Wine Quality Analysis

Authored by

Greg Labavitch, Varsha Sharma, Jonathan Gilbert, Brent Lang

Region of Northern Portugal

The northern Portugal region's climate and grape varieties contribute to the unique taste of these young red and white, slightly sparkling wines.

Overview

Wine Quality – “Vinho Verde” (literally “green wine” but translates as “young wine” with wine being released three to six months after the grapes are harvested)

The Vinho Verde region is diverse and versatile in wine styles and profiles, known for producing not only light and fresh wines, but also complex, structured, and mineral wines.

The data that we are working with contains information gathered from performing physicochemical tests on several thousand samples of wines from across the “Vinho Verde” region of northern Portugal. The vast majority of these wines are of medium (normal) quality, with few low (poor) and high (excellent) quality wines.

The data set consists exclusively of numerical data and is divided into two sets: red wine and white wine. The variables that we will be focusing on are fixed acidity, volatile acidity, and chloride levels.

Relevant Wine Definitions/Variables

01

Fixed Acidity

Fixed acidity is the naturally produced acids of the wine. The acids include tartaric, malic, citric, succinic, and lactic.

02

Volatile Acidity

Volatile acidity is the measurement of gaseous acids. Acetic acid is the primary volatile acid in wine and is associated with the smell and taste of vinegar.

03

Chlorides

Chloride indicates the level of sodium in the wine. The greater amount of sodium chloride is expressed through added saltiness. The chloride content is used to account for quality.

Defining Quality

General Definitions of Quality

- **Subtlety and complexity**
- **Aging potential**
- **Stylistic purity**
- **Varietal expression**
- **Ranking by experts**
- **Consumer acceptance**

For the purposes of analysis, the quality used describes chemical properties. In this project, the data is limited to describing the wine quality as number from a scale from 0 to 10. We consider low quality to be a score of less than 5, medium quality to be greater than 5 and less than 7, and high quality to be greater than 7. The data used may be helpful in research regarding many of these categories; however, it may not always be suitable without additional resources.

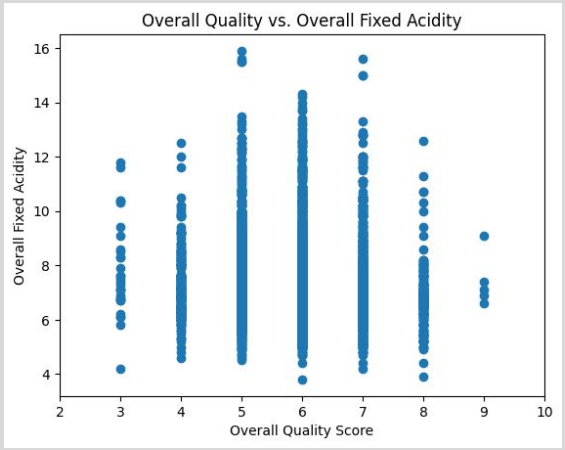
**Quality is an output variable based on sensory data.*

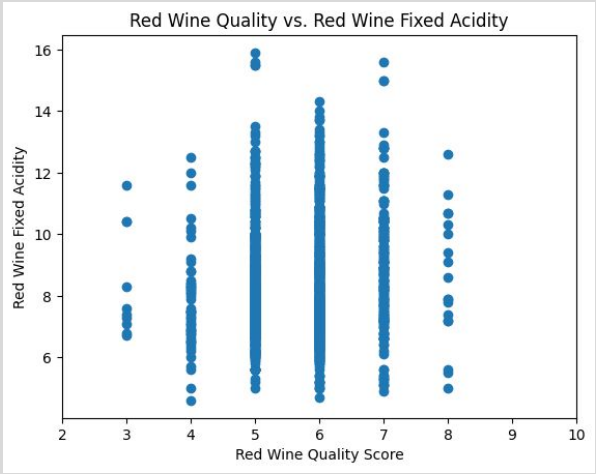
Project Objective

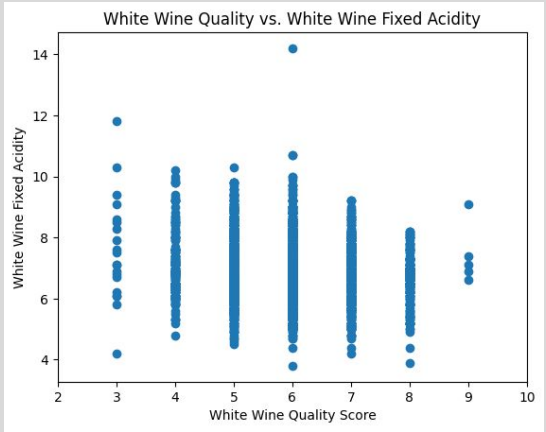
- *Analyze the quality of red and white wine*
- *Compare overall quality*
- *Analyze how wine quality is affected by fixed acidity, volatile acidity, and chloride levels*

