Week 4 Assignment   
Predicting Wine Quality with the UCI Red Wine Quality Dataset

Scott Schirkofsky  
MSDS 680 Machine Learning

Regis University

Predicting Wine Quality with the UCI Red Wine Quality Dataset

**Introduction**

This paper focusses on the Wine Quality Dataset from the UCI Machine Learning Depository. The objective is to use the wine’s chemical properties to predict a wine quality ranking.

**Load Libraries, Collect & Load Data**

All required libraries are loaded, and the data is uploaded from UCI.

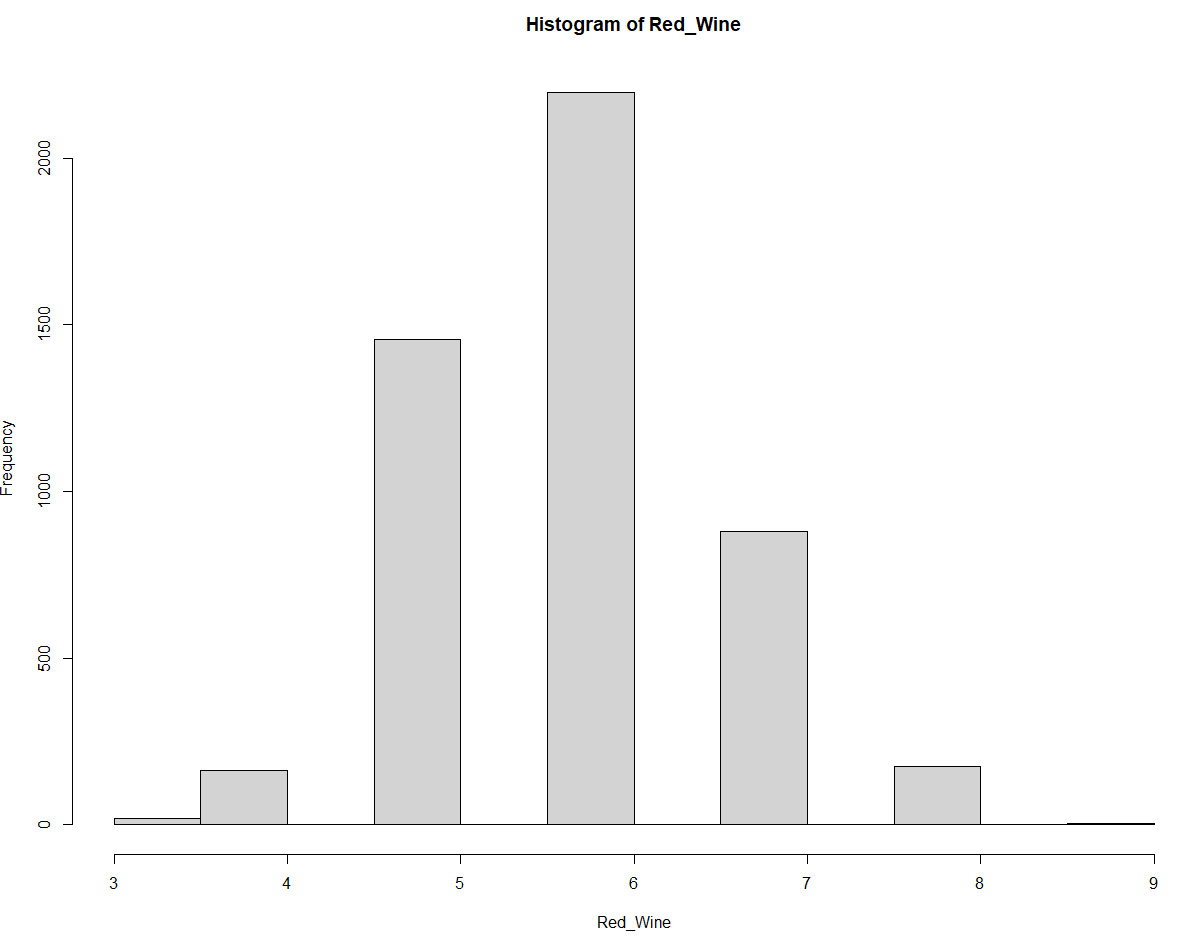


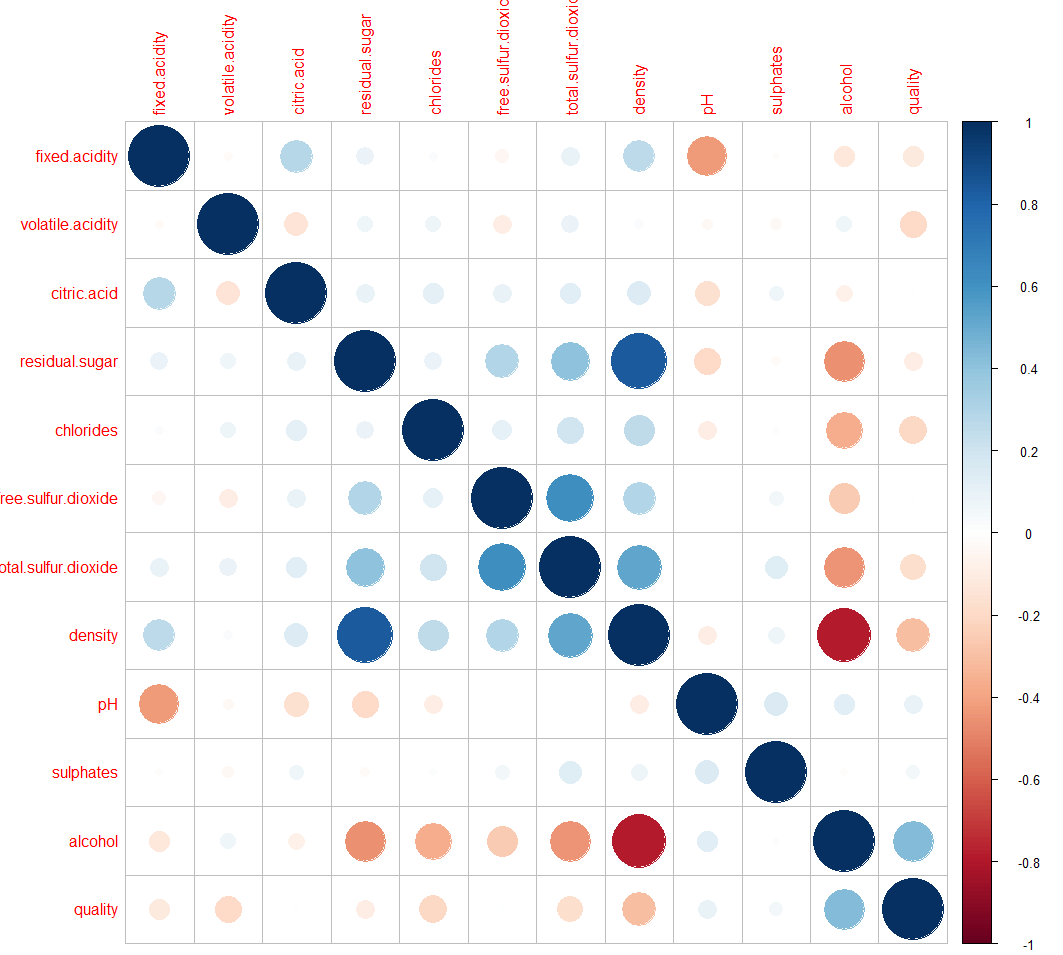
**Data Exploration**

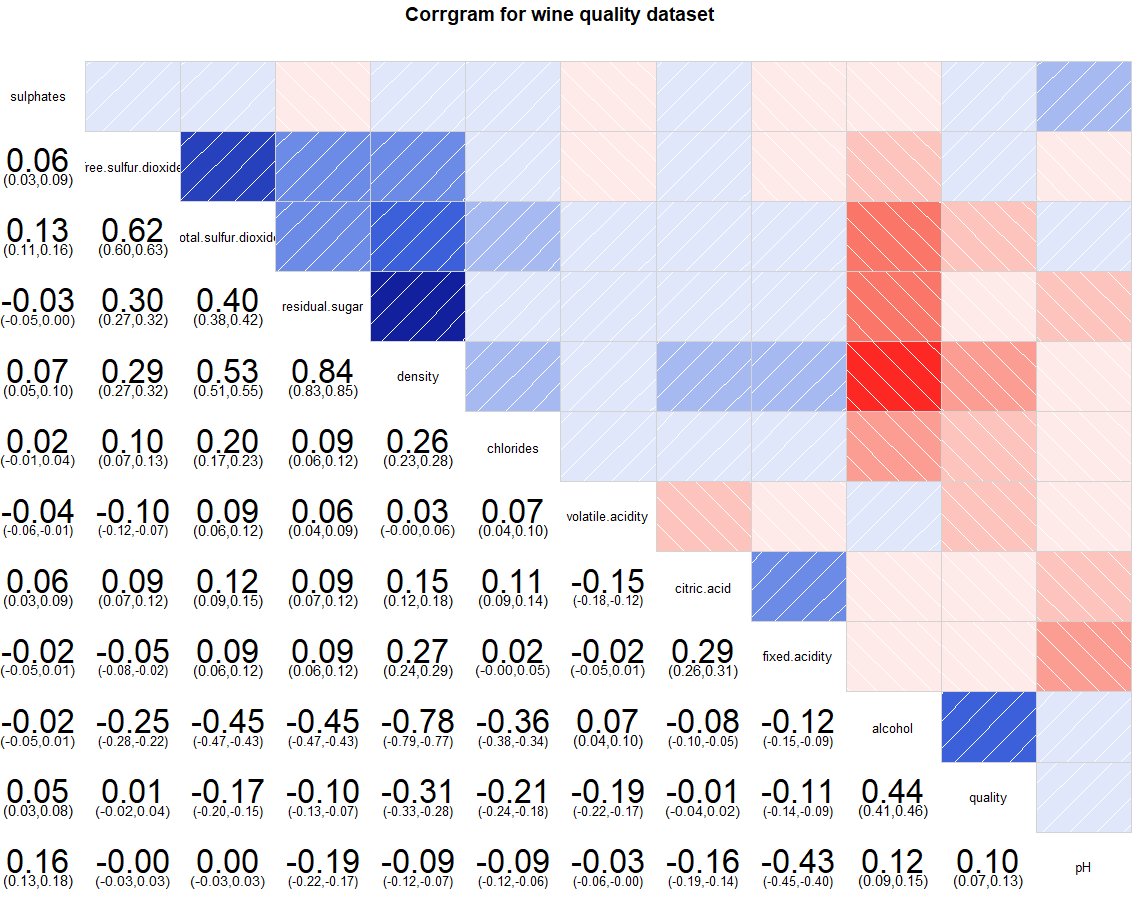
The dataset contains 4898 objects and 12 variables. All of the variables except Quality are numeric. Quality is an integer.



