Exercise 1

a) Descriptive statistics and normality checks for magnesium values follow. The mean and median values are very close, with the median (50% point) being 19.5 and the mean being just slightly below that. There is a slight positive skew of .213 indicating a small tail to the right, but we should still be able to use the standard deviation as a measure of spread here. The standard deviation of 3.34 indicates about 95% of magnesium values would be expected to fall between about 12.8 and 26.2 if the data is roughly normal.

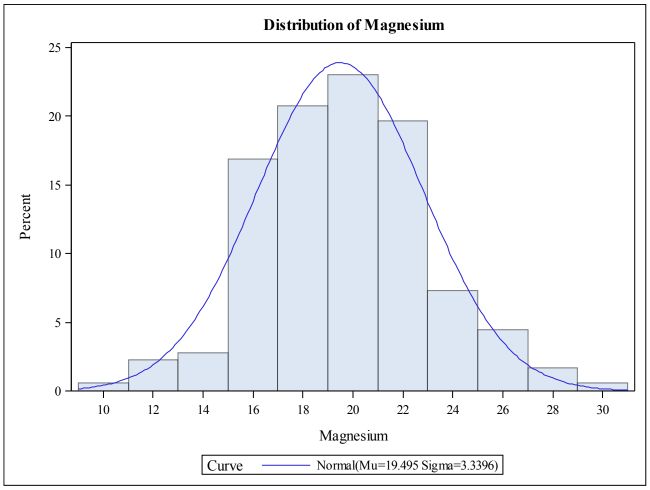
Variable: Magnesium

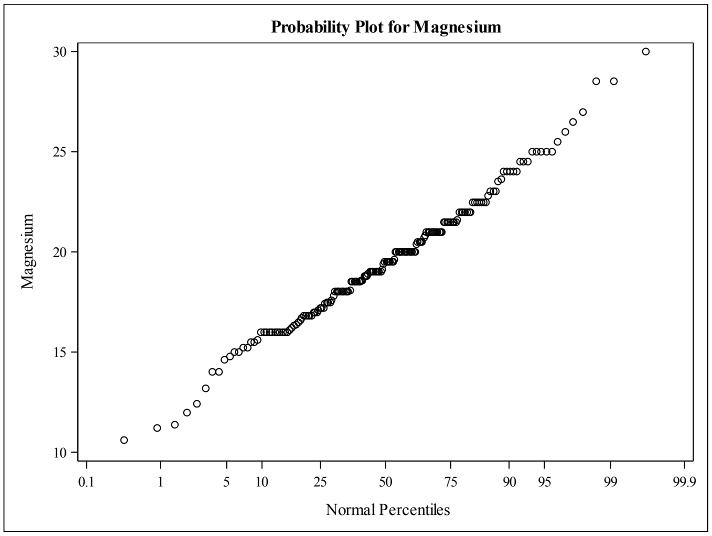
Moments			
N	178	**Sum Weights**	178
Mean	19.4949438	**Sum Observations**	3470.1
Std Deviation	3.33956377	**Variance**	11.1526862
Skewness	0.21304689	**Kurtosis**	0.48794154
Uncorrected SS	69623.43	**Corrected SS**	1974.02545
Coeff Variation	17.1304098	**Std Error Mean**	0.25031089

Basic Statistical Measures			
Location		**Variability**	
Mean	19.49494	**Std Deviation**	3.33956
Median	19.50000	**Variance**	11.15269
Mode	20.00000	**Range**	19.40000
		Interquartile Range	4.30000

From the following normality tests, we see no strong evidence that the magnesium values are far from normal. The Shapiro-Wilk test, specific to normality, is insignificant at a .05 level, as are the other 3 distributional goodness of fit tests. The histogram is very bell-shaped, and the probability plot is also pretty close to a straight line indicating that the data is close to normal. Based on these results, we would not reject normality and it would be fine to use tests that assume normality on the magnesium values as a whole.

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.990225	**Pr < W**	0.2639
Kolmogorov-Smirnov	**D**	0.063491	**Pr > D**	0.0793
Cramer-von Mises	**W-Sq**	0.072874	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.500758	**Pr > A-Sq**	0.2138





b) Now we consider the magnesium values for each cultivar. For cultivar 1, the mean and median are lower than in the combined data, with the mean being 17 and the median slightly lower at 16.8. The skewness is about the same as before with a small positive value of .206. We can again trust the standard deviation as a measure of spread and find that it is smaller than in the overall sample, with a value of 2.55.

Normality is again not rejected, as the p-values for all tests of normality are greater than .05; the histogram is still fairly bell-shaped; and despite a little more deviation from a straight line in the probability plot, the plot is still pretty straight and demonstrates no reason for concern about a normality assumption.

