

The Wine Industry: A Visual Exploration

Brian Wright
Dataviz Final Project

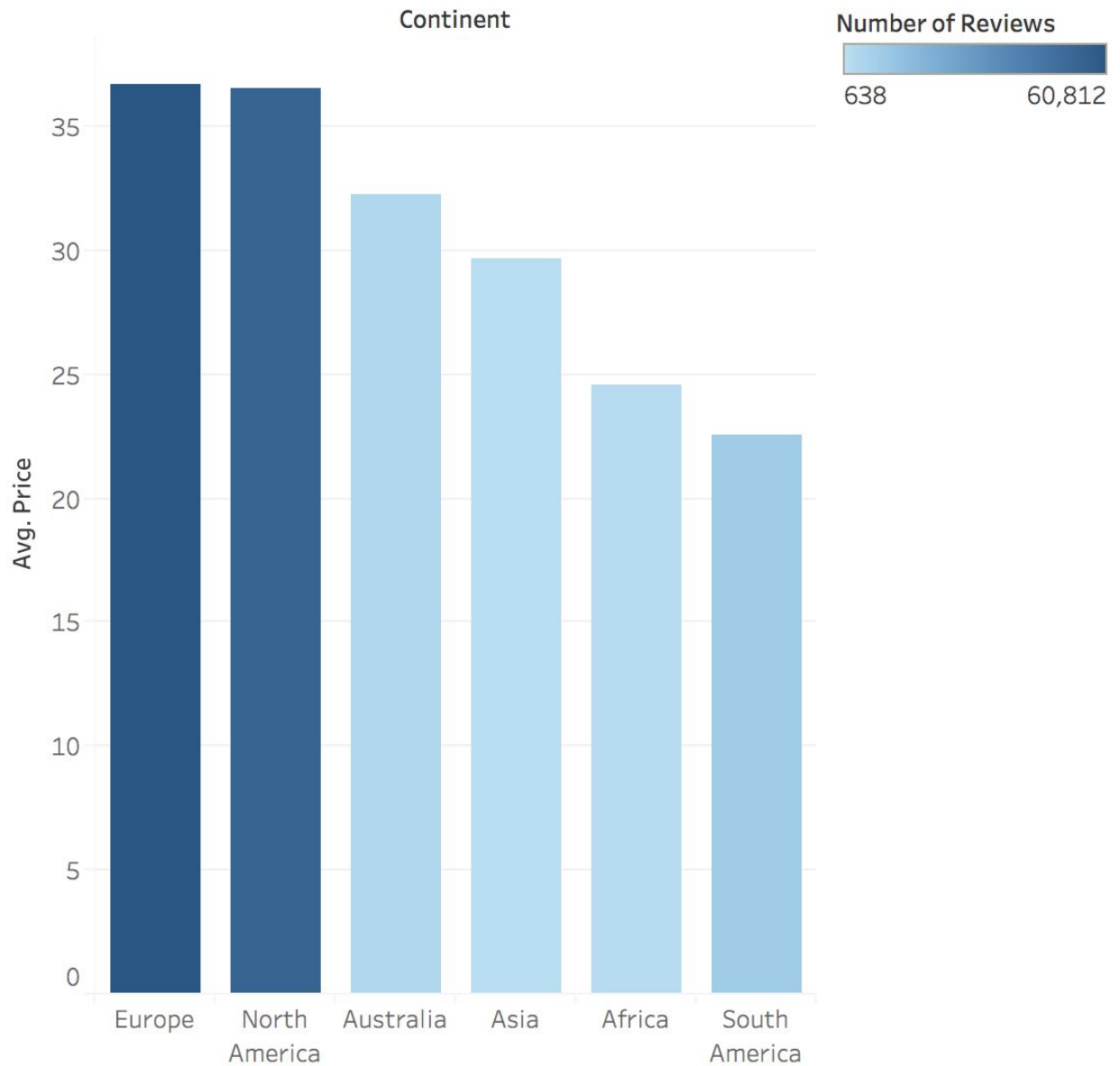
Introduction:

When assigned this final project, the trickiest part was choosing a dataset. With an open door, allowing us to choose any dataset we were interested in, it was tough to decide on one single dataset. I eventually decided to end up going with a dataset from Kaggle related to wine and wine reviews. The initial dataset was intended to be a classification machine learning problem, but the data was suited well for visualization as well. The csv contains 170,000 wine reviews scraped from various wine outlet or forum websites. The dataset included interesting columns related to the geographical location of the wine, the wine review in text, a numerical score, and the reviewer of the wine. So with all of this metadata and a bit of preprocessing and labeling, we had interesting information including sentiment scores, word counts, pricing and scoring, etc. As a wine fan myself, I was excited to explore pricing and scoring trends across various regions and wine varieties. My dataset was missing some information I needed for some graphs, however, including dates and import/export information, for the stacked area and connection plots. I was able to find some supplementary datasets revolving the consumption of alcohol by country, allowing me to fulfil the stacked area requirement. All datasets used will be cited at the end of the paper.

