Introdução à Análise de Componentes Principais (PCA)

Daniel Ikenaga - Jupter

Introdução à Análise de Componentes Principais (PCA)

- PCA - Principal Component Analysis
- Carl Pearsons (1901)

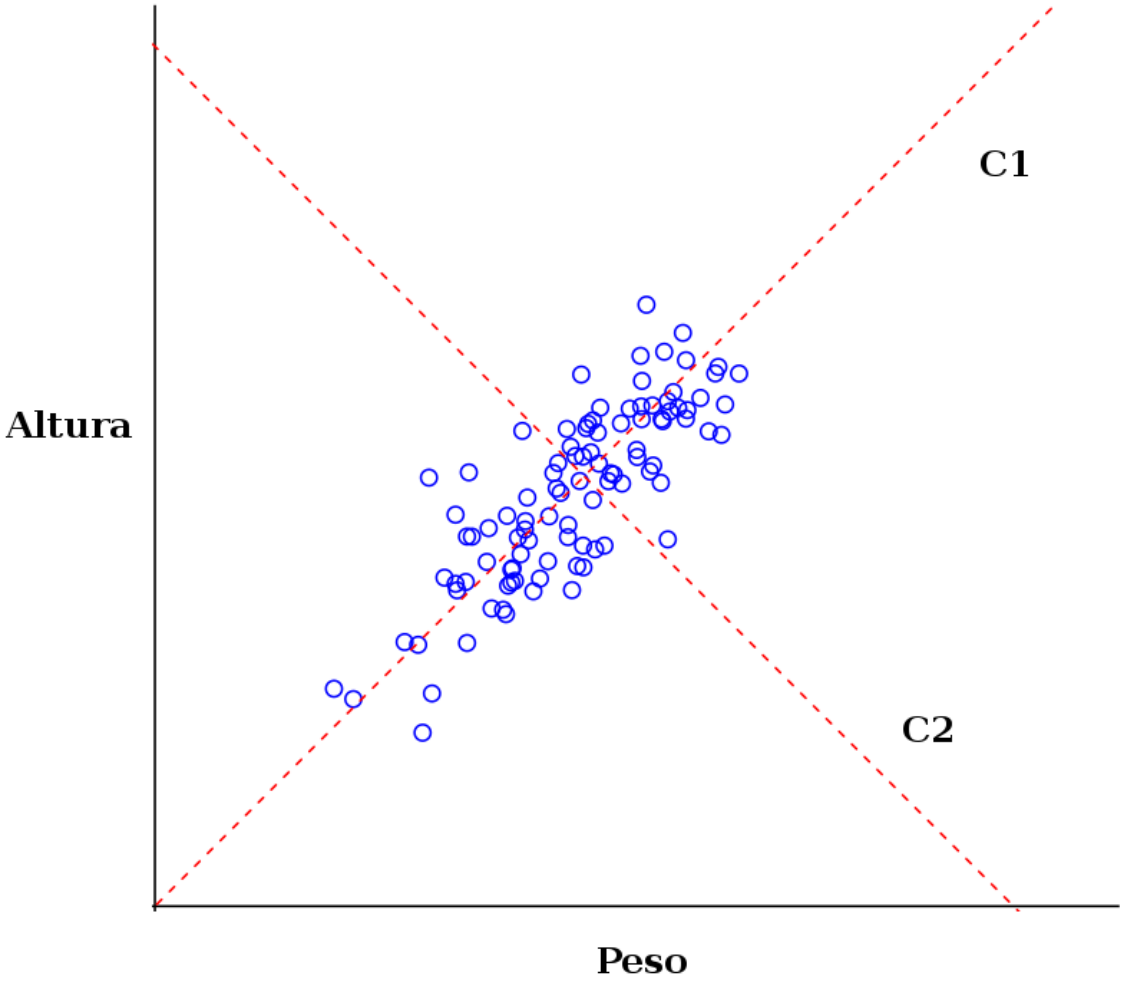
Objetivos

- Reduzir a dimensão dos dados
- Compreender a variância das variáveis
- Produzir escores
- Avaliar variáveis
- Permite agrupar elementos
- Gerar insights

Procedimento

- Criação de novas variáveis (componentes principais)
- Combinações lineares das variáveis originais