Fixed Acidity

Q: Is there a correlation between fixed acidity levels and wine quality? If so, what kind?



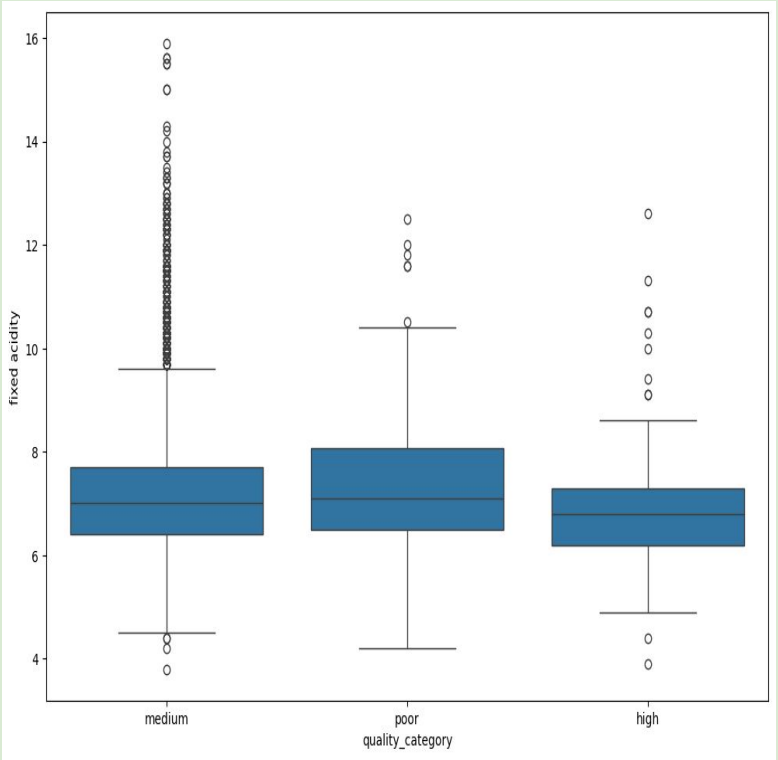




Box Plot for Fixed Acidity

Likely due to the fact that most of the wines are of medium quality, the set of medium quality wines contains many outliers for the fixed acidity data. Across the three quality categories, the fixed acidity tends to fall within a fairly small range.

