Parent and Provider Perceptions of Behavioral Healthcare in Pediatric Primary Care (PI: Andrew Riley; BDP2-262)

Benjamin Chan (chanb@ohsu.edu)

2018-07-05

# Import Andrew’s SPSS data

Map new names to variables.

|  |  |
| --- | --- |
| oldnames | newnames |
| record\_id | id |
| eng\_span | languageSurvey |
| children\_totv\_1 | totalChildren |
| oldest\_middle\_youngest | birthOrder |
| child\_sexv\_1 | childSex |
| child\_age\_years | childAge |
| child\_ethnicity | childEthnicity |
| child\_racev\_1\_\_\_1 | childRaceWhite |
| child\_racev\_1\_\_\_2 | childRaceAsian |
| child\_racev\_1\_\_\_3 | childRaceAfrAm |
| child\_racev\_1\_\_\_4 | childRaceAIAN |
| child\_racev\_1\_\_\_5 | childRaceNHPI |
| child\_racev\_1\_\_\_6 | childRaceOther |
| child\_racev\_1\_\_\_7 | childRaceNoResp |
| related\_child | childRelationship |
| gender | parentGender |
| parent\_sexv\_1 | parentSex |
| parent\_agev\_1 | parentAge |
| parent\_ethnicity | parentEthnicity |
| parent\_race\_\_\_1 | parentRaceWhite |
| parent\_race\_\_\_2 | parentRaceAsian |
| parent\_race\_\_\_3 | parentRaceAfrAm |
| parent\_race\_\_\_4 | parentRaceAIAN |
| parent\_race\_\_\_5 | parentRaceNHPI |
| parent\_race\_\_\_6 | parentRaceOther |
| parent\_race\_\_\_7 | parentRaceNoResp |
| marital\_status | parentMaritalStatus |
| parenting\_situationv\_1 | parentSituation |
| number\_parents | parentsNumber |
| parent\_to\_child\_ratio | parentChildRatio |
| zipcode\_classification\_combined | zipcodeClass |
| zipcode | zipcode |
| community\_type | community |
| distance | distance |
| parent\_educationv\_1 | parentEducation |
| annual\_income | income |
| internet | internet |
| ECBI\_intensity\_raw\_score | ECBI\_intensity\_raw\_score |
| ECBI\_intensity\_T\_score | ECBI\_intensity\_T\_score |
| ECBI\_intensity\_clinical\_cutoff | ECBI\_intensity\_clinical\_cutoff |
| ECBI\_problem\_raw\_score | ECBI\_problem\_raw\_score |
| ECBI\_problem\_T\_score | ECBI\_problem\_T\_score |
| ECBI\_problem\_clinical\_cutoff | ECBI\_problem\_clinical\_cutoff |
| ECBI\_Opp | ECBI\_Opp |
| ECBI\_Inatt | ECBI\_Inatt |
| ECBI\_Cond | ECBI\_Cond |
| MAPS\_PP | MAPS\_PP |
| MAPS\_PR | MAPS\_PR |
| MAPS\_WM | MAPS\_WM |
| MAPS\_SP | MAPS\_SP |
| MAPS\_HS | MAPS\_HS |
| MAPS\_LC | MAPS\_LC |
| MAPS\_PC | MAPS\_PC |
| MAPS\_POS | MAPS\_POS |
| MAPS\_NEG | MAPS\_NEG |
| SEPTI\_nurturance | SEPTI\_nurturance |
| SEPTI\_n\_clinical\_cutoff | SEPTI\_n\_clinical\_cutoff |
| SEPTI\_discipline | SEPTI\_discipline |
| SEPTI\_d\_clinical\_cutoff | SEPTI\_d\_clinical\_cutoff |
| SEPTI\_play | SEPTI\_play |
| SEPTI\_p\_clinical\_cutoff | SEPTI\_p\_clinical\_cutoff |
| SEPTI\_routine | SEPTI\_routine |
| SEPTI\_r\_clinical\_cutoff | SEPTI\_r\_clinical\_cutoff |
| SEPTI\_total | SEPTI\_total |
| SEPTI\_total\_clin\_cutoff | SEPTI\_total\_clin\_cutoff |
| PCB1\_Total | PCB1\_Total |
| PCB1\_CondEmot | PCB1\_CondEmot |
| PCB1\_DevHab | PCB1\_DevHab |
| PCB2\_Tot | PCB2\_Tot |
| PCB3\_Total | PCB3\_Total |
| PBC3\_PCPonly | PCB3\_PCPonly |
| PCB3\_Person | PCB3\_Person |
| PCB3\_Resource | PCB3\_Resource |

## Warning: package 'bindrcpp' was built under R version 3.4.4

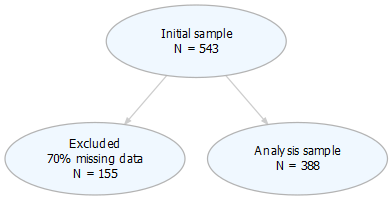
Remove certain predictor variables:

* Clinical cutoffs
* Raw scores
* Total scores

## [1] "ECBI\_intensity\_raw\_score" "ECBI\_intensity\_clinical\_cutoff"  
## [3] "ECBI\_problem\_raw\_score" "ECBI\_problem\_clinical\_cutoff"   
## [5] "SEPTI\_n\_clinical\_cutoff" "SEPTI\_d\_clinical\_cutoff"   
## [7] "SEPTI\_p\_clinical\_cutoff" "SEPTI\_r\_clinical\_cutoff"   
## [9] "SEPTI\_total" "SEPTI\_total\_clin\_cutoff"

Build analysis data set. Exclude if missing any dependent variable, PCB1\_Total, PCB2\_Tot, PCB3\_Total. Exclude rows if there are a high proportion of row-wise NA.

## PCB1\_Total PCB2\_Tot PCB3\_Total   
## Min. :18.00 Min. : 6.00 Min. :15.00   
## 1st Qu.:58.00 1st Qu.:22.00 1st Qu.:39.00   
## Median :71.00 Median :25.00 Median :48.00   
## Mean :67.85 Mean :24.53 Mean :47.58   
## 3rd Qu.:81.00 3rd Qu.:28.00 3rd Qu.:57.00   
## Max. :90.00 Max. :30.00 Max. :75.00



figures/flowChart.png

<https://uc-r.github.io/hc_clustering> <http://www.sthda.com/english/wiki/factoextra-r-package-easy-multivariate-data-analyses-and-elegant-visualization>

## Warning: package 'cluster' was built under R version 3.4.4

## Warning: package 'ggdendro' was built under R version 3.4.4

## Warning: package 'factoextra' was built under R version 3.4.4

## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ

## Warning: package 'dendextend' was built under R version 3.4.4

##   
## ---------------------  
## Welcome to dendextend version 1.8.0  
## Type citation('dendextend') for how to cite the package.  
##   
## Type browseVignettes(package = 'dendextend') for the package vignette.  
## The github page is: https://github.com/talgalili/dendextend/  
##   
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues  
## Or contact: <tal.galili@gmail.com>  
##   
## To suppress this message use: suppressPackageStartupMessages(library(dendextend))  
## ---------------------

##   
## Attaching package: 'dendextend'

## The following object is masked from 'package:ggdendro':  
##   
## theme\_dendro

## The following object is masked from 'package:stats':  
##   
## cutree

##   
## To cite package 'factoextra' in publications use:  
##   
## Alboukadel Kassambara and Fabian Mundt (2017). factoextra:  
## Extract and Visualize the Results of Multivariate Data Analyses.  
## R package version 1.0.5.  
## https://CRAN.R-project.org/package=factoextra  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {factoextra: Extract and Visualize the Results of Multivariate Data Analyses},  
## author = {Alboukadel Kassambara and Fabian Mundt},  
## year = {2017},  
## note = {R package version 1.0.5},  
## url = {https://CRAN.R-project.org/package=factoextra},  
## }

## [1] 345 63

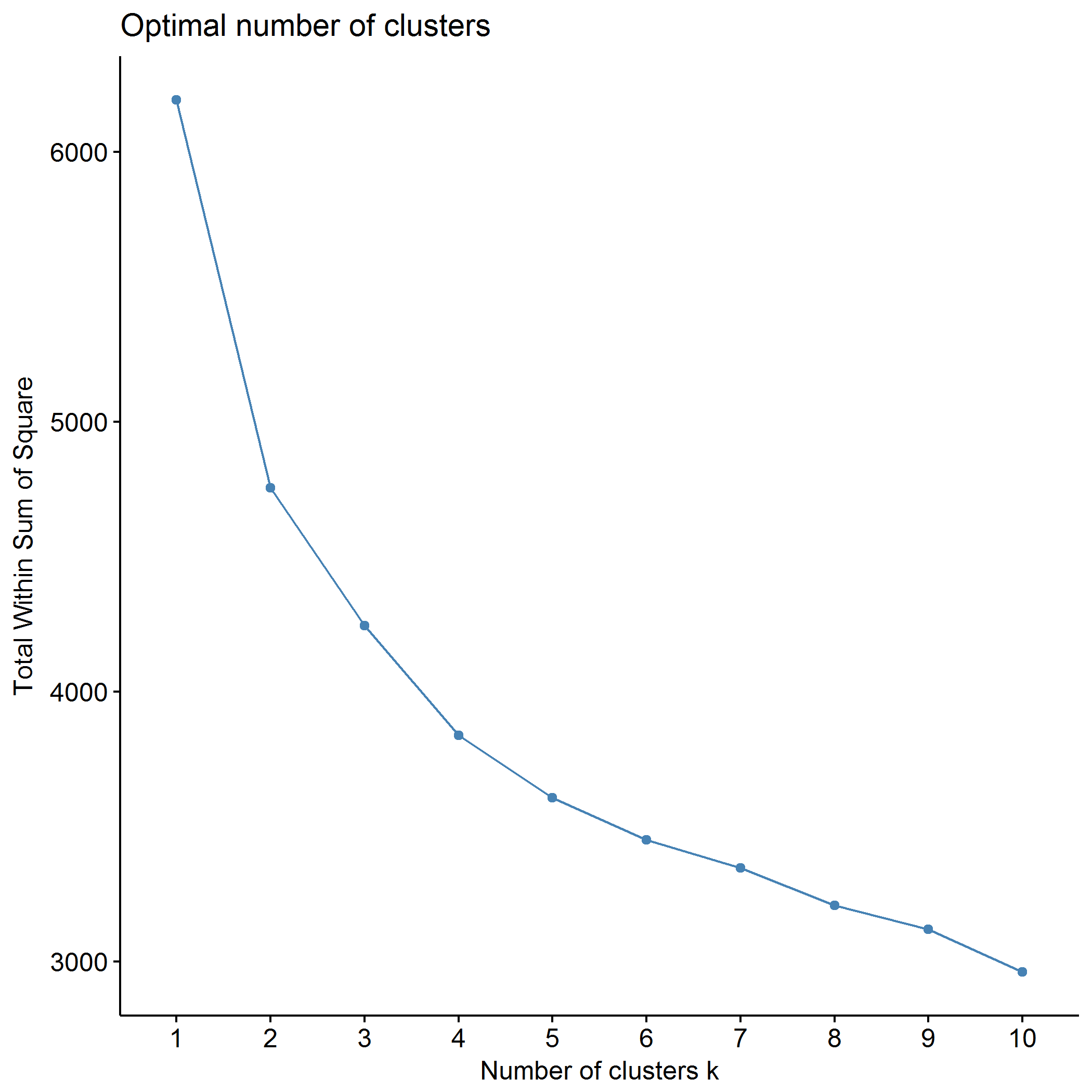
## [1] 345 18

## [1] 345 8

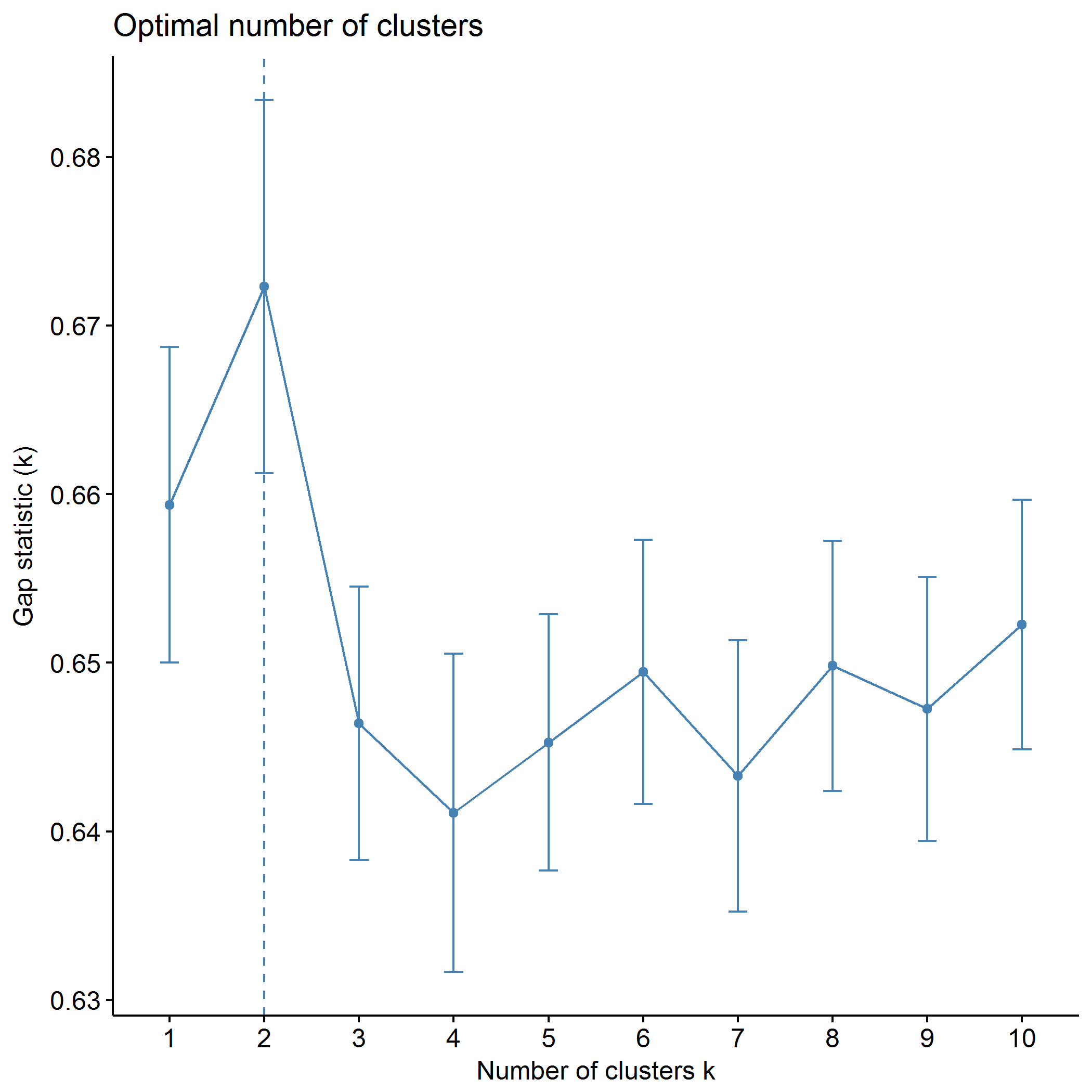
## [1] 345 18

## NULL

## K-means clustering

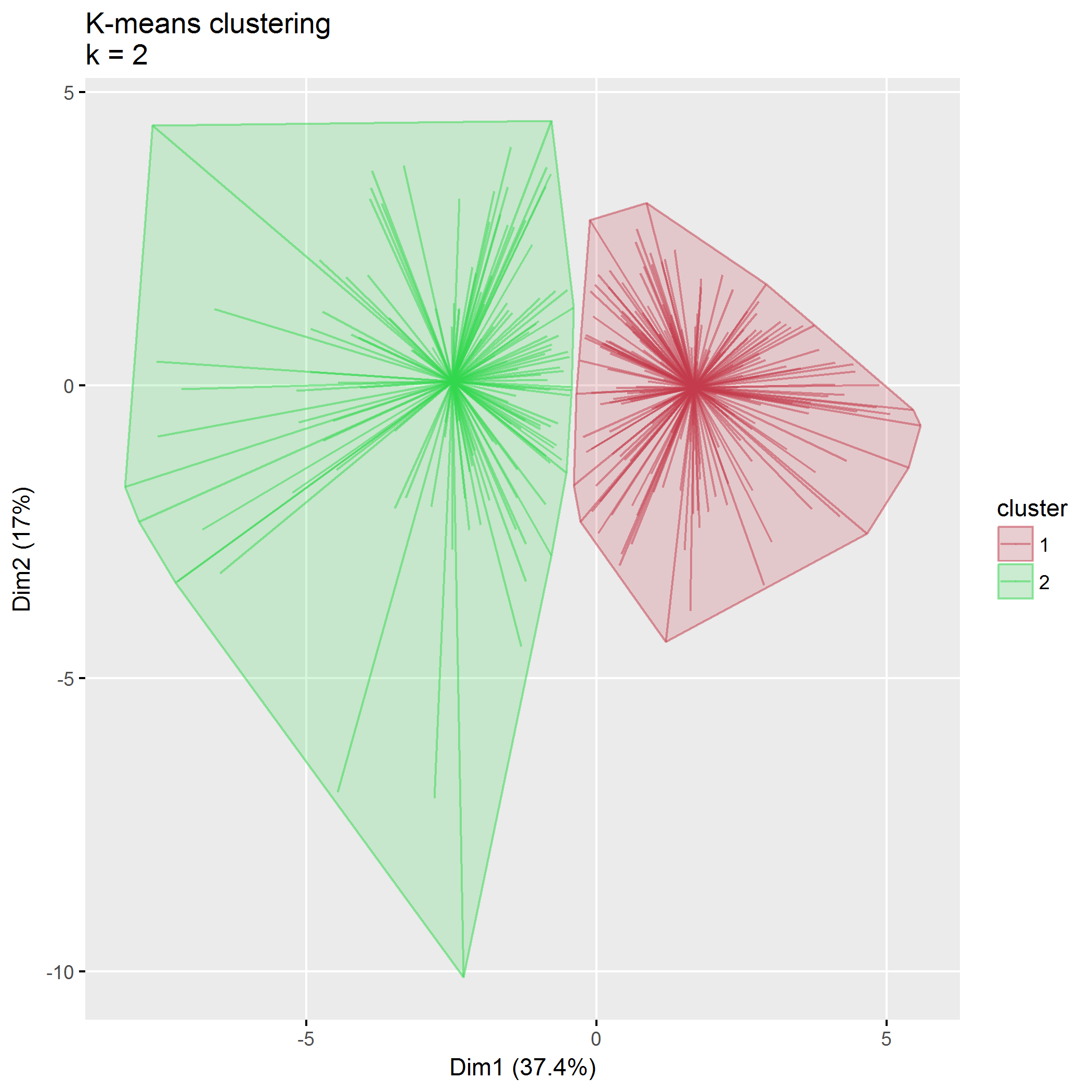


## Clustering k = 1,2,..., K.max (= 10): .. done  
## Bootstrapping, b = 1,2,..., B (= 500) [one "." per sample]:  
## .................................................. 50   
## .................................................. 100   
## .................................................. 150   
## .................................................. 200   
## .................................................. 250   
## .................................................. 300   
## .................................................. 350   
## .................................................. 400   
## .................................................. 450   
## .................................................. 500



plot of chunk predictors\_kmeans

## .  
## 1 2   
## 206 139



plot of chunk predictors\_kmeans

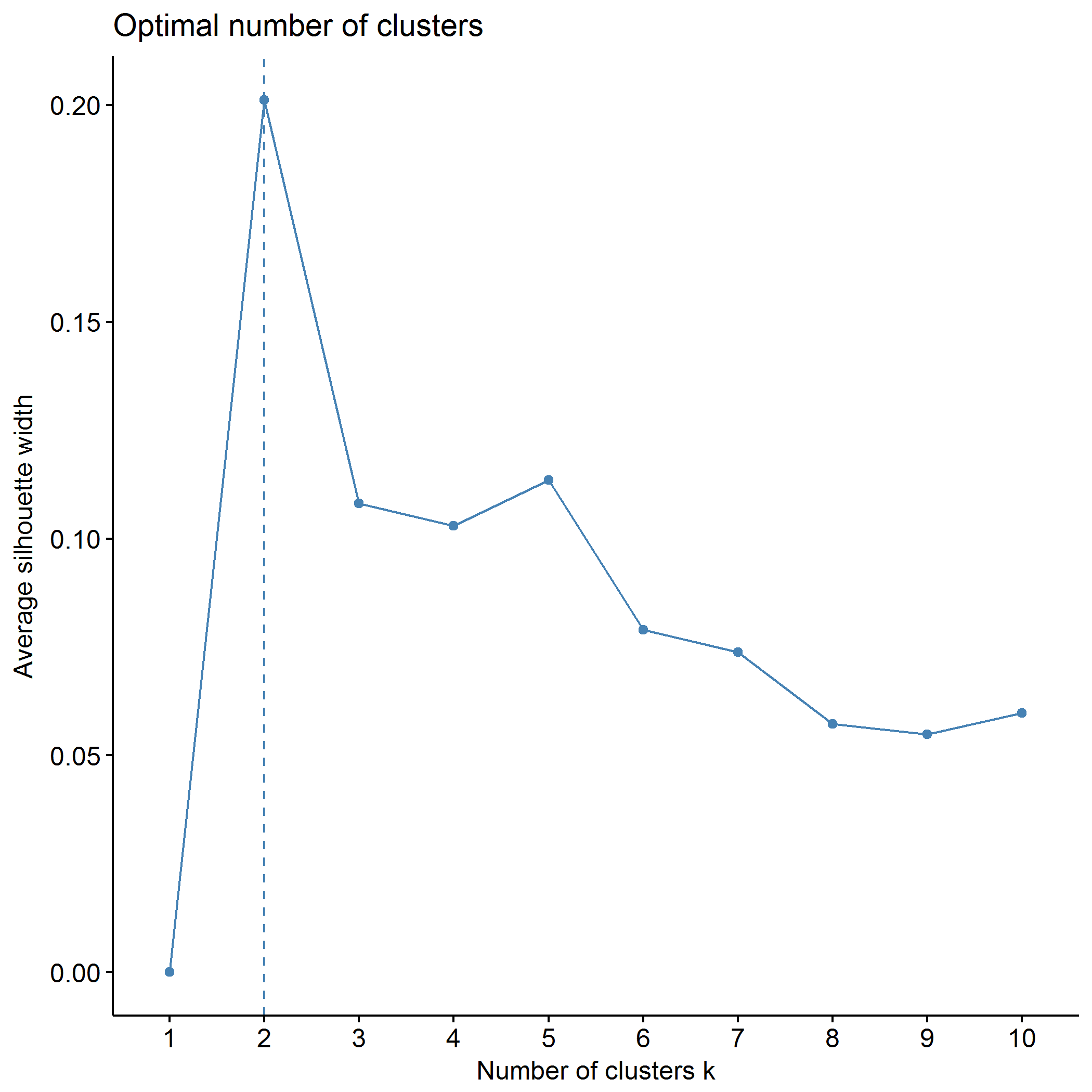
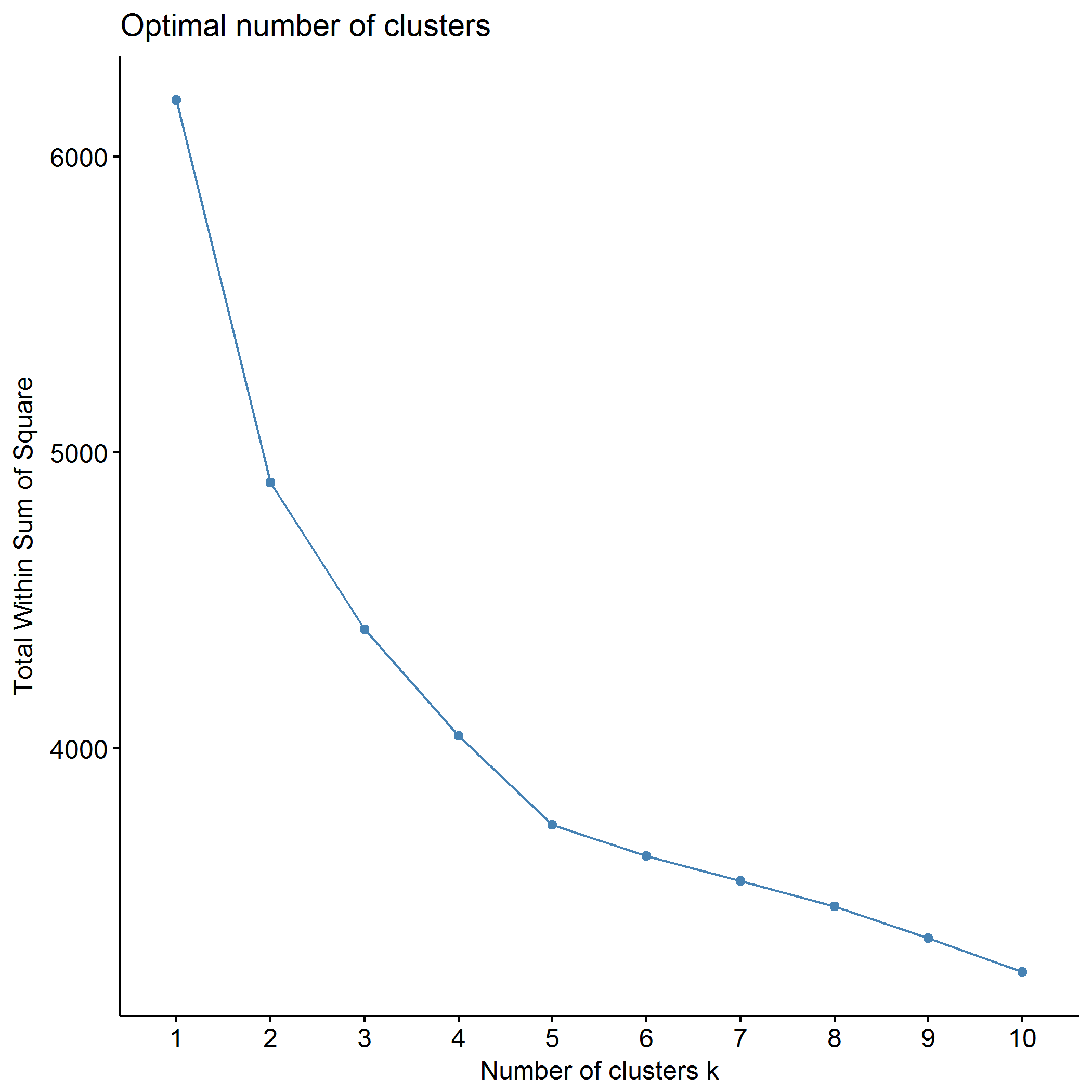
## Within cluster sum of squares, cluster 1: 2116.72  
## Within cluster sum of squares, cluster 2: 2638.50

