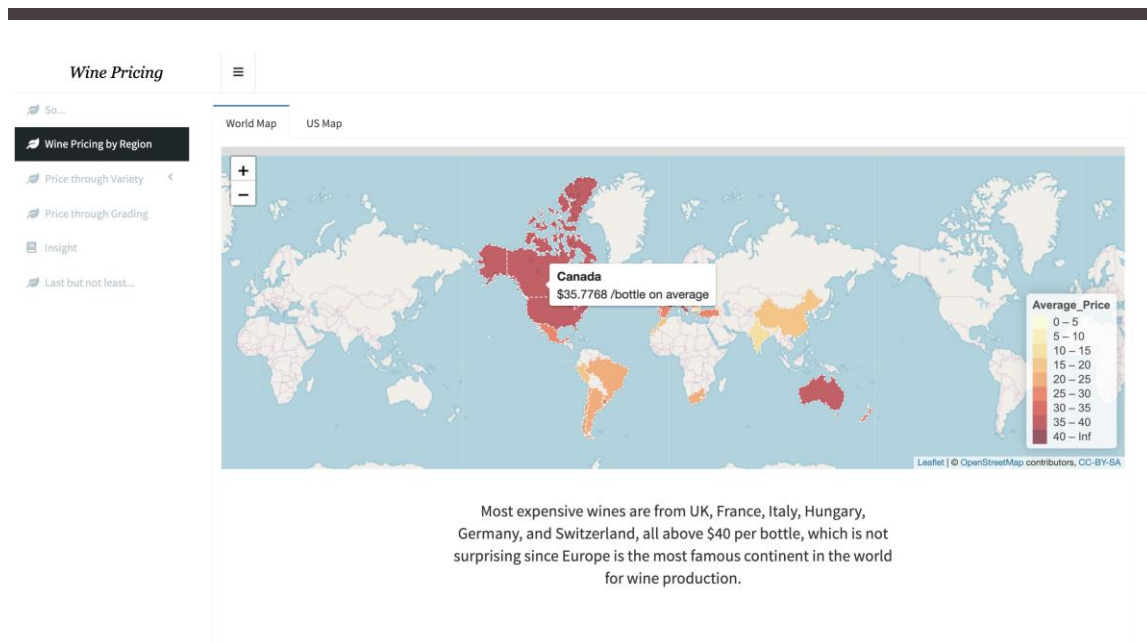


The Secret behind Quality Wine Pricing

Process Notebook

BY HE LIU, MENGRAN XIA, KAREN XU



Introduction

Our project explores different aspects of wine characteristics that might determine its price.

- Does grape variety influence the price of wine?
 - Does flavor complexity influence the price of wine?
 - Does age of different grape varieties influence the price of wine?
- Does expert rating on the wine correlate with its price?

Initial Project Plan: Data

We retrieved our data, “wine review” from Kaggle.com. The useful variables were the wine title, winery of production, city, state, country, price, expert reviews, and expert ratings. Flavors and age of wines could not be derived from the dataset directly, hence we did some manipulation.

- Flavor: In order to retrieve flavor information for each variety, we did text mining on each review to gain the most mentioned flavors for each variety.
- Age: We extracted the production date information from wine titles and subtracted the year of production from 2019.

We also eliminated some useless variables for our projects including taster name and their Twitter names. Since we were interested in providing some choropleth maps, we did eliminate entries that did not provide any information on wine production cities nor states.

Initial Project Plant: Plots

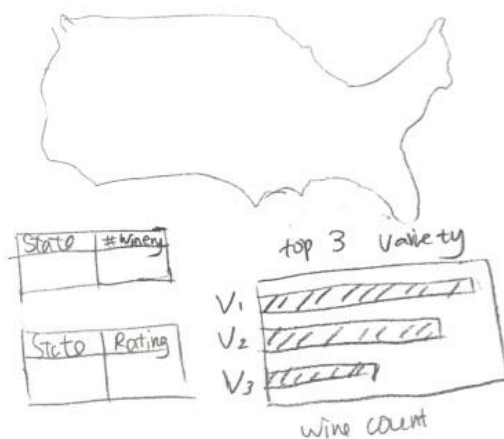
1. General information

We started with providing the most general information of wine price using choropleth maps to demonstrate the price difference between countries and within US between states.

