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

Benjamin Chan (chanb@ohsu.edu)

2018-07-10

# 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 |

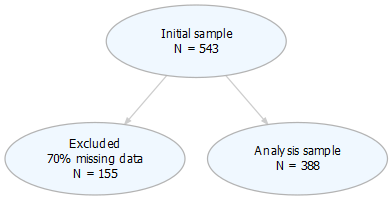
## Warning: package 'bindrcpp' was built under R version 3.4.4

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.



figures/flowChart.png

# Cluster analysis

Use divisive hierarchical clustering (DIANA). See [Divisive Hierarchical Clustering Essentials](http://www.sthda.com/english/articles/28-hierarchical-clustering-essentials/94-divisive-hierarchical-clustering-essentials/).

## Warning: package 'cluster' 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

##   
## 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},  
## }

Use the **manhattan** metric.

## Cluster on PCB metrics

**Clustering on PCB metrics isn’t terrible.**

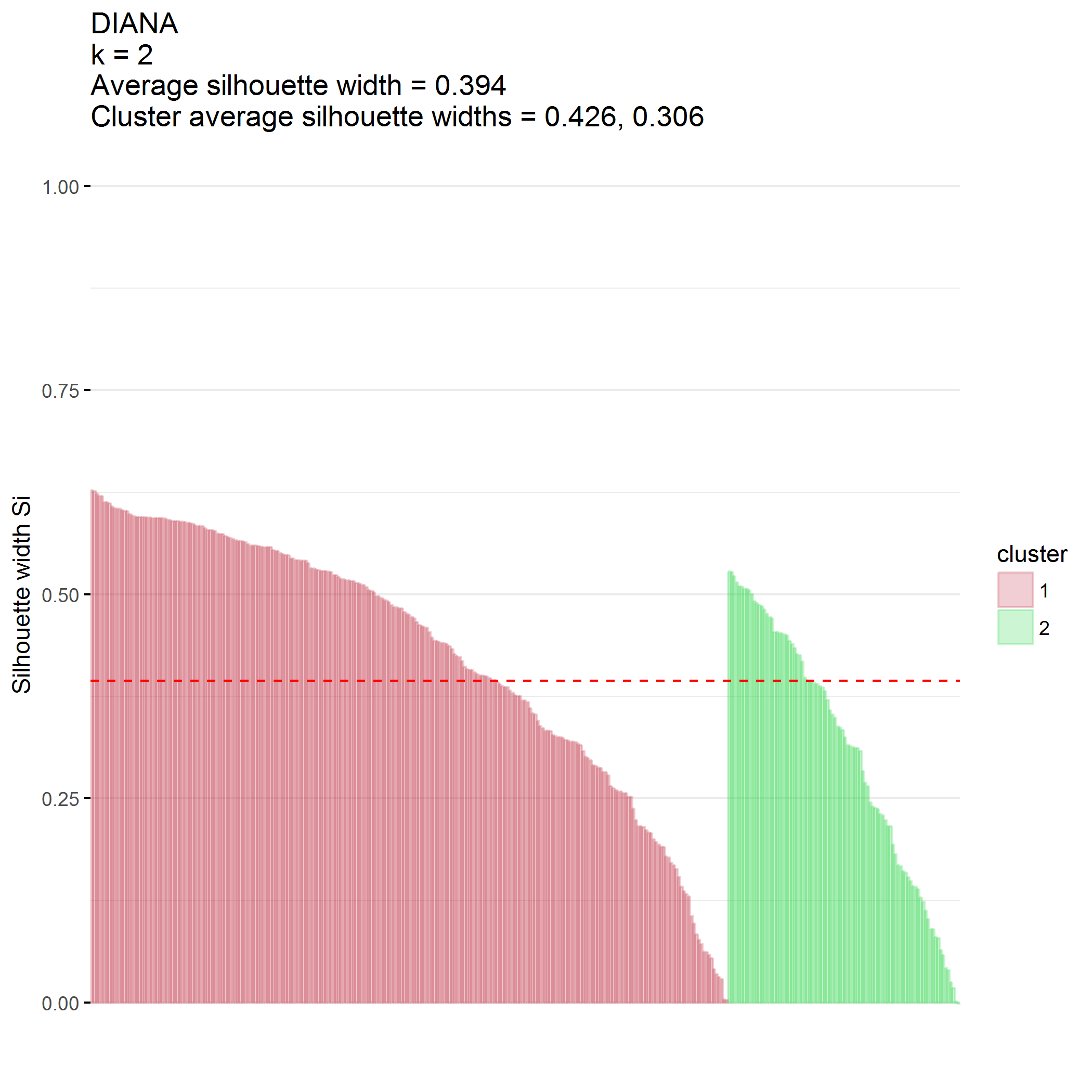
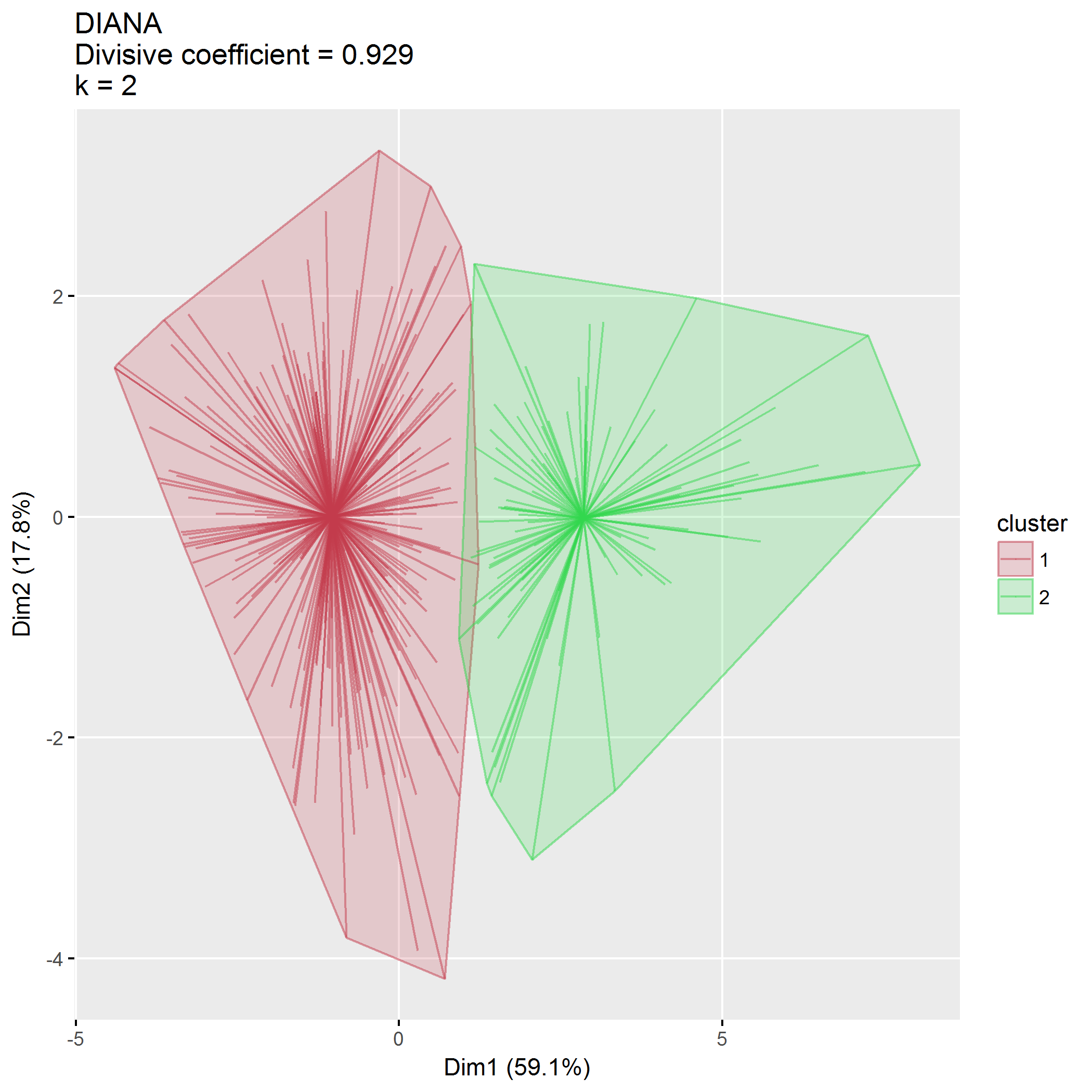
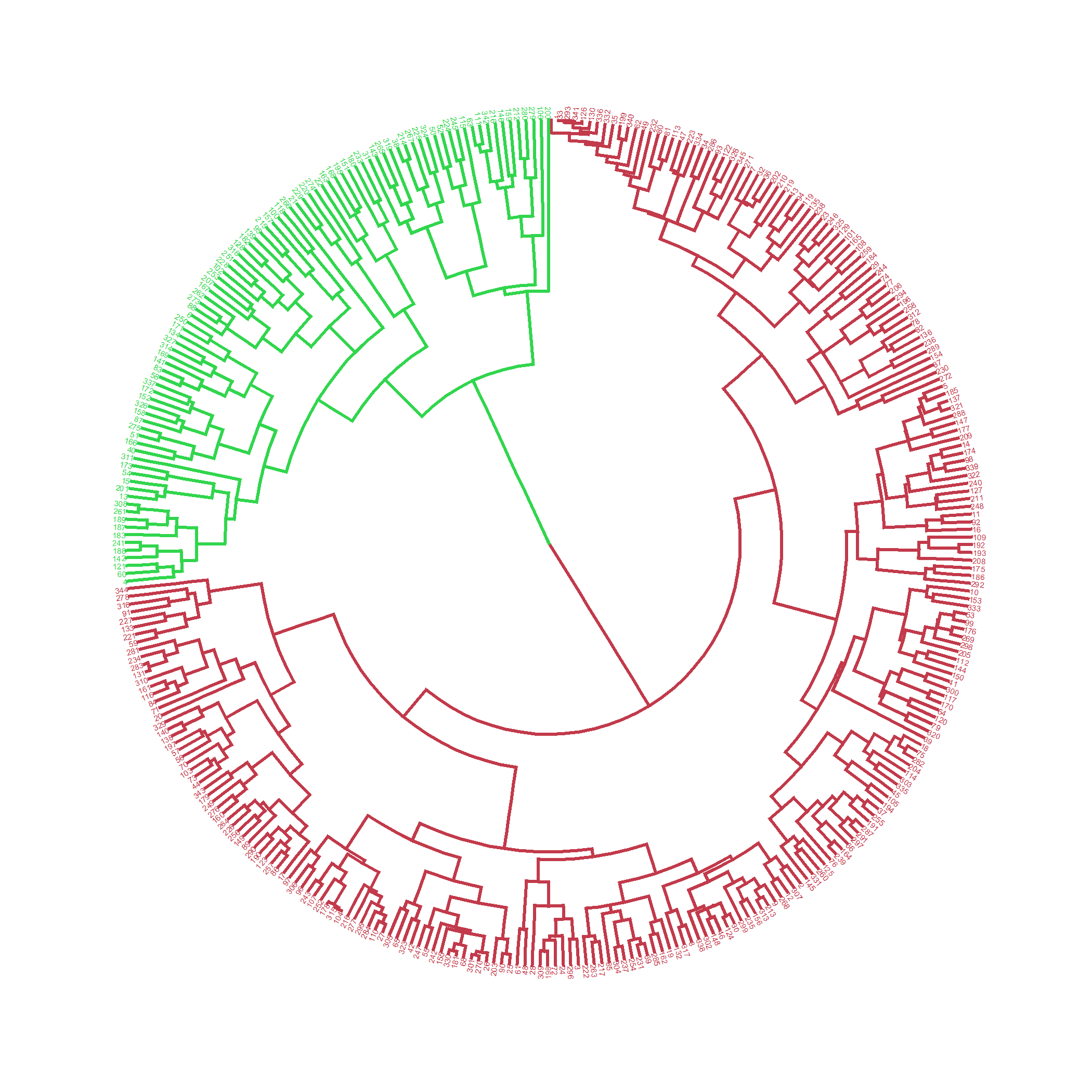
* Cluster 1 () has high PCB scores on all domains
* Cluster 2 () has low PCB scores on all domains

