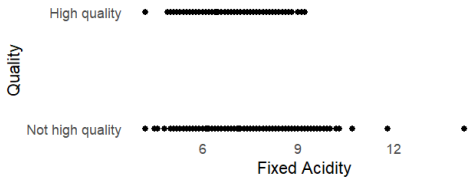

Predicting Wine Quality
Using Bayesian Approach

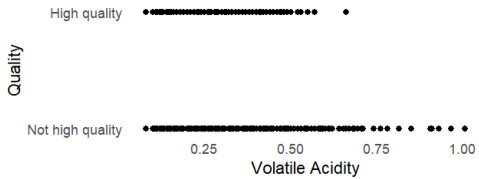
CONTENTS

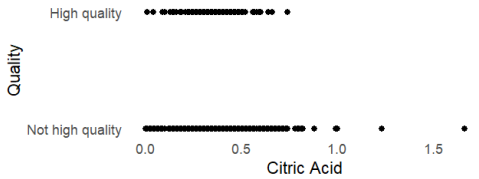
Data Descriptive

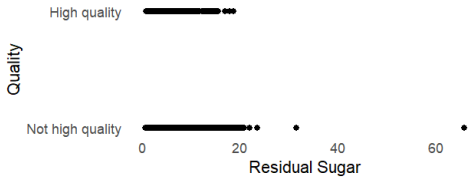
- Physicochemical and sensory data on Portuguese "Vinho Verde" white wine
  Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Wine Quality [Dataset]
- Approximately 5,000 observations, 11 independent variables and "Quality" as the dependent variable
- Wines rated 7 and above are labeled “High Quality” (1) and wines rated 6 and below are labeled “Not High Quality” (0)
- 70% of data used to train the model and 30% reserved for testing predictions of wine quality

