Principal Component Analysis

Dataset dados.ctrl

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

### 1. Study of the outliers

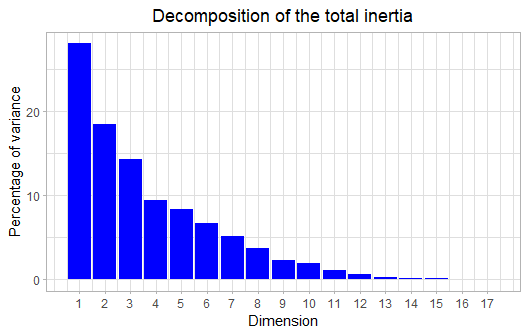
The analysis of the graphs does not detect any outlier.

### 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 **46.46%** of the total dataset inertia ; that means that 46.46% of the individuals (or variables) cloud total variability is explained by the plane. This is an intermediate percentage and the first plane represents a part of the data variability. This value is greater than the reference value that equals **38.38%**, the variability explained by this plane is thus significant (the reference value is the 0.95-quantile of the inertia percentages distribution obtained by simulating 2083 data tables of equivalent size on the basis of a normal distribution).

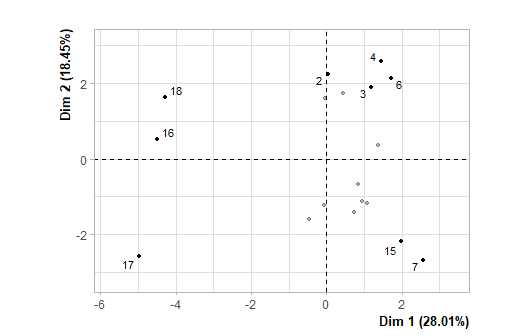
From these observations, it may be interesting to consider the next dimensions which also express a high percentage of the total inertia.



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

An estimation of the right number of axis to interpret suggests to restrict the analysis to the description of the first 3 axis. These axis present an amount of inertia greater than those obtained by the 0.95-quantile of random distributions (60.68% against 51.86%). This observation suggests that only these axis are carrying a real information. As a consequence, the description will stand to these axis.

### 3. Description of the plane 1:2

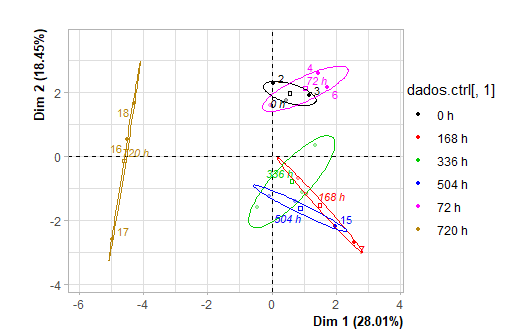


**Figure 3.1 - Individuals factor map (PCA)** *The labeled individuals are those with the higher contribution to the plane construction.*

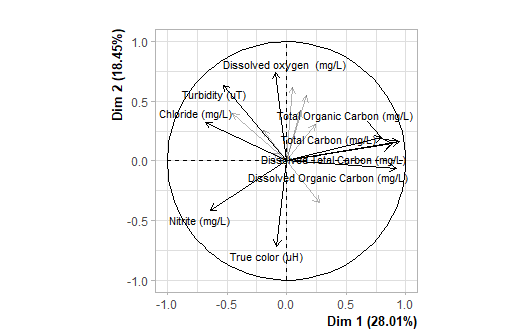
The Wilks test p-value indicates which variable factors are the best separated on the plane (i.e. which one explain the best the distance between individuals).

## dados.ctrl[, 1]   
## 7.546454e-07

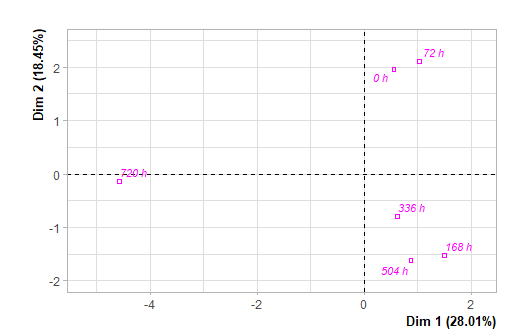
There only is one possible qualitative variable to illustrate the distance between individuals : *dados.ctrl[, 1]*.



