

# DATA SCIENCE FINAL PROJECT

Jobert Jay R. Pandan

2022-12-17

## MODEL 3

**Comparison of three clustering techniques: K-Means Clustering, Heirarchecal Clustering and Model-based Clustering.**

*Libraries needed for the modeling*

```
library(readr)
library(dplyr)
library(ggplot2)
library(stringr)
library(cluster)
library(factoextra)
library(gridExtra)
library(tidyverse)
library(dplyr)
pacman::p_load(tidyverse)
pacman::p_load(bestNormalize)
library(caret)
library(mclust)
library(fpc)
```

## Data Acquisition

Data used in this modeling is radiomics\_completedata.csv.

```
df = read.csv("radiomics_completedata.csv")
```

Since we all know that the data is not normal, we do some transfromation, similar in the previous models.

```
df$Failure.binary = as.factor(df$Failure.binary)
newdf1 = df %>% select_if(is.numeric)
tempDF=apply(newdf1,2,orderNorm)
tempDF=lapply(tempDF, function(x) x$x.t)
tempDF=tempDF%>%as.data.frame()
normalized = cbind(df[c('Failure.binary')], tempDF)
```

## K-Means Clustering

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

```
## K-means clustering with 3 clusters of sizes 50, 103, 44
```

```
##
```

```
## Cluster means:
```

```
##   Failure.binary   Failure Entropy_cooc.W.ADC GLNU_align.H.PET Min_hist.PET
## 1      0.3000000  0.08209356      0.06141728    -0.091994140   0.85812682
```

## 2	0.3495146	0.06533738	-0.04771216	-0.009427188	-0.43614178
## 3	0.3636364	-0.24619905	0.04189900	0.126606986	0.04582414
##	Max_hist.PET	Mean_hist.PET	Variance_hist.PET	Standard_Deviation_hist.PET	
## 1	0.87619837	0.86965533	0.4852431		0.86141787
## 2	-0.46164331	-0.45835946	-0.3328592		-0.43567962
## 3	0.08498506	0.08473313	0.2277805		0.04100243
##	Skewness_hist.PET	Kurtosis_hist.PET	Energy_hist.PET	Entropy_hist.PET	
## 1	0.7993239	-0.03993469	0.8193998		1.2518393
## 2	-0.1562938	0.04875016	0.1318335		-0.5421416
## 3	-0.5424529	-0.06873937	-1.2397463		-0.1534404
##	AUC_hist.PET	H_suv.PET	Volume.PET	X3D_surface.PET	ratio_3ds_vol.PET
## 1	1.26010348	0.8782050	0.5234776	0.5377336	0.913001395
## 2	-0.09783915	-0.3528293	-0.4469966	-0.3135307	-0.002503013
## 3	-1.20290321	-0.1720189	0.4515175	0.1228859	-1.031642260
##	ratio_3ds_vol_norm.PET	irregularity.PET	tumor_length.PET	Compactness_v1.PET	
## 1	0.9194300	1.2601035	1.0133370		1.01671038
## 2	-0.0514460	-0.2024322	-0.3575160		-0.04050121
## 3	-0.9243764	-0.9580603	-0.3146062		-1.06054229
##	Compactness_v2.PET	Spherical_disproportion.PET	Sphericity.PET	Asphericity.PET	
## 1	0.681694797		0.9194300	0.834642568	0.90736711
## 2	-0.331242192		-0.0514460	-0.408748276	-0.04630539
## 3	0.000756093		-0.9243764	0.008385093	-0.92270228
##	Center_of_mass.PET	Max_3D_diam.PET	Major_axis_length.PET		
## 1	0.7289097	0.8104716	0.8723031		
## 2	-0.2828044	-0.5662795	-0.5482485		
## 3	-0.1662871	0.4046184	0.2921464		
##	Minor_axis_length.PET	Least_axis_length.PET	Elongation.PET	Flatness.PET	
## 1	1.03916907		0.8915085	1.2336004	1.1995823
## 2	-0.54506409		-0.5442359	-0.2537675	-0.3405795
## 3	0.09507154		0.2609288	-0.8077720	-0.5658959
##	Max_cooc.L.PET	Average_cooc.L.PET	Variance_cooc.L.PET	Entropy_cooc.L.PET	
## 1	0.8463483	1.1824572	0.9546964		1.2601035
## 2	0.1175010	-0.3623110	-0.1757674		-0.4970677
## 3	-1.2368186	-0.4955642	-0.6734268		-0.2683455
##	DAVE_cooc.L.PET	DVAR_cooc.L.PET	DENT_cooc.L.PET	SAVE_cooc.L.PET	
## 1	1.1285694	0.9883594	1.2601035		1.1821908
## 2	-0.2723810	-0.1908866	-0.3104527		-0.3629185
## 3	-0.6448461	-0.6762875	-0.7051941		-0.4938394
##	SVAR_cooc.L.PET	SENT_cooc.L.PET	ASM_cooc.L.PET	Contrast_cooc.L.PET	
## 1	0.9910248	1.2601035	0.8094807		0.7982773
## 2	-0.1976191	-0.2139714	0.1318891		-0.1506873
## 3	-0.6635562	-0.9310482	-1.2286049		-0.5543879
##	Dissimilarity_cooc.L.PET	Inv_diff_cooc.L.PET	Inv_diff_norm_cooc.L.PET		
## 1	1.1285694	1.2476933		1.2601035	
## 2	-0.2723810	-0.2737576		-0.2641384	
## 3	-0.6448461	-0.7769917		-0.8136118	
##	IDM_cooc.L.PET	IDM_norm_cooc.L.PET	Inv_var_cooc.L.PET	Correlation_cooc.L.PET	
## 1	1.1842495	1.2601035	1.1893919		1.0050778
## 2	-0.1967642	-0.1986930	-0.1926182		-0.3347999
## 3	-0.8851310	-0.9668136	-0.9006801		-0.3583977
##	Autocorrelation_cooc.L.PET	Tendency_cooc.L.PET	Shade_cooc.L.PET		
## 1	0.8960104	0.9910248	0.46295135		
## 2	-0.2521613	-0.1976191	-0.06384699		
## 3	-0.4279069	-0.6635562	-0.37662109		

```

## Prominence_cooc.L.PET IC1_.L.PET IC2_.L.PET Coarseness_vdif_.L.PET
## 1 0.70058703 -0.57073804 1.2562010 0.8110514
## 2 -0.06485971 0.28929559 -0.1824053 0.1388886
## 3 -0.64429095 -0.02864872 -1.0005069 -1.2467745
## Contrast_vdif_.L.PET Busyness_vdif_.L.PET Complexity_vdif_.L.PET
## 1 0.67242321 0.5626713 1.1186189
## 2 -0.02509042 -0.3707759 -0.1974261
## 3 -0.70538289 0.2285534 -0.8090014
## Strength_vdif_.L.PET SRE_align.L.PET LRE_align.L.PET GLNU_align.L.PET
## 1 0.4987130 1.26010348 1.2601035 0.4467533
## 2 0.1337268 -0.06234094 -0.2798127 -0.3993524
## 3 -0.8797615 -1.28600130 -0.7769197 0.4271735
## RLNU_align.L.PET RP_align.L.PET LGRE_align.L.PET HGRE_align.L.PET
## 1 0.4109397 1.26010348 0.9931653282 0.9236862
## 2 -0.4020416 -0.06596095 0.0003066528 -0.2605452
## 3 0.4741659 -1.27752719 -1.1293148102 -0.4397308
## LGSRE_align.L.PET HGSRE_align.L.PET LGHRE_align.L.PET HGLRE_align.L.PET
## 1 0.99751061 0.9215848 0.9631837618 0.9364447
## 2 0.00150105 -0.2532377 -0.0002880774 -0.2914542
## 3 -1.13704860 -0.4544490 -1.0938526391 -0.3818739
## GLNU_norm_align.L.PET RLNU_norm_align.L.PET GLVAR_align.L.PET
## 1 1.03096731 1.26010348 0.9868800
## 2 0.05022954 -0.08924613 -0.2126497
## 3 -1.28913655 -1.22301869 -0.6236610
## RLVAR_align.L.PET Entropy_align.L.PET SZSE.L.PET LZSE.L.PET LGLZE.L.PET
## 1 1.01691811 1.2601035 1.2601035 1.1491524 1.008605590
## 2 -0.02323398 -0.4815782 -0.1120901 -0.4485566 -0.002220365
## 3 -1.10120014 -0.3046051 -1.1695429 -0.2558248 -1.140945043
## HGLZE.L.PET SZLGE.L.PET SZHGE.L.PET LZLGE.L.PET LZHGE.L.PET GLNU_area.L.PET
## 1 0.9357872 1.025279294 0.9346034 0.86889376 0.8735299 0.4579289
## 2 -0.2601593 0.001671419 -0.2327100 -0.02472047 -0.3085215 -0.4001837
## 3 -0.4543853 -1.169002746 -0.5172963 -0.92951089 -0.2704267 0.4164200
## ZSNU.L.PET ZSP.L.PET GLNU_norm.L.PET ZSNU_norm.L.PET GLVAR_area.L.PET
## 1 0.4303770 1.2601035 1.03034483 1.2601035 1.0033531
## 2 -0.4102506 -0.1545715 0.04986928 -0.1800621 -0.2146953
## 3 0.4712947 -1.0700979 -1.28758586 -1.0104268 -0.6375918
## ZSVAR.L.PET Entropy_area.L.PET Max_cooc.H.PET Average_cooc.H.PET
## 1 0.8970301 1.2601035 0.5503672 1.2601035
## 2 -0.2132431 -0.5135166 0.1787787 -0.2414238
## 3 -0.5201697 -0.2298400 -1.0439220 -0.8667845
## Variance_cooc.H.PET Entropy_cooc.H.PET DAVE_cooc.H.PET DVAR_cooc.H.PET
## 1 1.2346999 1.1893755 1.2585499 1.2504238
## 2 -0.5534078 -0.5127561 -0.4670489 -0.4579410
## 3 -0.1075906 -0.1512476 -0.3368512 -0.3489379
## DENT_cooc.H.PET SAVE_cooc.H.PET SVAR_cooc.H.PET SENT_cooc.H.PET
## 1 1.2062372 1.2601035 1.2395185 0.9803265
## 2 -0.5101347 -0.2968597 -0.5086056 -0.1373389
## 3 -0.1765452 -0.7370141 -0.2179443 -0.7925095
## ASM_cooc.H.PET Contrast_cooc.H.PET Dissimilarity_cooc.H.PET
## 1 0.5645624 1.1982041 1.2585499
## 2 0.2172291 -0.4438519 -0.4670489
## 3 -1.1500618 -0.3225785 -0.3368512
## Inv_diff_cooc.H.PET Inv_diff_norm_cooc.H.PET IDM_cooc.H.PET
## 1 1.0670858 1.2601035 0.92681301

