Chickadee Pathogen Data Statistical Analysis

Junsoo Park

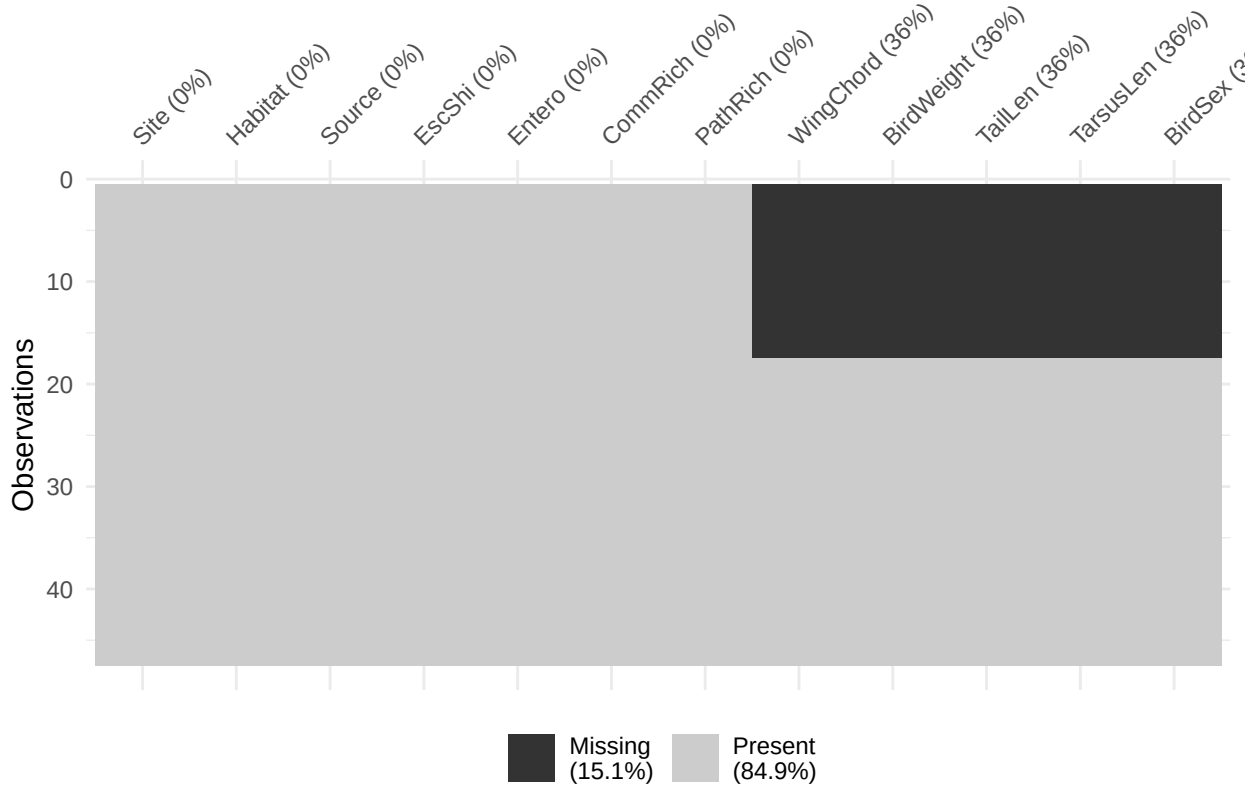
2023-09-02

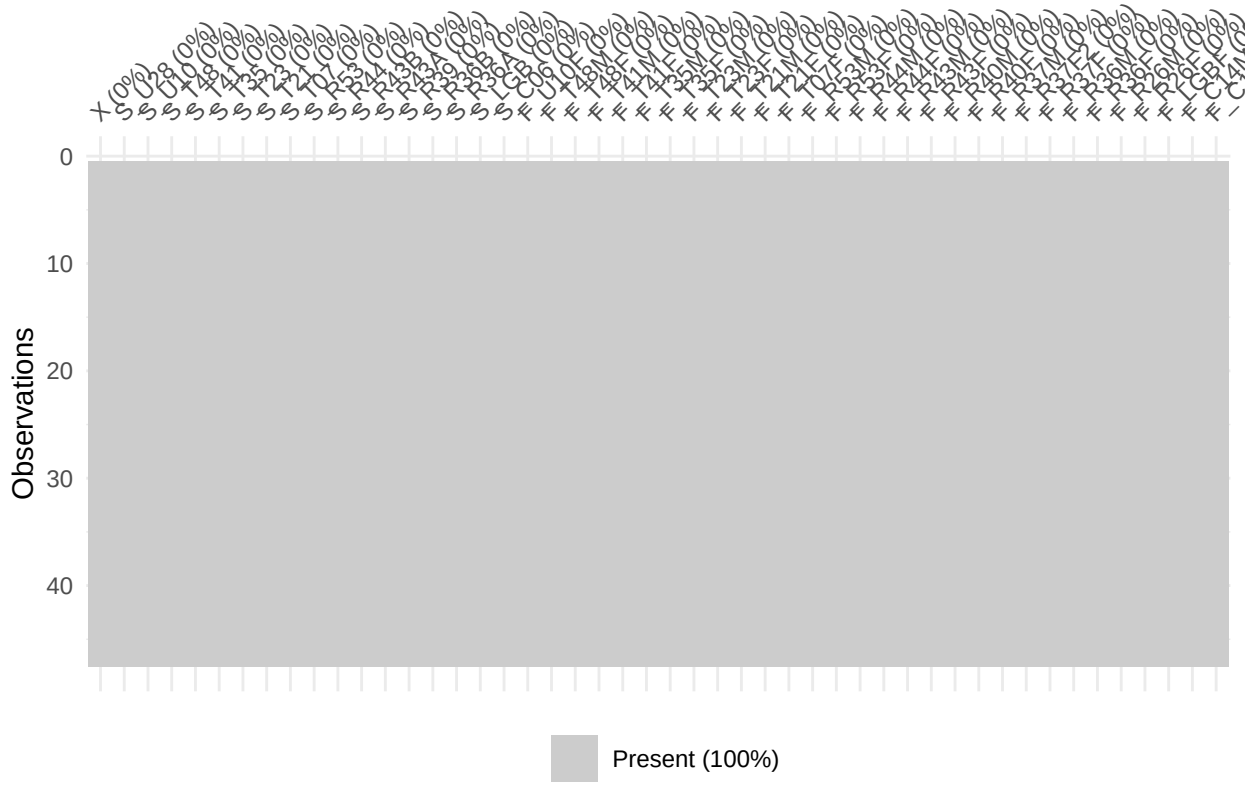
TRU 4001 - Biostatistics - Final Project

Introduction

Data from nest boxes at 47 mountain chickadee sites was collected to examine microbial community compositions associated with chickadees in urban, semi-urban, and rural environments. DNA sequencing was used to determine the relative abundance of different microbial taxa present in swabs from either chickadee nests or feathers in nest boxes in these three habitat types. Nest boxes were set up to encourage nesting in the sample sites. For the feather data, physical characteristics of the birds were recorded. The data was also used to determine two different measures of microbial species richness (alpha diversity) at each site.

1. Numerical Summary





The data consists of 47 observations with 12 variables: Site, Habitat, Source, EscShi, Entero, CommRich, PathRich, WingChord, BirdWeight, TailLen, TarsusLen, and BirdSex. Below are the summary statistics for key variables:

Categorical Variables:

Habitat: 3 types (Rural: 20, Semi-urban: 14, Urban: 13)

Source: 2 types (Feather: 30, Nest: 17)

CommRich: 2 types (High: 19, Low: 28)

BirdSex: 2 types (F: 17, M: 13) [missing: 17]

Numerical Variables:

EscShi: Min: 0, 1st Quartile: 25, Median: 112, Mean: 577.9, 3rd Quartile: 432.5, Max: 6520

Entero: Min: 0, 1st Quartile: 9, Median: 20, Mean: 92.74, 3rd Quartile: 48.5, Max: 687

PathRich: Min: 11, 1st Quartile: 36, Median: 41, Mean: 44.3, 3rd Quartile: 53.5, Max: 88

Variables with Missing Data:

WingChord: 30 non-missing (Min: 60, Median: 65, Max: 69) [missing: 17]

BirdWeight: 30 non-missing (Min: 10, Median: 11, Max: 17) [missing: 17]

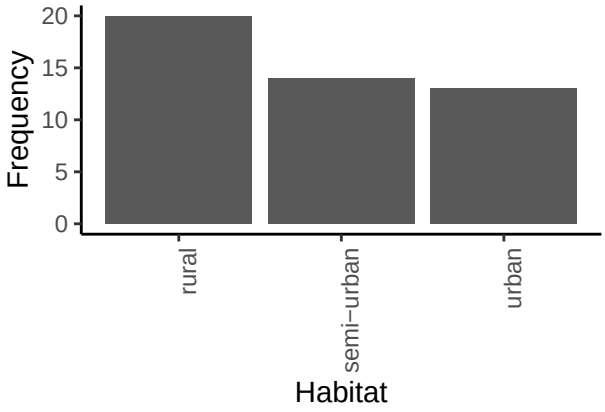
TailLen: 30 non-missing (Min: 53, Median: 57, Max: 62) [missing: 17]

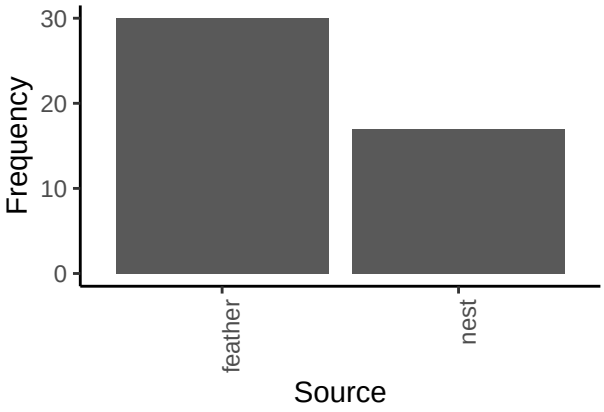
TarsusLen: 30 non-missing (Min: 16.8, Median: 18.25, Max: 19.5) [missing: 17]

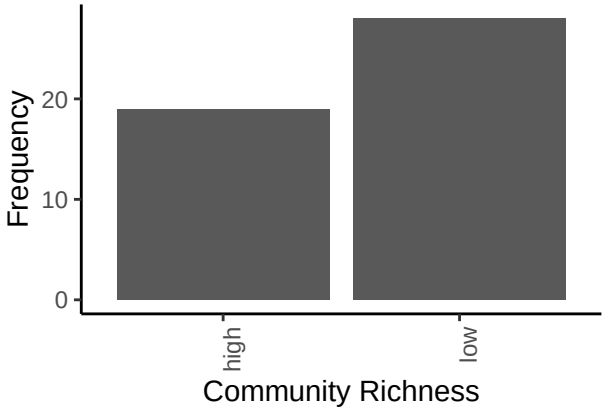
We find that 15.1% of the chickadee dataset is missing, while none of the dissimilarities dataset is.

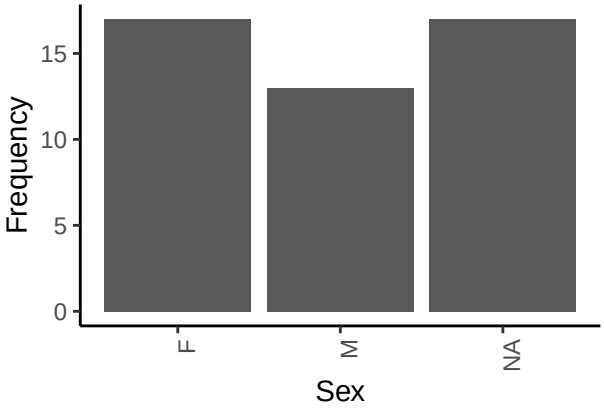
The data shows a variety of bird habitats and sources, with most coming from feathers. The data also suggests wide variations in the numerical variables, like EscShi and Entero. There are also quite a few missing values for variables related to bird physical features, which should be taken into account in any subsequent analysis.