## Between SS / Total SS: 1436.78 / 6192.00 = 23.20%

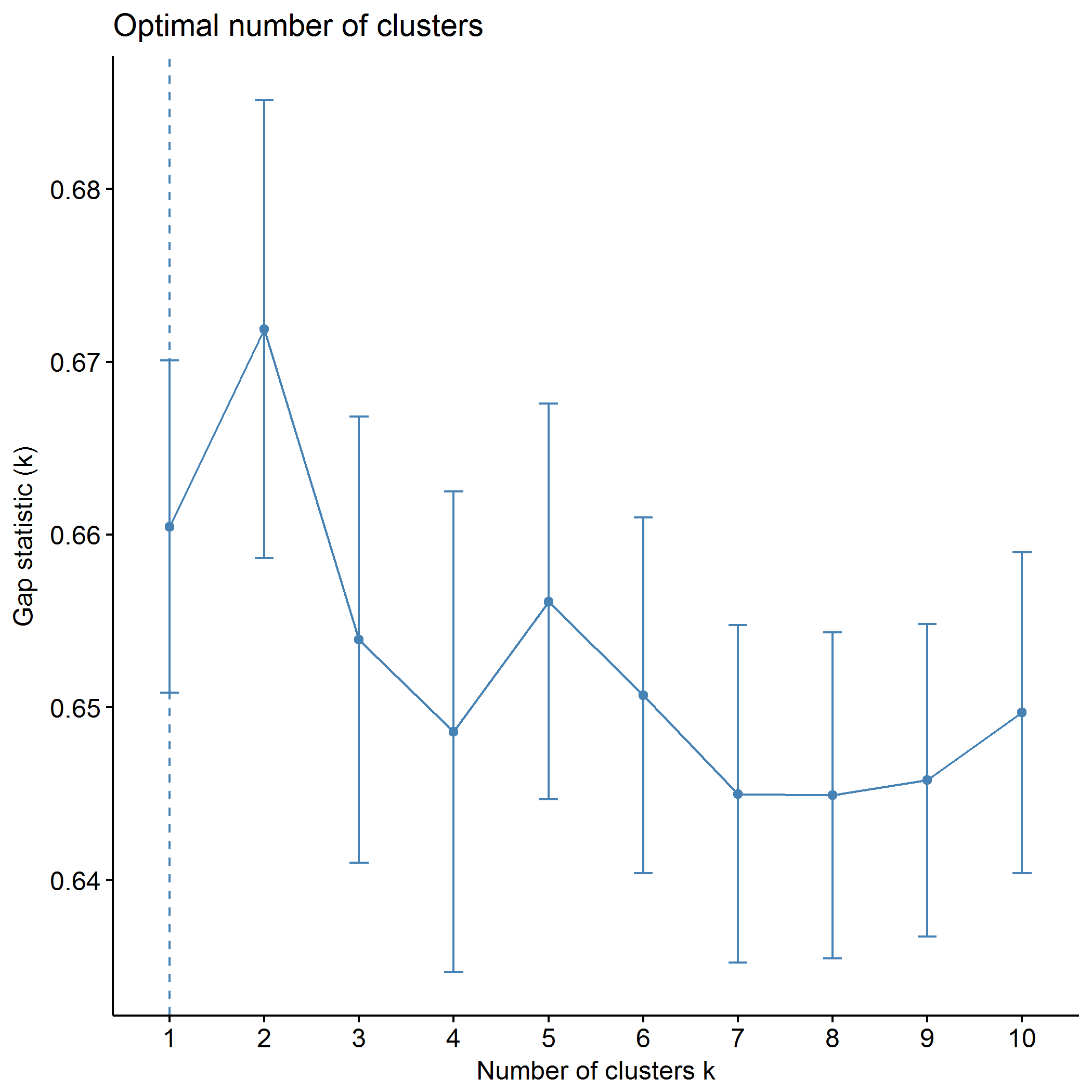
## Total within SS: 4755.22

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| ECBI\_intensity\_T\_score | -0.43 | 0.63 |
| ECBI\_problem\_T\_score | -0.42 | 0.62 |
| ECBI\_Opp | -0.43 | 0.63 |
| ECBI\_Inatt | -0.21 | 0.31 |
| ECBI\_Cond | -0.36 | 0.53 |
| MAPS\_PP | 0.33 | -0.48 |
| MAPS\_PR | 0.35 | -0.52 |
| MAPS\_WM | 0.34 | -0.50 |
| MAPS\_SP | 0.43 | -0.64 |
| MAPS\_HS | -0.43 | 0.64 |
| MAPS\_LC | -0.36 | 0.54 |
| MAPS\_PC | -0.30 | 0.44 |
| MAPS\_POS | 0.46 | -0.68 |
| MAPS\_NEG | -0.48 | 0.71 |
| SEPTI\_nurturance | 0.47 | -0.69 |
| SEPTI\_discipline | 0.45 | -0.67 |
| SEPTI\_play | 0.37 | -0.55 |
| SEPTI\_routine | 0.41 | -0.61 |

## Partitioning around medoids (PAM)

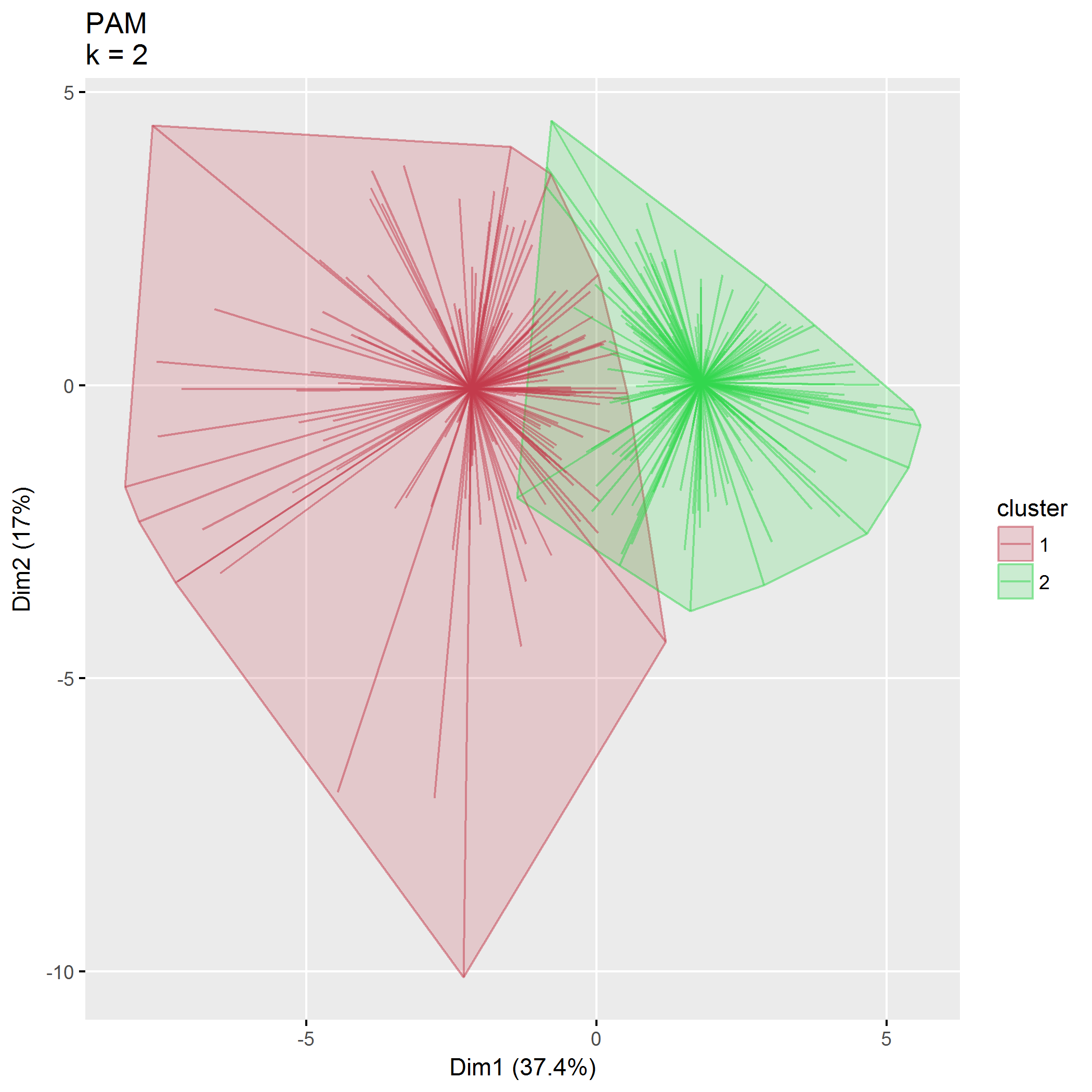


## Clustering k = 1,2,..., K.max (= 10): .. done  
## Bootstrapping, b = 1,2,..., B (= 500) [one "." per sample]:  
## .................................................. 50   
## .................................................. 100   
## .................................................. 150   
## .................................................. 200   
## .................................................. 250   
## .................................................. 300   
## .................................................. 350   
## .................................................. 400   
## .................................................. 450   
## .................................................. 500



plot of chunk predictors\_pam

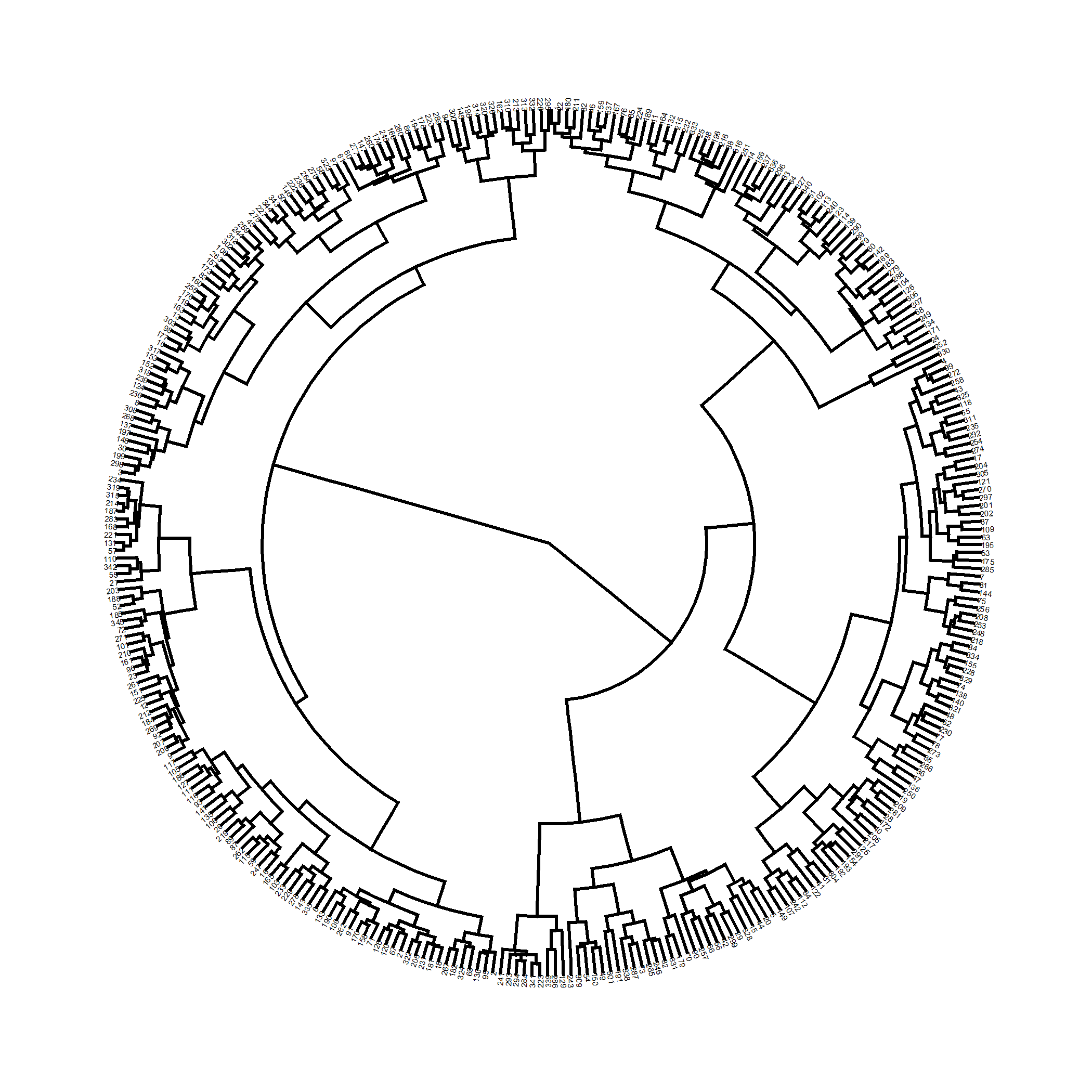
## .  
## 1 2   
## 157 188



plot of chunk predictors\_pam

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| size | max\_diss | av\_diss | diameter | separation |
| 157 | 41.80 | 14.97 | 65.48 | 6.16 |
| 188 | 23.96 | 11.43 | 36.86 | 6.16 |
|  | 321 | 298 |
| ECBI\_intensity\_T\_score | 0.19 | -0.38 |
| ECBI\_problem\_T\_score | 0.86 | -0.71 |
| ECBI\_Opp | 0.71 | -0.06 |
| ECBI\_Inatt | -0.60 | -0.12 |
| ECBI\_Cond | -0.09 | -0.39 |
| MAPS\_PP | 0.18 | 0.77 |
| MAPS\_PR | 0.01 | 0.47 |
| MAPS\_WM | 0.13 | 0.78 |
| MAPS\_SP | -0.63 | 0.97 |
| MAPS\_HS | 0.49 | -0.42 |
| MAPS\_LC | 0.43 | -0.22 |
| MAPS\_PC | -0.04 | -0.74 |
| MAPS\_POS | -0.12 | 0.95 |
| MAPS\_NEG | 0.38 | -0.62 |
| SEPTI\_nurturance | -0.35 | 0.42 |
| SEPTI\_discipline | -1.24 | 0.90 |
| SEPTI\_play | -1.11 | -0.82 |
| SEPTI\_routine | -0.51 | 0.08 |

## Agglomerative hierarchical clustering (AGNES)



plot of chunk predictors\_agnes

Correlation between cophenetic distance and the original distance is 0.417.

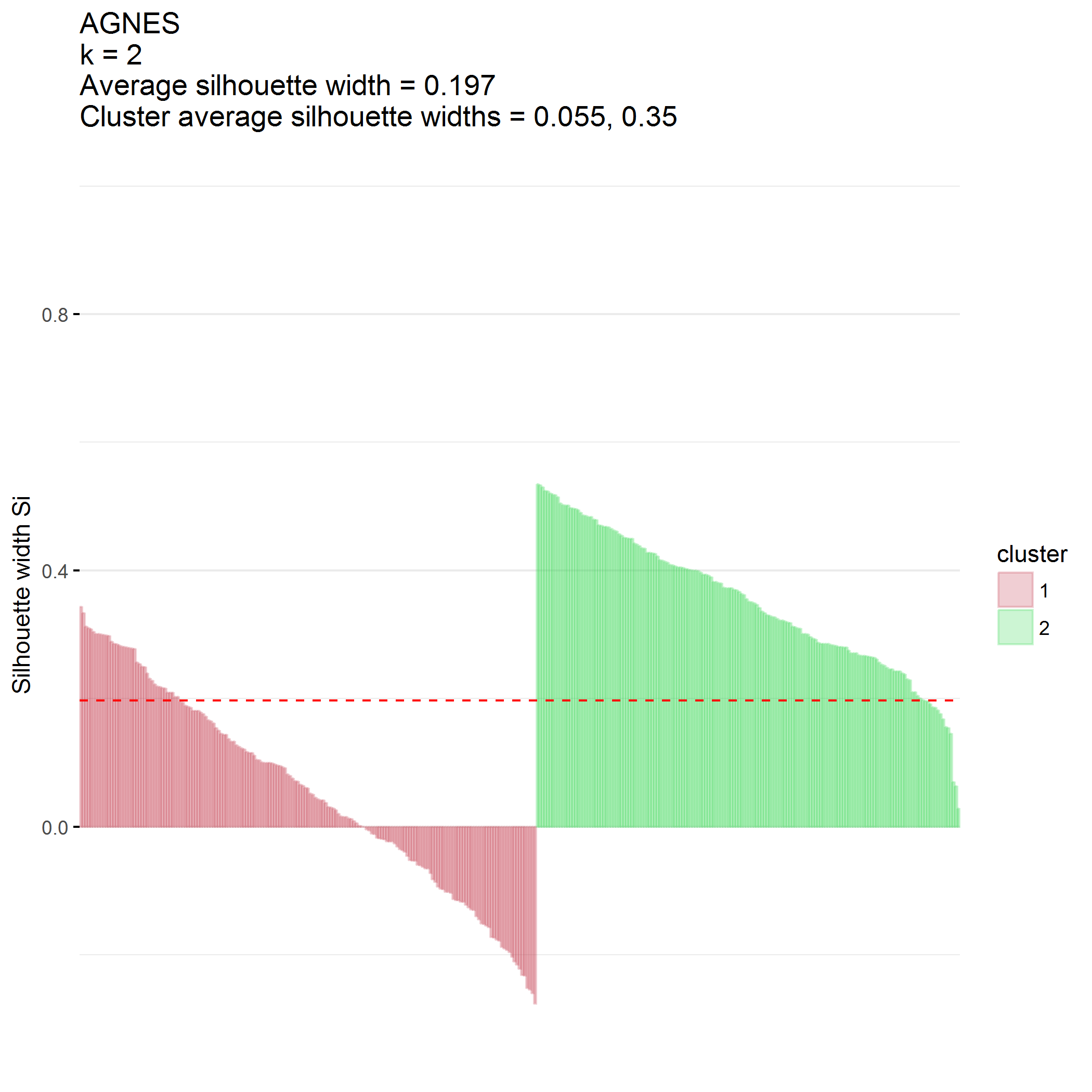
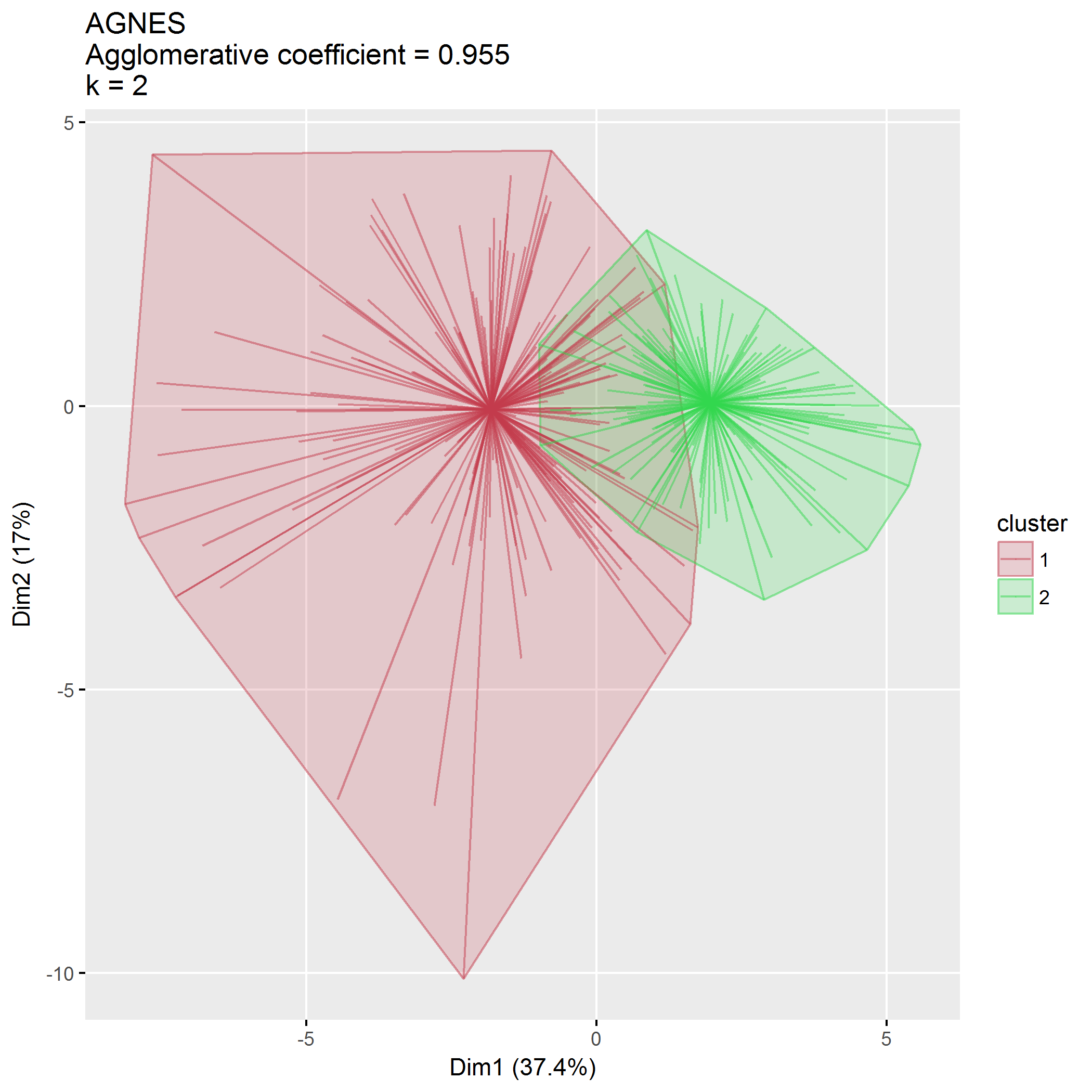
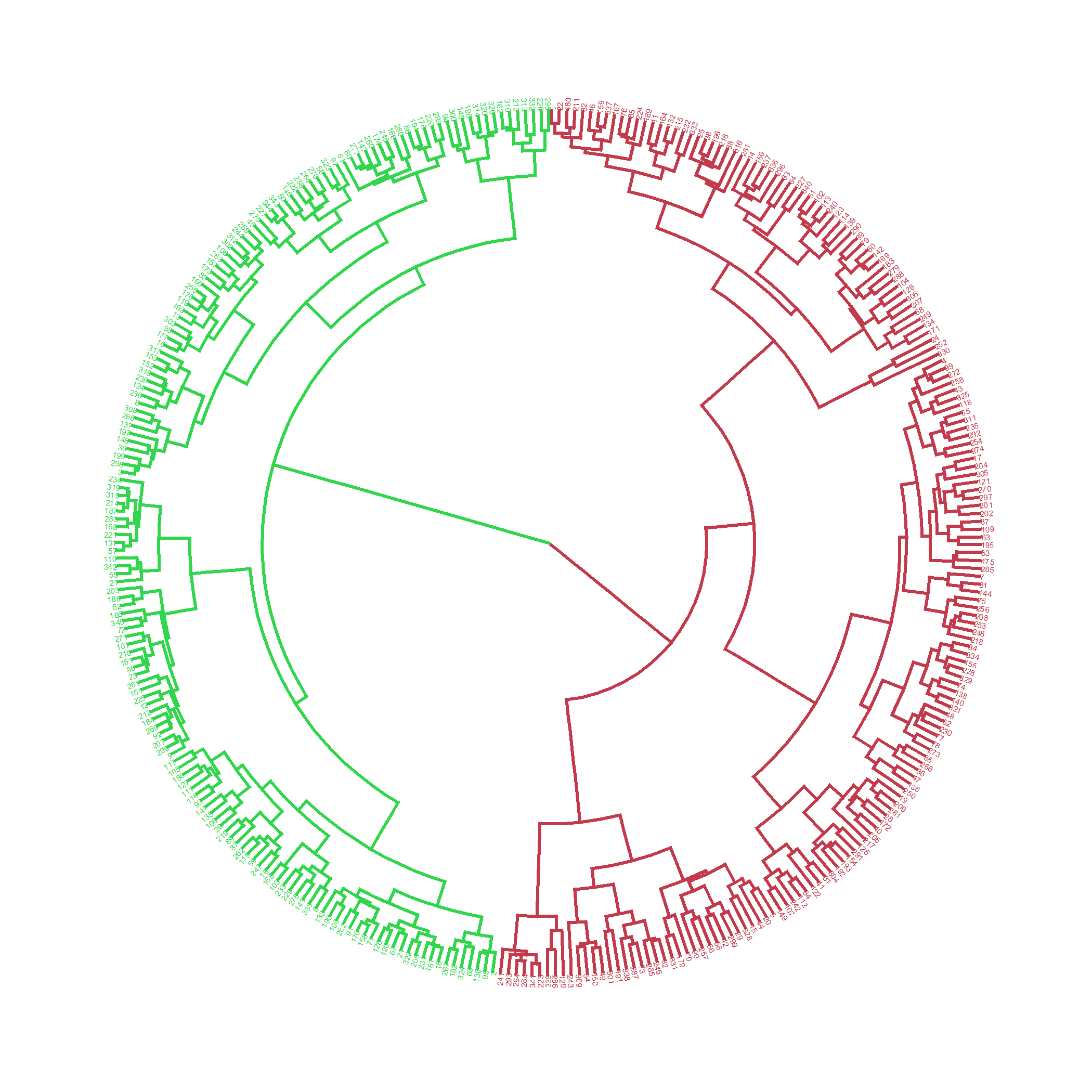
The closer the value of the correlation coefficient is to 1, the more accurately the clustering solution reflects your data. Values above 0.75 are felt to be good.

