

sommelieR

Seoyoon Cho, Paloma Hauser, Taylor Lagler, Mike Nodzenski,
Bryce Rowland

4/24/2019

Introduction

- ▶ Wine tasting can range from a casual pastime to a lucrative profession
- ▶ For professional sommeliers, considerable time and training is required to adequately rate wine quality
- ▶ Intuitively, we might expect expert ratings to reflect the underlying chemical composition of the wines

Project Aim

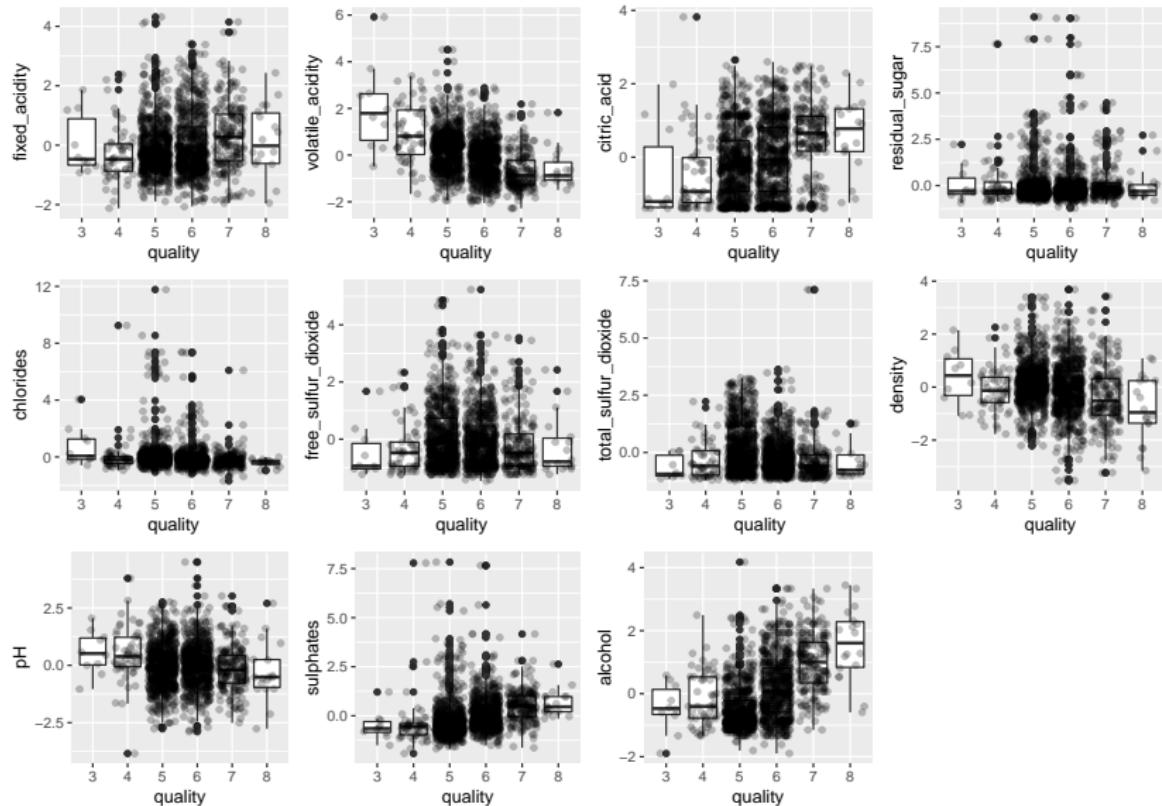
To determine how accurately expert wine quality ratings can be predicted using a set of easily measured chemical components.



