Principal Component Analysis (PCA) with FactoMineR (Wine dataset)

Magalie Houée-Bigot & François Husson

Import data

Upload the Expert Wine dataset on your computer.

```r
setwd("C:/users/houee/") # select the current directory
wine <- read.table("data_PCA_ExpertWine.csv",header=TRUE, sep=";", dec=".", row.names=1)
```

`header=TRUE` : indicates that the file contains the names of the variables

`sep=";"` : indicates the fields separator (usually ";" or "," for csv files)

`row.names=1` : indicates the column of the table which contains the row names

It is important to check that the import is well done

```r
summary(wine)
```

```
##        Label        O.fruity        O.passion         O.citrus
##   Sauvignon:5   Min.   :2.143   Min.   :0.5000   Min.   :1.000
##   Vouvray  :5   1st Qu.:4.170   1st Qu.:0.7083   1st Qu.:2.857
##                 Median :4.339   Median :1.6429   Median :3.536
##                 Mean   :4.412   Mean   :1.8667   Mean   :3.662
##                 3rd Qu.:5.143   3rd Qu.:2.9643   3rd Qu.:4.896
##                 Max.   :5.625   Max.   :4.0000   Max.   :5.714
##   O.Intensity.before.shaking O.Intensity.after.shaking   Expression
##   Min.   :3.857              Min.   :4.250               Min.   :4.429
##   1st Qu.:6.062              1st Qu.:6.562               1st Qu.:6.000
##   Median :6.938              Median :7.312               Median :6.440
##   Mean   :6.523              Mean   :6.950               Mean   :6.274
##   3rd Qu.:7.125              3rd Qu.:7.594               3rd Qu.:6.929
##   Max.   :7.750              Max.   :8.000               Max.   :7.143
##   O.candied.fruit    O.vanilla        O.wooded        O.mushroom
##   Min.   :0.6667   Min.   :0.5000   Min.   :0.3333   Min.   :0.2857
##   1st Qu.:1.1250   1st Qu.:0.5417   1st Qu.:0.3750   1st Qu.:0.3333
##   Median :2.9286   Median :0.9881   Median :0.9286   Median :0.8929
##   Mean   :2.5810   Mean   :1.9762   Mean   :2.1357   Mean   :1.1238
##   3rd Qu.:3.5119   3rd Qu.:1.8333   3rd Qu.:1.7500   3rd Qu.:1.6667
##   Max.   :4.7143   Max.   :6.2857   Max.   :7.7500   Max.   :3.0000
##     O.plante        O.flower         O.alcohol        Typicity
##   Min.   :1.857   Min.   :2.833   Min.   :1.667   Min.   :0.625
##   1st Qu.:3.167   1st Qu.:3.393   1st Qu.:2.658   1st Qu.:2.393
##   Median :4.330   Median :3.929   Median :3.238   Median :4.009
##   Mean   :4.217   Mean   :4.487   Mean   :3.095   Mean   :3.913
##   3rd Qu.:4.612   3rd Qu.:5.598   3rd Qu.:3.400   3rd Qu.:5.844
##   Max.   :6.500   Max.   :6.571   Max.   :4.286   Max.   :6.833
##     Grade        Surface.feeling   Freshness        Oxidation
```

```
## Min.  :2.429    Min.  :2.500    Min.  :4.286    Min.  :0.2857
## 1st Qu.:3.857   1st Qu.:2.975   1st Qu.:4.976   1st Qu.:0.4286
## Median :4.143   Median :3.650   Median :5.792   Median :1.0833
## Mean  :4.926    Mean  :3.623    Mean  :5.733    Mean  :1.3929
## 3rd Qu.:6.679   3rd Qu.:4.108   3rd Qu.:6.576   3rd Qu.:1.7679
## Max.  :7.429    Max.  :4.750    Max.  :6.875    Max.  :4.0000
##    Smoothness    Attack.intensity   Sweetness         Acidity
## Min.  :4.143    Min.  :5.143    Min.  :3.000    Min.  :4.375
## 1st Qu.:4.268   1st Qu.:5.464   1st Qu.:3.308   1st Qu.:5.438
## Median :5.214   Median :5.786   Median :3.500   Median :5.795
## Mean  :4.988    Mean  :5.843    Mean  :3.904    Mean  :5.796
## 3rd Qu.:5.500   3rd Qu.:6.357   3rd Qu.:3.844   3rd Qu.:6.312
## Max.  :6.000    Max.  :6.571    Max.  :7.875    Max.  :6.750
##    Bitterness     Astringency    Aroma.intensity Aroma.persistency
## Min.  :2.500    Min.  :1.429    Min.  :4.857    Min.  :5.143
## 1st Qu.:3.812   1st Qu.:2.375   1st Qu.:6.143   1st Qu.:5.714
## Median :4.062   Median :2.750   Median :6.500   Median :6.238
## Mean  :3.975    Mean  :2.851    Mean  :6.460    Mean  :6.100
## 3rd Qu.:4.219   3rd Qu.:3.125   3rd Qu.:7.107   3rd Qu.:6.571
## Max.  :5.000    Max.  :4.250    Max.  :7.286    Max.  :6.714
## Visual.intensity Odor.preference Overall.preference
## Min.  :3.000     Min.  :4.400     Min.  :4.125
## 1st Qu.:3.719    1st Qu.:5.000    1st Qu.:5.000
## Median :4.562    Median :5.214    Median :5.312
## Mean  :4.912     Mean  :5.277     Mean  :5.252
## 3rd Qu.:6.188    3rd Qu.:5.850    3rd Qu.:5.500
## Max.  :7.375     Max.  :6.143     Max.  :6.429
```
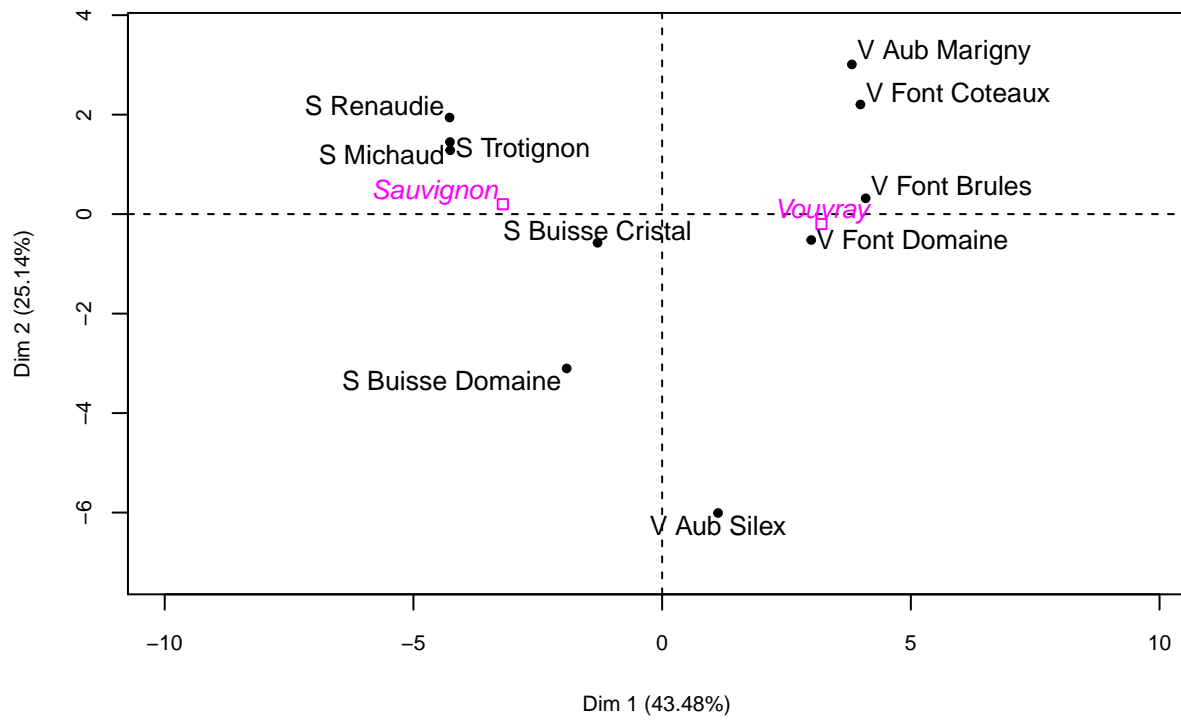
Loading FactoMineR

```
library(FactoMineR)
```

