

TASK1

Introduction

The provided data contained information about a flycatcher population each year from 1990 to 2022. The data contained each year's current population density, the previous year's population density, and the previous year's predation rate. The question was what the effects of the previous predation rate and previous population density were on the current year's population density. Between the previous density and the previous predation we already know that there is correlation.

Workflow

For the testing general linear models (GLM) were built to examine these effects. The analysis was made by R. The used package was car.

First, a model was built to test the effect of the previous density and previous predation rate. It gave the gross parameter estimates for both variables. For statistical results, the Type III Anova was used.

Secondly, two separate models was built to check the individual effects of each variable. The Type III Anova was also used for the statistical result.

Next, a model was built to test the effect of previous predation on the previous year. The residuals of this model were added to the end of the dataset.

Finally, a model was built to test the effect of the previous predation and previous residual density. It gave the gross parameter estimate for the previous predation and the net parameter estimate for the previous density. Type III Anova was used for the statistical results.

Interpretation

The collinear model (containing both previous density and previous predation effects) showed the net estimates for both variables. The statistical results (*Table 1*) showed that both previous predation and previous density have statistically significant effects ($\text{Pr}(>F)$ is 0.03828 and 4.605e-07 <0.05 respectively) and the combination has a marginally significant effect ($\text{Pr}(>F)$ is 0.05337). The parameter estimates (*Table2*) for previous density and previous predation had positive change (net estimates = 2.9968988 and 0.9179258) but the combination had a negative change (estimates= -1.3362067).

Table1: Collinear model ANOVA:

Anova Table (Type II tests)					
Response: dens	Sum Sq	Df	F value	Pr(>F)	
prevpred	3.894	1	4.7121	0.03828	*
prevdens	34.437	1	41.6703	4.605e-07	***
prevpred:prevdens	3.352	1	4.0561	0.05337	.
Residuals	23.966	29			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table2: Coefficient of collinear model:

(Intercept)	prevpred	prevdens	prevpred:prevdens
0.7099150	2.9968988	0.9179258	-1.3362067

For adjusting the collinear model, for each variable, a new, separate model was built to check their effect on the current density and get the gross estimates for both variables. Separate model for previous predator showed statistically significance (*Table3*) and had a negative changing effect (*Table4*), gross parameter estimates was -4.320271.

Table3: Separate1 model for previous predator on current density anova:

Anova Table (Type II tests)					
Response:	dens	Sum Sq	Df	F value	Pr(>F)
prevpred	20.748	1	10.415	0.002948	**
Residuals	61.755	31			
<hr/>					
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table4: Coefficient of the separate1 model

Intercept)	prevpred
5.142993	-4.320271

Separate model for the previous predator also showed a statistically significant effect (*Table5*) ($\text{Pr}(>\text{F}) = 5.07e-08 < 0.05$) and slightly positive change (*Table6*), gross parameter estimates was 0.7872702.

Table5: Separate2 model for previous density on current density ANOVA:

Anova Table (Type II tests)					
Response:	dens	Sum Sq	Df	F value	Pr(>F)
prevdens	51.291	1	50.942	5.07e-08	***
Residuals	31.212	31			
<hr/>					
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Table6: Coefficient of the separate2 model:

(Intercept)	prevdens
0.9109078	0.7872702

A model also built to check the effect of previous density on previous predation. The statistical analysis (*Table7*) showed significant effect ($\text{Pr}(>F) = 0.02857$). From this model, the residuals added to the end of the dataset. It was used for making the residual model.

Table7: Effect of previous density on previous predation model ANOVA:

```
Anova Table (Type II tests)

Response: prevdens
          Sum Sq Df F value Pr(>F)
prevpred 12.031  1 5.2735 0.02857 *
Residuals 70.724 31
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The residual model is built to check the effect of previous predation and previous residual density on current density. This model shows the correlated effect of the two variables on predation. The statistical analysis (*Table8*) showed that both variables have significant effects ($\text{Pr}(>F) = 4.415e-05$ and $9.184e-07$ respectively). It also gives the gross parameter estimate for previous predation and net parameter estimates for previous density (*Table9*). The previous predation has a negative change (gross estimates= -4.3202713), while the previous density has a slightly positive change (net estimates= 0.6977991).

Table8: Model with residuals ANOVA:

```
Anova Table (Type II tests)

Response: dens
          Sum Sq Df F value    Pr(>F)
prevpred 20.748  1 22.785 4.415e-05 ***
resld    34.437  1 37.818 9.184e-07 ***
Residuals 27.318 30
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table9: Coefficient of the model:

(Intercept)	prevpred	resld
5.1429925	-4.3202713	0.6977991

TASK2

Introduction

The provided data contained information about 4 different populations of *Asellus aquaticus*. It contained data about their behaviour rate (exploration) and different parameters that may affect it. The question is how these given parameters (sex, light) affect these population's behaviour(exploration).