```

```

## 2      -0.1145474      -0.1151856      -0.05599947
## 3      -0.9444524      -1.1622969      -0.92210693
##   IDM_norm_cooc.H.PET Inv_var_cooc_.H.PET Correlation_cooc.H.PET
## 1      1.26010348      0.93988938      1.0155183
## 2      -0.08588108      0.03014244      -0.3674791
## 3      -1.23089598      -1.13861682      -0.2937629
##   Autocorrelation_cooc.H.PET Tendency_cooc.H.PET Shade_cooc.H.PET
## 1      1.2468553      1.21525719      -0.60793428
## 2      -0.2311291      -0.55484894      0.28955287
## 3      -0.8758287      -0.08212314      0.01301747
##   Prominence_cooc.H.PET IC1_d.H.PET IC2_d.H.PET Coarseness_vdif.H.PET
## 1      0.93025143 -0.1763575 1.1955390 0.8000853
## 2      -0.47574640 0.3320062 -0.4234805 0.1365392
## 3      0.05657518 -0.5767901 -0.3672377 -1.2288138
##   Contrast_vdif.H.PET Busyness_vdif.H.PET Complexity_vdif.H.PET
## 1      0.62214341 0.4186155 0.96031165
## 2      -0.05874494 -0.4356625 -0.07505258
## 3      -0.56946460 0.5441468 -0.91557198
##   Strength_vdif.H.PET SRE_align.H.PET LRE_align.H.PET RLNU_align.H.PET
## 1      0.3354073 1.2601035 1.0641749 0.4119913
## 2      0.2374769 -0.3701632 -0.2153213 -0.4136037
## 3      -0.9370565 -0.5654173 -0.7052421 0.5000366
##   RP_align.H.PET LGRE_align.H.PET HGRE_align.H.PET LGSRE_align.H.PET
## 1      1.2601035 0.8113038 1.2442073 0.8113038
## 2      -0.3896171 0.1184110 -0.2735091 0.1187439
## 3      -0.5198775 -1.1991256 -0.7736121 -1.1999048
##   HGSRE_align.H.PET LGHRE_align.H.PET HGLRE_align.H.PET GLNU_norm_align.H.PET
## 1      1.2601035 0.8135873 0.8807012 0.82073753
## 2      -0.3249637 0.1172528 -0.1800196 0.03756653
## 3      -0.6712252 -1.1990100 -0.5793873 -1.02059612
##   RLNU_norm_align.H.PET GLVAR_align.H.PET RLVAR_align.H.PET Entropy_align.H.PET
## 1      1.2601035 1.19494406 0.62368460 1.26010348
## 2      -0.4317955 -0.54757978 -0.01909491 -0.57137386
## 3      -0.4211417 -0.07605649 -0.66403304 -0.09440151
##   SZSE.H.PET LZSE.H.PET LGLZE.H.PET HGLZE.H.PET SZLGE.H.PET SZHGE.H.PET
## 1 1.2232981 0.39252039 0.8109439 1.2392449 0.8108628 1.1985268
## 2 -0.4030065 -0.07461144 0.1185232 -0.3368316 0.1195423 -0.3805777
## 3 -0.4467099 -0.27138729 -1.1989792 -0.6197406 -1.2012727 -0.4710644
##   LZLGE.H.PET LZHGE.H.PET GLNU_area.H.PET ZSNU.H.PET ZSP.H.PET
## 1 0.4585734 0.349850688 0.4825644 0.3391641 1.0420572
## 2 0.1834368 0.002522648 -0.4063182 -0.3657920 -0.3878684
## 3 -0.9505149 -0.403462889 0.4027853 0.4708721 -0.2761913
##   GLNU_norm.H.PET ZSNU_norm.H.PET GLVAR_area.H.PET ZSVAR_H.PET
## 1 0.84722728 1.0696557 1.1796629 0.30451604
## 2 0.01917366 -0.3840552 -0.5267189 0.02642073
## 3 -1.00764207 -0.3164794 -0.1075251 -0.40788949
##   Entropy_area.H.PET Max_cooc.W.PET Average_cooc.W.PET Variance_cooc.W.PET
## 1 1.26010348 0.6427321 0.8516085 0.4647485
## 2 -0.58988871 0.2202280 -0.4613212 -0.3253769
## 3 -0.05105993 -1.2459112 0.1121741 0.2335546
##   Entropy_cooc.W.PET DAVE_cooc.W.PET DVAR_cooc.W.PET DENT_cooc.W.PET
## 1 1.2449097 0.867756082 0.4904128 1.2360752
## 2 -0.5584923 -0.424119492 -0.3222303 -0.5229929
## 3 -0.1072905 0.006738718 0.1970246 -0.1803521