## [1] 345 8

## [1] "PCB1\_Total" "PCB1\_CondEmot" "PCB1\_DevHab" "PCB2\_Tot"   
## [5] "PCB3\_Total" "PCB3\_PCPonly" "PCB3\_Person" "PCB3\_Resource"

## cluster size ave.sil.width  
## 1 1 253 0.43  
## 2 2 92 0.31

* Hopkins statistic is 0.254
* Analysis identified clusters
* Divisive coefficient is 0.929
* Average silhouette width is 0.394



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| cluster | n | PCB1\_Total\_mean | PCB1\_CondEmot\_mean | PCB1\_DevHab\_mean |
| 1 | 253 | 74.0 | 53.7 | 20.3 |
| 2 | 92 | 47.1 | 33.6 | 13.5 |

|  |  |  |
| --- | --- | --- |
| cluster | n | PCB2\_Tot\_mean |
| 1 | 253 | 25.9 |
| 2 | 92 | 20.0 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| cluster | n | PCB3\_Total\_mean | PCB3\_PCPonly\_mean | PCB3\_Person\_mean | PCB3\_Resource\_mean |
| 1 | 253 | 51.7 | 4.4 | 17.4 | 30 |
| 2 | 92 | 34.2 | 3.3 | 11.9 | 19 |

## Cluster on ECBI, MAPS, SEPTI metrics

**Clustering on ECBI, MAPS, and SEPTI metrics isn’t terrible.**

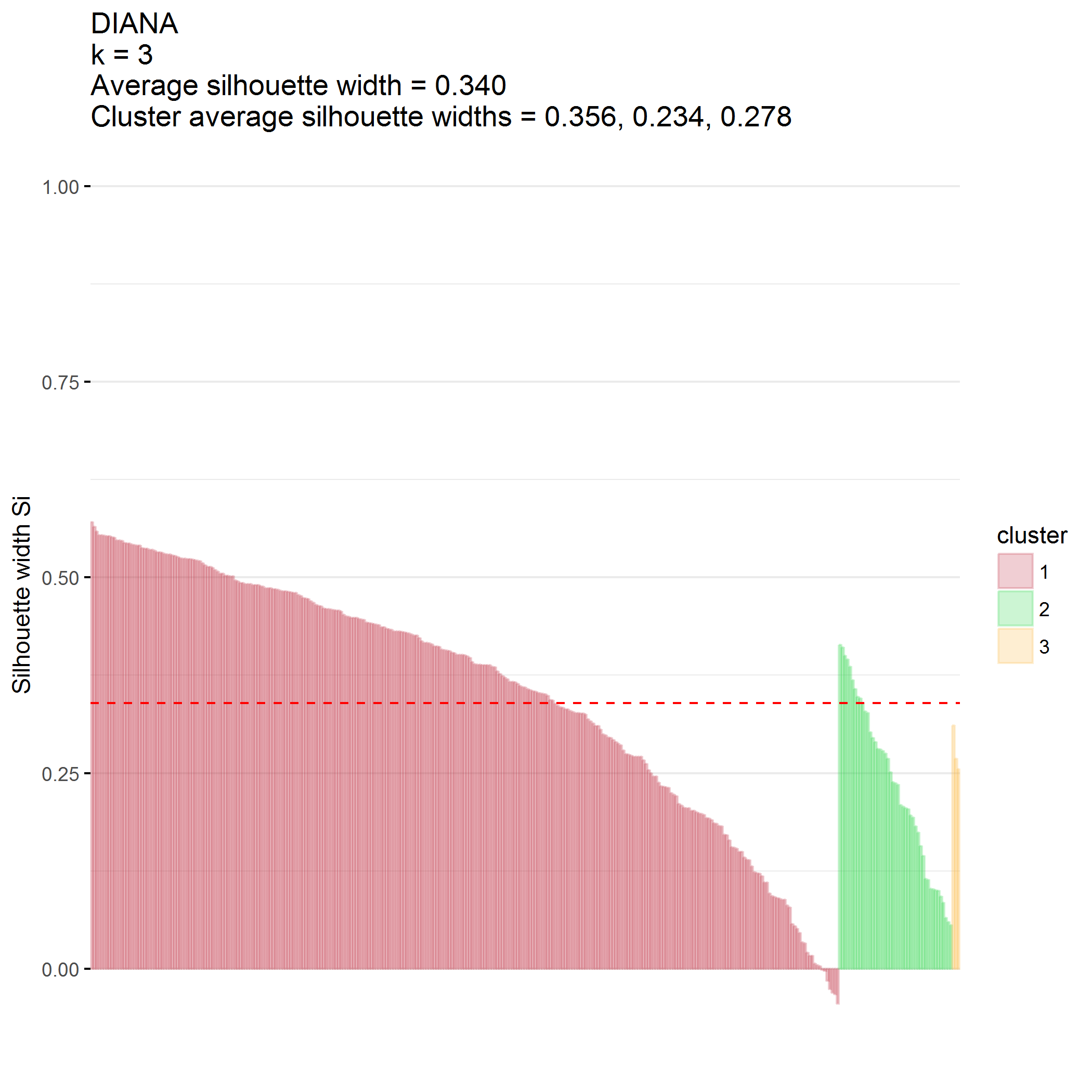
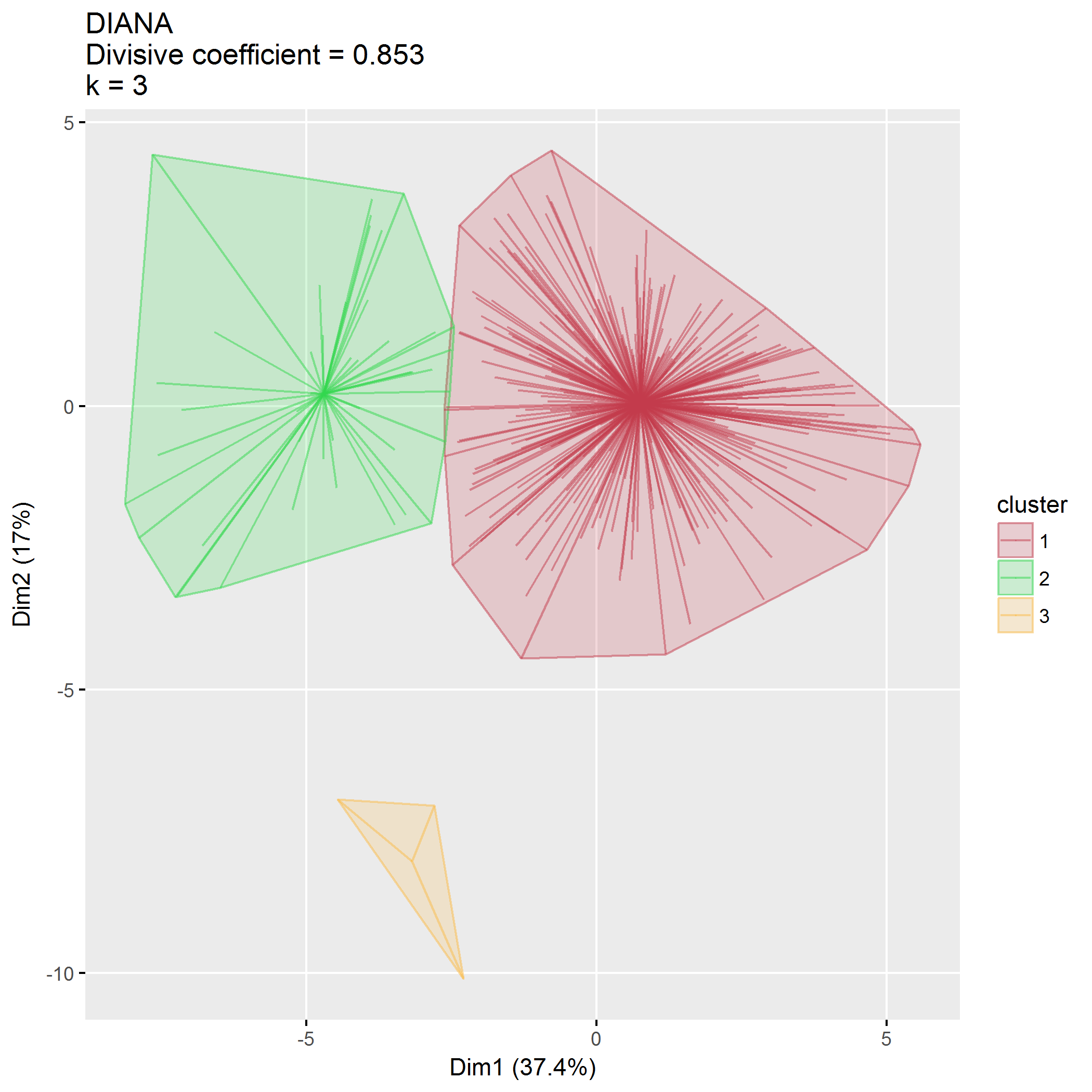
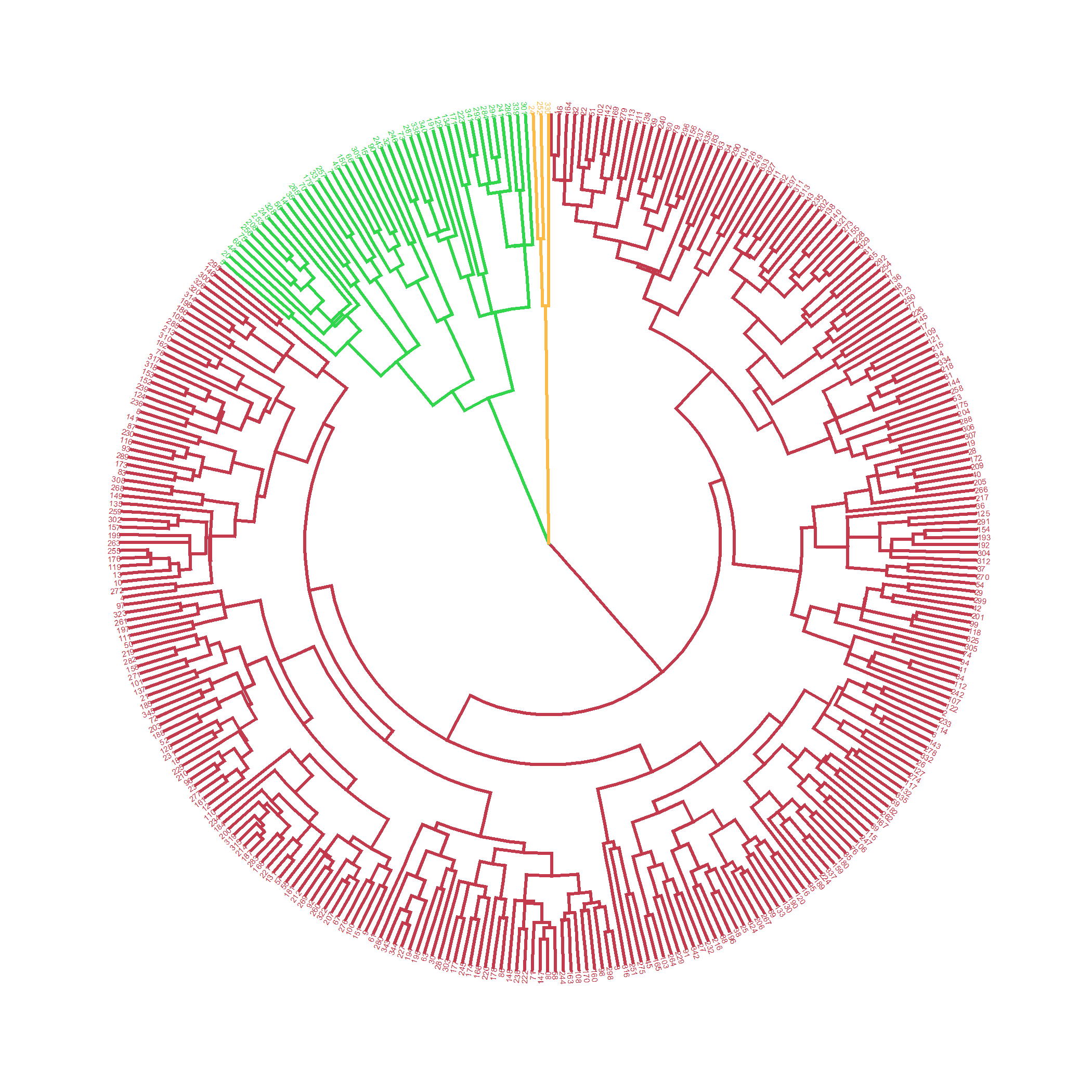
* Cluster 1 () has high positive MAPS scores and high SEPTI scores
* Cluster 2 () has high negative MAPS scores and high ECBI scores
* Cluster 3 () has low ECBI scores

