

Statistical Analysis Report

PennCHOP Microbiome Program

September 21, 2018

Contents

Sample size	2
Heatmap based on bacteria taxonomic assignments	3
Alpha diversity	4
Beta diversity	6
Ordination based on Bray-Curtis distance	6
Ordination based on Jaccard distance	7
Test of differential abundance	8
Analyses on imaging data	11
Bray-Curtis (microbiome) and Manhattan (Nmf)	11
Bray-Curtis (microbiome) and Euclidean (Nmf)	12
Jaccard (microbiome) and Manhattan (Nmf)	13
Jaccard (microbiome) and Euclidean (Nmf)	14

Sample size

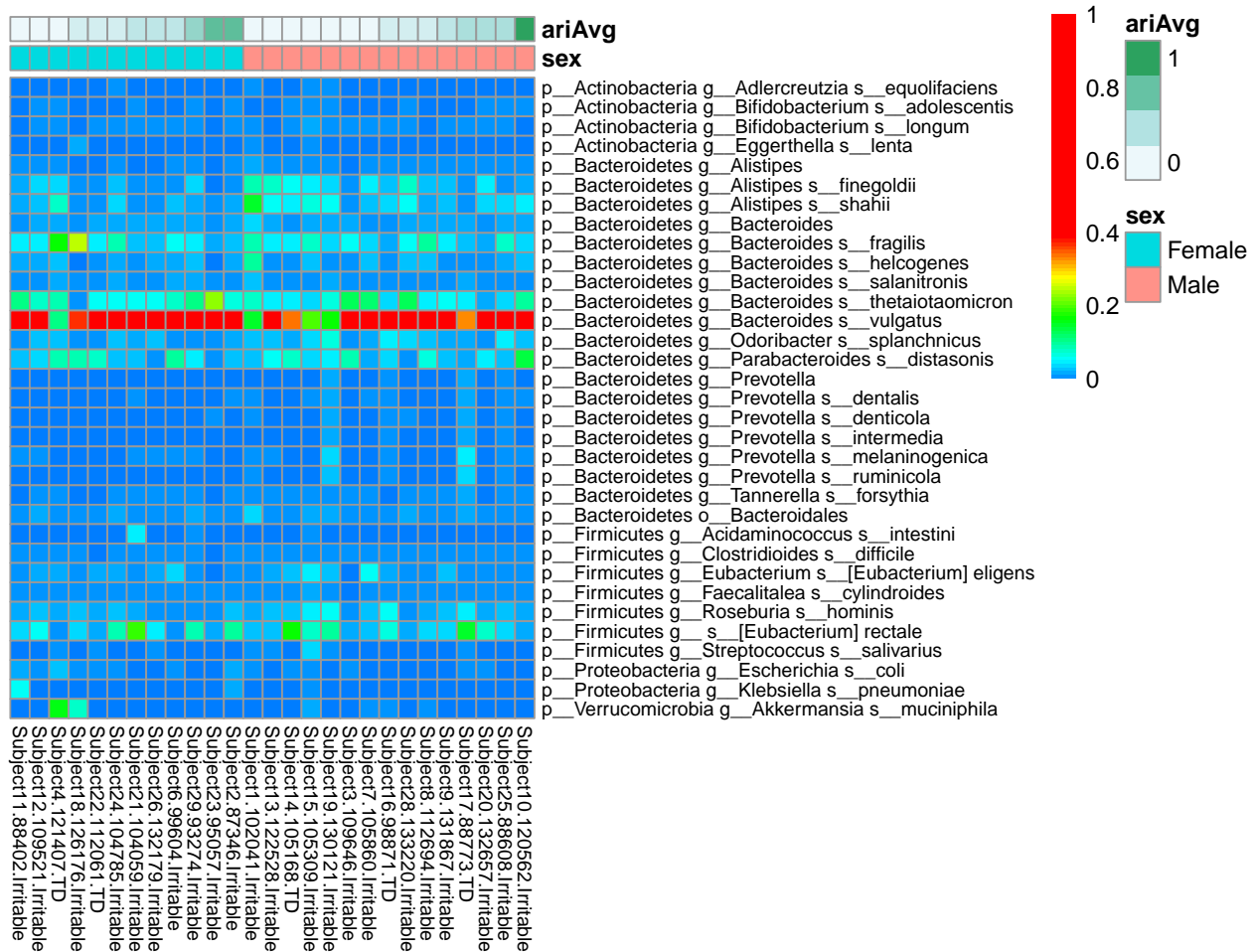
Table 1: Study group vs. sex

	Female	Male	Sum
Irritable	10	12	22
TD	2	3	5
Sum	12	15	27

Heatmap based on bacteria taxonomic assignments

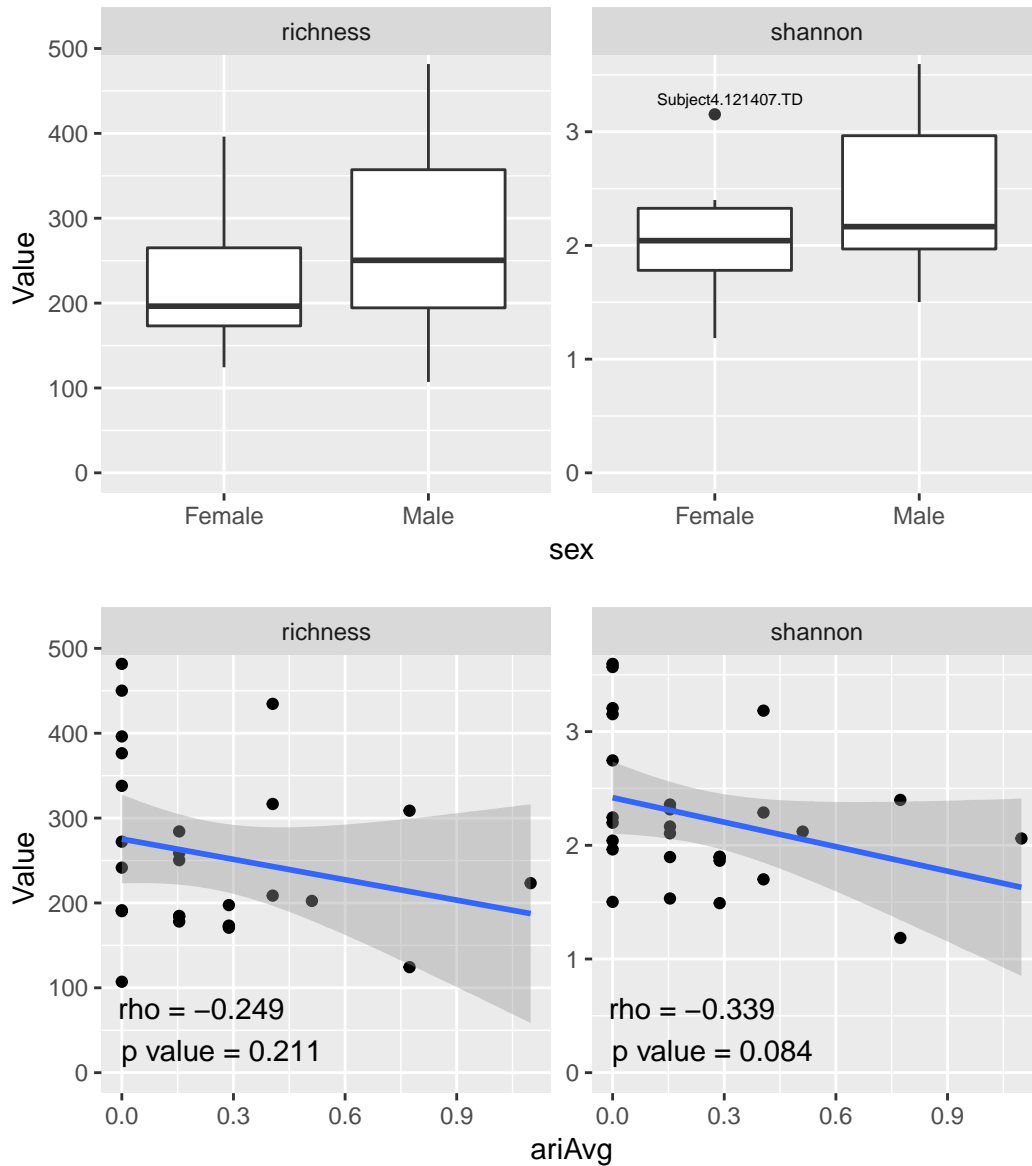
Each column of the heatmap represents one sample and each row represents one taxon. Taxa were included in the chart if the abundance in any sample exceeded 1%.

