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

Benjamin Chan (chanb@ohsu.edu)

2018-07-03

# 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 |

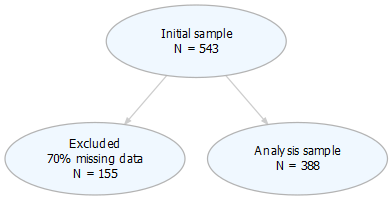
Remove certain predictor variables:

* 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>

##   
## 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

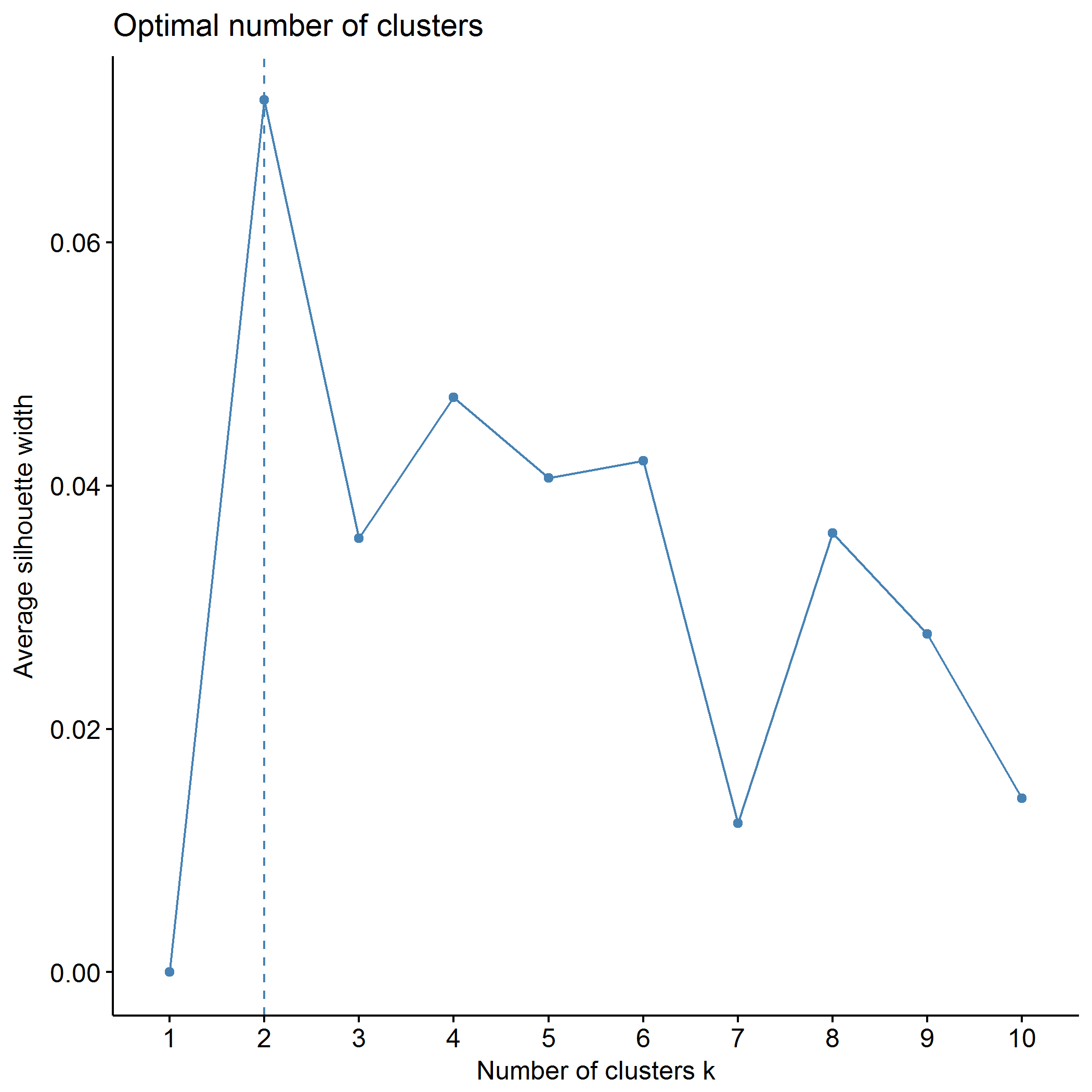
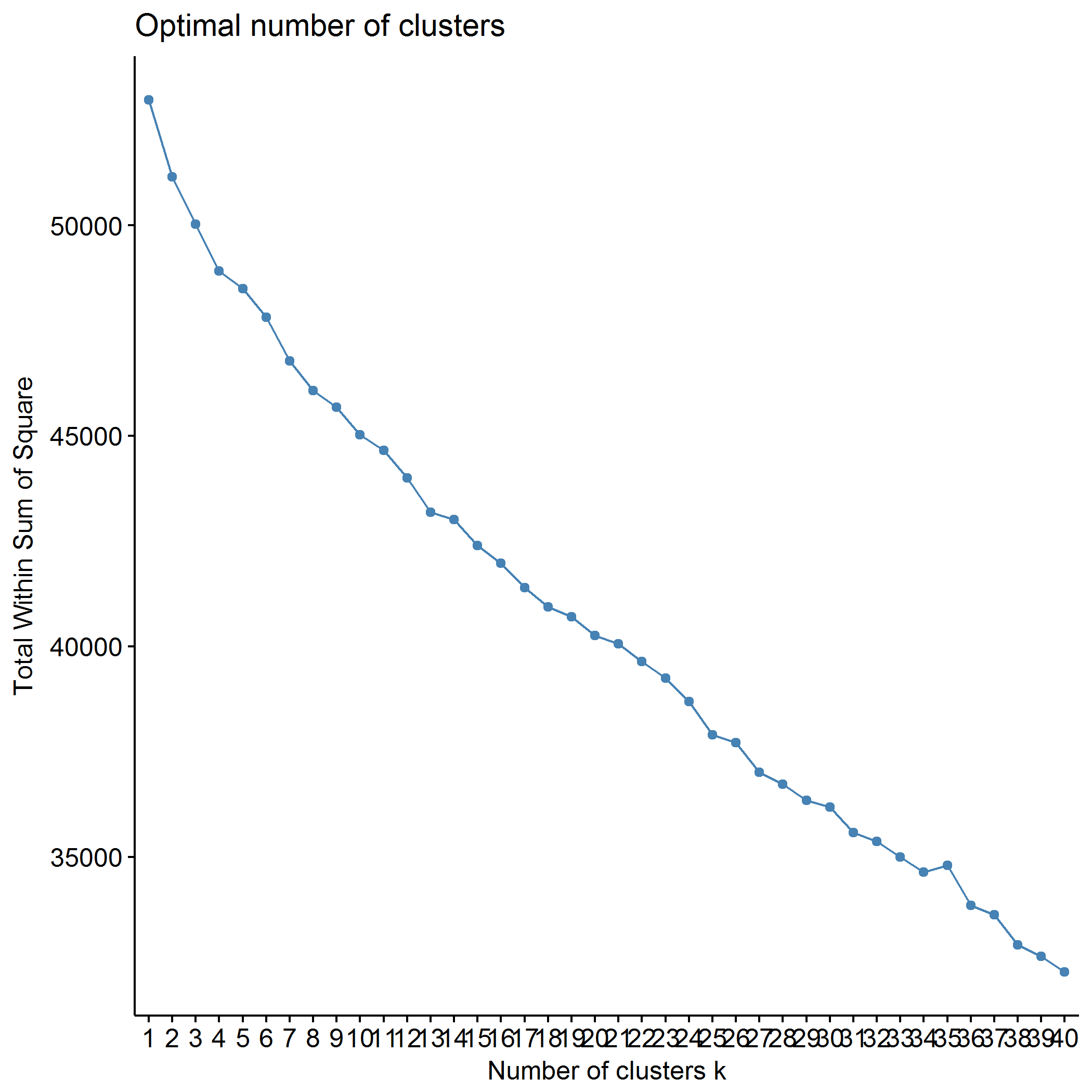
## [1] 345 54

## [1] 345 8

## [1] 345 154

## NULL

## K-means clustering



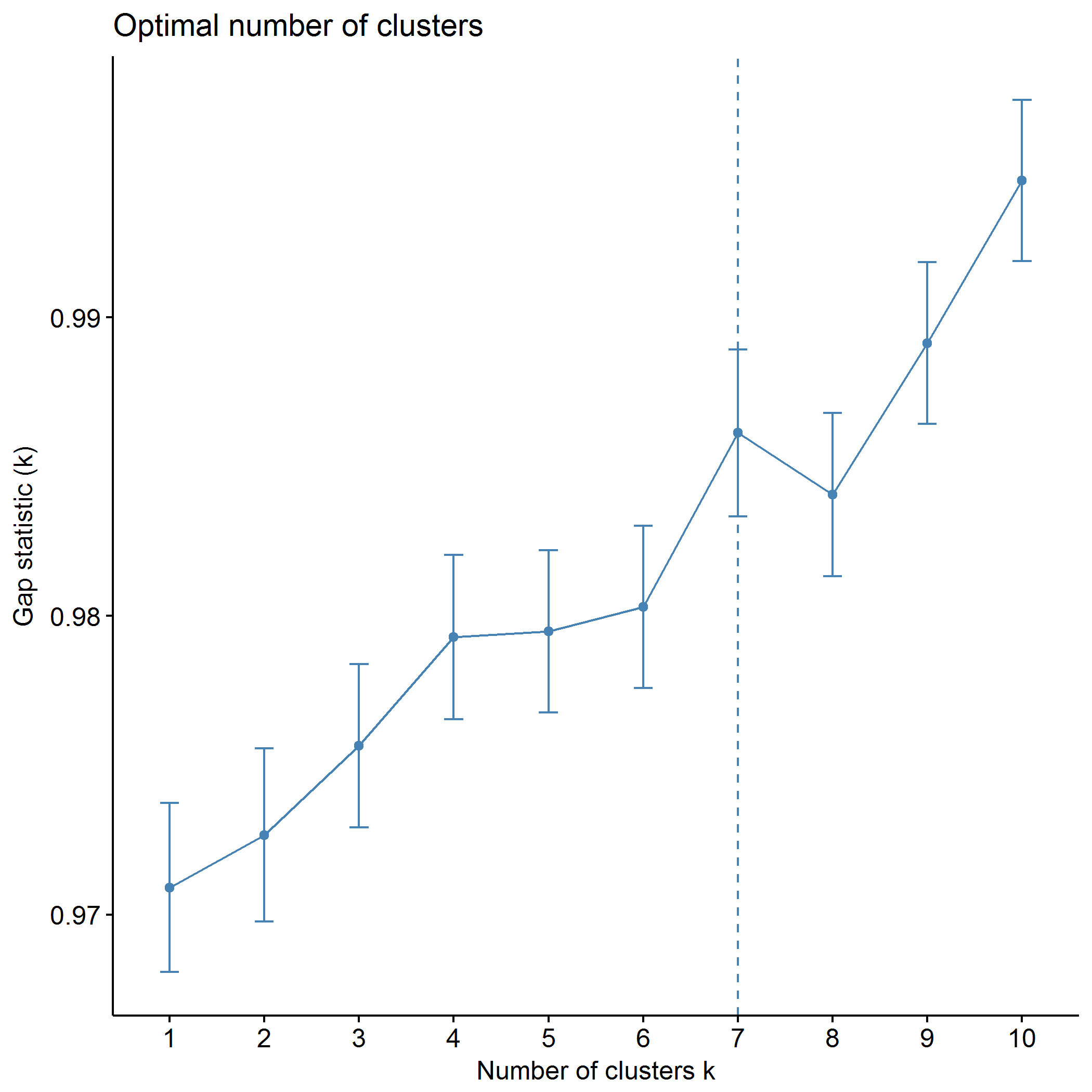
## Clustering k = 1,2,..., K.max (= 10): .. done  
## Bootstrapping, b = 1,2,..., B (= 500) [one "." per sample]:  
## ...................

## Warning: did not converge in 10 iterations

## ............................... 50   
## .................................................. 100   
## .................................................. 150   
## .................................................. 200   
## .................................................. 250   
## .................................................. 300   
## .................................................. 350   
## .................................................. 400   
## ......

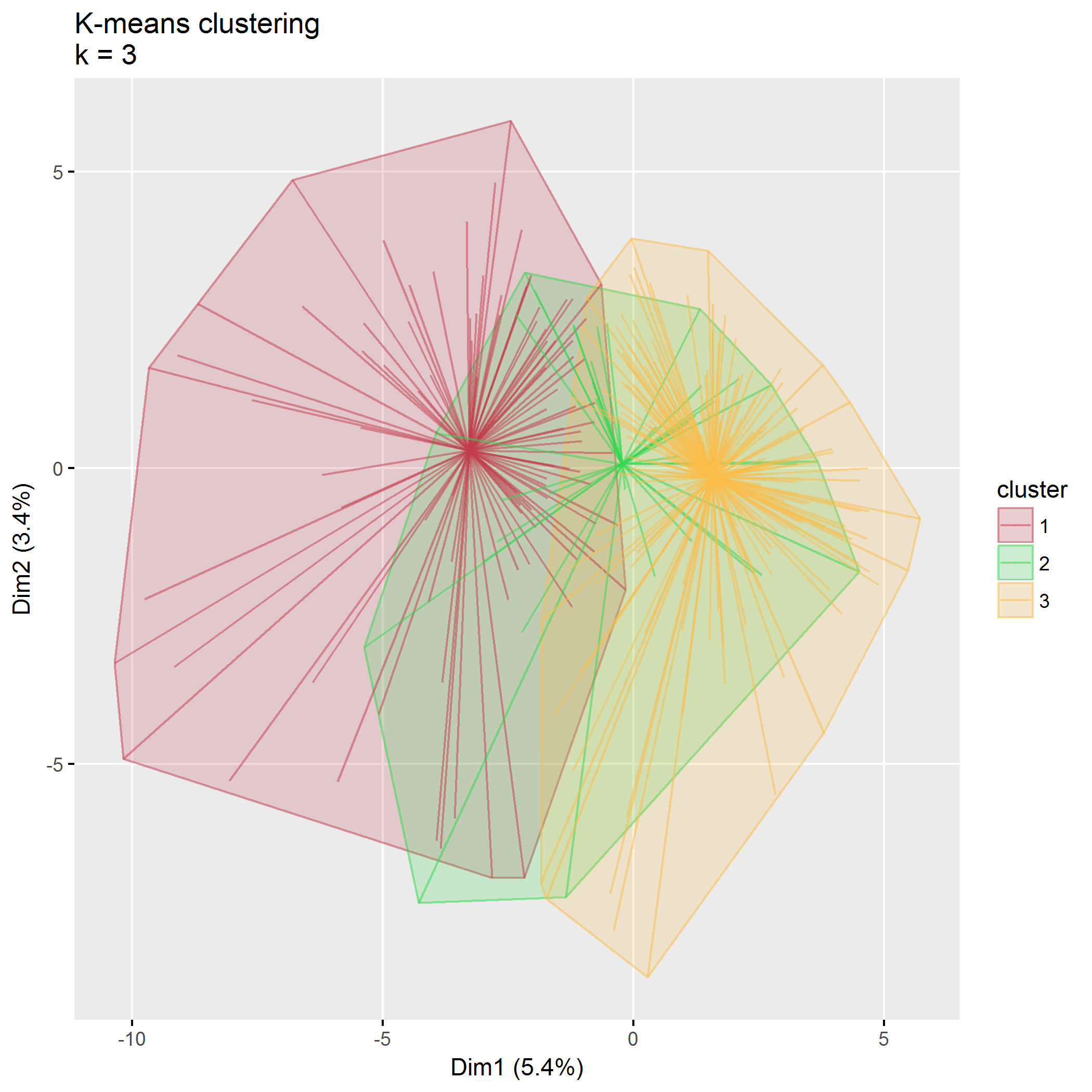
## Warning: did not converge in 10 iterations

## ............................................ 450   
## .................................................. 500



plot of chunk predictors\_kmeans

## .  
## 1 2 3   
## 97 41 207



plot of chunk predictors\_kmeans

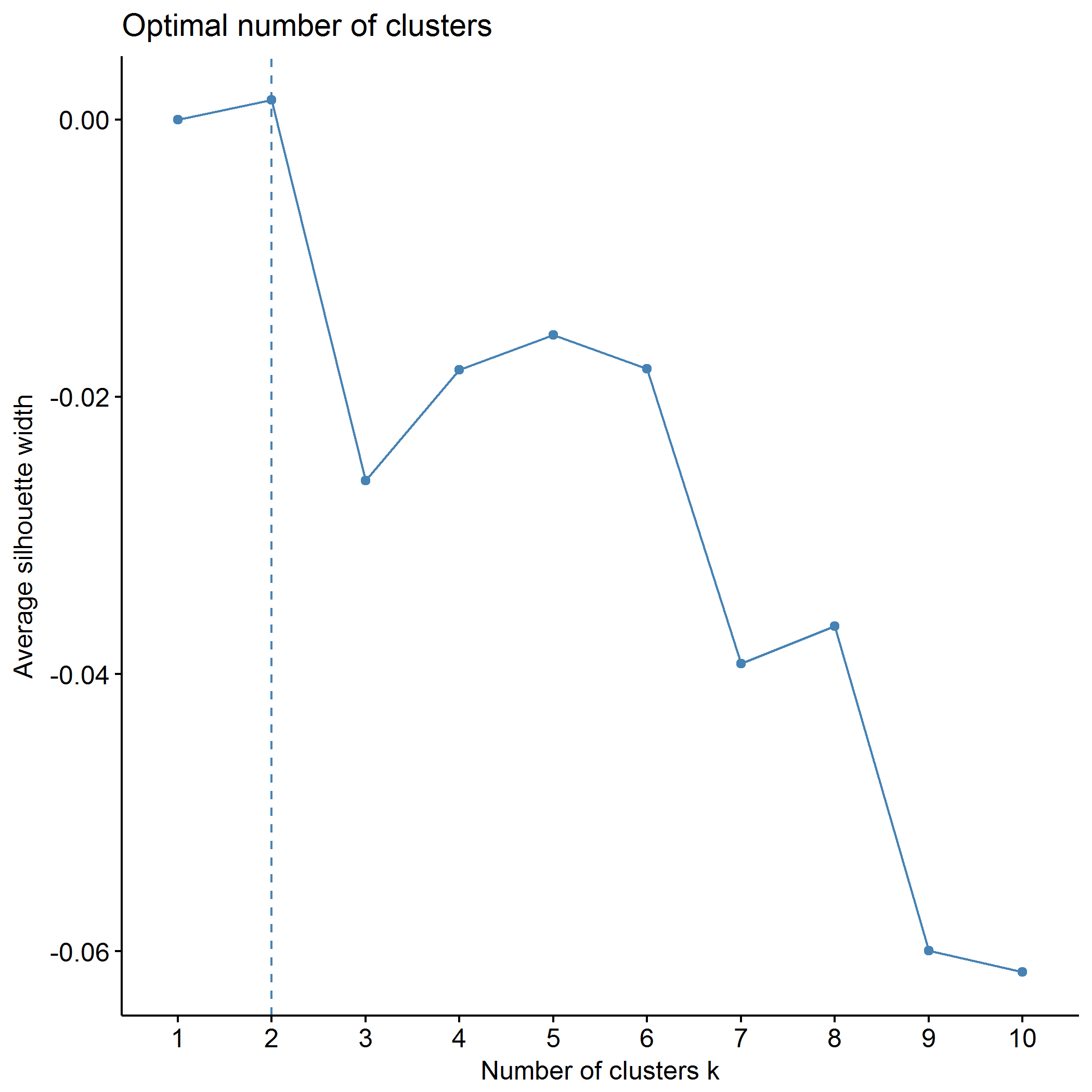
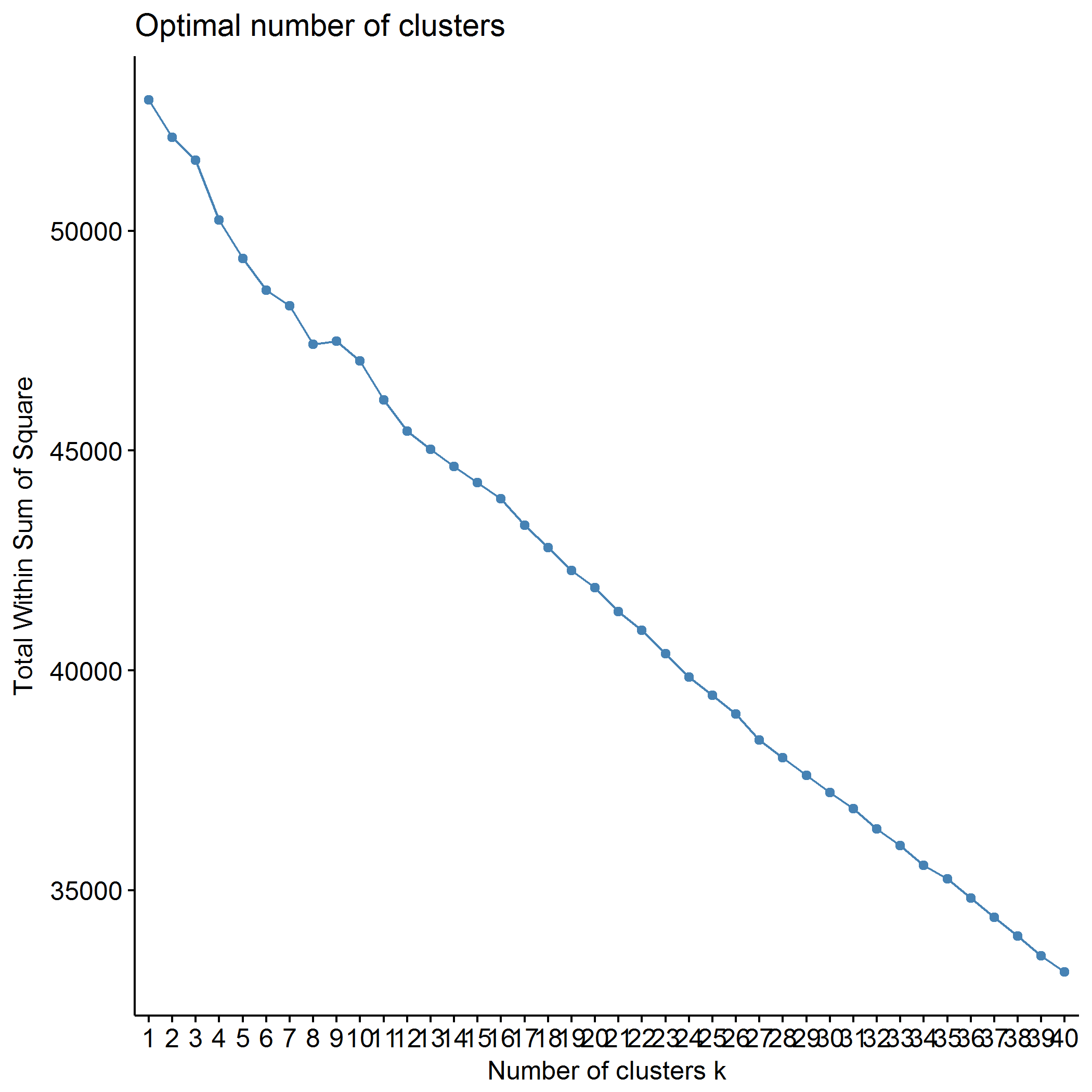
## Within cluster sum of squares, cluster 1: 17104.40  
## Within cluster sum of squares, cluster 2: 7267.38  
## Within cluster sum of squares, cluster 3: 25793.53