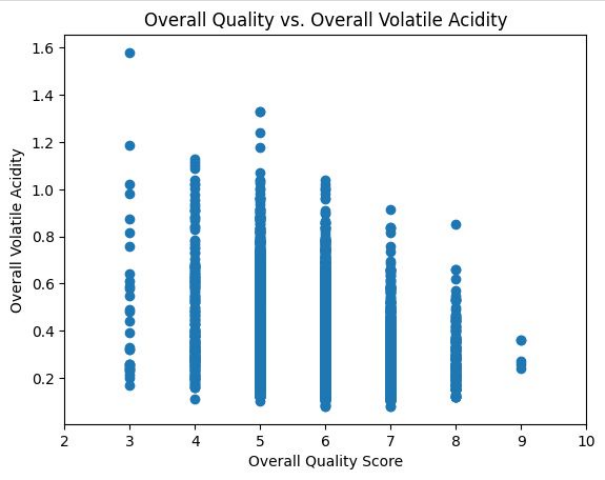
Fixed Acidity Levels

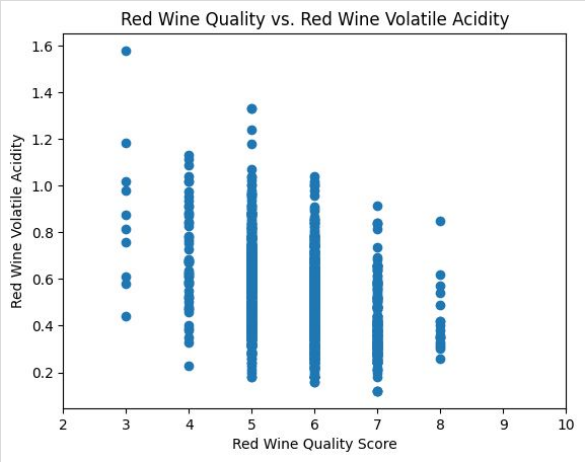


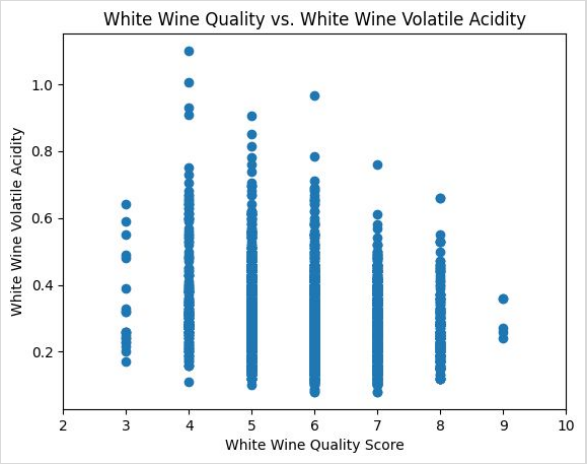
Quality Category

Volatile Acidity

Q: Is there a correlation between volatile acidity levels and wine quality? If so, what kind?



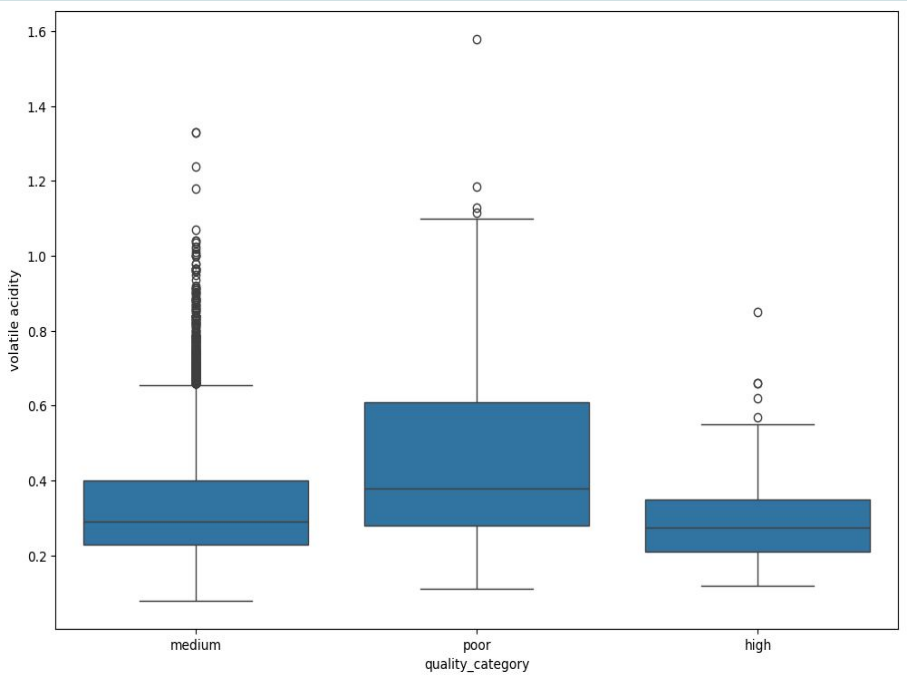




Box Plot for Volatile Acidity

The volatile acidity data is highly spread out in the poor wines and more concentrated in the medium and high quality wines. The outliers are far more numerous among the medium wines.