PCA with supplementary variables

```
res <- PCA(wine,quali.sup=1,quanti.sup=29:30,graph=FALSE)
par(cex=0.8,cex.main=1,cex.lab=0.8,cex.axis=0.8) # change graphical parameters (font size)
plot(res,choix="ind") # individuals factor map
```
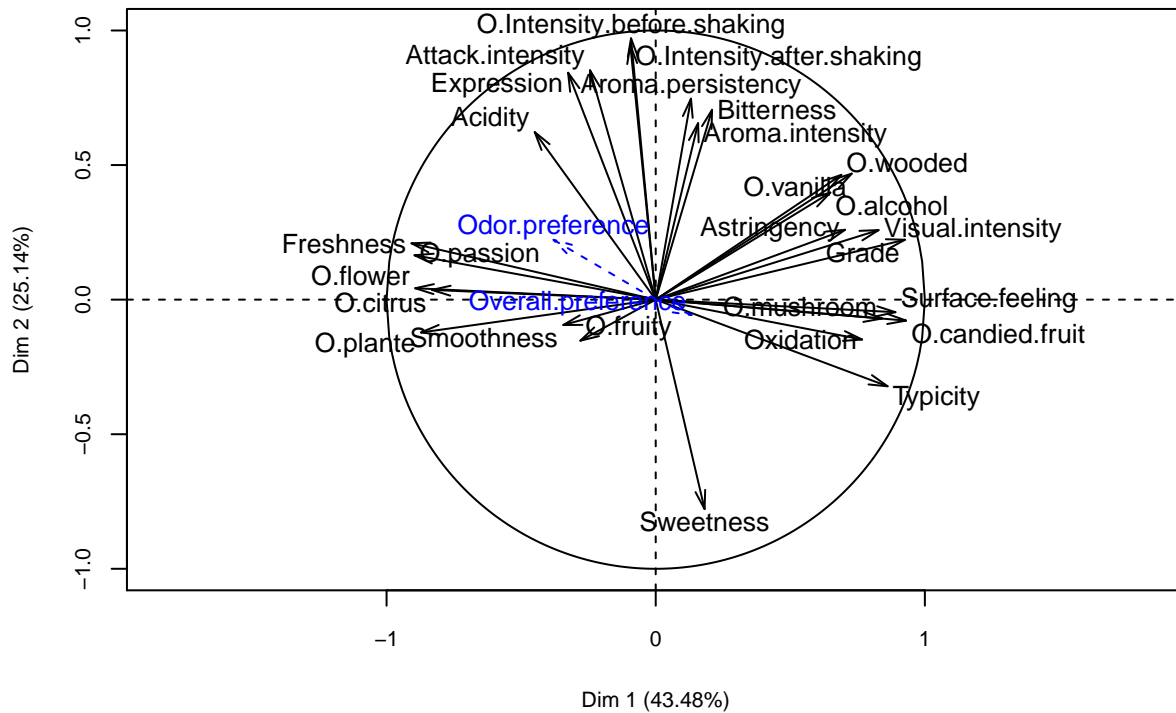
Individuals factor map (PCA)


```
plot(res,choix="var") # variables factor map
```

Variables factor map (PCA)

Summary of outputs

Outputs can be summarized with the function `summary`.

```
summary(res, nbelements=Inf)  ## to print all the elements
```

```
##
## Call:
## PCA(X = wine, quanti.sup = 29:30, quali.sup = 1, graph = FALSE)
##
##
## Eigenvalues
##                         Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
## Variance               11.740   6.787   2.741   2.030   1.612   0.840
## % of var.              43.482  25.137  10.151   7.518   5.970   3.110
## Cumulative % of var.   43.482  68.619  78.770  86.288  92.258  95.367
##
##                         Dim.7   Dim.8   Dim.9
## Variance                0.642   0.366   0.242
## % of var.               2.379   1.356   0.897
## Cumulative % of var.   97.747  99.103 100.000
##
## Individuals
##                           Dist    Dim.1    ctr   cos2    Dim.2    ctr
## S Michaud              |  5.405 |  -4.265 15.492 0.623 |  1.450  3.097
```

