Analysis Correlations of features associated with genes DE and detailed PCA 125 patients

Carla Casanova

2022-06-14

```
library(factoextra)
library(corrplot)
library(FactoMineR)
library(NMF)
library(RColorBrewer)
```

```
load("/Users/carlacasanovasuarez/Documents/Master Bioinformatics UAB/Prácticas
Radiomics/Radiomic features/Results_rfeatures/R objects/rdr_assay_scaled.rda")
load("/Users/carlacasanovasuarez/Documents/Master Bioinformatics UAB/Prácticas
Radiomics/Radiomic features/Results_rfeatures/R objects/sputum_eset_countsOK.rda")
load("/Users/carlacasanovasuarez/Documents/Master Bioinformatics UAB/Prácticas
Radiomics/Radiomic features/Results_rfeatures/R objects/sputum_eset_phenoOK.rda")
```

```
head(rdr_assay, 3)
```

```
##                              30442       85931       4428      41420
## Elongation.original      -0.8292798  0.57463917  1.4786649 -0.4289525
## Flatness.original        -1.3271903  0.93615221  0.1014230  1.3444300
## LeastAxisLength.original -0.3553029 -0.08535293 -0.7748625  1.7224070
##                              15405       26383      55856    92849      40819
## Elongation.original      -0.5223833 0.005841054  0.6091809 1.143732 -0.8826891
## Flatness.original         1.6197135 0.865417524 -1.5204424 1.433299  0.3719586
## LeastAxisLength.original  0.6486760 0.129955023 -1.0959976 1.611491  1.2635395
##                              77811      14804      51797      55255      29240
## Elongation.original      -0.7405845 -1.6697908  1.2625101 -0.6116792 -0.4826862
## Flatness.original        -1.3846652 -0.2980319  0.5118313 -0.5790064  2.1745543
## LeastAxisLength.original -0.6941301  0.8547479 -0.7307843  0.5846189  1.5960623
##                             95106      87587     80068      73795   10788
## Elongation.original      0.4298360  0.2577669 -1.125272 -1.2737981 1.04647
## Flatness.original        1.1724278 -1.2230098  1.421167  0.0727714 2.17705
## LeastAxisLength.original 0.1872222 -0.9677312  2.532617  0.9101333 1.62036
##                              47780    66276       3269      58758      42518
## Elongation.original       0.6916887 1.286194  1.2808403 -1.6736516  0.8032563
## Flatness.original        -0.7372629 1.026036  0.7510787 -0.9425739 -1.0696448
## LeastAxisLength.original -1.3571905 1.259442 -0.8614411 -0.5891942 -1.4743710
##                              98007      90488       45977      82970
## Elongation.original       0.9460872 -0.8456865  2.09470593 -0.2241601
## Flatness.original         0.4406275  0.2536706 -0.06807934 -1.1167328
## LeastAxisLength.original -1.1514928  0.8541716 -0.96783371 -0.5413949
##                               1466       38458       75451       12443
## Elongation.original       1.2601573 -0.07072068 -1.61938281 -1.13215083
```

```
## Flatness.original        -0.6640676 -1.51005753 -0.09751104  0.07358346
## LeastAxisLength.original -2.0305398 -1.22398113 -0.75980559  0.68918350
##                               30940       4925      15902      71391     74976
## Elongation.original      -0.5541348 -0.1896107 -1.3331527 -1.4791068  1.820599
## Flatness.original         0.1821081  1.4822234 -1.1340990 -1.0413768 -1.191841
## LeastAxisLength.original  0.4864675  0.2161385  0.1567745  0.1310087 -1.584660
##                              93472      85953      41442      96931    26404
## Elongation.original      1.0186731  0.8497634 -0.4106150  2.0591743 1.152393
## Flatness.original        0.4541745 -0.3231623  0.5984982  0.8098340 4.005021
## LeastAxisLength.original 0.6349061 -0.7383804  1.3399676 -0.3566548 3.065986
##                               85352       3848       14826      92270
## Elongation.original      -0.3926316 -0.1729709 -0.54189550 -1.2992826
## Flatness.original         0.9421409 -0.1624292  1.38514733 -0.1406693
## LeastAxisLength.original  1.0922928 -0.8835705  0.08597909  0.8418388
##                               47758       3247      40240      51217
## Elongation.original      -0.5216068 -1.3974239 -1.0702889  0.3943931
## Flatness.original        -1.5577573 -0.9568438 -0.1105678 -0.5941881
## LeastAxisLength.original -0.2595194 -0.4880271  0.6149842 -0.4405684
##                               88210      91669      39639     46556
## Elongation.original      -0.90743560 0.03915172 -0.4827684 0.4577083
## Flatness.original         0.08460186 1.19882467  1.0578664 1.0928048
## LeastAxisLength.original  1.10017648 1.24174565  1.3569317 0.7087421
##                                83549     50015       5504      24000
## Elongation.original      -0.804460105  1.051174 -1.2599194 -1.1839048
## Flatness.original        -0.004254626 -2.470431  0.6056674  0.2754346
## LeastAxisLength.original -0.003393030 -2.552729  1.0953973  1.5867633
##                               60993       90466      82948     38437      67910
## Elongation.original       0.1424956  0.02424732  0.2385426 0.9084779 -0.4862622
## Flatness.original         0.4155954 -0.53082327 -1.1196725 1.4607546 -0.1794514
## LeastAxisLength.original -0.2828021  0.72277991  0.3725146 0.2915283  0.1373594
##                               23399       30317      48813      96783
## Elongation.original      -0.4595873  0.66596566  0.8547475 -1.2461428
## Flatness.original        -0.3128051 -0.21179033 -0.8220422  0.1273786
## LeastAxisLength.original  0.3823280 -0.01526712 -1.0589477  0.3390800
##                                3701     22197    44152     88062       6558
## Elongation.original      -1.1318863 0.9428124 1.204460 0.4381599 -0.7398427
## Flatness.original        -1.8242938 1.9348145 2.391823 2.6557726  0.4739143
## LeastAxisLength.original  0.2197063 1.5250100 1.894141 2.5224733  1.0361961
##                               60845       48064     66561      85057     51523
## Elongation.original      -1.4162502 -0.06666286 1.1387810  2.0705575 0.3277708
## Flatness.original         0.9355742  0.47079335 0.8988818 -0.4171988 1.5489982
## LeastAxisLength.original  1.2839033  0.47480792 1.2681948 -0.1401579 0.1958828
##                               24907     83855      94832      98291       21607
## Elongation.original       0.8621163 0.2809112 -1.2292837 -0.9015037  0.56957467
## Flatness.original        -0.6195458 0.2898502  0.2488953  0.3876710 -0.05227464
## LeastAxisLength.original -1.1680295 0.7334565  0.4726829 -0.4314852  0.41275441
##                               87472      62364      99357     73342      32743
## Elongation.original      -0.5264626 0.03804299 -0.4638809 1.7176453  0.4050058
## Flatness.original         0.1294606 1.67297269  0.1830202 0.5994512  0.1370055
## LeastAxisLength.original  0.2588511 1.78103225  0.5127459 0.1258829 -0.5679314
##                               25224       43720     62217       36202
## Elongation.original      -0.3125049  0.30159990 0.7464043  1.10247340
## Flatness.original        -1.2502352  0.30599093 0.4913851  1.43500654
## LeastAxisLength.original -0.5953601 -0.04987364 0.2490570 -0.01607101
```

```
##                                39660      83412     75893      12886      60856
## Elongation.original      -1.2467418 -1.2775881 0.9897145  1.2036763  0.4191224
## Flatness.original        -0.3623748 -1.1773274 0.2211943  0.3176731 -0.3650902
## LeastAxisLength.original  0.5097345 -0.9667502 0.5068700 -0.8571246  0.2340758
##                                82811       1307     12285     67774     41759
## Elongation.original       0.484615489  0.9452282 -1.999940 0.2403159 0.9187287
## Flatness.original         1.232657738  1.0170458 -1.963327 0.9188092 1.3297018
## LeastAxisLength.original -0.006661428 -1.1258914 -1.973731 0.2592670 1.2282852
##                                60255      70484      25973      44469
## Elongation.original      -0.9753002 -0.5223515 -0.6936350  1.2511047
## Flatness.original         0.6255008 -1.3473326 -0.8126154  0.4811139
## LeastAxisLength.original  0.5234387 -0.7409427 -0.3500249 -0.7810927
##                                99958      36950      55446      15744
## Elongation.original       0.001946951 0.07436336  2.3460426  0.5257634
## Flatness.original        -1.187148422 0.97434938 -0.3152815  0.4937721
## LeastAxisLength.original -1.892860688 1.36300506 -1.0074319 -0.1751736
##                                74691      15143      89128       40557     51534
## Elongation.original      -1.1900936 -0.9957399 -1.5274097 -0.07286195 -0.325766
## Flatness.original         0.7244930 -0.1722804 -0.4674297  0.01446868 -1.171972
## LeastAxisLength.original -0.3326143  0.3881390 -0.3239862  0.59138359  0.072593
##                                 2963      72888      53157       1127      23082
## Elongation.original      -0.2518089 -0.4194247 -0.8563466 -0.7955098  0.1212896
## Flatness.original         1.1945531  1.2148009 -1.5483665 -0.5224356 -0.3008626
## LeastAxisLength.original  1.0933157  1.2067433 -0.8773386 -0.4856435 -1.0609077
##                                41578      97067     85489      26098       3542
## Elongation.original      -0.08288337 -0.3289904 1.0513348 -0.9261862 -0.8763397
## Flatness.original         0.12628321 -0.6319421 0.3213083 -0.5998131  1.8727552
## LeastAxisLength.original -0.32110373 -1.4852088 0.5301211  0.6475418  1.1509766
##                                77527
## Elongation.original       0.2149637
## Flatness.original        -1.1022768
## LeastAxisLength.original -2.0307580
```

