DUE: Assignment 2**

Week 5 | Statistical data analysis

- What are correlations and what are some of their pitfalls in neuroscience?
- **DUE: Reading Quiz 2**

Week 6 | Time series analysis and signal processing

How do we perform analyses on open-source datasets for electrophysiology (single cell, EEG, and MEG) and imaging (two-photon) data? Students will perform time series analyses on electrophysiology and imaging data, and learn about the various types of signal processing used for different data types.

- Field potential and/or calcium data
- **DUE: Assignment 3**

Week 7 | Spike sorting, PCA

- Introduction to Neuropixels dataset
- **DUE: Reading Quiz 3**

Week 8 | Parameterizing heterogeneous datasets

Modern neuroscience incorporates various types of data, both physiological and behavioral. This portion of this course will address how we integrate diverse types of data (e.g., analog signals, video, trial pulses, text data, etc.) to address an experimental question.

- Handling of multiple data types in a Neuropixels dataset (running data, pupil data, electrophysiology data)
- **DUE: Assignment 4**

Week 9 | Modeling single trial data
With many simultaneously recorded neurons, we can avoid signal averaging and actually observe patterns in signal trial data. Here, we'll explore different approaches to investigating single trials of population data.

- **DUE: Reading Quiz 4**

Week 10 | Final projects
In the last two weeks of the course, students will work on final projects in which they use any of the open source datasets that we have encountered to address pressing questions in neuroscience, such as the network properties of the connectome, the role of different cell types in the cortex, or the representation of language?

- **DUE: Assignment 5**