```

```

## SAVE_cooc.W.PET SVAR_cooc.W.PET SENT_cooc.W.PET ASM_cooc.W.PET
## 1 0.8504316 0.4522152 1.2579126 0.7028941
## 2 -0.4622211 -0.3284163 -0.4561071 0.1984999
## 3 0.1156181 0.2549118 -0.3617410 -1.2634130
## Contrast_cooc.W.PET Dissimilarity_cooc.W.PET Inv_diff_cooc.W.PET
## 1 0.5126415 0.867756082 1.1660786
## 2 -0.3223858 -0.424119492 -0.1655483
## 3 0.1721288 0.006738718 -0.9375557
## Inv_diff_norm_cooc.W.PET IDM_cooc.W.PET IDM_norm_cooc.W.PET
## 1 1.2601035 0.98192759 1.2601035
## 2 -0.2461061 -0.08797305 -0.1918482
## 3 -0.8558238 -0.90988988 -0.9828363
## Inv_var_cooc.W.PET Correlation_cooc.W.PET Autocorrelation_cooc.W.PET
## 1 1.0699412 1.0064888 0.4690750
## 2 -0.1190777 -0.3422077 -0.3619782
## 3 -0.9370922 -0.3426602 0.3143182
## Tendency_cooc.W.PET Shade_cooc.W.PET Prominence_cooc.W.PET IC1_d.W.PET
## 1 0.4522152 0.1987828 0.2388296 -0.2466266
## 2 -0.3284163 -0.1930359 -0.2585297 0.3870571
## 3 0.2549118 0.2259899 0.3337971 -0.6258080
## IC2_d.W.PET Coarseness_vdif.W.PET Contrast_vdif.W.PET Busyness_vdif.W.PET
## 1 1.2462317 0.7396460 0.7732209 0.4575578
## 2 -0.3970734 0.1742902 -0.2217484 -0.1646606
## 3 -0.4866597 -1.2485044 -0.3595673 -0.1344966
## Complexity_vdif.W.PET Strength_vdif.W.PET SRE_align.W.PET LRE_align.W.PET
## 1 0.3702140 0.56761023 1.2601035 1.2126937
## 2 -0.3064288 -0.09767275 -0.2821763 -0.2476012
## 3 0.2966242 -0.41636860 -0.7713867 -0.7984492
## GLNU_align.W.PET RLNU_align.W.PET RP_align.W.PET LGRE_align.W.PET
## 1 0.4934063 0.4137762 1.2601035 0.77452284
## 2 -0.3661553 -0.4103442 -0.3195602 0.04873235
## 3 0.2964473 0.4903783 -0.6838744 -0.99421758
## HGRE_align.W.PET LGSRE_align.W.PET HGSRE_align.W.PET LGHRE_align.W.PET
## 1 0.4723914 0.81795589 0.4620087 0.62455287
## 2 -0.3648013 0.03592996 -0.3575084 0.07771776
## 3 0.3171582 -1.01360410 0.3118848 -0.89164937
## HGLRE_align.W.PET GLNU_norm_align.W.PET RLNU_norm_align.W.PET
## 1 0.4926271 0.82019209 1.2601035
## 2 -0.3794703 0.09070163 -0.3832069
## 3 0.3285020 -1.14436073 -0.5348833
## GLVAR_align.W.PET RLVAR_align.W.PET Entropy_align.W.PET SZSE.W.PET
## 1 0.4832532 0.69530904 1.26010348 1.2601035
## 2 -0.3400889 0.03912658 -0.57369063 -0.3785226
## 3 0.2469658 -0.88171590 -0.08897816 -0.5458488
## LZSE.W.PET LGLZE.W.PET HGLZE.W.PET SZLGE.W.PET SZHGE.W.PET LZLGE.W.PET
## 1 0.61848970 0.80618711 0.4775674 0.92689401 0.4673123 0.3729370
## 2 -0.05493204 0.03609125 -0.3612973 0.01377567 -0.3411788 0.1221971
## 3 -0.57423830 -1.00060806 0.3030741 -1.08553622 0.2676319 -0.7098443
## LZHGE.W.PET GLNU_area.W.PET ZSNU.W.PET ZSP.W.PET GLNU_norm.W.PET
## 1 0.5380657 0.4985108 0.3924770 1.2531151 0.84183325
## 2 -0.3855402 -0.3914134 -0.3951841 -0.4241199 0.07906742
## 3 0.2910763 0.3497736 0.4790935 -0.4311684 -1.14171833
## ZSNU_norm.W.PET GLVAR_area.W.PET ZSVAR.W.PET Entropy_area.W.PET Min_hist.ADC
## 1 1.2455332 0.4768281 0.39447957 1.26010348 0.5356019

