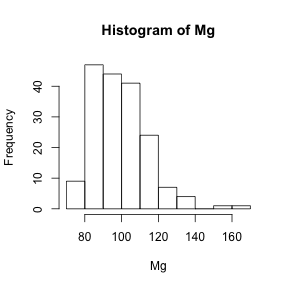
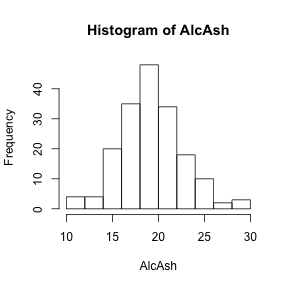
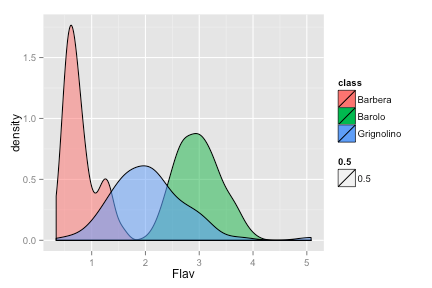
# Executive Summary

With the Wine dataset we will be applying a supervised learning to predict the class of the unknown wine. In this initial project we will perform an EDA on the dataset to understand the distributions of the variables with the context of classification in mind. Within our EDA we have two tasks: **univariate and bivariate analysis**.

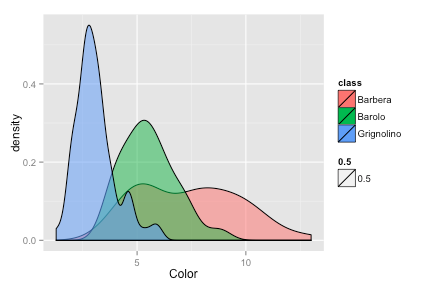
The univariate EDA was performed using histograms of each variable. The results show that there are some variables that are normally distributed, while others appear to be log-normal in distribution. For example, “AlcAsh” shows a normal distribution whereas “Mg” could potentially benefit from a log transformation if necessary:



The bivariate EDA was then performed to see the relationship between the explanatory and target variables. This was completed using Kernel Density Estimates, with each density estimate broken down by the target classes. The results were positive, with some variables showing good breakdown between the classes. For example “Flav” shows excellent separation between the “Barbera” and “Barolo” classes, as show below:



In addition, the “Color” variable could be used to separate the “Grignolino” class from the others:



Overall, these results show that an algorithm should be able to separate the three classes easily using this variable.