## Between SS / Total SS: 2810.69 / 52976.00 = 5.31%

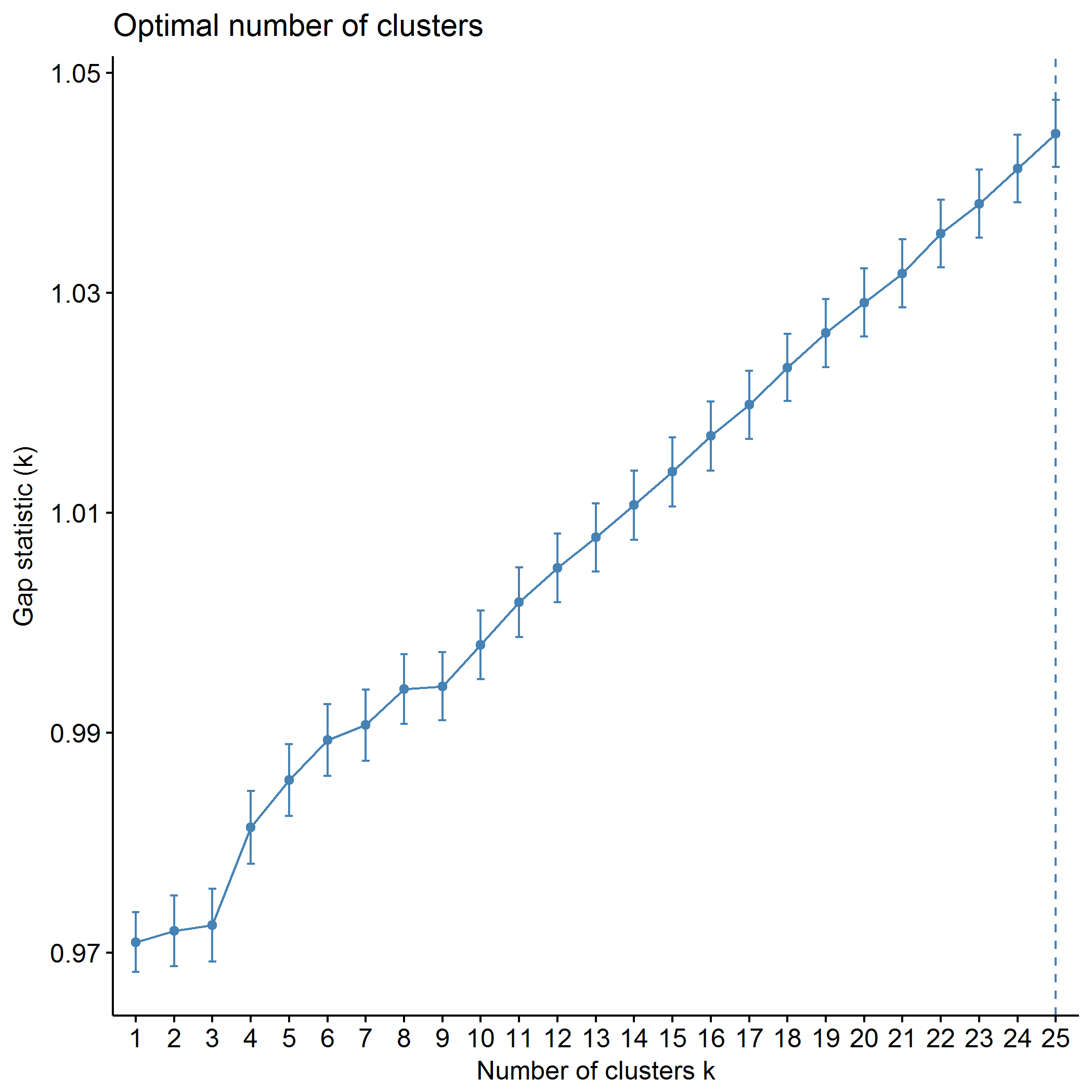
## Total within SS: 50165.31

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| languageSurveyEnglish | 0.05 | 0.05 | -0.04 |
| languageSurveySpanish | -0.05 | -0.05 | 0.04 |
| totalChildren | -0.06 | 0.10 | 0.01 |
| birthOrderOldest | 0.18 | -0.24 | -0.04 |
| birthOrderMiddle | -0.10 | -0.08 | 0.06 |
| birthOrderYoungest | -0.03 | 0.32 | -0.05 |
| childSexMale | -0.05 | -0.11 | 0.05 |
| childAge | 0.23 | 0.16 | -0.14 |
| childEthnicityNot Hispanic/Latino | -0.14 | -0.50 | 0.17 |
| childEthnicityUnknown | -0.13 | 0.43 | -0.02 |
| childEthnicityPrefer not to respond | 0.18 | -0.01 | -0.08 |
| childRaceWhite1 | -0.68 | 0.24 | 0.27 |
| childRaceAsian1 | 0.67 | -0.26 | -0.26 |
| childRaceAfrAm1 | 0.18 | 0.19 | -0.12 |
| childRaceAIAN1 | -0.09 | 0.33 | -0.03 |
| childRaceNHPI1 | 0.15 | 0.03 | -0.08 |
| childRaceOther1 | 0.29 | -0.09 | -0.12 |
| childRaceNoResp1 | 0.17 | 0.01 | -0.08 |
| childRelationshipBiological or adoptive father | 0.44 | -0.24 | -0.16 |
| childRelationshipGrandparent | -0.05 | 0.40 | -0.05 |
| childRelationshipOther | 0.19 | -0.08 | -0.08 |
| parentGenderFemale | -0.47 | 0.12 | 0.20 |
| parentGenderTransgender | -0.05 | -0.05 | 0.04 |
| parentGenderOther | 0.14 | -0.05 | -0.05 |
| parentGenderPrefer not to respond | -0.09 | 0.69 | -0.09 |
| parentSexMale | 0.44 | -0.27 | -0.15 |
| parentAge | 0.16 | -0.24 | -0.03 |
| parentEthnicityNot Hispanic/Latino | -0.24 | -0.15 | 0.14 |
| parentEthnicityUnknown | 0.10 | 0.05 | -0.06 |
| parentEthnicityPrefer not to respond | 0.20 | -0.06 | -0.08 |
| parentRaceWhite1 | -0.67 | 0.25 | 0.26 |
| parentRaceAsian1 | 0.67 | -0.27 | -0.26 |
| parentRaceAfrAm1 | 0.10 | 0.05 | -0.06 |
| parentRaceAIAN1 | -0.15 | 0.33 | 0.01 |
| parentRaceNHPI1 | 0.08 | 0.03 | -0.04 |
| parentRaceOther1 | 0.26 | -0.19 | -0.08 |
| parentRaceNoResp1 | 0.19 | 0.03 | -0.09 |
| parentMaritalStatusWidowed | -0.05 | 0.40 | -0.05 |
| parentMaritalStatusDivorced | 0.02 | 1.05 | -0.22 |
| parentMaritalStatusSeparated | -0.13 | 0.98 | -0.13 |
| parentMaritalStatusRemarried | 0.08 | -0.11 | -0.02 |
| parentMaritalStatusNever married | -0.06 | 0.84 | -0.14 |
| parentSituationCouple parenting with spouse or partner in the same household | 0.26 | -2.54 | 0.38 |
| parentSituationCo-parenting in separate households | -0.09 | 1.21 | -0.20 |
| parentsNumber | 0.26 | -2.54 | 0.38 |
| parentChildRatio | 0.07 | -0.93 | 0.15 |
| zipcodeClass2 | -0.36 | 0.20 | 0.13 |
| zipcode91020 | 0.14 | -0.05 | -0.05 |
| zipcode91204 | 0.14 | -0.05 | -0.05 |
| zipcode91206 | 0.14 | -0.05 | -0.05 |
| zipcode91210 | 0.14 | -0.05 | -0.05 |
| zipcode91402 | 0.14 | -0.05 | -0.05 |
| zipcode97003 | -0.09 | -0.09 | 0.06 |
| zipcode97006 | 0.18 | 0.19 | -0.12 |
| zipcode97007 | -0.08 | -0.08 | 0.05 |
| zipcode97008 | 0.10 | -0.13 | -0.02 |
| zipcode97023 | -0.05 | -0.05 | 0.04 |
| zipcode97027 | -0.08 | -0.08 | 0.05 |
| zipcode97032 | -0.05 | -0.05 | 0.04 |
| zipcode97034 | 0.06 | 0.24 | -0.08 |
| zipcode97035 | 0.14 | -0.05 | -0.05 |
| zipcode97045 | -0.05 | -0.05 | 0.04 |
| zipcode97056 | -0.05 | -0.05 | 0.04 |
| zipcode97060 | 0.14 | -0.05 | -0.05 |
| zipcode97062 | 0.14 | -0.05 | -0.05 |
| zipcode97068 | -0.05 | -0.05 | 0.04 |
| zipcode97071 | -0.08 | -0.08 | 0.05 |
| zipcode97078 | -0.08 | 0.24 | -0.01 |
| zipcode97086 | -0.05 | -0.05 | 0.04 |
| zipcode97089 | 0.19 | -0.08 | -0.08 |
| zipcode97101 | -0.05 | 0.40 | -0.05 |
| zipcode97116 | -0.08 | -0.08 | 0.05 |
| zipcode97123 | 0.13 | -0.09 | -0.04 |
| zipcode97124 | 0.05 | 0.29 | -0.08 |
| zipcode97140 | 0.06 | 0.24 | -0.08 |
| zipcode97141 | 0.14 | -0.05 | -0.05 |
| zipcode97201 | 0.18 | -0.11 | -0.06 |
| zipcode97202 | 0.32 | -0.08 | -0.13 |
| zipcode97203 | 0.14 | -0.05 | -0.05 |
| zipcode97206 | -0.07 | -0.14 | 0.06 |
| zipcode97209 | 0.02 | 0.17 | -0.04 |
| zipcode97210 | -0.05 | 0.40 | -0.05 |
| zipcode97211 | -0.01 | 0.12 | -0.02 |
| zipcode97212 | -0.11 | -0.11 | 0.07 |
| zipcode97213 | 0.02 | -0.09 | 0.01 |
| zipcode97214 | 0.22 | -0.12 | -0.08 |
| zipcode97215 | -0.05 | -0.05 | 0.04 |
| zipcode97217 | 0.13 | -0.09 | -0.04 |
| zipcode97219 | 0.17 | -0.04 | -0.07 |
| zipcode97220 | -0.08 | -0.08 | 0.05 |
| zipcode97221 | -0.08 | -0.08 | 0.05 |
| zipcode97222 | -0.09 | 0.17 | 0.01 |
| zipcode97223 | 0.05 | -0.12 | 0.00 |
| zipcode97224 | 0.19 | -0.08 | -0.08 |
| zipcode97225 | 0.05 | 0.08 | -0.04 |
| zipcode97227 | -0.05 | -0.05 | 0.04 |
| zipcode97229 | 0.23 | -0.20 | -0.07 |
| zipcode97230 | -0.08 | -0.08 | 0.05 |
| zipcode97232 | -0.08 | 0.24 | -0.01 |
| zipcode97233 | 0.19 | -0.08 | -0.08 |
| zipcode97236 | -0.05 | -0.05 | 0.04 |
| zipcode97239 | 0.05 | -0.18 | 0.01 |
| zipcode97266 | -0.05 | -0.05 | 0.04 |
| zipcode97267 | -0.09 | 0.17 | 0.01 |
| zipcode97321 | -0.05 | -0.05 | 0.04 |
| zipcode97325 | -0.05 | -0.05 | 0.04 |
| zipcode97429 | 0.14 | -0.05 | -0.05 |
| zipcode97527 | -0.08 | -0.08 | 0.05 |
| zipcode97701 | -0.21 | 0.16 | 0.07 |
| zipcode97702 | -0.26 | -0.16 | 0.15 |
| zipcode97703 | 0.01 | -0.17 | 0.03 |
| zipcode97707 | -0.05 | -0.05 | 0.04 |
| zipcode97734 | -0.09 | -0.09 | 0.06 |
| zipcode97738 | -0.05 | -0.05 | 0.04 |
| zipcode97741 | -0.05 | 0.26 | -0.03 |
| zipcode97753 | -0.09 | 0.17 | 0.01 |
| zipcode97754 | -0.15 | 0.01 | 0.07 |
| zipcode97756 | -0.25 | 0.18 | 0.08 |
| zipcode97759 | -0.08 | -0.08 | 0.05 |
| zipcode97760 | 0.02 | -0.09 | 0.01 |
| zipcode98632 | -0.05 | -0.05 | 0.04 |
| zipcode98660 | 0.14 | -0.05 | -0.05 |
| zipcode98683 | -0.05 | 0.40 | -0.05 |
| zipcode98685 | -0.05 | -0.05 | 0.04 |
| communitySuburban | -0.03 | 0.06 | 0.00 |
| communityRural | -0.24 | 0.08 | 0.10 |
| distance | 0.00 | 0.00 | 0.00 |
| parentEducationVocational school/some college | 0.10 | 0.30 | -0.11 |
| parentEducationCollege | -0.12 | -0.27 | 0.11 |
| parentEducationGraduate/professional school | 0.15 | -0.35 | 0.00 |
| income$25,001-$49,999 | 0.09 | 0.41 | -0.12 |
| income$50,000-$79,999 | -0.04 | -0.16 | 0.05 |
| income$80,000-$119,999 | -0.16 | -0.44 | 0.16 |
| income$120,000-$149,999 | 0.25 | -0.32 | -0.05 |
| income$150,000 or more | -0.04 | -0.40 | 0.10 |
| internet | -0.03 | -0.29 | 0.07 |
| ECBI\_intensity\_T\_score | 0.62 | -0.03 | -0.28 |
| ECBI\_problem\_T\_score | 0.63 | 0.01 | -0.29 |
| ECBI\_Opp | 0.58 | 0.09 | -0.29 |
| ECBI\_Inatt | 0.31 | -0.17 | -0.11 |
| ECBI\_Cond | 0.53 | 0.08 | -0.26 |
| MAPS\_PP | -0.44 | -0.38 | 0.28 |
| MAPS\_PR | -0.55 | 0.06 | 0.24 |
| MAPS\_WM | -0.54 | 0.06 | 0.24 |
| MAPS\_SP | -0.60 | -0.05 | 0.29 |
| MAPS\_HS | 0.86 | -0.37 | -0.33 |
| MAPS\_LC | 0.74 | 0.23 | -0.39 |
| MAPS\_PC | 0.58 | -0.08 | -0.25 |
| MAPS\_POS | -0.67 | -0.10 | 0.34 |
| MAPS\_NEG | 0.96 | -0.09 | -0.43 |
| SEPTI\_nurturance | -0.84 | 0.17 | 0.36 |
| SEPTI\_discipline | -0.69 | -0.09 | 0.34 |
| SEPTI\_play | -0.63 | 0.19 | 0.26 |
| SEPTI\_routine | -0.83 | 0.05 | 0.38 |

## Partitioning around medoids (PAM)

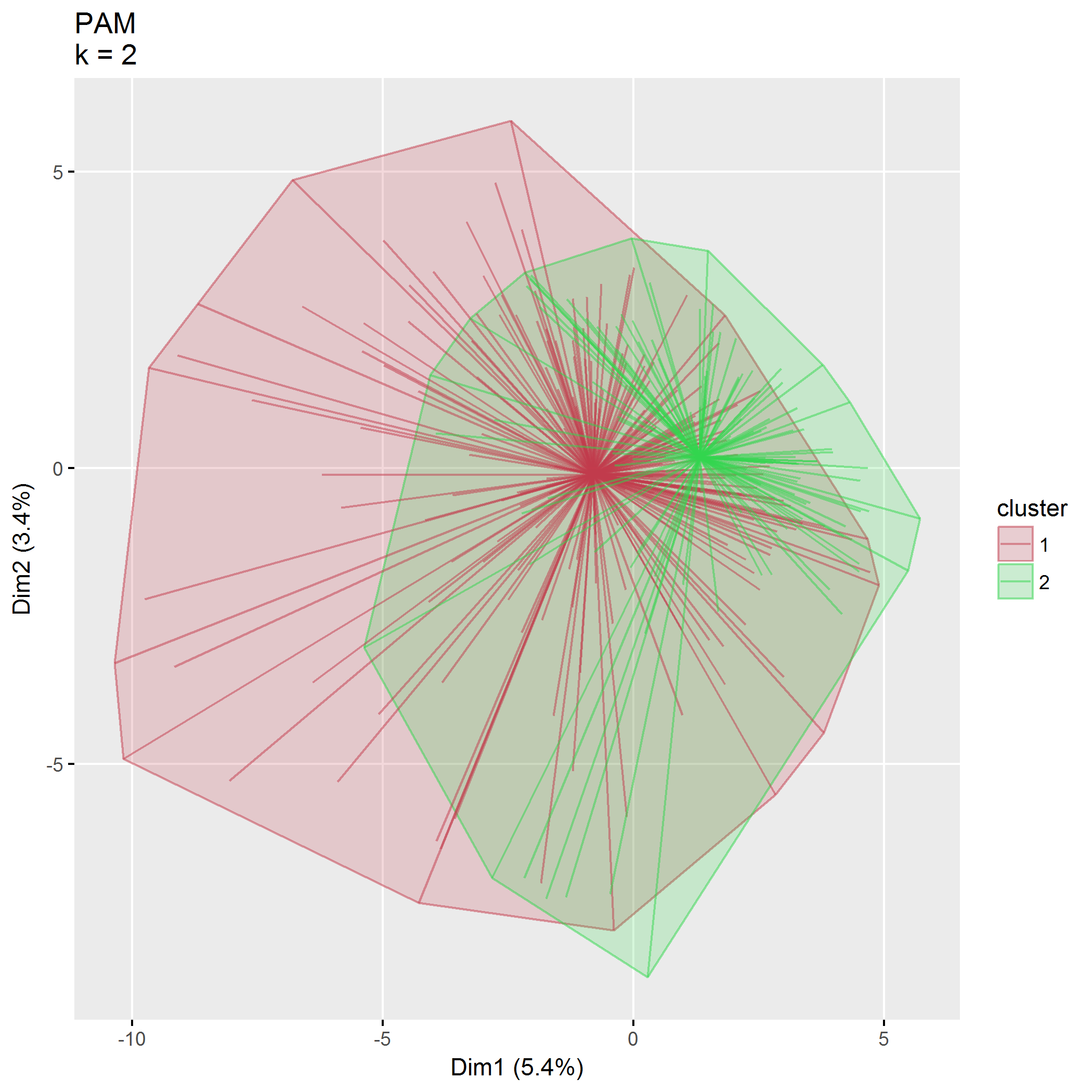


## Clustering k = 1,2,..., K.max (= 25): .. done  
## Bootstrapping, b = 1,2,..., B (= 500) [one "." per sample]:  
## .................................................. 50   
## .................................................. 100   
## .................................................. 150   
## .................................................. 200   
## .................................................. 250   
## .................................................. 300   
## .................................................. 350   
## .................................................. 400   
## .................................................. 450   
## .................................................. 500



plot of chunk predictors\_pam

## .  
## 1 2   
## 214 131

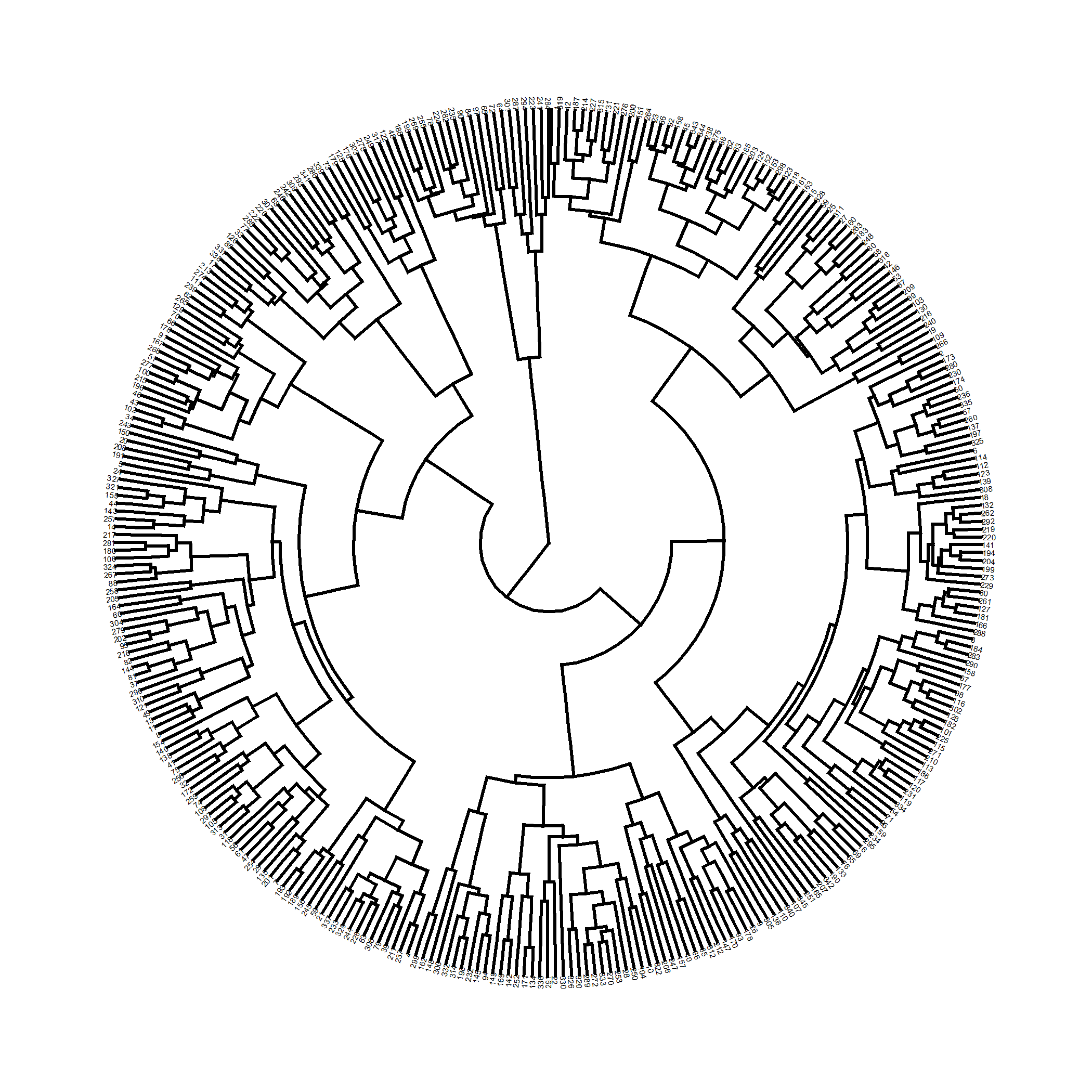


plot of chunk predictors\_pam

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| size | max\_diss | av\_diss | diameter | separation |
| 214 | 104.21 | 49.54 | 161.32 | 14.38 |
| 131 | 81.93 | 43.70 | 135.56 | 14.38 |

|  |  |  |
| --- | --- | --- |
|  | 292 | 298 |
| languageSurveyEnglish | 0.05 | 0.05 |
| languageSurveySpanish | -0.05 | -0.05 |
| totalChildren | -0.99 | 0.87 |
| birthOrderOldest | -0.62 | -0.62 |
| birthOrderMiddle | -0.38 | -0.38 |
| birthOrderYoungest | -0.61 | 1.63 |
| childSexMale | 0.92 | 0.92 |
| childAge | -1.25 | -0.10 |
| childEthnicityNot Hispanic/Latino | 0.54 | 0.54 |
| childEthnicityUnknown | -0.13 | -0.13 |
| childEthnicityPrefer not to respond | -0.29 | -0.29 |
| childRaceWhite1 | 0.49 | 0.49 |
| childRaceAsian1 | -0.34 | -0.34 |
| childRaceAfrAm1 | -0.20 | -0.20 |
| childRaceAIAN1 | -0.15 | -0.15 |
| childRaceNHPI1 | -0.14 | -0.14 |
| childRaceOther1 | -0.21 | -0.21 |
| childRaceNoResp1 | -0.27 | -0.27 |
| childRelationshipBiological or adoptive father | -0.39 | -0.39 |
| childRelationshipGrandparent | -0.05 | -0.05 |
| childRelationshipOther | -0.08 | -0.08 |
| parentGenderFemale | 0.44 | 0.44 |
| parentGenderTransgender | -0.05 | -0.05 |
| parentGenderOther | -0.05 | -0.05 |
| parentGenderPrefer not to respond | -0.09 | -0.09 |
| parentSexMale | -0.41 | -0.41 |
| parentAge | -0.76 | 0.39 |
| parentEthnicityNot Hispanic/Latino | 0.48 | 0.48 |
| parentEthnicityUnknown | -0.13 | -0.13 |
| parentEthnicityPrefer not to respond | -0.26 | -0.26 |
| parentRaceWhite1 | 0.54 | 0.54 |
| parentRaceAsian1 | -0.35 | -0.35 |
| parentRaceAfrAm1 | -0.13 | -0.13 |
| parentRaceAIAN1 | -0.15 | -0.15 |
| parentRaceNHPI1 | -0.14 | -0.14 |
| parentRaceOther1 | -0.19 | -0.19 |
| parentRaceNoResp1 | -0.27 | -0.27 |
| parentMaritalStatusWidowed | -0.05 | -0.05 |
| parentMaritalStatusDivorced | -0.22 | -0.22 |
| parentMaritalStatusSeparated | -0.13 | -0.13 |
| parentMaritalStatusRemarried | -0.11 | -0.11 |
| parentMaritalStatusNever married | -0.43 | -0.43 |
| parentSituationCouple parenting with spouse or partner in the same household | 0.38 | 0.38 |
| parentSituationCo-parenting in separate households | -0.20 | -0.20 |
| parentsNumber | 0.38 | 0.38 |
| parentChildRatio | 1.47 | -0.83 |
| zipcodeClass2 | -0.62 | 1.62 |
| zipcode91020 | -0.05 | -0.05 |
| zipcode91204 | -0.05 | -0.05 |
| zipcode91206 | -0.05 | -0.05 |
| zipcode91210 | -0.05 | -0.05 |
| zipcode91402 | -0.05 | -0.05 |
| zipcode97003 | -0.09 | -0.09 |
| zipcode97006 | -0.20 | -0.20 |
| zipcode97007 | -0.08 | -0.08 |
| zipcode97008 | -0.13 | -0.13 |
| zipcode97023 | -0.05 | -0.05 |
| zipcode97027 | -0.08 | -0.08 |
| zipcode97032 | -0.05 | -0.05 |
| zipcode97034 | -0.08 | -0.08 |
| zipcode97035 | -0.05 | -0.05 |
| zipcode97045 | -0.05 | -0.05 |
| zipcode97056 | -0.05 | -0.05 |
| zipcode97060 | -0.05 | -0.05 |
| zipcode97062 | -0.05 | -0.05 |
| zipcode97068 | -0.05 | -0.05 |
| zipcode97071 | -0.08 | -0.08 |
| zipcode97078 | -0.08 | -0.08 |
| zipcode97086 | -0.05 | -0.05 |
| zipcode97089 | -0.08 | -0.08 |
| zipcode97101 | -0.05 | -0.05 |
| zipcode97116 | -0.08 | -0.08 |
| zipcode97123 | -0.09 | -0.09 |
| zipcode97124 | -0.12 | -0.12 |
| zipcode97140 | -0.08 | -0.08 |
| zipcode97141 | -0.05 | -0.05 |
| zipcode97201 | -0.11 | -0.11 |
| zipcode97202 | -0.21 | -0.21 |
| zipcode97203 | -0.05 | -0.05 |
| zipcode97206 | -0.14 | -0.14 |
| zipcode97209 | -0.09 | -0.09 |
| zipcode97210 | -0.05 | -0.05 |
| zipcode97211 | -0.11 | -0.11 |
| zipcode97212 | -0.11 | -0.11 |
| zipcode97213 | -0.09 | -0.09 |
| zipcode97214 | -0.12 | -0.12 |
| zipcode97215 | -0.05 | -0.05 |
| zipcode97217 | -0.09 | -0.09 |
| zipcode97219 | -0.18 | -0.18 |
| zipcode97220 | -0.08 | -0.08 |
| zipcode97221 | -0.08 | -0.08 |
| zipcode97222 | -0.09 | -0.09 |
| zipcode97223 | -0.12 | -0.12 |
| zipcode97224 | -0.08 | -0.08 |
| zipcode97225 | -0.12 | -0.12 |
| zipcode97227 | -0.05 | -0.05 |
| zipcode97229 | -0.20 | -0.20 |
| zipcode97230 | -0.08 | -0.08 |
| zipcode97232 | -0.08 | -0.08 |
| zipcode97233 | -0.08 | -0.08 |
| zipcode97236 | -0.05 | -0.05 |
| zipcode97239 | -0.18 | -0.18 |
| zipcode97266 | -0.05 | -0.05 |
| zipcode97267 | -0.09 | -0.09 |
| zipcode97321 | -0.05 | -0.05 |
| zipcode97325 | -0.05 | -0.05 |
| zipcode97429 | -0.05 | -0.05 |
| zipcode97527 | -0.08 | -0.08 |
| zipcode97701 | 2.72 | -0.37 |
| zipcode97702 | -0.26 | -0.26 |
| zipcode97703 | -0.17 | -0.17 |
| zipcode97707 | -0.05 | -0.05 |
| zipcode97734 | -0.09 | -0.09 |
| zipcode97738 | -0.05 | -0.05 |
| zipcode97741 | -0.17 | -0.17 |
| zipcode97753 | -0.09 | -0.09 |
| zipcode97754 | -0.15 | -0.15 |
| zipcode97756 | -0.43 | 2.34 |
| zipcode97759 | -0.08 | -0.08 |
| zipcode97760 | -0.09 | -0.09 |
| zipcode98632 | -0.05 | -0.05 |
| zipcode98660 | -0.05 | -0.05 |
| zipcode98683 | -0.05 | -0.05 |
| zipcode98685 | -0.05 | -0.05 |
| communitySuburban | 1.09 | 1.09 |
| communityRural | -0.48 | -0.48 |
| distance | 0.01 | -0.34 |
| parentEducationVocational school/some college | -0.49 | 2.02 |
| parentEducationCollege | 1.22 | -0.82 |
| parentEducationGraduate/professional school | -0.62 | -0.62 |
| income$25,001-$49,999 | -0.56 | -0.56 |
| income$50,000-$79,999 | 1.67 | -0.60 |
| income$80,000-$119,999 | -0.44 | 2.27 |
| income$120,000-$149,999 | -0.32 | -0.32 |
| income$150,000 or more | -0.40 | -0.40 |
| internet | 0.16 | 0.16 |
| ECBI\_intensity\_T\_score | 0.47 | -0.38 |
| ECBI\_problem\_T\_score | 0.23 | -0.71 |
| ECBI\_Opp | 0.49 | -0.06 |
| ECBI\_Inatt | 0.36 | -0.12 |
| ECBI\_Cond | -0.54 | -0.39 |
| MAPS\_PP | -0.70 | 0.77 |
| MAPS\_PR | 0.47 | 0.47 |
| MAPS\_WM | -0.52 | 0.78 |
| MAPS\_SP | -0.10 | 0.97 |
| MAPS\_HS | -0.42 | -0.42 |
| MAPS\_LC | 0.21 | -0.22 |
| MAPS\_PC | -0.74 | -0.74 |
| MAPS\_POS | -0.26 | 0.95 |
| MAPS\_NEG | -0.43 | -0.62 |
| SEPTI\_nurturance | -0.60 | 0.42 |
| SEPTI\_discipline | -0.42 | 0.90 |
| SEPTI\_play | 0.46 | -0.82 |
| SEPTI\_routine | -0.71 | 0.08 |

## Agglomerative hierarchical clustering (AGNES)



plot of chunk predictors\_agnes

Correlation between cophenetic distance and the original distance is 0.383.