## [1] 345 18

## [1] "ECBI\_intensity\_T\_score" "ECBI\_problem\_T\_score"   
## [3] "ECBI\_Opp" "ECBI\_Inatt"   
## [5] "ECBI\_Cond" "MAPS\_PP"   
## [7] "MAPS\_PR" "MAPS\_WM"   
## [9] "MAPS\_SP" "MAPS\_HS"   
## [11] "MAPS\_LC" "MAPS\_PC"   
## [13] "MAPS\_POS" "MAPS\_NEG"   
## [15] "SEPTI\_nurturance" "SEPTI\_discipline"   
## [17] "SEPTI\_play" "SEPTI\_routine"

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

* Hopkins statistic is 0.288
* Analysis identified clusters
* Divisive coefficient is 0.853
* Average silhouette width is 0.340



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| cluster | n | ECBI\_intensity\_T\_score\_mean | ECBI\_problem\_T\_score\_mean | ECBI\_Opp\_mean | ECBI\_Inatt\_mean | ECBI\_Cond\_mean |
| 1 | 297 | 52.5 | 52.2 | 32.3 | 13.2 | 14.3 |
| 2 | 45 | 62.3 | 64.6 | 43.2 | 15.8 | 24.4 |
| 3 | 3 | 37.7 | 44.3 | 17.7 | 5.3 | 8.0 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cluster | n | MAPS\_PP\_mean | MAPS\_PR\_mean | MAPS\_WM\_mean | MAPS\_SP\_mean | MAPS\_HS\_mean | MAPS\_LC\_mean | MAPS\_PC\_mean | MAPS\_POS\_mean | MAPS\_NEG\_mean |
| 1 | 297 | 4.1 | 4.6 | 4.7 | 4.5 | 2.0 | 1.9 | 1.4 | 4.5 | 1.8 |
| 2 | 45 | 3.6 | 4.0 | 4.1 | 3.6 | 3.0 | 2.7 | 2.5 | 3.8 | 2.7 |
| 3 | 3 | 2.4 | 2.9 | 3.2 | 2.1 | 1.5 | 1.9 | 1.9 | 2.7 | 1.8 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| cluster | n | SEPTI\_nurturance\_mean | SEPTI\_discipline\_mean | SEPTI\_play\_mean | SEPTI\_routine\_mean |
| 1 | 297 | 38.1 | 24.3 | 32.8 | 29.4 |
| 2 | 45 | 33.4 | 18.8 | 25.0 | 23.5 |
| 3 | 3 | 26.3 | 19.3 | 25.7 | 24.0 |

## Cluster on ECBI, MAPS, SEPTI metrics and demographics

**Clustering on ECBI, MAPS, and SEPTI metrics isn’t terrible.**

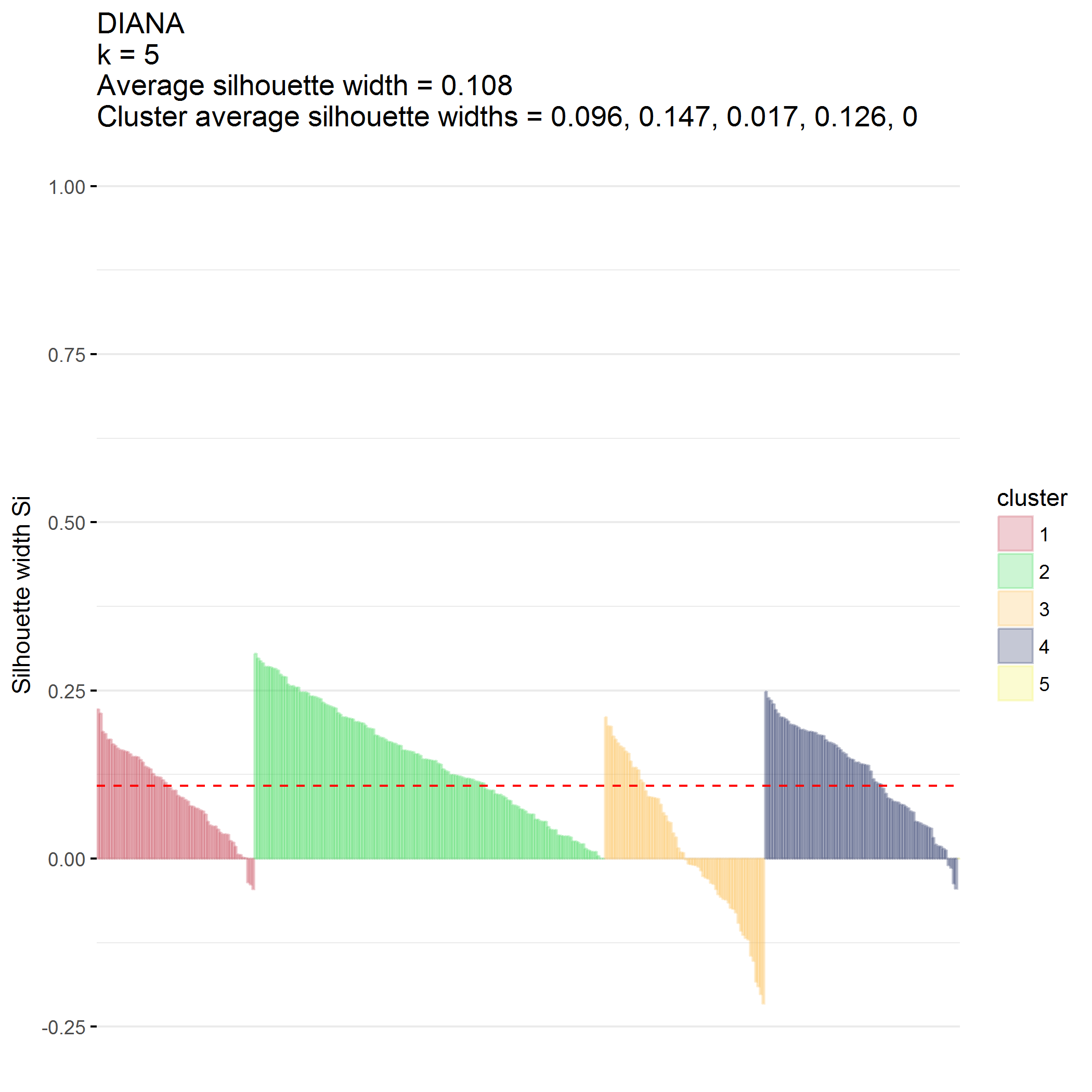
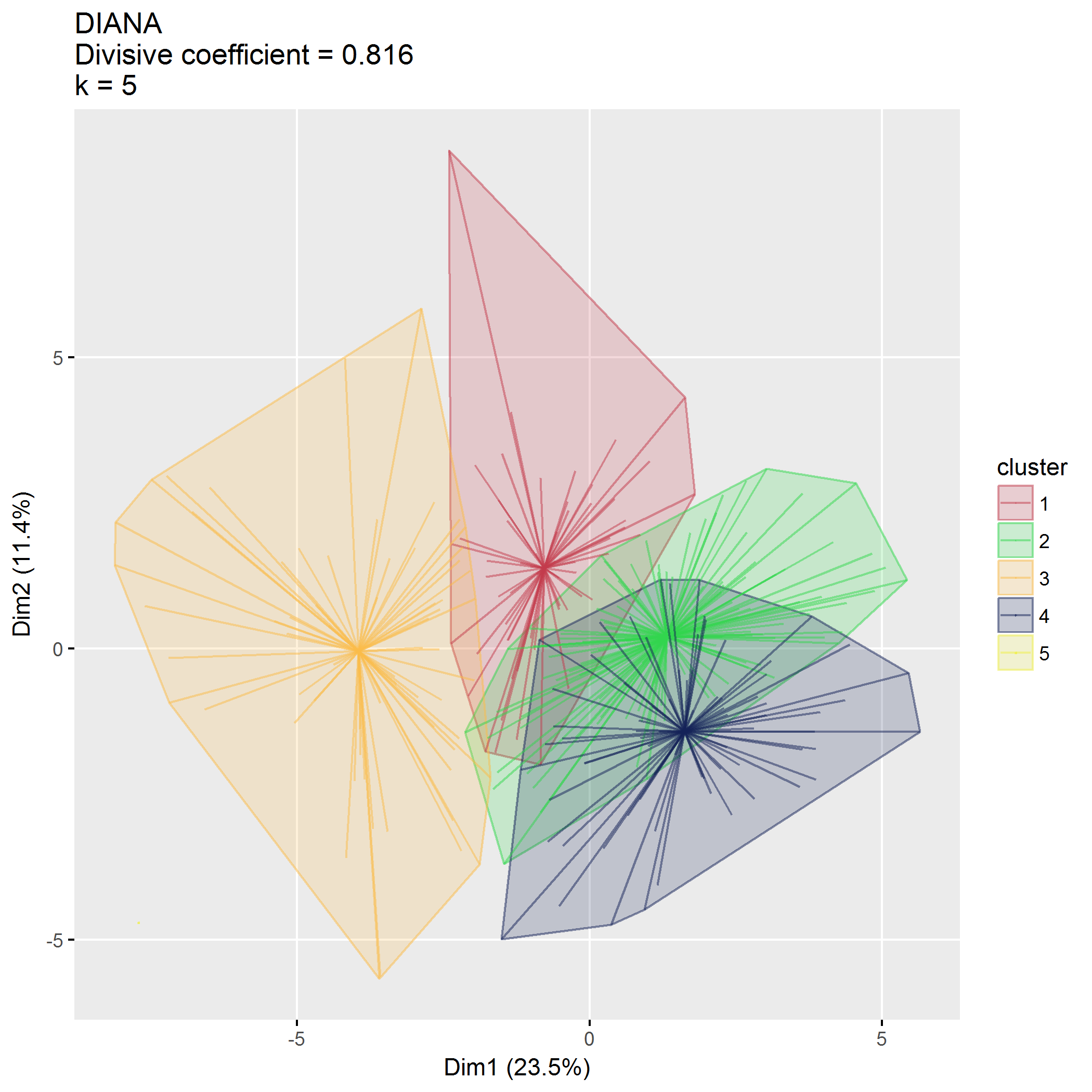
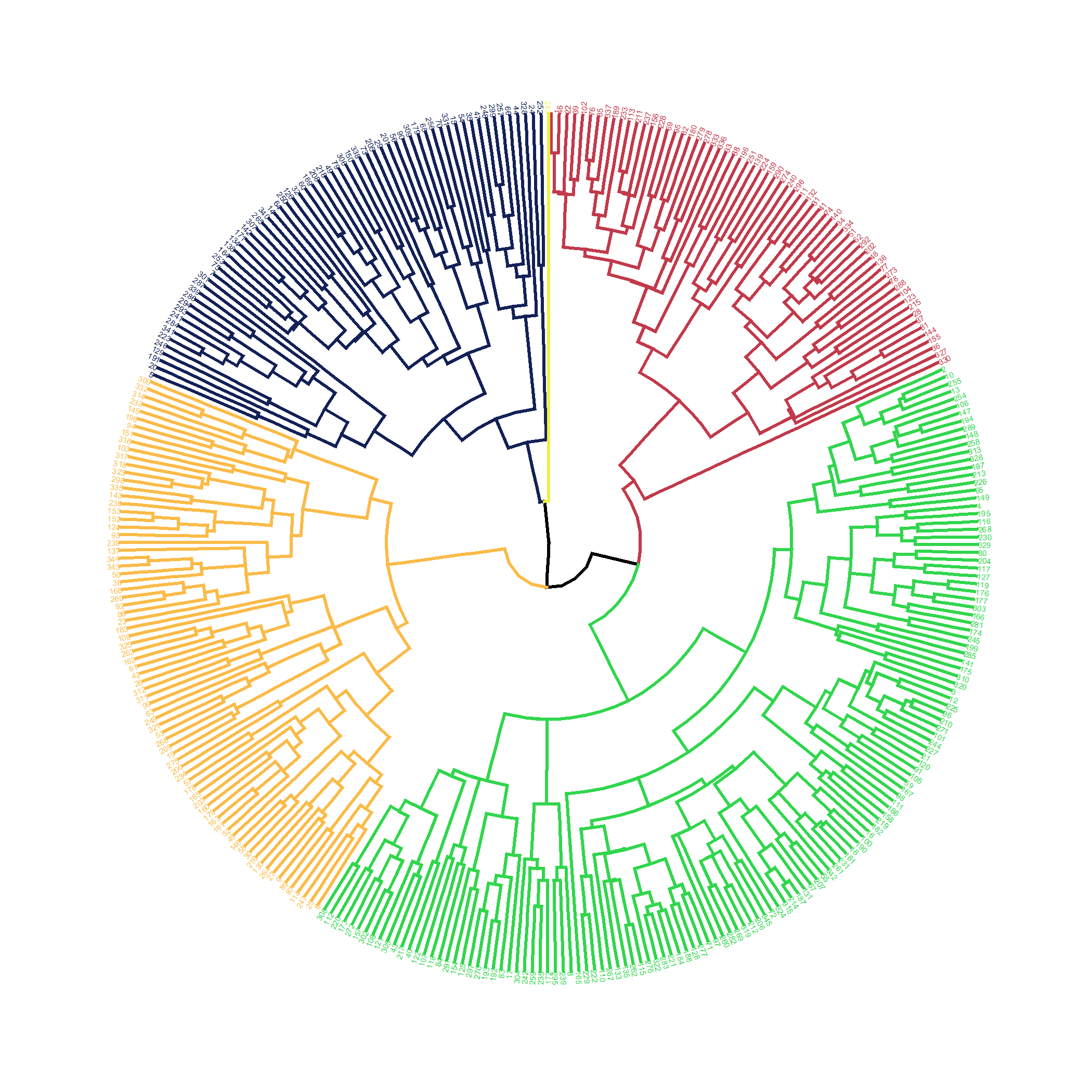
* languageSurvey dropped from consideration;
* Cluster 1 () has high positive MAPS scores and high SEPTI scores
* Cluster 2 () has high negative MAPS scores and high ECBI scores
* Cluster 3 () has low ECBI scores

