A Handbook of Statistical Analyses Using R — 2nd Edition

Brian S. Everitt and Torsten Hothorn

CHAPTER 17

Multidimensional Scaling: British Water Voles and Voting in US Congress

17.1 Introduction

17.2 Multidimensional Scaling

17.3 Analysis Using R

We can apply classical scaling to the distance matrix for populations of water voles using the R function `cmdscale`. The following code finds the classical scaling solution and computes the two criteria for assessing the required number of dimensions as described above.

```
R> data("watervoles", package = "HSAUR2")
R> voles_mds <- cmdscale(watervoles, k = 13, eig = TRUE)
R> voles_mds$eig
```
```
 [1]  7.36e-01  2.63e-01  1.49e-01  6.99e-02  2.96e-02
 [6]  1.93e-02  1.39e-16 -1.14e-02 -1.28e-02 -2.85e-02
[11] -4.25e-02 -5.26e-02 -7.41e-02 -1.10e-01
```

Note that some of the eigenvalues are negative. The criterion P_2 can be computed by

```
R> sum(abs(voles_mds$eig[1:2]))/sum(abs(voles_mds$eig))
```
```
[1] 0.625
```

and the criterion suggested by Mardia et al. (1979) is

```
R> sum((voles_mds$eig[1:2])^2)/sum((voles_mds$eig)^2)
```
```
[1] 0.922
```

The two criteria for judging number of dimensions differ considerably, but both values are reasonably large, suggesting that the original distances between the water vole populations can be represented adequately in two dimensions. The two-dimensional solution can be plotted by extracting the coordinates from the `points` element of the `voles_mds` object; the plot is shown in Figure 17.1. The `textplot` function from package **wordcloud** can be used to annotate the plot with non-overlapping text.

We shall now apply non-metric scaling to the voting behaviour shown in Table **??**. Non-metric scaling is available with function `isoMDS` from package **MASS** (Venables and Ripley, 2002):

```
R> library("MASS")
R> data("voting", package = "HSAUR2")
R> voting_mds <- isoMDS(voting)
```

```
R> x <- voles_mds$points[,1]
R> y <- voles_mds$points[,2]
R> plot(x, y, xlab = "Coordinate 1", ylab = "Coordinate 2",
+       xlim = range(x)*1.2, type = "n")
R> textplot(x, y, words = colnames(watervoles), new = FALSE)
```
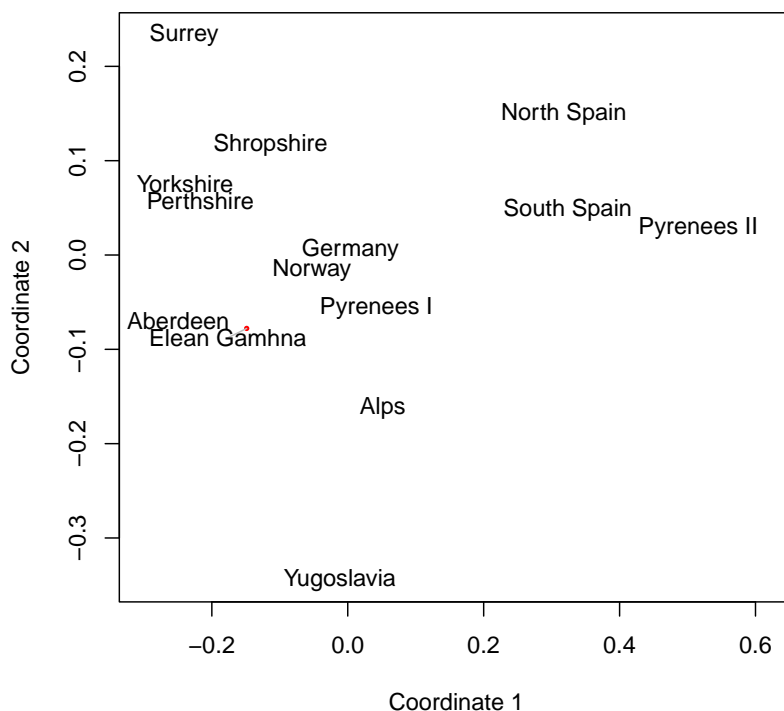

