

# Classification avec FactoMineR sur les données du cours

*François Husson*

Script et sorties R permettant de retrouver les graphes et sorties du cours. Le jeu de données doit être téléchargé et sauvegardé dans un répertoire connu (dans mon cas, le fichier a été sauvegardé dans le répertoire C:/husson).

```
library(FactoMineR)
```

## Importation des données

```
setwd("C:/users/husson/") # permet de placer la session R où se trouve le jeu de données
temperature <- read.table("AnaDo_JeuDonnees_TemperatFrance.csv",
  header=TRUE, sep=";", dec=".", row.names=1, fileEncoding="latin1")
```

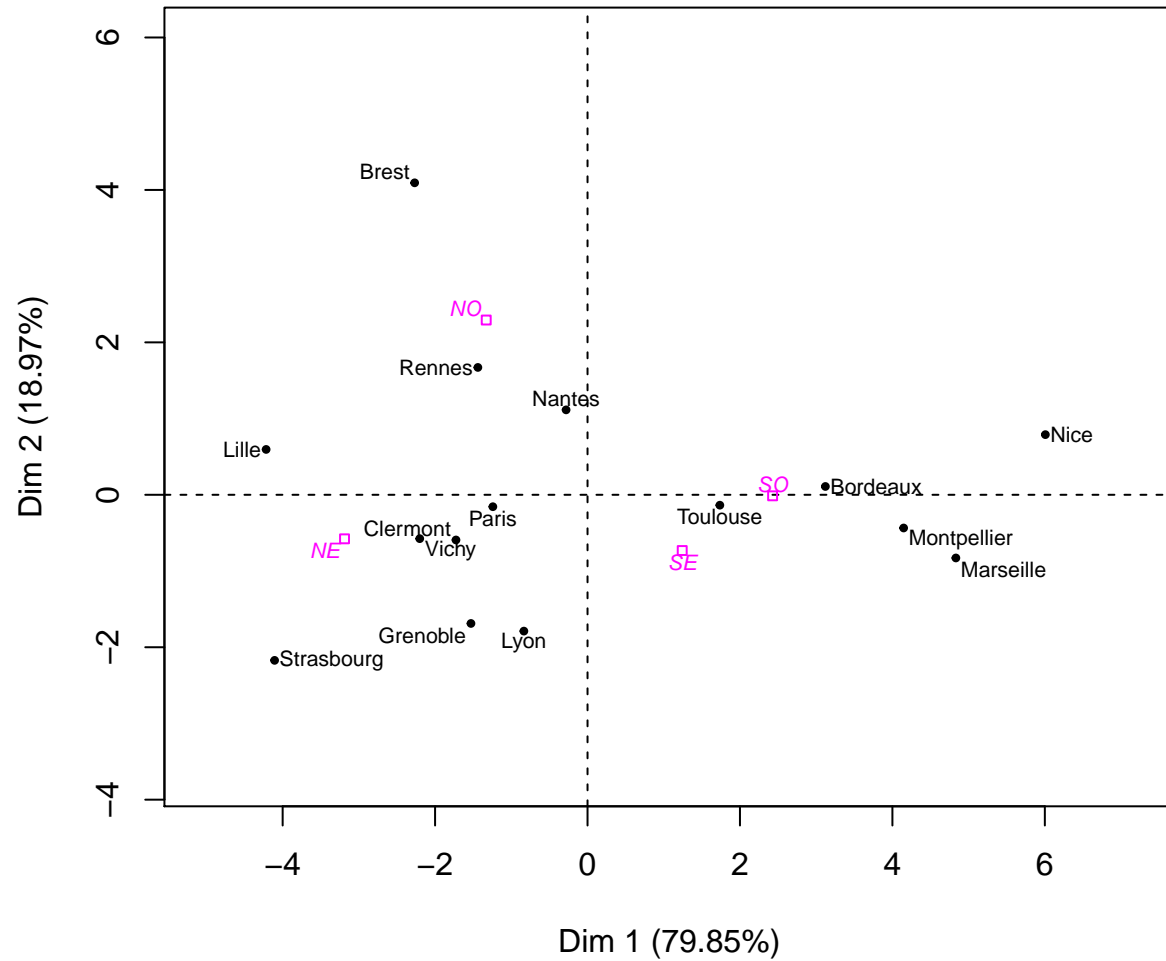
```
summary(temperature)
```

##	Janv	Févr	Mars	Avri	
##	Min. :0.400	Min. :1.500	Min. : 5.600	Min. : 8.90	
##	1st Qu.:2.400	1st Qu.:3.350	1st Qu.: 7.550	1st Qu.:10.00	
##	Median :4.700	Median :5.300	Median : 7.800	Median :10.70	
##	Mean :3.973	Mean :4.833	Mean : 8.233	Mean :10.98	
##	3rd Qu.:5.550	3rd Qu.:6.200	3rd Qu.: 9.550	3rd Qu.:12.20	
##	Max. :7.500	Max. :8.500	Max. :10.800	Max. :13.30	
##	Mai	Juin	juil	Août	
##	Min. :11.60	Min. :14.40	Min. :15.60	Min. :16.00	
##	1st Qu.:13.70	1st Qu.:17.15	1st Qu.:18.90	1st Qu.:18.45	
##	Median :14.30	Median :17.50	Median :19.40	Median :19.10	
##	Mean :14.43	Mean :17.83	Mean :19.83	Mean :19.57	
##	3rd Qu.:15.35	3rd Qu.:19.00	3rd Qu.:20.90	3rd Qu.:20.95	
##	Max. :16.80	Max. :20.80	Max. :23.30	Max. :22.80	
##	Sept	Octo	Nove	Déce	
##	Min. :14.70	Min. : 9.50	Min. : 4.900	Min. :1.300	
##	1st Qu.:15.85	1st Qu.:11.30	1st Qu.: 6.600	1st Qu.:3.450	
##	Median :16.40	Median :11.60	Median : 7.800	Median :5.400	
##	Mean :16.99	Mean :12.32	Mean : 7.927	Mean :4.847	
##	3rd Qu.:18.45	3rd Qu.:13.55	3rd Qu.: 9.050	3rd Qu.:6.350	
##	Max. :20.30	Max. :16.00	Max. :11.500	Max. :8.200	
##	Lati	Long	Moye	Ampl	Région
##	Min. :43.18	Min. :-4.290	Min. : 9.72	Min. :10.20	NE:3
##	1st Qu.:43.96	1st Qu.: 0.460	1st Qu.:10.86	1st Qu.:14.95	NO:3
##	Median :45.47	Median : 3.050	Median :11.18	Median :16.20	SE:7
##	Mean :46.04	Mean : 2.583	Mean :11.81	Mean :15.91	SO:2
##	3rd Qu.:48.15	3rd Qu.: 4.875	3rd Qu.:13.01	3rd Qu.:17.45	
##	Max. :50.38	Max. : 7.450	Max. :14.84	Max. :18.60	

