# MODEL 3 CLUSTERING TECHNIQUE

REY P. PENDANG

2022-12-16

## Compare the following clustering technique;

1 K-Means:

K-means clustering is the most common partitioning algorithm. K-means reassigns each data in the dataset to only one of the new clusters formed

2 Hierarchical:

Hierarchical clustering is separating data into groups based on some measure of similarity, finding a way to measure how they're alike and different, and further narrowing down the data.

3 Model Based:

Model-based clustering is a statistical approach to data clustering. The observed (multivariate) data is assumed to have been generated from a finite mixture of component models., without considering the binary output and categorical variables in the dataset.

```
# Helper packages

library(dplyr)         # for data wrangling
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)     # for filtering
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.4.0     v purrr   0.3.5
## v tibble  3.1.8     v stringr 1.4.1
## v tidyr   1.2.1     v forcats 0.5.2
## v readr   2.1.3
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)         # load dataset
library(bestNormalize) # for normalizing the dataset
library(ggplot2)       # data visualization
```

```
library(stringr)       # for string functionality
library(gridExtra)     # for manipulaiting the grid
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(mclust)        # for model-based clustering
```

```
## Package 'mclust' version 6.0.0
## Type 'citation("mclust")' for citing this R package in publications.
##
## Attaching package: 'mclust'
##
## The following object is masked from 'package:purrr':
##
##     map
```

```
library(cluster)       # for general clustering algorithms
library(factoextra)    # for visualizing cluster results
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

## Load the data set

The data frame output of data reprocessing converted into to "csv", which will be used for entire project.

```
dt <- read_csv("normalRad.CSV")
```

```
## Rows: 197 Columns: 431
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (1): Institution
## dbl (430): Failure.binary, Failure, Entropy_cooc.W.ADC, GLNU_align.H.PET, Mi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(dt)

head(dt)
```

```
## # A tibble: 6 x 431
##   Institution Failure.~1 Failure Entro~2 GLNU_~3 Min_h~4 Max_h~5 Mean_~6 Varia~7
##   <chr>            <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 A                    0   1.15     12.9  -0.433  -0.270 -0.257   -0.192  0.0509
## 2 A                    1  -0.533    12.2  -1.02    0.671  0.405    0.490  0.687
## 3 A                    0   2.24     12.8   0.179  -1.41  -1.57    -1.53  -1.57
## 4 A                    1  -0.140    13.5   2.00   -0.218  0.0764  -0.153  0.0127
## 5 A                    0   0.787    12.6   0.153  -1.06  -1.15    -1.45  -1.91
## 6 A                    1  -2.80     13.2   0.391  -1.57  -1.91    -1.72  -1.84
## # ... with 422 more variables: Standard_Deviation_hist.PET <dbl>,
## #   Skewness_hist.PET <dbl>, Kurtosis_hist.PET <dbl>, Energy_hist.PET <dbl>,
```

```
## #   Entropy_hist.PET <dbl>, AUC_hist.PET <dbl>, H_suv.PET <dbl>,
## #   Volume.PET <dbl>, X3D_surface.PET <dbl>, ratio_3ds_vol.PET <dbl>,
## #   ratio_3ds_vol_norm.PET <dbl>, irregularity.PET <dbl>,
## #   tumor_length.PET <dbl>, Compactness_v1.PET <dbl>, Compactness_v2.PET <dbl>,
## #   Spherical_disproportion.PET <dbl>, Sphericity.PET <dbl>, ...
```

## Standardizing the Data

Standardizing data in R can be done by using the standard scale () command and adequate values with various columns of data. It would be best to use the min-max normalization and min-max scaling procedures for large or small data, including the standard scale () command.

```
df <- scale(dt[c(3:431)])
```

## Check for null and missing values

Using *sum(is.n())*.function, We can determine if any missing values in our data.

#The result shows either *True* or *False*. If True, omit the missing values using *na.omit()*

#[1] FALSE

#Thus, our data has no missing values.

```
sum(is.na(df))
```

```
## [1] 0
```

# 1 K-MEANS

```
kmeans(df, centers = 3, iter.max = 100, nstart = 100)
```

```
## K-means clustering with 3 clusters of sizes 44, 103, 50
##
## Cluster means:
##       Failure Entropy_cooc.W.ADC GLNU_align.H.PET Min_hist.PET Max_hist.PET
## 1 -0.24638949         0.05200710      0.126697686   0.04585696   0.08504594
## 2  0.06537718        -0.04573823     -0.009433942  -0.43645422  -0.46197403
## 3  0.08214575         0.04845450     -0.092060043   0.85874157   0.87682607
##   Mean_hist.PET Variance_hist.PET Standard_Deviation_hist.PET Skewness_hist.PET
## 1    0.08479383         0.2279436                   0.0410318        -0.5428415
## 2   -0.45868782        -0.3330976                  -0.4359917        -0.1564058
## 3    0.87027835         0.4855907                   0.8620350         0.7998965
##   Kurtosis_hist.PET Energy_hist.PET Entropy_hist.PET AUC_hist.PET  H_suv.PET
## 1       -0.06878862      -1.2406345       -0.1535503  -1.20376496 -0.1721421
## 2        0.04878509       0.1319279       -0.5425300  -0.09790924 -0.3530821
## 3       -0.03996330       0.8199868        1.2527361   1.26100621  0.8788341
##   Volume.PET X3D_surface.PET ratio_3ds_vol.PET ratio_3ds_vol_norm.PET
## 1  0.4518410       0.1229739      -1.032381320            -0.92503862
## 2 -0.4473168      -0.3137553      -0.002504806            -0.05148286
## 3  0.5238526       0.5381188       0.913655462             0.92008868
##   irregularity.PET tumor_length.PET Compactness_v1.PET Compactness_v2.PET
## 1       -0.9587467       -0.3148322        -1.06130280         0.00075632
## 2       -0.2025772       -0.3577728        -0.04053042        -0.33148005
## 3        1.2610062        1.0140643         1.01743912         0.68218335
```

```
##   Spherical_disproportion.PET Sphericity.PET Asphericity.PET Center_of_mass.PET
## 1                  -0.92503862      0.0083911     -0.92336330         -0.1664062
## 2                  -0.05148286     -0.4090411     -0.04633856         -0.2830070
## 3                   0.92008868      0.8352405      0.90801714          0.7294319
##   Max_3D_diam.PET Major_axis_length.PET Minor_axis_length.PET
## 1       0.4049083             0.2923557            0.09513965
## 2      -0.5666852            -0.5486413           -0.54545457
## 3       0.8110522             0.8729280            1.03991352
##   Least_axis_length.PET Elongation.PET Flatness.PET Max_cooc.L.PET
## 1             0.2611157     -0.8083506   -0.5663013     -1.2377046
## 2            -0.5446258     -0.2539493   -0.3408235      0.1175852
## 3             0.8921472      1.2344841    1.2004416      0.8469546
##   Average_cooc.L.PET Variance_cooc.L.PET Entropy_cooc.L.PET DAVE_cooc.L.PET
## 1         -0.4959192          -0.6739093         -0.2685377      -0.6453081
## 2         -0.3625706          -0.1758933         -0.4974238      -0.2725761
## 3          1.1833043           0.9553804          1.2610062       1.1293779
##   DVAR_cooc.L.PET DENT_cooc.L.PET SAVE_cooc.L.PET SVAR_cooc.L.PET
## 1      -0.6767720      -0.7056993      -0.4941932      -0.6640316
## 2      -0.1910234      -0.3106751      -0.3631785      -0.1977607
## 3       0.9890675       1.2610062       1.1830377       0.9917348
##   SENT_cooc.L.PET ASM_cooc.L.PET Contrast_cooc.L.PET Dissimilarity_cooc.L.PET
## 1      -0.9317152     -1.2294851          -0.5547851               -0.6453081
## 2      -0.2141247      0.1319836          -0.1507953               -0.2725761
## 3       1.2610062      0.8100606           0.7988492                1.1293779
##   Inv_diff_cooc.L.PET Inv_diff_norm_cooc.L.PET IDM_cooc.L.PET
## 1          -0.7775483               -0.8141946     -0.8857651
## 2          -0.2739537               -0.2643276     -0.1969052
## 3           1.2485871                1.2610062      1.1850979
##   IDM_norm_cooc.L.PET Inv_var_cooc.L.PET Correlation_cooc.L.PET
## 1          -0.9675063         -0.9013253             -0.3586545
## 2          -0.1988353         -0.1927562             -0.3350398
## 3           1.2610062          1.1902440              1.0057979
##   Autocorrelation_cooc.L.PET Tendency_cooc.L.PET Shade_cooc.L.PET
## 1                 -0.4282134          -0.6640316      -0.37689090
## 2                 -0.2523420          -0.1977607      -0.06389272
## 3                  0.8966523           0.9917348       0.46328300
##   Prominence_cooc.L.PET  IC1_.L.PET IC2_.L.PET Coarseness_vdif_.L.PET
## 1           -0.64475251 -0.02866924  -1.001224             -1.2476681
## 2           -0.06490617  0.28950284  -0.182536              0.1389880
## 3            0.70108893 -0.57114691   1.257101              0.8116326
##   Contrast_vdif_.L.PET Busyness_vdif_.L.PET Complexity_vdif_.L.PET
## 1           -0.7058882            0.2287172             -0.8095809
## 2           -0.0251084           -0.3710415             -0.1975675
## 3            0.6729049            0.5630744              1.1194203
##   Strength_vdif_.L.PET SRE_align.L.PET LRE_align.L.PET GLNU_align.L.PET
## 1           -0.8803918      -1.2869226      -0.7774763        0.4274795
## 2            0.1338226      -0.0623856      -0.2800131       -0.3996385
## 3            0.4990703       1.2610062       1.2610062        0.4470734
##   RLNU_align.L.PET RP_align.L.PET LGRE_align.L.PET HGRE_align.L.PET
## 1        0.4745056     -1.2784424    -1.1301238429       -0.4400459
## 2       -0.4023296     -0.0660082     0.0003068724       -0.2607318
## 3        0.4112341      1.2610062     0.9938768247        0.9243479
##   LGSRE_align.L.PET HGSRE_align.L.PET LGHRE_align.L.PET HGLRE_align.L.PET
## 1      -1.137863173        -0.4547746     -1.0946362665        -0.3821475
```