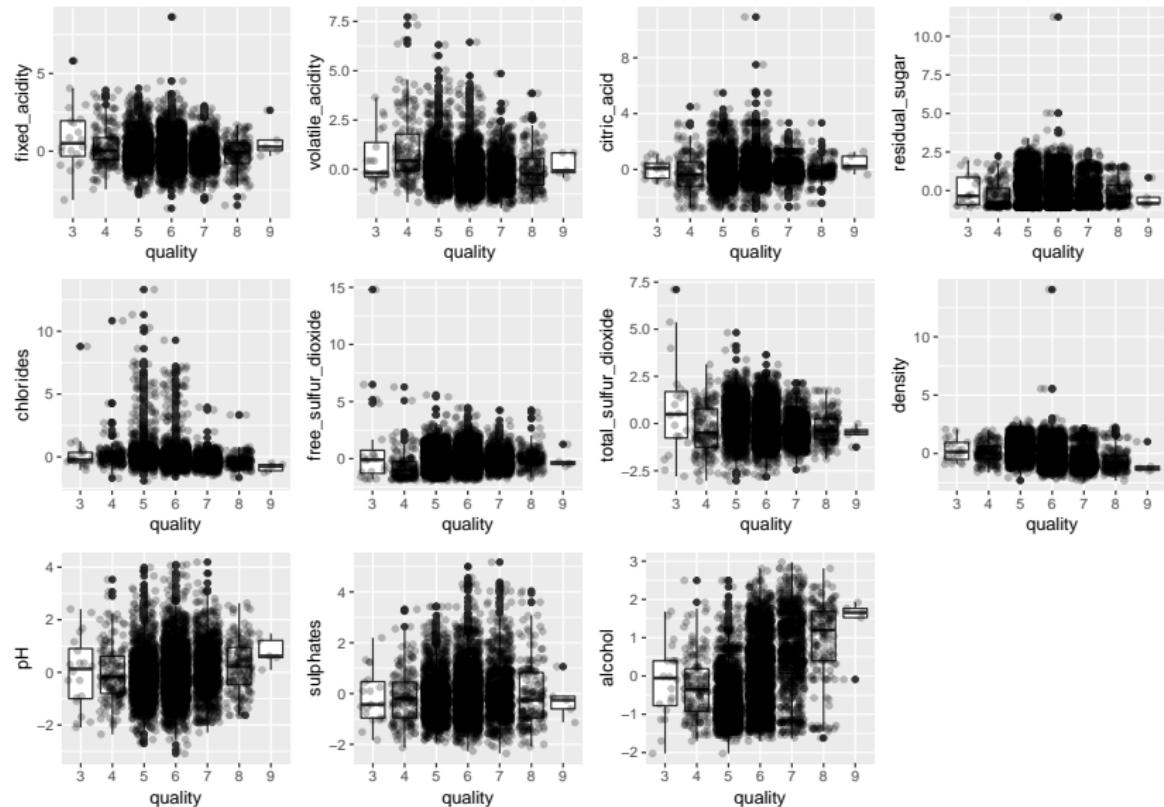
Data

- ▶ 2 datasets of expert quality ratings of red and white Vinho Verde wines
 - ▶ source: <http://archive.ics.uci.edu/ml/datasets/wine+quality>
 - ▶ 1599 red wines
 - ▶ 4898 white wines
- ▶ Outcome variable: wine quality
 - ▶ ordinal variable theoretically ranging from 0-10
 - ▶ Observed ratings range from 3-9
 - ▶ 0 is poor, 10 is excellent
 - ▶ classes are unbalanced
- ▶ Predictor variables: 11 physiochemical wine components:
 - ▶ fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol

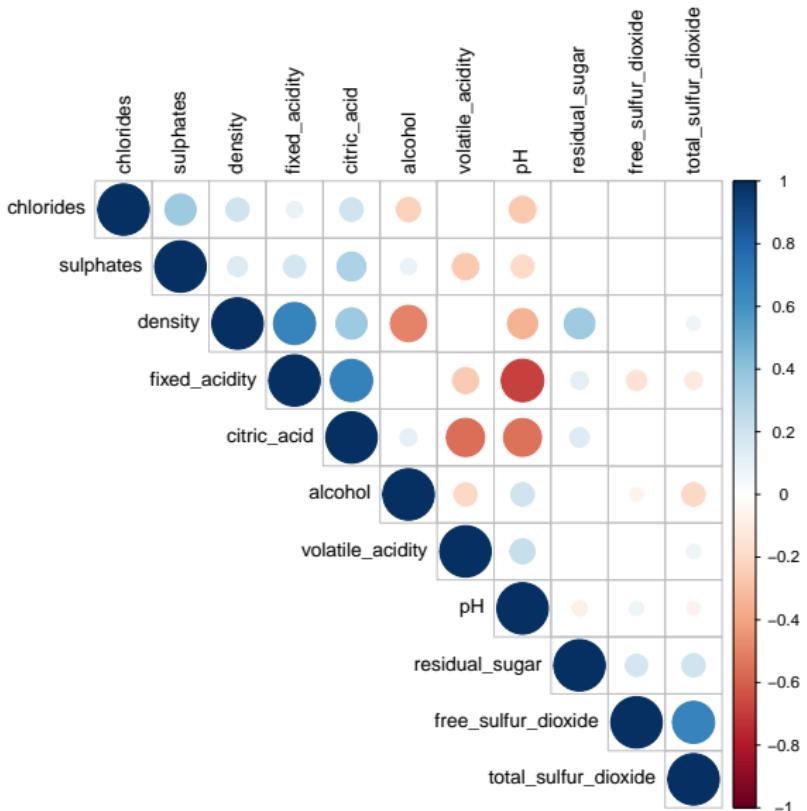
Looking At the Red Wine Data



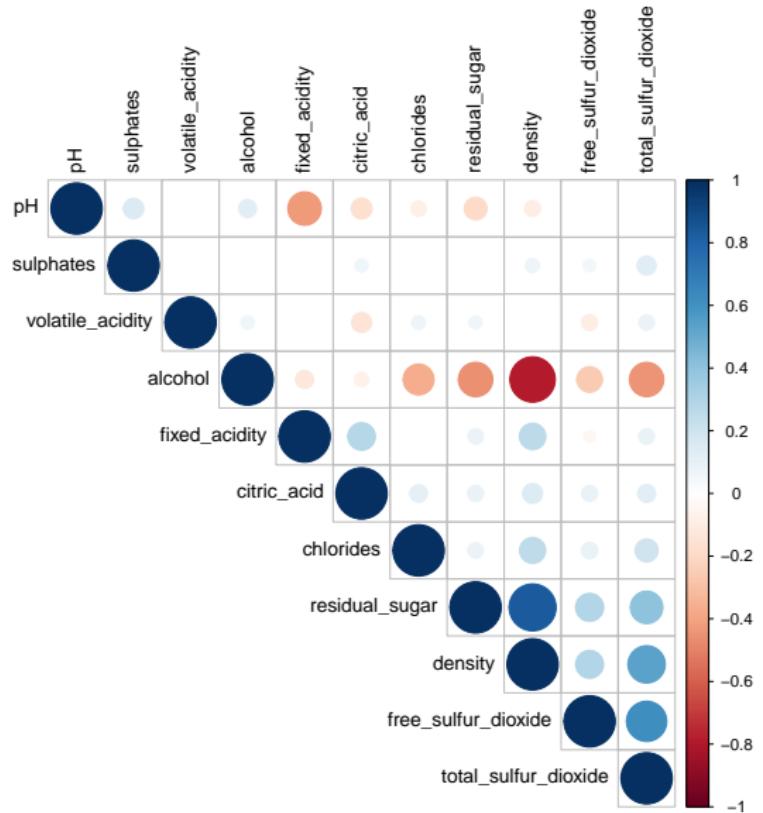
Looking At the White Wine Data



Correlations- Red Wine



Correlations- White Wine



Methods

- ▶ The red and white wine datasets were split into training and testing sets
 - ▶ 80% of available data sampled for training models
 - ▶ Remaining 20% used to evaluate accuracy
 - ▶ Relative frequencies of quality ratings in full dataset were preserved in training and testing data
- ▶ Modeling Approach
 - ▶ Machine learning
 - ▶ Random forest classification of quality ratings
 - ▶ Likelihood based
 - ▶ linear regression
 - ▶ partial proportional odds model
 - ▶ multinomial model