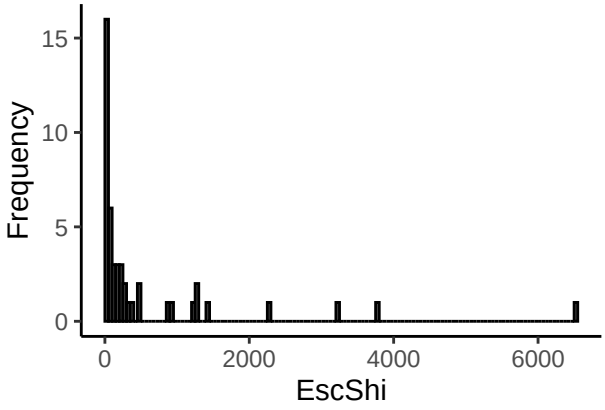
2. Graphical Summary

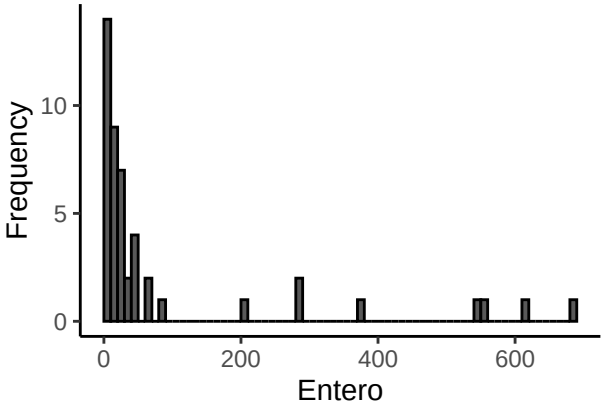


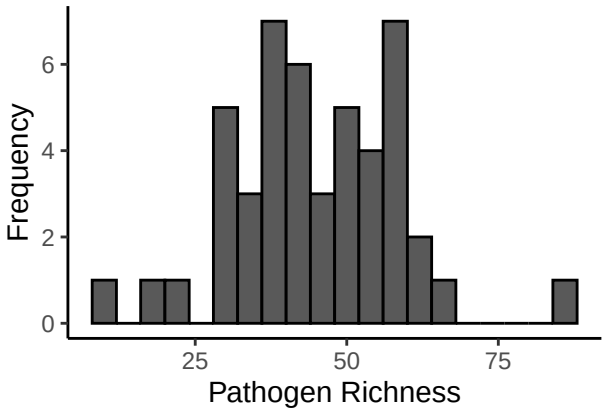


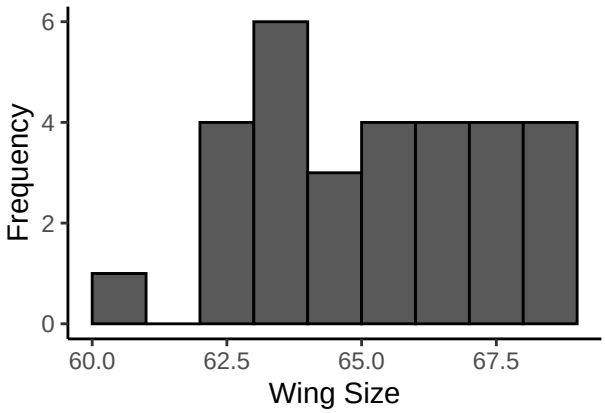


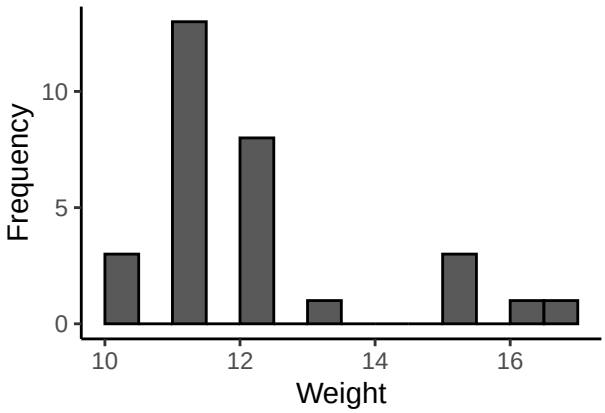


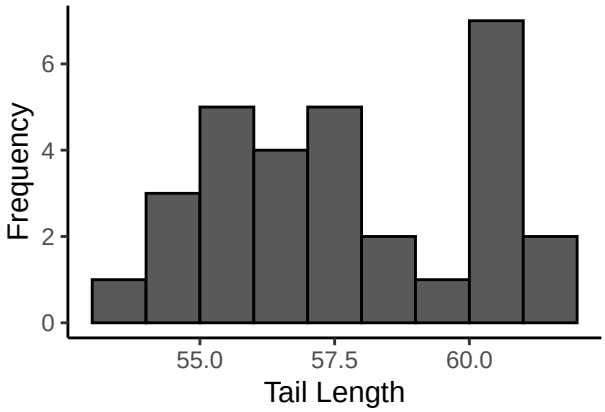


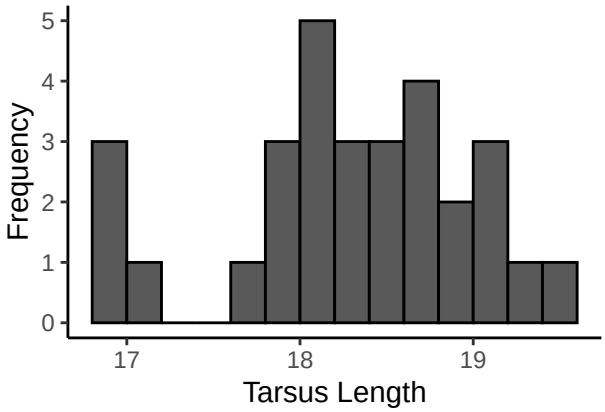


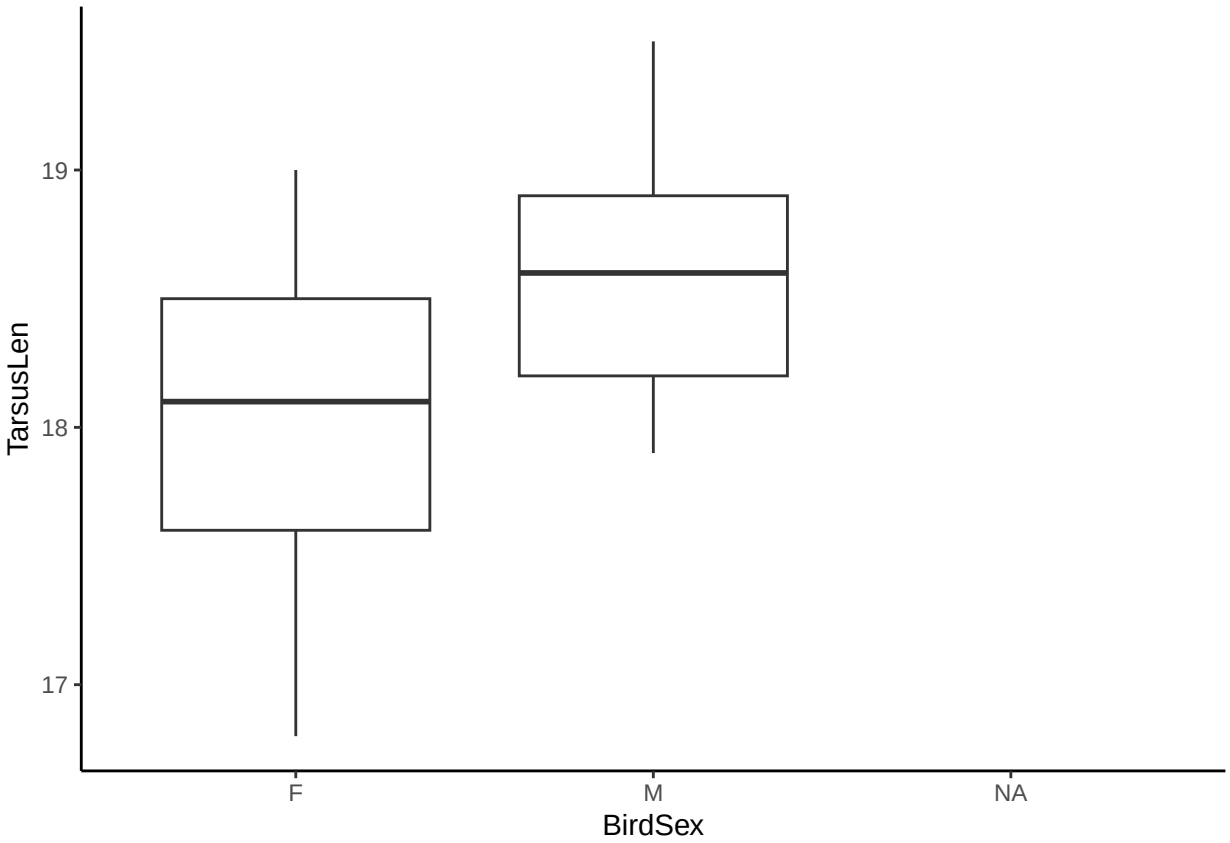


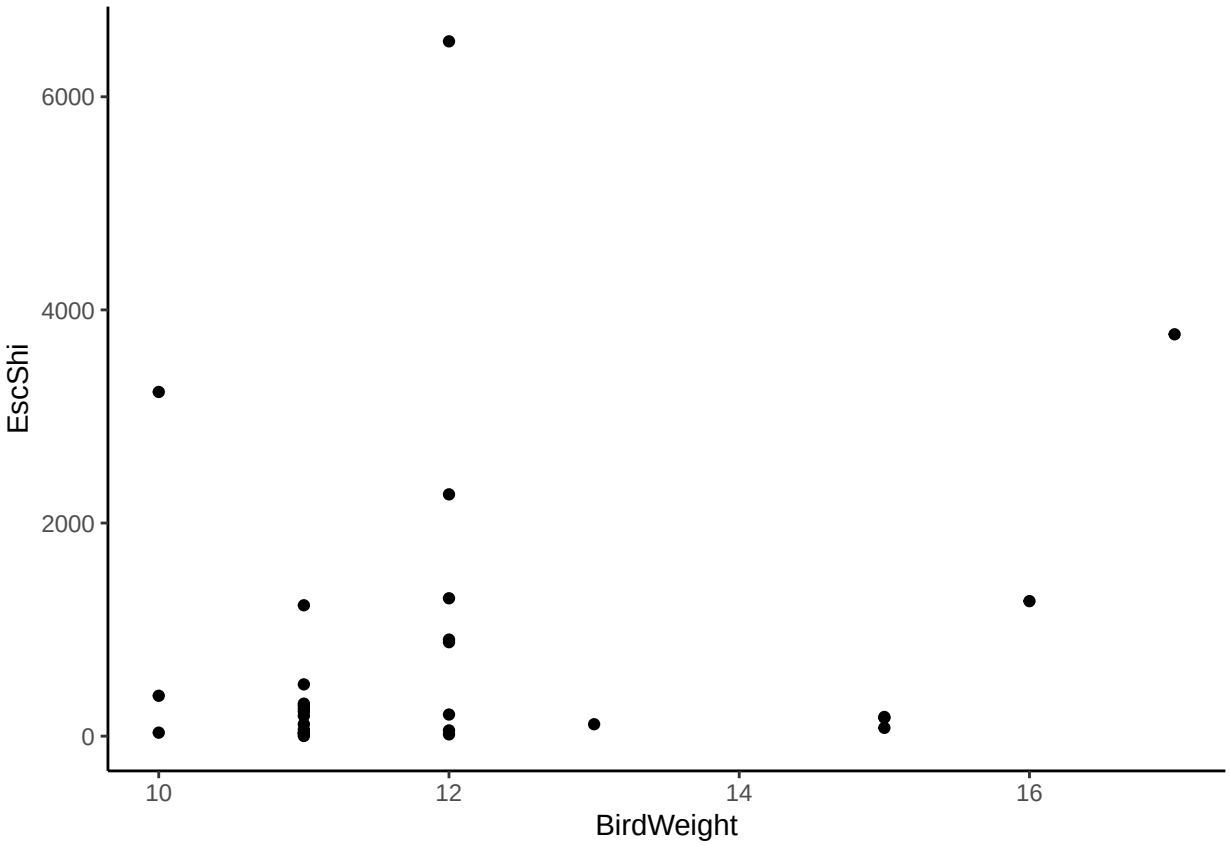

Frequency
0
2
4
6
60.0
62.5
65.0
67.5
Wing Size

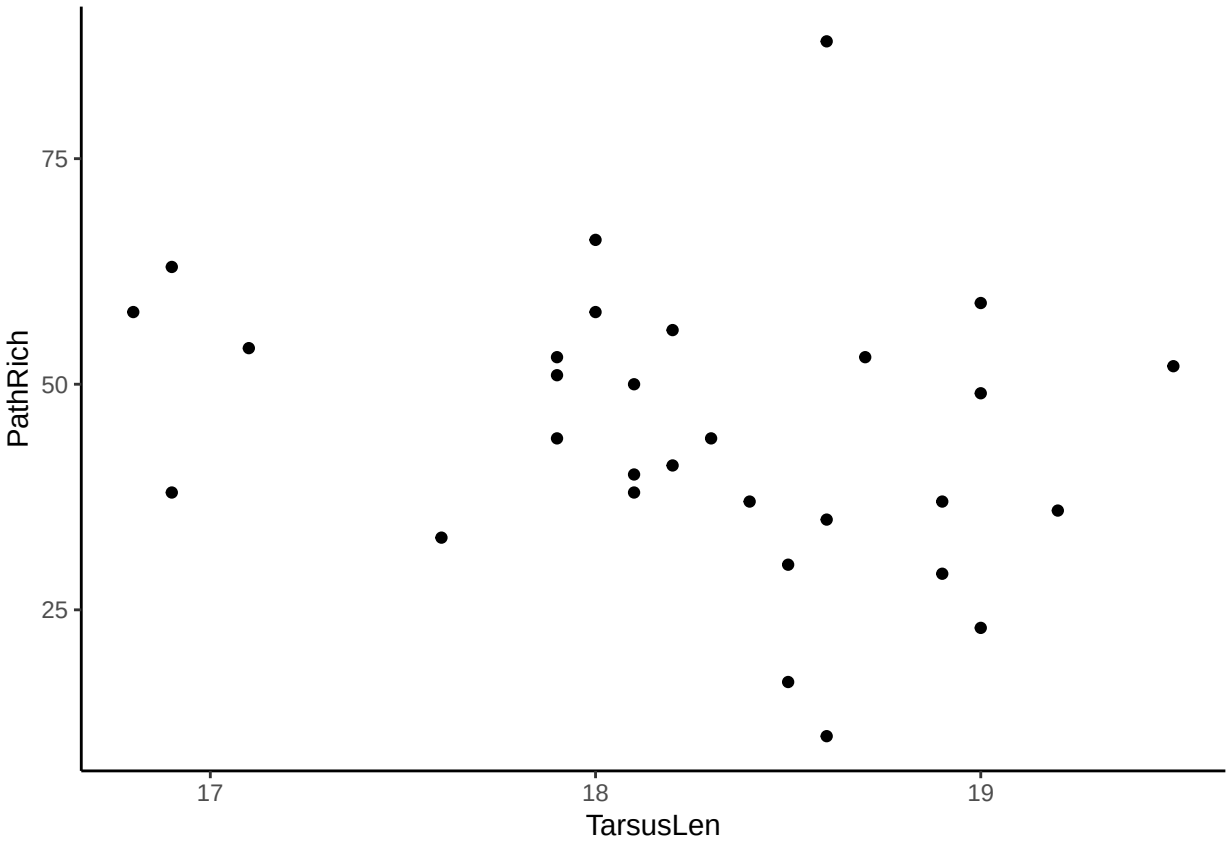

Frequency
0
5
10
10
12
14
16
Weight

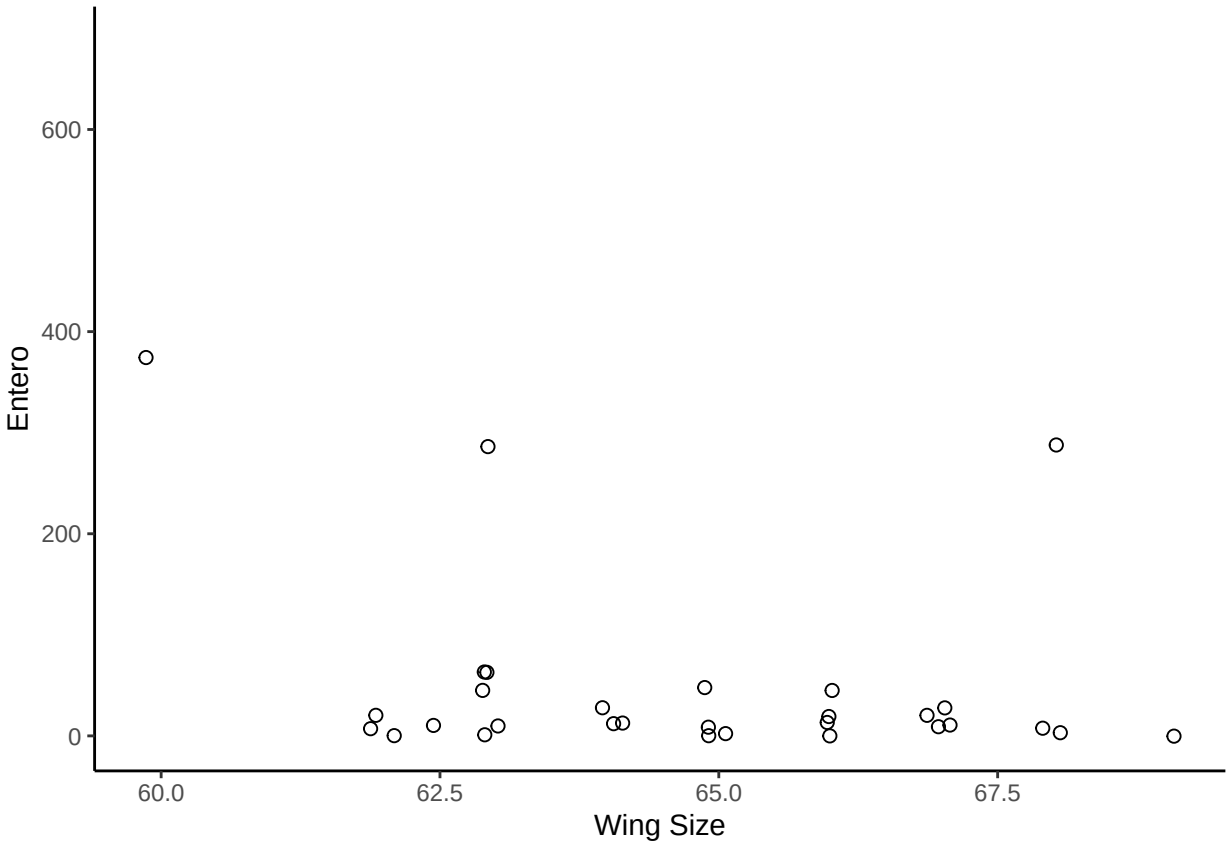

Frequency
0
2
4
6
55.0
57.5
60.0
Tail Length

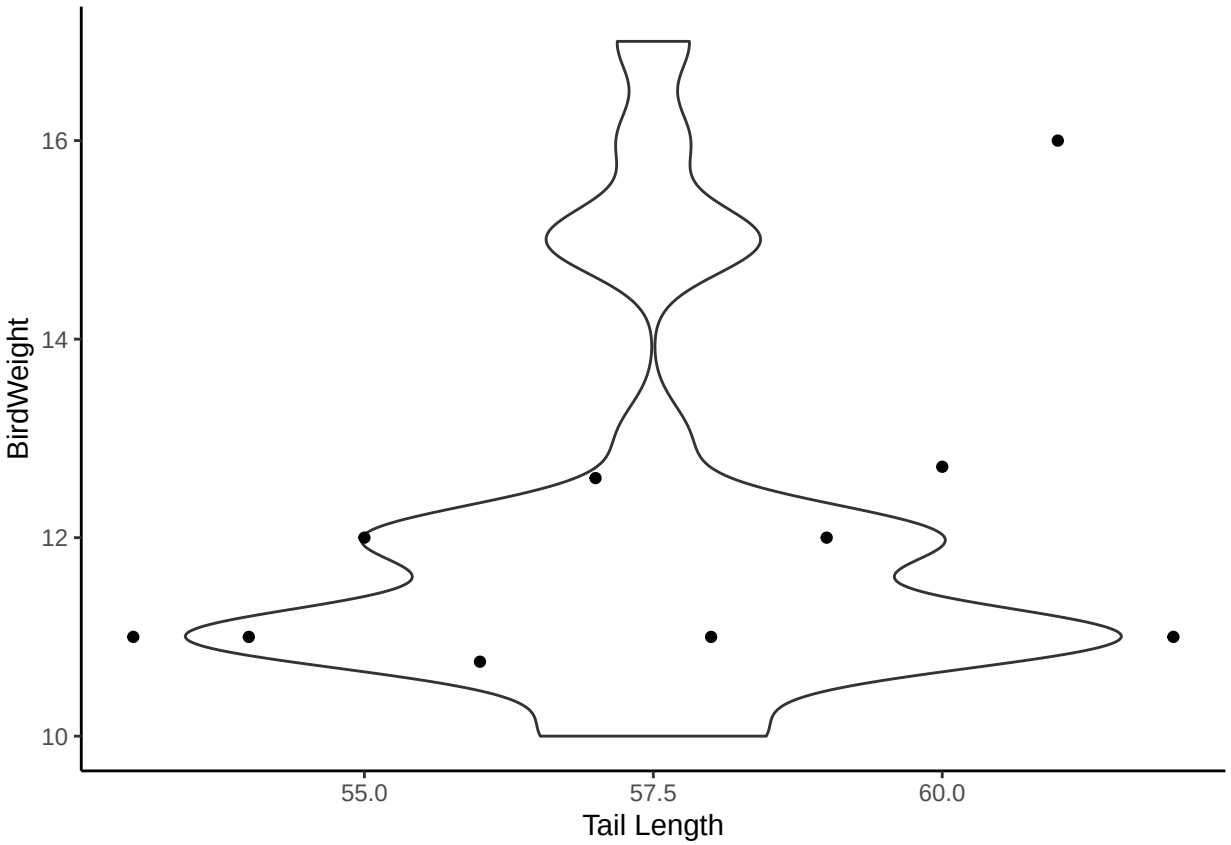

Frequency
0
1
2
3
4
5
17
18
19
Tarsus Length


19
18
17
TarsusLen
F
M
NA
BirdSex


6000
4000
2000
0
EscShi
10
12
14
16
BirdWeight


75
50
25
PathRich
17
18
19
TarsusLen


600
400
200
0
Entero
60.0
62.5
65.0
67.5
Wing Size



Upon visual inspection, we find that *Escherichia/Shigella* and *Enterococcus* distributions are right-skewed, and that pathogen richness, wing size, tail length, and tarsus length are approximately normally distributed.

3. Two-Sample *t*-Test For *Escherichia/Shigella* Between Nest and Feathers

In the two-sample t-test conducted to determine whether the mean abundance of bacteria identified as genus *Escherichia/Shigella* differs significantly between nests and feathers, our null hypothesis is that the true difference in means between the feather group and the nest group is equal to 0. The alternative hypothesis is that this difference is not equal to 0.

Based on the test results for the natural logarithm-transformed variable, the test statistic t is 4.0756, and the p-value is 0.0001842. Given that the p-value is less than the alpha level of 0.05, we reject the null hypothesis.

We conclude that there is a statistically significant difference in the mean abundance of bacteria identified as genus *Escherichia/Shigella* between nests and feathers, based on this data.

4. Mann-Whitney U-Test For *Escherichia/Shigella* Between Nest and Feathers

In the Mann-Whitney U-test conducted to determine whether the population distribution of the abundance of bacteria identified as genus *Escherichia/Shigella* differs significantly between nests and feathers, our null hypothesis is that the true location shift between the feather group and the nest group is equal to 0. The alternative hypothesis is that this location shift is not equal to 0.

Based on the test results for the untransformed data, the test statistic W is 405.5, and the p-value is 0.0008952. Since the p-value is less than the alpha level of 0.05, we reject the null hypothesis.

We conclude that there is statistically significant evidence to suggest that the distributions of abundance for bacteria identified as genus *Escherichia/Shigella* differ between nests and feathers, based on this data.

The Mann-Whitney U-test is useful in this context as it does not require the normal distribution assumption that the two-sample t-test does. Therefore, it allows us to extend our conclusions to the original, untransformed population, providing further evidence for a difference in bacterial abundance between the two groups.

5. Two-Sample *t*-Test For *Enterococcus* Between Nest and Feathers

In the two-sample t-test conducted to assess whether the mean abundance of bacteria identified as genus *Enterococcus* differs significantly between nests and feathers, our null hypothesis is that the true difference in means between the feather group and the nest group is equal to 0. The alternative hypothesis is that this difference is not equal to 0.

Based on the results of the test on the natural logarithm transformed data, the test statistic t is -2.6195, and the p-value is 0.01196. Since the p-value is less than the alpha level of 0.05, we reject the null hypothesis.

We conclude that there is statistically significant evidence to suggest that the mean abundance of bacteria identified as genus *Enterococcus* differs between nests and feathers, based on this data. Additionally, the test indicates that the mean abundance of this bacterial genus is higher in nests compared to feathers.

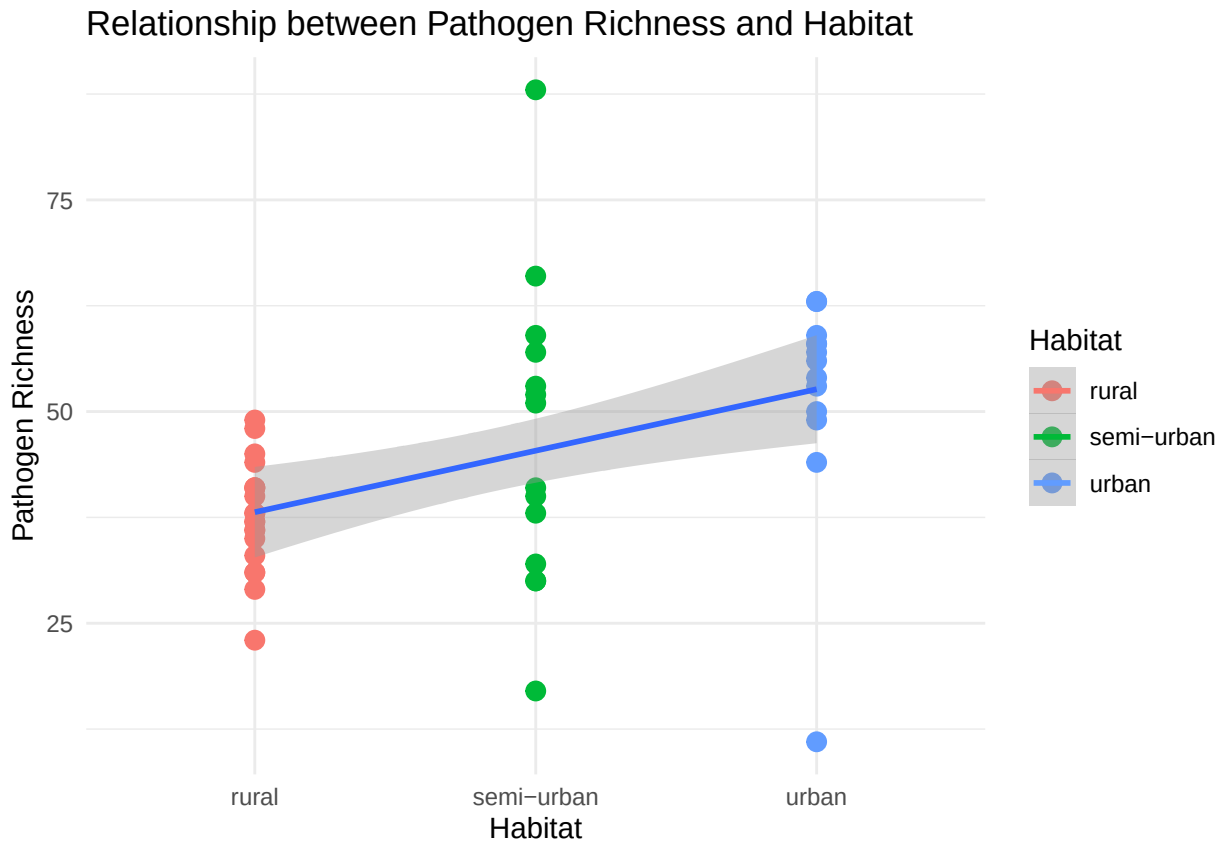
6. Mann-Whitney U-Test For *Enterococcus* Between Nest and Feathers

In the Mann-Whitney U-test to examine whether the population distribution of the abundance of bacteria identified as genus *Enterococcus* varies significantly between nests and feathers, our null hypothesis is that the true location shift between the feather and nest groups is equal to 0. The alternative hypothesis is that this location shift is not equal to 0.

From the test results, the test statistic W is 148, and the p-value is 0.01826. Given that the p-value is less than the alpha level of 0.05, we reject the null hypothesis.

We conclude that there is statistically significant evidence to suggest that the population distribution of the abundance of bacteria identified as genus *Enterococcus* differs between nests and feathers. Furthermore, based on this evidence and previous data, we infer that the abundance of *Escherichia/Shigella* is higher in feathers, while the abundance of *Enterococcus* is higher in nests.

7. Analysis of Variance for Pathogen Richness and Habitat



In the analysis of variance (ANOVA), we assess whether the mean pathogen richness is significantly related to habitat. The null hypothesis is that there is no relationship between habitat and pathogen richness, while the alternative hypothesis is that such a relationship does exist.

According to the linear model results, the F-statistic is 5.275 with a corresponding p-value of 0.008844. Given that this p-value is below the alpha level of 0.05, we reject the null hypothesis.

The model shows statistically significant coefficients for both semi-urban and urban habitats, with p-values of 0.04777 and 0.00305 respectively. This suggests that the type of habitat is a meaningful predictor of pathogen richness.

The Multiple R^2 value of 0.1934 indicates that approximately 19.34% of the variability in pathogen richness is explained by habitat. While statistically significant, the model accounts for less than one-fifth of the variance, suggesting that other factors may also be influential.

We conclude that habitat is a statistically significant predictor of pathogen richness, although it explains only a limited portion of the variability. The trend in the estimates of coefficients implies that pathogen richness tends to increase in more densely populated human environments, such as semi-urban and urban areas, based on this data.

8. Kruskal-Wallis Test for Pathogen Richness and Habitat

In the Kruskal-Wallis test, we assess whether mean pathogen richness varies significantly across different habitats. The null hypothesis is that there are no differences in mean pathogen richness across habitats, while the alternative hypothesis is that at least one habitat exhibits a difference in mean pathogen richness.

The Kruskal-Wallis X^2 test statistic is 13.587, with 2 degrees of freedom, and a p-value of 0.001121. Given that the p-value is well below the alpha level of 0.05, we reject the null hypothesis.

We conclude that there is a statistically significant difference in mean pathogen richness across habitats, based on this data. This evidence aligns with the findings from the linear model, supporting the significance of habitat as a factor influencing pathogen richness.

9. Tukey-Kramer Test for Pathogen Richness and Specfic Habitat

In the Tukey-Kramer post-hoc test, we assess which specific habitats differ significantly in terms of mean pathogen richness. The test provides pairwise comparisons between the three habitat types: rural, semi-urban, and urban.

