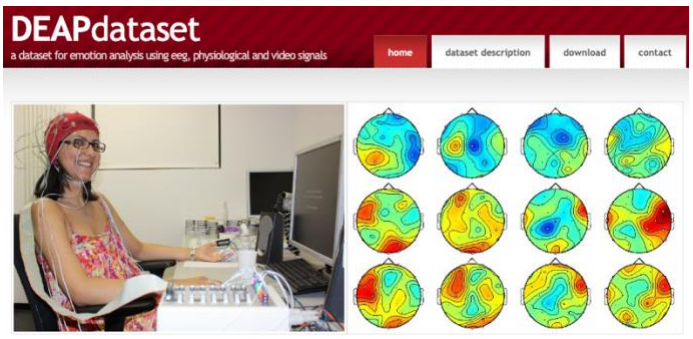
UVM Spring 2024 Brain-Computer Interfaces
Project 3: **Public Dataset Wrangling**

Overview:



We have spent this course loading, visualizing, and analyzing publicly available brain imaging datasets. In this final project, you will choose your own dataset and analysis based on your team's interests.

In this project, you are assuming the role of a BCI developer who is deciding how their device should work. You will use existing data of your choice to answer a question of your choice about how that device should work. You will then present your work in writing and submit your analysis code.

Your submission will be in three parts:

1. Propose an analysis *(due early, see syllabus for dates):*
    a. **Find a dataset** including publicly available brain activity data
    b. **Create a README** to clarify how to load the data
    c. **Design an analysis** to characterize or optimize a brain signal that could be of interest to a brain-computer interface (BCI) developer
2. **Carry out that analysis** using Python code *(draft due before final class, final due after, see syllabus for dates):*
    a. **Load** the data
    b. **Plot** the brain signal of interest
    c. **Characterize or optimize** the BCI control signal
3. Describe a BCI Device (*due after final class, see syllabus for dates):*
    a. **Describe your results in writing** such that another scientist could replicate it
    b. Based on your results, **describe a complete BCI protocol & application** using this brain signal
    c. Outline the remaining **technical and ethical challenges** your group would need to address to bring it to market.

Details of each submission are below.

Submission 1: Propose an Analysis – due early

Many deidentified brain-imaging datasets are freely available online. Here are a few extensive lists:

- Frontiers Datasets for BCI Applications: 2021 journal issue detailing various BCI datasets
- Meagmohit's EEG-Datasets repository: a self-proclaimed "list of all public EEG-datasets."
- BNCI Horizon 2020 datasets: European Commission's project to foster BCI collaboration.

- BCI Competition datasets: Annual competition to see who can classify BCI data best

I strongly recommend EEG or ECoG – they will look like the data you've seen so far in class. Your dataset does not have to come from these lists. If you find one that interests you but you're unable to access it, let me know early and we might be able to email the authors.

Identify a dataset that looks interesting to you and write code to load the data from one or more subjects/runs into Python. Many datasets are in MATLAB (.mat) format, so you can make use of this class' loadmat function or search for help online on how to load the dataset.

Some datasets have comprehensive documentation, but others are incomplete or inaccurate. **Create your own comprehensive README** describing the data and its format. This document should contain all the information someone needs to start working with the data, just like the READMEs you've used in class (e.g., EEG data size and format, event times and types).

Finding a workable dataset can be challenging. You may want to narrow your focus to datasets that do not require preprocessing beyond the filtering discussed in this course. Note that some datasets are beyond help – it's just too difficult to extract the data or understand how it's organized. Don't be afraid to abandon a poor dataset and find a better one.

Once you've found a workable dataset, you will **write a proposal for a meaningful analysis** you can perform on it. Please read the rest of this document to understand what is expected in the final results. Your formal 1-page proposal should address the prompt below, borrowed from the National Science Foundation Graduate Research Fellowship Program.

> *Describe the research idea, your general approach, as well as any unique resources that may be needed for accomplishing the research goal. Address the potential of the research to advance knowledge and understanding within science as well as the potential for broader impacts on society.*

Please also include a brief statement outlining the **anticipated contributions of each group member**, keeping in mind their interests and abilities.

Submit the README and writeup. This will serve mostly as a chance to seek advice and course corrections from the instructor & TA.

Submission 2: Analysis Code – draft due before final class, final due after

Load the data using the information detailed in your README. Then **produce a plot that shows the signal of interest** (one that the user could modulate to control a BCI), including error bars or another measure of variability/confidence. This usually involves epoching the data around task events and contrasting two or more types of trials. Depending on the type of brain signal, this plot could be an ERP, mean FFT, bar plot of band power, or scalp maps. Be sure that your plot shows the presence of the BCI control signal as clearly as possible.

Finally, **perform an analysis** that includes skills we have developed in class. The goal should be justifying this signal as a viable BCI control signal, quantifying its performance in a figure of merit, and/or optimizing choices like epoch times. This can be calculating statistics, producing predictions with a heuristic, training a machine learning classifier, or anything else that would be meaningful and challenging to you and your group.

In your final coding product, please create a **module** with *at least one function for each part of the assignment*, then a **test script** that calls each of these functions in turn to complete the overall task. You are strongly encouraged to make your functions flexible and customizable so they could be used for other purposes too. If you need to modify a function from your old code, be sure to document the change in your writeup.