Workflow

For the testing general linear models (GLM) were built to examine these effects. The analysis was made by R. The used packages were car, lme4, MASS, and emmeans.

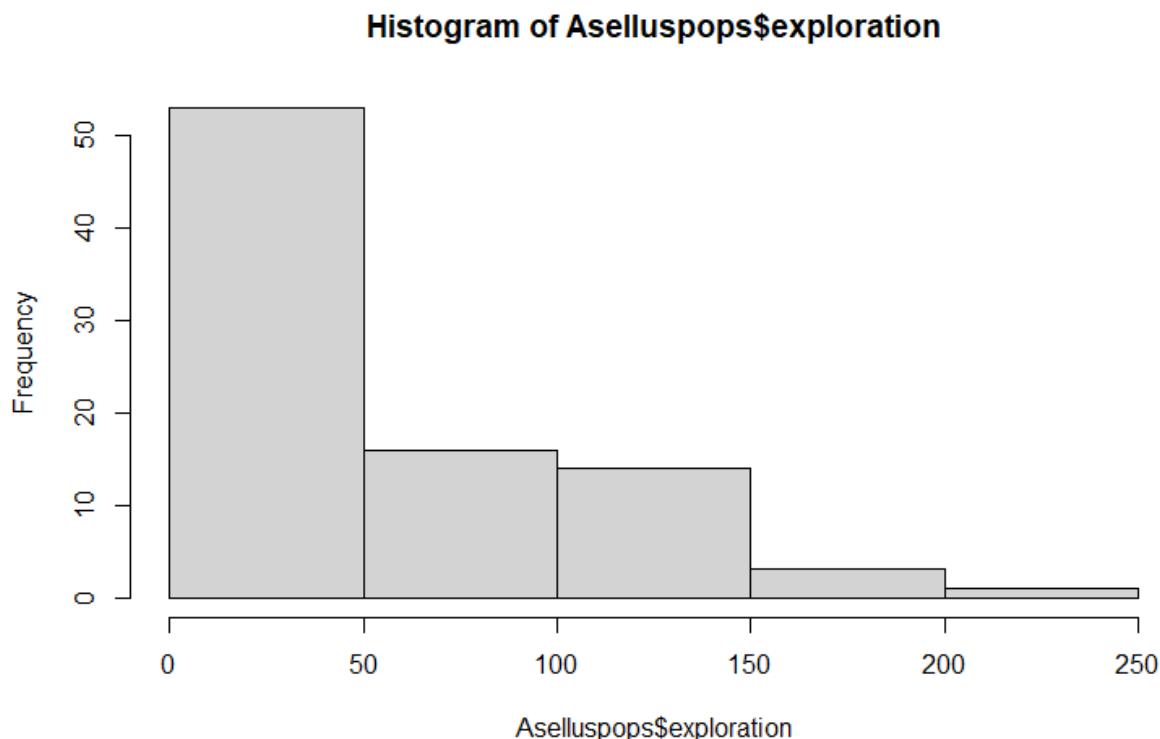
The model was built by using "exploration" as a dependent variable and "population", "sex" and "light.regime" as an independent variable. The model didn't contain interaction terms.

For the statistical result, the Type III Anova was used.