Variable: Magnesium

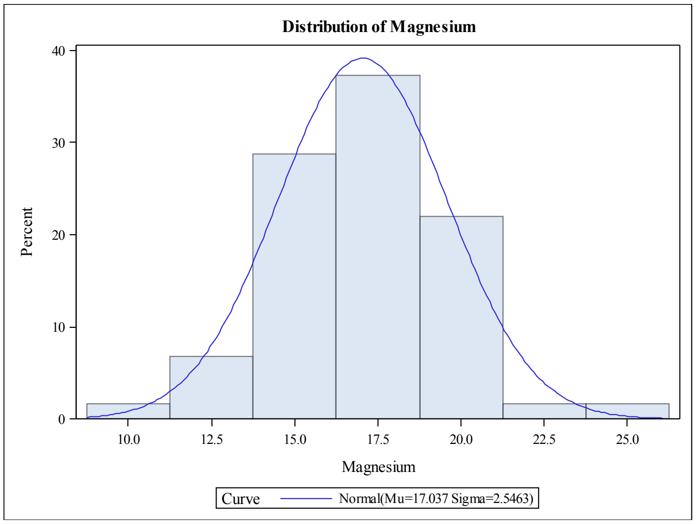
Alcohol=1

Moments			
N	59	**Sum Weights**	59
Mean	17.0372881	**Sum Observations**	1005.2
Std Deviation	2.54632245	**Variance**	6.48375804
Skewness	0.20588305	**Kurtosis**	1.19850312
Uncorrected SS	17501.94	**Corrected SS**	376.057966
Coeff Variation	14.9455854	**Std Error Mean**	0.33150295

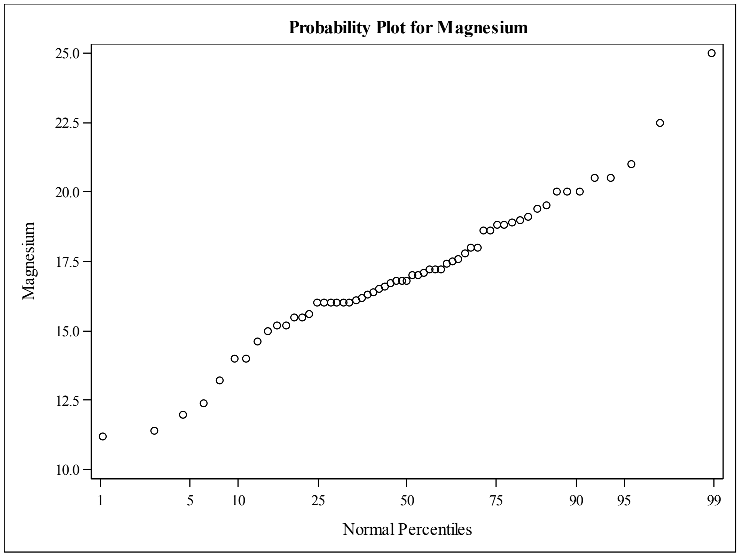
Basic Statistical Measures			
Location		**Variability**	
Mean	17.03729	**Std Deviation**	2.54632
Median	16.80000	**Variance**	6.48376
Mode	16.00000	**Range**	13.80000
		Interquartile Range	2.80000

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.973147	**Pr < W**	0.2161
Kolmogorov-Smirnov	**D**	0.104581	**Pr > D**	0.1059
Cramer-von Mises	**W-Sq**	0.09571	**Pr > W-Sq**	0.1288
Anderson-Darling	**A-Sq**	0.559479	**Pr > A-Sq**	0.1457

Alcohol=1



Alcohol=1



For cultivar 2, the mean and median are just slightly above those for the overall sample. The skewness is a bit higher at .43, indicating a more noticeable right tail in the cultivar 2 distribution, but the skewness is still not terribly strong and using the standard deviation as a measure of spread should still be fine. The standard deviation is 3.35 is almost the same as for the overall sample.

The tests for normality all have p-values greater than .07. While the sample distribution is farther from normal that the overall sample and that for cultivar 1, we still do not see a statistically significant difference. The histogram again looks pretty bell-shaped and the probability plot reasonably straight, so we again determine a normality assumption is not unreasonable.

Variable: Magnesium

Alcohol=2

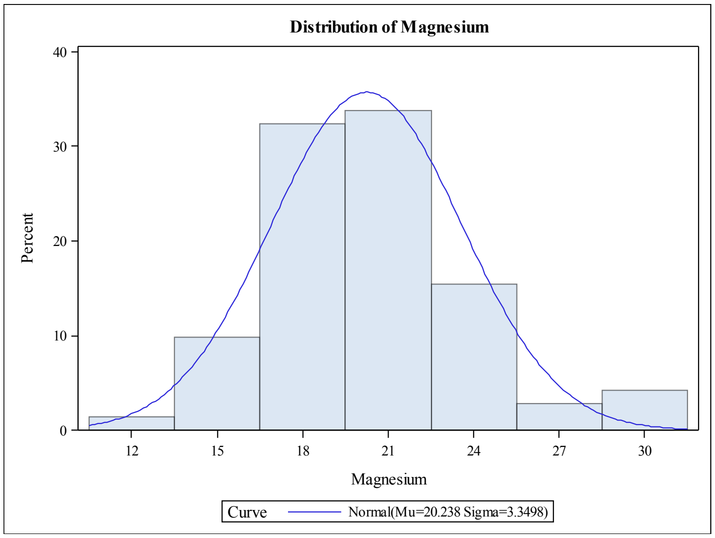
Moments			
N	71	**Sum Weights**	71
Mean	20.2380282	**Sum Observations**	1436.9
Std Deviation	3.34977041	**Variance**	11.2209618
Skewness	0.43078349	**Kurtosis**	1.22117379
Uncorrected SS	29865.49	**Corrected SS**	785.467324
Coeff Variation	16.5518616	**Std Error Mean**	0.39754461