Componentes Principais



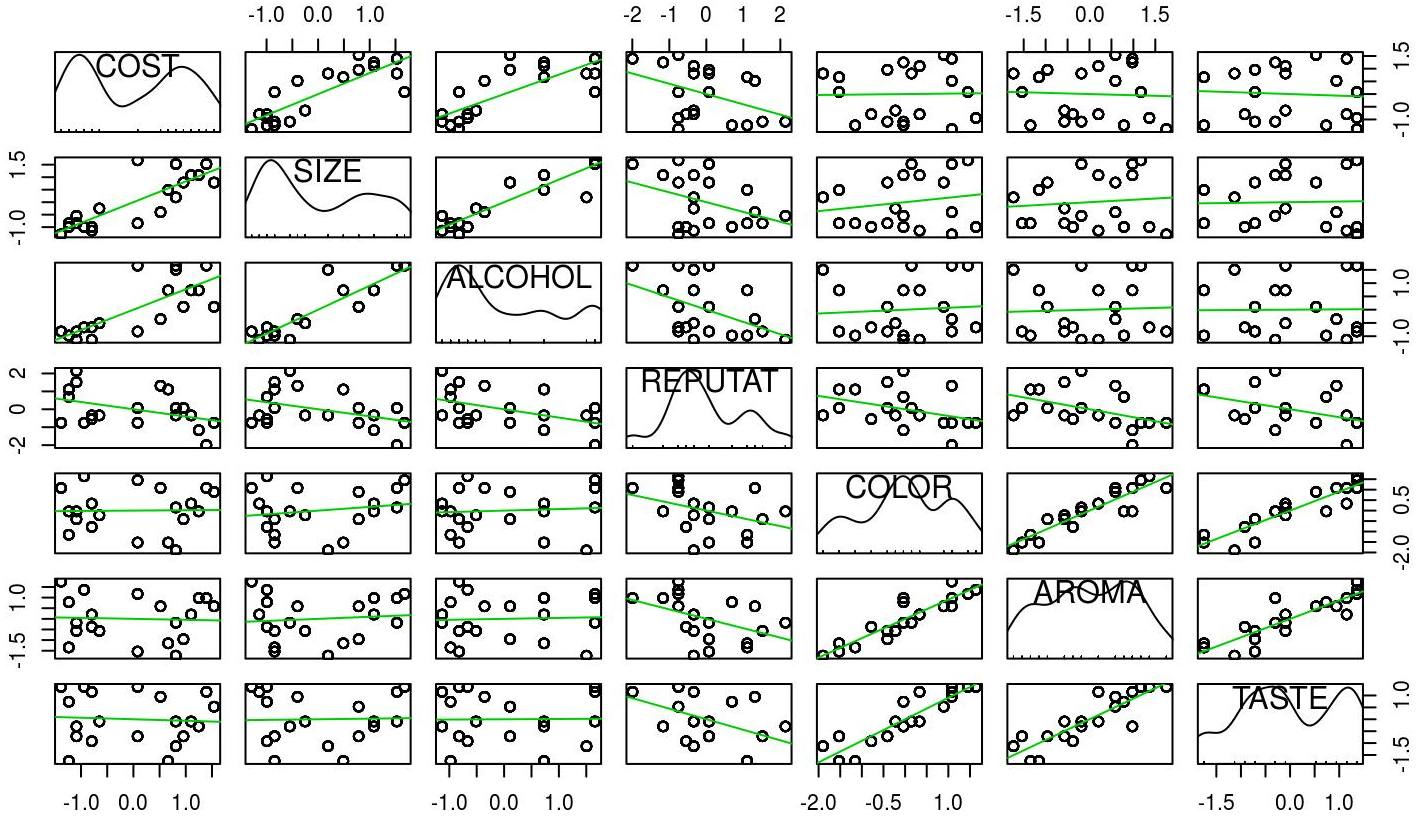
200 avaliações de packs de Cerveja

Summary

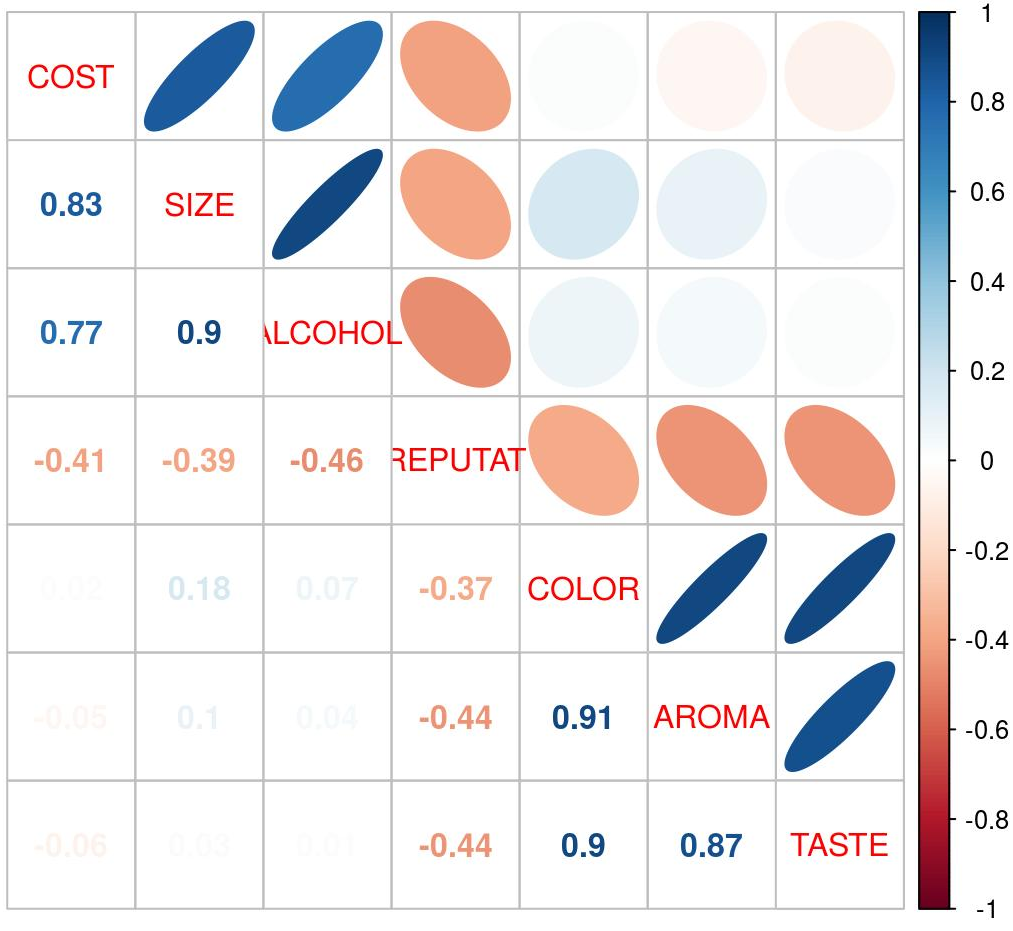
```
##       COST                SIZE               ALCOHOL             REPUTAT
##  Min.   :-1.37900   Min.   :-1.2895   Min.   :-1.1318   Min.   :-1.9934
##  1st Qu.:-0.97770   1st Qu.:-0.8819   1st Qu.:-0.8217   1st Qu.:-0.7540
##  Median : 0.08026   Median :-0.3261   Median :-0.4341   Median :-0.3408
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.84637   3rd Qu.: 0.8597   3rd Qu.: 0.7287   3rd Qu.: 0.7953
##  Max.   : 1.53952   Max.   : 1.6749   Max.   : 1.6590   Max.   : 2.1380
```