Corralation analysis: features with positive DE

```
pos_rf <- list("Sphericity.original", "Variance.original", "Autocorrelation.original",
    "ClusterShade.original", "Contrast.original", "DifferenceAverage.original",
    "Id.original",
    "Idm.original", "Idmn.original", "Idn.original", "InverseVariance.original",
    "JointAverage.original", "SumAverage.original", "HighGrayLevelEmphasis.original",
    "LongRunLowGrayLevelEmphasis.original", "LowGrayLevelRunEmphasis.original",
    "RunPercentage.original",
    "LargeAreaEmphasis.original", "LowGrayLevelEmphasis.original",
    "LargeAreaHighGrayLevelEmphasis.original",
    "LargeAreaLowGrayLevelEmphasis.original", "SmallAreaEmphasis.original",
    "SmallAreaHighGrayLevelEmphasis.original",
    "SmallAreaLowGrayLevelEmphasis.original", "ZoneVariance.original",
    "DependenceNonUniformity.original",
    "HighGrayLevelRunEmphasis.original", "LargeDependenceEmphasis.original",
    "LargeDependenceLowGrayLevelEmphasis.original",
    "Busyness.original", "Coarseness.original", "Complexity.original",
    "Strength.original",
    "X10Percentile.original", "InterquartileRange.original")
```

Subset original data of patients with transcriptomics:

```
# Prepare data in object: rdr_assay.scaled
rownames(rdr_assay)[rownames(rdr_assay) == "10Percentile.original"] <-
"X10Percentile.original"
rownames(rdr_assay)[rownames(rdr_assay) == "90Percentile.original"] <-
"X90Percentile.original"

# Store an array with values of features associated to genes DE
rdr_positive <- rdr_assay[unlist(pos_rf), ]

# Change labels to visualize better
rownames(rdr_positive) <- substr(rownames(rdr_positive), start = 1, stop =
nchar(rownames(rdr_positive)) -
    9)
```

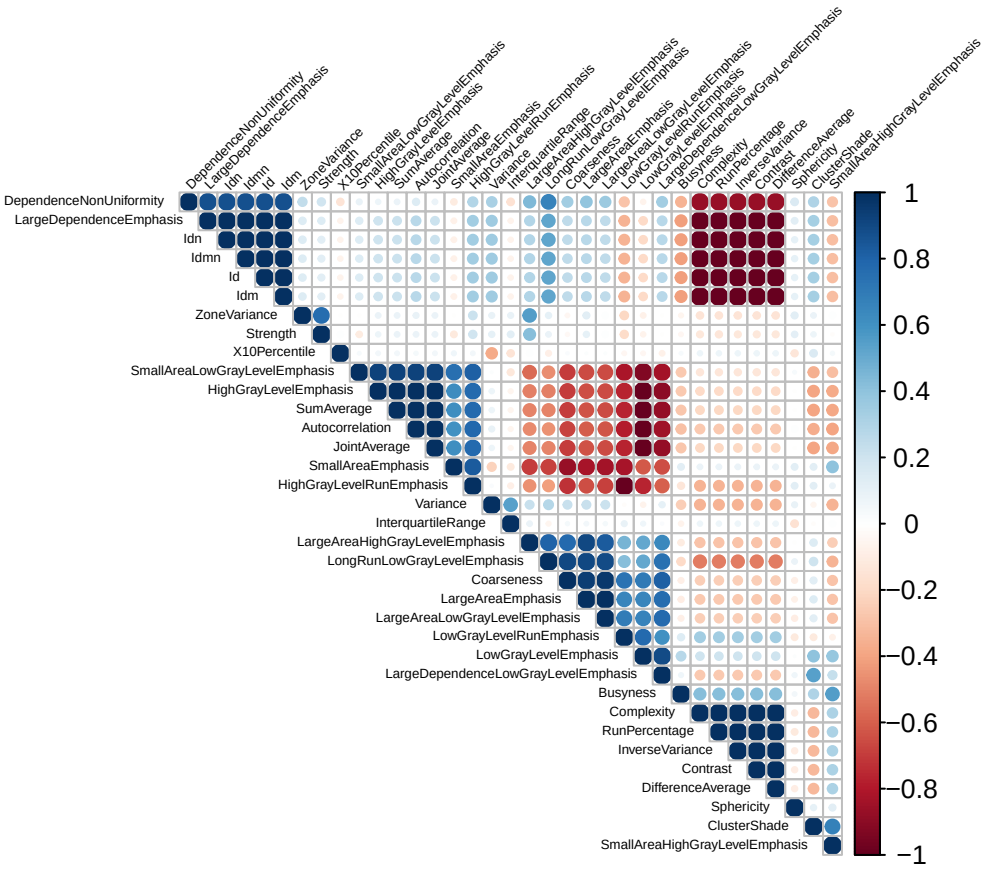
See correlations between features positively associated to genes DE:

```
cor.mat <- round(cor(t(rdr_positive)), 2)

head(cor.mat, 4)
```

```
##                 Sphericity Variance Autocorrelation ClusterShade Contrast
## Sphericity            1.00     0.09            0.00         0.10    -0.10
## Variance              0.09     1.00            0.09        -0.06    -0.35
## Autocorrelation       0.00     0.09            1.00        -0.37    -0.27
## ClusterShade          0.10    -0.06           -0.37         1.00    -0.33
##                 DifferenceAverage   Id  Idm Idmn  Idn InverseVariance
## Sphericity                  -0.10 0.10 0.10 0.10 0.10           -0.10
## Variance                    -0.35 0.35 0.35 0.35 0.35           -0.35
## Autocorrelation             -0.27 0.27 0.27 0.27 0.27           -0.27
## ClusterShade                -0.33 0.33 0.33 0.33 0.33           -0.33
##                 JointAverage SumAverage HighGrayLevelEmphasis
## Sphericity             -0.01      -0.01                 -0.01
## Variance                0.07       0.07                  0.07
## Autocorrelation         1.00       1.00                  1.00
## ClusterShade           -0.40      -0.40                 -0.40
##                 LongRunLowGrayLevelEmphasis LowGrayLevelRunEmphasis
## Sphericity                             0.08                   -0.12
## Variance                               0.28                   -0.02
## Autocorrelation                       -0.46                   -0.78
## ClusterShade                           0.20                   -0.12
##                 RunPercentage LargeAreaEmphasis LowGrayLevelEmphasis
## Sphericity              -0.12             -0.08                 0.01
## Variance                -0.35              0.22                -0.07
## Autocorrelation         -0.27             -0.61                -1.00
## ClusterShade            -0.33              0.14                 0.40
##                 LargeAreaHighGrayLevelEmphasis LargeAreaLowGrayLevelEmphasis
## Sphericity                               -0.01                         -0.10
## Variance                                  0.22                          0.21
## Autocorrelation                          -0.48                         -0.63
## ClusterShade                              0.15                          0.13
##                 SmallAreaEmphasis SmallAreaHighGrayLevelEmphasis
## Sphericity                   0.12                          0.13
## Variance                    -0.23                         -0.35
```

```
## Autocorrelation              0.60                          -0.40
## ClusterShade                 0.12                           0.67
##                 SmallAreaLowGrayLevelEmphasis ZoneVariance
## Sphericity                               0.03         0.13
## Variance                                 0.01         0.09
## Autocorrelation                          0.92         0.09
## ClusterShade                            -0.36         0.08
##                 DependenceNonUniformity HighGrayLevelRunEmphasis
## Sphericity                         0.16                    0.12
## Variance                           0.32                    0.02
## Autocorrelation                    0.14                    0.78
## ClusterShade                       0.30                    0.12
##                 LargeDependenceEmphasis LargeDependenceLowGrayLevelEmphasis
## Sphericity                         0.12                                0.06
## Variance                           0.35                                0.10
## Autocorrelation                    0.27                               -0.85
## ClusterShade                       0.33                                0.54
##                 Busyness Coarseness Complexity Strength X10Percentile
## Sphericity          0.05      -0.09      -0.10     0.11         -0.14
## Variance           -0.25       0.22      -0.35     0.10         -0.38
## Autocorrelation    -0.30      -0.67      -0.27     0.08          0.04
## ClusterShade        0.30       0.13      -0.33     0.07          0.16
##                 InterquartileRange
## Sphericity                   -0.16
## Variance                      0.54
## Autocorrelation              -0.06
## ClusterShade                 -0.02
```

```
# pdf('Radiomic features from DE analysis correlation matrix.pdf')

corrplot(cor.mat, type = "upper", order = "hclust", tl.col = "black", tl.srt = 45,
    tl.cex = 0.4)
```




```
# dev.off()
```

