Grouping and Ordering of Sites by Rind Thicknesses in the River Valleys of Northwest Tasmania

By Nicholas Forbes-Smith

Aim

To order 15 sites in the River Valleys of Northwest Tasmania by rind thickness and to identify which of those sites can be grouped together due to similar rind thickness measurements.

Background

The data set contains measurements of 50 rind thicknesses from each of 15 sites in the river valleys of Northwest Tasmania.

The recorded variables are:

- *Site Number* – The identifying number used for each of the 15 sites

- *Thickness* – Measurement of rind thickness

Findings

1. In terms of decreasing average rind thickness, the sites are in the following order:
   - **6, 16, 15, 1, 9, 7, 2, 3, 5, 11, 8, 4, 14, 13, 12**   [Table 01]

2. If a conservative grouping of sites is required, the following sites should be grouped [Table 09].

Grouping Structure 1

Group	Sites in Group
A	1, 15
B	2, 3, 5
C	4, 8, 11
D	12, 13, 14
E	6
F	7
G	9
H	16

3. A more liberal grouping of sites would include the following groupings, in addition to the groups presented in the first grouping structure [Table 12 + Table 13].

- *Put site 9 in Group A with sites 1 and 15*

- *Put site 7 in Group B with sites 2, 3 and 5*

Grouping Structure 2

Group	Sites in Group
A	1, *9*, 15
B	2, 3, 5, *7*
C	4, 8, 11
D	12, 13, 14
E	6
F	16

4. Additionally, sites 6 and 16 can be grouped for some sacrifice in accuracy of the grouping structure [Table 13].

- It should be kept in mind that sites 6 and 16 have the largest rind thickness on average [Table 01] and they also have highest variation in rind thicknesses recorded [Figure 03]. Therefore, if more rind thickness measurements are taken for these sites, it may show that this grouping isn't appropriate.

- *Put site 6 and site 16 into Group E*

Grouping Structure 3

| Group	Sites in Group
A	1, 9, 15
B	2, 3, 5, 7
C	4, 8, 11
D	12, 13, 14
E	6, *16*

Statistical Appendix

Computational Methods

- All statistical analyses were undertaken using Jupyter lab (alpha 0.26.5) with R (v 3.4.1 on x86_64-w64-mingw32 platform) and RStudio (v 1.0.153). All were installed as part of the Anaconda3 distribution (v 4.4.0)

- Libraries—MASS (boxcox), ggplot2 (t-interval plot), car (levene.test)

- Functions used during analyses:
  - read.csv() was used to import CSV files into R
  - sort() and tapply() were used to sort sites by mean and median rind thickness
  - lm() was used for regression
  - anova() was used to perform nested F-tests
  - summary() was used to print linear regression summary tables
  - influence.measures() was used to print influence measures summary table
  - kruskal.test() used to perform Kruskal-Wallis rank sum tests
  - TukeyHSD(), pairwise.t.test() and a custom Games Howell function were used for post-hoc testing
  - plot() was used to display residual diagnostic plots
  - plot(), plot.design()and legend() were used for all 2D plots
  - ggplot(), geom_point(), geom_errorbar(), ggtitle() were used for the T-interval plot
  - boxcox() was used to perform box cox transformations
  - oneway.test() was used to perform Welch's ANOVA
  - bartlett.test() and leveneTest() were used to perform test of homogeneity of variances

- All the code used for statistical analyses is available at:
  https://github.com/nf-s/Grouping-and-Ordering-of-Sites-in-the-River-Valleys-of-Northwest-Tasmania-by-Rind-Thicknesses

Results

Sites by Mean and Median Rind Thickness

SiteNum	Mean Thickness		SiteNum	Median Thickness
6	10.48		6	10.50
16	6.33		16	5.80
15	3.16		15	2.80
1	2.39		1	2.40
9	1.49		9	1.50
7	1.11		7	1.00
2	0.88		2	0.80
3	0.75		3	0.80
5	0.73		5	0.80
11	0.57		11	0.40
8	0.40		8	0.30
4	0.34		4	0.20
14	0.25		12	0.20
13	0.22		13	0.20
12	0.20		14	0.20

