## UC San Diego: Neural Data Science

## Final Project

Your final project should integrate two (or more!) 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.

#### The broad goals of this project are to:

* Identify questions that are addressable by publicly available datasets
* Integrate datasets to address a scientific question
* Implement the technical skills you’ve learned
* Work effectively with a team

Non-exhaustive list of possible datasets/tools:

* LISC (<https://lisc-tools.github.io/lisc/auto_tutorials/index.html>)
* Neuroelectro (<https://neuroelectro.org/api/docs/>)
* Allen RNAseq (<https://celltypes.brain-map.org/rnaseq/search>)
* Allen Microarray (<https://human.brain-map.org/>)
* Allen Cell Types (<https://celltypes.brain-map.org/data>)
* Allen Connectivity (<https://connectivity.brain-map.org/>)
* Allen Brain Observatory (<https://observatory.brain-map.org/visualcoding>)
* Open Datasets in Electrophysiology (<https://github.com/openlists/ElectrophysiologyData>)

### Proposal

In approximately one page, describe your main question and approach that you will take. Your proposal should have the following elements:

#### I. Group member names, and your group name (optional)

#### II. Experimental question

One sentence that describes the question you’ll address in your proposal.

#### III. Background

In 2-3 paragraphs, describe the motivation behind your question. What’s the big picture, and why is it interesting? Are there published papers addressing aspects of your question? You should cite at least **three primary references** in your background. You are welcome to replicate published papers using publicly available data.

##### IV. Approach

Which datasets will you use, and how will you interact with these datasets? Do you need to compute any metrics from these data? In addition, decide on two different ways that you will compare your, *e.g.*, cell types or brain regions. In other words, what plots will you make with your data? In one paragraph, describe what you’re hoping to ultimately compare.

You may organize your “Approach” section as paragraphs (1-2), or as a list. Thinking of this part sequentially will ultimately help you design your analysis and convert that into items in the GitHub Project Kanban board. For example, after pulling in the data from the SDK, then what will you do?

### Project

Your final product for this project will be a Jupyter Notebook that walks us through the steps of your analysis to answer your experimental question about the difference between two different cell types or brains or whatever you choose to focus on. Ultimately, you should generate three different plots that concisely address your experimental question.

You should integrate the feedback given on your proposal into your product, and reach out to Dr. Voytek or our TA if you have any questions or concerns.

In addition to parts that wrangle, analyze, and visualize the data, your notebook should have a **background section** (similar to the one in your proposal, but updated as your project evolves) and a **short discussion section** that summarizes your findings.

#### Project Checklist

1. **Introduction & Background (20%)**
   * Overview:
     + Write a clear summary of what you did
     + Limit overview to 3-4 sentences
   * Research Question:
     + Include your *specific* data science question
     + Make sure what you’re measuring (variables) to answer your question is clear
   * Background & Prior Work
     + Include explanation of what work has been done previously
     + Include citations or links to previous work
     + Include an explanation of dataset(s) used (*i.e.*, how the data was collected, features/variables included, number of observations, information in dataset)
   * Hypothesis
     + Include your team’s hypothesis
     + Ensure that this hypothesis is clear to readers
     + Explain why you think this will be the outcome (what was your thinking?)
2. **Data Analysis (60%)**
   * Data Wrangling
     + Pull your data into Python and explain steps taken to isolate the data you need
   * Data Visualization
     + Include at least three visualizations
     + Clearly label all axes on plots
     + Type of all plots appropriate given data displayed
     + Interpretation of each visualization included in text
   * Data Analysis & Results
     + Appropriate analysis performed
     + Output of analysis interpreted and interpretation included in notebook
3. **Conclusion & Discussion (20%)**
   * Discussion of your results and how they address your experimental question
   * Limitations of analysis discussed
   * What additional experiments would be interesting, and what data would you need?

#### Submission

This single notebook should include all the code you used for all components of the project (wrangling, visualization, analysis). Before submitting, edit all text for clarity, and ensure that code is clearly commented. Your final notebook will be submitted on datahub.