Basic Statistical Measures			
Location		Variability	
Mean	20.23803	Std Deviation	3.34977
Median	20.00000	Variance	11.22096
Mode	18.00000	Range	19.40000
		Interquartile Range	4.00000

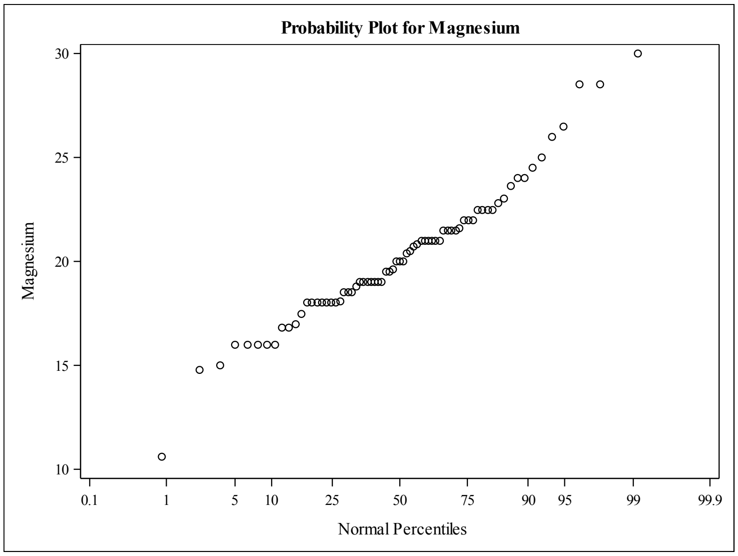
Note: The mode displayed is the smallest of 2 modes with a count of 7.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.968778	Pr < W	0.0740
Kolmogorov-Smirnov	D	0.083016	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.099353	Pr > W-Sq	0.1152
Anderson-Darling	A-Sq	0.690754	Pr > A-Sq	0.0722

Alcohol=2



Alcohol=2



For cultivar 3, we see the highest mean and median values, roughly 1.5 to 2 higher than in the overall sample. We also see a positive skewness just slightly above the skewness of cultivar 2 and a standard deviation a little less than that for cultivar 1. If normality is not rejected, we can still use standard deviation as our measure of spread for cultivar 3.

As we look at the normality tests, we see that Shapiro-Wilk is insignificant at a .05 level, Cramer-von Mises and Anderson-Darling are close to .05 but still insignificant at a .05 level, and Kolmogorov-Smirnov is very significant at a .05 level. We see the same story in the plots. The histogram is starting to skew away from the bell curve with some more concentrated weight on the left side, and the probability plot shows more dips and peaks around a straight line. This indicates that we are seeing more deviation from normality than before and this sample is approaching the point at which we would need to reject normality, but it will still be OK to assume normality when we perform tests on this data.

Variable:
Magnesium

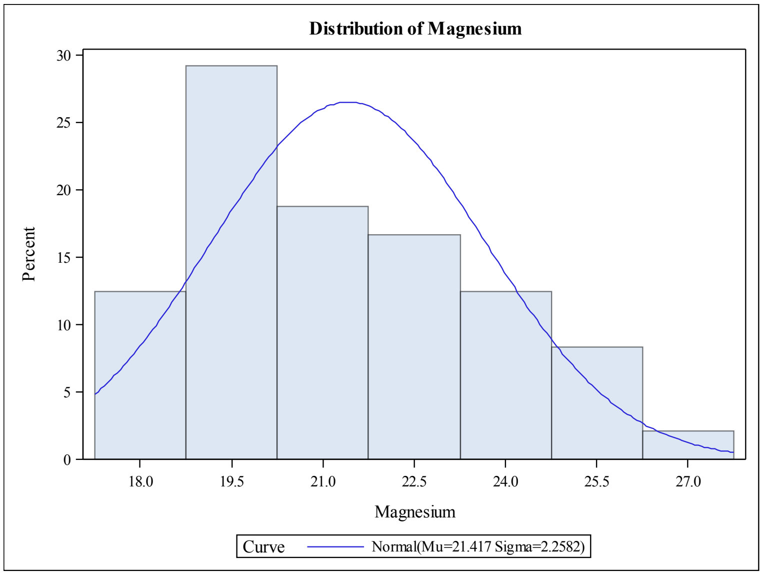
Alcohol=3

Moments			
N	48	**Sum Weights**	48
Mean	21.4166667	**Sum Observations**	1028
Std Deviation	2.25816093	**Variance**	5.09929078
Skewness	0.46792981	**Kurtosis**	-0.5241882
Uncorrected SS	22256	**Corrected SS**	239.666667
Coeff Variation	10.5439421	**Std Error Mean**	0.32593746