**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.ctrl[, 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 *15* and *7* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *17*, *16* and *18* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *15* and *7* stand (characterized by a positive coordinate on the axis) is sharing :

* high values for the variable *True.color.(uH)*.
* low values for the variables *Dissolved.oxygen..(mg/L)*, *Turbidity.(uT)*, *Nitrate.(mg/L)* and *pH* (variables are sorted from the weakest).

The group in which the individuals *17*, *16* and *18* stand (characterized by a negative coordinate on the axis) is sharing :

* high values for the variables *Chloride.(mg/L)*, *Nitrite.(mg/L)* and *Turbidity.(uT)* (variables are sorted from the strongest).
* low values for the variables *Total.Carbon.(mg/L)*, *Dissolved.Total.Carbon.(mg/L)*, *Dissolved.Organic.Carbon.(mg/L)* and *Total.Organic.Carbon.(mg/L)* (variables are sorted from the weakest).

Note that the variable *720 h* is highly correlated with this dimension (correlation of 0.98). This variable could therefore summarize itself the dimension 1.

The **dimension 2** opposes individuals such as *6*, *4*, *2* and *3* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *15* and *7* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

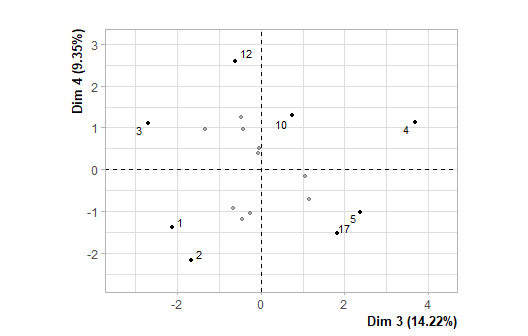
The group in which the individuals *6*, *4*, *2* and *3* stand (characterized by a positive coordinate on the axis) is sharing :

* high values for the variables *Dissolved.oxygen..(mg/L)* and *pH* (variables are sorted from the strongest).
* low values for the variable *True.color.(uH)*.

The group in which the individuals *15* and *7* stand (characterized by a negative coordinate on the axis) is sharing :

* high values for the variable *True.color.(uH)*.
* low values for the variables *Dissolved.oxygen..(mg/L)*, *Turbidity.(uT)*, *Nitrate.(mg/L)* and *pH* (variables are sorted from the weakest).

### 4. Description of the dimension 3

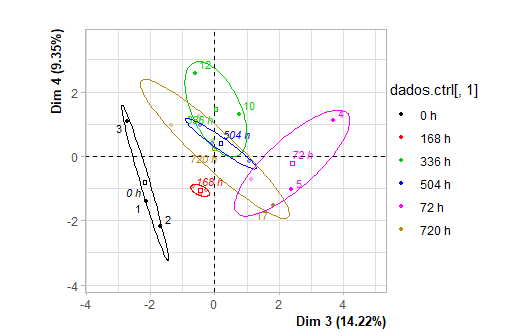


**Figure 4.1 - Individuals factor map (PCA)** *The labeled individuals are those with the higher contribution to the plane construction.*

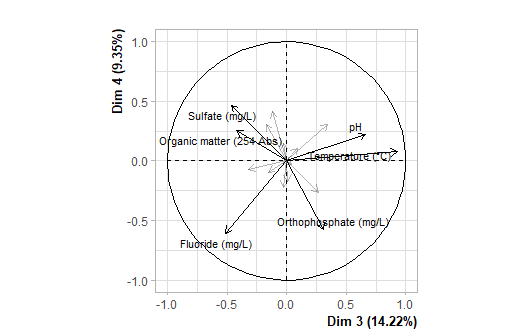
The Wilks test p-value indicates which variable factors are the best separated on the plane (i.e. which one explain the best the distance between individuals).