The wine quality ratings frequency is shown on a histogram below. The levels that are shown representing the ratings we created from converting Quality to a factor.







**Data Transformation / Preparation**

Data are checked for NA values, none are found. The Quality predictor is converted to a factor so that it can be used to categorize the data. Nine levels are used.





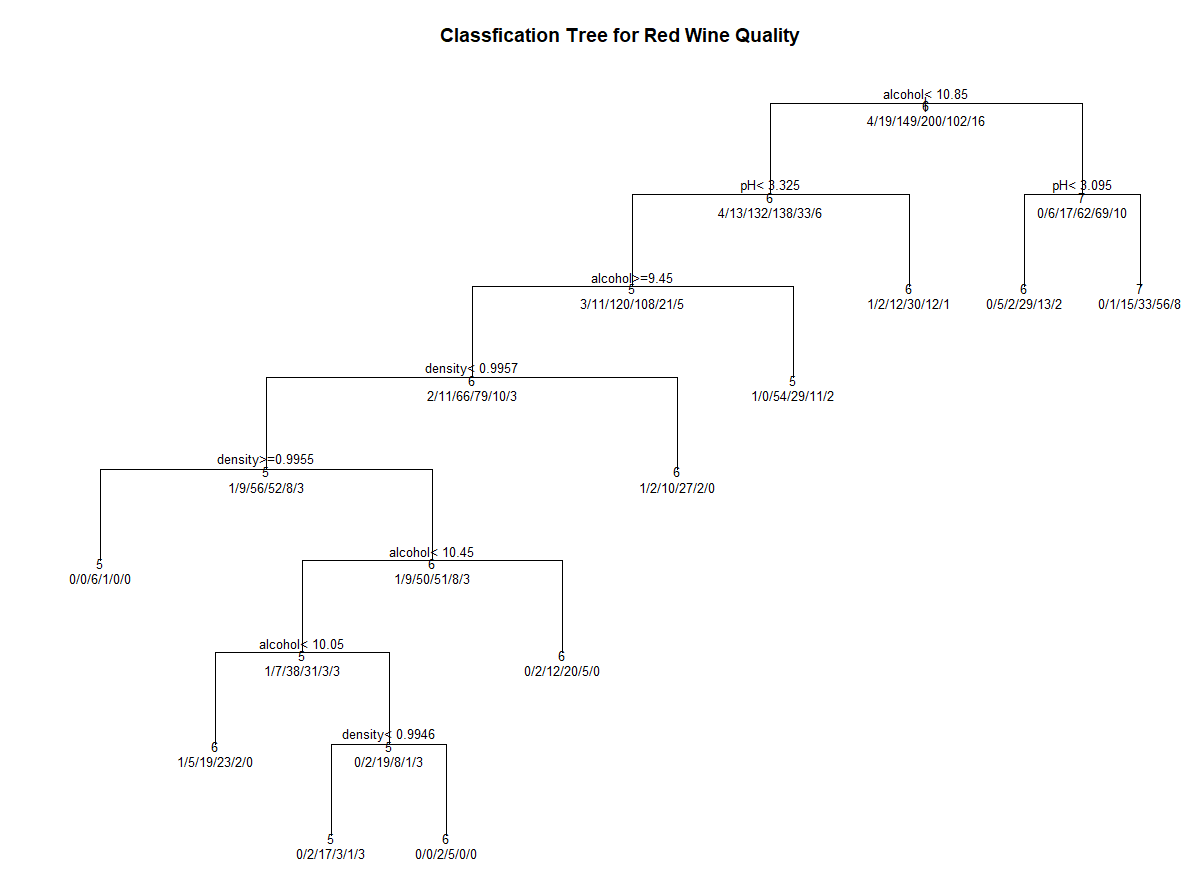
**Train and Predict**



**Create decision tree**







**Prune Tree**



**Conclusion**