Principal Component Analysis

Dataset dados.trat

This dataset contains 18 individuals and 6 variables, 1 qualitative variable is considered as illustrative.

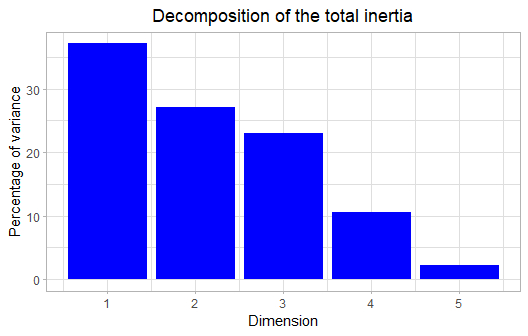
### 1. Study of the outliers

The analysis of the graphs does not detect any outlier.

### 2. Inertia distribution

The inertia of the first dimensions shows if there are strong relationships between variables and suggests the number of dimensions that should be studied.

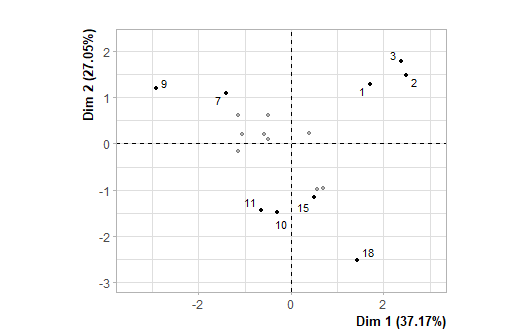
The first two dimensions of analyse express **64.23%** of the total dataset inertia ; that means that 64.23% of the individuals (or variables) cloud total variability is explained by the plane. The inertia observed on the first plane is smaller than the reference value that equals **67.66%**, therefore low in comparison (the reference value is the 0.95-quantile of the inertia percentages distribution obtained by simulating 413 data tables of equivalent size on the basis of a normal distribution). Moreover, the inertia projected on the first dimension is smaller than the reference value **41.92%**. The variability expressed by the analyse is thus **not** significant.



**Figure 2 - Decomposition of the total inertia**

An estimation of the right number of axis to interpret suggests to not interpret the analysis at all. Indeed, the amount of inertia of the first axis is not higher than that obtained by the 0.95-quantile of random distributions (37.17% against 41.92%). This observation suggests that no axis is carrying a real information. As a consequence, the description will stand to these axis.

### 3. Description of the dimension 1

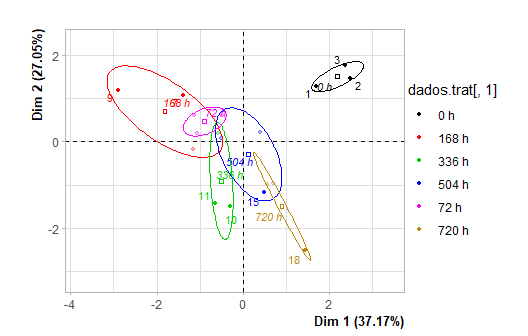


**Figure 3.1 - Individuals factor map (PCA)** *The labeled individuals are those with the higher contribution to the plane construction.*

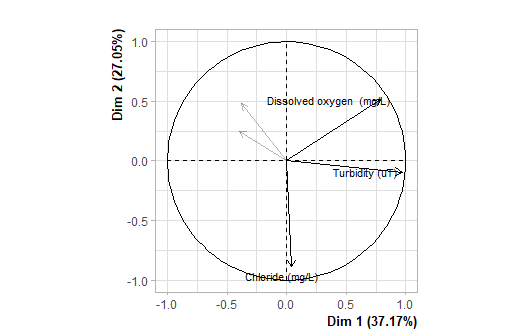
The Wilks test p-value indicates which variable factors are the best separated on the plane (i.e. which one explain the best the distance between individuals).

## dados.trat[, 1]   
## 2.075009e-07

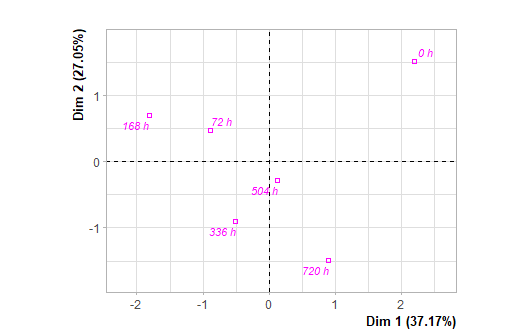
There only is one possible qualitative variable to illustrate the distance between individuals : *dados.trat[, 1]*.



