Wine Quality Analysis Utilizing Multivariate Statistical Techniques

KONG Yunchuan, PAN Shenyi, ZENG Qinghao, ZHOU Zhihao*

May 2014

Introduction

Wine is a kind of alcoholic beverage mainly made from fermented grapes. Once viewed as a precious and luxury drink in the medieval age, wine is nowadays increasingly enjoyed by a wider range of consumers worldwide. According to the statistics from BKWineMagazine, global wine consumption has been stable around 250 million hectoliters per year since 2000. Many studies have demonstrated that moderate consumption of wine has positive medical benefits, especially in reducing the chances of cardiovascular illnesses such as heart failure.

In today's wine industry, the quality factor, which is the quality ratings received from critics through wine tasting, has become the most significant index in wine making and selling process. Usually, a high wine rating implies high wine quality, which further indicates the possible increased sales for the wine. Wine quality is not absolute due to the fact that the quality rating mainly depends on the judging critics. Nevertheless, some physical properties of wine could also be determining factors of wine quality, suggesting multivariate analysis on these properties may be used to provide some valuable insights in order to optimize the chances of success in both marketing and selling [2].

According to colors, wine can mainly be classified into two groups: red wine and white wine. Therefore, in this project, we are utilizing different multivariate techniques mentioned in *Applied multivariate statistical analysis* such as outlier detection, Principal Component Analysis, Correlation Analysis, Factor Analysis, Discriminant Analysis, and Logistic Regression, to investigate the relationship among properties of these two kinds of wine, and to perform a prediction on the wine quality given the observed properties, which is difficult to measure in a quantitative way, but is the major concern in determining the price of wine. All of our implementations of the above statistical analysis are based on R programs.

Data Description

The dataset used in this project is related to the red and white variants of the Portuguese "Vinho Verde" wine. It was first used in *Modelling wine preferences by data mining from physicochemical properties* by P. Cortez, A.

*Corresponding SIDs: 1155014473, 1155014429, 1155014454, 1155014412

Cerdeira, F. Almeida, T. Matos and J. Reis in 2009[1]. The dataset consists of 1599 red wine cases and 4898 white wine cases. For each observation, values of 12 attributes are recorded, which are fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality respectively. The first 11 attributes are input variables based on physicochemical tests, while the last 1 attribute is the output variable measuring wine quality based on sensory data in the range 1 to 10. The data summary table (Table 13), box plot (Figure 5), and density plot (Figure 6) of the different attributes in these two wine datasets are attached in the appendix. From these plots, it can be seen that the distributions of most attributes in these two datasets are skewed to the right, and also widely different from each other, which may be ascribed to the distinct physicochemical properties of these two different kinds of wine. Notwithstanding these differences, several attributes do have similar mean and close standard deviation, illustrating that a test on the equality of variance could be conducted for further investigation.

In order to test the equality of the covariance matrices between red and white wine, Box's M test can be performed to check the equality. The test results shows the p-value of the test is 0, which means that there is sufficient evidence to believe that the covariance matrices are dissimilar.

Multivariate Normality Test and Outlier Detection

In Statistics, multivariate normal distribution is a very important distribution possessing several desirable properties which would bring convenience for subsequent analysis. Therefore, given the wine data, we would like to apply multivariate normality test to check for their similarity to the multivariate normal distribution. As is shown in the following QQ chi-square plots for the two datasets, both datasets to some extent diverge from the theoretical line in their tail part. Accordingly, we may conclude that there are some evidence against the normality assumption.

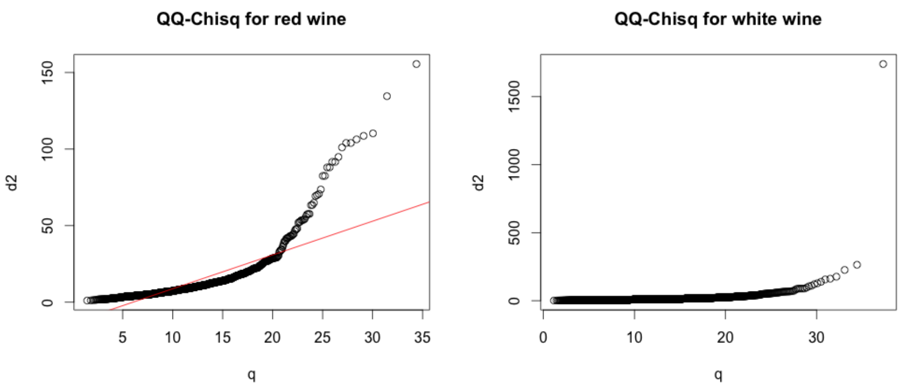

Figure 1: QQ Chi-square plot

Additionally, in multivariate analysis, some anomaly observations or outliers which do not conform to an expected pattern or other items in a dataset may lead to some analytical error and bias. To avoid these demerits, an outlier detection should be performed to clean the data and facilitate succeeding analysis. In this project, the method based on Mahalanobis Distance is used for detecting multivariate outliers.

