

Greg Finkelberg: HW6

R packages

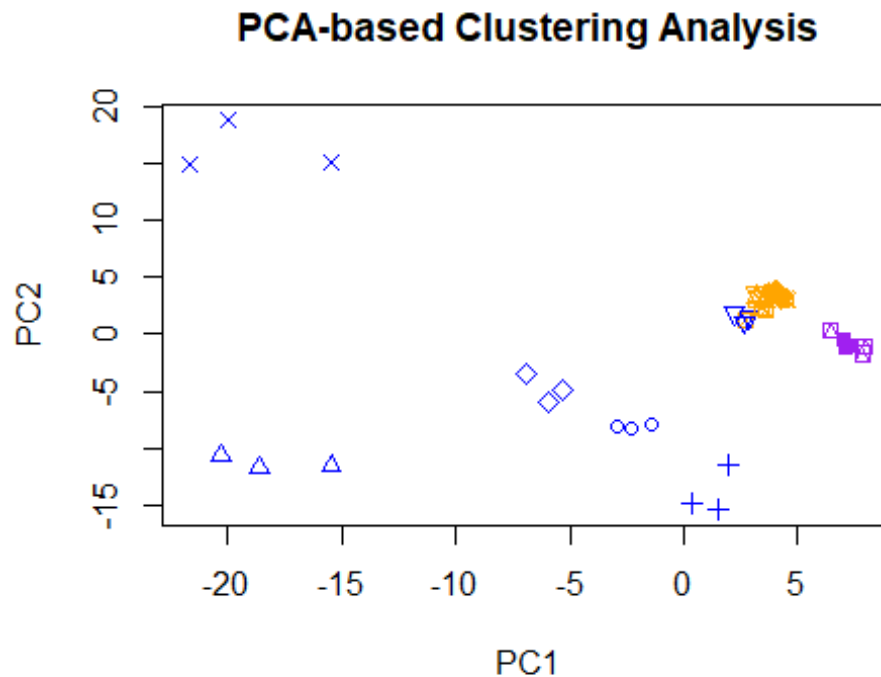
```
library(vegan)
```

Load dataset

```
load("C:/Users/Greg/Downloads/tobacco_clr.Rdata")
```

PCA cluster analysis

```
pca_result <- prcomp(tobacco_clr$data, scale. = TRUE)
plot(pca_result$x[, 1], pca_result$x[, 2], col = tobacco_clr$sample.color,
     pch = tobacco_clr$sample.pch, main = "PCA-based Clustering Analysis", xlab =
"PC1", ylab = "PC2")
```



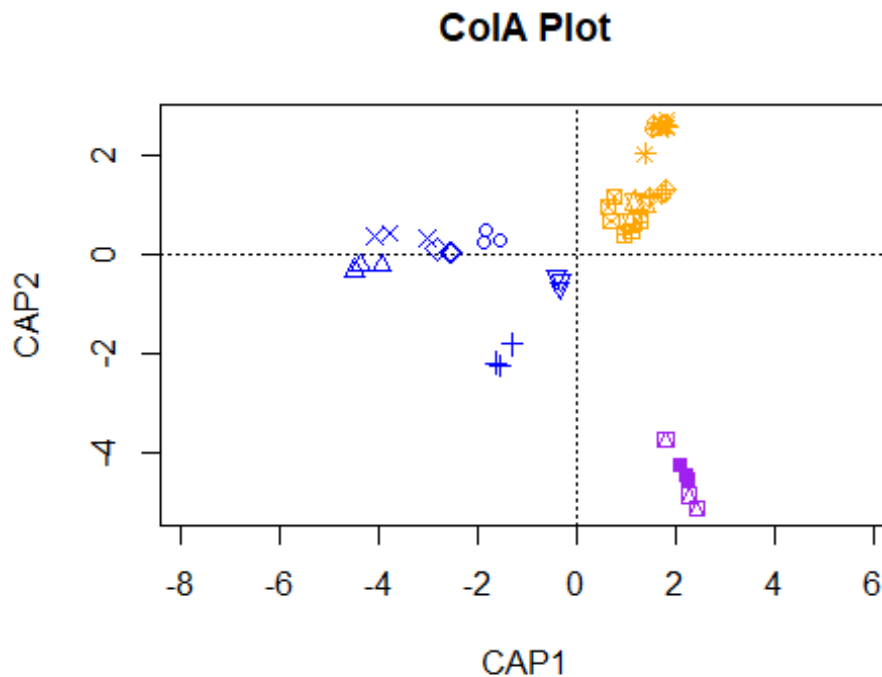
OTU sample points seem to be closely clustered by shapes across the board. Orange and purple data points are very tightly clustered, but the blue points seem to have much more variability and are not as closely related. Additionally, orange and purple data points contribute much more to PC1 and PC2 vs blue data points.

CoIA analysis

```
coia_result <- capscale(tobacco_clr$data ~ tobacco_clr$H, distance =
"euclidean")
```

```
plot(coia_result, display = "sites", type = "n", main = "CoIA Plot")

points(coia_result, display = "sites", col = tobacco_clr$sample.color, pch =
tobacco_clr$sample.pch)
```



Orange and purple data points are closely clustered together while there is much more variability with the blue data points. Orange data points are positively associated with CAP1 and CAP2. Purple data points are positively associated with CAP1 and negatively associated with CAP2. Blue data points are negatively associated with CAP1, but do not appear to have a trend with CAP2.

Comparing plots

Both plots show much of the same trends. Orange and purple data points are closely clustered in both, and blue points are not closely clustered in both. Shapes also appear to be closely clustered in both plots. We see that trend in all three colors. The major difference I found is a very negative relationship between the purple data points and CAP2, while the same points had no strong relationship with PC2 in either direction. This change is likely due to the inclusion of the similarity matrix into the analysis.