The chart is colored white if taxa were not observed in the sample, dark blue if taxa were observed at very low abundance. This allows the reader to quickly survey presence/absence. Abundance values exceeding 40% are colored red, indicating an extremely dominant species.



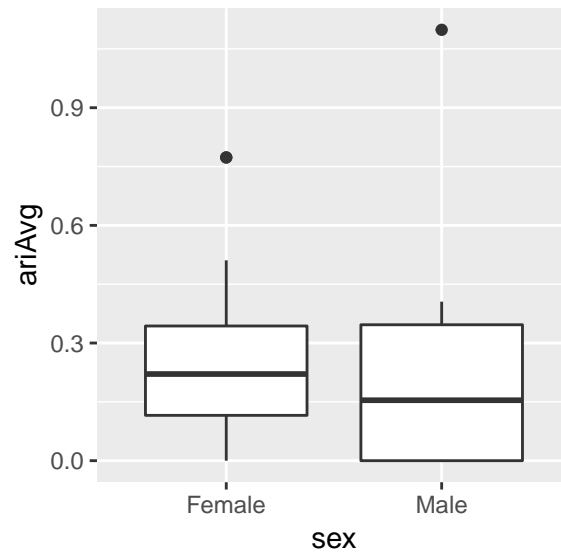
Alpha diversity

Alpha diversity was assessed by richness (expected number of observed species out of rarefying sample size of 10000) and Shannon index.



To test the association between alpha diversity and explanatory variables, a linear regression model was implemented. Sex/ariAvg were used as covariates.

Alpha diversity	Term	Estimate	p-value
richness	Female → Male	49.82	0.2097
richness	ariAvg	-67.71	0.3364
shannon	Female → Male	0.32	0.177
shannon	ariAvg	-0.64	0.1376