Agglomerative coeffficient using the Ward method is 0.955.

### clusters

## cluster size ave.sil.width  
## 1 1 179 0.05  
## 2 2 166 0.35

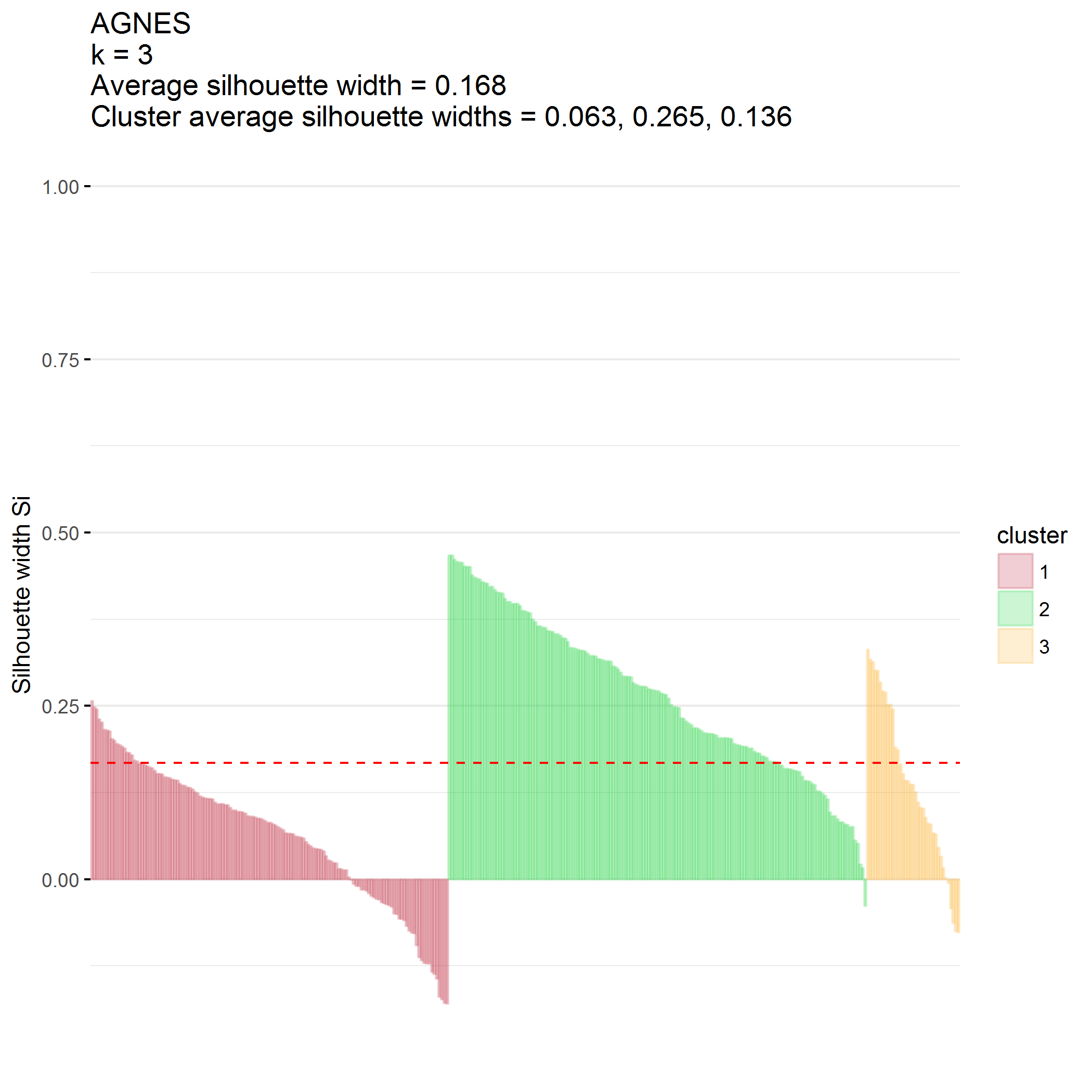
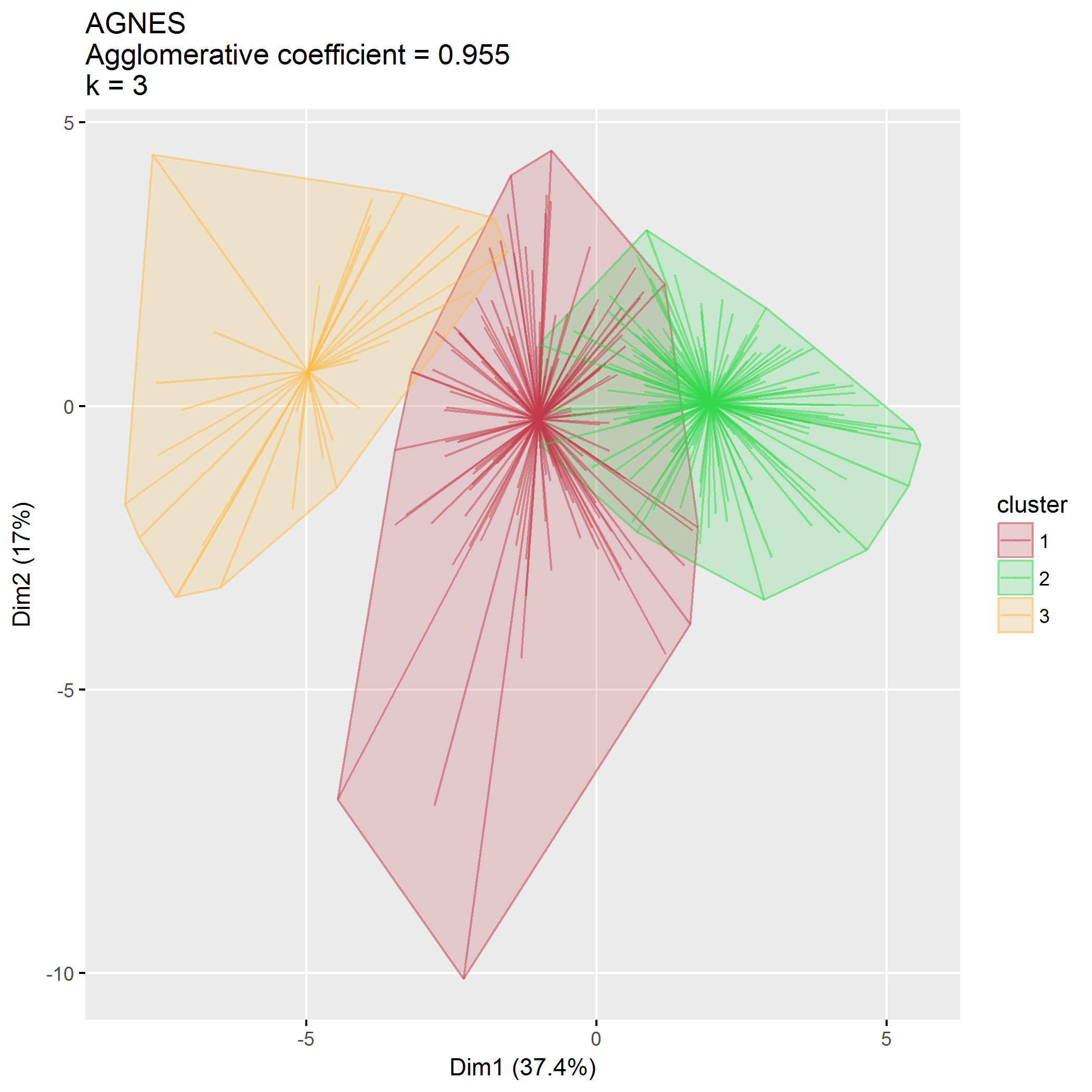
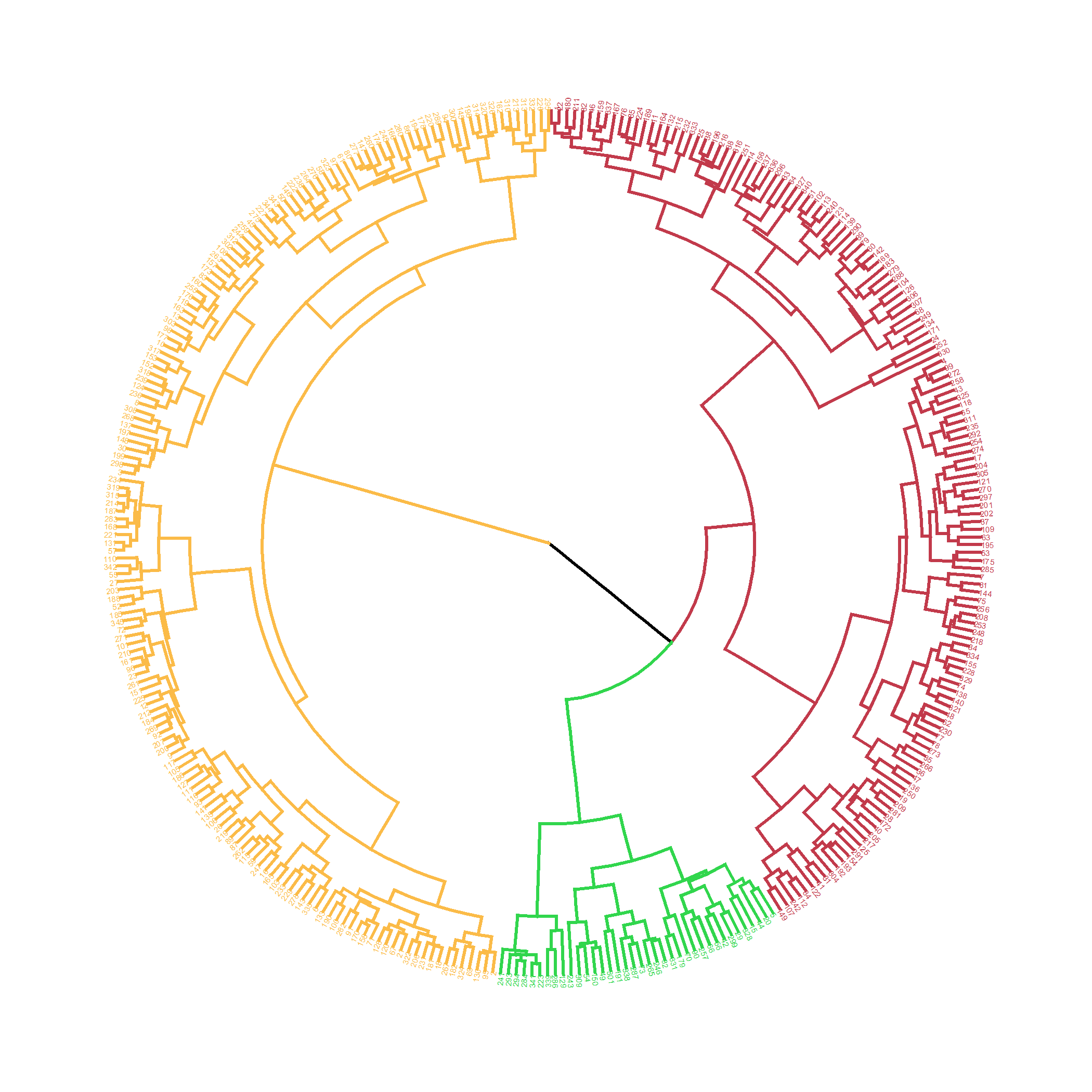
## .  
## 1 2   
## 179 166



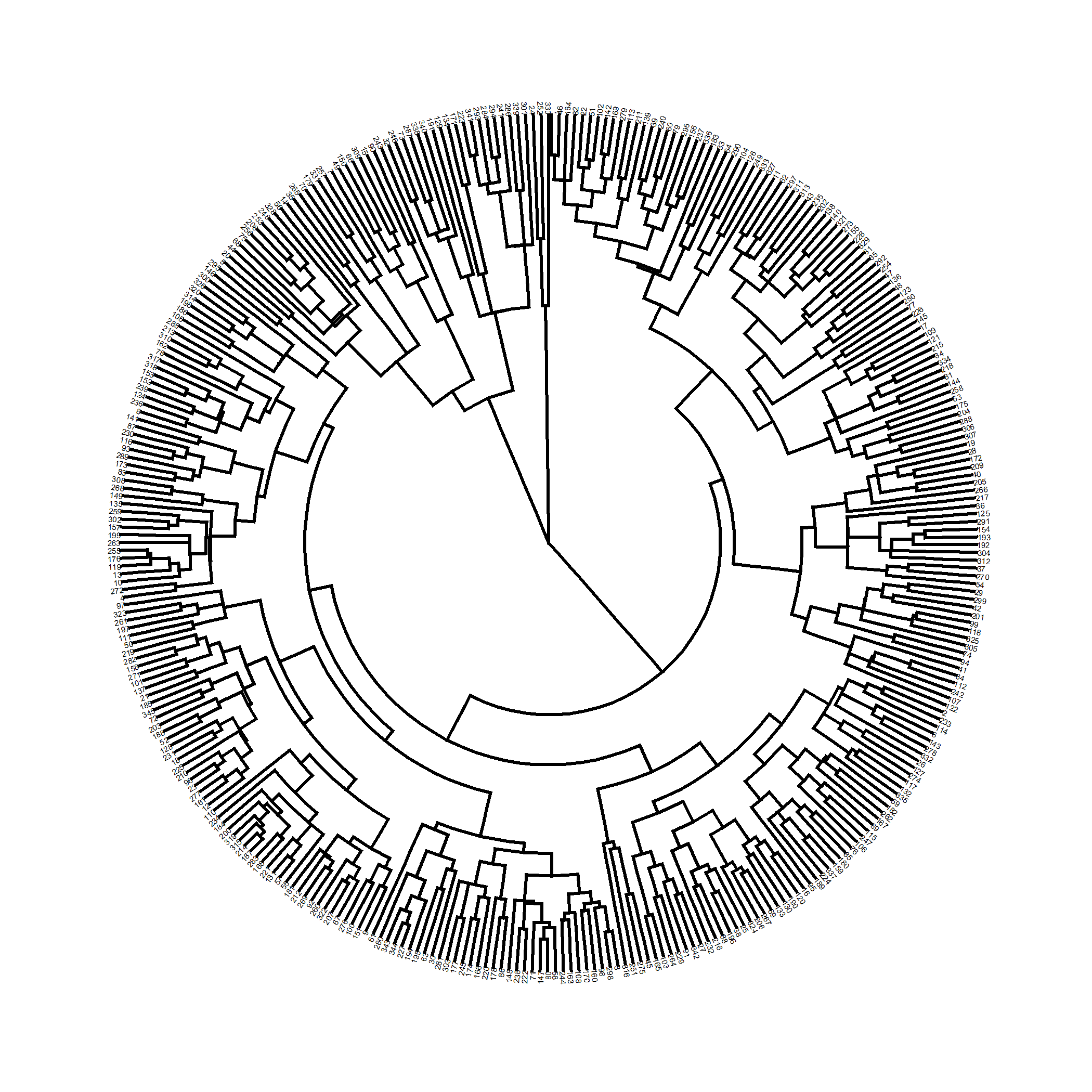
### clusters

## cluster size ave.sil.width  
## 1 1 142 0.06  
## 2 2 166 0.26  
## 3 3 37 0.14

## .  
## 1 2 3   
## 142 166 37



## Divisive hierarchical clustering (DIANA)



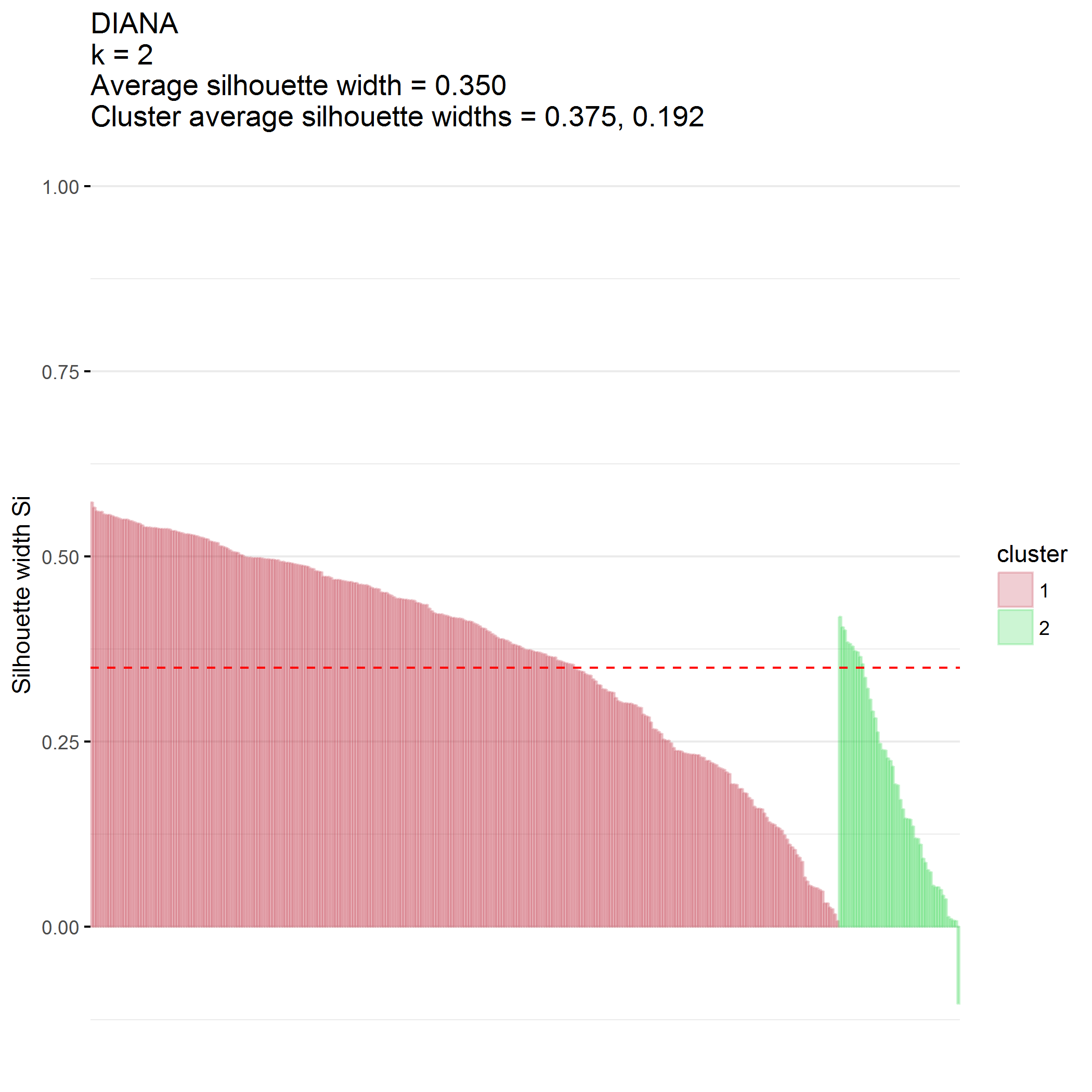
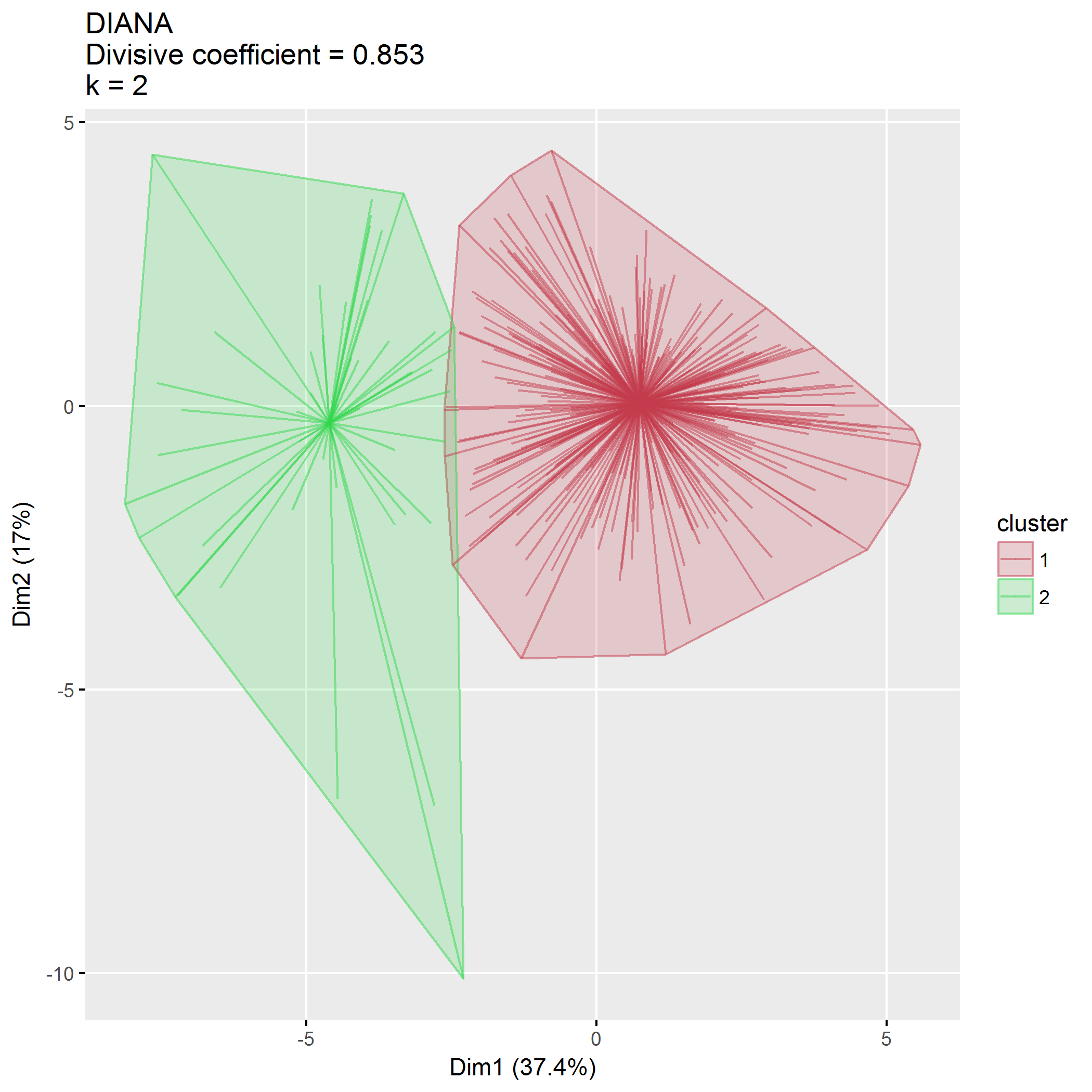
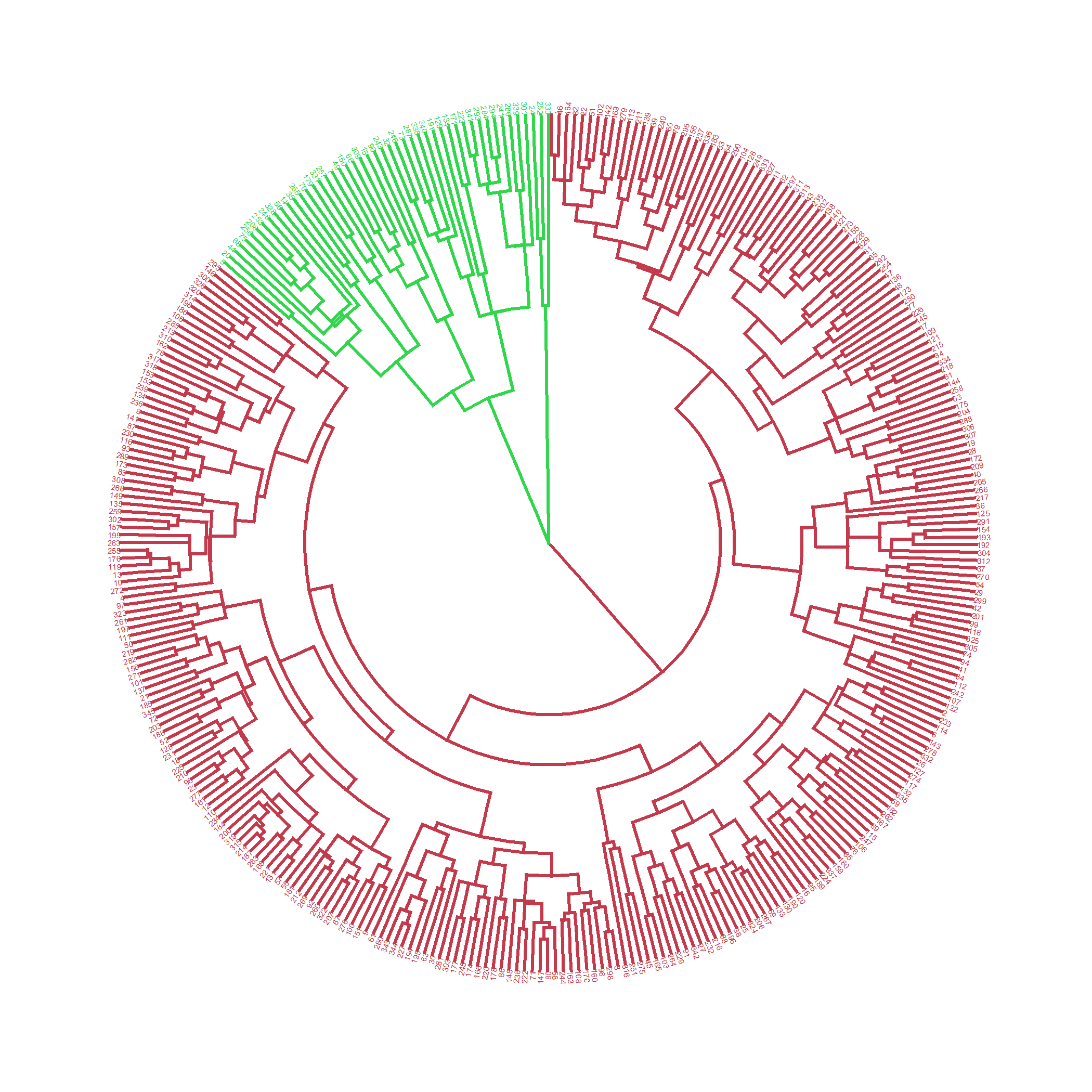
plot of chunk predictors\_diana

Divisive coefficient is 0.853.