Summary of Data:

1. What is the average price of wine by continent? How about the number of wine reviews by continent?

Average Wine Price by Country

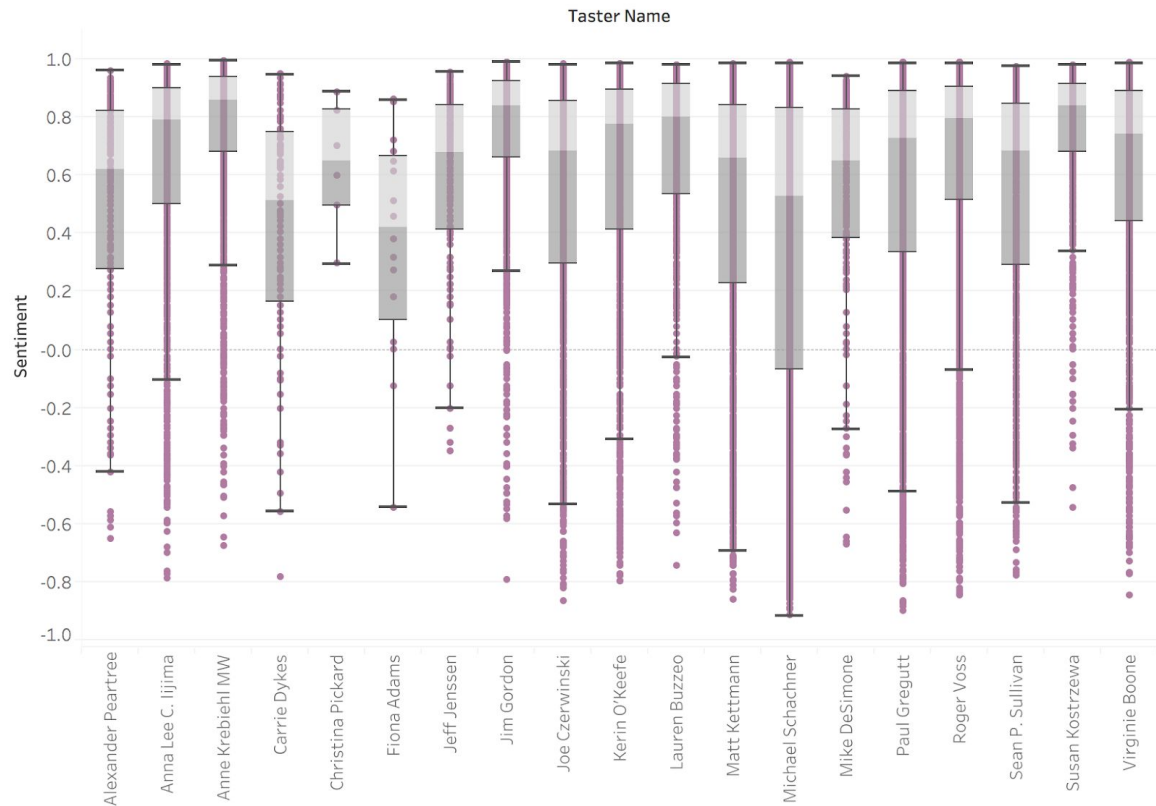


Average of Price for each Continent. Color shows sum of Number of Records. The view is filtered on Continent, which excludes NA.

With the exception of South America, countries with more wine reviews tend to also have more expensive wine, which makes sense.

2. What does the distribution of scores assigned by the reviewers look like?

Distribution of Sentiment by Reviewer

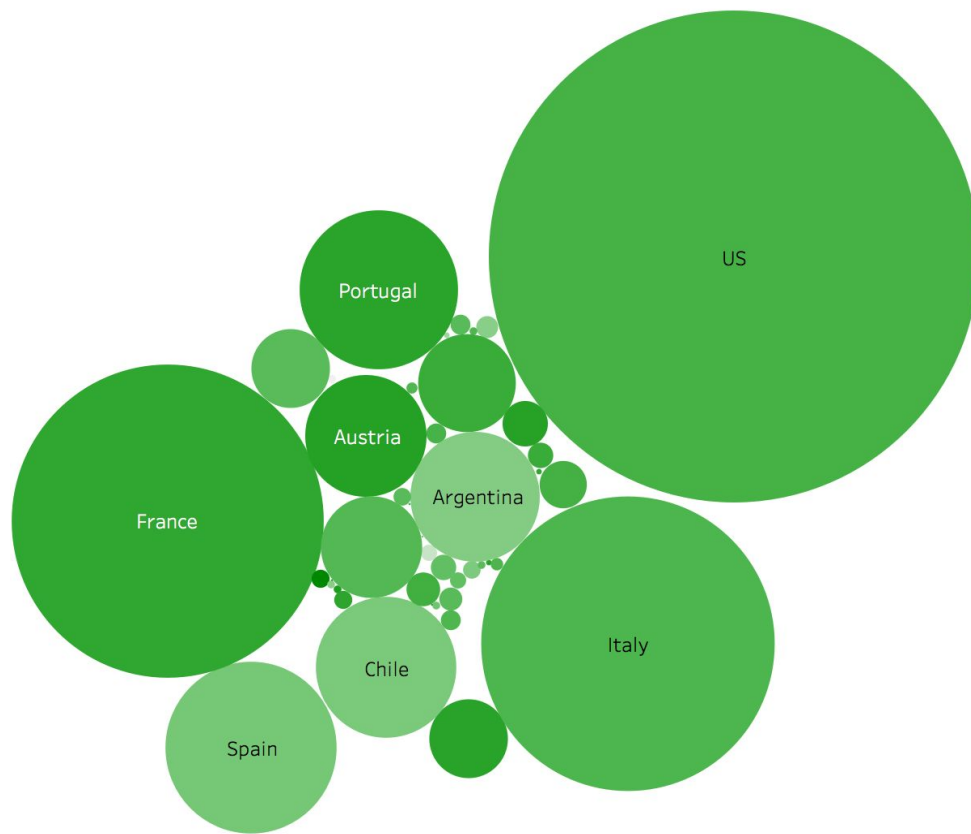


Sentiment for each Taster Name. The view is filtered on Taster Name, which excludes Null.

This box plot is interesting because you can get a sense of how varied the reviews are from person to person. Some reviewers tend to give consistently similar reviews, while some range from strong praise to harsh criticism.