```

```

## 2      -0.4342701      -0.3293288  0.05364306      -0.58783229  -0.1360750
## 3      -0.3987918      0.2290786 -0.57384578      -0.05587383  -0.2839460
## Max_hist.ADC Mean_hist.ADC Variance_hist.ADC Standard_Deviation_hist.ADC
## 1      1.2549000      1.2496503      0.69944832      1.1308474
## 2      -0.4109585      -0.4404749      -0.09465534      -0.2637073
## 3      -0.4635840      -0.3889453      -0.57324808      -0.6677390
## Skewness_hist.ADC Kurtosis_hist.ADC Energy_hist.ADC Entropy_hist.ADC
## 1      0.44277242      0.29487646      0.8087608      1.2601035
## 2      -0.18754348      -0.16659367      0.1306280      -0.4216926
## 3      -0.06412824      0.05489374      -1.2248338      -0.4447917
## AUC_hist.ADC Volume.ADC X3D_surface.ADC ratio_3ds_vol.ADC
## 1      1.2601035  0.5074105      0.62015938      1.0660761
## 2      -0.2619403 -0.4230570      -0.27514913      -0.1973241
## 3      -0.8187563  0.4137352      -0.06062747      -0.7495328
## ratio_3ds_vol_norm.ADC irregularity.ADC Compactness_v1.ADC Compactness_v2.ADC
## 1      1.2601035      1.2601035      1.06168964      1.1019793
## 2      -0.3696801      -0.3302445      0.03139817      -0.2320236
## 3      -0.5665482      -0.6588633      -1.27996576      -0.7091030
## Spherical_disproportion.ADC Sphericity.ADC Asphericity.ADC Center_of_mass.ADC
## 1      1.2601035      1.2601035      1.1338612      0.45153340
## 2      -0.3696801      -0.2587123      -0.3555244      -0.08723432
## 3      -0.5665482      -0.8263138      -0.4562283      -0.30889853
## Max_3D_diam.ADC Major_axis_length.ADC Minor_axis_length.ADC
## 1      0.9501040      1.0970495      0.9851300
## 2      -0.3641393      -0.4002567      -0.3810801
## 3      -0.2272466      -0.3096826      -0.2273920
## Least_axis_length.ADC Elongation.ADC Flatness.ADC Max_cooc.L.ADC
## 1      0.9153330      1.2566904      1.2282154      0.90280643
## 2      -0.3766661      -0.4154356      -0.4102787      0.08247576
## 3      -0.1584101      -0.4555603      -0.4352743      -1.21898466
## Average_cooc.L.ADC Variance_cooc.L.ADC Entropy_cooc.L.ADC DAVE_cooc.L.ADC
## 1      1.2528828      0.8853611      1.2601035      1.1650991
## 2      -0.4657713      -0.1418733      -0.3448756      -0.2975032
## 3      -0.3334022      -0.6739796      -0.6246132      -0.6275483
## DVAR_cooc.L.ADC DENT_cooc.L.ADC SAVE_cooc.L.ADC SVAR_cooc.L.ADC
## 1      0.8752124      1.2601035      1.2528828      0.8583253
## 2      -0.1346799      -0.3056692      -0.4658976      -0.1328135
## 3      -0.6792861      -0.7163919      -0.3331065      -0.6644654
## SENT_cooc.L.ADC ASM_cooc.L.ADC Contrast_cooc.L.ADC Dissimilarity_cooc.L.ADC
## 1      1.0157352      0.8370442      0.8134852      1.1650991
## 2      -0.3227867      0.1096301      -0.1425577      -0.2975032
## 3      -0.3986302      -1.2078330      -0.5907004      -0.6275483
## Inv_diff_cooc.L.ADC Inv_diff_norm_cooc.L.ADC IDM_cooc.L.ADC
## 1      1.2578935      1.260102      1.2115347
## 2      -0.2643399      -0.170323      -0.2353613
## 3      -0.8106287      -1.033224      -0.8257846
## IDM_norm_cooc.L.ADC Inv_var_cooc.L.ADC Correlation_cooc.L.ADC
## 1      1.2601035      1.2193186      1.0207723
## 2      -0.1048688      -0.2448072      -0.2079255
## 3      -1.1864475      -0.8125180      -0.6732339
## Autocorrelation_.L.ADC Tendency_cooc.L.ADC Shade_.L.ADC Prominence_cooc.L.ADC
## 1      1.0565602      0.8583253  0.23784080      0.536314764
## 2      -0.3806305      -0.1328135  -0.08926511      0.009639142
## 3      -0.3096152      -0.6644654  -0.06131213      -0.632012951

```

```

## IC1_.L.ADC IC2_.L.ADC Coarseness_vdif_.L.ADC Contrast_vdif_.L.ADC
## 1 -0.5804176 1.2488160 0.7057617 0.67346133
## 2 0.4221620 -0.2429369 0.1831016 -0.00426411
## 3 -0.3286775 -0.8504159 -1.2306293 -0.75531507
## Busyness_vdif_.L.ADC Complexity_vdif_.L.ADC Strength_vdif_.L.ADC
## 1 0.6863390 1.1388120 0.39993735
## 2 -0.1266996 -0.2878726 -0.04138263
## 3 -0.4833384 -0.6202209 -0.35760129
## SRE_align.L.ADC LRE_align.L.ADC GLNU_align.L.ADC RLNU_align.L.ADC
## 1 1.26010162 1.2601035 0.487428372 0.50549851
## 2 -0.08697493 -0.2825775 -0.240760507 -0.22537592
## 3 -1.22833517 -0.7704476 0.009702583 -0.04684558
## RP_align.L.ADC LGRE_align.L.ADC HGRE_align.L.ADC LGSRE_align.L.ADC
## 1 1.2601035 0.7915610 1.1312349 0.7928837
## 2 -0.1056257 0.1448884 -0.3944579 0.1441628
## 3 -1.1846756 -1.2386718 -0.3621041 -1.2384754
## HGSRE_align.L.ADC LGHRE_align.L.ADC HGLRE_align.L.ADC GLNU_norm_align.L.ADC
## 1 1.1331911 0.7785402 1.1307935 1.16390802
## 2 -0.3912582 0.1518783 -0.4119247 -0.04113242
## 3 -0.3718172 -1.2402380 -0.3207144 -1.22633549
## RLNU_norm_align.L.ADC GLVAR_align.L.ADC RLVAR_align.L.ADC Entropy_align.L.ADC
## 1 1.2601035 0.9358820 1.0768529 1.2601035
## 2 -0.1603942 -0.1602669 -0.0982178 -0.2949903
## 3 -1.0564678 -0.6883319 -0.9937756 -0.7413903
## SZSE.L.ADC LZSE.L.ADC LGLZE.L.ADC HGLZE.L.ADC SZLGE.L.ADC SZHGE.L.ADC
## 1 1.2601035 1.2004418 0.8011667 1.1481754 0.8030950 1.1433289
## 2 -0.1625277 -0.3953429 0.1399974 -0.4018017 0.1376303 -0.3949377
## 3 -1.0514733 -0.4386766 -1.2381380 -0.3641634 -1.2347891 -0.3747242
## LZLGE.L.ADC LZHGE.L.ADC GLNU_area.L.ADC ZSNU.L.ADC ZSP.L.ADC
## 1 0.6995465 1.0756933 0.492957756 0.50995654 1.2601035
## 2 0.1872096 -0.4080677 -0.238004422 -0.22780845 -0.2352542
## 3 -1.2331796 -0.2671294 -0.003032552 -0.04621719 -0.8812271
## GLNU_norm.L.ADC ZSNU_norm.L.ADC GLVAR_area.L.ADC ZSVAR.L.ADC
## 1 1.15753016 1.2601012 0.9490632 0.7539999
## 2 -0.03057441 -0.2520807 -0.1626997 -0.1999883
## 3 -1.24380273 -0.8418370 -0.6976156 -0.3886638
## Entropy_area.L.ADC Max_cooc.H.ADC Average_cooc.H.ADC Variance_cooc.H.ADC
## 1 1.2601035 0.8206280 1.2601035 1.2601035
## 2 -0.3140165 0.1226214 -0.4131354 -0.4428532
## 3 -0.6968516 -1.2195763 -0.4648233 -0.3952568
## Entropy_cooc.H.ADC DAVE_cooc.H.ADC DVAR_cooc.H.ADC DENT_cooc.H.ADC
## 1 1.2601035 1.2601035 1.2575931 1.2600955
## 2 -0.5067629 -0.4362605 -0.4222198 -0.4035056
## 3 -0.2456499 -0.4106896 -0.4407048 -0.4873658
## SAVE_cooc.H.ADC SVAR_cooc.H.ADC SENT_cooc.H.ADC ASM_cooc.H.ADC
## 1 1.2601035 1.2601035 1.2601035 0.8094090
## 2 -0.4162940 -0.3757314 -0.4055264 0.1276736
## 3 -0.4574294 -0.5523827 -0.4826353 -1.2187170
## Contrast_cooc.H.ADC Dissimilarity_cooc.H.ADC Inv_diff_cooc.H.ADC
## 1 1.2119605 1.2601035 1.2597865
## 2 -0.4341280 -0.4362605 -0.1075456
## 3 -0.3609737 -0.4106896 -1.1798197
## Inv_diff_norm_cooc.H.ADC IDM_cooc.H.ADC IDM_norm_cooc.H.ADC
## 1 1.2601035 1.23864077 1.26010348

