Introduction à R
Délai médecins

Pascal Bessonneau

06/2015

Table des matières

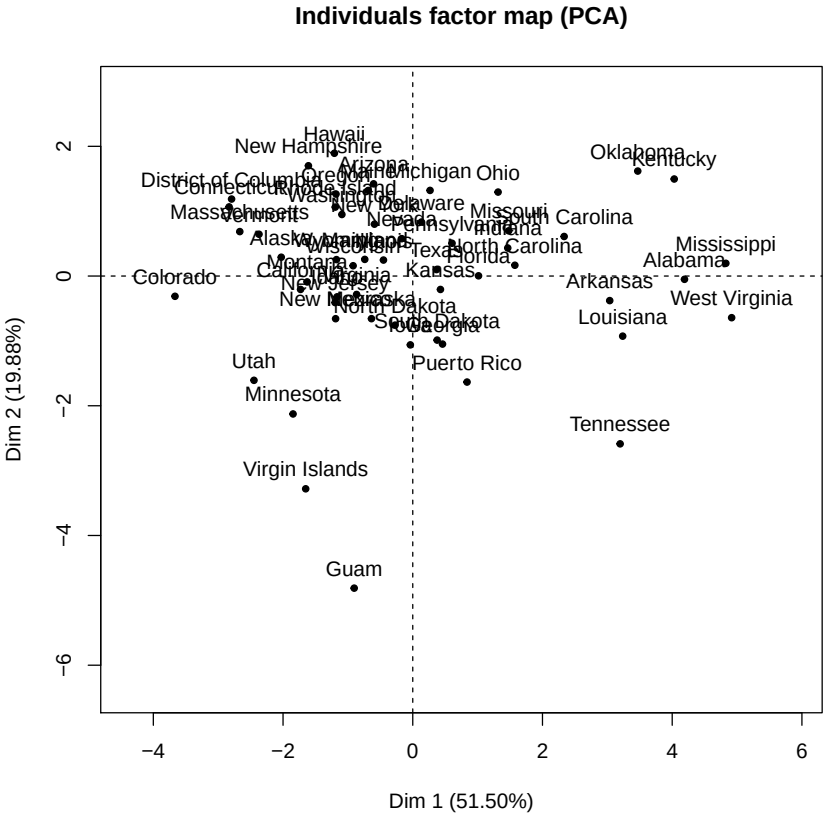
1 Fichier de santé des Etats-Unis

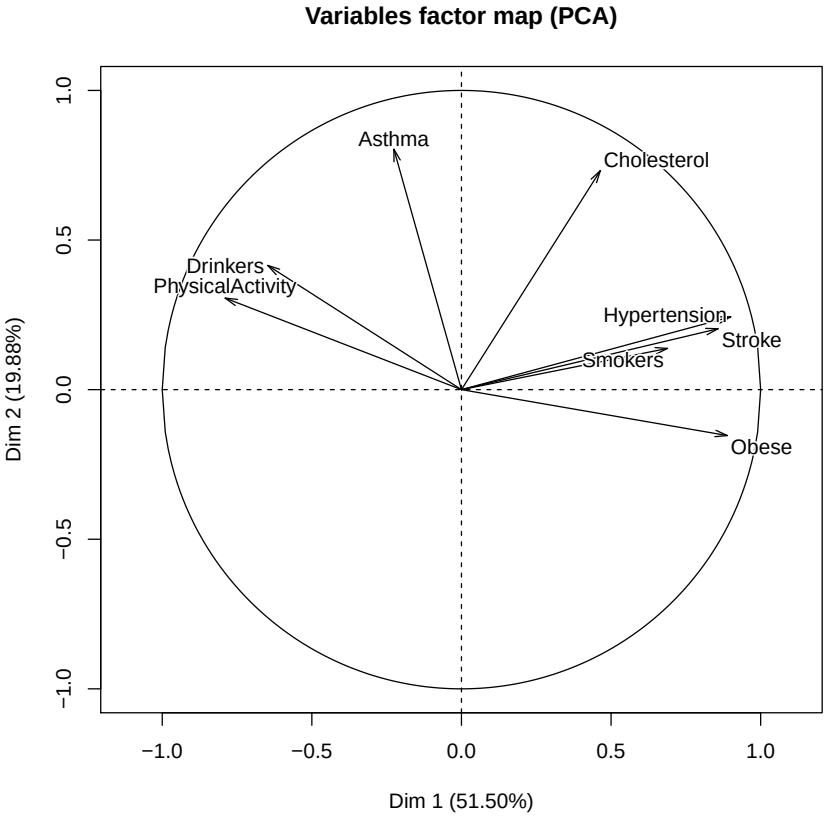
1. Préparer la *data.frame* pour la classification
2. Faire une classification après une ACP avec le fichier *USHealth.csv* avec
   les variables : "Smokers","PhysicalActivity","Obese","Stroke","Asthma","Drinkers","Hypertension","C
3. laisser parler son imagination