PCA

```
# Change labels to visualize better
rownames(rdr_assay) <- substr(rownames(rdr_assay), start = 1, stop =
nchar(rownames(rdr_assay)) -
    9)

# Perform PCA for all features, but only 125 patients
res.pca <- PCA(t(rdr_assay), graph = FALSE)
```

Variables contribution

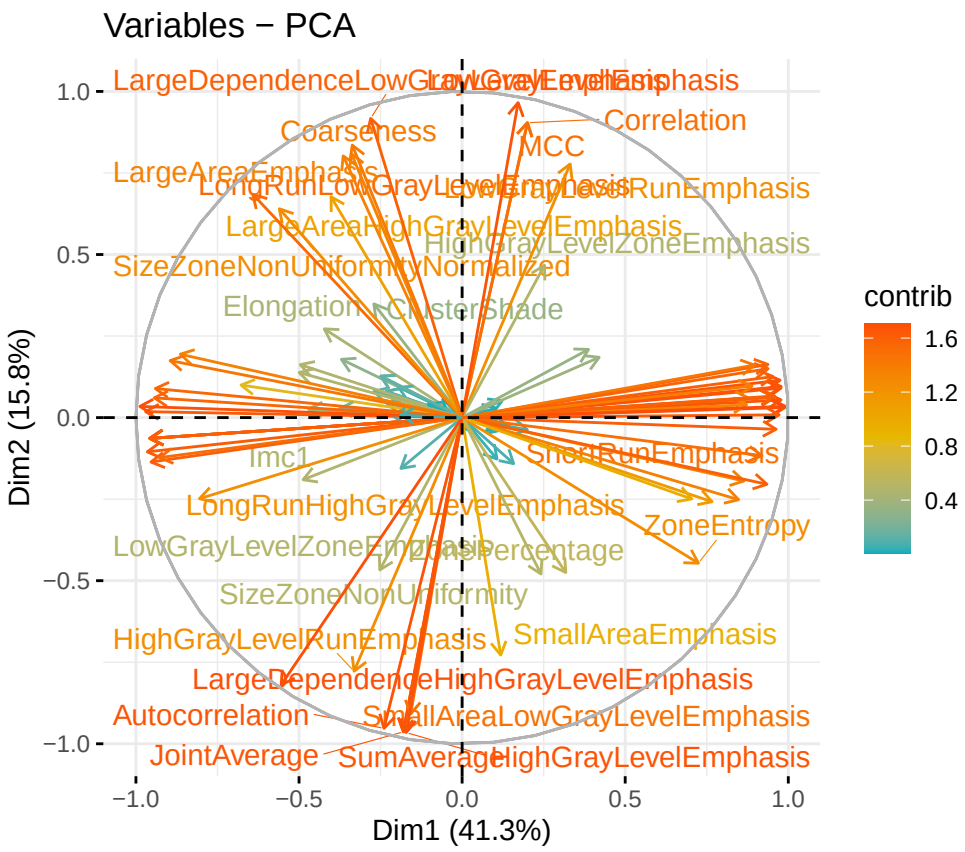
```
# Extract the results for variables
var <- get_pca_var(res.pca)
var
```

```
## Principal Component Analysis Results for variables
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates for the variables"
## 2 "$cor"     "Correlations between variables and dimensions"
## 3 "$cos2"    "Cos2 for the variables"
## 4 "$contrib" "contributions of the variables"
```

```
head(var$cor)
```

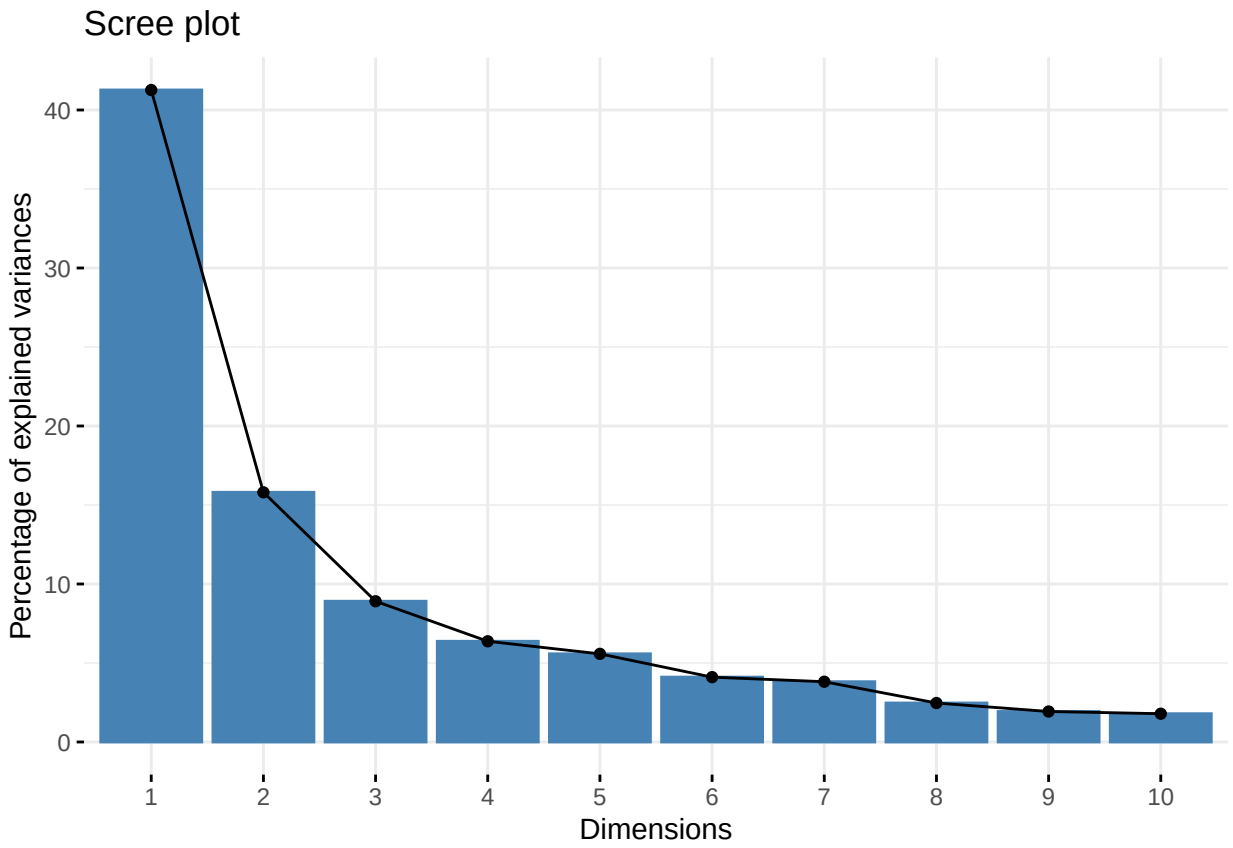
```
##                                Dim.1       Dim.2      Dim.3       Dim.4
## Elongation                -0.42271997  0.27203111 0.05235467 -0.07268764
## Flatness                  -0.05939099  0.05613864 0.24931915 -0.01818358
## LeastAxisLength            0.02764643 -0.04283511 0.70664866 -0.08359625
## MajorAxisLength            0.10699159 -0.12712578 0.62958073 -0.09868416
## Maximum2DDiameterColumn   -0.23013232  0.04330301 0.81480576 -0.16907432
## Maximum2DDiameterRow      -0.27012443  0.08076097 0.70034454 -0.18039511
##                                Dim.5
## Elongation                -0.1493979
## Flatness                  -0.2192308
## LeastAxisLength            0.1006798
## MajorAxisLength            0.4129329
## Maximum2DDiameterColumn    0.3299912
## Maximum2DDiameterRow       0.2876013
```

```
# pdf('Contribution variables PCA 125 patients.pdf')

# Graph of variables: default plot
fviz_pca_var(res.pca, col.var = "contrib", gradient.cols = c("#00AFBB", "#E7B800",
    "#FC4E07"), repel = TRUE)  # Avoid text overlapping
```