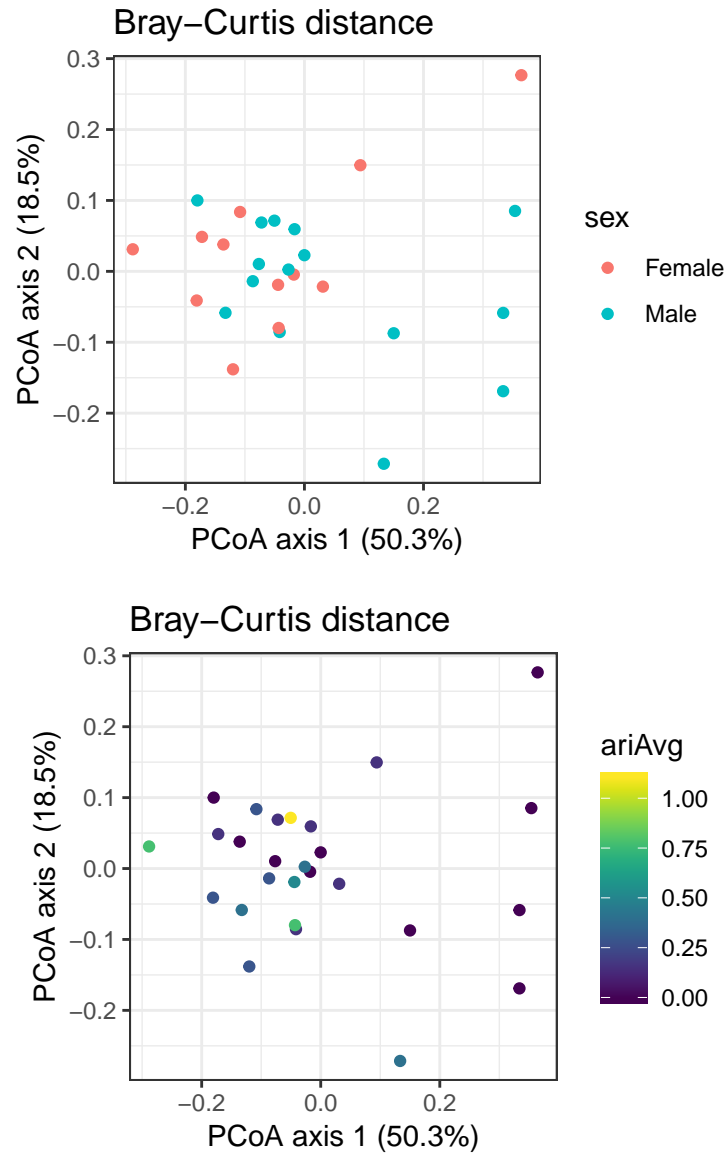


```
##
## Welch Two Sample t-test
##
## data:  s$ariAvg by s$sex
## t = 0.70561, df = 24.428, p-value = 0.4871
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1490703  0.3041687
## sample estimates:
## mean in group Female    mean in group Male
##      0.2818920          0.2043428
```

Beta diversity

Beta diversity (similarity between samples) was assessed by Bray-Curtis distance and Jaccard distance.

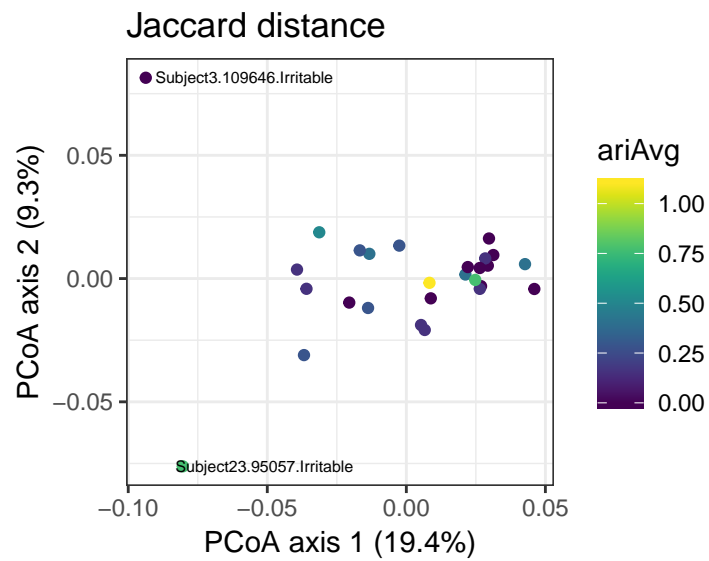
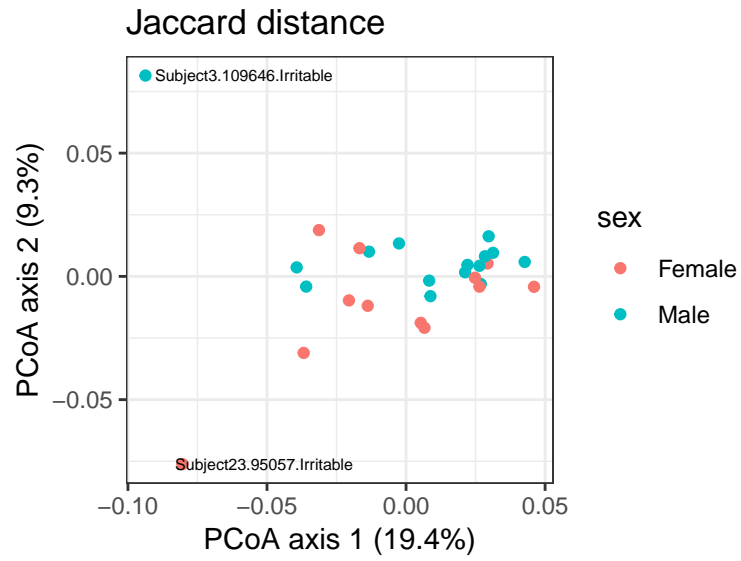
Ordination based on Bray-Curtis distance



The association between the centroid positions and sex/ariAvg was tested using a permutational MANOVA.

Term	Df	R^2	p -value
sex	1	0.064	0.125
ariAvg	1	0.077	0.092
Residuals	24	0.859	NA
Total	26	1	NA

Ordination based on Jaccard distance



The association between the centroid positions and sex/ariAvg was tested using a permutational MANOVA.