## dados.ctrl[, 1]   
## 0.002149993

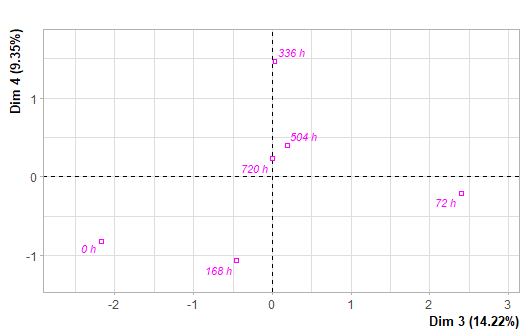
There only is one possible qualitative variable to illustrate the distance between individuals : *dados.ctrl[, 1]*.



**Figure 4.2 - 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.ctrl[, 1].



**Figure 4.3 - Variables factor map (PCA)** *The labeled variables are those the best shown on the plane.*



**Figure 4.4 - Qualitative factor map (PCA)** *The labeled factors are those the best shown on the plane.*

The **dimension 3** opposes individuals such as *5*, *4* and *17* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *2*, *12*, *1*, *3* and *10* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *5*, *4* and *17* stand (characterized by a positive coordinate on the axis) is sharing :

* high values for the variables *Temperature.(°C)*, *pH* and *Orthophosphate.(mg/L)* (variables are sorted from the strongest).
* low values for the variable *Organic.matter.(254.Abs)*.

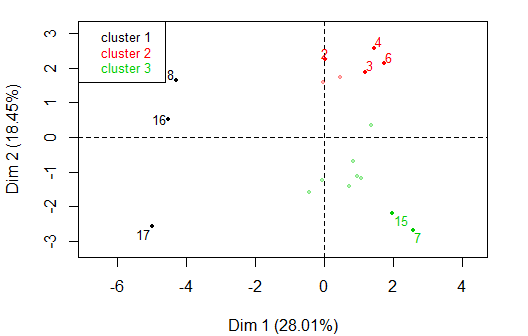
The group in which the individuals *2* and *1* stand (characterized by a negative coordinate on the axis) is sharing :

* high values for the variable *Fluoride.(mg/L)*.
* low values for the variable *pH*.

The group in which the individuals *12*, *3* and *10* stand (characterized by a negative coordinate on the axis) is sharing :

* low values for the variable *Orthophosphate.(mg/L)*.

### 5. Classification



**Figure 5 - Ascending Hierarchical Classification of the individuals.** *The classification made on individuals reveals 3 clusters.*

The **cluster 1** is made of individuals such as *16*, *17* and *18*. This group is characterized by :

* high values for the variables *Chloride.(mg/L)*, *Nitrite.(mg/L)* and *Turbidity.(uT)* (variables are sorted from the strongest).
* low values for the variables *Total.Carbon.(mg/L)*, *Dissolved.Total.Carbon.(mg/L)*, *Dissolved.Organic.Carbon.(mg/L)* and *Total.Organic.Carbon.(mg/L)* (variables are sorted from the weakest).

The **cluster 2** is made of individuals such as *2*, *3*, *4* and *6*. This group is characterized by :

* high values for the variables *Dissolved.oxygen..(mg/L)*, *pH*, *Fluoride.(mg/L)* and *Orthophosphate.(mg/L)* (variables are sorted from the strongest).
* low values for the variable *True.color.(uH)*.

The **cluster 3** is made of individuals such as *7* and *15*. This group is characterized by :

* low values for the variables *Dissolved.oxygen..(mg/L)*, *Turbidity.(uT)* and *pH* (variables are sorted from the weakest).