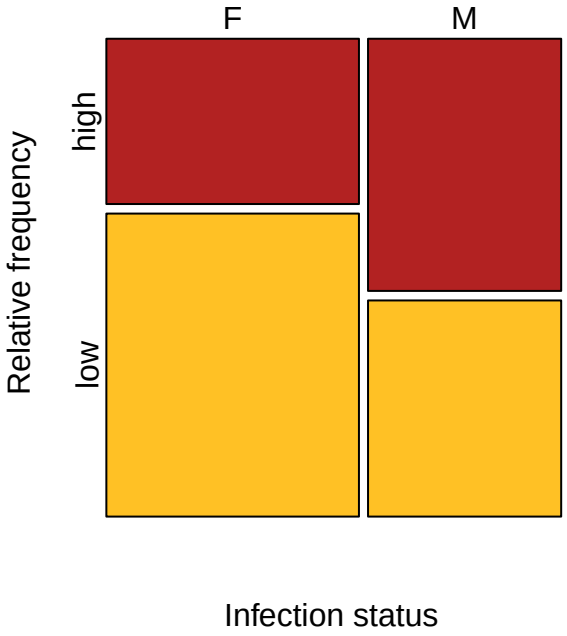
Rural vs. Semi-Urban: The estimated mean difference is -9.06, with a p-value of 0.1155. Since this p-value exceeds the alpha level of 0.05, the difference in pathogen richness between rural and semi-urban habitats is not statistically significant.

Rural vs. Urban: The estimated mean difference is -14.27, with a p-value of 0.0084. This p-value is less than the alpha level of 0.05, indicating a statistically significant difference in pathogen richness between rural and urban habitats.

Semi-Urban vs. Urban: The estimated mean difference is -5.21, with a p-value of 0.5445. As this p-value is greater than 0.05, the difference in pathogen richness between semi-urban and urban habitats is not statistically significant.

Based on the Tukey-Kramer test, we can conclude that pathogen richness is significantly higher in urban areas compared to rural areas. However, no statistically significant differences in pathogen richness were found between rural and semi-urban habitats, or between semi-urban and urban habitats. These findings are consistent with those from the linear regression model and the Kruskal-Wallis test, further substantiating the influence of habitat on pathogen richness.

10. Contingency Table Analysis and Chi-Squared Test for Community Richness and Sex



In this contingency table analysis using Pearson's Chi-squared test, we assess whether community richness on feathers is independent of the sex of mountain chickadees. The null hypothesis is that community richness and the sex of the bird are independent variables, while the alternative hypothesis is that they are not independent.

According to the test results, the Chi-squared statistic X^2 is 1.0325 with 1 degree of freedom, and the p-value is 0.3096. Given that the p-value exceeds the alpha level of 0.05, we fail to reject the null hypothesis.

We conclude that there is insufficient evidence to suggest that community richness on feathers is dependent on the sex of mountain chickadees, based on this data.

11. Linear Regression of Morphological Features and Pathogen Richness

In this analysis, simple linear regression was used to examine the linear relationship between four bird features (Wing Chord, Bird Weight, Tail Length, and Tarsus Length) and pathogen richness on feathers. Additionally, multiple linear regression was conducted to examine the combined effect of these predictors.

Wing Chord: With an R^2 of 0.0344 and a p-value of 0.3262, Wing Chord is not a statistically significant predictor of pathogen richness.

Bird Weight: Similarly, Bird Weight shows a low R^2 of 0.0130 and a high p-value of 0.5492, indicating it is also not a statistically significant predictor.

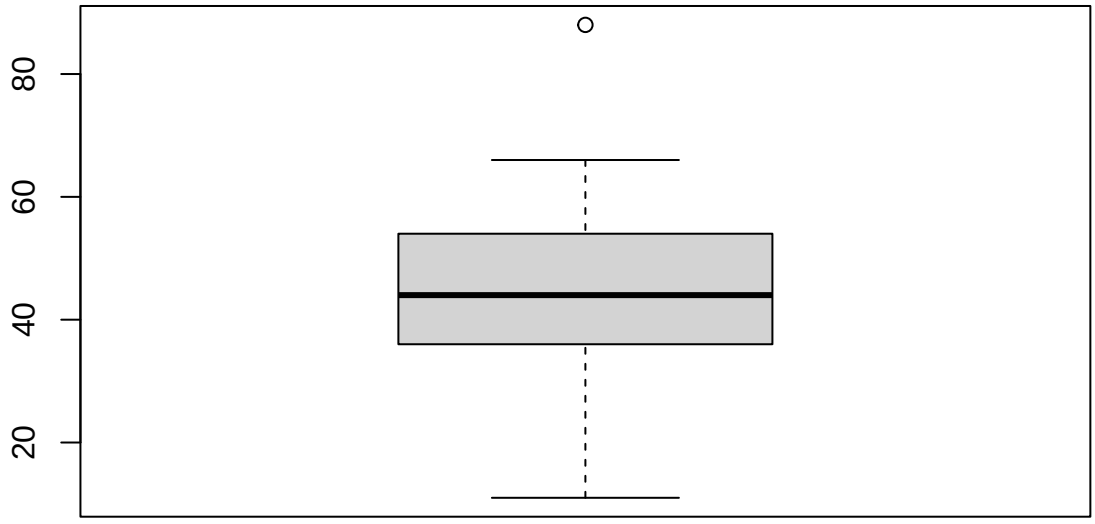
Tail Length: Among the features, Tail Length had the highest R^2 value of 0.1139 and the lowest p-value of 0.0682, which approaches significance. While not statistically significant at the alpha = 0.05 level, this variable shows the most promise as a predictor among those tested.

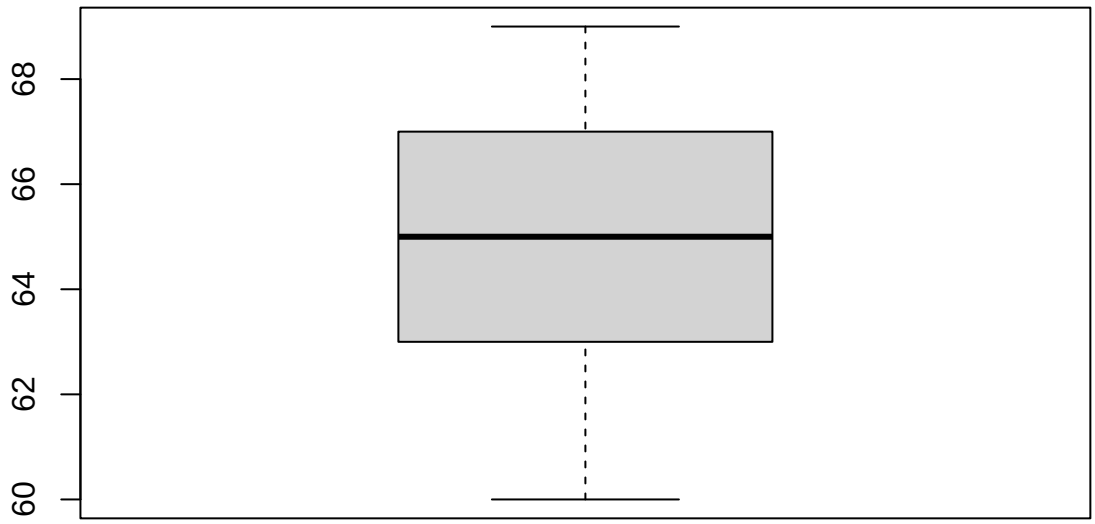
Tarsus Length: With an R^2 value of 0.0587 and a p-value of 0.1971, Tarsus Length was also not found to be a statistically significant predictor but does have a negative relationship with pathogen richness.

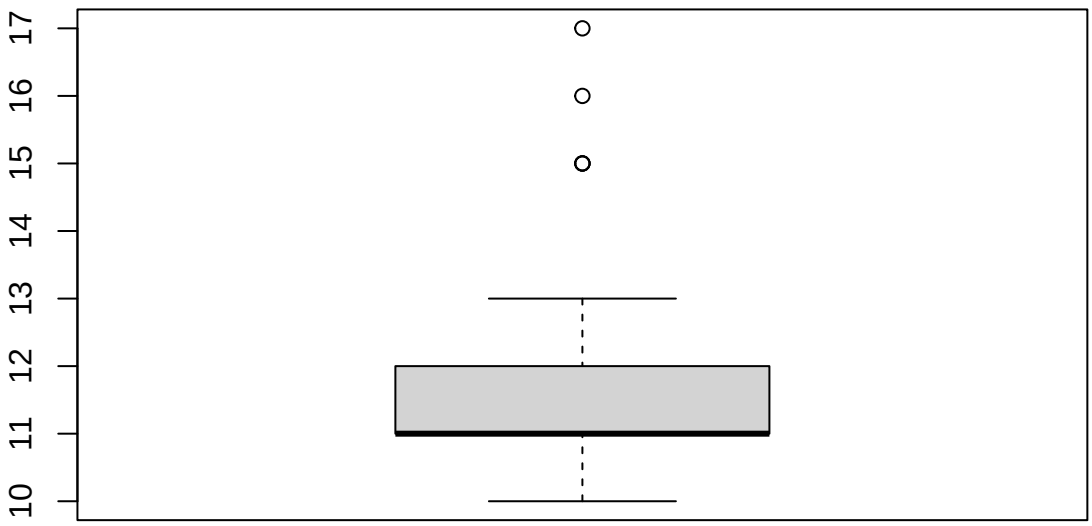
None of the single predictors are statistically significant based on their p-values exceeding the alpha level of 0.05.

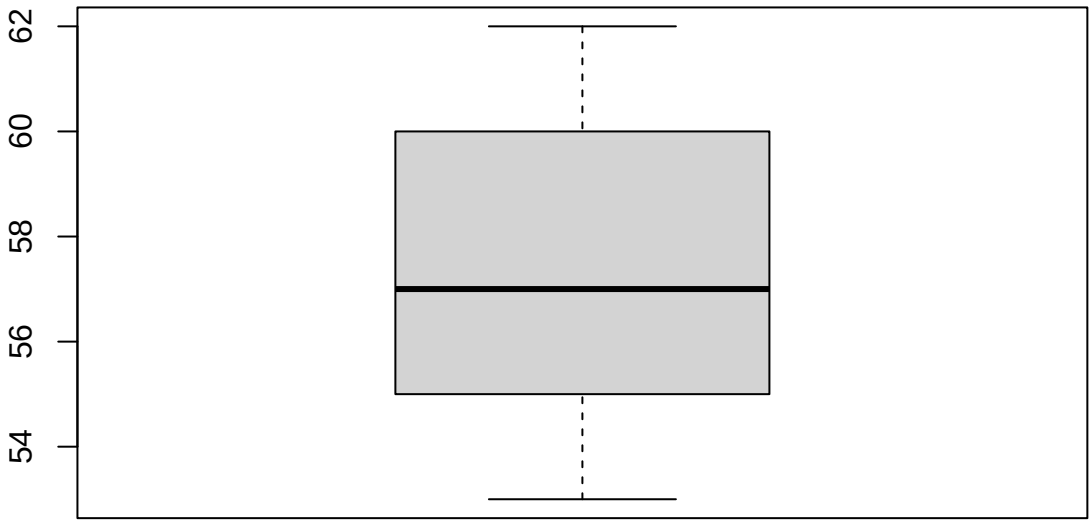
The multiple regression model yielded a p-value of 0.3015 and an R^2 of 0.1708, indicating that collectively, the predictors are not statistically significant at explaining the variability in pathogen richness. The multiple regression model also failed to reach statistical significance, supporting the null hypothesis that none of these predictors significantly influence pathogen richness on feathers.

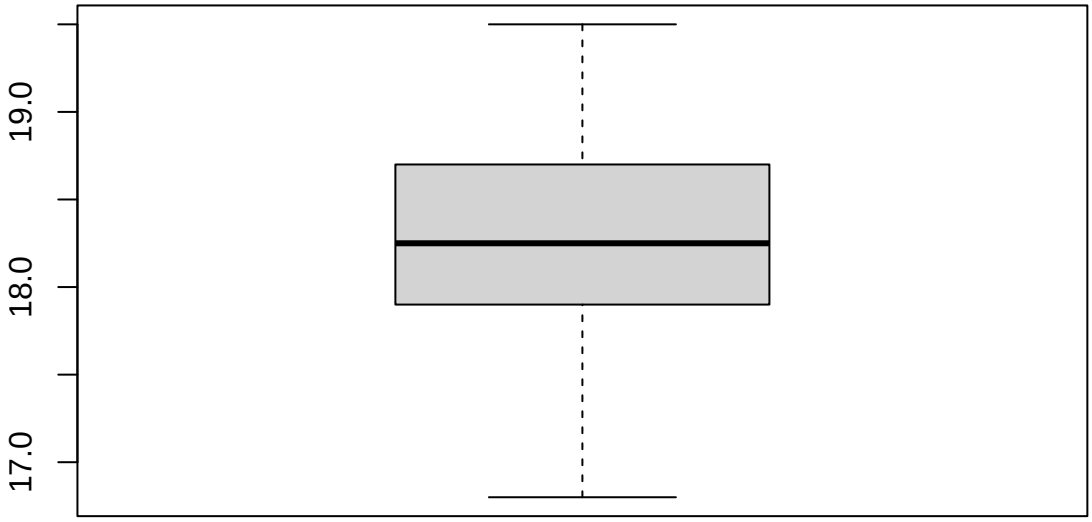
12. Linear Regression of Morphological Features and Pathogen Richness With Outliers Removed