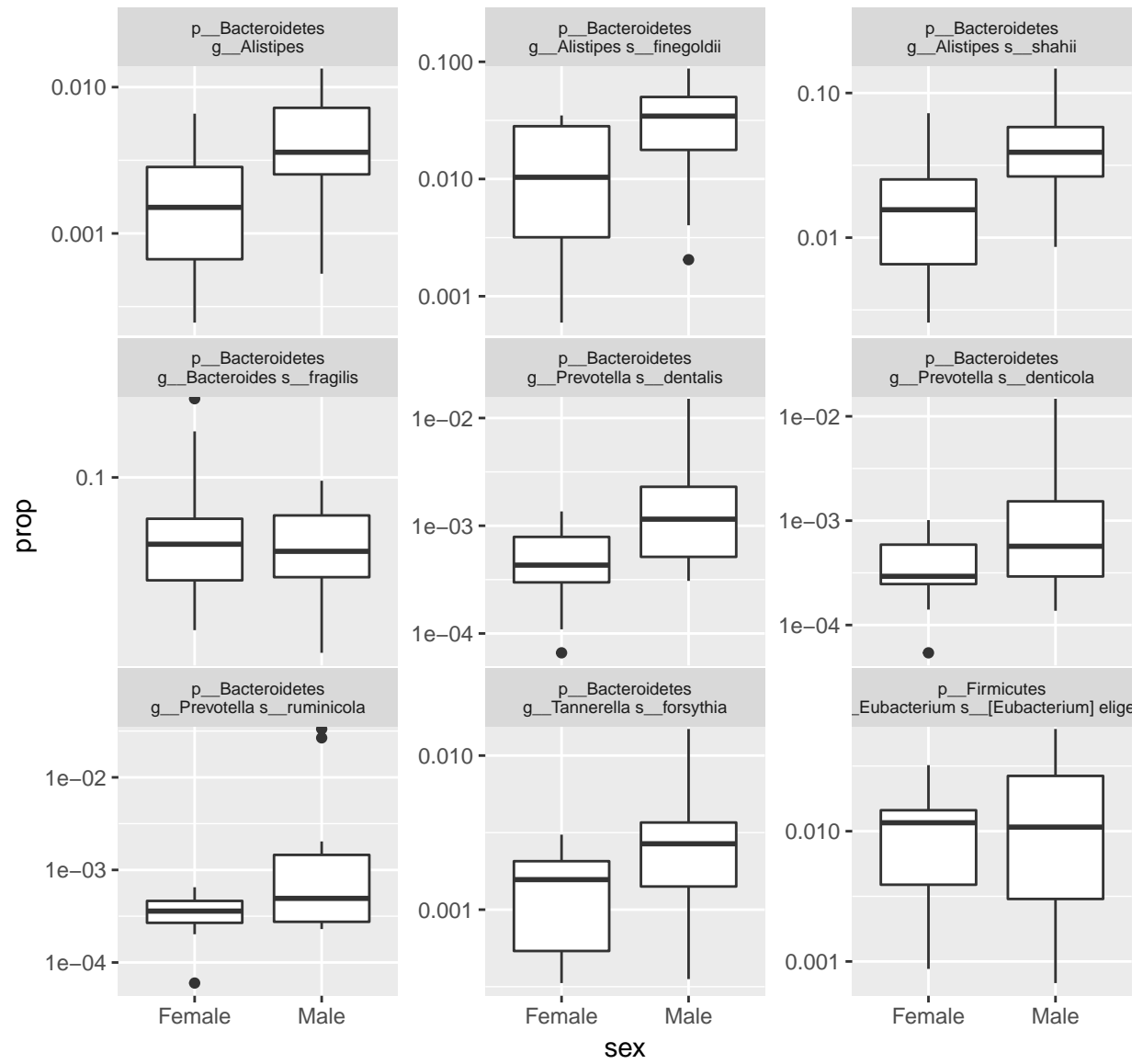
Term	Df	R^2	p -value
sex	1	0.041	0.267
ariAvg	1	0.04	0.324
Residuals	24	0.919	NA
Total	26	1	NA