## [1] 345 29

## [1] "zipcodeClass1" "zipcodeClass2"   
## [3] "communitySuburban" "communityRural"   
## [5] "distance" "income$25,001-$49,999"   
## [7] "income$50,000-$79,999" "income$80,000-$119,999"   
## [9] "income$120,000-$149,999" "income$150,000 or more"   
## [11] "internet" "ECBI\_intensity\_T\_score"   
## [13] "ECBI\_problem\_T\_score" "ECBI\_Opp"   
## [15] "ECBI\_Inatt" "ECBI\_Cond"   
## [17] "MAPS\_PP" "MAPS\_PR"   
## [19] "MAPS\_WM" "MAPS\_SP"   
## [21] "MAPS\_HS" "MAPS\_LC"   
## [23] "MAPS\_PC" "MAPS\_POS"   
## [25] "MAPS\_NEG" "SEPTI\_nurturance"   
## [27] "SEPTI\_discipline" "SEPTI\_play"   
## [29] "SEPTI\_routine"

## cluster size ave.sil.width  
## 1 1 63 0.10  
## 2 2 140 0.15  
## 3 3 64 0.02  
## 4 4 77 0.13  
## 5 5 1 0.00

* Hopkins statistic is 0.272
* Analysis identified clusters
* Divisive coefficient is 0.816
* Average silhouette width is 0.108



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| cluster | n | ECBI\_intensity\_T\_score\_mean | ECBI\_problem\_T\_score\_mean | ECBI\_Opp\_mean | ECBI\_Inatt\_mean | ECBI\_Cond\_mean |
| 1 | 63 | 52.8 | 52.5 | 32.3 | 13.1 | 14.2 |
| 2 | 140 | 52.9 | 52.5 | 32.5 | 13.7 | 14.3 |
| 3 | 64 | 59.7 | 61.8 | 40.6 | 15.2 | 21.7 |
| 4 | 77 | 50.5 | 50.0 | 30.5 | 11.9 | 13.6 |
| 5 | 1 | 78.0 | 84.0 | 53.0 | 23.0 | 41.0 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cluster | n | MAPS\_PP\_mean | MAPS\_PR\_mean | MAPS\_WM\_mean | MAPS\_SP\_mean | MAPS\_HS\_mean | MAPS\_LC\_mean | MAPS\_PC\_mean | MAPS\_POS\_mean | MAPS\_NEG\_mean |
| 1 | 63 | 3.8 | 4.2 | 4.4 | 4.0 | 2.1 | 2.0 | 1.2 | 4.1 | 1.8 |
| 2 | 140 | 4.3 | 4.7 | 4.8 | 4.7 | 1.9 | 1.8 | 1.3 | 4.6 | 1.7 |
| 3 | 64 | 3.6 | 4.0 | 4.1 | 3.8 | 2.8 | 2.6 | 2.3 | 3.9 | 2.6 |
| 4 | 77 | 4.2 | 4.7 | 4.8 | 4.6 | 1.9 | 1.8 | 1.5 | 4.6 | 1.8 |
| 5 | 1 | 4.3 | 5.0 | 3.0 | 4.7 | 3.6 | 3.9 | 1.8 | 4.2 | 3.1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| cluster | n | SEPTI\_nurturance\_mean | SEPTI\_discipline\_mean | SEPTI\_play\_mean | SEPTI\_routine\_mean |
| 1 | 63 | 35.0 | 21.9 | 29.4 | 28.1 |
| 2 | 140 | 39.3 | 25.2 | 34.6 | 30.1 |
| 3 | 64 | 33.5 | 19.2 | 25.8 | 23.7 |
| 4 | 77 | 39.1 | 25.6 | 33.8 | 30.4 |
| 5 | 1 | 29.0 | 13.0 | 11.0 | 20.0 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| cluster | n | distance\_mean | distance\_median | distance\_range |
| 1 | 63 | 5.7 | 5 | 1.0-20.0 |
| 2 | 140 | 6.5 | 5 | 0.4-30.0 |
| 3 | 64 | 12.3 | 9 | 1.0-150.0 |
| 4 | 77 | 17.5 | 14 | 1.0-130.0 |
| 5 | 1 | 10.0 | 10 | 10.0-10.0 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | zipcodeClass | n | pct |
| 1 | 1 | 55 | 0.87 |
| 1 | 2 | 8 | 0.13 |
| 2 | 1 | 139 | 0.99 |
| 2 | 2 | 1 | 0.01 |
| 3 | 1 | 53 | 0.83 |
| 3 | 2 | 11 | 0.17 |
| 4 | 1 | 2 | 0.03 |
| 4 | 2 | 75 | 0.97 |
| 5 | 1 | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | community | n | pct |
| 1 | Urban | 30 | 0.48 |
| 1 | Suburban | 30 | 0.48 |
| 1 | Rural | 3 | 0.05 |
| 2 | Urban | 55 | 0.39 |
| 2 | Suburban | 78 | 0.56 |
| 2 | Rural | 7 | 0.05 |
| 3 | Urban | 27 | 0.42 |
| 3 | Suburban | 29 | 0.45 |
| 3 | Rural | 8 | 0.12 |
| 4 | Urban | 10 | 0.13 |
| 4 | Suburban | 21 | 0.27 |
| 4 | Rural | 46 | 0.60 |
| 5 | Rural | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | income | n | pct |
| 1 | $25,000 or less | 7 | 0.11 |
| 1 | $25,001-$49,999 | 9 | 0.14 |
| 1 | $50,000-$79,999 | 16 | 0.25 |
| 1 | $80,000-$119,999 | 13 | 0.21 |
| 1 | $120,000-$149,999 | 7 | 0.11 |
| 1 | $150,000 or more | 11 | 0.17 |
| 2 | $25,000 or less | 13 | 0.09 |
| 2 | $25,001-$49,999 | 28 | 0.20 |
| 2 | $50,000-$79,999 | 32 | 0.23 |
| 2 | $80,000-$119,999 | 23 | 0.16 |
| 2 | $120,000-$149,999 | 14 | 0.10 |
| 2 | $150,000 or more | 30 | 0.21 |
| 3 | $25,000 or less | 7 | 0.11 |
| 3 | $25,001-$49,999 | 26 | 0.41 |
| 3 | $50,000-$79,999 | 12 | 0.19 |
| 3 | $80,000-$119,999 | 6 | 0.09 |
| 3 | $120,000-$149,999 | 8 | 0.12 |
| 3 | $150,000 or more | 5 | 0.08 |
| 4 | $25,000 or less | 7 | 0.09 |
| 4 | $25,001-$49,999 | 20 | 0.26 |
| 4 | $50,000-$79,999 | 31 | 0.40 |
| 4 | $80,000-$119,999 | 14 | 0.18 |
| 4 | $120,000-$149,999 | 3 | 0.04 |
| 4 | $150,000 or more | 2 | 0.03 |
| 5 | $25,000 or less | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | internet | n | pct |
| 1 | 1 | 63 | 1.00 |
| 2 | 0 | 1 | 0.01 |
| 2 | 1 | 139 | 0.99 |
| 3 | 0 | 2 | 0.03 |
| 3 | 1 | 62 | 0.97 |
| 4 | 0 | 5 | 0.06 |
| 4 | 1 | 72 | 0.94 |
| 5 | 0 | 1 | 1.00 |