## Annexes

dimdesc(res, axes = 1:3)

$Dim.1  
$quanti  
 correlation p.value  
Total Carbon (mg/L) 0.9445637 3.766407e-09  
Dissolved Organic Carbon (mg/L) 0.9181036 7.850606e-08  
Dissolved Total Carbon (mg/L) 0.8859229 1.002486e-06  
Total Organic Carbon (mg/L) 0.7985561 7.088577e-05  
Turbidity (uT) -0.5302707 2.358944e-02  
Nitrite (mg/L) -0.6403510 4.199706e-03  
Chloride (mg/L) -0.6798832 1.907735e-03  
  
$quali  
 R2 p.value  
dados.ctrl[, 1] 0.9021827 1.126425e-05  
  
$category  
 Estimate p.value  
dados.ctrl[, 1]=720 h -4.583027 7.67611e-09  
  
attr(,"class")  
[1] "condes" "list "   
  
$Dim.2  
$quanti  
 correlation p.value  
Dissolved oxygen (mg/L) 0.7384976 0.0004651375  
Turbidity (uT) 0.6367644 0.0044877414  
pH 0.6153995 0.0065561279  
Conductivity (µS/cm) 0.5449703 0.0193456184  
True color (uH) -0.7194543 0.0007636790  
  
$quali  
 R2 p.value  
dados.ctrl[, 1] 0.7348687 0.003462072  
  
$category  
 Estimate p.value  
dados.ctrl[, 1]=72 h 2.103629 0.02329963  
dados.ctrl[, 1]=0 h 1.961979 0.03655498  
  
attr(,"class")  
[1] "condes" "list "   
  
$Dim.3  
$quanti  
 correlation p.value  
Temperature (°C) 0.9269219 3.246289e-08  
pH 0.6603912 2.854421e-03  
Fluoride (mg/L) -0.5128357 2.953035e-02  
  
$quali  
 R2 p.value  
dados.ctrl[, 1] 0.7397917 0.003119113  
  
$category  
 Estimate p.value  
dados.ctrl[, 1]=72 h 2.404729 0.001476561  
dados.ctrl[, 1]=0 h -2.169174 0.005662502  
  