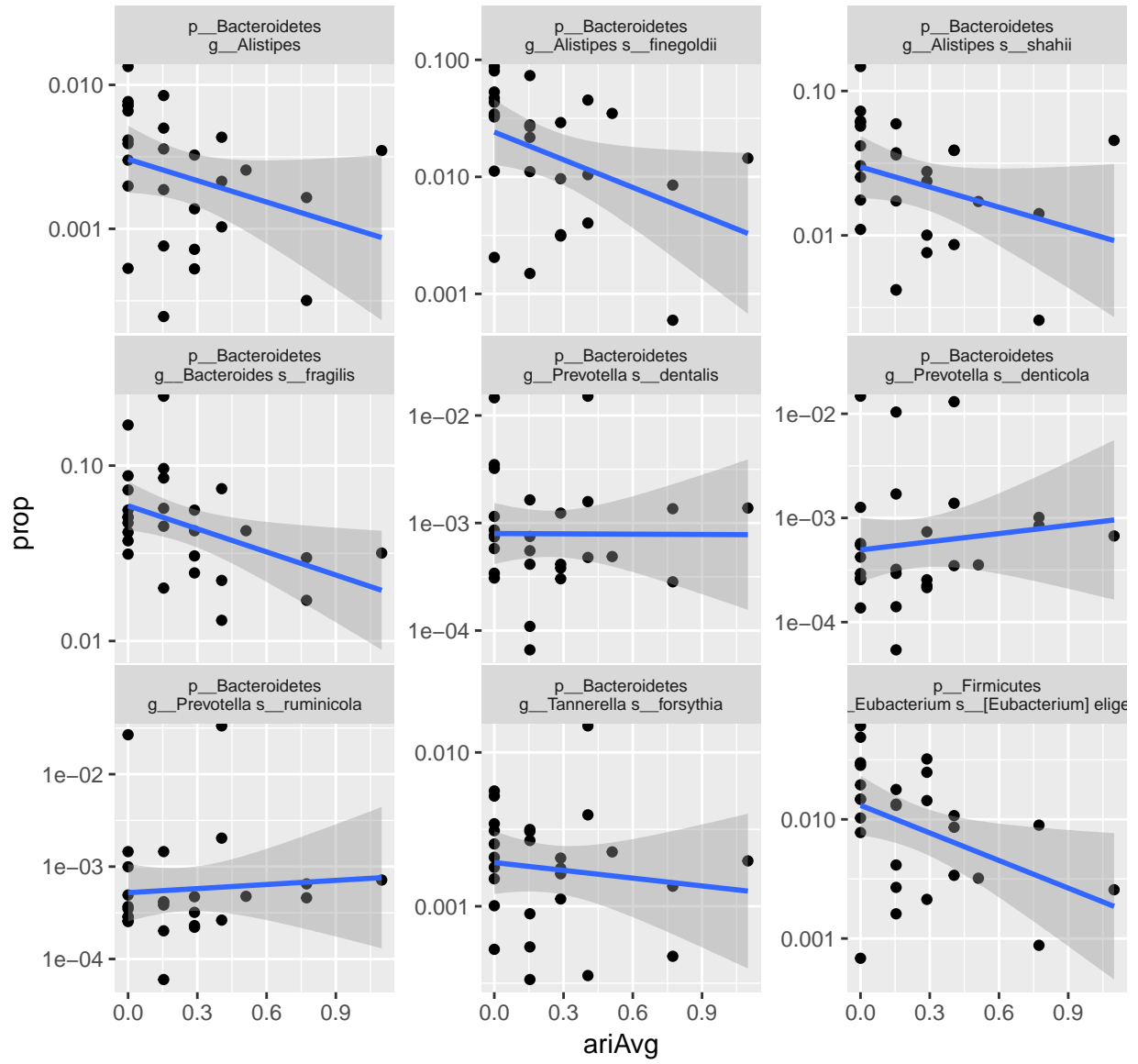
Test of differential abundance

A linear model was implemented to test the association between $\log(\text{proportion})$ of taxa and sex/ariAvg. Taxa were tested if the abundance in any sample exceeded 1%. The following table summarizes the test result: taxa that have a nominal p -value < 0.05 are present.

Taxa	Term	Estimate	p -value	FDR
p__Bacteroidetes g__Alistipes s__shahii	Female \rightarrow Male	0.94	0.008399	0.3035
p__Bacteroidetes g__Prevotella s__dentalis	Female \rightarrow Male	1.16	0.01477	0.3035
p__Bacteroidetes g__Bacteroides s__fragilis	ariAvg	-1.08	0.01722	0.3035
p__Bacteroidetes g__Alistipes	Female \rightarrow Male	0.92	0.01906	0.3035
p__Bacteroidetes g__Alistipes s__finegoldii	Female \rightarrow Male	1.04	0.02696	0.3035
p__Firmicutes g__ Eubacterium s[Eubacterium] eligens	ariAvg	-1.77	0.03307	0.3035
p__Bacteroidetes g__Prevotella s__ruminicola	Female \rightarrow Male	1.12	0.03387	0.3035
p__Bacteroidetes g__Prevotella s__denticola	Female \rightarrow Male	1.1	0.03679	0.3035
p__Bacteroidetes g__Tannerella s__forsythia	Female \rightarrow Male	0.69	0.04916	0.3475

Plots describing the abundance of above 9 taxa with nominal p -value < 0.05 are given below.



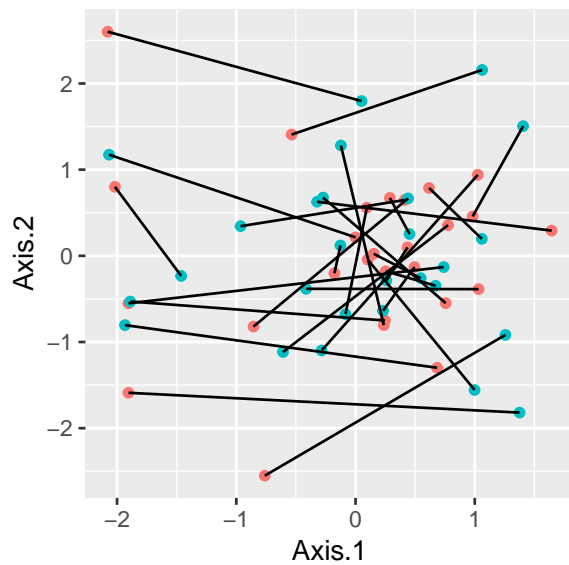


Analyses on imaging data

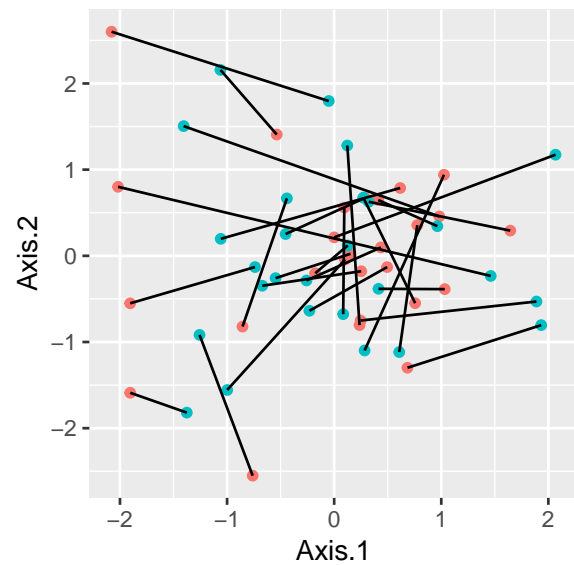
Bray-Curtis (microbiome) and Manhattan (Nmf)

We applied Bray-Curtis distances to microbiome data and Manhattan distances to imaging data. The p -value from Mantel's test for similarity between two distance matrices was obtained. To visualize the comparison between two distance measures, the scaled PCoA plot based on microbiome data (red) is given and the scaled PCoA plot based on imaging data (blue) is overlaid. Additional plots are created after flipping x and y axes of imaging data PCoA. The line segments represent the same sample.

