Doug Kirkman

N328 - D3 Final Project (Wine Reviews)

4.4.2021

<https://www.kaggle.com/zynicide/wine-reviews>

For this project, I wanted to take a look at the dataset and create a visualization that related the price of wines to their actual scores. The initial idea that I had for the visualization was to create a sort of multi-bar chart with the average score for each price range being shown and with the maximum and minimum values for the price range shown. However, I found this to be far too unnecessarily complicated, so I wanted to go for a simpler approach in the end. This led me to instead going to a scraper plot of the two variables instead, as that would also lead to finding out the same information, and might make more sense for the average person to understand.

One design choice that I considered and eventually decided against was to go for a clustering of the data. This would in essence split the data into 3 different colors on the visualization, with each color representing a certain cluster range. However, I decided against this for two different reasons. The first reason is that while this would be possible, I felt after taking a look at the scatterplot that in reality there wasn’t much clustering or spacing out of the data to design clusters around. The second reason is that by using the dots with opacity, it actually has a similar effect on the visualization. For example, in the scatter plot it is clear that the price range for $10-20 is much darker blue than the rest of the graph, meaning that this would be a sort of most popular cluster. Creating opacity in the points overall had a similar and in my opinion easier to understand effect than the regular clustering method.

The overall design choices I made for this project were selecting the method at which to make the scatterplot interactable and what axis ranges should be used. For the intractability, being able to mouse-hover for exact points I think was the most important part of this project, because it adds a lot to the user experience. For the axis, while there are a lot of points that lie at price of zero, overall the only real change was setting the starting wine points at 75, which made the visualization a lot clearer to understand.

One of the questions that I think was answered was that there isn’t really a clear relationship between price and points given. I think that I was really expecting to see some sort of linear relationship between these two variables, but as it turns out there really isn’t much of an increase in points as the price increases. While disappointing, it still answers the question. I still believe that there might be some sort of correlation between the variables that could be found using more in depth data analysis tools, but that is outside the scope of this visualization.