```
## 2       0.001502125        -0.2534191     -0.0002882839        -0.2916630
## 3       0.998225215         0.9222450      0.9638737792         0.9371156
##   GLNU_norm_align.L.PET RLNU_norm_align.L.PET GLVAR_align.L.PET
## 1           -1.29006007           -1.22389485        -0.6241077
## 2            0.05026552           -0.08931007        -0.2128020
## 3            1.03170588            1.26100621         0.9875870
##   RLVAR_align.L.PET Entropy_align.L.PET SZSE.L.PET LZSE.L.PET  LGLZE.L.PET
## 1       -1.10198903          -0.3048233 -1.1703808 -0.2560081 -1.141762407
## 2       -0.02325062          -0.4819232 -0.1121704 -0.4488779 -0.002221956
## 3        1.01764663           1.2610062  1.2610062  1.1499756  1.009328147
##   HGLZE.L.PET  SZLGE.L.PET SZHGE.L.PET LZLGE.L.PET LZHGE.L.PET GLNU_area.L.PET
## 1  -0.4547108 -1.169840211  -0.5176669 -0.93017679  -0.2706204       0.4167183
## 2  -0.2603457  0.001672616  -0.2328767 -0.02473818  -0.3087426      -0.4004704
## 3   0.9364576  1.026013796   0.9352729  0.86951623   0.8741557       0.4582570
##   ZSNU.L.PET  ZSP.L.PET GLNU_norm.L.PET ZSNU_norm.L.PET GLVAR_area.L.PET
## 1  0.4716323 -1.0708645     -1.28850827      -1.0111507       -0.6380485
## 2 -0.4105445 -0.1546823      0.04990501      -0.1801911       -0.2148491
## 3  0.4306853  1.2610062      1.03108296       1.2610062        1.0040719
##   ZSVAR.L.PET Entropy_area.L.PET Max_cooc.H.PET Average_cooc.H.PET
## 1  -0.5205423         -0.2300047     -1.0446699         -0.8674055
## 2  -0.2133959         -0.5138845      0.1789068         -0.2415968
## 3   0.8976727          1.2610062      0.5507615          1.2610062
##   Variance_cooc.H.PET Entropy_cooc.H.PET DAVE_cooc.H.PET DVAR_cooc.H.PET
## 1          -0.1076677         -0.1513560      -0.3370926      -0.3491879
## 2          -0.5538043         -0.5131234      -0.4673835      -0.4582690
## 3           1.2355844          1.1902276       1.2594515       1.2513196
##   DENT_cooc.H.PET SAVE_cooc.H.PET SVAR_cooc.H.PET SENT_cooc.H.PET
## 1      -0.1766716      -0.7375421      -0.2181004      -0.7930772
## 2      -0.5105001      -0.2970724      -0.5089699      -0.1374373
## 3       1.2071013       1.2610062       1.2404064       0.9810288
##   ASM_cooc.H.PET Contrast_cooc.H.PET Dissimilarity_cooc.H.PET
## 1     -1.1508857          -0.3228096               -0.3370926
## 2      0.2173848          -0.4441699               -0.4673835
## 3      0.5649668           1.1990625                1.2594515
##   Inv_diff_cooc.H.PET Inv_diff_norm_cooc.H.PET IDM_cooc.H.PET
## 1          -0.9451290               -1.1631295    -0.92276752
## 2          -0.1146295               -0.1152681    -0.05603959
## 3           1.0678502                1.2610062     0.92747697
##   IDM_norm_cooc.H.PET Inv_var_cooc_.H.PET Correlation_cooc.H.PET
## 1          -1.2317778         -1.13943252             -0.2939734
## 2          -0.0859426          0.03016403             -0.3677423
## 3           1.2610062          0.94056271              1.0162458
##   Autocorrelation_cooc.H.PET Tendency_cooc.H.PET Shade_cooc.H.PET
## 1                 -0.8764561         -0.08218198        0.0130268
## 2                 -0.2312947         -0.55524643        0.2897603
## 3                  1.2477485          1.21612779       -0.6083698
##   Prominence_cooc.H.PET IC1_d.H.PET IC2_d.H.PET Coarseness_vdif.H.PET
## 1            0.05661571  -0.5772033  -0.3675008            -1.2296950
## 2           -0.47608722   0.3322440  -0.4237839             0.1366371
## 3            0.93091786  -0.1764838   1.1963955             0.8006591
##   Contrast_vdif.H.PET Busyness_vdif.H.PET Complexity_vdif.H.PET
## 1         -0.56987256           0.5445366           -0.91622789
## 2         -0.05878702          -0.4359746           -0.07510634
## 3          0.62258911           0.4189154            0.96099961
```

```
##   Strength_vdif.H.PET SRE_align.H.PET LRE_align.H.PET RLNU_align.H.PET
## 1          -0.9377278      -0.5658223      -0.7057473        0.5003948
## 2           0.2376470      -0.3704284      -0.2154755       -0.4139000
## 3           0.3356476       1.2610062       1.0649372        0.4122865
##   RP_align.H.PET LGRE_align.H.PET HGRE_align.H.PET LGSRE_align.H.PET
## 1     -0.5202500       -1.1999846       -0.7741663        -1.2007644
## 2     -0.3898962        0.1184958       -0.2737050         0.1188289
## 3      1.2610062        0.8118850        1.2450986         0.8118850
##   HGSRE_align.H.PET LGHRE_align.H.PET HGLRE_align.H.PET GLNU_norm_align.H.PET
## 1        -0.6717061        -1.1998696        -0.5798024           -1.02132727
## 2        -0.3251965         0.1173371        -0.1801486            0.03759345
## 3         1.2610062         0.8141708         0.8813322            0.82132550
##   RLNU_norm_align.H.PET GLVAR_align.H.PET RLVAR_align.H.PET Entropy_align.H.PET
## 1            -0.4214434       -0.07611098       -0.66450875         -0.09446914
## 2            -0.4321049       -0.54797206       -0.01910859         -0.57178319
## 3             1.2610062        1.19580011        0.62413140          1.26100621
##   SZSE.H.PET LZSE.H.PET LGLZE.H.PET HGLZE.H.PET SZLGE.H.PET SZHGE.H.PET
## 1 -0.4470299 -0.2715817  -1.1998381  -0.6201846  -1.2021333  -0.4714018
## 2 -0.4032952 -0.0746649   0.1186081  -0.3370729   0.1196279  -0.3808504
## 3  1.2241745  0.3928016   0.8115249   1.2401327   0.8114437   1.1993854
##   LZLGE.H.PET  LZHGE.H.PET GLNU_area.H.PET ZSNU.H.PET  ZSP.H.PET
## 1  -0.9511959 -0.403751927       0.4030738  0.4712095 -0.2763891
## 2   0.1835682  0.002524456      -0.4066093 -0.3660541 -0.3881463
## 3   0.4589019  0.350101318       0.4829101  0.3394070  1.0428037
##   GLNU_norm.H.PET ZSNU_norm.H.PET GLVAR_area.H.PET ZSVAR_H.PET
## 1      -1.0083639      -0.3167062       -0.1076021 -0.40818170
## 2       0.0191874      -0.3843304       -0.5270962  0.02643966
## 3       0.8478342       1.0704220        1.1805080  0.30473420
##   Entropy_area.H.PET Max_cooc.W.PET Average_cooc.W.PET Variance_cooc.W.PET
## 1        -0.05109651     -1.2468037          0.1122544           0.2337219
## 2        -0.59031130      0.2203858         -0.4616517          -0.3256100
## 3         1.26100621      0.6431925          0.8522186           0.4650814
##   Entropy_cooc.W.PET DAVE_cooc.W.PET DVAR_cooc.W.PET DENT_cooc.W.PET
## 1         -0.1073673     0.006743546       0.1971657      -0.1804813
## 2         -0.5588924    -0.424423328      -0.3224611      -0.5233676
## 3          1.2458016     0.868377736       0.4907641       1.2369607
##   SAVE_cooc.W.PET SVAR_cooc.W.PET SENT_cooc.W.PET ASM_cooc.W.PET
## 1       0.1157009       0.2550944      -0.3620001     -1.2643186
## 2      -0.4625522      -0.3286516      -0.4564338      0.1986421
## 3       0.8510408       0.4525392       1.2588138      0.7033977
##   Contrast_cooc.W.PET Dissimilarity_cooc.W.PET Inv_diff_cooc.W.PET
## 1           0.1722521              0.006743546          -0.9382274
## 2          -0.3226168             -0.424423328          -0.1656669
## 3           0.5130087              0.868377736           1.1669140
##   Inv_diff_norm_cooc.W.PET IDM_cooc.W.PET IDM_norm_cooc.W.PET
## 1               -0.8564369    -0.91054172          -0.9835407
## 2               -0.2462824    -0.08803608          -0.1919857
## 3                1.2610062     0.98263103           1.2610065
##   Inv_var_cooc.W.PET Correlation_cooc.W.PET Autocorrelation_cooc.W.PET
## 1         -0.9377635             -0.3429057                  0.3145434
## 2         -0.1191630             -0.3424528                 -0.3622375
## 3          1.0707077              1.0072099                  0.4694111
##   Tendency_cooc.W.PET Shade_cooc.W.PET Prominence_cooc.W.PET IC1_d.W.PET
## 1           0.2550944        0.2261518             0.3340363  -0.6262563
```

```
## 2          -0.3286516       -0.1931742            -0.2587149   0.3873344
## 3           0.4525392        0.1989252             0.2390007  -0.2468033
##   IC2_d.W.PET Coarseness_vdif.W.PET Contrast_vdif.W.PET Busyness_vdif.W.PET
## 1  -0.4870084            -1.2493989          -0.3598249          -0.1345930
## 2  -0.3973578             0.1744151          -0.2219072          -0.1647786
## 3   1.2471245             0.7401759           0.7737748           0.4578856
##   Complexity_vdif.W.PET Strength_vdif.W.PET SRE_align.W.PET LRE_align.W.PET
## 1             0.2968367         -0.41666689      -0.7719393      -0.7990212
## 2            -0.3066483         -0.09774272      -0.2823784      -0.2477785
## 3             0.3704793          0.56801686       1.2610062       1.2135624
##   GLNU_align.W.PET RLNU_align.W.PET RP_align.W.PET LGRE_align.W.PET
## 1        0.2966597        0.4907296     -0.6843643      -0.99492983
## 2       -0.3664176       -0.4106382     -0.3197891       0.04876726
## 3        0.4937597        0.4140726      1.2610062       0.77507770
##   HGRE_align.W.PET LGSRE_align.W.PET HGSRE_align.W.PET LGHRE_align.W.PET
## 1        0.3173854        -1.0143302         0.3121082       -0.89228814
## 2       -0.3650626         0.0359557        -0.3577645        0.07777343
## 3        0.4727298         0.8185419         0.4623397        0.62500029
##   HGLRE_align.W.PET GLNU_norm_align.W.PET RLNU_norm_align.W.PET
## 1         0.3287373           -1.14518054            -0.5352664
## 2        -0.3797422            0.09076661            -0.3834814
## 3         0.4929801            0.82077967             1.2610062
##   GLVAR_align.W.PET RLVAR_align.W.PET Entropy_align.W.PET SZSE.W.PET
## 1         0.2471427       -0.88234771          -0.0890419 -0.5462398
## 2        -0.3403325        0.03915467          -0.5741016 -0.3787938
## 3         0.4835994        0.69580736           1.2610062  1.2610062
##    LZSE.W.PET LGLZE.W.PET HGLZE.W.PET SZLGE.W.PET SZHGE.W.PET LZLGE.W.PET
## 1 -0.57464968 -1.00132489   0.3032912 -1.08631389   0.2678236  -0.7103528
## 2 -0.05497139  0.03611711  -0.3615562  0.01378553  -0.3414232   0.1222846
## 3  0.61893278  0.80676466   0.4779095  0.92755803   0.4676471   0.3732041
##   LZHGE.W.PET GLNU_area.W.PET ZSNU.W.PET  ZSP.W.PET GLNU_norm.W.PET
## 1   0.2912848       0.3500241  0.4794367 -0.4314773     -1.14253625
## 2  -0.3858164      -0.3916938 -0.3954672 -0.4244237      0.07912406
## 3   0.5384512       0.4988679  0.3927581  1.2540128      0.84243633
##   ZSNU_norm.W.PET GLVAR_area.W.PET ZSVAR.W.PET Entropy_area.W.PET Min_hist.ADC
## 1      -0.3990774        0.2292427 -0.57425688        -0.05591386   -0.2864881
## 2      -0.4345812       -0.3295647  0.05368149        -0.58825341   -0.1380119
## 3       1.2464255        0.4771697  0.39476218         1.26100621    0.5364140
##   Max_hist.ADC Mean_hist.ADC Variance_hist.ADC Standard_Deviation_hist.ADC
## 1   -0.4641644    -0.3892240       -0.57365875                   -0.6682174
## 2   -0.4114837    -0.4407905       -0.09472315                   -0.2638962
## 3    1.2561211     1.2505455        0.69994940                    1.1316575
##   Skewness_hist.ADC Kurtosis_hist.ADC Energy_hist.ADC Entropy_hist.ADC
## 1       -0.06417418        0.05493307      -1.2257127       -0.4451103
## 2       -0.18767783       -0.16671301       0.1307215       -0.4219947
## 3        0.44308961        0.29508771       0.8093408        1.2610062
##   AUC_hist.ADC Volume.ADC X3D_surface.ADC ratio_3ds_vol.ADC
## 1   -0.8193435  0.4140316      -0.0606709        -0.7500699
## 2   -0.2621283 -0.4233601      -0.2753462        -0.1974654
## 3    1.2610066  0.5077740       0.6206037         1.0668403
##   ratio_3ds_vol_norm.ADC irregularity.ADC Compactness_v1.ADC Compactness_v2.ADC
## 1             -0.5669541       -0.6593353        -1.28088272         -0.7096110
## 2             -0.3699450       -0.3304811         0.03142066         -0.2321898
## 3              1.2610062        1.2610062         1.06245022          1.1027687
```