Methods: Linear Regression

- ▶ $Z = X^T \beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I)$
- ▶ $\hat{\beta} = (X^T X)^{-1} X^T Z$
- ▶ Predicted classification: The nearest integer to the predicted mean rating

Methods: Partial Proportional Odds Models

Three different approaches were considered:

- ▶ Non-proportional odds: $\text{logit}(P(Z \leq j|x)) = \alpha_j + x^T \beta_j$
- ▶ Proportional odds: $\text{logit}(P(Z \leq j|x)) = \alpha_j + x^T \beta$
- ▶ Partial proportional odds:
$$\text{logit}(P(Z \leq j|x)) = \alpha_j + x_*^T \beta + x_{**}^T \beta_j$$
- ▶ $\alpha_j > \alpha_i$ for $j > i$
- ▶ Log likelihood maximized using BFGS (default) or any available method in package optimx
- ▶ Predicted classification: the rating with the highest predicted probability

Methods: Multinomial Regression

- ▶ Response takes values in K classes
- ▶ Model: $\log \frac{Pr(G=k|X=x)}{P(G=K|X=x)} = X\beta_k$ for $k = 1, \dots, (K - 1)$
- ▶ Predictions:
$$P(G = k|X = x) = \frac{\exp(X\beta_k)}{1 + \sum_{l=1}^{K-1} \exp(X\beta_l)}, k = 1, \dots, K - 1$$
- ▶ Considered 3 Models: All Predictors with Linear Terms, Reduced Model with Linear Terms, and Reduced Model with All Second Order Terms
- ▶ Computational Challenges

Methods: Random Forest

- ▶ Ensembles of decision trees which are built by bootstrapping the observations
- ▶ Final prediction is made by a majority vote of the ensemble of trees

Variable Selection

- ▶ Random Forest: all physiochemical variables included
- ▶ Likelihood based models:
 - ▶ Fit the full model and hypothesized that we could find a reduced model.
 - ▶ Determined by correlations between predictors and OLS best subset.
 - ▶ Covariates for reduced model:
 - ▶ red wine: volatile acidity, total sulfur dioxide, pH, alcohol, sulphates
 - ▶ white wine: pH, volatile_acidity, residual_sugar, alcohol

Model Evaluation

We compared models on the following metrics:

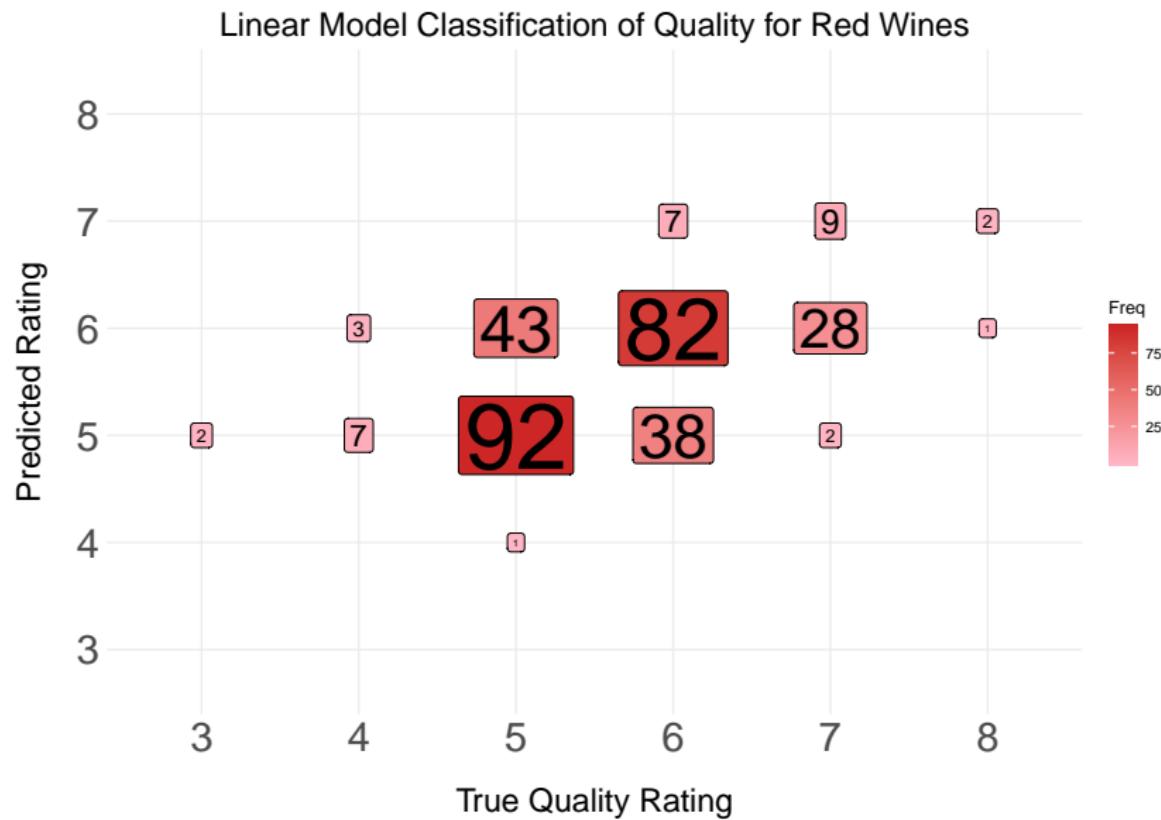
- ▶ Accuracy
 - ▶ Correct Classifications/Total
- ▶ Kappa
- ▶ Weighted Kappa - used to select the final model
 - ▶ A more useful version of Kappa for data with inherent ordering that penalizes misclassifications proportional to the distance from the true category.
 - ▶ For instance, when the true quality rating is 4, a prediction of 7 will be penalized more severely than a prediction of 5.