3. How does sentiment vary across countries? While also considering the number of reviews for a given country?

Exploring sentiment and number of reviews by country



Country. Color shows average of Sentiment. Size shows sum of Number of Records. The marks are labeled by Country.

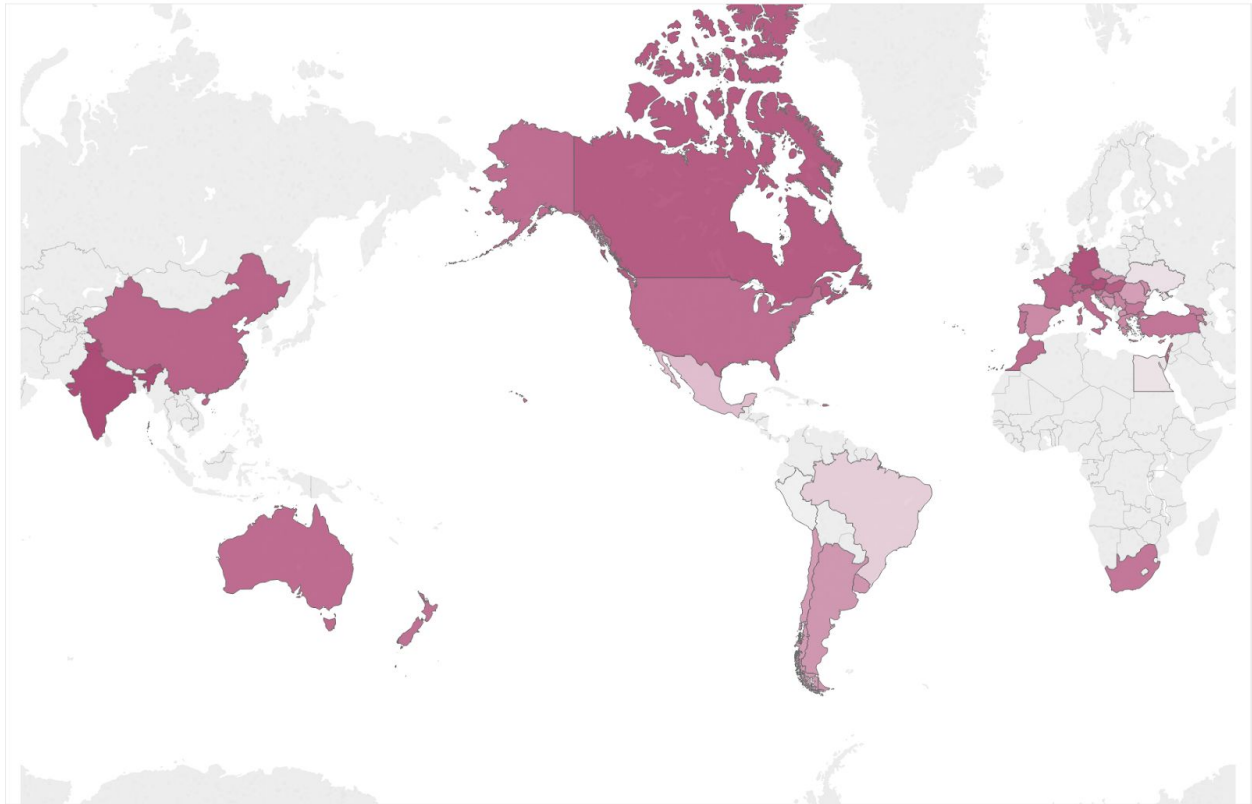
Average Sentiment Sco..



As I would expect, countries with more reviews, where wine would be more popular, tend to have a higher overall sentiment in their reviews. It is intuitive that as reviews are increasing in popularity, good product could be a cause.

4. What about average overall score across all wines in a map view?

Average Score by country



Map based on Longitude (generated) and Latitude (generated). Color shows average of Points. Details are shown for Country.