Mean Rind Thickness 95% t-interval
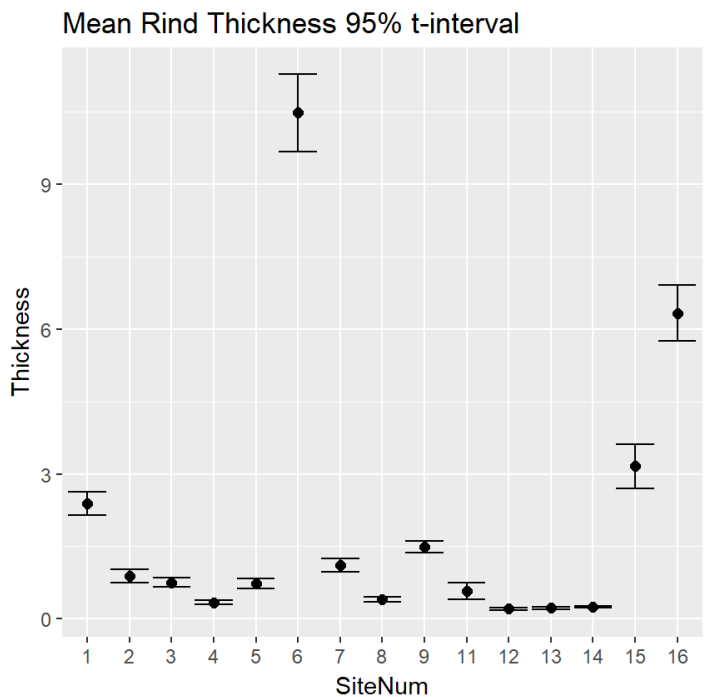
Table 01: *The table presents the Mean and Median thickness of each site in descending order, the spacing of sites represents the initial grouping structure. There is a clear order of sites as the mean and medians are in the same order.*

Figure 01: *This figure shows that there may be potential heterogeneous variance across groups. It also shows the large difference in rind thickness between sites with larger values, especially Site 6, 16, 15 and 1.*

To determine whether the differences in mean rind thickness between sites are statistically significant, we will perform a One-way ANOVA.

One-way ANOVA summary – No Transformation

Formula = Thickness ~ SiteNum

	Df	Sum Sq.	Mean Sq.	F Value	Pr(>F)
SiteNum	14	5724	408.9	369.8	**<0.001**
Residuals	735	813	1.1		

Table 02: *The ANOVA summary reports that there is at least one significant difference in mean rind thickness between sites.*

This result is questionable as the figure below highlights that the assumptions required for ANOVA haven't been met.