### clusters

## cluster size ave.sil.width  
## 1 1 297 0.38  
## 2 2 48 0.19

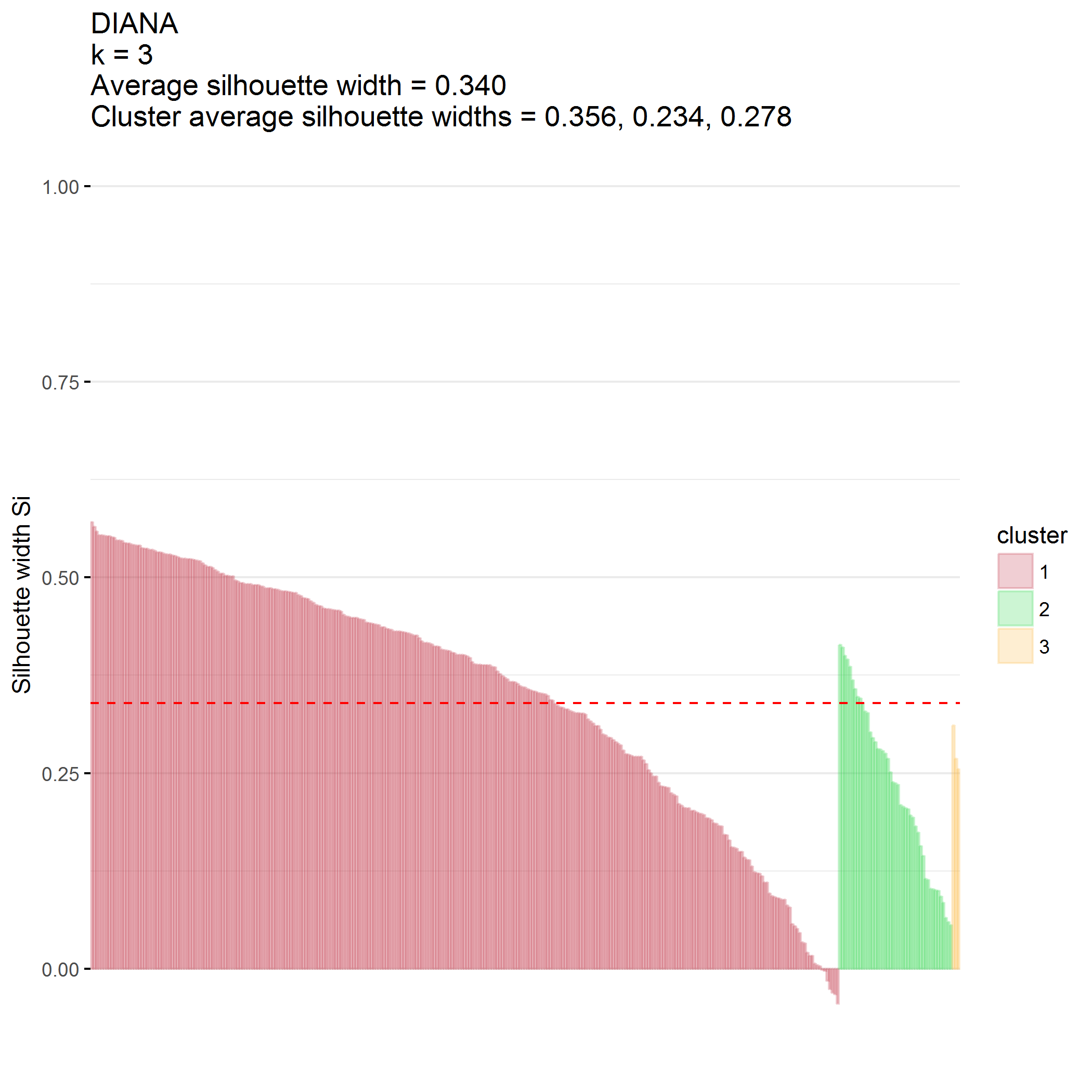
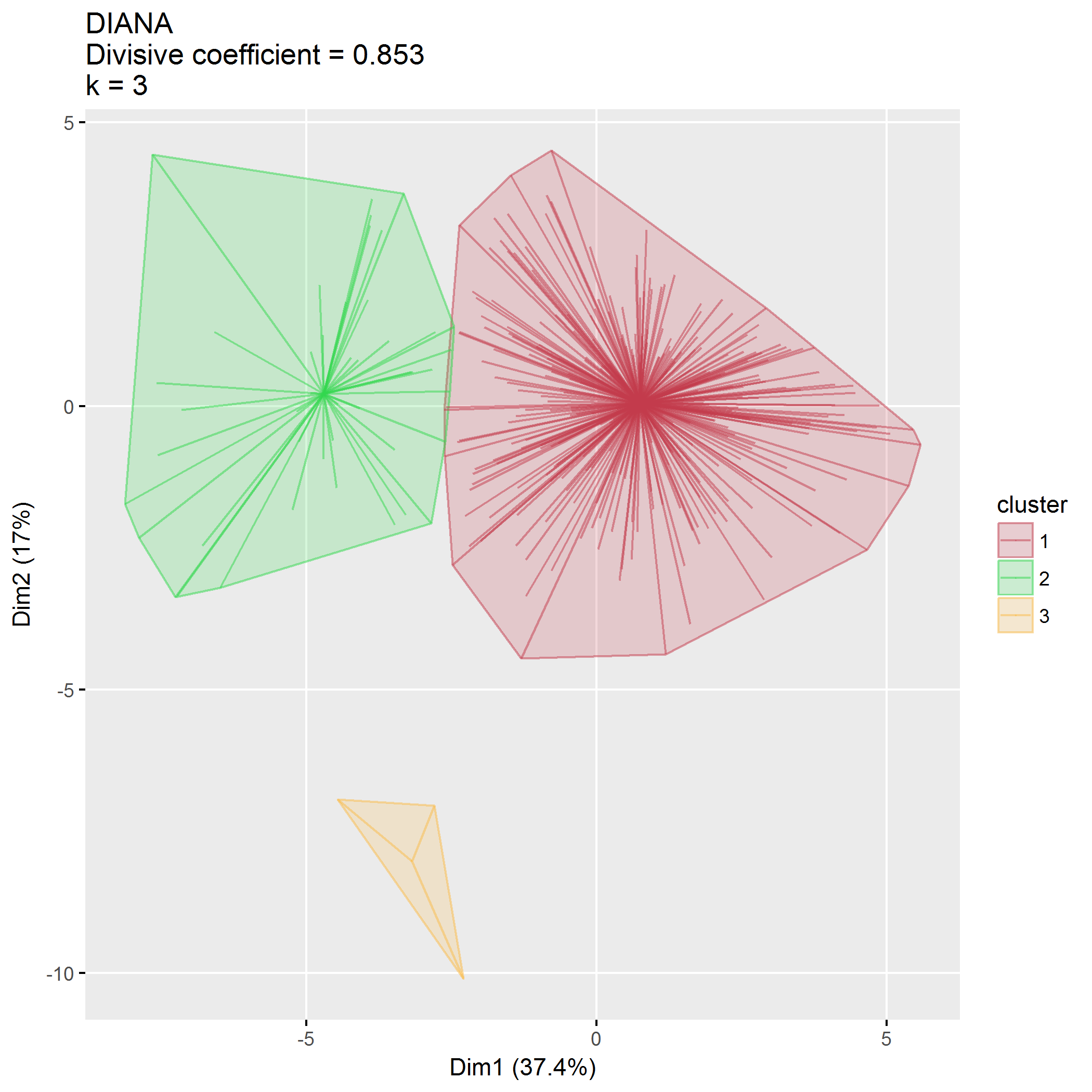
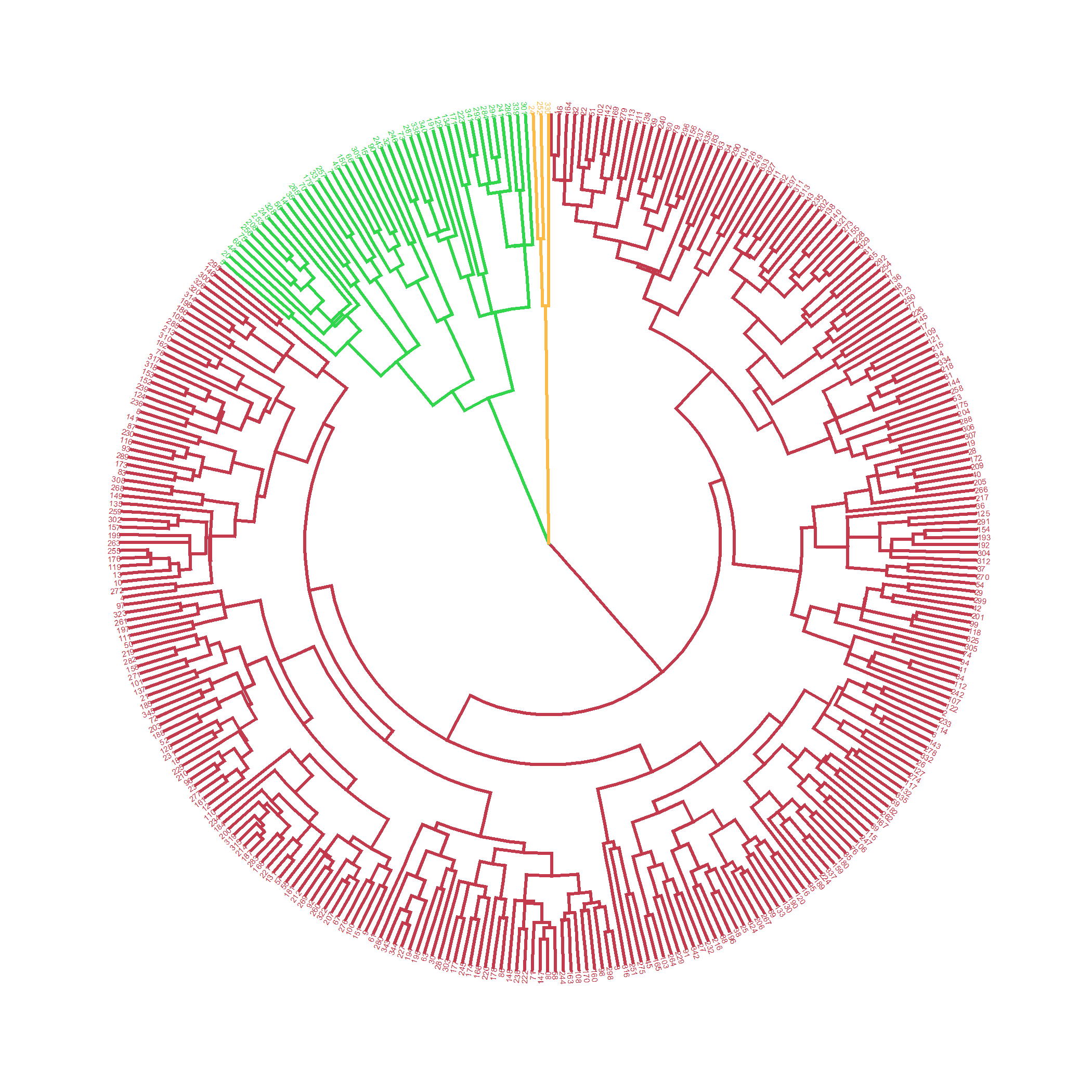
## .  
## 1 2   
## 297 48



### clusters

## cluster size ave.sil.width  
## 1 1 297 0.36  
## 2 2 45 0.23  
## 3 3 3 0.28

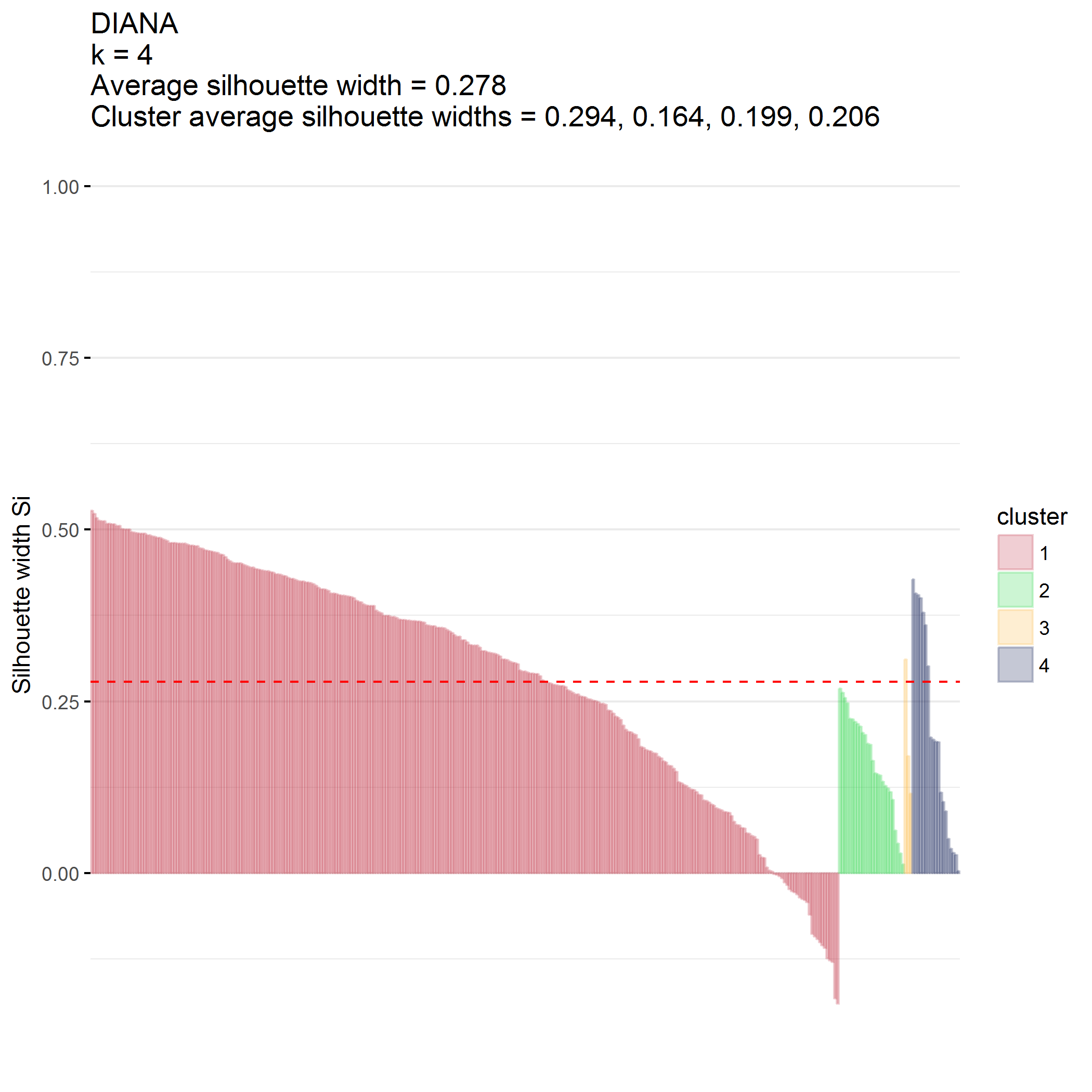
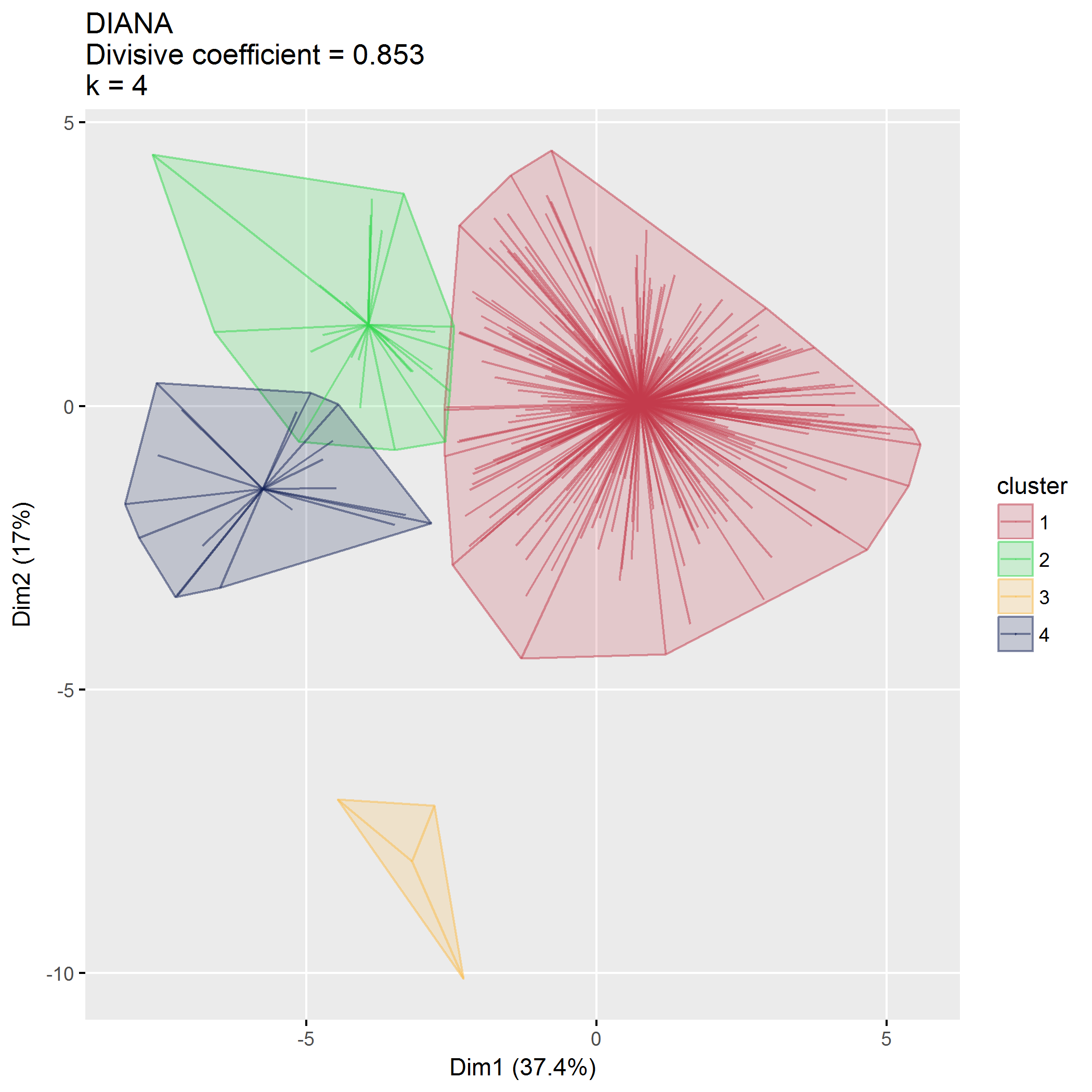
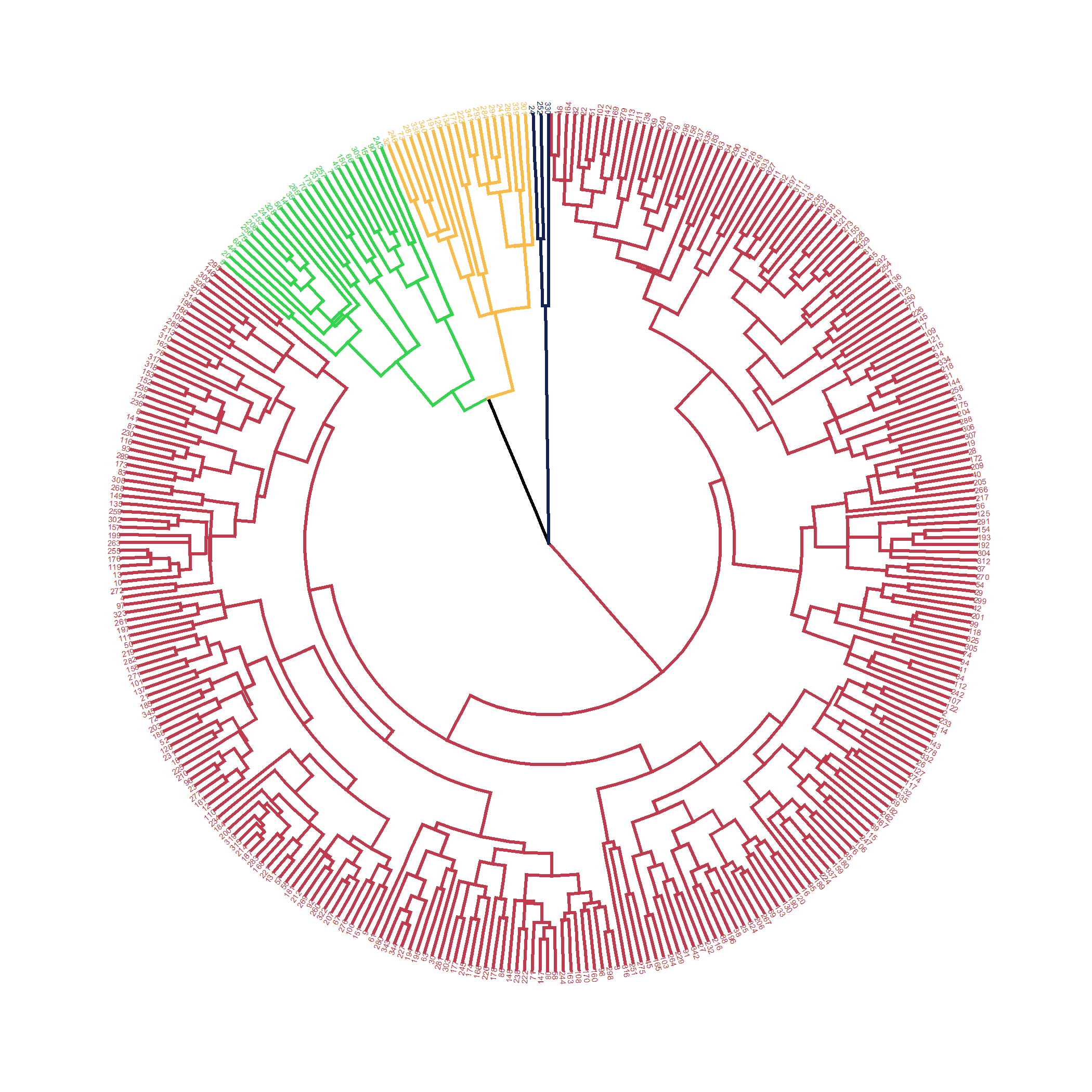
## .  
## 1 2 3   
## 297 45 3



