Read between the Wines.
a refined analysis. in red and white.

Dipika Subramaniam

April 2020

1 Abstract

Why should anybody study wine? It's the lowest on the priority list, especially given today's climate. However, that is exactly the reason I submit to you as to why we should. It's nice to take a step back.
Take a step back from the overwhelming issues of the modern day, and skim over some quaint, sleepy data that leads to delightful conclusions, data that can be used for light-hearted casual conversations and to quench a little interest in our everyday passing thoughts. We will explore the different variables in this dataset and see how they relate to the quality of wine. It's interesting to see that there are even variables related to quality, does this imply that scientifically, there is a highest quality wine? Could we establish solid correlations between different variables? For example: the closer the acidity to 3, the higher the quality.

1.1 Data

Here I'm using the Wine Quality Data Set [1] from UC Irvine. It has 3 files, where two of them are data sets for red wine and white wine, and the third file is essentially a README that includes attributions, explanations of the variables, etc.
Here is a project that uses the winequality dataset to do machine learning, but I will not be referencing it at all.
https://rstudio-pubs-static.s3.amazonaws.com/175762$_8$3$cf2d7b322c4c63bf9ba2487b79e77e.html$

I don't want the graphs to influence my analysis, and the libraries used aren't in the scope of my knowledge of R. The data is also uploaded on Kaggle, and other dataset websites.

2 Conclusion

I hope this is a fun, not serious project for you to sit back and glance through! Best enjoyed with a Cabernet Sauvignon.

References

[1] F. Almeida T. Matos P. Cortez, A. Cerdeira and J. Reis. *Modeling wine preferences by data mining from physicochemical properties.* In Decision Support Systems, Elsevier, 2009.