Results: Linear Model

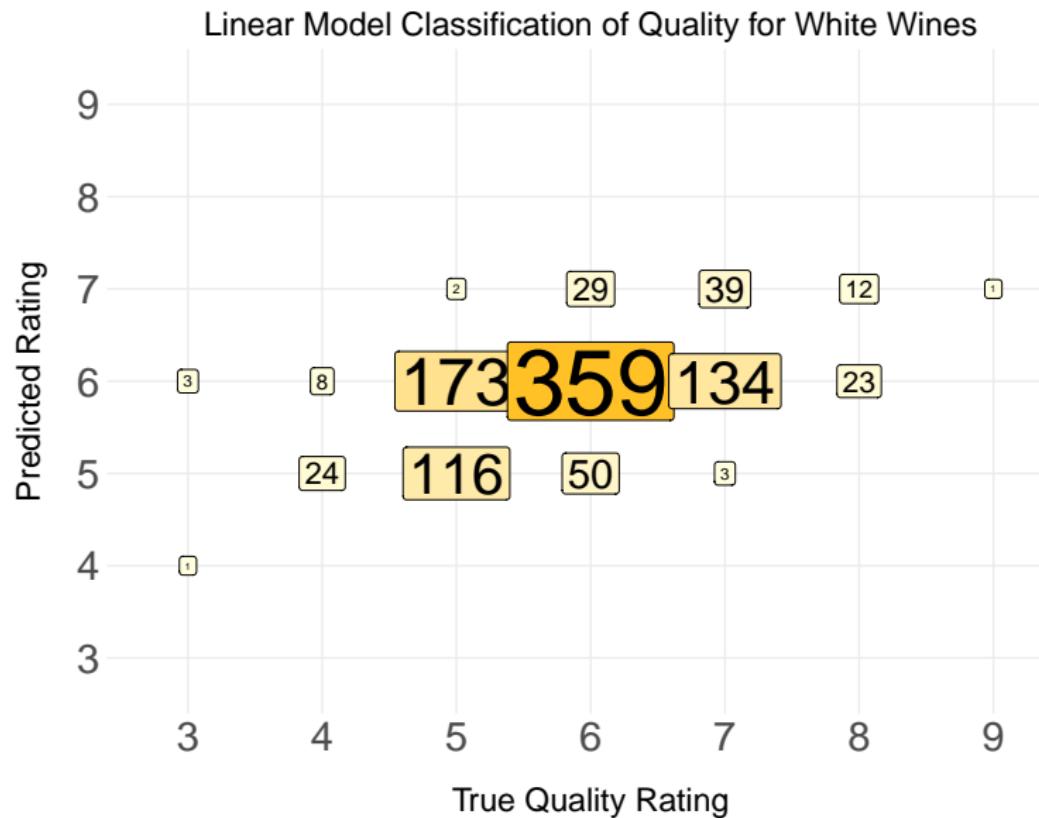
Table 1: Comparison of Full and Reduced Linear Regression Models for Red and White Wine Quality

	Overall Results			Percent Correct by Category						
	Accuracy	Kappa	Weighted Kappa	3	4	5	6	7	8	9
Full (Red)	57.73	0.2998	0.4996	0	0	67.65	64.57	23.08	0	NA
Reduced (Red)	56.15	0.2728	0.4545	0	0	65.44	63.78	20.51	0	NA
Full (White)	52.61	0.2162	0.4211	0	0	39.86	81.78	22.16	0	0
Reduced (White)	51.02	0.1975	0.3943	0	0	43.99	77.22	18.18	0	0

Results: Linear Model (Red Wine)



Results: Linear Model (White Wine)

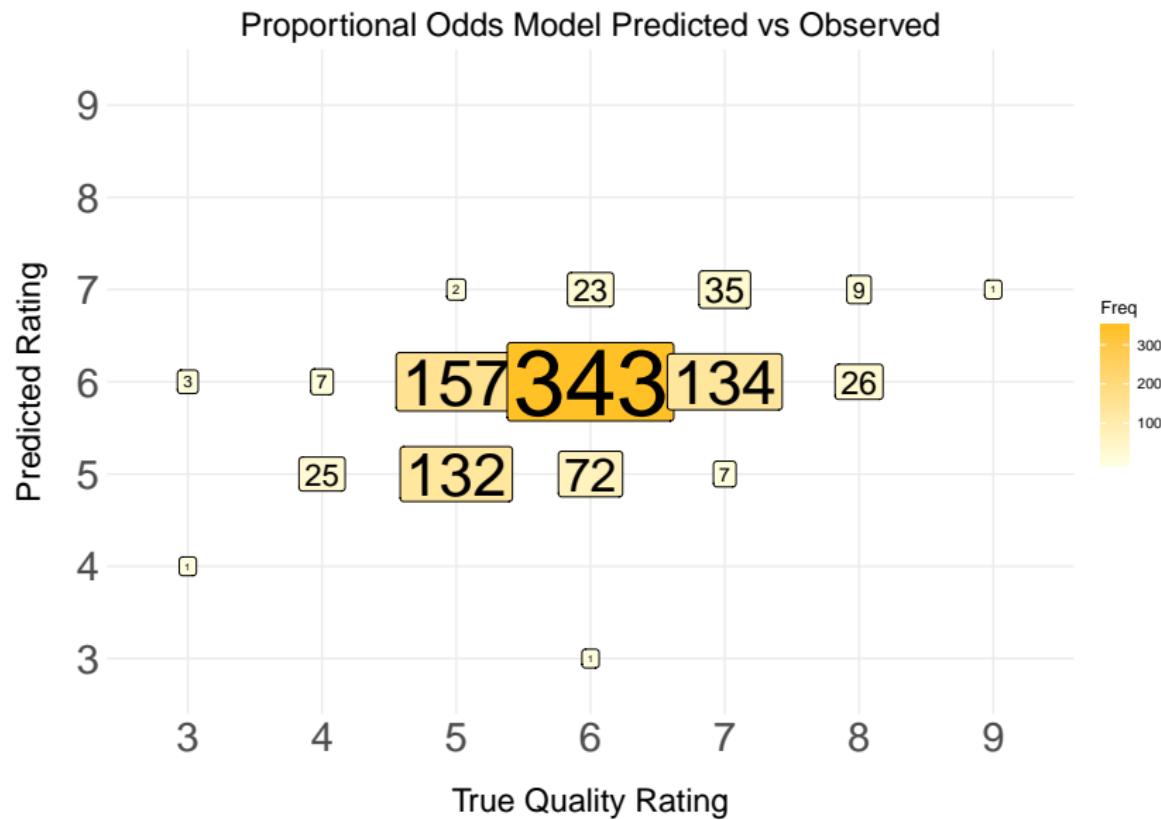


Results: Partial Proportional Odds Model (White Wine)