One-way ANOVA Residual Plot – Thickness ~ SiteNum
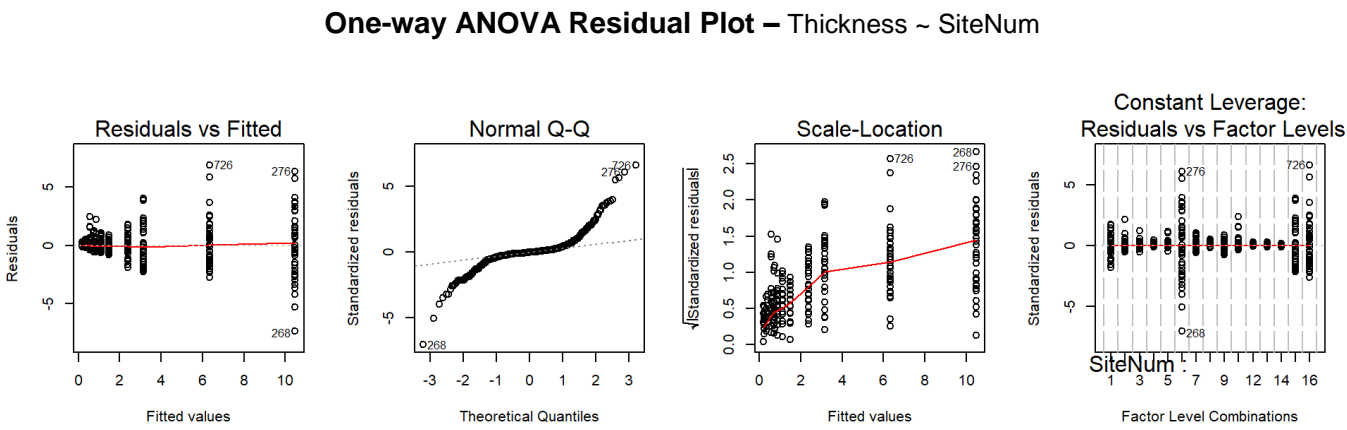
Figure 02:   *There are obvious issues with normality and heteroscedastic variance. The normal Q-Q plot shows that the untransformed data has violated the assumption of normally distributed residuals. The residual vs fitted plot shows that the assumption of homoscedasticity hasn't been met.*

There are also a few points with high standardised residuals - namely 268, 276 and 726. Points 276 and 268 are in both from site 6 and have +5 and -5 standardised residuals respectively, this further indicates that there is unequal variance among sites.

The following table has been included to allow comparison between post-hoc tests which require homoscedasticity and those which allow heteroscedasticity (Games Howell test and Pairwise t-tests with non-pooled standard deviation – See Table 09). We can't interpret the results of Tukey's HSD post-hoc test as it requires normally distributed residuals and homoscedastic variance, both of which aren't met.

Tukey's HSD Pairwise P-value Matrix – Thickness ~ SiteNum

Site#	1	2	3	4	5	6	7	8	9	11	12	13	14	15
2	<.01													
3	<.01	>.99												
4	<.01	.37	.82											
5	<.01	>.99	>.99	.88										
6	<.01	<.01	<.01	<.01	<.01									
7	<.01	>.99	.92	.02	.88	<.01								
8	<.01	.60	.95	>.99	.97	<.01	.05							
9	<.01	.19	.03	<.01	.02	<.01	.89	<.01						
11	<.01	.98	>.99	>.99	>.99	<.01	.39	>.99	<.01					
12	<.01	.08	.37	>.99	.45	<.01	<.01	>.99	<.01	.92				
13	<.01	.11	.44	>.99	.52	<.01	<.01	>.99	<.01	.95	>.99			
14	<.01	.15	.53	>.99	.61	<.01	<.01	>.99	<.01	.97	>.99	>.99		
15	.02	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	
16	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01

Table 03:   *The table shows that Tukey's HSD results report that there are quite a few groups which have no significance difference in mean rind thickness. In total there are 44 pairs of groups which have no significant difference (at a .95 confidence level). Note that these results cannot be interpreted as the assumptions required for Tukey's HSD have not been met.*

To confirm whether the assumption of homogeneity of variance among groups has been violated we will examine the variances among the sites and perform a Bartlett test and a Levene's test.