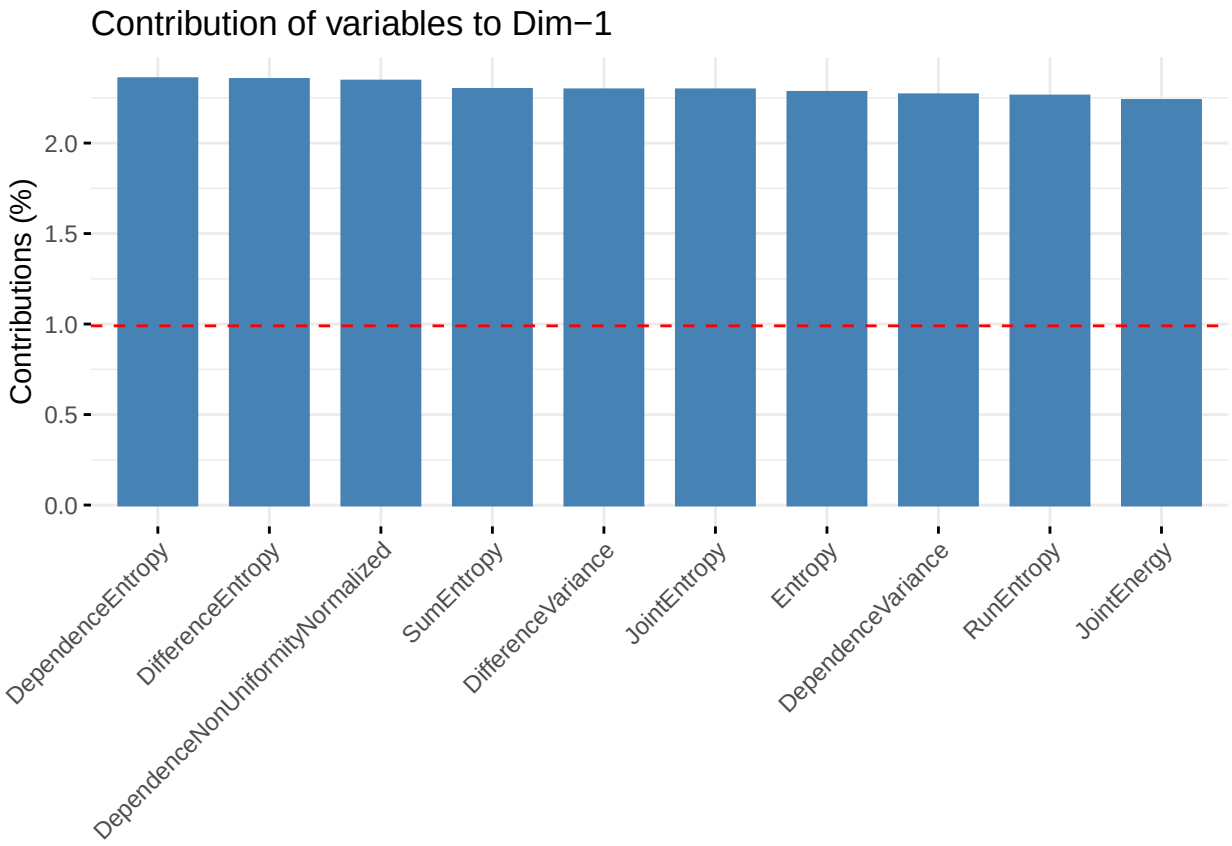



```
# dev.off()
```

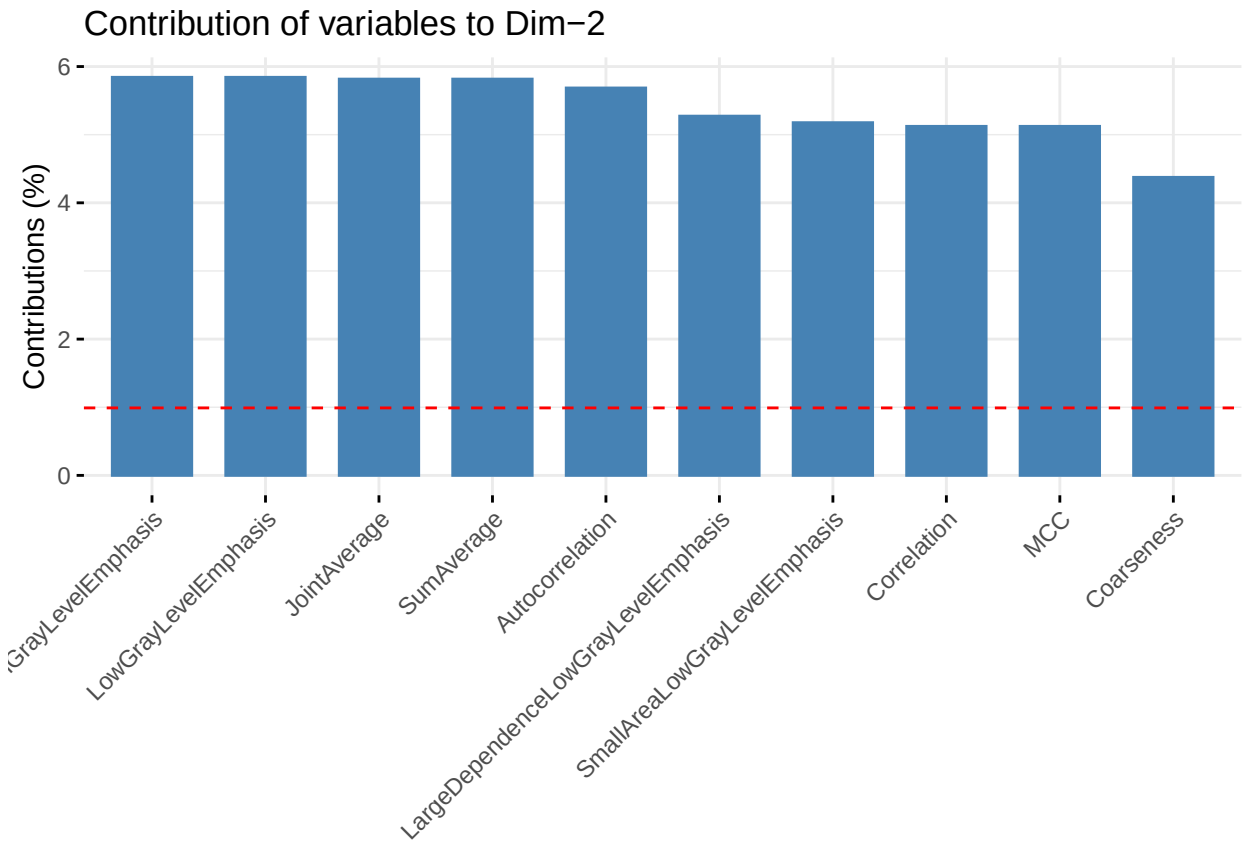
```
fviz_screeplot(res.pca, ncp = 10)
```