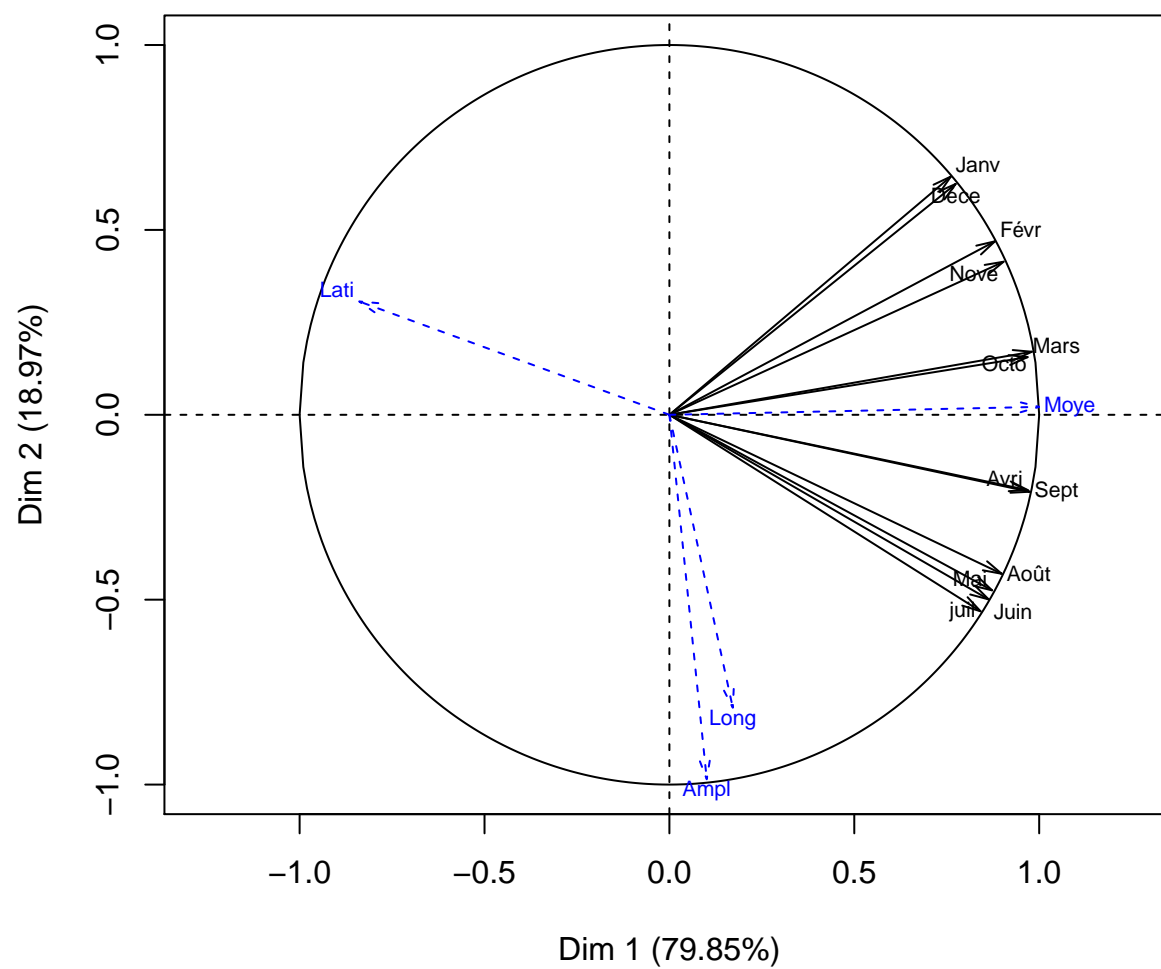
## L'ACP

```
res <- PCA(temperature, quanti.sup=13:16,quali.sup=17)
```

Individuals factor map (PCA)



