Design Sprint

INFO 4602

Jack Driscoll

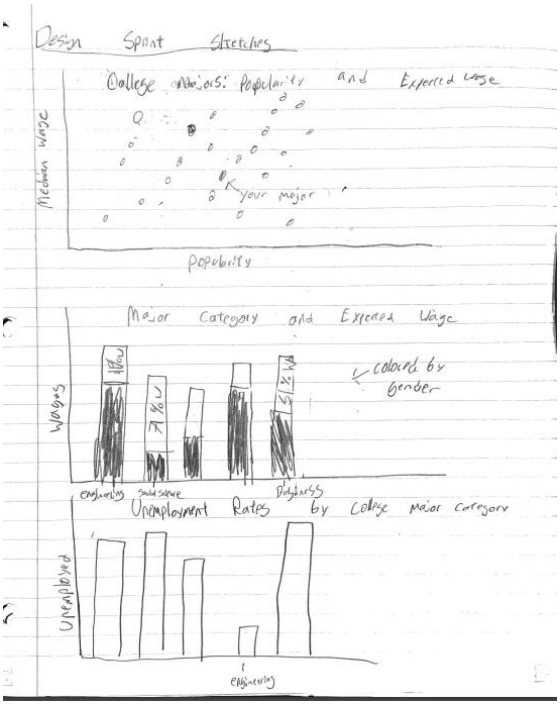
Description of Data

The dataset I chose comes from the American Community Survey (2010-2012). It's from a larger dataset containing information on college major choice as correlated to general frequency, employment status, demographic information, and various wage information (median, 75th percentile, etc…). I am using a specific csv titled "recent grads" that focuses on people who graduated in the last 5 years. The dataset and more information can be found at https://www.kaggle.com/datasets/tunguz/college-majors?resource=download.

Visualization Audience

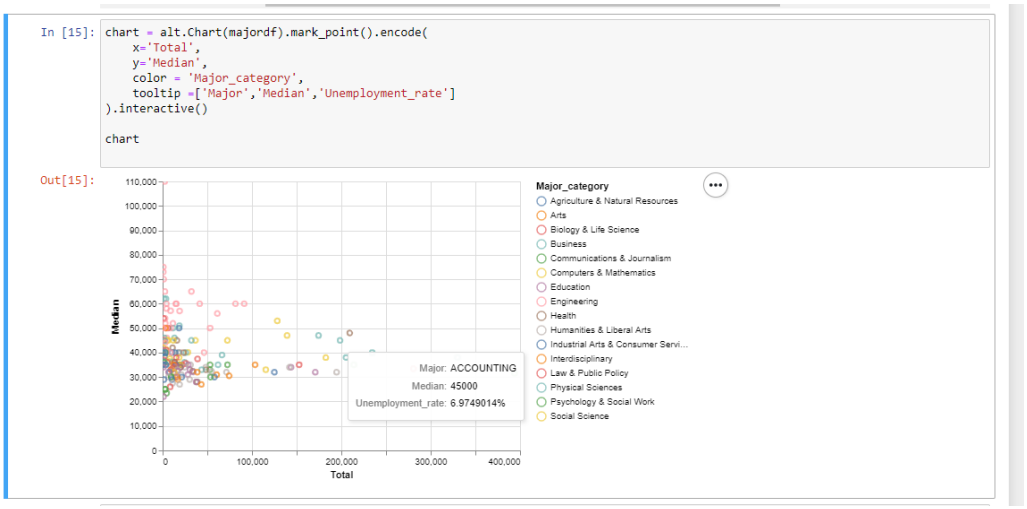
Ideally, I would like to cater this visualization to freshman and sophomore undergraduate students who are considering what major to choose. I would like to create something that can help students determine more about their current/prospective major in that it informs them of general demographic and wage information. Ultimately, it could provide insight into what they could expect to be paid upon graduating with a specific degree.

Sketches



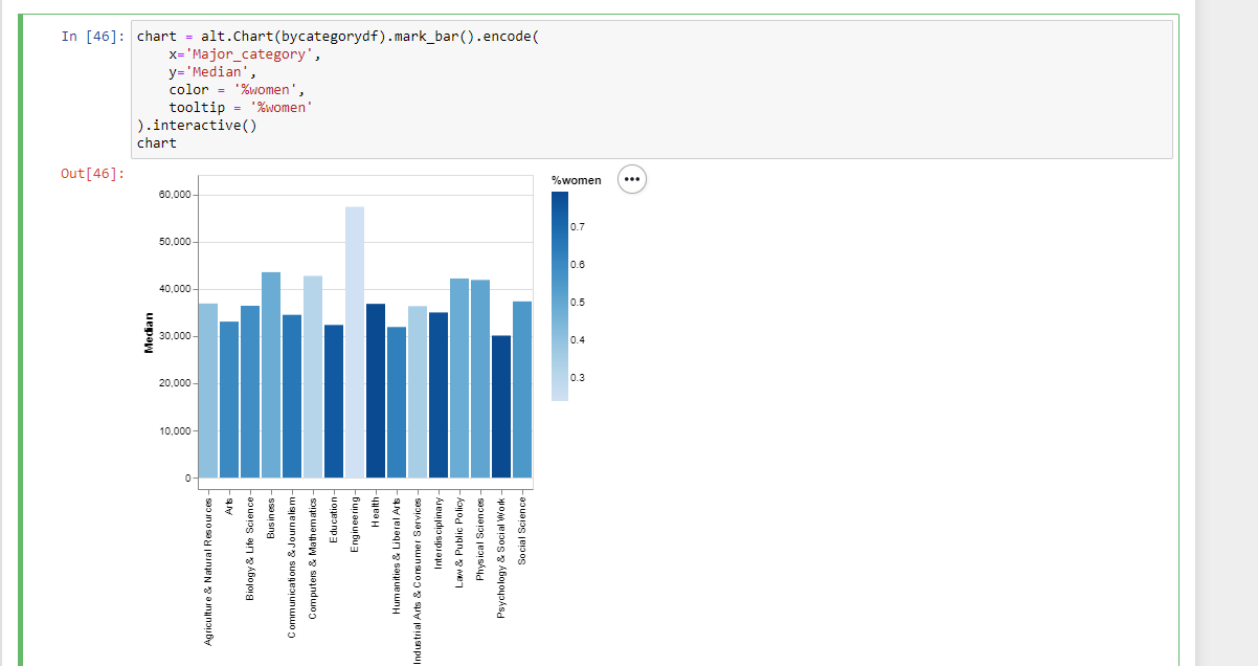
Implementation of Sketches (Altair)

Sketch 1:



The chart above is interactive. You can hover over a point and see what major it represents.

Sketch 2:



I used the groupby function to aggregate the percent of women who make up each major category.

This chart represents the %women that make up a major. The darker the blue, the more women in that field/major.

Reflection

I think the first vis came out really well, and is effective at conveying the spread of majors by popularity compared to median wage. I also think its specifically informative, as you can find your own major pretty easily and get a gauge for where you stand.

In the second vis I think it would be more optimal to have each bar color coded to the percentage of women that make up each major (see the hand drawn sketch 2 for an idea of what I am talking about). I was trying (for about an hour) with altair to encode a scale or a range of values that would allow me to visualize this, but ultimately gave up as I concluded my data wasn't in an optimal format for this. The second implementation was as close as I got.