Plot of sites – Variance of Sites & Boxplots of Rind Thicknesses by Site
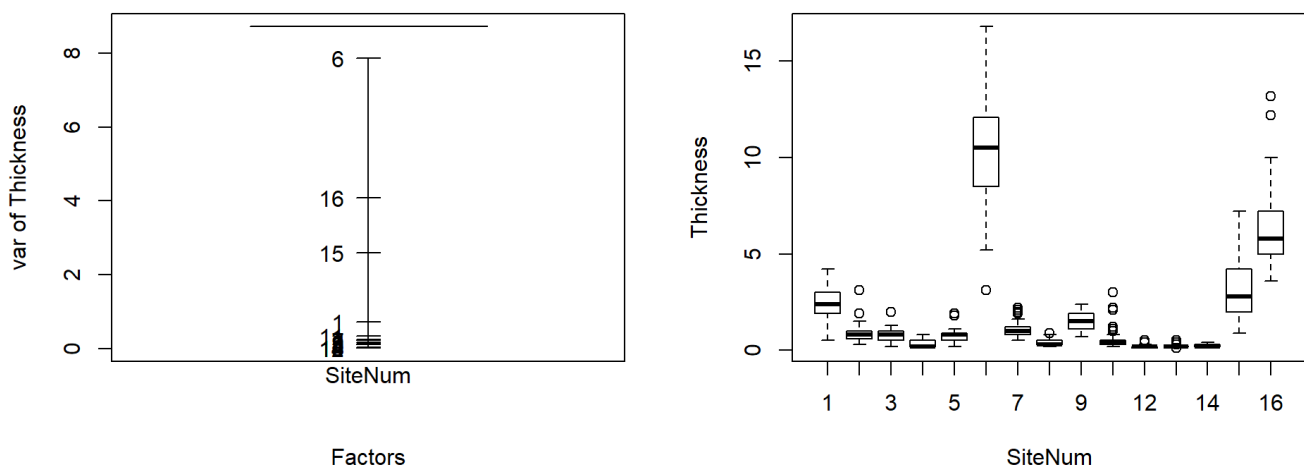
Figure 03:   *The left plot shows that there is very large difference in variance in rind thickness between the different sites. Site 6 has highest variance, with 7.874, and Site 14 has the lowest variance, with 0.005. The variance of Site 6 is over 1457 times higher than Site 14, which is way too high. The box plot illustrates the dramatic differences in variance of thickness between the sites. It also shows that variance of sites increases as the mean rind thickness increases.*

Bartlett and Levene tests – Thickness ~ SiteNum

	Bartlett test				Levene test		
Df	*K-squared*	*P-value*		*Df*	*F Value*	*P-value*
14	1439.5	**<0.001**		14	45.0	**<0.001**

Table 04: *Bartlett's and Leveve's test both test for homogeneity of variance. The significant p-values reject the null hypothesis that there is homogeneous variance among groups. This confirms observations made from the previous two figures, the assumption of homoscedasticity has been violated and therefore the results of the ANOVA and Tukey's HSD cannot be interpreted.*

To fix issues of heteroscedasticity and lack of normality in distribution of residuals we will try to apply a transformation on the response variable (Rind Thickness). To identify possible transformations, we will perform a box cox transformation.
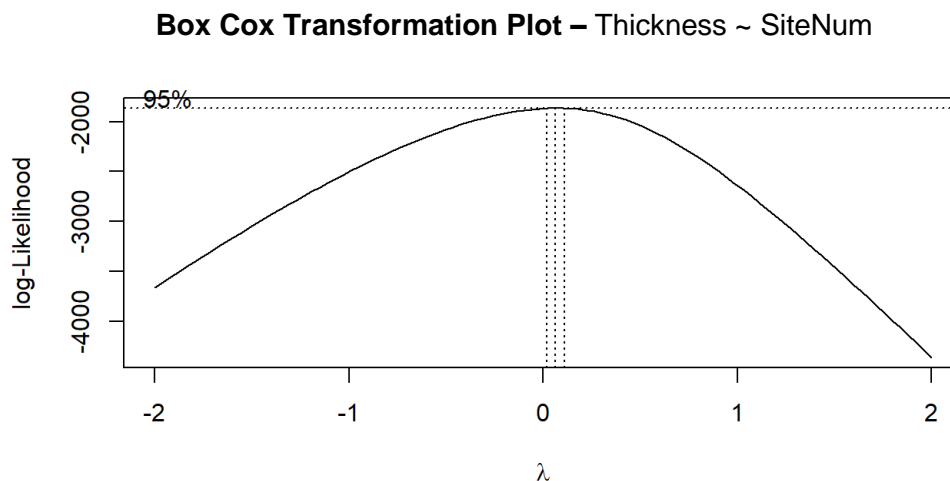
Box Cox Transformation Plot – Thickness ~ SiteNum

Figure 04: *The box cox plot suggests a log transformation on the response variable.*

We will transform the Thickness variable with natural log and re-complete previous analyses, starting with One-way ANOVA.

