Principal Component Analysis

Dataset a

This dataset contains 36 individuals and 6 variables.

### 1. Study of the outliers

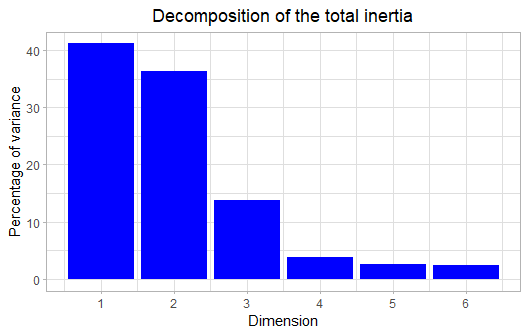
The analysis of the graphs does not detect any outlier.

### 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 **77.5%** of the total dataset inertia ; that means that 77.5% of the individuals (or variables) cloud total variability is explained by the plane. This percentage is high and thus the first plane represents an important part of the data variability. This value is strongly greater than the reference value that equals **53.2%**, the variability explained by this plane is thus highly significant (the reference value is the 0.95-quantile of the inertia percentages distribution obtained by simulating 1815 data tables of equivalent size on the basis of a normal distribution).

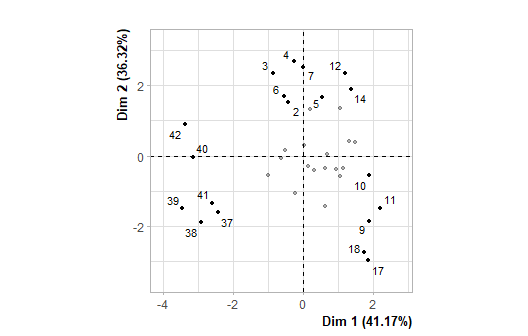
From these observations, it is probably not useful to interpret the next dimensions.



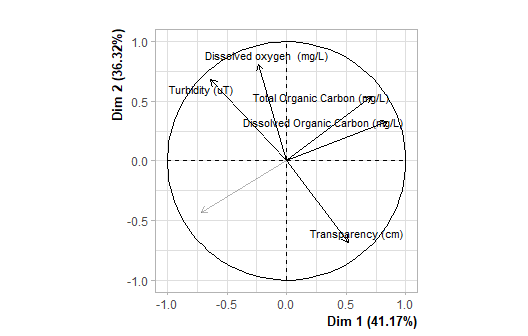
**Figure 2 - Decomposition of the total inertia**

An estimation of the right number of axis to interpret suggests to restrict the analysis to the description of the first 2 axis. These axis present an amount of inertia greater than those obtained by the 0.95-quantile of random distributions (77.5% against 53.2%). 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.*



**Figure 3.2 - Variables factor map (PCA)** *The labeled variables are those the best shown on the plane.*

The **dimension 1** opposes individuals such as *18*, *11*, *9*, *17* and *10* (to the right of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *38*, *39*, *37*, *41*, *40* and *42* (to the left of the graph, characterized by a strongly negative coordinate on the axis).

The group in which the individuals *18*, *11*, *9*, *17* and *10* stand (characterized by a positive coordinate on the axis) is sharing :

* high values for the variable *Transparency.(cm)*.
* low values for the variables *Turbidity.(uT)* and *Dissolved.oxygen..(mg/L)* (variables are sorted from the weakest).

The group in which the individuals *38*, *39*, *37*, *41*, *40* and *42* stand (characterized by a negative coordinate on the axis) is sharing :

* high values for the variable *Time*.
* low values for the variables *Dissolved.Organic.Carbon.(mg/L)* and *Total.Organic.Carbon.(mg/L)* (variables are sorted from the weakest).

The **dimension 2** opposes individuals such as *14*, *12*, *5*, *3*, *4* and *7* (to the top of the graph, characterized by a strongly positive coordinate on the axis) to individuals such as *18*, *11*, *9*, *17* and *10* (to the bottom of the graph, characterized by a strongly negative coordinate on the axis).

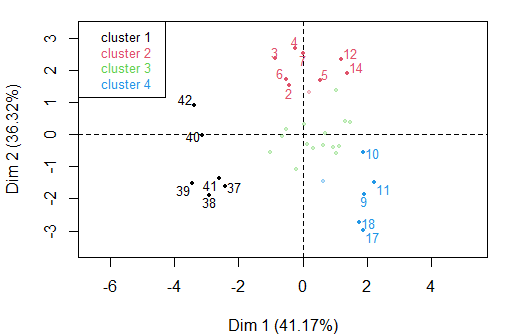
The group in which the individuals *14*, *12*, *5*, *3*, *4* and *7* stand (characterized by a positive coordinate on the axis) is sharing :

* high values for the variables *Dissolved.oxygen..(mg/L)*, *Total.Organic.Carbon.(mg/L)* and *Dissolved.Organic.Carbon.(mg/L)* (variables are sorted from the strongest).
* low values for the variables *Time* and *Transparency.(cm)* (variables are sorted from the weakest).