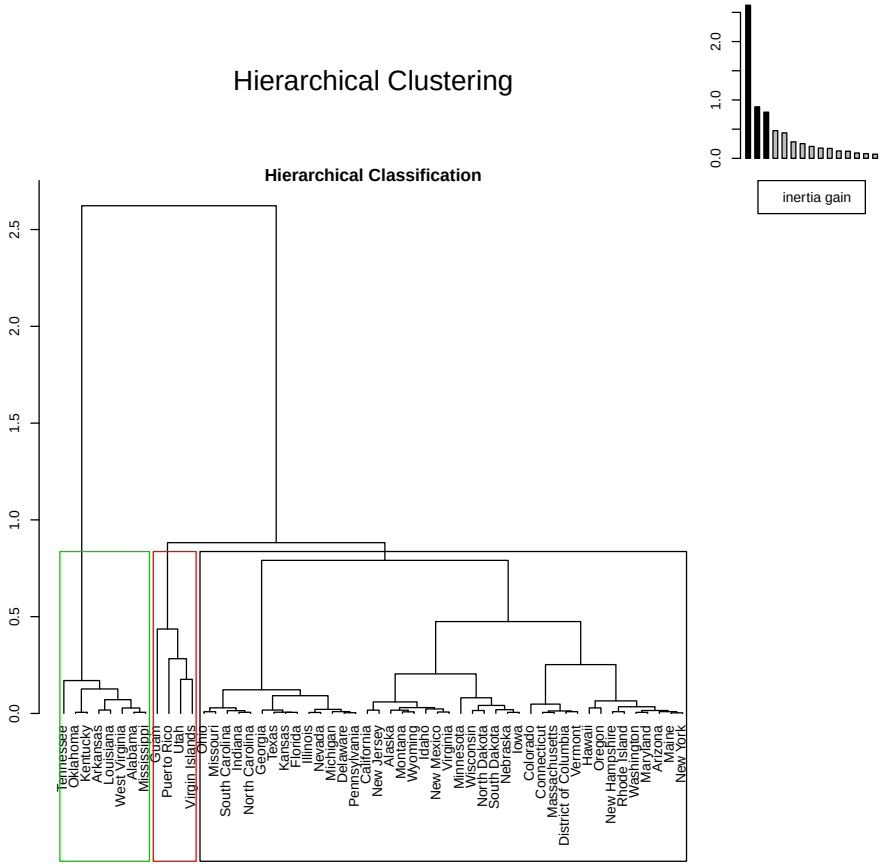
```
> usa <- read.csv2("data/USHealth.csv")
> rownames(usa) <- usa$State.
> usa <- usa[,-1]
> colnames(usa)

##  [1] "Smokers"
##  [2] "Smoke.everyday"
##  [3] "PhysicalActivity"
##  [4] "HighPhysicalActivity"
##  [5] "LimitedActivities"
##  [6] "Obese"
##  [7] "Stroke"
##  [8] "CoronaryDisease"
##  [9] "HeartAttack"
## [10] "Asthma"
## [11] "AsthmaCurrent"
## [12] "Arthritis"
## [13] "BingeDrinkers"
## [14] "HeavyDrinkers"
## [15] "Drinkers"
## [16] "Hypertension"
## [17] "Good.or.Better.Health"
## [18] "Consume.5.or.more.times.per.day"
## [19] "SpecialEquipment"
## [20] "Diabete"
## [21] "Unable.to.work"
## [22] "Cholesterol"
## [23] "White"
## [24] "Black"
## [25] "Hispanic"
## [26] "Other"
## [27] "X18.24.years"
## [28] "X35.44.years"
## [29] "X65..years"
## [30] "NoneChildren"
## [31] "TwoOrMoreChildren"
## [32] "Male"
## [33] "Female"
## [34] "Less15000"
## [35] "Mid1"
## [36] "Mid2"
## [37] "Mid3"
## [38] "More50000"
## [39] "Emplyd"
## [40] "Self.emplyd"
## [41] "No.work"
## [42] "Homemkr"
## [43] "Student"
## [44] "Retired"
## [45] "FluVaccination"
## [46] "HealthCoverage"
## [47] "CholesterolChecked"

> usa1 <- usa[,c("Smokers","PhysicalActivity","Obese","Stroke",
+                "Asthma","Drinkers","Hypertension","Cholesterol")]
> usa1 <- apply(usa1,2,scale)
> rownames(usa1) <- rownames(usa)
```

```
> pca_usa <- PCA(usa1)
> hcpc <- HCPC(pca_usa,nb.clust = 3)
>
> hcpc$desc.var

## $quanti.var
##                       Eta2      P-value
## Stroke           0.6184586 2.134574e-11
## Drinkers         0.5965167 8.882819e-11
## Hypertension     0.5450457 1.897481e-09
## Smokers          0.5423660 2.204027e-09
## Obese            0.5159436 9.222486e-09