```
##   Spherical_disproportion.ADC Sphericity.ADC Asphericity.ADC Center_of_mass.ADC
## 1                  -0.5669541     -0.8269058      -0.4565552        -0.30911982
## 2                  -0.3699450     -0.2588976      -0.3557791        -0.08729681
## 3                   1.2610062      1.2610062       1.1346735         0.45185687
##   Max_3D_diam.ADC Major_axis_length.ADC Minor_axis_length.ADC
## 1      -0.2274094            -0.3099045            -0.2275549
## 2      -0.3644002            -0.4005435            -0.3813531
## 3       0.9507846             1.0978355             0.9858357
##   Least_axis_length.ADC Elongation.ADC Flatness.ADC Max_cooc.L.ADC
## 1            -0.1585236     -0.4558866   -0.4355862    -1.21985793
## 2            -0.3769359     -0.4157332   -0.4105727     0.08253485
## 3             0.9159888      1.2575906    1.2290955     0.90345319
##   Average_cooc.L.ADC Variance_cooc.L.ADC Entropy_cooc.L.ADC DAVE_cooc.L.ADC
## 1         -0.3336411          -0.6744625         -0.6250607      -0.6279979
## 2         -0.4661050          -0.1419750         -0.3451227      -0.2977163
## 3          1.2537804           0.8859954          1.2610062       1.1659338
##   DVAR_cooc.L.ADC DENT_cooc.L.ADC SAVE_cooc.L.ADC SVAR_cooc.L.ADC
## 1      -0.6797727      -0.7169051      -0.3333451      -0.6649414
## 2      -0.1347764      -0.3058882      -0.4662314      -0.1329086
## 3       0.8758394       1.2610062       1.2537804       0.8589402
##   SENT_cooc.L.ADC ASM_cooc.L.ADC Contrast_cooc.L.ADC Dissimilarity_cooc.L.ADC
## 1      -0.3989158     -1.2087035          -0.5911236               -0.6279979
## 2      -0.3230180      0.1097121          -0.1426598               -0.2977163
## 3       1.0164629      0.8376521           0.8140680                1.1659338
##   Inv_diff_cooc.L.ADC Inv_diff_norm_cooc.L.ADC IDM_cooc.L.ADC
## 1          -0.8112094               -1.0339652     -0.8263762
## 2          -0.2645293               -0.1704452     -0.2355299
## 3           1.2587946                1.2610066      1.2124027
##   IDM_norm_cooc.L.ADC Inv_var_cooc.L.ADC Correlation_cooc.L.ADC
## 1          -1.1872975         -0.8131001              -0.6737162
## 2          -0.1049439         -0.2449825              -0.2080744
## 3           1.2610062          1.2201921               1.0215036
##   Autocorrelation_.L.ADC Tendency_cooc.L.ADC Shade_.L.ADC Prominence_cooc.L.ADC
## 1             -0.3098370          -0.6649414  -0.06135605          -0.632465720
## 2             -0.3809032          -0.1329086  -0.08932906           0.009646047
## 3              1.0573171           0.8589402   0.23801119           0.536698976
##   IC1_.L.ADC IC2_.L.ADC Coarseness_vdif_.L.ADC Contrast_vdif_.L.ADC
## 1 -0.3289130 -0.8510251             -1.2315126         -0.755856168
## 2  0.4224645 -0.2431110              0.1832338         -0.004267165
## 3 -0.5808334  1.2497107              0.7062694          0.673943788
##   Busyness_vdif_.L.ADC Complexity_vdif_.L.ADC Strength_vdif_.L.ADC
## 1           -0.4836846             -0.6206652          -0.35785747
## 2           -0.1267904             -0.2880789          -0.04141227
## 3            0.6868307              1.1396278           0.40022386
##   SRE_align.L.ADC LRE_align.L.ADC GLNU_align.L.ADC RLNU_align.L.ADC
## 1     -1.22921678      -0.7709996      0.009709534      -0.04687914
## 2     -0.08703695      -0.2827799     -0.240932986      -0.22553738
## 3      1.26100689       1.2610062      0.487777562       0.50586065
##   RP_align.L.ADC LGRE_align.L.ADC HGRE_align.L.ADC LGSRE_align.L.ADC
## 1     -1.1855243       -1.2395592       -0.3623635         -1.239364
## 2     -0.1057014        0.1449922       -0.3947405          0.144266
## 3      1.2610062        0.7921281        1.1320453          0.793452
##   HGSRE_align.L.ADC LGHRE_align.L.ADC HGLRE_align.L.ADC GLNU_norm_align.L.ADC
## 1        -0.3720836        -1.2411265        -0.3209441           -1.22721402
```

```
## 2        -0.3915385         0.1519871        -0.4122198           -0.04116189
## 3         1.1340029         0.7790980         1.1316036            1.16474183
##   RLNU_norm_align.L.ADC GLVAR_align.L.ADC RLVAR_align.L.ADC Entropy_align.L.ADC
## 1            -1.0572248        -0.6888250       -0.99448887          -0.7419215
## 2            -0.1605091        -0.1603817       -0.09828869          -0.2952016
## 3             1.2610066         0.9365524        1.07762492           1.2610062
##   SZSE.L.ADC LZSE.L.ADC LGLZE.L.ADC HGLZE.L.ADC SZLGE.L.ADC SZHGE.L.ADC
## 1 -1.0522267 -0.4389908  -1.2390250  -0.3644243  -1.2356743  -0.3749926
## 2 -0.1626442 -0.3956262   0.1400977  -0.4020896   0.1377293  -0.3952206
## 3  1.2610065  1.2013018   0.8017407   1.1489979   0.8036711   1.1441480
##   LZLGE.L.ADC LZHGE.L.ADC GLNU_area.L.ADC ZSNU.L.ADC  ZSP.L.ADC GLNU_norm.L.ADC
## 1  -1.2340635  -0.2673208    -0.003034725 -0.0462503 -0.8818584     -1.24469430
## 2   0.1873437  -0.4083600    -0.238174927 -0.2279717 -0.2354227     -0.03059644
## 3   0.7000477   1.0764639     0.493310907  0.5103219  1.2610062      1.15835965
##   ZSNU_norm.L.ADC GLVAR_area.L.ADC ZSVAR.L.ADC Entropy_area.L.ADC
## 1      -0.8424410       -0.6981154  -0.3889422         -0.6973508
## 2      -0.2522613       -0.1628163  -0.2001315         -0.3142415
## 3       1.2610063        0.9497431   0.7545401          1.2610062
##   Max_cooc.H.ADC Average_cooc.H.ADC Variance_cooc.H.ADC Entropy_cooc.H.ADC
## 1     -1.2204511         -0.4651563          -0.3955399         -0.2458259
## 2      0.1227091         -0.4134314          -0.4431704         -0.5071259
## 3      0.8212163          1.2610062           1.2610062          1.2610062
##   DAVE_cooc.H.ADC DVAR_cooc.H.ADC DENT_cooc.H.ADC SAVE_cooc.H.ADC
## 1      -0.4109838      -0.4410205      -0.4877151      -0.4577571
## 2      -0.4365730      -0.4225223      -0.4037945      -0.4165922
## 3       1.2610062       1.2584940       1.2610060       1.2610062
##   SVAR_cooc.H.ADC SENT_cooc.H.ADC ASM_cooc.H.ADC Contrast_cooc.H.ADC
## 1      -0.5527784      -0.4829811     -1.2196578          -0.3612323
## 2      -0.3760006      -0.4058169      0.1277874          -0.4344390
## 3       1.2610062       1.2610062      0.8100568           1.2128288
##   Dissimilarity_cooc.H.ADC Inv_diff_cooc.H.ADC Inv_diff_norm_cooc.H.ADC
## 1               -0.4109838          -1.180666               -1.1778345
## 2               -0.4365730          -0.107623               -0.1089863
## 3                1.2610062           1.260690                1.2610062
##   IDM_cooc.H.ADC IDM_norm_cooc.H.ADC Inv_var_cooc.H.ADC Correlation_cooc.H.ADC
## 1    -1.20639848         -1.21325796        -1.23708979             -0.6759406
## 2    -0.08635799         -0.09385398        -0.07470963             -0.2054811
## 3     1.23952812          1.26100621         1.24254086              1.0181188
##   Autocorrelation_cooc.H.ADC Tendency_cooc.H.ADC Shade_cooc.H.ADC
## 1                 -0.5596337          -0.5527784       -0.0402483
## 2                 -0.3730721          -0.3760006       -0.1719959
## 3                  1.2610062           1.2610062        0.3897301
##   Prominence_cooc.H.ADC IC1_d.H.ADC IC2_d.H.ADC Coarseness_vdif.H.ADC
## 1            -0.5278320  -0.2932829  -0.7751245            -1.2283655
## 2            -0.3866573   0.3519148  -0.2796836             0.1823866
## 3             1.2610062  -0.4668555   1.2582578             0.7052453
##   Contrast_vdif.H.ADC Busyness_vdif.H.ADC Complexity_vdif.H.ADC
## 1          -0.4519574          -0.2333788            -0.3779299
## 2          -0.4189158          -0.1962940            -0.4478493
## 3           1.2606890           0.6097390             1.2551479
##   Strength_vdif.H.ADC SRE_align.H.ADC LRE_align.H.ADC GLNU_align.H.ADC
## 1         -0.31471124     -1.28600125      -1.0985272      -0.05173676
## 2         -0.03718193     -0.06277931      -0.1428652      -0.22768808
## 3          0.35354067      1.26100647       1.2610062       0.51456579
```

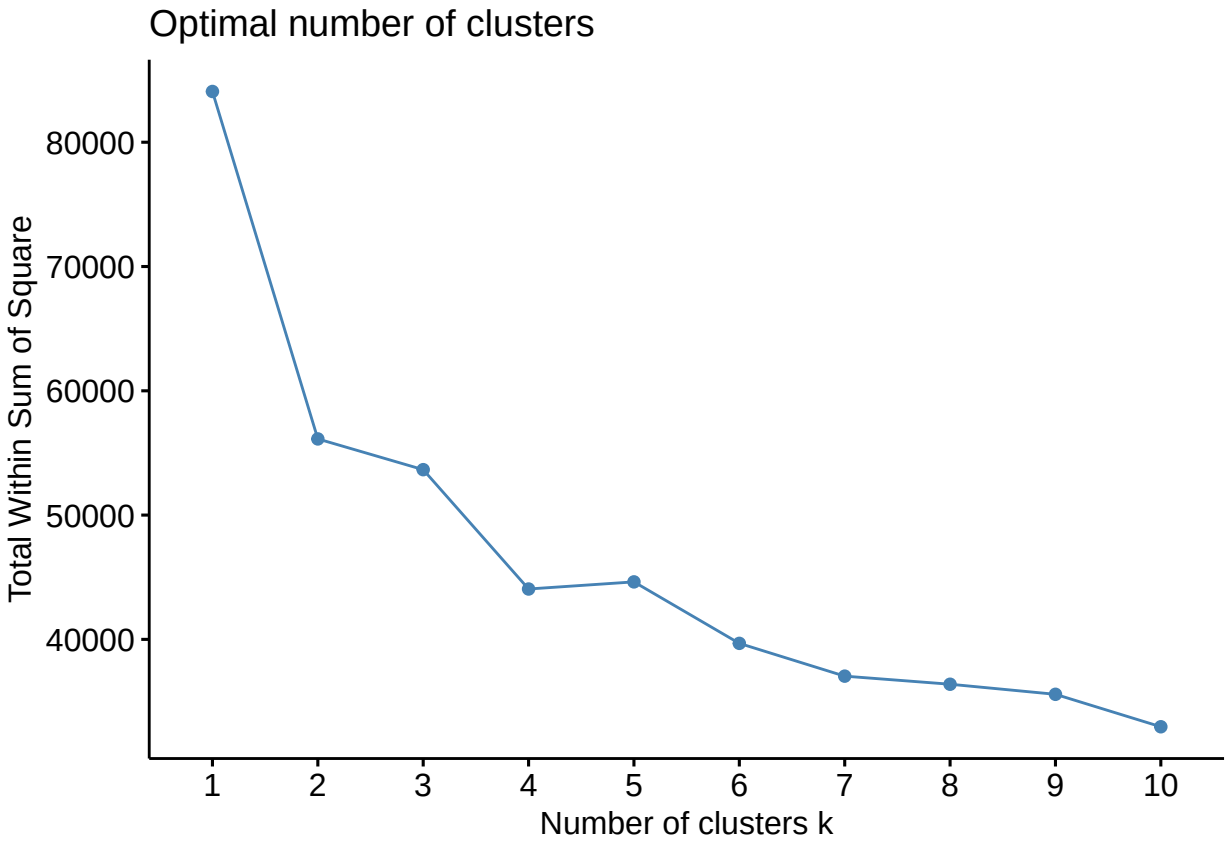
```
##   RLNU_align.H.ADC RP_align.H.ADC LGRE_align.H.ADC HGRE_align.H.ADC
## 1      -0.04867056    -1.27361922      -1.27765222       -0.5639565
## 2      -0.23013578    -0.06806875       0.03987414       -0.3712255
## 3       0.51690981     1.26100654       1.04219323        1.2610062
##   LGSRE_align.H.ADC HGSRE_align.H.ADC LGHRE_align.H.ADC HGLRE_align.H.ADC
## 1       -1.27663363        -0.4695871      -1.272360535        -0.3919663
## 2        0.04337941        -0.4115386       0.007475531        -0.4446970
## 3        1.03407602         1.2610062       1.104277676         1.2610062
##   GLNU_norm_align.H.ADC RLNU_norm_align.H.ADC GLVAR_align.H.ADC
## 1           -1.25810344             -1.202406        -0.4587517
## 2            0.05820082             -0.098490        -0.4161673
## 3            0.98723733              1.261007         1.2610062
##   RLVAR_align.H.ADC Entropy_align.H.ADC SZSE.H.ADC LZSE.H.ADC LGLZE.H.ADC
## 1      -1.234720843          -0.7049199 -1.1335874 -0.7280354 -1.26510163
## 2       0.006019786          -0.3110083 -0.1278884 -0.3011638  0.04242133
## 3       1.074153583           1.2610065  1.2610069  1.2610685  1.02590149
##   HGLZE.H.ADC SZLGE.H.ADC SZHGE.H.ADC  LZLGE.H.ADC LZHGE.H.ADC GLNU_area.H.ADC
## 1  -0.6239460 -1.25721240  -0.4628732 -1.178895154  -0.5645359      -0.0463249
## 2  -0.3455989  0.05397137  -0.4144067 -0.003059701  -0.3705110      -0.2300161
## 3   1.2610062  0.99516588   1.2610062  1.043730718   1.2600443       0.5145991
##    ZSNU.H.ADC  ZSP.H.ADC GLNU_norm.H.ADC ZSNU_norm.H.ADC GLVAR_area.H.ADC
## 1 -0.03811053 -0.9997149     -1.26118455      -0.8740728       -0.3962385
## 2 -0.23482238 -0.1850766      0.05953909      -0.2387489       -0.4428720
## 3  0.51727138  1.2610069      0.98719188       1.2610069        1.2610062
##   ZSVAR.H.ADC Entropy_area.H.ADC Max_cooc.W.ADC Average_cooc.W.ADC
## 1 -0.87026837         -0.6180253     -1.2235870         -0.3007142
## 2 -0.02620128         -0.3481281      0.1290821         -0.3723147
## 3  0.81981081          1.2610062      0.8108473          1.0315968
##   Variance_cooc.W.ADC DAVE_cooc.W.ADC DVAR_cooc.W.ADC DENT_cooc.W.ADC
## 1         -0.57415096      -0.6388969     -0.61163767      -0.7149309
## 2         -0.08008876      -0.2961704     -0.09076978      -0.3067316
## 3          0.67023569       1.1723403      0.72522689       1.2610062
##   SAVE_cooc.W.ADC SVAR_cooc.W.ADC SENT_cooc.W.ADC ASM_cooc.W.ADC
## 1      -0.3121548     -0.52230929      -0.3572574     -1.2193996
## 2      -0.3680932     -0.07792817      -0.3284037      0.1276739
## 3       1.0329682      0.62016420       0.9908981      0.8100633
##   Contrast_cooc.W.ADC Dissimilarity_cooc.W.ADC Inv_diff_cooc.W.ADC
## 1          -0.5852988               -0.6388969          -1.0941138
## 2          -0.1102574               -0.2961704          -0.1119249
## 3           0.7421931                1.1723403           1.1933853
##   Inv_diff_norm_cooc.W.ADC IDM_cooc.W.ADC IDM_norm_cooc.W.ADC
## 1               -1.0332091    -1.15194943           -1.184432
## 2               -0.1707681    -0.09023132           -0.106168
## 3                1.2610062     1.19959202            1.261006
##   Inv_var_cooc.W.ADC Correlation_cooc.W.ADC Autocorrelation_cooc.W.ADC
## 1        -1.13055941             -0.6715404                 -0.2375357
## 2        -0.09551775             -0.2088642                 -0.2458585
## 3         1.19165885              1.0212158                  0.7154998
##   Tendency_cooc.W.ADC Shade_cooc.W.ADC Prominence_cooc.W.ADC IC1_d.W.ADC
## 1         -0.52230929      -0.01899192            -0.4123323 -0.09228064
## 2         -0.07792817      -0.08233801             0.0284408  0.32940186
## 3          0.62016420       0.18632920             0.3042644 -0.59736087
##   IC2_d.W.ADC Coarseness_vdif.W.ADC Contrast_vdif.W.ADC Busyness_vdif.W.ADC
## 1  -0.8163227            -1.2320377         -0.67799661         -1.09537673
```