Table 2: Comparison of White Wine Proportional Odds Model Results

	Overall Results			Percent Correct by Category						
	Prediction Accuracy	Kappa	Weighted Kappa	3	4	5	6	7	8	9
Proportional Odds (F)	52.1472	0.2126	0.4053	0	0	45.3608	78.1321	19.8864	0	0
Proportional Odds (R)	51.7382	0.2108	0.3993	0	0	51.2027	74.9431	15.9091	0	0

Results: Proportional Odds Model (White Wine, Full)



Results: Partial Proportional Odds Model (Red Wine)

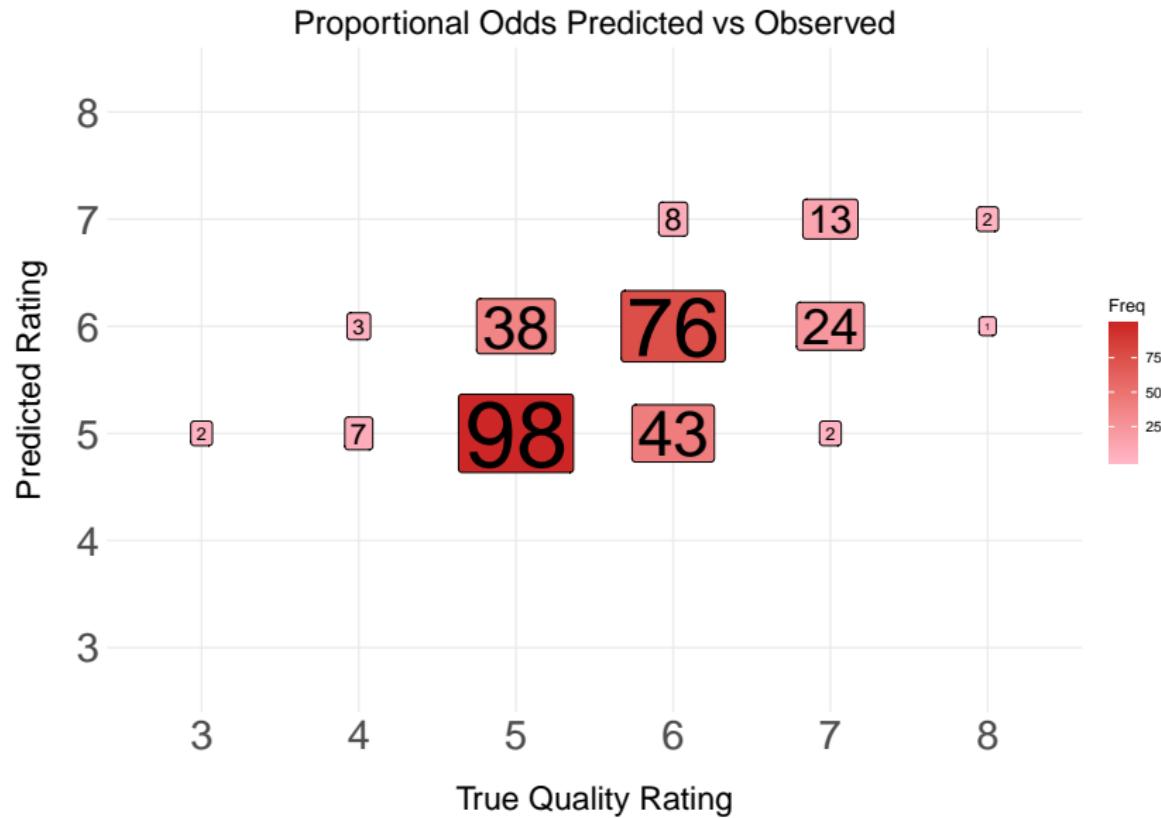
- ▶ Proportional, partial proportional models converged
- ▶ Results presented for partial proportional model with coefficient for total sulfur dioxide allowed to vary with the level of wine quality

Results: Comparison of Partial Proportional Odds Models (Red Wine)