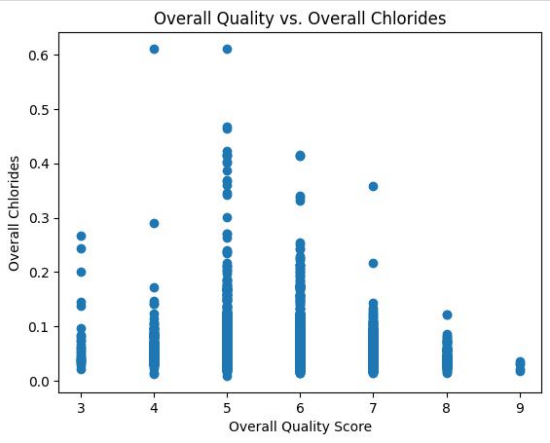
Volatile Acidity Levels

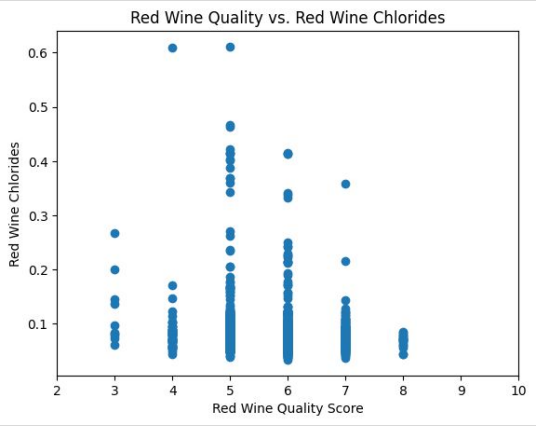


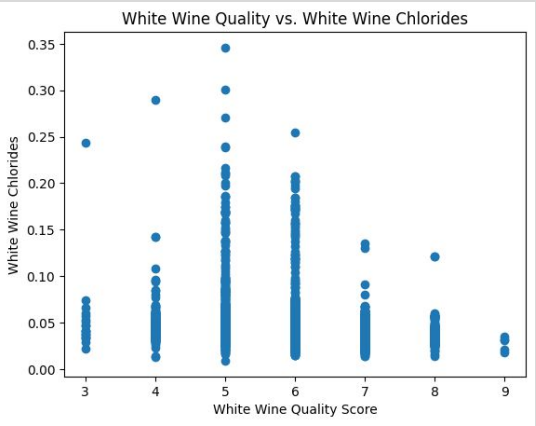
Quality Category

Chlorides

Q: Is there a correlation between chloride levels and wine quality? If so, what kind?



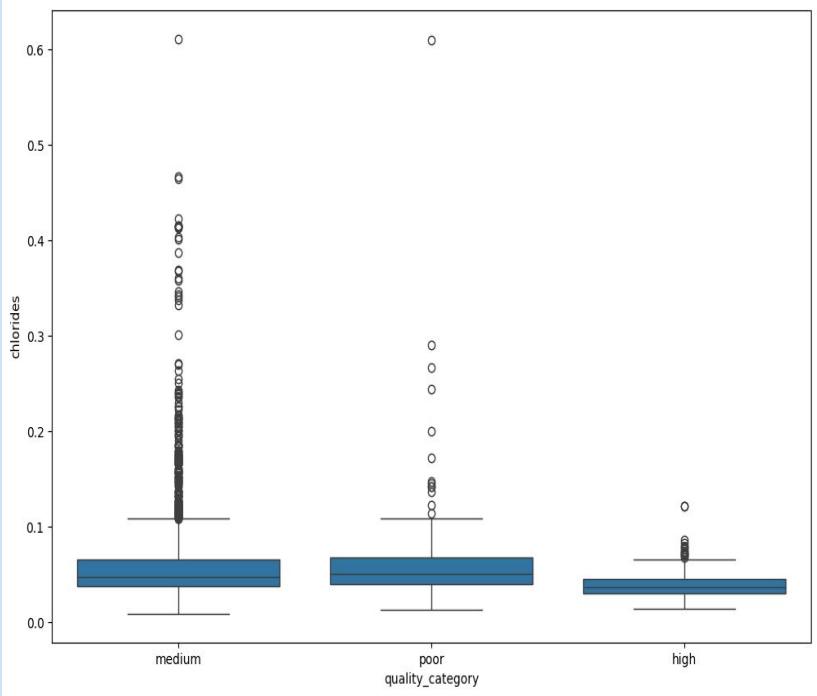




Box Plot for Chlorides

Chlorides box plot also illustrates a predominant distribution of medium values, accompanied by sparse presence of outliers in the poor and high quality wines, and many outliers among the medium quality wines.

