Design Rationale and Reflection for Assignment 5

Vicky Yeh
ID:2225544

Overview of your process

The dataset is related to student mental health, collected from university students. And this dataset is achieved from Kaggle. I chose this dataset for this assignment because I'm curious about the relationship between major, age, and mental health. Additionally, I wanted to explore working with a binary dataset rather than a continuous one, as we have practiced in class. For the ethical visualization, I tried to put combined information I would like to know in an interactive visualization, such as age, major, and treatment. Also, I filter some data to make it more clear but not deceptive. As for the deceptive one, I began by observing all the columns in the dataset first. After doing so, I noticed that the distribution of data across different ages is uneven. This uneven distribution could potentially mislead viewers into believing that certain age groups are more prone to anxiety, while it might not be true in reality.

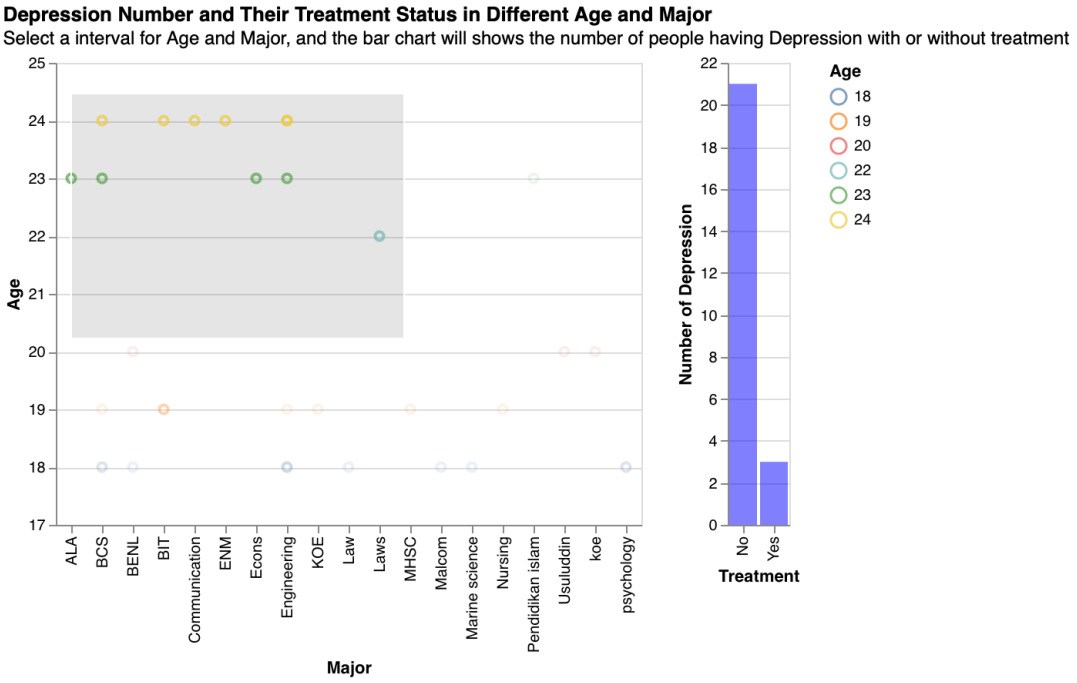
Link to dataset: https://www.kaggle.com/datasets/shariful07/student-mental-health

Ethical description - Visualization 1

This visualization shows the depression number and their treatment status in different range of age and major. In the scatter plot, x-axis is the major, and the y-axis is the age. It has utilized filter to show the student with depression only. The color variation highlights age differences, while the color shade represents the quantity of individuals within each age-major group. When selecting specific interval in the scatter plot, a corresponding bar chart will show the number of students with depression and the number of students with and without treatment. The question is "Is there a difference in people with depression seeking for treatment or not across different age groups or majors?", and I think by the interacting with the interval to select different age and major, viewer can answer this question easily.

Link: https://observablehq.com/d/3d22f89b6b37d039

Screenshot:

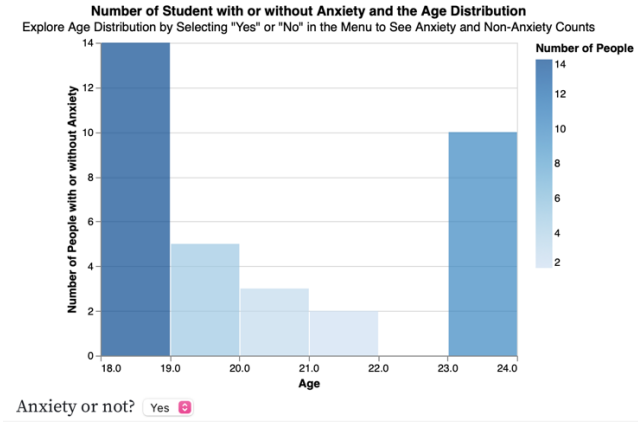


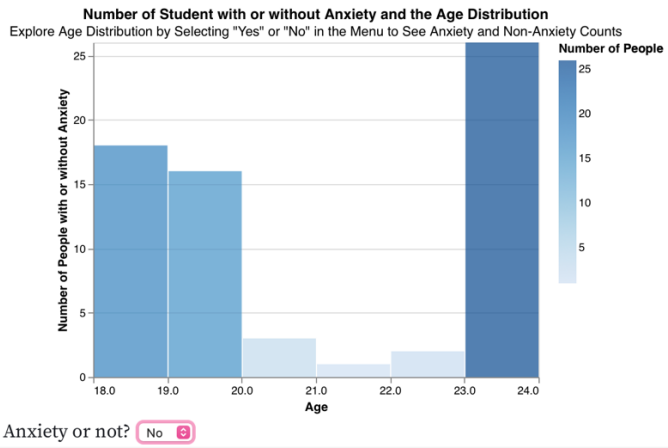
Deceptive description - Visualization 2

This visualization shows the age Distribution of student with or without anxiety. In the bar plot, x-axis is the age, and the y-axis means the number of people with or without anxiety in that age. The color is to highlight the age differences. When selecting “Yes” or “No” in the menu, a corresponding bar chart will show the distribution of students age with anxiety or without anxiety. The question is “What is the age distribution of people with or without Anxiety? “. This visualization will deceive the viewer because the original distribution of age from all collected data is uneven. It may lead viewers to believe that certain age groups have a higher proportion of anxiety. However, they might overlook the total number of students in each age group. Also, I make the two bar charts with different range of y-axis, which might also lead to some incorrect conclusion if viewers don’t observe them carefully.

Link: https://observablehq.com/d/ba6cc31c6c617d90

Screenshot:





Learning reflection

In this assignment, I learn lots of skills in writing observable + Vega-Lite such as many interactive functions like selected interval and menu. I also realize that the hardest part of this assignment is not the coding itself. Instead, it takes me a lot of time to understand the dataset and try to interpret it. For the ethical one, I tried my best to combine all the important information together and convey it by using the interactive tool. Actually, I encountered some technical problem in the process. For example, when using selected interval, I fail to find a way to connect it to three different bar charts simultaneously. So, I only selected the one with the most important information. As for the deceptive visualization, it took me much more effort on how to mislead readers with accurate information, without just using a simple transformation of the scale. The method I ultimately came up with is using the unevenness feature of the data distribution to make readers misinterpreting the results. Although this visualization looks simple, I think it has achieved the goal. Overall, this assignment makes me focus more on the dataset itself. Overall, this assignment taught me the importance of dataset comprehension and thoughtful design in data visualization, rather than just focusing on creating visually appealing charts without meaningful insights.