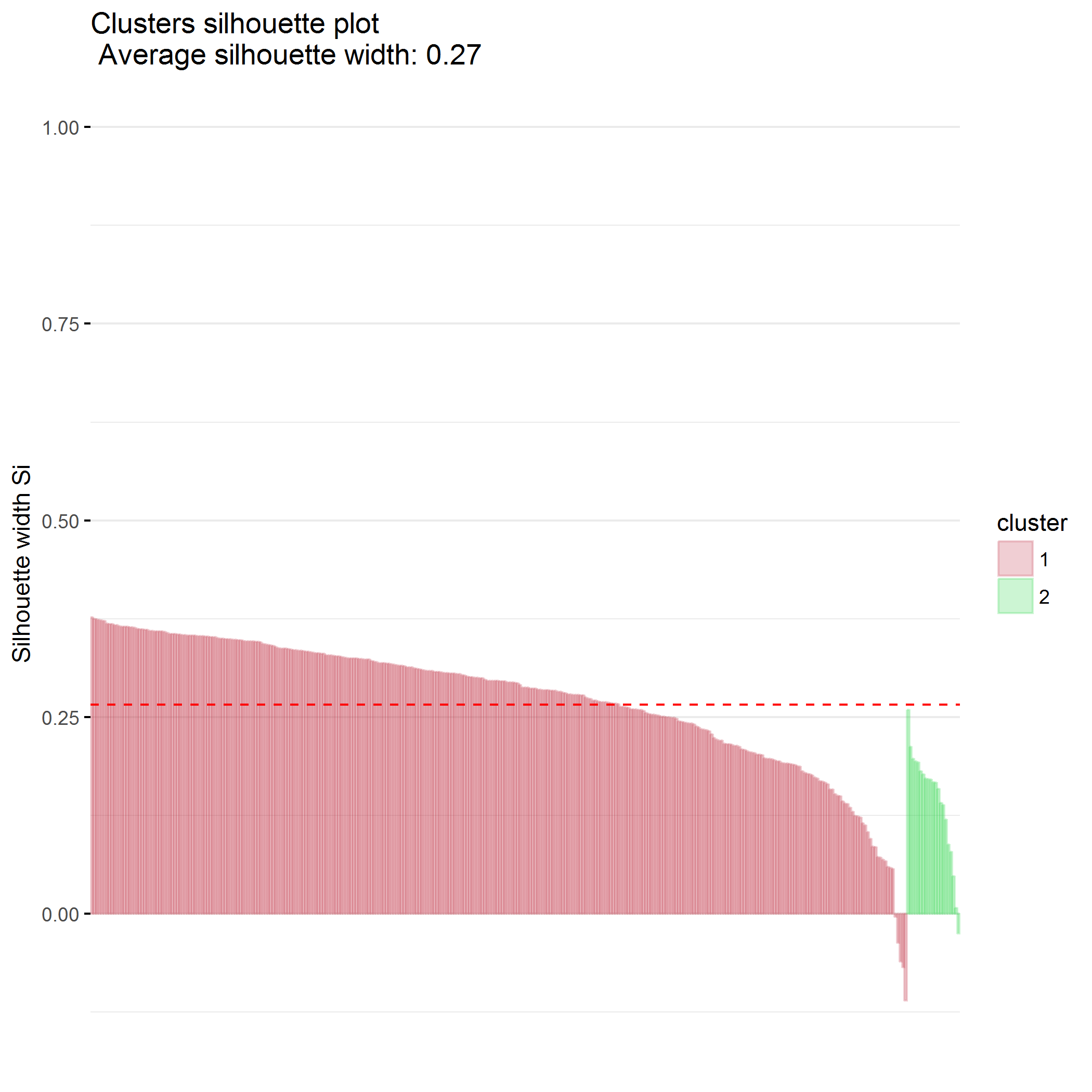
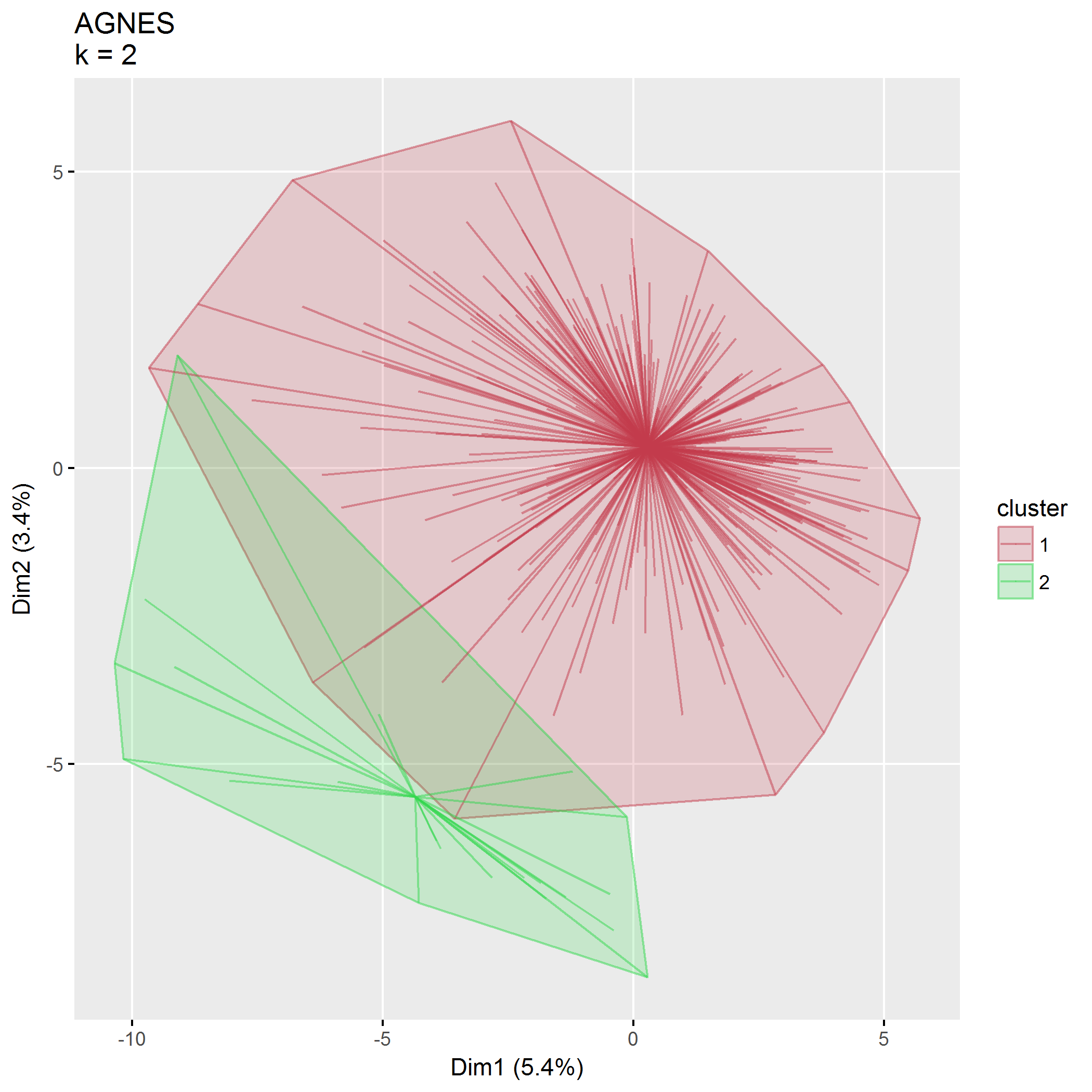
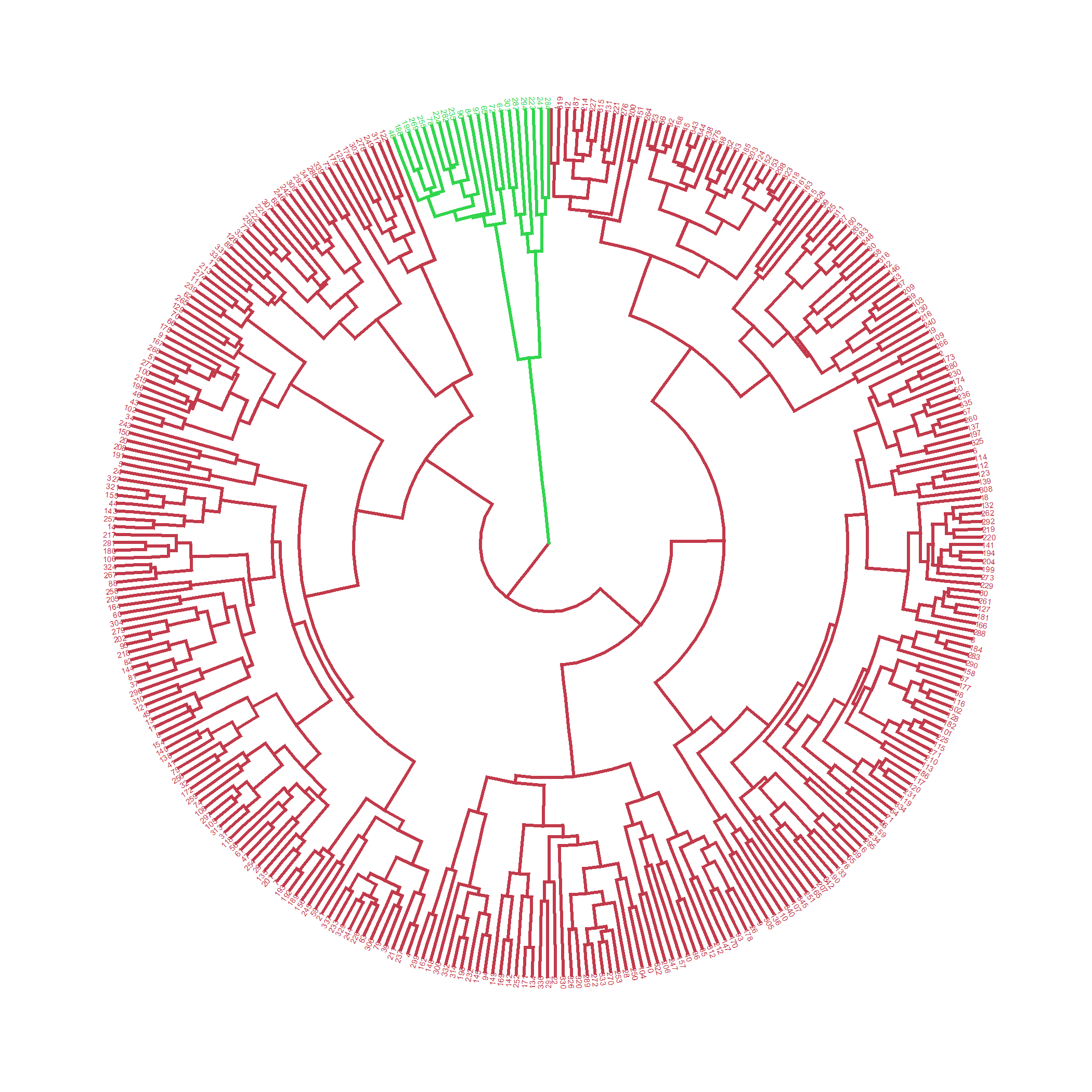
The closer the value of the correlation coefficient is to 1, the more accurately the clustering solution reflects your data. Values above 0.75 are felt to be good.

Agglomerative coeffficient using the Ward method is 0.887.

### clusters

## cluster size ave.sil.width  
## 1 1 324 0.27  
## 2 2 21 0.14

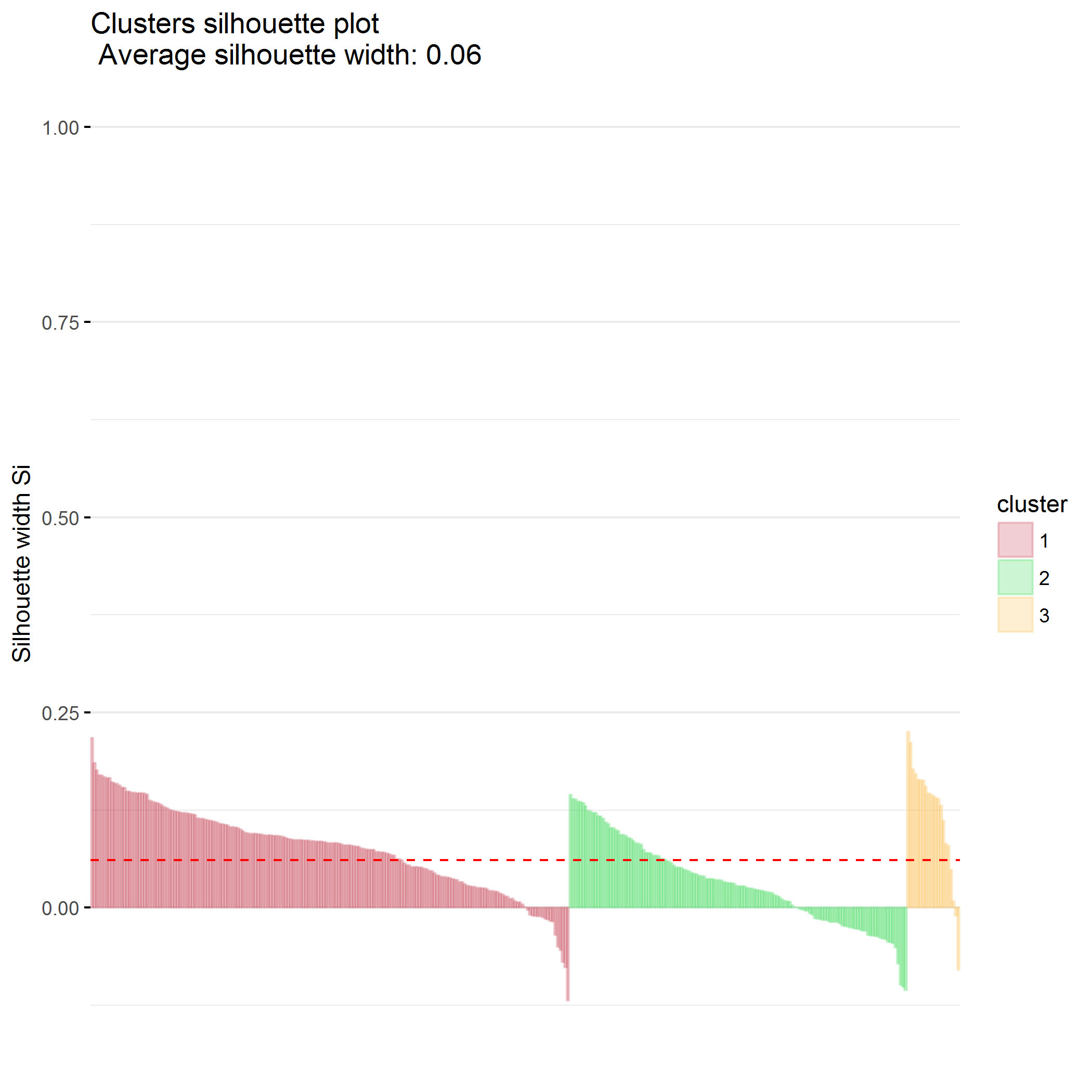
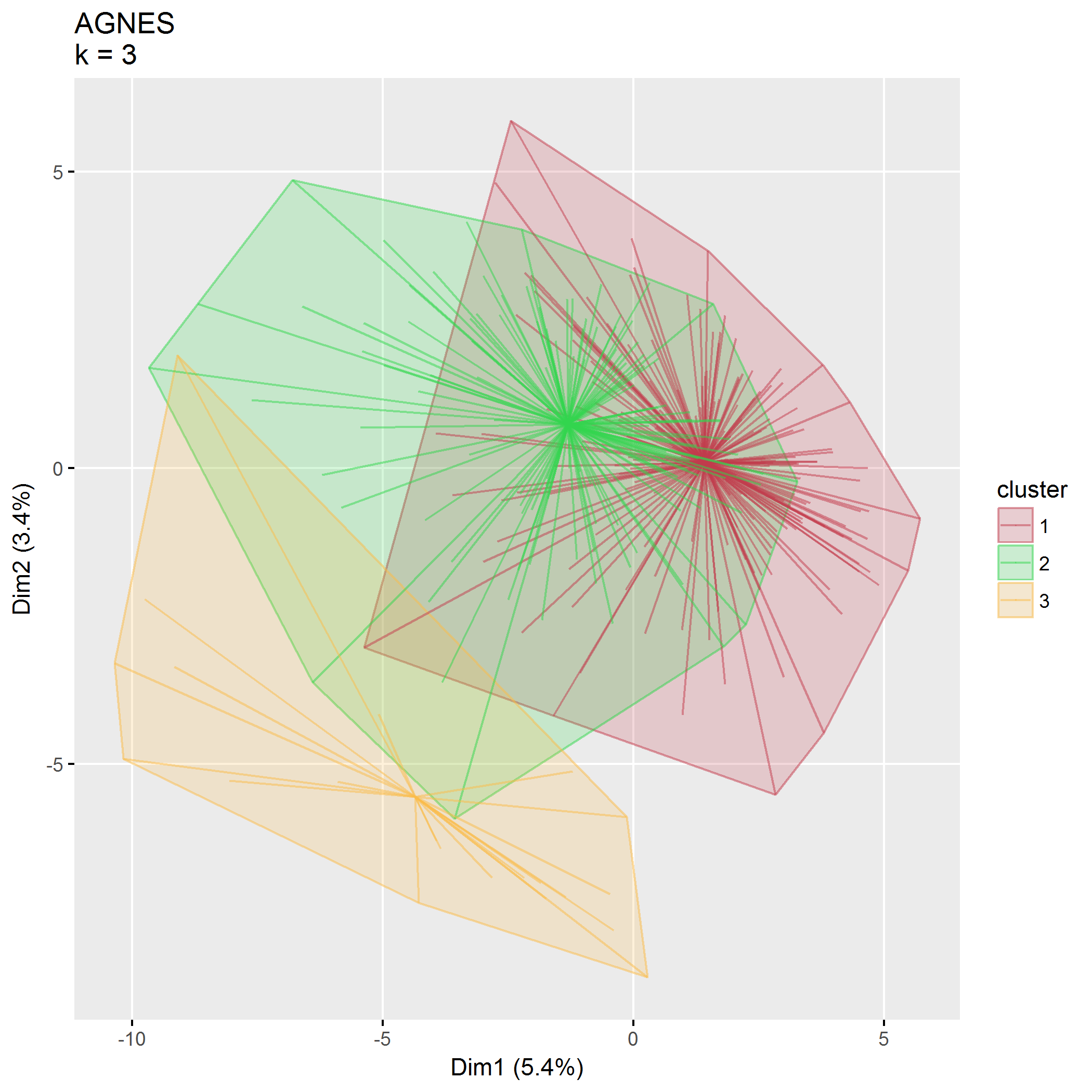
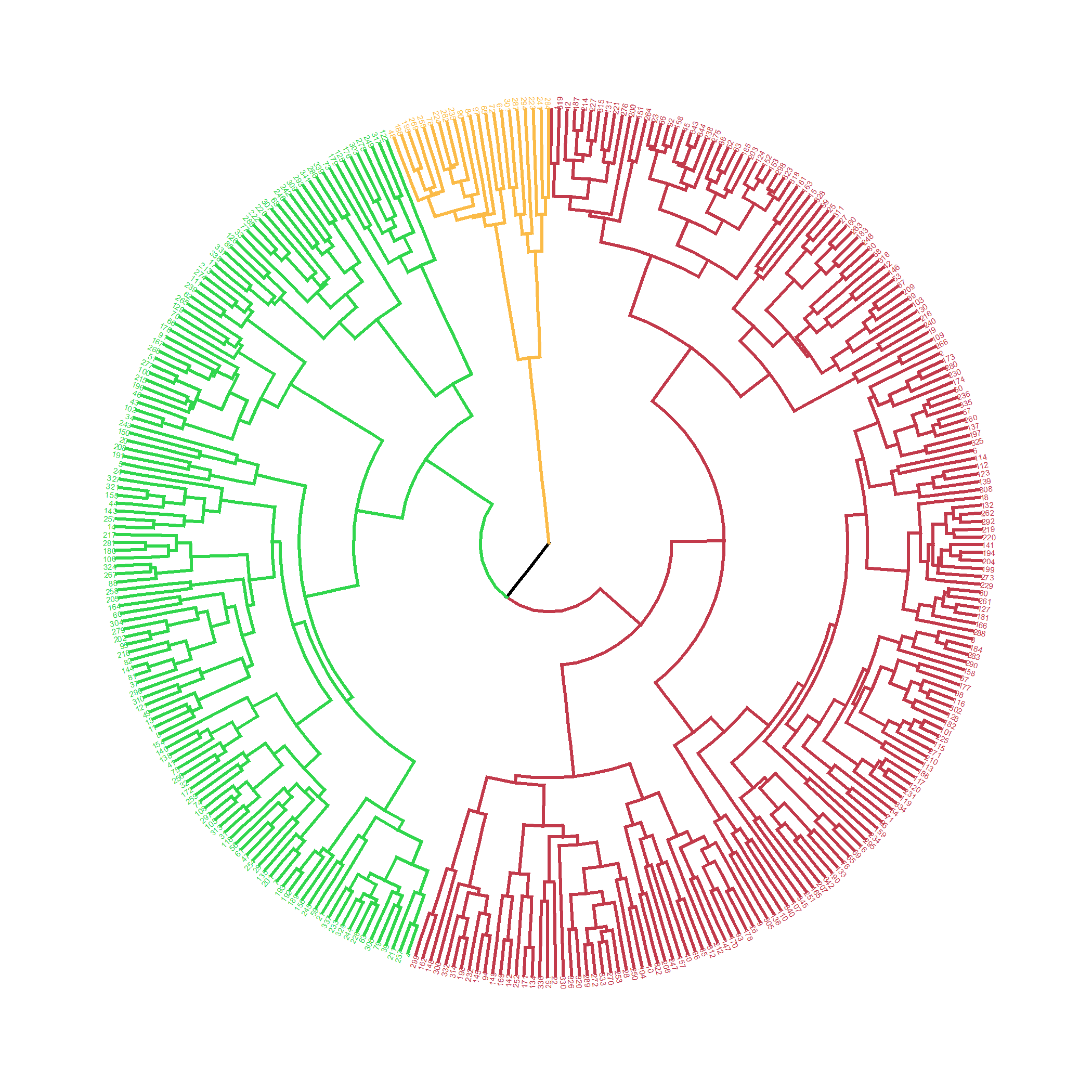
## .  
## 1 2   
## 324 21



### clusters

## cluster size ave.sil.width  
## 1 1 190 0.08  
## 2 2 134 0.03  
## 3 3 21 0.12

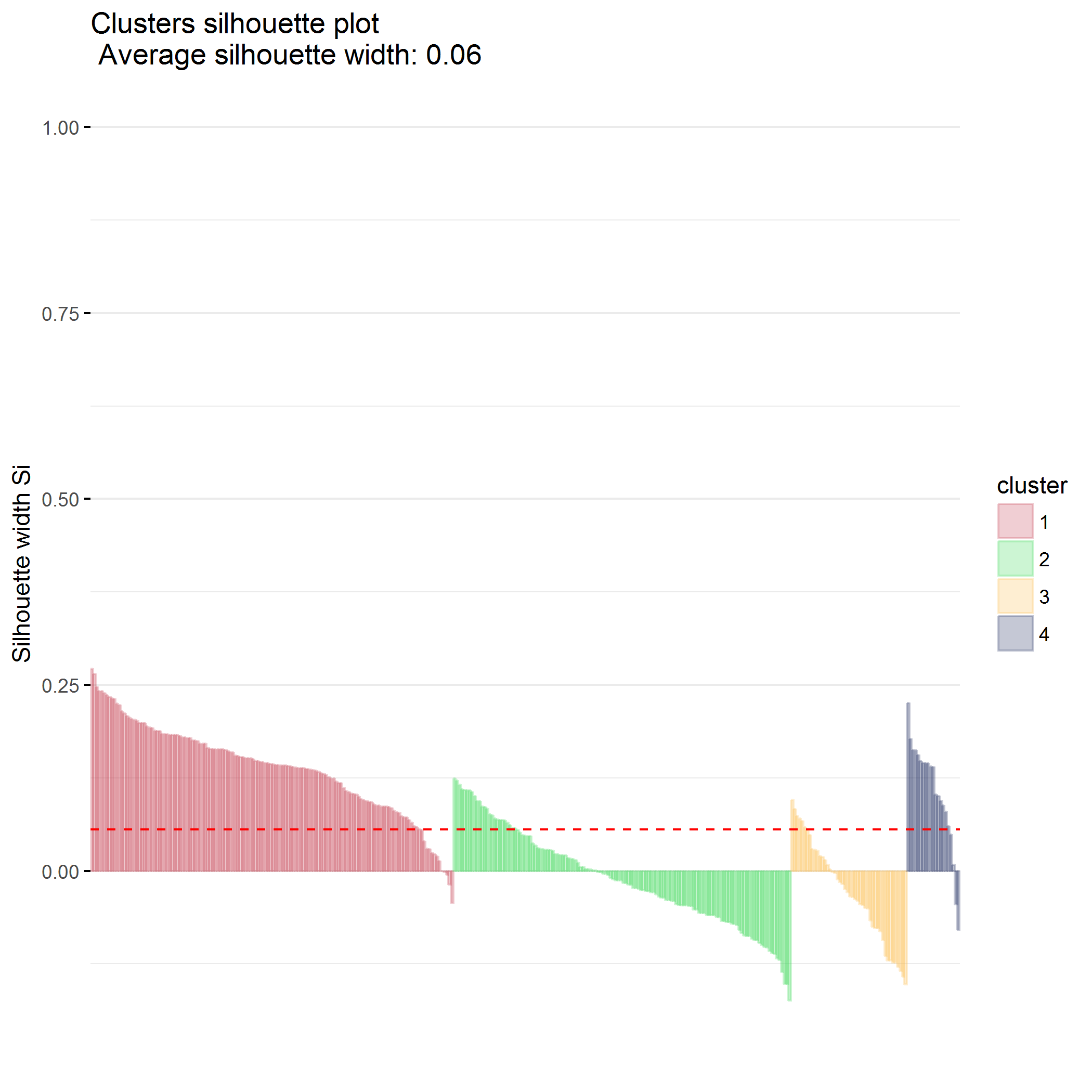
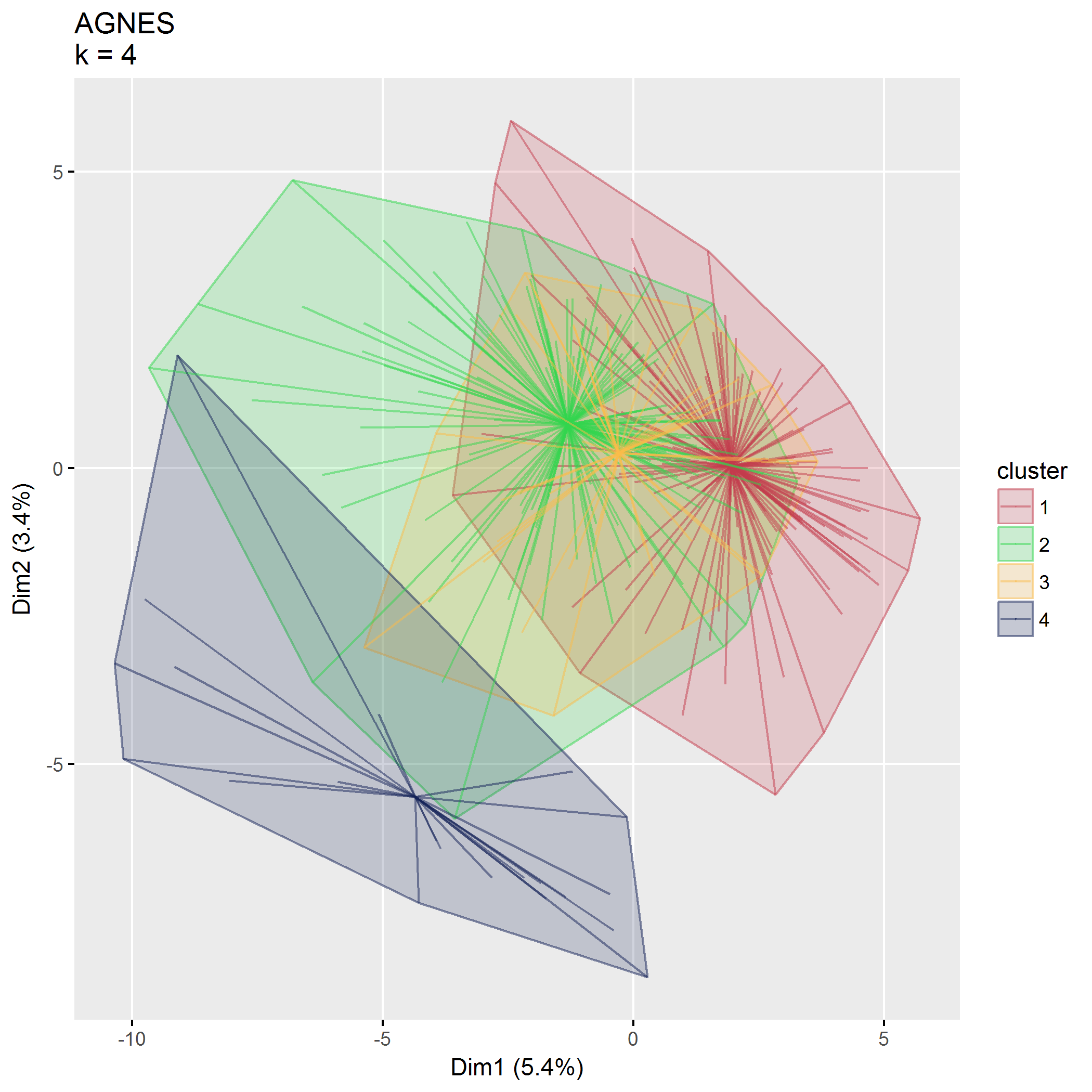
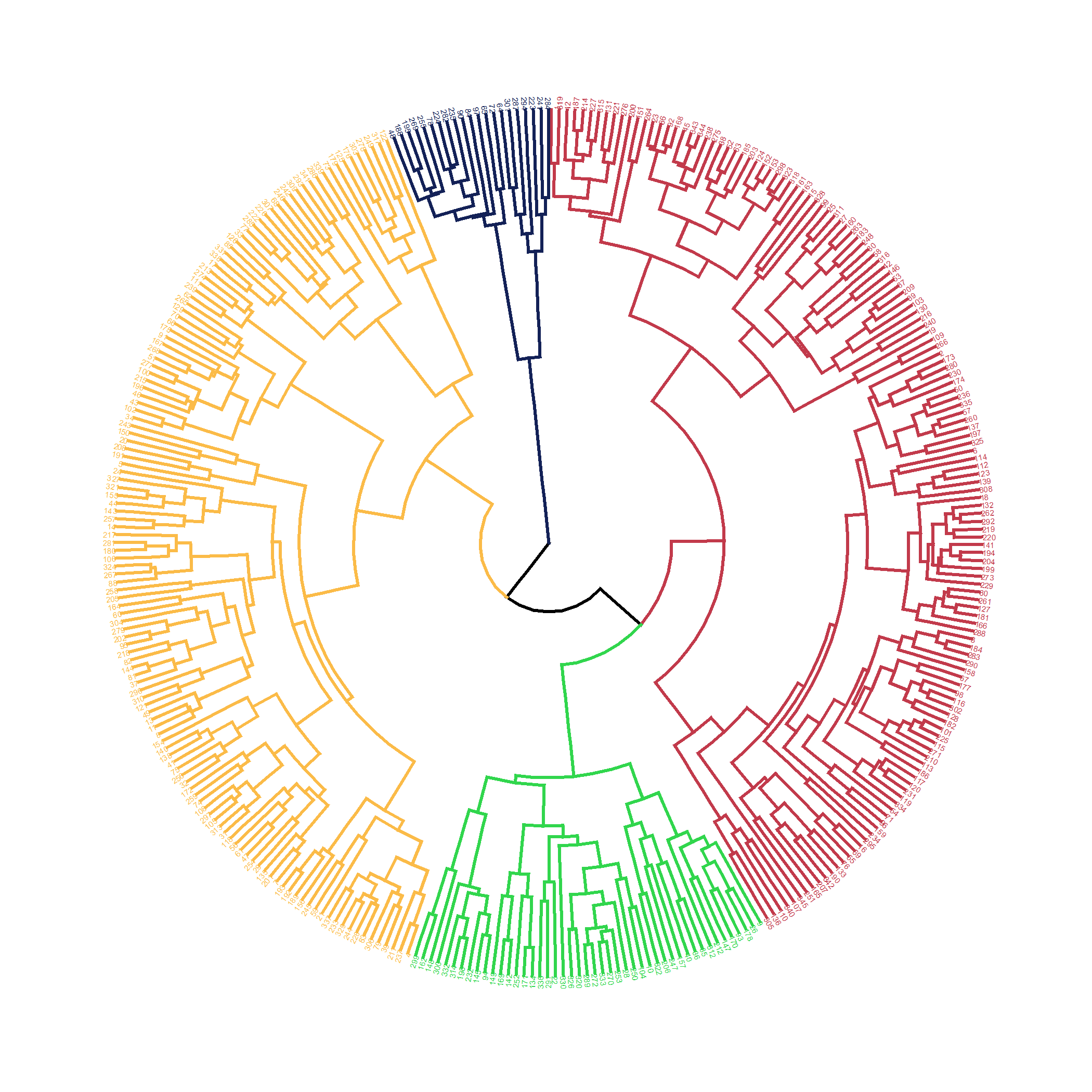
## .  
## 1 2 3   
## 190 134 21