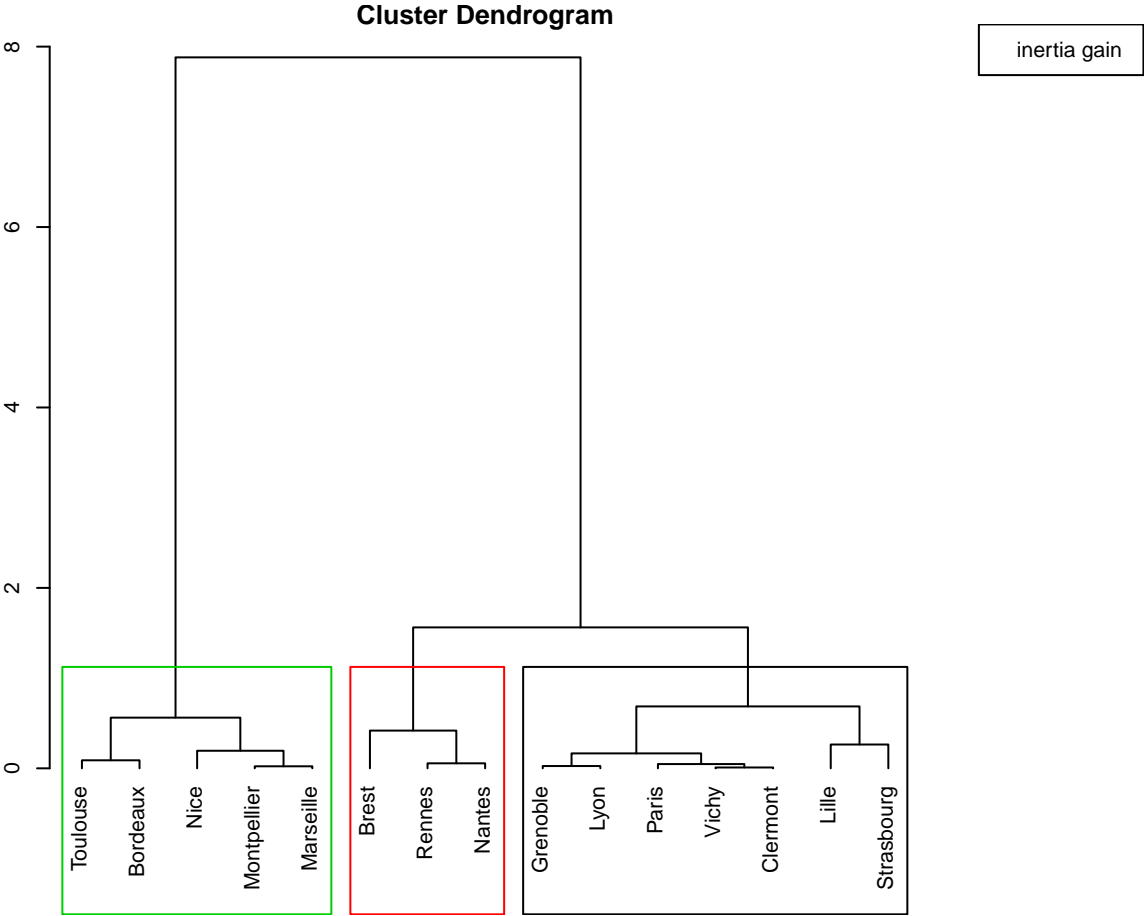
## Variables factor map (PCA)



