Analyse en composantes principales
(TP Splus)

Analyse en composantes principales des données Forsythia

54 variétés de Forsythia ont été mesurées pour 9 caractères morphologiques floraux :

- LP = longueur du pédoncule floral
- Lcal = longueur du calice
- lsep = largeur des sépales
- LTC = longueur totale de la corolle (tube + limbe)
- Pgl = plus grande largeur du limbe
- LTP = longueur du pistil (le forsythia est hétérostylé avec des formes brevi ou longistylées)
- LTE = longueur totale de l'étamine
- LF = longueur du filet
- Del = différence de hauteur entre le stigmate et les anthères
  Le signe négatif indique que la variété est brévistylée (stigmate sous les anthères)

Le but de cette analyse est d'essayer d'étudier la structuration de la diversité des différentes variétés étudiées.

Sous Unix :

Unix> vi forsy (visualisation du fichier de données forsy avec l'éditeur vi)

```
       LP Lcal lsep   LTC   PGl  LTP  LTE   LF   Del
 C18  6.30 4.45 1.75 21.10  6.50 3.80 5.70 4.10 -3.15
 C17  6.95 6.75 2.53 18.30  7.95 3.30 6.70 4.30 -4.20
 C16  7.35 6.70 2.42 16.95  7.15 3.30 7.00 4.70 -4.45
 C27  8.55 6.20 3.26 26.35  9.30 4.40 6.30 4.25 -3.45
....................................................
....................................................
 C51 10.10 7.45 3.01 25.55 11.25 8.65 4.20 1.55  2.95
 C52 12.80 8.05 2.42 17.55  7.70 4.40 6.80 4.20 -4.25
 C53  7.65 5.10 2.57 22.10  6.55 8.00 3.45 1.45  3.55
 C54  4.40 4.50 2.75 13.45  5.10 6.50 2.75 1.15  2.55
 C56 11.50 7.60 3.87 29.90 12.60 5.40 5.05 2.00 -1.20
```

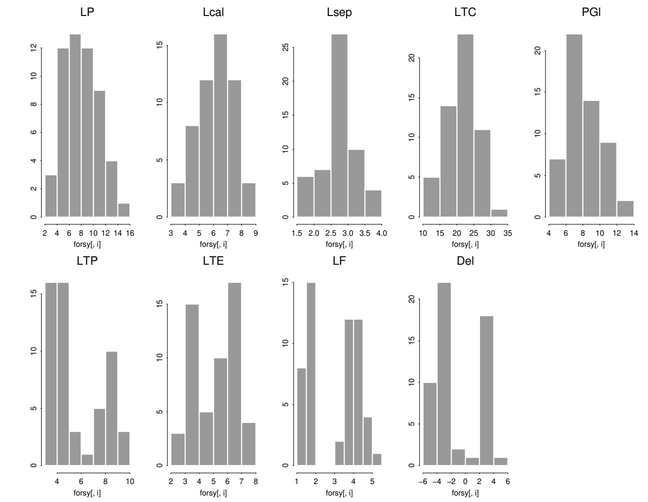
Passer sous Splus :

Récupération du fichier de données sous forme de dataframe

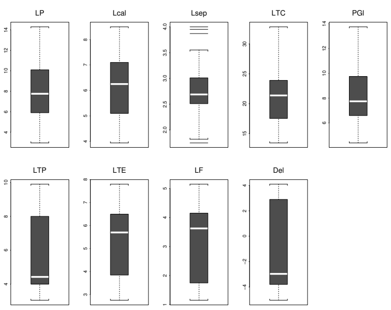
```
>forsy_read.table("forsy",header =T) #on rapatrie le fichier
 et les noms de lignes et de colonnes ; on l'appelle forsy

>forsy[1:10,] #visualisation des 10 premières lignes du fichier
               de données forsy
```

Représentation graphique des données de forsy

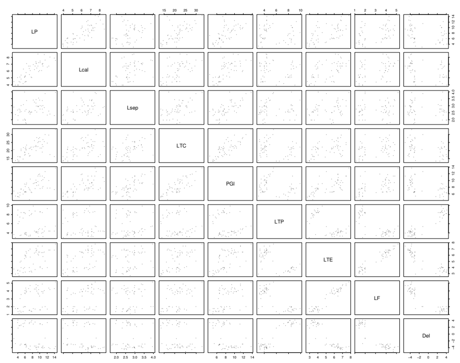
```
>motif() #ouverture d'une fenêtre graphique
>par(mfrow=c(2,5)) #préparation de la fenêtre graphique :
                    division en 10 (2 x 5)
>for (i in 1:9){hist(forsy[,i],main = names(forsy[i]))} #boucle
 pour sortir les histogrammes de toutes les variables
```




```
> par(mfrow=c(2,5))
> for (i in 1:9){boxplot(forsy[,i],main=names(forsy[i]))}
  #représentation en boxplot
```




```
> pairs(forsy)#représentation des biplots 2 à 2
```