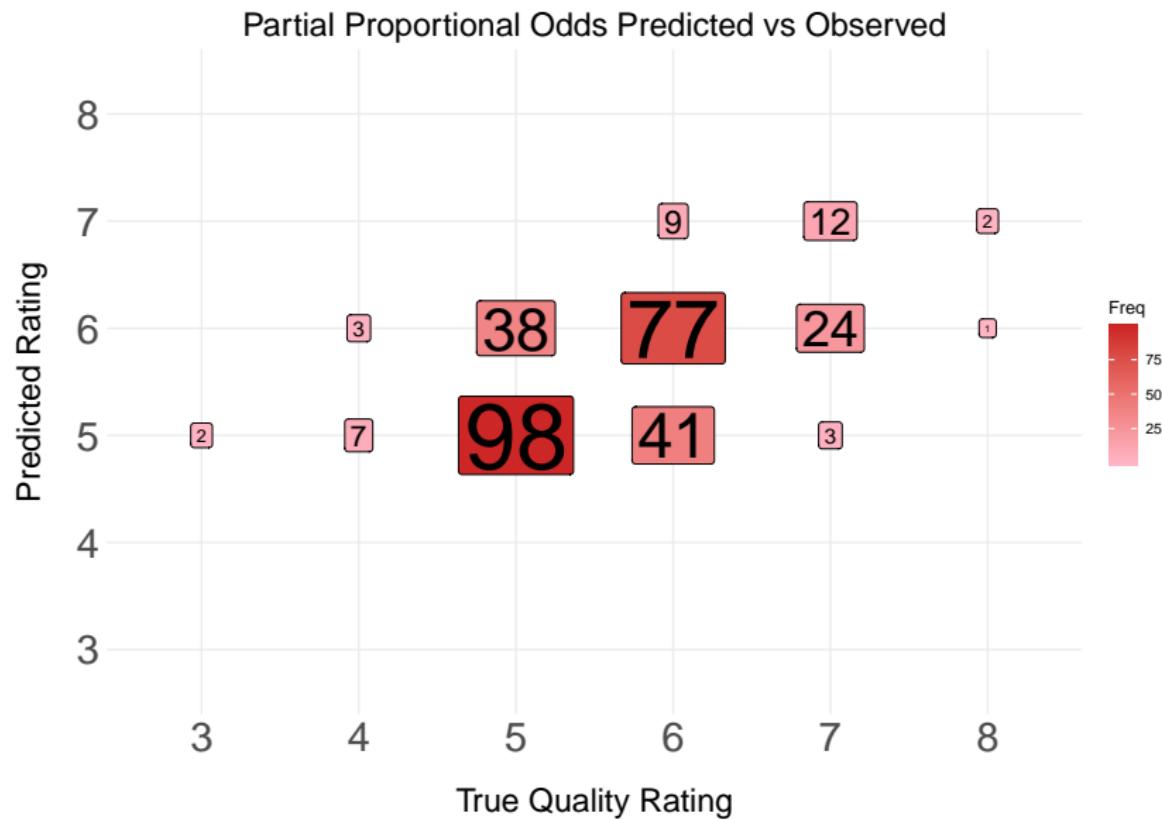
Table 3: Comparison of Red Wine Proportional Odds Model Results

	Overall Results			Percent Correct by Category					
	Prediction Accuracy	Kappa	Weighted Kappa	3	4	5	6	7	8
Proportional Odds (F)	58.9905	0.3232	0.5258	0	0	72.0588	59.8425	33.3333	0
Proportional Odds (R)	58.0442	0.3065	0.4707	0	0	72.0588	59.8425	25.6410	0
Partial Proportional Odds (F)	58.9905	0.3233	0.5162	0	0	72.0588	60.6299	30.7692	0
Partial Proportional Odds (R)	58.3596	0.3117	0.4742	0	0	72.0588	59.8425	28.2051	0

Results: Proportional Odds Model (Red Wine, Full)



Results: Partial Proportional Odds Model (Red Wine, Full)

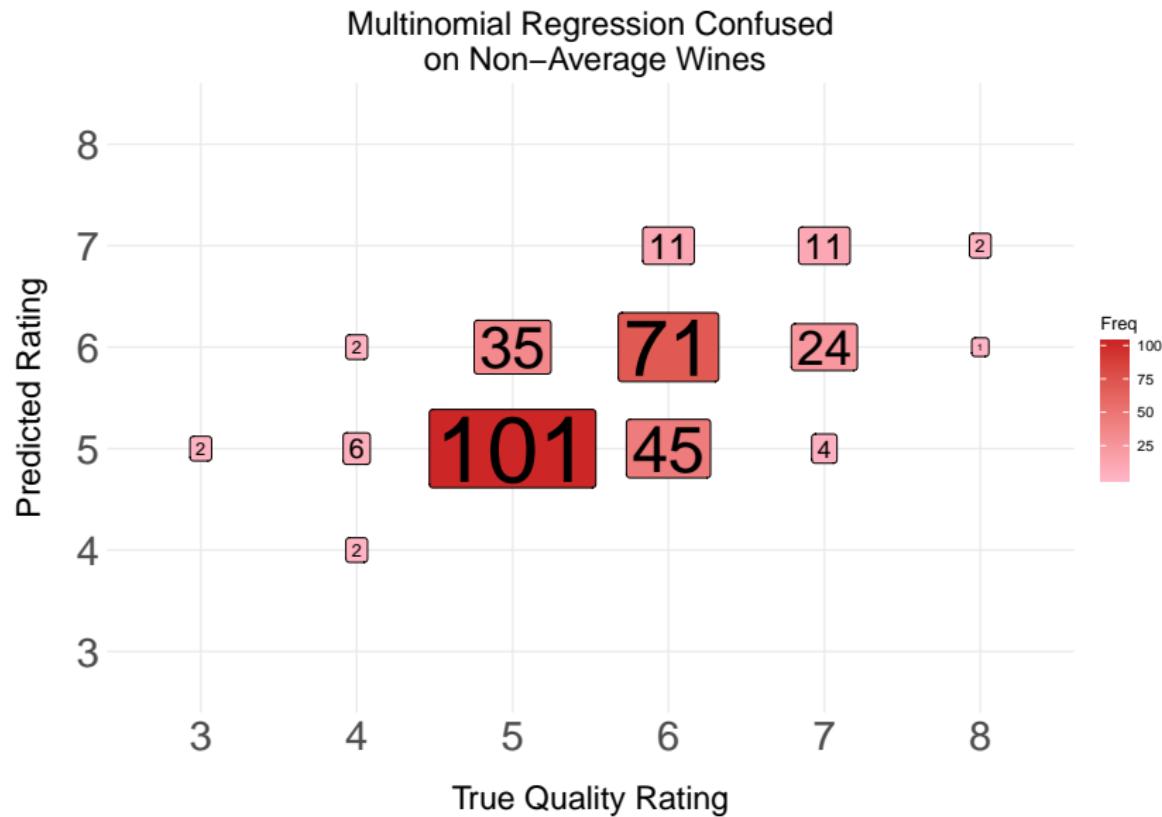


Results: Multinomial Regression (Red Wine Quality Classification)

Table 4: Comparison of Multinomial Regression Models for Red Wine Quality

Model	Accuracy	Kappa	Weighted Kappa	3	4	5	6	7	8
Full Model (Linear Terms)	58.3596	0.3158	0.5227	0	20	74.2647	55.9055	28.2051	0
Reduced Model (Linear Terms)	58.3596	0.3193	0.4952	0	10	72.7941	56.6929	33.3333	0
Reduced Model (Second Order Terms)	55.2050	0.2702	0.4789	0	0	67.6471	55.1181	33.3333	0