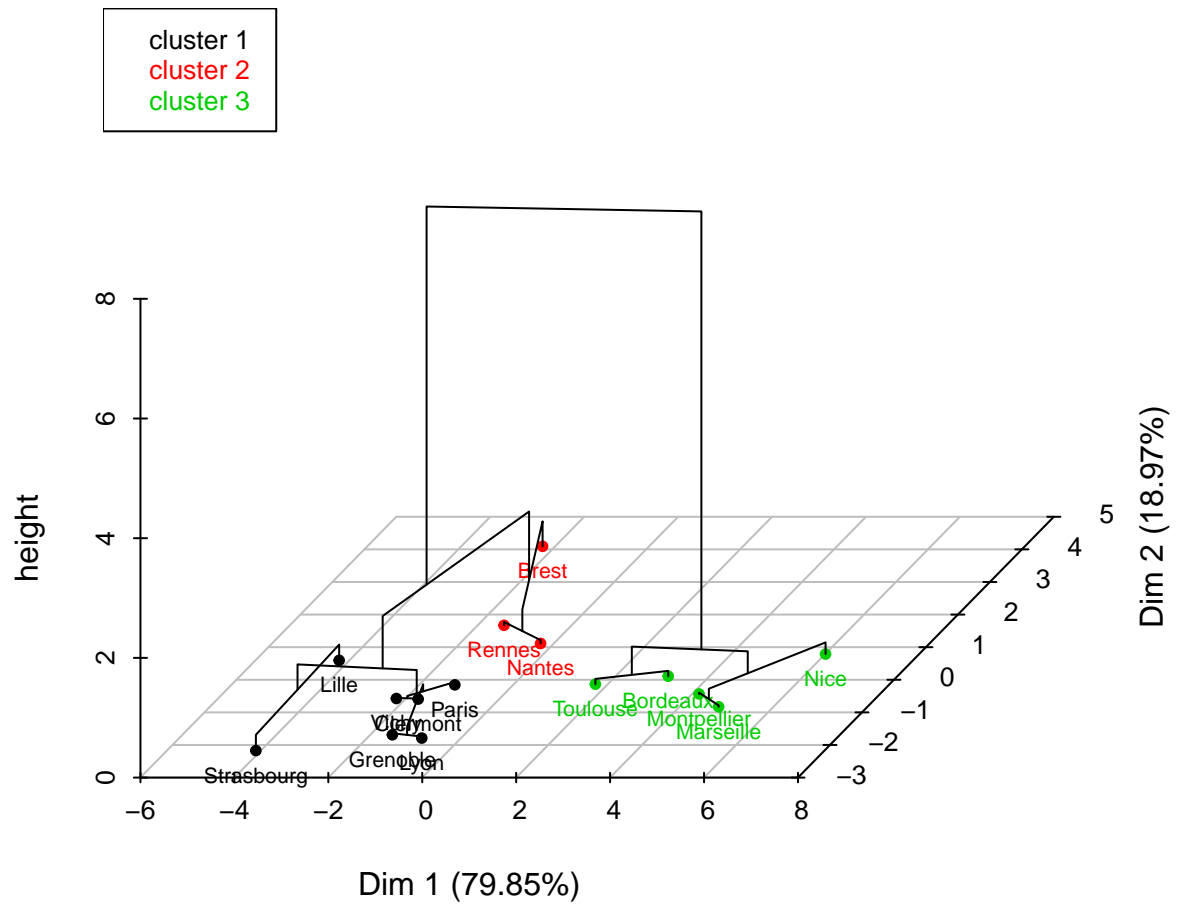
# Construction de la classification ascendante hiérarchique

```
res.hcpc <- HCPC(res)
```