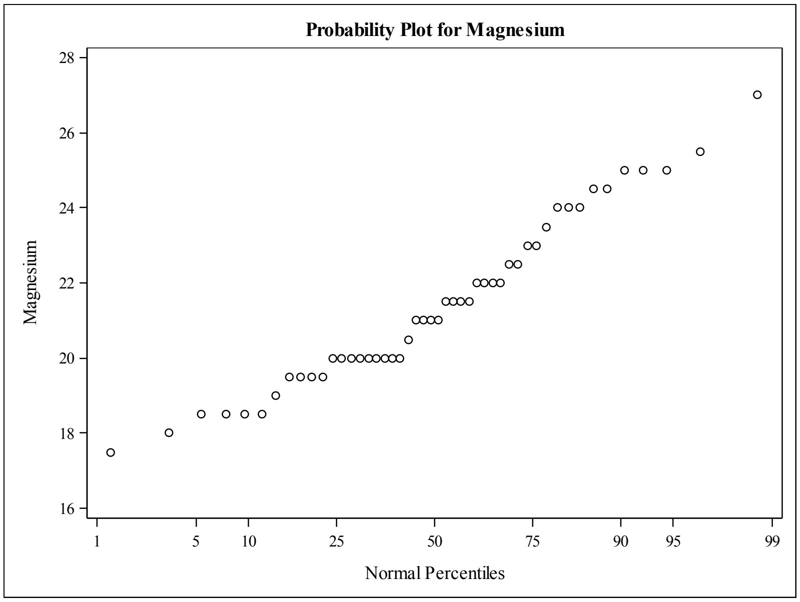
Basic Statistical Measures			
Location		**Variability**	
Mean	21.41667	**Std Deviation**	2.25816
Median	21.00000	**Variance**	5.09929
Mode	20.00000	**Range**	9.50000
		Interquartile Range	3.00000

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.959762	**Pr < W**	0.0987
Kolmogorov-Smirnov	**D**	0.151453	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.124029	**Pr > W-Sq**	0.0515
Anderson-Darling	**A-Sq**	0.729006	**Pr > A-Sq**	0.0545

Alcohol=3



Alcohol=3


Probability Plot for Magnesium
Magnesium
28
26
24
22
20
18
16
1
5
10
25
50
75
90
95
99
Normal Percentiles

Exercise 2

a) In Exercise 1, we found that normality was not unreasonable for magnesium values as a whole. Thus, we choose a one-sided t test to check whether cultivar 2 has a mean value significantly higher than 20. The p-value for that test is .2756, so we would not reject the null hypothesis that the population mean magnesium value for cultivar 2 is 20, and we conclude that the true population mean magnesium value for cultivar 2 is not significantly different from 20.

Variable: Magnesium

DF	t Value	Pr > t
70	0.60	0.2756

b) In Exercise 1, we concluded that magnesium values for cultivars 1 and 3 were each reasonably close to normal, and so we use a one-sided t test to see if magnesium levels are significantly higher in cultivar 3 than in cultivar 1 wines. From the folded F test, we conclude that the variances for the two populations are not significantly different, and it is fine to use the pooled estimate of variance for our t test. The hypothesis test is highly significant indicating that magnesium levels are significantly higher in cultivar 3 wines than in cultivar 1 wines. Though not asked for in the exercise, the pooled estimate of the difference indicates magnesium levels would be expected to be about 4.38 higher in cultivar 3 wines than in cultivar 1 wines.

Variable: Magnesium

Alcohol	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		17.0373	16.3737	17.7009	2.5463	2.1555	3.1115
3		21.4167	20.7610	22.0724	2.2582	1.8798	2.8285
Diff (1-2)	**Pooled**	-4.3794	-Infty	-3.5983	2.4216	2.1337	2.8000
Diff (1-2)	**Satterthwaite**	-4.3794	-Infty	-3.6078			

Variable: Magnesium

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	105	-9.30	<.0001
Satterthwaite	Unequal	104.19	-9.42	<.0001

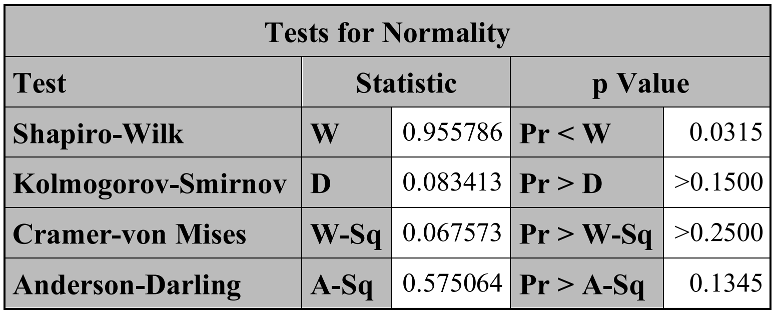
Equality of Variances				
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Folded F	58	47	1.27	0.3970

Exercise 3

a) As we look at the normality checks for color intensity for cultivars 1 and 3, we start to see some conflicting results. For cultivar 1, Shapiro-Wilk is significant, but the other distributional tests are not. Looking at the histogram, the distribution does not look very normal, and in the probability plot the plot still looks to largely follow a straight line with some deviation at the edges. We should be a little cautious about assuming normality for cultivar 1 color intensities.

