Multidimensional Scaling

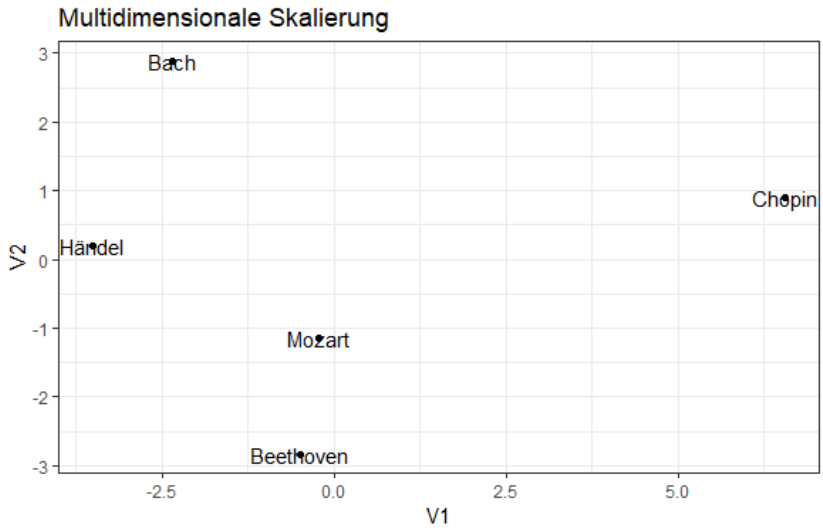
Question 1:

A study was conducted to investigate how the works of famous composers are judged by listeners. To do this, a study participant was asked to rank all pairs of the five composers *Bach, Mozart, Beethoven, Chopin* and *Händel*. The most similar pair was given a 1, the second most similar pair a 2, etc. A total of 10 pairs were compared. Then, the data was analyzed using multidimensional scaling (MDS).

a) What is the general aim of using an MDS? In addition, please state the practical advantage of using an MDS over a principal component analysis to graphically display the results of a cluster analysis graphically.

b) Briefly describe in your own words (no formulas!) which calculation steps are undertaken in a metric MDS. As data, you are given an $n \times p$ data matrix of the examined objects with all variables already standardized .

c) (i) Applying MDS in R yields the following output:




```
$points
                [,1]       [,2]
Händel    -3.4986510  0.1877229
```

```
Bach      -2.3515949  2.8932054
Beethoven -0.4864350 -2.8317511
Chopin     6.5536394  0.8904531
Mozart    -0.2169585 -1.1396304


$eig
[1]  61.00  18.52  4.03  -6.55  0.00
```

How would you interpret this output?

d) To obtain the "most suitable" dimension to reduce the data to for MDS, one may proceed the same way as for PCA - using eigenvalues. In this specific case, how many dimensions would you reduce to?

Question 2: MDS in R

In this task, you are to perform both a classic (metric) and a non-metric MDS in `R`.
We use the data set `gardenflowers` from the package `HSAUR3` as a basis. This represents a distance matrix for 18 types of garden flowers, which was calculated on the basis of various characteristics. Since it is unclear whether the distances represent accurate objective distance measurements or are based on subjective, more *ordinal* assessments, we apply both types of MDS discussed.

a) Load the data set into `R`. Get an initial overview of the distance matrix by plotting it with `levelplot()` (package `lattice`). Can you already recognize the first similar groups?

b) Perform a classic MDS using the function `cmdscale()` (`stats`). Is a restriction to two dimensions justifiable? Plot the data on two dimensions and interpret the result.

c) Perform a non-metric MDS using the function `isoMDS()` (`MASS`). Assess again whether a restriction to two dimensions is justifiable and interpret the result using a two-dimensional plot.

d) Assuming you have the raw data as a data set:

(i) How could you check whether the content of the two dimensions used for the plots can be interpreted?

(ii) Additionally, assume that the distance matrix contains Euclidean distances. How would the results change if you were to perform a principal component analysis (on the original, standardized variables) instead of a metric MDS (on the pairwise distances) in subtask b)?