STAT 550 Group Project

Agro-thermal Heat Treatment on Grapevine

Naitong Chen (21539151)
Shirley Cui (24697154)
Shannon Edie (17609181, group leader)

Contributions:

Naitong: Missing data analysis, two-way analysis
Shirley: EDA, complete case analysis, subgroup analysis
Shannon: Censored data analysis, repeated measures analysis

Abstract

In this report, the effects of agro-thermal heat treatment on various measures of grapevine quality and productivity in the Okanagan valley are investigated. Specifically, the effects of agro-thermal heat treatment are investigated in a Chardonnay and a Merlot vineyard using data from both 2019 and 2020. With the design of the study that includes a control group and a blocking factor based on the location of the grapevines, a number of parametric and non-parametric two-way analyses are conducted. Agro-thermal heat treatment does not have a statistically significant effect on most of the measures considered in the study. However, the interaction effect of heat and the blocking factor are shown to be significant for a few measures from the Merlot vineyard in 2019. This suggests that the agro-thermal heat treatment may have a signifcant impact on grapevine quality and productivity under some specific geospatial conditions.

Table of Contents

List of Tables

List of Figures

1. Introduction

This study investigates the effects of agro-thermal heat treatment on various measures of grapevine quality and productivity. Agro-thermal heat-treatment technology in grapevines (https://agrothermalsystems.com/) has been shown to increase yield, decrease the use of agro-chemicals used to fight pests, enhance wine quality and improve profits, but this technology has not been tested in the Okanagan Valley or Canada. Heat treatment is applied by driving a tractor through the vineyard rows that blows extreme heat into the canopy.

The goal of this study is to identify whether agro-thermal heat treatment influences various measures of yield, growth, and quality of Merlot and Chardonnay grapes grown at vineyards in the Okanagan Valley in the 2019 and 2020 growing seasons, analyzing the growing seasons both separately and jointly.

During the growing seasons of 2019 and 2020, heat treatment was applied to a Merlot and a Chardonnay vineyard in the Okanagan valley six times during the growing season with application ten days apart from each other. Treatments were: (1) heat, and (2) control (no heat). Outcome variables (further described in table 1) were recorded at specific time points. A blocking factor was also incorporated to help control for geospatial factors such as soil saturation or sunlight exposure that may influence outcome variables.

This report includes exploratory data analysis; handling of missing data; implementation and interpretation of several models used to analyze the effects of treatment, blocking factor, and interactions between treatment and blocking factor on the outcome variables; and a discussion of the implication of the findings in this study.

All R code used in this data analysis can be viewed at https://github.com/secedie/STAT550_Grapevine.

2. Exploratory data analysis

This section takes a closer look at the outcome variables and the relationships between them. We describe the data structure and use some data visualization methods including correlation plots and side-by-side boxplots to help look at the data before making any assumptions.

The data are collected as a two way design with one of the main factors being the block factor: grapevines are grouped in geospatial blocks; half of each block receives the agro-thermal treatment, and the other half does not; representative vines are then randomly sampled from each block to consistently collect data from so that there are an equal number of vines receiving each block/treatment combination. Vines selected for sampling are kept consistent from year to year. Outcome variables are collected for both 2019 and 2020. The Chardonnay datasets contain data for 40 grapevines (five blocks, four vines per treatment within each block), and the Merlot datasets contain data for 60 grapevines (six blocks, five vines per treatment within each block).

2.1 Response variable description

The list of variables and their descriptions are shown in table 1. Note that 50% veraison is determined by estimating the percentage of berries on a treatment vine that have changed from green to red. This assessment is done three times and the 50% veraison is calculated via regression analysis. The raw data for 2019 is missing because berries were already red (100%) on the first assessment day and 50% value could not be determined.

2.2 Correlation among outcome variables

Based on the description of outcome variables in the table 1, it is likely that some may be correlated. To investigate the relationships among these variables, correlation plots have been generated for all measurements of both years in the Chardonnay and Merlot datasets.

High correlation between two outcome variables (¿90% correlation) indicates that running tests on both variables may be redundant. Moderate correlation may indicate that an analysis of treatment effects on the variables can be investigated simultaneously using a two-way MANOVA.

Table 1. Descriptions of all outcome variables recorded in this study. The vineyard for which each variable is recorded (Chardonnay or Merlot), along with the year(s) in which each variable is recorded (2019, 2020, or both 2019 and 2020), is also noted.

Variable	Definition and collection method	Vineyard	Year
# of clusters	the cluster count for each vine at the time of harvest in October	Both	Both
Yield (kg)	yield in kg for each vine at the time of harvest in October	Both	Both
Average cluster weight (kg)	total grapevine yield divided by number of grape clusters	Both	Both
Leaf greenness (SPAD)	average value of ten randomly subsampled leaves measured in July	Merlot	Both
50% veraison (days after August 1st)	time by which half of the berries on a treatment vine have changed from green to red	Merlot	Both
Average berry weight (kg)	average weight of thirty randomly subsampled berries measured in October	Merlot	Both
Average # of berries per cluster	cluster weight (in grams) divided by the berry weight (in grams)	Merlot	Both
Berry quality: pH, TA, and Brix	measurements from the same set of randomly subsampled berries, may influence quality of wine	Merlot	Both
Brown seed color	visually accessed seed color change once grapes have matured	Merlot	2020
Pruning weight (kg)	the pruned cane weight after harvest	Merlot	Both
Ravaz index	yield divided by pruning weight	Merlot	Both
Fruitfulness	the number of clusters emerged in Spring divided by the number of new shoots	Merlot	2020
50% Bloom (days after June 1st)	time by which half of flowers on a treatment vine have started to bloom	Merlot	2020

Most of the correlations are acceptable, but in Chardonnay 2020, the correlation between the number of clusters and the yield seems to be quite high (corr = 0.83), which indicates we may need to conduct MANOVA to incorporate this relation into the model (Figs. 1 and fig. 2).

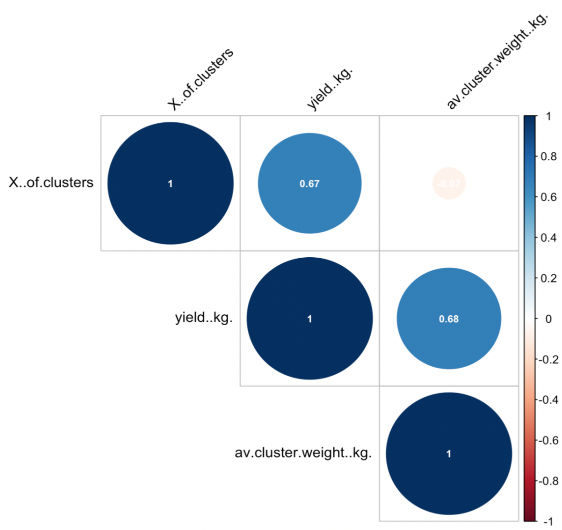


Fig. 1. Correlation plot for Chardonnay 2019

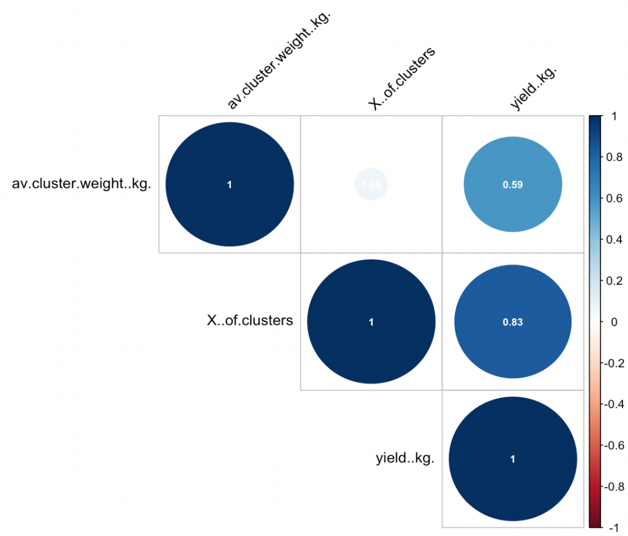


Fig. 2. Correlation plot for Chardonnay 2020

Correlation plots of outcome variables for Merlot 2019 and 2020 are included in Appendix A[5]. From the plots, most of the correlations are not high. However, the cluster weight is highly correlated with berry.cluster variable in both years (corr = 0.94 in 2019 as shown in fig. 12, corr = 0.88 in 2020 as shown in fig. 14). This is reasonable because the variable berry.cluster is calculated based on the ratio of cluster weight and berry weight. Yield is also found to be highly correlated with cluster weight in 2019 (corr=0.76), and highly correlated with cluster number in 2020 (corr = 0.92). Both correlations can also be reasonably explained as that yield itself depends on the number of clusters and cluster weight.

With the existence of high correlations among some outcome variables in both vineyards, we could consider to use MANOVA to analyze the data as mentioned above. However, since each outcome variable is of interest, and that the size of our samples may be too small to handle MANOVA models which contain a large number of parameters to be estimated, we perform two-way ANOVA on each outcome variable in both vineyards. The high correlation, however, indicates that any conclusions regarding these three variables should be similar.

2.3 Distribution of outcome variables

Potential differences in control and treatment group were investigated by plotting the outcome variables according to treatment and block (Appendix A[5]). Plots indicating potentially significant findings are discussed here. Preliminary data exploration included running t-tests to detect within-block differences in treatment versus control (shown in the boxplots).

For Chardonnay 2019, outcome variables with different treatments are distributed similarly among different blocks, except that there might be a difference in the average of cluster weight between heat and control group in the third block. The plot can be found in appendix A[5]

In Chardonnay 2020, yield is found to be quite different between the treatment groups in the first block, as shown in fig. 3. Both p-values from t-test and Wilcoxon test provide a significant result, which may indicate that there is a significant influence of heat to the yield inside this specific block.

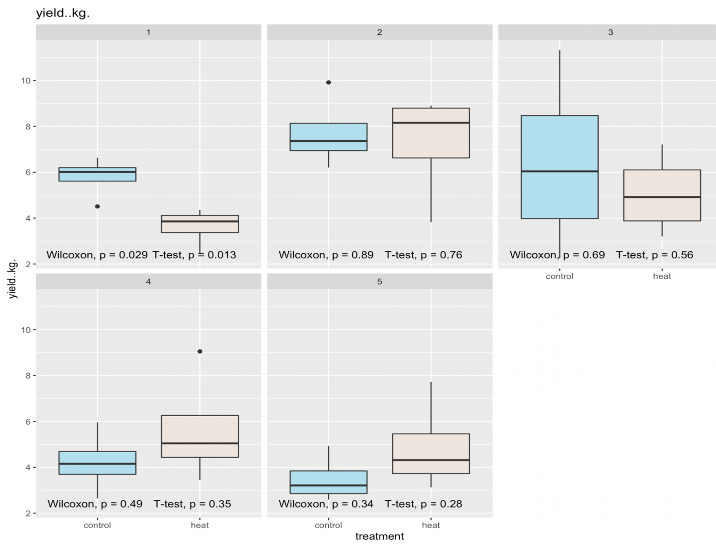


Fig. 3. Side-by-side boxplots of yield in Chardonnay 2020, divided by block.

For Merlot 2019, we have removed the side-by-side box plot for veraison. Veraison and bloom will be discussed in section 3.1 as censored data. Based on the plots shown in appendix A, several outcome variables seem to have quite different distributions in the Merlot 2019 dataset. For example, cluster weight, berry weight, berry PH and some other variables are found to have different distributions based on the p-values from the t-test and Wilcox test. fig. 4 shows in detail how the distributions of cluster weight under heat and control seem to be substantially different in block 5.

The Merlot 2020 dataset containe many missing values due to zero yield that year. Some plots became less informative after dropping the missing values, as shown in appendix A. Among all the distributions of outcome variables, the measurements of SPAD seem to have important difference in distribution under treatment groups in block 2, as shown in fig. 5

2.4 Preliminary Conclusion

For most outcome variables, treatment effect do not appear significantly influence the outcome measurements. However, within particular blocks, some variables have distinct distributions under different treatment groups. This may indicate an important role of interaction between geographic area and treatments (heat/control), but a confirmatory analysis needs to be carried out to justify these observed differences.

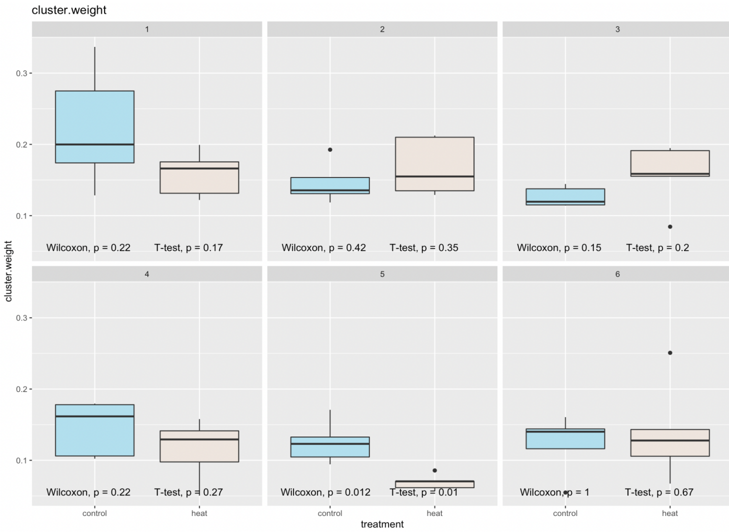


Fig. 4. Side-by-side boxplot of cluster weight in Merlot 2019

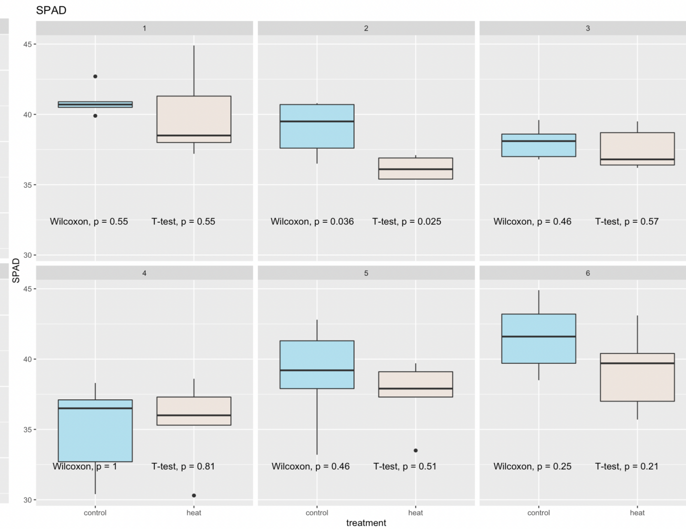


Fig. 5. Side-by-side boxplot of SPAD in Merlot 2020

3. Handling missing data

There are a number of response variables that contain missing values (fig. 6). In addition, 43.3% of the veraison values are also missing from Merlot 2019. Depending on the reason why these values are missing, simply ignoring them may cause the analysis result to be biased. Below, we discuss the outcome variables with missing values, and how we handle these missing data.

bloom	veraison	yield related measures	berry quality related measures	brown seed color	pruning weight	Ravaz index
11.7%	13.3%	11.7%	13.3%	26.7%	13.3%	38.3%

Fig. 6. List of response variables and their missing rates in Merlot 2020

3.1 Censored data

Censored data is a special type of missing data where we have some information about the missing value, but not all the information. This is the case for Bloom and Veraison.