Average Score

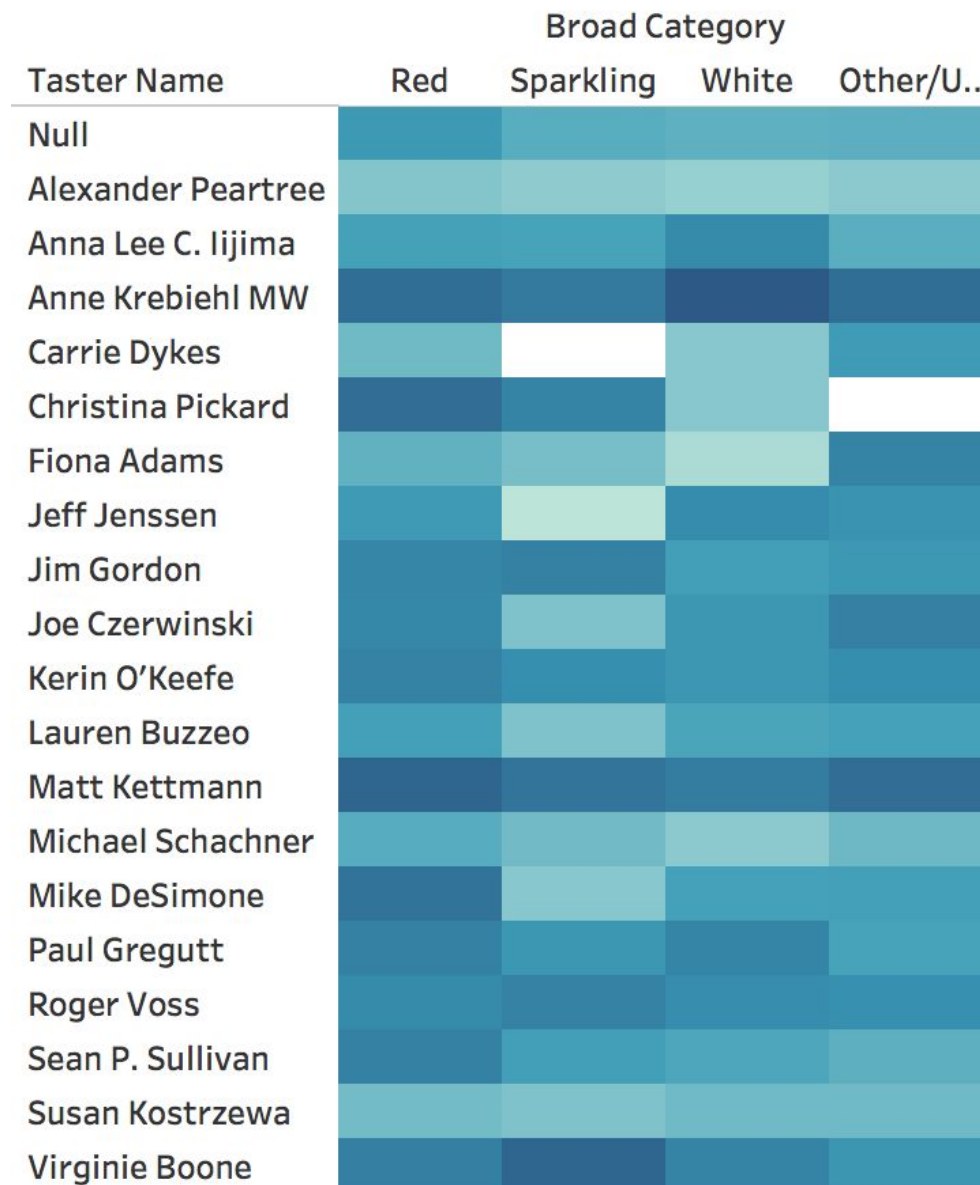


83.563

91.581

5. How does sentiment of reviewer change across different wine types?

Sentiment by reviewer and wine type



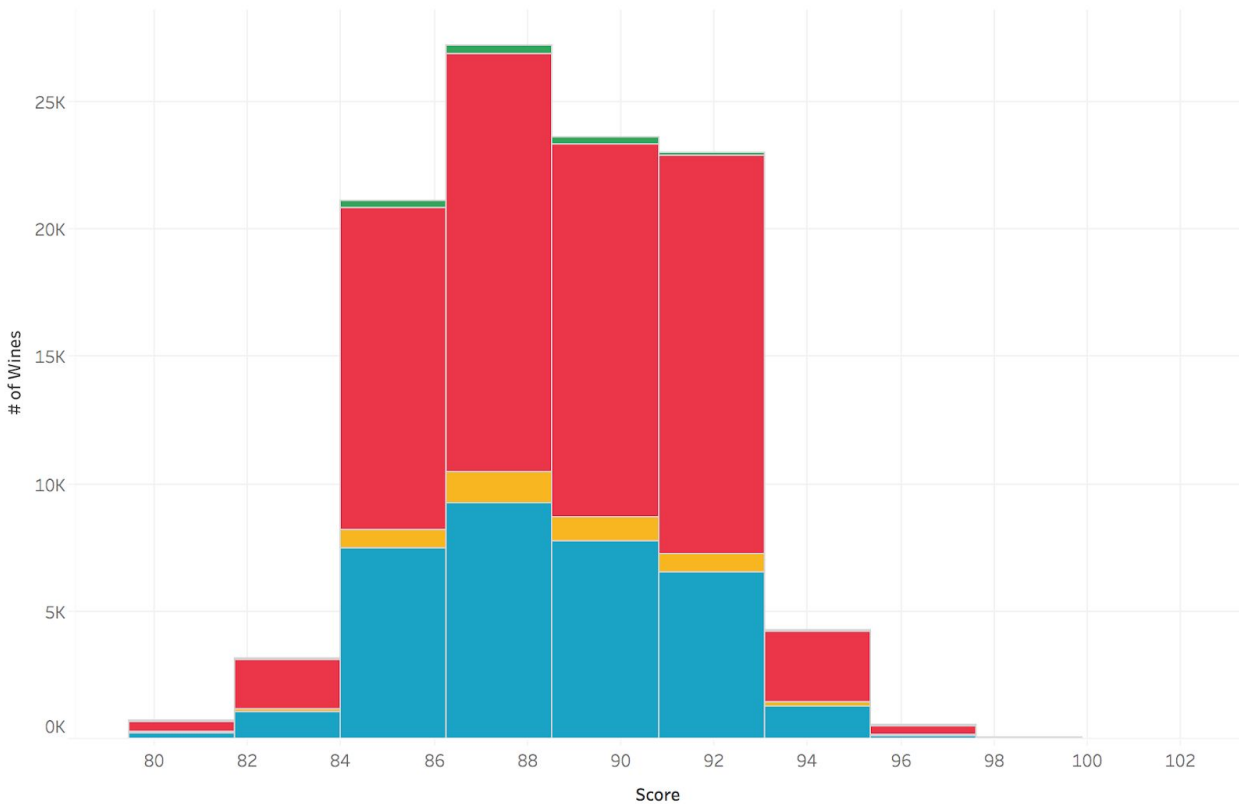
Average of Points (color) broken down by Broad Category vs. Taster Name.

Average Points



6. What does an overall distribution of ratings look like? Segmented by type of wine?

Rating Distribution - By Wine Type



The trend of count of Points for Points (bin). Color shows details about Broad Category. The data is filtered on Taster Name, which excludes Null.