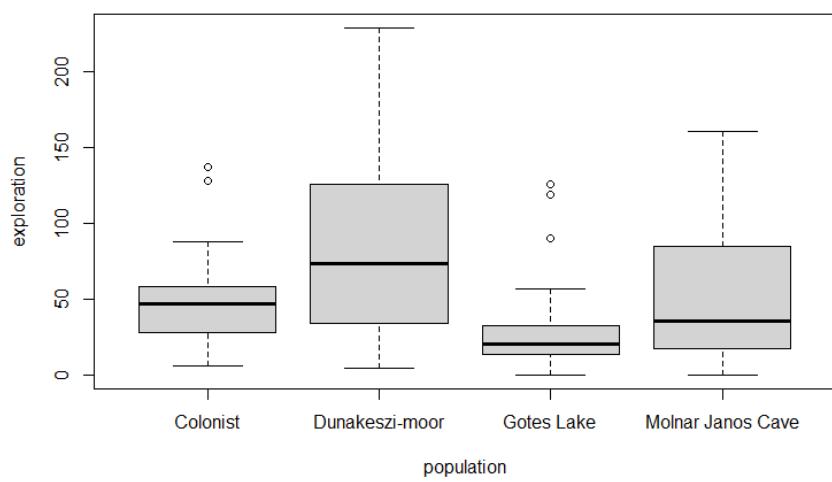
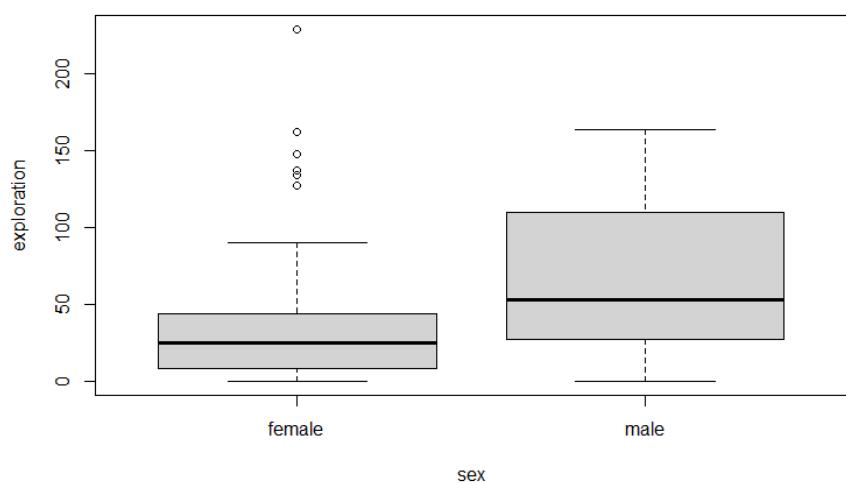
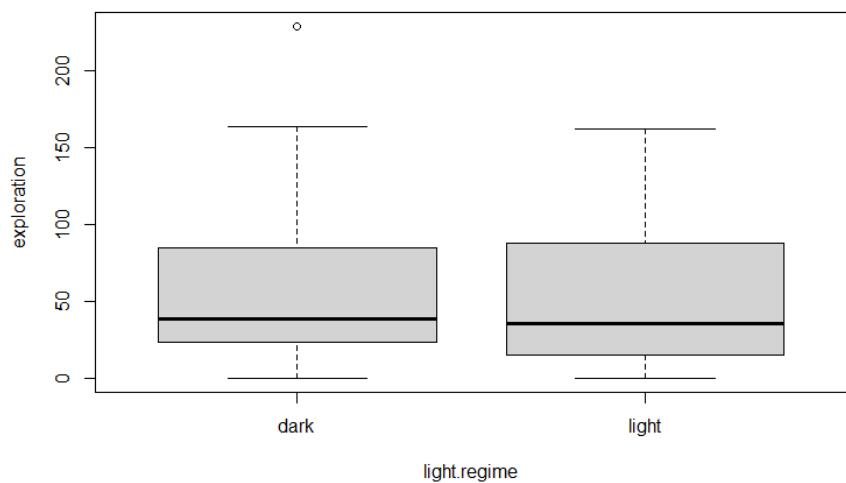
Interpretation

First, the frequency of the exploration rates of *A. aquaticus* population visualised on histogram (*Plot1*) and the exploration also visualized by each categorical variable separately on boxplots (*Plot2*).

Plot1: Visualising the frequency of the different exploration rates:



Plot2: Boxplots for each variable (light.regime, sex and population) on exploration:



A model was built to check the effect of all three dependent variables (population, sex, and light regime) on the exploration distribution (*Table 1*). The model did not contain interactional terms.

The statistical analysis (*Table 2*) showed statistically significant effect ($P < 0.05$) for population and sex variables ($\text{Pr}(>\text{ChiSq})$ is 0.001715 and 0.032741 respectively). For the significant variables, the estimated marginal means were checked combined with Tukey posthoc test.

As for the population (*Table 3*), the estimated marginal means was 51.8 for Colonist, 75.8 for Dunakeszi-moor, 28.2 for Gotes Lake and 59.0 for Molnar Janos Cave. The pairwise difference was significant between Gotes Lake and Dunakeszi-moor ($p\text{-value} = 0.0004 < 0.05$), and Gotes Lake and Molnar Janos Cave ($p\text{-value} = 0.046 < 0.05$). As for the sex (*Table 4*), the estimated marginal means was 40.7 for female and 62.8 for males.

Conclusion:

According to the statistical analysis, females have a lower exploration rate than males. The Gotes Lake population has also a lower exploration rate than the populations of Dunakeszi-moor and Molnar Janos Cave.