At each step, **use good coding practices,** including informative and frequent comments for collaborators, well documented functions, print commands to tell users what is happening, good variable naming, properly labeled axes and plots, and readable code. Your objective here is to make it possible for a collaborator (like your TA & instructor) to run, read, and modify your code without talking to you about it.

Submission 3: Written Description – due after our final class

Your final writeup will be **due the night before the final exam period** and outlined as described below. This description and rubric borrows heavily from Dr. Ryan McGinnis and his excellent Wearable Sensing course.

i. **Introduction**: in a few sentences, summarize what your BCI would be designed to do, what dataset you chose, and how the analysis and discussion you're about to describe will bring that BCI closer to reality. Here's a rough outline of an effective intro:
    a. Provide some context to orient readers unfamiliar with your type of BCI
    b. State the need for your work (a gap between your goals and existing results)
    c. Indicate what you have done to address this need
    d. Preview the rest of the paper to prepare readers for its structure
ii. **Methods**: summarize the analysis you carried out and the methods you used. This should include sufficient detail for other scientists to reproduce the experiments you describe. For example:
    a. The number of subjects and any information available about them
    b. The way data was collected (e.g., instrumentation employed, protocol used for data collection)
    c. how that data was processed & analyzed
    d. the statistical tests used to check your findings
iii. **Results**: describe your findings in both verbal and graphical form. Display important figures you created and any statistics or figures of merit you produced. This should align with your methods section and be free of interpretation – just the facts. Figures & tables

should be clearly readable, captioned, and referenced in the text so readers know how & when to interpret them.

iv. **Discussion**: interpret your results by placing them back in context. Then describe a BCI that could be developed using this signal. Be sure to include how this design is influenced by the results of your analysis. Specifically, describe:
    a. A potential user of your BCI system and the need this BCI would fill.
    b. Include a figure of the protocol (what the participant would see). Be specific about the BCI protocol and timing of events on the screen.
    c. Provide a description or flow chart of the real-time analyses that the BCI would have to perform to produce its predictions.
    d. Describe the assistive technology or other application this BCI would inform (the actions the BCI would take when it makes its predictions).
    e. Estimate the information transfer rate (ITR) or (if another would be more meaningful) another figure of merit based on your analyses or online research about similar signals. If you are using other people's work to describe any of the above, be sure to use only peer-reviewed journal articles and **cite your sources**.
v. **Conclusions & Challenges**: this should not be a simple summary/recap. In this section:
    a. Interpret your findings more abstractly. Describe how much you have addressed the need described in the introduction.
    b. Describe future challenges, both technical and ethical, that stand in the way of bringing this BCI to market.
vi. **Contributions** of each group member
    a. If these do not match those in your proposal, please explain.

Please explain your reasoning, and **use what we've learned in class** about the brain to inform your explanations whenever possible.

Your **target audience** is a potential business partner who has very little background in BCIs, and your **objective** is to convince them that you are making sound choices in your BCI design. Please keep it brief – **no more than 4 pages**, please.

This project is designed to challenge you to apply what you've learned in class with much less step-by-step guidance than you have in the labs. Please **start early and ask questions!**

Grading:

The parts containing code will be graded as follows (100 points total):

- Does it work as instructed under default conditions? 50 points
- Does it allow and adapt to different choices? 25 points
- Does it follow good coding practices? 25 points

The parts that involve communication will be graded according to the following rubrics:

UVM Spring 2024 Brain-Computer Interfaces

Project 3: **Public Dataset Wrangling**

Proposal: 100 points	**Possible Points**
README	**20**
Identifies workable dataset, shows evidence of loading & examining	10
Complete description of dataset format & elements	10
Communication	**30**
Innovative & impactful	10
Professional design/communication	10
Appropriate citations & references provided	10
Content Assessment	**50**
Research idea is clearly conveyed	10
Approach is clear and achievable	10
Resources list needed for experiment is complete	10
Potential to advance knowledge or society clearly articulated	10
Anticipated contributions of each group member	10

Final Report: 250 points	**Possible Points**
Communication	**40**
Innovation & impact	10
Professional design/communication	10
Appropriate citations & references provided	10
Contributions of each group member	10
Introduction	**20**
Includes context to orient the reader to your BCI	5
Describes the question/problem addressed by the manuscript	10
Provides a preview of the structure of the manuscript	5
Methods	**50**
Information about subjects clearly presented	10
Data collection, including instrumentation, described clearly	10
Data processing described clearly	20
Appropriate statistical tests selected and described correctly	10
Results	**40**
Results align with Introduction and Methods sections	10
Appropriate data is reported in tabular and graphical form	10
Tables and figures have informative captions and axis labels	10
Text used to introduce and provide context for figures and tables	10
Discussion	**50**
Potential BCI user & needs clearly described	10
BCI protocol clearly described	10
BCI predictions & application clearly described	10
Implications of reported results discussed in context of eventual BCI	20
Conclusion & Challenges	**50**
Most important outcome clearly stated.	10
Technical challenges stated clearly & completely	20
Ethical challenges stated clearly & completely	20