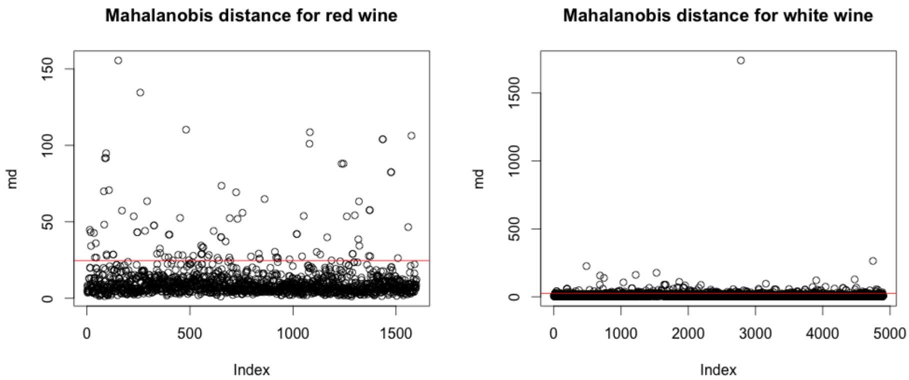

Figure 2: Mahalanobis Distance plot

From the plots above, it can be seen that a finite number of observations in each dataset are deemed to be outliers, which will be eliminated in further analysis. The detailed information of outlier detection is listed in Table 1 below.

	No. of raw observations	No. of outliers	No. of remaining observations
Red wine	1599	102	1497
White wine	4898	220	4678

Table 1: Outlier cleaning information

Partial Correlation Analysis

The heat maps for correlation matrices of these two wine datasets are attached in the appendix (Figure 7). As is seen from plots, there are some strong correlations between several attributes in both datasets. According to principles of Chemistry, we infer that fixed acidity, volatile acidity and citric acid could have certain effect on the pH, while free sulfur dioxide could influence total sulfur dioxide level. Therefore, fixed acidity, volatile acidity, citric acid, and free sulfur dioxide are partitioned to a matrix Z, while the other 8 attributes residual sugar, chlorides, total sulfur dioxide, density, pH, sulphates, alcohol, and quality are partitioned to a matrix X. The unconstrained correlation matrix of X is denoted as S11, while the partial correlation matrix of X given the fixed matrix Z is denoted as S11.2. The resulted difference matrices between correlation matrix and partial correlation matrix (namely S11-S11.2) for both wine datasets are attached in the appendix (Table 14).

For the red wine case, it can be seen from the table that the correlations of total sulfur dioxide with other attributes do not have any significant changes between correlation matrix and partial correlation matrix. However, the correlation between pH and density changes significantly from -0.3440 to 0.2527, while pH's correlations with other attributes do not vary drastically. Consequently, fixed acidity, volatile acidity, and citric acid (the fixed matrix Z) might affect the correlation between pH and density for the red wine case. In contrast, for white wine, the difference between the partial correlation and correlation is not significant, which means that the fixed acidity, volatile acidity, and citric acid (the fixed matrix Z) do not have any dramatic impacts on the correlation matrix of the other 8 attributes, namely X.

Factor Analysis

In order to further analyze the correlation structure between the attributes, we will apply factor analysis to the wine quality data, with the aim to investigate whether there are some latent factors which could affect the wine quality and to determine the most suitable number of factors.

First, we will utilize factor analysis technique for red wine data. Among all the p-values of possible models from 1-factor to 7-factor model, the largest p-value implies that the corresponding model would be the most suitable one. Nonetheless, the result shows that all the p-values of these models are very close to 0. Under this circumstance, the 7-factor model (Table 15) with the largest p-value is chosen because it might contain more information than other models. Table 2 illustrates the meanings of the factors in the 7-factor model using the correlations with the attributes.

	Positive correlation	Negative correlation	Interpretation
Factor 1	alcohol quality	density	Concentration of alcohol, which would decrease density and improve quality
Factor 2	free sulfur dioxide total sulfur dioxide		Concentration of sulfur dioxide
Factor 3	pH	fixed acidity	Alkalinity or some chemical substances with strong alkalinity
Factor 4	citric acid	volatile acidity	Amount of citric acid, which has negative impact on volatile acidity
Factor 5	sulphates quality	volatile acidity	Concentration of sulphates, which has negative impact on volatile acidity and is important to wine quality
Factor 6	fixed acidity density		Some chemical substances in red wine that is acid and its density is high
Factor 7	residual sugar density		Residual sugar that has high density

Table 2: Interpretation of factors for red wine

Similarly, for white wine, we apply the same analysis method and again find that all the p-values are close to 0, which suggests the 7-factor model, which is the model with the most factors, is chosen. However, Factor 7 in this model does not have any significant correlations with any other attributes except for sulphates, which could not give further latent information. As a result, we choose 6-factor model (Table 15) for further discussion. The following Table 3 interprets the meanings of the factors in the 6-factor model utilizing the correlations with the attributes.