The Bloom and Veraison variables are both measurements of when an event (50% bloom or veraison) occurs in time. Bloom and veraison were measured at three separate time points (hereafter T1, T2, and T3), and a linear regression was used to approximate the 50% bloom/veraison time for each grapevine in the study. Depending on whether the vine produced any berries, the missing value is known to be in a certain range. In other words, these missing values are partially known. Data of this nature are known as censored data. A visualization of this censored data can be seen in Appendix A3.

The bloom and veraison data are missing because either the plant had already reached 100% bloom or veraison before T1. Other times, these data are missing because the plant did not produce any blooms or berries. In either case, we have a little bit of information about when 50% bloom or veraison occurred– either before T1 or after T3, respectively. Thus, instead of interpreting 50% Bloom and Veraison as occurring on fixed dates, we can interpret these events (50% bloom or veraison) as occurring within certain time intervals. There are three cases:

1. The event did not yet occur because no berries or flowers grew. This is known as right-censored data. In this case, the event happened in the interval $(T3, \infty)$

2. The event already occurred prior to T1. This is known as left-censored data. In this case, the event happened sometime in the interval $(-\infty, T1)$.

3. The event was observed. In this case, the ”interval” is a fixed number indicating when the event happened.

This analysis could be further modified by more accurate biological interpretations of the intervals– for example, putting the earliest date that the vines may have realistically reached 50% veraison for left-censored data.

3.2 Non-censored data

Outcome variables that contained missing data that could not be interpreted as censored include yield-related measures, berry quality measures, and pruning weight and Ravaz index.

For all columns that contain missing values and are related to yield (yield, cluster weight, berries/cluster, berry weight), the reason they are missing is that the vine had no yield. Consequently, these data were not truly ”missing”– they have a ”true” value of zero to indicate zero yield.

Berry quality related measures (including berry brix, TA, and pH; and brown seed color) are also missing due to zero yield. However, the quality of the berry may actually be related to the other variables in the dataset.

Pruning weight was missing because the winery had started to prune canes before data could be collected; these data are missing completely at random.

These reasons for missing do not necessarily imply that ignoring these missing values would result in biased analysis. However, by removing the missing measurements, the balanced design of the experiment would be broken. In other words, there would no longer be an equal number of measurements under each treatment and blocking combination. This imbalanced structure would likely complicate interpreting the results of the statistical analysis.

Multiple imputation is used to deal with these missing values. Multiple imputation involves using the observed values to make several predictions on each of the missing values. This approach is suitable for this study as it does not remove any of the observed values, so the balanced structure stays intact. By having multiple predicted values on each value that is missing, we can reflect the additional uncertainties introduced by not observing these missing data. Finally, once the predictions are made on each missing value, we construct several of complete datasets. Standard analyses can be carried out on each of these complete datasets and final conclusions can be drawn after the results from each of the complete datasets are pooled together.

For simplicity, linear regression models are used to make predictions of the missing values. And as suggested by [1], the final conclusions of the subsequent analyses will be drawn after combining five imputed datasets.

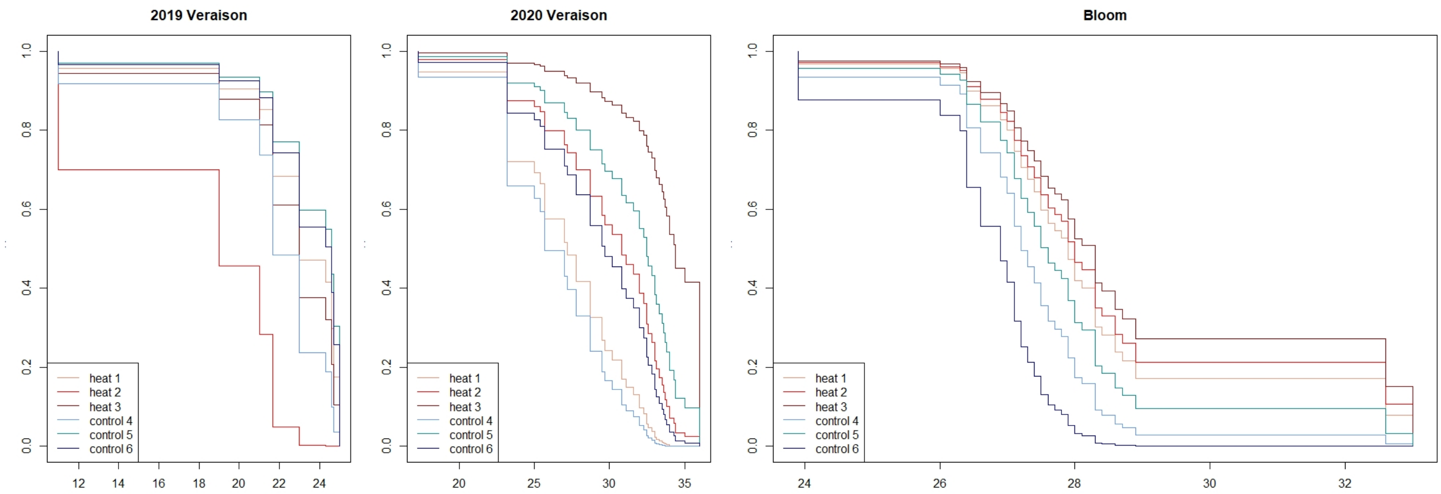
4. Statistical analysis

4.1 Censored data analysis

We use a Cox Proportional Hazards (Cox PH) model to investigate the effects of treatment and blocking on the censored variables: veraison and bloom. This model is commonly used in clinical data analysis to model patient survival in clinical studies. A Cox PH model is expressed by a hazard function, $h(t)$, that can be interpreted as the risk of the event (in this case, ”risk” of 50% veraison or bloom) at time t, based on a set of covariates $x_1, ..., x_n$:

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)$$

In this study, the covariates include the treatment, block, and interaction effects. The Cox PH model estimates coefficients for these covariates, and then much like for a linear regression model, we can look at the significance of the coefficients to see if the covariates are significant. The specific R package used to implement the Cox PH models uses bootstrapping so we can make these inferences on the coefficients.



(a) Merlot 2019 and 2020 veraison Cox PH model

(b) Merlot 2020 bloom Cox PH model

Fig. 7. Survival functions for the fitted Cox PH models constructed for 50% bloom and 50% veraison for the Merlot 2019 and 2020 datasets. "Survival" indicates the probability that, at time t, the event (50% bloom or veraison) has not yet occurred.

Inferences on the effect of treatment, block, and interaction are based on bootstrapped estimates of the coefficients for these covariates. The interaction effect between treatment and block significantly is shown to affect v raison in the 2020 Merlot dataset (Z=-2.211, p=0.027). Otherwise, there is no significant evidence indicating that the block, treatment, or interaction effects influenced bloom or veraison. Coefficients for block, treatment, and interaction effects for veraison in the 2019 dataset range from 0.42 to 0.98; for veraison in the 2020 dataset range from 0.10 to 0.91, and for bloom in the 2020 dataset range from 0.14 to 0.98.

Cox PH models are dependent on the proportional hazards assumption, which can be roughly checked by seeing if the estimates for the survival curves overlap. Based on the curves in section 4.1, these curves do not appear to overlap, so we may assume that the proportional hazards assumption roughly holds.

4.2 Non-censored data analysis

In each of the four datasets, there are both the treatment factor and blocking factor to be considered, and there are an equal number of measurements for each response under each treatment/block combination. A two-way layout fits the objective of analyzing the effect of heat treatment. As mentioned in section 2, since each outcome variable is of individual interest, and the sample sizes may be too small to handle multivariate analysis, a univariate two-way ANOVA model is most appropriate.

4.2.1 Assumption checks

Two-way ANOVA models assume that the measurements under each treatment/block combination have an equal variance and are normally distributed. These assumptions are checked before fitting two-way ANOVA models to the data to know whether an ANOVA is an appropriate model.

From the boxplots in section 2, the equal variance assumptions may have been slightly violated. We ran Bartlett and Levene tests on the data to check equal variance assumptions. Bartlett tests indicated potential breach of equal variance for yield in Merlot 2020 (p=0.0039); for cluster weight, berry pH, and pruning weight (p=0.029, p-0.0036, p=0.017, respectively); and for number of clusters for the Chardonnay 2020 data.

To check normality assumptions, a set of Shapiro-Wilkes tests are conducted on each response. The Q-Q plots of each treatment/block combination that show a significant (5% significance level) departure from being normally distributed from the Shapiro-Wilkes test are shown in Appendix B [5]. Note that the red dots in the Q-Q plots for Merlot 2020 indicate the imputed values.

Both assumptions are slightly violated for a number of responses in each of the four datasets, so two-way ANOVA models may not be fully appropriate. To check the conclusions drawn from our analyses, non-parametric tests are also conducted. By checking whether the conclusions from both tests are consistent, we can further determine the reliability of these conclusions.

The choice of non-parametric test is the Aligned Rank Transform (ART) ANOVA model described in [2]. The ART ANOVA model is suitable because it does not impose any assumptions on the distribution, variance, or the size of the data. It is also capable of testing both the main factors and the interaction effects. The test statistics are also shown to be approximately follow F distributions. Then under the presence of multiple imputed datasets, methods that combine parametric two-way ANOVA models' test results can also be used justifiably the ART ANOVA models.

4.2.2 Two-way analysis

The blocking factor is significantly related to several outcome variables at the 5% significance level. This confirms the intention behind the design of the study, which isolates the blocking effects from the treatment effects. We focus on mostly the treatment and interaction effects below.

For Chardonnay 2019 and 2020, since there are no missing values, standard two-way ANOVAs and ART ANOVAs are carried out. The p-values for the treatment, blocking, and interaction effects of each response variable are shown in fig. 15. The p-values in the parametric and non-parametric models are mostly similar, indicating that the conclusions between the models are consistent. There are no significant results at the 5% level for any of the responses tested.

For the Merlot 2019 dataset, both the berries per cluster and berry pH have moderate evidence indicating that the treatment may have a significant influence on these variables (fig. 8), although neither of these treatment effects are significant at the 5% significance level under both ART and ANOVA models.

The interaction effects of Berry weight, berry pH, pruning weight, and Ravaz index are significant at the 5% significance level under both ART and ANOVA models.

These results align with our observation from exploratory data analysis that some of the treatment effects and interaction effects may be significant.

	treatment	block	interaction
SPAD	0.27	0.00	0.53
cluster.number	0.28	0.47	0.99
yield	0.08	0.01	0.28
cluster.weight	0.24	0.00	0.06
berries.cluster	0.06	0.00	0.05
berry.weight	0.19	0.00	0.00
berry.TA	0.49	0.02	0.44
berry.pH	0.09	0.00	0.01
berry.Brix	0.25	0.11	0.71
pruning.weight	0.80	0.04	0.03
Ravaz.index	0.11	0.42	0.01

(a) p-values from two-way ANOVA on responses with missing values in Merlot 2020

	treatment	block	interaction
SPAD	0.27	0.00	0.64
cluster.number	0.29	0.38	0.98
yield	0.19	0.01	0.49
cluster.weight	0.19	0.00	0.05
berries.cluster	0.05	0.00	0.06
berry.weight	0.25	0.01	0.01
berry.TA	0.42	0.03	0.43
berry.pH	0.03	0.00	0.00
berry.Brix	0.58	0.19	0.83
pruning.weight	0.88	0.02	0.01
Ravaz.index	0.17	0.34	0.01

(b) p-values from ART ANOVA on responses in Merlot 2019

Fig. 8. p-values from two-way analysis on responses in Merlot 2019

Out of all of the responses without missing values in the Merlot 2020 dataset (fig. 9), only SPAD shows a significant treatment effect under the ART ANOVA model at the 5% significance level. However, the corresponding p-value in the ANOVA model is slightly about 5%. Interaction effects are not significantly related to any outcome variable.

Since there are missing values in the Merlot 2020 dataset, it is necessary to address how the parametric and non-parametric ANOVA results from each of the five imputed datasets are pooled. The approach from [3] is used to obtain a set of pooled p-values for each response variable based on the fact that both the test statistics of the ART and ANOVA models follow F distributions approximately.

Among these response variables, pruning weight indicates moderate evidence that treatment and interaction effects may be significant, although p-values are not below 5% under both models (fig. 9).

The power of these tests are another important consideration. The power of these tests, or the ability for these tests to identify a significant finding when a significant finding truly exists, is primarily dependent on sample size and expected effect size. The expected effect size in these tests is the ratio of the desired difference in means to the standard deviation of the control group. The observed effect sizes in this data are quite small (around 0.05-0.2); this, combined with small sample size, means that power for this study may be low. For reference, a contour plot was constructed which depicts the power of a two-way ANOVA test based on expected effect size, given the sample sizes in the Merlot 2019 dataset (fig. 16).

	treatment	block	interaction
fruitfulness	0.14	0.13	0.31
SPAD	0.06	0.00	0.69
cluster.number	0.62	0.24	0.32
yield	0.76	0.20	0.20
cluster.weight	0.69	0.07	0.18
berries.cluster	0.61	0.16	0.22
berry.weight	1.00	0.16	0.82

(a) p-values from two-way ANOVA on responses with missing values in Merlot 2020

	treatment	block	interaction
fruitfulness	0.08	0.08	0.14
SPAD	0.05	0.00	0.66
cluster.number	0.67	0.37	0.52
yield	0.80	0.40	0.39
cluster.weight	0.70	0.11	0.18
berries.cluster	0.35	0.14	0.17
berry.weight	0.91	0.25	0.90

(b) p-values from ART ANOVA on responses without missing values in Merlot 2020

	treatment	block	interaction
berry.Brix	0.21	0.43	0.06
berry.TA	0.90	0.12	0.25
berry.pH	0.87	0.67	0.70
brown.seed.color	0.79	0.03	0.24
pruning.weight	0.07	0.21	0.10
Ravaz.index	0.82	0.00	0.94

(c) p-values from two-way ANOVA on responses with missing values in Merlot 2020

	treatment	block	interaction
berry.Brix	0.18	0.38	0.12
berry.TA	0.82	0.09	0.20
berry.pH	0.98	0.48	0.68
brown.seed.color	0.64	0.03	0.26
pruning.weight	0.10	0.29	0.05
Ravaz.index	0.93	0.01	0.84

(d) p-values from ART ANOVA on responses with missing values in Merlot 2020

Fig. 9. p-values from two way analysis on responses in Merlot 2020

4.2.3 Complete-case comparison

Since multiple imputation techniques are carried out to analyze the data in Merlot 2020, before deriving conclusions on the outcome variables with missing values based on the imputed Merlot 2020 datasets, we run a complete-case test as an additional reference to see how sensitive our analysis is to the outcome variables with missing values.

A complete randomized block design is created by first removing all the missing values under the particular column variable. To balance the design, the average of the remaining response values under each treatment and block are taken for simplicity. Note there is only one measurement in each block under each treatment in current situation; hence, a randomized block design does not consider interaction effect, resulting in a loss of information.

The results from the ANOVA are consistent with previous conclusions in terms of the treatment factor, but not with previous conclusions regarding the blocking factor (fig. 10). Blocking factor is significant for brown seed color and Ravaz index in the previous analyses, but not significant in this complete-case design analysis. This may be caused by loss of information from complete case design as discussed above.