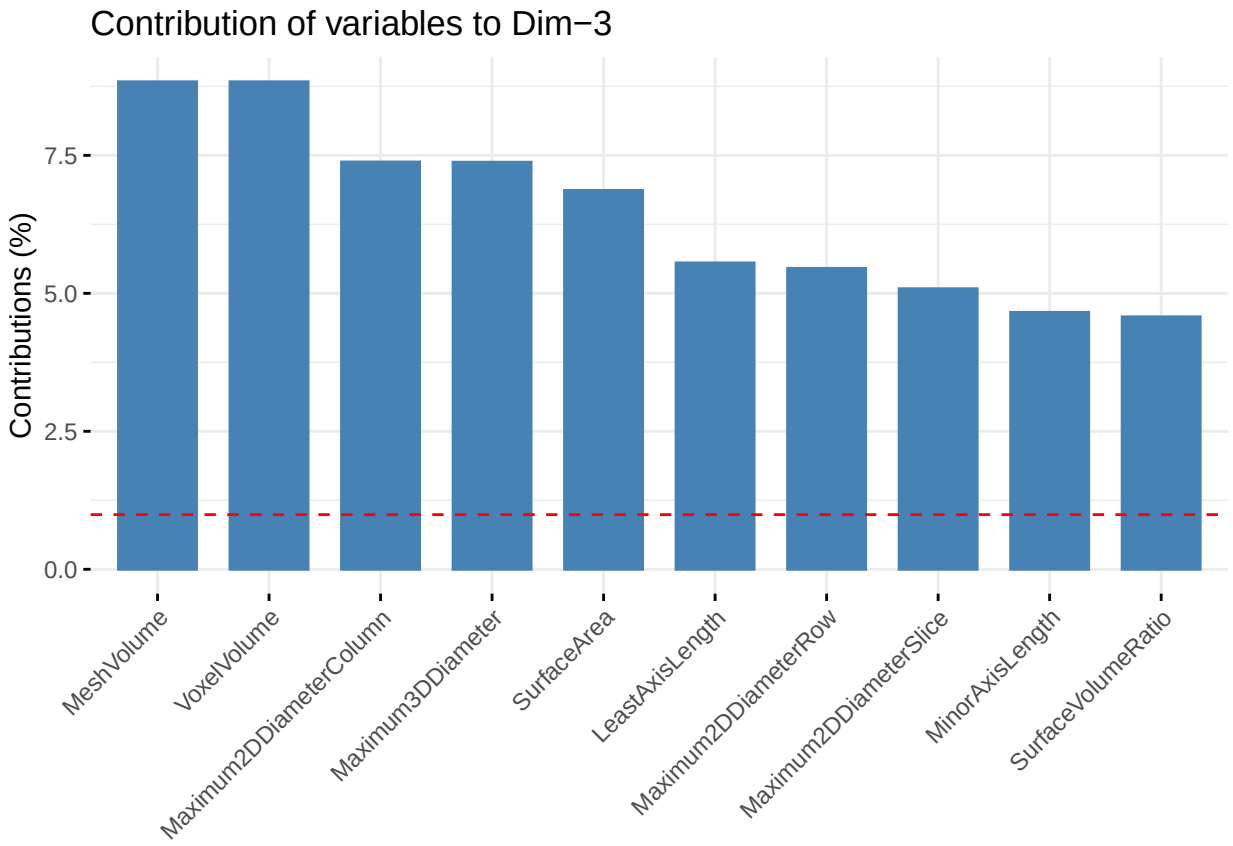

```
# pdf('Contribution variables PC1-3.pdf')

# Contributions of variables to PC1
fviz_contrib(res.pca, choice = "var", axes = 1, top = 10)
```




```
# Contributions of variables to PC2
fviz_contrib(res.pca, choice = "var", axes = 2, top = 10)
```



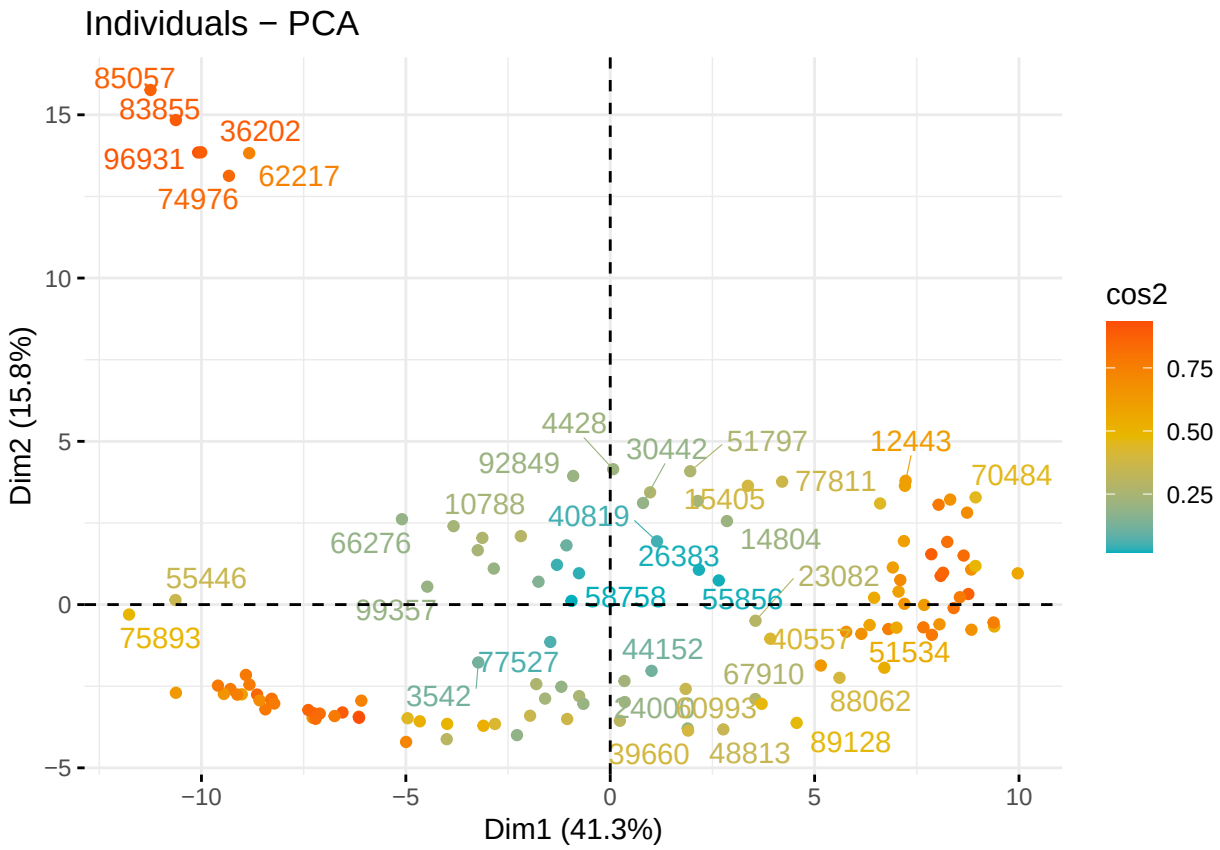

```
# Contributions of variables to PC3
fviz_contrib(res.pca, choice = "var", axes = 3, top = 10)
```




```
# dev.off()
```

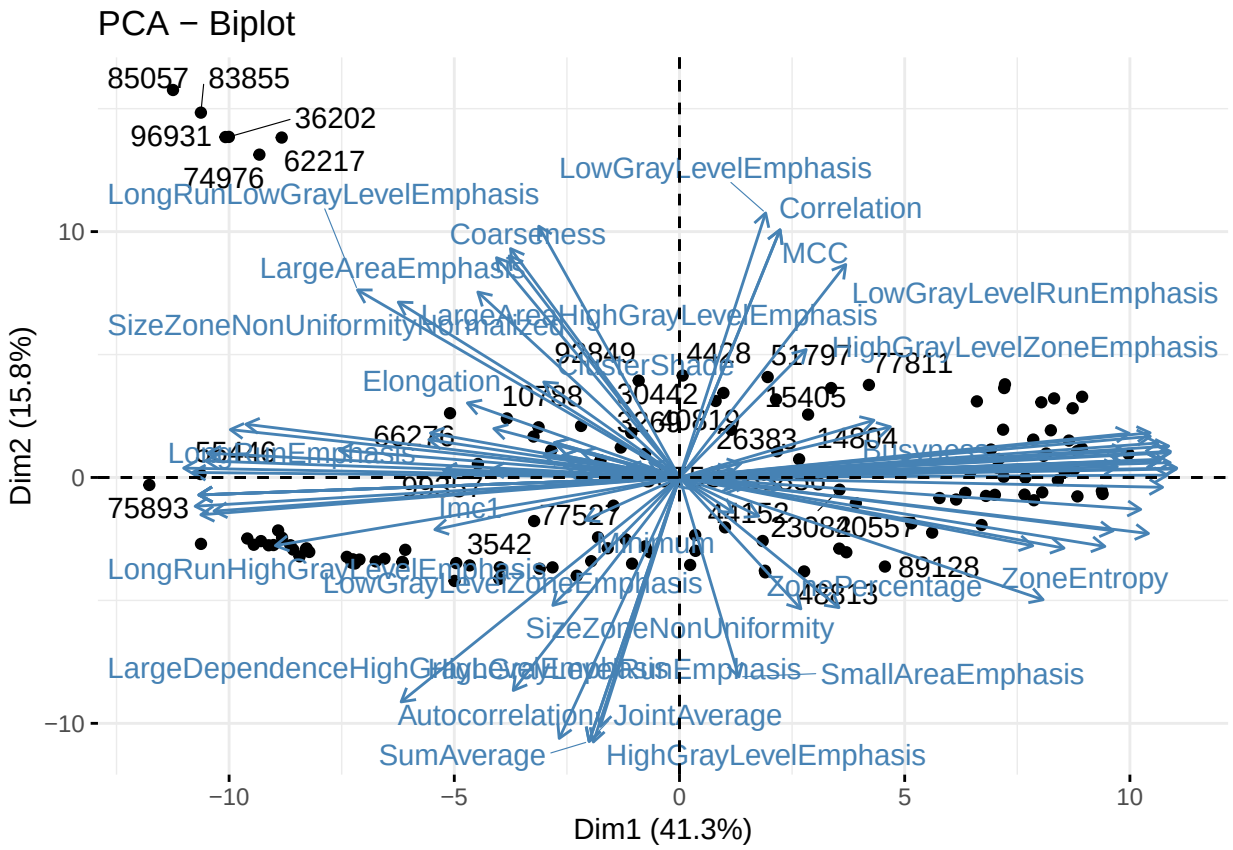
Individuals

```
# pdf('Contribution individuals PCA 125 patients.pdf')

# Graph of individuals 1. Use repel = TRUE to avoid overplotting 2. Control
# automatically the color of individuals using the cos2 cos2 = the quality of
# the individuals on the factor map Use points only 3. Use gradient color
fviz_pca_ind(res.pca, col.ind = "cos2", gradient.cols = c("#00AFBB", "#E7B800",
"#FC4E07"),
    repel = TRUE  # Avoid text overlapping (slow if many points)
)
```

