# Non-supervised methods

**TASK 1 - Multidimensional reduction:** File `nhanes.Rdata` contains two tables (`nhanes.nut` and `nhanes.air`) including variables about nutrients and air pollution obtained from NHANES project. The file also contains two objects describing the column names of those tables (`nut.desc` and `air.desc`, respectively). These tables can be loaded into R by executing

```
load(‘‘data_exercises/nhanes.Rdata’’)
```

1. Perform a principal component analysis of nutrient variables (columns 1:29 of table `nhanes.nut`) using variable  as the grouping variable and determine those variables that are associated with each category (`normal` and `hypercol`). Use the default method of the `ord` function and do not forget to scale the data.

**TASK 2 - Multidimensional reduction:** File `nci60.Rdata` contains miRNA, mRNA and protein data of melanoma, leukemia and CNS disease. Data are encapsulated in a list where each components stands for a given omic data (NOTE: features are in rows and samples in columns). Data corresponds to cells lines from the NCI-60 panel available at TCGA project. 21 cell lines are providing information about 537 miRNAs, 12,895 gene expression and 7,016 proteins. We are interested in obtaining omic profiles to characterize those diseases. NOTE: The vector `cancer` is a factor variable indicating the type of cancer of each sample.

1. Load data into R and select miRNA table by executing

```
load(’’data_exercises/nci60.Rdata’’)
miRNA <- nci60$miRNA
```

2. Perform a PCA using of miRNA dataset and give the top-5 features associated with each tumor.
3. How much variability is explained by the first two axes?
4. Determine how many axes are necessary to be selected to properly reduce the dimensionallity of this data.

**TASK 3 - Hierarchical analysis:** Using the same data, perform clustering analysis (use the thechnique you prefer) and assess whether these clusters can be used to predict cancer types (variable `cancer`).