The group in which the individuals *18*, *11*, *9*, *17* and *10* stand (characterized by a negative coordinate on the axis) is sharing :

* high values for the variable *Transparency.(cm)*.
* low values for the variables *Turbidity.(uT)* and *Dissolved.oxygen..(mg/L)* (variables are sorted from the weakest).

### 4. Classification



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

The **cluster 1** is made of individuals such as *37*, *38*, *39*, *40*, *41* and *42*. This group is characterized by :

* high values for the variables *Time* and *Turbidity.(uT)* (variables are sorted from the strongest).
* low values for the variables *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*, *5*, *6*, *7*, *12* and *14*. This group is characterized by :

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

The **cluster 3** is made of individuals sharing :

* high values for the variables *Dissolved.Organic.Carbon.(mg/L)* and *Total.Organic.Carbon.(mg/L)* (variables are sorted from the strongest).
* low values for the variable *Dissolved.oxygen..(mg/L)*.

The **cluster 4** is made of individuals such as *9*, *10*, *11*, *17* and *18*. This group is characterized by :

* high values for the variable *Transparency.(cm)*.
* low values for the variables *Turbidity.(uT)* and *Dissolved.oxygen..(mg/L)* (variables are sorted from the weakest).

## Annexes

dimdesc(res, axes = 1:2)

$Dim.1  
$quanti  
 correlation p.value  
Dissolved Organic Carbon (mg/L) 0.8493992 5.780098e-11  
Total Organic Carbon (mg/L) 0.7119145 1.123485e-06  
Transparency (cm) 0.5157643 1.283946e-03  
Turbidity (uT) -0.6354039 3.133530e-05  
Time -0.7180715 8.207208e-07  
  
attr(,"class")  
[1] "condes" "list "   
  
$Dim.2  
$quanti  
 correlation p.value  
Dissolved oxygen (mg/L) 0.8117344 1.889122e-09  
Turbidity (uT) 0.6810781 4.835254e-06  
Total Organic Carbon (mg/L) 0.5404389 6.674790e-04  
Dissolved Organic Carbon (mg/L) 0.3307773 4.878284e-02  
Time -0.4350279 8.012254e-03  
Transparency (cm) -0.6825633 4.525116e-06  
  
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 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
[33] 1 1 1 1  
  
$call$X  
 Dim.1 Time Turbidity (uT) Transparency (cm)  
2 -0.423152876 0 6.71 49.0  
3 -0.846043864 0 7.74 47.0  
4 -0.247325805 0 6.66 46.0  
5 0.530136475 0 6.55 50.0  
6 -0.531183699 0 6.50 49.0  
7 -0.001708808 0 6.95 54.0  
9 1.893130191 72 3.59 98.0  
10 1.888074655 72 4.31 92.0  
11 2.204674488 72 3.59 95.0  
12 1.191919444 72 6.25 47.0  
13 0.208138896 72 5.91 54.0  
14 1.372565847 72 5.97 56.0  
16 0.635357475 168 2.95 80.0  
17 1.867030718 168 2.67 115.0  
18 1.754101711 168 1.49 94.0  
19 1.475048721 168 4.59 51.0  
20 0.941931427 168 3.59 53.0  
21 1.128864442 168 3.83 56.0  
23 0.142190685 336 4.81 51.5  
24 1.049017122 336 4.23 67.0  
25 0.620266102 336 5.07 57.0  
26 1.045839281 336 5.99 56.0  
27 0.040301445 336 5.51 58.0  
28 1.314392897 336 5.31 54.0  
30 -0.638043585 504 5.19 52.0  
31 -0.221389765 504 4.12 62.0  
32 -1.008071439 504 5.78 58.0  
33 -0.524253460 504 5.77 52.5  
34 0.326363850 504 4.50 59.0  
35 0.685523626 504 4.26 49.0  
37 -2.424921471 720 6.06 61.0  
38 -2.918827852 720 5.35 59.0  
39 -3.443895110 720 6.51 52.5  
40 -3.129569890 720 7.27 51.0  
41 -2.586367356 720 5.17 55.0  
42 -3.370114518 720 10.10 36.5  
 Dissolved oxygen (mg/L) Total Organic Carbon (mg/L)  
2 6.30 17.20  
3 7.90 17.00  
4 9.20 17.40  
5 5.85 17.60  
6 7.20 17.20  
7 8.90 17.50  
9 2.39 17.20  
10 4.50 18.50  
11 2.74 17.50  
12 5.96 24.80  
13 6.26 18.70  
14 6.30 21.20  
16 4.44 15.00  
17 2.32 15.70  
18 2.45 15.10  
19 3.22 19.60  
20 3.24 17.90  
21 3.06 18.10  
23 3.19 17.20  
24 3.33 18.50  
25 2.73 17.50  
26 4.85 24.80  
27 4.70 18.70  
28 2.89 21.20  
30 4.79 16.30  
31 3.53 16.70  
32 3.70 16.70  
33 4.34 18.30  
34 3.91 17.60  
35 3.80 19.10  
37 3.54 10.50  
38 4.03 9.78  
39 3.64 10.80  
40 6.30 12.10  
41 4.73 11.40  
42 4.20 14.60  
 Dissolved Organic Carbon (mg/L)  
2 13.80  
3 14.00  
4 15.90  
5 17.60  
6 13.40  
7 16.50  
9 15.60  
10 17.00  
11 17.30  
12 16.60  
13 15.00  
14 18.80  
16 13.60  
17 14.60  
18 15.00  
19 18.70  
20 15.90  
21 16.60  
23 15.60  
24 17.00  
25 17.30  
26 16.60  
27 15.00  
28 18.80  
30 15.00  
31 14.00  
32 12.70  
33 14.70  
34 16.90  
35 18.20  
37 11.90  
38 9.55  
39 8.22  
40 11.00  
41 10.50  
42 12.10