## Cholesterol      0.4533757 2.046760e-07
## PhysicalActivity 0.3972829 2.471125e-06
##
## $quanti
## $quanti$`1`
##                v.test Mean in category
## Stroke      -2.143158        -1.031128
## Smokers     -3.022893        -1.454390
## Drinkers    -3.224836        -1.551550
## Cholesterol -4.394967        -2.114530
##              Overall mean sd in category
## Stroke      -2.163907e-16      0.2585800
## Smokers     -1.183146e-16      1.7761304
## Drinkers    -7.581384e-17      0.9019578
## Cholesterol  1.519168e-16      1.3736341
##             Overall sd      p.value
## Stroke       0.9906975 3.210038e-02
## Smokers      0.9906975 2.503711e-03
## Drinkers     0.9906975 1.260447e-03
## Cholesterol  0.9906975 1.107893e-05
##
## $quanti$`2`
##                      v.test Mean in category
## Drinkers           5.531542        0.5102381
## PhysicalActivity   4.489469        0.4141157
## Smokers           -2.762740       -0.2548395
## Stroke            -3.930407       -0.3625469
## Hypertension      -3.988071       -0.3678659
## Obese             -4.659871       -0.4298339
##                    Overall mean
## Drinkers          -7.581384e-17
## PhysicalActivity   1.106368e-16
## Smokers           -1.183146e-16
## Stroke            -2.163907e-16
## Hypertension      -4.741578e-16
## Obese              1.496360e-16
##                   sd in category Overall sd
## Drinkers               0.5384500  0.9906975
## PhysicalActivity       0.6217485  0.9906975
## Smokers                0.4809968  0.9906975
## Stroke                 0.5895187  0.9906975
## Hypertension           0.5829012  0.9906975
## Obese                  0.7633896  0.9906975
##                       p.value
## Drinkers         3.174281e-08
```

