Data Mural Project

This project will bring awareness to the current climate for the disabled students' community on campus and hopefully inspire data scientists, researchers, and administrators to create and use datasets that better capture the individual's identity and mental self, which in turn allows for research that better serves the community. It highlights how we should bring the subjects of these surveys, in particular disabled students, to the frontlines of data collection and analysis. In focus groups, students will be asked to engage with data visualizations from university surveys. We would be working with the Student Experience in the Research University (SERU) Consortium survey and the University of California Undergraduate Experience Survey (UCUES). They will determine if they feel accurately and holistically represented and acknowledged in these datasets and survey questions. Students' perspectives on these topics will be portrayed in a community-created, dynamic, and accessible data mural, a mural that visualizes data, which we hope will continue to spark conversations after its creation. This art installation will be housed in the new Disability Cultural Community Center.

One additional step we can take is to analyze students' transcribed responses with natural language processing methods, particularly using the Latent Dirichlet Allocation topic modeling algorithm, Word2vec algorithm, and sentiment analysis using the VADER (Valence Aware Dictionary for Sentiment Reasoning) model. This approach would help us analyze students' thoughts and feelings in response to the UCUES and SERU data visualizations. This step, however, is not needed to create the data mural.

Links:

Original Project Proposal:
https://docs.google.com/document/d/1oUR2KeiC3qEMtuel94_dVEG7EQHFeGzeQYnkLRuBVUY/edit

Final Prospectus for Honors Thesis:
https://docs.google.com/document/d/1sNyzZCsjqjiGbCVuRegstYTqPa-VbUuIH9iRrTprtW8/edit

Presentation for the 2021 CSTMS undergraduate research symposium:
https://docs.google.com/presentation/d/1DcNgFQTKcqsFoHVr1dI2Oeb_0qWjvjBTqDnBmkalS4M/edit#slide=id.ga25f85cae5_0_0

Datasets:

SERU Consortium Report: The Experiences of Undergraduate Students with Physical, Learning, Neurodevelopmental, and Cognitive Disabilities During the Pandemic

UCUES Data: https://www.universityofcalifornia.edu/infocenter/ucues-data-tables-2020

- https://opa.berkeley.edu/campus-surveys/survey-results-reporting-analysis/uc-undergraduate-experience-survey-ucues#:~:text=Administered%20from%20March%20until%20July,aspirations%2C%20experiences%20with%20academic%20and
- https://opa.berkeley.edu/can-items-one-module-be-correlated-items-different-module

Focus Groups:

- Relevant links:
  - https://cphs.berkeley.edu/training.html
  - https://cphs.berkeley.edu/

Original Step by Step Plan

1. Generate and collect visualizations
    a. Use visualizations from the SERU Consortium survey
    b. Generate visualizations from the University of California Undergraduate Experience Survey dataset. Example question from the survey to focus on: I feel valued as an individual at this institution and I feel that I belong at this university (across demographics available)
        i. Create datasets that students can interact with using Plotly
        ii. Create general visualizations using seaborn
2. Create focus groups of students to engage with datasets - in-person/over zoom
    a. Collect qualitative data on their reactions to the datasets
    b. Examples of questions to ask:
        i. Which forms of visualization did they find most impactful?
        ii. Which forms of visualization did they find easy to engage with?
        iii. How do they feel about the data presented
        iv. Can they relate to the responses shown in the data
        v. What do they think is missing from the data presented
        vi. What aspects of the dataset do they feel are most valuable
        vii. Do they feel represented in the data
        viii. What questions should be asked that would better represent you
        ix. How do you want your data presented
        x. How do you see yourself as a participant in the surveys/ dataset?

    EXTRA: Use natural language processing techniques to analyze student's thoughts and feelings, create a summary of their collective experience that doesn't undermine individuality
    c. Topic modeling: return a series of word-probability pairs that quantify how well a word captures the topic of a response
    d. Word2vec: return a list of words for each interview question ordered by which words are predicted to best capture the overall sentiment of the question topic
    e. Sentiment analysis gauges students' approval of the datasets presented in regards to how well they feel represented in the data

3. Use the results of the analysis of the focus group data to create a dynamic data mural
    a. As a group, analyze personal data/ individual responses and group data. Then collectively decide what message to share with representation in data as the overall topic
    b. Group exercises to inspire different ways of representing the story
    c. Ideas that resonate most are combined and transformed into a single design
    d. The dynamic mural is established in public space as a community