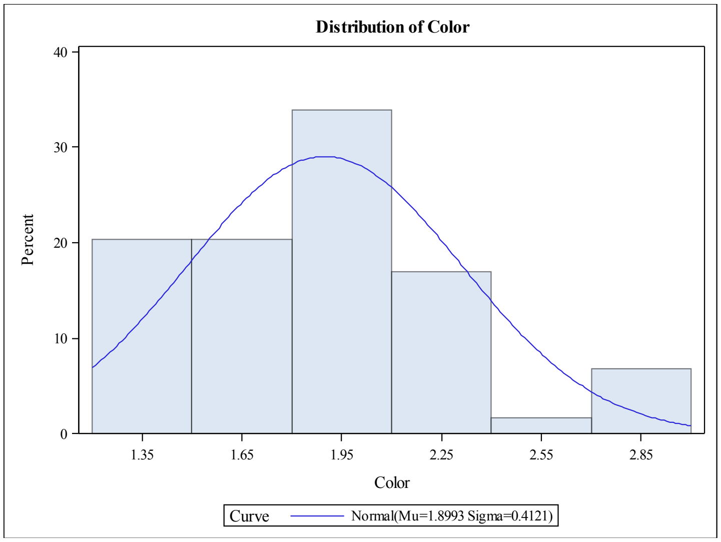
Variable: Color

Alcohol=1

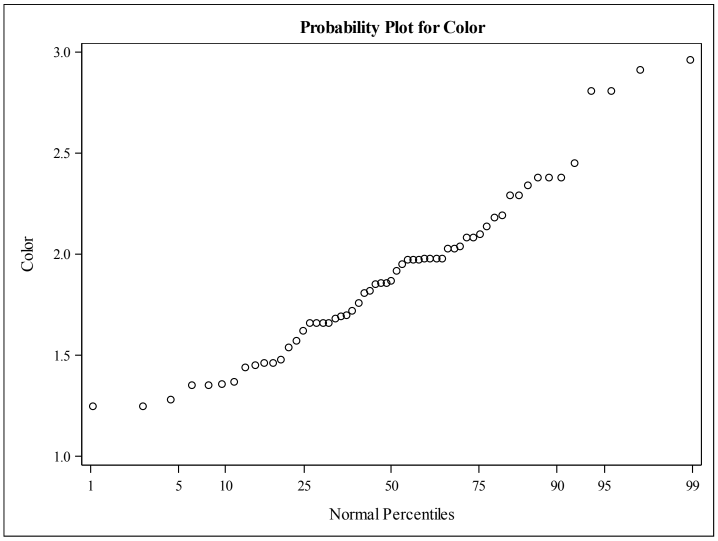


Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.955786	**Pr < W**	0.0315
Kolmogorov-Smirnov	**D**	0.083413	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.067573	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.575064	**Pr > A-Sq**	0.1345

Alcohol=1



Alcohol=1



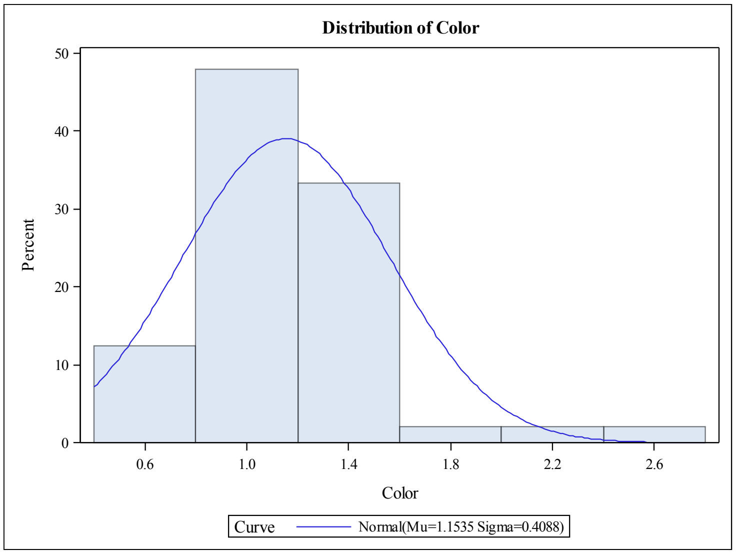
For cultivar 3, all but Kolmogorov-Smirnov solidly reject normality. The histogram and probability plot also indicate a long right tail. We should clearly reject normality for cultivar 3.

Variable: Color

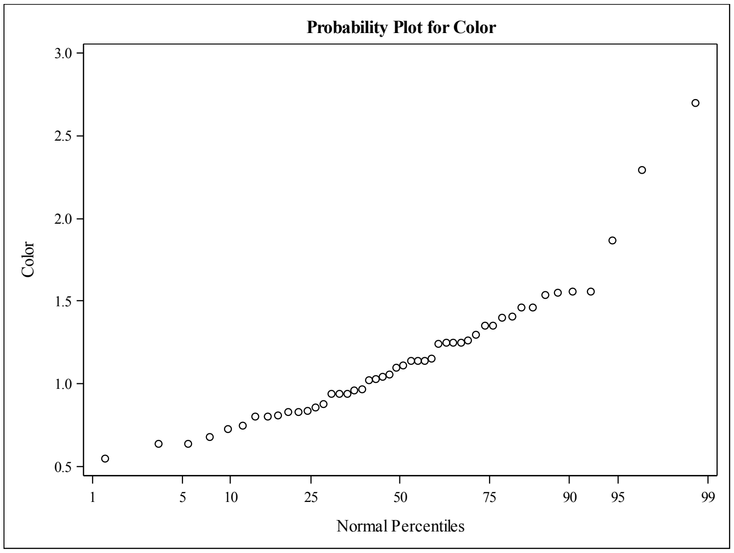
Alcohol=3

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.887202	**Pr < W**	0.0002
Kolmogorov-Smirnov	**D**	0.107623	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.144202	**Pr > W-Sq**	0.0276
Anderson-Darling	**A-Sq**	1.09874	**Pr > A-Sq**	0.0067