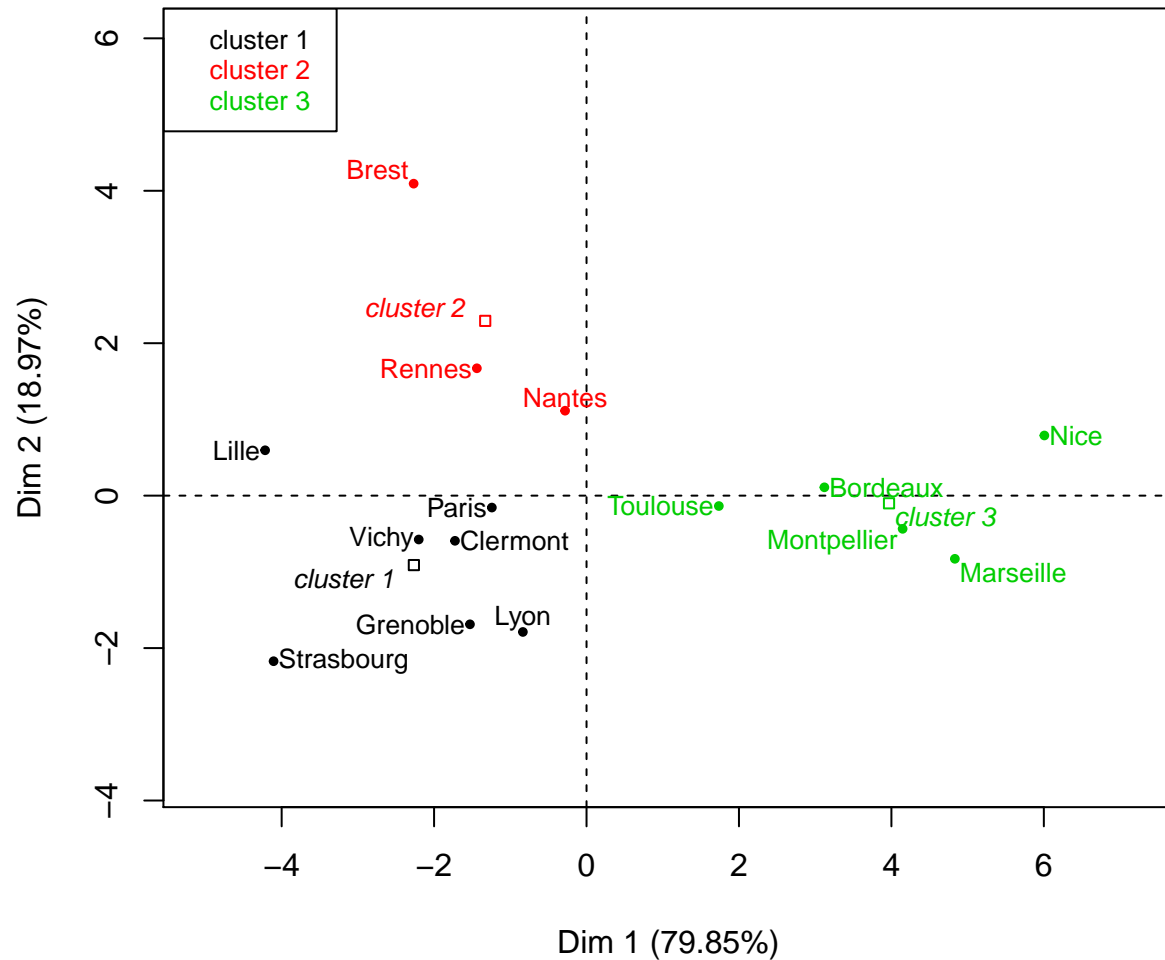
# Hierarchical clustering



## Hierarchical clustering on the factor map



## Factor map



# Résultats de la classification

```
names(res.hcpc)
```

```
## [1] "data.clust" "desc.var" "desc.axes" "call" "desc.ind"
```

```
res.hcpc$data.clust
```

```
##          Janv Févr Mars Avri  Mai  Juin  juil  Août  Sept  Octo  Nove  Déce
## Bordeaux  5.6  6.6 10.3 12.8 15.8 19.3 20.9 21.0 18.6 13.8  9.1  6.2
## Brest     6.1  5.8  7.8  9.2 11.6 14.4 15.6 16.0 14.7 12.0  9.0  7.0
## Clermont  2.6  3.7  7.5 10.3 13.8 17.3 19.4 19.1 16.2 11.2  6.6  3.6
## Grenoble  1.5  3.2  7.7 10.6 14.5 17.8 20.1 19.5 16.7 11.4  6.5  2.3
## Lille     2.4  2.9  6.0  8.9 12.4 15.3 17.1 17.1 14.7 10.4  6.1  3.5
## Lyon      2.1  3.3  7.7 10.9 14.9 18.5 20.7 20.1 16.9 11.4  6.7  3.1
## Marseille 5.5  6.6 10.0 13.0 16.8 20.8 23.3 22.8 19.9 15.0 10.2  6.9
```

```

## Montpellier 5.6 6.7 9.9 12.8 16.2 20.1 22.7 22.3 19.3 14.6 10.0 6.5
## Nantes      5.0 5.3 8.4 10.8 13.9 17.2 18.8 18.6 16.4 12.2 8.2 5.5
## Nice        7.5 8.5 10.8 13.3 16.7 20.1 22.7 22.5 20.3 16.0 11.5 8.2
## Paris       3.4 4.1 7.6 10.7 