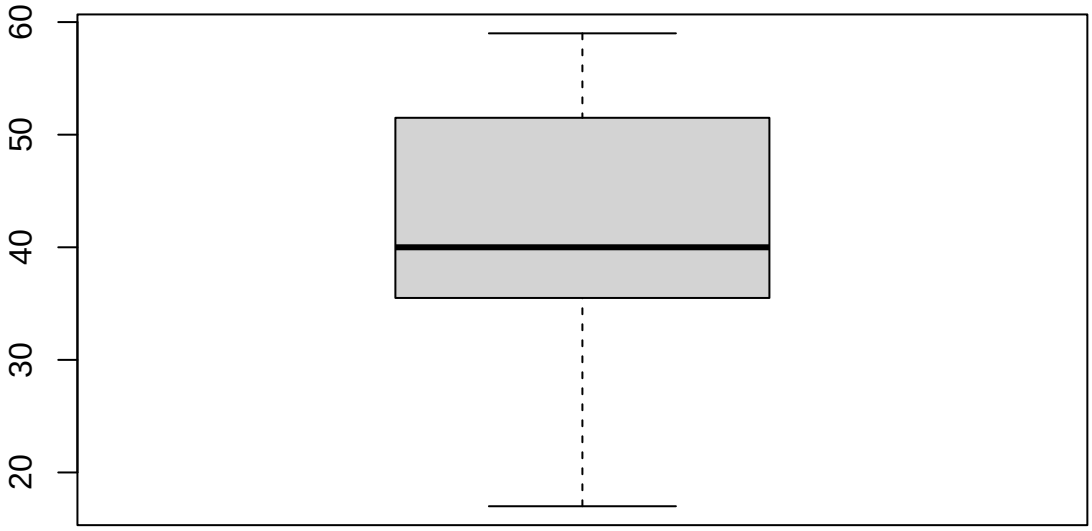


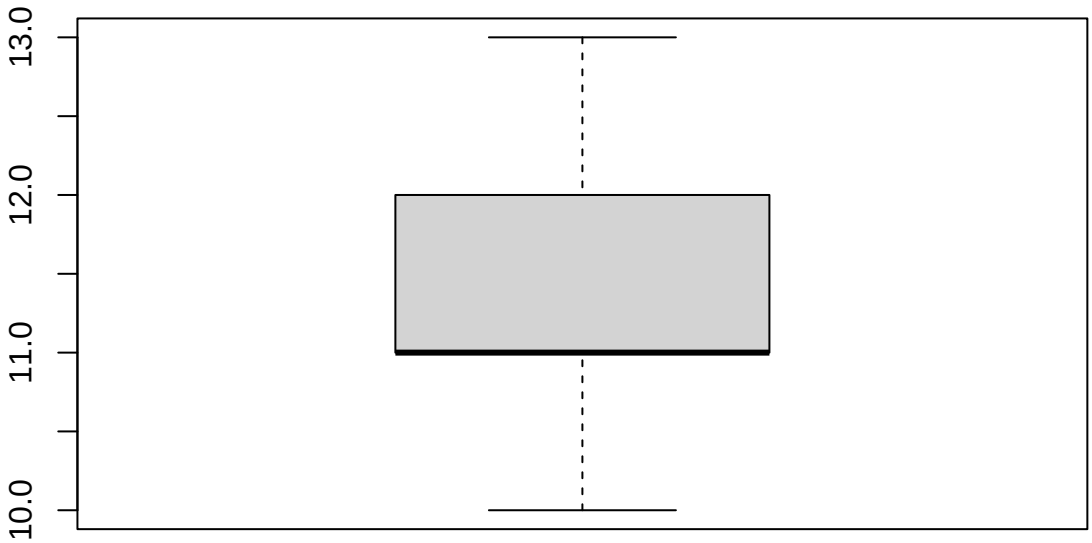


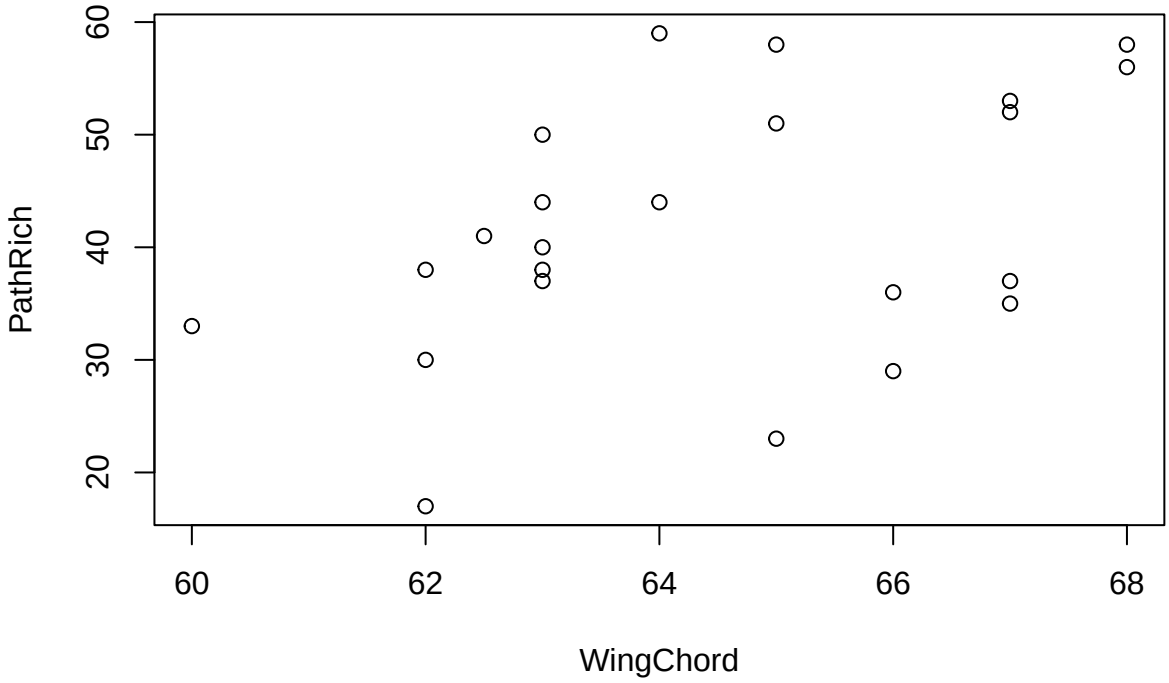

17
16
15
14
13
12
11
10

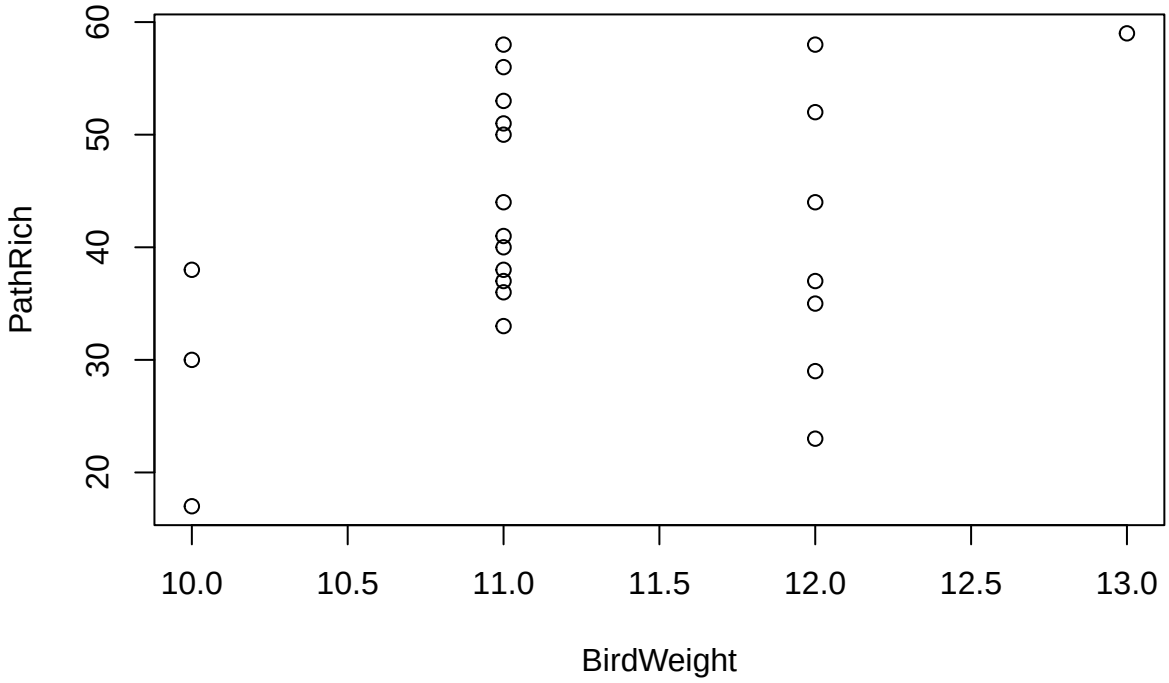

62
60
58
56
54

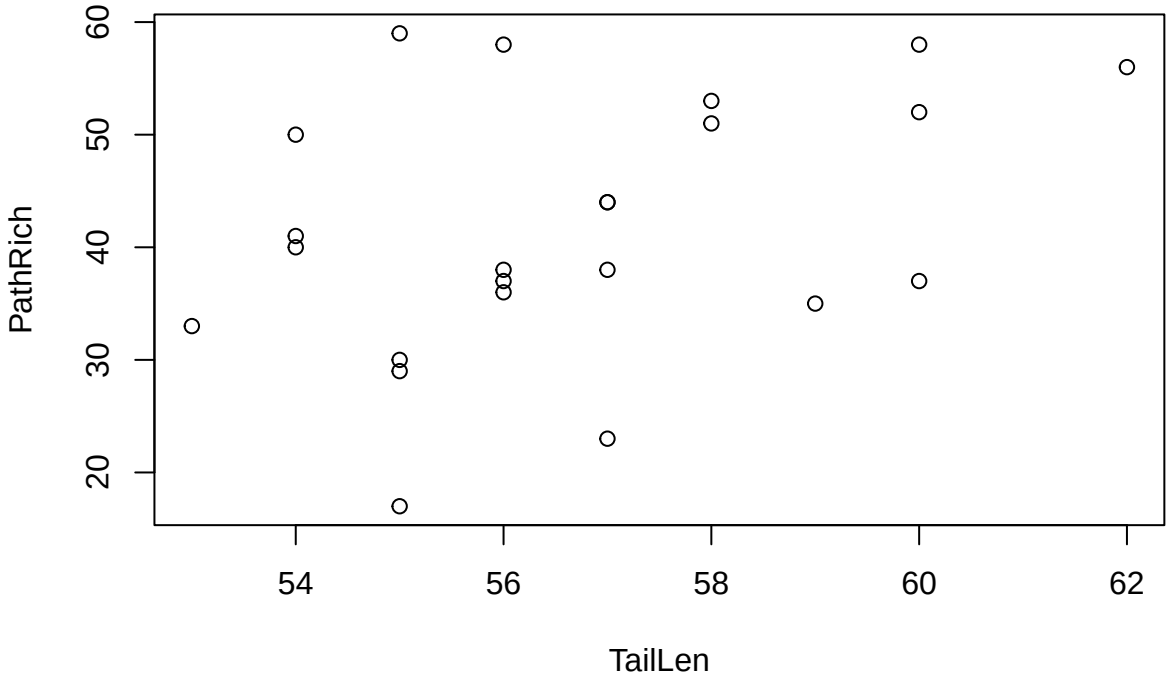

19.0
18.0
17.0

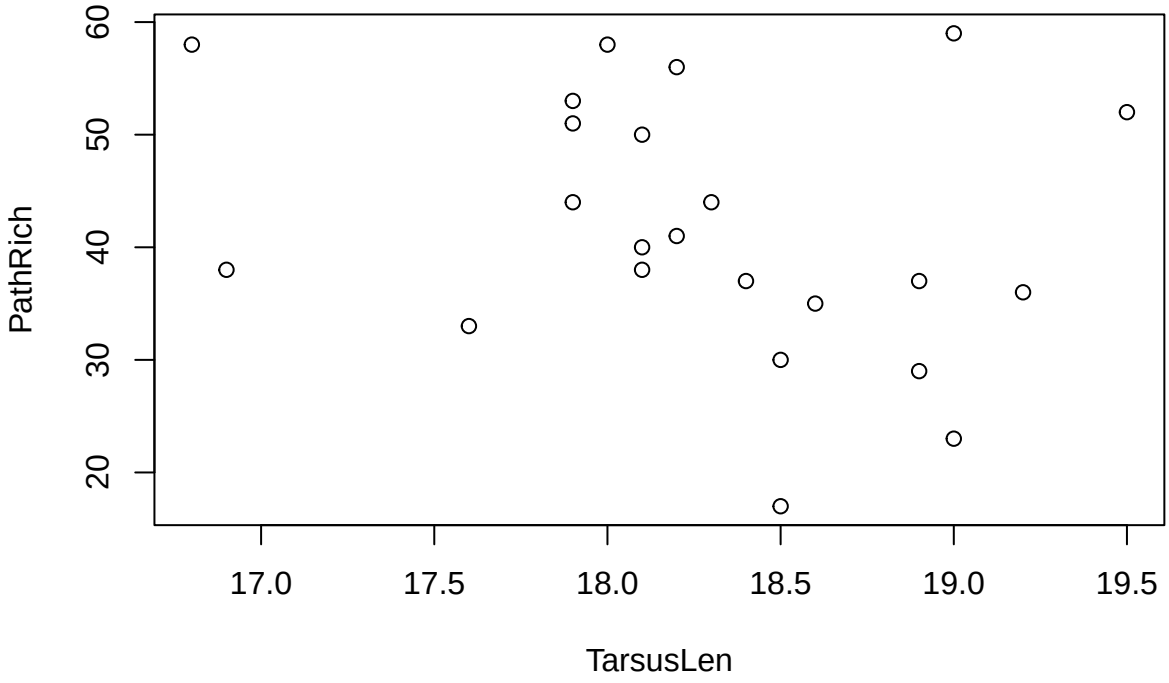

60
50
40
30
20

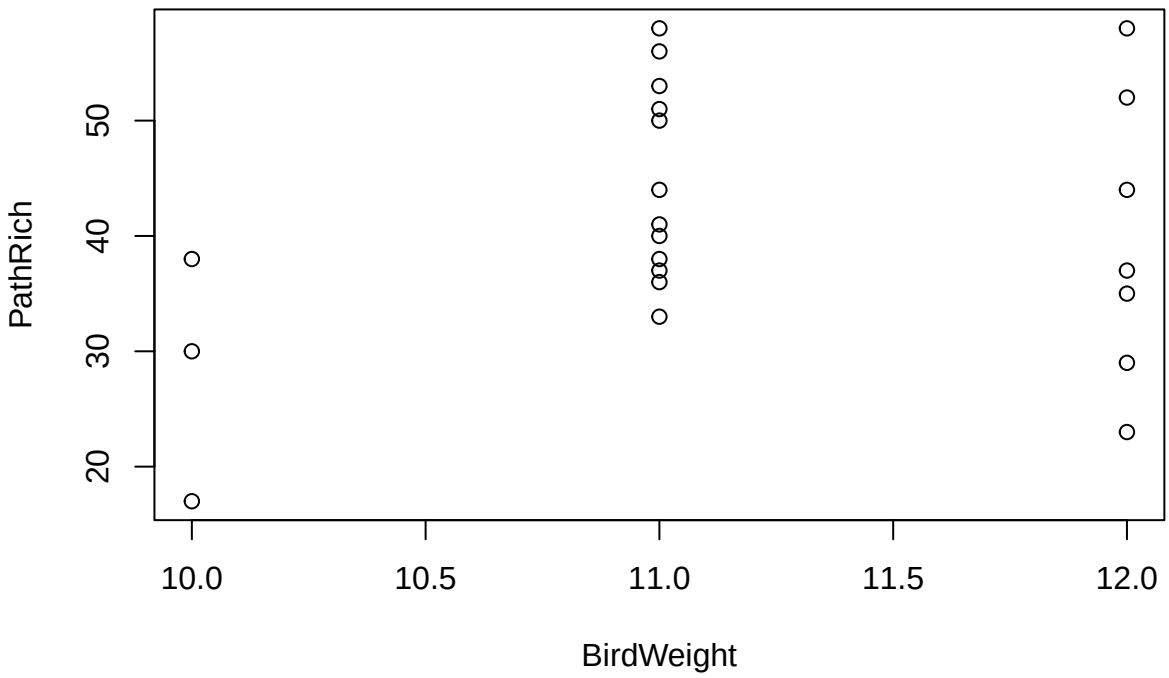

13.0
12.0
11.0
10.0


PathRich
20
30
40
50
60
60
62
64
66
68
WingChord


PathRich
20
30
40
50
60
10.0
10.5
11.0
11.5
12.0
12.5
13.0
BirdWeight


PathRich
20
30
40
50
60
54
56
58
60
62
TailLen


PathRich
20
30
40
50
60
17.0
17.5
18.0
18.5
19.0
19.5
TarsusLen



After identifying and removing the most extreme outliers from our dataset, we refit our simple linear regression models and found that two predictor variables remained statistically significant.

Wing Chord: The model estimates a coefficient of 2.2607 for Wing Chord, with a p-value of 0.0301, which is significant at the 0.05 level. The Multiple R^2 value for this model is 0.2141.

Tail Length: The model estimates a coefficient of 2.1883 for Tail Length, with a p-value of 0.0307, also significant at the 0.05 level. The Multiple R^2 value for this model is 0.2128.

When we used all four predictors in a multiple linear regression model, the overall model was significant with a p-value of 0.006673, and a Multiple R^2 value of 0.5476. This suggests that our model explains approximately 54.76% of the variability in the response variable, which is a substantial improvement from the individual models.

Therefore, after excluding the most extreme outliers, we reject the null hypothesis that the predictors have no effect on the response variable for both the individual and multiple regression models. We conclude that there exists at least one predictor that has a statistically significant effect on the response variable.

13. Multiple Linear Regression of Morphological Features and Pathogen Richness to Find Best Two Predictors

To identify the best multiple linear regression model for predicting pathogen richness on feathers from two predictors among Wing Chord, Bird Weight, Tail Length, and Tarsus Length, we evaluated several models based on their p-values, Multiple R^2, Adjusted R^2, and Residual Standard Error (RSE).

Based on these criteria, three models emerged as significant:

Wing Chord + Tarsus Length Model:

P-value: 0.000625 (Significant)

Multiple R^2: 0.54

Adjusted R^2: 0.4916

RSE: 7.946

Bird Weight + Tarsus Length Model:

P-value: 0.04415 (Significant)

Multiple R^2: 0.2799

Adjusted R^2: 0.2042

RSE: 9.941

Tail Length + Tarsus Length Model

P-value: 0.003174 (Significant)

Multiple R^2: 0.4542

Adjusted R^2: 0.3968

RSE: 8.655

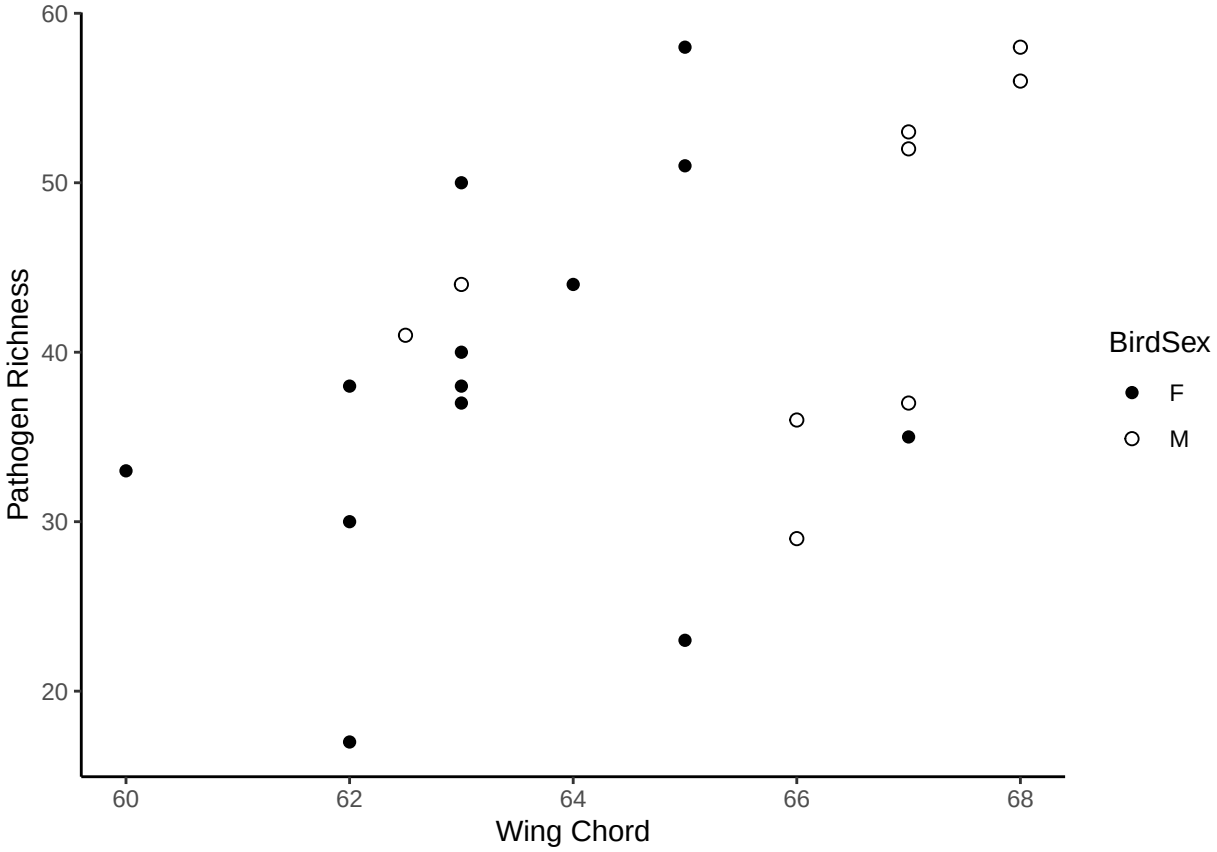
The model that includes Wing Chord and Tarsus Length as predictors stands out as the best model based on all the criteria we considered. It not only has the lowest p-value but also the highest Multiple R^2 and Adjusted R^2 values, along with the lowest RSE. This suggests that the model is both statistically significant and explains a substantial portion of the variance in pathogen richness.

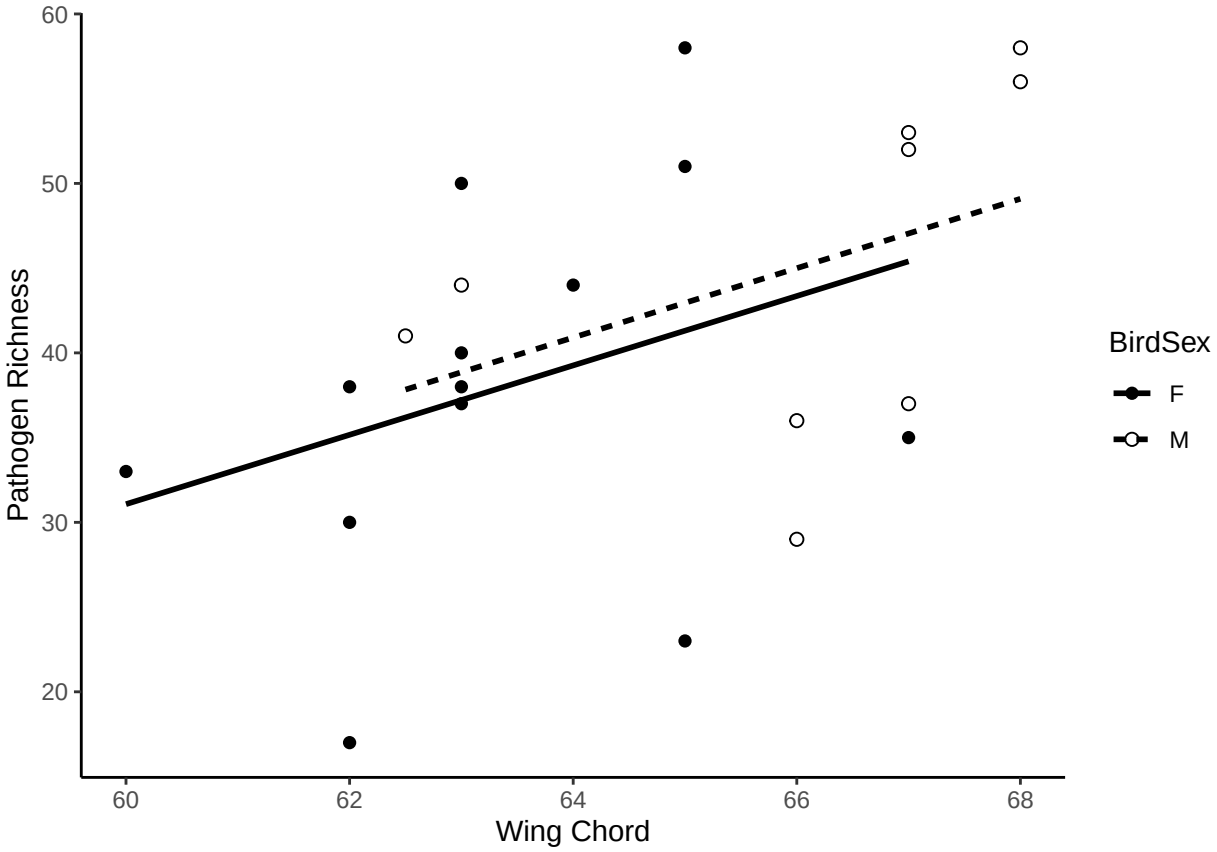
This finding is consistent with our single linear regression analysis, which also identified Wing Chord as a significant predictor. Although Tail Length was identified as a significant predictor in the single linear regression, the multiple regression analysis showed that a model including Wing Chord and Tarsus Lenth is superior based on our evaluation criteria.

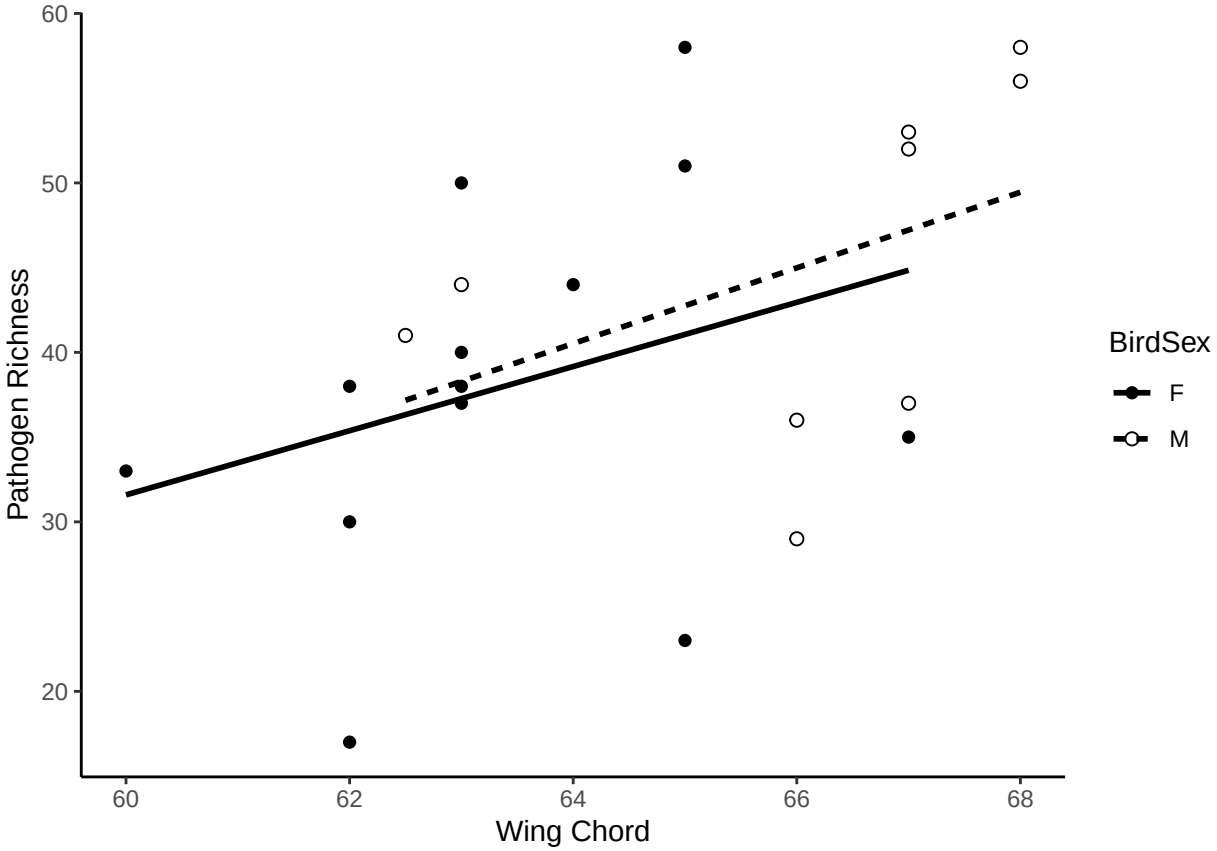
14. Two-Sample *t*-Test for Pathogen Richness on Feathers between Sexes

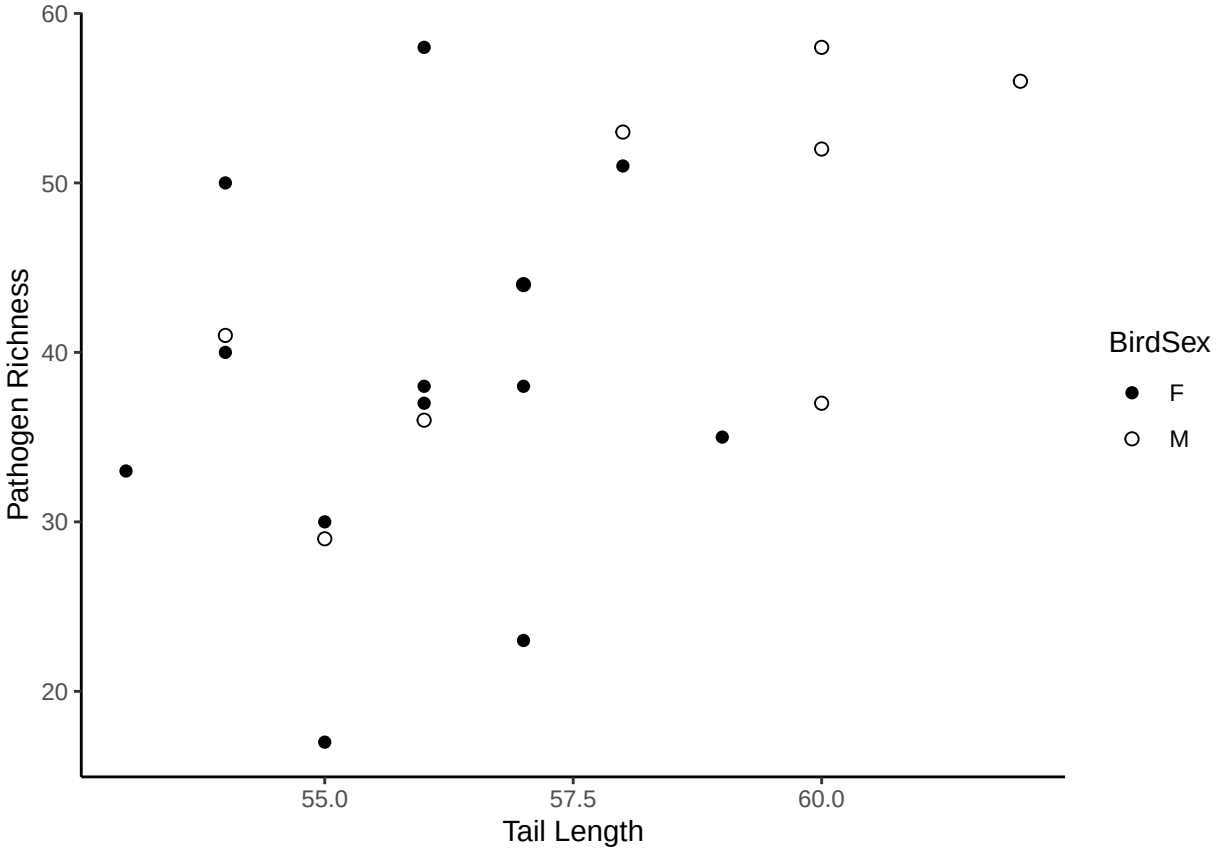
In this study, we assessed whether there is a significant difference in the mean pathogen richness on feathers between male and female birds. To do so, we used a two-sample t-test. The null hypothesis is that there is no significant difference between the means, while the alternative hypothesis asserts otherwise. Our analysis yielded the t-value of -1.5165 with 20 degrees of freedom, and a p-value of 0.1451. We find that the p-value exceeds the alpha level of 0.05. Thus, there is insufficient evidence to reject the null hypothesis. We conclude that there is no statistically significant difference in the mean pathogen richness on feathers between male and female birds based on this data.