Séparation des variétés brévi et longistylées

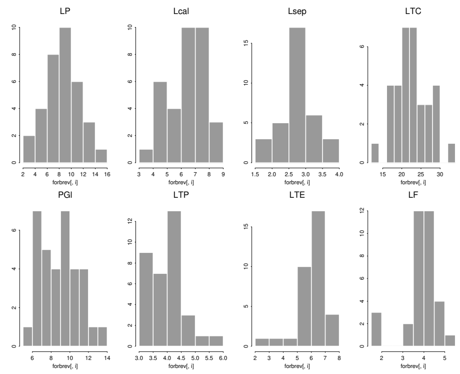
```
> forsyor_forsy[order(forsy$Del),] #on ordonne forsy par la
                                   variable Del
> forsyor[1:8,]
        LP Lcal lsep   LTC   PGl  LTP  LTE   LF   Del
C11   9.50 6.70 2.99 25.10 11.35 4.00 7.45 4.70 -5.05
C37   9.20 6.45 2.55 28.05  8.50 4.05 6.80 4.00 -4.50
C16   7.35 6.70 2.42 16.95  7.15 3.30 7.00 4.70 -4.45
C52  12.80 8.05 2.42 17.55  7.70 4.40 6.80 4.20 -4.25
C17   6.95 6.75 2.53 18.30  7.95 3.30 6.70 4.30 -4.20
C10  14.00 7.85 2.30 16.45  7.35 4.45 6.85 4.30 -4.20
 C4  10.90 6.70 3.00 22.55 10.35 4.05 6.40 3.95 -4.10
C35   9.10 7.10 2.97 22.45  9.45 4.30 6.75 4.45 -4.00
> forsyor[50:54,]
      LP Lcal lsep   LTC  PGl  LTP  LTE   LF  Del
C53 7.65 5.10 2.57 22.10 6.55 8.00 3.45 1.45 3.55
C23 8.75 6.75 2.67 22.50 7.55 8.75 3.70 1.70 3.60
C15 7.85 5.05 2.54 20.75 6.35 8.20 3.30 1.40 4.00
C48 6.05 5.25 2.00 17.55 8.25 8.20 3.30 2.00 4.00
C33 4.05 3.95 2.00 16.35 6.25 9.05 3.60 1.75 4.10
> forbrev_forsyor[forsyor$Del<0,] #création du fichier des brevi
> forlong_forsyor[forsyor$Del>0,] #création du fichier des longi
> forbrev_forbrev[,-9] #élimination de la variable Del
> forlong_forlong[,-9] #élimination de la variable Del
```

ACP sur les brévistylés

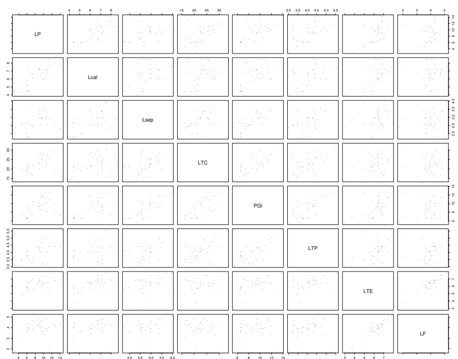
```
> library(multidim,first=T) #on appelle la librairie multidim
> search() #on verifie la position de multidim
[1] ".Data"
[2] "/bao/prive/splus3.2/library/multidim/.Data"
[3] "/bao/prive/splus3.2/splus/.Functions"
[4] "/bao/prive/splus3.2/stat/.Functions"
[5] "/bao/prive/splus3.2/s/.Functions"
[6] "/bao/prive/splus3.2/s/.Datasets"
[7] "/bao/prive/splus3.2/stat/.Datasets"
[8] "/bao/prive/splus3.2/splus/.Datasets"
```