14.3 17.5 19.1 18.7 16.0 11.4 7.1 4.3
## Rennes     4.8 5.3 7.9 10.1 13.1 16.2 17.9 17.8 15.7 11.6 7.8 5.4
## Strasbourg 0.4 1.5 5.6 9.8 14.0 17.2 19.0 18.3 15.1 9.5 4.9 1.3
## Toulouse   4.7 5.6 9.2 11.6 14.9 18.7 20.9 20.9 18.3 13.3 8.6 5.5
## Vichy       2.4 3.4 7.1 9.9 13.6 17.1 19.3 18.8 16.0 11.0 6.6 3.4
##            Lati Long Moye Ampl Région clust
## Bordeaux   44.50 -0.34 13.33 15.4      SO      3
## Brest       48.24 -4.29 10.77 10.2      NO      2
## Clermont    45.47 3.05 10.94 16.8      SE      1
## Grenoble    45.10 5.43 10.98 18.6      SE      1
## Lille       50.38 3.04 9.73 14.7      NE      1
## Lyon        45.45 4.51 11.36 18.6      SE      1
## Marseille  43.18 5.24 14.23 17.8      SE      3
## Montpellier 43.36 3.53 13.89 17.1      SE      3
## Nantes      47.13 -1.33 11.69 13.8      NO      2
## Nice        43.42 7.15 14.84 15.2      SE      3
## Paris       48.52 2.20 11.18 15.7      NE      1
## Rennes     48.05 -1.41 11.13 13.1      NO      2
## Strasbourg 48.35 7.45 9.72 18.6      NE      1
## Toulouse   43.36 1.26 12.68 16.2      SO      3
## Vichy       46.08 3.26 10.72 16.9      SE      1