### clusters

## cluster size ave.sil.width  
## 1 1 297 0.29  
## 2 2 26 0.16  
## 3 3 3 0.20  
## 4 4 19 0.21

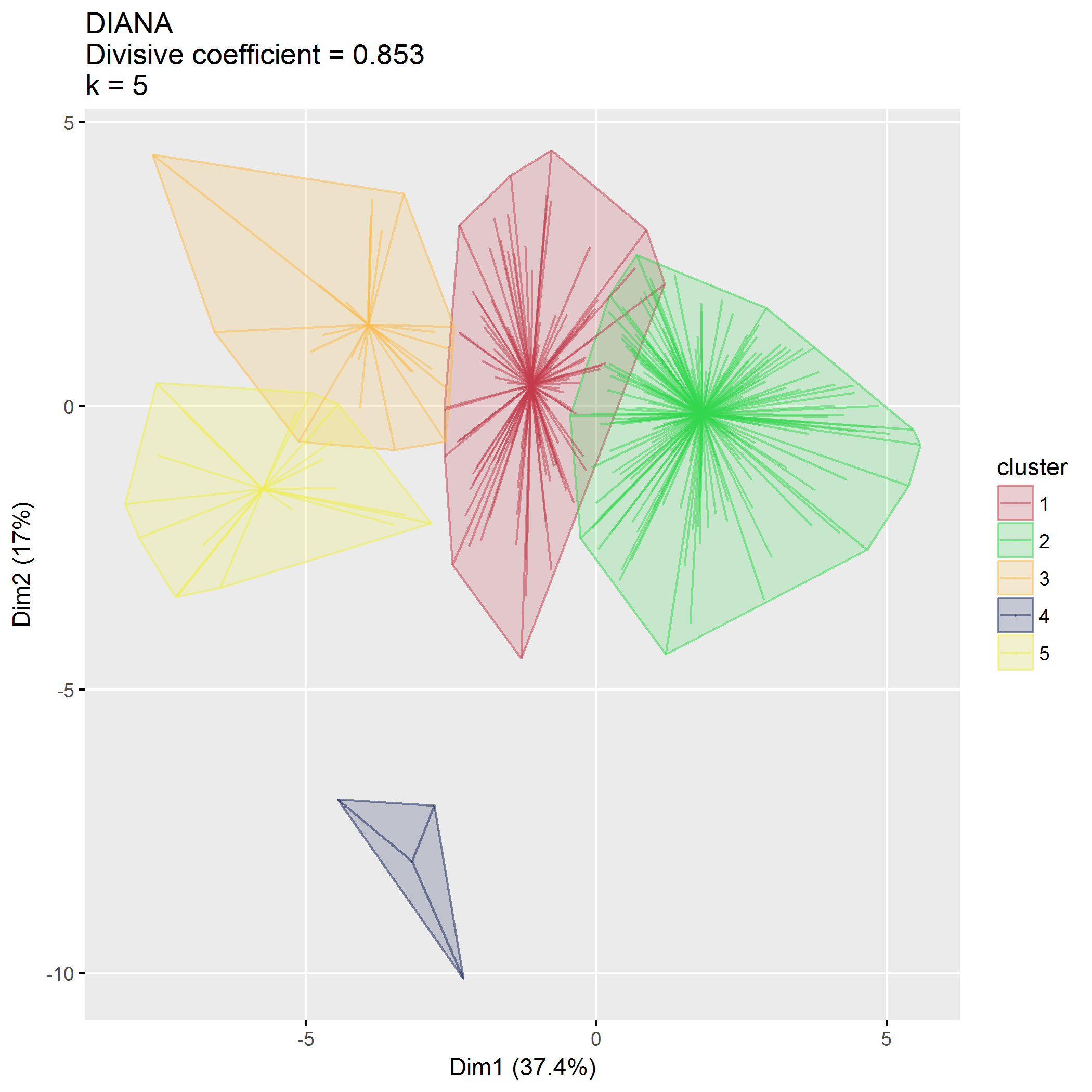
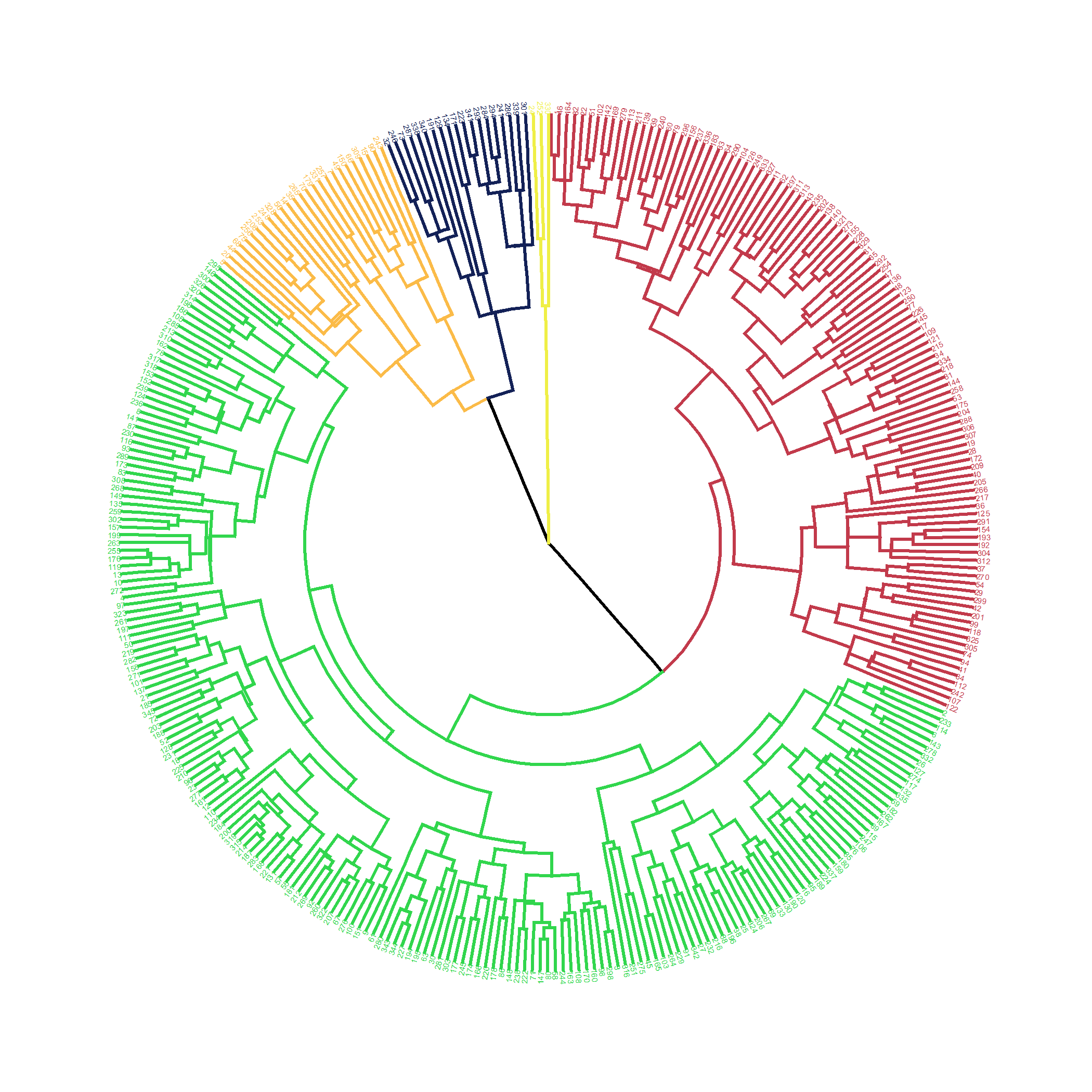
## .  
## 1 2 3 4   
## 297 26 3 19



### clusters

## cluster size ave.sil.width  
## 1 1 108 0.09  
## 2 2 189 0.22  
## 3 3 26 0.08  
## 4 4 3 0.20  
## 5 5 19 0.20

## .  
## 1 2 3 4 5   
## 108 189 26 3 19



## Compare

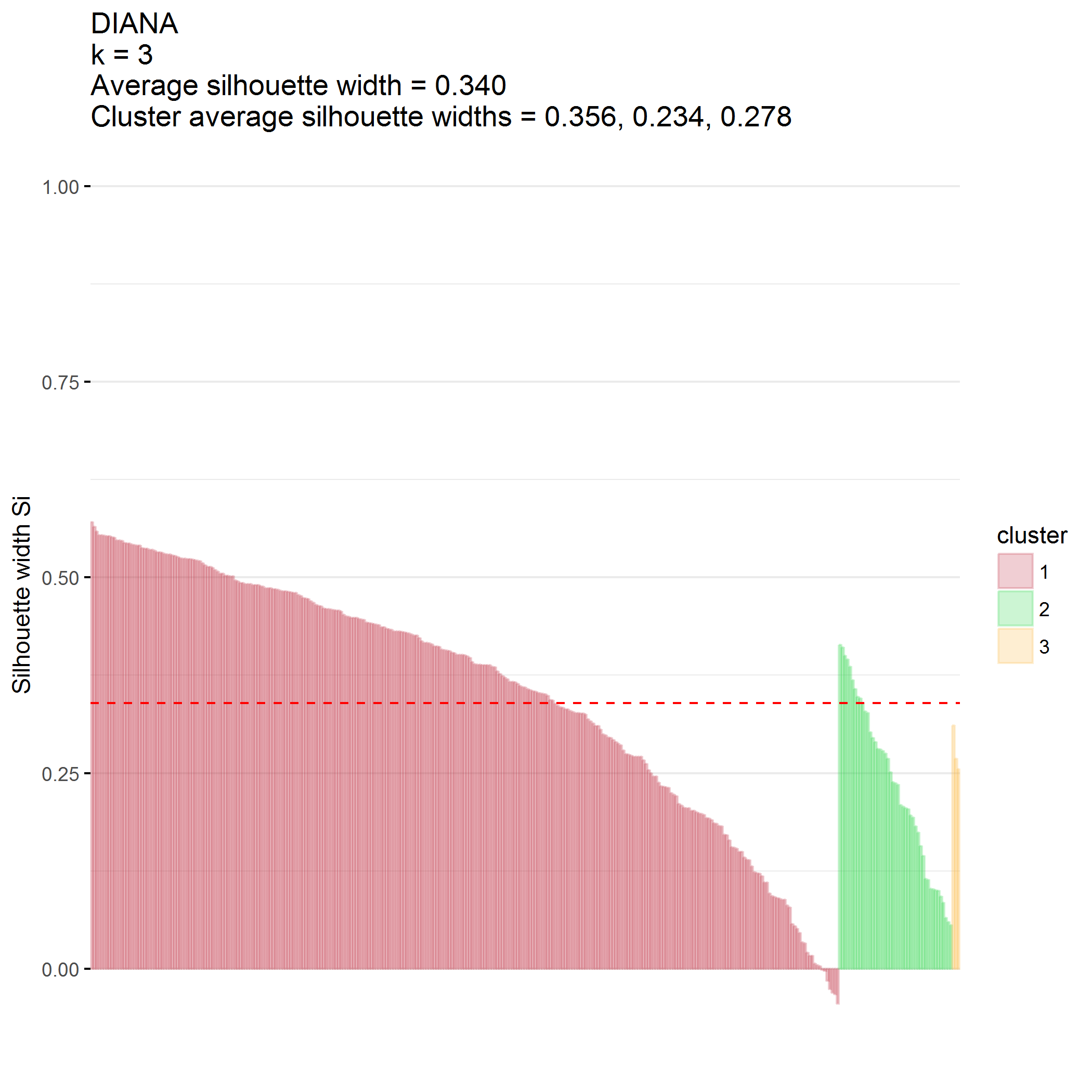
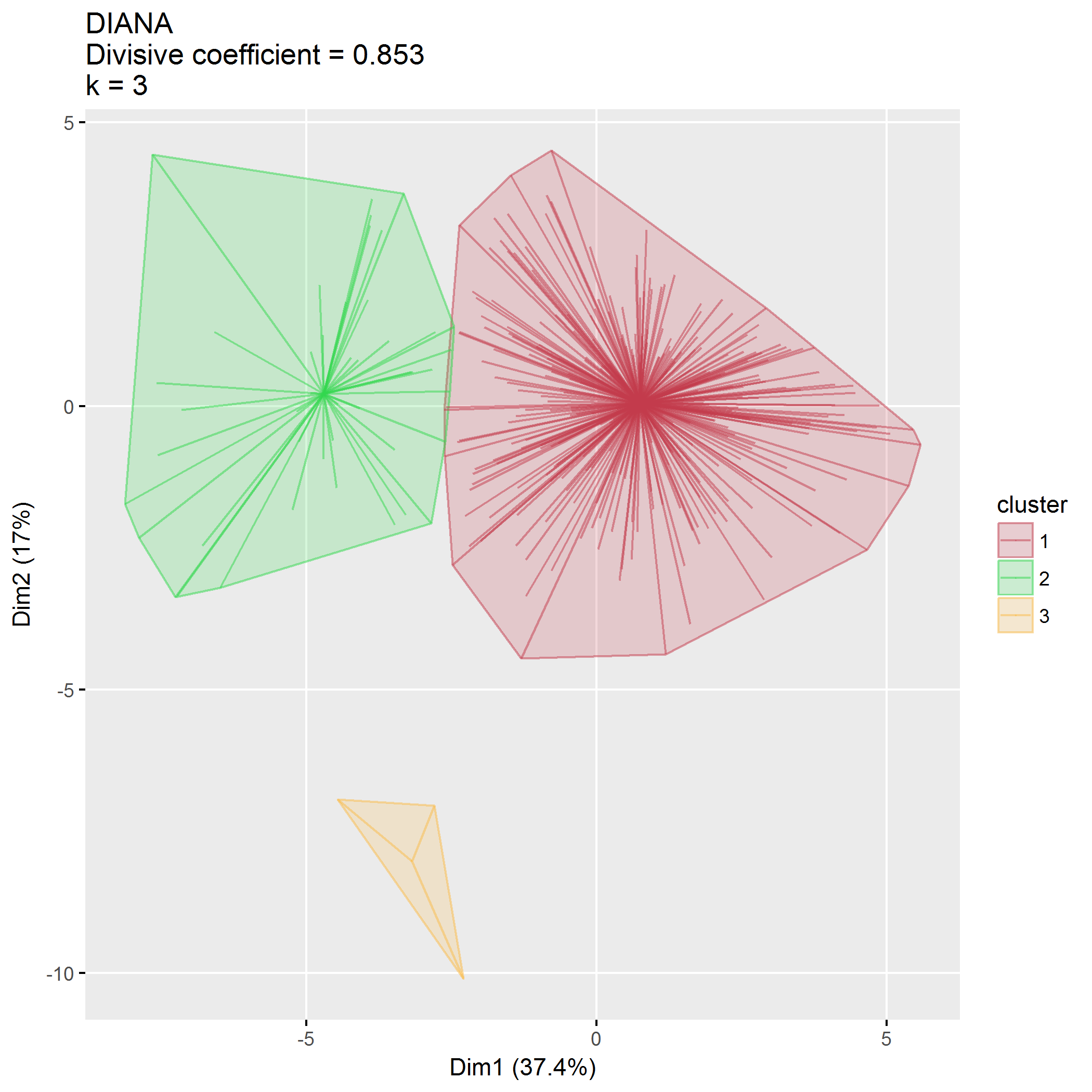
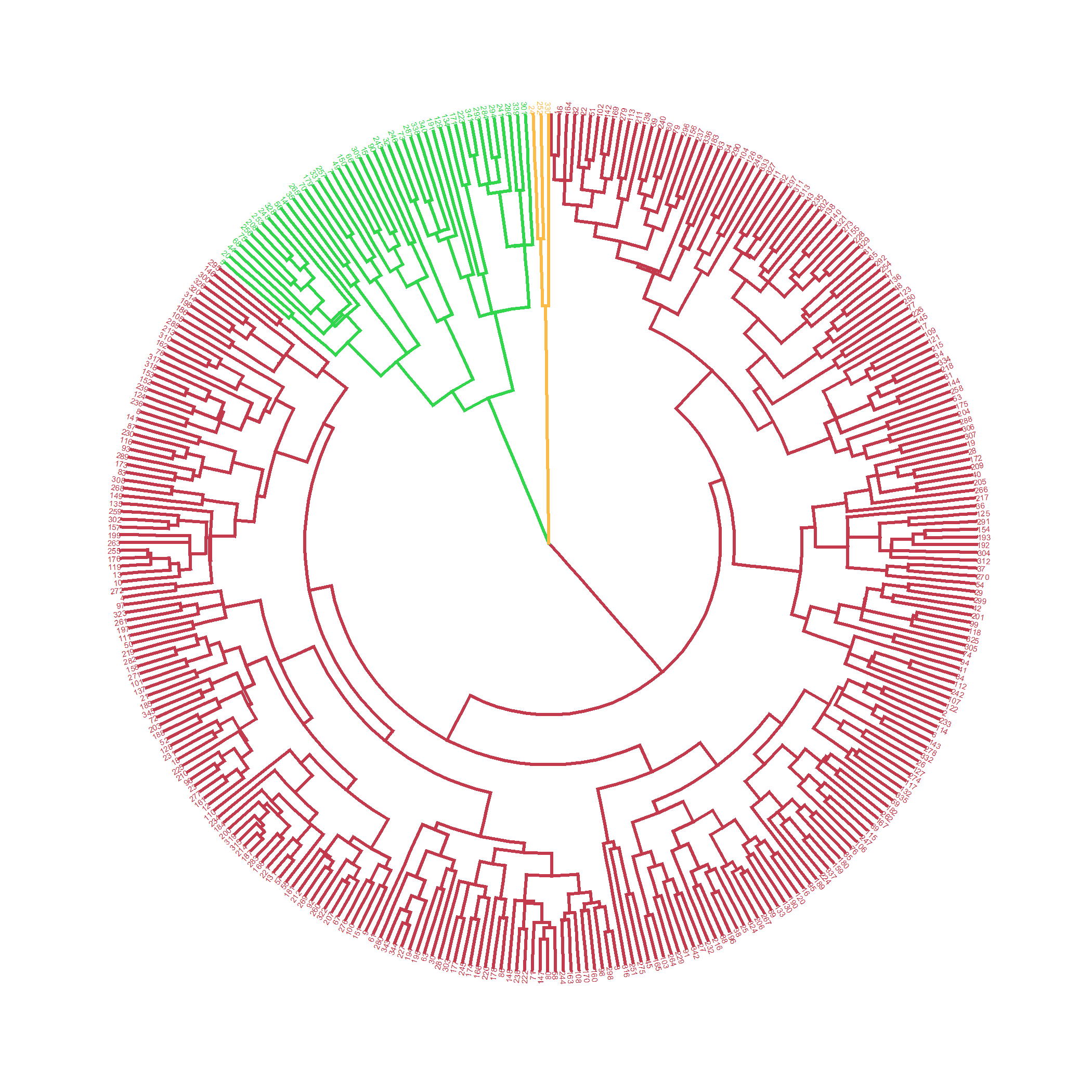
Comparison between agnes and diana doesn’t give much insight.

*Do not evaluate*

## Final cluster identification

Use DIANA,

## .  
## 1 2 3   
## 297 45 3



<https://uc-r.github.io/hc_clustering> <http://www.sthda.com/english/wiki/factoextra-r-package-easy-multivariate-data-analyses-and-elegant-visualization>

##   
## To cite package 'factoextra' in publications use:  
##   
## Alboukadel Kassambara and Fabian Mundt (2017). factoextra:  
## Extract and Visualize the Results of Multivariate Data Analyses.  
## R package version 1.0.5.  
## https://CRAN.R-project.org/package=factoextra  
##   
## A BibTeX entry for LaTeX users is  
##   
## @Manual{,  
## title = {factoextra: Extract and Visualize the Results of Multivariate Data Analyses},  
## author = {Alboukadel Kassambara and Fabian Mundt},  
## year = {2017},  
## note = {R package version 1.0.5},  
## url = {https://CRAN.R-project.org/package=factoextra},  
## }

## [1] 345 63

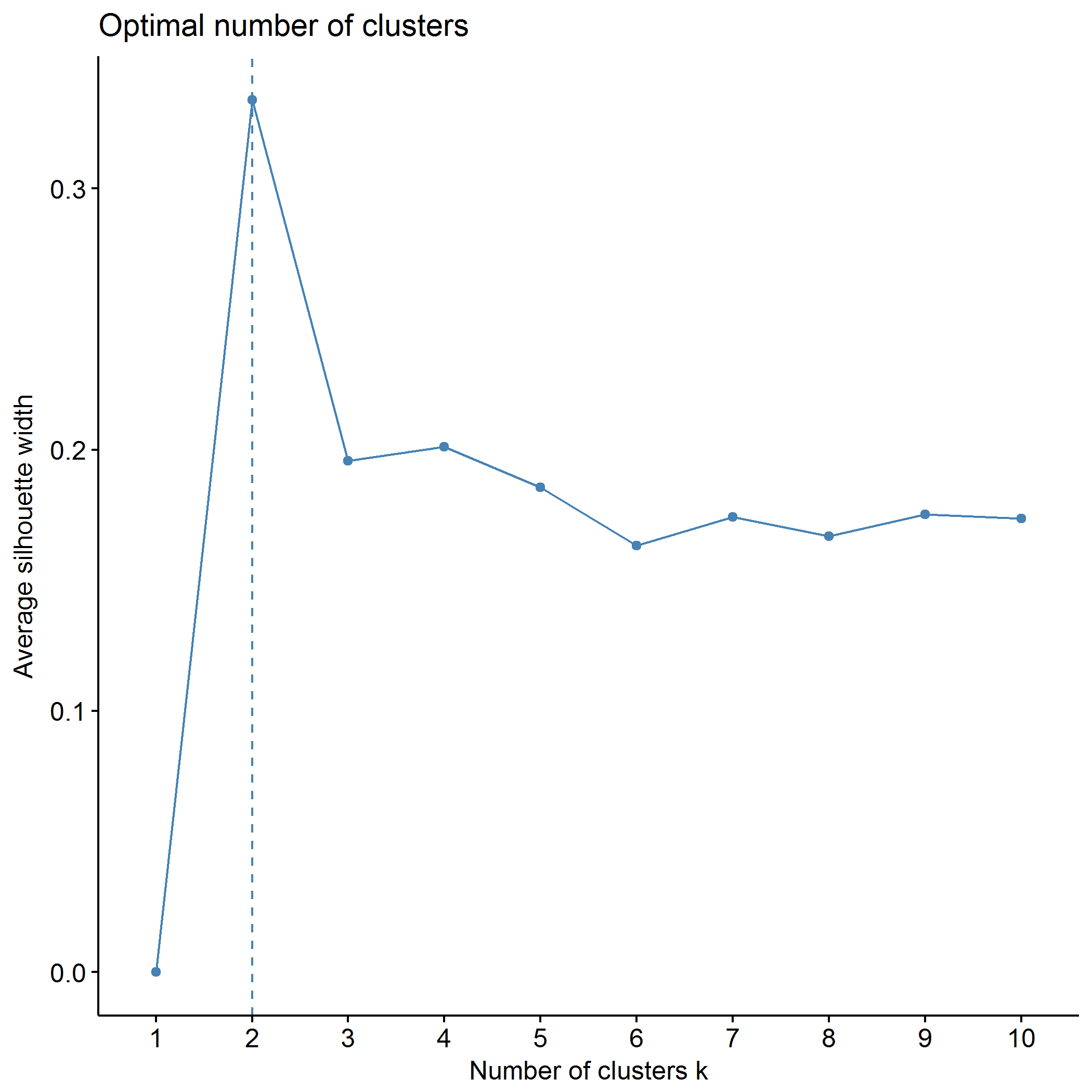
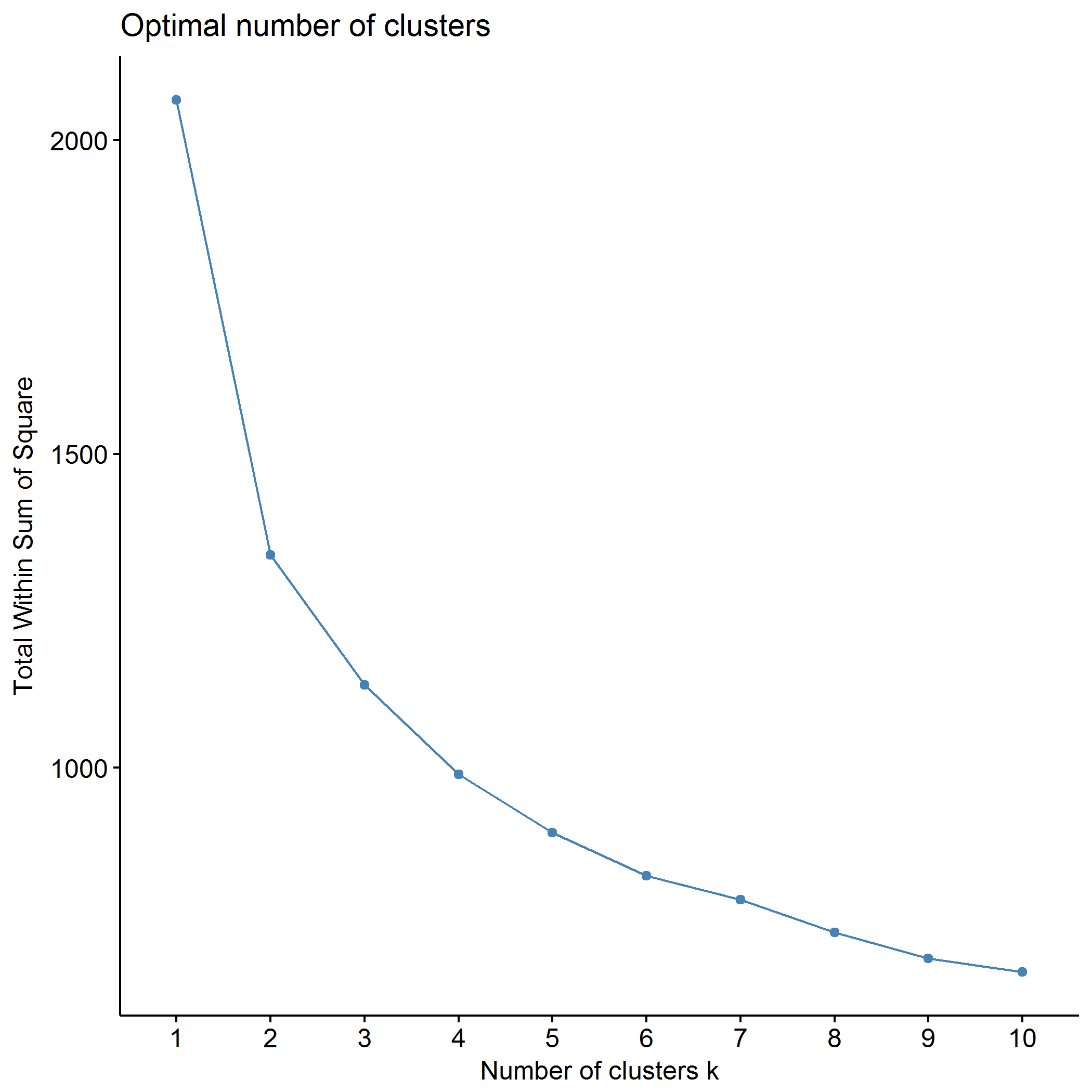
## [1] 345 54