### clusters

## cluster size ave.sil.width  
## 1 1 144 0.14  
## 2 2 134 -0.01  
## 3 3 46 -0.03  
## 4 4 21 0.10

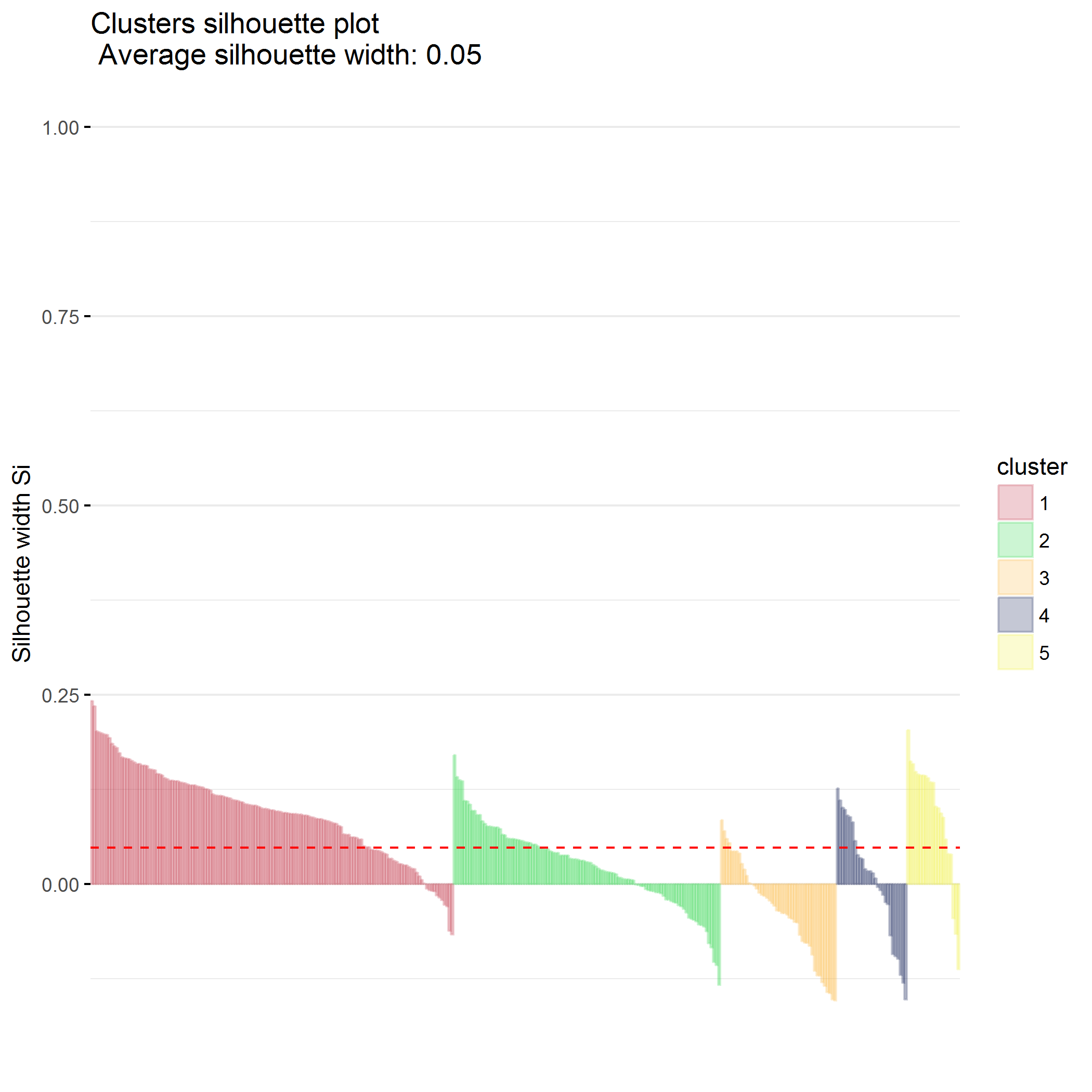
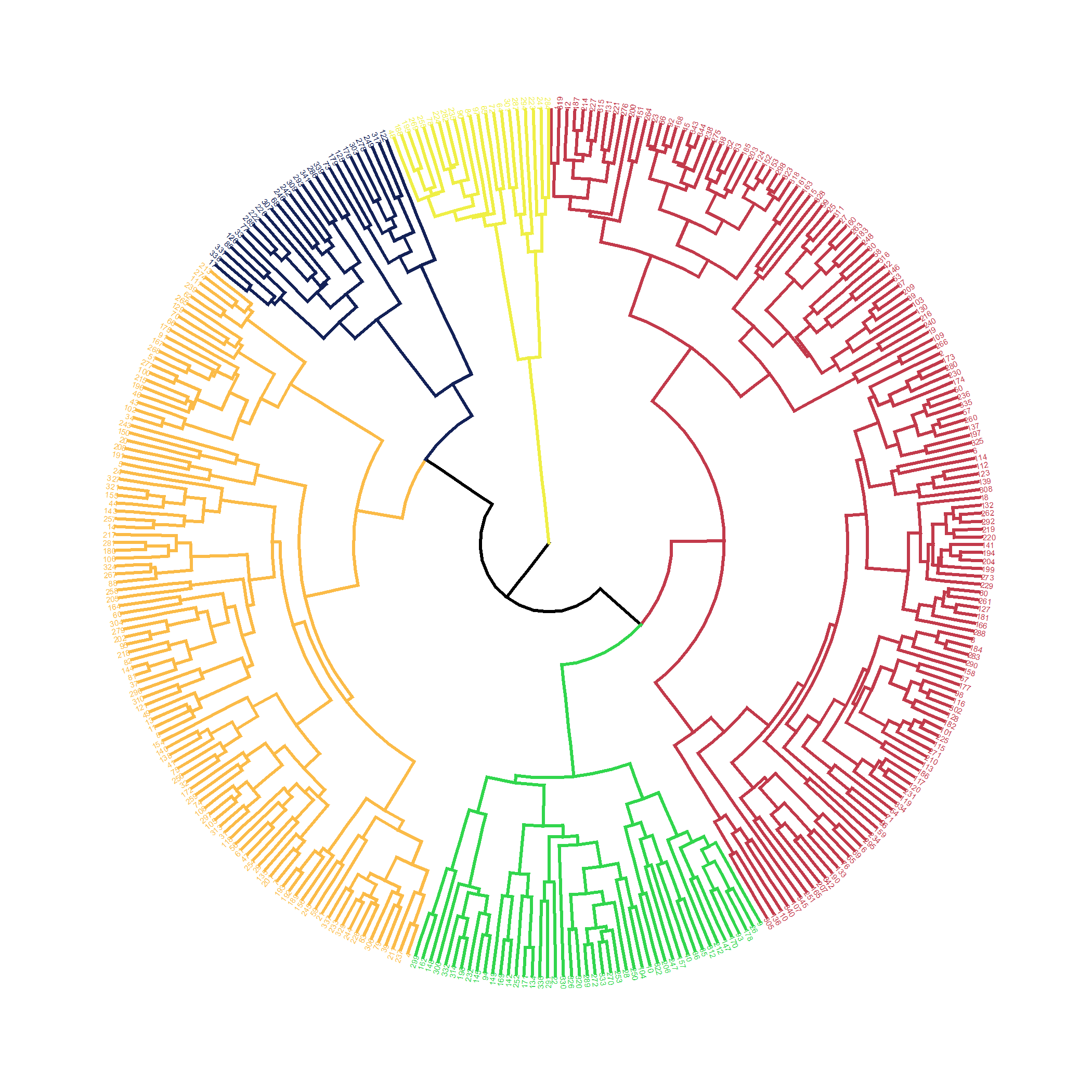
## .  
## 1 2 3 4   
## 144 134 46 21



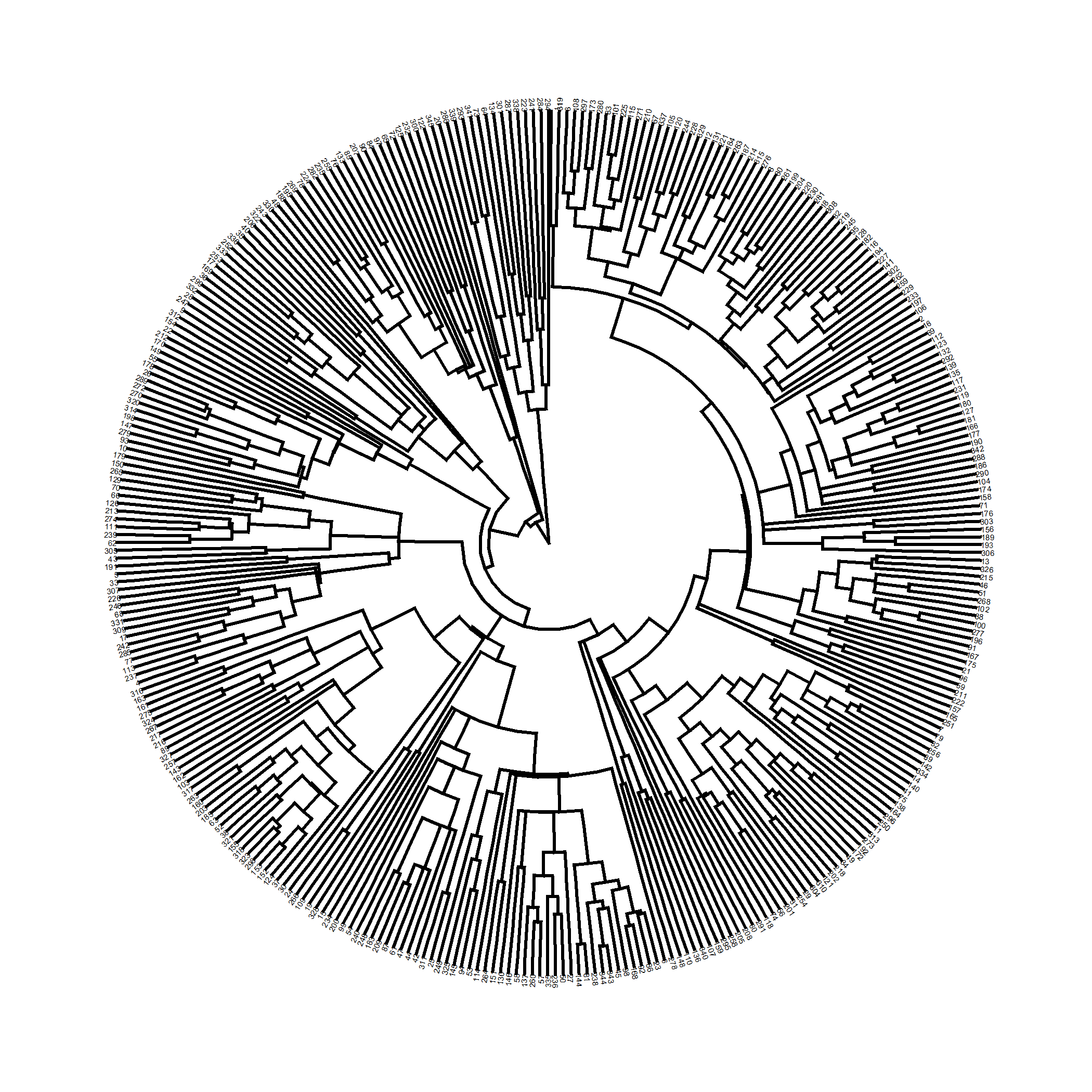
### clusters

## cluster size ave.sil.width  
## 1 1 144 0.09  
## 2 2 106 0.02  
## 3 3 46 -0.04  
## 4 4 28 0.00  
## 5 5 21 0.09

## .  
## 1 2 3 4 5   
## 144 106 46 28 21



## Divisive hierarchical clustering (DIANA)



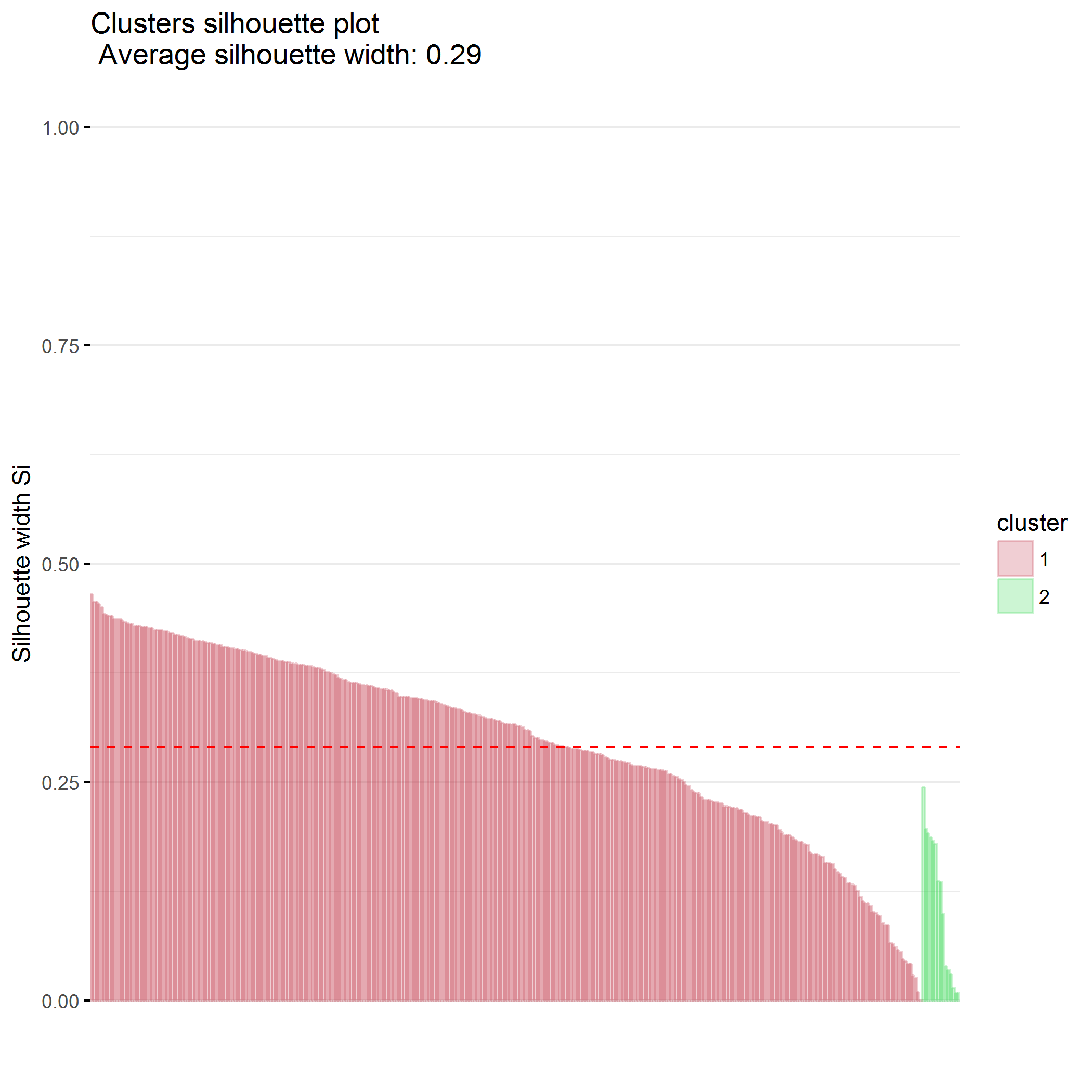
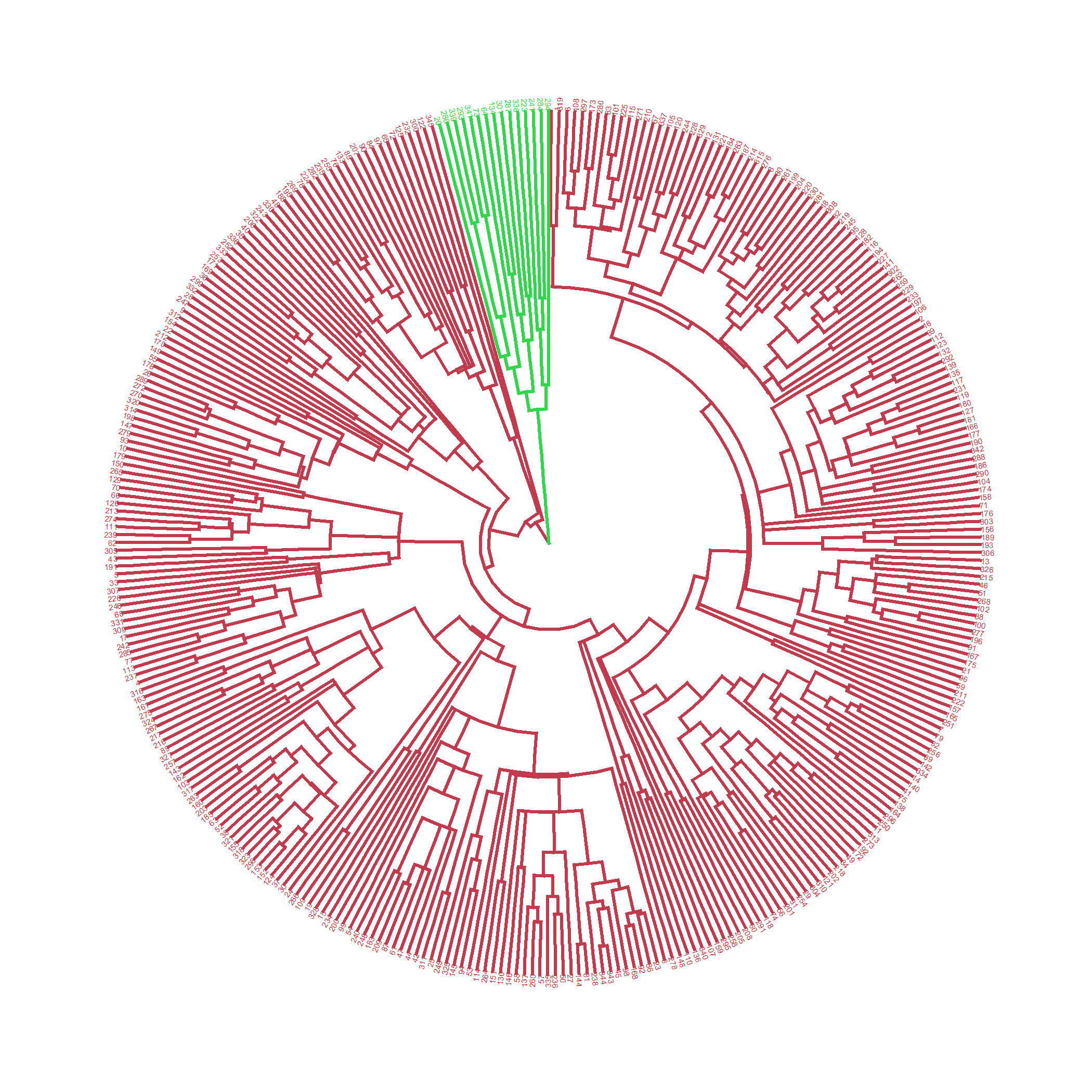
plot of chunk predictors\_diana

Divisive coeffficient is 0.727.

### clusters

## cluster size ave.sil.width  
## 1 1 330 0.30  
## 2 2 15 0.11

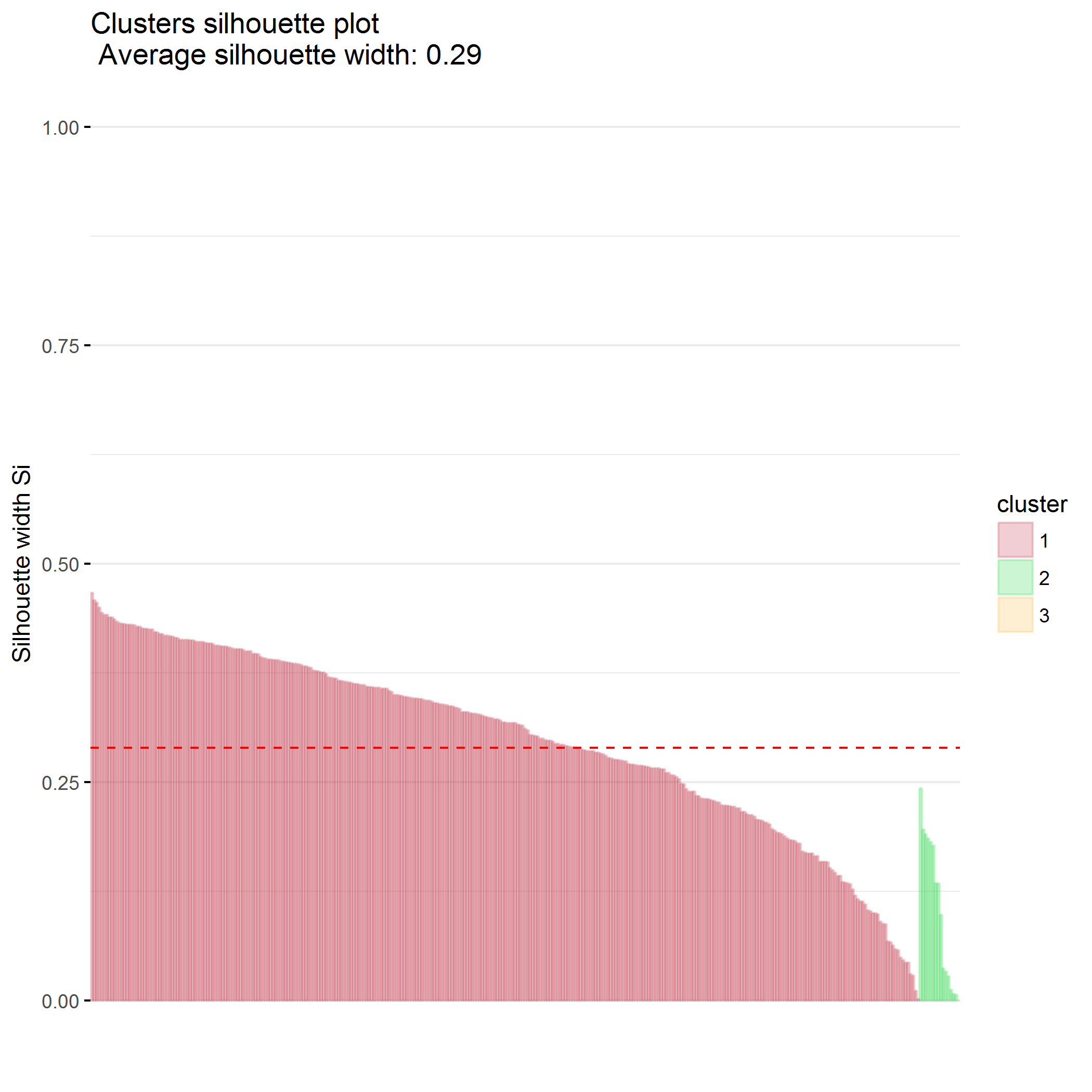
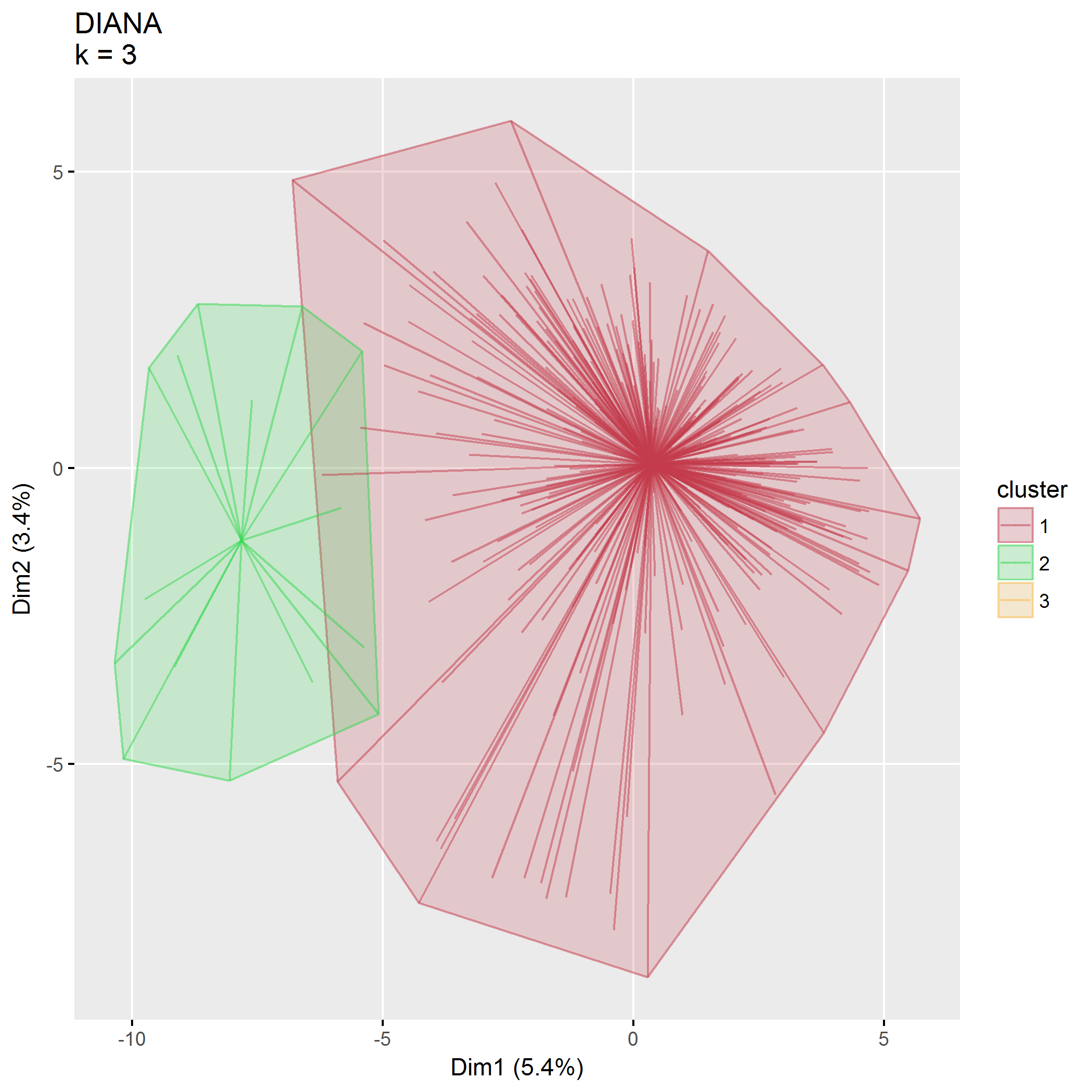
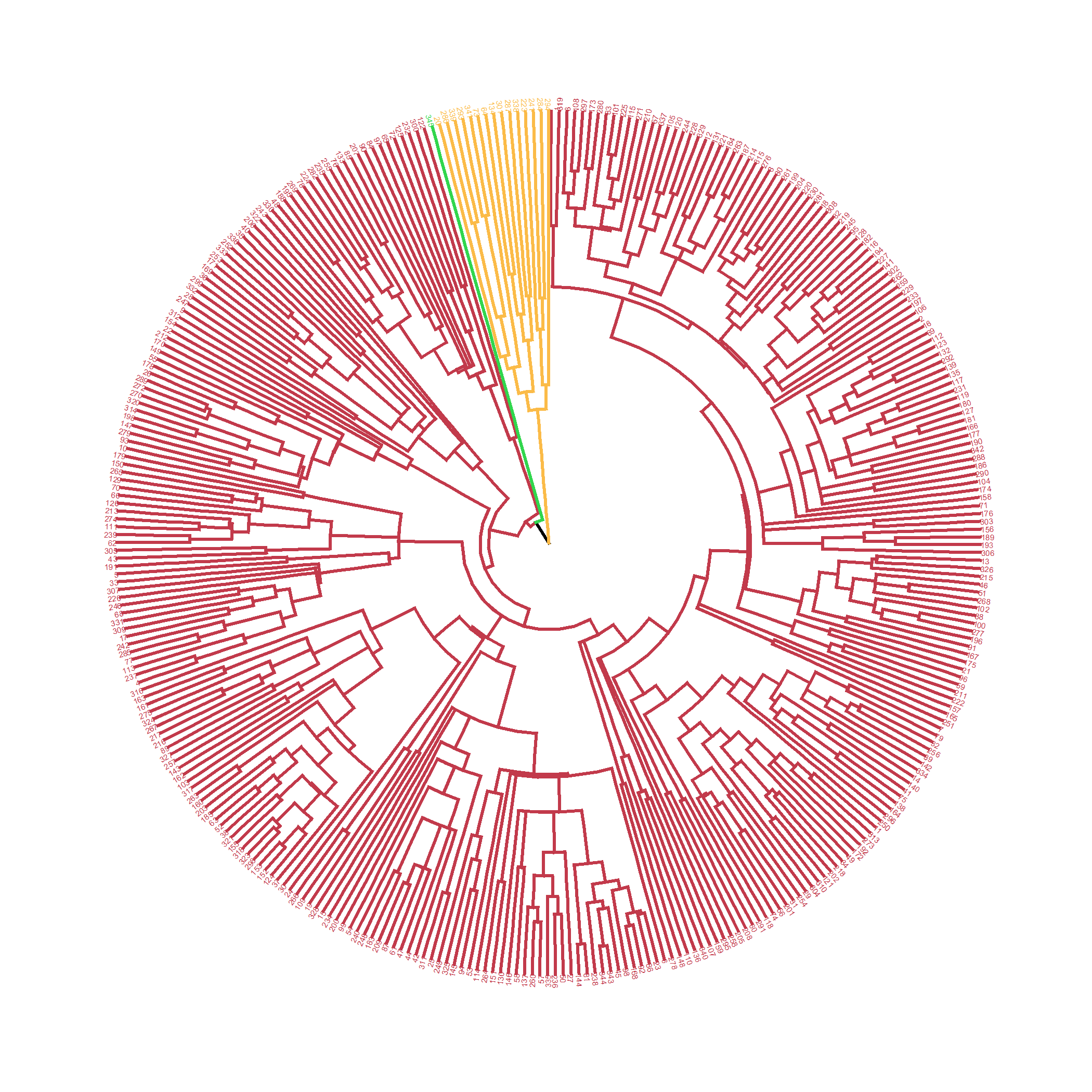
## .  
## 1 2   
## 330 15