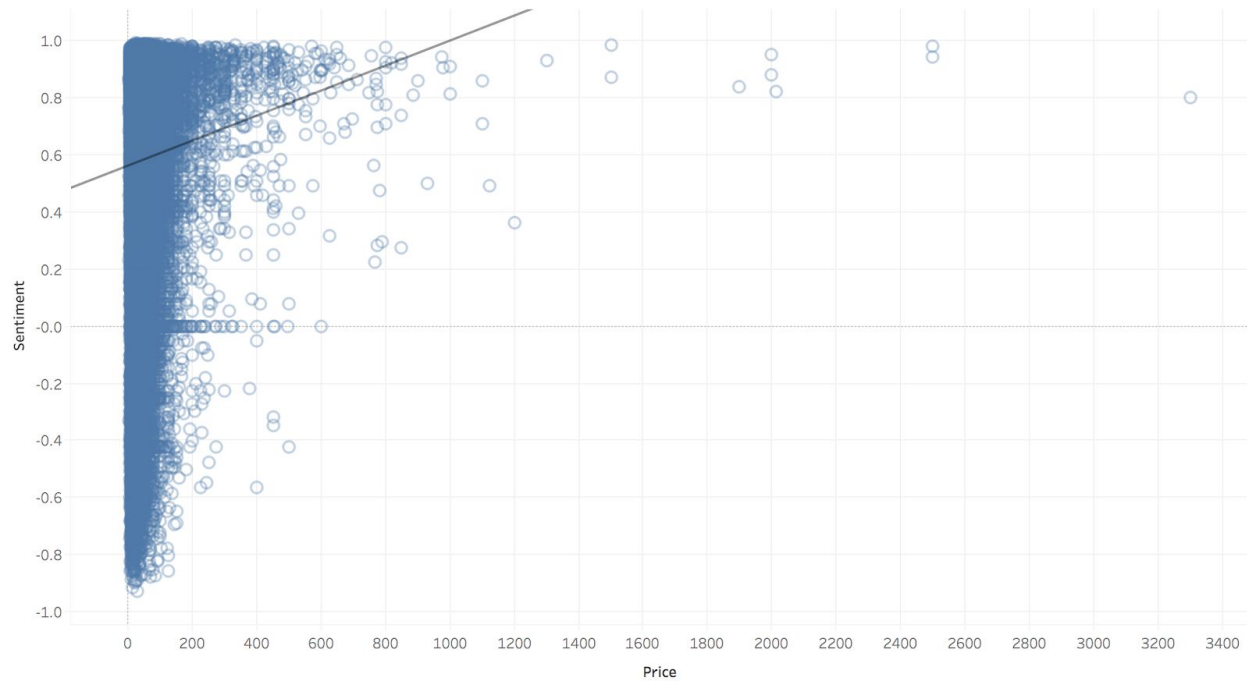
Broad Category

- Other/Unknown
- Red
- Sparkling
- White

It would appear that wine critics are overall high scoring, which the mean clustered about 88-90 overall points. The distribution is quite normal also.

7. Is there a relationship between price of wine and sentiment of the review?

Sentiment of Review over Price

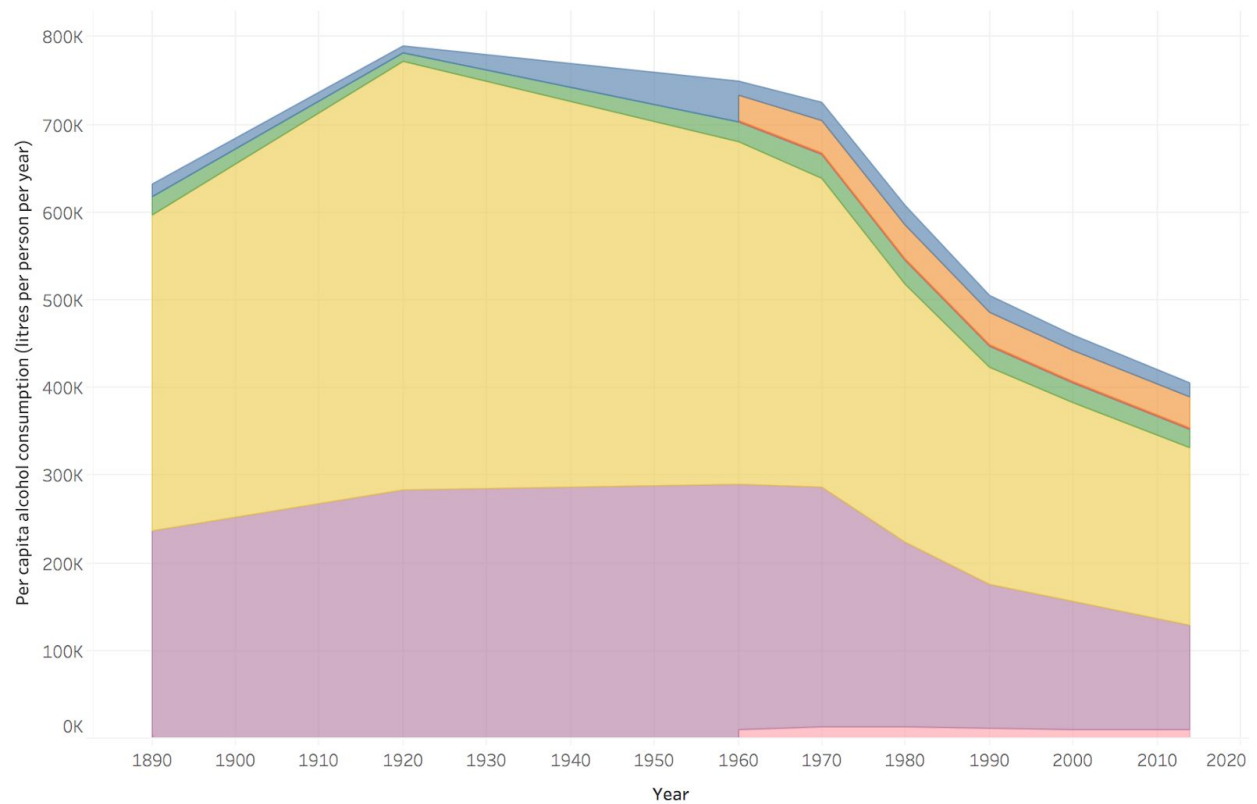


Price vs. Sentiment.