```

## 2	-0.1089083	-0.08629617	-0.09378679	
## 3	-1.1769913	-1.20553484	-1.21238942	
##	Inv_var_cooc.H.ADC	Correlation_cooc.H.ADC	Autocorrelation_cooc.H.ADC	
## 1	1.2416511	1.0173899	1.2601035	
## 2	-0.0746561	-0.2053340	-0.3728050	
## 3	-1.2362039	-0.6754567	-0.5592331	
##	Tendency_cooc.H.ADC	Shade_cooc.H.ADC	Prominence_cooc.H.ADC	IC1_d.H.ADC
## 1	1.2601035	0.38945107	1.2601035	-0.4665213
## 2	-0.3757314	-0.17187278	-0.3863805	0.3516626
## 3	-0.5523827	-0.04021948	-0.5274542	-0.2930730
##	IC2_d.H.ADC	Coarseness_vdif.H.ADC	Contrast_vdif.H.ADC	Busyness_vdif.H.ADC
## 1	1.2573571	0.7047403	1.2597865	0.6093025
## 2	-0.2794834	0.1822561	-0.4186159	-0.1961535
## 3	-0.7745696	-1.2274857	-0.4516339	-0.2332118
##	Complexity_vdif.H.ADC	Strength_vdif.H.ADC	SRE_align.H.ADC	LRE_align.H.ADC
## 1	1.2542493	0.35328758	1.26010348	1.2601035
## 2	-0.4475287	-0.03715531	-0.06273425	-0.1427629
## 3	-0.3776594	-0.31448595	-1.28508015	-1.0977408
##	GLNU_align.H.ADC	RLNU_align.H.ADC	RP_align.H.ADC	LGRE_align.H.ADC
## 1	0.51419743	0.51653976	1.26010348	1.04142924
## 2	-0.22752508	-0.22997103	-0.06802005	0.03984021
## 3	-0.05169972	-0.04863572	-1.27270722	-1.27672651
##	HGRE_align.H.ADC	LGSRE_align.H.ADC	HGSRE_align.H.ADC	LGHRE_align.H.ADC
## 1	1.2601035	1.03332111	1.2601035	1.103486399
## 2	-0.3709597	0.04334384	-0.4112440	0.007470147
## 3	-0.5635528	-1.27571074	-0.4692509	-1.271448875
##	HGLRE_align.H.ADC	GLNU_norm_align.H.ADC	RLNU_norm_align.H.ADC	
## 1	1.2601035	0.98645175	1.26010348	
## 2	-0.4443787	0.05815368	-0.09841951	
## 3	-0.3916857	-1.25710430	-1.20154494	
##	GLVAR_align.H.ADC	RLVAR_align.H.ADC	Entropy_align.H.ADC	SZSE.H.ADC
## 1	1.2601035	1.073384618	1.2601035	1.2601035
## 2	-0.4158694	0.006015476	-0.3107856	-0.1277969
## 3	-0.4584233	-1.233836931	-0.7044152	-1.1327756
##	LGLZE.H.ADC	HGLZE.H.ADC	SZLGE.H.ADC	SZHGE.H.ADC
## 1	1.02516668	1.2601035	0.99445300	1.2601035
## 2	0.04239107	-0.3453515	0.05393276	-0.4141100
## 3	-1.26419521	-0.6234993	-1.25631169	-0.4625418
##	GLNU_area.H.ADC	ZSNU.H.ADC	ZSP.H.ADC	GLNU_norm.H.ADC
## 1	0.51423069	0.5168978	1.2601035	0.98604257
## 2	-0.22985144	-0.2346340	-0.1849442	0.05936074
## 3	-0.04629173	-0.0380691	-0.9989990	-1.25998052
##	GLVAR_area.H.ADC	ZSVAR.H.ADC	Entropy_area.H.ADC	Max_cooc.W.ADC
## 1	1.2601035	0.81922295	1.2601035	0.8102563
## 2	-0.4425550	-0.02618112	-0.3478789	0.1289851
## 3	-0.3959549	-0.86964158	-0.6175828	-1.2227040
##	Average_cooc.W.ADC	Variance_cooc.W.ADC	DAVE_cooc.W.ADC	DVAR_cooc.W.ADC
## 1	1.0308583	0.66975588	1.1715010	0.7247077
## 2	-0.3720482	-0.08003143	-0.2959584	-0.0907048
## 3	-0.3004989	-0.57373994	-0.6384395	-0.6111998
##	DENT_cooc.W.ADC	SAVE_cooc.W.ADC	SVAR_cooc.W.ADC	SENT_cooc.W.ADC
## 1	1.2601035	1.0322287	0.61972023	0.9901887
## 2	-0.3065120	-0.3678297	-0.07787238	-0.3281686
## 3	-0.7144191	-0.3119314	-0.52193538	-0.3570017



```

## ASM_cooc.W.ADC Contrast_cooc.W.ADC Dissimilarity_cooc.W.ADC
## 1 0.8094624 0.7416618 1.1715010
## 2 0.1275774 -0.1101784 -0.2959584
## 3 -1.2185004 -0.5848798 -0.6384395
## Inv_diff_cooc.W.ADC Inv_diff_norm_cooc.W.ADC IDM_cooc.W.ADC
## 1 1.1925307 1.2601035 1.19873300
## 2 -0.1118448 -0.1706458 -0.09016671
## 3 -1.0933303 -1.0324694 -1.15112454
## IDM_norm_cooc.W.ADC Inv_var_cooc.W.ADC Correlation_cooc.W.ADC
## 1 1.260103 1.19080576 1.0204847
## 2 -0.106092 -0.09544937 -0.2087147
## 3 -1.183584 -1.12975006 -0.6710596
## Autocorrelation_cooc.W.ADC Tendency_cooc.W.ADC Shade_cooc.W.ADC
## 1 0.7149876 0.61972023 0.18619581
## 2 -0.2456825 -0.07787238 -0.08227907
## 3 -0.2373656 -0.52193538 -0.01897833
## Prominence_cooc.W.ADC IC1_d.W.ADC IC2_d.W.ADC Coarseness_vdif.W.ADC
## 1 0.30404482 -0.59693323 1.2601035 0.7345515
## 2 0.02841938 0.32916605 -0.2632300 0.1693522
## 3 -0.41203710 -0.09221458 -0.8157384 -1.2311552
## Contrast_vdif.W.ADC Busyness_vdif.W.ADC Complexity_vdif.W.ADC
## 1 0.67024557 0.98696733 0.4927289
## 2 -0.03593965 -0.01151753 -0.1147268
## 3 -0.67751125 -1.09459202 -0.2913541
## Strength_vdif.W.ADC SRE_align.W.ADC LRE_align.W.ADC GLNU_align.W.ADC
## 1 0.57971814 1.26010348 1.2601035 0.56925565
## 2 -0.04365878 -0.05916419 -0.1166881 -0.28907241
## 3 -0.55656938 -1.29343778 -1.1587796 0.02981081
## RLNU_align.W.ADC RP_align.W.ADC LGRE_align.W.ADC HGRE_align.W.ADC
## 1 0.50409113 1.26010348 0.7918681 0.7331058
## 2 -0.22807570 -0.06046407 0.1437900 -0.2416115
## 3 -0.03892635 -1.29039489 -1.2364486 -0.2674843
## LGSRE_align.W.ADC HGSRE_align.W.ADC LGHRE_align.W.ADC HGLRE_align.W.ADC
## 1 0.7942434 0.7328334 0.7780140 0.7398451
## 2 0.1423437 -0.2403440 0.1509209 -0.2440651
## 3 -1.2357620 -0.2701417 -1.2374136 -0.2693989
## GLNU_norm_align.W.ADC RLNU_norm_align.W.ADC GLVAR_align.W.ADC
## 1 0.93830670 1.26010348 0.70985644
## 2 0.07440259 -0.07235316 -0.09716138
## 3 -1.24042617 -1.26256386 -0.57920910
## RLVAR_align.W.ADC Entropy_align.W.ADC SZSE.W.ADC LZSE.W.ADC LGLZE.W.ADC
## 1 1.00970174 1.2601035 1.26010348 1.2601035 0.7975267
## 2 0.03268583 -0.4311495 -0.09510682 -0.2380116 0.1396824
## 3 -1.22390939 -0.4226539 -1.20929938 -0.8747723 -1.2332639
## HGLZE.W.ADC SZLGE.W.ADC SZHGE.W.ADC LZLGE.W.ADC LZHGE.W.ADC GLNU_area.W.ADC
## 1 0.7333844 0.8019663 0.7316883 0.7154832 0.7564613 0.57283746
## 2 -0.2415063 0.1359966 -0.2401366 0.1826938 -0.2542061 -0.29214726
## 3 -0.2680471 -1.2296821 -0.2693260 -1.2407187 -0.2645417 0.03293853
## ZSNU.W.ADC ZSP.W.ADC GLNU_norm.W.ADC ZSNU_norm.W.ADC GLVAR_area.W.ADC
## 1 0.49253950 1.2601035 0.91985828 1.260103 0.7168983
## 2 -0.22612770 -0.1099147 0.08398451 -0.152807 -0.1003412
## 3 -0.03035959 -1.1746353 -1.24189406 -1.074228 -0.5797675
## ZSVAR.W.ADC Entropy_area.W.ADC
## 1 1.02259359 1.2601035