```

```
res.hcpc$desc.var
```

```

## $test.chi2
##           p.value df
## Région 0.001700272 6
##
## $category
## $category$`1`
## NULL
##
## $category$`2`
##           Cla/Mod Mod/Cla Global      p.value    v.test
## Région=NO      100      100      20 0.002197802 3.062113
##
## $category$`3`
## NULL
##
##
## $quanti.var
##           Eta2      P-value
## Moye 0.8364881 1.911157e-05
## Octo 0.8362199 1.930043e-05
## Sept 0.8300795 2.406988e-05
## Févr 0.8227293 3.103285e-05
## Mars 0.8126389 4.325897e-05
## Janv 0.8117941 4.444265e-05
## Nove 0.8082999 4.962875e-05
## Avri 0.7928986 7.890384e-05
## Déce 0.7870853 9.316098e-05

```

```

## Août 0.7863781 9.503312e-05
## Juin 0.7241197 4.408838e-04
## Mai 0.7163772 5.205314e-04
## juil 0.7156365 5.287407e-04
## Ampl 0.6464252 1.953833e-03
## Lati 0.6395987 2.191381e-03
## Long 0.6016129 3.997896e-03
##
## $quanti
## $quanti$`1`
##      v.test Mean in category Overall mean sd in category Overall sd
## Sept -2.046173      15.942857      16.986667      0.7384485      1.785447
## Avri -2.107041      10.157143      10.980000      0.6366366      1.366846
## Moye -2.603924      10.661429      11.812667      0.6202863      1.547408
## Octo -2.811187      10.900000      12.320000      0.6611678      1.767937
## Mars -2.854433       7.028571       8.233333      0.8066054      1.477235
## Nove -3.152395       6.357143       7.926667      0.6543419      1.742591
## Févr -3.250132       3.157143       4.833333      0.7631835      1.805055
## Déce -3.283930       3.071429       4.846667      0.9113793      1.892042
## Janv -3.355280       2.114286       3.973333      0.8757504      1.939232
##      p.value
## Sept 0.0407393978
## Avri 0.0351139969
## Moye 0.0092163304
## Octo 0.0049359135
## Mars 0.0043113787
## Nove 0.0016193729
## Févr 0.0011535126
## Déce 0.0010237026
## Janv 0.0007928483
##
## $quanti$`2`
##      v.test Mean in category Overall mean sd in category Overall sd
## Mai -2.016380      12.866667      14.433333      0.9533566      1.453578
## Août -2.021201      17.466667      19.566667      1.0873004      1.943765
## Juin -2.051475      15.933333      17.833333      1.1585431      1.732692
## juil -2.183050      17.433333      19.833333      1.3474255      2.056750
## Long -2.875244      -2.343333       2.583333      1.3768886      3.205624
## Ampl -2.952095      12.366667      15.913333      1.5584893      2.247626
##      p.value
## Mai 0.043760297
## Août 0.043258932
## Juin 0.040220723
## juil 0.029032132
## Long 0.004037156
## Ampl 0.003156256
##
## $quanti$`3`
##      v.test Mean in category Overall mean sd in category Overall sd
## Sept 3.398358      19.280      16.986667      0.7547185      1.785447
## Moye 3.387674      13.794      11.812667      0.7415821      1.547408
## Avri 3.329339      12.700      10.980000      0.5796551      1.366846
## Octo 3.322273      14.540      12.320000      0.9414882      1.767937
## Mars 3.235769      10.040       8.233333      0.5238320      1.477235