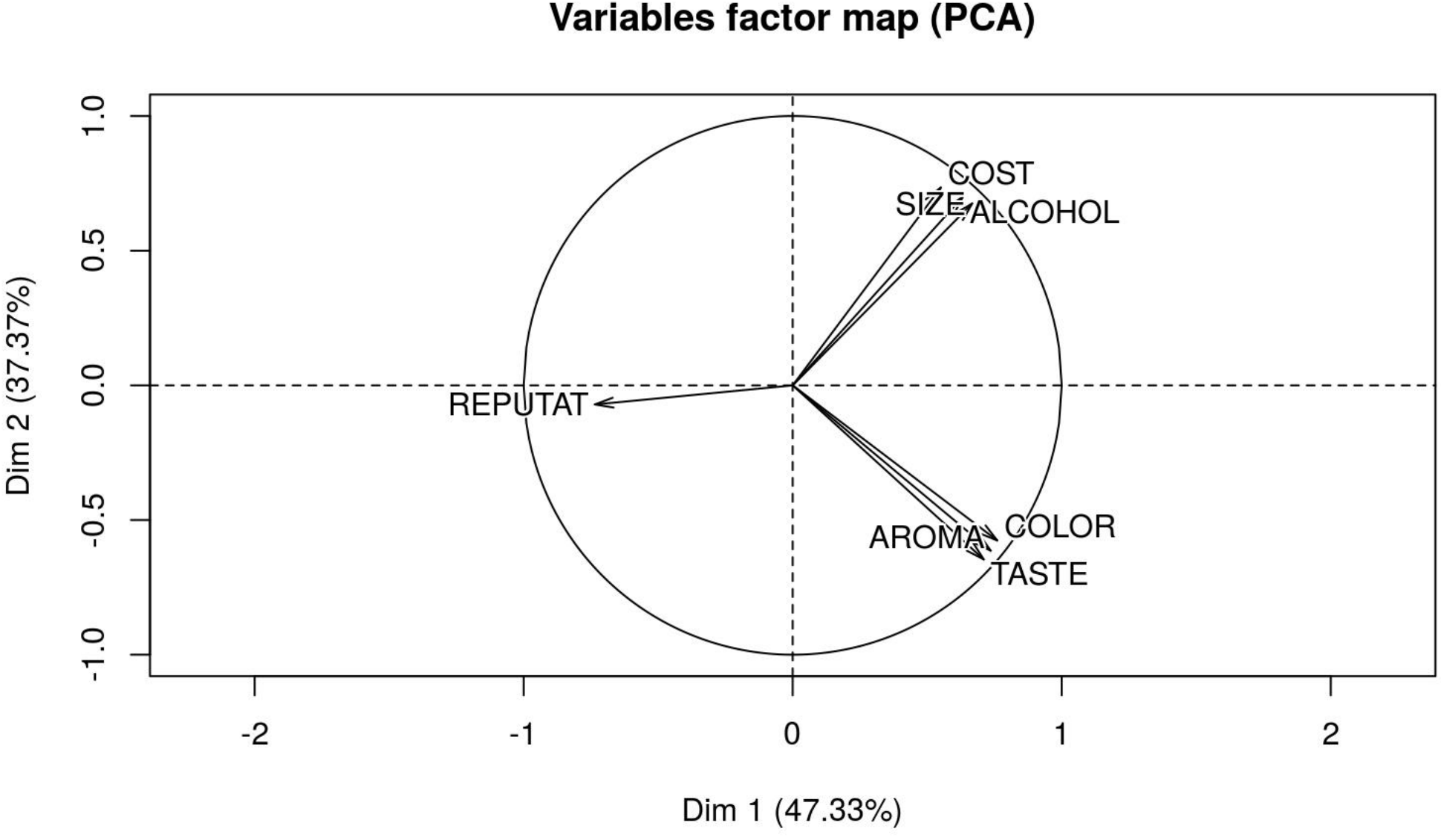
```
##       COLOR               AROMA                TASTE
##  Min.   :-1.89765   Min.   :-1.733093   Min.   :-1.75233
##  1st Qu.:-0.50232   1st Qu.:-0.668064   1st Qu.:-0.71545
##  Median :-0.03721   Median : 0.009682   Median :-0.09332
##  Mean   : 0.00000   Mean   : 0.000000   Mean   : 0.00000
##  3rd Qu.: 0.93952   3rd Qu.: 0.832659   3rd Qu.: 0.99541
##  Max.   : 1.63719   Max.   : 1.752457   Max.   : 1.35831
```

Scatter Plot Matrix



Corrplot


COST
0.83
SIZE
0.77
0.9
ALCOHOL
-0.41
-0.39
-0.46
REPUTAT
0.02
0.18
0.07
-0.37
COLOR
-0.05
0.1
0.04
-0.44
0.91
AROMA
-0.06
0.03
0.01
-0.44
0.9
0.87
TASTE
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