```
## 2  -0.2634185             0.1694735         -0.03596539         -0.01152585
## 3   1.2610062             0.7350778          0.67072573          0.98767477
##   Complexity_vdif.W.ADC Strength_vdif.W.ADC SRE_align.W.ADC LRE_align.W.ADC
## 1            -0.2915629         -0.55696810     -1.29436439      -1.1596097
## 2            -0.1148090         -0.04369006     -0.05920658      -0.1167717
## 3             0.4930819          0.58013345      1.26100621       1.2610062
##   GLNU_align.W.ADC RLNU_align.W.ADC RP_align.W.ADC LGRE_align.W.ADC
## 1       0.02983217      -0.03895424    -1.29131932       -1.2373352
## 2      -0.28927950      -0.22823909    -0.06050738        0.1438929
## 3       0.56966346       0.50445226     1.26100621        0.7924356
##   HGRE_align.W.ADC LGSRE_align.W.ADC HGSRE_align.W.ADC LGHRE_align.W.ADC
## 1       -0.2676759        -1.2366483        -0.2703353        -1.2383055
## 2       -0.2417845         0.1424456        -0.2405162         0.1510334
## 3        0.7336310         0.7948126         0.7333584         0.7785801
##   HGLRE_align.W.ADC GLNU_norm_align.W.ADC RLNU_norm_align.W.ADC
## 1        -0.2695919            -1.2413161           -1.26346860
## 2        -0.2442399             0.0744557           -0.07240495
## 3         0.7403751             0.9389794            1.26100657
##   GLVAR_align.W.ADC RLVAR_align.W.ADC Entropy_align.W.ADC  SZSE.W.ADC
## 1       -0.57962404       -1.22478921          -0.4229567 -1.21016596
## 2       -0.09723098        0.03271082          -0.4314584 -0.09517497
## 3        0.71036498        1.01043022           1.2610062  1.26100648
##   LZSE.W.ADC LGLZE.W.ADC HGLZE.W.ADC SZLGE.W.ADC SZHGE.W.ADC LZLGE.W.ADC
## 1 -0.8753990  -1.2341484  -0.2682391  -1.2305634  -0.2695190  -1.2416075
## 2 -0.2381821   0.1397825  -0.2416793   0.1360943  -0.2403086   0.1828247
## 3  1.2610062   0.7980986   0.7339098   0.8025415   0.7322124   0.7159958
##   LZHGE.W.ADC GLNU_area.W.ADC  ZSNU.W.ADC  ZSP.W.ADC GLNU_norm.W.ADC
## 1  -0.2647312      0.03296213 -0.03038134 -1.1754768     -1.24278422
## 2  -0.2543883     -0.29235655 -0.22628969 -0.1099935      0.08404481
## 3   0.7570032      0.57324783  0.49289235  1.2610062      0.92051780
##   ZSNU_norm.W.ADC GLVAR_area.W.ADC ZSVAR.W.ADC Entropy_area.W.ADC
## 1      -1.0749981       -0.5801828 -1.07632587         -0.5280456
## 2      -0.1529165       -0.1004131 -0.03697058         -0.3865660
## 3       1.2610062        0.7174119  1.02332617          1.2610062
##
## Clustering vector:
##   [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##  [38] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##  [75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 2 2 2 2 2 2
## [112] 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 3
## [149] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
## [186] 3 3 3 3 3 3 3 3 3 3 3 3
##
## Within cluster sum of squares by cluster:
## [1] 10412.77 24997.15 13419.94
##  (between_SS / total_SS =  41.9 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

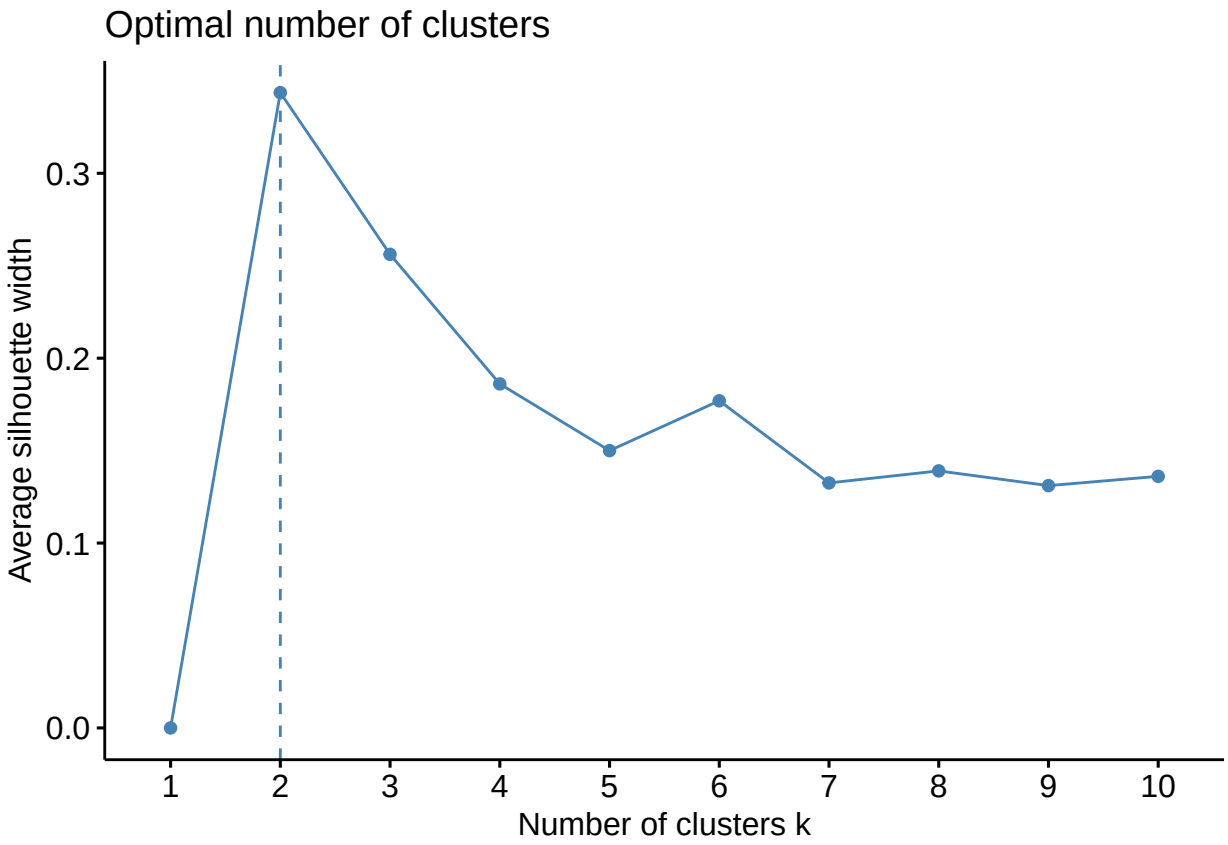
```
clusters <- kmeans(df, centers = 3, iter.max = 100, nstart = 100)
```