As I was expecting, with price increasing, the overall trend in sentiment is a positive increase. This is expected, but I was wondering if there might be a negative effect from reviewers maybe saying the wine was not worth the high price or something.

8. How was European alcohol consumption varied over time?

Alcohol Consumption Over Time



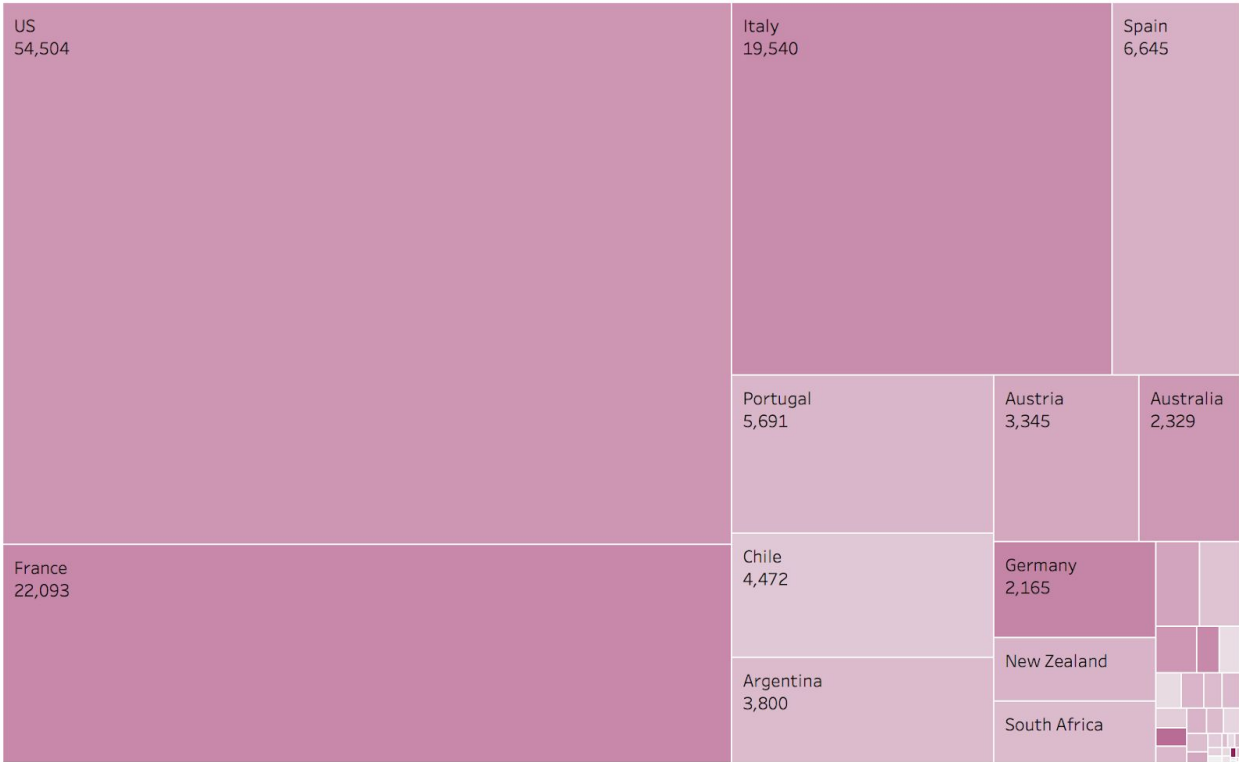
The plot of sum of Per capita alcohol consumption (litres per person per year) for Year. Color shows details about Code.