```

```
## 2 -0.03694412          -0.3862893
## 3 -1.07555535          -0.5276676
##
## Clustering vector:
## [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 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
## [75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 2 3 3 3 2 2 2 3 3 3 3 2 2 2 2 2
## [112] 2 3 3 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 2 3 3 3 3 3 1
## [149] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [186] 1 1 1 1 1 1 1 1 1 1 1 1 1
##
## Within cluster sum of squares by cluster:
## [1] 13412.45 24980.82 10409.39
## (between_SS / total_SS =  41.9 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
clustering_kmeans <- kmeans(normalized, centers = 3, iter.max = 100, nstart = 100)
```

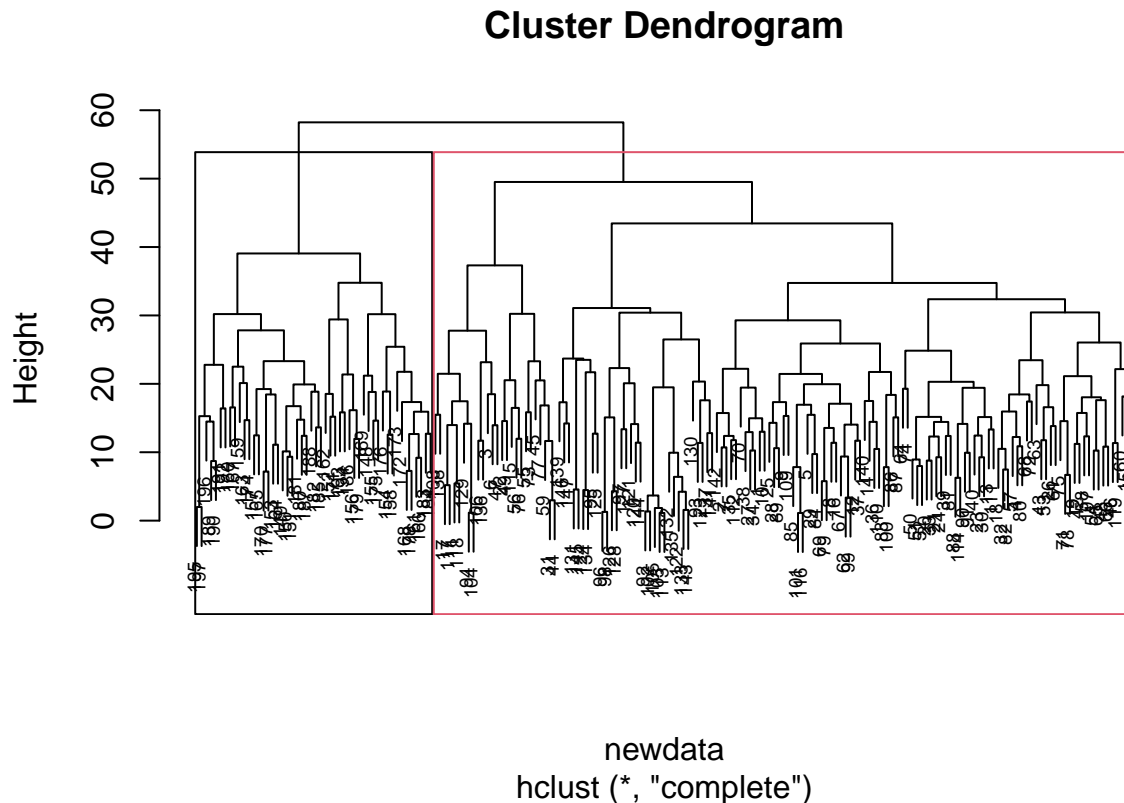
The three clusters are made which are 50, 103 and 44 and within the clusters the sum of square is 41.9% which tells the quality of partition.