### clusters

## cluster size ave.sil.width  
## 1 1 329 0.30  
## 2 2 15 0.11  
## 3 3 1 0.00

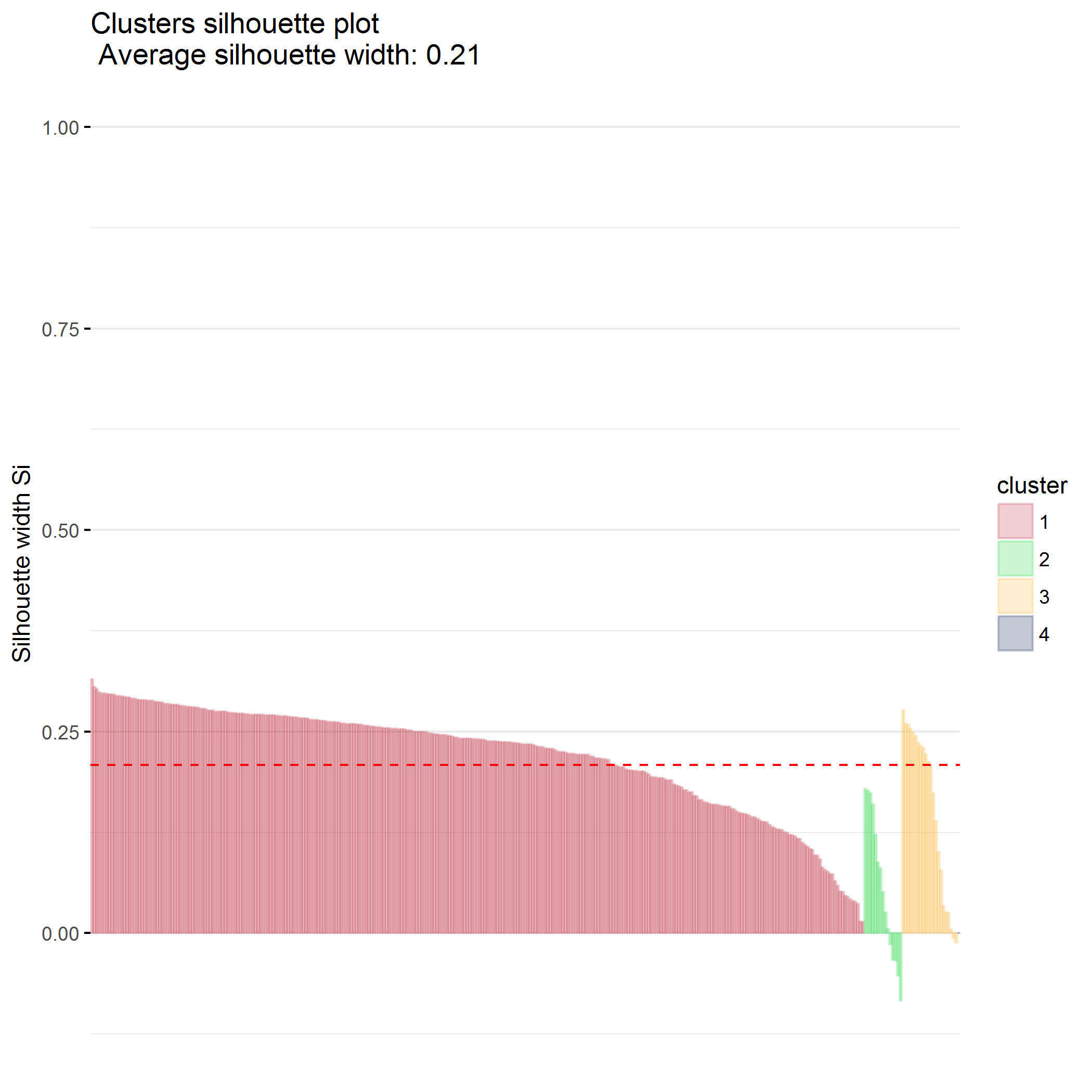
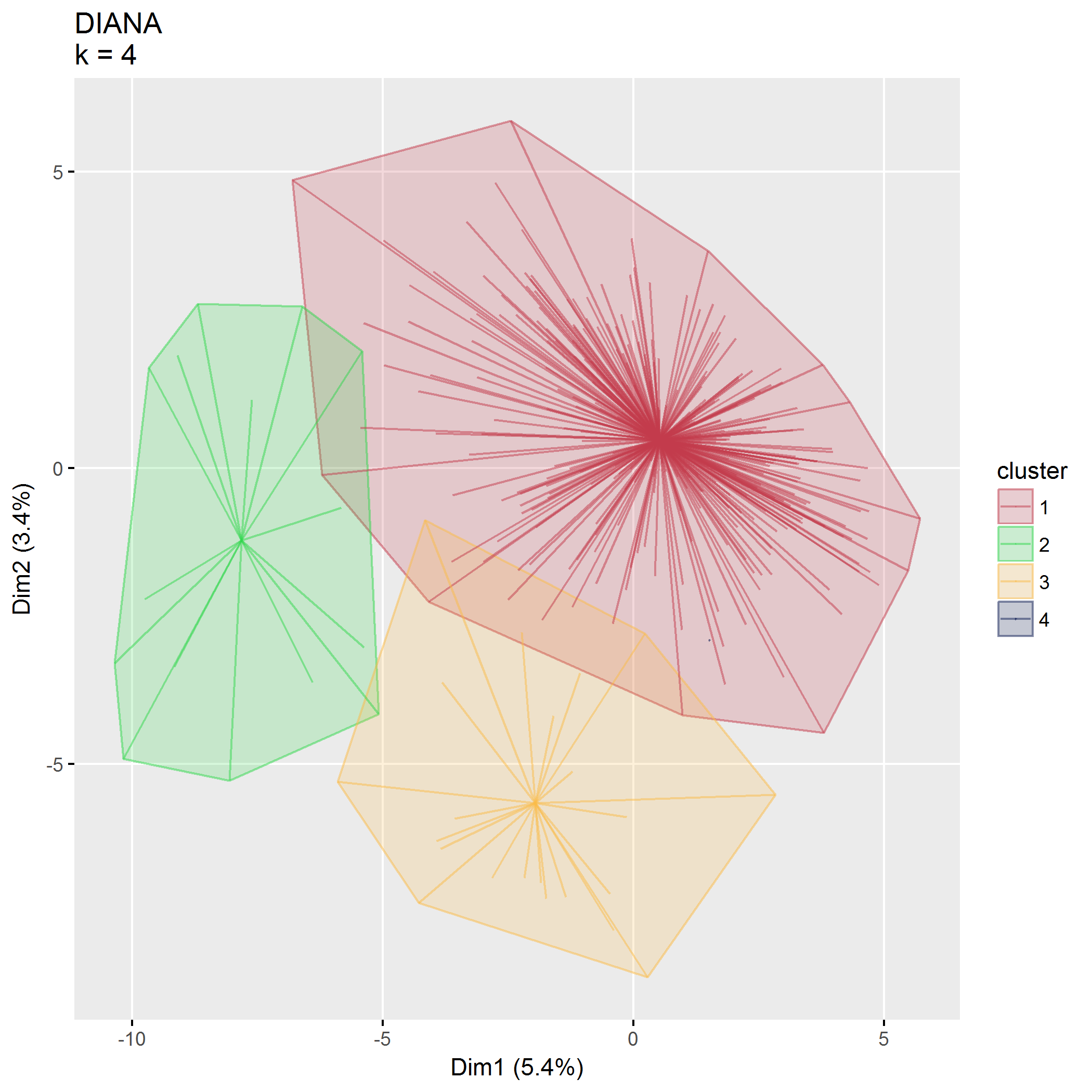
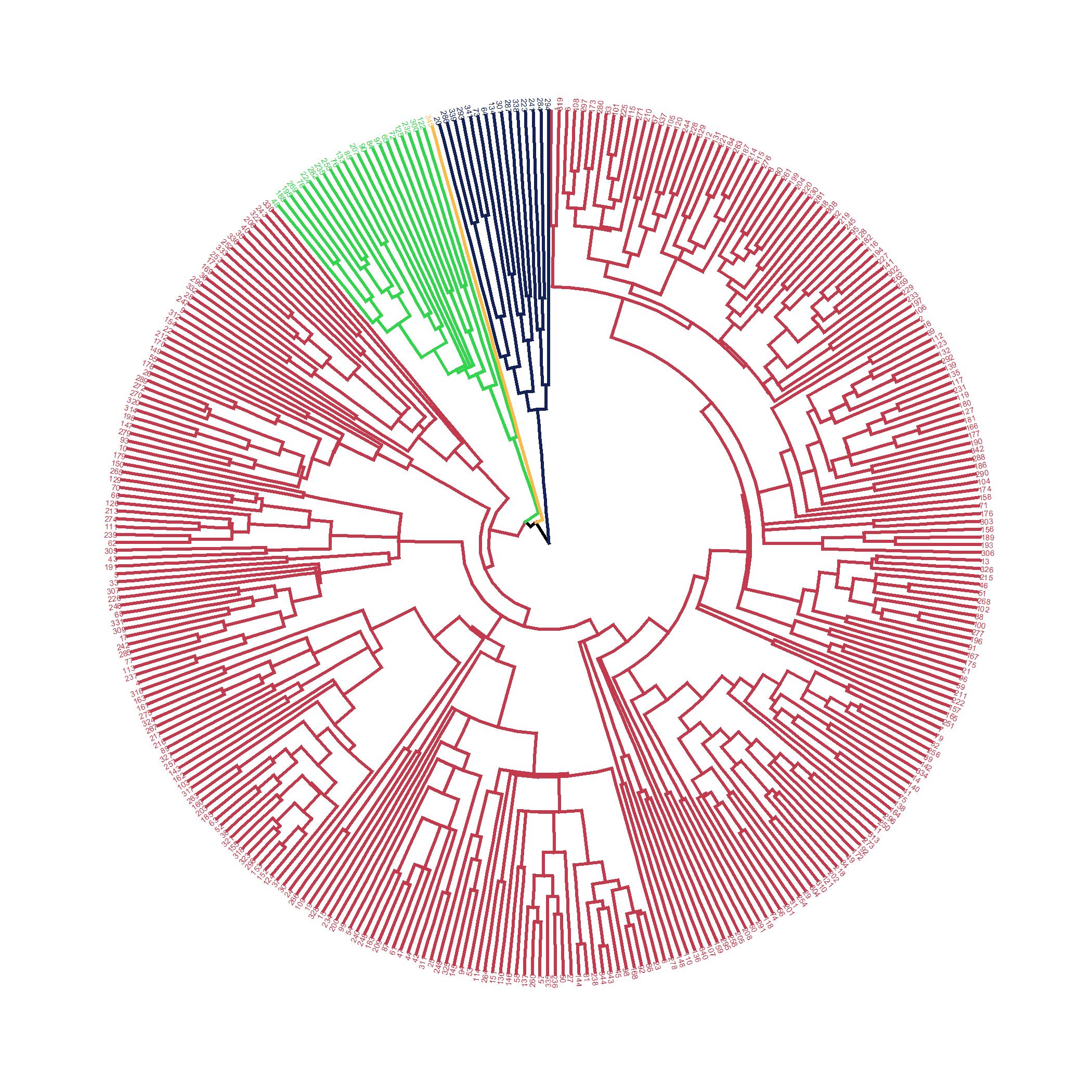
## .  
## 1 2 3   
## 329 15 1



### clusters

## cluster size ave.sil.width  
## 1 1 307 0.22  
## 2 2 15 0.06  
## 3 3 22 0.16  
## 4 4 1 0.00

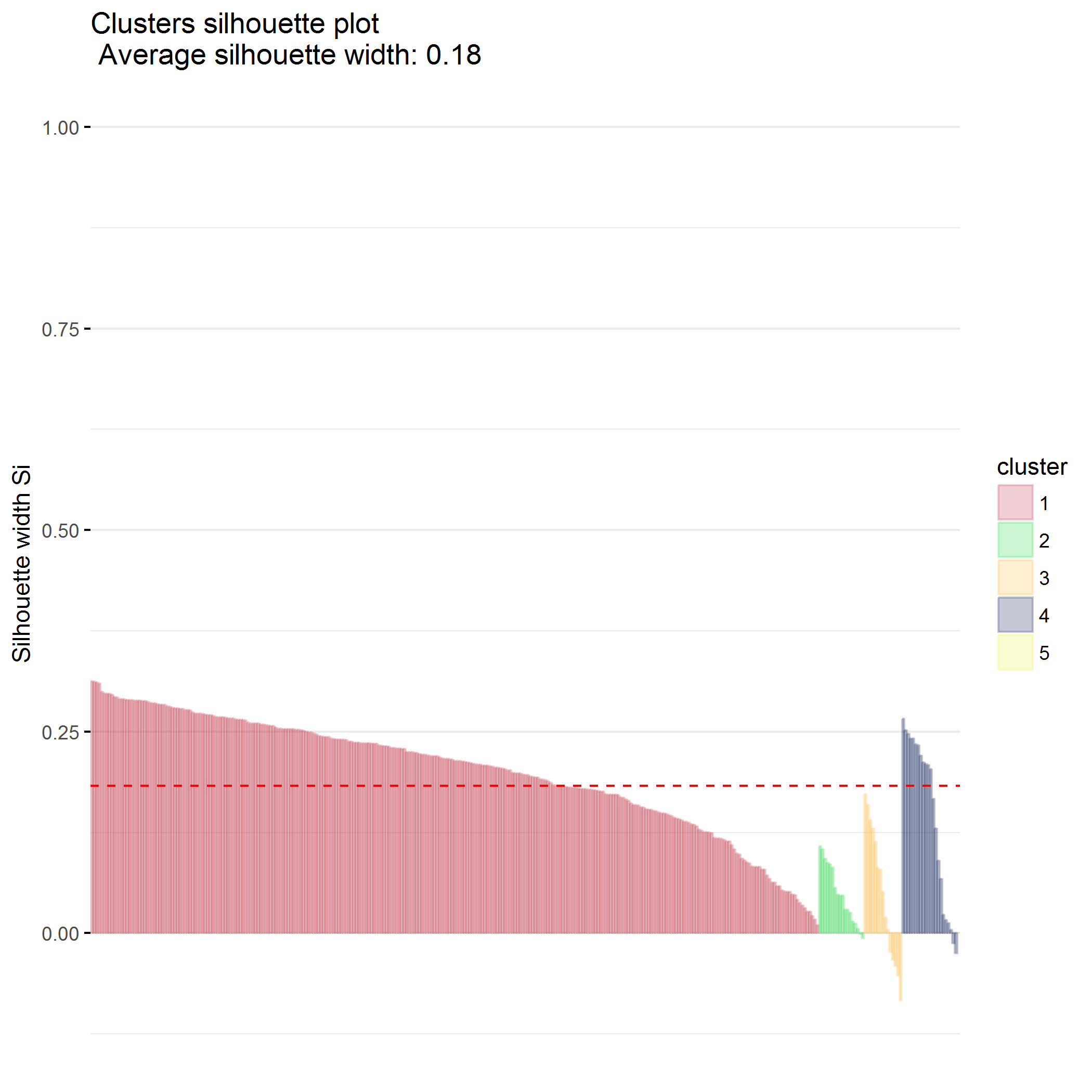
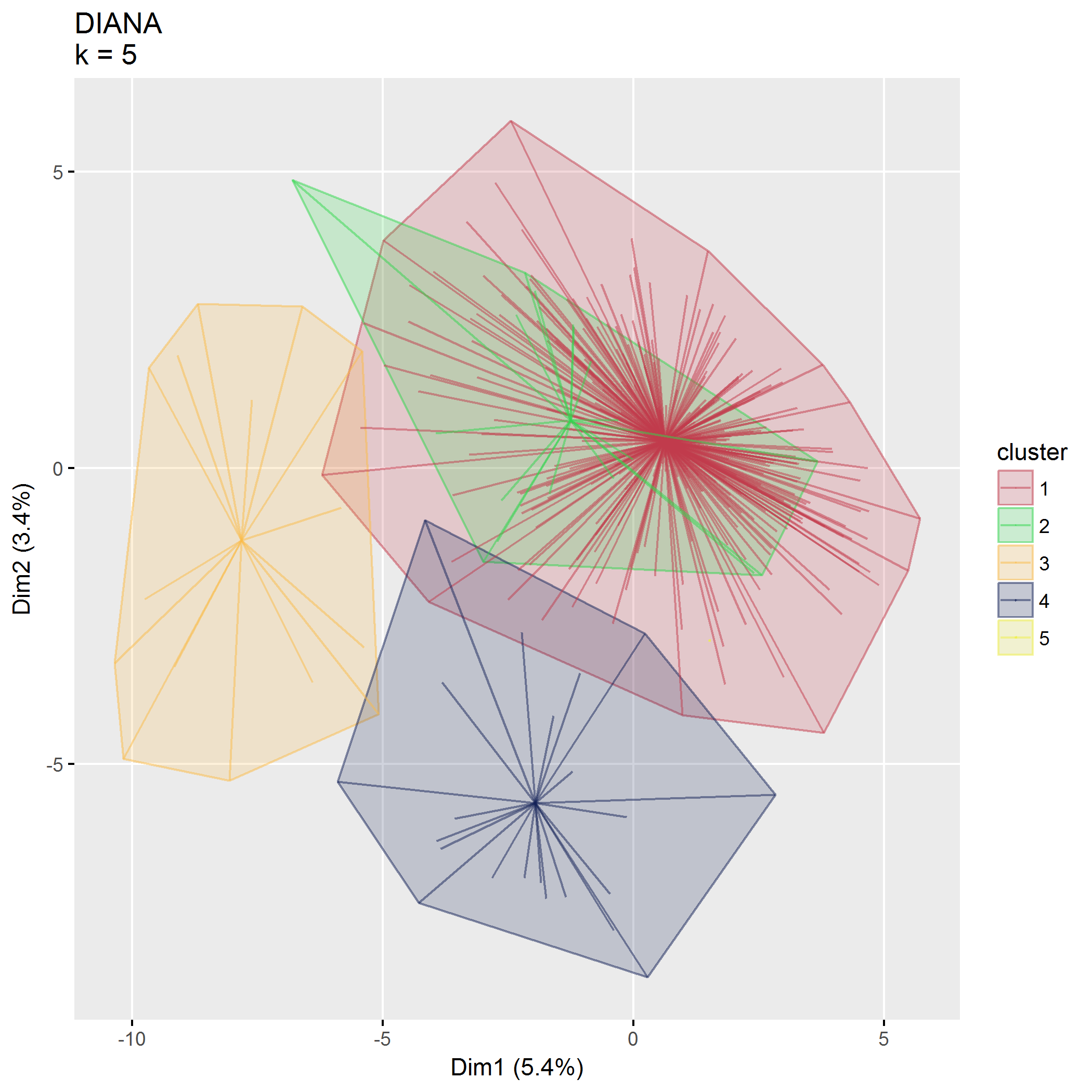
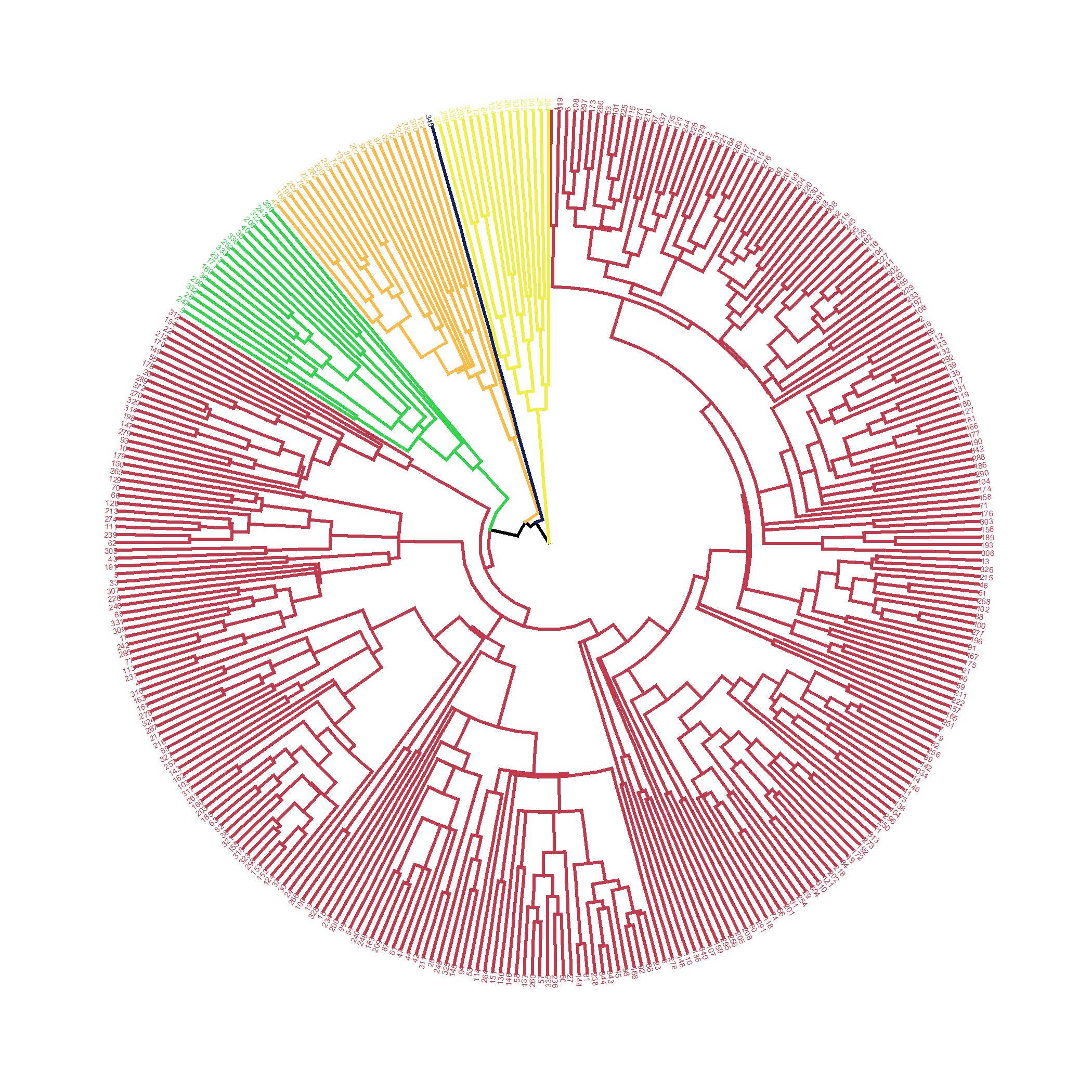
## .  
## 1 2 3 4   
## 307 15 22 1