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.ctrl[, 1] Turbidity (uT) True color (uH) Temperature (°C) pH  
11 -0.44952931 dados.ctrl[, 1]=336 h 5.51 84.5258 29.2 8.81  
18 -4.28215089 dados.ctrl[, 1]=720 h 10.10 77.0056 28.9 9.30  
16 -4.50222182 dados.ctrl[, 1]=720 h 7.27 77.0056 29.0 8.82  
4 1.43655976 dados.ctrl[, 1]=72 h 6.25 65.7253 29.9 9.97  
1 0.45355146 dados.ctrl[, 1]=0 h 6.55 65.7253 28.6 8.78  
14 0.72874437 dados.ctrl[, 1]=504 h 4.50 73.2455 29.2 8.92  
17 -4.96470858 dados.ctrl[, 1]=720 h 5.17 88.2859 29.6 8.80  
12 0.94109186 dados.ctrl[, 1]=336 h 5.31 84.5258 29.0 8.96  
15 1.97156441 dados.ctrl[, 1]=504 h 4.26 77.0056 28.8 8.91  
10 1.36346377 dados.ctrl[, 1]=336 h 5.99 73.2455 29.4 8.91  
5 -0.04103369 dados.ctrl[, 1]=72 h 5.91 69.4854 29.4 9.64  
9 1.07701505 dados.ctrl[, 1]=168 h 3.83 92.0460 29.0 8.72  
2 0.03568333 dados.ctrl[, 1]=0 h 6.50 65.7253 28.8 8.84  
6 1.71779914 dados.ctrl[, 1]=72 h 5.97 80.7657 29.3 9.58  
13 -0.07455115 dados.ctrl[, 1]=504 h 5.77 77.0056 28.9 8.82  
8 0.83899703 dados.ctrl[, 1]=168 h 3.59 69.4854 28.9 8.49  
3 1.18023319 dados.ctrl[, 1]=0 h 6.95 65.7253 28.7 8.92  
7 2.56949205 dados.ctrl[, 1]=168 h 4.59 95.8061 28.9 8.63  
 Conductivity (µS/cm) Dissolved oxygen (mg/L) Organic matter (254 Abs) Nitrite (mg/L)  
11 359 4.70 0.266 0.7857  
18 437 4.20 0.262 0.8377  
16 365 6.30 0.261 0.7757  
4 405 5.96 0.258 0.7650  
1 399 5.85 0.260 0.8025  
14 361 3.91 0.260 0.7814  
17 366 4.73 0.259 0.9984  
12 369 2.89 0.261 0.7803  
15 370 3.80 0.265 0.7995  
10 413 4.85 0.265 0.7717  
5 407 6.26 0.257 0.8071  
9 412 3.06 0.261 0.7597  
2 399 7.20 0.260 0.7656  
6 408 6.30 0.260 0.7470  
13 361 4.34 0.263 0.7954  
8 408 3.24 0.256 0.7960  
3 401 8.90 0.268 0.7393  
7 408 3.22 0.273 0.7977  
 Nitrate (mg/L) Orthophosphate (mg/L) Sulfate (mg/L) Fluoride (mg/L) Chloride (mg/L)  
11 0.9393 0.2104 4.0427 1.5510 59.4973  
18 0.9852 0.4059 5.0883 1.4414 72.2762  
16 0.8414 0.4964 5.3734 1.5756 76.5410  
4 0.7011 0.7062 4.5881 1.3360 60.8607  
1 1.0615 0.5352 4.8274 2.1755 58.3756  
14 0.7615 0.5699 4.3737 1.1738 50.4481  
17 0.7647 0.5118 4.3349 1.4624 64.3679  
12 0.5211 0.3660 5.5282 1.4243 69.3126  
15 0.2815 0.4326 4.7030 1.3601 52.9002  
10 0.8912 0.2588 4.3441 1.4597 59.0873  
5 0.6737 0.7605 4.1154 1.5626 58.4088  
9 0.5822 0.6784 4.5149 1.9637 63.3046  
2 0.8710 0.6535 4.5177 2.3617 58.4847  
6 0.7532 0.9436 4.7204 1.8501 61.5443  
13 0.5947 0.5980 4.6155 1.3369 57.9144  
8 0.6592 0.6788 5.1632 1.5498 51.9858  
3 0.5204 0.4501 5.1925 1.7332 62.2655  
7 0.6046 0.6523 4.1092 1.6512 48.6389  
 Total Organic Carbon (mg/L) Dissolved Organic Carbon (mg/L) Total Carbon (mg/L)  
11 18.7 15.0 35.20000  
18 14.6 12.1 27.10000  
16 12.1 11.0 25.40000  
4 24.8 16.6 41.30000  
1 17.6 17.6 36.83089  
14 17.6 16.9 34.70000  
17 11.4 10.5 24.30000  
12 21.2 18.8 37.80000  
15 19.1 18.2 37.40000  
10 24.8 16.6 41.30000  
5 18.7 15.0 35.20000  
9 18.1 16.6 36.00000  
2 17.2 13.4 35.89976  
6 21.2 18.8 37.80000  
13 18.3 14.7 35.30000  
8 17.9 15.9 37.20000  
3 17.5 16.5 36.96703  
7 19.6 18.7 37.80000  
 Dissolved Total Carbon (mg/L)  
11 31.3  
18 23.8  
16 23.4  
4 32.9  
1 37.8  
14 34.3  
17 23.2  
12 36.3  
15 35.9  
10 32.9  
5 31.3  
9 34.6  
2 36.6  
6 36.3  
13 31.5  
8 32.4  
3 40.8  
7 36.4

**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.ctrl[,.1] 8.41761e-05 10  
  
Description of each cluster by the categories  
=============================================  
$`1`  
 Cla/Mod Mod/Cla Global p.value v.test  
dados.ctrl[,.1]=720 h 100 100 16.66667 0.00122549 3.23288  
  