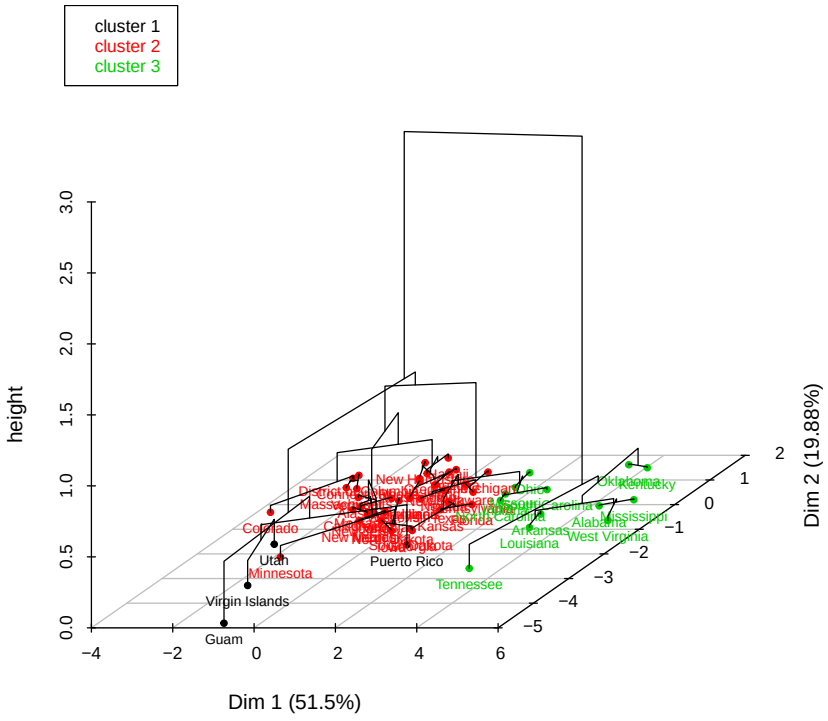
```
## PhysicalActivity 7.140105e-06
## Smokers          5.731832e-03
## Stroke           8.480234e-05
## Hypertension     6.661279e-05
## Obese            3.164074e-06
##
## $quanti$`3`
##                     v.test Mean in category
## Stroke            5.582539        1.3491343
## Hypertension      5.318807        1.2853981
## Obese             5.192017        1.2547565
## Smokers           4.852962        1.1728171
## Cholesterol       2.843245        0.6871281
## Drinkers         -4.033661       -0.9748162
## PhysicalActivity -4.257818       -1.0289884
##                   Overall mean
## Stroke           -2.163907e-16
## Hypertension     -4.741578e-16
## Obese             1.496360e-16
## Smokers          -1.183146e-16
## Cholesterol       1.519168e-16
## Drinkers         -7.581384e-17
## PhysicalActivity  1.106368e-16
##                  sd in category Overall sd
## Stroke                0.7387902  0.9906975
## Hypertension          0.6481763  0.9906975
## Obese                 0.4612090  0.9906975
## Smokers               0.4864159  0.9906975
## Cholesterol           0.7897587  0.9906975
## Drinkers              0.7546344  0.9906975
## PhysicalActivity      0.7345809  0.9906975
##                       p.value
## Stroke           2.370318e-08
## Hypertension     1.044496e-07
## Obese            2.080281e-07
## Smokers          1.216308e-06
## Cholesterol      4.465671e-03
## Drinkers         5.491457e-05
## PhysicalActivity 2.064319e-05
##
##
## attr(,"class")
## [1] "catdes" "list "
```