	Positive correlation	Negative correlation	Interpretation
Factor 1	chlorides, density	alcohol quality	Concentration of alcohol and chlorides, this factor increases the density and downgrade the quality
Factor 2	residual sugar density		Residual sugar with high density
Factor 3	free sulfur dioxide total sulfur dioxide		Concentration of sulfur dioxide
Factor 4	fixed acid	pH	Acidity or some chemical substance with strong acidity
Factor 5	pH sulphates		Certain hydrosulphates containing HSO_4^- with alkalinity
Factor 6	volatile acidity alcohol		Substances with volatile acidity and alcohol that affect the quality

Table 3: Interpretation of factors for white wine

Principal Component Analysis (PCA)

Having observed the correlation between the 11 attributes that may potentially affect the quality of wine, heuristically we can see that some of the attributes are actually of the same type, say "free sulfur dioxide" and "total sulfur dioxide". Thus we are motivated to perform Principal Component Analysis on the data which implements data dimension reduction and facilities interpretation. Again, we performed PCA for both white wine and red wine respectively.

We start PCA with the white wine data. Table 4 gives the variance table of all the 11 principal components, and Figure 3 (left) shows the corresponding scree plot. From the table and the plot, one can easily examine that the first 3 PCs are enough to represent the whole variance of the data, since the 3 components explain 0.917, 0.072, and 0.010 of the total variance respectively, and the cumulative sum of the proportion attains 0.999, which means the other 8 PCs are actually not significant. Hence we conclude that the dimension can be reduced to three with almost all the information of the dataset preserved.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Proportion of Variance	0.917	0.0721	0.0101	5.10E-04	3.21E-04	8.97E-06
Cumulative Proportion	0.917	0.9891	0.9991	1.00E+00	1.00E+00	1.00E+00

(continued)	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11
Proportion of Variance	5.96E-06	5.06E-06	3.45E-06	6.36E-08	8.97E-11
Cumulative Proportion	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00

Table 4: Variance table of PCs for white wine

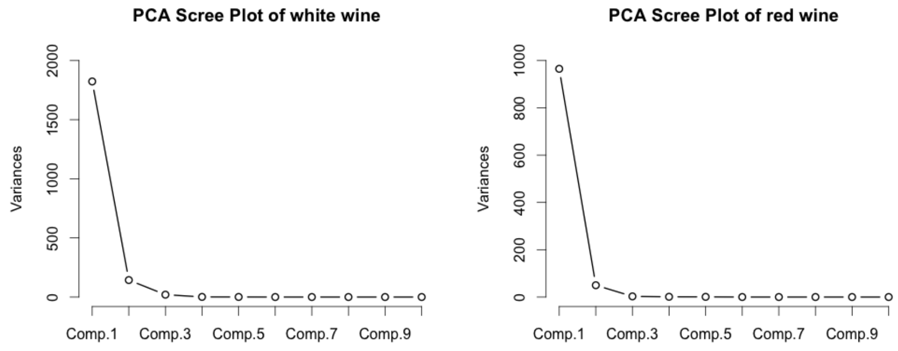

Figure 3: Scree plot of PCs

We now go to the loadings of these three principal components shown in Table 5. The first two components contain only two non-zero coefficients, namely "free sulfur dioxide" and "total sulfur dioxide". However, the coefficients of these two attributes are of the same sign in the first component while in the second component they have opposite signs. More precisely, known "free sulfur dioxide" is a chemical form of sulfur dioxide, we interpret the first component as "total effect of sulfur dioxide" since the coefficient of "total sulfur dioxide" is 0.97 which dominates. While the second component indicates "effect of the proportion of free sulfur dioxide" given the dominant coefficient 0.97 of "free sulfur dioxide" in the second PC, which is actually the contrast between free sulfur dioxide and total sulfur dioxide. Finally, the third component is solely dominated by "residual sugar", suggesting that residual sugar is another essential chemical effect in addition to sulfur dioxide.

	Comp.1	Comp.2	Comp.3
fixed.acidity			
volatile.acidity			
citric.acid			
residual.sugar			-0.993
chlorides			
free.sulfur.dioxide	0.249	0.967	
total.sulfur.dioxide	0.967	-0.251	
density			
pH			
sulphates			
alcohol			

Table 5: PC loadings of white wine

Similar to the analysis for white wine, Table 6 gives the variance table of all the 11 principal components for red wine, and Figure 3 (right) is the corresponding scree plot. Again we discovered the first three PCs are good enough to represent the whole with explaination of the total variance 0.946, 0.049, and 0.003 respectively. The cumulative sum of the proportion then gets 0.998, meaning almost all information is preserved at this stage.