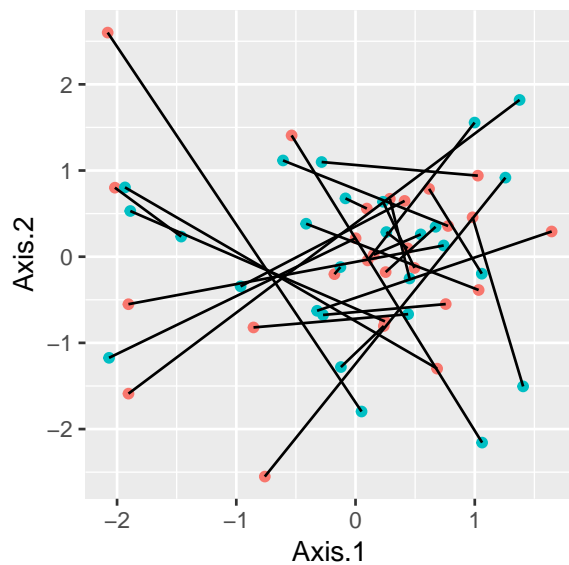
```
## Mantel's permutation test for similarity p-value: 0.243
```



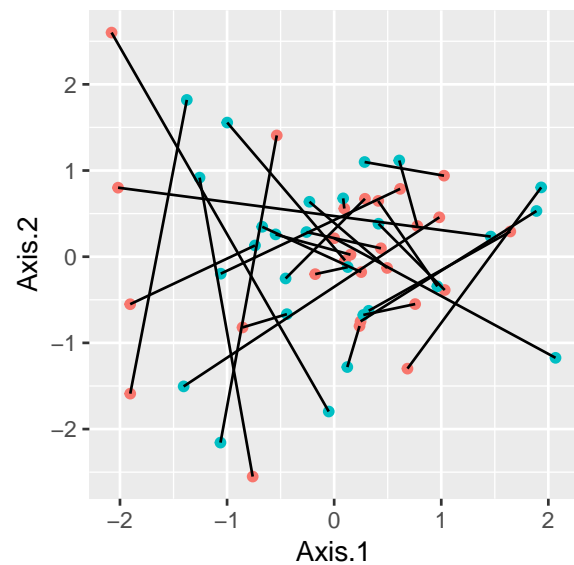
source ● Microbiome ● Nmf



source ● Microbiome ● Nmf_x_ref



source ● Microbiome ● Nmf_y_ref

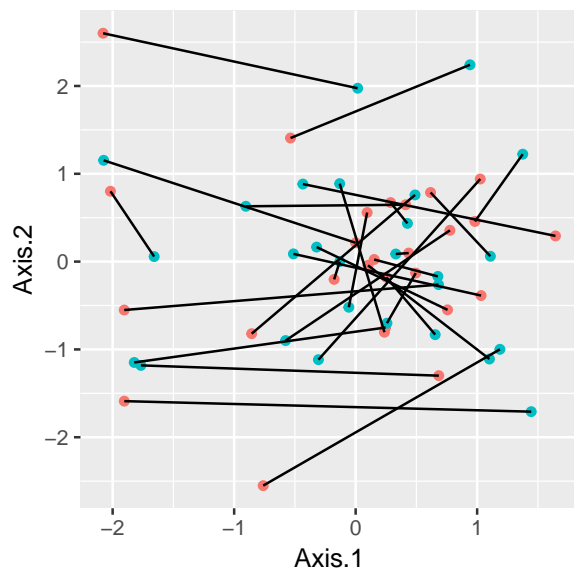


source ● Microbiome ● Nmf_xy_ref

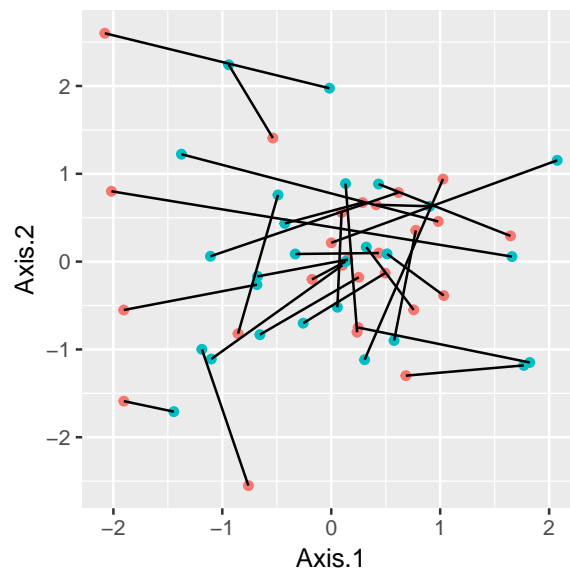
Bray-Curtis (microbiome) and Euclidean (Nmf)

We applied Bray-Curtis distances to microbiome data and Euclidean distances to imaging data. The p -value from Mantel's test for similarity between two distance matrices was obtained. To visualize the comparison between two distance measures, the scaled PCoA plot based on microbiome data (red) is given and the scaled PCoA plot based on imaging data (blue) is overlaid. Additional plots are created after flipping x and y axes of imaging data PCoA. The line segments represent the same sample.

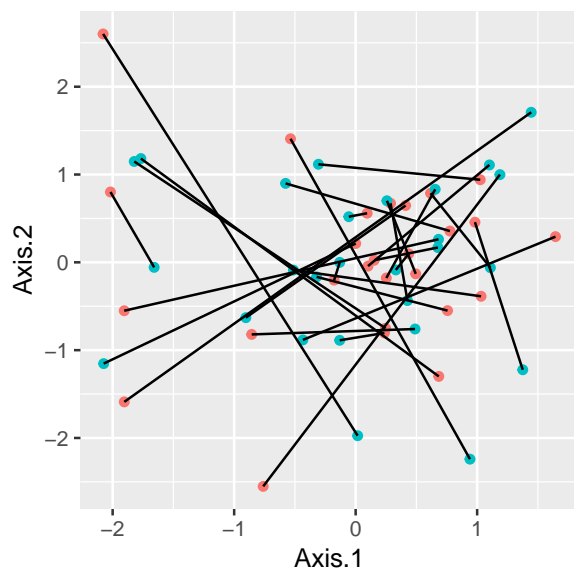
```
## Mantel's permutation test for similarity p-value: 0.145
```