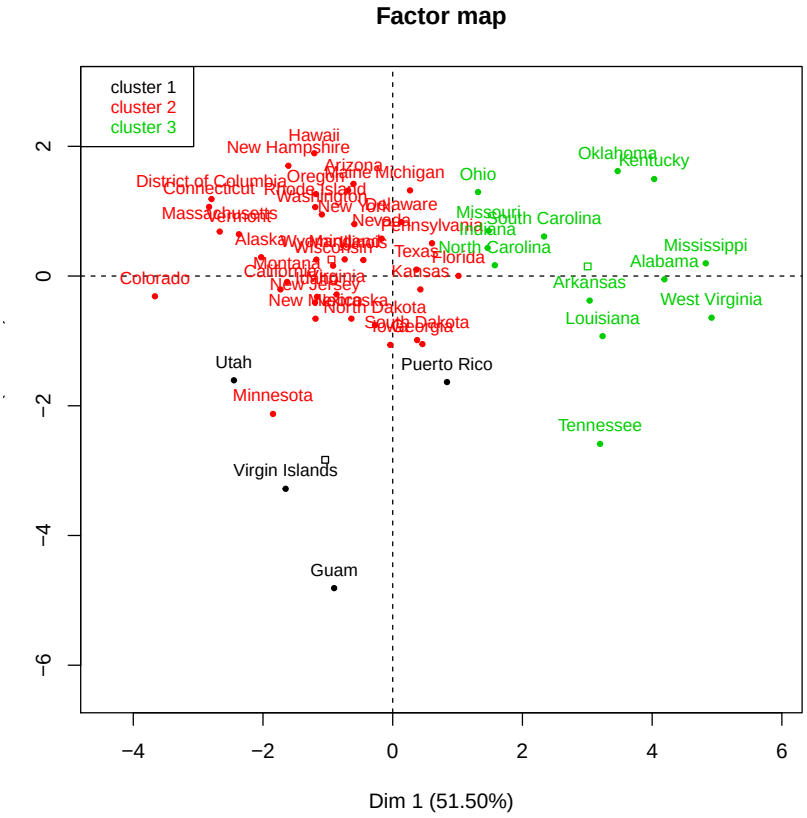

Individuals factor map (PCA)
Dim 2 (19.88%)
Dim 1 (51.50%)
Hawaii
New Hampshire
Arizona
Michigan
Ohio
Oklahoma
Kentucky
Missouri
North Carolina
Indiana
Mississippi
Alabama
Texas
Florida
Kansas
Arkansas
West Virginia
Louisiana
Colorado
Utah
Minnesota
Puerto Rico
Tennessee
Virgin Islands
Guam


Variables factor map (PCA)
Asthma
Cholesterol
Drinkers
PhysicalActivity
Hypertension
Stroke
Smokers
Obese
Dim 2 (19.88%)
Dim 1 (51.50%)
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0


Hierarchical Clustering
Hierarchical Classification
0.0
0.5
1.0
1.5
2.0
2.5
Tennessee
Oklahoma
Kentucky
Arkansas
Louisiana
West Virginia
Alabama
Mississippi
Guam
Puerto Rico
Utah
Virgin Islands
Ohio
Missouri
South Carolina
Indiana
North Carolina
Georgia
Texas
Kansas
Florida
Illinois
Nevada
Michigan
Delaware
Pennsylvania
California
New Jersey
Alaska
Montana
Wyoming
Idaho
New Mexico
Virginia
Minnesota
Wisconsin
North Dakota
South Dakota
Nebraska
Iowa
Colorado
Connecticut
Massachusetts
District of Columbia
Vermont
Hawaii
Oregon
New Hampshire
Rhode Island
Washington
Maryland
Arizona
Maine
New York
0.0
1.0
2.0
inertia gain

Hierarchical clustering on the factor map




Factor map
cluster 1
cluster 2
cluster 3
Hawaii
New Hampshire
Arizona
Oregon
Maine
Michigan
Ohio
Oklahoma
Kentucky
District of Columbia
Connecticut
Rhode Island
Washington
Delaware
New York
Nevada
Missouri
South Carolina
Massachusetts
Vermont
Alaska
Pennsylvania
Indiana
Wisconsin
Texas
North Carolina
Florida
Mississippi
Alabama
Montana
Kansas
Colorado
New Jersey
Arkansas
West Virginia
North Dakota
South Dakota
Louisiana
Utah
Puerto Rico
Minnesota
Tennessee
Virgin Islands
Guam
cluster 1
cluster 2
cluster 3
2
0
−2
−4
−6
−4
−2
0
2
4
6
Dim 1 (51.50%)