## Cluster on parent factors

**Clustering on parent factors is terrible.**

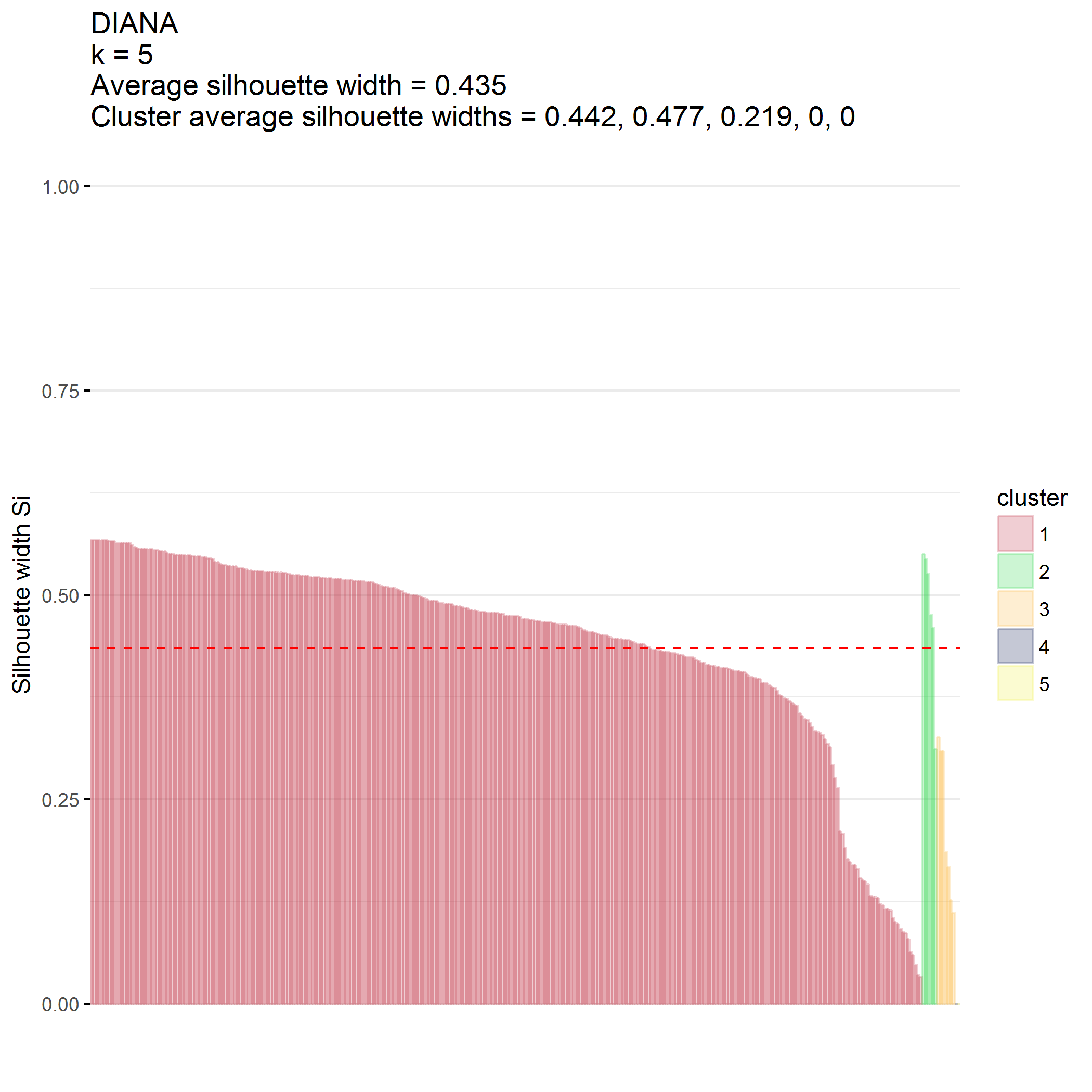
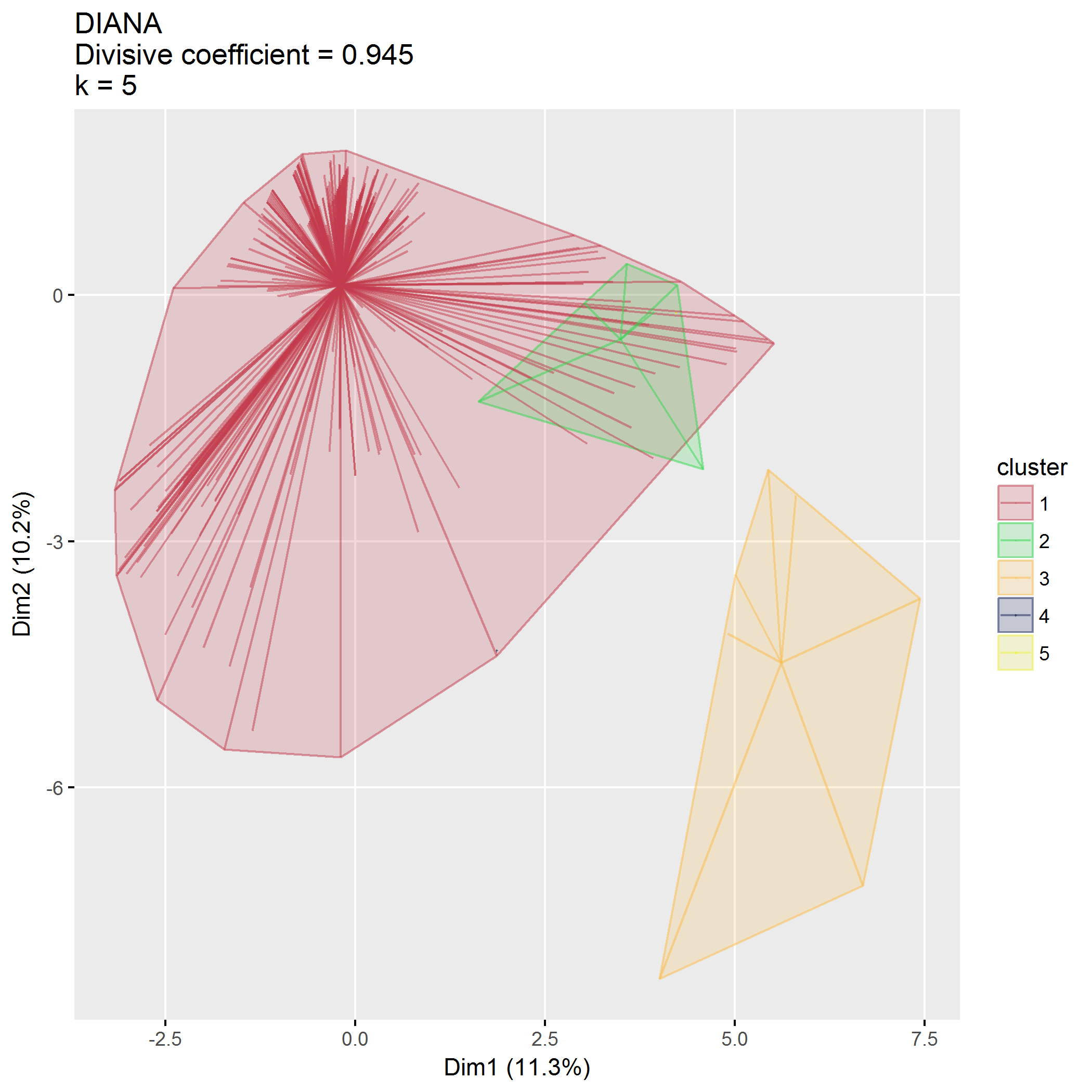
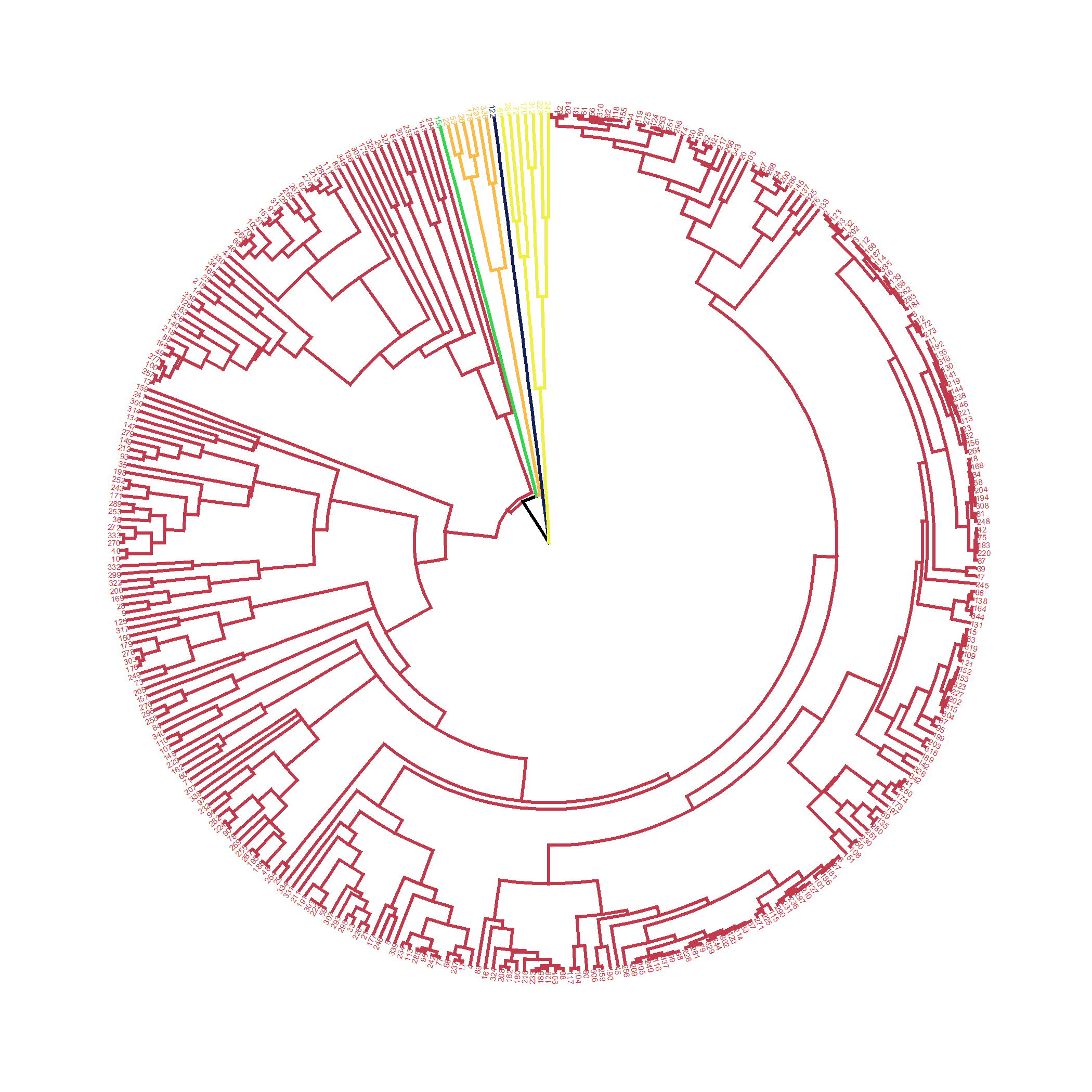
* First split, with , produces one very small cluster ().