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Proportion of Variance	0.946	0.049	0.0027	0.00104	0.00073	0.0000333
Cumulative Proportion	0.946	0.995	0.9982	0.99921	0.99994	0.9999688

(continued)	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11
Proportion of Variance	0.000014	0.0000087	7.97E-06	5.12E-07	4.38E-10
Cumulative Proportion	0.999983	0.9999915	1.00E+00	1.00E+00	1.00E+00

Table 6: Variance table of PCs for red wine

As shown in Table 7 , the loadings of the first two components are very similar to those of white wine, and thus we conclude the same interpretation for these two loadings in the red wine case. However, in the third component, there are two non-zero coefficients: "residual sugar", which solely appears in the third PC of white wine, and "fixed acidity". Noting that the signs of the two coefficients are identical, this PC could be explained as the total effect on red wine quality under fixed acidity and residual sugar, though the fixed acidity dominates this component obviously with a coefficient of 0.985.

	Comp.1	Comp.2	Comp.3
fixed.acidity			0.985
volatile.acidity			
citric.acid			
residual.sugar			0.113
chlorides			
free.sulfur.dioxide	-0.222	0.974	
total.sulfur.dioxide	-0.975	-0.221	
density			
pH			
sulphates			
alcohol			

Table 7: PC loadings of red wine

Discriminant Analysis

Linear Discriminant Analysis (LDA)

In order to further investigate the wine quality, we aim to find a decision or classification rule, which is a function of wine attributes, in order to facilitate the prediction of wine quality. In this section, Linear Discriminant Analysis, which consists of discriminant functions that appear based on a linear combination of predictive variables providing the best discrimination between groups, will be used to predict binary wine quality based on the first 11 input attributes. The response, wine quality, will be divided into dichotomous groups. Observations with quality lower than 6 will be classified as low quality (labeled as 0) while the remaining ones with quality higher than or equal to 6 will fall into high quality class (labeled as 1). The predictors include all 11 attributes in the cleaned data set.

	Number of observations in red wine	Number of observations in white wine
Total	1497	4678
High quality	804	3152
Low quality	693	1526

Table 8: Summary of wine types

Assumptions for LDA

Before implementing the model, it is crucial to check whether the corresponding assumptions hold. Generally, discriminant analysis is quite sensitive to outliers and the size of the smallest group must be larger than the number of predictor variables. Since we have deleted outliers previously and the numbers of observations within both quality groups for two wine types are much larger than 11, we may safely assert the accordance with these two assumptions. Following are two more complex assumptions.

1. Multivariate normality: Referring to previous analysis, the normality assumption does not hold well for neither red wine nor white wine. Since it has been suggested that discriminant analysis is relatively robust to slight violations of these assumptions and it has also been shown that discriminant analysis may still be reliable when applying to dichotomous variables, where multivariate normality is often violated[4]. Thus we suspect this assumption can be somehow relaxed in this case and we will perform discriminant analysis to examine its effect.

2. Homogeneity of covariance: Linear discriminant analysis assumes homogeneity of variance-covariance matrices. Here we employed Box's M test for two wines, with

$$H_0 : \Sigma_{low\ quality} = \Sigma_{high\ quality}$$
$$H_1 : \Sigma_{low\ quality} \neq \Sigma_{high\ quality}$$

The resulting p-value is nearly 1 under both case. Thus we accept the null hypothesis stating that the low and high quality wine groups have homogeneous covariance structure, which holds in both red and white wine.

Model building and testing for LDA

We first directly perform linear discriminant analysis separately to each wine type with proportional prior probabilities incorporated. From the output, the resulting LDA model offers around 75.55% and 75.39% classification accuracy rate for red and wine correspondingly. In order to gain more accurate assessment of the predicting capacity of the LDA model, we further integrate the cross-validation technique. Each training data is treated as the test data and excluded from training data. The final classification accuracy rates are quite similar to previous case with rate 75.28% and 75.33%, as shown in Table 9, indicating the reliability of the LDA model in predicting wine quality class. Moreover, we may conclude the violations of the normality assumption are not fatal in this case.

LDA for red wine with cross validation

Prediction\True	0	1
0	518	195
1	175	609

LDA for white wine with cross validation

Prediction\True	0	1
0	739	367
1	787	2785

Table 9: Classification table with cross validation

Canonical Discriminant Analysis (CDA)

Canonical discriminant analysis is a dimension-reduction technique related to canonical correlation and principal component analysis, which derives canonical variables (linear combinations of all the predictors) given a classification variable and several interval variables that summarize between-class variation in much the same way that principal components summarize total variation. After performing the CDA on red wine and white wine, as illustrated in Figure 4 and Table 10, we obtain a pair of slight lower classification accuracy rates, which are around 75.01% and 72.08%. Considering the massive reduction in dimension, these minor lose in classification accuracy is acceptable. And this result is also consistent with the results from principle component analysis, which shows the first few loadings massively capture the variation within the data.