4

```
## S Renaudie            |  5.003 | -4.273 15.556   0.730 |  1.942  5.557
## S Trotignon           |  4.812 | -4.260 15.457   0.784 |  1.284  2.429
## S Buisse Domaine       |  4.723 | -1.919  3.137   0.165 | -3.102 14.180
## S Buisse Cristal       |  3.804 | -1.298  1.435   0.116 | -0.576  0.489
## V Aub Silex           |  6.566 |  1.124  1.077   0.029 | -6.009 53.197
## V Aub Marigny         |  5.957 |  3.815 12.394   0.410 |  3.010 13.346
## V Font Domaine        |  4.167 |  2.998  7.655   0.517 | -0.521  0.399
## V Font Brules         |  5.350 |  4.092 14.262   0.585 |  0.318  0.149
## V Font Coteaux        |  5.590 |  3.986 13.536   0.509 |  2.204  7.158
##                         cos2   Dim.3    ctr   cos2
## S Michaud             0.072 |  1.422  7.373   0.069 |
## S Renaudie            0.151 |  0.269  0.264   0.003 |
## S Trotignon           0.071 | -0.301  0.331   0.004 |
## S Buisse Domaine       0.431 | -1.130  4.660   0.057 |
## S Buisse Cristal       0.023 | -0.752  2.063   0.039 |
## V Aub Silex           0.838 | -0.789  2.272   0.014 |
## V Aub Marigny         0.255 | -2.315 19.556   0.151 |
## V Font Domaine        0.016 |  2.189 17.484   0.276 |
## V Font Brules         0.004 |  3.114 35.378   0.339 |
## V Font Coteaux        0.155 | -1.706 10.619   0.093 |
##
## Variables
##                             Dim.1    ctr    cos2    Dim.2    ctr    cos2
## O.fruity                 | -0.280  0.668   0.078 | -0.153  0.343   0.023 |
## O.passion                | -0.896  6.840   0.803 |  0.165  0.401   0.027 |
## O.citrus                 | -0.831  5.889   0.691 |  0.036  0.019   0.001 |
## O.Intensity.before.shaking | -0.092  0.072   0.008 |  0.971 13.887   0.943 |
## O.Intensity.after.shaking | -0.094  0.076   0.009 |  0.945 13.171   0.894 |
## Expression               | -0.325  0.902   0.106 |  0.843 10.472   0.711 |
## O.candied.fruit          |  0.930  7.371   0.865 | -0.078  0.089   0.006 |
## O.vanilla                |  0.728  4.520   0.531 |  0.468  3.230   0.219 |
## O.wooded                 |  0.690  4.052   0.476 |  0.464  3.166   0.215 |
## O.mushroom               |  0.841  6.030   0.708 | -0.070  0.072   0.005 |
## O.plante                 | -0.872  6.476   0.760 | -0.122  0.219   0.015 |
## O.flower                 | -0.894  6.803   0.799 |  0.042  0.027   0.002 |
## O.alcohol                |  0.649  3.585   0.421 |  0.398  2.332   0.158 |
## Typicity                 |  0.862  6.322   0.742 | -0.322  1.527   0.104 |
## Grade                    |  0.926  7.309   0.858 |  0.223  0.729   0.050 |
## Surface.feeling          |  0.891  6.758   0.793 | -0.047  0.032   0.002 |
## Freshness                | -0.907  7.014   0.823 |  0.209  0.646   0.044 |
## Oxidation                |  0.767  5.010   0.588 | -0.148  0.324   0.022 |
## Smoothness               | -0.343  1.005   0.118 | -0.094  0.131   0.009 |
## Attack.intensity         | -0.244  0.508   0.060 |  0.852 10.702   0.726 |
## Sweetness                |  0.182  0.281   0.033 | -0.778  8.919   0.605 |
## Acidity                  | -0.450  1.722   0.202 |  