Alcohol=3



Alcohol=3



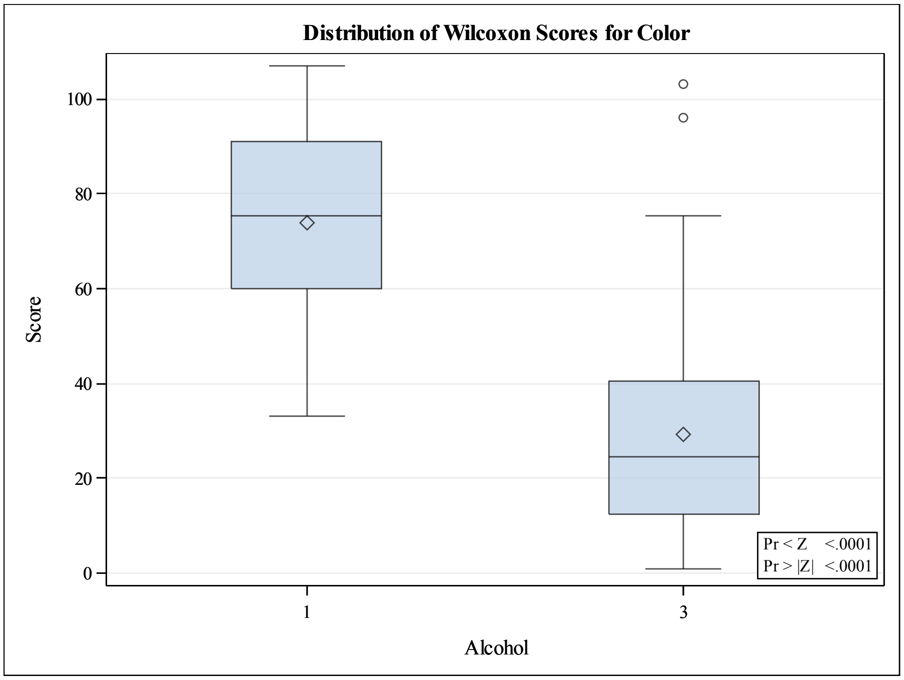
b) Having rejected normality for cultivar 3, we need to use a Wilcoxon rank sum test to compare color intensities for cultivar 1 and 3. Here the statistic is based on the sum for cultivar 3 and the left-sided test is highly significant indicating that color intensities for cultivar 3 are lower than for cultivar 1. We can also see this in the box plot as well. Consumers who prefer greater color intensity would tend to prefer cultivar 1 wines.

Wilcoxon Scores (Rank Sums) for Variable Color
Classified by Variable Alcohol

Alcohol	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	59	4367.0	3186.0	159.614423	74.016949
3	48	1411.0	2592.0	159.614423	29.395833

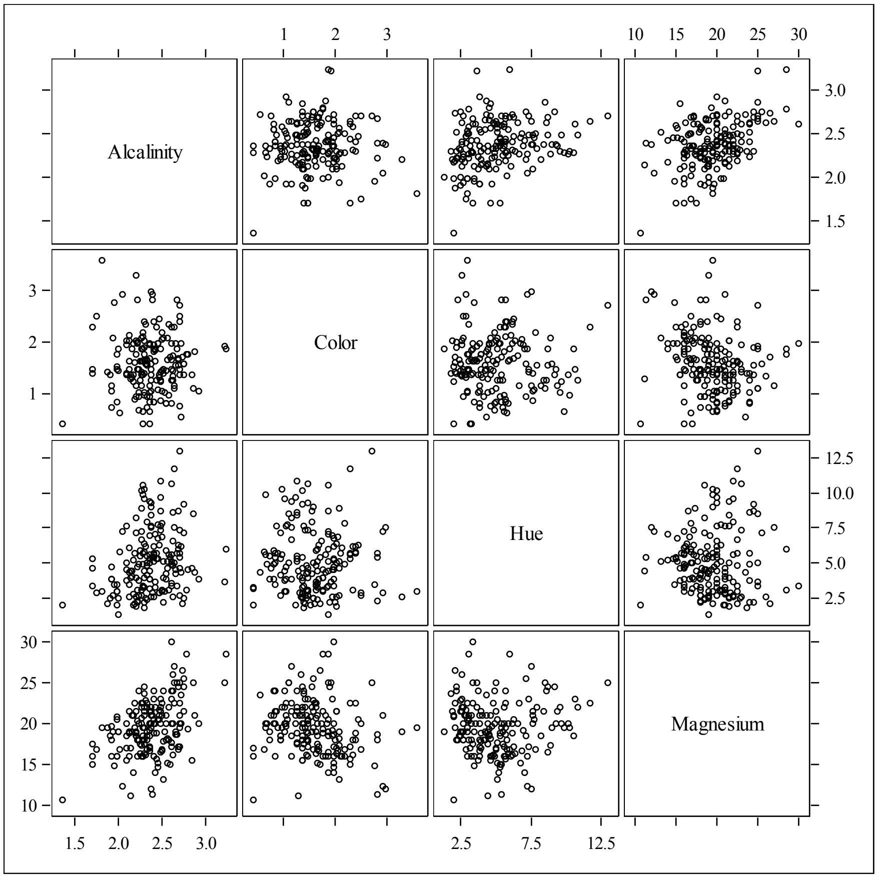
Average scores were used for ties.