Chloride Levels



Quality Category

Bar Chart: Red and White Overall

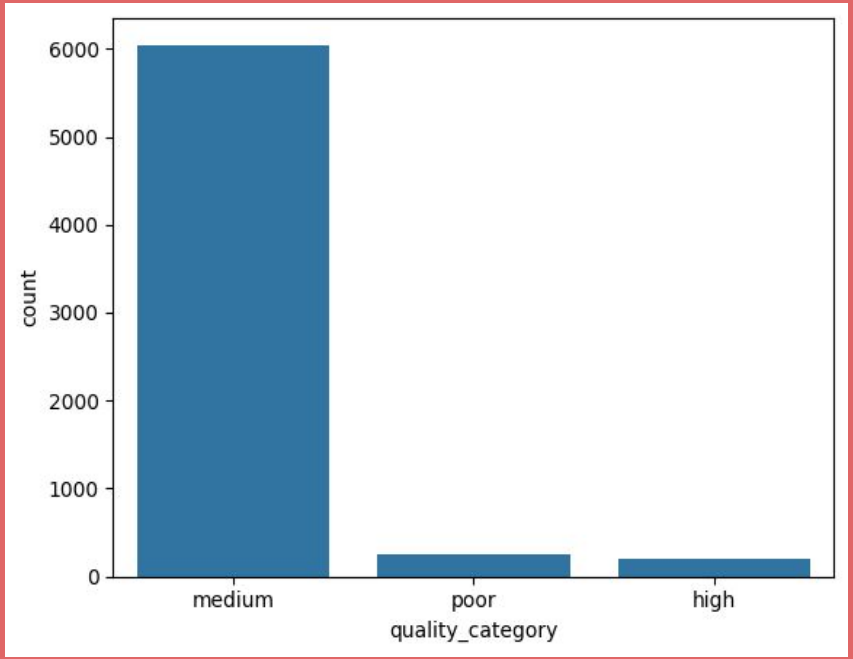
**CAVEAT: The wine classes are ordered and not balanced (e.g. there are many more normal (medium) wines than excellent or poor ones).*