## [1] 345 30

## [1] "totalChildren"   
## [2] "parentGenderMale"   
## [3] "parentGenderFemale"   
## [4] "parentGenderTransgender"   
## [5] "parentGenderOther"   
## [6] "parentGenderPrefer not to respond"   
## [7] "parentSexMale"   
## [8] "parentAge"   
## [9] "parentEthnicityNot Hispanic/Latino"   
## [10] "parentEthnicityUnknown"   
## [11] "parentEthnicityPrefer not to respond"   
## [12] "parentRaceWhite1"   
## [13] "parentRaceAsian1"   
## [14] "parentRaceAfrAm1"   
## [15] "parentRaceAIAN1"   
## [16] "parentRaceNHPI1"   
## [17] "parentRaceOther1"   
## [18] "parentRaceNoResp1"   
## [19] "parentMaritalStatusWidowed"   
## [20] "parentMaritalStatusDivorced"   
## [21] "parentMaritalStatusSeparated"   
## [22] "parentMaritalStatusRemarried"   
## [23] "parentMaritalStatusNever married"   
## [24] "parentSituationCouple parenting with spouse or partner in the same household"  
## [25] "parentSituationCo-parenting in separate households"   
## [26] "parentsNumber"   
## [27] "parentChildRatio"   
## [28] "parentEducationVocational school/some college"   
## [29] "parentEducationCollege"   
## [30] "parentEducationGraduate/professional school"

## cluster size ave.sil.width  
## 1 1 330 0.44  
## 2 2 6 0.48  
## 3 3 7 0.22  
## 4 4 1 0.00  
## 5 5 1 0.00

* Hopkins statistic is 0.078
* Analysis identified clusters
* Divisive coefficient is 0.945
* Average silhouette width is 0.435



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| cluster | n | totalChildren\_mean | totalChildren\_median | parentAge\_mean | parentAge\_median |
| 1 | 330 | 2.0 | 2.0 | 33.6 | 33.0 |
| 2 | 6 | 1.7 | 1.5 | 30.0 | 27.5 |
| 3 | 7 | 3.4 | 3.0 | 39.3 | 36.0 |
| 4 | 1 | 5.0 | 5.0 | 31.0 | 31.0 |
| 5 | 1 | 2.0 | 2.0 | 43.0 | 43.0 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | parentGender | n | pct |
| 1 | Male | 51 | 0.15 |
| 1 | Female | 278 | 0.84 |
| 1 | Transgender | 1 | 0.00 |
| 2 | Female | 4 | 0.67 |
| 2 | Prefer not to respond | 2 | 0.33 |
| 3 | Male | 1 | 0.14 |
| 3 | Female | 5 | 0.71 |
| 3 | Prefer not to respond | 1 | 0.14 |
| 4 | Other | 1 | 1.00 |
| 5 | Female | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | parentSituation | n | pct |
| 1 | Single parenting | 24 | 0.07 |
| 1 | Couple parenting with spouse or partner in the same household | 299 | 0.91 |
| 1 | Co-parenting in separate households | 7 | 0.02 |
| 2 | Single parenting | 5 | 0.83 |
| 2 | Couple parenting with spouse or partner in the same household | 1 | 0.17 |
| 3 | Single parenting | 1 | 0.14 |
| 3 | Co-parenting in separate households | 6 | 0.86 |
| 4 | Couple parenting with spouse or partner in the same household | 1 | 1.00 |
| 5 | Single parenting | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | parentEducation | n | pct |
| 1 | High school or less | 38 | 0.12 |
| 1 | Vocational school/some college | 62 | 0.19 |
| 1 | College | 135 | 0.41 |
| 1 | Graduate/professional school | 95 | 0.29 |
| 2 | High school or less | 3 | 0.50 |
| 2 | Vocational school/some college | 2 | 0.33 |
| 2 | Graduate/professional school | 1 | 0.17 |
| 3 | Vocational school/some college | 3 | 0.43 |
| 3 | College | 3 | 0.43 |
| 3 | Graduate/professional school | 1 | 0.14 |
| 4 | High school or less | 1 | 1.00 |
| 5 | Vocational school/some college | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | parentEthnicity | n | pct |
| 1 | Hispanic/Latino | 35 | 0.11 |
| 1 | Not Hispanic/Latino | 272 | 0.82 |
| 1 | Unknown | 4 | 0.01 |
| 1 | Prefer not to respond | 19 | 0.06 |
| 2 | Not Hispanic/Latino | 6 | 1.00 |
| 3 | Not Hispanic/Latino | 2 | 0.29 |
| 3 | Unknown | 2 | 0.29 |
| 3 | Prefer not to respond | 3 | 0.43 |
| 4 | Hispanic/Latino | 1 | 1.00 |
| 5 | Not Hispanic/Latino | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | parentRaceWhite | n | pct |
| 1 | 0 | 73 | 0.22 |
| 1 | 1 | 257 | 0.78 |
| 2 | 1 | 6 | 1.00 |
| 3 | 0 | 4 | 0.57 |
| 3 | 1 | 3 | 0.43 |
| 4 | 0 | 1 | 1.00 |
| 5 | 1 | 1 | 1.00 |