One-way ANOVA Summary – With Log Transformation

Formula = log(Thickness) ~ SiteNum					
	Df	*Sum Sq.*	*Mean Sq.*	*F Value*	*Pr(>F)*
SiteNum	14	1044.9	74.63	378.9	**<0.001**
Residuals	735	114.8	0.2		

Table 05: *The summary table present equivalent results as the ANOVA without the log transformation (in Table 02), the very small p-value indicates there is at least one significant difference in mean rind thickness between sites.*

**One-way ANOVA Residual Plot** – log(Thickness) ~ SiteNum
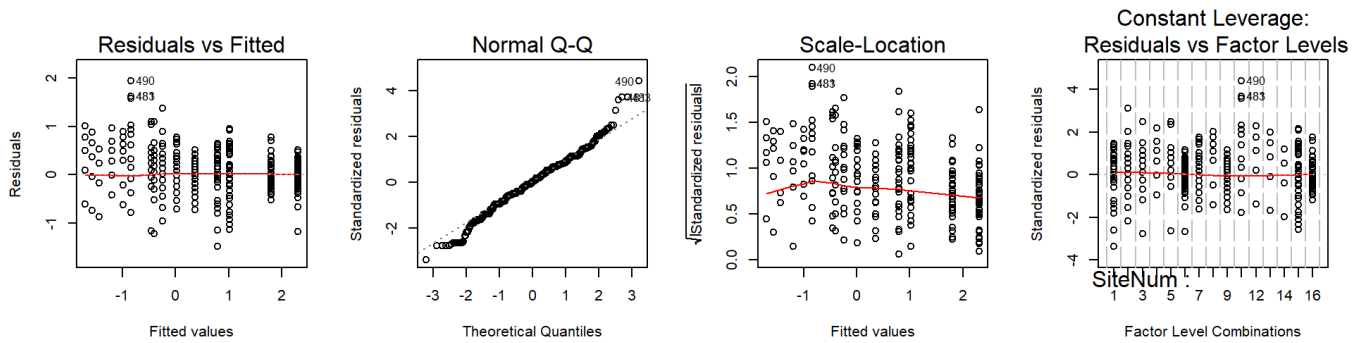
Figure 05: *The log transformation on the response variable has significantly improved normality of residuals, but the residuals vs fitted plot still shows evidence of heteroscedastic variance. There are also a few outliers which have high standardized residuals and deviate from the normal Q-Q line, namely points 490, 483 and 481 which are all from site 11.*

To confirm whether the transformation has lowered heteroscedasticity to an acceptable level for ANOVA we need to examine the variance among groups and perform a Bartlett and a Levene's test on the transformed data.