## DETERMINING AND VISUALIZING OPTIMAL NUMBER OF CLUSTERS

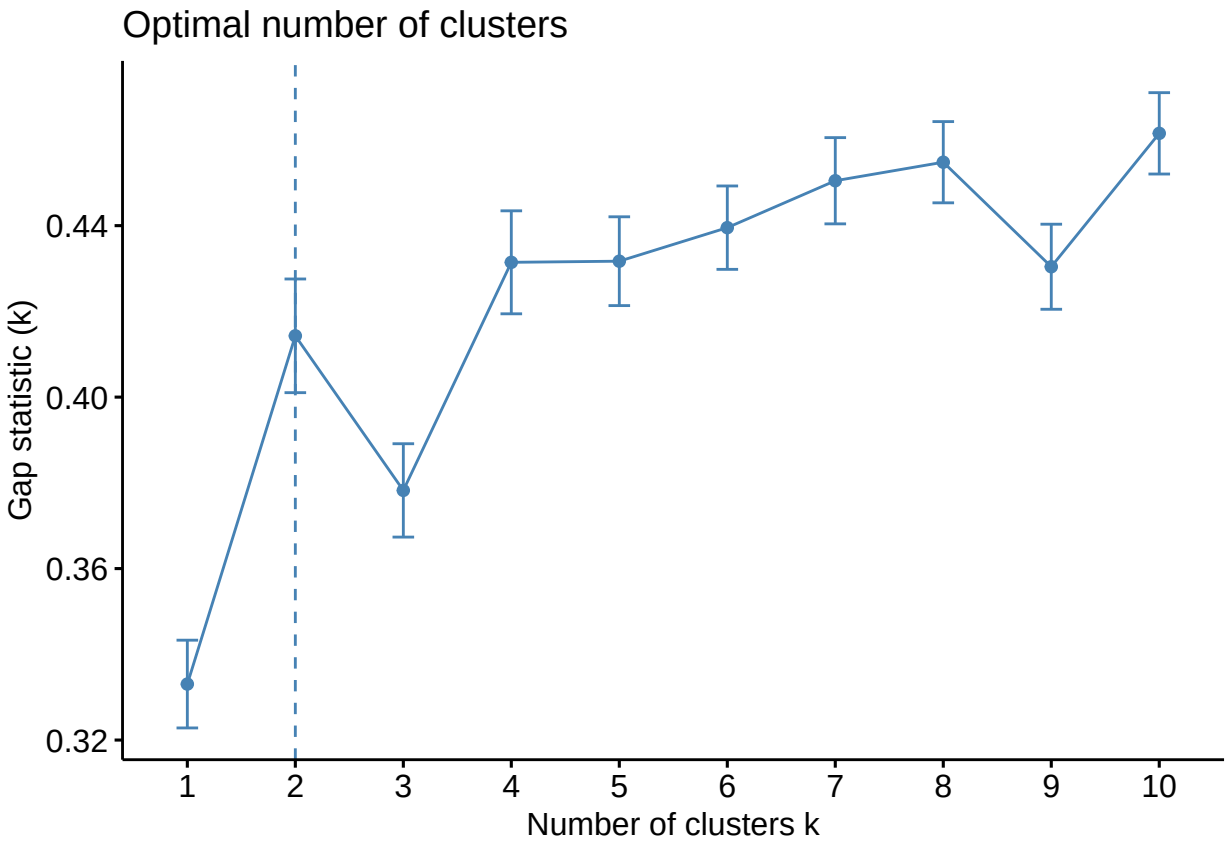
```
set.seed(123) # Determining Optimal Number of Clusters

fviz_nbclust(df, kmeans, method = "wss")
```



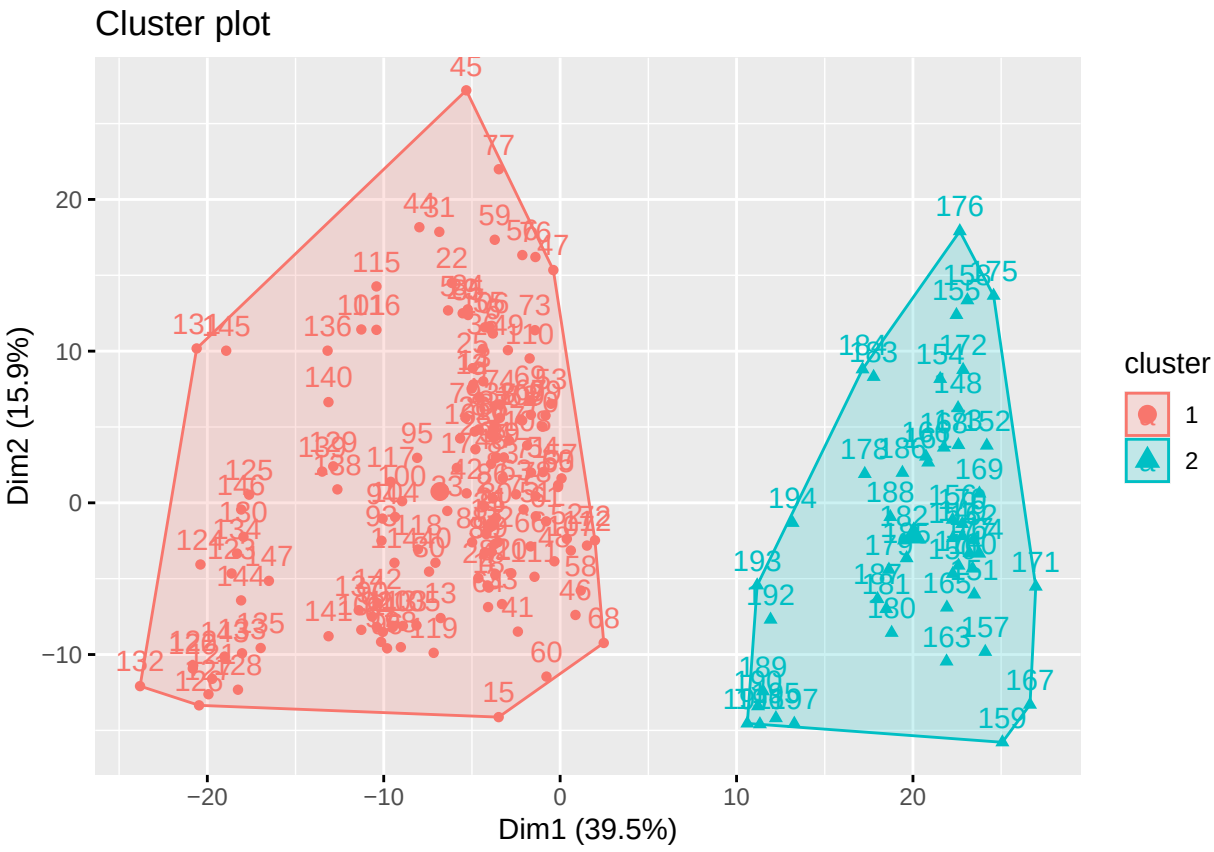

```
fviz_nbclust(df, kmeans, method = "silhouette")
```




```
fviz_nbclust(df, kmeans, method = "gap_stat")
```




```
clusters <- kmeans(df, centers = 2, iter.max = 100, nstart = 100)
fviz_cluster(kmeans(df, centers = 2, iter.max = 100, nstart = 100), data = df)
```



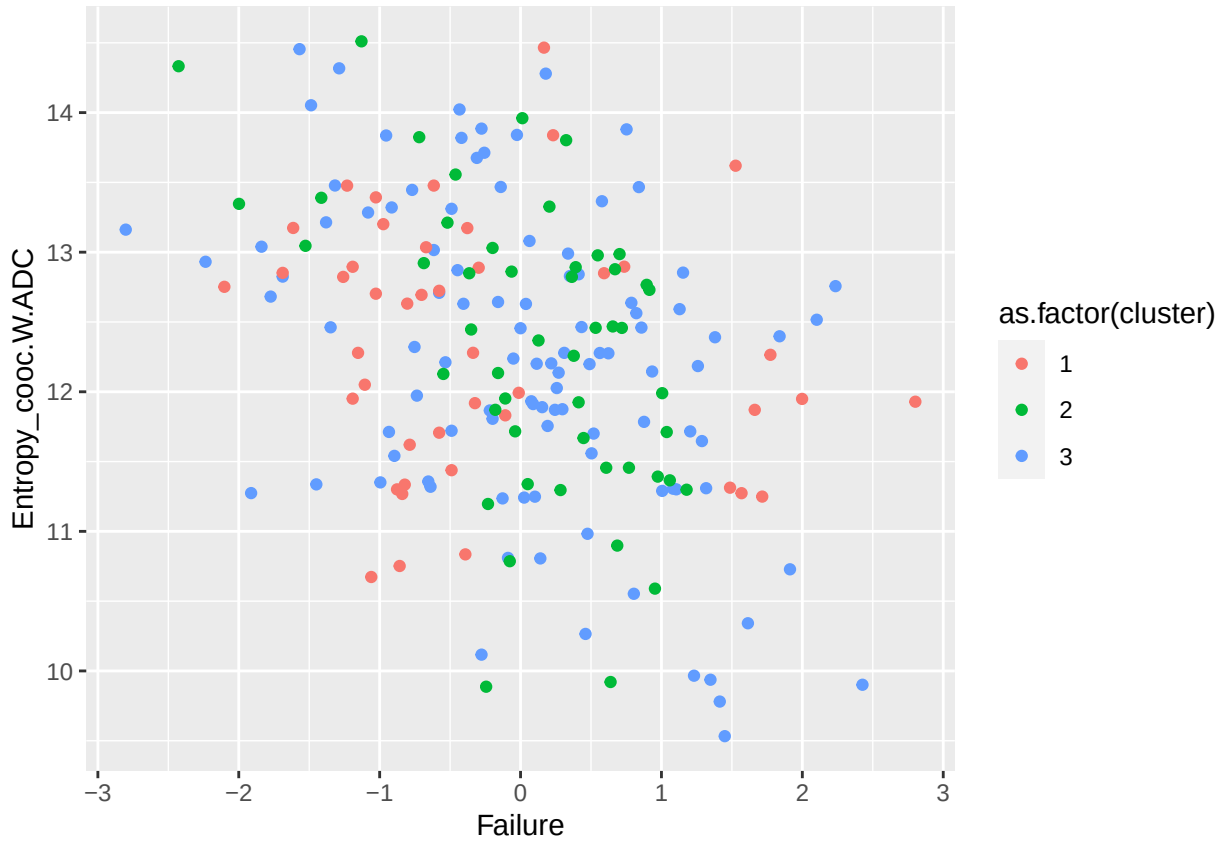

## Quality of k-means partition

```
clusters$betweenss / clusters$totss
```

```
## [1] 0.3324795
```

## Visualizing clusters using original variables

```
clusters <- kmeans(df, centers = 3, iter.max = 100, nstart = 100)
dt <- dt |> mutate(cluster = clusters$cluster)
dt |> ggplot(aes(x = Failure, y = Entropy_cooc.W.ADC, col = as.factor(cluster))) + geom_point()
```