0.623  5.710   0.388 |
## Bitterness               |  0.209  0.373   0.044 |  0.705  7.330   0.498 |
## Astringency              |  0.703  4.215   0.495 |  0.259  0.992   0.067 |
## Aroma.intensity          |  0.158  0.211   0.025 |  0.656  6.333   0.430 |
## Aroma.persistency        |  0.131  0.146   0.017 |  0.747  8.212   0.557 |
## Visual.intensity         |  0.828  5.841   0.686 |  0.258  0.984   0.067 |
##                             Dim.3    ctr    cos2
## O.fruity                  0.575 12.048   0.330 |
## O.passion                -0.091  0.301   0.008 |
## O.citrus                  0.403  5.939   0.163 |
```
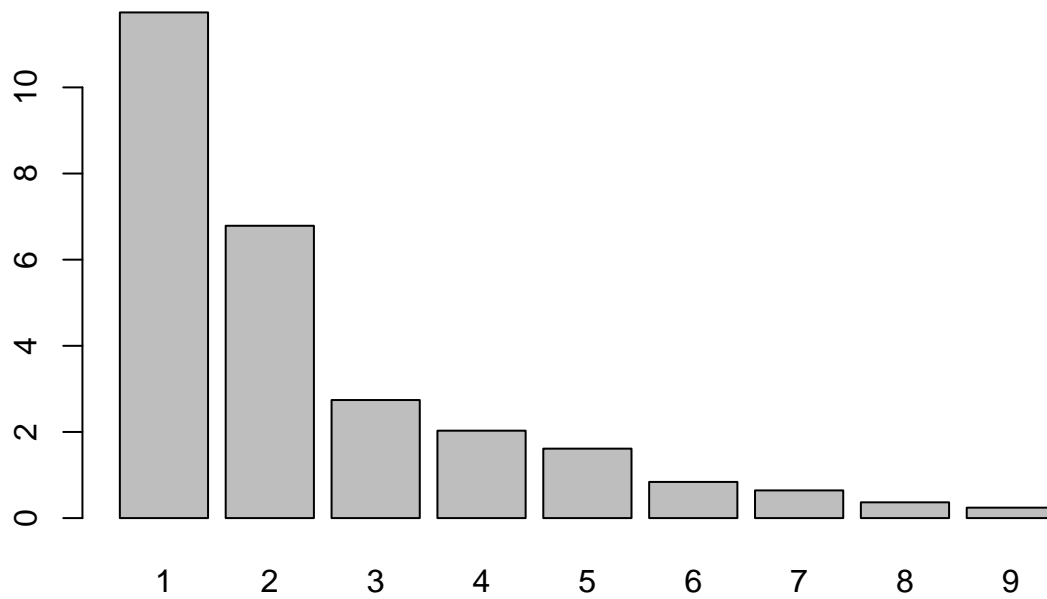
```
## O.Intensity.before.shaking  0.087  0.276  0.008 |
## O.Intensity.after.shaking   0.179  1.165  0.032 |
## Expression                  0.318  3.697  0.101 |
## O.candied.fruit             0.222  1.806  0.049 |
## O.vanilla                  -0.460  7.720  0.212 |
## O.wooded                   -0.519  9.832  0.269 |
## O.mushroom                  0.402  5.907  0.162 |
## O.plante                    0.260  2.469  0.068 |
## O.flower                    0.191  1.332  0.037 |
## O.alcohol                  -0.056  0.116  0.003 |
## Typicity                    0.208  1.580  0.043 |
## Grade                       0.181  1.200  0.033 |
## Surface.feeling             0.026  0.025  0.001 |
## Freshness                  -0.123  0.555  0.015 |
## Oxidation                   0.578 12.203  0.334 |
## Smoothness                 -0.397  5.755  0.158 |
## Attack.intensity           -0.225  1.848  0.051 |
## Sweetness                  -0.180  1.181  0.032 |
## Acidity                    -0.064  0.148  0.004 |
## Bitterness                 -0.077  0.216  0.006 |
## Astringency                -0.530 10.263  0.281 |
## Aroma.intensity             0.405  5.979  0.164 |
## Aroma.persistency           0.182  1.208  0.033 |
## Visual.intensity            0.379  5.232  0.143 |
##
## Supplementary continuous variables
##                        Dim.1   cos2    Dim.2   cos2    Dim.3   cos2
## Odor.preference     | -0.380  0.145 |  0.222  0.049 | -0.496  0.246
## Overall.preference  |  0.134  0.018 | -0.058  0.003 |  0.559  0.312
##
## Odor.preference          |
## Overall.preference       |
##
## Supplementary categories
##                       Dist    Dim.1   cos2 v.test    Dim.2   cos2
## Sauvignon           | 3.237 | -3.203  0.979 -2.804 |  0.200  0.004
## Vouvray             | 3.237 |  3.203  0.979  2.804 | -0.200  0.004
##                      v.test   Dim.3   cos2 v.test
## Sauvignon            0.230 | -0.099  0.001 -0.179 |
## Vouvray             -0.230 |  0.099  0.001  0.179 |
```

Bar chart of eigenvalues

```
barplot(res$eig[,1],main="Eigenvalues",names.arg=1:nrow(res$eig))
```

Eigenvalues

Drawing wines according to the label

```
plot(res,habillage="Label")
```

Individuals factor map (PCA)

Dim 2 (25.14%)

Dim 1 (43.48%)

Sauvignon
Vouvray

S Renaudie
S Michaud
S Trotignon
Sauvignon
S Buisse Cristal
S Buisse Domaine

V Aub Marigny
V Font Coteaux
V Font Brules
Vouvray
V Font Domaine
V Aub Silex