

# CS/INFO 3300 Project 2 Report

Joseph Mo, Andrea Ruggiero, Xiaoxu Guo

## **Data Description**

We used the wine reviews dataset, which was available on Kaggle from the WineEnthusiast. The dataset consisted of a vast list, where each record was an individual wine brand. Information such as title, description, province, and price are accounted for on the list. The columns represent these attributes. For our visualizations, we filtered out only the information we would need from the dataset: the title, province, variety, points and price of each wine. These were all accessible through a .csv file in the Kaggle query called wine\_data.csv.

## **Visual Design Rationale**

We decided to create four graphs showing relationship between different categorizations of wine:

- 1) Our first graph shows the relationship between the average ratings associated with the top five most common varietals in the nine most popular wine producing countries. In this graph, the marks include the different lines representing average ratings of each country listed. The visual channels include the x and y positioning of the lines, the color difference of lines, and a legend that differentiates the data shown by the graph through hovering over specific colors. When implementing this graph, an ordinal scale was used because the data needed to be separated by different colors. Moreover, the colors used in this visualization are predominantly brighter colors to represent wine-like colors.
- 2) Our second graph shows the relationship between average wine ratings per state, as well as total wineries included in those states. The marks included in this visualization consists of the individual states on the map separated by path elements. The visual channels consisted of the legend representing the data based off color hue, and the color hue of the actual map, which varied due to different average wine ratings per state. When implementing this graph, we used an ordinal scale to represent the different color hues in each state.
- 3) Next, we have the visualization showing the relationships between the price and rating in New York State. It is not surprising that higher-priced wines do tend to have higher ratings. Initially, the extremely high-priced wines graphed on a linear scale make it difficult to see the relationship for the majority of the wines reviewed, so We've plotted the price on a logarithmic scale, which shows a much more direct relationship between price and wine rating. The points are color coded based on the varietals. We also add a legend at the bottom depicts top 5 common varietal names in NY and color coding, and hovering over a varietal the scatterplot will filter to show the breakdown of that specific variety.
- 4) Our last visualization represents the mapping of specific ratings to wine bottles as well as the price of those bottles. The marks included in this visualization are images of a wine bottle, and a background image that mimics a shelf holding the bottles. As for visual channels, the different ratings labeled on each wine bottle allows a user to randomly

select a wine bottle associated with that respective rating. Also, it is not explicitly shown at first glance, but the pricing of wine bottles seems to have a relationship in that higher rated bottles tend to be more expensive than lower rated bottles.

### **Interactive Elements**

- 1) In the first graph, the interactive elements incorporated into the visualization consists of mouseovers of each line, representing the average rating points, and a legend at the bottom that separates countries individually so that a user may clearly see which lines belong to which country. To make them discoverable and interesting, a text box is clearly shown whenever a user hovers over a line. Also, the legend clearly differentiates country by setting the width of other lines to zero so that a user only sees the line representing the specific country being hovered over.
- 2) Next, the interactivity incorporated into this data visualization consisted of a hover option, which displayed the average wine rating, the name of the state, and the total number of wineries in that respective state. We thought displaying the total number of wineries and average wine ratings was a good idea because it allows us as well as other users to see if there is any relationship between the number of wineries and the rating number of a state. Moreover, we thought this was also a great opportunity to incorporate two different datasets to see if there is any relationship between the two.
- 3) In the scatterplot, the first interactive elements we have is hovering over a point on the graph will show the wine name, rating, price, and tasting notes. Then, clicking on the circles on the scatterplot will show the data points which have the same varietals and hide the others. If the user clicks that circle again, all the points will be restored. We also add a zoom function in the chart, so that the overlapped circles could be distinguished. The legend below depicts top 5 common varietal names in NY and color coding, and hovering over a varietal the scatterplot will filter to show data points of that specific variety.
- 4) Finally, for the last visualization, the interactive elements consisted of clickable wine bottles that displayed an image of a random wine bottle from that respective rating. At first, the wine bottle ratings were visualized using buttons to click, but we found that actually adding the number within the bottle enhanced the visualization more. Moreover, the design rationale for the background image of a shelf was meant to represent an object that looked as if it were actually holding up the wine bottles. We thought it actually fit in smoothly with the visualization and made it more aesthetically pleasing rather than having a plain white background.

### **The Story**

We ultimately chose to make visualizations based on Wine Reviews Reports because we were curious to see whether there would be a positive correlation between high-priced wines and high ratings, which states and countries are more expensive, and which states and regions tend to give higher scores to the same kind of wine. As a result, all of our graphs incorporated some sort of visualization to check if in fact higher ratings resulted in higher priced wines. However,

after completing our visualizations, there seemed to be some trend involving higher ratings with higher priced wines, but for the most part the two aspects varied. Ultimately, we wanted to convey to our users visualizations that are sufficient enough in data, and aesthetically pleasing to follow along with. Because this was our goal, we figured incorporating the last graph involving the random rated wines was a more fun and enjoyable aspect both for us to implement and users to play around with.

### **Team Contributions**

Joseph Mo

- Provided potential data chart ideas
- Drew concept sketches
- Contacted and attended TA meeting
- Agreed to make chart based on Wine Reviews data
- Completed first two charts
- Attended project demo
- Edited rationale
- ~ 25 hours

Andrea Ruggiero

- Provided potential data chart ideas
- Drew concept sketches
- Contacted and attended TA meeting
- Agreed to make chart based on Wine Reviews data
- Completed map data chart
- Attended project demo
- Edited rationale
- ~25 hours

Xiaoxu Guo

- Provided potential data chart ideas
- Drew concept sketches
- Contacted and attended TA meeting
- Agreed to make chart based on Wine Reviews data
- Completed scatterplot data chart
- Attended project demo
- Edited rationale
- ~25 hours