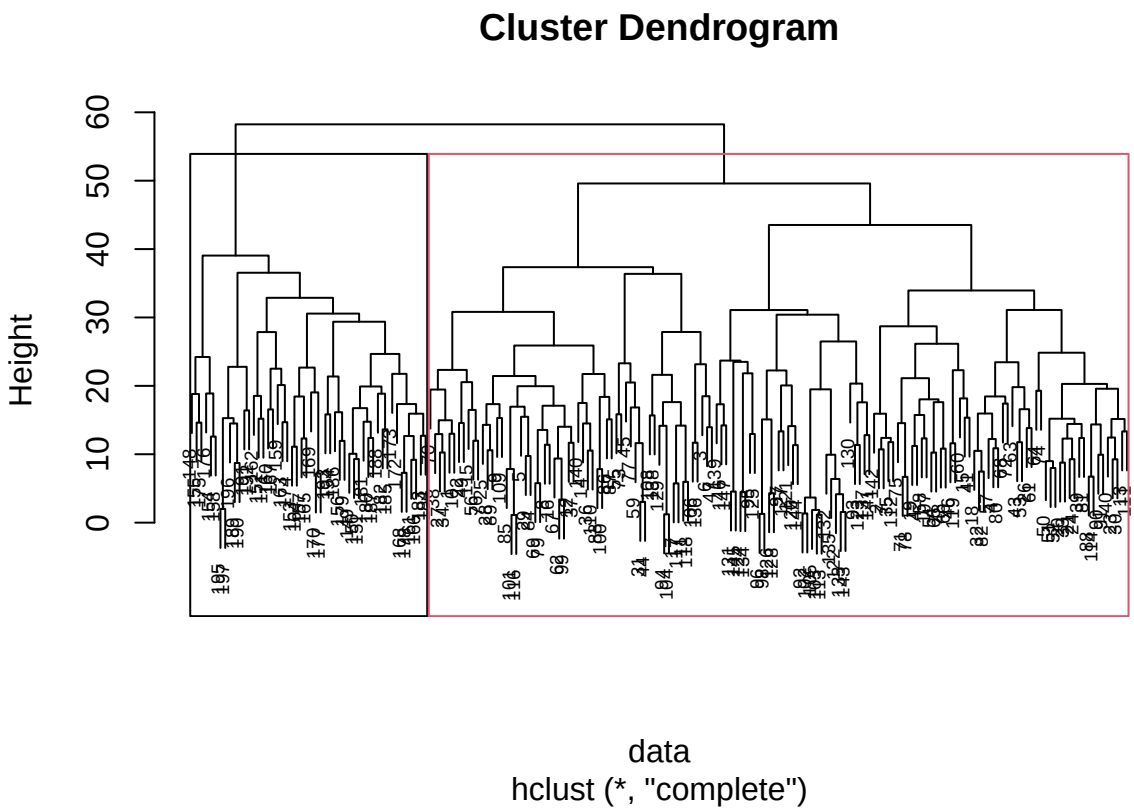



## 2 Hierarchical

```
dts <- dt%>%
  select_if(is.numeric) %>%  # select numeric columns
  select(-Failure.binary) %>%    # remove target column
  mutate_all(as.double) %>%  # coerce to double type
  scale()
data <- dist(dts, method = "euclidean")
```

## Completing Linkage

```
ct1 <- hclust(data, method = "complete")
plot(ct1, cex = 0.6)
rect.hclust(ct1, k = 2, border = 1:4)
```




## Computing maximum linkage clustering with agnes

```
set.seed(123)
ct2 <- agnes(dts, method = "complete")
ct2$ac
```

```
## [1] 0.8072963
```

## Computing divisive hierarchical clustering

```
ct3 <- diana(dts)
```

## Divise coefficient

```
ct3$dc
```

```
## [1] 0.7915983
```

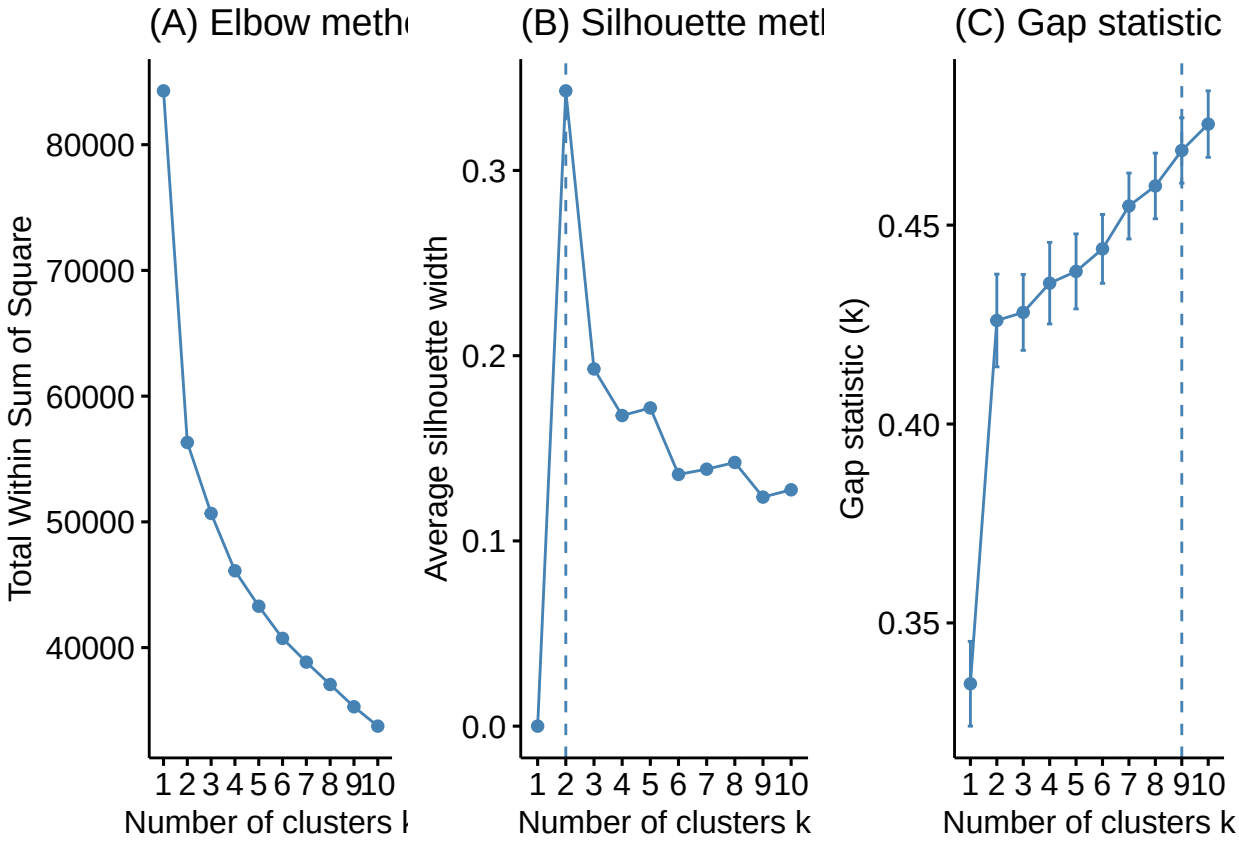
## Plotting cluster results

```
plot1 <- fviz_nbclust(dts, FUN = hcut, method = "wss",
                      k.max = 10) +
```

```
  ggtitle("(A) Elbow method")
plot2 <- fviz_nbclust(dts, FUN = hcut, method = "silhouette",
                      k.max = 10) +
  ggtitle("(B) Silhouette method")
plot3 <- fviz_nbclust(dts, FUN = hcut, method = "gap_stat",
                      k.max = 10) +
  ggtitle("(C) Gap statistic")
```

## Display the plot side by side

```
gridExtra::grid.arrange(plot1, plot2, plot3, nrow = 1)
```




## Ward's Method

```
ct4 <- hclust(data, method = "ward.D2" )

#  Cuting tree into 4 groups

sub_grp <- cutree(ct4, k = 8)

# Number of members in each cluster

table(sub_grp)
```

```
## sub_grp
##  1  2  3  4  5  6  7  8
## 71 33 12 21 10 19 22  9
```

# 3 Model Based

## Applying gmm model with 3 components

```
model1 <- Mclust(df[,1:10], G=3)
summary(model1)
```

```
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## ----------------------------------------------------
##
## Mclust VEE (ellipsoidal, equal shape and orientation) model with 3 components:
##
##  log-likelihood   n df       BIC      ICL
##       -1074.583 197 89 -2619.371 -2638.94
##
## Clustering table:
##   1   2   3
## 111  50  36
```

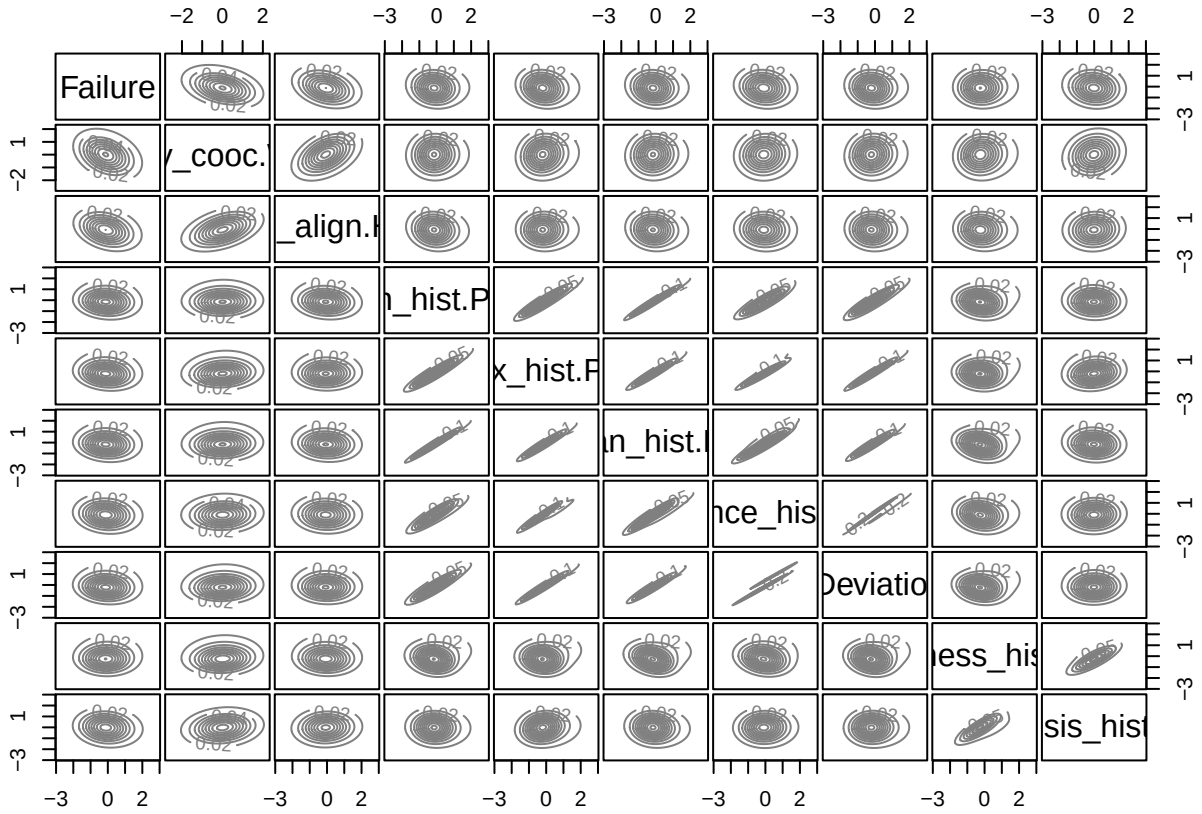
```
model2 = Mclust(df, 1:9)

summary(model2)
```

```
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## ----------------------------------------------------
##
## Mclust VEI (diagonal, equal shape) model with 9 components:
##
##  log-likelihood   n   df       BIC       ICL
##       -77829.42 197 4306 -178408.3 -178408.3
##
## Clustering table:
##  1  2  3  4  5  6  7  8  9
## 32 29 25 11 21 41 19 10  9
```