**Figure 5 - List of variables characterizing the dimensions of the analysis.**

res.hcpc$desc.var

Link between the cluster variable and the quantitative variables  
================================================================  
 Eta2 P-value  
Time 0.8850190 4.070333e-15  
Transparency.(cm) 0.8638687 5.998404e-14  
Dissolved.oxygen..(mg/L) 0.7272269 3.739921e-09  
Turbidity.(uT) 0.6535286 1.634644e-07  
Total.Organic.Carbon.(mg/L) 0.6287742 4.845121e-07  
Dissolved.Organic.Carbon.(mg/L) 0.6195027 7.141389e-07  
  
Description of each cluster by quantitative variables  
=====================================================  
$`1`  
 v.test Mean in category  
Time 4.422978 720.000000  
Turbidity.(uT) 2.397825 6.743333  
Total.Organic.Carbon.(mg/L) -4.477431 11.530000  
Dissolved.Organic.Carbon.(mg/L) -4.617272 10.545000  
 Overall mean sd in category  
Time 300.000000 0.000000  
Turbidity.(uT) 5.301389 1.657444  
Total.Organic.Carbon.(mg/L) 17.082778 1.550430  
Dissolved.Organic.Carbon.(mg/L) 15.026944 1.345285  
 Overall sd p.value  
Time 251.236940 9.734949e-06  
Turbidity.(uT) 1.591036 1.649272e-02  
Total.Organic.Carbon.(mg/L) 3.281183 7.554657e-06  
Dissolved.Organic.Carbon.(mg/L) 2.568207 3.888179e-06  
  
$`2`  
 v.test Mean in category Overall mean  
Dissolved.oxygen..(mg/L) 4.892435 7.096667 4.567500  
Turbidity.(uT) 2.749705 6.582222 5.301389  
Transparency.(cm) -2.047178 50.222222 60.472222  
Time -3.752312 24.000000 300.000000  
 sd in category Overall sd p.value  
Dissolved.oxygen..(mg/L) 1.2086540 1.765736 9.959624e-07  
Turbidity.(uT) 0.5209986 1.591036 5.964902e-03  
Transparency.(cm) 3.3920750 17.101797 4.064065e-02  
Time 33.9411255 251.236940 1.752110e-04  
  
$`3`  
 v.test Mean in category  
Dissolved.Organic.Carbon.(mg/L) 2.283802 16.200000  
Total.Organic.Carbon.(mg/L) 2.230734 18.546667  
Dissolved.oxygen..(mg/L) -2.498014 3.685333  
 Overall mean sd in category  
Dissolved.Organic.Carbon.(mg/L) 15.02694 1.6800794  
Total.Organic.Carbon.(mg/L) 17.08278 2.0645796  
Dissolved.oxygen..(mg/L) 4.56750 0.6755528  
 Overall sd p.value  
Dissolved.Organic.Carbon.(mg/L) 2.568207 0.02238314  
Total.Organic.Carbon.(mg/L) 3.281183 0.02569878  
Dissolved.oxygen..(mg/L) 1.765736 0.01248913  
  
$`4`  
 v.test Mean in category Overall mean  
Transparency.(cm) 5.444793 95.66667 60.472222  
Dissolved.oxygen..(mg/L) -2.138944 3.14000 4.567500  
Turbidity.(uT) -3.660714 3.10000 5.301389  
 sd in category Overall sd p.value  
Transparency.(cm) 10.3387083 17.101797 5.186559e-08  
Dissolved.oxygen..(mg/L) 0.9496140 1.765736 3.244024e-02  
Turbidity.(uT) 0.8891756 1.591036 2.515133e-04

**Figure 6 - List of variables characterizing the clusters of the classification.**