## [1] 345 6

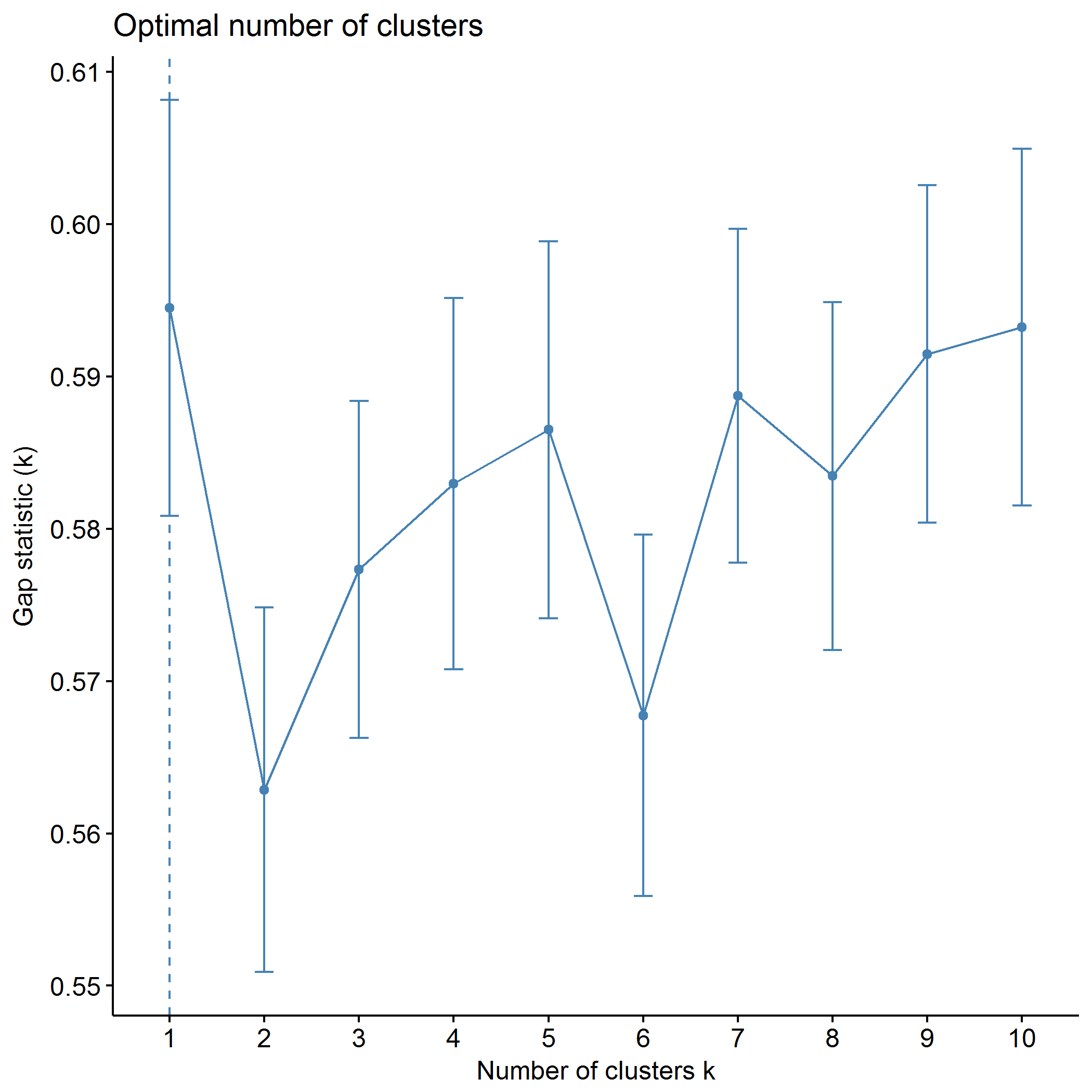
## [1] 345 6

## NULL

## K-means clustering

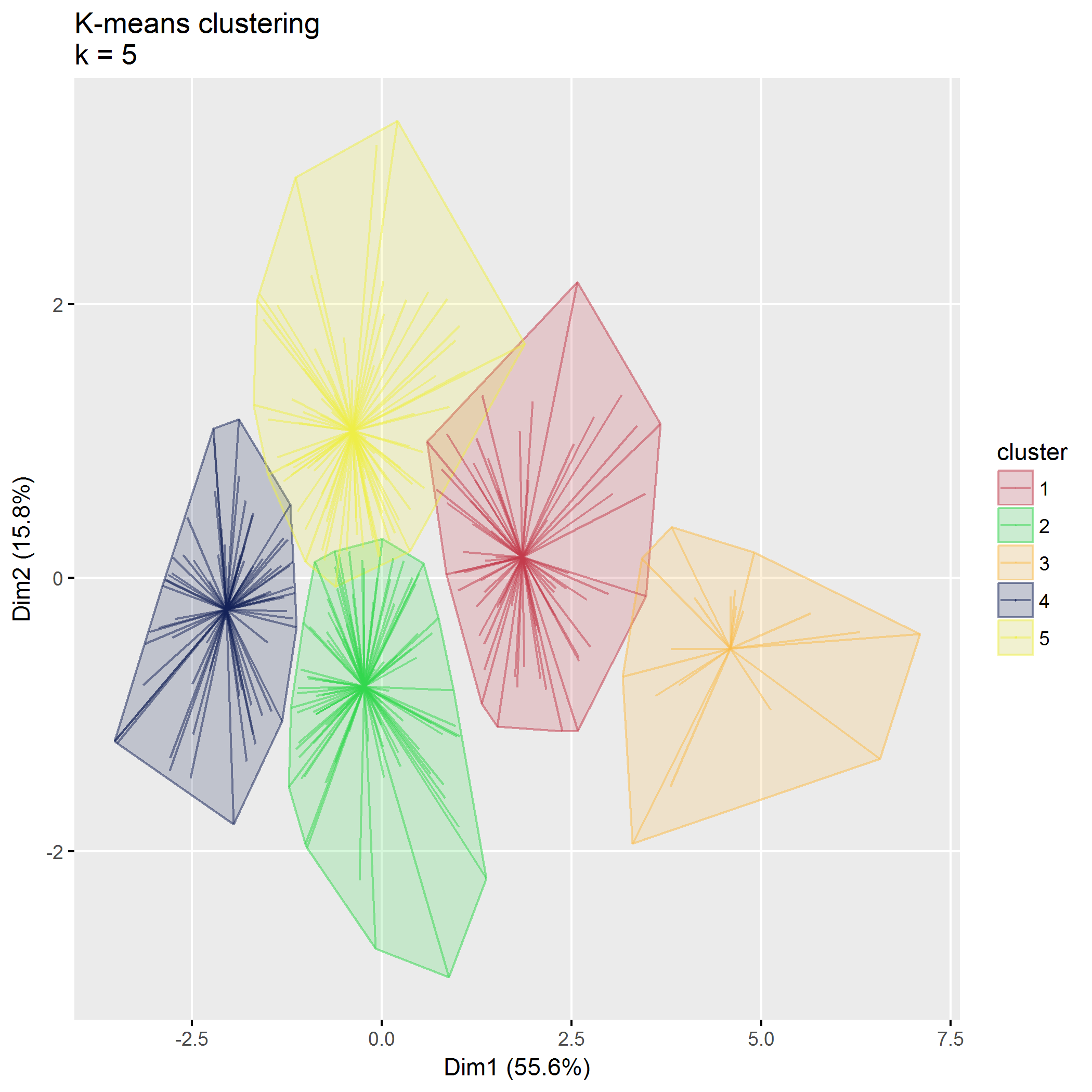


## Clustering k = 1,2,..., K.max (= 10): .. done  
## Bootstrapping, b = 1,2,..., B (= 500) [one "." per sample]:  
## .................................................. 50   
## .................................................. 100   
## .................................................. 150   
## .................................................. 200   
## .................................................. 250   
## .................................................. 300   
## .................................................. 350   
## .................................................. 400   
## .................................................. 450   
## .................................................. 500



plot of chunk PCB\_kmeans

## .  
## 1 2 3 4 5   
## 67 96 19 76 87



plot of chunk PCB\_kmeans

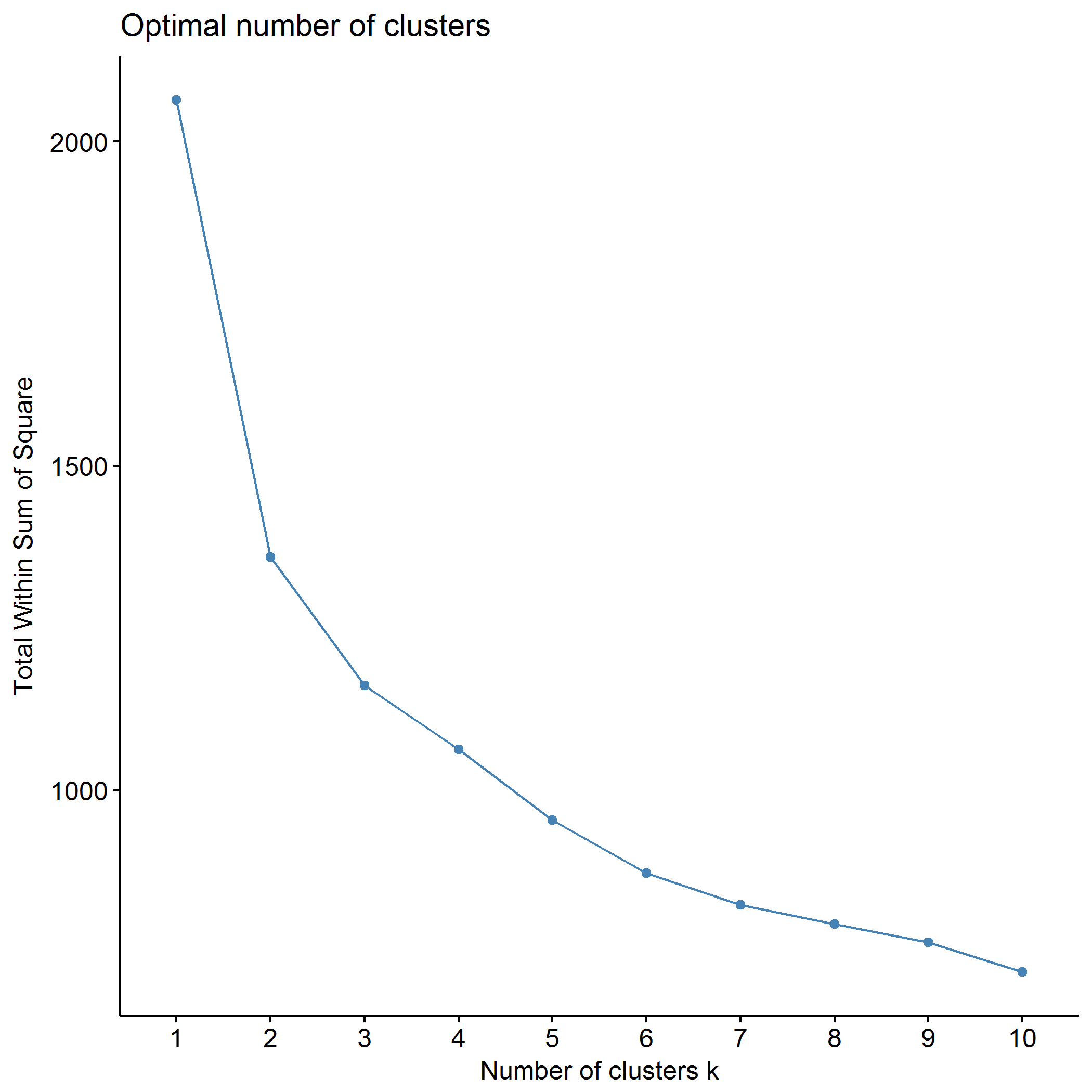
## Within cluster sum of squares, cluster 1: 232.49  
## Within cluster sum of squares, cluster 2: 231.94  
## Within cluster sum of squares, cluster 3: 72.29  
## Within cluster sum of squares, cluster 4: 122.39  
## Within cluster sum of squares, cluster 5: 237.95

## Between SS / Total SS: 1166.94 / 2064.00 = 56.54%

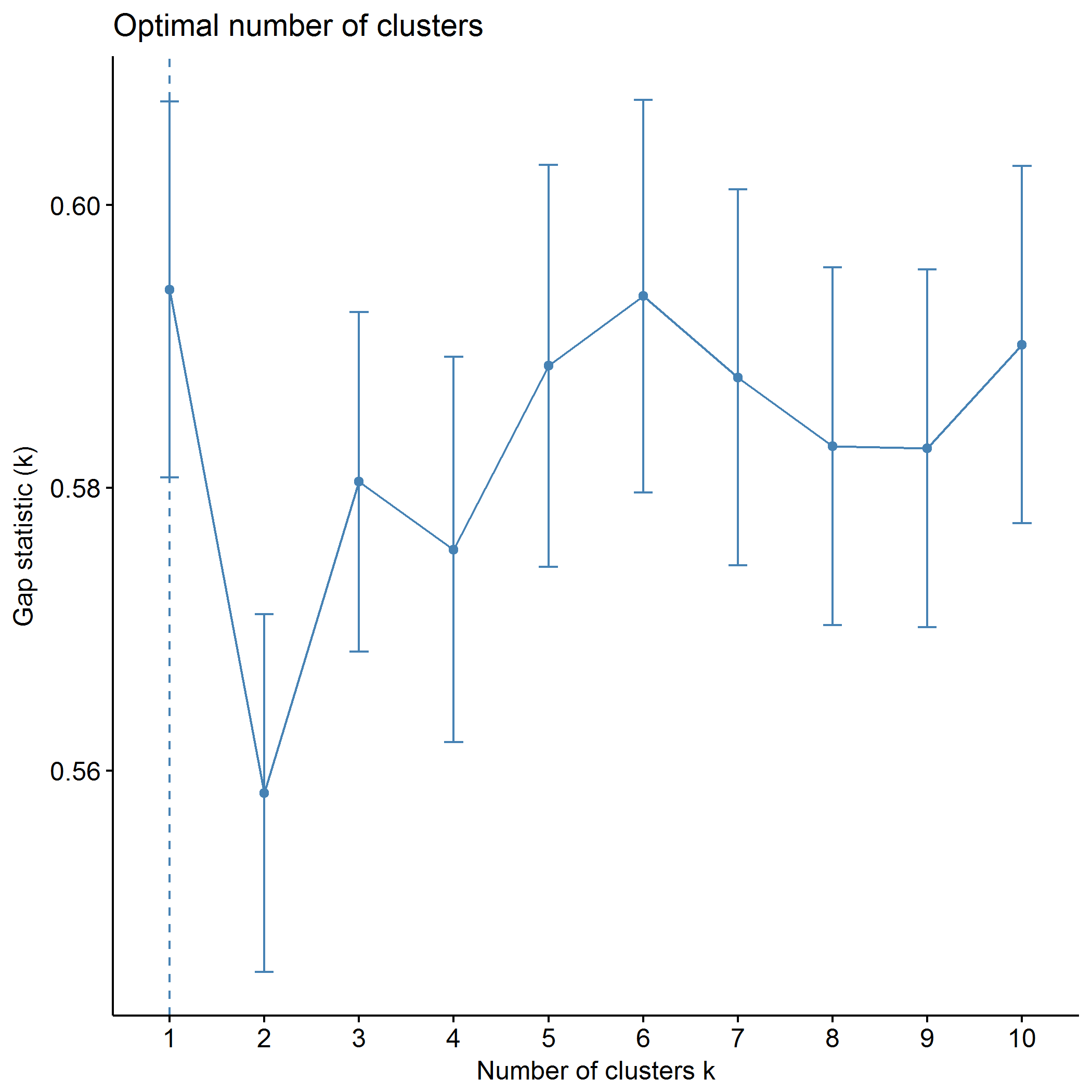
## Total within SS: 897.06

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| PCB1\_CondEmot | -1.01 | -0.10 | -1.96 | 0.90 | 0.53 |
| PCB1\_DevHab | -1.00 | -0.14 | -2.02 | 0.96 | 0.53 |
| PCB2\_Tot | -0.66 | 0.02 | -2.36 | 0.86 | 0.26 |
| PCB3\_PCPonly | -0.38 | -0.14 | -2.19 | 0.62 | 0.38 |
| PCB3\_Person | -0.62 | 0.48 | -1.42 | 0.80 | -0.44 |
| PCB3\_Resource | -0.78 | 0.66 | -1.19 | 0.89 | -0.64 |

## Partitioning around medoids (PAM)



## Clustering k = 1,2,..., K.max (= 10): .. done  
## Bootstrapping, b = 1,2,..., B (= 500) [one "." per sample]:  
## .................................................. 50   
## .................................................. 100   
## .................................................. 150   
## .................................................. 200   
## .................................................. 250   
## .................................................. 300   
## .................................................. 350   
## .................................................. 400   
## .................................................. 450   
## .................................................. 500



plot of chunk PCB\_pam

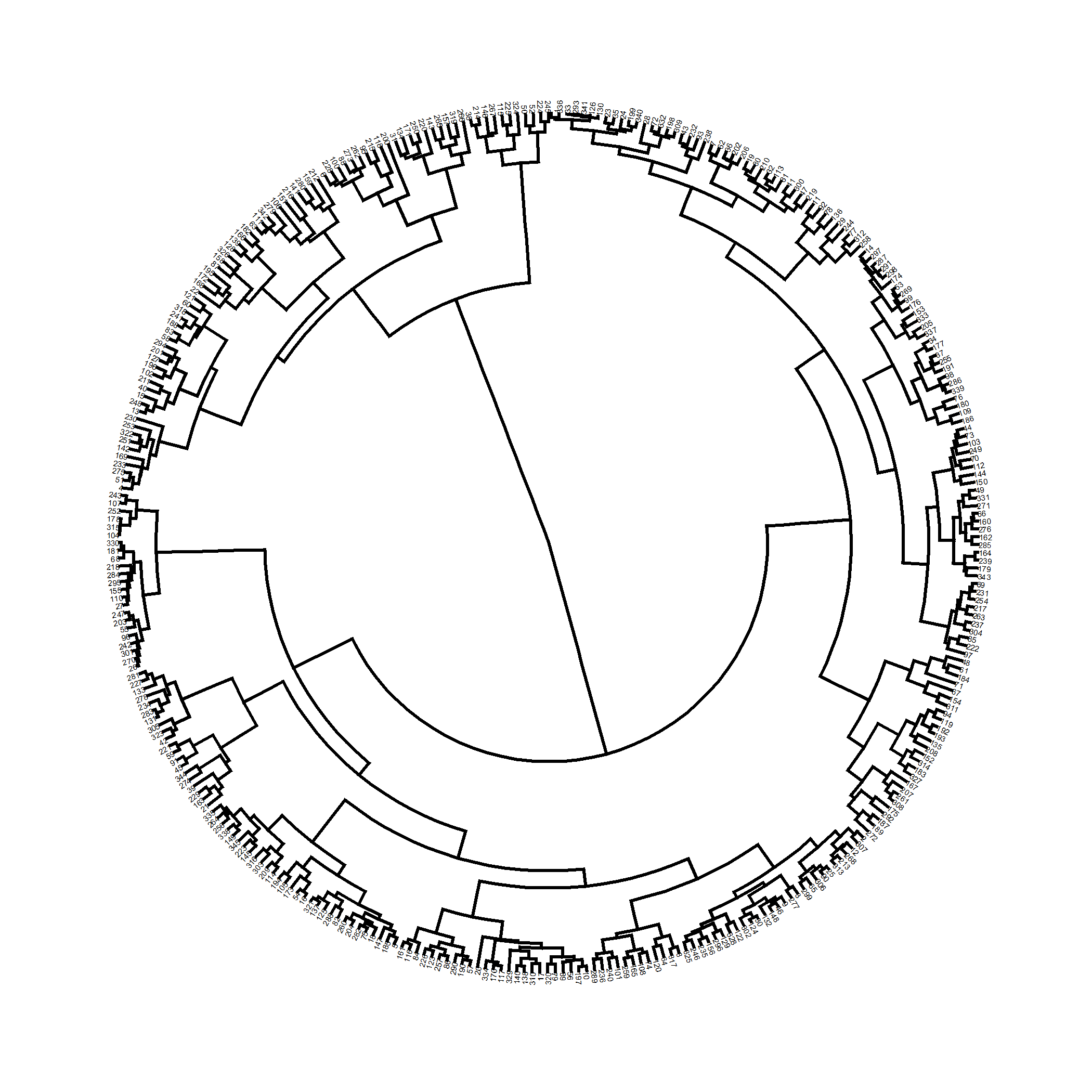
## .  
## 1 2   
## 193 152



plot of chunk PCB\_pam

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| size | max\_diss | av\_diss | diameter | separation |
| 193 | 7.04 | 3.31 | 13.49 | 0.51 |
| 152 | 14.27 | 4.48 | 18.18 | 0.51 |
|  | 156 | 4 |
| PCB1\_CondEmot | 0.59 | -0.57 |
| PCB1\_DevHab | 0.35 | -0.33 |
| PCB2\_Tot | 0.58 | -0.52 |
| PCB3\_PCPonly | 0.89 | -0.09 |
| PCB3\_Person | 0.22 | -0.60 |
| PCB3\_Resource | 0.69 | -0.81 |

## Agglomerative hierarchical clustering (AGNES)



plot of chunk PCB\_agnes

Correlation between cophenetic distance and the original distance is 0.565.