Représentations graphiques préliminaires sur le groupe des brévistylés

```
> par(mfrow=c(2,4)) #on recrée une fenêtre adaptée
> for(i in 1:8){hist(forbrev[,i],main=names(forbrev[i]))}
  #histogrammes des 8 variables
```

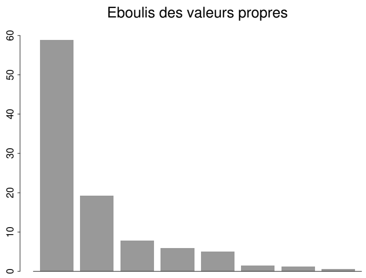



```
> pairs(forbrev)
```



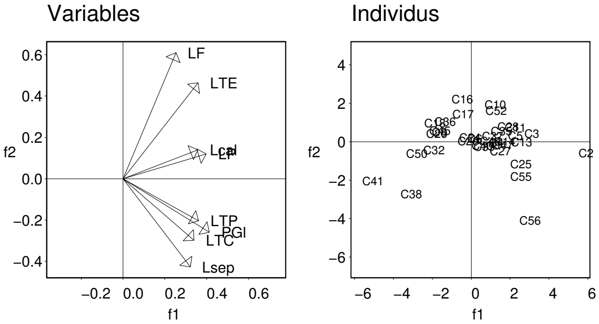
ACP sur le groupe des brévistylés

```
> forbrev.acp_acp(forbrev) #lancement de l'ACP sur les brévi
Contributions des lignes (en millièmes)

C11 C37 C16 C52 C17 C10 C4 C35 C5 C44 C28 C13 C24 C40 C3 C47  C2 C20 C14 C6
 18   9  25  27  15  36  5   6 22   3  17  20   5  12 32  13 119   8   7 17
C39 C27 C50 C18 C8 C29 C45 C36 C32 C55 C25 C41 C38 C56
 11   8  46  34 22  25  22  31  25  31  24 135  76  92

Contributions des colonnes (en millièmes)
  LP Lcal lsep LTC PGl LTP LTE  LF
 125  125  125 125 125 125 125 125

1: On continue ?
2: Arret
Selection: 1
Inertie totale:  8
> prexpl(forbrev.acp)
Pourcentage d'inertie expliquée
            f1 f2 f3 f4 f5 tot
inertie exp 59 19  8  6  5  97
inertie cum 59 78 86 92 97  97
> par(mfrow=c(1,1)) #on recadre la fenêtre graphique sur
                     toute la page
> eboulis(forbrev.acp) #éboulis des valeurs propres
                        correspondant aux axes
```

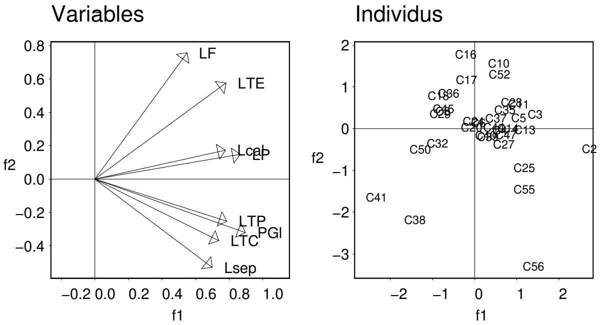



```
> contri(forbrev.acp)

Cosinus carré individus axes principaux

     f1  f2  f3  f4  f5 Axes >5  TOT
C11 608 110  91  85  34      72 1000
C37 125  42  79 160 426     168 1000
C16 151 724  11 110   0       5 1000
C52  72 359 384  91  91       4 1000
C17 223 510  29 206   8      25 1000
C10  48 385 314 132 102      19 1000
..........
..........

Contribution des variables aux facteurs

       f1   f2   f3   f4   f5  f>5  TOT VARI
  LP  159   14  146  169    3  229  125  232
Lcal  128   20  499   25   20  133  125   44
lsep  104  182    1  395   55  100  125    8
 LTC  115   89  123   62  508   42  125  543
 PGl  170   67   14   39    1  263  125  119
 LTP  131   42   30  262  410   55  125   10
 LTE  129  215   37   19    2  109  125   27
  LF   64  372  149   29    1   70  125   16
     1000 1000 1000 1000 1000 1000 1000 1000
```

```
> biplot(forbrev.acp) #on choisit un biplot isométrique ligne

1: Biplot isometrique ligne ?
2: Biplot isometrique colonne ?
3: Biplot a_l'hollandaise ?
Selection: 1
```