Plot of sites – Variance of Sites & Boxplots of Rind Thicknesses by Site

[With log transformation on thickness variable]
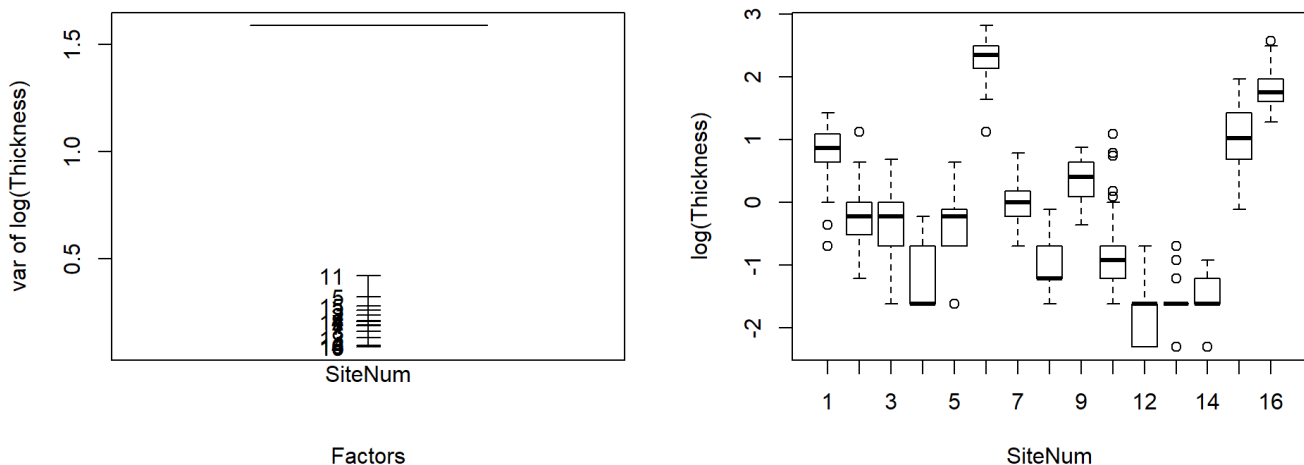
Figure 06: *The difference in variance between groups has decreased when compared with Figure 03. The site with the lowest variance is now site 16, with 0.09, and the site with the highest variance is now site 11, with 0.42. The variance of site 16 is 4.9 times than the variance of site 11, this is sometimes considered too high to assume homoscedasticity. Figure 05 showed that site 11 had a few outlying points, the box plot shows that site 11 has more outliers than any other site and the widest range of points (absolute range not IQR).*

Bartlett and Levene tests – Thickness ~ SiteNum

	Bartlett test				Levene test		
	Df	*K-squared*	*P-value*		*Df*	*F Value*	*P-value*
	14	82.55	**<0.001**		14	3.03	**<0.001**

Table 06:   *There is an improvement in K-squared and F values compared with non-transformed data (Table 04). However, the p-values are still highly significant, which indicate that there is heteroscedasticity between sites and therefore the results of the ANOVA and Tukey's HSD cannot be interpreted.*

Welch's ANOVA doesn't require homoscedasticity, but it does require normality of residuals. This allows us to interpret the output if we accept that assumption of normally distributed residuals has been met.

Welch One-way ANOVA Summary – log(Thickness) ~ SiteNum

Formula = log(Thickness) ~ SiteNum				
	Df	*Denom df*	*F Value*	*Pr(>F)*
SiteNum	14	279.66	604.73	**<0.001**

Table 07:   *The Welch ANOVA also reports a significant p-value, which further states that there are differences in mean rind thickness among sites.*

As we cannot fully confirm that the assumption of normality has been met we will also perform a Kruskal-Wallis Ranked sum test, which is a distribution free test.

Formula = log(Thickness) ~ SiteNum			
	Kruskal-Wallis chi-squared	*Df*	*Pr(>F)*
SiteNum	637.4	14	**<0.001**

Table 08:   *The rank-sum test also reports a highly significant p-value, which confirms the findings of the ANOVA and Welch ANOVA. The lack of difference in p-values between the tests suggest that the assumption of normality may not have be violated.*

As the data is still heteroscedastic, Tukey's HSD post hoc test can't be used. Instead we will use a Games-Howell test and pairwise t-tests with non-pooled standard deviations (and a Holm p-value adjustment).