Figure 17.1   Two-dimensional solution from classical multidimensional scaling of distance matrix for water vole populations.

and we again depict the two-dimensional solution (Figure 17.3). The Figure suggests that voting behaviour is essentially along party lines, although there is more variation among Republicans. The voting behaviour of one of the Republicans (Rinaldo) seems to be closer to his democratic colleagues rather than to the voting behaviour of other Republicans.

```
R> library("ape")
R> st <- mst(watervoles)
R> plot(x, y, xlab = "Coordinate 1", ylab = "Coordinate 2",
+       xlim = range(x)*1.2, type = "n")
R> for (i in 1:nrow(watervoles)) {
+       w1 <- which(st[i, ] == 1)
+       segments(x[i], y[i], x[w1], y[w1])
+ }
R> textplot(x, y, words = colnames(watervoles), new = FALSE)
```
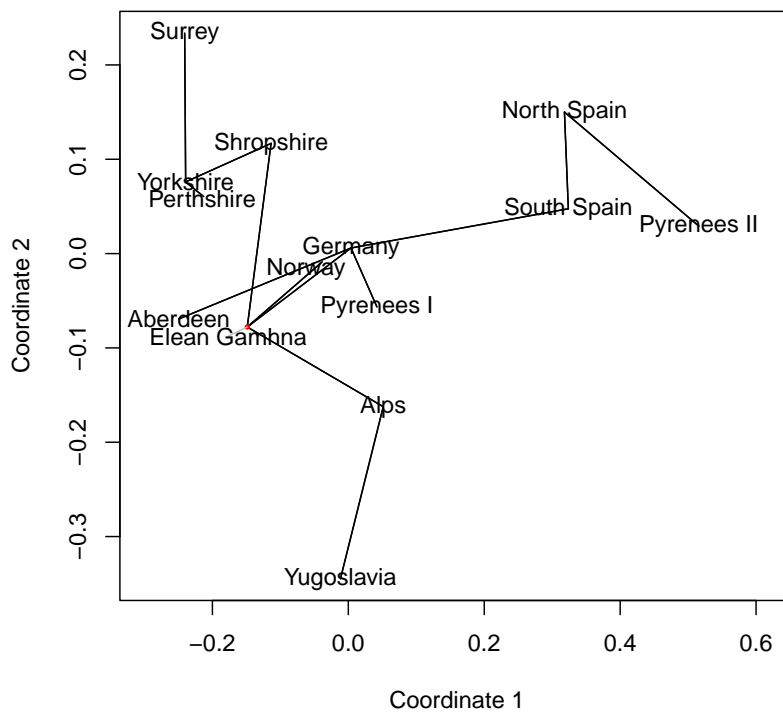