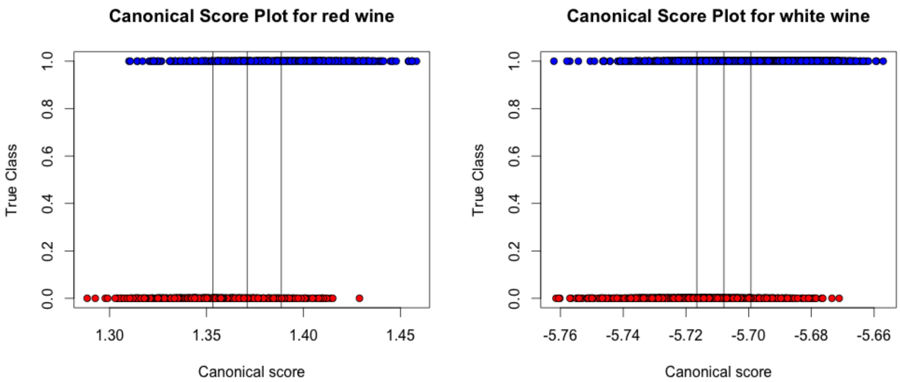

Figure 4: Canonical score plots

CDA for red wine

Prediction\True	0	1
0	540	221
1	153	583

CDA for white wine

Prediction\True	0	1
0	1126	906
1	400	2246

Table 10: Classification table for CDA

Logistic Regression

In the previous section, discriminant analysis was employed for the purpose of wine quality classification and prediction with some promising outcomes. As an extension, we are using logistic regression in this section for the same purpose. There are 11 attributes that may be useful to predict the wine quality. Again as we defined previously, observations with quality lower than 6 will be classified as low quality, while those with quality higher than or equal to 6 will fall into high quality class. For both the datasets of white and red wine, the first 80% of the observations are used for model building as training dataset while the latter 20% are retained for quality prediction and accuracy checking as testing dataset.

For each dataset, firstly a full model regressing quality on all the 11 attributes is fitted. Then we perform the stepwise variable selection based on AIC to find the optimal regression model. The coefficients of final models for the white and red wine are summarized in Table 11.

Coefficients for white wine:

(Intercept)	fixed acidity	volatile acidity	residual sugar	free sulfur dioxide
425.5879	0.2106	-5.7947	0.2151	0.0157
	density	pH	sulphates	alcohol
	-442.7833	2.1536	2.5016	0.537

Coefficients for red wine:

(Intercept)	fixed acidity	volatile acidity	citric acid	residual sugar
-10.72224	0.23195	-3.47426	-2.19293	-0.14151
	free.sulfur.dioxide	total.sulfur.dioxide	sulphates	alcohol
	0.02547	-0.02128	3.4801	0.97526

Table 11: Logsitic regression models

Using the above logistic regression, we are then able to perform wine quality prediction with the testing data. The prediction results are listed in Table 12. From the outcomes, the prediction accuracy rate for white and red wine are 74.79% and 61% respectively, suggesting that logistic regression model is also dependable in predicting wine quality class.

White wine	Predicted	
True value	0	1
0	116	159
1	77	584

Red wine	Predicted	
True value	0	1
0	144	6
1	111	39

Table 12: Results of prediction of logistic regression

Conclusion

In this project, we utilized different multivariate techniques to analyze the red and wine quality data. Through Partial Correlation Analysis and Factor Analysis, we analyzed the covariance structure of the attributes in each dataset and obtained some latent factors which might better model the covariance structure. By performing Principal Component Analysis, we successfully achieved data dimension reduction and found some principal components to facilitate the data interpretation. Additionally, Discriminant Analysis and Logistic Regression were also employed for the class prediction of wine quality. The relatively high prediction accuracy rate indicates that as an alternative to the quality ratings judged by wine critics, some physicochemical properties of wine can also be used for quality prediction with favorable outcomes. Moreover, further studies might be conducted to improve the accuracy rate of quality prediction.

Acknowledgments

The work for this project is evenly allocated among the four group members, we are grateful for the instructions of Professor Leung Pui-Lam and the assistance from Mr. Li Yuanbo and Mr. Chen Kun.

References

[1] K. Bache and M. Lichman. UCI machine learning repository, 2013.

[2] Paulo Cortez, Antonio Cerdeira, Fernando Almeida, Telmo Matos, and Jose Reis. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547 – 553, 2009. Smart Business Networks: Concepts and Empirical Evidence.

[3] Richard Arnold Johnson and Dean W Wichern. *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 2002.

[4] William R Klecka. *Discriminant analysis*. Number 19. Sage, 1980.