## Heirarchical Clustering

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

Plot of dendrogram:

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



Above is the heirarchy of clusters, where x-axis is the distance matrix and y-axis as height.

```
clustering_heir
```

```
##
## Call:
## hclust(d = newdata, method = "complete")
##
## Cluster method   : complete
## Distance         : euclidean
## Number of objects: 197
```

In creating the model, the clustering method is complete, the distance is euclidean and no, of objects are 197.

## Model-based Clustering

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

```
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VEI (diagonal, equal shape) model with 3 components:
##
##   log-likelihood    n df         BIC          ICL
##   -2200.968 197 44 -4634.396 -4642.41
```

```
##
## Clustering table:
##  1  2  3
## 95 44 58
```

Model-based clustering with 3 components has size of 95, 44 and 58 with Bayesian information Criteria of -1462.41.

```
Clustering_mbased1 <- Mclust(normalized[,1:10], G=2) # due to lower RAM
summary(Clustering_mbased1)
```

```
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
##
## Mclust VVE (ellipsoidal, equal orientation) model with 2 components:
##
## log-likelihood   n df      BIC      ICL
##      -1037.511 197 86 -2529.378 -2533.265
##
## Clustering table:
##   1   2
##  66 131
```

We can conclude that 3 components is better than 2 components.

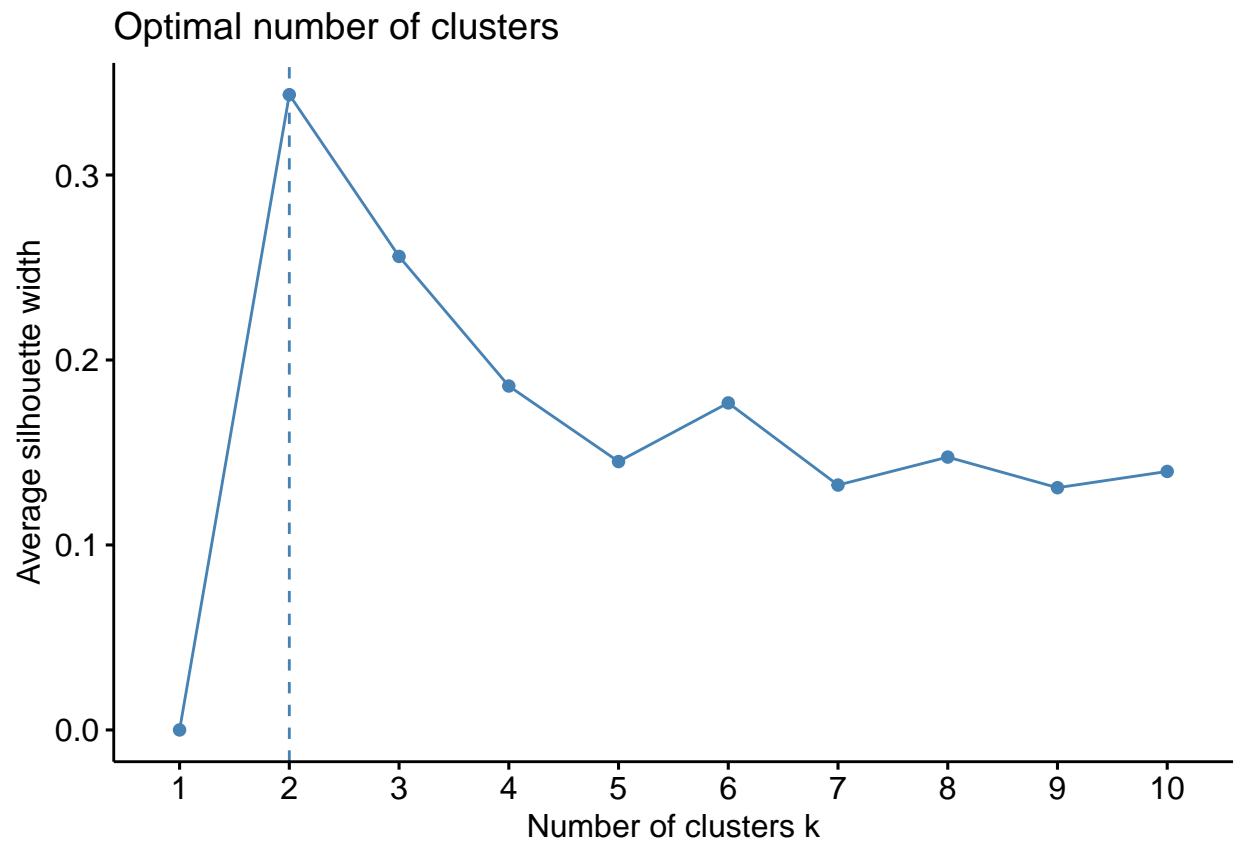
## Comparison of different clustering techniques.

For comparison, we used the average silhouette width.

The average distance of points in the cluster to which it was assigned, and is the average distance of points in the nearest cluster to which it was not assigned.