Code

- AUS
- AUT
- CAN
- CHE
- DEU
- FRA
- ITA
- NZL

9. What does a visual representation of the wines producers by country look like?

Biggest Wine Producers

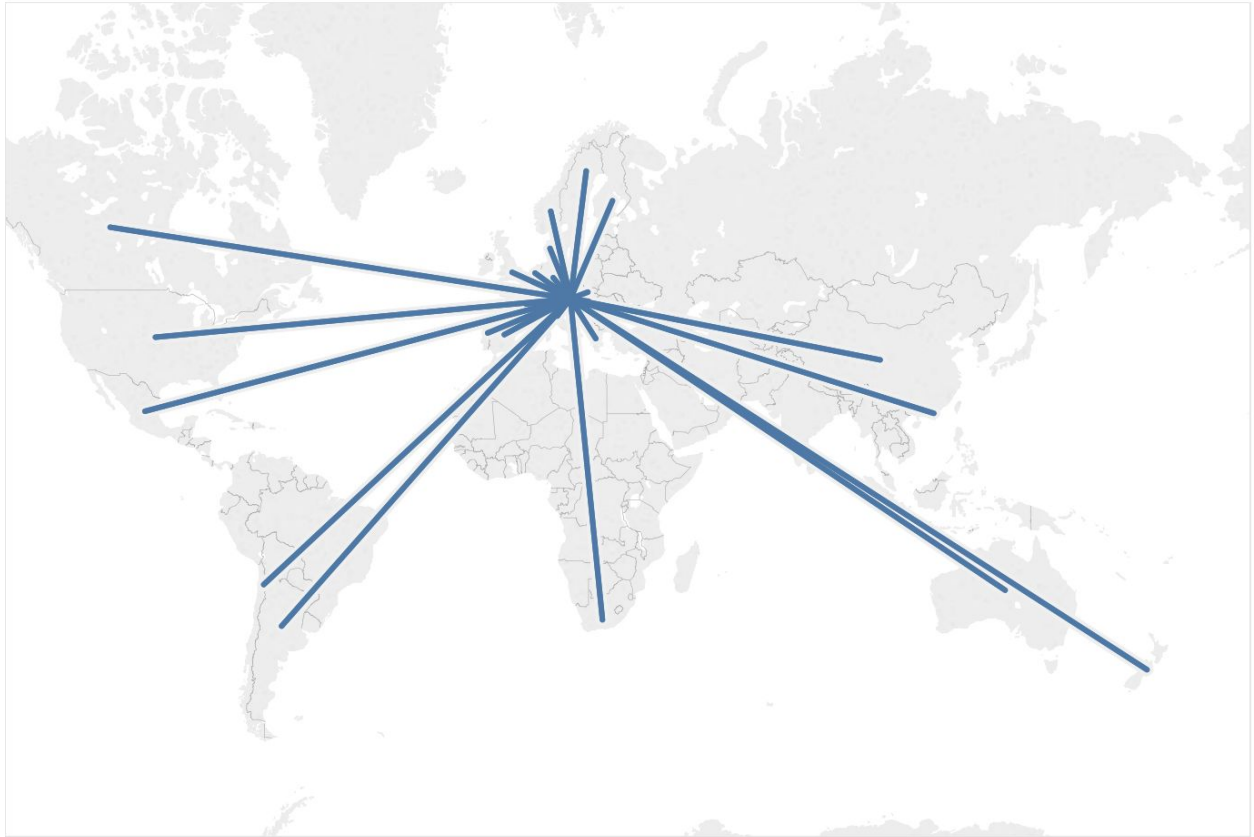


Country and sum of Number of Records. Color shows average of Price. Size shows sum of Number of Records. The marks are labeled by Country and sum of Number of Records.



10. Which countries does Austria trade wine with? (For context see storyline)

Austria's Import and Export Wine Partners



Map based on Longitude (generated) and Latitude (generated). Details are shown for Job ID.

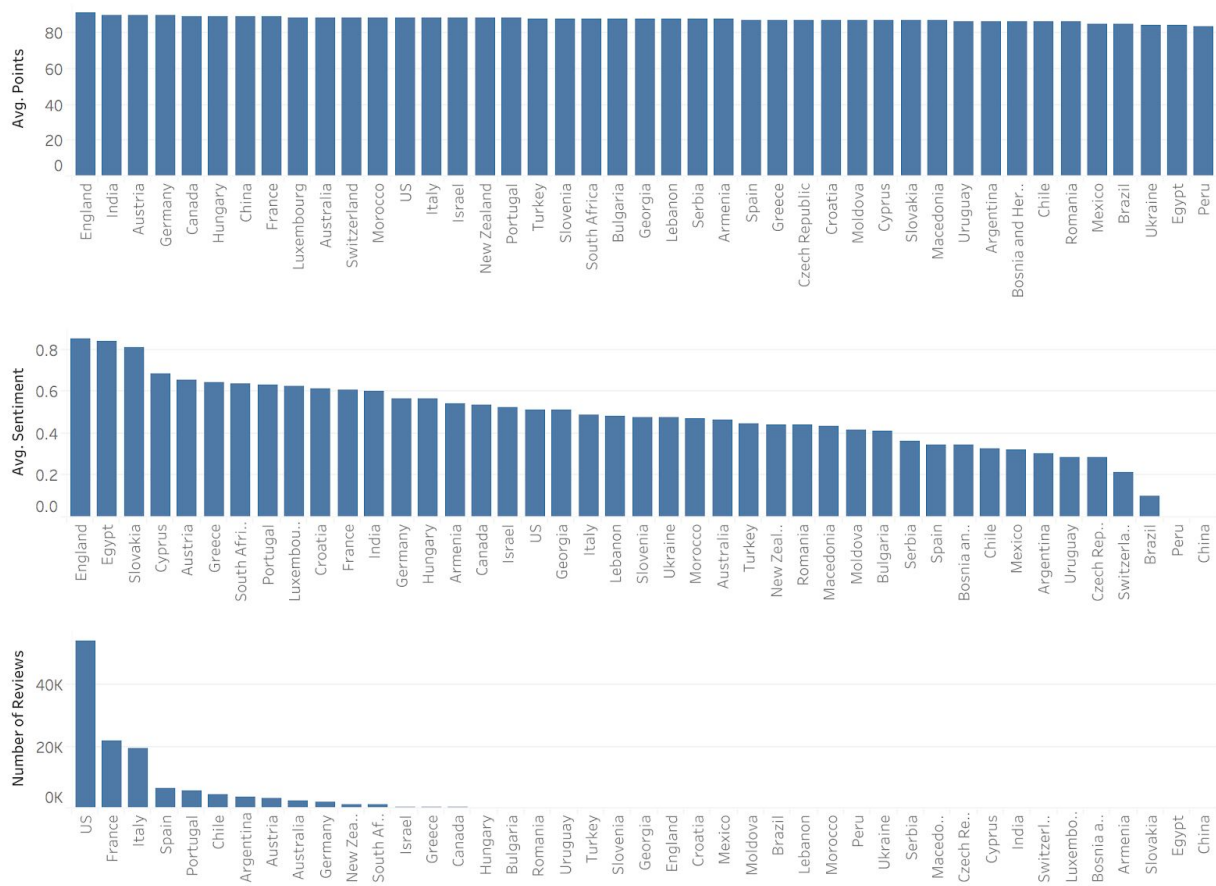
For an interactive graph please see the link below:

https://brianwright2.github.io/msds622_final/interactive_bar/my_index.html

The Wine Titans and the Next in Line:

My natural inclination was to explore not just the top producers of quantity of wine, but of quality wine also. I began by analyzing the relative average Wine Score (1-100), sentiment of the wine review (-1 - 1), and the total amount of documented reviews. The results just from this exploration were very interesting. Here are three bar charts showing the respective metrics mentioned above.