Games-Howell and Pairwise t-test post hoc tests – log(Thickness) ~ SiteNum

Games-Howell Pairwise P-value Matrix

Site#	1	2	3	4	5	6	7	8	9	11	12	13	14	15
2	<.01													
3	<.01	.96												
4	<.01	<.01	<.01											
5	<.01	.79	1.00	<.01										
6	<.01	<.01	<.01	<.01	<.01									
7	<.01	.25	<.01	<.01	<.01	<.01								
8	<.01	<.01	<.01	.56	<.01	<.01	<.01							
9	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01						
11	<.01	<.01	.02	**.11**	**.13**	<.01	<.01	.98	<.01					
12	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01				
13	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	.96			
14	<.01	<.01	<.01	**.12**	<.01	<.01	<.01	<.01	<.01	<.01	**.07**	.82		
15	.54	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	
16	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01

Pairwise t-test P-value Matrix (with Holm p-value adjustment)

Site#	1	2	3	4	5	6	7	8	9	11	12	13	14	15
2	<.01													
3	<.01	.48												
4	<.01	<.01	<.01											
5	<.01	.29	.60	<.01										
6	<.01	<.01	<.01	<.01	<.01									
7	<.01	.05	<.01	<.01	<.01	<.01								
8	<.01	<.01	<.01	.16	<.01	<.01	<.01							
9	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01						
11	<.01	<.01	<.01	**.02**	**.02**	<.01	<.01	.48	<.01					
12	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01				
13	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	.48			
14	<.01	<.01	<.01	**.02**	<.01	<.01	<.01	<.01	<.01	<.01	**.01**	.29		
15	.16	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	
16	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01

Table 09:   The underlined p-values indicate the tests which produced different results between the two post hoc tests (with a confidence level of 95%). The Games-Howell test reported 13 insignificant differences in rind thickness between pairs of sites and the pairwise t-tests reported 9 insignificant differences. Tukey's HSD, in Table 03, reported 44 insignificant differences, but Tukey's HSD was performed on untransformed data and didn't have required assumptions met.

There is a major difference in number of insignificantly different pairs of sites between Tukey's HSD results (Table 03) and the results above. As Tukey's HSD uses a pooled standard deviation in its testing, and there is heterogeneous variance across groups, the pooled standard deviation may not accurately represent the variance of groups it is testing. This means that it is more likely to cause Type 2 errors with sites with larger variance than the pooled variance and more likely to cause Type 1 errors with sites that

have smaller variance. The Games-Howell test and pairwise t-test both use separate measures of variance for their testing, which results in much more accurate results for comparisons of groups with unequal variances.

To construct the initial grouping of sites we used the Games-Howell test results as it provided more mutually exclusive groupings of sites with insignificant differences. All further possible grouping structures were determined using the 95% confidence intervals for difference in mean, also from the Games-Howell results.

Games-Howell Summary Table – log(Thickness) ~ SiteNum

Insignificant Differences in Rind Thickness Between Sites

Sites	Std. Error	t	df	p	Mean Difference	95% CI Upper	95% CI Lower
1:15	0.07	2.37	94.75	0.54	0.23	0.57	-0.11
2:03	0.07	1.57	97.82	0.96	-0.16	0.19	-0.50
2:05	0.08	2.01	95.72	0.79	-0.21	0.16	-0.58
2:07	0.07	2.83	96.74	0.25	0.26	0.58	-0.06
3:05	0.08	0.52	96.78	>.99	-0.06	0.32	-0.43
4:08	0.06	2.35	96.47	0.56	0.20	0.50	-0.10
4:11	0.08	3.20	87.98	0.11	0.36	0.75	-0.03
4:14	0.06	3.15	85.42	0.12	-0.24	0.03	-0.51
5:11	0.09	3.13	96.35	0.13	-0.38	0.04	-0.80
8:11	0.08	1.45	81.83	0.98	0.16	0.53	-0.22
12:13	0.06	1.57	93.44	0.96	0.13	0.42	-0.16
12:14	0.06	3.37	85.59	0.07	0.26	0.53	-0.01
13:14	0.05	1.96	95.10	0.82	0.13	0.36	-0.10

Table 10:   *The shaded rows within each cluster of rows represent the initial grouping structure of the sites, that is the first group contains sites 1 and 15, the second contains 2, 3 and 5, etc. These groups only include sites which are insignificantly different to all other sites in the same group (to a 95% confidence level). Consequently, the grouping of sites 2, 3 and 5 doesn't include 7, as site 7 is only insignificantly different from site 2, not from site 3 and 5.*