$`2`  
 Cla/Mod Mod/Cla Global p.value v.test  
dados.ctrl[,.1]=72 h 100 50 16.66667 0.0245098 2.249042  
dados.ctrl[,.1]=0 h 100 50 16.66667 0.0245098 2.249042  
  
$`3`  
NULL  
  
  
Link between the cluster variable and the quantitative variables  
================================================================  
 Eta2 P-value  
Total.Carbon.(mg/L) 0.8483910 7.168665e-07  
Dissolved.Total.Carbon.(mg/L) 0.7852649 9.756013e-06  
Dissolved.oxygen..(mg/L) 0.7040784 1.080997e-04  
Dissolved.Organic.Carbon.(mg/L) 0.6640274 2.800778e-04  
Total.Organic.Carbon.(mg/L) 0.5571241 2.223900e-03  
Turbidity.(uT) 0.5050658 5.118066e-03  
Chloride.(mg/L) 0.4976610 5.721064e-03  
Nitrite.(mg/L) 0.4024738 2.102215e-02  
True.color.(uH) 0.3822128 2.699552e-02  
pH 0.3490585 3.995441e-02  
  
Description of each cluster by quantitative variables  
=====================================================  
$`1`  
 v.test Mean in category Overall mean sd in category Overall sd  
Chloride.(mg/L) 2.802318 71.061700 60.3452167 5.04330240 7.05138428  
Nitrite.(mg/L) 2.571154 0.870600 0.7947611 0.09384608 0.05438803  
Turbidity.(uT) 2.192194 7.513333 5.7788889 2.02000550 1.45888423  
Total.Organic.Carbon.(mg/L) -3.077493 12.700000 18.3555556 1.37355985 3.38857922  
Dissolved.Organic.Carbon.(mg/L) -3.338980 11.200000 15.7166667 0.66833126 2.49427121  
Dissolved.Total.Carbon.(mg/L) -3.574223 23.466667 32.8722222 0.24944383 4.85223952  
Total.Carbon.(mg/L) -3.794951 25.600000 35.1943154 1.15181017 4.66173119  
 p.value  
Chloride.(mg/L) 0.0050736884  
Nitrite.(mg/L) 0.0101360255  
Turbidity.(uT) 0.0283654883  
Total.Organic.Carbon.(mg/L) 0.0020874967  
Dissolved.Organic.Carbon.(mg/L) 0.0008408660  
Dissolved.Total.Carbon.(mg/L) 0.0003512691  
Total.Carbon.(mg/L) 0.0001476724  
  
$`2`  
 v.test Mean in category Overall mean sd in category Overall sd  
Dissolved.oxygen..(mg/L) 3.246325 6.745000 4.9838889 1.0568467 1.5816276  
pH 2.335744 9.288333 8.9900000 0.4597977 0.3723797  
Fluoride.(mg/L) 2.217807 1.836517 1.6093889 0.3481219 0.2985767  
Orthophosphate.(mg/L) 2.049307 0.674850 0.5504667 0.1586484 0.1769557  
True.color.(uH) -2.549043 68.858717 76.7967056 5.4991239 9.0791015  
 p.value  
Dissolved.oxygen..(mg/L) 0.001169054  
pH 0.019504598  
Fluoride.(mg/L) 0.026568000  
Orthophosphate.(mg/L) 0.040432061  
True.color.(uH) 0.010801913  
  
$`3`  
 v.test Mean in category Overall mean sd in category Overall sd  
pH -2.140648 8.796667 8.990000 0.1475730 0.3723797  
Turbidity.(uT) -2.719437 4.816667 5.778889 0.8130464 1.4588842  
Dissolved.oxygen..(mg/L) -3.141284 3.778889 4.983889 0.6862044 1.5816276  
 p.value  
pH 0.032302474  
Turbidity.(uT) 0.006539314  
Dissolved.oxygen..(mg/L) 0.001682086

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