Wilcoxon Two-Sample Test	
Statistic	1411.0000
Normal Approximation	
Z	-7.3959
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	<.0001
t Approximation	
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	<.0001
Z includes a continuity correction of 0.5.	



Exercise 4

a) From the scatter plots for the entire data set, there are no obvious nonlinear trends or concerning evidence of non-constant variance. The plots mostly look like very spread out points. Based on the plots, Pearson correlation will be fine here.



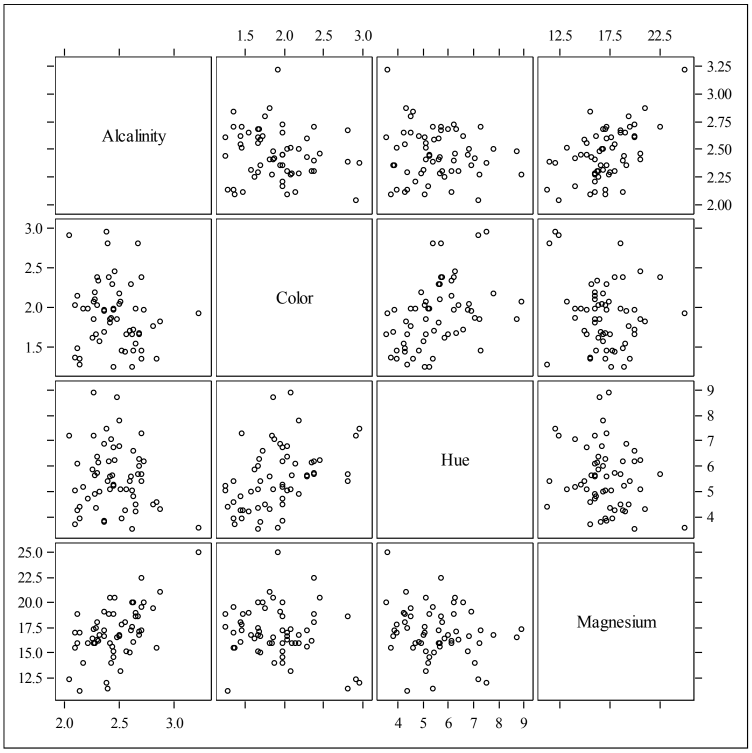
From the correlation matrix, we find 3 statistically significant correlations. The correlation between alkalinity and hue is statistically significant and has a positive estimate of .26. This indicates a small tendency for alkalinity to increase as hue increases. Alkalinity and magnesium have a higher positive correlation of .44, indicating a larger, but still small to moderate, tendency for those two variables to increase and decrease together. The correlation between magnesium and color is negative and fairly small in magnitude indicating a small tendency for one to increase as the other decreases.

Pearson Correlation Coefficients, N = 178
Prob > |r| under H0: Rho=0

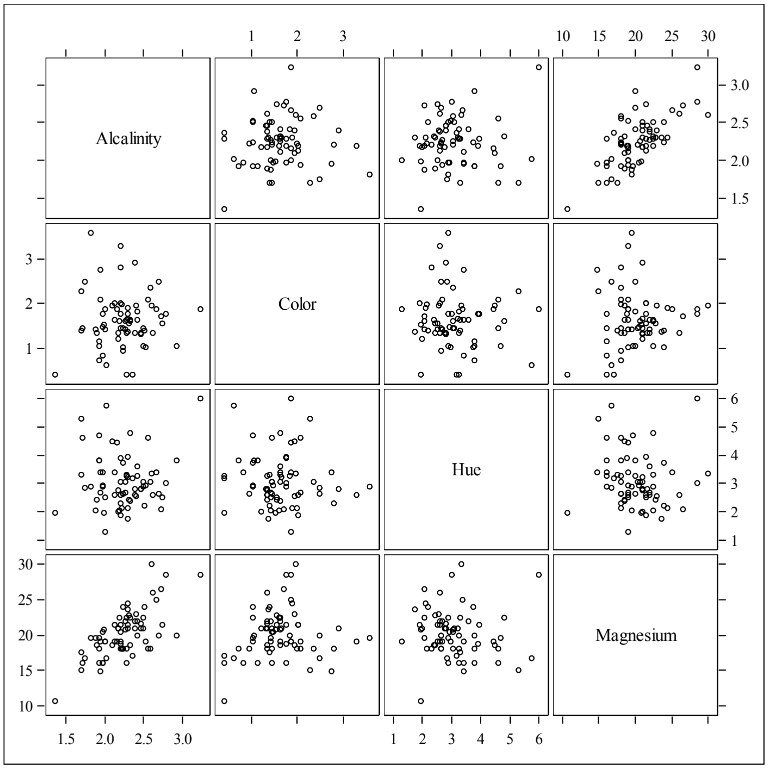
	Alcalinity	Color	Hue	Magnesium
Alcalinity	1.00000	0.00965 0.8983	0.25889 0.0005	0.44337 <.0001
Color	0.00965 0.8983	1.00000	-0.02525 0.7380	-0.19733 0.0083
Hue	0.25889 0.0005	-0.02525 0.7380	1.00000	0.01873 0.8040
Magnesium	0.44337 <.0001	-0.19733 0.0083	0.01873 0.8040	1.00000

b) Checking scatter plots for each cultivar, we see no concerning nonlinear trends or high variability of variances for any of the pairs of variables. We may notice, though, that linear trends are becoming more apparent for some variable pairs within cultivars. Pearson correlation should again be used for each cultivar.