Appendix

	Red Wine				White Wine			
Attribute	Min	Mean	Max	s.d.	Min	Mean	Max	s.d.
fixed acidity	4.6	8.3196	15.9	1.7411	3.8	6.8548	14.2	0.8439
volatile acidity	0.12	0.5278	1.58	0.1791	0.08	0.2782	1.1	0.1008
citric acid	0	0.271	1	0.1948	0	0.3342	1.66	0.121
residual sugar	0.9	2.5388	15.5	1.4099	0.6	6.3914	65.8	5.0721
Chlorides	0.012	0.0875	0.611	0.0471	0.009	0.0458	0.346	0.0218
free sulfur dioxide	1	15.8749	72	10.4602	2	35.3081	289	17.0071
total sulfur dioxide	6	46.4678	289	32.8953	9	138.3607	440	42.4981
density	0.99007	0.9967	1.00369	0.0019	0.98711	0.994	1.03898	0.003
pH	2.74	3.3111	4.01	0.1544	2.72	3.1883	3.82	0.151
sulphates	0.33	0.6581	2	0.1695	0.22	0.4898	1.08	0.1141
alcohol	8.4	10.423	14.9	1.0657	8	10.5143	14.2	1.2306

Table 13: Data summary

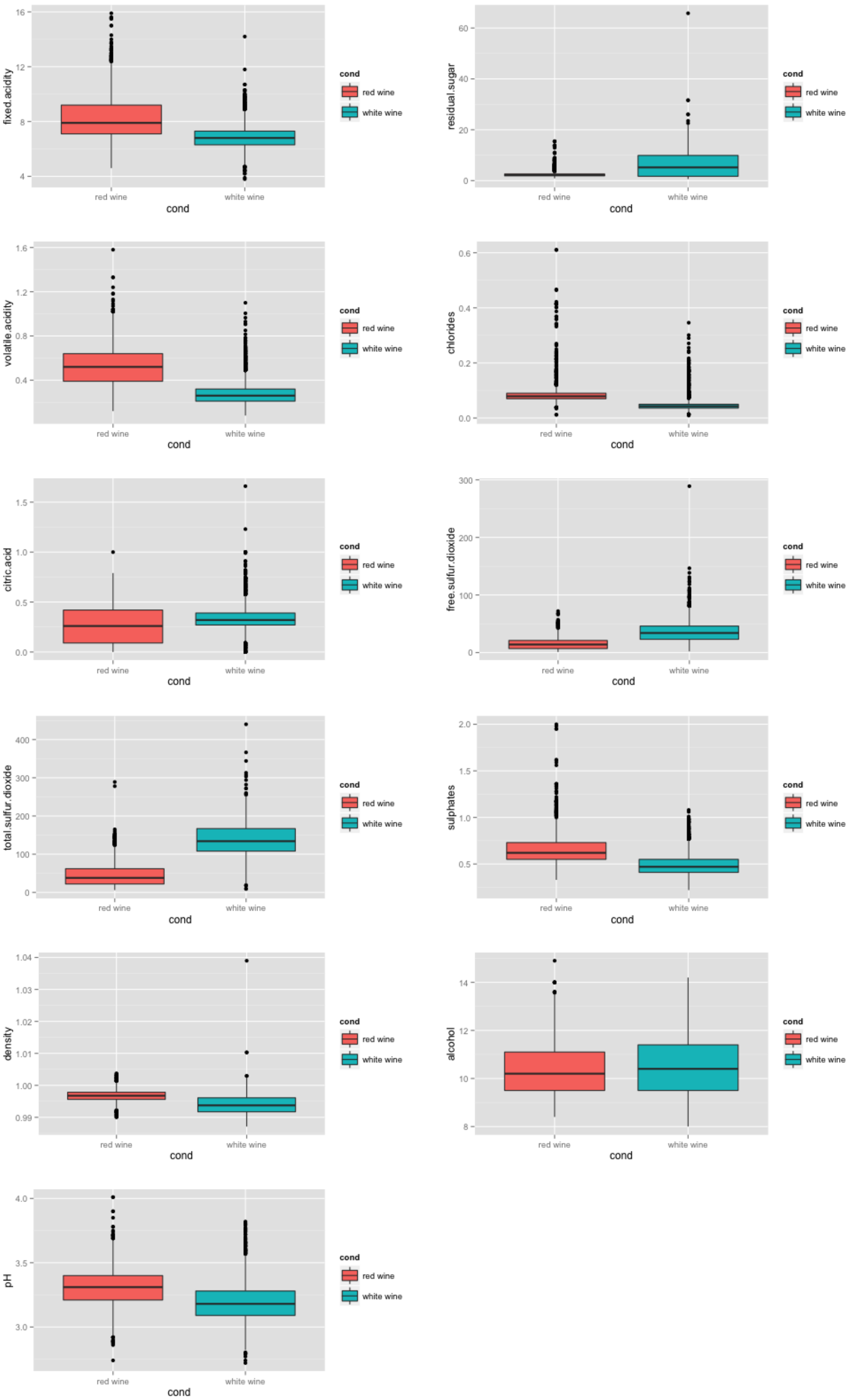

Figure 5: Box plot for raw dataset

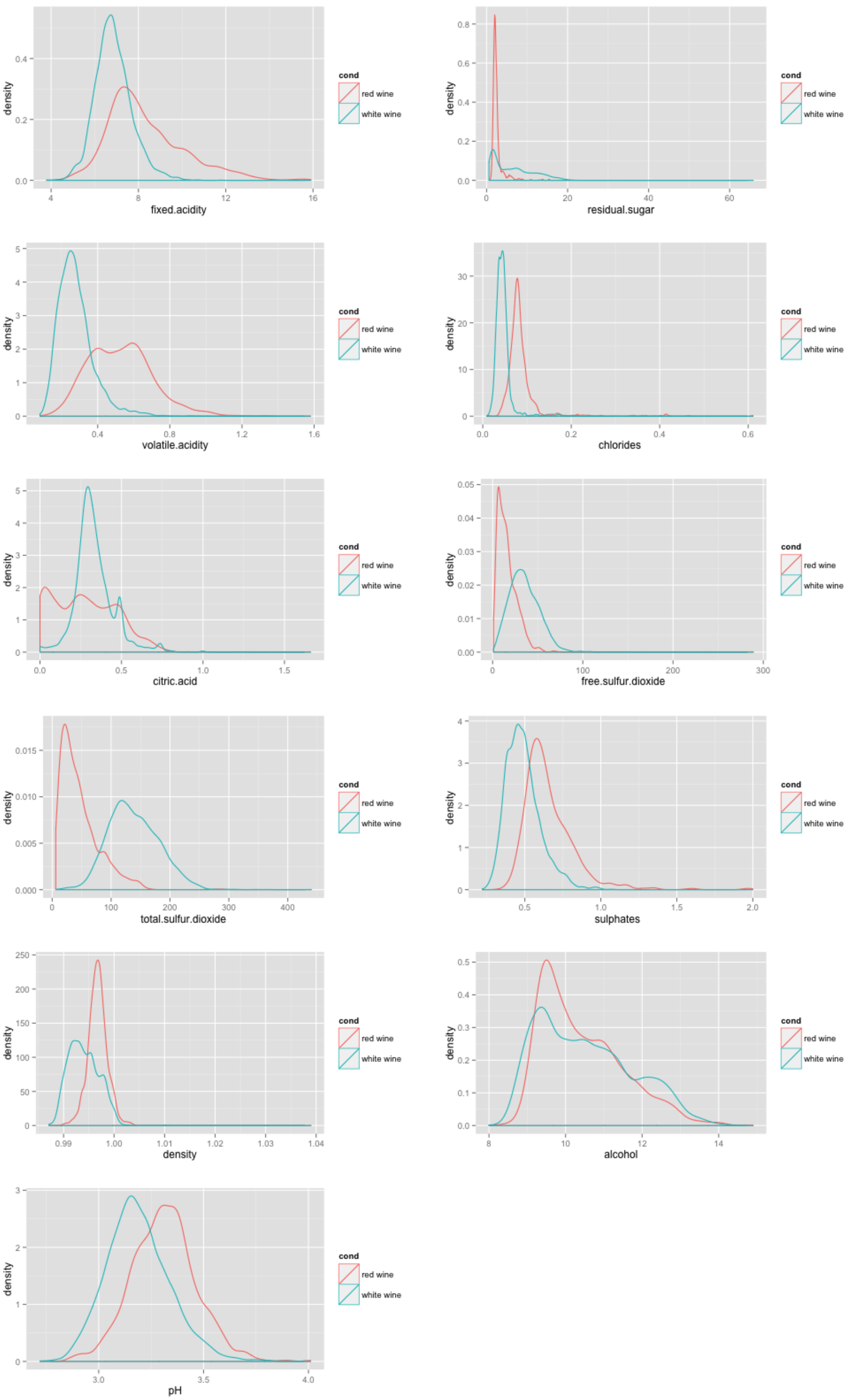

Figure 6: Density plot for raw dataset

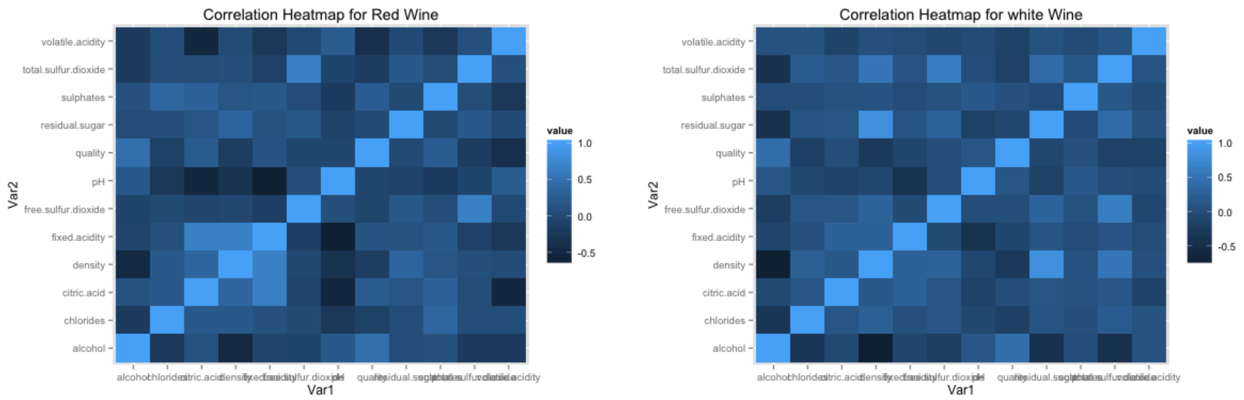

Figure 7: Correlation heat maps

Red wine:

	residual sugar	chlorides	total sulfur dioxide	density	pH	sulphates	alcohol	quality
residual sugar	0	0.047	0.062	0.021	-0.113	0.01	-0.035	-0.029
chlorides	0.047	0	-0.019	0.058	-0.07	-0.017	-0.021	-0.021
