Solution Sheet 7

The full R code will be made available in a separate file on the course homepage.

1. **a)** The vegetation types are reasonably well seperated. Type 1 (circle) and type 2 (triangle) are probably the most difficult to tell apart.

   b) In comparison to **a)**, the separation worsens. Type 1 (circle) and type 2 (triangle) are no longer separated. Working with standardized data increases the influence of uncommon species on the dissimilarity, which may be inappropriate since the counts of rare species tend to be less reliably determined.

   c) Metric MDS using ℓ^2-dissimilarities is expected to give very similar results as PCA. Here, we used non-metric scaling, and ℓ^1-dissimilarities, but still, the result is quite similar.

2. **a)** setosa is well separated from the other two species.

   b) virginica and versicolor are not well separated. Instead, two observations that reside at the margins of the point clouds form a cluster.

   c) Using single or centroid linkage does not properly separate the species virginica and versicolor. In contrast, using complete linkage, the species are better separated.

3. **a)** The location of the initial centers determines which one of the two large point clouds will be split in half.

   b) Using the logarithms of the means of the 3 species as initial centers, the 3 species can be well separated.