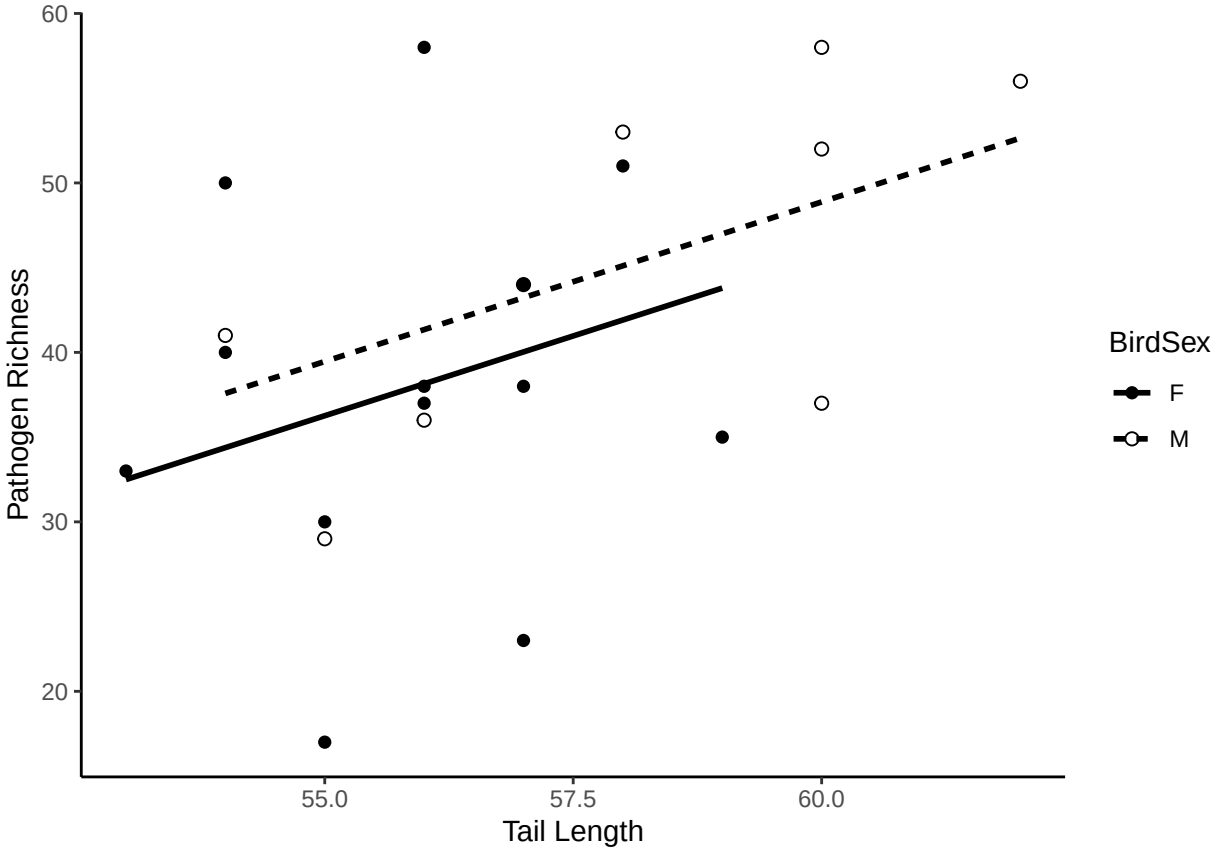
15. Analysis of Covariance on Linear Models between Sexes

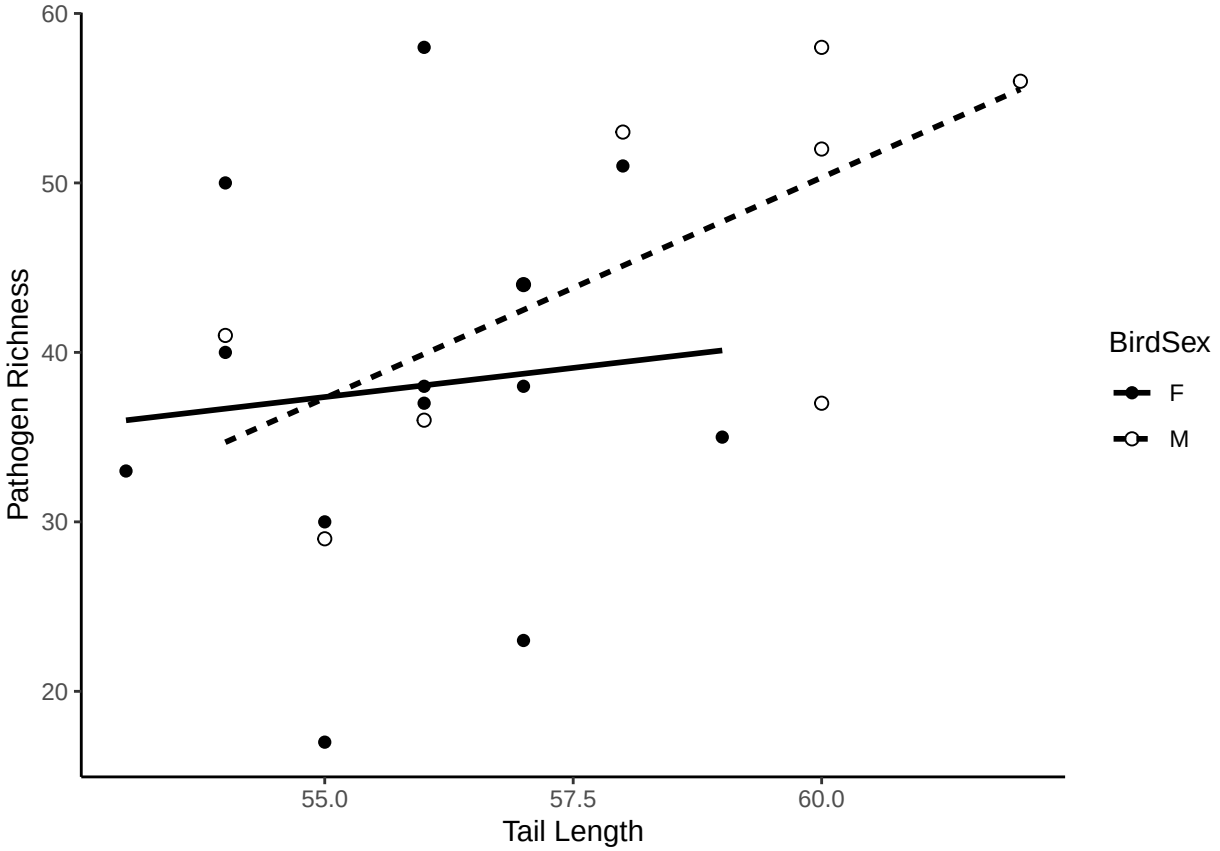





Pathogen Richness
60
50
40
30
20
60
62
64
66
68
Wing Chord
BirdSex
F
M


60
50
40
30
20
Pathogen Richness
55.0
57.5
60.0
Tail Length
BirdSex
F
M


60
50
40
30
20
Pathogen Richness
55.0
57.5
60.0
Tail Length
BirdSex
F
M

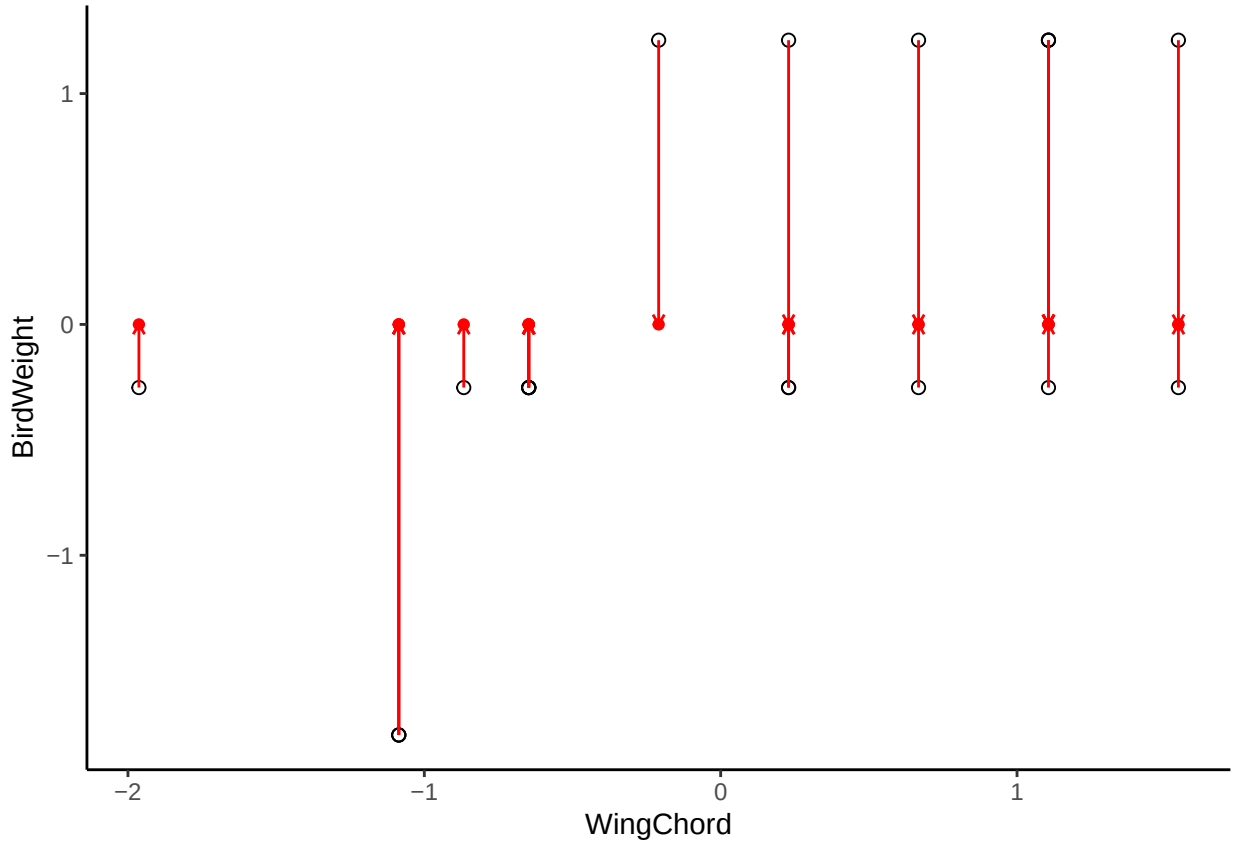


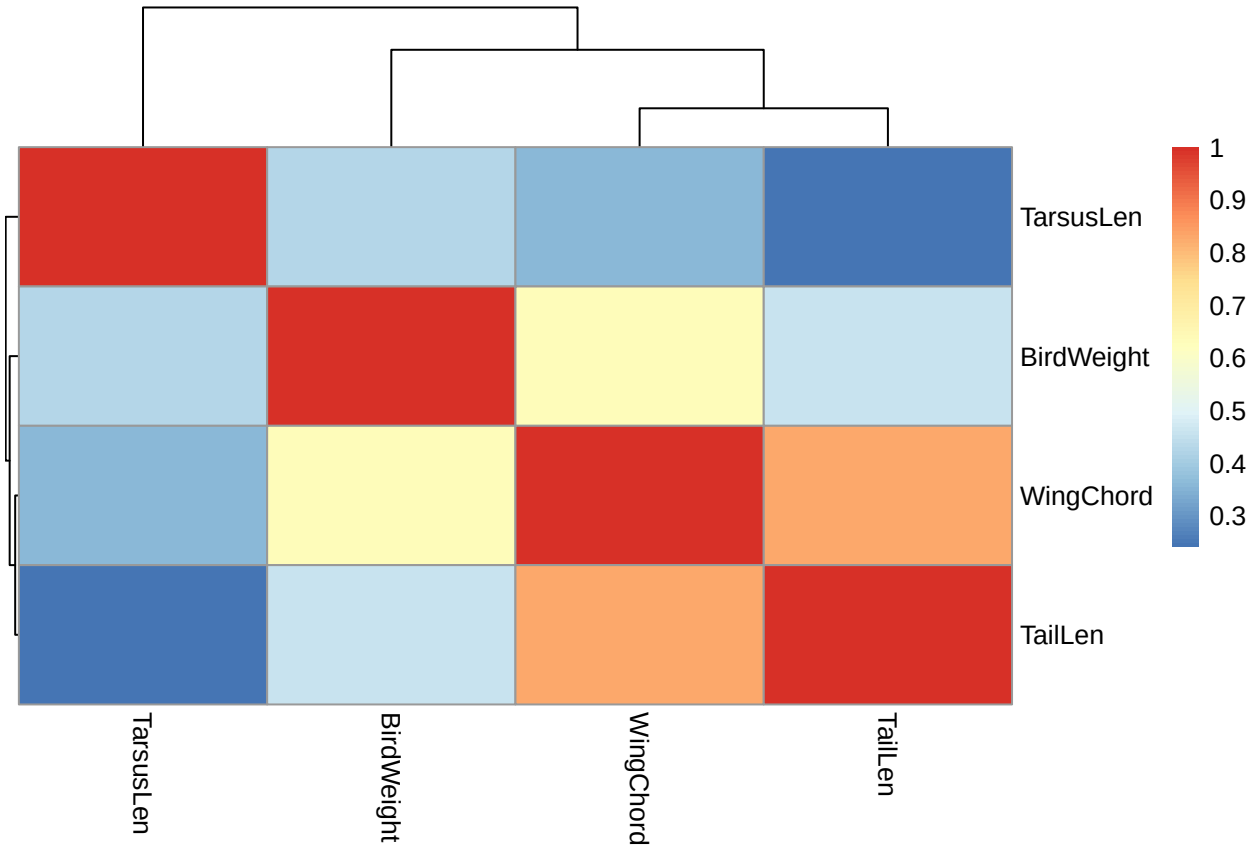
In this study, we used analysis of covariance to investigate whether the linear association in the model differs for male and female birds.

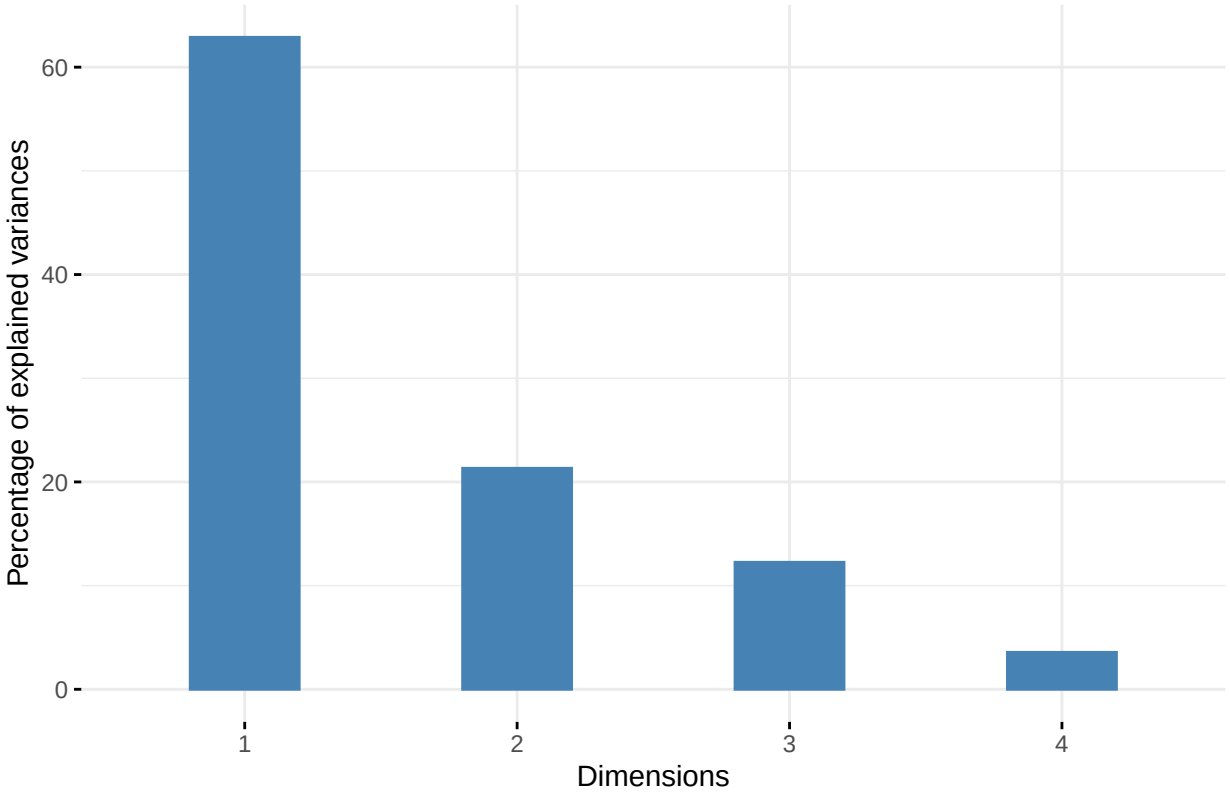
For the model focusing on Wing Chord and Bird Sex, the p-value for the effect of Bird Sex was 0.771, suggesting no significant difference between sexes. Additionally, an interaction term between Wing Chord and Bird Sex was added to the model, resulting in an F-statistic of 0.0179 and an associated p-value of 0.895. Both of these p-values are above the 0.05 significance level, indicating no significant interaction. Similarly, in the model that incorporated Tail Length and BirdSex, the p-value for the Bird Sex effect was 0.5282. When the interaction term between Tail Length and Bird Sex was considered, it yielded an F-statistic of 0.7468 and a p-value of 0.399, both of which also exceed the 0.05 significance level. In both cases, the Analysis of Variance tables suggested that including the interaction term did not significantly improve the fit of the models, as evidenced by their respective p-values of 0.895 and 0.3989.