**Figure 3.2 - Individuals factor map (PCA)** *The labeled individuals are those with the higher contribution to the plane construction.* *The individuals are coloured after their category for the variable* dados.trat[, 1].



**Figure 3.3 - Variables factor map (PCA)** *The labeled variables are those the best shown on the plane.*



**Figure 3.4 - Qualitative factor map (PCA)** *The labeled factors are those the best shown on the plane.*

The **dimension 1** opposes individuals such as *2*, *1* and *3* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *9* and *7* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *2*, *1* and *3* stand (characterized by a positive coordinate on the axis) is sharing :

* high values for the variables *Dissolved.oxygen..(mg/L)* and *Turbidity.(uT)* (variables are sorted from the strongest).

The group in which the individuals *9* and *7* stand (characterized by a negative coordinate on the axis) is sharing :

* low values for the variable *Turbidity.(uT)*.

Note that the variable *Turbidity (uT)* is highly correlated with this dimension (correlation of 0.94). This variable could therefore summarize itself the dimension 1.

The **dimension 2** opposes individuals such as *2*, *1*, *3*, *9* and *7* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *15*, *18*, *10* and *11* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *9* and *7* stand (characterized by a positive coordinate on the axis) is sharing :

* low values for the variable *Turbidity.(uT)*.

The group in which the individuals *2*, *1* and *3* stand (characterized by a positive coordinate on the axis) is sharing :

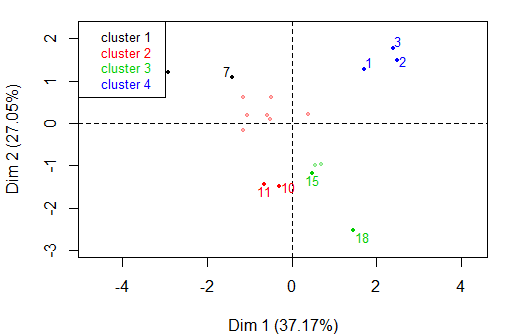
* high values for the variables *Dissolved.oxygen..(mg/L)* and *Turbidity.(uT)* (variables are sorted from the strongest).

The group in which the individuals *15*, *18*, *10* and *11* stand (characterized by a negative coordinate on the axis) is sharing :

* high values for the variable *Chloride.(mg/L)*.
* low values for the variable *Dissolved.Organic.Carbon.(mg/L)*.

### 4. Description of the plane -1:0

### 5. Classification



**Figure 5 - Ascending Hierarchical Classification of the individuals.** *The classification made on individuals reveals 4 clusters.*

The **cluster 1** is made of individuals such as *7* and *9*. This group is characterized by :

* high values for the variable *True.color.(uH)*.
* low values for the variable *Turbidity.(uT)*.

The **cluster 2** is made of individuals such as *10* and *11*. This group is characterized by :

* high values for the variable *Dissolved.Organic.Carbon.(mg/L)*.

The **cluster 3** is made of individuals such as *15* and *18*. This group is characterized by :

* low values for the variable *Dissolved.Organic.Carbon.(mg/L)*.

The **cluster 4** is made of individuals such as *1*, *2* and *3*. This group is characterized by :

* high values for the variables *Dissolved.oxygen..(mg/L)* and *Turbidity.(uT)* (variables are sorted from the strongest).

## Annexes

dimdesc(res, axes = 1:0)

$Dim.1  
$quanti  
 correlation p.value  
Turbidity (uT) 0.9670656 6.276745e-11  
Dissolved oxygen (mg/L) 0.7898294 9.661315e-05  
  
$quali  
 R2 p.value  
dados.trat[, 1] 0.8944496 1.758909e-05  
  
$category  
 Estimate p.value  
dados.trat[, 1]=0 h 2.190067 0.0007834697  
dados.trat[, 1]=168 h -1.816221 0.0090771805  
  
attr(,"class")  
[1] "condes" "list "   
  
$<NA>  
$quanti  
 correlation p.value  
Dissolved oxygen (mg/L) -0.5030248 0.03334882  
  
attr(,"class")  
[1] "condes" "list "   
  