Quality category is based on quality score

Quality score < 5 = poor

Quality score (>= 5 and <= 7) = medium

Quality score > 7 = high



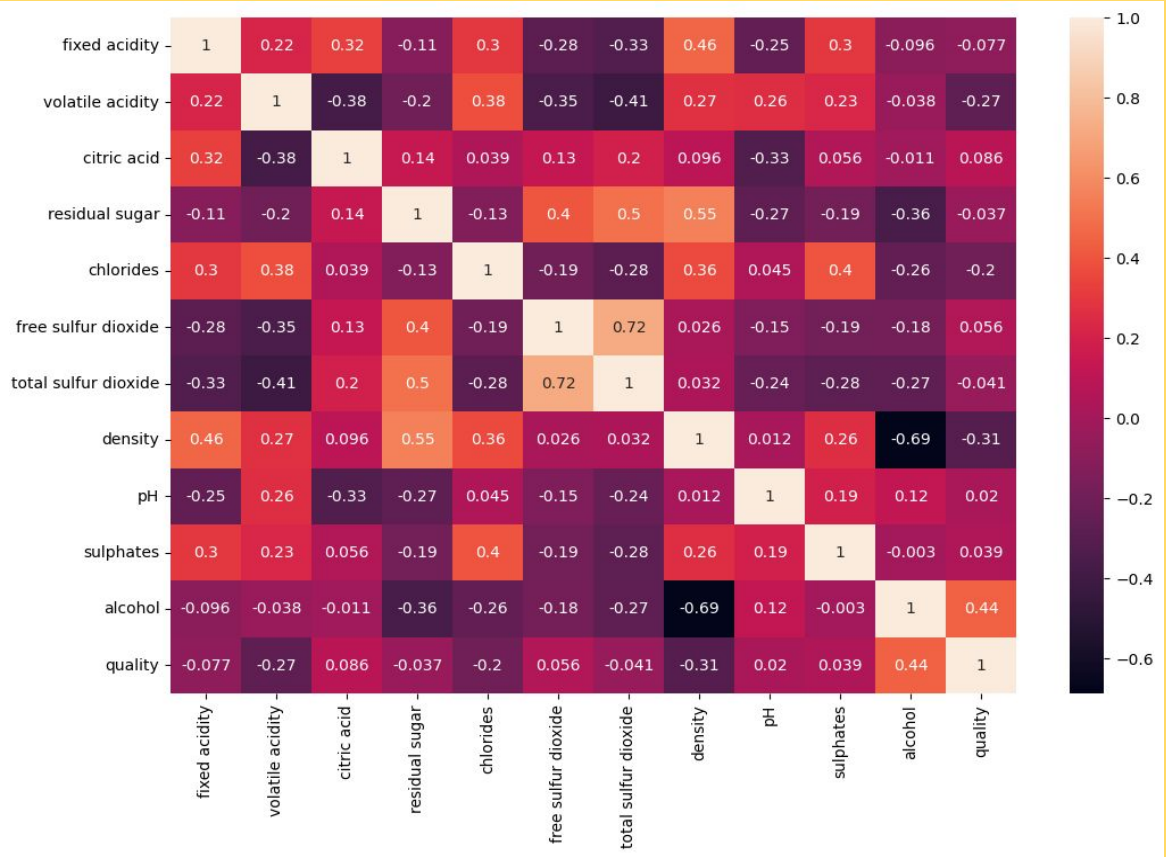
Heat Map

The data has been analyzed using a heat map to visually represent patterns and correlation among variables in our dataset.

These variables appear to have a positive correlation with quality: free sulphur dioxide, citric acid, alcohol and pH.

The variables that appear to have a negative correlation with quality are chlorides, volatile acidity, fixed acidity, residual sugar, total sulphur dioxide and density.

The heat map indicates a low negative correlation between our chosen variables and wine quality.



Conclusions

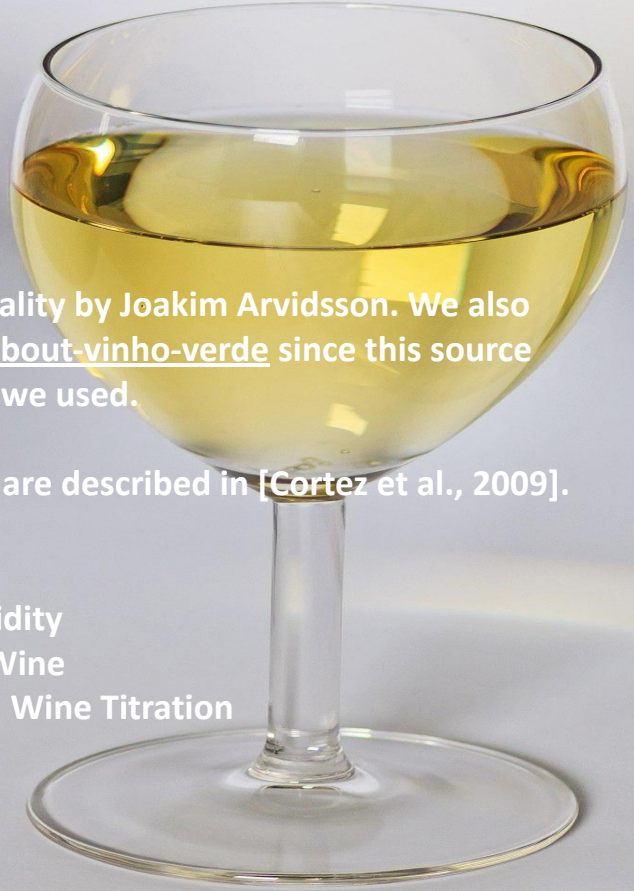
- Fixed acidity and quality: no correlation
- Volatile acidity and quality: low negative correlation
- Chloride levels and quality: low negative correlation

The heat map more or less confirms these findings. However, the heat map considers all points, while the conclusions are drawn after ignoring the points in the scatter plots that represent low and high-quality wines. If the heat map were reconfigured without these points, we may be able to draw better conclusions.

Digging Deeper/Improvements for next time

- Strip out the high and poor quality wines
- Drawing conclusions about how the quality of the wines is affected by our chosen variables may be easier when the statistics aren't affected by a low number of excellent and poor quality wines.
- We also need to improve the labels and add titles to our box plots.

Sources



The source for our data is Kaggle under the title Wine Quality by Joakim Arvidsson. We also drew information from https://www.vinhoverde.pt/en/about-vinho-verde since this source was included with the page where we found the csv files we used.

This dataset is publicly available for research. The details are described in [Cortez et al., 2009].

Source definitions:
Fixed Acidity; Agrovin Techniques for Correcting Wine Acidity
Volatile Acidity; Penn State Extension Volatile Acidity in Wine
Chloride; Mantech Inc. Application Note #105-Chloride in Wine Titration

Images: General Google search


Questions?