Supplementary tables:

Table 1: Summary of the model:

```

Call:
glm.nb(formula = exploration ~ population + sex + light.regime,
       data = Asellusrops, init.theta = 1.187510693, link = log)

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                         3.8123    0.3157 12.077 <2e-16 ***
populationDunakeszi-moor           0.3802    0.3114  1.221   0.2220
populationGotes Lake                -0.6071   0.3156 -1.924   0.0544 .
populationMolnar Janos Cave        0.1306    0.3397  0.385   0.7005
sexmale                             0.4333    0.2013  2.153   0.0314 *
light.regime.light                 -0.1634   0.2012 -0.812   0.4167
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.1875) family taken to be 1)

Null deviance: 120.42 on 86 degrees of freedom
Residual deviance: 101.67 on 81 degrees of freedom
AIC: 871.07

Number of Fisher Scoring iterations: 1

Theta:  1.188
Std. Err.: 0.177
2 x log-likelihood: -857.073

```

Table2: Anova test for the model:

```
Analysis of Deviance Table (Type III tests)

Response: exploration
          LR Chisq Df Pr(>chisq)
population   15.1228  3   0.001715 **
sex          4.5593  1   0.032741 *
light.regime 0.6475  1   0.421004
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table3: Est. marginal means/Post hoc test for the significant result (population):

```
$`emmeans of population`
population      response   SE  df asymp.LCL asymp.UCL
Colonist        51.8 13.45 Inf   31.1    86.2
Dunakeszi-moor 75.8 13.02 Inf   54.1   106.1
Gotes Lake     28.2  5.09 Inf   19.8    40.2
Molnar Janos Cave 59.0 12.92 Inf   38.4    90.7

Results are averaged over the levels of: sex, light.regime
Confidence level used: 0.95
Intervals are back-transformed from the log scale

$`pairwise differences of population`
 1           ratio   SE  df null z.ratio p.value
Colonist / (Dunakeszi-moor) 0.684 0.213 Inf   1 -1.221 0.6134
Colonist / Gotes Lake       1.835 0.579 Inf   1  1.924 0.2180
Colonist / Molnar Janos Cave 0.878 0.298 Inf   1 -0.385 0.9807
(Dunakeszi-moor) / Gotes Lake 2.684 0.669 Inf   1  3.962 0.0004
(Dunakeszi-moor) / Molnar Janos Cave 1.284 0.358 Inf   1  0.896 0.8070
Gotes Lake / Molnar Janos Cave 0.478 0.136 Inf   1 -2.600 0.0460

Results are averaged over the levels of: sex, light.regime
P value adjustment: tukey method for comparing a family of 4 estimates
Tests are performed on the log scale
```

Table4: Est. marginal means/Post hoc test for the significant result (sex):

```
$`emmeans of sex`
sex      response   SE  df asymp.LCL asymp.UCL
female   40.7 6.20 Inf   30.2    54.9
male     62.8 8.73 Inf   47.8    82.4

Results are averaged over the levels of: population, light.regime
Confidence level used: 0.95
Intervals are back-transformed from the log scale

$`pairwise differences of sex`
 1           ratio   SE  df null z.ratio p.value
female / male 0.648 0.131 Inf   1 -2.153 0.0314