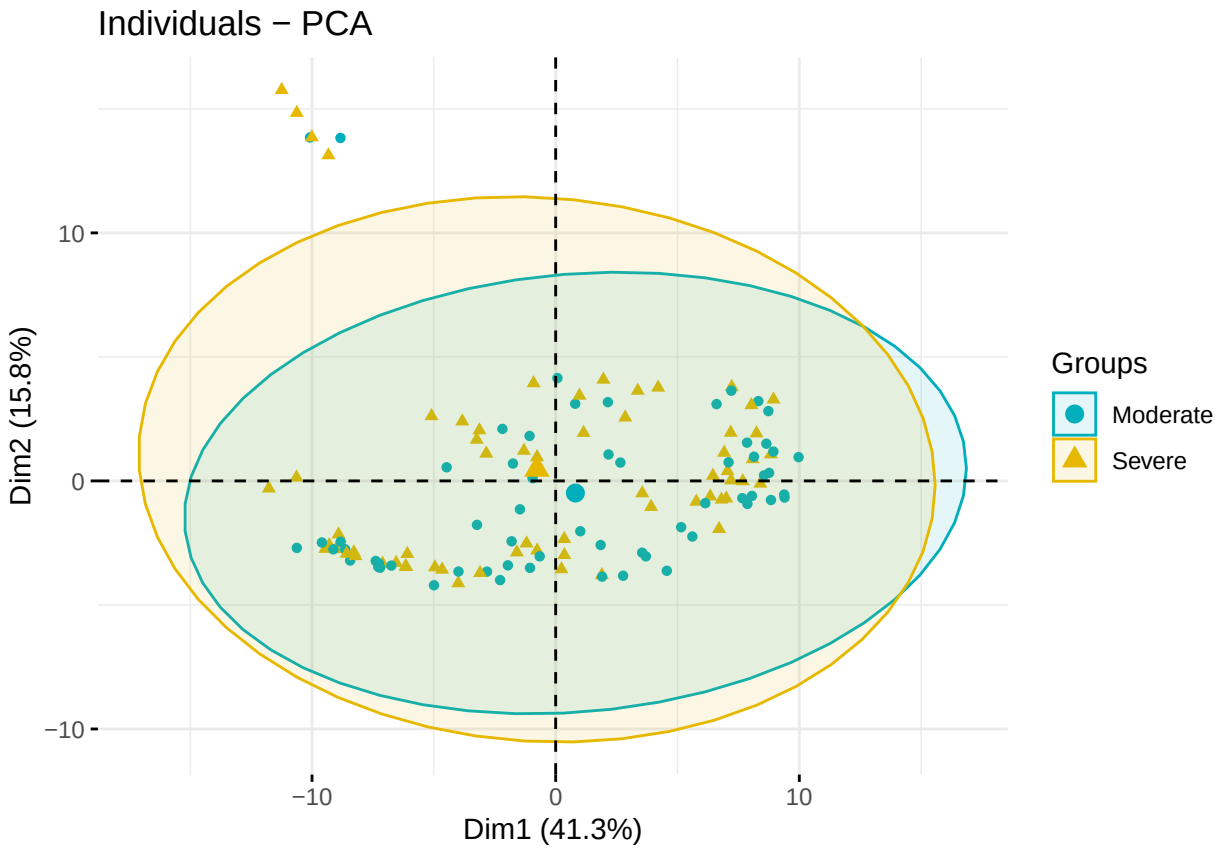



```
# dev.off()
```

```
# Biplot of individuals and variables
fviz_pca_biplot(res.pca, repel = TRUE)
```




```r
#pdf("Contribution individuals by group.pdf")

# Visualize
# Use habillage to specify groups for coloring
fviz_pca_ind(res.pca,
             label = "none", # hide individual labels
             habillage = phenoDataFrame$GROUP_bin, # color by groups
             palette = c("#00AFBB", "#E7B800"),
             addEllipses = TRUE # Concentration ellipses
             )
```




```
#dev.off()
```

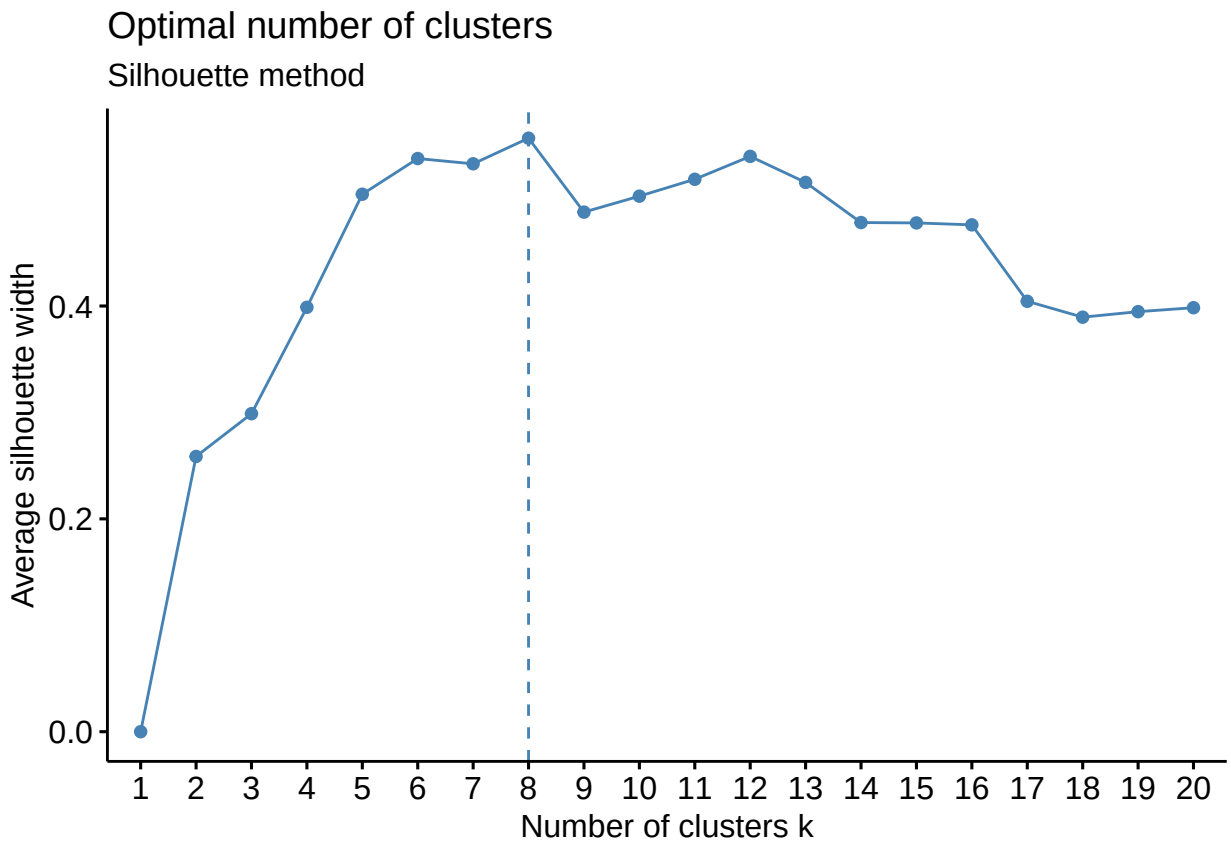
HCL

Variables must be columns and observations row:

