

# **Wines of the World**

Undergraduate Group 2

Bryson Gundry, Damion Anderson, Heather Wright

CPCS4030, Fall 2020

## **Project Report**

### **Introduction**

Shopping for wine is a daunting task with the various prices and selections available. The goal of our project was to explore the various wineries around the world and identify the wines that are the highest rated for the lowest prices. We began at the global level, identifying which countries produced the best wine. Then, since we are all located in the United States, we narrowed our focus on American wines, creating a visualization which displays the average points per state and then the average prices. Finally, we identified Texas as our primary state of interest. Based on our dataset, we identified McPherson as the best vineyard in terms of points per price. They sell a \$13 bottle of wine with a 88 point rating.

### **Dataset**

Our dataset came from Kaggle.com, a website that provides free datasets for data scientists. We chose a dataset that had over 130,000 wine reviews. The data for the dataset was collected from Winemag.com contains a collection of reviews written by wine experts all over the globe. It is important to note that Winemag.com only collects reviews for wines scoring above 80 on a 100 point scale. There are many variables in this dataset including: the name of the wine, country, province, vineyard, winery, points (out of 100), price (\$), the review, the reviewer's name, and many more. In order to use the dataset, we utilized RStudio to clean the dataset in preparation for the visualization we created in D3.

### **Design solution**

Average Rating of Wines Worldwide

We selected a geometry that demonstrates which countries produce the best wines. The attributes used are the Points [Quantitative] for each wine, and the Countries [Categorical] in the world. The channels used are the colors, which depend on the countries ratings and correspond with the hue (darker represents a higher point average). A dark red represents countries that produce the best wines and yellow represents countries that produce the worst wines, and a mix represents countries that produce quality wine somewhere in the middle. The user is also able to interact with the visualization using a tooltip which displays the average rating for the country. We created a geometry (map), since we wanted to begin with a broader scope and focus on wines from a global perspective. This is an effective approach as it gives the viewer an overview of the story, before delving into the details represented in the later visualizations. From this visualization, we were able to see that wine quality is in fact dependent on geographical location. For example, North American countries produced higher quality wines as opposed to South American countries.

#### Dynamic Average Price and Points per State

For the second visualization, we narrowed our focus to the United States. We made this decision because the US had a high average rating in the previous visualization and we are all located there. We chose to create a dynamic bar chart that illustrates the average points depending on state and then the average price. The user is able to sort the data to further illustrate how a state's location changes depending on the attribute selected. The attributes we chose are the Points [Quantitative], Price [Quantitative], and State [Categorical]. The channels are the length of the bars which depends on the average rating or price, and the horizontal position which depends on the state. In addition, the horizontal position of the state can be sorted in descending order depending on the attribute selected. We selected this dynamic design for our bar chart since it allows us to illustrate which states produce the best wines and which states produce the most expensive wines. This dynamic visualization allows the user to get information quickly as they can choose whether the graph illustrates price or points. They are also able to sort the chart which helps the user further digest the information provided. From this visualization, we discovered that the majority of states have similar locations

when sorting based on price or points. For example, California is ranked 4th for quality and 1st for price.

However, there are some outliers like Nevada which has a lower quality and then a higher price.

### Points vs Price in Texas

For the final visualization, we utilized a scatter plot to demonstrate the points vs price for wineries in Texas. We wanted to focus on Texas since it was in the middle of the pack for both price and points in the last visualization. It also has a good array of wineries available. We chose a scatter plot since there was an appropriate number of datapoints for this type of visualization. It also allows the user to quickly spot trends in the data depending on the location and color of each point. The attributes are the Points [Quantitative], the Price [Quantitative] and the Vineyard [Categorical]. The channels are the colors for the vineyard, the vertical position for the points, and the horizontal positions for the price. The graph utilizes force stimulation to ensure that the points do not overlap. The user is also able to use a tooltip to see the exact price, points, and winery that represent each individual datapoint. This is an effective visualization as the user is able to quickly locate the best quality wine for the lowest price. The wines that satisfy this criteria are located in the upper left corner of the graph. For example, in Texas McPherson winery is the best in terms of points and price. They sell a \$13 bottle of wine with a 88 point rating.