Fig. 10. p-values from randomized block complete-case analyses on responses with missing values in Merlot 2020

	berry.Brix	berry.TA	berry.pH	brown.seed.color	pruning.weight	Ravaz.index
treatment	0.27	0.84	0.71	0.80	0.23	0.87
block	0.83	0.31	0.11	0.45	0.56	0.19

4.2.4 Repeated measures analysis

The 2019 and 2020 data are combined into one analysis by using a two-way repeated-measures ANOVA. This model is very similar to the two-way ANOVA implemented earlier, but this model uses both 2019 and 2020 data, and takes into account within-individual and within-year variation. A standard two-way ANOVA cannot be used to combine the 2019 and 2020 data because it does not model the correlation between the repeated measures for each vine, and because the data violates the assumption of independence (naturally, a measurement from the same vine in 2019 will not be independent from a measurement from that same vine in 2020).

Results from the repeated-measures ANOVA for the Chardonnay dataset are consistent with the findings of the two-way ANOVA and ART models for separate years (2). No significant interaction or treatment effects are observed, but blocking is significant for all three variables observed.

Most results for the repeated-measures ANOVA for the Merlot dataset are consistent with the findings of the two-way ANOVA and ART models for separate years (3), particularly with the findings from 2019.

Blocking is significant for virtually every factor except for cluster number and Brix berry quality measure. There is moderate evidence that berry pH and yield may be influenced by the treatment, but this finding was not significant at the 5% significance level. Finally, interaction effects are significant for berries per cluster, berry weight, and berry pH

There are some inconsistencies in conclusions from the repeated-measures analysis. There is moderate evidence that treatment significantly affects the number of berries per cluster. There is also no evidence that treatment significantly influenced SPAD. This may be due to the fact that repeated-measures models further isolated the effects under consideration from the within individual variation.

Results for the outcome variables with missing data are based on pooled results from the imputed datasets, similar to the approach above. Results for Ravaz index and pruning weight must still be computed, as the 2020 Ravax index and pruning weight data was received only recently.

4.2.5 Subgroup analysis

For the four outcome variables with significant (at the 5% significance level) interaction effects found by the ANOVA and ART models, we conduct a subgroup analysis to further study their pairwise differences. Specifically, we use pairwise t-tests to quantify the difference of pairs of interaction effects.

Because the p-values are similar and the conclusions are consistent between both the parametric and non-parametric models, the slight violation of the ANOVA assumptions likely does not greatly impact these particular

	block	interaction	treatment
Number of clusters	0.015	0.142	0.134
Average cluster weight (kg)	0.069	0.350	0.527
Yield (kg)	0.027	0.759	0.559

Table 2. The p-values for a two-way repeated-measures ANOVA on the Chardonnay data are displayed below.

Var	block	block:treatment	treatment
SPAD	0.001	0.533	0.266
berries.cluster	0.004	0.049	0.064
berry.Brix	0.086	0.708	0.247
berry.TA	0.014	0.431	0.486
berry.pH	0.000	0.004	0.079
berry.weight	0.003	0.003	0.195
cluster.number	0.467	0.988	0.282
cluster.weight	0.001	0.064	0.237
yield	0.009	0.285	0.082

Table 3. The p-values for a two-way repeated-measures ANOVA on the Merlot data are displayed below.

models. Thus, we proceed with multiple comparisons using Tukey's method, which has assumptions similar to those of the ANOVA. Significant pairwise comparisons are reported in fig. 11.

Although the Ravaz index has significant interaction effects, none of the pairs in multiple comparison are significant. Looking at the boxplots of the Ravaz index variable in Appendix A [5], it is clear that its variance is much higher than the other three responses that have significant interaction effects. In addition, to keep the overall significance level at 5%, Tukey's method may be overly conservative as we simultaneously compare 66 pairs of interaction effects. These two factors together may explain why we observe no significant pairs of interaction effects.

However, overall, these results still suggest that heating has some significant effect, either positive or negative, under some specific geospatial conditions.

group1	group2	estimate	conf.low	conf.high	p.adj
heat:1	heat:5	-0.38	-0.68	-0.07	0.00
heat:1	control:6	-0.34	-0.65	-0.04	0.01
heat:2	heat:5	-0.33	-0.64	-0.03	0.02
heat:2	control:6	-0.30	-0.60	0.00	0.05
heat:3	heat:5	-0.30	-0.61	0.00	0.05
control:4	heat:5	-0.35	-0.65	-0.04	0.01
control:4	control:6	-0.31	-0.62	-0.01	0.04
heat:5	heat:6	0.36	0.05	0.66	0.01
control:6	heat:6	0.32	0.02	0.63	0.03

(a) Merlot 2019 berry weight

group1	group2	estimate	conf.low	conf.high	p.adj
heat:1	heat:5	0.21	0.06	0.36	0.00
control:2	heat:5	0.23	0.08	0.38	0.00
heat:2	heat:5	0.21	0.07	0.36	0.00
control:3	heat:5	0.17	0.03	0.32	0.01
heat:3	heat:5	0.15	0.00	0.30	0.05
control:4	heat:5	0.24	0.09	0.39	0.00
heat:5	control:6	-0.21	-0.36	-0.06	0.00
heat:5	heat:6	-0.17	-0.32	-0.03	0.01

(b) Merlot 2019 berry pH

group1	group2	estimate	conf.low	conf.high	p.adj
control:1	control:6	-0.33	-0.65	-0.02	0.03

(c) Merlot 2019 pruning weight

Fig. 11. Pairwise comparisons for significant interaction effects

5. Discussion

Except for berry weight, berry pH, pruning weight, and Ravaz index in Merlot 2019, neither of the treatment nor the interaction effects for all outcome variables are statistically significant at a 5% significance level. For the four responses mentioned above, the interaction effects are significant at the 5% significance level, but the treatment effects are not. This suggests that for Merlot in the Okanagan Valley, the combination of agro-thermal treatment and some specific geospatial features combined has some influence on the four variables listed above.

High rates of missing data in certain response variables, the presence of outliers, and minor violations in the assumptions for the statistical tests used may affect the accuracy of the conclusions drawn in this study.

The sample sizes are small. In addition to making multivariate analysis difficult, small sample size also affects the power of the tests. Current sample sizes, and observed small differences in treatment/block effect may have led to a small power for our test.

As a direction for future analysis, mixed effect models may be used to test the effect of heating after combining both years' measurements. This is possible because the measurements were taken on the same set of vines in both years. Instead of fitting models on each year separately, with a mixed effect model we may better isolate the treatment effects by further accounting for the within individual variation. Mixed effects models may be more appropriate for this dataset than the repeated-measures ANOVA because they can better handle both blocking factors and missing data.

References

[1] Rubin DB (2004) *Multiple imputation for nonresponse in surveys*. Vol. 81 (John Wiley & Sons), .

[2] Toothaker LE, Newman D (1994) Nonparametric competitors to the two-way anova. *Journal of Educational Statistics* 19(3):237–273.

[3] Grund S, Lüdtke O, Robitzsch A (2016) Pooling anova results from multiply imputed datasets. *Methodology* .

Appendix A: Supplementary Plots

Part I: Additional plots in exploratory data analysis

1. Correlation plots for outcomes variables in Merlot vineyard

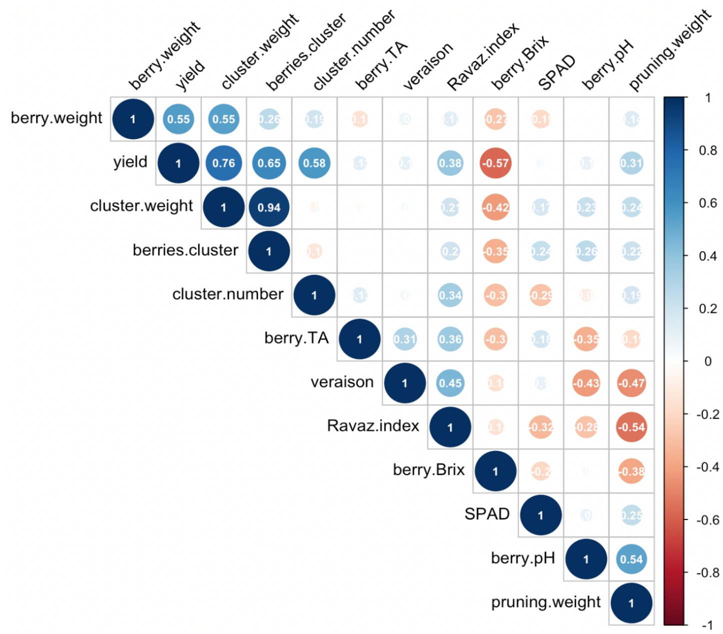


Fig. 12. Correlation plot for Merlot 2019

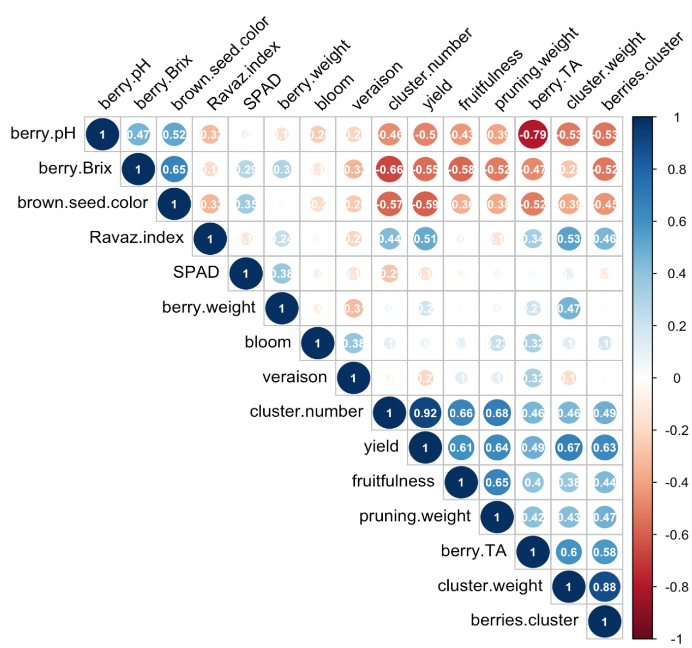
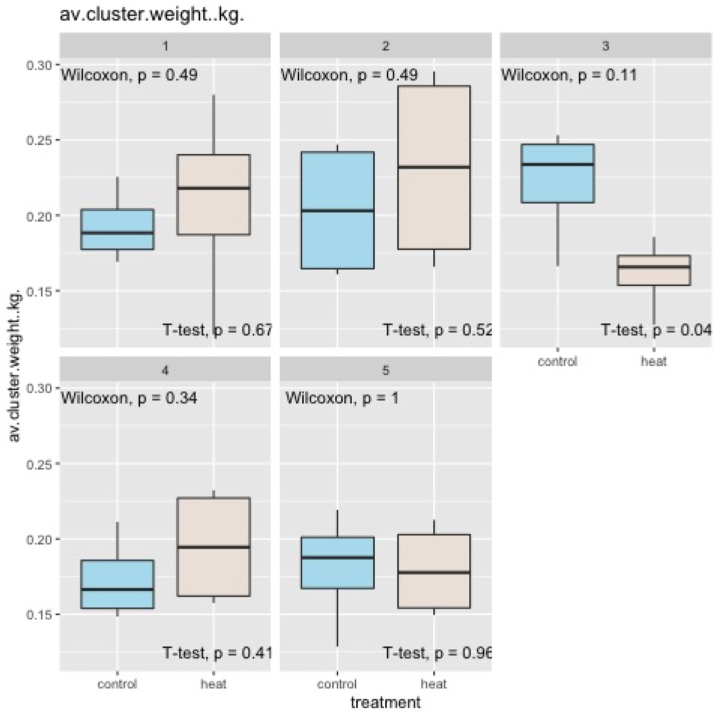
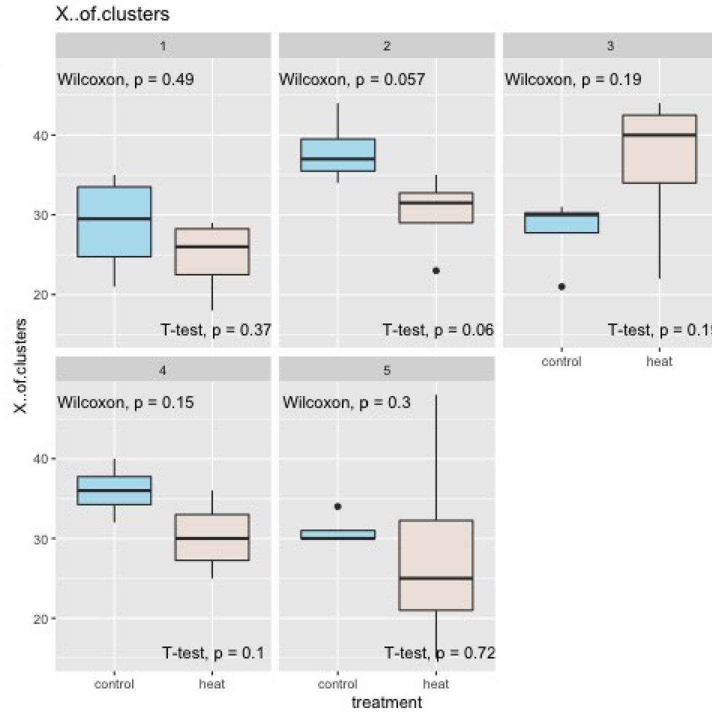


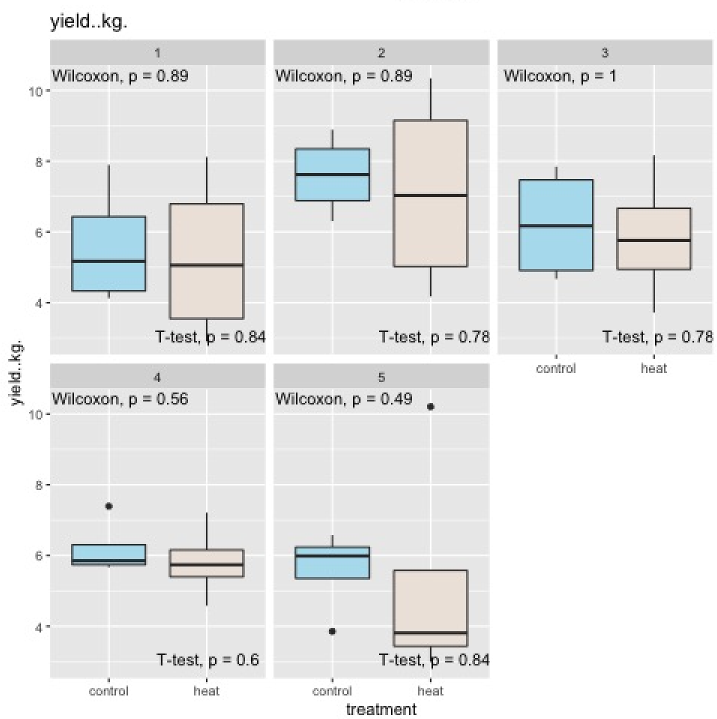
Fig. 13. Correlation plot for Merlot 2020