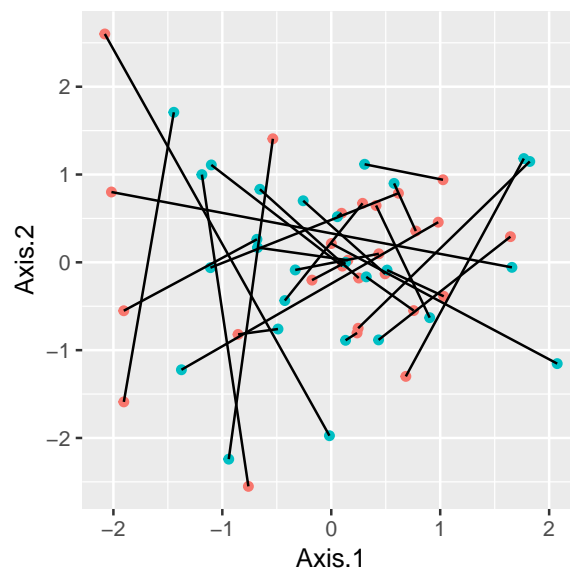
source ● Microbiome ● Nmf



source ● Microbiome ● Nmf_x_ref



source ● Microbiome ● Nmf_y_ref

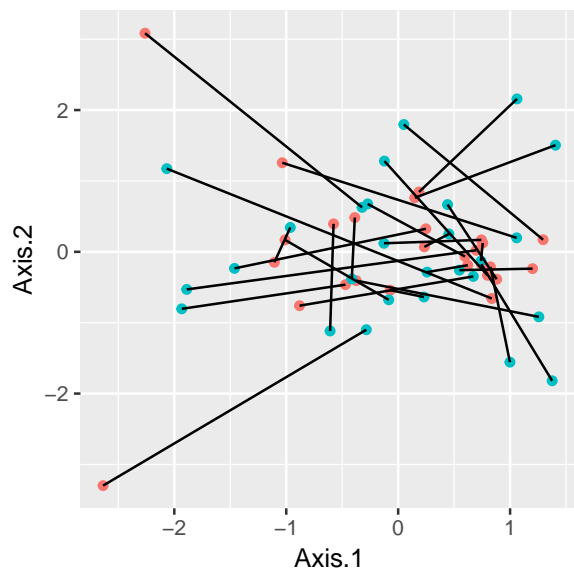


source ● Microbiome ● Nmf_xy_ref

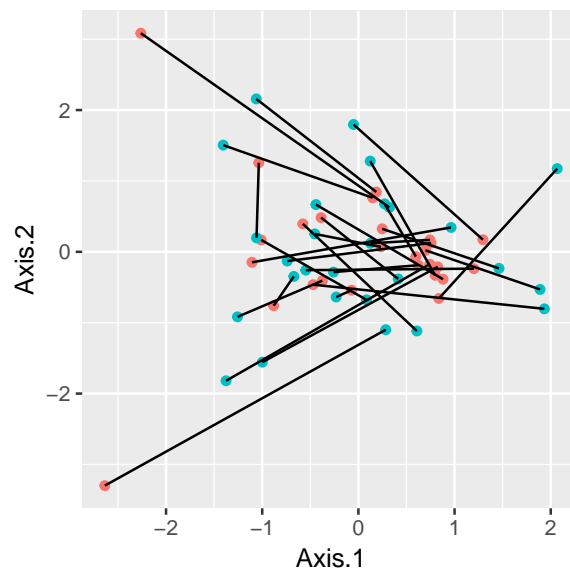
Jaccard (microbiome) and Manhattan (Nmf)

We applied Jaccard distances to microbiome data and Manhattan distances to imaging data. The p -value from Mantel's test for similarity between two distance matrices was obtained. To visualize the comparison between two distance measures, the scaled PCoA plot based on microbiome data (red) is given and the scaled PCoA plot based on imaging data (blue) is overlaid. Additional plots are created after flipping x and y axes of imaging data PCoA. The line segments represent the same sample.

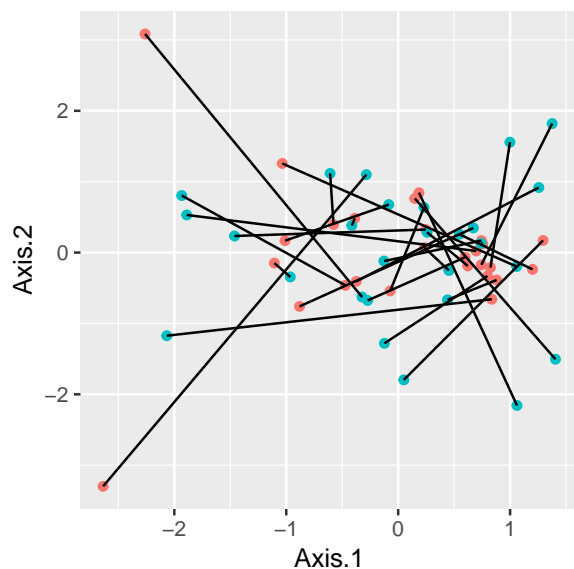
```
## Mantel's permutation test for similarity p-value: 0.291
```