```



```

## Août 3.176011 21.900 19.566667 0.7924645 1.943765
## Juin 3.003021 19.800 17.833333 0.7266361 1.732692
## Mai 2.997205 16.080 14.433333 0.6910861 1.453578
## Nove 2.965719 9.880 7.926667 0.9987993 1.742591
## juil 2.915782 22.100 19.833333 1.0039920 2.056750
## Févr 2.882633 6.800 4.833333 0.9402127 1.805055
## Déce 2.535689 6.660 4.846667 0.8957678 1.892042
## Janv 2.464888 5.780 3.973333 0.9239048 1.939232
## Lati -2.953993 43.564 46.039333 0.4748726 2.217038
## p.value
## Sept 0.0006779159
## Moye 0.0007048800
## Avri 0.0008705234
## Octo 0.0008928734
## Mars 0.0012131534
## Août 0.0014931507
## Juin 0.0026731405
## Mai 0.0027246782
## Nove 0.0030197627
## juil 0.0035479803
## Févr 0.0039436601
## Déce 0.0112226447
## Janv 0.0137056069
## Lati 0.0031369098
##
##
## attr("class")
## [1] "catdes" "list "

```

```
res.hcpc$desc.axes
```

```

## $quanti.var
##      Eta2      P-value
## Dim.1 0.8347343 2.037493e-05
## Dim.2 0.6335647 2.420938e-03
##
## $quanti
## $quanti$`1`
##      v.test Mean in category Overall mean sd in category Overall sd
## Dim.2 -2.113402      -0.9110454 3.645232e-16      0.9270734      1.508780
## Dim.1 -2.561180      -2.2651406 6.402286e-16      1.2606742      3.095445
##      p.value
## Dim.2 0.03456641
## Dim.1 0.01043172
##
## $quanti$`2`
##      v.test Mean in category Overall mean sd in category Overall sd
## Dim.2 2.843227      2.292998 3.645232e-16      1.293129      1.50878
##      p.value
## Dim.2 0.004465921
##
## $quanti$`3`
##      v.test Mean in category Overall mean sd in category Overall sd
## Dim.1 3.392217      3.968786 6.402286e-16      1.45853      3.095445

```

```
##           p.value
## Dim.1 0.0006932945
##
##
## attr("class")
## [1] "catdes" "list "
```

```
res.hcpc$desc.ind
```

```
## $para
## Cluster: 1
##      Vichy  Clermont  Grenoble      Paris      Lyon
## 0.4280303 0.6685057 1.1837153 1.3393108 1.6800950
## -----
## Cluster: 2
##      Rennes      Nantes      Brest
## 0.6405312 1.5857132 2.0450969
## -----
## Cluster: 3
## Montpellier      Bordeaux      Marseille      Nice      Toulouse
## 0.4189157 1.1413337 1.1929910 2.2421547 2.2558797
##
## $dist
## Cluster: 1
## Strasbourg      Lyon      Grenoble      Lille      Vichy
## 5.271307 4.112164 4.021064 3.379704 3.019088
## -----
## Cluster: 2
##      Brest      Nantes      Rennes
## 5.005932 2.850147 2.719915
## -----
## Cluster: 3
##      Nice      Marseille  Montpellier      Bordeaux      Toulouse
## 7.497414 6.923052 6.124694 4.997409 3.930113
```