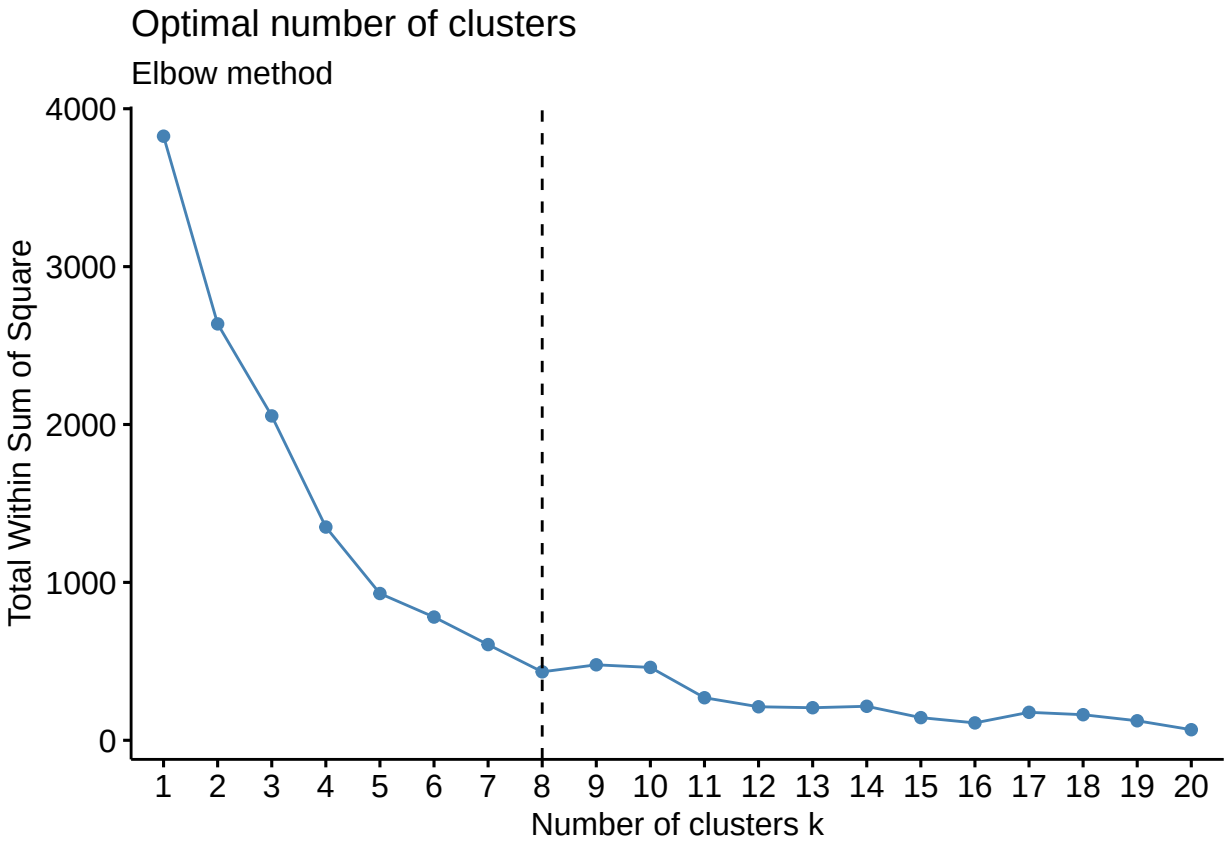
```
# pdf('Heatmap positively correlated features with genes DE.pdf')

# Colors
mypalette <- brewer.pal(11, "PiYG")
morecols <- colorRampPalette(mypalette)

# Heatmap
aheatmap(rdr_positive, col = rev(morecols(37)), main = "Rfeatures with genes DE",
    scale = "none")
```


```
# dev.off()
```

```
set.seed(123)

fviz_nbclust(rdr_positive, kmeans, method = "silhouette", k.max = 20) + labs(subtitle =
"Silhouette method")
```



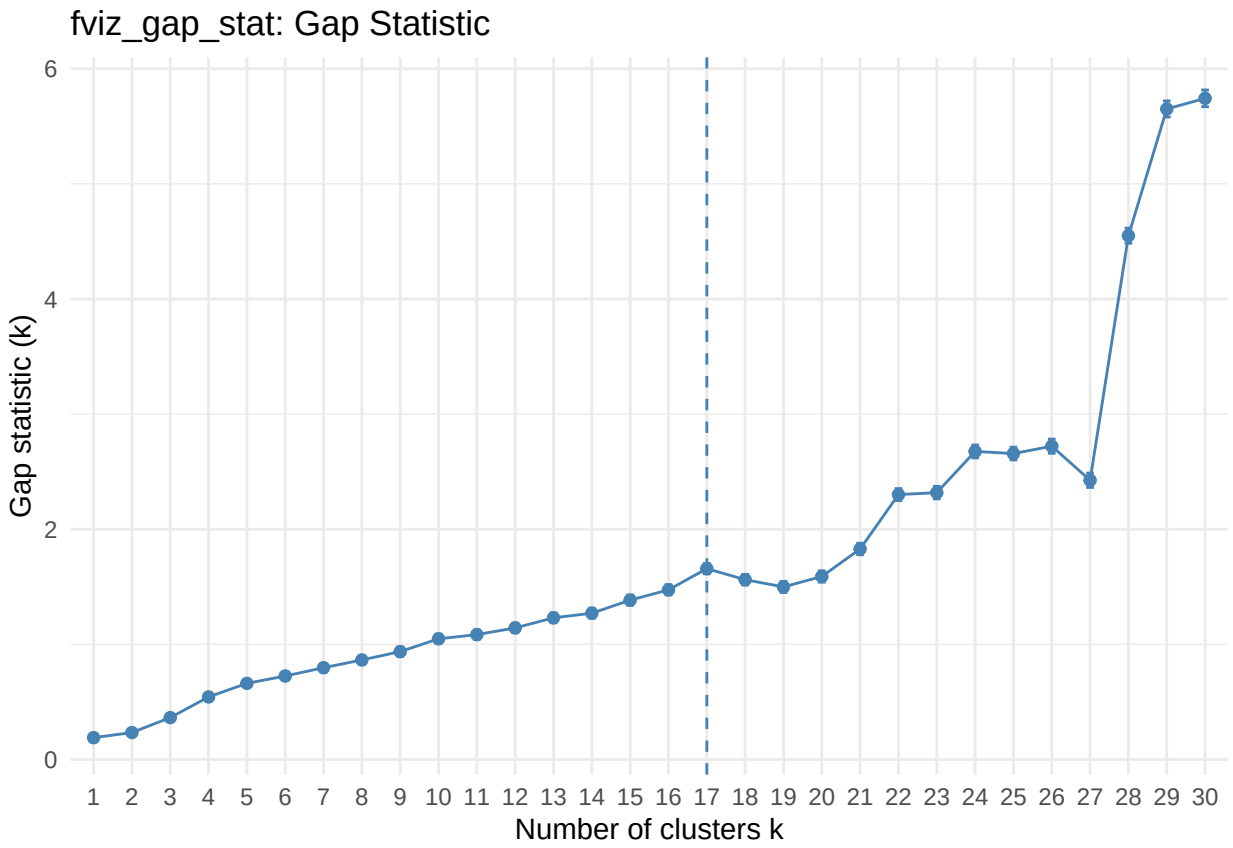

```
fviz_nbclust(rdr_positive, kmeans, method = "wss", k.max = 20) + geom_vline(xintercept =
8,
    linetype = 2) + labs(subtitle = "Elbow method")
```