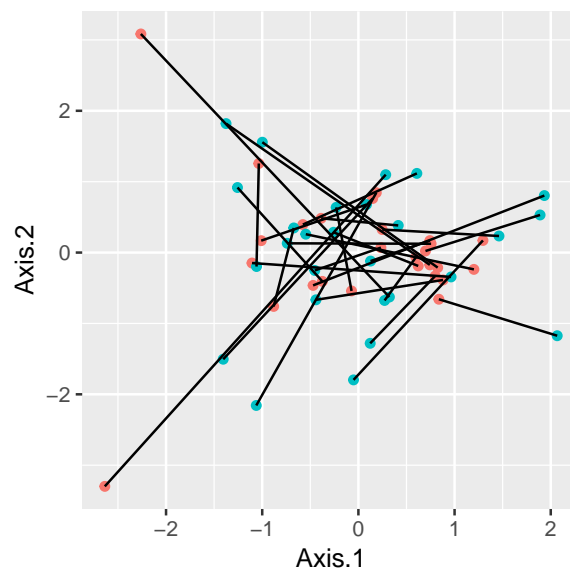
source ● Microbiome ● Nmf



source ● Microbiome ● Nmf_x_ref



source ● Microbiome ● Nmf_y_ref

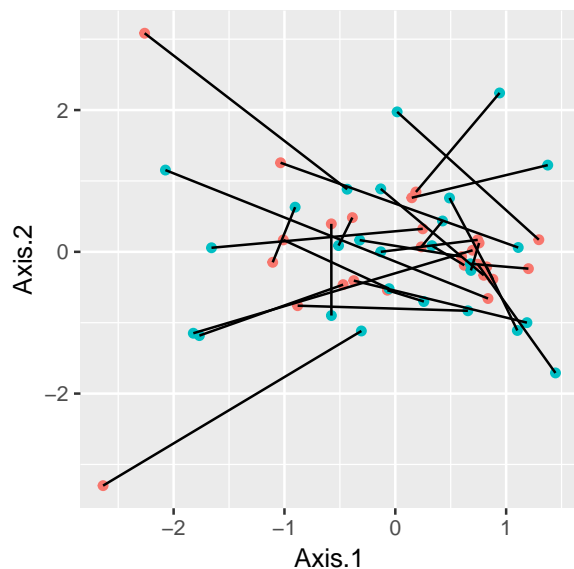


source ● Microbiome ● Nmf_xy_ref

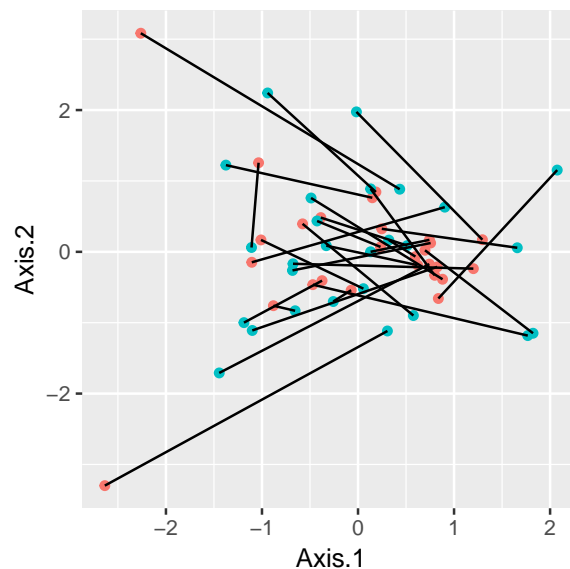
Jaccard (microbiome) and Euclidean (Nmf)

We applied Jaccard distances to microbiome data and Euclidean distances to imaging data. The p -value from Mantel's test for similarity between two distance matrices was obtained. To visualize the comparison between two distance measures, the scaled PCoA plot based on microbiome data (red) is given and the scaled PCoA plot based on imaging data (blue) is overlaid. Additional plots are created after flipping x and y axes of imaging data PCoA. The line segments represent the same sample.

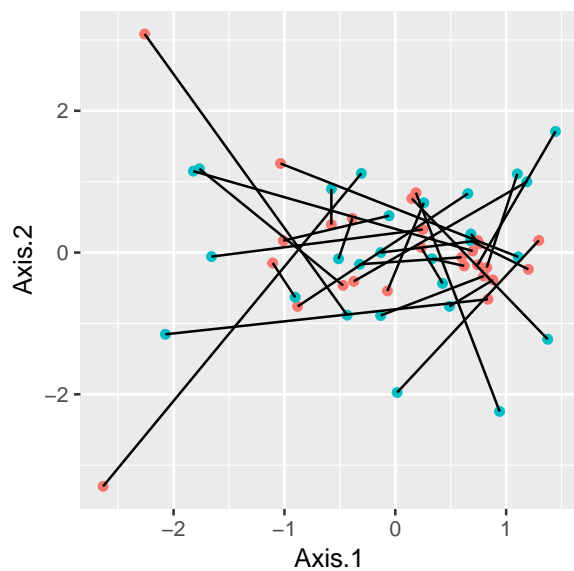
```
## Mantel's permutation test for similarity p-value: 0.335
```



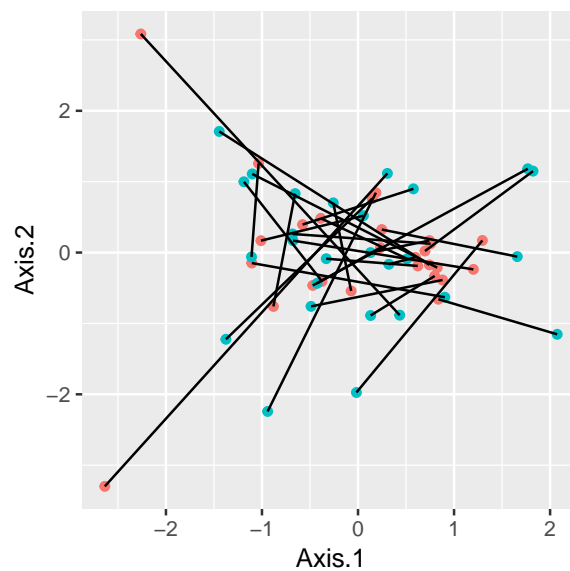
source Microbiome Nmf



source Microbiome Nmf_x_ref



source Microbiome Nmf_y_ref



source Microbiome Nmf_xy_ref