Results are averaged over the levels of: population, light.regime
Tests are performed on the log scale
```

TASK3

Introduction

The provided data is about the coverage of different plant species from a field experiment. The data contained two types of treatment (Treat1, Treat2), the number of the parallel (Block 1-6) and the coverage of each plant species in percentage (0-100).

For the testing, a multivariate approach was used. The analysis and visualization were made by R. The used packages were vegan, ggplot2, ggforce, ggrepel, dplyr, dendextend, devtools and pairwiseAdonis.

Interpretation

After importing the data, the characters are converted to factors (1-3 columns). The data remained unstandardized due to all coverage data were on the same scale and represented the same phenomenon, therefore the differences between the measures have important implication.

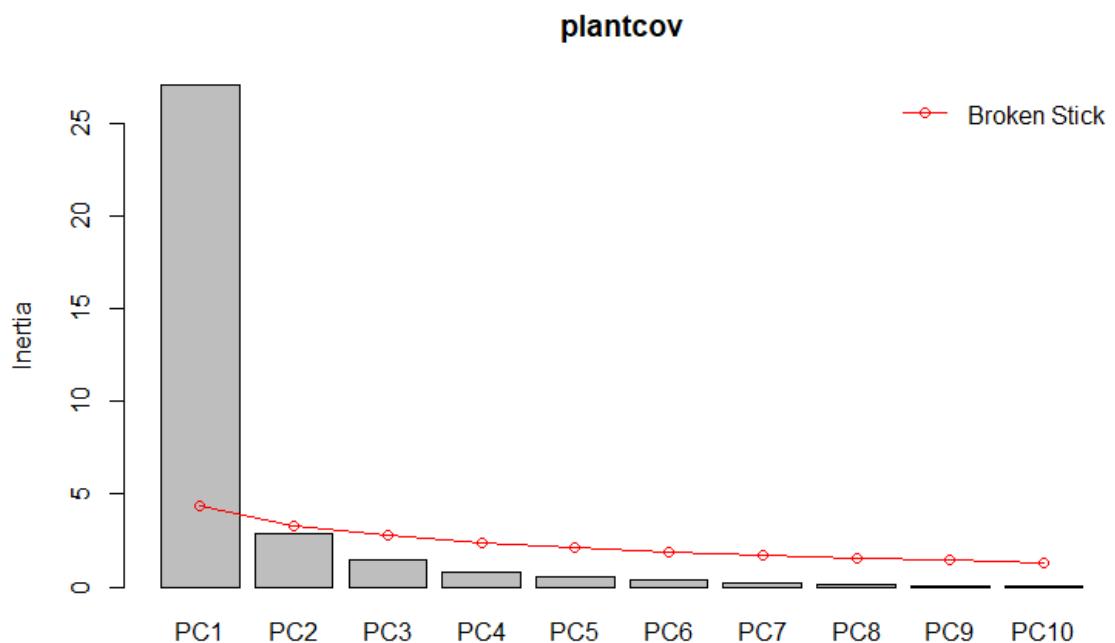