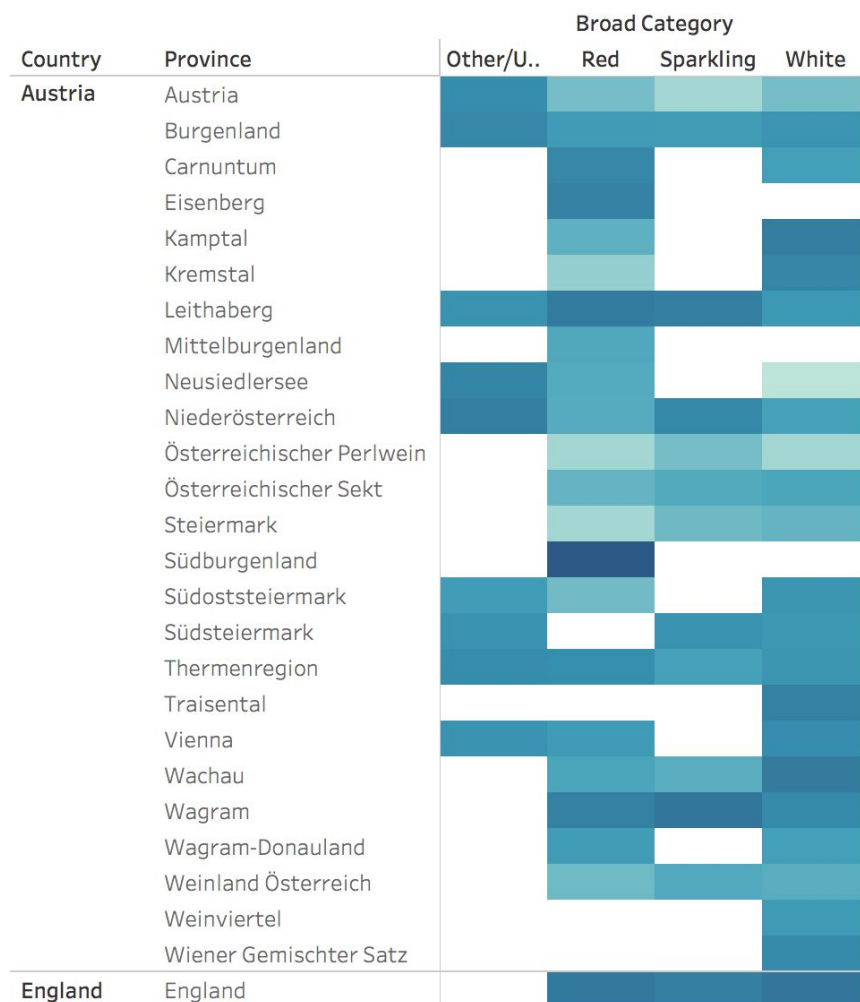
Average Score on Wine Reviews, Average Sentiment of Wine Reviews, and Total Wine Reviews, by country (2017)



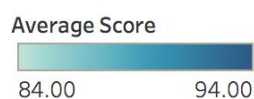
What I found most interesting here was that England lead all countries not only in average score, but in average sentiment of the review. This shocked me as while I figured England had some wine presence, I knew it was not on the magnitude or status of countries like Italy, France, etc. I was a bit suspicious on the sample size portion, and upon review, it appears that they are right around the middle of a steep right tail, implying that their small sample size could be a sign that their data is not truly representative. On the other hand, countries like US, France, Italy, Spain, Portugal,

Chile, Argentina, and Austria have quite large presences in the industry. With that being said, it is worth noting that Austria placed fifth in overall sentiment, and third in average score. The only countries ahead had concerning small sample sizes. While this was some good exploration, I was hoping to go a bit deeper. I decided to break down the scores for Austria by province, and also by type of wine. This was done in an attempt to understand which provinces of Austria in particular are driving this wine revolution, and what their expertise is in. I attempted to do the same with England, but they did not have a province or regional level in the dataset for England, unfortunately.

Average Scores Across Provinces



Average of Points (color) broken down by Broad Category vs. Country and Province. The view is filtered on Country, which keeps Austria and England.



We can see the province Südburgenland has a high score of 94 on average for Reds, and Leithaberg has a strong sparkling wine, as does Wagram. Further, Wachau, Kramptal, and Kremstal seem to be leading the pack in the white wine category. England, when broken down by wine type, is remarkably consistent and high-scoring.

Conclusion:

The conclusion to be drawn here is that Austria is a sleeping giant in the wine industry. While they do not seem to have the same size as powerhouses such as US, or even France or Italy, they have astounding quality. They lead the top ten wine producing countries in both overall average score, and overall average sentiment of the review. This means that not only do the critics give them the highest scores, they also feel the need to gush about their wines the most. Austria will be a force in the wine market in the coming years. England also has the potential to become a contender.

Link to your github page with this analysis and code:

[www.github.com/brianwright2/msds622_final](https://github.com/brianwright2/msds622_final)

Citations:

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

<https://www.austrianwine.com/press-multimedia/statistics/>

<https://ourworldindata.org/alcohol-consumption>