2. Side-by-side boxplots

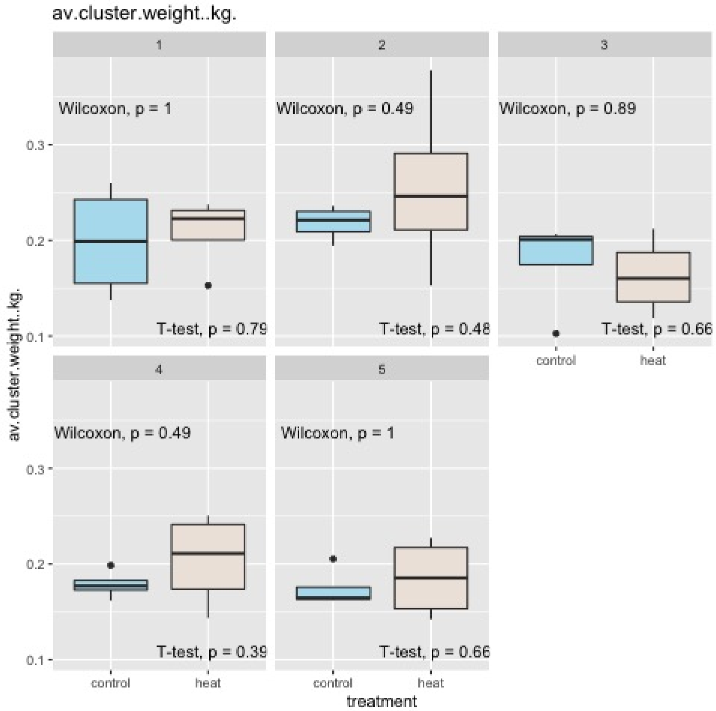
2.1 Chardonnay 2019: side-by-side boxplots under treatment and block combination

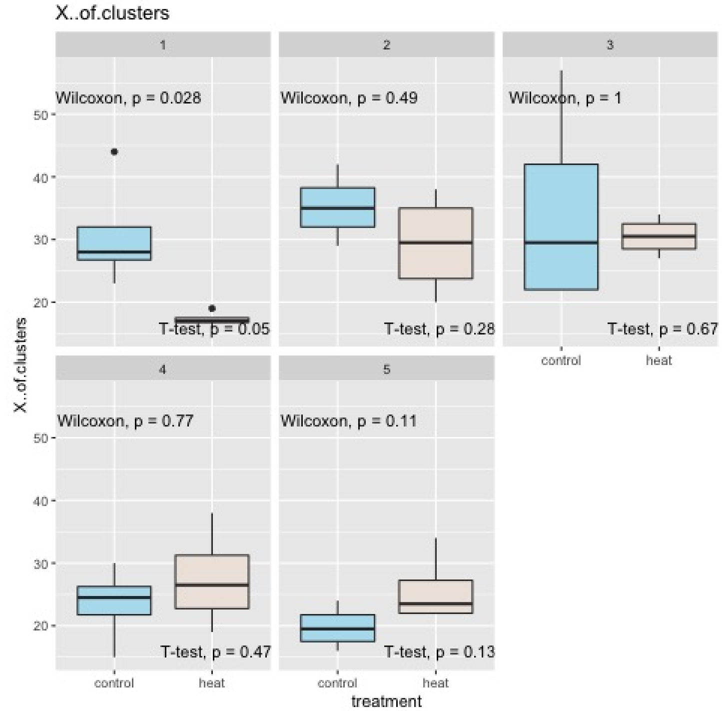


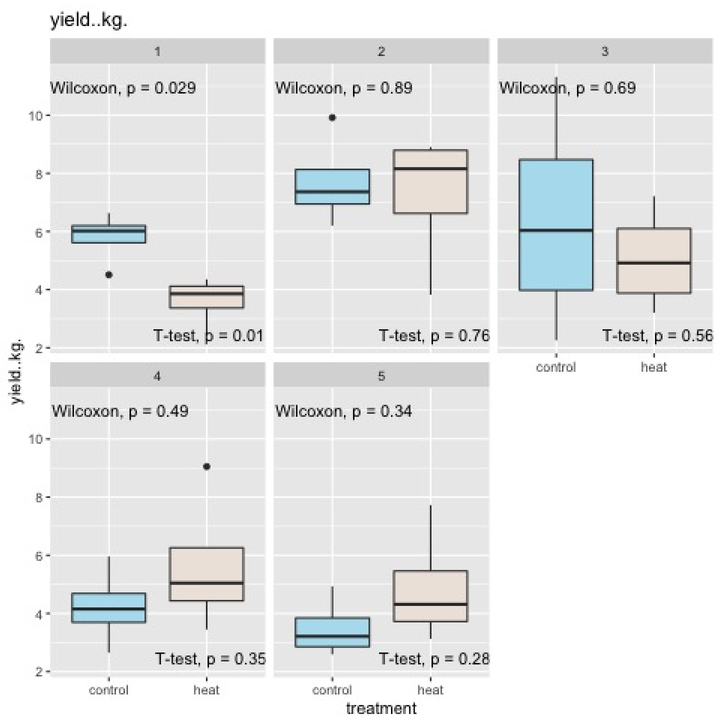




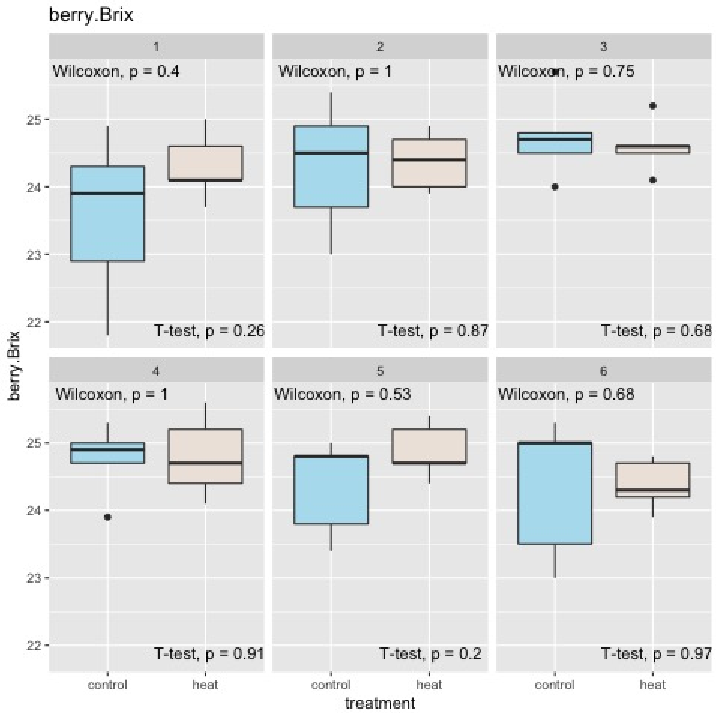
2.2 Chardonnay 2020: side-by-side boxplots under treatment and block combination

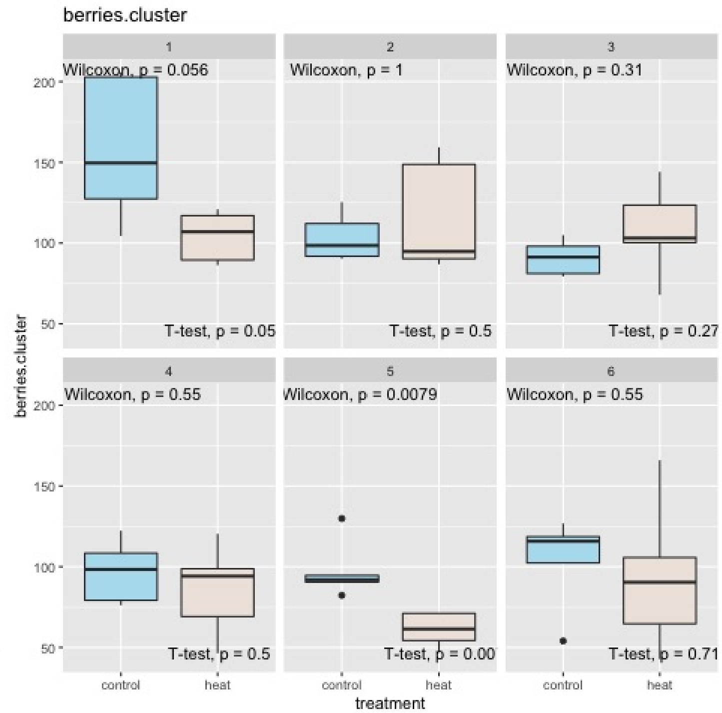


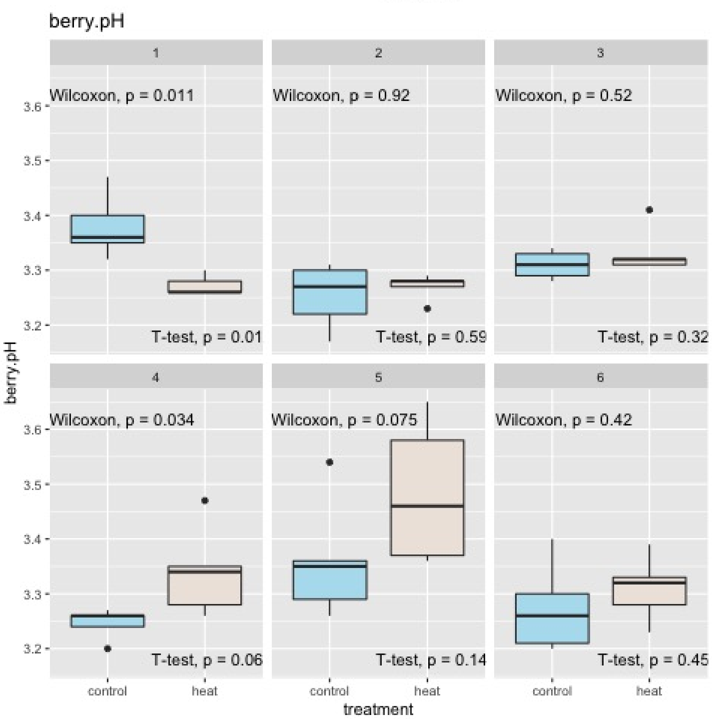


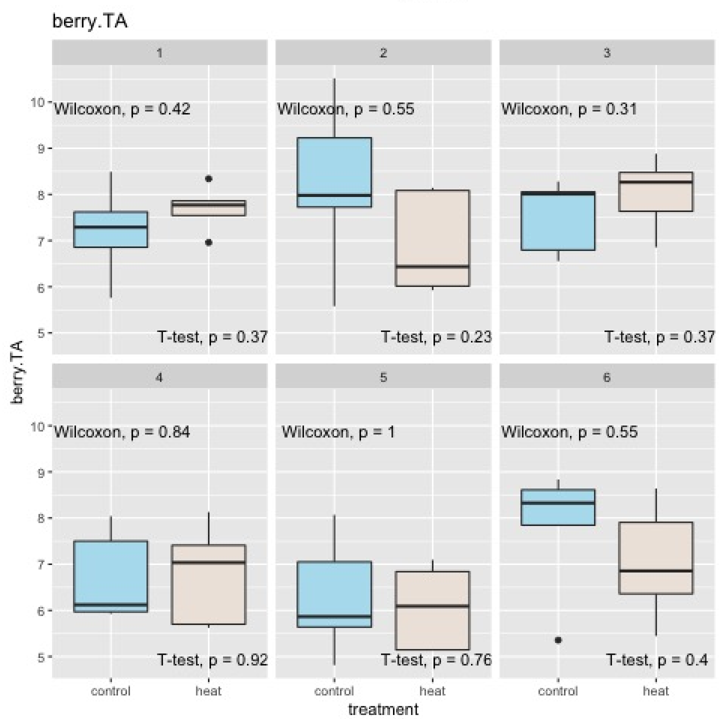


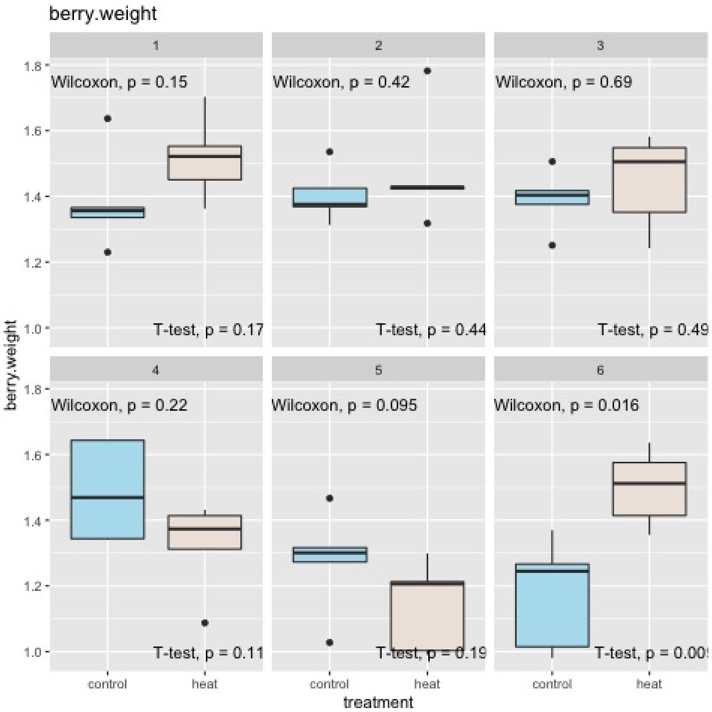
2.3 Merlot 2019: side-by-side boxplots under treatment and block combination