### clusters

## cluster size ave.sil.width  
## 1 1 289 0.20  
## 2 2 18 0.05  
## 3 3 15 0.05  
## 4 4 22 0.15  
## 5 5 1 0.00

## .  
## 1 2 3 4 5   
## 289 18 15 22 1



## Compare

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

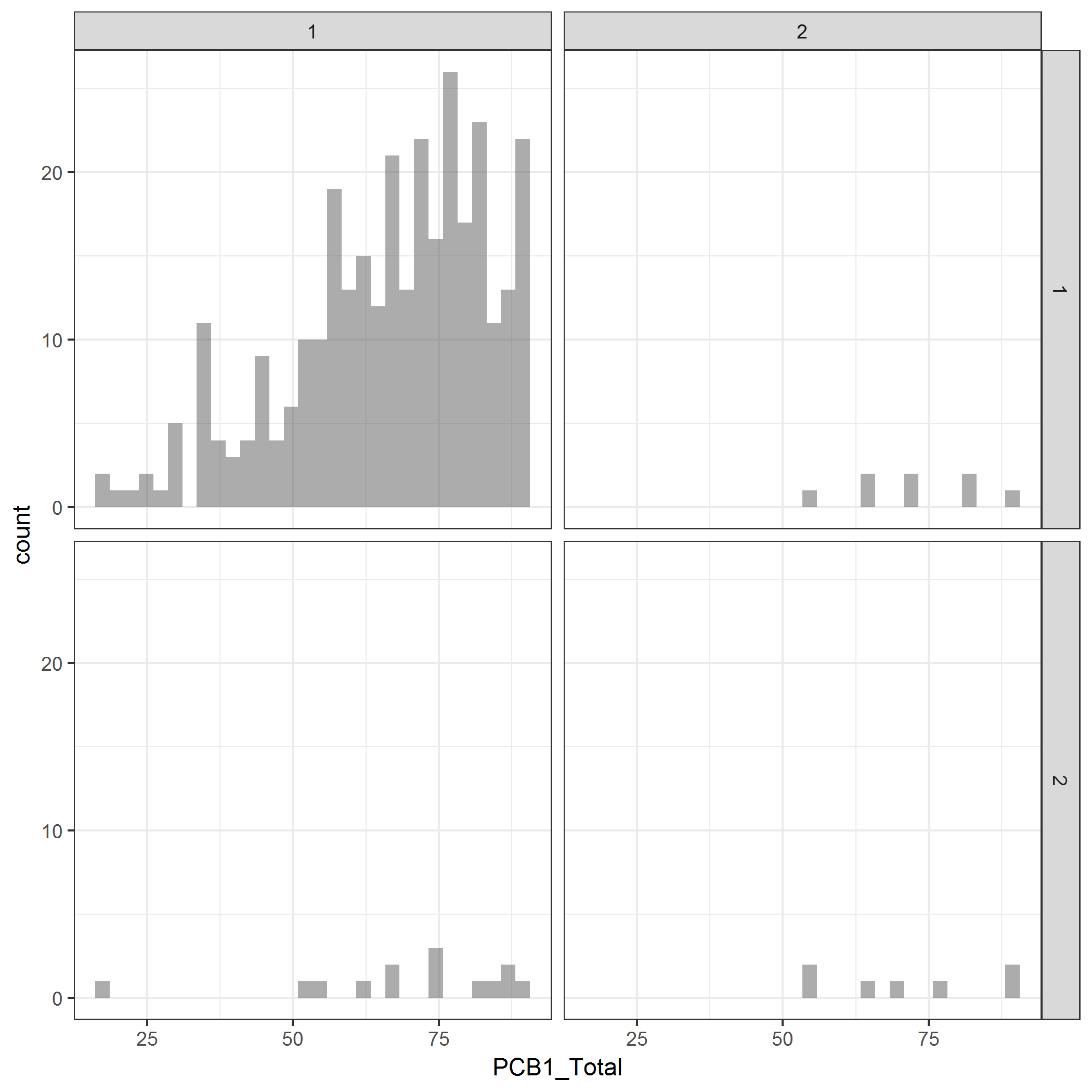
*Do not evaluate*

## Examine clusters

* clusters seems optimal using AGNES
* clusters seems optimal using DIANA

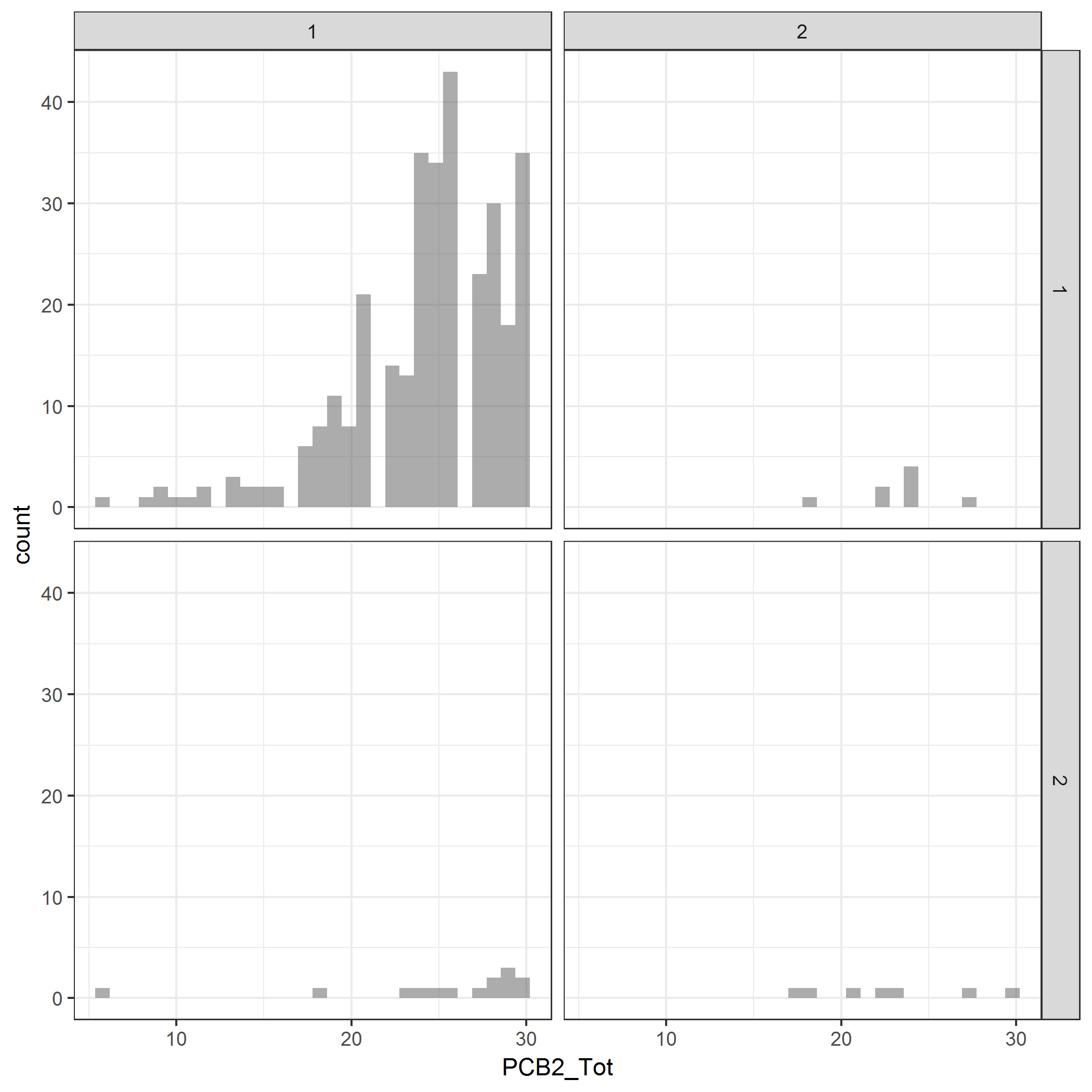
##   
## 1 2  
## 1 316 8  
## 2 14 7

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



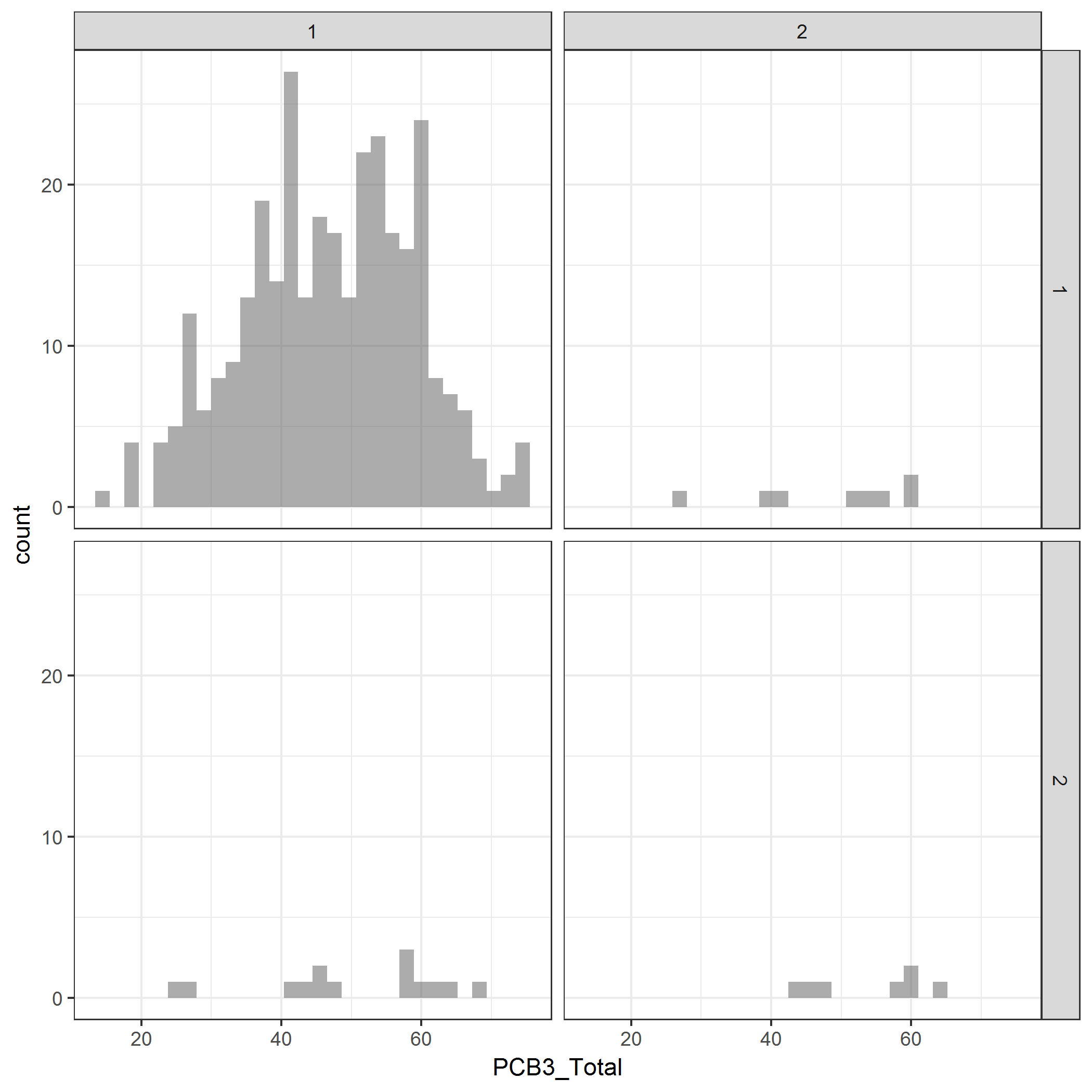
plot of chunk predictors\_PCB

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



plot of chunk predictors\_PCB

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

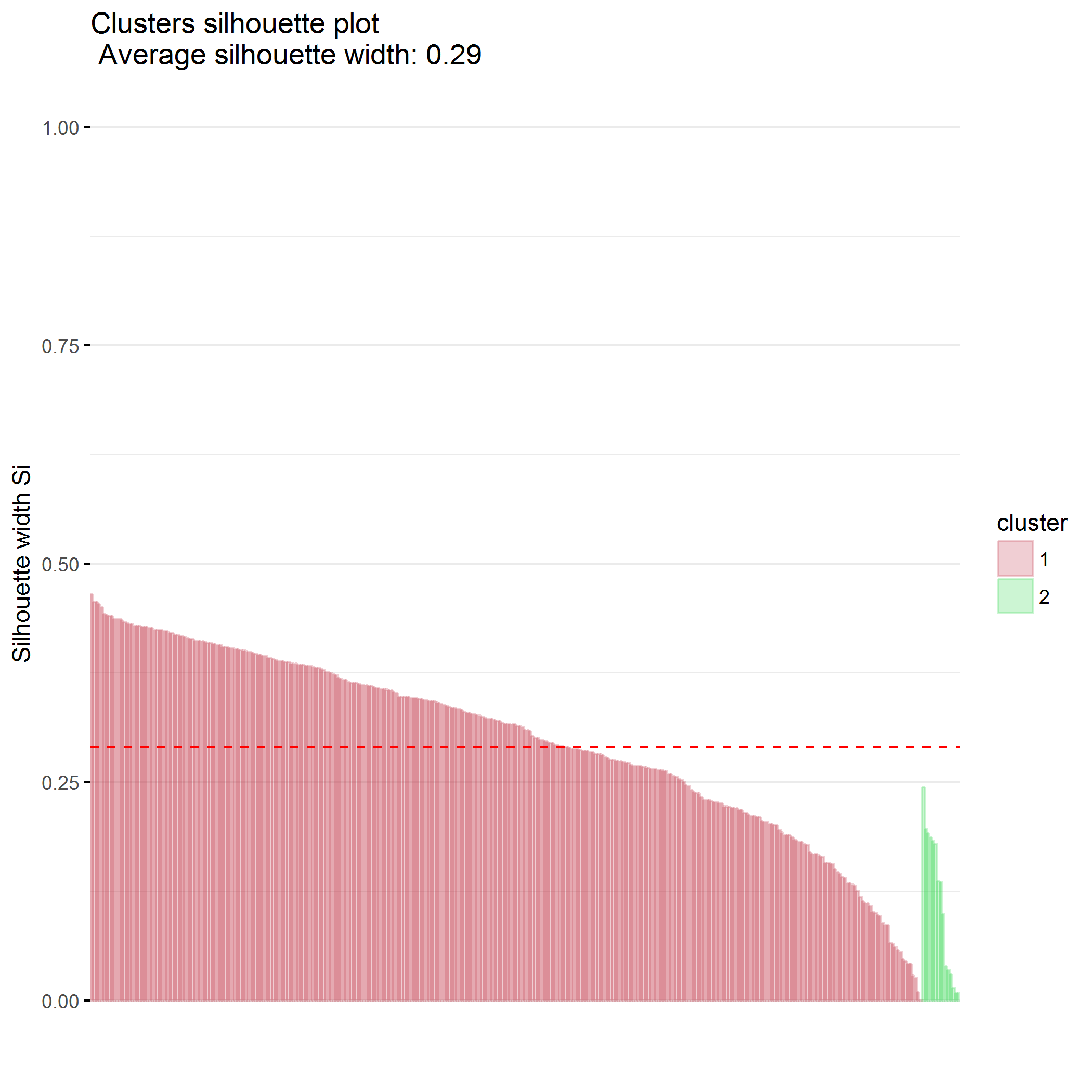
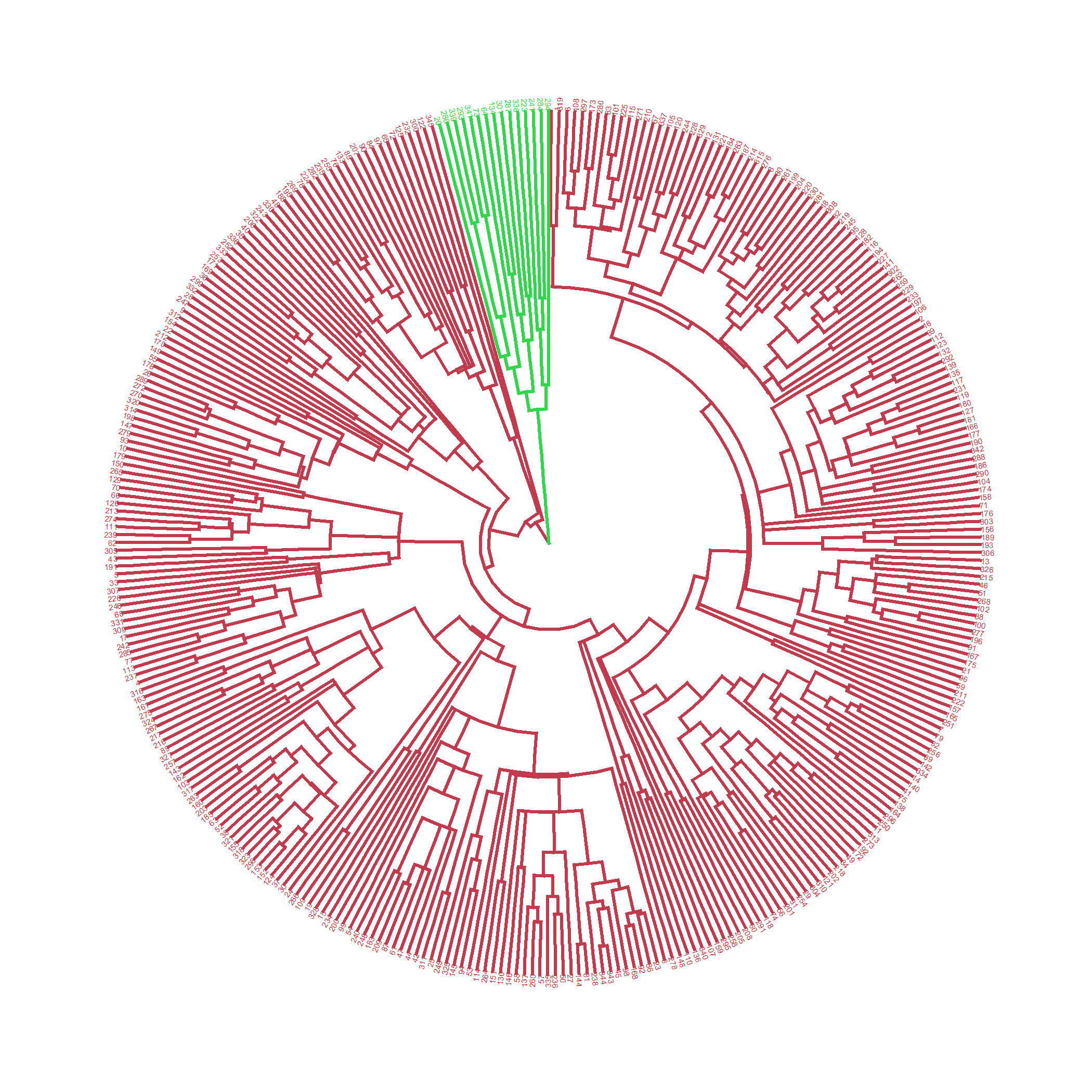


plot of chunk predictors\_PCB

## Final cluster identification

Use DIANA,

## .  
## 1 2   
## 330 15



<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

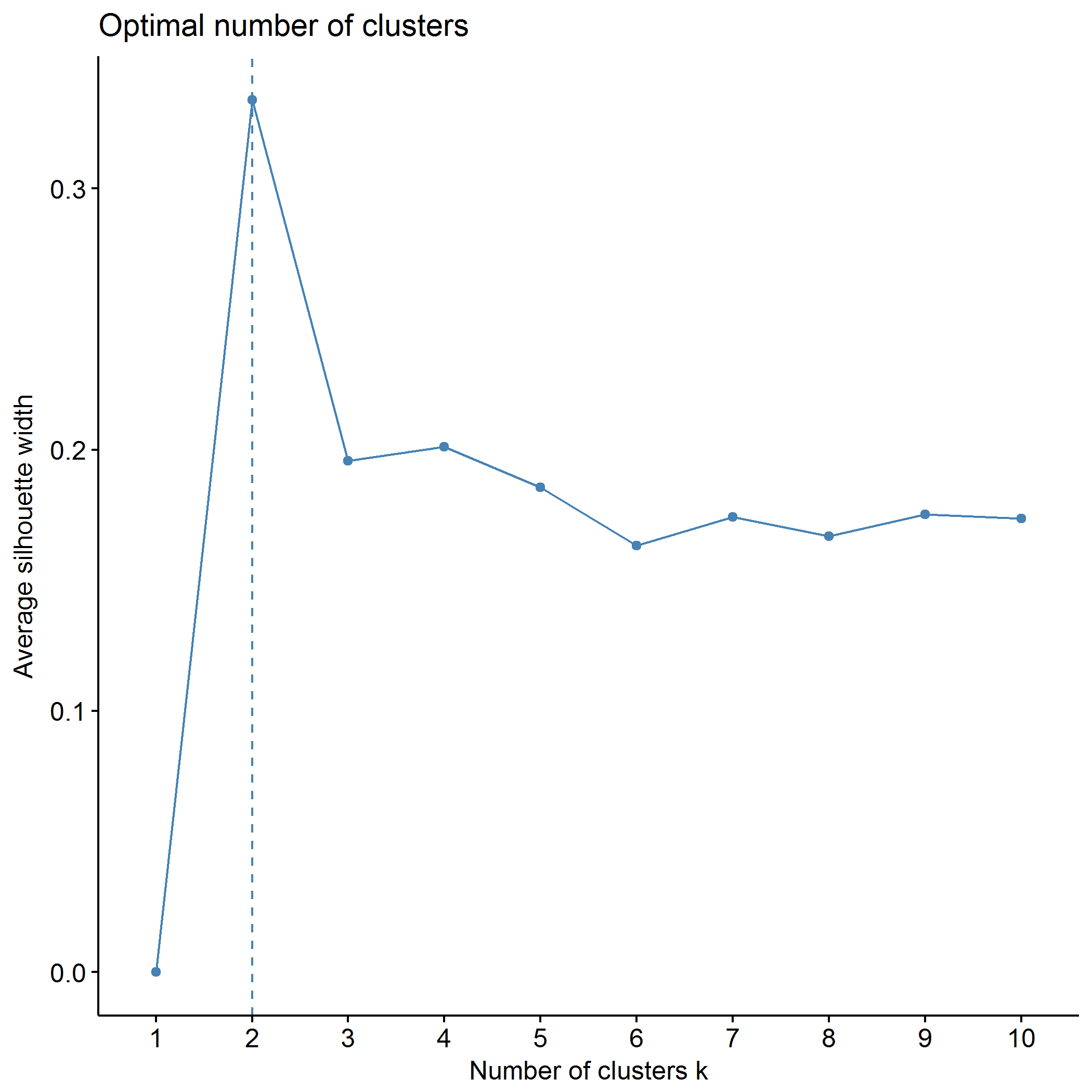
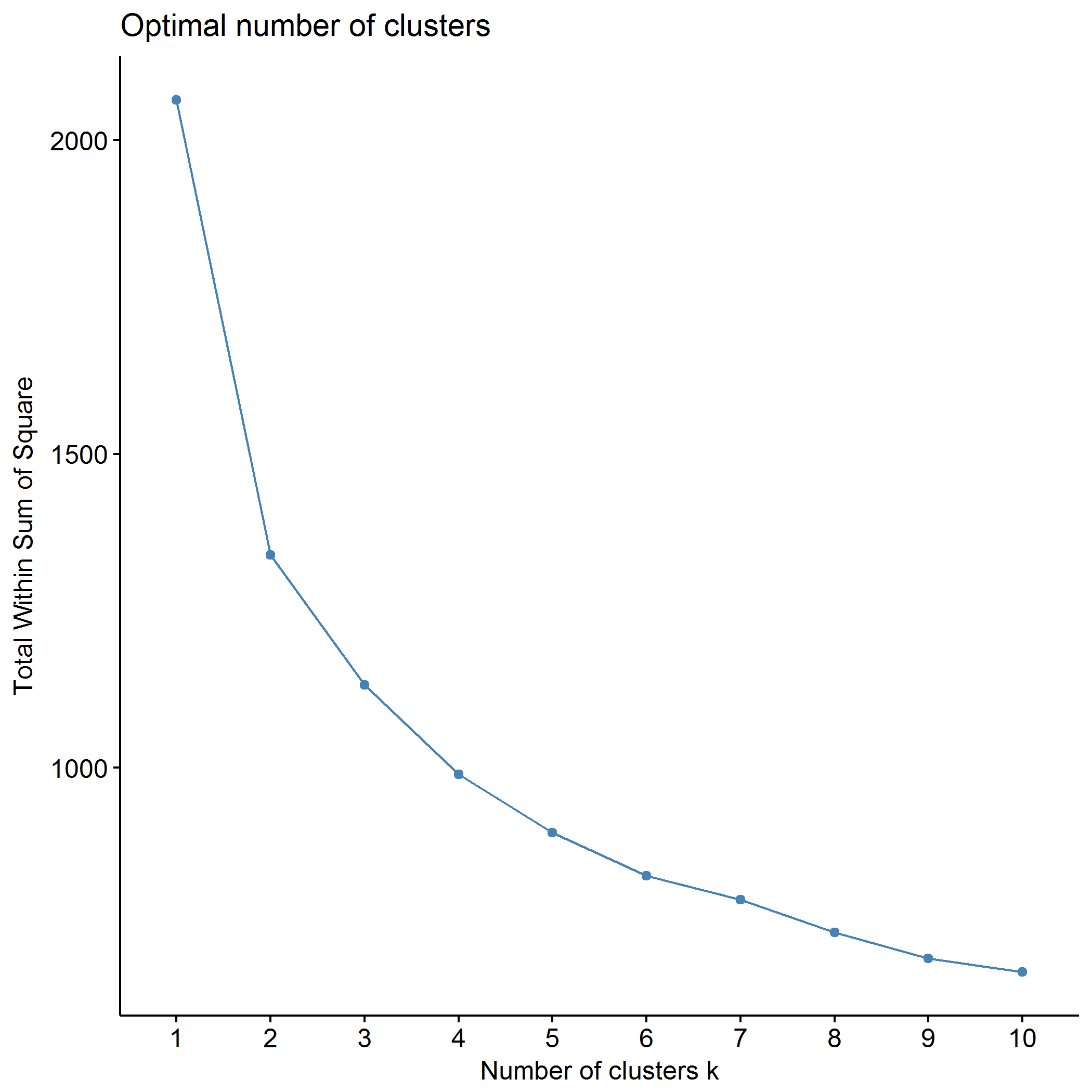
## [1] 345 54

## [1] 345 6

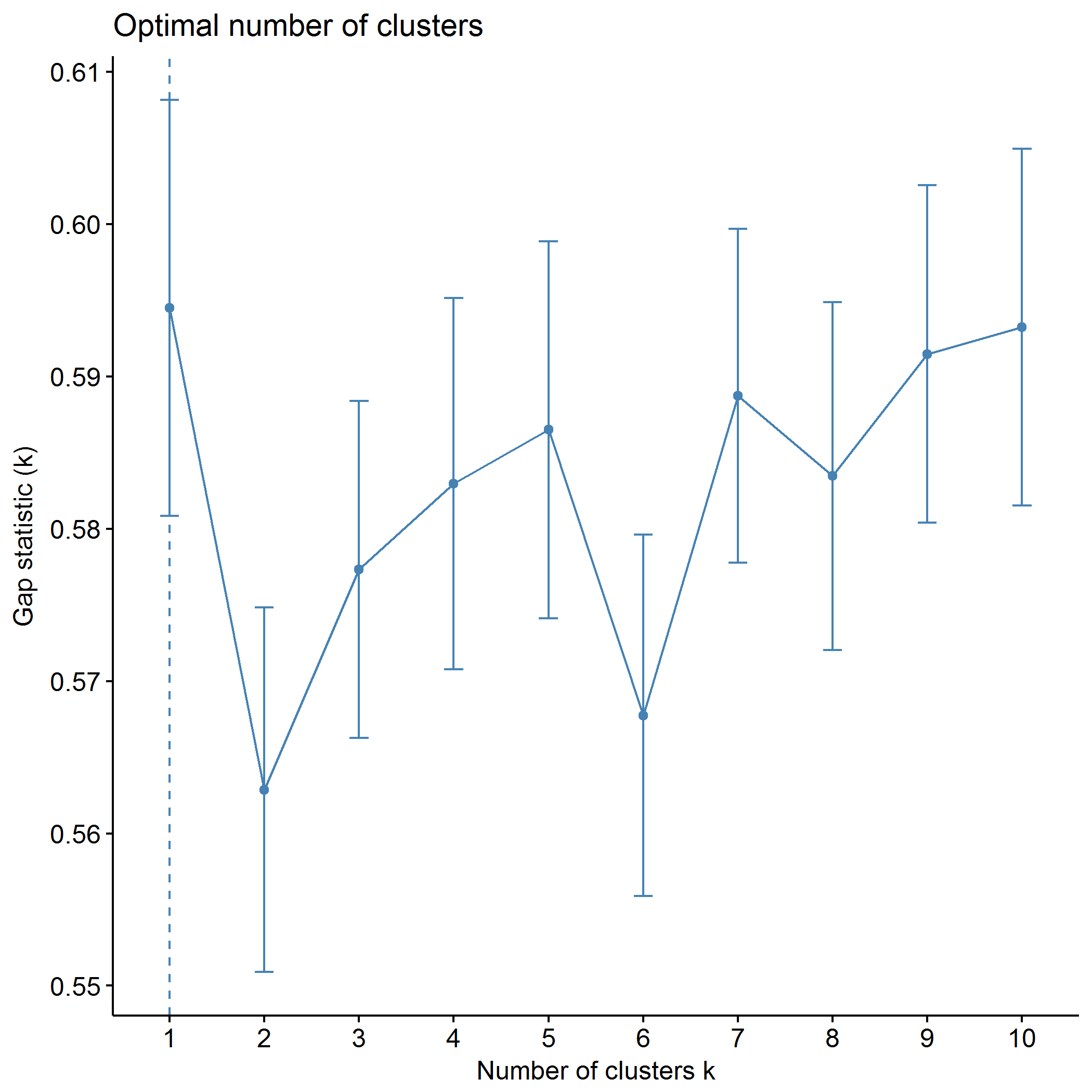
## [1] 345 6

## NULL

## K-means clustering

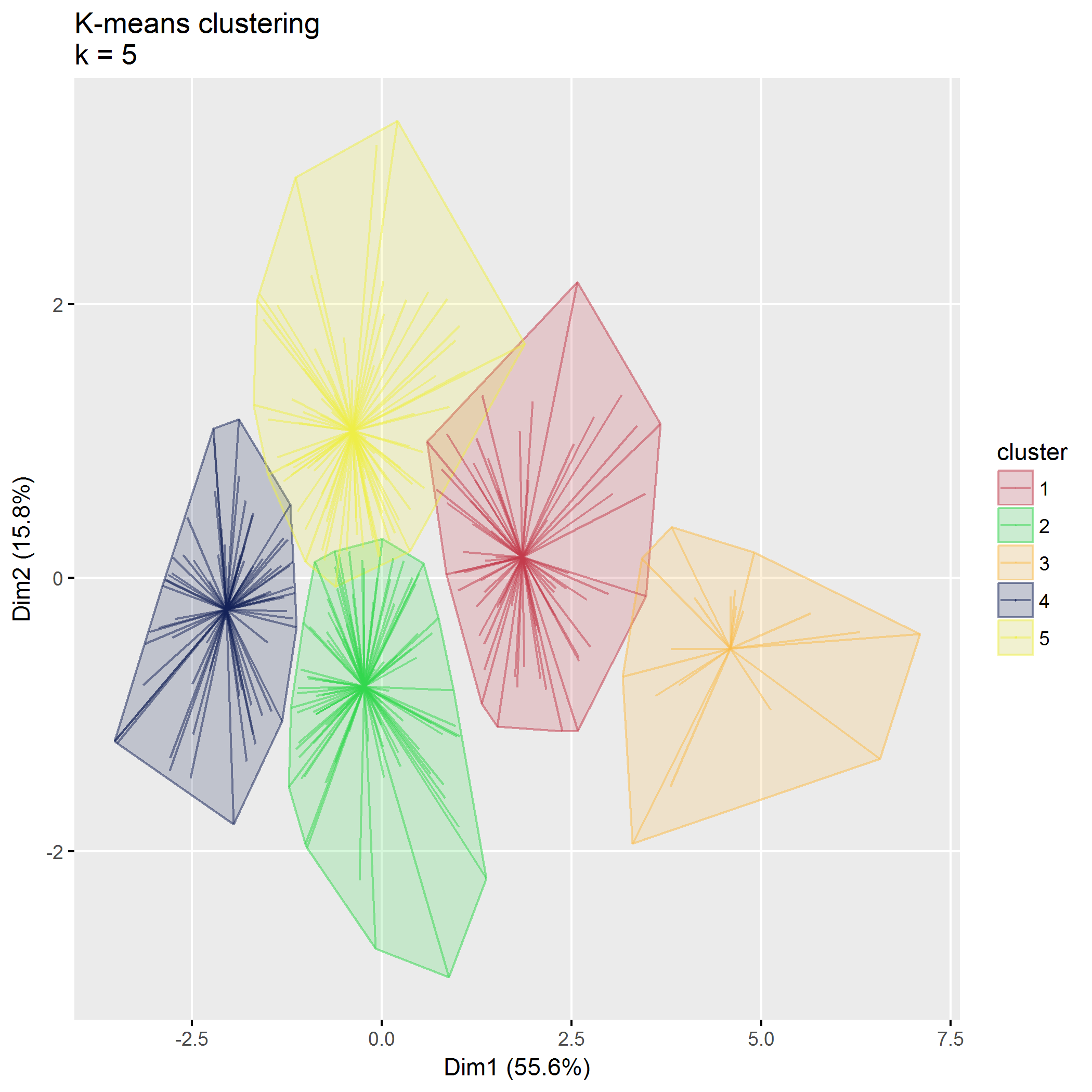


## Clustering k = 1,2,..., K.max (= 10): .. done  
## Bootstrapping, b = 1,2,..., B (= 500) [one "." per sample]:  
## .................................................. 50   
## .................................................. 100   
## .................................................. 150   
##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

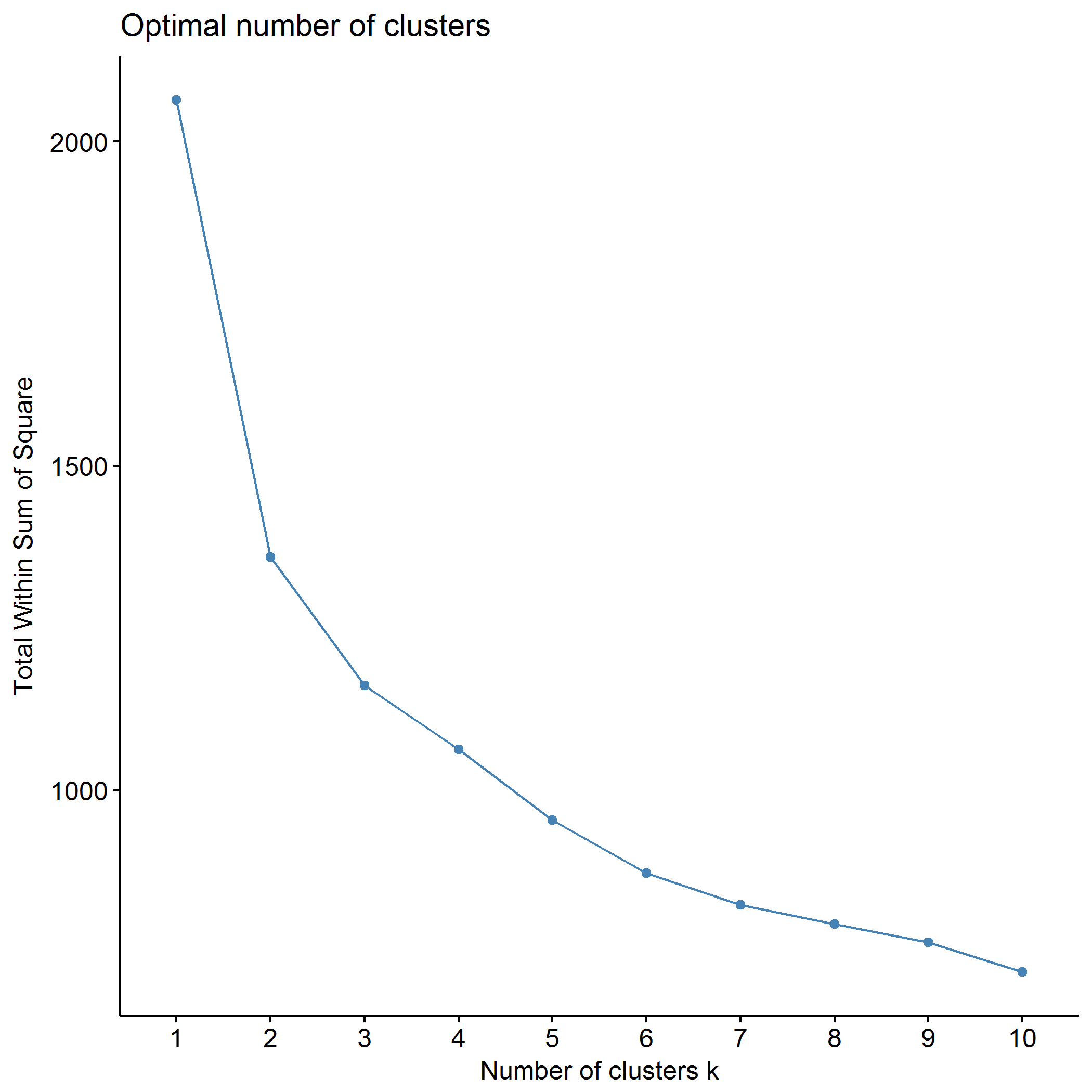
## Within cluster sum of squares, cluster 1: 232.49  
## Within cluster sum of squares, cluster 2: 231.94  
## Within cluster sum of squares, cluster 3: 72.29  
## Within cluster sum of squares, cluster 4: 122.39  
## Within cluster sum of squares, cluster 5: 237.95