Before visualisation, ordination was used to reduce the dimensions of the data. It was done by PCA because of the focus on identifying the patterns/relationships of the variables. Furthermore, the data is assumed to be linear and contains more variables than observations (48 obs. of 60 variables).

The first 2 components (PC1 and PC2) were used from the PCA for further analysis. The PC1 explains 80.07%, and the PC2 explains 8.56% of the data (*Table1*). After the reduction, we can explain the 88.63% through the 2D visualization. As for eigenvalue (the presence of the total amount of variance explained by the given component) is 27.08 for PC1 and 2.89 for PC2 (*Plot1*). The explained variable of PC1 is on (*Plot2*), and the explained variable of PC2 is on (*Plot3*). The biplot of PCA1 and PCA2 shows the relationship between the variables and the observations (*Plot4*).

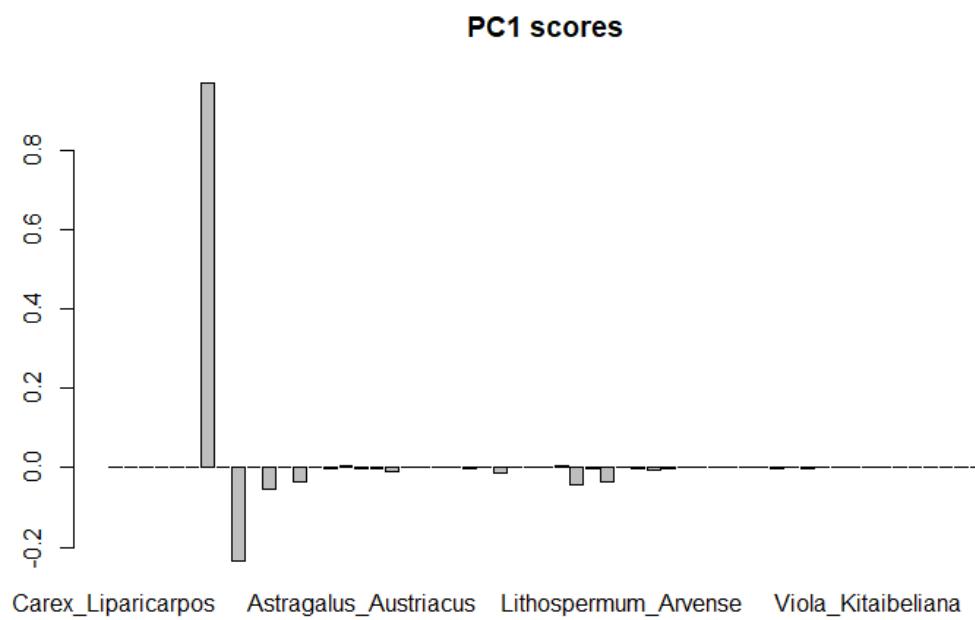
Table1: The goodness of the PCA (for simplicity only the first 3 components are shown):

Importance of components:					
	PC1	PC2	PC3	...	
Eigenvalue	27.0766	2.89269	1.45604	...	
Proportion Explained	0.8007	0.08555	0.04306	...	