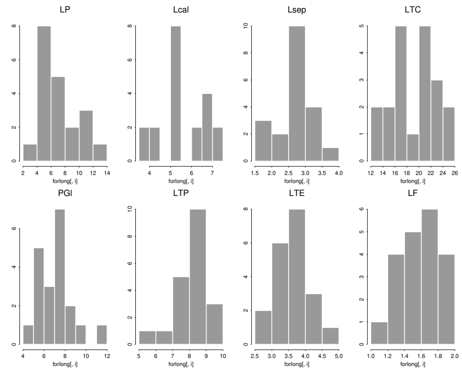



```
> biplot(forbrev.acp)

1: Biplot isometrique ligne ?
2: Biplot isometrique colonne ?
3: Biplot a_l'hollandaise ?
Selection: 2
```

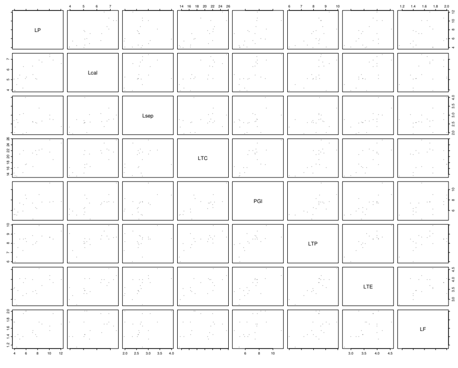


Représentations graphiques préliminaires sur le groupe des longistylés

```
> par(mfrow=c(2,4))
> for (i in 1:8){hist(forlong[,i],main=names(forlong[i]))}
```

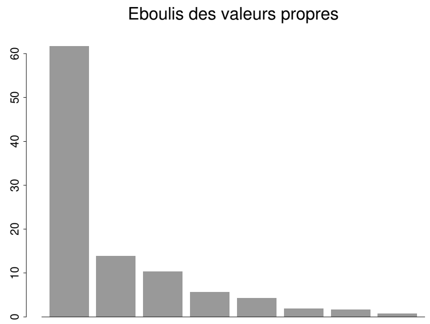



```
> pairs(forlong)
```



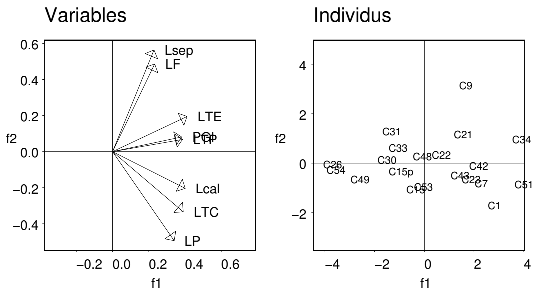
ACP sur les longistylés

```
> forlong.acp_acp(forlong) #lancement de l'ACP sur les longistyl
Contributions des lignes (en millièmes)

C26 C21 C30 C54 C22 C42 C34 C51 C7 C43 C49 C9 C1 C15p C31 C53 C23 C15 C48 C33
111  27  25 116  24  40  91 108 34  14  62 85 66   25  34  13  20  16  41  51

Contributions des colonnes (en millièmes)

  LP Lcal lsep LTC PGl LTP LTE  LF
 125  125  125 125 125 125 125 125

1: On continue ?
2: Arret
Selection: 1
Inertie totale:  8
> prexpl(forlong.acp)
Pourcentage d'inertie expliquée

            f1 f2 f3 f4 f5 tot
inertie exp 62 14 10  6  4  96
inertie cum 62 76 86 92 96  96
> par(mfrow=c(1,1))
> eboulis(forlong.acp)
```

