Wine Cultivar, Alcohol, and Ash Analysis

Loren Grooms

01/07/2020

Introduction

Wine is a beverage produced and enjoyed by humans across the globe for thousands of years. A number of factors play into the taste and quality of wine, from magnesium levels to protein content. This project aims to analyze the relationship between the alcohol and ash content of Italian wine made from three different grapes. First, we will see if alcohol and ash content are accurate predictors for grape variety using K Nearest Neighbors analysis. Then, we will use logistic regression to visualize this relationship per grape variety.

Method

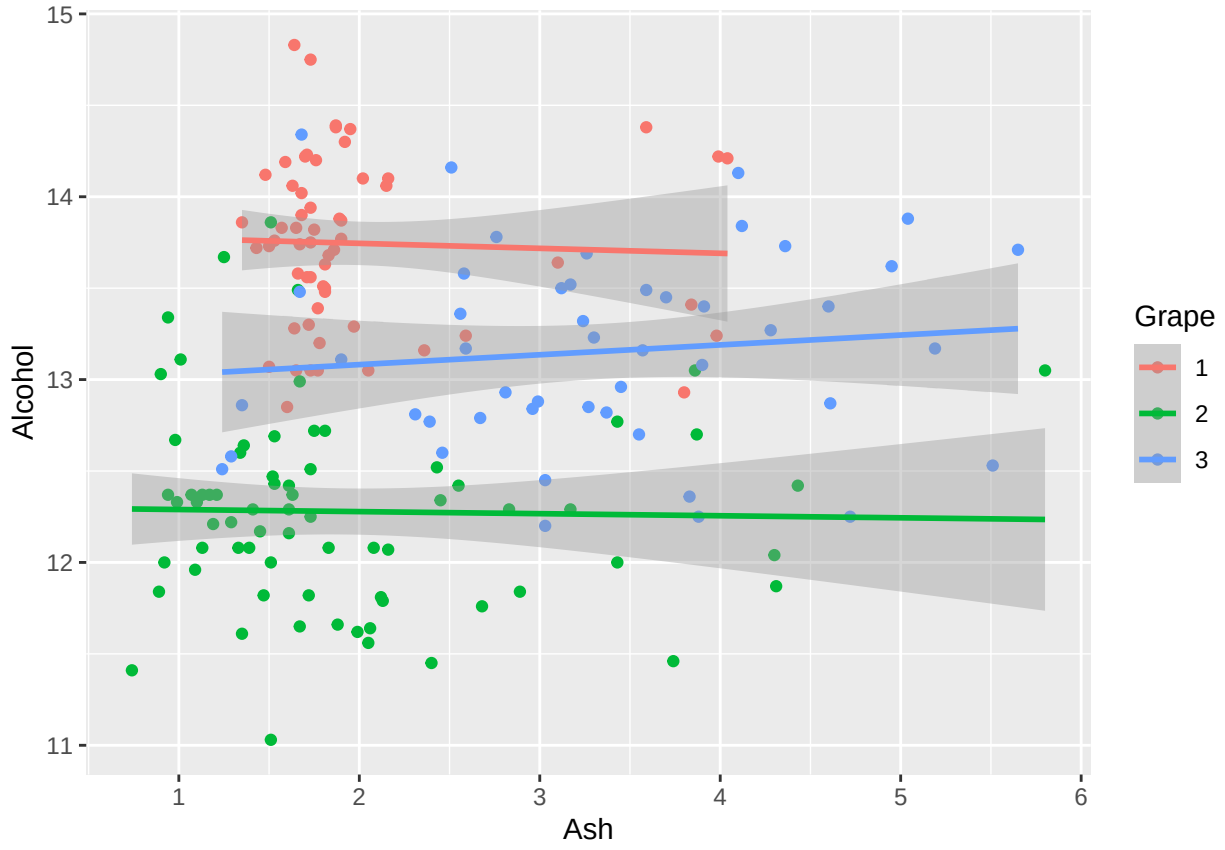
This analysis begins by checking for and installing any missing packages onto the user's system. Next, the data file is downloaded from UCI's online data repository and cleaned by removing delimiters, creating a dataframe, eliminating unused columns, and converting all datatypes to numeric. Before using our kNN analysis, we must normalize the data using a standard normalization function. After applying this to the entire frame, we set 66% of the data as our training set and the remaining 33% as the test set. We then separate our prediction target (type of grape) into separate training and test variables. Next, we apply the kNN function from the "class" package to form groups of predicted values. These groups indicate regions of data with similar attributes which are predicted to contain the same grape type. Finally, we will use a validation table from the "gmodels" package to calculate the accuracy of these predictions, and then graph a logistical model to visualize the relationship between alcohol and ash content per grape.

Results

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  61
##
##
##                | kWine
##    testGrape |         1 |         2 |         3 | Row Total |
## -------------|-----------|-----------|-----------|-----------|
##            1 |        30 |         0 |         0 |        30 |
##              |     1.000 |     0.000 |     0.000 |     0.492 |
##              |     1.000 |     0.000 |     0.000 |           |
##              |     0.492 |     0.000 |     0.000 |           |
## -------------|-----------|-----------|-----------|-----------|
##            2 |         0 |        19 |         0 |        19 |
##              |     0.000 |     1.000 |     0.000 |     0.311 |
##              |     0.000 |     1.000 |     0.000 |           |
##              |     0.000 |     0.311 |     0.000 |           |
## -------------|-----------|-----------|-----------|-----------|
##            3 |         0 |         0 |        12 |        12 |
##              |     0.000 |     0.000 |     1.000 |     0.197 |
##              |     0.000 |     0.000 |     1.000 |           |
##              |     0.000 |     0.000 |     0.197 |           |
## -------------|-----------|-----------|-----------|-----------|
## Column Total |        30 |        19 |        12 |        61 |
##              |     0.492 |     0.311 |     0.197 |           |
## -------------|-----------|-----------|-----------|-----------|
##
##
```

The observations from this table tell us that the kNN method of analysis accurately predicted 61 out of 61 grape cultivars using the alcohol and ash content of the wine in question. This indicates that these wine elements vary significantly according to the grape types used in their creation. We can now run a logistic analysis to see how alcohol content relates to ash content in each grape cultivar.

```
## `geom_smooth()` using formula 'y ~ x'
```



This graph breaks down each grape type and how each affects its wine. Grape 1 leads to a higher alcohol, low ash wine with a slight negative correlation between the two. Grape 2, while producing a lower alcohol/higher ash wine than grape 1, displays a similar negative relationship. Grape 3, however, leads to varied alcohol content and high ash, and a more significant positive correlation as well.

Conclusion

Using several analysis methods we have successfully utilized the alcohol and ash content of wine to predict its constituent grape. This analysis could be expanded to the other variables in the original dataset to find similar trends.