Cumulative Proportion	0.8007	0.88628	0.92934	...	

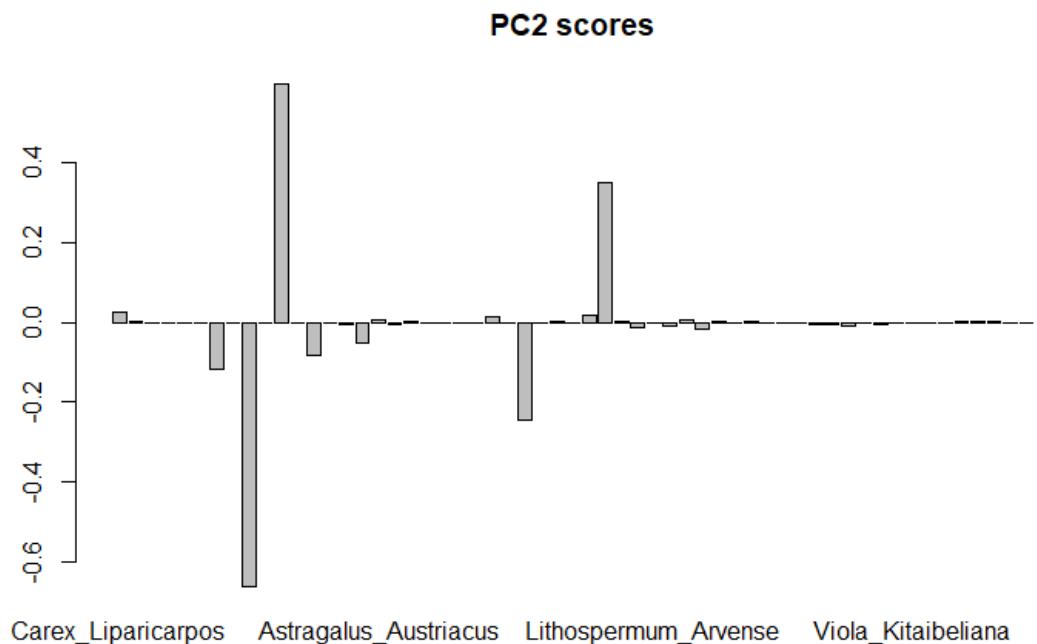
Plot1.: Variances explained by each principal component.



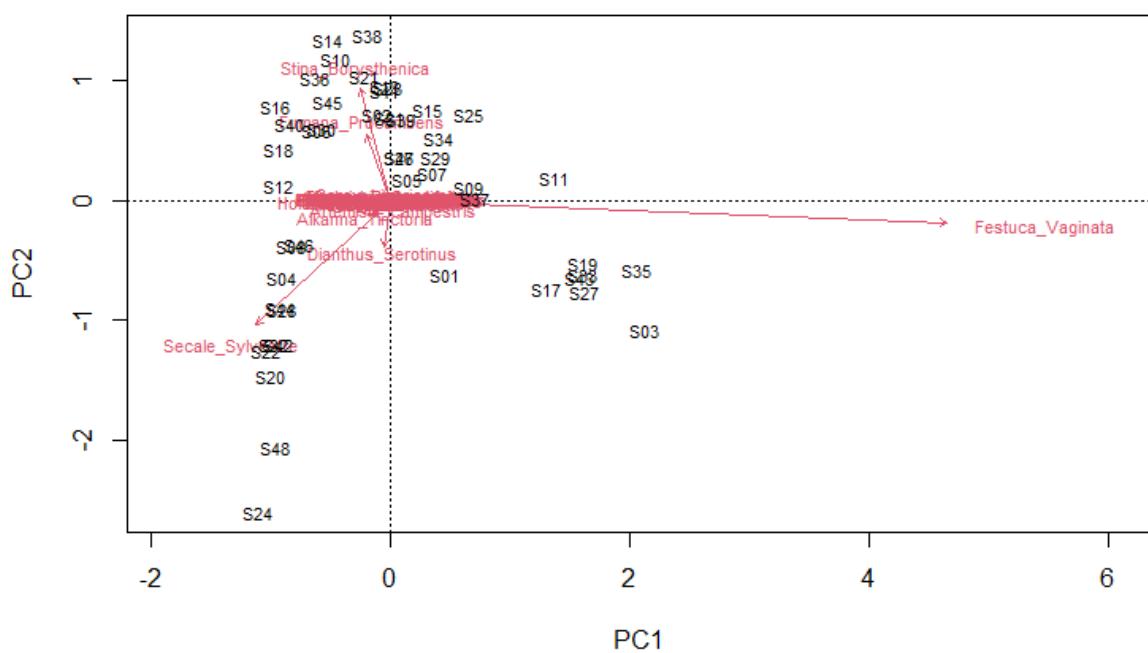
Plot2: Variables explained by PC1



Plot3.: Variables explained by PC2



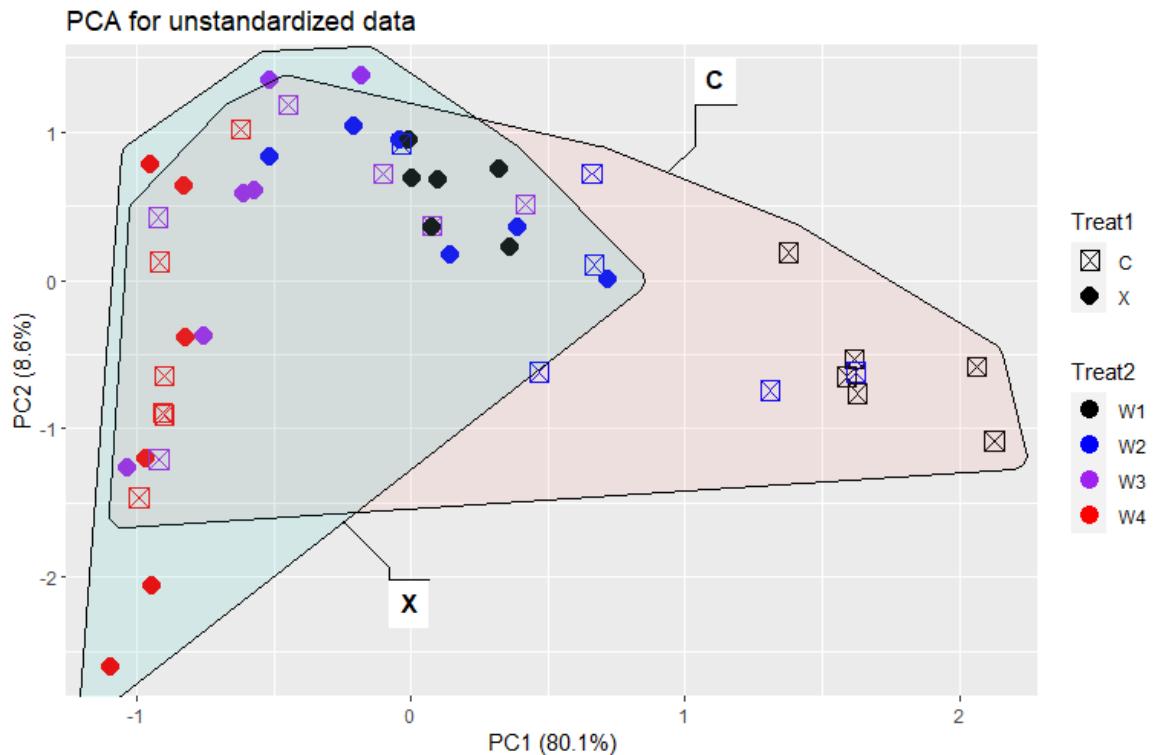
Plot4: Biplot of the PCA1 and PCA2



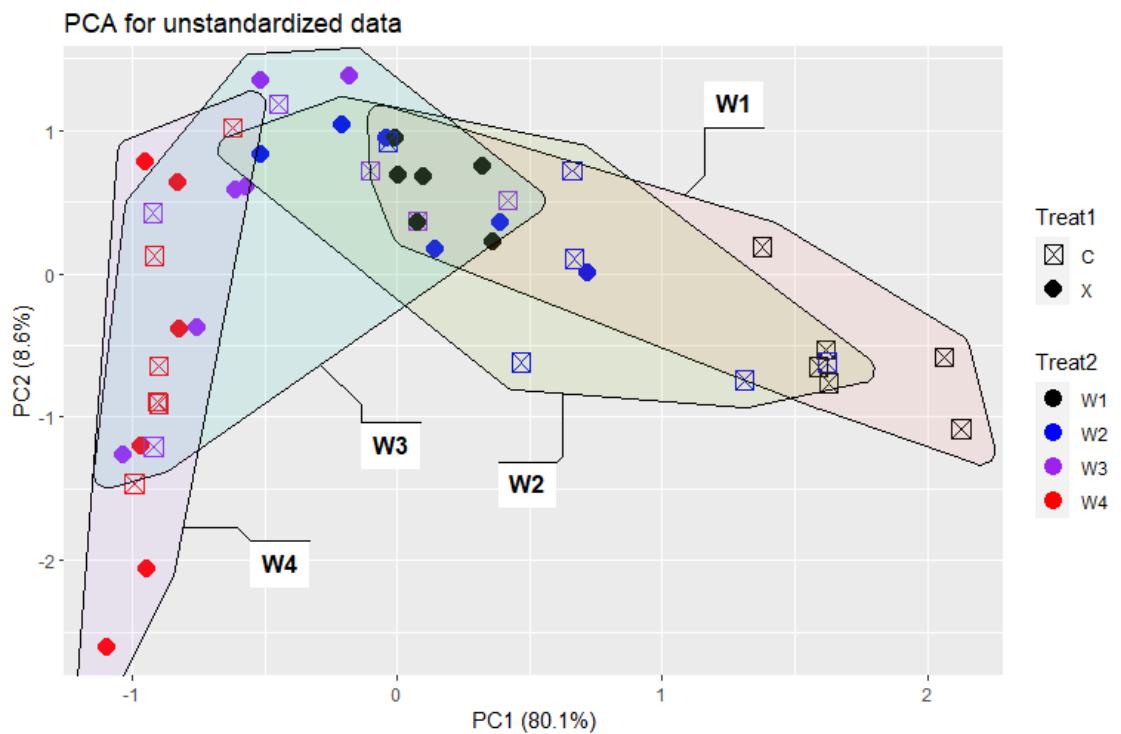
Visualization

The dataset is visualized with ggplot in 2D (reduced by PCA). *Plot5* highlights the Treatment1, *Plot6* highlights the Treatment2, and the *Plot7* highlights the Block (parallels) with the hulls. We can see the distribution of each factor on each plot.

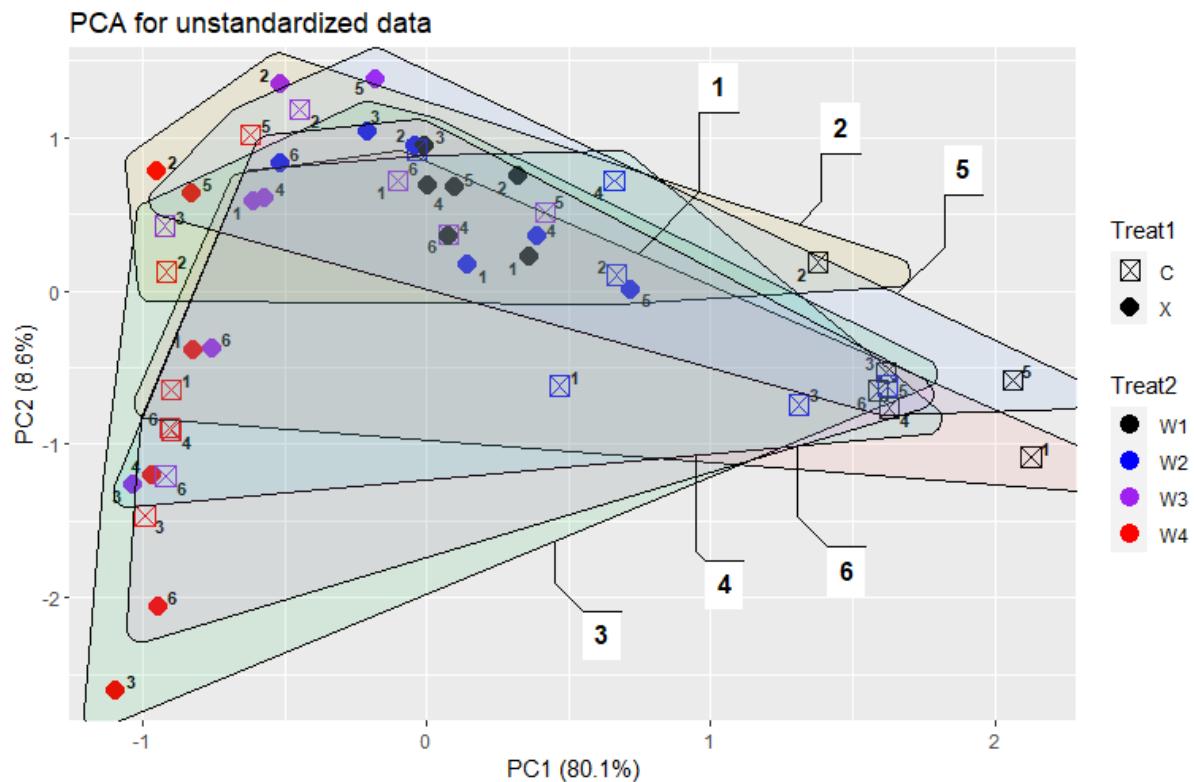
Plot5.: Scatterplot for the 2 different treatments (Treatment1 circled)



Plot6.: Scatterplot for the 2 different treatments (Treatment2 circled)



Plot7.: Scatterplot for the 2 different treatment in parallels/blocks (blocks are circled)



Statistical analysis

The statistical analysis is tested permanova (euclidean method, 10000 permutations). The test used separately for treatment1, treatment2, and blocks. For significant results posthoc test (pairwise adonis) was used.

Results

According to permanova both treatment1 and treatment2 had a statistically significant difference between the samples ($Pr(>F) < 0.05$), but no significant difference for the blocks. Posthoc test showed (Table2) that in Treatment2, W1 and W2 were not statistically different from each other, but W1/W2 differed from W3 and W4. Also, W3 and W4 differed from each other. As for Treatment1, it had two-components, posthoc test was not applied.

Table2: pairwise adonis posthoc test (Treatment2):

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
1 w2 vs w3	1	0.7376544	6.801313	0.23614594	0.001	0.006	*
2 w2 vs w1	1	0.1267657	1.576507	0.06686771	0.165	0.990	
3 w2 vs w4	1	2.3947117	21.646252	0.49594755	0.001	0.006	*
4 w3 vs w1	1	1.1073261	10.440657	0.32183864	0.001	0.006	*
5 w3 vs w4	1	0.8249870	6.053652	0.21578837	0.006	0.036	.
6 w1 vs w4	1	2.7841512	25.724206	0.53901800	0.001	0.006	*