## Cluster on child factors

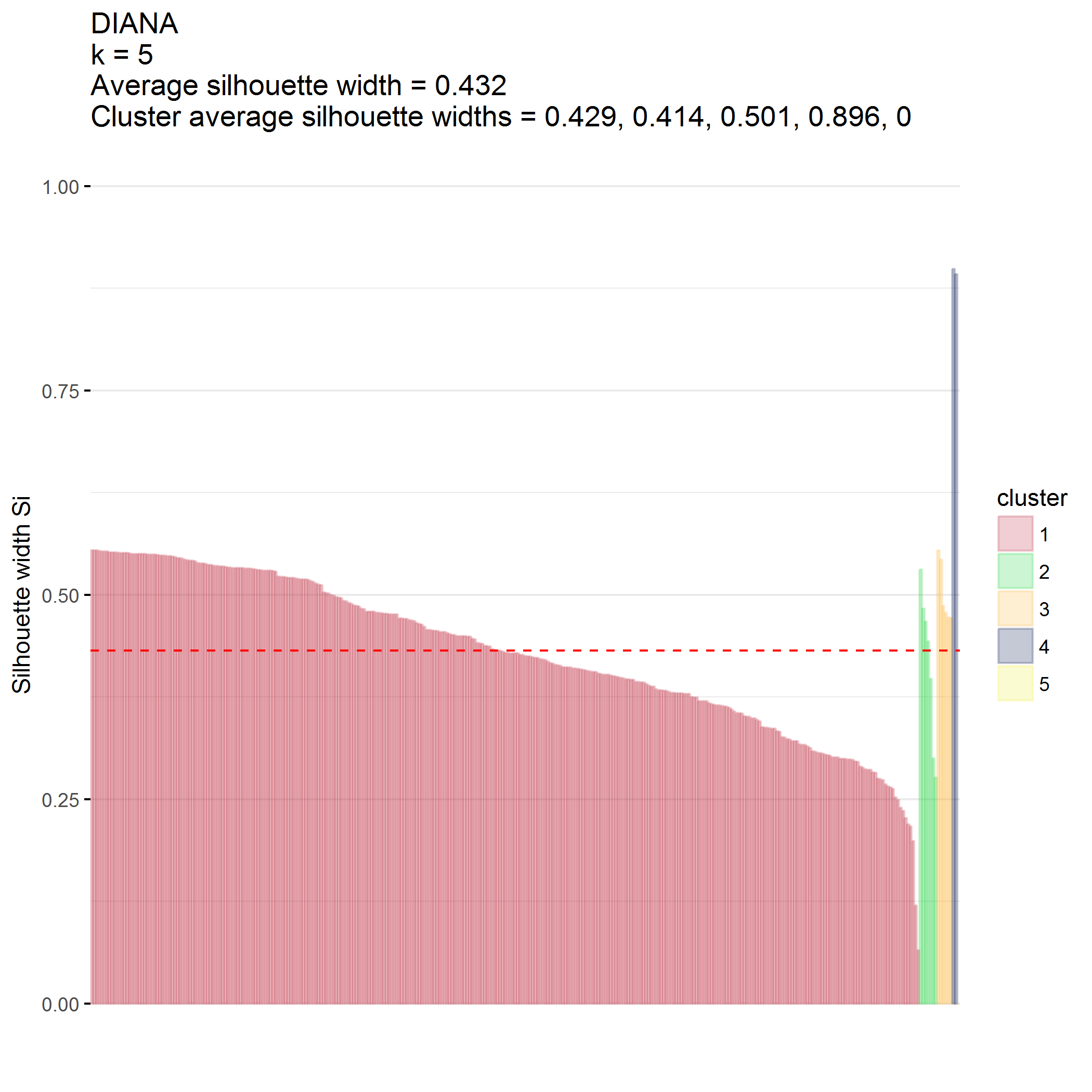
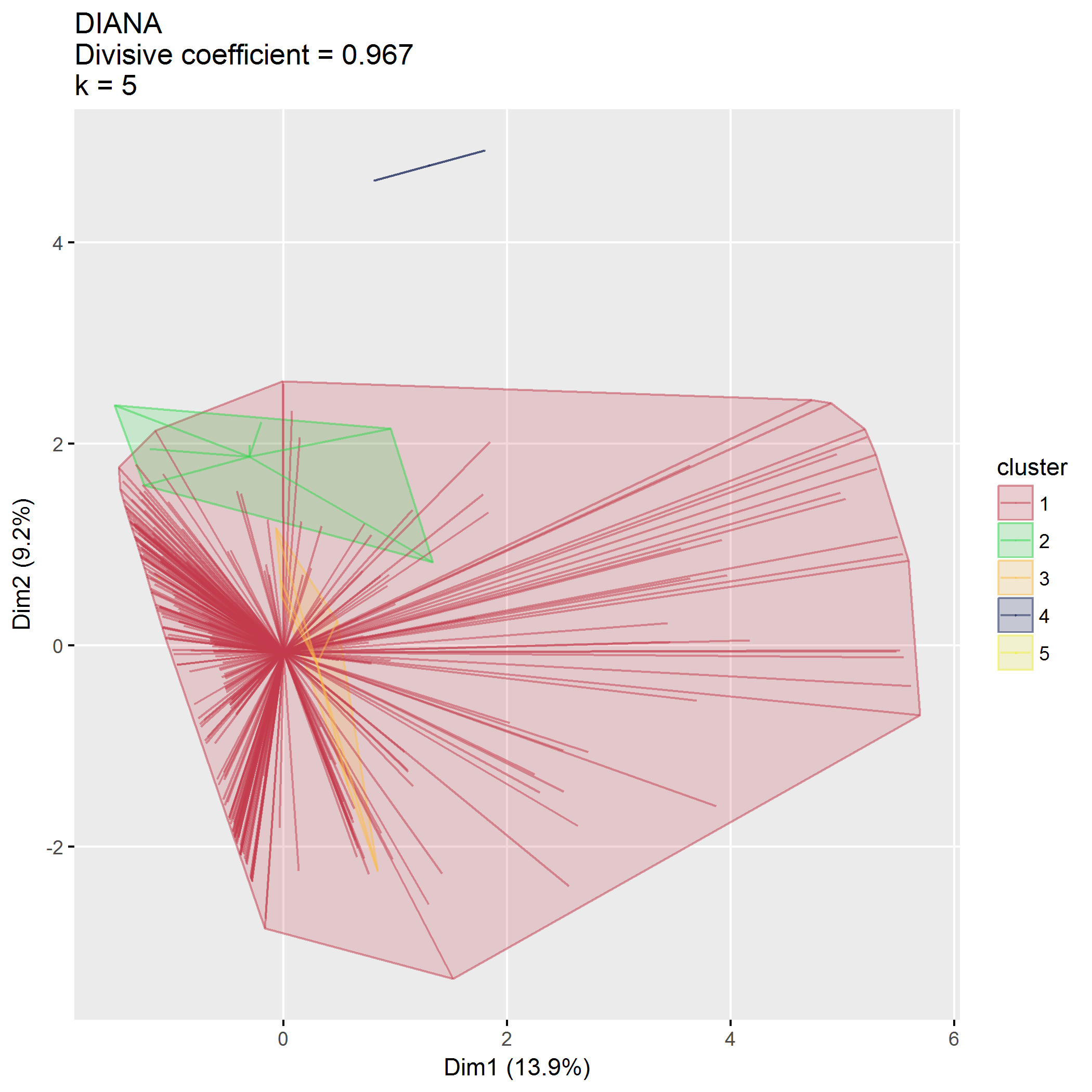
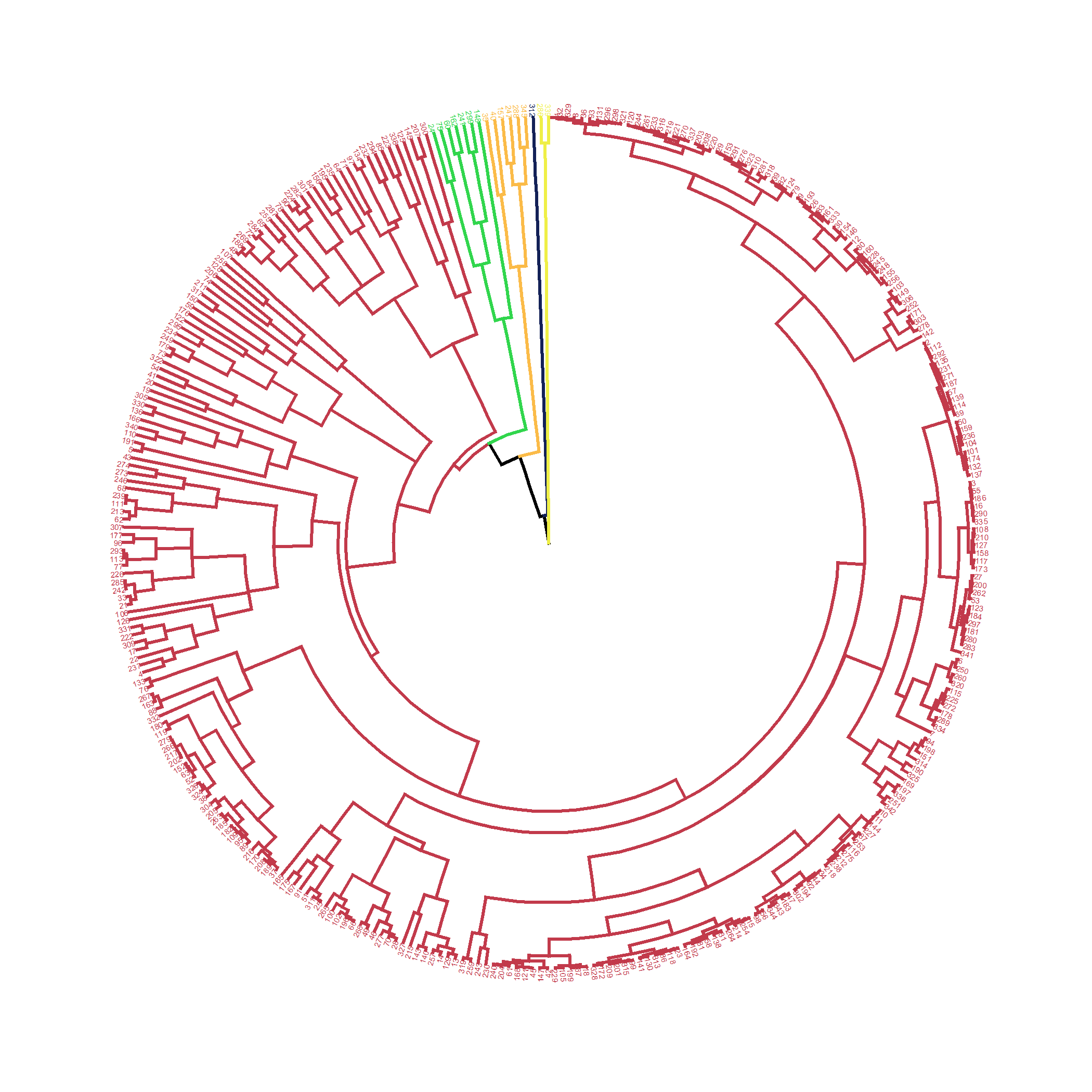
**Clustering on child factors is terrible.** First split, with , produces one very small cluster.

## [1] 345 19

## [1] "birthOrderOnly child"   
## [2] "birthOrderOldest"   
## [3] "birthOrderMiddle"   
## [4] "birthOrderYoungest"   
## [5] "childSexMale"   
## [6] "childAge"   
## [7] "childEthnicityNot Hispanic/Latino"   
## [8] "childEthnicityUnknown"   
## [9] "childEthnicityPrefer not to respond"   
## [10] "childRaceWhite1"   
## [11] "childRaceAsian1"   
## [12] "childRaceAfrAm1"   
## [13] "childRaceAIAN1"   
## [14] "childRaceNHPI1"   
## [15] "childRaceOther1"   
## [16] "childRaceNoResp1"   
## [17] "childRelationshipBiological or adoptive father"  
## [18] "childRelationshipGrandparent"   
## [19] "childRelationshipOther"

## cluster size ave.sil.width  
## 1 1 329 0.43  
## 2 2 7 0.41  
## 3 3 6 0.50  
## 4 4 2 0.90  
## 5 5 1 0.00

* Hopkins statistic is 0.065
* Analysis identified clusters
* Divisive coefficient is 0.967
* Average silhouette width is 0.432



|  |  |  |  |
| --- | --- | --- | --- |
| cluster | n | childAge\_mean | childAge\_median |
| 1 | 329 | 3.5 | 3.4 |
| 2 | 7 | 4.5 | 4.9 |
| 3 | 6 | 4.0 | 4.2 |
| 4 | 2 | 4.7 | 4.7 |
| 5 | 1 | 4.8 | 4.8 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | birthOrder | n | pct |
| 1 | Only child | 109 | 0.33 |
| 1 | Oldest | 90 | 0.27 |
| 1 | Middle | 38 | 0.12 |
| 1 | Youngest | 92 | 0.28 |
| 2 | Only child | 1 | 0.14 |
| 2 | Oldest | 4 | 0.57 |
| 2 | Middle | 2 | 0.29 |
| 3 | Only child | 2 | 0.33 |
| 3 | Oldest | 2 | 0.33 |
| 3 | Middle | 1 | 0.17 |
| 3 | Youngest | 1 | 0.17 |
| 4 | Middle | 2 | 1.00 |
| 5 | Youngest | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | childSex | n | pct |
| 1 | Female | 149 | 0.45 |
| 1 | Male | 180 | 0.55 |
| 2 | Female | 4 | 0.57 |
| 2 | Male | 3 | 0.43 |
| 3 | Female | 4 | 0.67 |
| 3 | Male | 2 | 0.33 |
| 4 | Male | 2 | 1.00 |
| 5 | Female | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | childEthnicity | n | pct |
| 1 | Hispanic/Latino | 42 | 0.13 |
| 1 | Not Hispanic/Latino | 261 | 0.79 |
| 1 | Prefer not to respond | 26 | 0.08 |
| 2 | Hispanic/Latino | 4 | 0.57 |
| 2 | Not Hispanic/Latino | 3 | 0.43 |
| 3 | Unknown | 6 | 1.00 |
| 4 | Hispanic/Latino | 1 | 0.50 |
| 4 | Not Hispanic/Latino | 1 | 0.50 |
| 5 | Not Hispanic/Latino | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | childRaceWhite | n | pct |
| 1 | 0 | 63 | 0.19 |
| 1 | 1 | 266 | 0.81 |
| 2 | 0 | 2 | 0.29 |
| 2 | 1 | 5 | 0.71 |
| 3 | 1 | 6 | 1.00 |
| 4 | 0 | 2 | 1.00 |
| 5 | 1 | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | childRelationship | n | pct |
| 1 | Biological or adoptive mother | 285 | 0.87 |
| 1 | Biological or adoptive father | 44 | 0.13 |
| 2 | Biological or adoptive mother | 6 | 0.86 |
| 2 | Biological or adoptive father | 1 | 0.14 |
| 3 | Biological or adoptive mother | 6 | 1.00 |
| 4 | Other | 2 | 1.00 |
| 5 | Grandparent | 1 | 1.00 |