total sulfur dioxide	0.062	-0.019	0	-0.112	0.15	0.04	0.043	-0.003
density	0.021	0.058	-0.112	0	-0.597	0.052	0.087	0.104
pH	-0.113	-0.07	0.15	-0.597	0	-0.211	-0.032	-0.12
sulphates	0.01	-0.017	0.04	0.052	-0.211	0	0.043	0.089
alcohol	-0.035	-0.021	0.043	0.087	-0.032	0.043	0	0.03
quality	-0.029	-0.021	-0.003	0.104	-0.12	0.089	0.03	0

White wine:

	residual sugar	chlorides	total sulfur dioxide	density	pH	sulphates	alcohol	quality
residual sugar	0	0.039	0.17	0.011	-0.02	0.023	-0.051	-0.006
chlorides	0.039	0	0.019	0.016	-0.027	0.004	-0.013	-0.002
total sulfur dioxide	0.17	0.019	0	0.105	-0.092	0	-0.059	0.029
density	0.011	0.016	0.105	0	-0.127	0.009	-0.008	0.015
pH	-0.02	-0.027	-0.092	-0.127	0	-0.01	0.043	0.044
sulphates	0.023	0.004	0	0.009	-0.01	0	-0.015	0.011
alcohol	-0.051	-0.013	-0.059	-0.008	0.043	-0.015	0	-0.032
quality	-0.006	-0.002	0.029	0.015	0.044	0.011	-0.032	0

Table 14: Difference between correlation matrices and partial correlation matrices (S11-S11.2)

7-factor model for red wine

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
fixed acidity	-0.089	-0.12	-0.49	0.329	0.162	0.763	0.123
volatile acidity	-0.057	0.039	0.116	-0.515	-0.519	0.012	0.144
citric acid	0.004	-0.028	-0.28	0.854	0.27	0.293	0.164