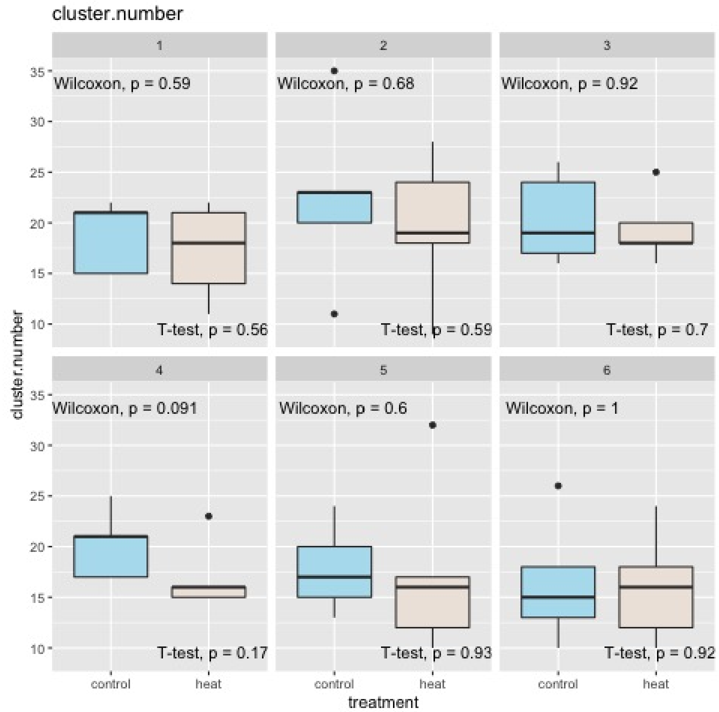


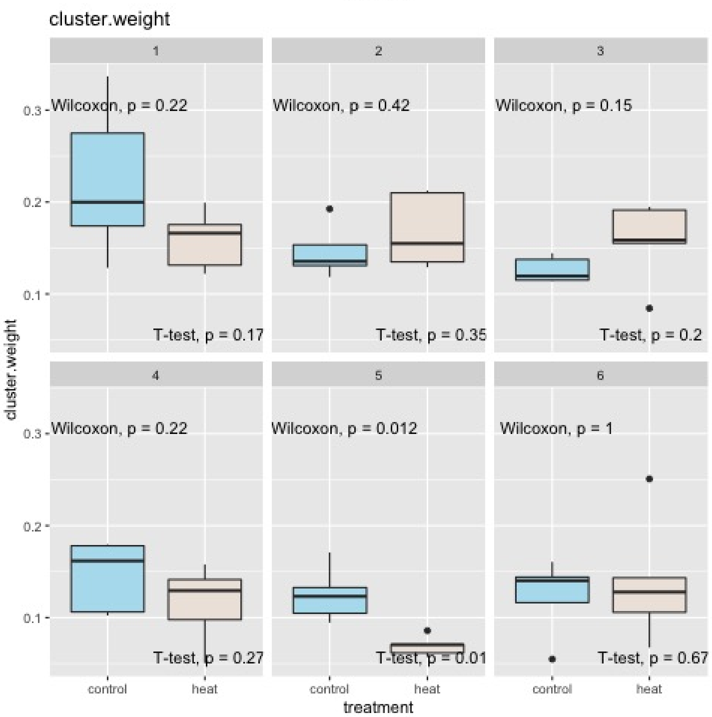


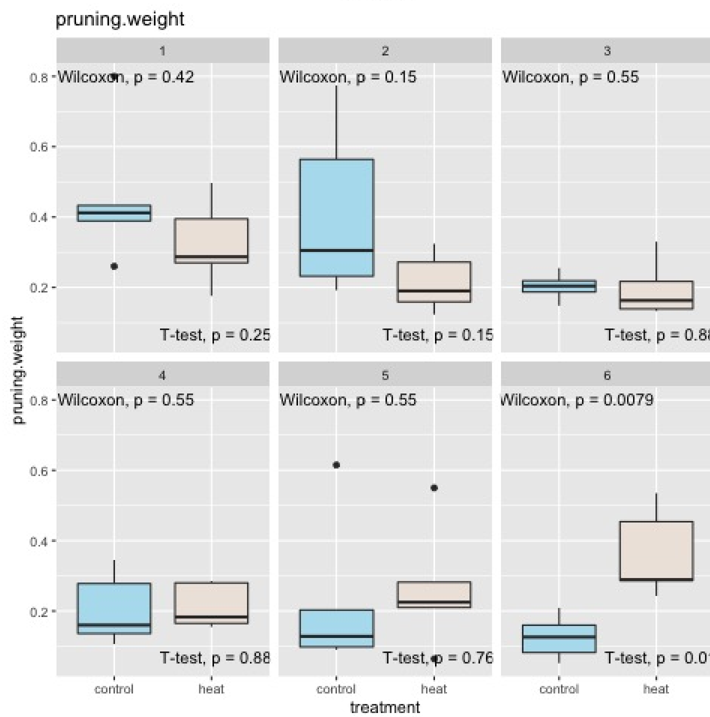


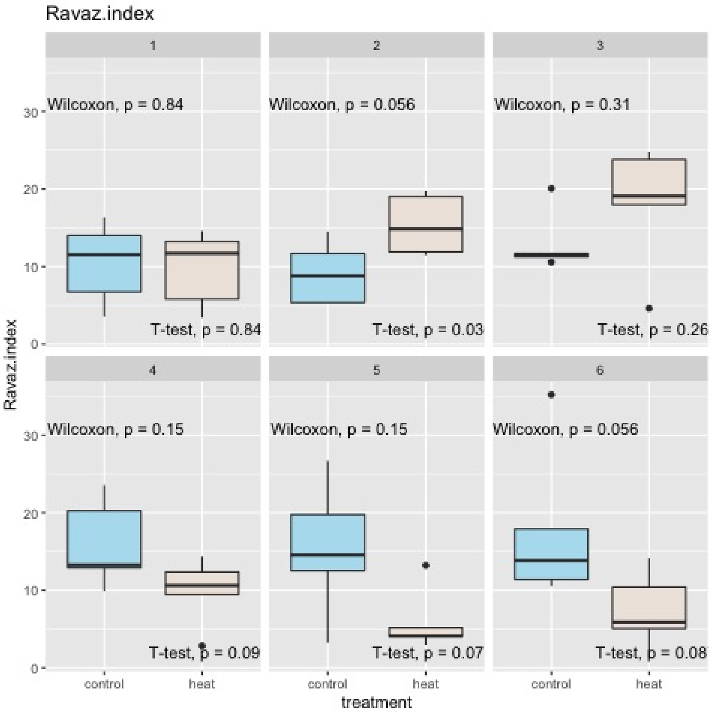


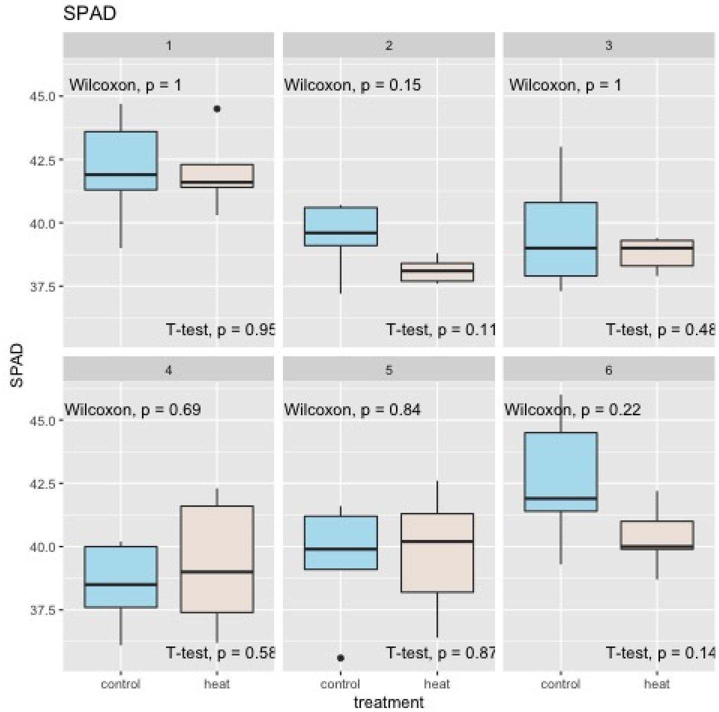

berry.weight
Wilcoxon, p = 0.15
T-test, p = 0.17
Wilcoxon, p = 0.42
T-test, p = 0.44
Wilcoxon, p = 0.69
T-test, p = 0.49
Wilcoxon, p = 0.22
T-test, p = 0.11
Wilcoxon, p = 0.095
T-test, p = 0.19
Wilcoxon, p = 0.016
berry.weight
treatment
control
heat

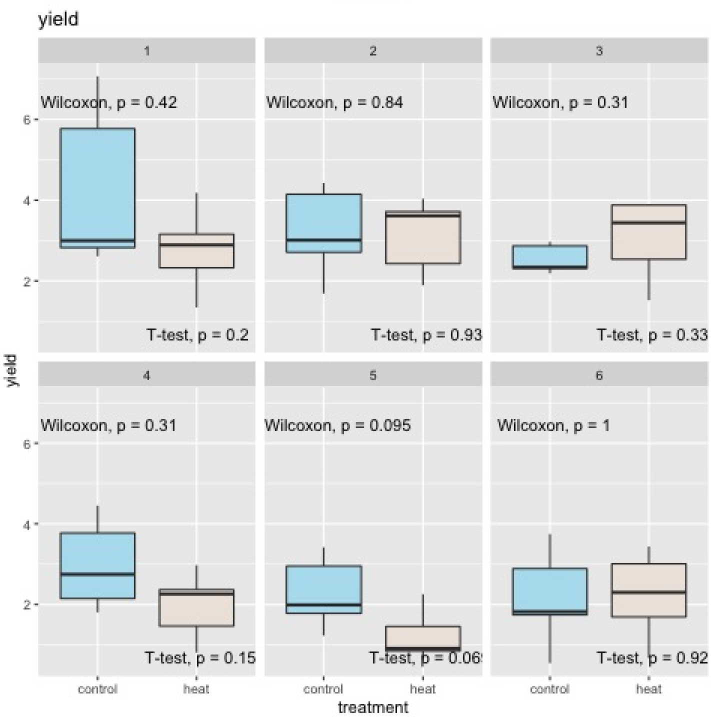

cluster.number
Wilcoxon, p = 0.59
T-test, p = 0.56
Wilcoxon, p = 0.68
T-test, p = 0.59
Wilcoxon, p = 0.92
T-test, p = 0.7
Wilcoxon, p = 0.091
T-test, p = 0.17
Wilcoxon, p = 0.6
T-test, p = 0.93
Wilcoxon, p = 1
T-test, p = 0.92
cluster.number
treatment
control
heat


cluster.weight
Wilcoxon, p = 0.22
T-test, p = 0.17
Wilcoxon, p = 0.42
T-test, p = 0.35
Wilcoxon, p = 0.15
T-test, p = 0.2
Wilcoxon, p = 0.22
T-test, p = 0.27
Wilcoxon, p = 0.012
T-test, p = 0.01
Wilcoxon, p = 1
T-test, p = 0.67
cluster.weight
treatment
control
heat

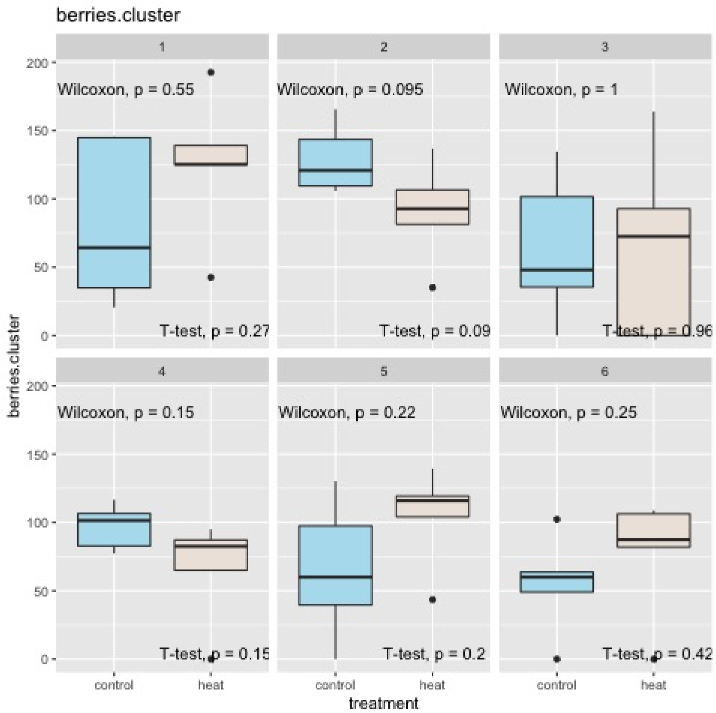

pruning.weight
Wilcoxon, p = 0.42
T-test, p = 0.25
Wilcoxon, p = 0.15
T-test, p = 0.15
Wilcoxon, p = 0.55
T-test, p = 0.88
Wilcoxon, p = 0.55
T-test, p = 0.88
Wilcoxon, p = 0.55
T-test, p = 0.76
Wilcoxon, p = 0.0079
T-test, p = 0.01
pruning.weight
treatment
control
heat

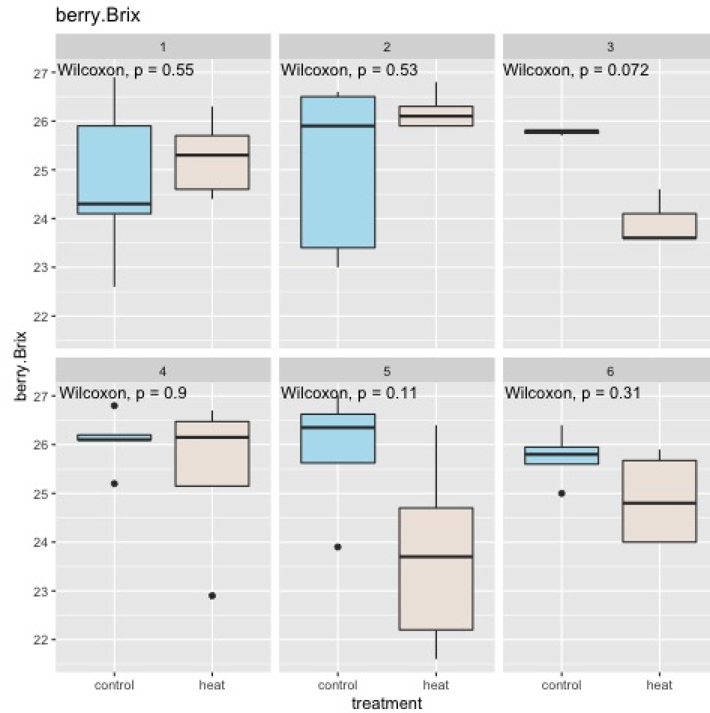

Ravaz.index
Wilcoxon, p = 0.84
T-test, p = 0.84
Wilcoxon, p = 0.056
T-test, p = 0.03
Wilcoxon, p = 0.31
T-test, p = 0.26
Wilcoxon, p = 0.15
T-test, p = 0.09
Wilcoxon, p = 0.15
T-test, p = 0.07
Wilcoxon, p = 0.056
T-test, p = 0.08
Ravaz.index
treatment
control
heat

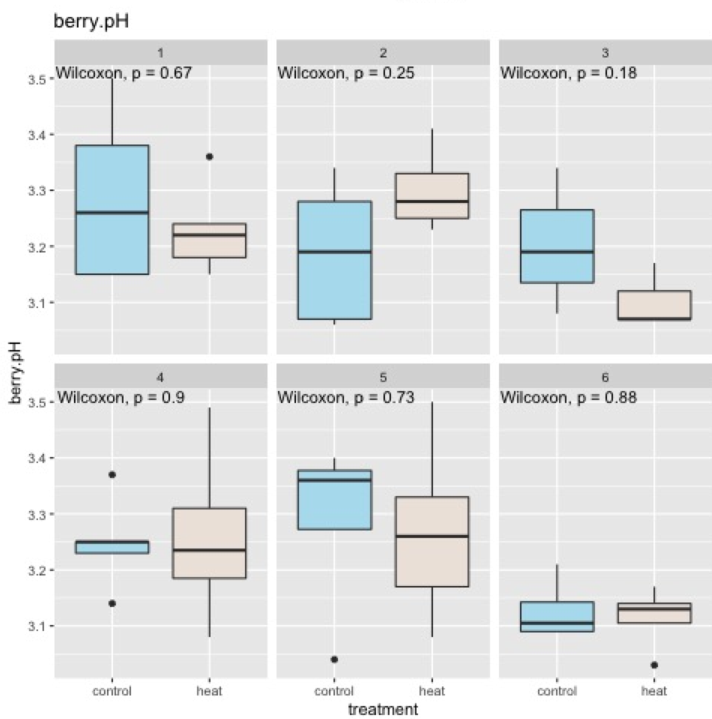

SPAD
Wilcoxon, p = 1
T-test, p = 0.95
Wilcoxon, p = 0.15
T-test, p = 0.11
Wilcoxon, p = 1
T-test, p = 0.48
Wilcoxon, p = 0.69
T-test, p = 0.58
Wilcoxon, p = 0.84
T-test, p = 0.87
Wilcoxon, p = 0.22
T-test, p = 0.14
SPAD
treatment
control
heat