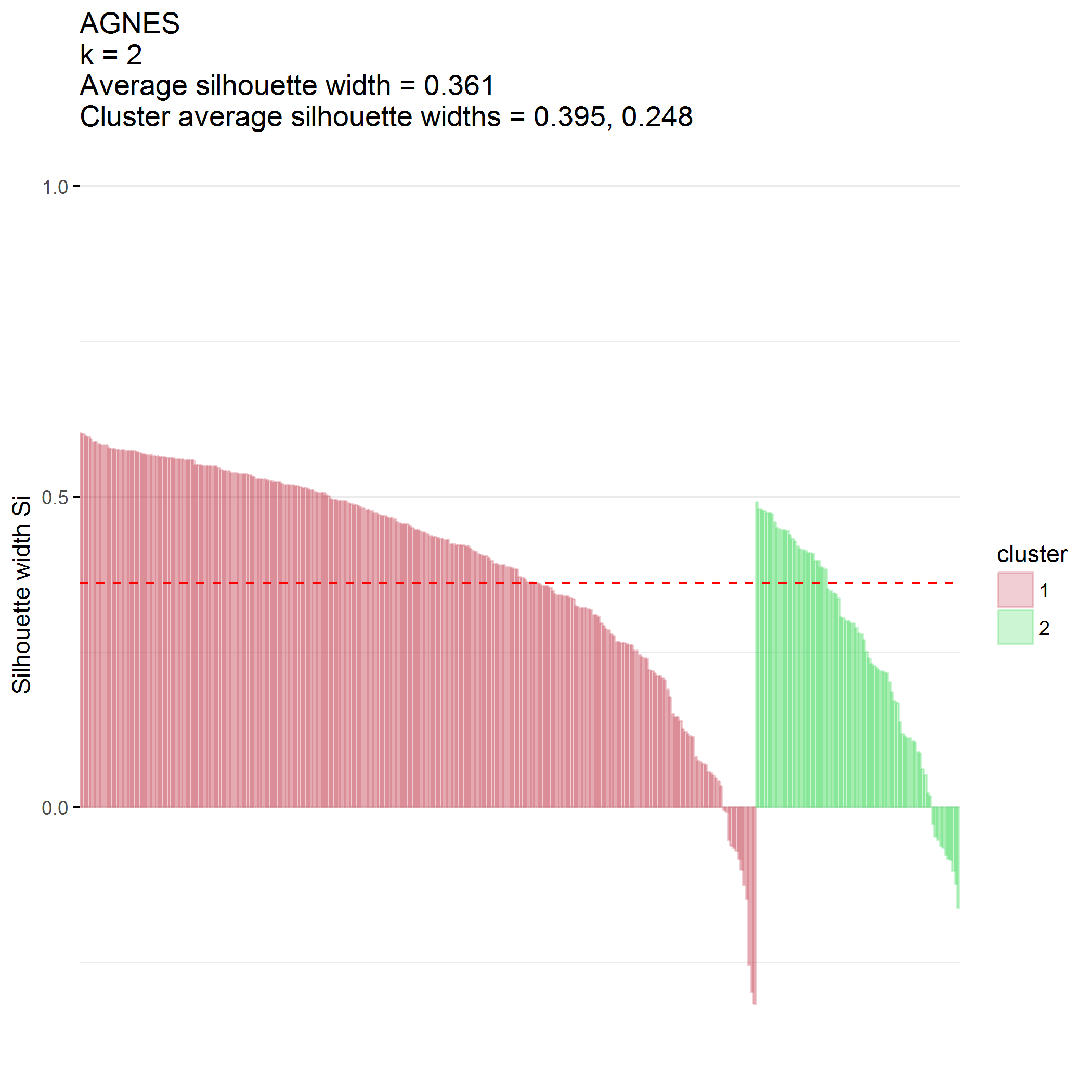
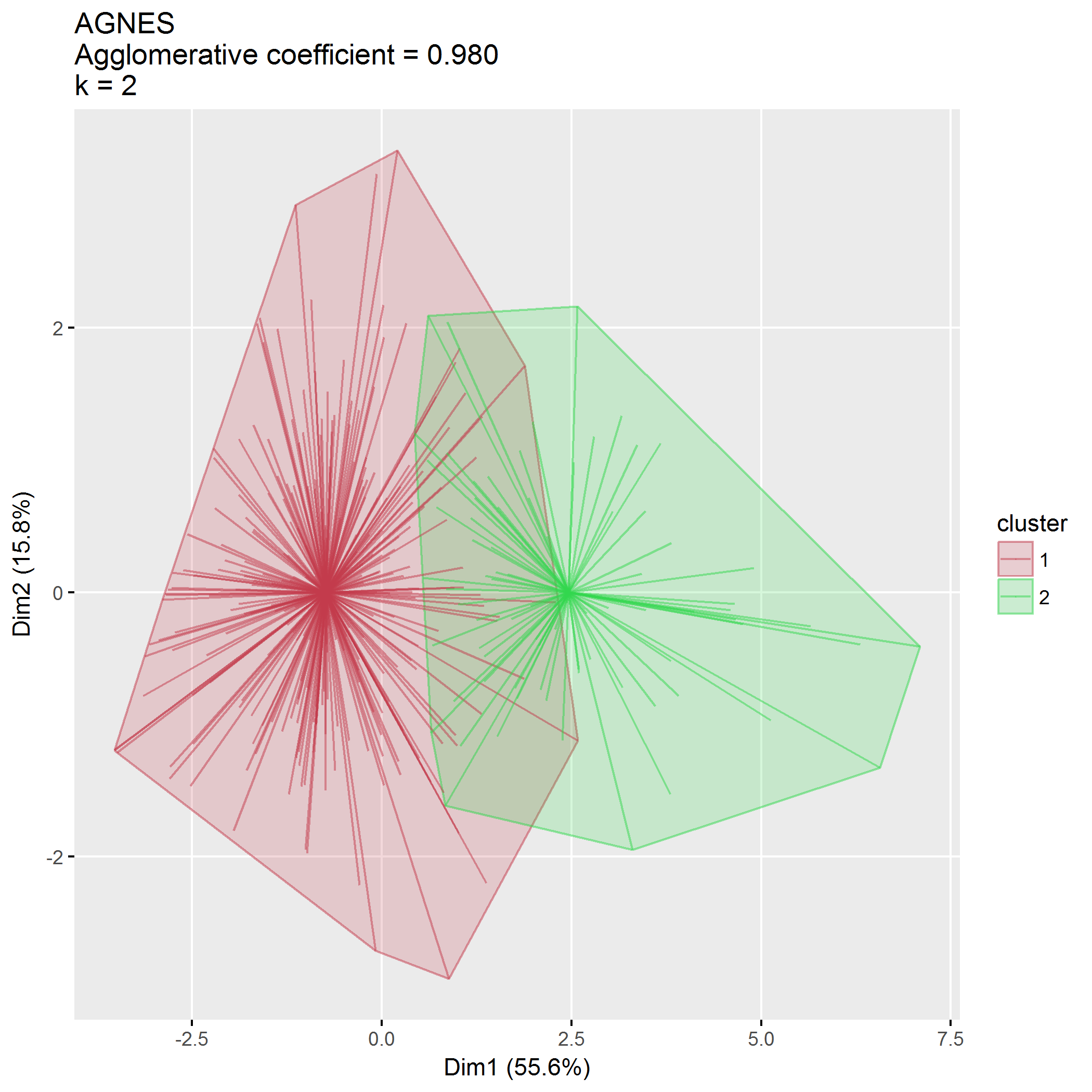
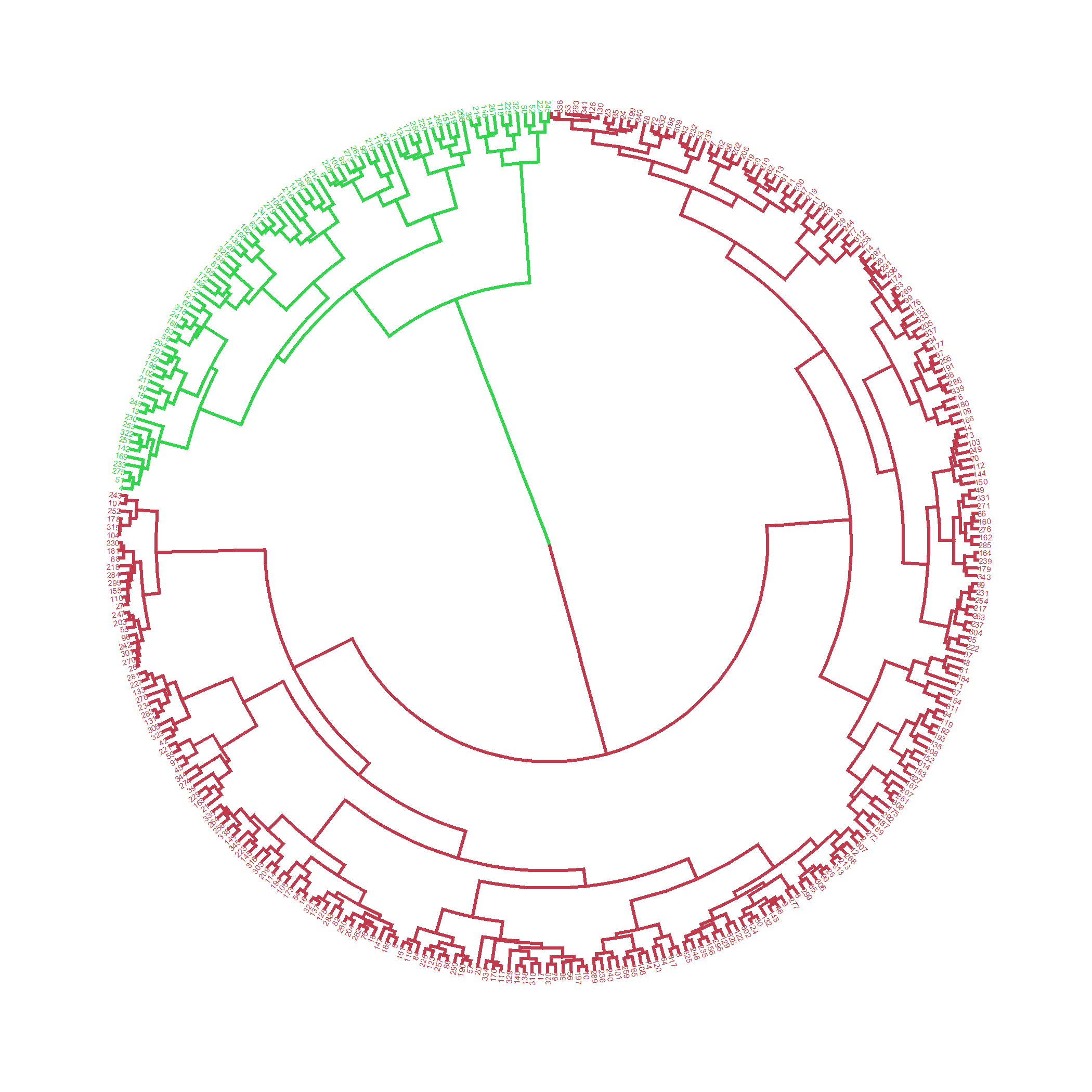
The closer the value of the correlation coefficient is to 1, the more accurately the clustering solution reflects your data. Values above 0.75 are felt to be good.

Agglomerative coeffficient using the Ward method is 0.980.

### clusters

## cluster size ave.sil.width  
## 1 1 265 0.39  
## 2 2 80 0.25

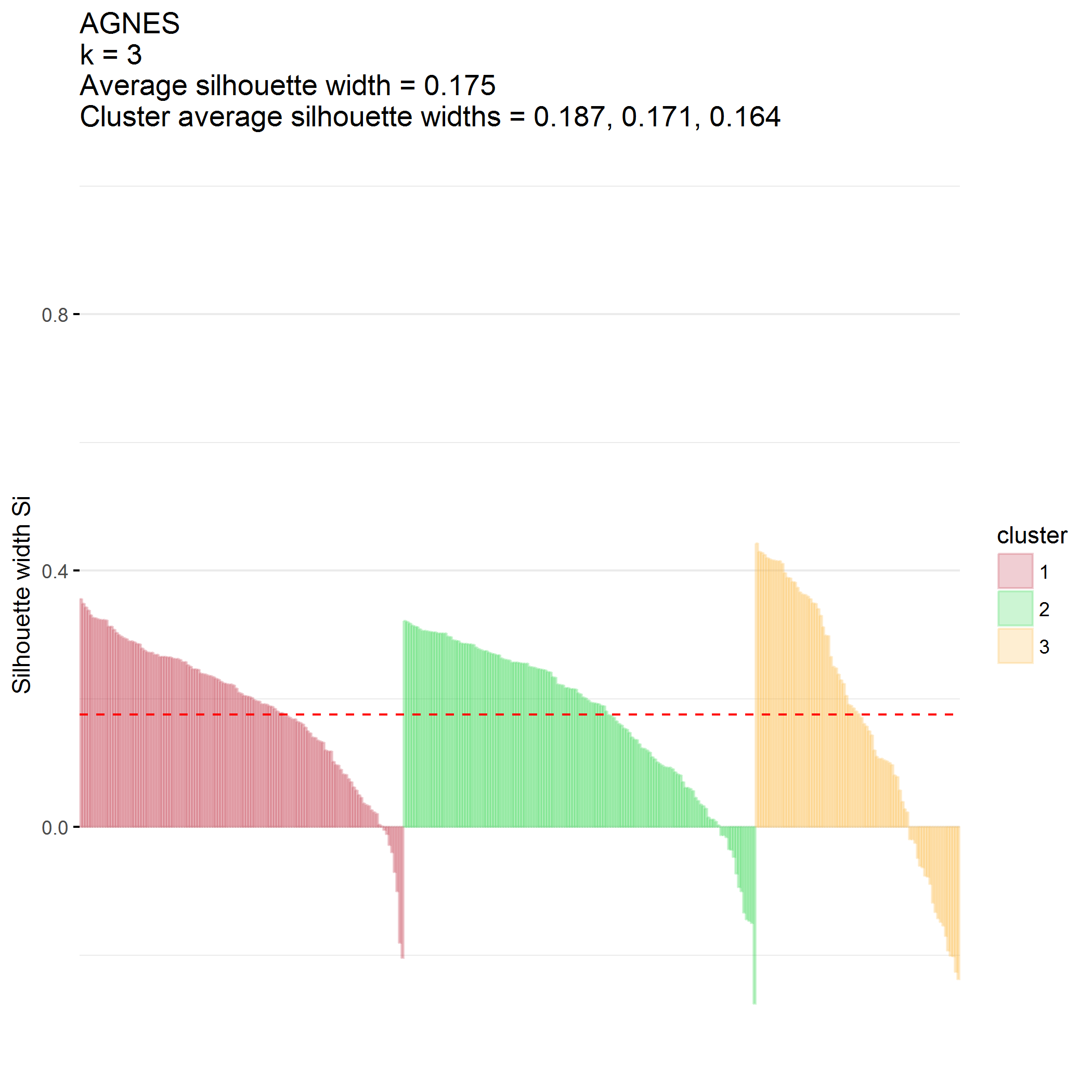
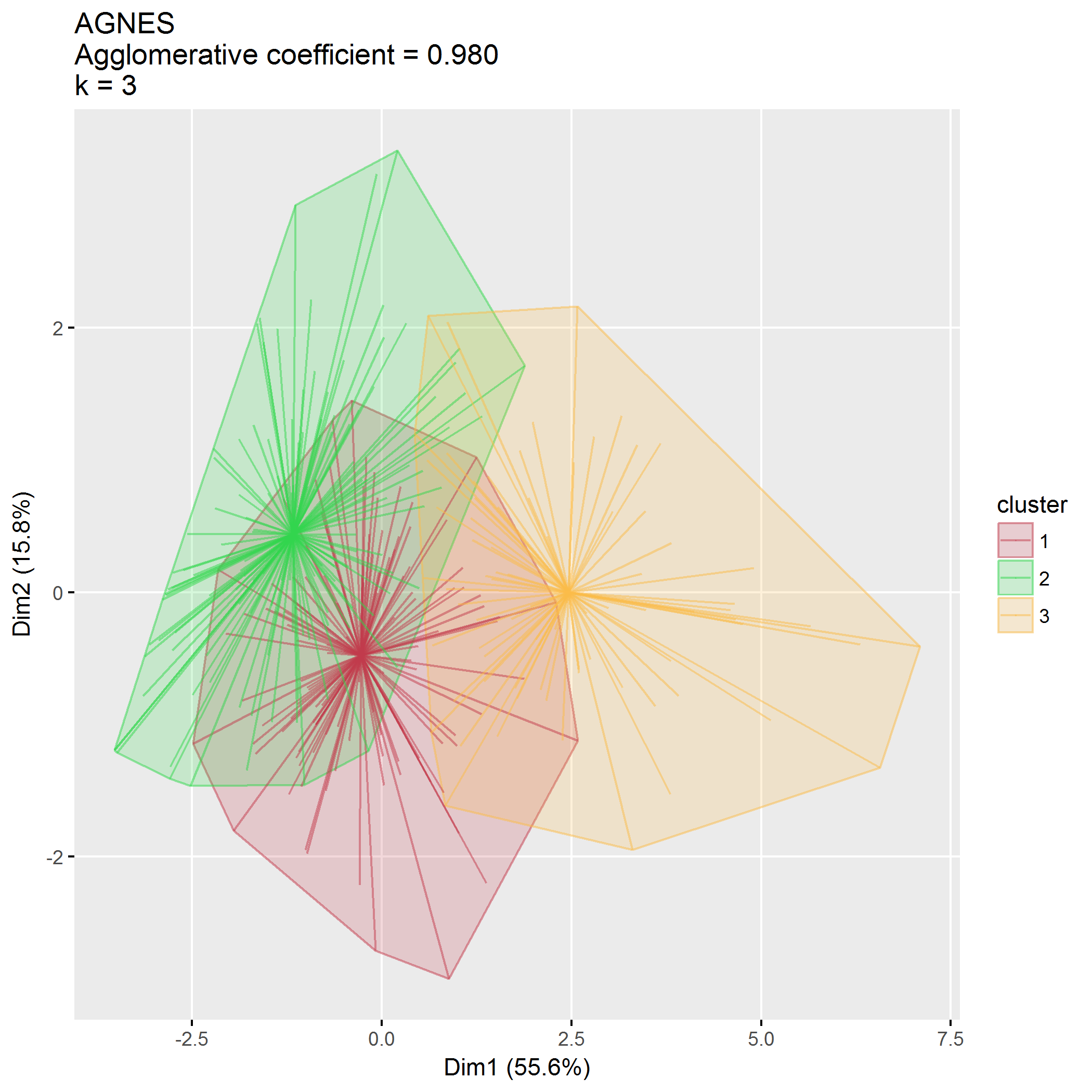
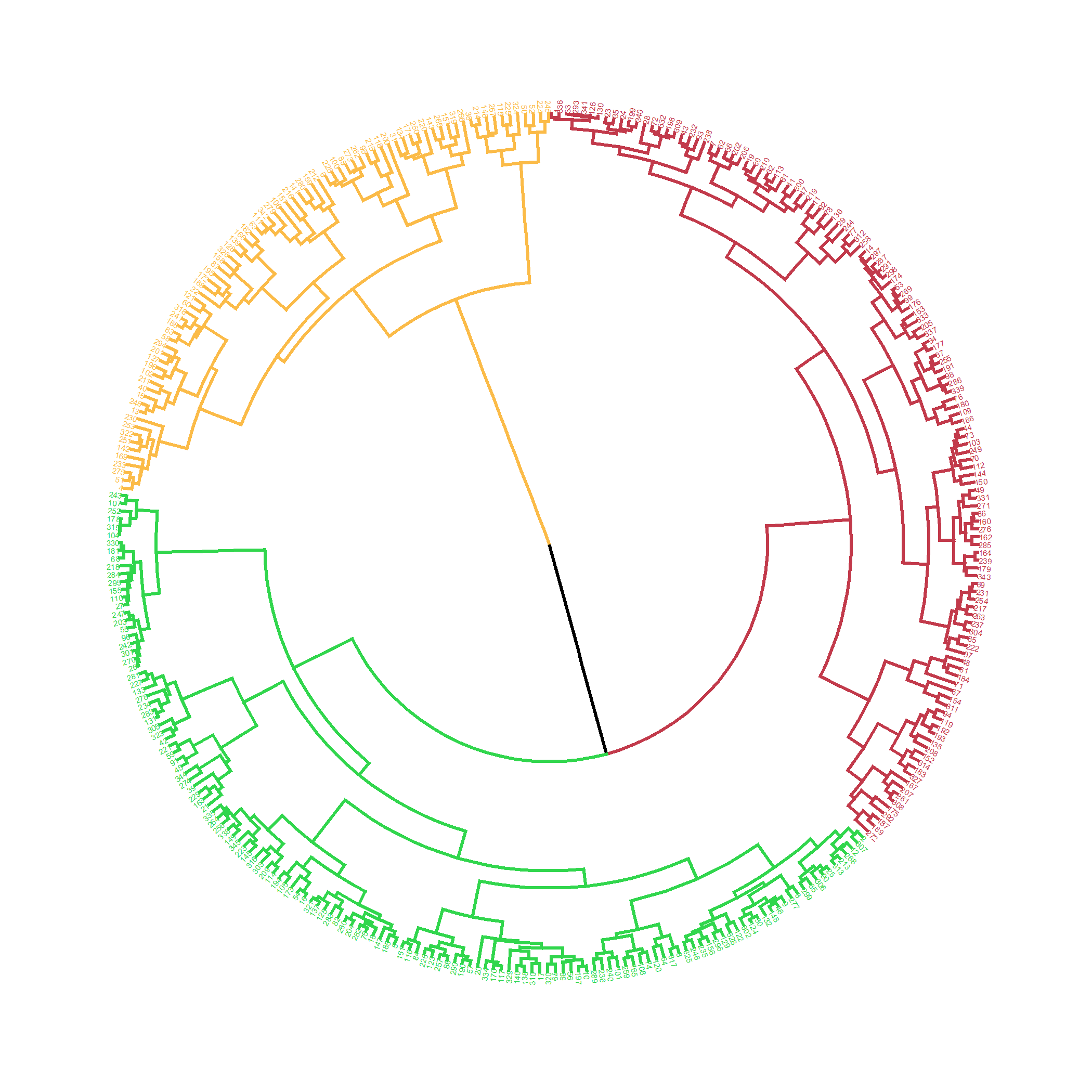
## .  
## 1 2   
## 265 80