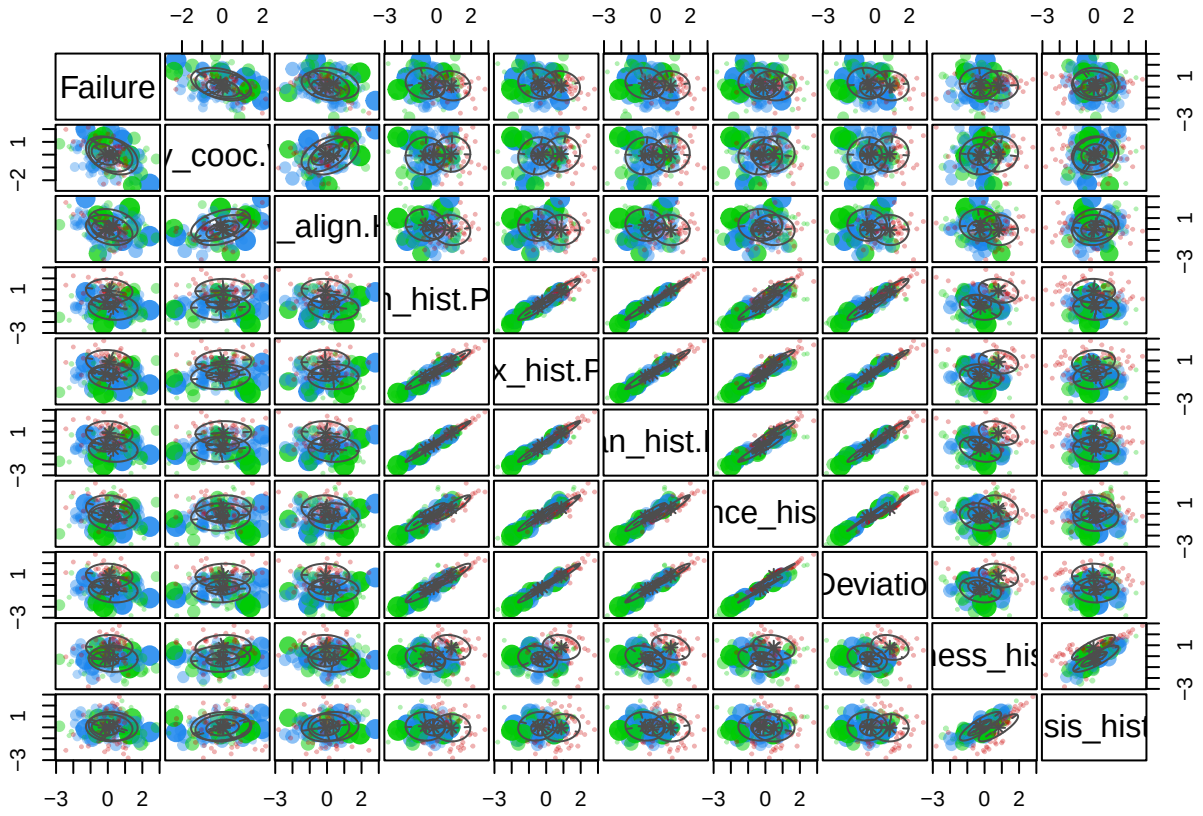
Thus, this shows 3 optimal number of clusters with BIC -2632.206. A negative zone with the highest value indicates the preferred model, In general, the lower the BIC value, the better. Plot the results with BIC, density and uncertainty.

## We Plot the results

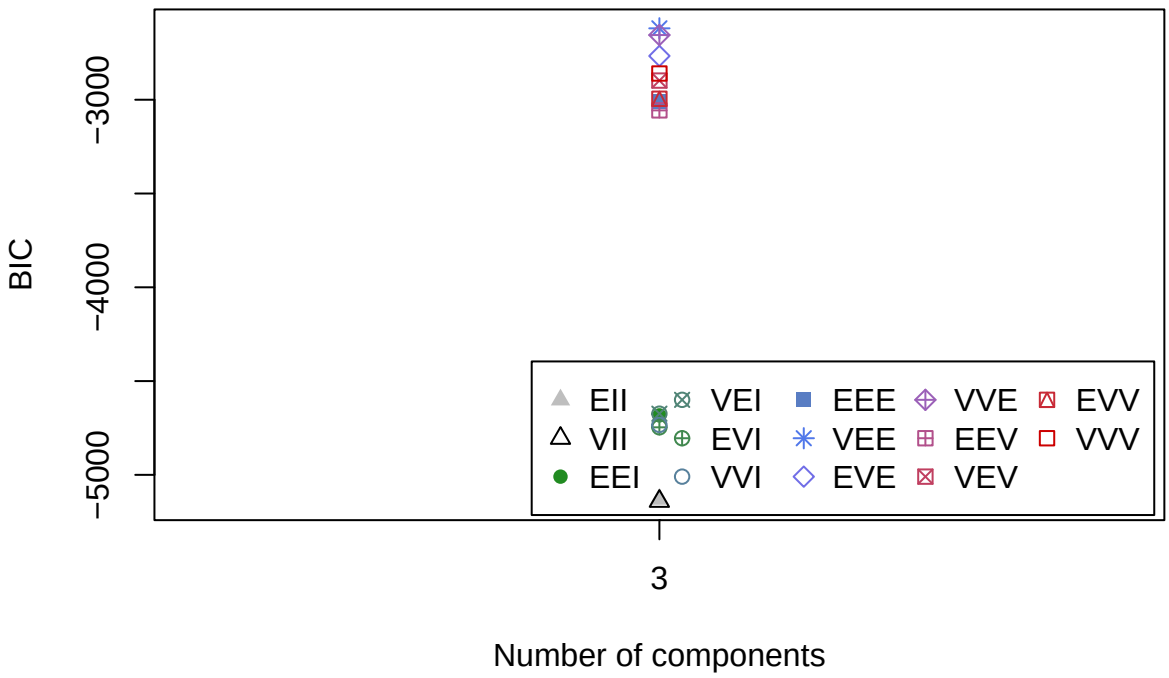
```
plot(model1, what = "density")
```



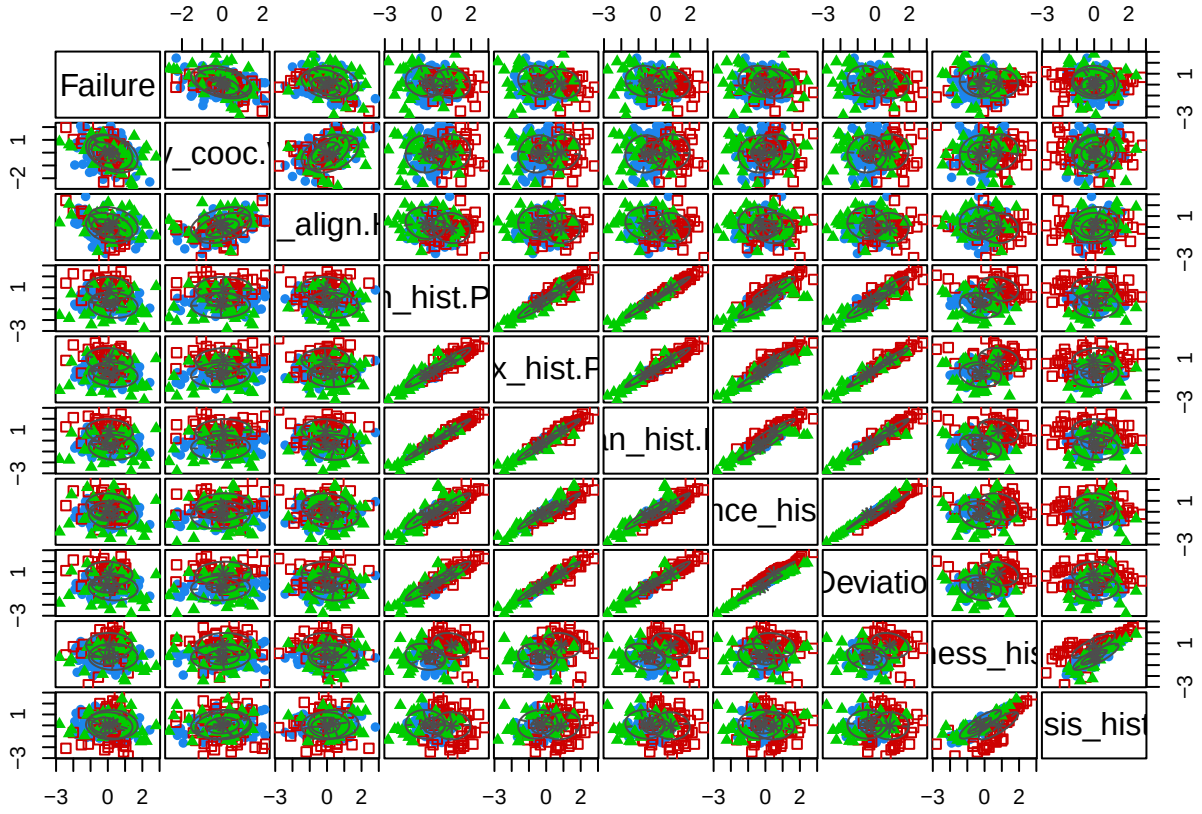

```
plot(model1, what = "uncertainty")
```



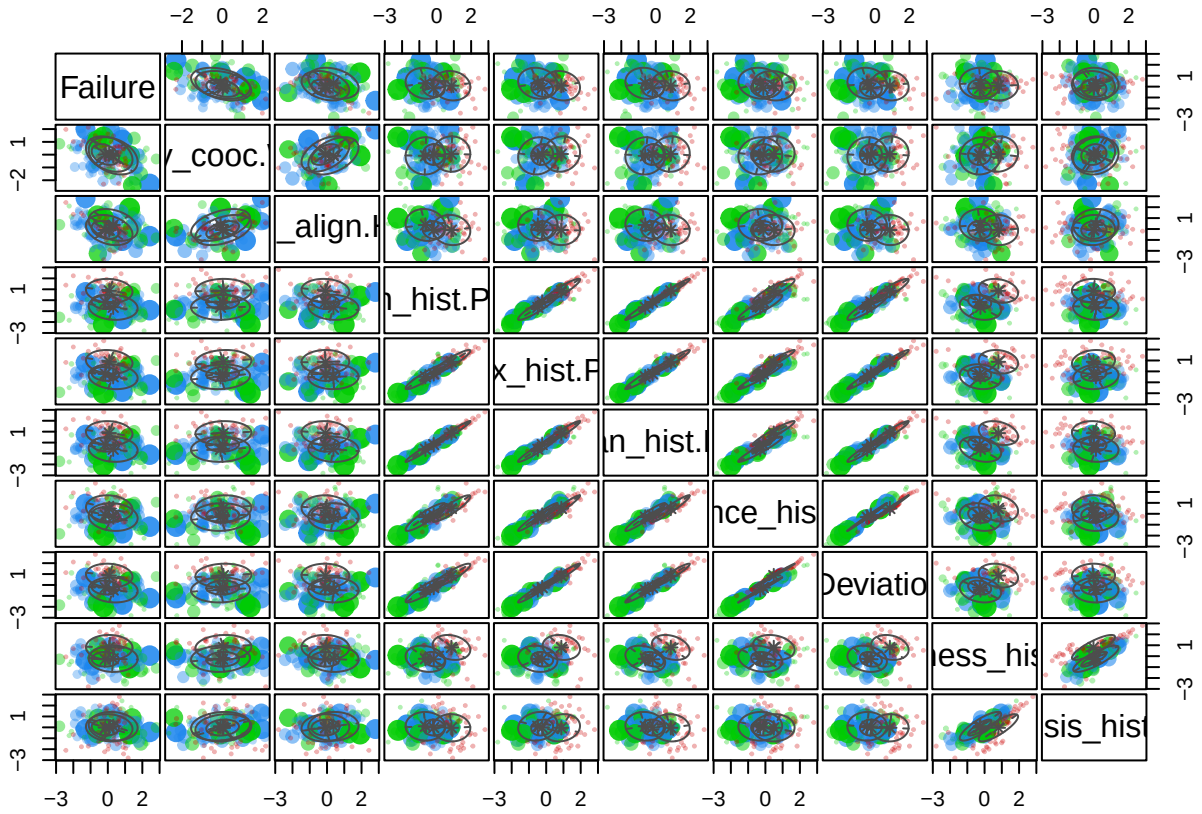

```
legend_args <- list(x = "bottomright", ncol = 5)
plot(model1, what = 'BIC', legendArgs = legend_args)
```




```
plot(model1, what = 'classification')
```