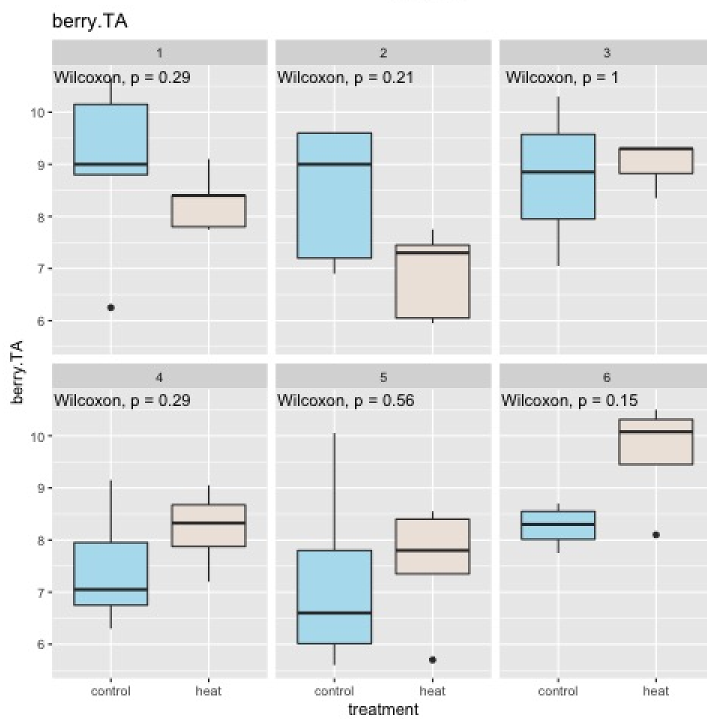

yield
Wilcoxon, p = 0.42
T-test, p = 0.2
Wilcoxon, p = 0.84
T-test, p = 0.93
Wilcoxon, p = 0.31
T-test, p = 0.33
Wilcoxon, p = 0.31
T-test, p = 0.15
Wilcoxon, p = 0.095
T-test, p = 0.06
Wilcoxon, p = 1
T-test, p = 0.92
yield
treatment
control
heat

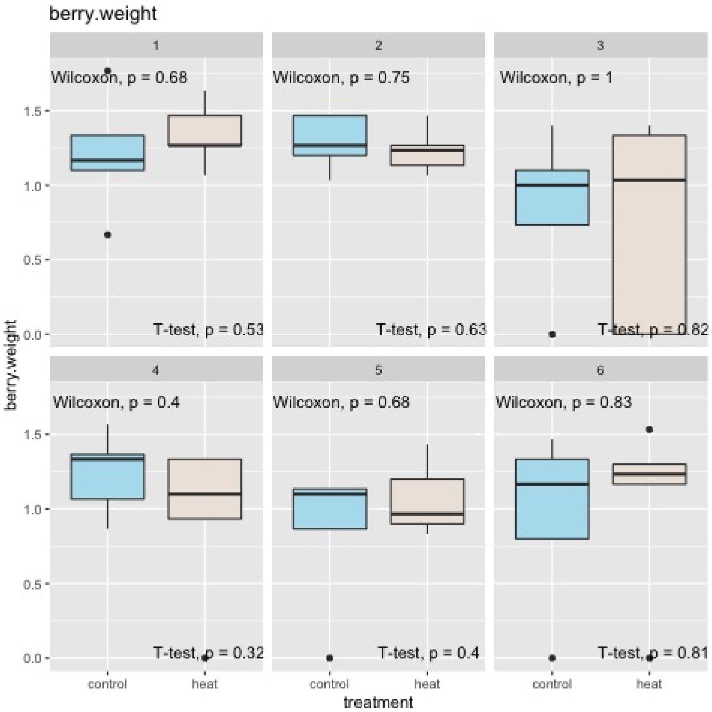
2.4 Merlot 2020: side-by-side boxplots under treatment/block combination

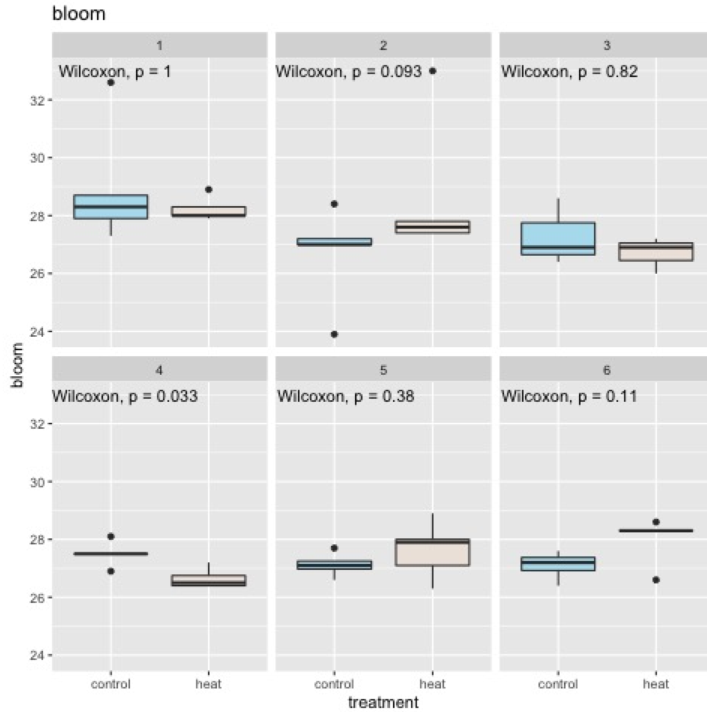


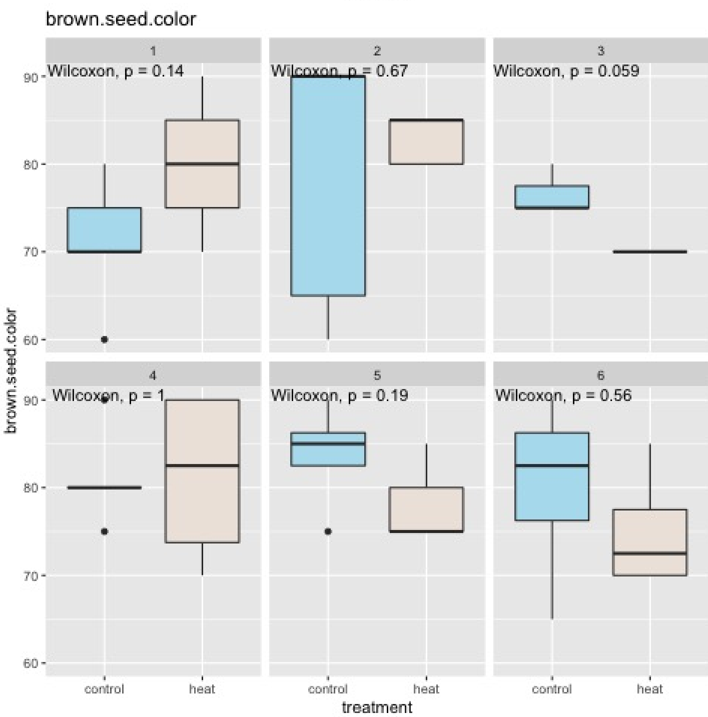


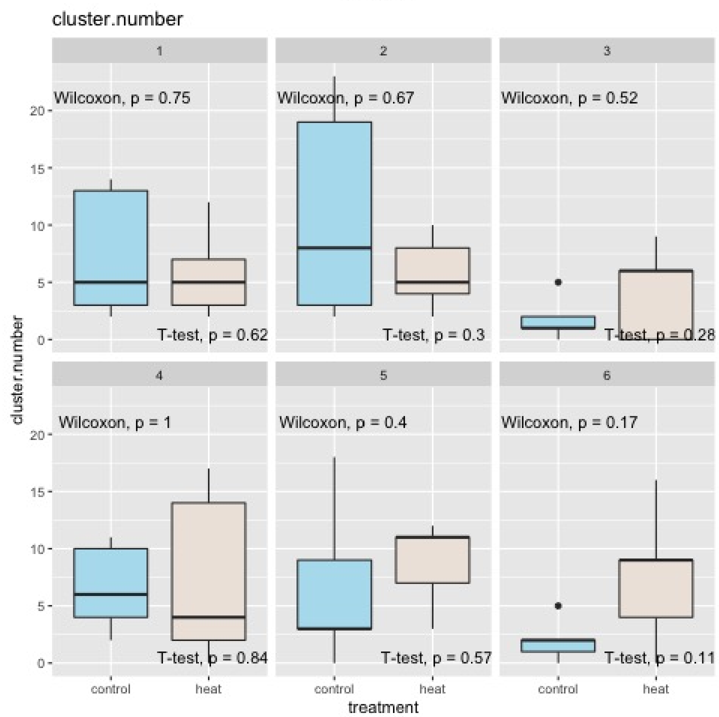


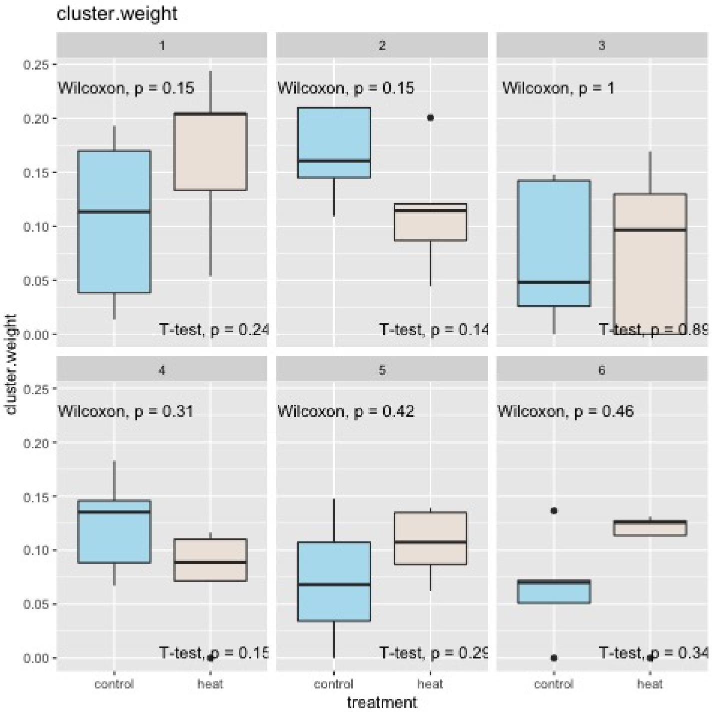


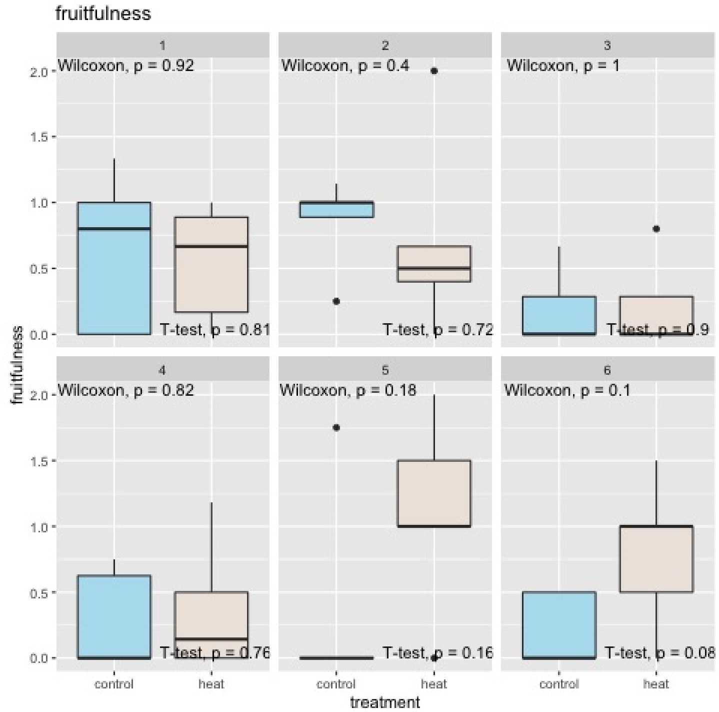

berry.weight
1
2
3
4
5
6
Wilcoxon, p = 0.68
Wilcoxon, p = 0.75
Wilcoxon, p = 1
Wilcoxon, p = 0.4
Wilcoxon, p = 0.68
Wilcoxon, p = 0.83
T-test, p = 0.53
T-test, p = 0.63
T-test, p = 0.82
T-test, p = 0.32
T-test, p = 0.4
T-test, p = 0.81
berry.weight
control
heat
treatment

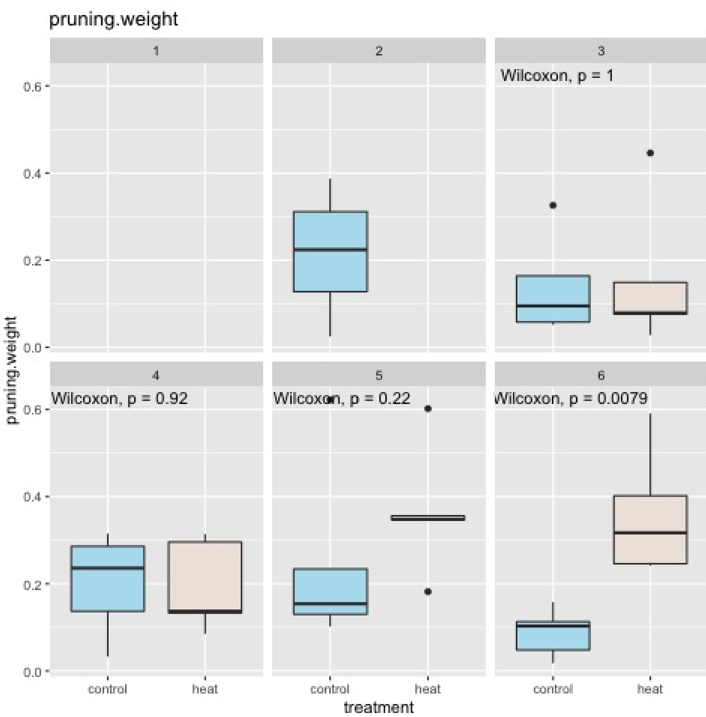

bloom
1
2
3
4
5
6
Wilcoxon, p = 1
Wilcoxon, p = 0.093
Wilcoxon, p = 0.82
Wilcoxon, p = 0.033
Wilcoxon, p = 0.38
Wilcoxon, p = 0.11
bloom
control
heat
treatment