Figure 17.2   Minimum spanning tree for the `watervoles` data.

```
R> x <- voting_mds$points[,1]
R> y <- voting_mds$points[,2]
R> plot(x, y, xlab = "Coordinate 1", ylab = "Coordinate 2",
+        xlim = range(voting_mds$points[,1])*1.2, type = "n")
R> textplot(x, y, words = colnames(voting), new = FALSE)
R> voting_sh <- Shepard(voting[lower.tri(voting)],
+                        voting_mds$points)
```
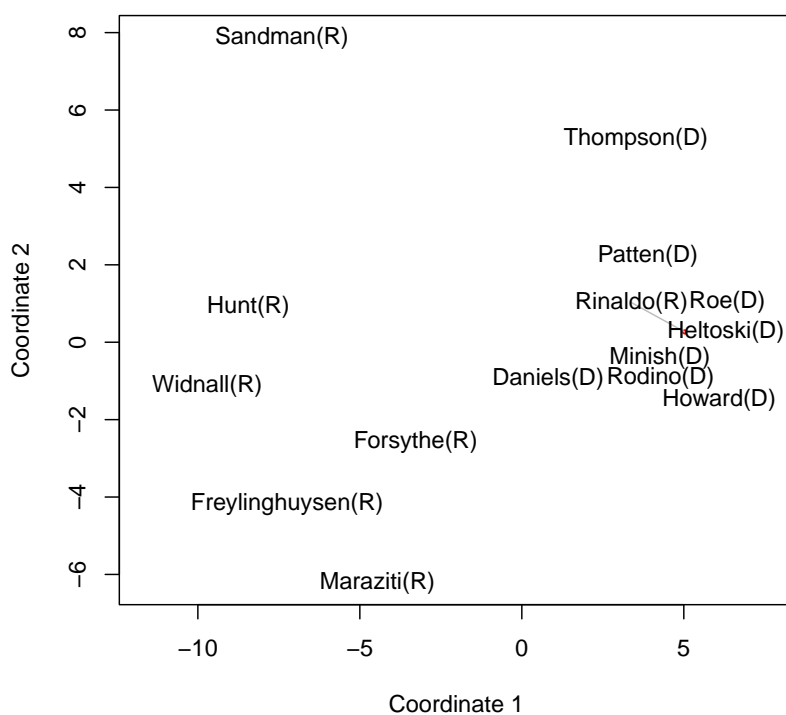

Figure 17.3 Two-dimensional solution from non-metric multidimensional scaling of distance matrix for voting matrix.

```
R> plot(voting_sh, pch = ".", xlab = "Dissimilarity",
+       ylab = "Distance", xlim = range(voting_sh$x),
+       ylim = range(voting_sh$x))
R> lines(voting_sh$x, voting_sh$yf, type = "S")
```
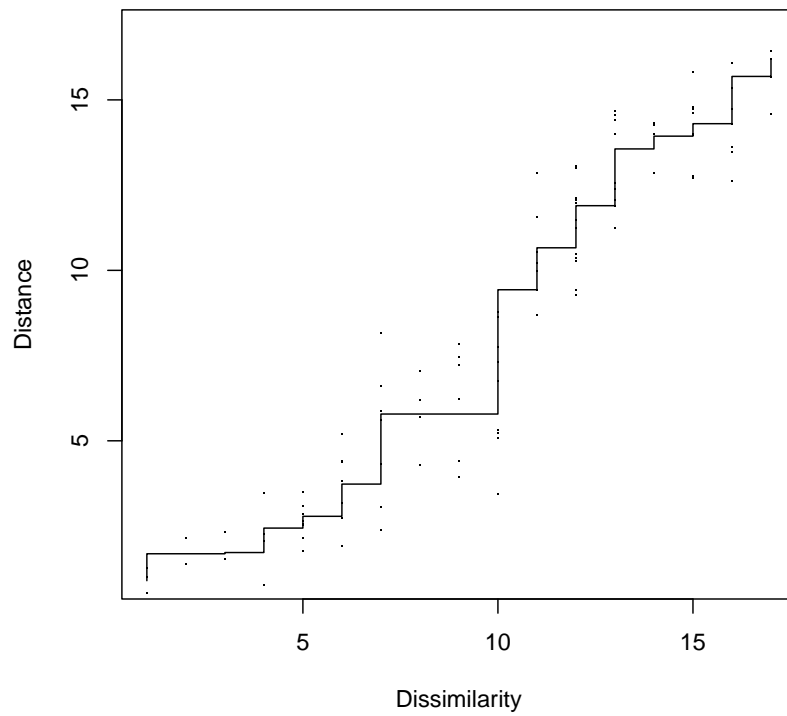

Figure 17.4   The Shepard diagram for the `voting` data shows some discrepancies between the original dissimilarities and the multidimensional scaling solution.

Bibliography

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, London, UK: Academic Press.

Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, New York, USA: Springer-Verlag, 4th edition, URL `http://www.stats.ox.ac.uk/pub/MASS4/`, ISBN 0-387-95457-0.