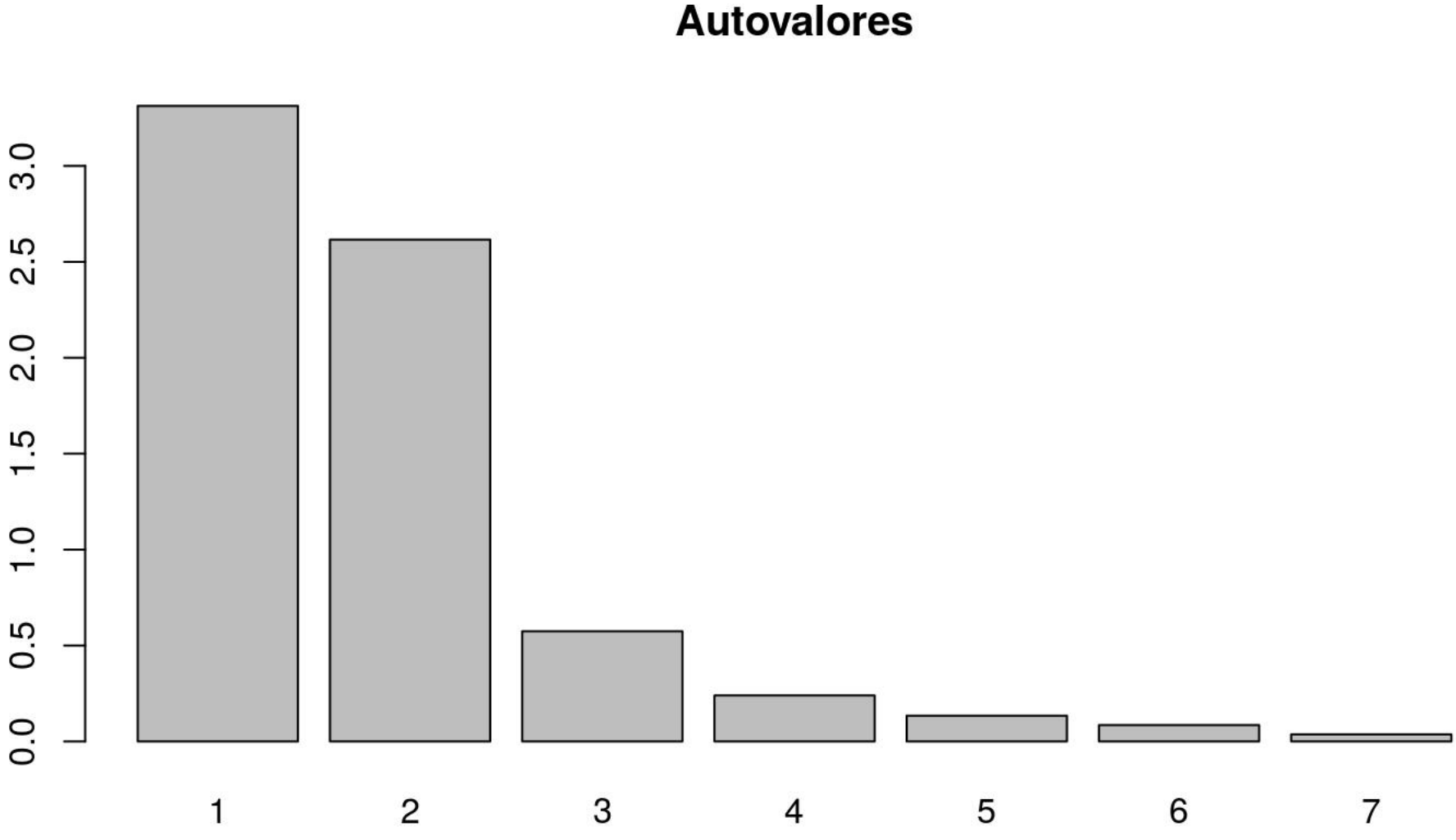

Variables factor map (PCA)
COST
SIZE
ALCOHOL
REPUTAT
AROMA
COLOR
TASTE
1.0
0.5
0.0
-0.5
-1.0
Dim 2 (37.37%)
-2
-1
0
1
2
Dim 1 (47.33%)