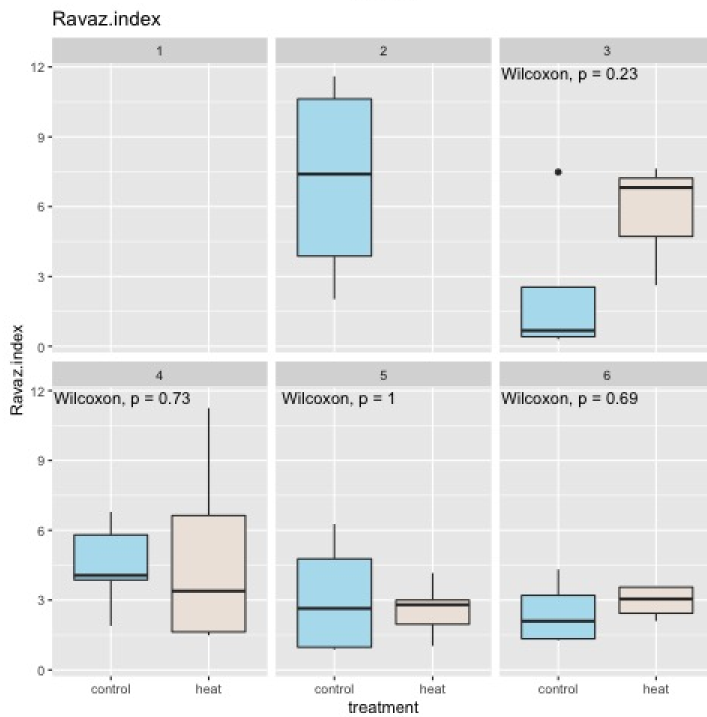

brown.seed.color
1
2
3
4
5
6
Wilcoxon, p = 0.14
Wilcoxon, p = 0.67
Wilcoxon, p = 0.059
Wilcoxon, p = 1
Wilcoxon, p = 0.19
Wilcoxon, p = 0.56
brown.seed.color
control
heat
treatment

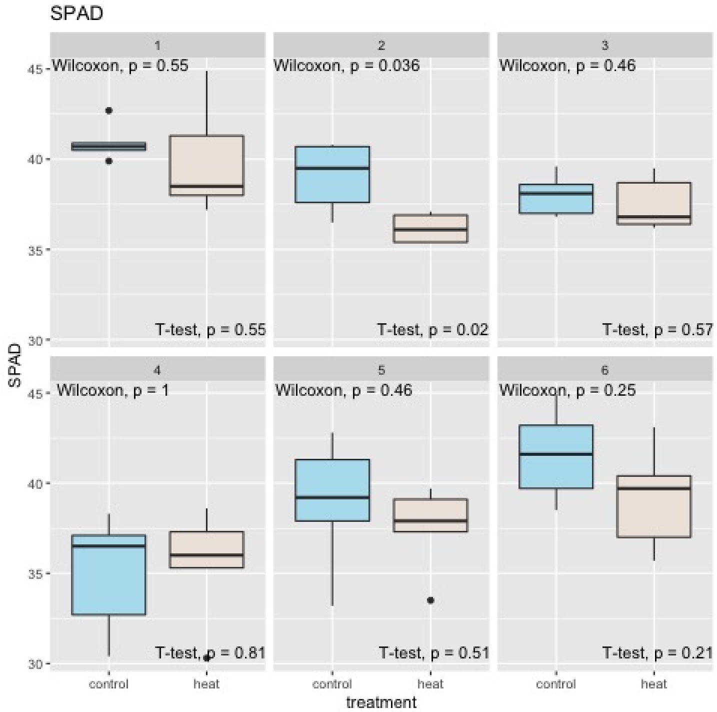

cluster.number
1
2
3
4
5
6
Wilcoxon, p = 0.75
Wilcoxon, p = 0.67
Wilcoxon, p = 0.52
Wilcoxon, p = 1
Wilcoxon, p = 0.4
Wilcoxon, p = 0.17
T-test, p = 0.62
T-test, p = 0.3
T-test, p = 0.28
T-test, p = 0.84
T-test, p = 0.57
T-test, p = 0.11
cluster.number
control
heat
treatment

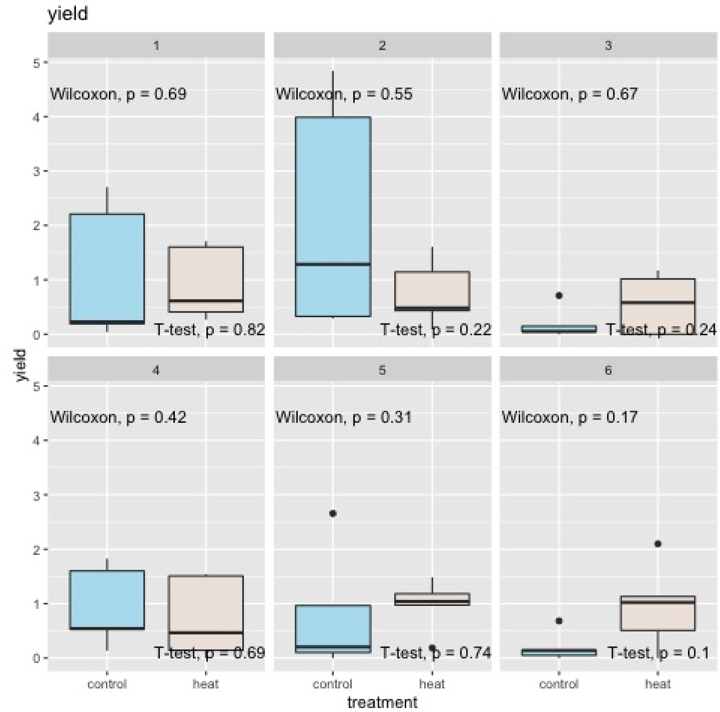

cluster.weight
1
2
3
4
5
6
Wilcoxon, p = 0.15
T-test, p = 0.24
Wilcoxon, p = 0.15
T-test, p = 0.14
Wilcoxon, p = 1
T-test, p = 0.89
Wilcoxon, p = 0.31
T-test, p = 0.15
Wilcoxon, p = 0.42
T-test, p = 0.29
Wilcoxon, p = 0.46
T-test, p = 0.34
cluster.weight
control
heat
treatment


fruitfulness
1
2
3
4
5
6
Wilcoxon, p = 0.92
T-test, p = 0.81
Wilcoxon, p = 0.4
T-test, p = 0.72
Wilcoxon, p = 1
T-test, p = 0.9
Wilcoxon, p = 0.82
T-test, p = 0.76
Wilcoxon, p = 0.18
T-test, p = 0.16
Wilcoxon, p = 0.1
T-test, p = 0.08
fruitfulness
control
heat
treatment


pruning.weight
1
2
3
4
5
6
Wilcoxon, p = 1
Wilcoxon, p = 0.92
Wilcoxon, p = 0.22
Wilcoxon, p = 0.0079
pruning.weight
control
heat
treatment


Ravaz.index
1
2
3
4
5
6
Wilcoxon, p = 0.23
Wilcoxon, p = 0.73
Wilcoxon, p = 1
Wilcoxon, p = 0.69
Ravaz.index
control
heat
treatment





3. Censored Data Visualization

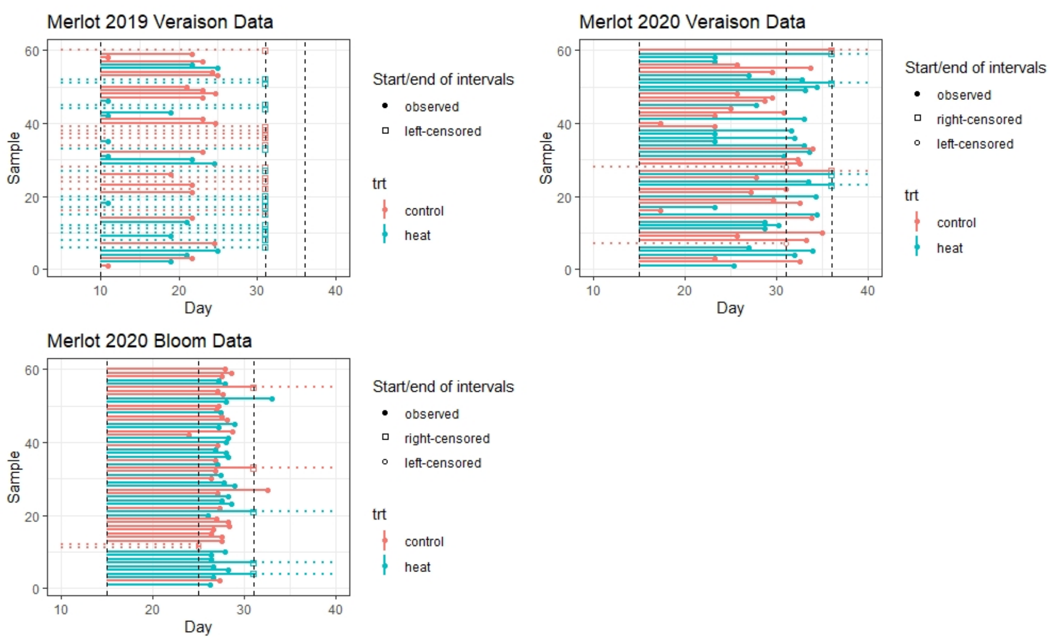
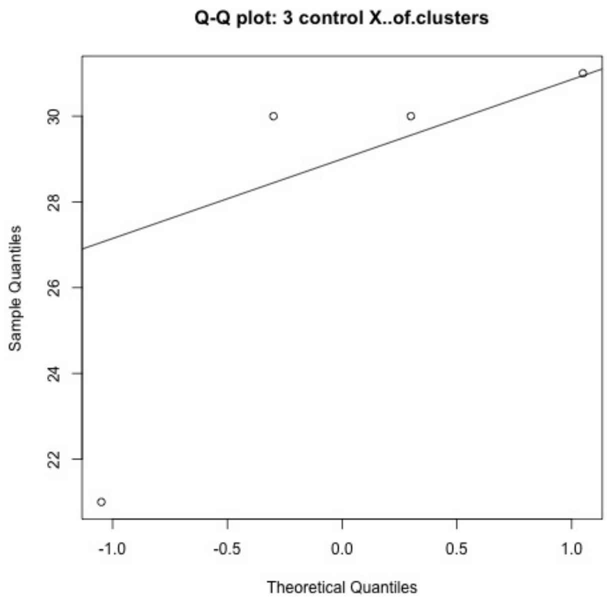


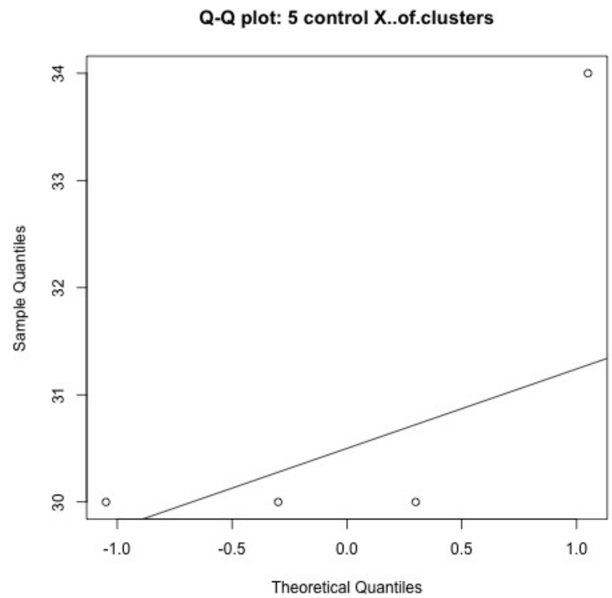
Fig. 14. Visual representation of censored bloom and Veraison data. On each plot, three vertical dotted lines are drawn, indicating ”study start date”, ”T1”, and ”T3” from left to right. Vines are monitored and their 50% bloom or veraison is recorded if it is observed (solid lines followed by a solid dot). If the vine is left-censored (the plant was already 100% in bloom or ripened by T1), then the event happened somewhere along the dotted line ending in time T1. If the vine is right-censored (50% bloom or veraison was not observed by T3), then the event happened somewhere along the dotted line to the right of T3.

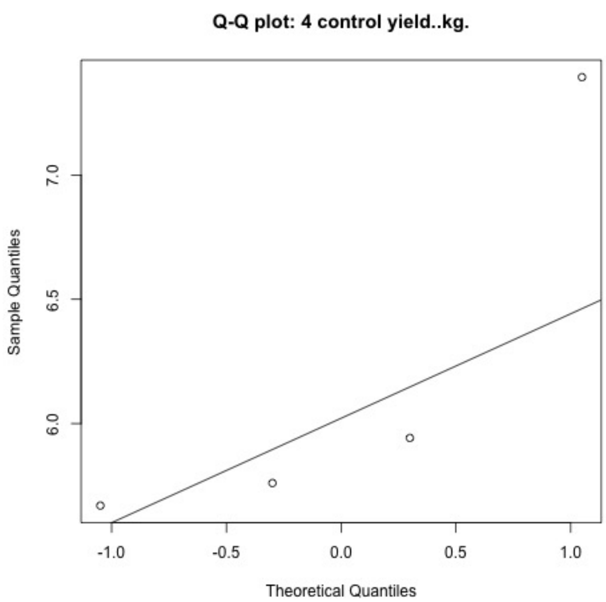
Part II: Normality assumption checks for two-way ANOVA

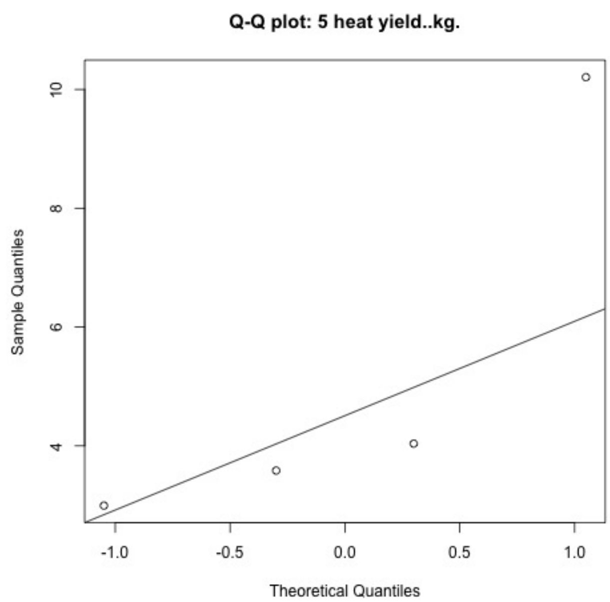
The following plots show the treatment/block combinations whose observations show significant departure from the normal distribution for each response variable in corresponding dataset. Significant departure is defined by having a p-value of 5% or less from the Shapiro-Wilks test of normality.