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

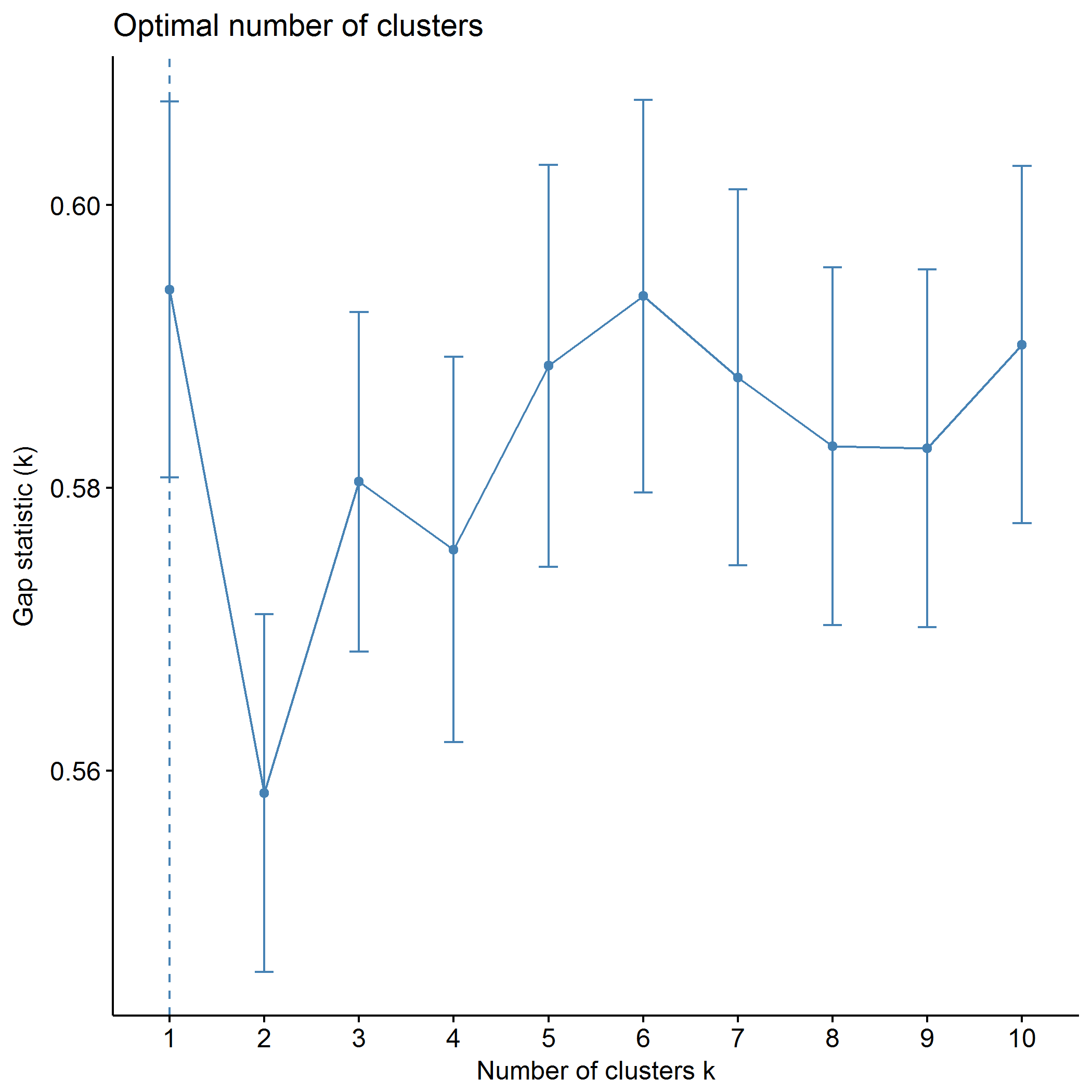
## Total within SS: 897.06

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 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

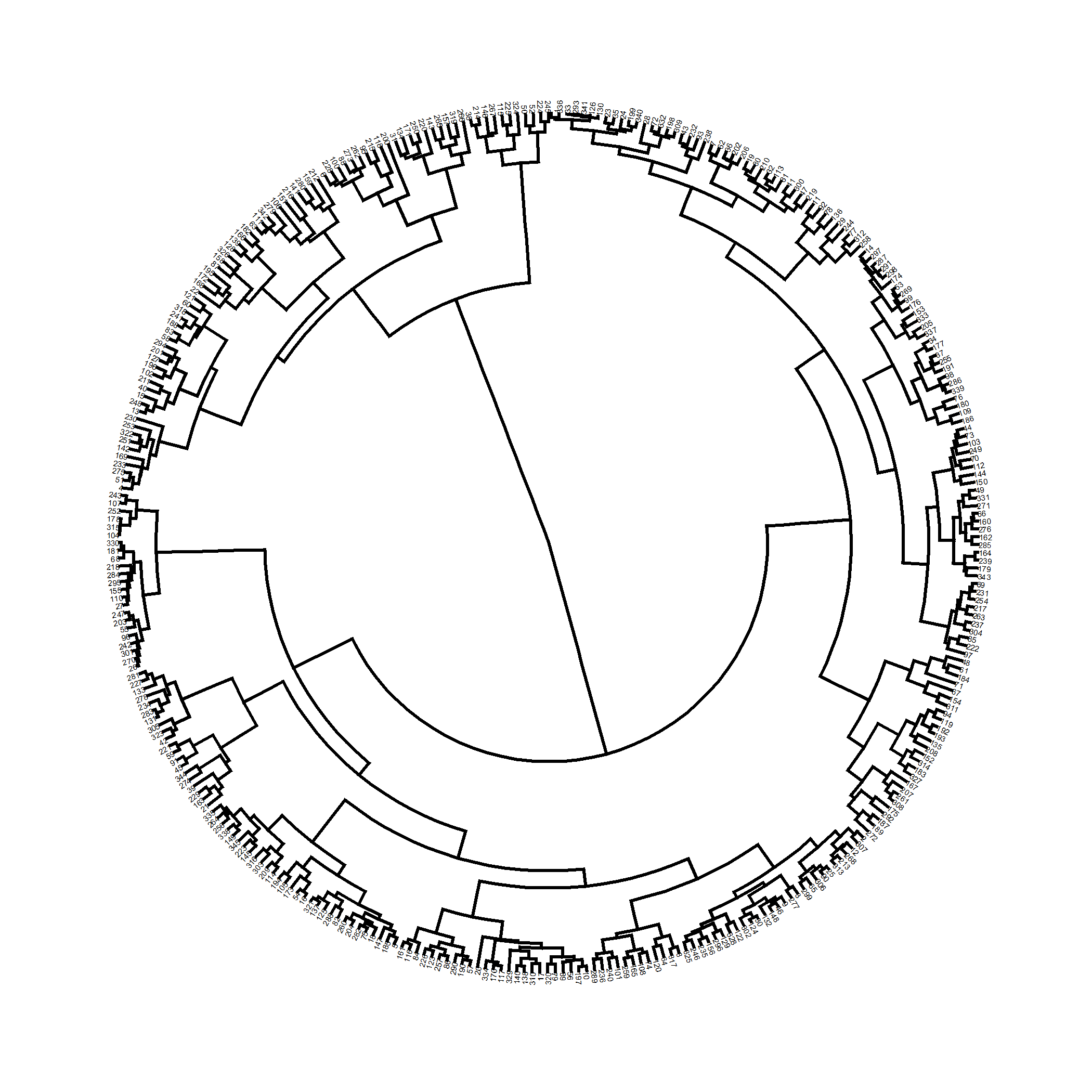


plot of chunk PCB\_pam

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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.

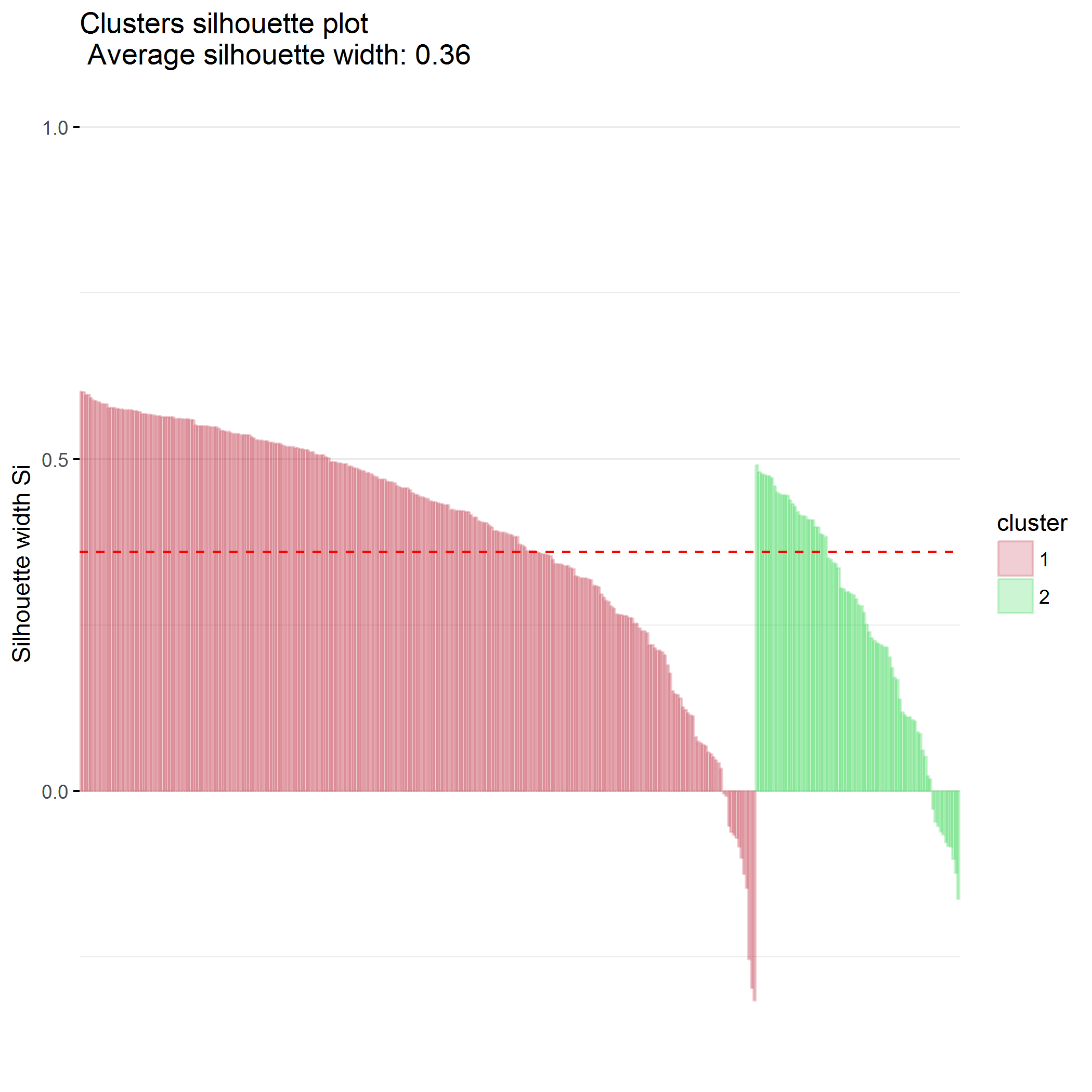
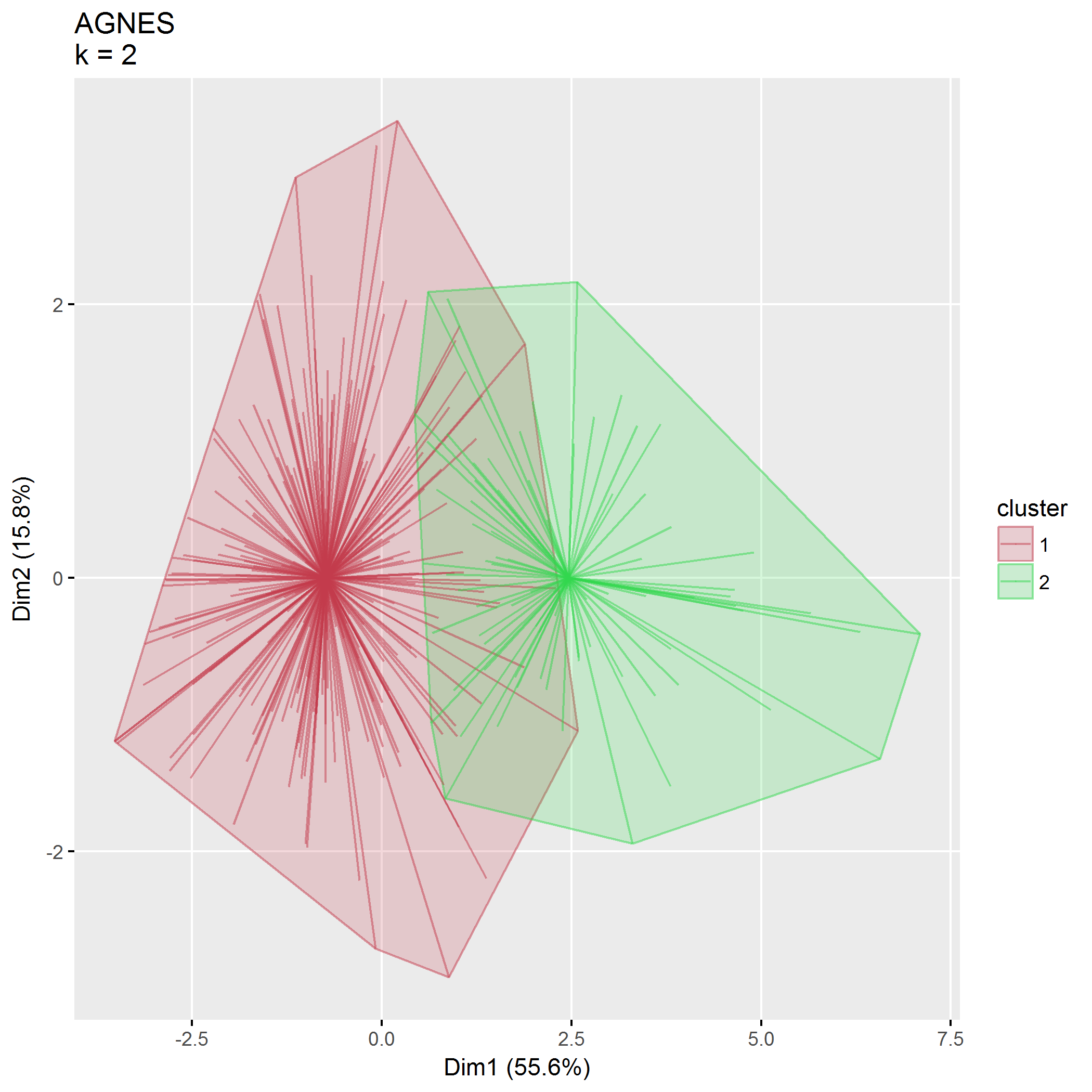
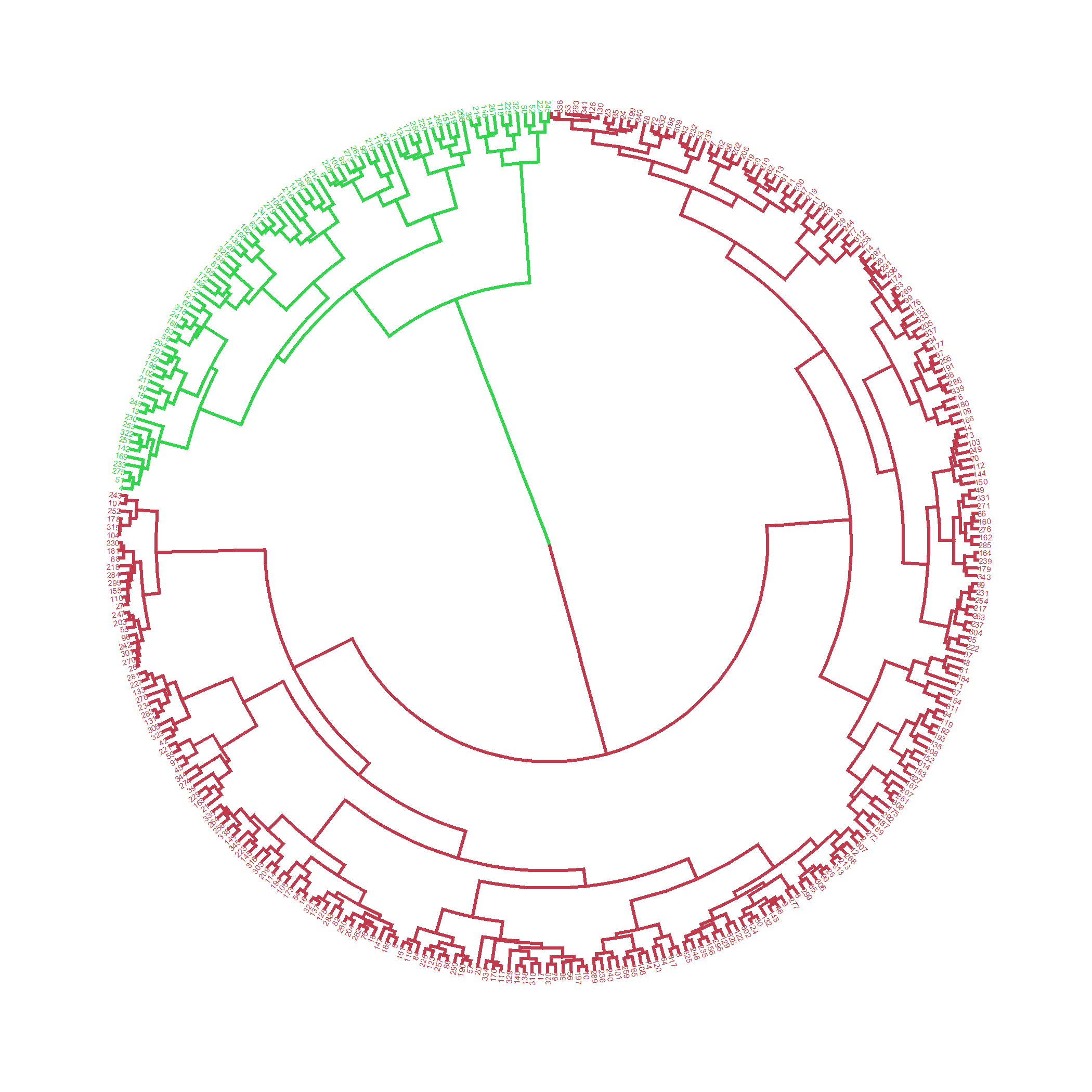
The closer the value of the correlation coefficient is to 1, the more accurately the clustering solution reflects your data. Values above 0.75 are felt to be good.

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

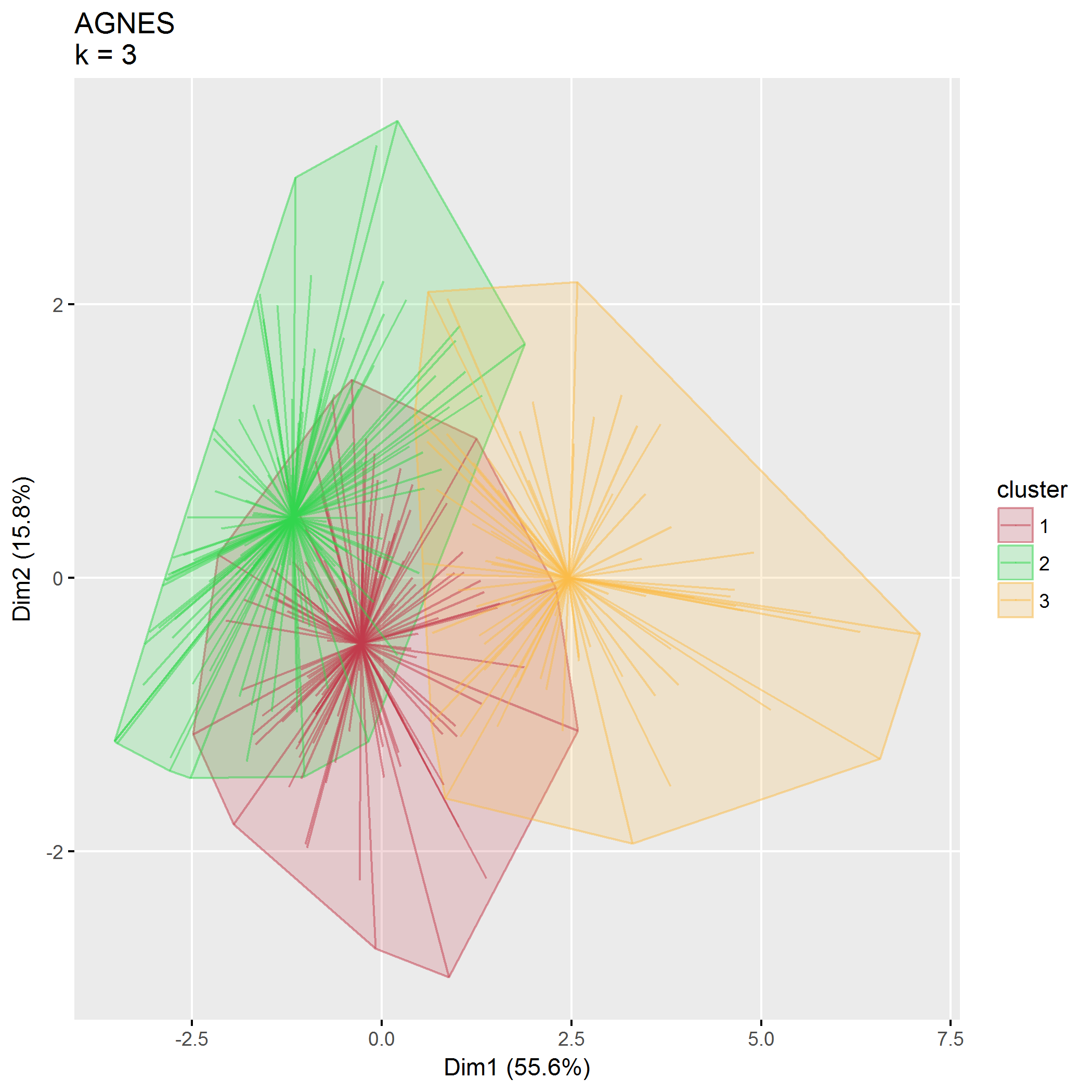
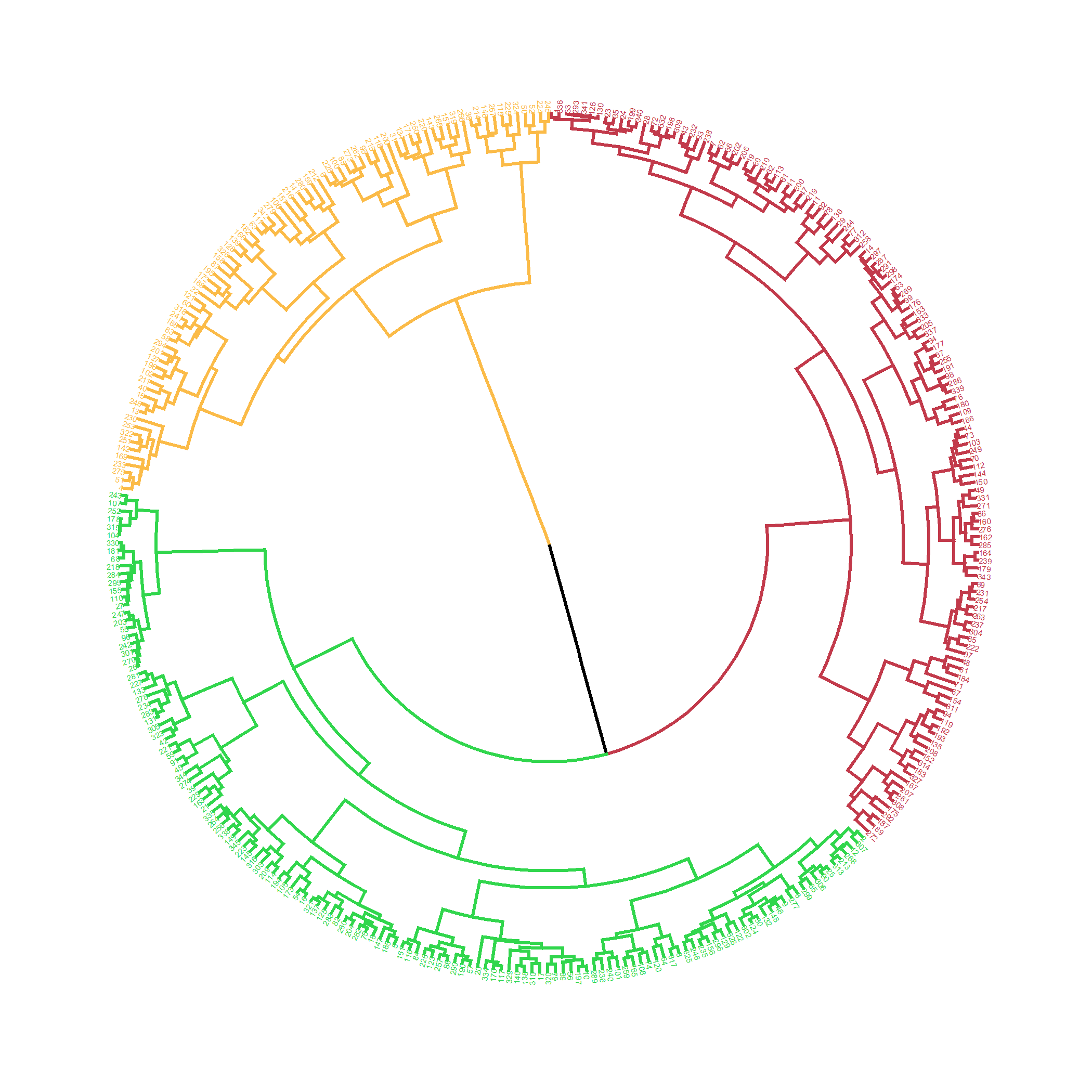
## .  
## 1 2   
## 265 80