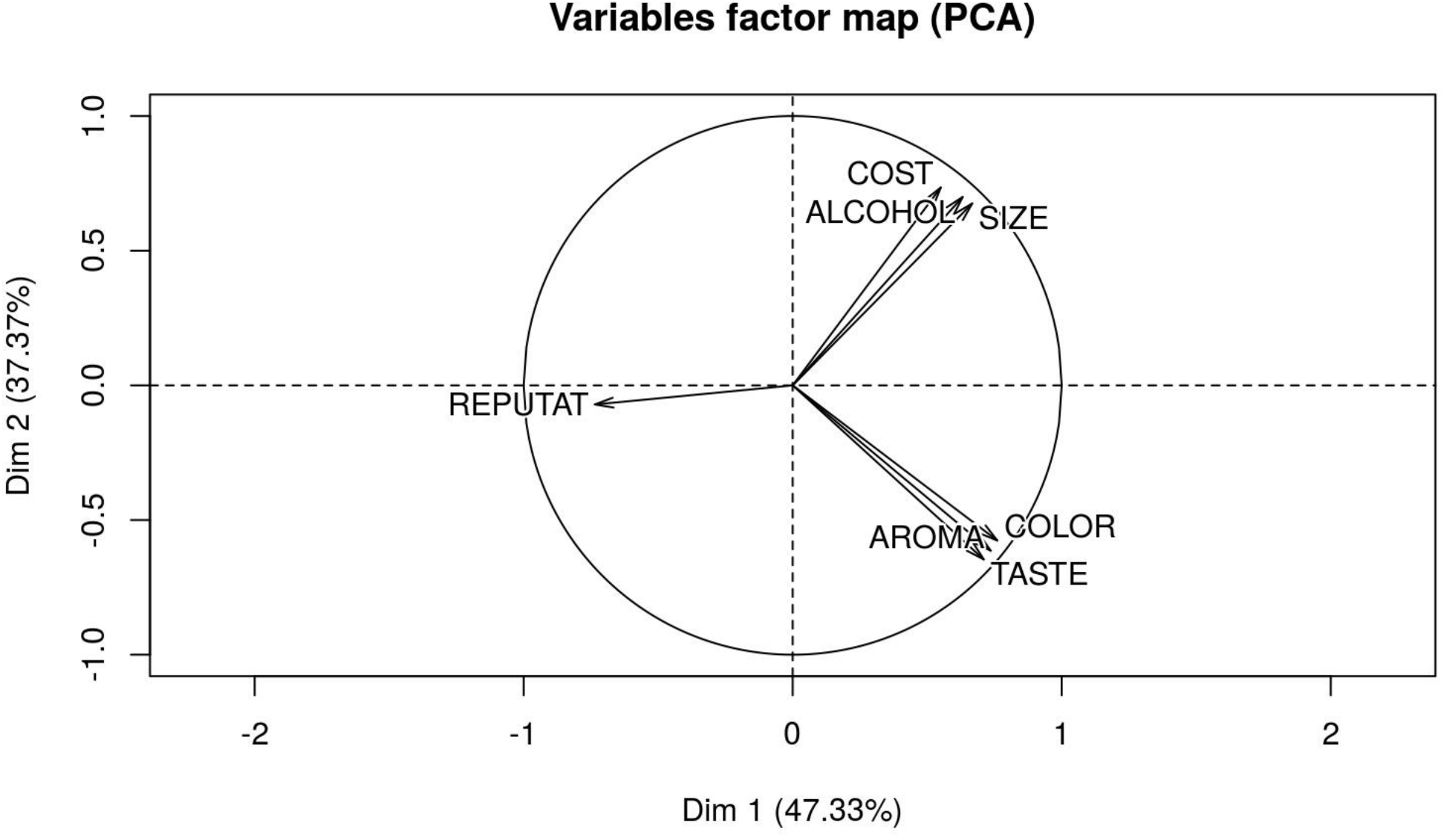
Porcentagem da variância

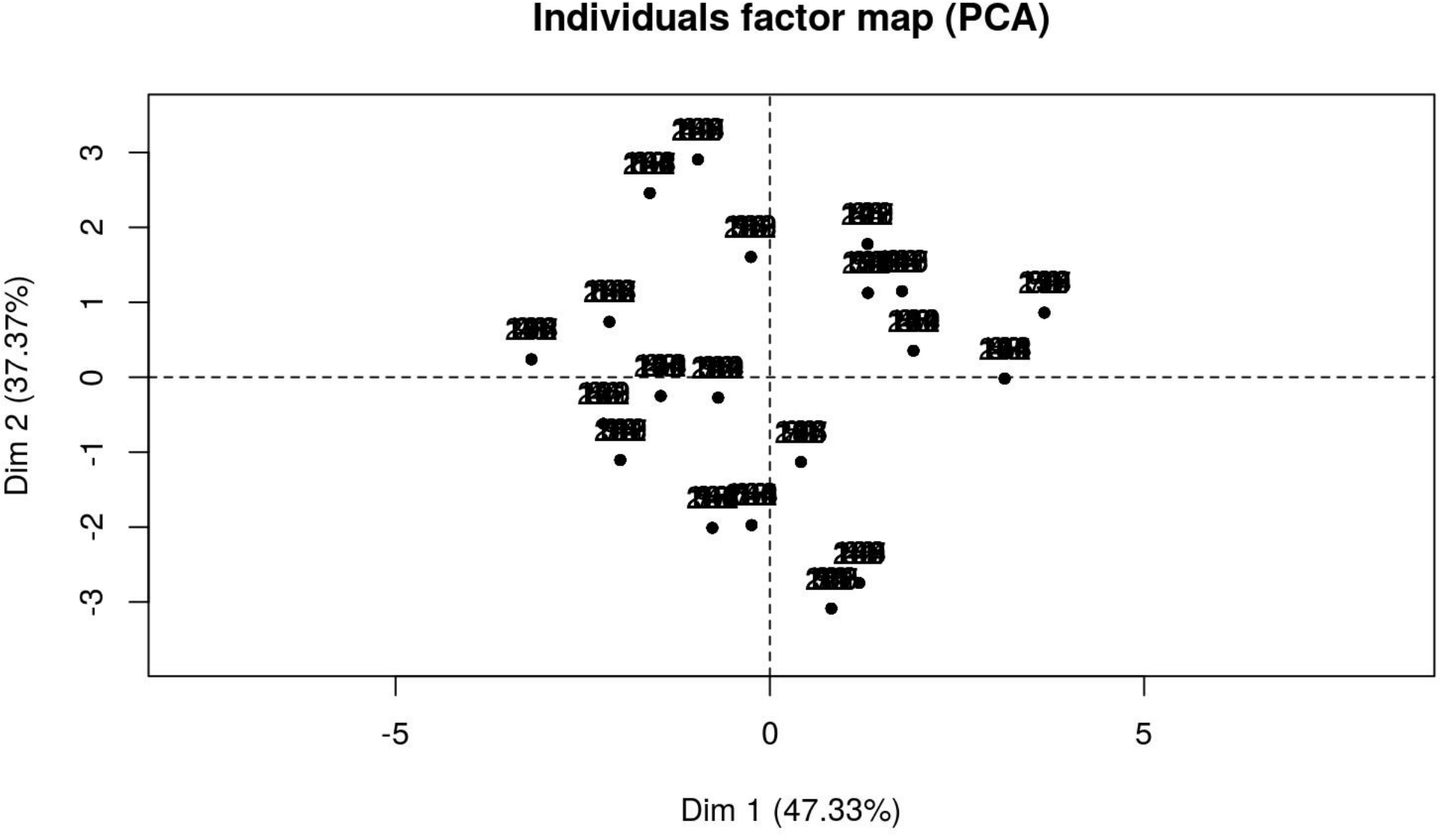
```
##          Autovalor Percentual da Variância Perc. Acumulado da Variância
## comp 1 3.31289019              47.3270028                      47.32700
## comp 2 2.61581560              37.3687943                      84.69580
## comp 3 0.57462888               8.2089839                      92.90478
## comp 4 0.23987960               3.4268514                      96.33163
## comp 5 0.13445601               1.9208001                      98.25243
## comp 6 0.08544310               1.2206157                      99.47305
## comp 7 0.03688663               0.5269518                     100.00000
```