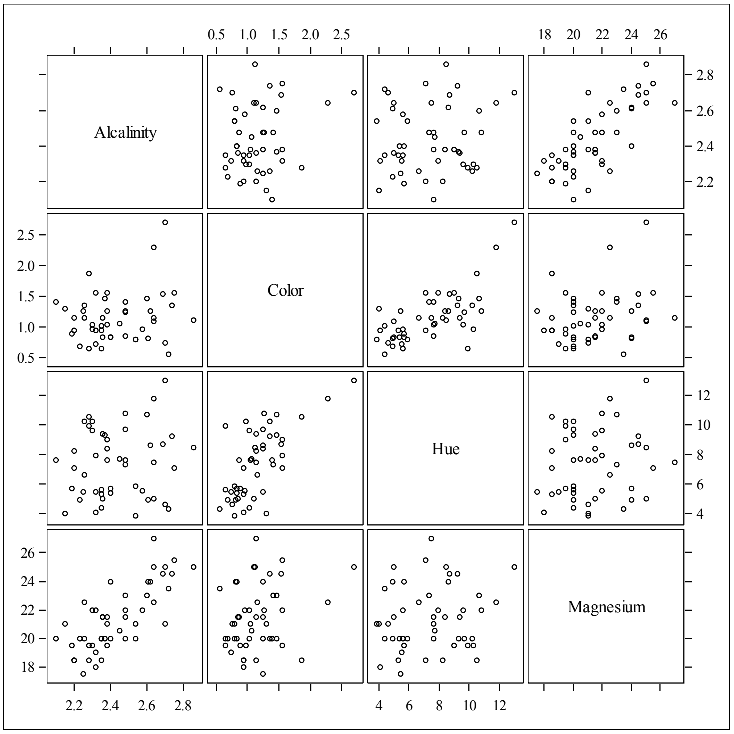
Alcohol=1



Alcohol=2



Alcohol=3



Significant correlations within cultivars are noticeably different from what we saw in the combined sample. For cultivar 1, we see moderate statistically significant correlations between alkalinity and magnesium and between hue and color intensity.

Alcohol=1

Pearson Correlation Coefficients, N = 59 Prob > |r| under H0: Rho=0				
	Alcalinity	**Color**	**Hue**	**Magnesium**
Alcalinity	1.00000	-0.14547 0.2716	-0.12422 0.3486	0.54933 <.0001
Color	-0.14547 0.2716	1.00000	0.42470 0.0008	-0.17363 0.1885
Hue	-0.12422 0.3486	0.42470 0.0008	1.00000	-0.21095 0.1088
Magnesium	0.54933 <.0001	-0.17363 0.1885	-0.21095 0.1088	1.00000

For cultivar 2, there is only one statistically significant correlation, a moderate to strong positive correlation between alkalinity and magnesium.

Alcohol=2

Pearson Correlation Coefficients, N = 71
Prob > |r| under H0: Rho=0

	Alcalinity	Color	Hue	Magnesium
Alcalinity	1.00000	0.04296 0.7221	0.06025 0.6177	0.69526 <.0001
Color	0.04296 0.7221	1.00000	-0.07376 0.5410	0.10884 0.3663
Hue	0.06025 0.6177	-0.07376 0.5410	1.00000	-0.08586 0.4765
Magnesium	0.69526 <.0001	0.10884 0.3663	-0.08586 0.4765	1.00000

Cultivar 3's statistically significant correlations are similar to those for cultivar 1, but much stronger. There is a stronger positive relationship between hue and color and between magnesium and alkalinity for cultivar 3 wine than for cultivar 1 wines.

Alcohol=3

Pearson Correlation Coefficients, N = 48
Prob > |r| under H0: Rho=0

	Alcalinity	Color	Hue	Magnesium
Alcalinity	1.00000	0.19383 0.1868	0.12515 0.3967	0.75852 <.0001
Color	0.19383 0.1868	1.00000	0.68491 <.0001	0.26340 0.0705
Hue	0.12515 0.3967	0.68491 <.0001	1.00000	0.16062 0.2755
Magnesium	0.75852 <.0001	0.26340 0.0705	0.16062 0.2755	1.00000

The only correlation consistent across all cultivars and in the sample as a whole is the positive relationship between alkalinity and magnesium, though the magnitude of that correlation varies by sampled group. The other statistically significant correlations for the overall sample are not seen within any of the cultivar subsamples; this can happen when there is a noticeable difference in variable magnitudes for the correlated variables across groups but not within groups.