residual sugar	0.024	0.075	-0.014	0.013	0.026	0.069	0.642
chlorides	-0.234	0.003	-0.126	0.036	-0.043	0.045	0.227
free sulfur dioxide	0.006	0.77	0.061	-0.072	0.07	-0.022	-0.009
total sulfur dioxide	-0.139	0.89	-0.028	0.045	-0.133	-0.038	0.132
density	-0.611	-0.005	-0.08	0.103	0.059	0.611	0.477
pH	0.143	0.032	0.937	-0.221	-0.042	-0.197	-0.076
sulphates	0.043	0.046	0.022	0.129	0.607	0.095	0.072
alcohol	0.912	-0.13	0.094	0.073	0.301	-0.058	0.103
quality	0.359	-0.113	-0.069	0.073	0.571	-0.003	-0.039

6-factor model for white wine

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
fixed acidity	0.112	0.011	-0.068	0.975	-0.161	0.035
volatile acidity	0.064	0.015	-0.005	-0.091	-0.072	0.593
citric acid	-0.011	0.067	0.114	0.313	0	-0.147
residual sugar	0.186	0.934	0.258	0.045	-0.133	0.039
chlorides	0.492	0.103	0.129	0.021	0.084	0.044
free sulfur dioxide	0.044	0.191	0.668	0.01	0.021	-0.14
total sulfur dioxide	0.308	0.148	0.898	0.128	0.156	0.177
density	0.607	0.71	0.256	0.221	0.086	-0.049
pH	-0.055	-0.072	-0.047	-0.298	0.758	-0.028
sulphates	0.044	-0.004	0.077	0.04	0.355	-0.04
alcohol	-0.904	-0.282	-0.209	-0.026	0.035	0.228
quality	-0.545	0.002	0.038	-0.021	0.041	-0.213

Table 15: Factor loadings