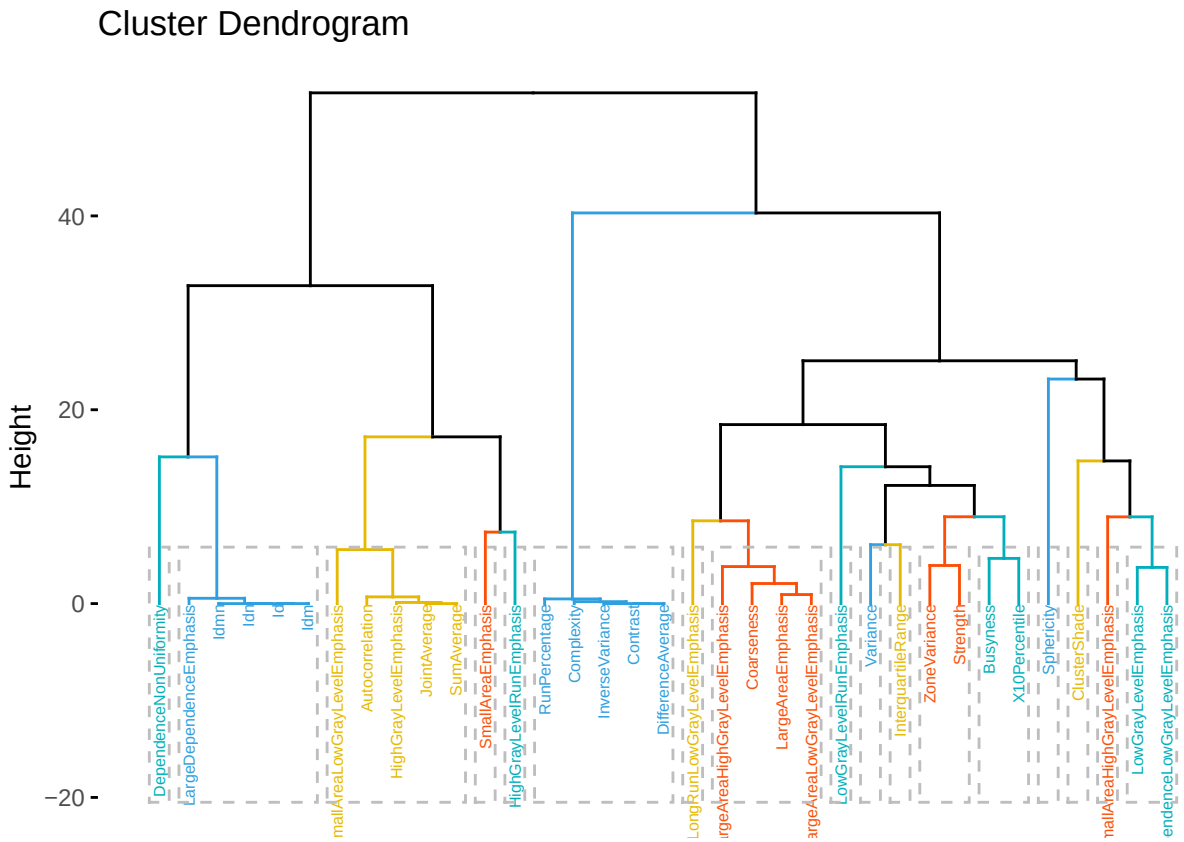
Gap statistics measures how different the total within intra-cluster variation can be between observed data and reference data with a **random uniform distribution**. A large gap statistics means the clustering structure is very far away from the random uniform distribution of points. The number of clusters can be chosen as the smallest value of k such that the gap statistic is within one **standard deviation** of the gap at k+1. In your case, when k = 12, its value is greater than the value that k = 13 minus one standard deviation.

`maxSE(f, SE.f)` is included as an argument of `fviz_gap_stat()` (unlike `fivz_nbclust`) and it determines the location of the maximum of f, taking a "1-SE rule" into account for the *SE* methods. The default method `firstSEmax` looks for the smallest k such that its value $f(k)$ is not more than 1 standard error away from the first local maximum.

```
set.seed(123)

gap_stat <- clusGap(rdr_positive, FUN = kmeans, nstart = 25, K.max = 30, B = 100)
fviz_gap_stat(gap_stat) + theme_minimal() + ggtitle("fviz_gap_stat: Gap Statistic")
```

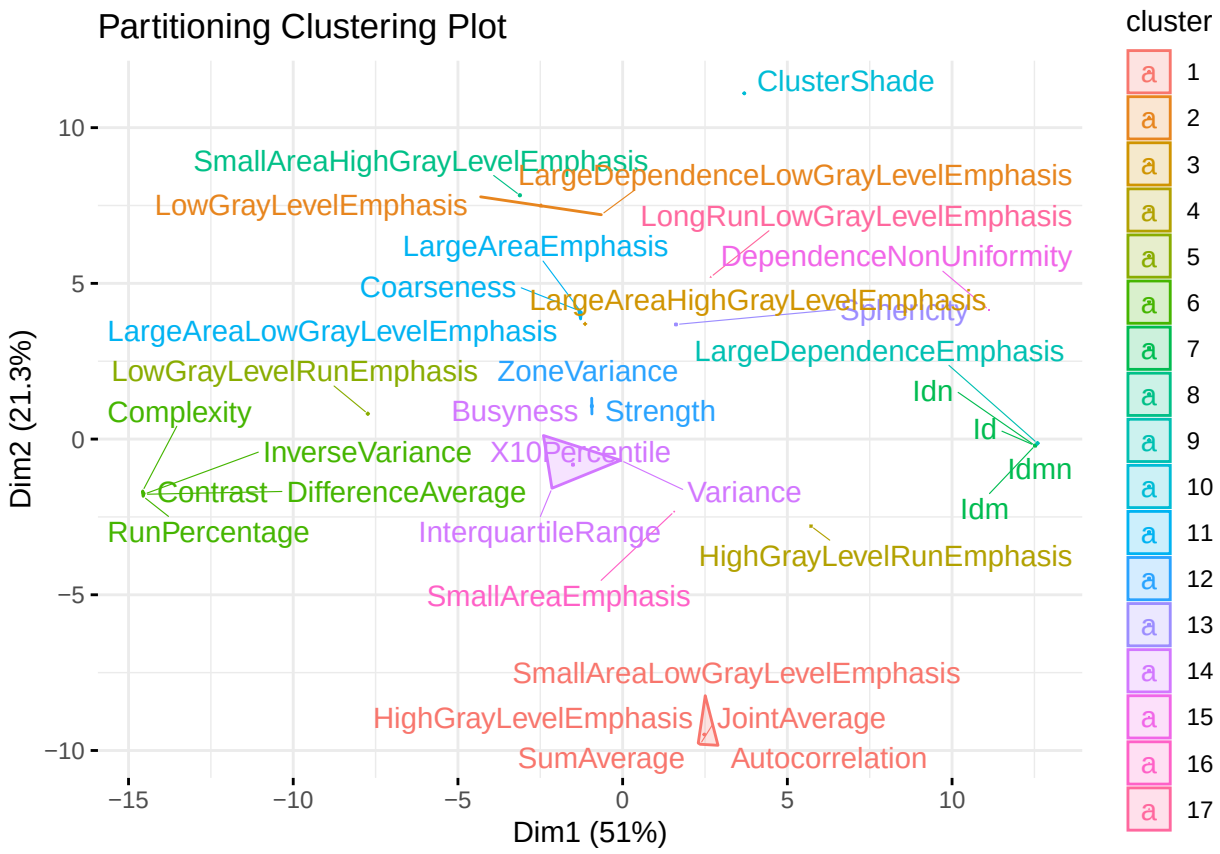



```
# pdf('Dendogram positive features.pdf')

# Compute hierarchical clustering and cut into 4 clusters
res <- hcut(rdr_positive, k = 17, stand = TRUE)
# Visualize
fviz_dend(res, rect = TRUE, cex = 0.4, lwd = 0.5, labels_track_height = 20, k_colors =
c("#00AFBB",
    "#2E9FDF", "#E7B800", "#FC4E07"))
```




```
# dev.off()
```

```
k.res <- kmeans(rdr_positive, centers = 17, nstart = 25)

fviz_cluster(k.res, geom = "text", data = rdr_positive, ggtheme = theme_minimal(),
    main = "Partitioning Clustering Plot", repel = TRUE)
```


Partitioning Clustering Plot
cluster
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
ClusterShade
SmallAreaHighGrayLevelEmphasis
LargeDependenceLowGrayLevelEmphasis
LowGrayLevelEmphasis
LongRunLowGrayLevelEmphasis
LargeAreaEmphasis
DependenceNonUniformity
Coarseness
LargeAreaHighGrayLevelEmphasis
Sphericity
LargeAreaLowGrayLevelEmphasis
LargeDependenceEmphasis
LowGrayLevelRunEmphasis
ZoneVariance
Idn
Complexity
Busyness
Strength
Id
InverseVariance
X10Percentile
Idmn
Contrast
DifferenceAverage
Variance
Idm
RunPercentage
InterquartileRange
HighGrayLevelRunEmphasis
SmallAreaEmphasis
SmallAreaLowGrayLevelEmphasis
HighGrayLevelEmphasis
JointAverage
SumAverage
Autocorrelation
10
5
0
−5
−10
−15
−10
−5
0
5
10
Dim2 (21.3%)
Dim1 (51%)