**Questions of Interest**

1. *How does the chemical makeup of a wine -- measured through fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol content -- as well as an expert’s quality score of the wine, and type of wine, relate to the wine’s alcohol content? (Regression)*
2. *Given the chemical makeup of a wine, as well as the type of wine, can a determination be made about whether a wine tastes good or bad?(Classification)*

**Prediction of Alcohol Content**

The analysis supports the prediction that a wine’s chemical makeup, its type, and an expert’s quality ranking can sufficiently explain wine alcohol content. The most important predictors for a wine’s alcohol content appear to be density and residual sugar. These predictors make sense when the chemistry of wine is considered. Since alcohol is less dense than water, the concentration of alcohol will strongly influence the density; and since the process of fermentation converts sugars to alcohol, it is also logical that residual sugar is a strong predictor of alcohol content. As sugar content increases, alcohol content should also increase considering that there is more sugar to convert.

**Commentary**

The Bootstrap Aggregation (bagging) model produced the best results out of the models tested, including the Random Forest model that was expected to be a superior predictor. The bagging model had a test mean square error of 0.18, compared to 0.21 for Random Forest and 0.52 for the initial tree and pruned tree models. Density was the most important feature of the model using both metrics. It appears that the Out of Bag (OOB) error decreased by an average of over 250% when density was removed and that the decrease in training Residual Sum of Squares (RSS) fell by an average of over 2500. This feature was the most important and was followed distantly by residual sugar. An interesting point of note is the unimportance as a variable of free sulfur dioxide and not type in explaining quality. It appeared that in general, white and red wines are more closely related to a certain quality.

**(2)**

**Classification of Good and Bad Wines**

A determination can be made about whether a wine will be ranked as “good” or “bad” based on the chemical makeup and type. Each model we used found the most significant predictor to be alcohol content, but other predictors were found to be highly significant as well. Some other important components of a wine which were effective in predicting quality include density and volatile acidity. Given a priori knowledge of wine and some basic knowledge of chemistry, these findings are consistent with what we expected. This strengthens our model and allows us to predict effectively whether a wine tastes good or bad given some basic information about the wine.

**Commentary**

The Random Forest model produced the best results out of the models tried. This model had a test mean square error of 0.1896, compared to 0.1945 for the Bagging model and 0.2739 for the initial tree and pruned classification tree models. Alcohol was the most important feature of the model using both metrics. Perhaps intuitively, the least important variable is type. As wines are complicated beverages, a lot goes into making them and even two white wines can differ greatly.