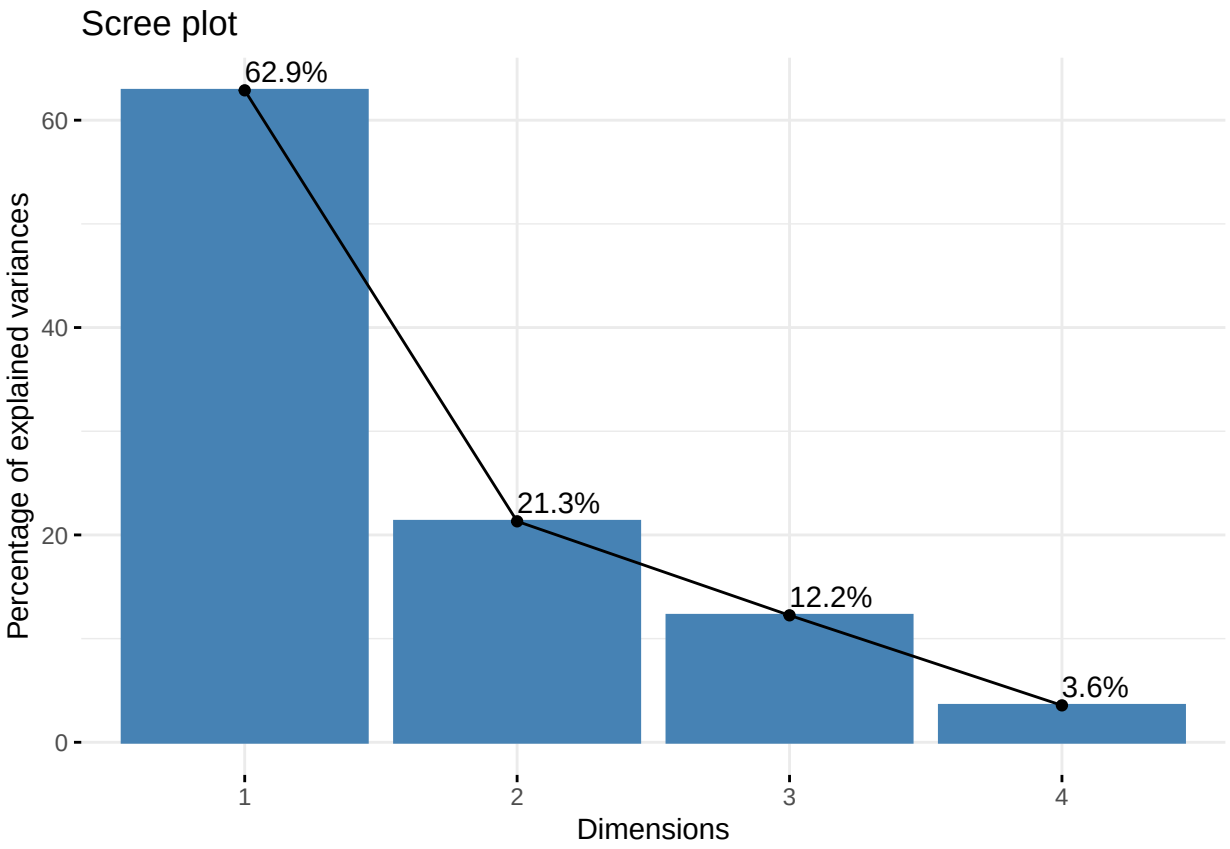
Given these findings, there is insufficient evidence to reject the null hypothesis, which is that there is no difference in the linear associations between pathogen richness and either Wing Chord or Tail Length, for male and female birds. Therefore, we conclude that the linear associations do not significantly differ between sexes in this dataset.

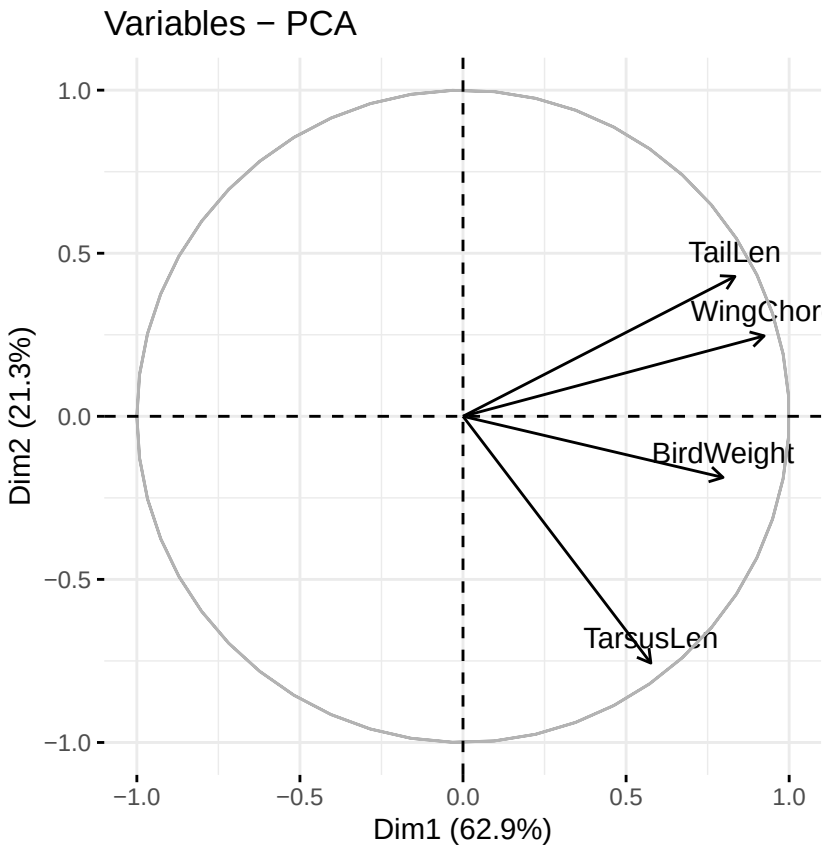
16. Principal Component Analysis on Morphological Features

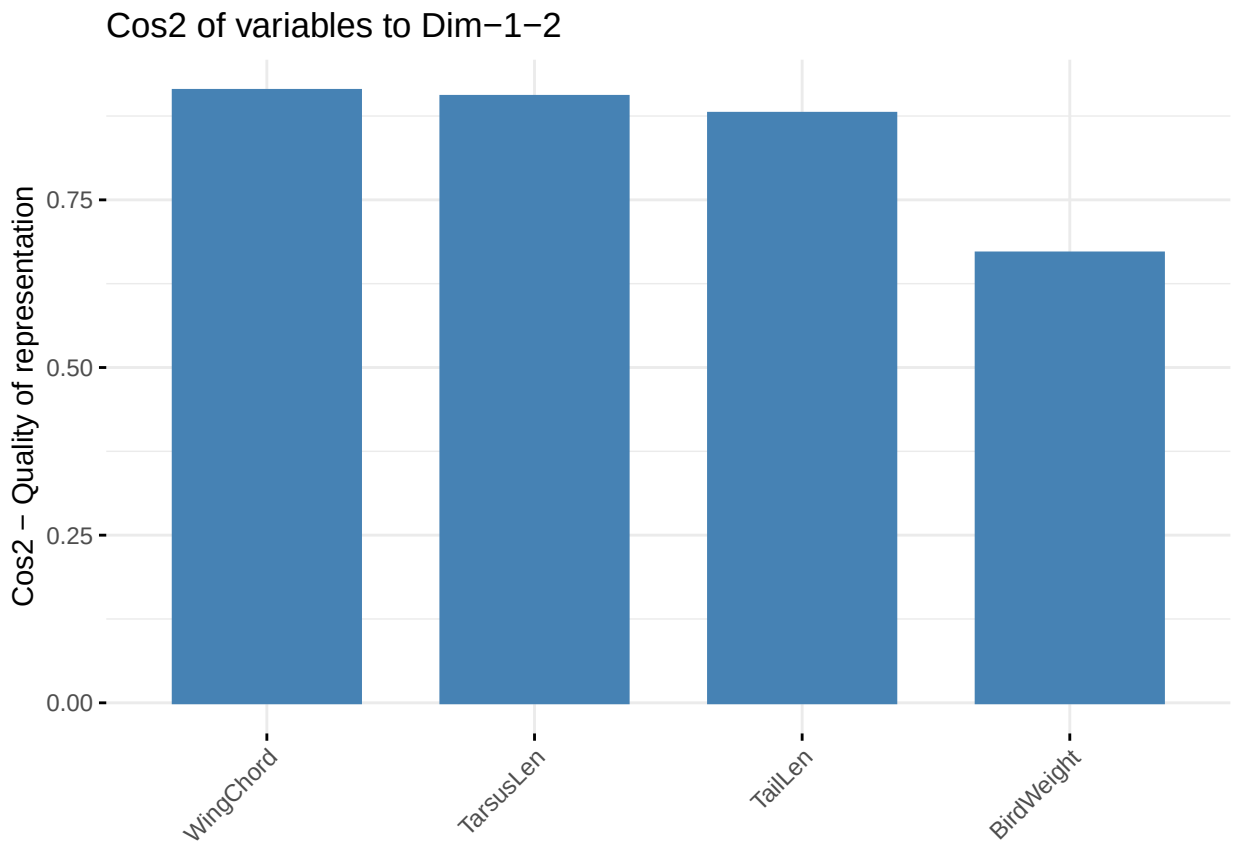


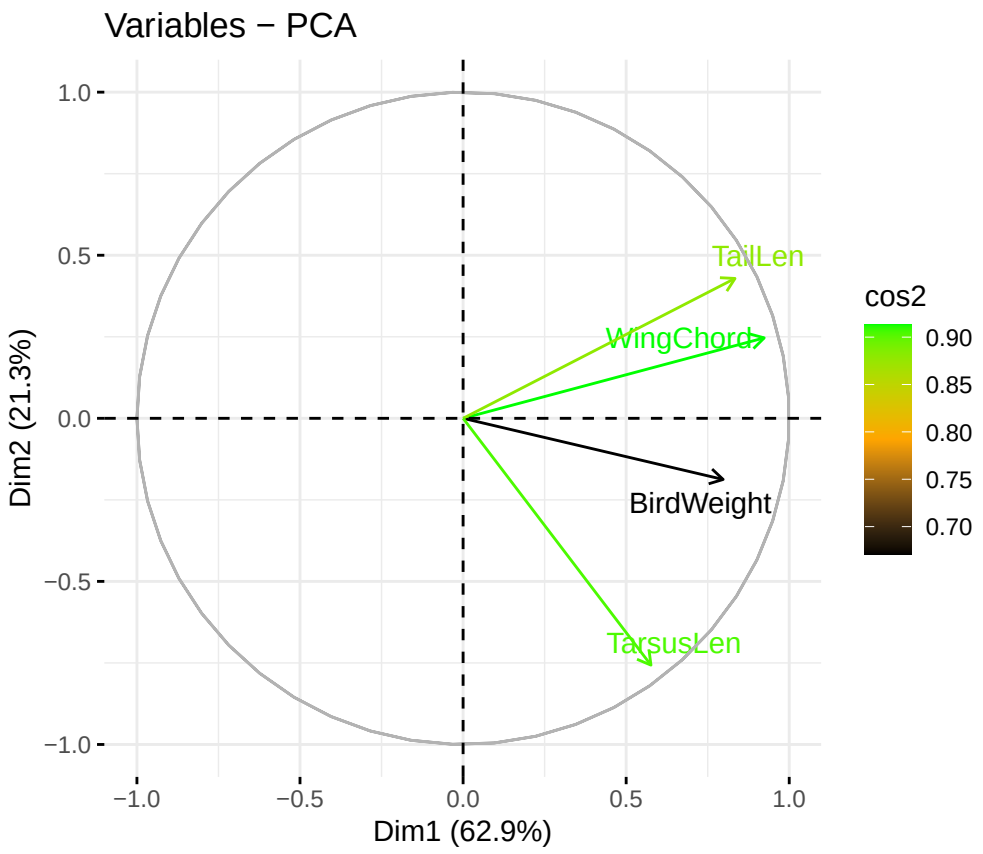


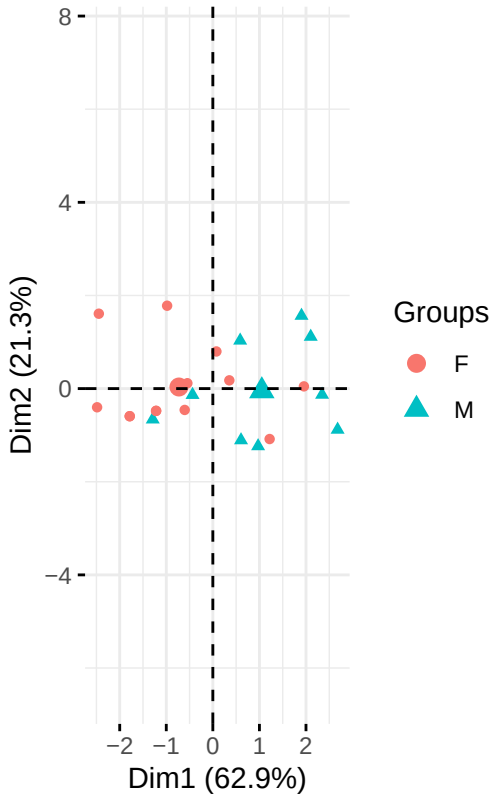

Percentage of explained variances
60
40
20
0
1
2
3
4
Dimensions

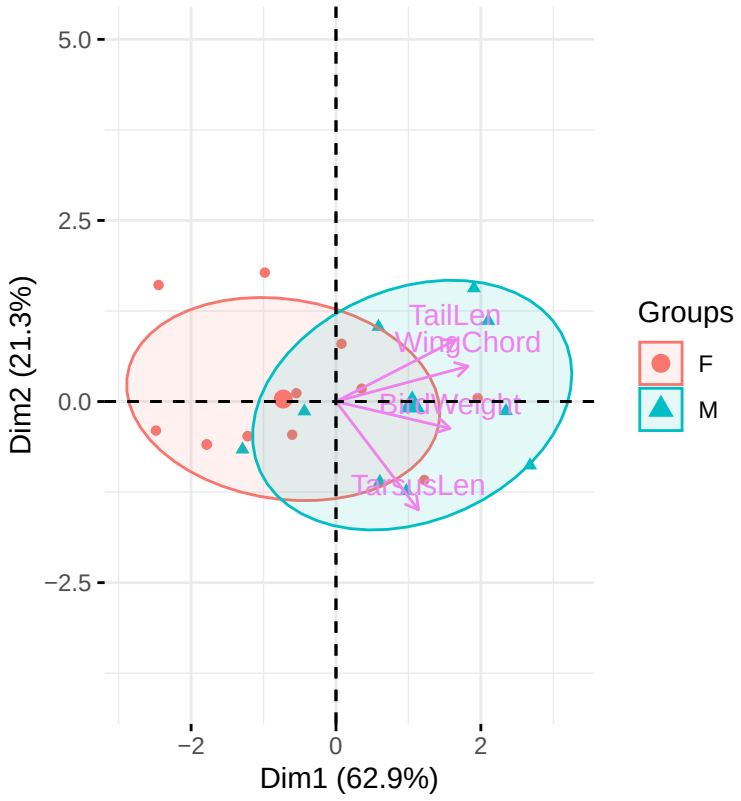

Scree plot
62.9%
21.3%
12.2%
3.6%
Percentage of explained variances
60
40
20
0
1
2
3
4
Dimensions


Variables − PCA
1.0
0.5
0.0
−0.5
−1.0
Dim2 (21.3%)
TailLen
WingChor
BirdWeight
TarsusLen
−1.0
−0.5
0.0
0.5
1.0
Dim1 (62.9%)


Cos2 of variables to Dim−1−2
Cos2 − Quality of representation
0.75
0.50
0.25
0.00
WingChord
TarsusLen
TailLen
BirdWeight


Variables − PCA
1.0
0.5
0.0
−0.5
−1.0
Dim2 (21.3%)
−1.0
−0.5
0.0
0.5
1.0
Dim1 (62.9%)
TailLen
WingChord
BirdWeight
TarsusLen
cos2
0.90
0.85
0.80
0.75
0.70


8
4
0
−4
Dim2 (21.3%)
−2
−1
0
1
2
Dim1 (62.9%)
Groups
F
M

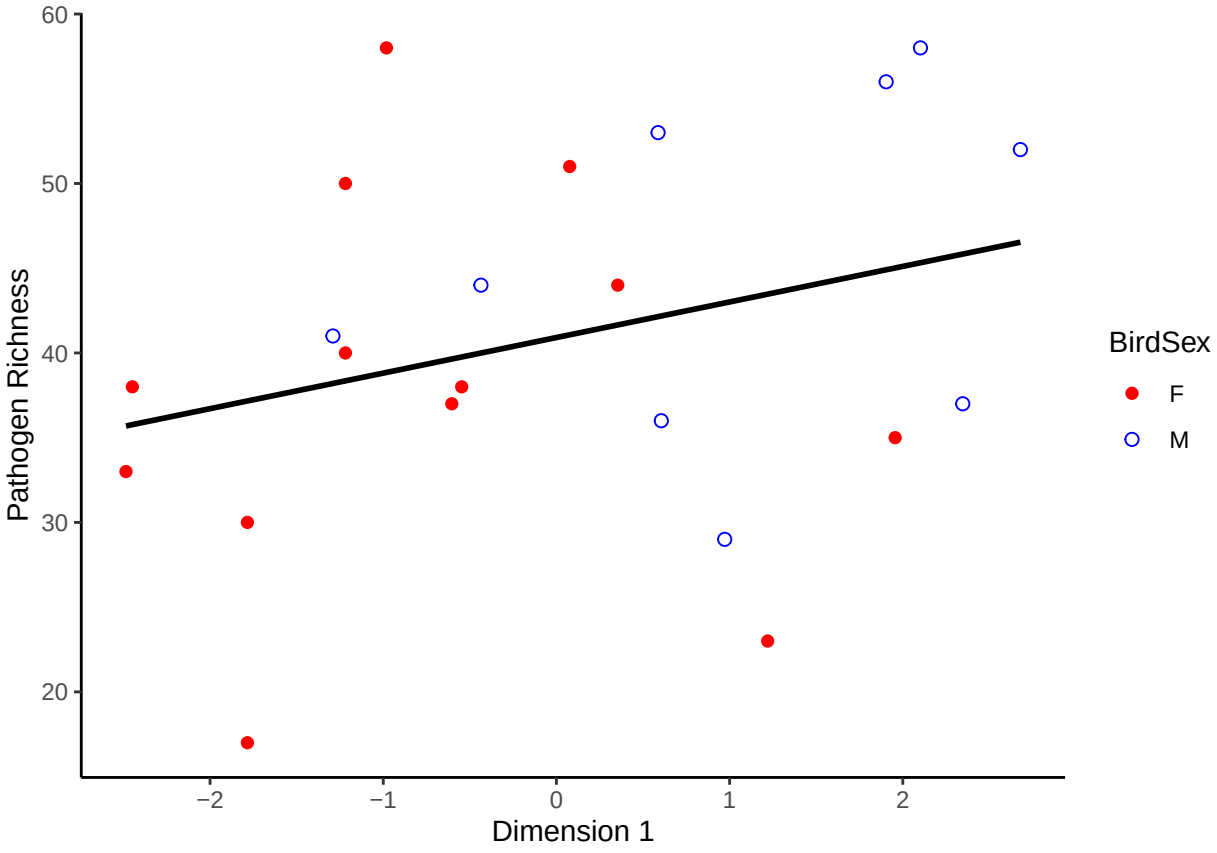


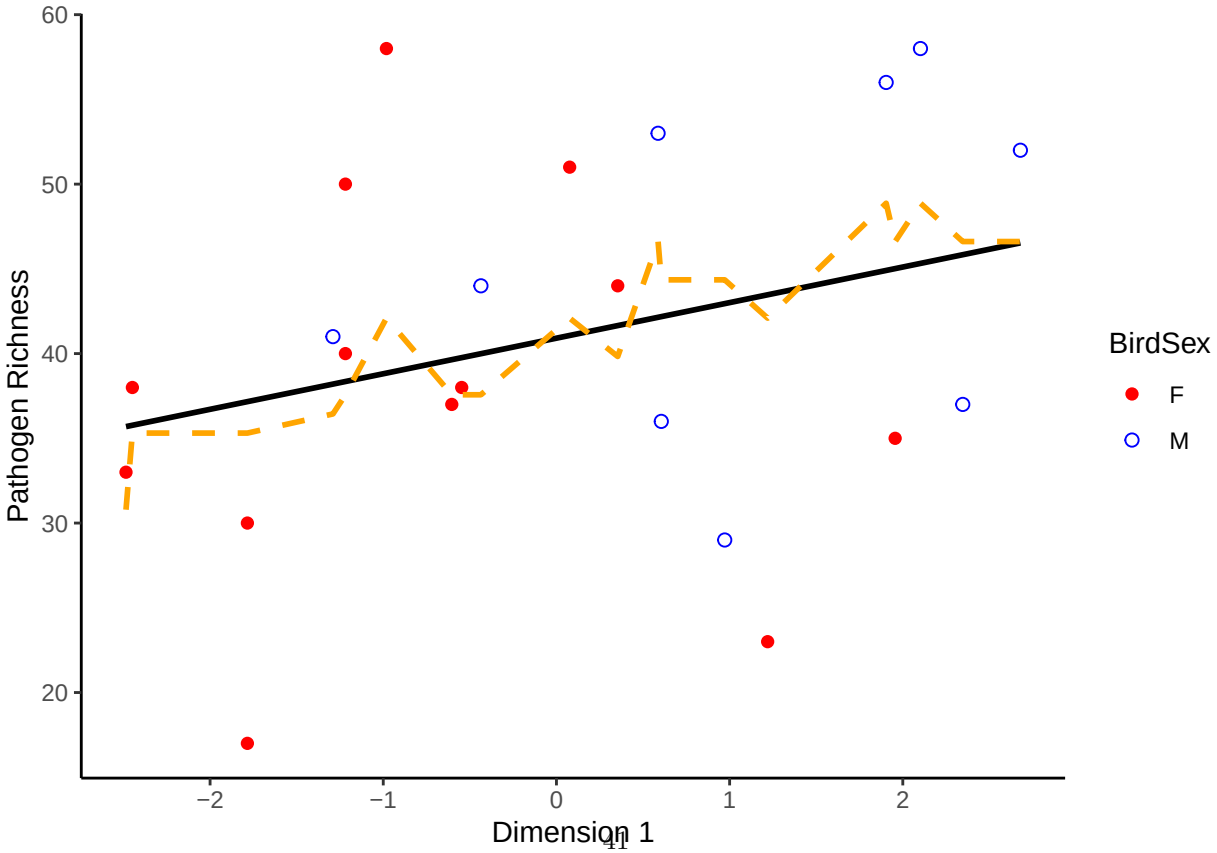
In this study, Principal Component Analysis (PCA) was applied to four variables: Wing Chord, Bird Weight, Tail Len, and Tarsus Length. The first principal component (PC1) accounted for approximately 62.88% of the total variance, with a standard deviation of 1.586. The second principal component (PC2) accounted for 21.31% of the variance, having a standard deviation of 0.923. Together, these two components captured 84.19% of the total variance in the data.

The biplot revealed that the female group was generally lower on both PC1 and PC2 as compared to the male group. The variables Wing Chord, Tail Length, and Bird Weight are positively correlated to each other.