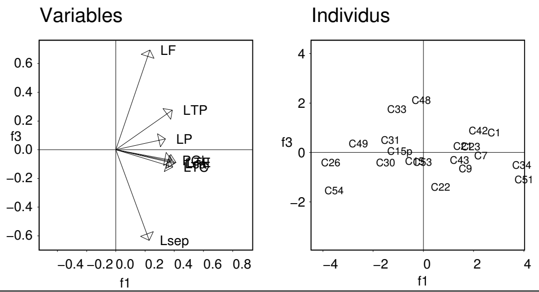



```
> contri(forlong.acp)

 Cosinus carré individus axes principaux

      f1  f2  f3  f4  f5 Axes >5  TOT
 C26 937   0   9  22  15      18 1000
 C21 323 315  16  11  19     316 1000
 C30 900   4  42   7   6      41 1000
 C54 831   4 126  18   0      20 1000
 C22  24  31 508 410   3      23 1000
 C42 514   2 129 306  16      33 1000
 C34 862  65  17  33   8      16 1000
 C51 766  43  69   3 114       6 1000
  C7 747 122   2  55  36      37 1000
 C43 500 104  36 248  18      94 1000
 C49 884  44  14   4  24      30 1000
  C9 147 738  30  27   9      50 1000
  C1 613 274  64   0  11      38 1000
C15p 526  28   1  41 403       1 1000
 C31 526 308  47   0  87      31 1000
 C53  87 450  73  31   2     357 1000
 C23 724 142  18  20  11      86 1000
 C15 207 445  44  25   2     277 1000
 C48  33  11 691  20 220      25 1000
 C33 253  47 385 300   2      12 1000
 TOT 617 139 103  56  43      42 1000
..........
..........

Contribution des variables aux facteurs

       f1   f2   f3   f4   f5  f>5  TOT VARI
  LP  115  243    6  203   26  163  125  245
Lcal  161   42    9  117   56  237  125   49
lsep   52  318  398   63   44   50  125    9
 LTC  153  111   15   48    7  258  125  546
 PGl  146    7    5    1  770   21  125  108
 LTP  150    5   76  313   75   81  125   33
 LTE  169   38    8  118   22  160  125    8
  LF   54  237  483  137    0   31  125    2
     1000 1000 1000 1000 1000 1000 1000 1000
```

```
> biplot(forlong.acp) #biplot isométrique ligne sur plan 1-2

1: Biplot isometrique ligne ?
2: Biplot isometrique colonne ?
3: Biplot a_l'hollandaise ?
Selection: 1
```




```
> biplot(forlong.acp,n1=1,n2=3) #biplot isométrique ligne sur
                                  plan 1-3

1: Biplot isometrique ligne ?
2: Biplot isometrique colonne ?
3: Biplot a_l'hollandaise ?
Selection:  1
```



Ajout d'une variable supplémentaire : la longueur des stomates des variétés

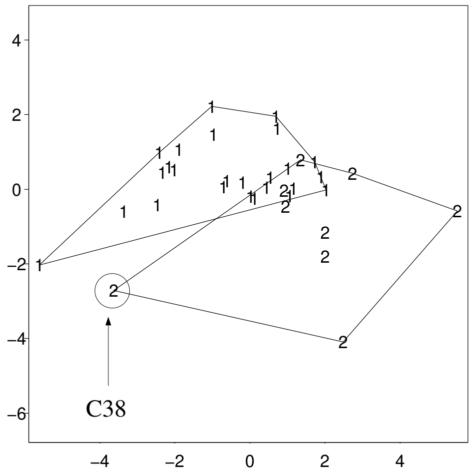
Jeu de données = brévistylées

```
> stomat_c(22.07,29.58,25.42,24.41,23.41,24.98,25.26,26.39,23.87,22.25,
           41.52,24.54,25.05,29.32,33.06,23.67,41.16,25.85,27.52,24.39,
           25.87,34.76,26.18,28.61,23.60,26.19,27.40,29.14,22.76,34.82,
           34.69,29.32,31.38,35.87)

> stomat # vecteur longueur des stomates des différentes variétés
           brévistylées

 [1] 22.07 29.58 25.42 24.41 23.41 24.98 25.26 26.39 23.87 22.25 41.52 24.54
[13] 25.05 29.32 33.06 23.67 41.16 25.85 27.52 24.39 25.87 34.76 26.18 28.61
[25] 23.60 26.19 27.40 29.14 22.76 34.82 34.69 29.32 31.38 35.87


> forbrev2.acp_varsup(forbrev.acp,stomat)
> round(forbrev2.acp$vsup,3)
       f1     f2     f3     f4     f5    f6     f7   f8
vs1 0.172 -0.346 -0.363 -0.018 -0.413 0.198 -1.037 0.92
attr(, "moy"):
[1] 28.06794
attr(, "sigma"):
[1] 5.029709


> corsup(forbrev2.acp)
Correlation variables sup./facteurs et cumuls

      f1  f2  f3  f4  f5 f>6
CorS 139 184  83   0  68 142
CorM 139 323 406 406 474 617
```