Results: Multinomial Regression - Full Model Confusion Matrix (Red Wine)

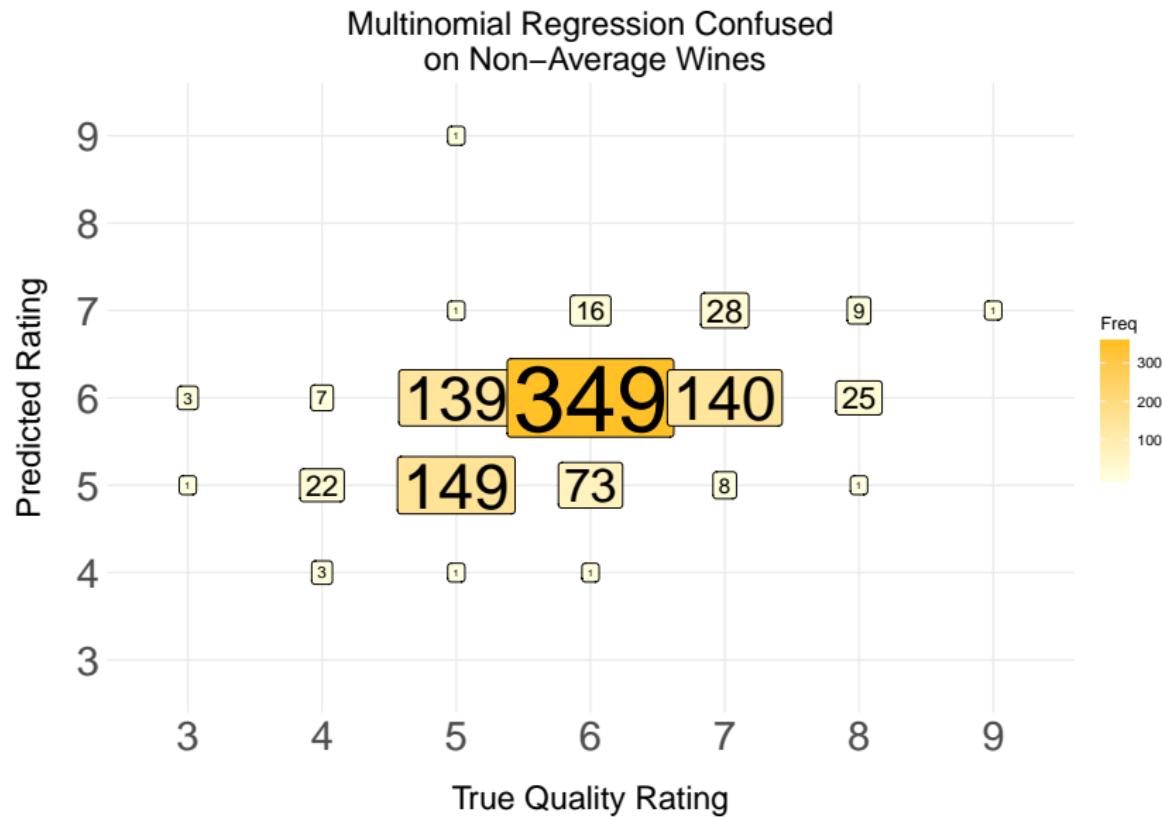


Results: Multinomial Regression (White Wine Quality Classification)

Table 5: Comparison of Multinomial Regression Models for White Wine Quality

Model	Accuracy	Kappa	Weighted Kappa	3	4	5	6	7	8	9
Full Model (Linear Terms)	54.0900	0.2451	0.4121	0	9.375	51.2027	79.4989	15.9091	0	0
Reduced Model (Linear Terms)	51.9427	0.2201	0.4093	0	3.125	54.6392	72.6651	16.4773	0	0
Reduced Model (Second Order Terms)	53.2720	0.2396	0.4010	0	3.125	54.2955	74.4875	19.8864	0	0

Results: Multinomial Regression - Full Model Confusion Matrix (White Wine)

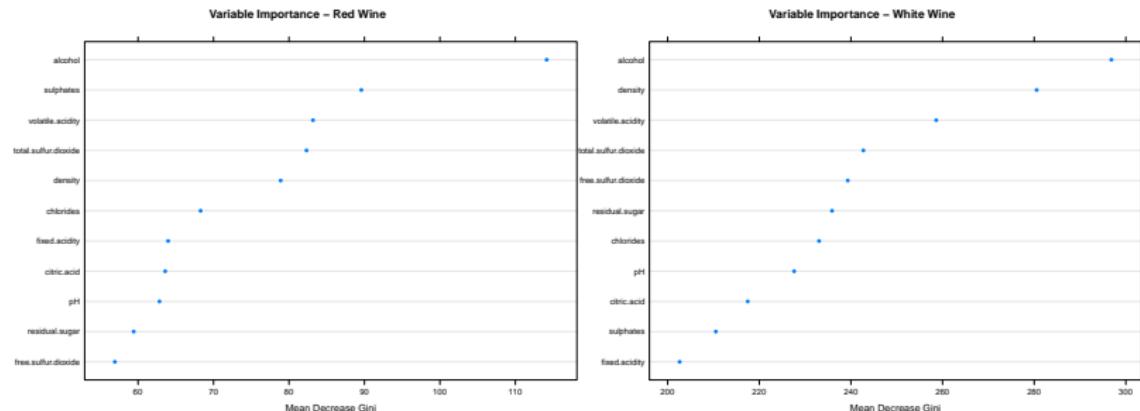


Results: Random Forest

Table 6: Random Forest Results for Red and White Wine

	Overall Results			Percent Correct by Category								
	Prediction Accuracy	Kappa	Weighted Kappa	3	4	5	6	7	8	9		
Red Wine	70.98	0.5263	0.6168	0	0	83.82	70.87	53.85	0.00	NA		
White Wine	67.28	0.4862	0.6542	0	25	67.35	80.87	47.73	42.86	0		