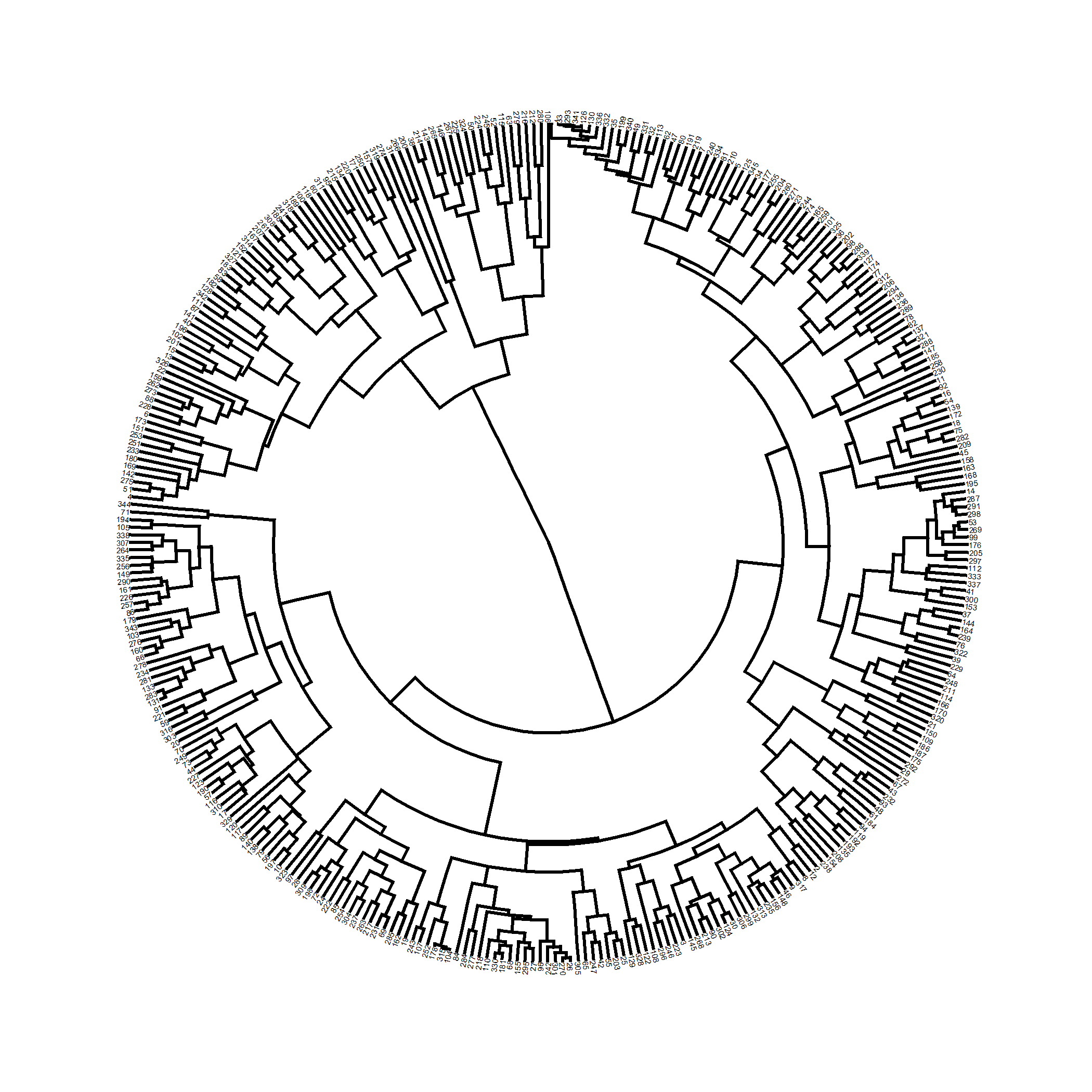
### clusters

## cluster size ave.sil.width  
## 1 1 127 0.19  
## 2 2 138 0.17  
## 3 3 80 0.16

## .  
## 1 2 3   
## 127 138 80



## Divisive hierarchical clustering (DIANA)



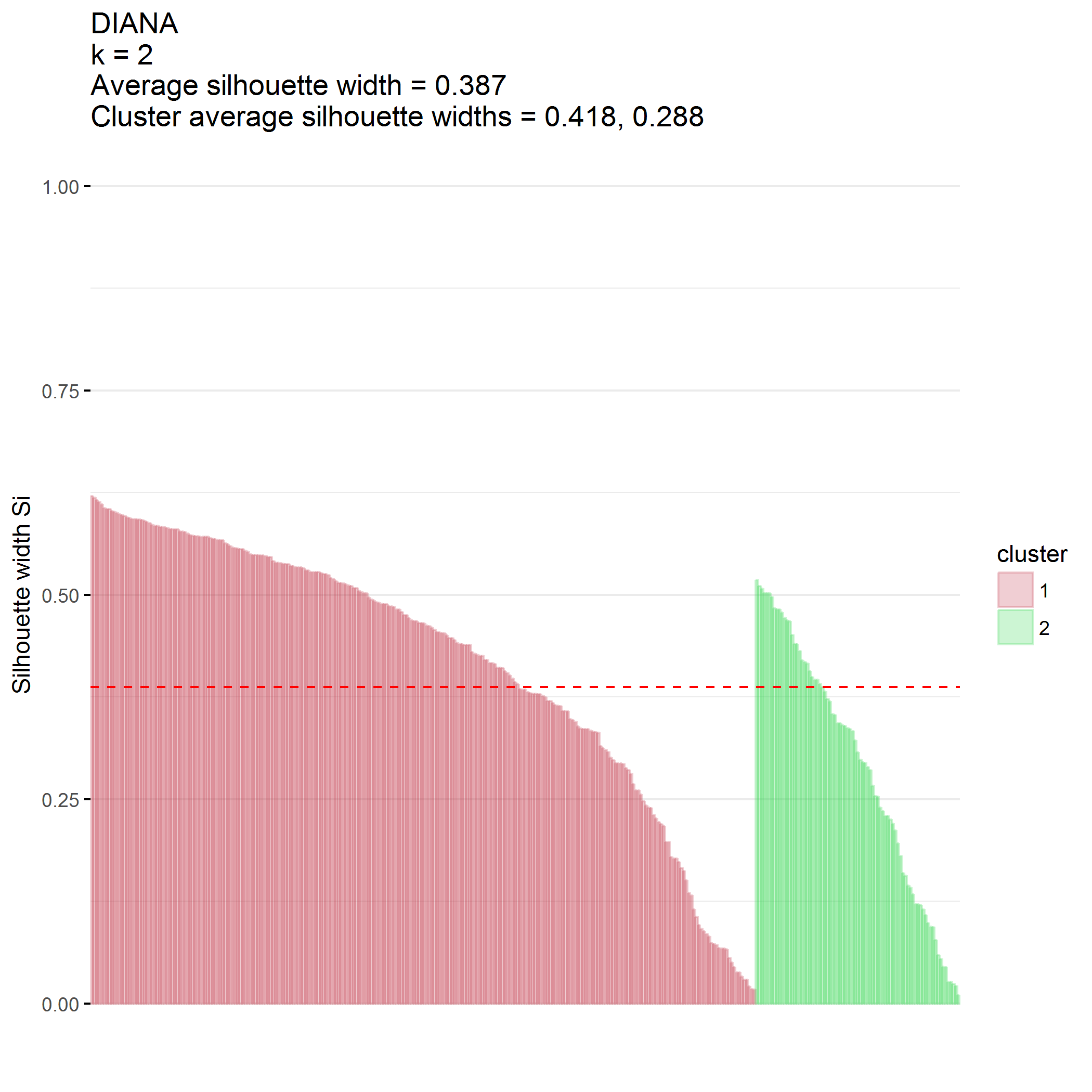
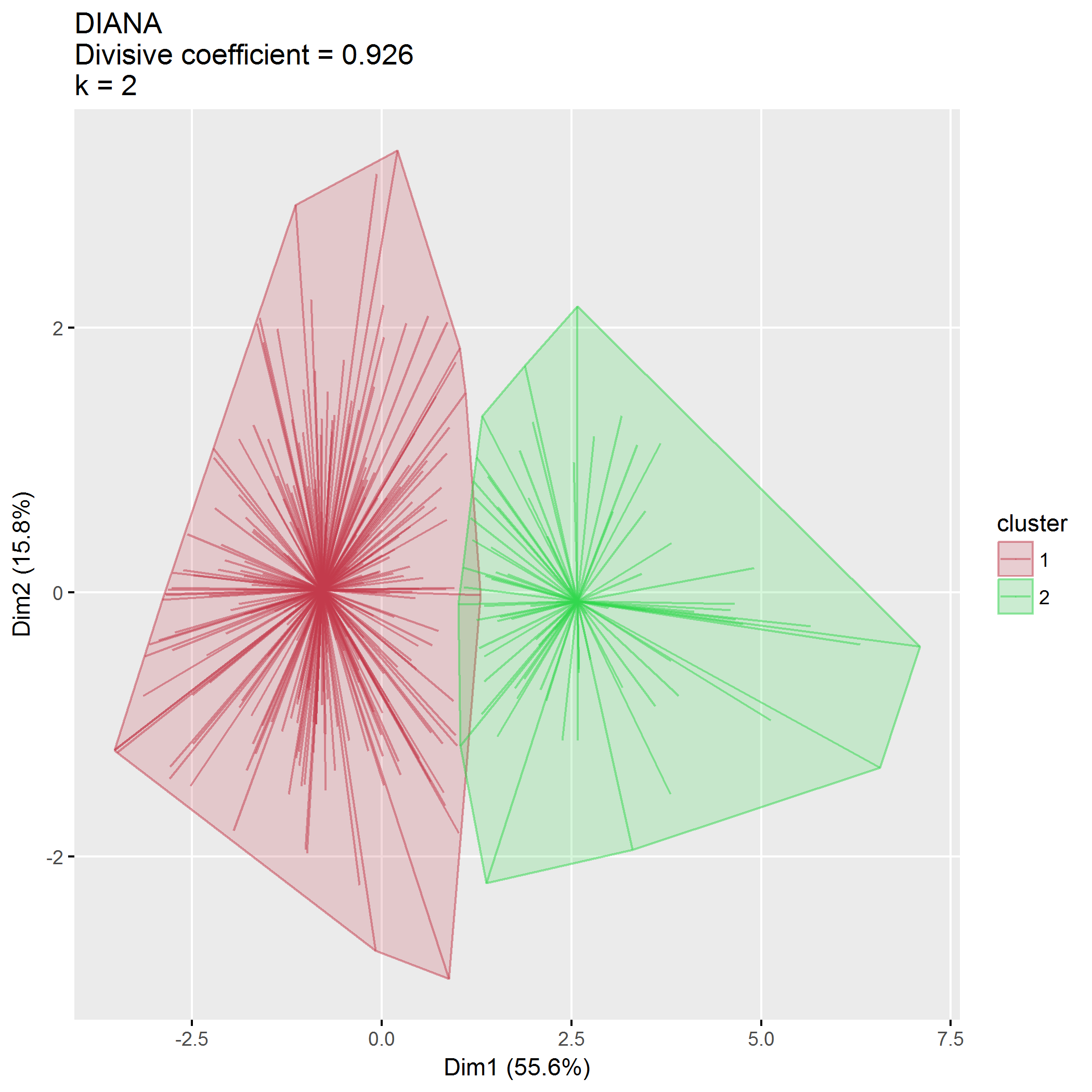
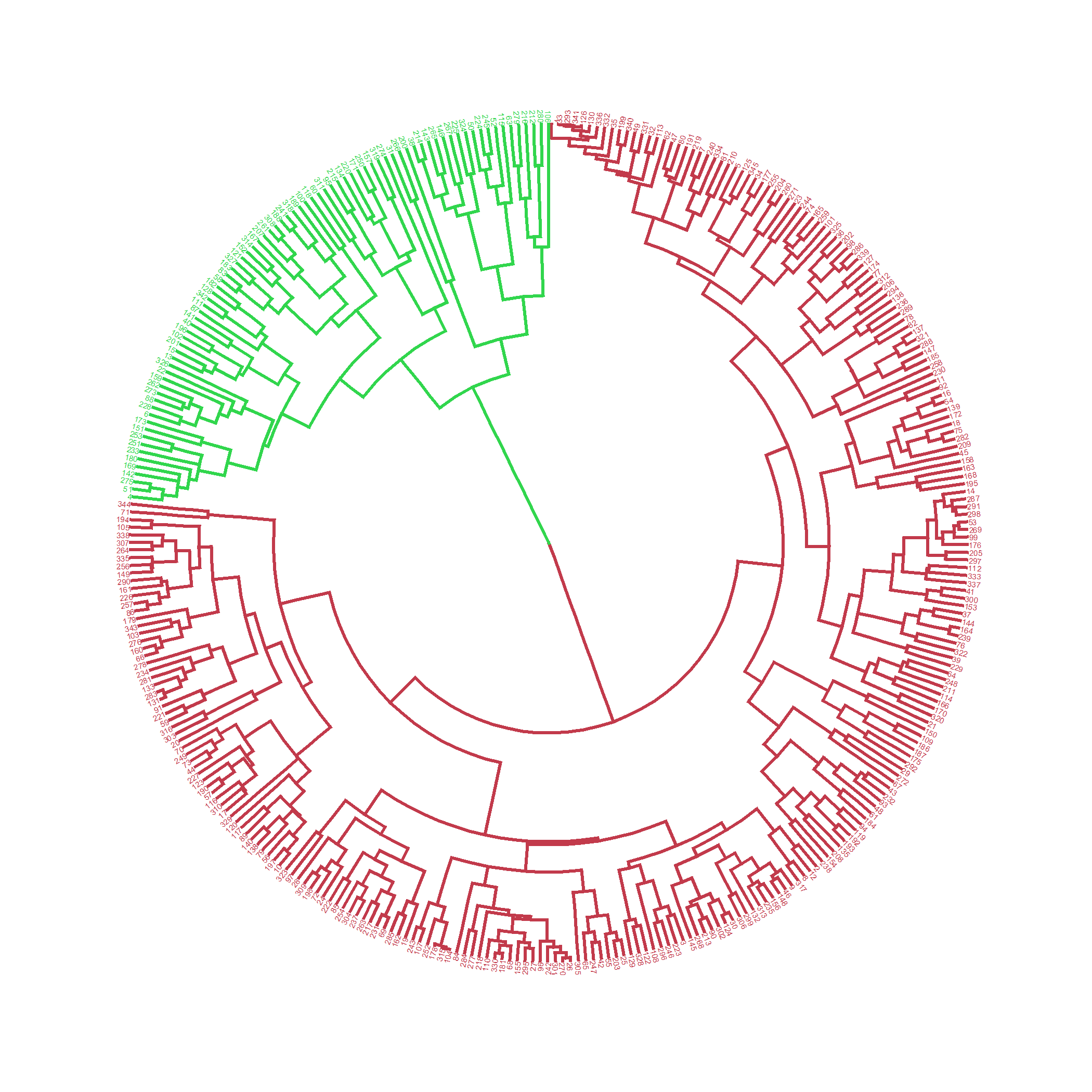
plot of chunk PCB\_diana

Divisive coefficient is 0.926.

### clusters

## cluster size ave.sil.width  
## 1 1 264 0.42  
## 2 2 81 0.29

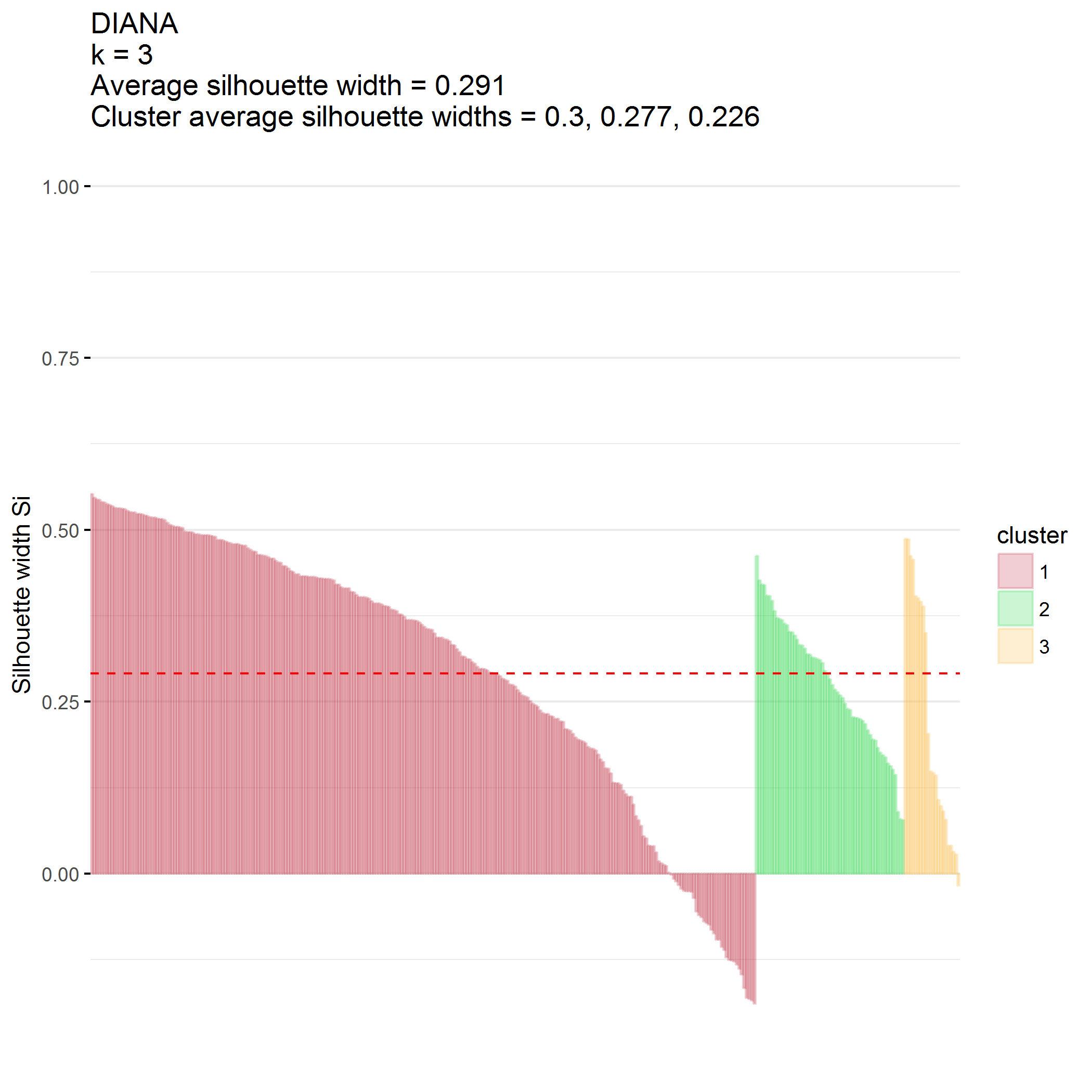
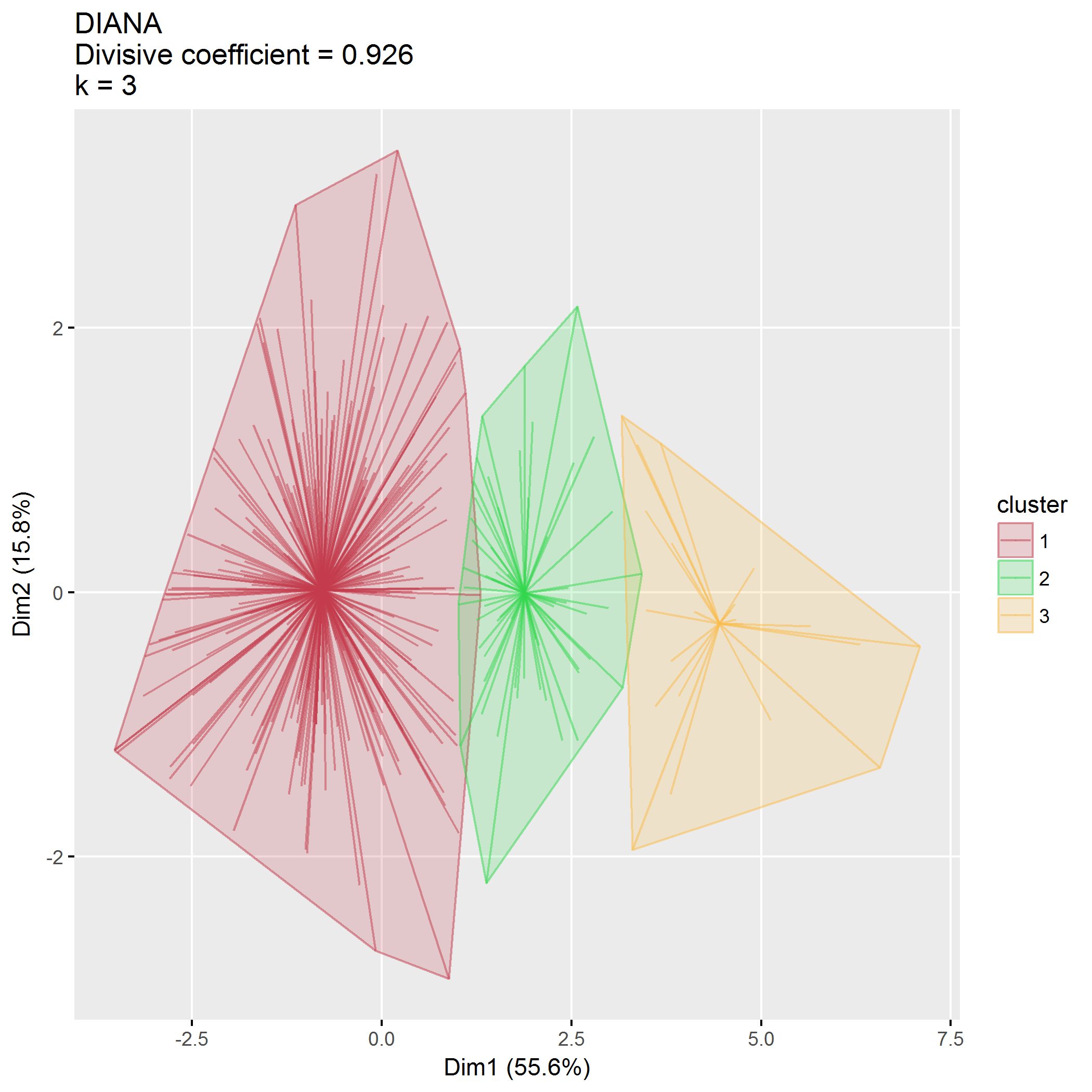
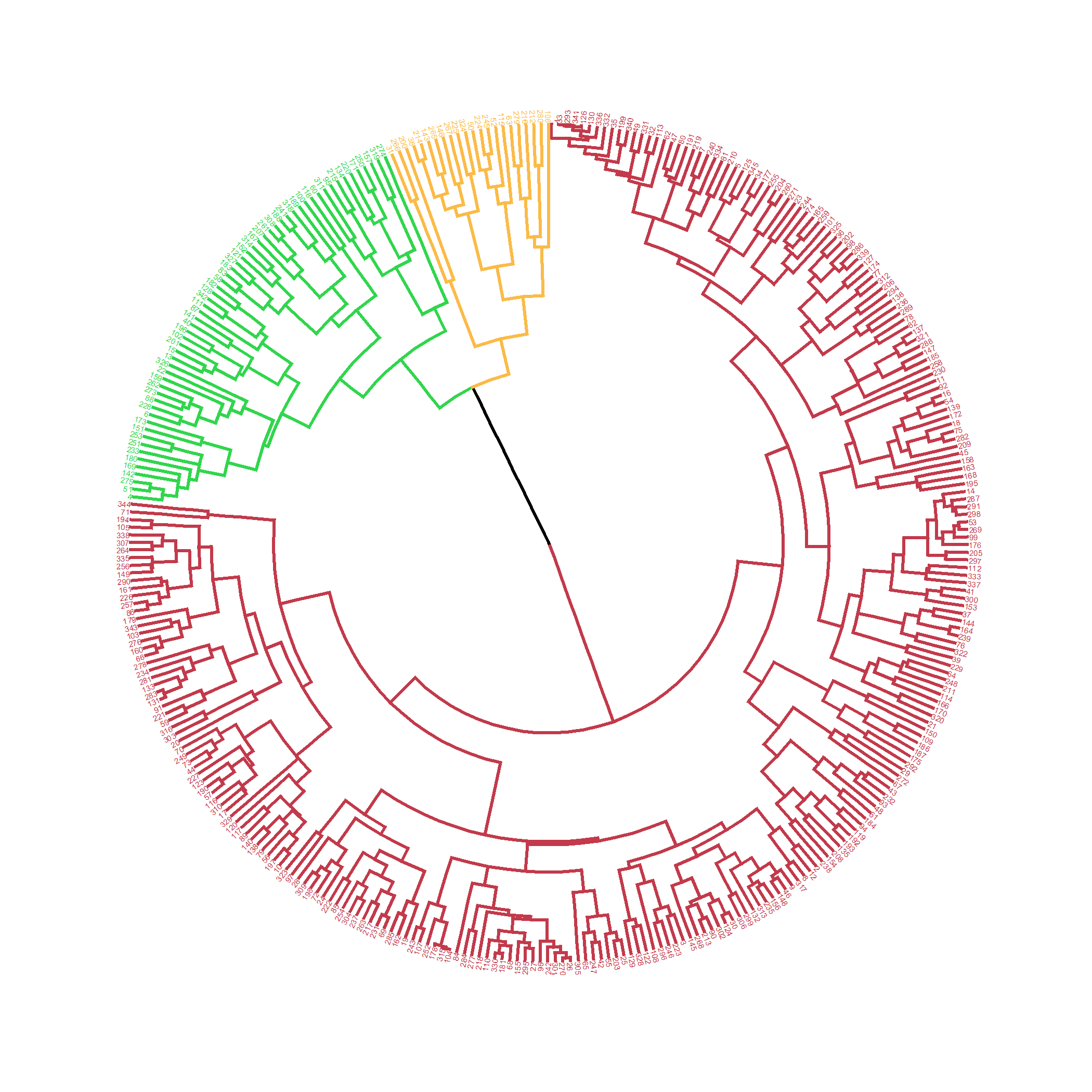
## .  
## 1 2   
## 264 81



### clusters

## cluster size ave.sil.width  
## 1 1 264 0.30  
## 2 2 59 0.28  
## 3 3 22 0.23

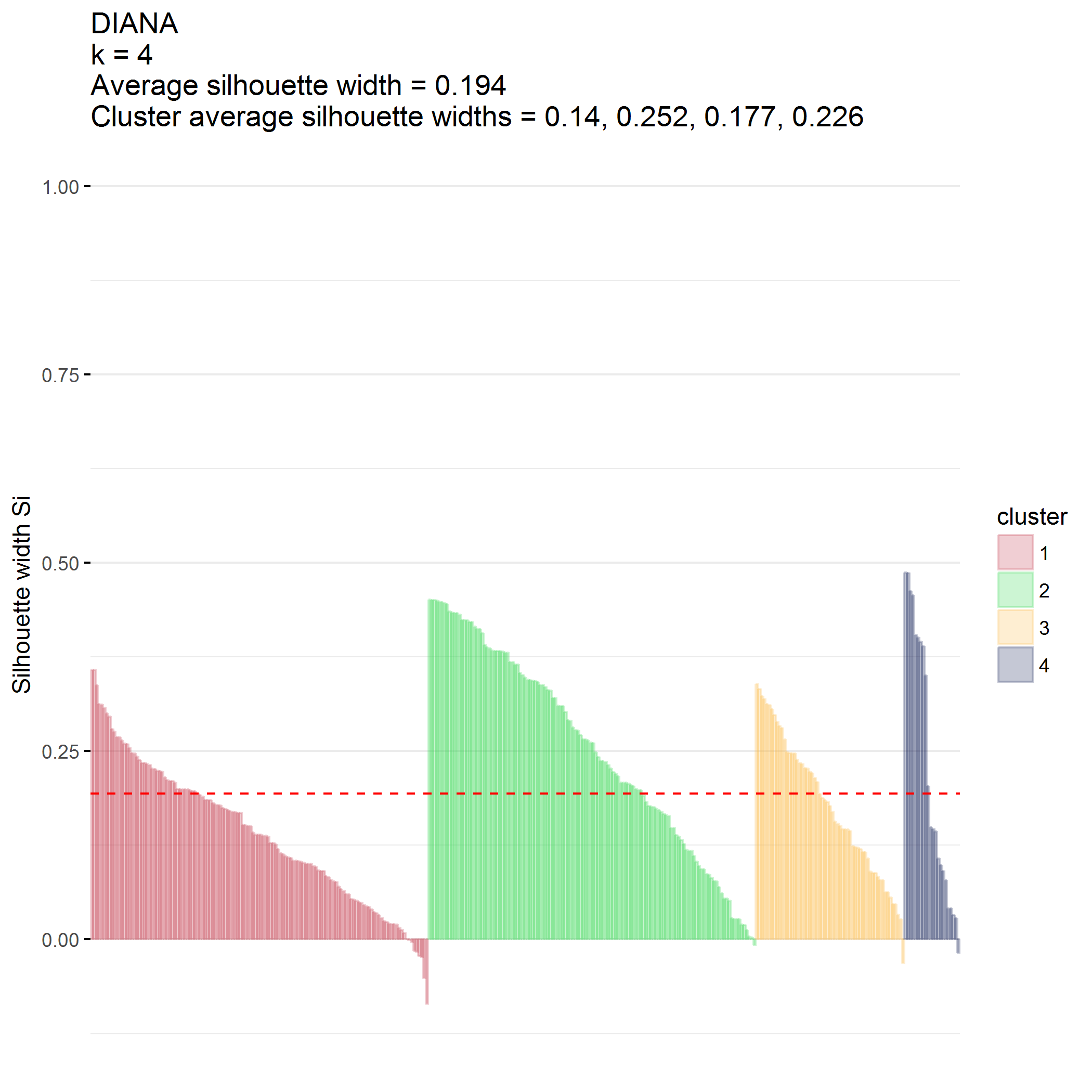
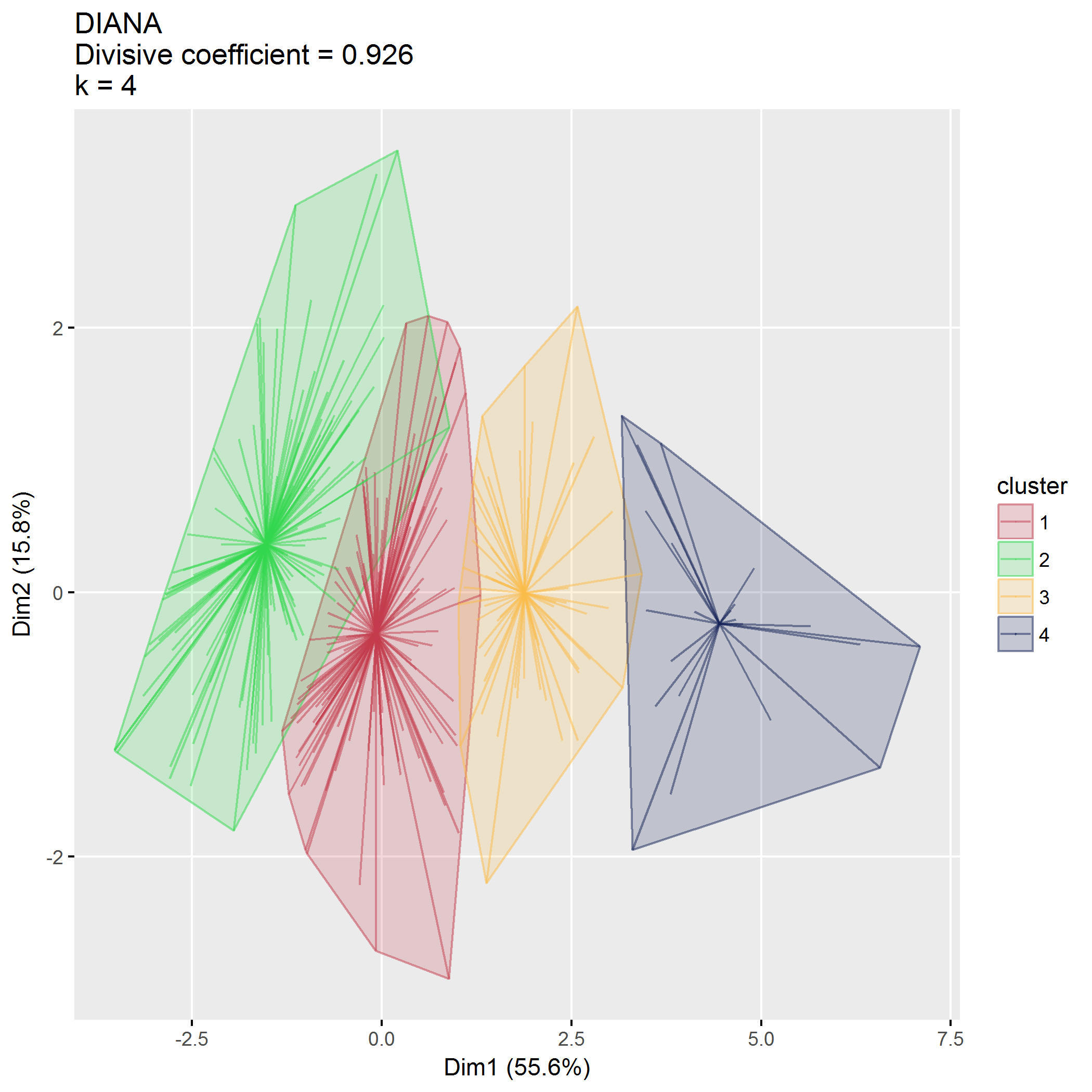
## .  
## 1 2 3   
## 264 59 22



### clusters

## cluster size ave.sil.width  
## 1 1 134 0.14  
## 2 2 130 0.25  
## 3 3 59 0.18  
## 4 4 22 0.23

## .  
## 1 2 3 4   
## 134 130 59 22



## Final cluster identification

Use DIANA,

## .  
## 1 2   
## 264 81

