

# Wine reviews data set: description file

April 11, 2024

The wine reviews has been adapted from a larger data set available on Kaggle: <https://www.kaggle.com/zynicide/wine-reviews>. The data set consists of  $n = 2500$  French wines that have been reviewed and rated by the Wine Enthusiast website. Wine ratings of  $\geq 90$  indicate that the wine is rated as “superior”. The data set includes the following variables:

**points** The points rating for each wine;

**superior** indicator for if the wine is judged to be superior, i.e., whether or not **points**  $\geq 90$ ;

**price** price of the wine in dollars;

**variety** variety of wine being rated;

**description** Free text, consisting of the review’s description of each wine.

The data set also includes indicator variables for the following terms: **Crisp**, **Dry**, **Finish**, **Firm**, **Fresh**, **Fruit**, **Full**, **Rich**, **Round**, **Soft**, **Sweet**. Each variable indicates whether or not the term was used in the wine review or not.