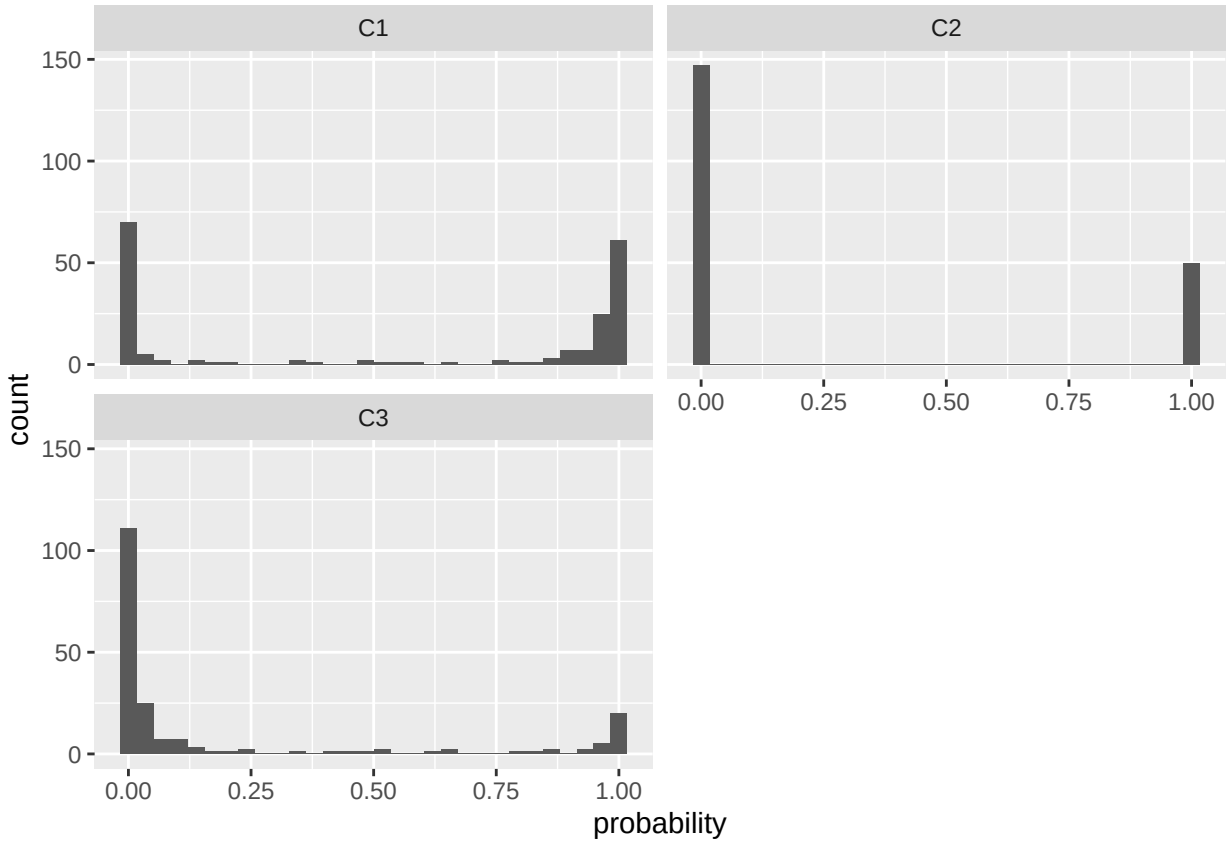

```
plot(model1, what = 'uncertainty')
```




Plotting the distribution for all observations. As clusters have more observations with middling levels of probability (i.e., 0.25–0.75), their clusters are usually less compact. Therefore, C3 is less compact than other clusters.

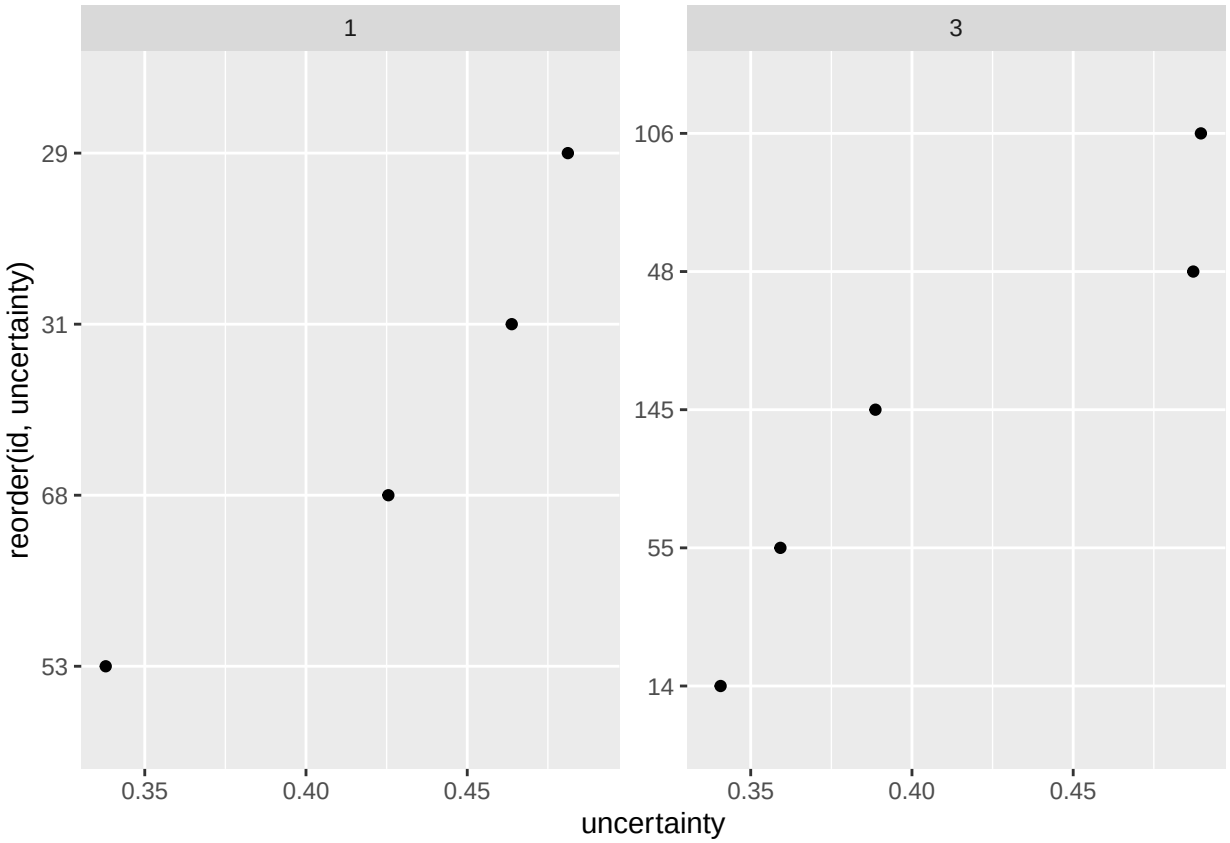
```
probabilities <- model1$z
colnames(probabilities) <- paste0('C', 1:3)

probabilities <- probabilities %>%
  as.data.frame() %>%
  mutate(id = row_number()) %>%
  tidyr::gather(cluster, probability, -id)

ggplot(probabilities, aes(probability)) +
  geom_histogram() +
  facet_wrap(~ cluster, nrow = 2)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



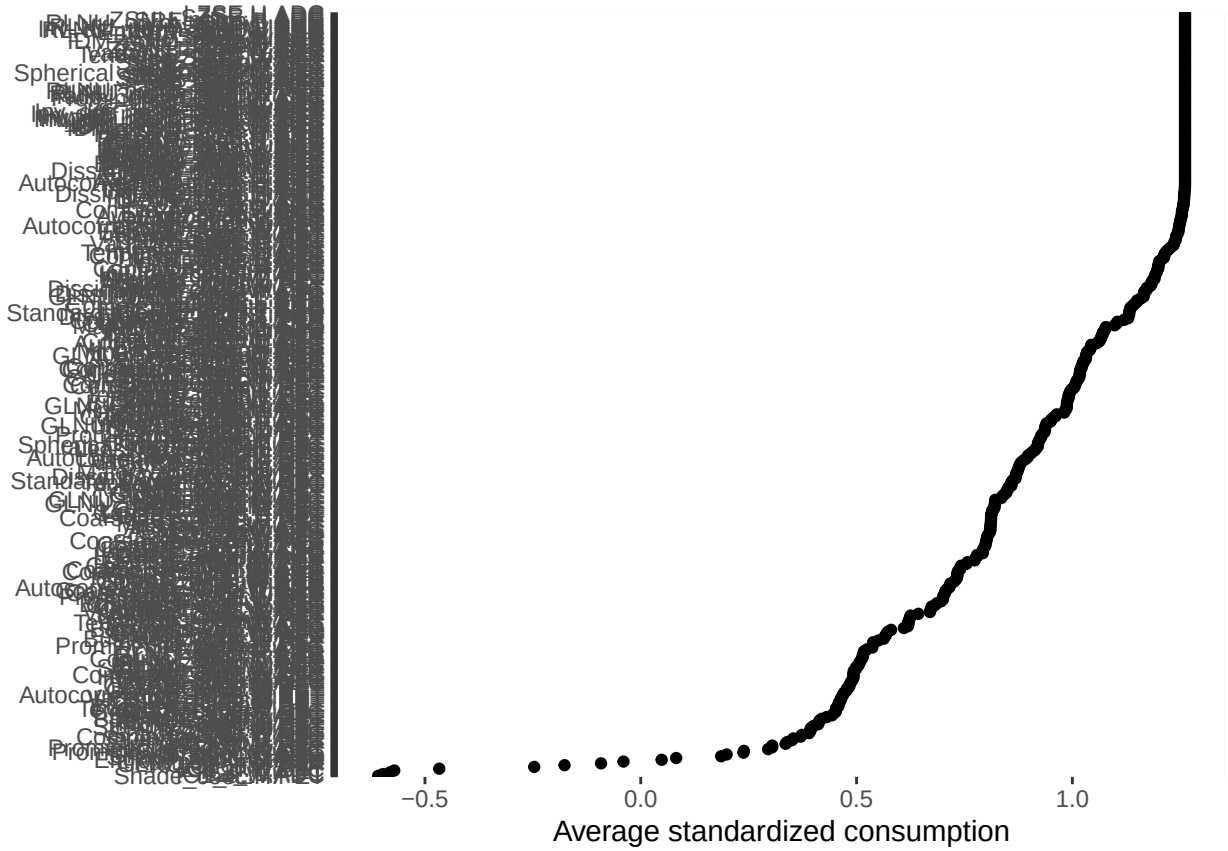

Ploting our the observations that are aligned to each cluster but their uncertainty of membership is greater than 0.25.

```
uncertainty <- data.frame(
  id = 1:nrow(df),
  cluster = model1$classification,
  uncertainty = model1$uncertainty
)
uncertainty %>%
  group_by(cluster) %>%
  filter(uncertainty > 0.25) %>%
  ggplot(aes(uncertainty, reorder(id, uncertainty))) +
  geom_point() +
  facet_wrap(~ cluster, scales = 'free_y', nrow = 1)
```




**Ploting the average standardized consumption for cluster 2 observations compared to all observations.**

```
clusterdt<- df %>%
  scale() %>%
  as.data.frame() %>%
  mutate(cluster = model1$classification) %>%
  filter(cluster == 2) %>%
  select(-cluster)

clusterdt%>%
  tidyr::gather(product, std_count) %>%
  group_by(product) %>%
  summarize(avg = mean(std_count)) %>%
  ggplot(aes(avg, reorder(product, avg))) +
  geom_point() +
  labs(x = "Average standardized consumption", y = NULL)
```




## Conclusion of the data set

Thus, using k-means clustering 2 is the best number of clusters with SSwithin = 33.2%. In Hierarchical, gap statistics suggest 9 clusters with 84.90% ac and 84.29%. Lastly, lastly, the model-based suggested 3 optimal number of clusters with BIC -2632.206.