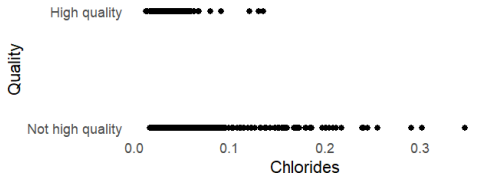
Visual Look

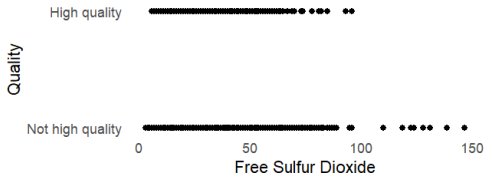

High quality
Not high quality
Quality
6
9
12
Fixed Acidity

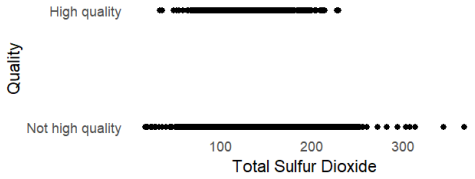

High quality
Not high quality
Quality
0.25
0.50
0.75
1.00
Volatile Acidity

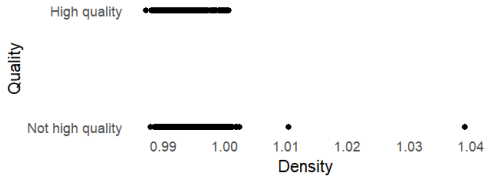

High quality
Not high quality
Quality
0.0
0.5
1.0
1.5
Citric Acid

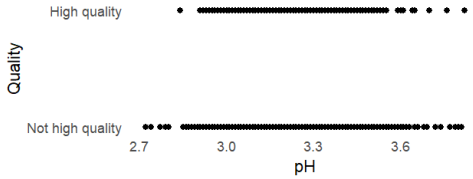

High quality
Not high quality
Quality
0
20
40
60
Residual Sugar

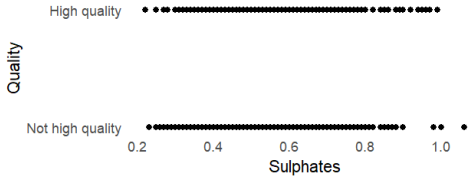

High quality
Not high quality
Quality
0.0
0.1
0.2
0.3
Chlorides

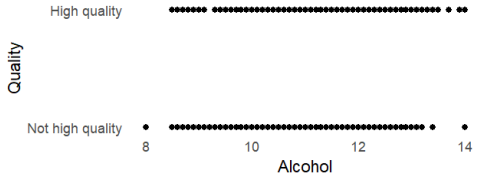

High quality
Not high quality
Quality
0
50
100
150
Free Sulfur Dioxide


High quality
Not high quality
Quality
100
200
300
Total Sulfur Dioxide

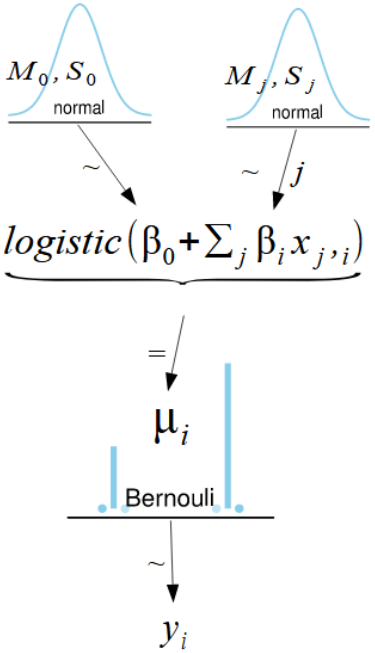

High quality
Not high quality
Quality
0.99
1.00
1.01
1.02
1.03
1.04
Density


High quality
Not high quality
Quality
2.7
3.0
3.3
3.6
pH


High quality
Not high quality
Quality
0.2
0.4
0.6
0.8
1.0
Sulphates


High quality
Not high quality
Quality
8
10
12
14
Alcohol

Model Specification



$Y \sim Bernoulli\ (\mu)$

Y = logistics (β_0 + β_1 Fixed acidity + β_2 Volatile acidity + β_3 citric acid + β_4 residual sugar + β_5 chlorides + β_6 free sulfur dioxide + β_7 total sulfur dioxide + β_8 density + β_9 pH + β_{10} sulphates + β_{11} alcohol)

Parameter Estimation

Non-informative priors

$\beta_0 \sim (0,2)$
$\beta_j \sim (0,2)$
Normal distribution with mean 0, high variance

Guess parameter:

Guess ~ (1,9)
Beta distribution ($\alpha=1$, $\beta=9$) for handling outliers, predominantly low but adaptable

Parameter Estimation

Non-informative priors

$\beta_0 \sim (0,2)$

$\beta_j \sim (0,2)$

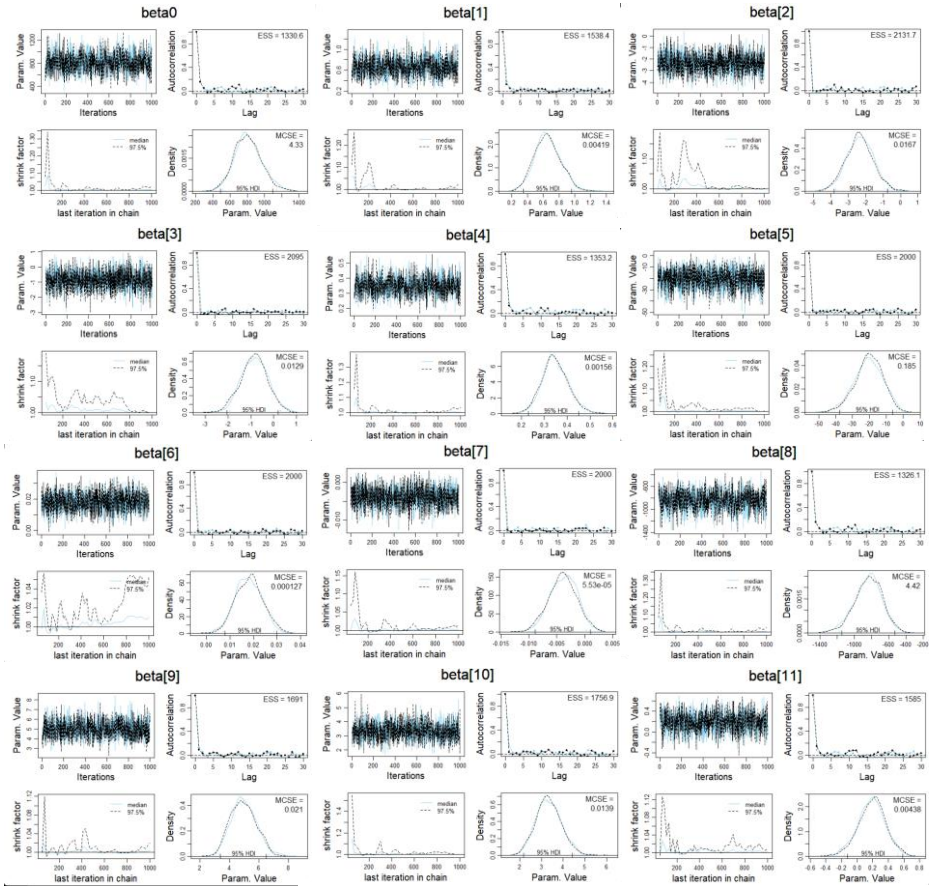
Normal distribution with mean 0, high variance

Guess parameter:

Guess ~ (1,9)

Beta distribution ($\alpha=1$, $\beta=9$) for handling outliers, predominantly low but adaptable

Diagnostics



Conclusion

```
$conf
          response
predicted   0   1
        0 994 161
        1 158 157
```

Approximately
80% correct predictions

All independent variables are significant