The purpose of this practical is to learn to apply multidimensional scaling (MDS) in R. The data set we will use in this practical comes from a study on breakfast cereals. The observations concern 43 cereals recorded in the file `Cereals.dat`. For each cereal, the following variables were registered: *Manufacturer*: The manufacturer (coded as G (General Mills), K (Kellogg) or Q (Quaker)), *Calories*: Amount of calories, *Protein*: Protein content, *Fat*: Fat content, *Sodium*: Sodium content, *Fiber*: Fiber content, *Carbohydrates*: Carbohydrate content, *Sugar*: Sugar content and *Potassium*: Potassium content.

Open a document for writing your report, and paste your numerical and graphical results into this document. Comment on your results, and also write your answers to the questions posed also in this document. Use the same numbering of items as used in this practical. You can include R instructions in your document. Upload your work as a .pdf file (e.g. PepeGonzalezMDS.pdf) to the web pages of the course on atenea.upc.edu no later than 3/4/2018.

1. Import the data file Cereals.dat in the R environment.
2. (1p) Compute the Euclidean distance matrix for the cereals, using the information on *calories* and all seven cereal components, and standardizing the variables prior to the calculation of the distance matrix. You can use the R functions `scale` and `dist` for this purpose. Paste the distance matrix of the first 5 specimens into your report.
3. (1p) Perform a metric MDS of the data, using the `cmdscale` program. Plot the two-dimensional solution, and label the cereals with a number or abbreviated name. Use a different colour or symbol to label each cereal according to its manufacturer.
4. (1p) Which pair of cereals is, according to the two-dimensional solution of the analysis, the most similar?
5. (1p) Which pair of cereals would you, according to the two-dimensional solution of the analysis, classify as most distinct?
6. (2p) Is it possible to find a configuration of the 43 cereals in k dimensions that will represent the original distance matrix exactly? Why or why not? If so, how many dimensions would be needed to obtain this exact representation?
7. (1p) Report the eigenvalues of the solution, and calcualte the goodness-of-fit of the two-dimensional solution.
8. (1p) Are there any zero eigenvalues? Can you explain these?
9. (3p) Compute the fitted distances according to the two-dimensional MDS solution. Graph fitted and observed distances and assess the goodness of fit by regression. What do you observe? Report the coefficient of determination of this regression.
10. (1p) Try now non-metric MDS with the `isoMDS` program. Plot the two-dimensional solution, labelling the points again with the name or number of the brand, and using different symbols for different manufacturers.
11. (1p) Which pair of cereals is, according to the two-dimensional solution of the non-metric analysis, most similar?

12. (2p) Compute the fitted distances according to the two-dimensional non-metric MDS solution. Graph fitted and observed distances and assess the goodness of fit by regression. What do you observe? Report the coefficient of determination of this regression.

13. (2p) Compute the stress for a 1, 2, 3, 4 and 5 dimensional solution. How many dimensions do you think are necessary to obtain a "good fit"?

14. (2p) Make a scatterplot matrix of the first two dimensions of the metric MDS solution and the non-metric MDS solution (use $k = 2$). Calculate the correlation matrix of these four variables and comment on your results.

15. (1p) We have used the Euclidean distance as a metric in this exercise, on the standardized variables. We could also have calculated the Euclidean distance matrix without prior standardization of the variables. Which approach do you think is preferable? Argue your answer.