- General Info (2 maps)
 - World map - Leaflet
 - Country Name (PopUp)
 - Country Average price per bottle of wine (PopUp & Color Fill)

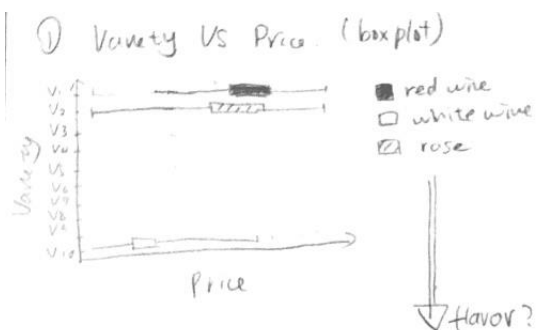


- US map - (Plot.ly)
 - States Name
 - States Average \$/bottle (PopUp & color)
 - Winery#, Expert Score & top 3 variety & count (click-on table & bar plot)



2. Variety vs. Price

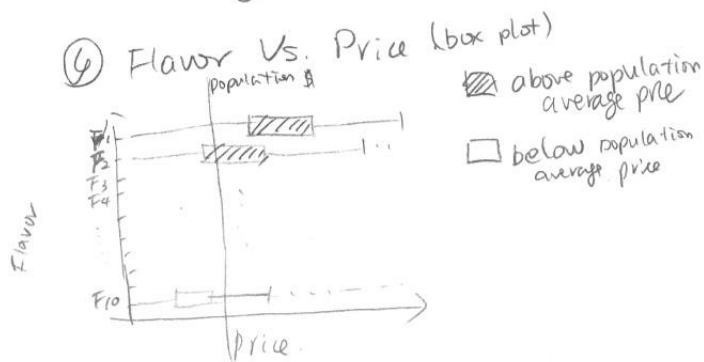
We wanted to explore whether different grape varieties would affect the wine price. We considered grape variety because grape was the only ingredient for wine, so presumably it should influence the quality of the wine, and hence the price. Then we investigated top flavors by variety to see if flavors vary significantly among varieties, using wordcloud and network. Lastly, we decided to group by flavor and calculate median price for each flavor and identify which flavors correlate with higher price.



② Flavor by Variety (Word cloud)

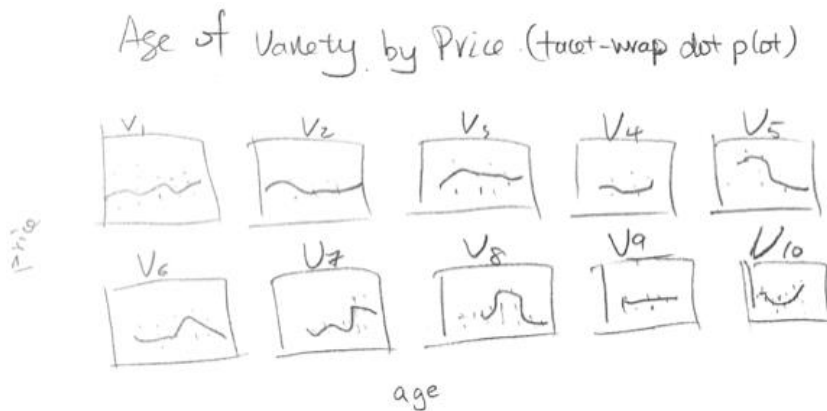


③ Flavor Vs Variety Network



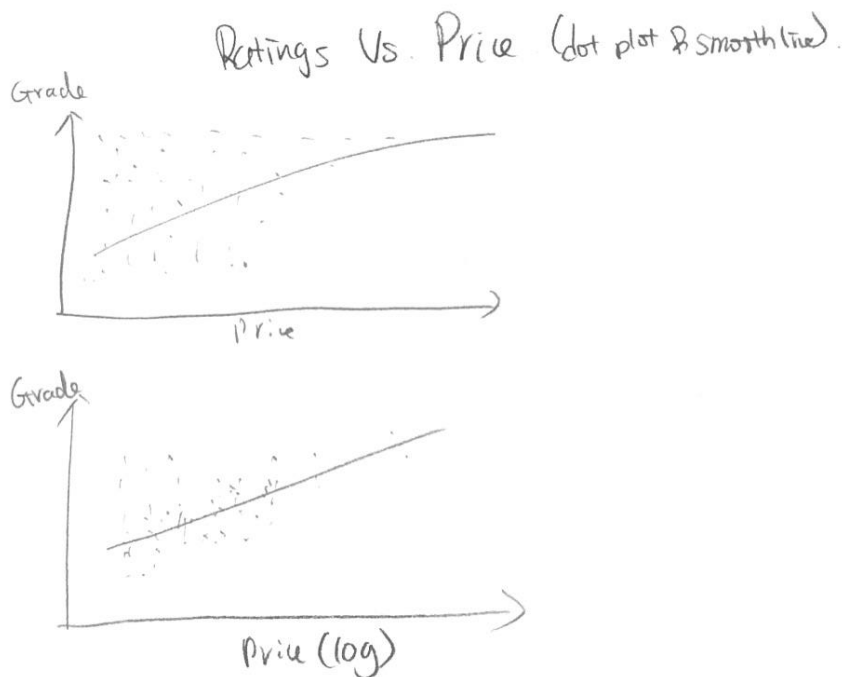
3. Age vs. Price by variety

The age of a wine might have influence on price as well. However, apparently some wine varieties taste better with age, whereas some wine varieties would not be valuable when stored too long.