Ajout d'une variable illustrative : niveau de ploïdie des variétés

Jeu de données = brévistylées

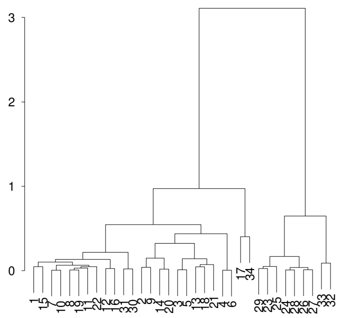
```
> chrom_c(2,2,2,2,2,2,4,2,2,2,4,2,2,2,4,2,4,2,2,2,2,4,2,2,2,2,2,2,2,
         4,4,2,4,4)
  chrom #vecteur représentant le niveau de ploïdie des
         différentes variétés brévistylées
 [1] 2 2 2 2 2 2 4 2 2 2 4 2 2 2 4 2 4 2 2 2 2 4 2 2 2 2 2 2 2 4 4 2 4 4

> chrom2_chrom/2 #pour que les chiffres correspondent exactement au nb
                   de classes on divise par 2 le vecteur précédent
> chrom2
 [1] 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 2 2

> plotcl(forbrev.acp$cmpr[,1:2],chrom2) #on place sur le plan
  1-2 du biplot les valeurs de chaque individu pour la
  variable illustrative
```



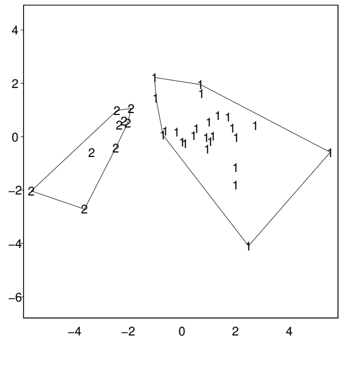
Classification automatique suite à une ACP

Sur le groupe des brévistylés

```
> ddbrev_dist(forbrev.acp$cmpr)
> cahdbrev_cahd(ddbrev)
> plclust(cahdbrev)
```

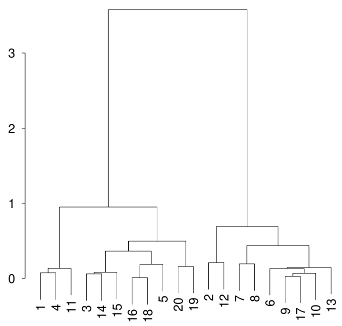


```
> pbrev_cutree(cahdbrev,2)
> plotcl(forbrev.acp$cmpr[,1:2],pbrev)
```

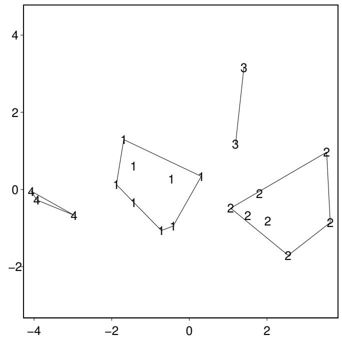


Sur le groupe des longistylés

```
> ddlong_dist(forlong.acp$cmpr)
> cahdlong_cahd(ddlong)
> plclust(cahdlong)
```



```
> plong_cutree(cahdlong,4) #coupure en 4 groupes
> plotcl(forlong.acp$cmpr[,1:2], plong) #représentation sur le
                                          plan 1-2
```



Test sur les résultats de l'ACP du groupe brévistylés de la classification nuées dynamiques ; fonction rkmeans

But : voir les différences de représentations en fonction des itérations
On remarque qu'il y a des groupes assez stables et d'autres beaucoup plus variables

```
> par(mfrow=c(2,2))
> resbrev_rkmeans(forbrev.acp$cmpr,5)

numeros des noyaux
 24 30 32 16 1

> plotcl(forbrev.acp$cmpr[,1:2], resbrev$cluster)
> resbrev_rkmeans(forbrev.acp$cmpr,5)

numeros des noyaux
 8 10 11 28 33

> plotcl(forbrev.acp$cmpr[,1:2], resbrev$cluster)
> resbrev_rkmeans(forbrev.acp$cmpr,5)

numeros des noyaux
 30 6 29 8 16

> plotcl(forbrev.acp$cmpr[,1:2], resbrev$cluster)
> resbrev_rkmeans(forbrev.acp$cmpr,5)

numeros des noyaux
 2 9 1 34 28

> plotcl(forbrev.acp$cmpr[,1:2], resbrev$cluster)
```