1. Chardonnay 2019

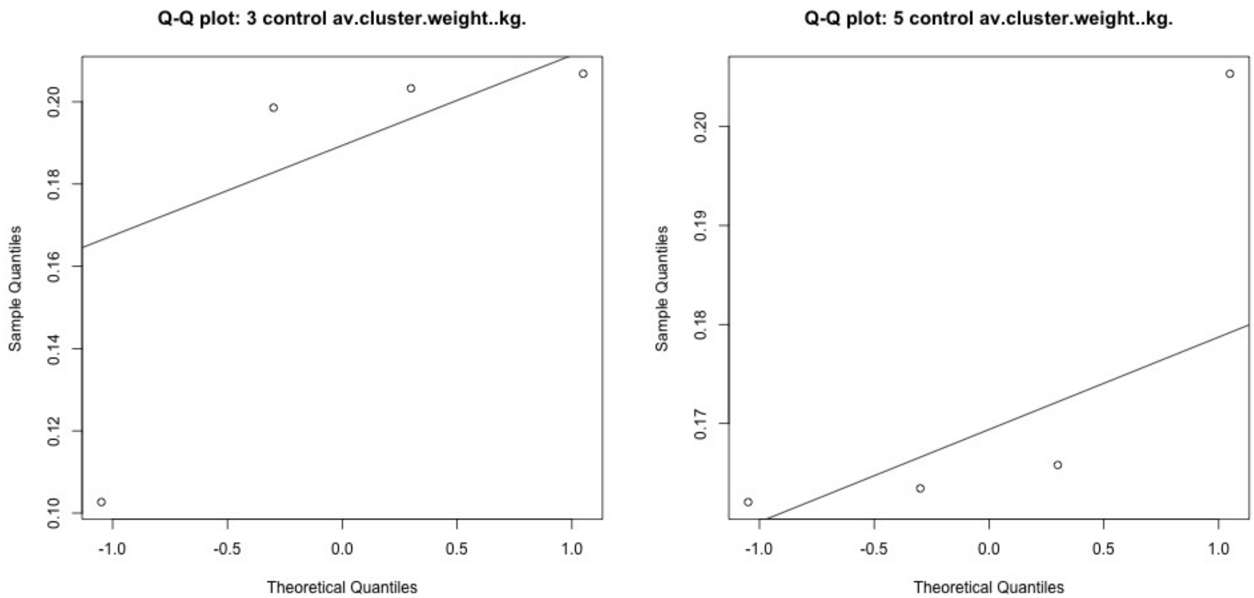




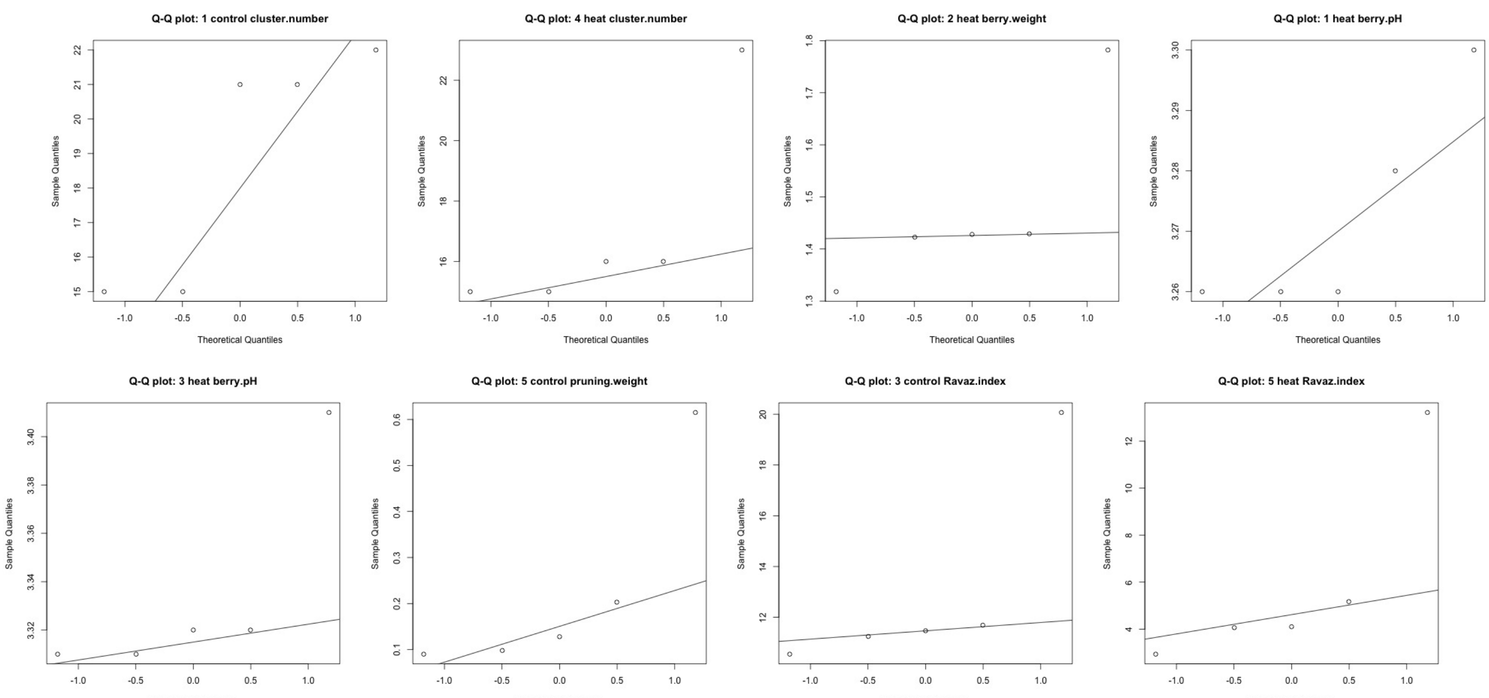




2. Chardonnay 2020

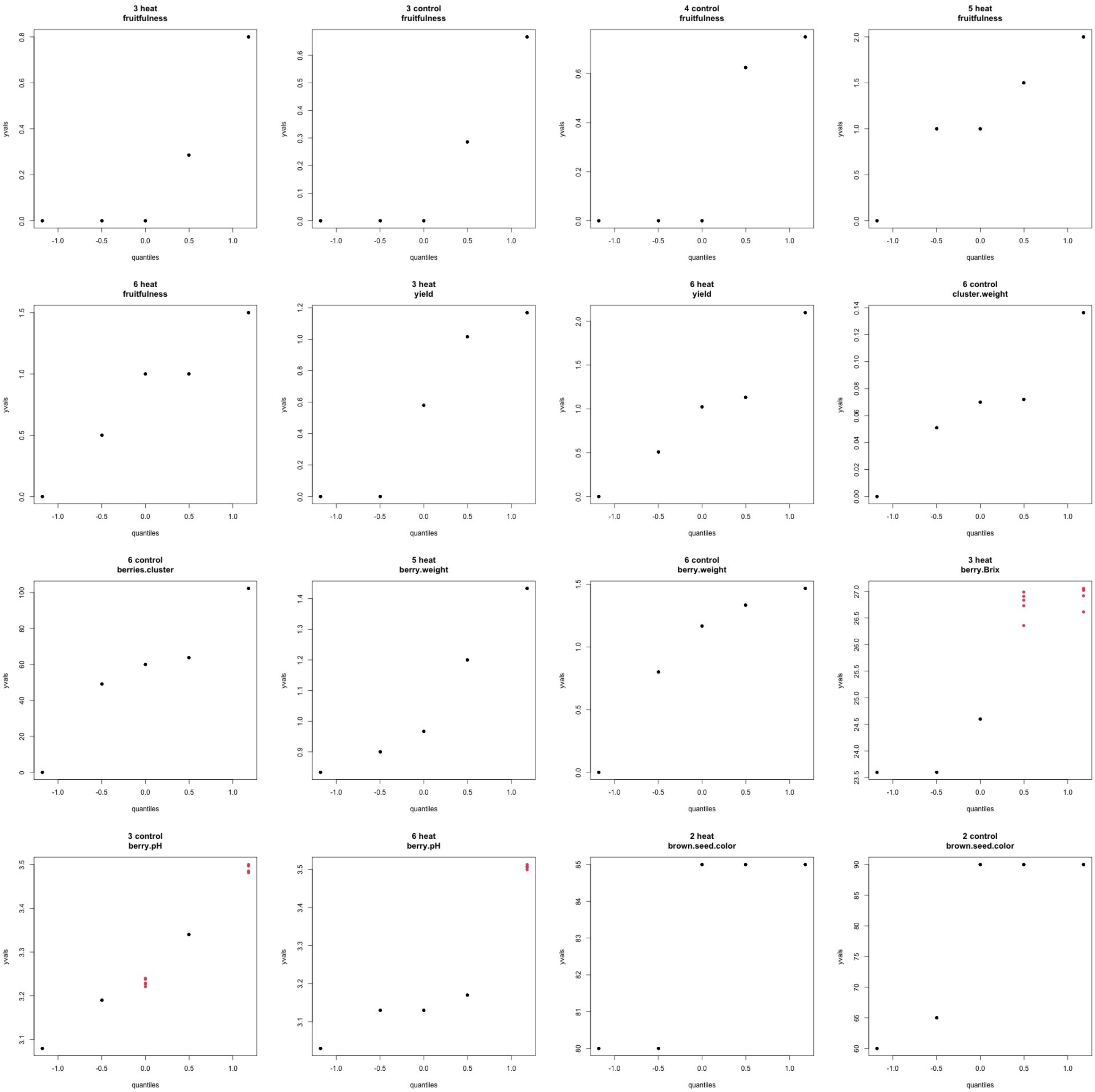


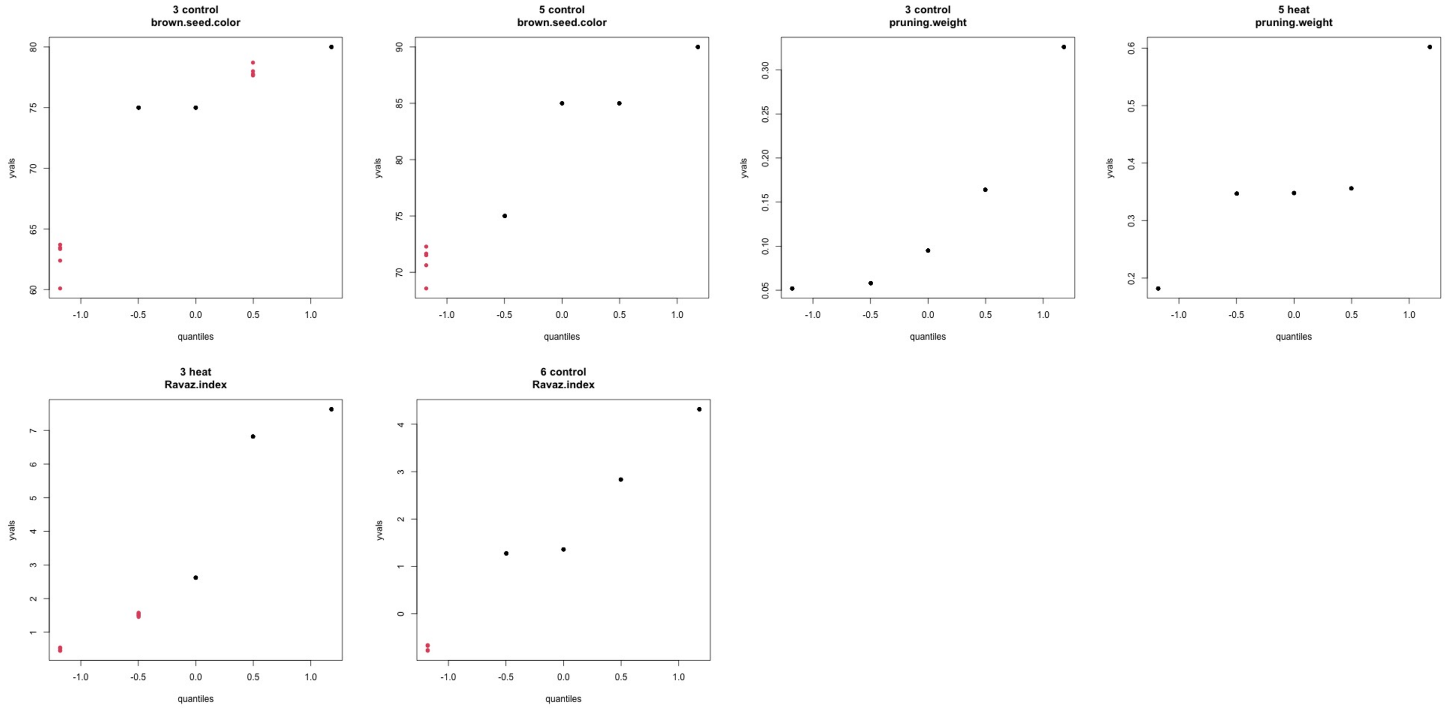
3. Merlot 2019



4. Merlot 2020

Note the red dots represent the imputed values for each measurement that is missing.





Part III: Additional Table of Results

	treatment	block	interaction
X..of.clusters	0.29	0.22	0.17
yield..kg.	0.55	0.29	1.00
av.cluster.weight..kg.	0.96	0.44	0.26

(a) p-values from two-way ANOVA on responses in Chardonnay 2019

	treatment	block	interaction
X..of.clusters	0.28	0.07	0.07
yield..kg.	0.33	0.27	1.00
av.cluster.weight..kg.	0.92	0.63	0.27

(b) p-values from ART ANOVA on responses in Chardonnay 2019

	treatment	block	interaction
X..of.clusters	0.29	0.04	0.12
yield..kg.	0.72	0.02	0.30
av.cluster.weight..kg.	0.37	0.08	0.85

(c) p-values from two-way ANOVA on responses in Chardonnay 2020

	treatment	block	interaction
X..of.clusters	0.33	0.04	0.04
yield..kg.	0.74	0.03	0.23
av.cluster.weight..kg.	0.40	0.07	0.83

(d) p-values from ART ANOVA on responses in Chardonnay 2020

Fig. 15. p-values from two way analysis on responses in Chardonnay 2019 and 2020

Part IV: Contour Map of Power

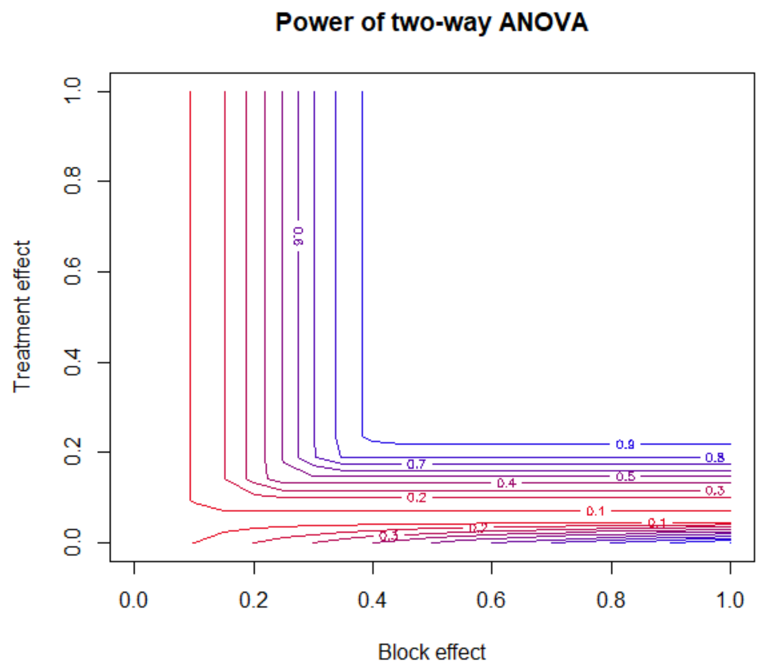


Fig. 16. Contour map of power