Unlike in Table 10, the pairwise p-value matrix, in Table 09, reports a significant difference for the pairs of sites 4:11, 4:14, 5:11 and 12:14. The Games-Howell confidence intervals (Table 10) for these pairs of sites are so close to 0 that I believe it would be reasonable for the actual difference in mean rind thickness to be 0. Although, I do find it somewhat concerning that Site 4 may not be different from 15, and that Site 5 may not be different from site 11, as this could create some overlap in mean rind thickness between the groups of sites (outlined in the table below). Some further analysis of the effects of the grouping structure would be needed to clarify this.

Grouping Structure 1

Group	Sites in Group
A	1, 15
B	2, 3, 5
C	4, 8, 11
D	12, 13, 14
E	6
F	7
G	9
H	16

Table 11:   This table shows the first grouping structure. When compared with the ordering of sites by mean rind thickness in Table 01, the groups seem sensible.

Table 10 shows that Site 7 has significant differences in rind thickness to sites 3 and 5, the table below shows that the confidence interval for differences in mean are quite close to 0 and therefore could reasonably be placed in Group B.

Games-Howell Summary Table – log(Thickness) ~ SiteNum

Differences in Rind Thickness Between Sites 3, 5 and 7

Sites	Std. Error	t	df	p	Mean Difference	95% CI Upper	95% CI Lower
3:07	0.07	4.42	95.67	0.002	0.42	0.75	0.09
5:07	0.07	4.68	91.63	0.001	0.47	0.82	0.12

Table 12:   As the confidence interval's lower bound is close to 0, I suggest that Site 7 can also be grouped with Sites 2, 3 and 5. This may result in some small loss in accuracy in comparing groups of sites.

Adding Site 7 into Group B will be included in the second grouping structure (Table 13). There are also a few other possible groupings to consider, which may result in some more lost accuracy when comparing the groups of Sites.

Games-Howell Summary Table – log(Thickness) ~ SiteNum

Differences in Rind Thickness - Possible Additional Groupings

Sites	Std. Error	t	df	p	Mean Difference	95% CI Upper	95% CI Lower
1:09	0.05	5.76	86.31	<.001	-0.43	-0.17	-0.69
9:15	0.06	7.71	77.27	<.001	0.66	0.96	0.36
6:16	0.04	8.47	97.83	<.001	-0.51	-0.30	-0.72

Table 13:   All the confidence intervals above are reasonably close to 0 and could allow further groupings to be made with some sacrifice in accuracy. Sites 6 and 16 have highest mean rind thickness (Table 01), so they may be worth grouping. Sites 15, 1 and 9 have the next highest mean rind thickness (Table 01), and sites 15 and 1 don't have a significant

difference in mean. The only problem with grouping these sites is that the analyses were completed on log transformed data, so the high positive rind thickness values have been pulled towards 0 and therefore the real-world differences in values may be underrepresented.

Due to the large raw (without the log transformation) difference in mean rind thickness between Sites 6 and 16, highlighted in Table 01, I am hesitant to group the 2 sites together. Therefore, I propose a second grouping structure (outline below) which does not include the grouping of Site 16 and 6 and a third grouping structure which does include the grouping. This will allow the researchers to determine whether they can afford to group these two sites.

Grouping Structure 2

Group #	Sites in Group
A	1, **_9_**, 15
B	2, 3, 5, **_7_**
C	4, 8, 11
D	12, 13, 14
E	6
F	16

Table 14: *This table reflects the changes outlined (highlighted in bold underline), excluding the grouping of Sites 6 and 16.*

Grouping Structure 3

Group #	Sites in Group
A	1, 9, 15
B	2, 3, 5, 7
C	4, 8, 11
D	12, 13, 14
E	**_6, 16_**

Table 15: *The third and final grouping structure, which includes all groups of the second grouping structure with the addition of grouping Sites 6 and 16.*