$call  
$call$num.var  
[1] 1  
  
$call$proba  
[1] 0.05  
  
$call$weights  
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
  
$call$X  
 Dim.1 dados.trat[, 1] Turbidity (uT) True color (uH) Dissolved oxygen (mg/L)  
5 -0.4913479 dados.trat[, 1]=72 h 4.31 80.7657 4.50  
11 -0.6551501 dados.trat[, 1]=336 h 4.23 69.4854 3.33  
9 -2.9054441 dados.trat[, 1]=168 h 1.49 125.8869 2.45  
16 0.6956074 dados.trat[, 1]=720 h 6.06 73.2455 3.54  
12 -0.5718773 dados.trat[, 1]=336 h 5.07 69.4854 2.73  
2 2.4872597 dados.trat[, 1]=0 h 7.74 77.0056 7.90  
6 -1.1371735 dados.trat[, 1]=72 h 3.59 61.9652 2.74  
1 1.7026656 dados.trat[, 1]=0 h 6.71 65.7253 6.30  
8 -1.1466194 dados.trat[, 1]=168 h 2.67 46.9248 2.32  
7 -1.3965992 dados.trat[, 1]=168 h 2.95 122.1268 4.44  
10 -0.2904939 dados.trat[, 1]=336 h 4.81 69.4854 3.19  
14 -0.4977769 dados.trat[, 1]=504 h 4.12 77.0056 3.53  
4 -1.0550812 dados.trat[, 1]=72 h 3.59 61.9652 2.39  
17 0.5615668 dados.trat[, 1]=720 h 5.35 88.2859 4.03  
13 0.3923641 dados.trat[, 1]=504 h 5.19 69.4854 4.79  
3 2.3802752 dados.trat[, 1]=0 h 6.66 65.7253 9.20  
18 1.4374064 dados.trat[, 1]=720 h 6.51 69.4854 3.64  
15 0.4904183 dados.trat[, 1]=504 h 5.78 77.0056 3.70  
 Chloride (mg/L) Dissolved Organic Carbon (mg/L)  
5 62.6528 17.00  
11 72.0511 17.00  
9 58.3456 15.00  
16 64.4523 11.90  
12 61.9534 17.30  
2 58.2736 14.00  
6 59.6134 17.30  
1 56.5814 13.80  
8 60.1321 14.60  
7 59.7072 13.60  
10 70.6972 15.60  
14 61.2889 14.00  
4 59.8171 15.60  
17 64.2314 9.55  
13 62.4493 15.00  
3 59.7708 15.90  
18 69.5783 8.22  
15 66.9576 12.70

**Figure 6 - List of variables characterizing the dimensions of the analysis.**

res.hcpc$desc.var

Link between the cluster variable and the categorical variables (chi-square test)  
=================================================================================  
 p.value df  
dados.trat[,.1] 0.0001996813 15  
  
Description of each cluster by the categories  
=============================================  
$`1`  
 Cla/Mod Mod/Cla Global p.value v.test  
dados.trat[,.1]=168 h 66.66667 100 16.66667 0.01960784 2.333769  
  
$`2`  
NULL  
  
$`3`  
 Cla/Mod Mod/Cla Global p.value v.test  
dados.trat[,.1]=720 h 100 75 16.66667 0.004901961 2.813407  
  
$`4`  
 Cla/Mod Mod/Cla Global p.value v.test  
dados.trat[,.1]=0 h 100 100 16.66667 0.00122549 3.23288  
  
  
Link between the cluster variable and the quantitative variables  
================================================================  
 Eta2 P-value  
True.color.(uH) 0.8365338 9.071372e-06  
Turbidity.(uT) 0.8267413 1.356305e-05  
Dissolved.oxygen..(mg/L) 0.7952423 4.295255e-05  
Dissolved.Organic.Carbon.(mg/L) 0.7308938 2.807390e-04  
Chloride.(mg/L) 0.4199724 4.860686e-02  
  
Description of each cluster by quantitative variables  
=====================================================  
$`1`  
 v.test Mean in category Overall mean sd in category Overall sd p.value  
True.color.(uH) 3.680169 124.0069 76.170022 1.88005 18.948465 0.0002330791  
Turbidity.(uT) -2.411089 2.2200 4.823889 0.73000 1.574305 0.0159049866  
  
$`2`  
 v.test Mean in category Overall mean sd in category Overall sd  
Dissolved.Organic.Carbon.(mg/L) 2.682511 15.93333 14.33722 1.184155 2.453274  
 p.value  
Dissolved.Organic.Carbon.(mg/L) 0.007307173  
  
$`3`  
 v.test Mean in category Overall mean sd in category Overall sd  
Dissolved.Organic.Carbon.(mg/L) -3.364062 10.5925 14.33722 1.793507 2.453274  
 p.value  
Dissolved.Organic.Carbon.(mg/L) 0.000768043  
  
$`4`  
 v.test Mean in category Overall mean sd in category Overall sd  
Dissolved.oxygen..(mg/L) 3.655700 7.800000 4.151111 1.1860298 1.840474  
Turbidity.(uT) 2.591723 7.036667 4.823889 0.4977505 1.574305  
 p.value  
Dissolved.oxygen..(mg/L) 0.0002564816  
Turbidity.(uT) 0.0095496759

**Figure 7 - List of variables characterizing the clusters of the classification.**