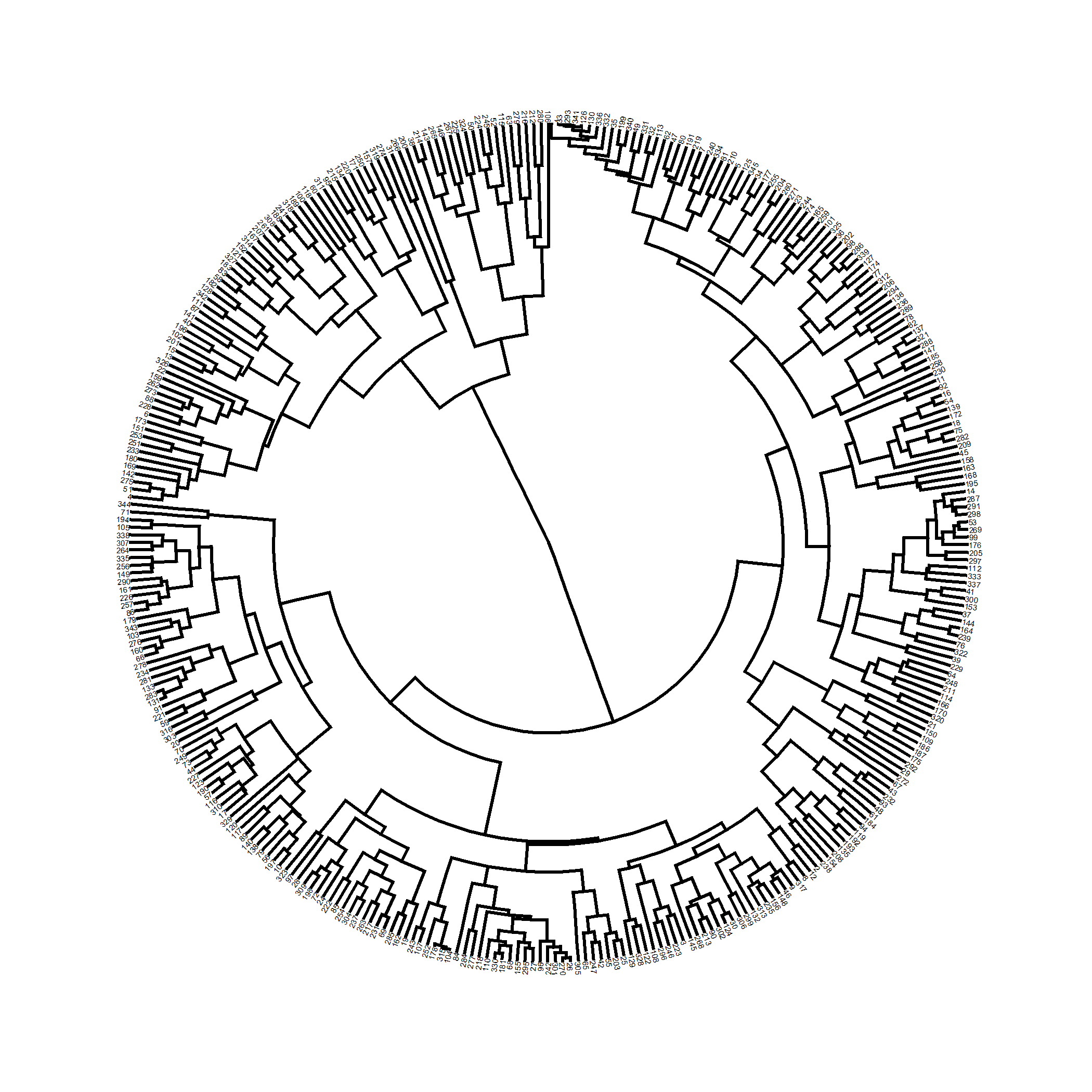
### clusters

## 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)



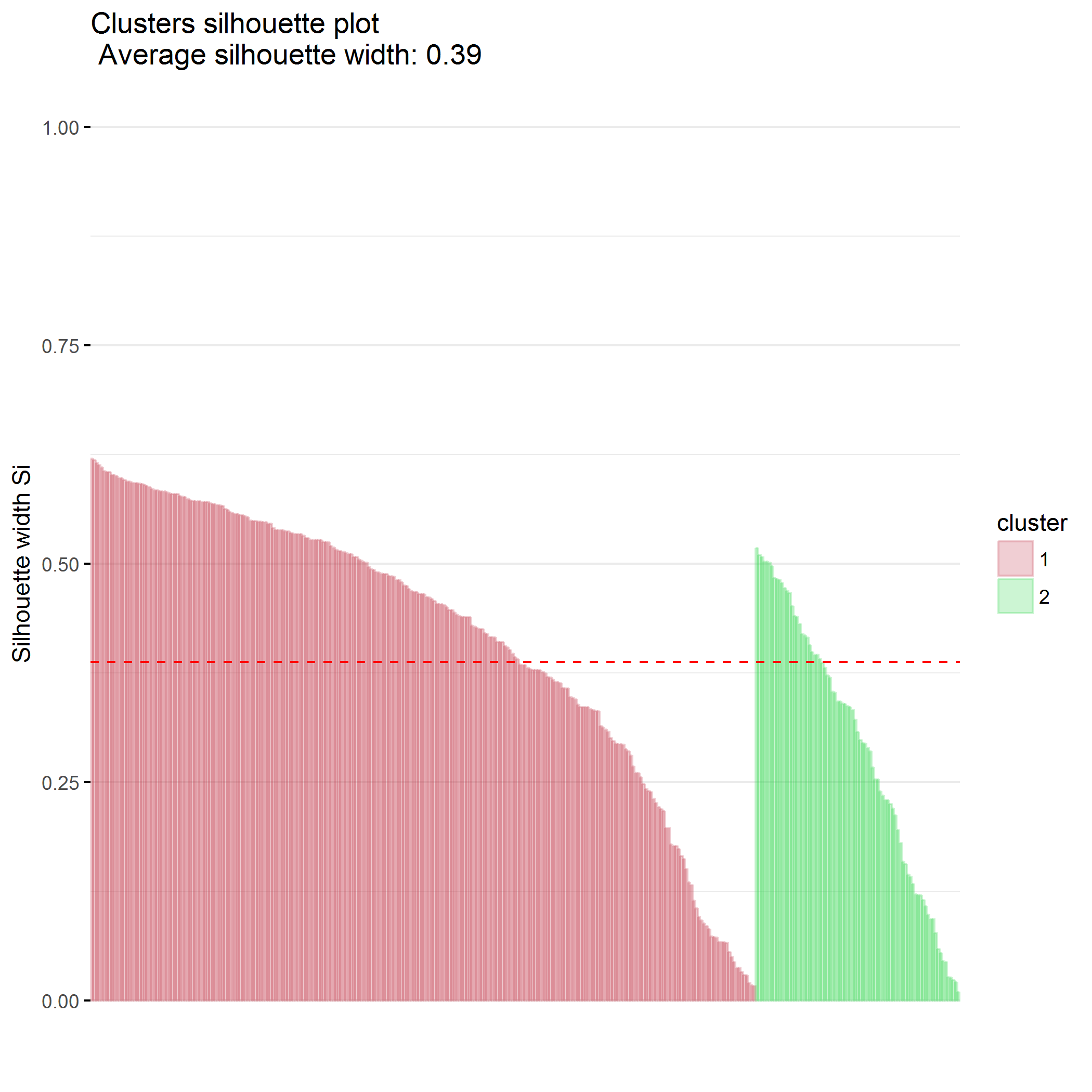
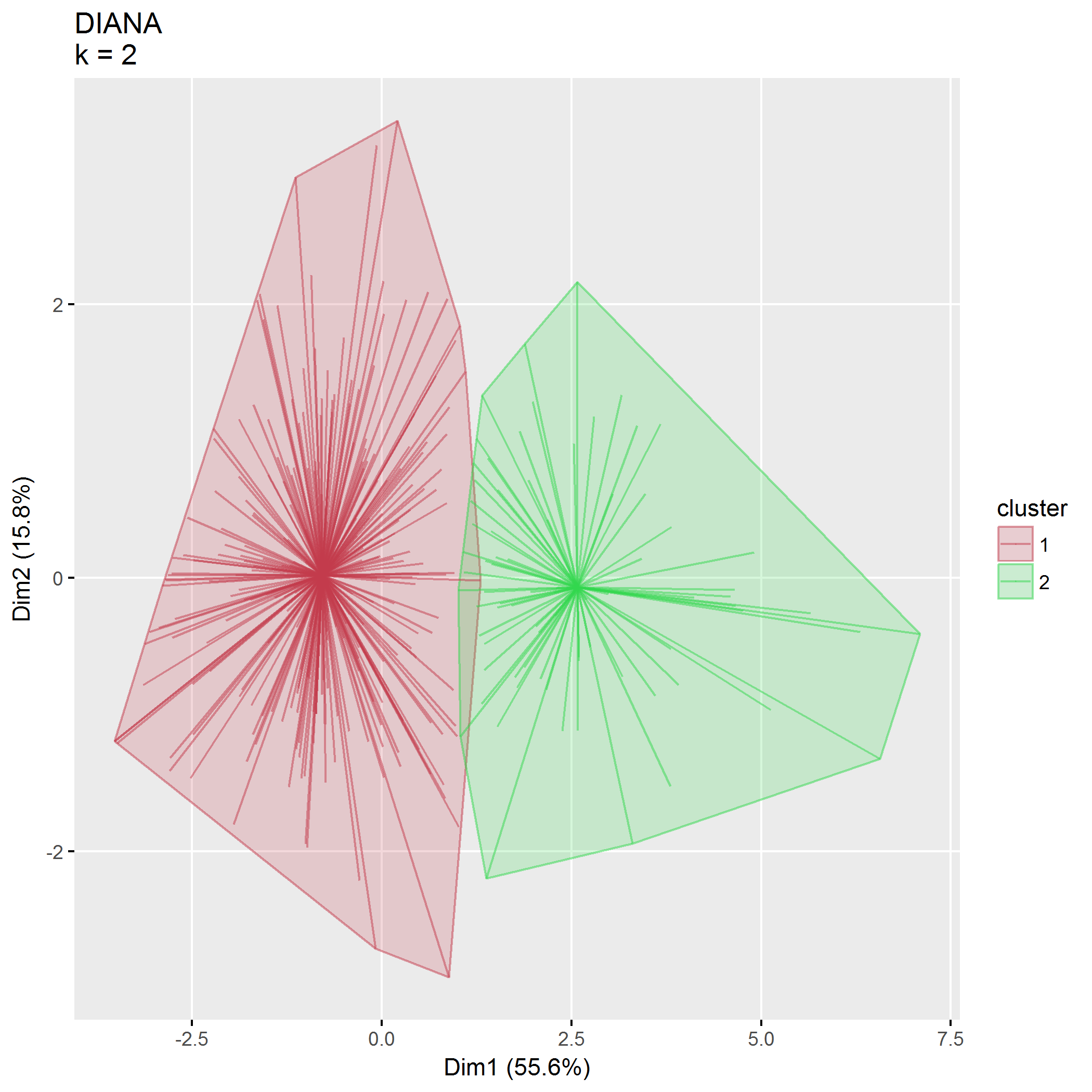
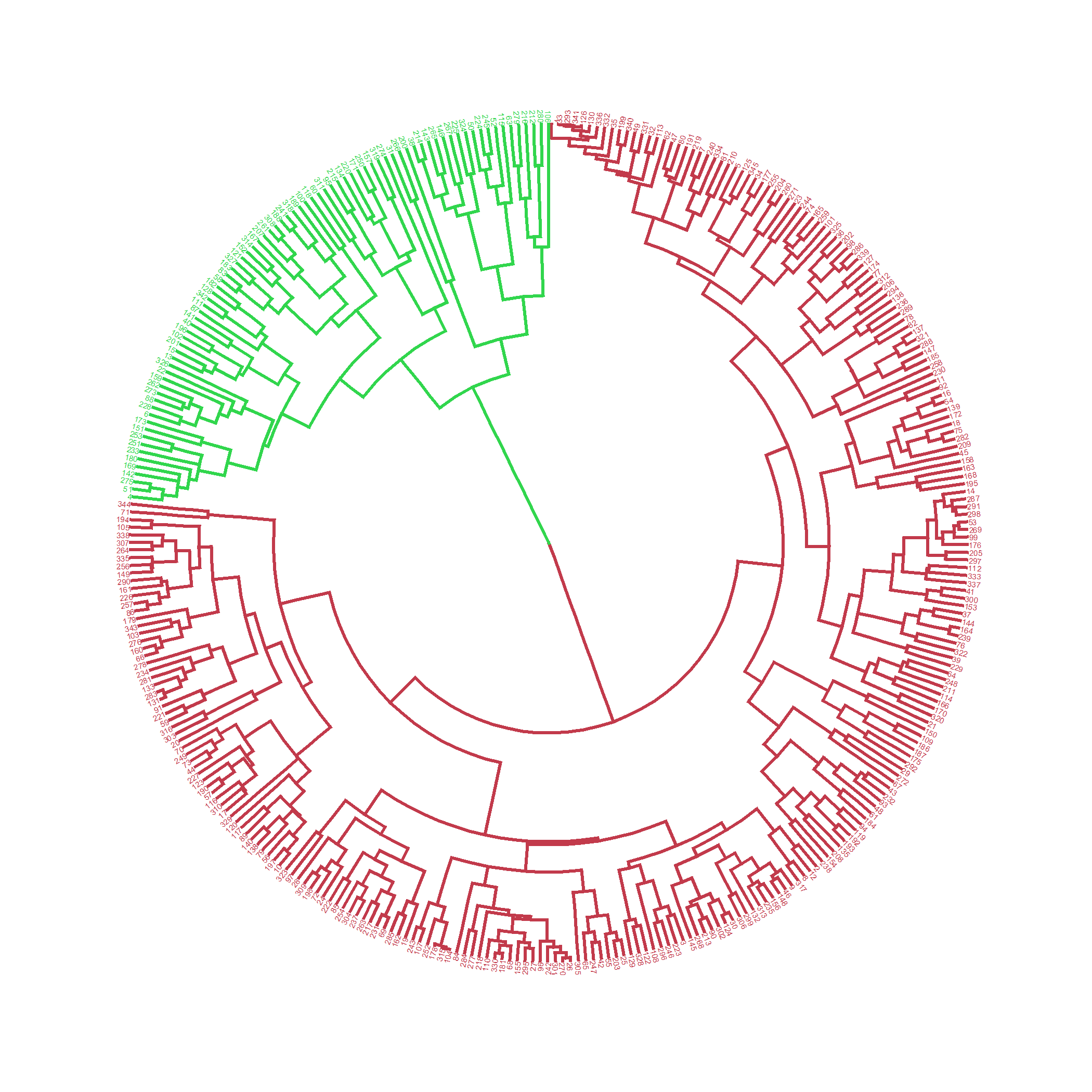
plot of chunk PCB\_diana

Divisive coeffficient is 0.926.

### clusters

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

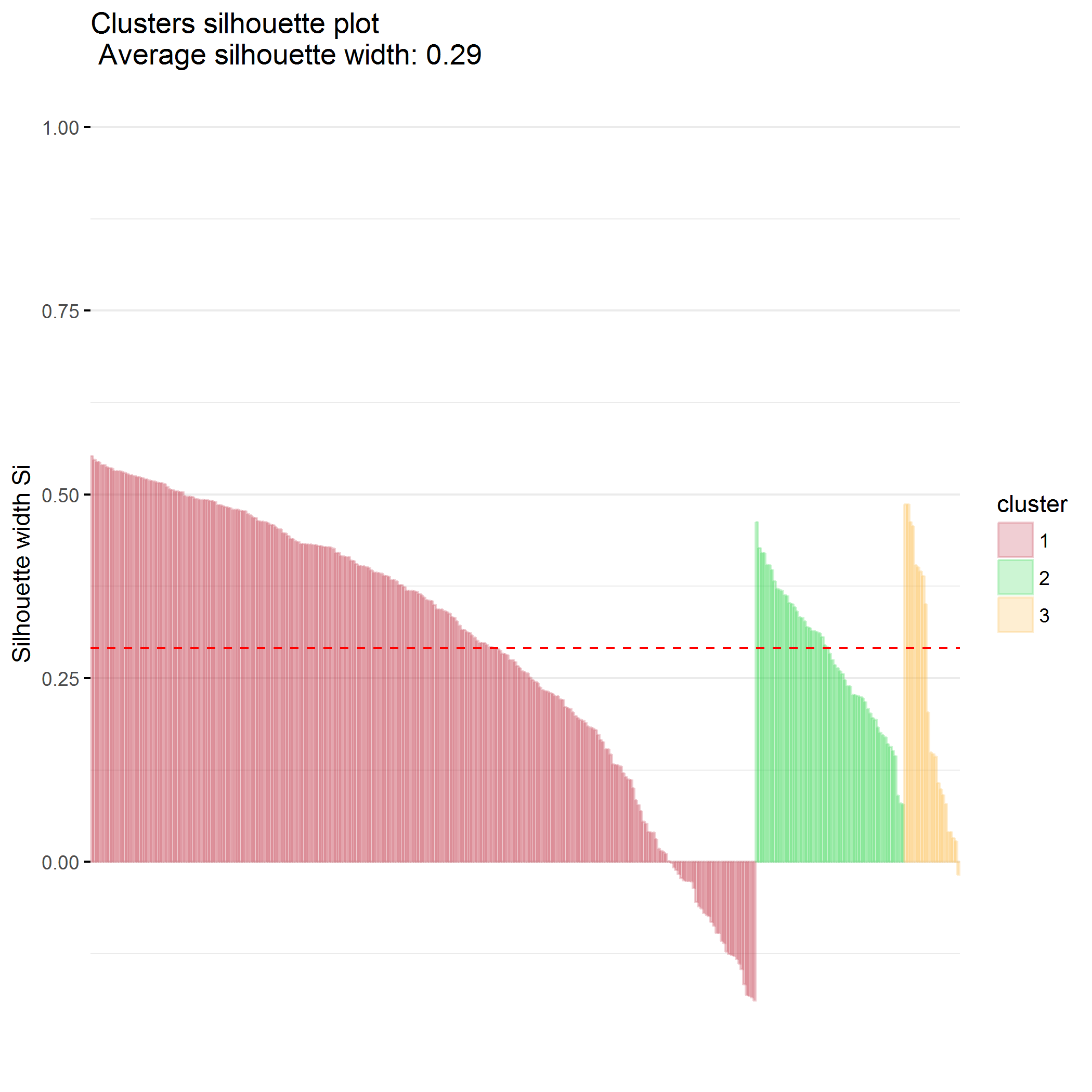
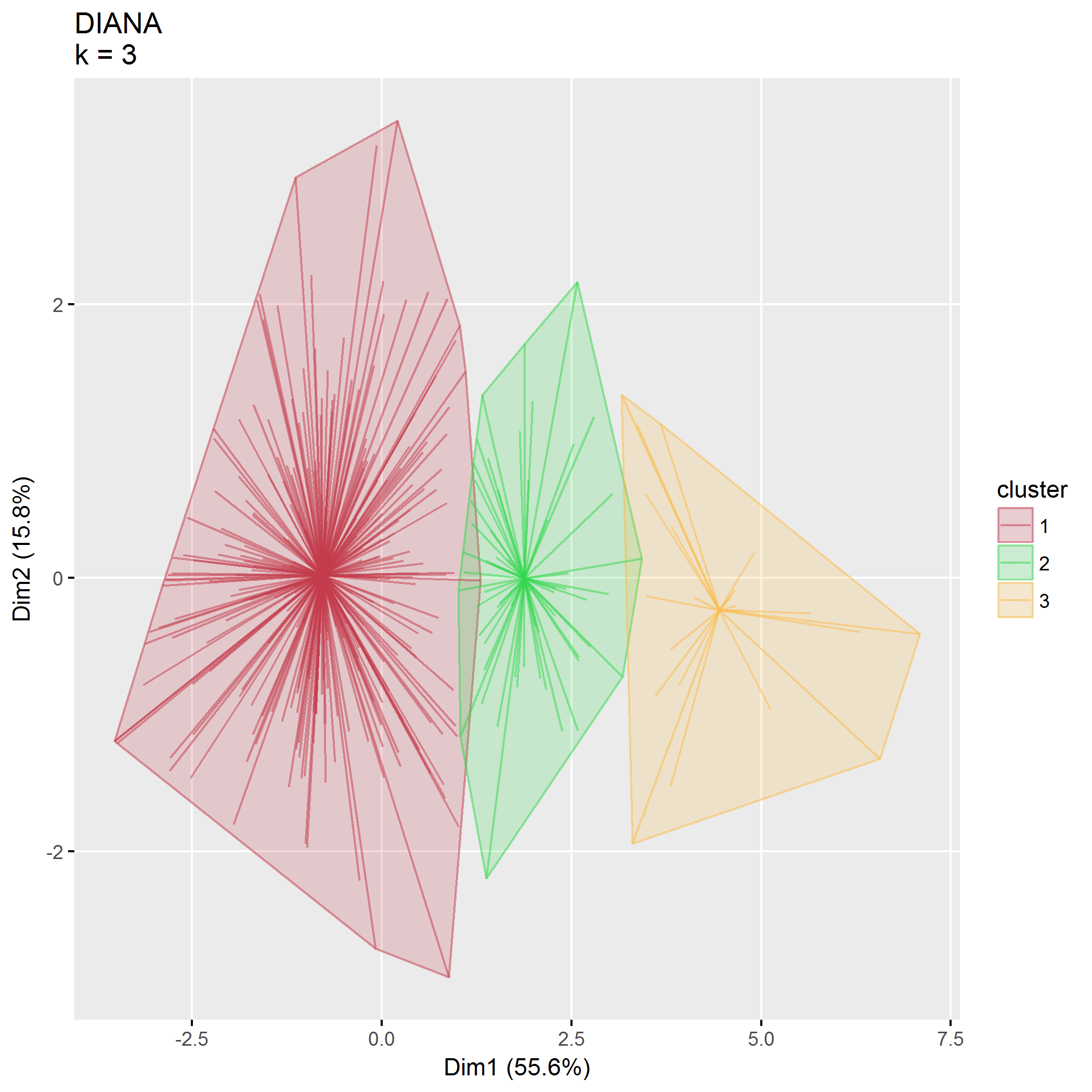
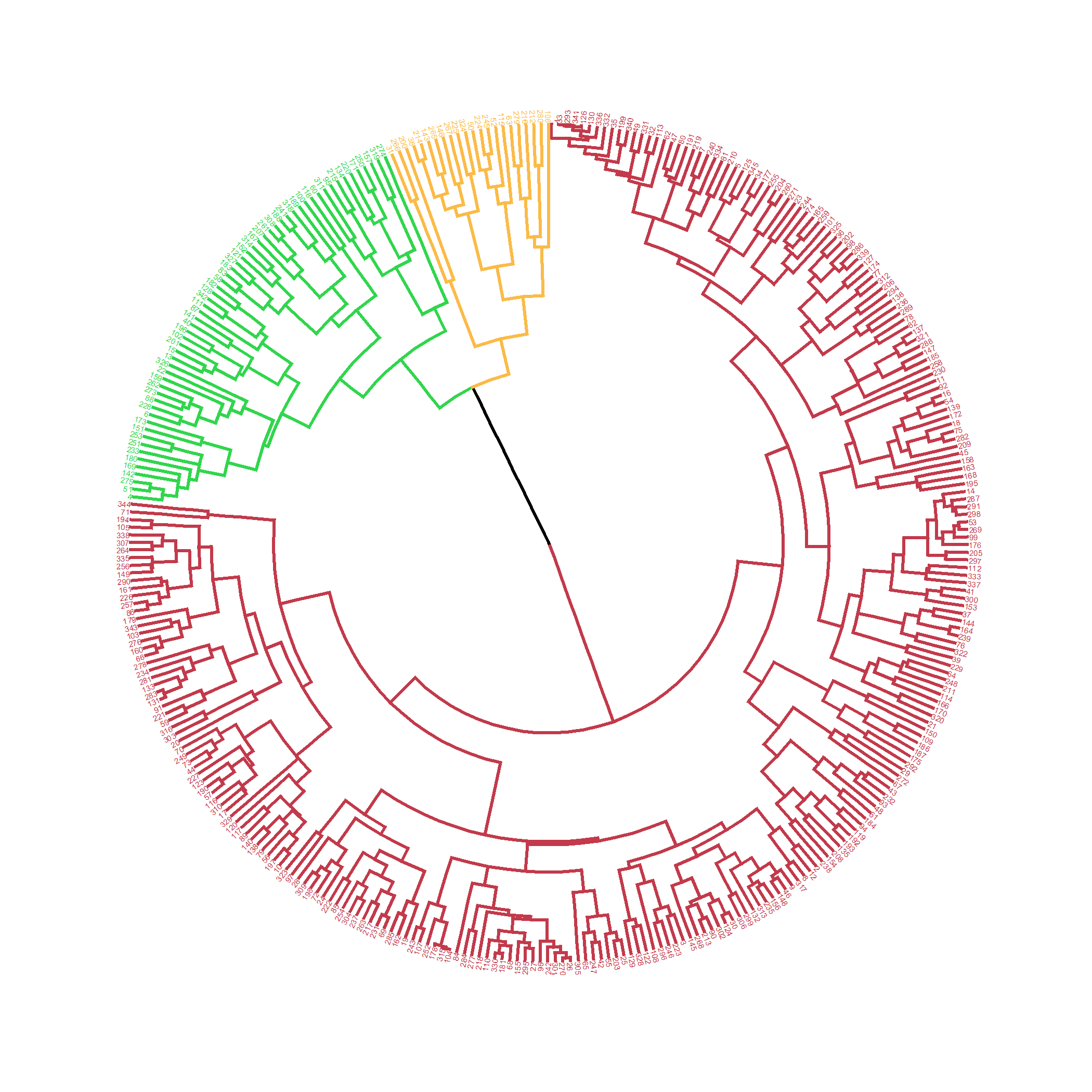
## .  
## 1 2   
## 264 81



### clusters

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

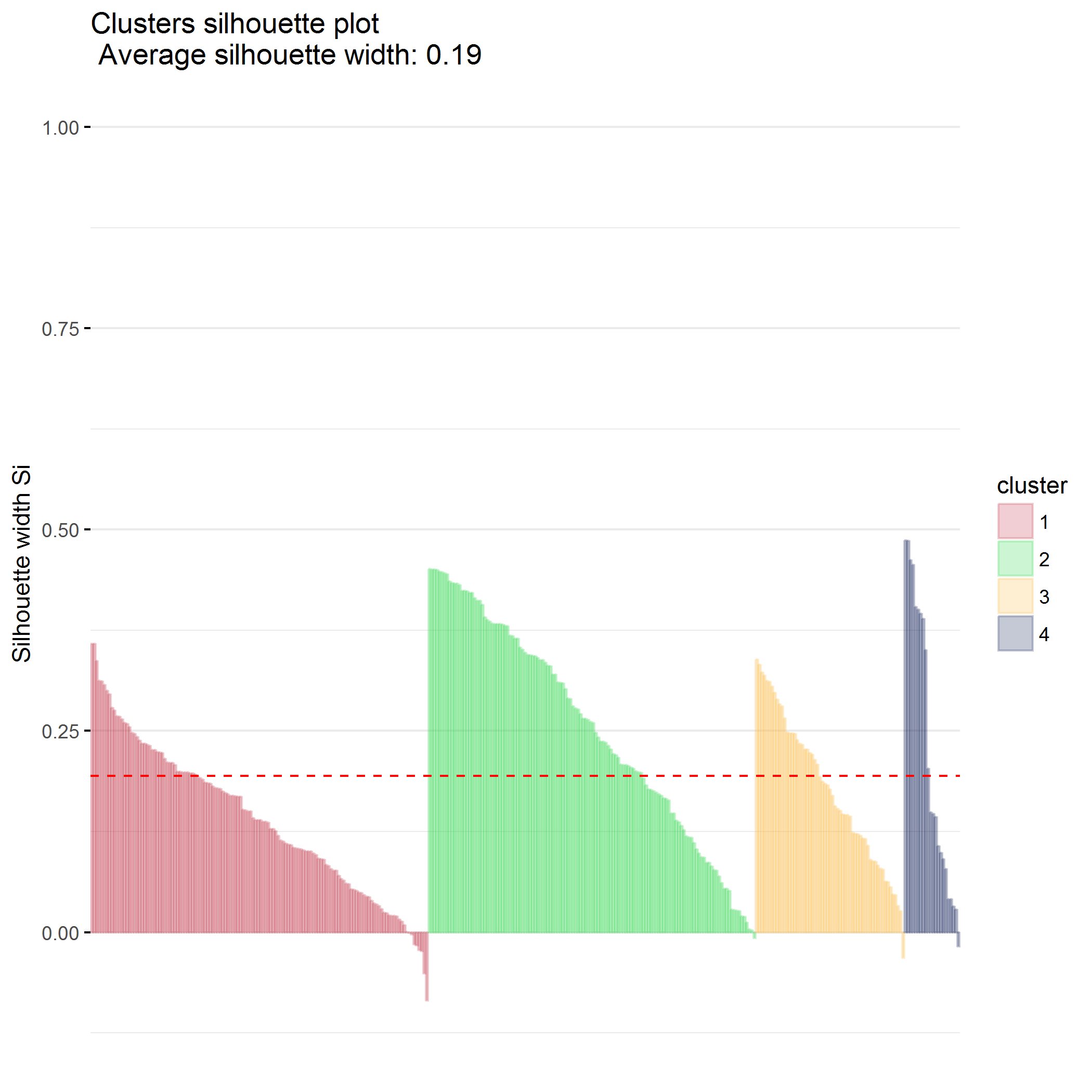
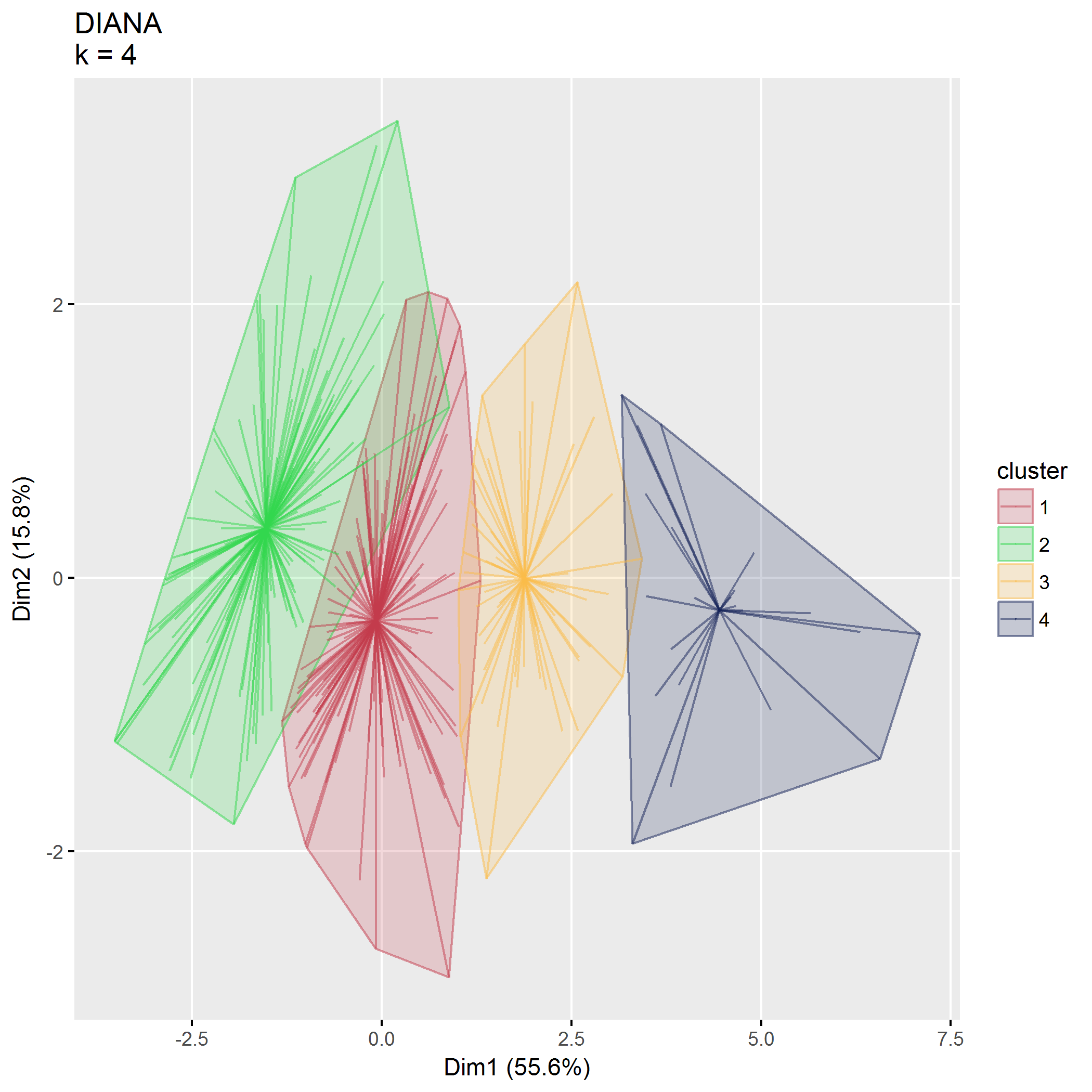
## .  
## 1 2 3   
## 264 59 22



### 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