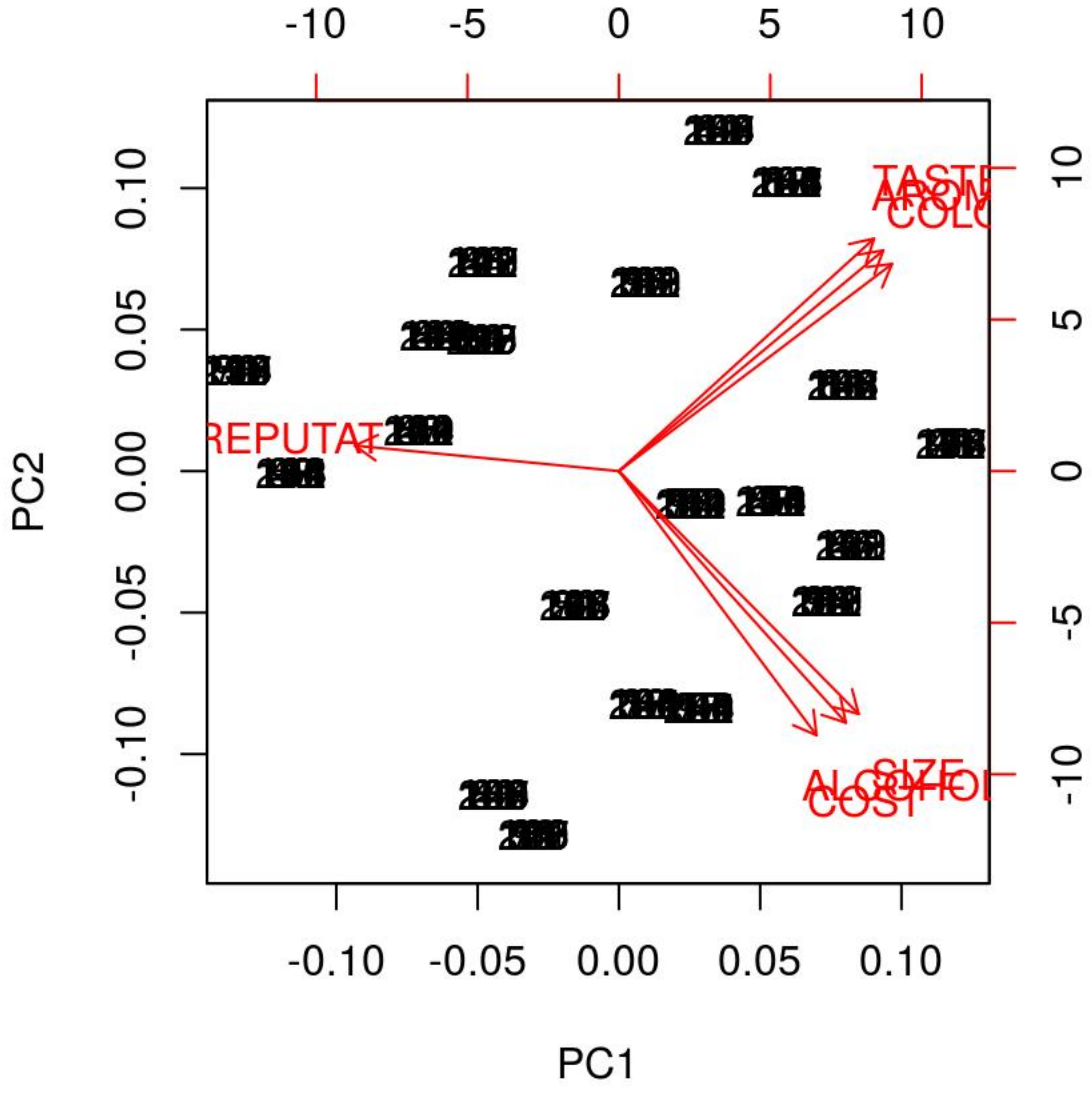
Correlação das Variáveis com as Componentes

```
##                Dim.1       Dim.2      Dim.3        Dim.4       Dim.5
## COST      0.5502440  0.73435109 0.06387144  0.384278545  0.01139728
## SIZE      0.6673536  0.67525974 0.23540802 -0.085283884 -0.07825990
## ALCOHOL   0.6319579  0.69949073 0.06561272 -0.283353683  0.10089370
## REPUTAT -0.7353913 -0.07088947 0.67013609  0.009228349  0.03738382
## COLOR     0.7604789 -0.57603888 0.23290472  0.041704566 -0.02067247
## AROMA     0.7360493 -0.61373125 0.08045650 -0.037167550 -0.21819982
## TASTE     0.7102767 -0.64635047 0.03204868  0.037961700  0.26188908
```


Autovalores
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4
5
6
7


Variables factor map (PCA)
COST
ALCOHOL
SIZE
REPUTAT
AROMA
COLOR
TASTE
Dim 2 (37.37%)
Dim 1 (47.33%)
1.0
0.5
0.0
-0.5
-1.0
-2
-1
0
1
2


Individuals factor map (PCA)
Dim 2 (37.37%)
3
2
1
0
-1
-2
-3
-5
0
5
Dim 1 (47.33%)


-10
-5
0
5
10
0.10
0.05
0.00
-0.05
-0.10
PC2
PC1
-0.10
-0.05
0.00
0.05
0.10
10
5
0
-5
-10
TASTE
COLO
REPUTAT
SIZE
COST

Recomendações em matemática/estatística

- Matrizes (combinação linear, autovalor e autovetor)
- Normalidade (normalização)
- Variância
- Covariância (ortogonalidade)

Obrigado!