In summary, the PCA indicates that Wing Chord, Bird Weight, and Tail Length are the major contributors to the first principal component and are positively correlated with each other. Tarsus Length, however, exhibits different behavior, especially in its significant contribution to the second principal component.

17. Simple Linear Regression Model for Pathogen Richness with the First Principal Component as a Predictor





In this study, we fit a a simple linear regression model with Pathogen Richness as a response variable and predictor variable equal to the first principal component, and compared it to the simple linear regression model with Wing Chord as a predictor.

In the first model, where Pathogen Richness is predicted by PC1, the coefficient for PC1 is 2.099, with a standard error of 1.499, resulting in a t-value of 1.40 and a p-value of 0.177. This indicates that, at the 5% significance level, PC1 is not a significant predictor of Pathogen Richness, as the p-value is greater than 0.05. The model explains only 8.92% of the variance in Pathogen Richness ($R^2 = 0.08924$), which is not a large amount, and the F-statistic of 1.96 with a p-value of 0.1769 further confirms that the model is not a good fit.

In the second and third models, where Pathogen Richness is predicted by Wing Chord, the coefficient for Wing Chord is 2.2607, with a standard error of 0.9686, resulting in a t-value of 2.334 and a p-value of 0.0301. This indicates that Wing Chord is a significant predictor of Pathogen Richness at the 5% significance level. These models explain 21.41% of the variance in Pathogen Richness ($R^2 = 0.2141$), which is higher than the PC1 model but still not very large. The F-statistic of 5.448 with a p-value of 0.03014 further confirms that the model is a better fit than the PC1 model but still only explains a small proportion of the variance in Pathogen Richness.

In conclusion, Wing Chord is a significant predictor of Pathogen Richness, whereas the first principal component (PC1) is not. However, both models explain a relatively small amount of variance in Pathogen Richness, indicating that other factors not included in these models are likely important in predicting pathogen richness.

18. Multiple Linear Regression Model for Pathogen Richness with the First Two Principal Components as Predictors

In this study, we fit a multiple linear regression model with Pathogen Richness as a response variable and predictor variables equal to the first two principal components, and compared the model with the multiple linear regression model with the morphological features as predictor variables.

In the study of the model with PC1 and PC2, PC1 has a coefficient of 2.099 and a p-value of 0.0781, indicating marginal significance. On the other hand, PC2 has a coefficient of 7.840 and a p-value of 0.000684, which is highly significant. The overall model explains 51.11% of the variance in Pathogen Richness ($R^2 = 0.5111$) and is statistically significant, with an F-statistic of 9.933 and a p-value of 0.001115. This suggests that both the first and the second principal components contribute to explaining pathogen richness in the data, with PC2 having a stronger effect than PC1.

In the study of the Multiple Linear Regression Model with Original Variables, Tarsus Length is highly significant with a p-value of 0.00369, while the other variables are not significant. The model explains 54.76% of the variance in Pathogen Richness ($R^2 = 0.5476$) and is statistically significant, with an F-statistic of 5.144 and a p-value of 0.006673.

Both models are statistically significant, but the model with the original variables (Wing Chord, Bird Weight, Tail Length, Tarsus Length) has a slightly higher R^2 (54.76%) compared to the model with PC1 and PC2 (51.11%).

In summary, using the first two principal components provides a lower dimensional, but nearly equally effective way of explaining variance in Pathogen Richness compared to using the original four variables. However, then the model with the original variables provides detail of biological features.

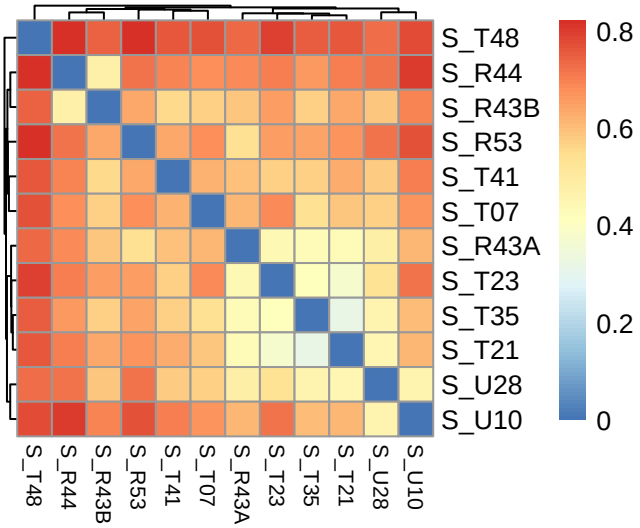
18.2. Simple Linear Regression Model for Pathogen Richness with the Second Principal Component as a Predictor

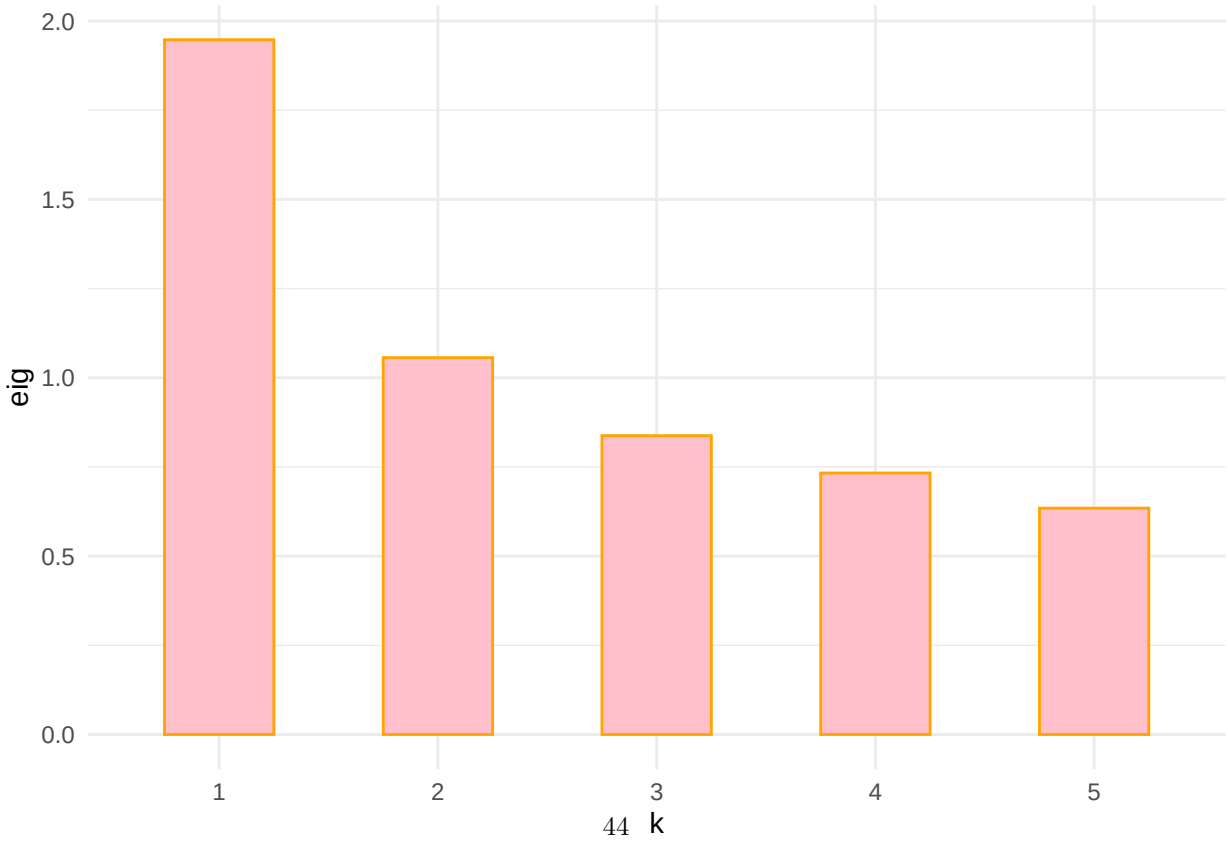
As the first principal component was not found to be statistically significant in prediction with linear regression, let us test the second principal component:

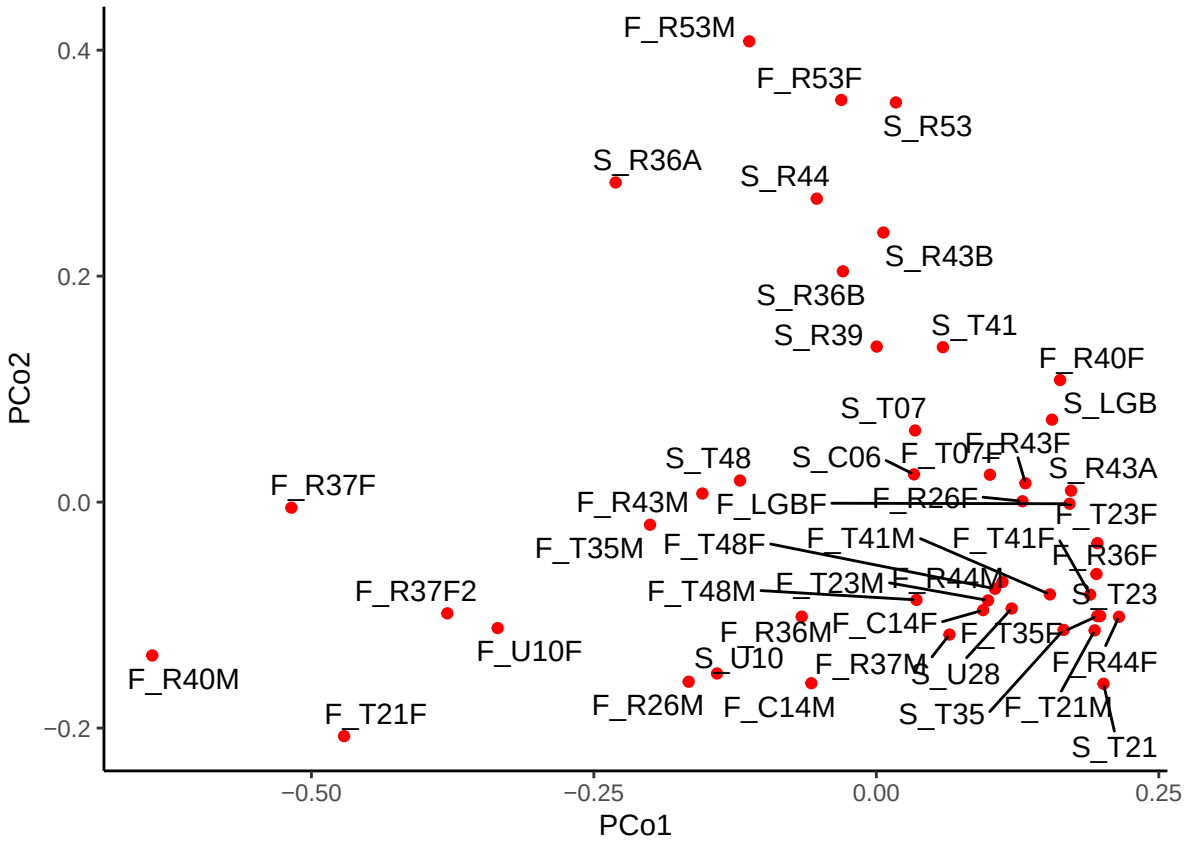
The F-statistic is 14.6 with a p-value of 0.00107, below the alpha level of 0.05. We reject the null hypothesis and conclude from this data that this model is statistically significant in explaining the variance in the response variable, pathogen richness. The predictor variable, the second principal component, has a t-value of 3.82, below the alpha level of 0.05. We conclude that the second principal component is statistically significant in predicting pathogen richness. The multiple R-squared value is 0.4219, and this suggests that approximately 42.19% of the variability in pathogen richness is explained by the second principal component.

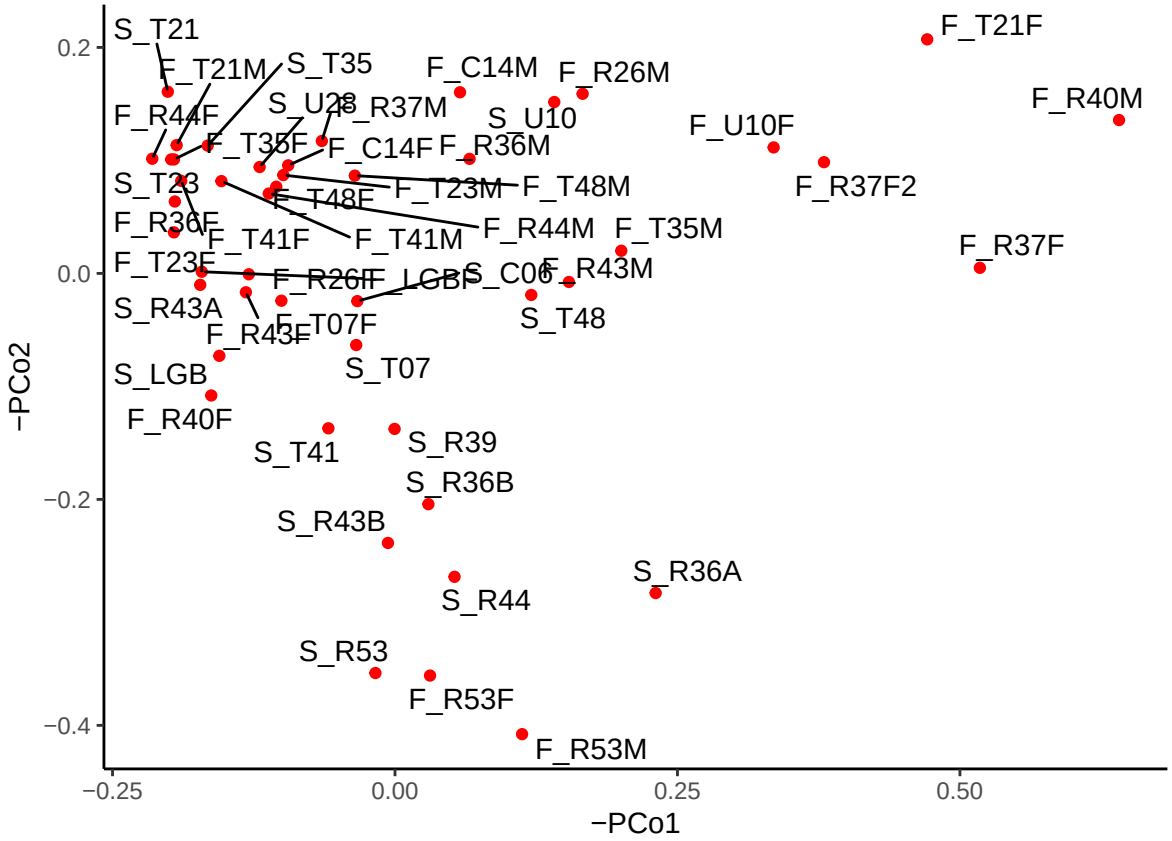
While the first principal component captured more of the variance in the predictor variables, this evidence demonstrates that this does not necessarily relate to how well the component predicts the response variable. While counter-intuitive at first glace - the second principal component in this case was superior to predicting pathogen richness.

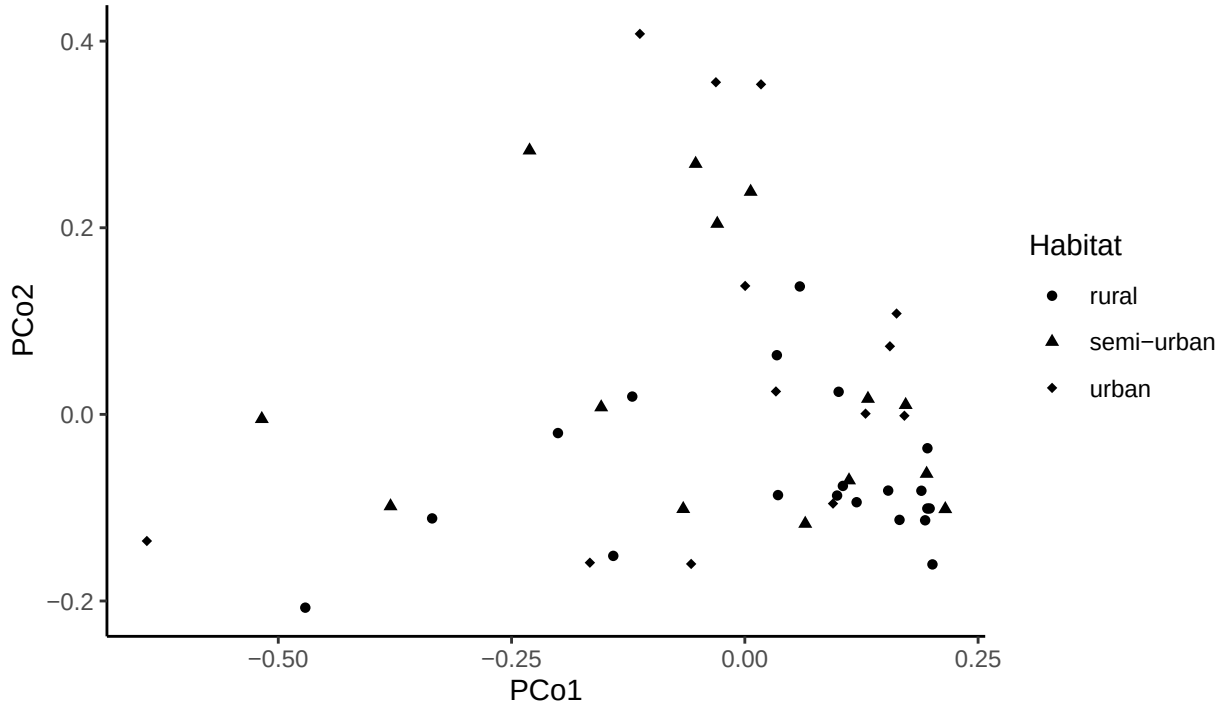
19. Metric Multidimensional Scaling (MDS) to Assess the Composition of Microbial Communities Relating to Habitat or Source