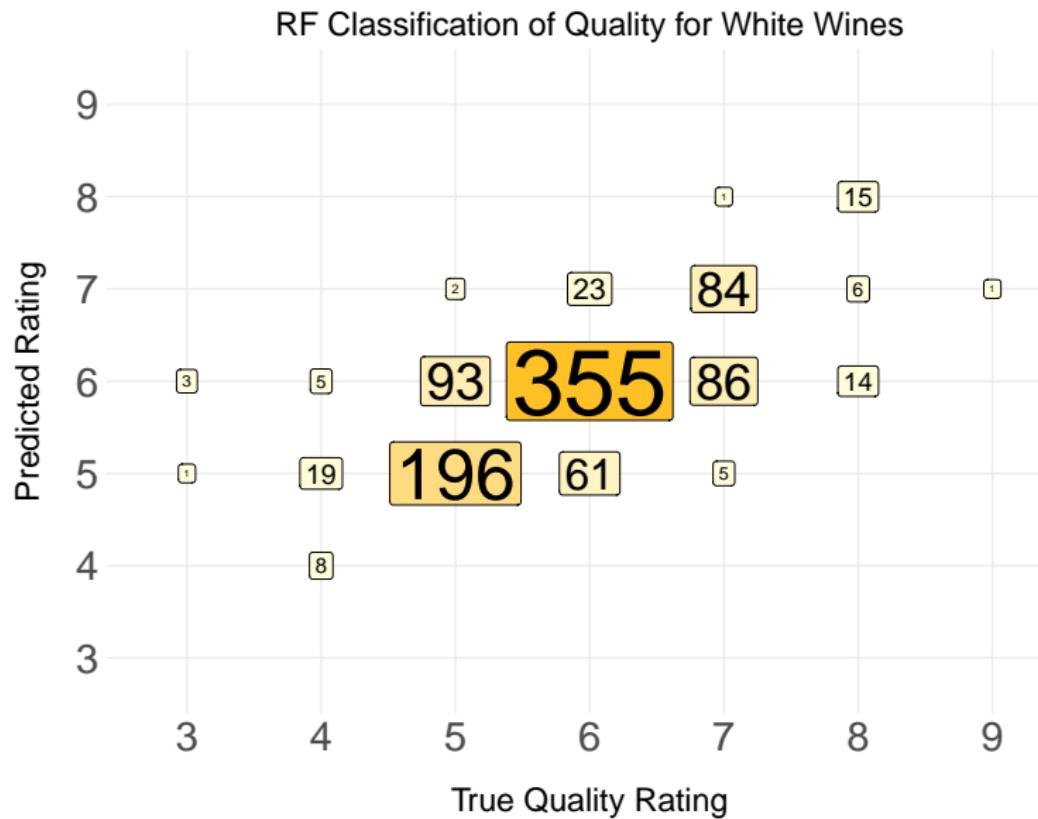
Results: Random Forest (Variable Importance)



Results: Random Forest (Red Wine)



Results: Random Forest (White Wine)



Comparison of Results: Red Wine

Table 7: Comparison of Results for Red Wine

Model	Overall Results			Percent Correct by Category							
	Prediction Accuracy	Kappa	Weighted Kappa	3	4	5	6	7	8		
Random Forest	70.9800	0.5263	0.6168	0	0	83.8200	70.8700	53.8500	0		
Proportional Odds	58.9905	0.3232	0.5258	0	0	72.0588	59.8425	33.3333	0		
Multinomial	58.3596	0.3158	0.5227	0	20	74.2647	55.9055	28.2051	0		
Partial Proportional Odds	58.9905	0.3233	0.5162	0	0	72.0588	60.6299	30.7692	0		
Linear Regression	57.7300	0.2998	0.4996	0	0	67.6500	64.5700	23.0800	0		

Comparison of Results: White Wine

Table 8: Comparison of Results for White Wine

Model	Overall Results				Percent Correct by Category						
	Prediction Accuracy	Kappa	Weighted Kappa		3	4	5	6	7	8	9
Random Forest	67.2800	0.4862	0.6542	0	25.000	67.3500	80.8700	47.7300	42.86	0	
Linear Regression	52.6100	0.2162	0.4211	0	0.000	39.8600	81.7800	22.1600	0.00	0	
Multinomial	54.0900	0.2451	0.4121	0	9.375	51.2027	79.4989	15.9091	0.00	0	
Proportional Odds	52.1472	0.2126	0.4053	0	0.000	45.3608	78.1321	19.8864	0.00	0	

Discussion: Random Forest

- ▶ Using random forests, we predicted expert wine quality ratings fairly accurately
 - ▶ White wine
 - ▶ Accuracy: 67.28%
 - ▶ Weighted Kappa: 0.6542
 - ▶ Red wine
 - ▶ Accuracy: 70.98%
 - ▶ Weighted Kappa: 0.6168
- ▶ Using the proposed cutoffs from Landis and Koch, these weighted Kappa values suggest moderate to substantial agreement with the expert ratings

Discussion: Likelihood Based Approaches

- ▶ Accuracies 10-15% lower and weighted kappas 0.1 to 0.2 lower
- ▶ No clear winner
 - ▶ However, the linear model showed a tendency to predict the mean value.
 - ▶ The multinomial model had higher accuracy for lower and higher quality wines
 - ▶ Accounting for the ordered nature of the ratings did not make a difference in predictions

Discussion: Limitations and Future Directions

- ▶ We speculate some misclassification was due to failure to measure other important chemicals.
- ▶ In the future, it would be interesting to characterize how individual chemicals relate to wine quality.
- ▶ There were only a small number of truly excellent wines (rating of 8 or higher) making it particularly difficult to predict these ratings. It would be worthwhile to examine whether various outlier detection algorithms are more suited to characterizing these wines than the approaches we took.

Bottom Line

Expert wine quality ratings can be predicted reasonably well using chemical components, but true wine connoisseurs are still better off consulting a sommelier.

References

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74.