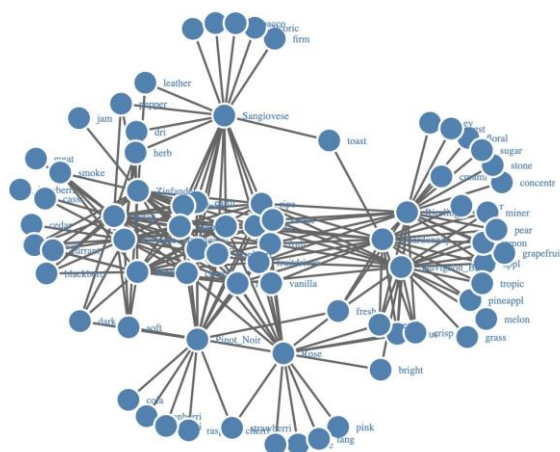
4. Ratings vs. Price

Would wines that received higher ratings tend to be more expensive?



Network of flavor and grape variety

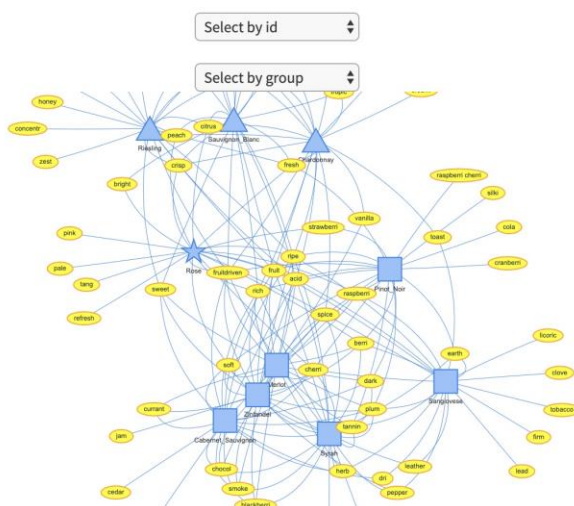
We were glad that most of our initial plots worked out. However, for the flavor vs. variety network plot, we plotted a simple network utilizing networkD3. However, since there were 20 flavors for each of the top 10 varieties on the plot, the simple network was not informative enough.



We wanted it to be more user-friendly and be able to provide some insights on what variety of wine to choose based on flavor preferences, or on what flavors would a particular variety contain.

We then advanced the plot to be interactive so that people could simply click on a specific flavor, and all the related varieties would be highlighted. Furthermore, there was a drop-down menu for users to choose specific varieties or flavors. Last but not the least, we also made nodes into different shapes to represent different colors of wine. Hence, our network plot was able to provide sufficient information yet still remained clear for users.

Our final network plot looked like this:



Hosting on Shiny

After we were introduced to Shiny app in our last two lectures, we decided to use Shiny to present our final project results. It was quite challenging but the results were very satisfying. We spent a long time trying to add click-on events on our US map so that when clicking on a specific state, data frames containing wineries/ratings/variety information could show up below the map. Fortunately, we figured it out utilizing plotly in the end.

Site is Up

We hosted all our final plots/results on shiny and the site is now up and running at:

<https://mengranxia.shinyapps.io/thesecretofwinepricing/>

We have tested many times and it should work fine! However, if there is any loading problems, try to refresh the page. If you have any problems, questions or suggestions, please email us at mx2205@columbia.edu.

Conclusion

- We investigate wine pricing from the wine characteristics perspective, which includes grape variety, wine flavors, wine ages, and wine review score.
- We find out that the most expensive grape varieties among popular varieties are Pinot Noir, Cabernet Sauvignon, and Syrah, which all happen to be red wines.
- By doing word cloud, we realize that red wines are more flavorful than white wines: it not only contains fruity flavors, but also other flavors like earth, spice, currant, and chocolate.
- The flavor versus price graph consolidates our view on flavors: the top most expensive flavors contain mostly extraordinary flavors that can be find in red wines.
- For some grape variety like Cabernet Sauvignon, Sauvignon Blanc, and Sangiovese, increasing age also contributes to increasing price, while for other grape varieties, age doesn't seem to be a contributing factor to price.
- Most importantly, we find that wine review scores correlate perfectly with wine price, which means that as customers, we do get what we pay for!