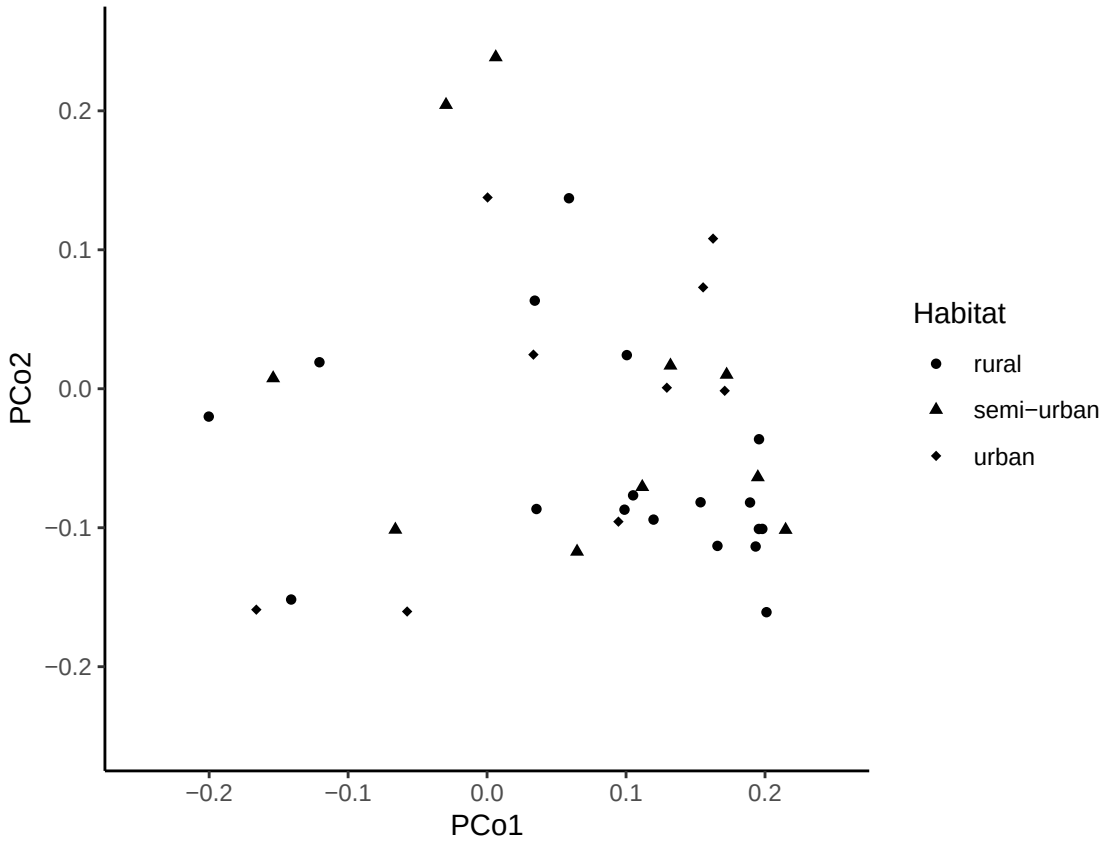


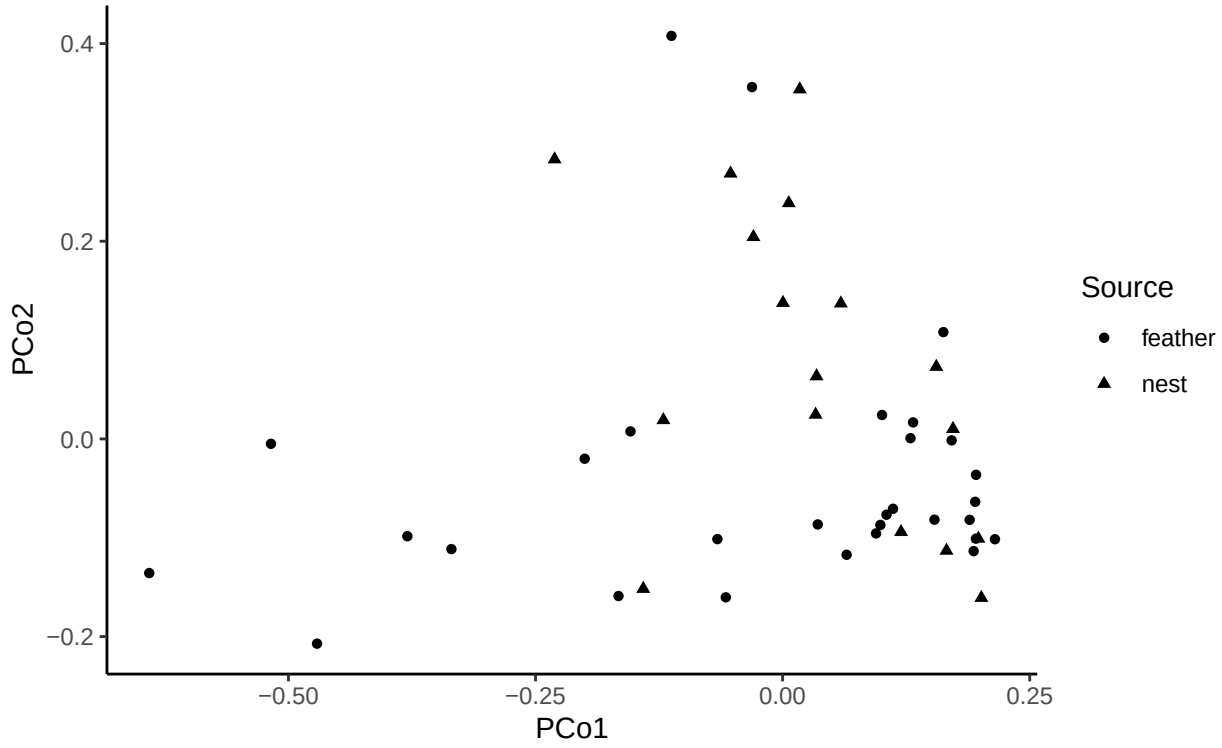


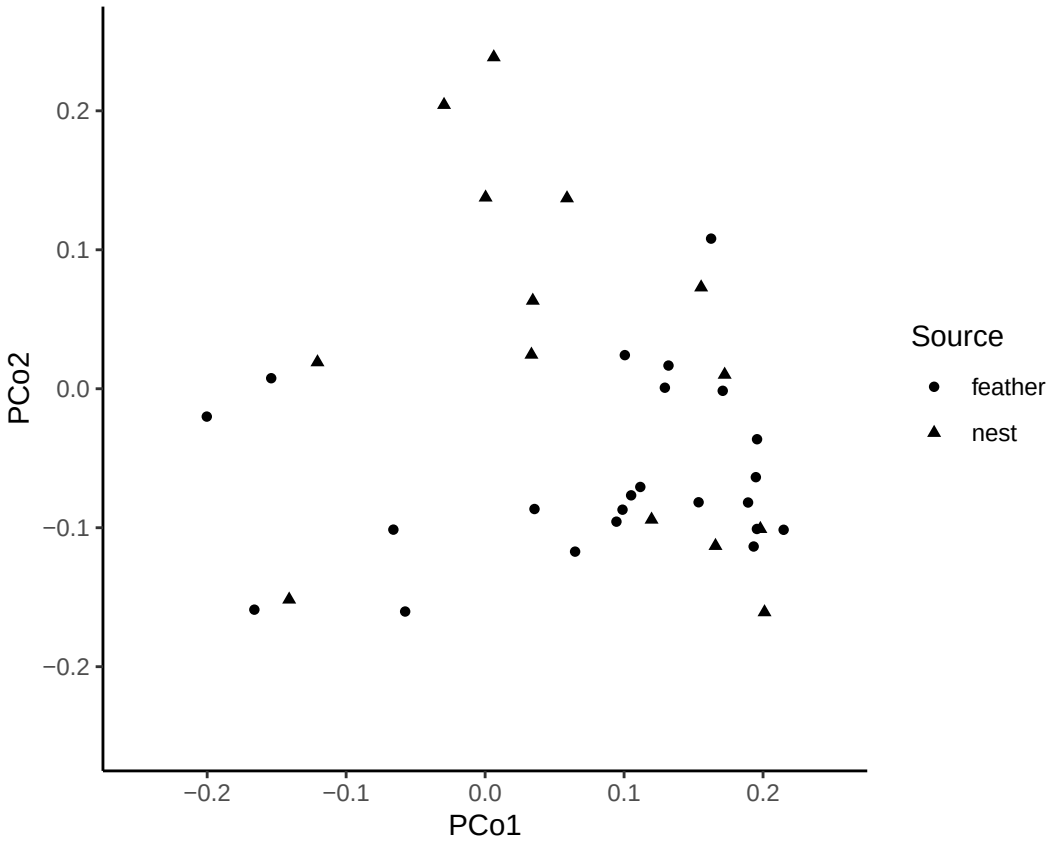

F_R53M
F_R53F
S_R53
S_R36A
S_R44
S_R43B
S_R36B
S_R39
S_T41
F_R40F
S_LGB
S_T07
F_R43F
S_T48
S_C06
F_T07F
S_R43A
F_R37F
F_R43M
F_LGBF
F_R26F
F_T23F
F_T35M
F_T48F
F_T41M
F_T41F
F_R36F
F_R37F2
F_T48M
F_T23M
F_R44M
S_T23
F_R36M
F_C14F
F_T35F
F_U10F
S_U10
F_R37M
S_U28
F_R44F
F_R40M
F_R26M
F_C14M
S_T35
F_T21M
F_T21F
S_T21
0.4
0.2
0.0
−0.2
−0.50
−0.25
0.00
0.25
PCo1
PCo2


S_T21
F_T21M
S_T35
F_C14M
F_R26M
F_T21F
F_R40M
F_R44F
S_U28
F_R37M
S_U10
F_U10F
F_T35F
F_C14F
F_R36M
F_R37F2
S_T23
F_T23M
F_T48M
F_T48F
F_R36F
F_T41F
F_T41M
F_R44M
F_T35M
F_R37F
F_T23F
F_R26F
F_LGBF
S_C06
F_R43M
S_R43A
F_T07F
S_T48
F_R43F
S_T07
S_LGB
F_R40F
S_T41
S_R39
S_R36B
S_R43B
S_R36A
S_R44
S_R53
F_R53F
F_R53M
0.2
0.0
−0.2
−0.4
−0.25
0.00
0.25
0.50
−PCo2
−PCo1


0.4
0.2
0.0
−0.2
PCo2
−0.50
−0.25
0.00
0.25
PCo1
Habitat
rural
semi−urban
urban

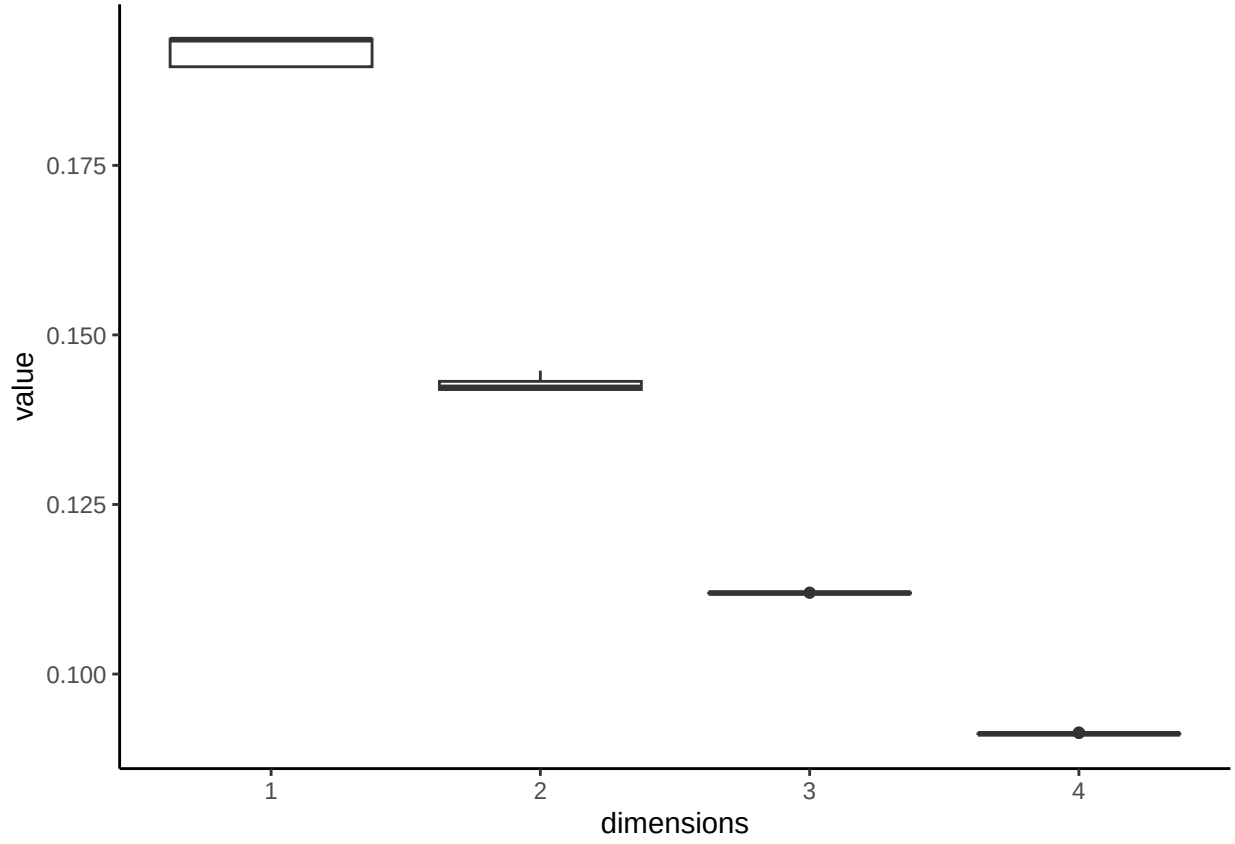

0.2
0.1
0.0
−0.1
−0.2
PCo2
−0.2
−0.1
0.0
0.1
0.2
PCo1
Habitat
rural
semi−urban
urban

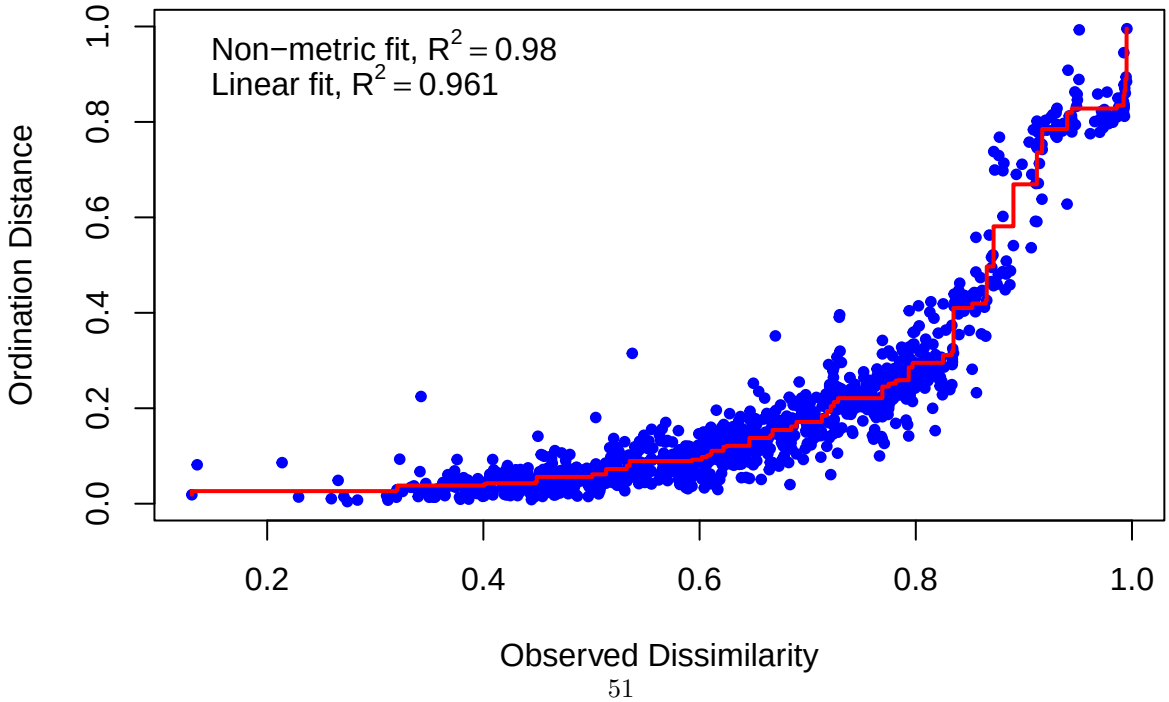

0.4
0.2
0.0
−0.2
PCo2
−0.50
−0.25
0.00
0.25
PCo1
Source
feather
nest

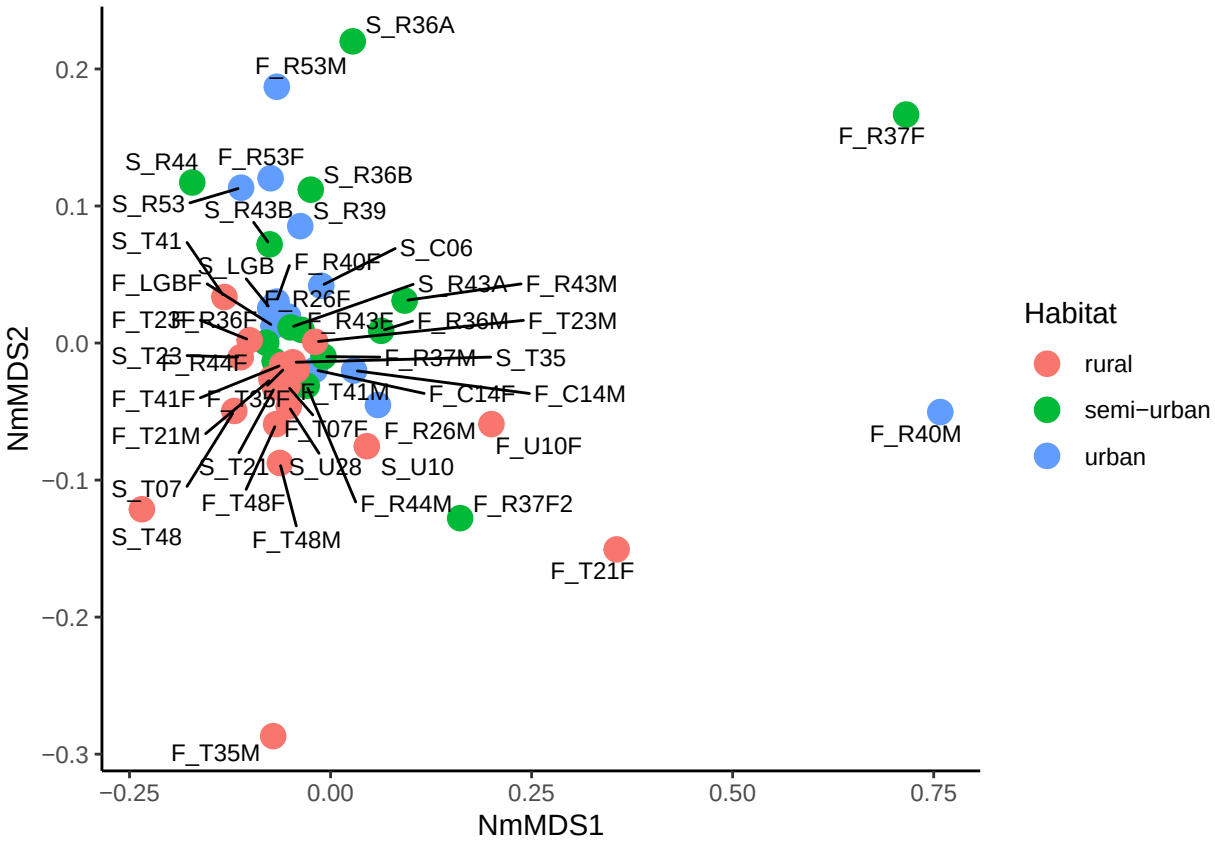


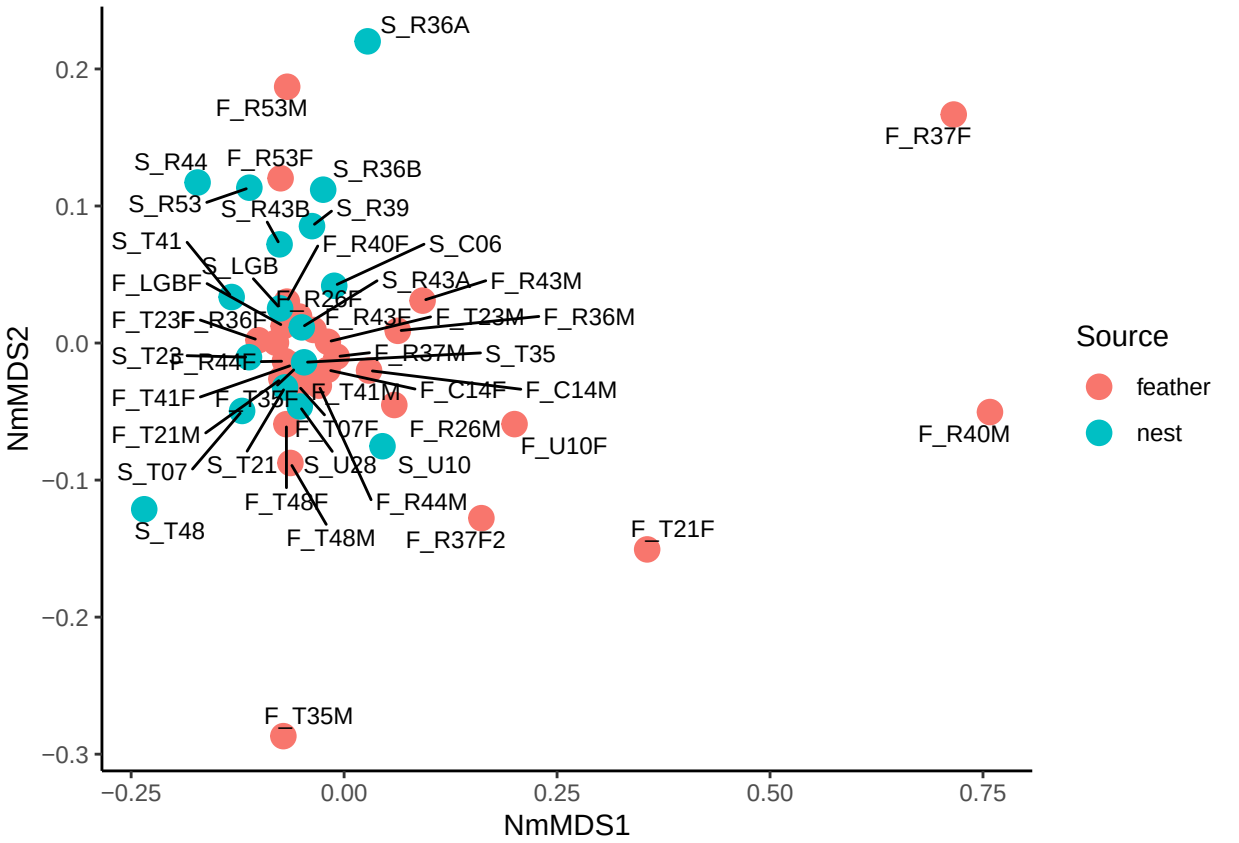
In this study, we applied metric multidimensional scaling (MDS) to the dissimilarities dataset to assess if the composition of the microbial communities appear to be related more to Habitat or Source. Upon visual inspection of the MDS, microbial community composition appears to be more closely associated with Source, as the different categories of Source are more closely clustered.

20. Nonmetric Multidimensional Scaling (NMDS) to Assess the Composition of Microbial Communities Relating to Habitat or Source






S_R36A
F_R53M
F_R37F
S_R44
F_R53F
S_R36B
S_R53
S_R43B
S_R39
S_T41
S_C06
S_LGB
F_R40F
F_LGBF
S_R43A
F_R43M
F_R26F
F_T23F
F_R36F
F_R43F
F_R36M
F_T23M
S_T23
F_R44F
F_R37M
S_T35
F_T41F
F_T35F
F_T41M
F_C14F
F_C14M
F_T21M
F_T07F
F_R26M
F_U10F
F_R40M
S_T21
S_U28
S_U10
S_T07
F_T48F
F_R44M
F_R37F2
S_T48
F_T48M
F_T21F
F_T35M
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.25
0.00
0.25
0.50
0.75
NmMDS1
NmMDS2
Habitat
rural
semi−urban
urban



In this study, we applied non-metric multidimensional scaling (NMDS) to the dissimilarities dataset to assess if the composition of the microbial communities appear to be related more to Habitat or Source.

The Shepard plot suggests that both the linear and non-metric fits are excellent fits, and indicates that the lower-dimensional representation of the data is capturing almost all of the dissimilarity structure in the original high-dimensional data. The higher R^2 value in the non-metric Fit suggests that the NMDS solution is a slightly better model, strictly considering R^2.

From these scatter plots, the composition of the microbial communities appear to be related more to Source, rather than Habitat, as the different categories of Source are more closely clustered.

References

1. www.tru.ca, Thompson Rivers University. "BIOL 4001: Biostatistics." *Thompson Rivers University*, http://www.tru.ca/distance/courses/biol4001.html. Accessed 20 Aug. 2023.

2. *Introduction to R.* https://www.zoology.ubc.ca/~bio501/R/workshops/workshops-intro.html. Accessed 20 Aug. 2023.

3. *Resources for The Analysis of Biological Data.* https://whitlockschluter3e.zoology.ubc.ca/index.html. Accessed 20 Aug. 2023.