For K-Means:

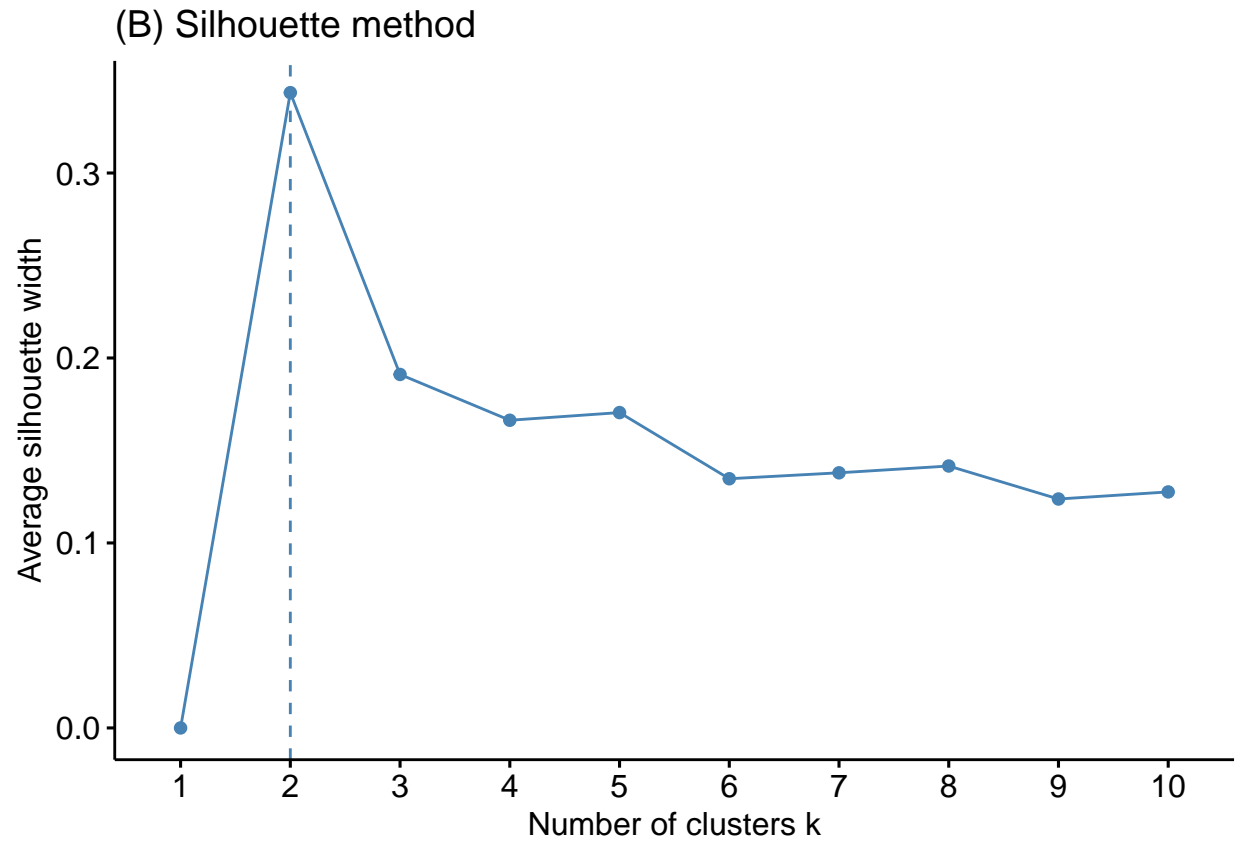
```
fviz_nbclust(normalized, kmeans, method = "silhouette")
```



Above plot, suggest that the optimal silhouette width is achieved in 2 clusters using K-means clustering.

For Heirarchical Clustering:

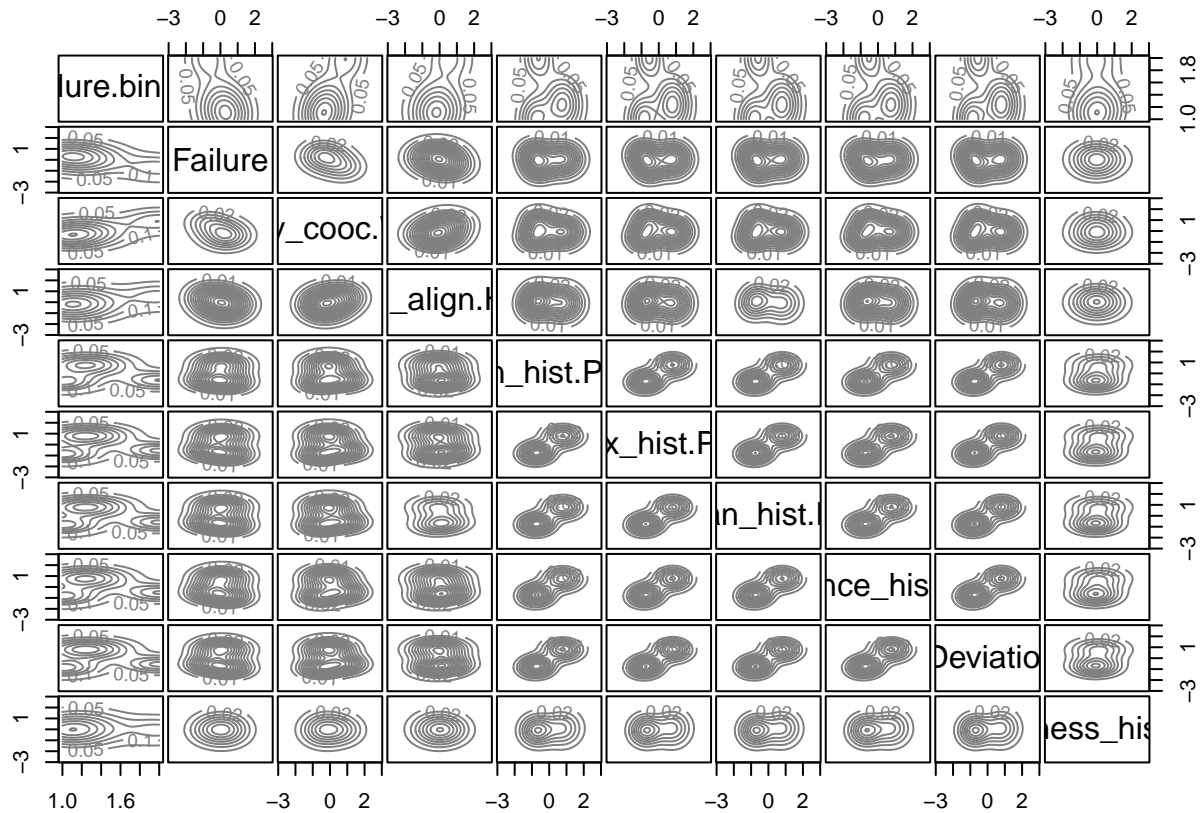
```
plot_heir <- fviz_nbclust(normalized, FUN = hcut, method = "silhouette",  
                        k.max = 10) +  
  ggtitle("(B) Silhouette method")  
plot_heir
```



Same with K-means, above plot suggest also that in order to obtain optimal silhouette width of approximately 0.33, cluster  $k=2$ .

For Model-based: *No silhouette plot available in model based.*

```
plot(Clustering_mbased, "density")
```



```
cs = cluster.stats(dist(normalized[,1:10]), Clustering_mbased$classification)
cs[c("avg.silwidth")]
```

```
## $avg.silwidth
## [1] 0.2090831
```

For model-based the average silhouette width is approximately 0.21.

Thus, we can say that when components are considered the model-based preformed better but if 2 cluster component were considered the other two is much better.