## Cluster on demographic factors

**Clustering on demographic factors is terrible.**

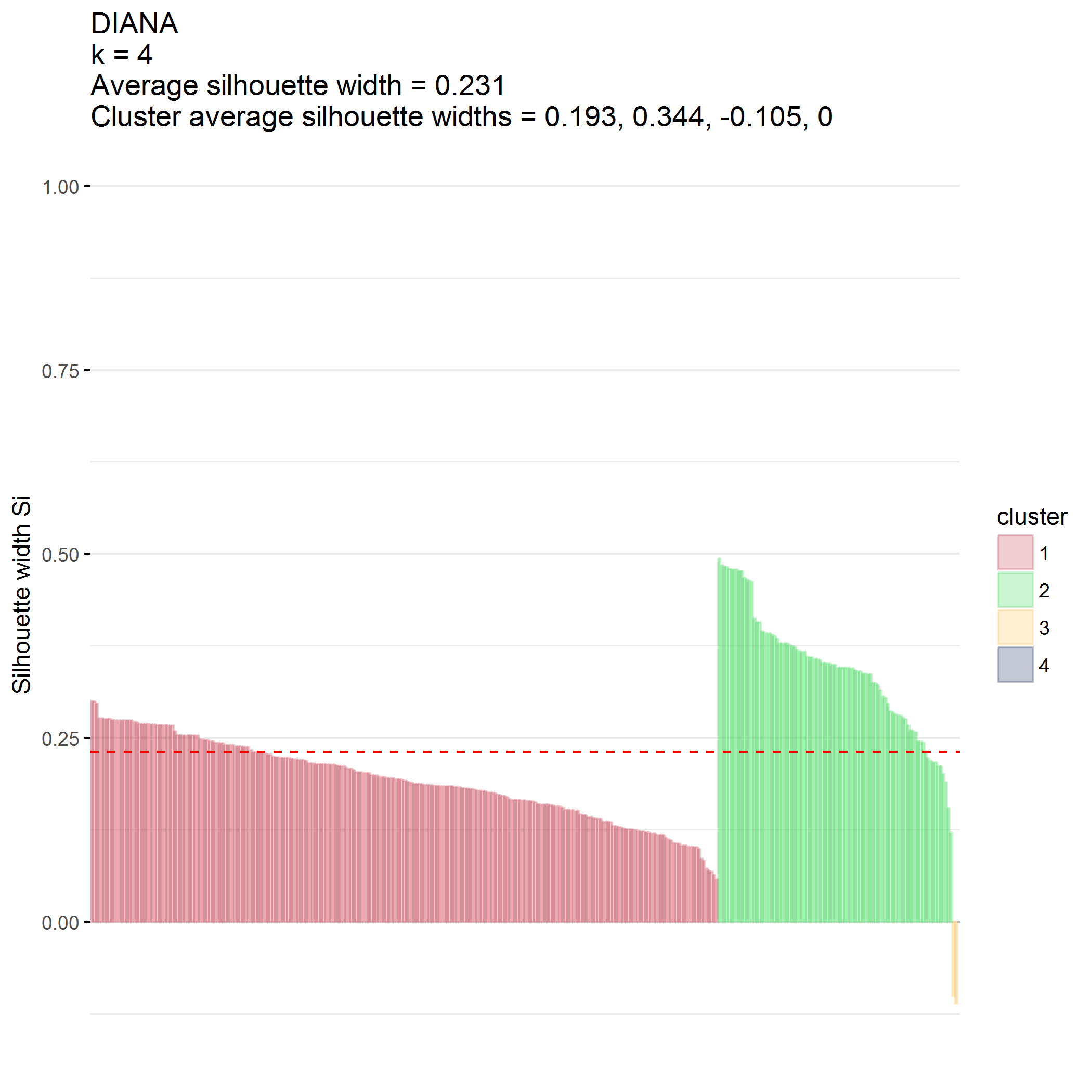
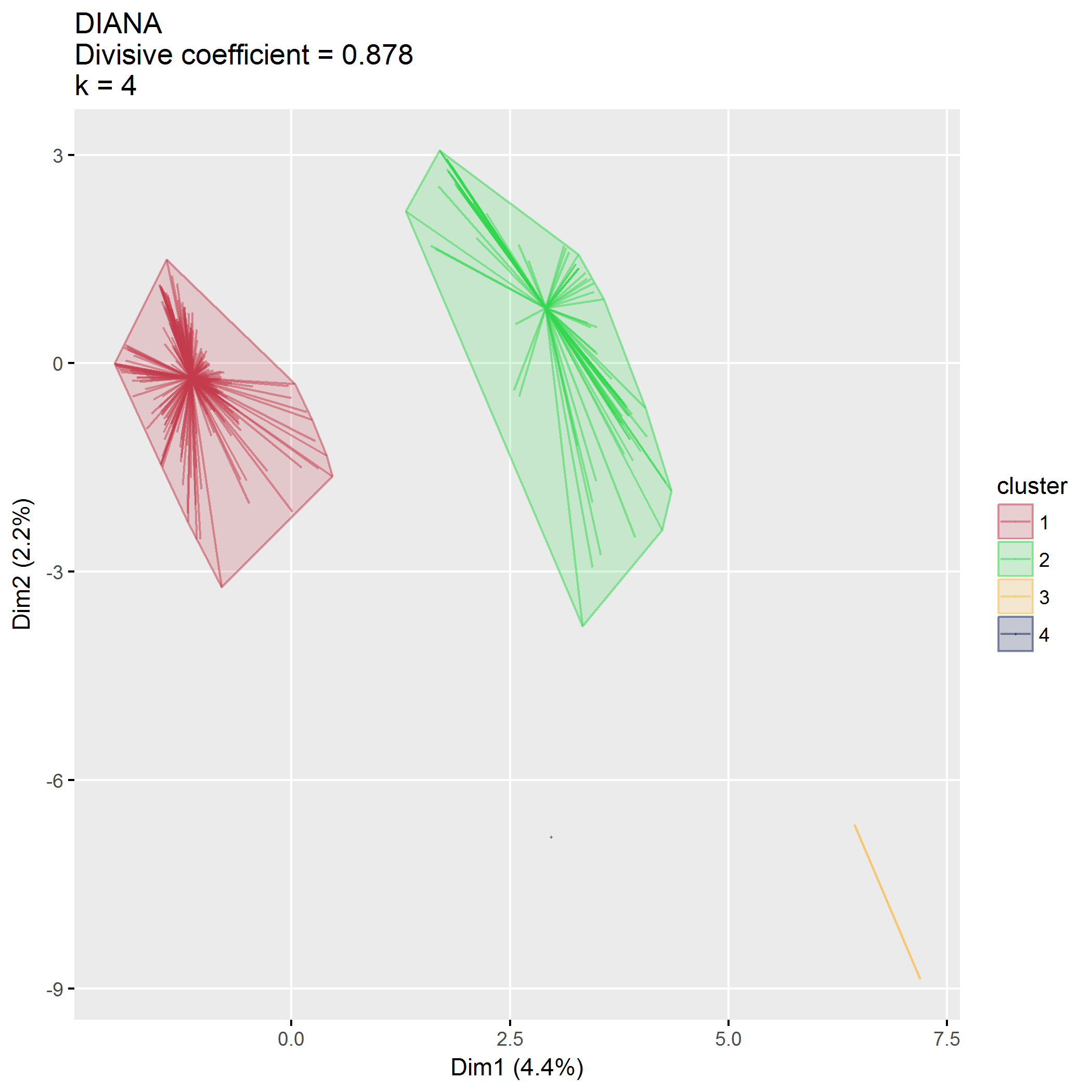
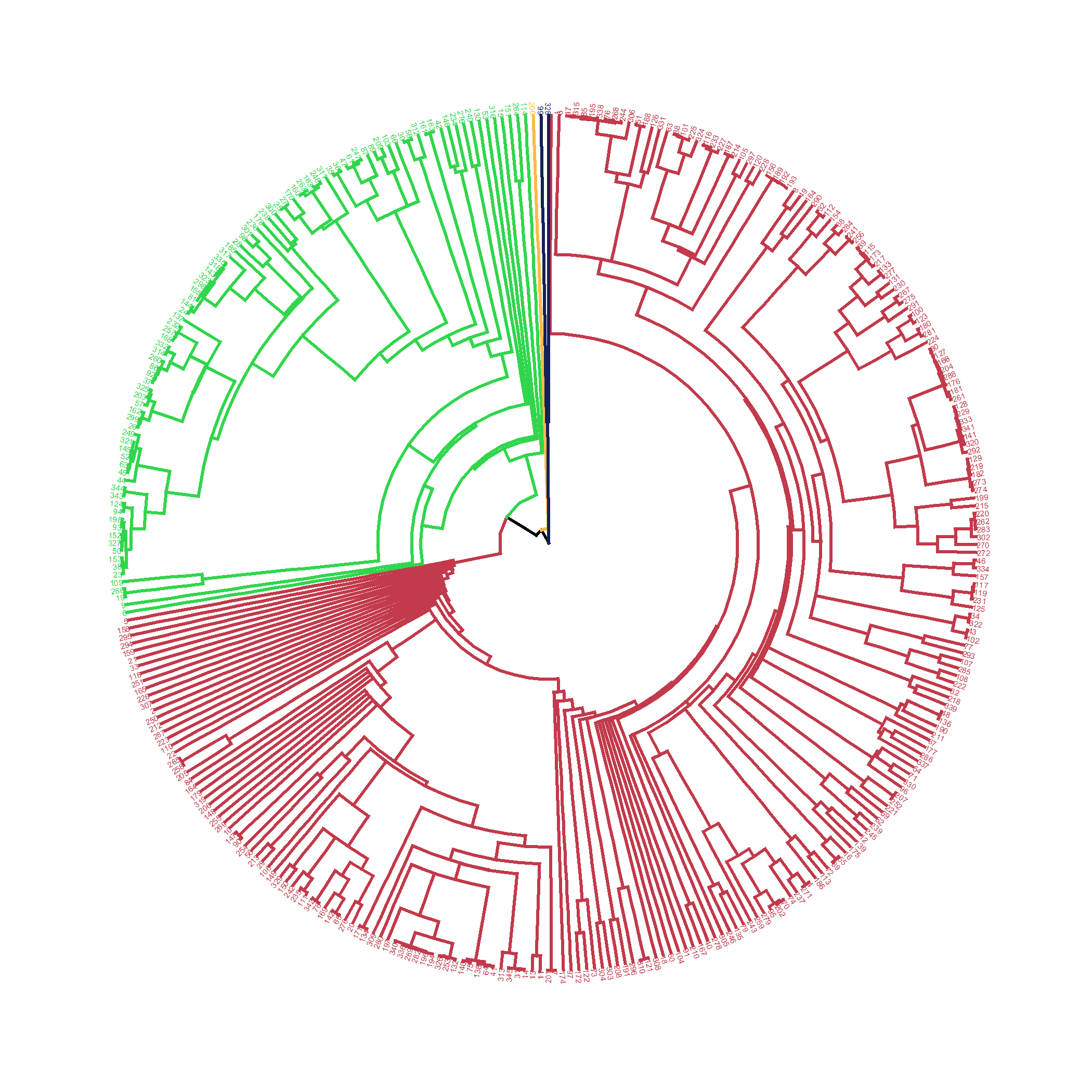
* languageSurvey dropped from consideration;
* First two splits, with , produces two very small clusters
* The next split, at , produces a good sized cluster
* The clustering between clusters 1 () and 2 () appears to have to do more with geography (distance, zipcode class, urban/rural)

## [1] 345 88

## [1] "zipcodeClass1" "zipcodeClass2"   
## [3] "zipcode91020" "zipcode91204"   
## [5] "zipcode91206" "zipcode91210"   
## [7] "zipcode91402" "zipcode97003"   
## [9] "zipcode97006" "zipcode97007"   
## [11] "zipcode97008" "zipcode97023"   
## [13] "zipcode97027" "zipcode97032"   
## [15] "zipcode97034" "zipcode97035"   
## [17] "zipcode97045" "zipcode97056"   
## [19] "zipcode97060" "zipcode97062"   
## [21] "zipcode97068" "zipcode97071"   
## [23] "zipcode97078" "zipcode97086"   
## [25] "zipcode97089" "zipcode97101"   
## [27] "zipcode97116" "zipcode97123"   
## [29] "zipcode97124" "zipcode97140"   
## [31] "zipcode97141" "zipcode97201"   
## [33] "zipcode97202" "zipcode97203"   
## [35] "zipcode97206" "zipcode97209"   
## [37] "zipcode97210" "zipcode97211"   
## [39] "zipcode97212" "zipcode97213"   
## [41] "zipcode97214" "zipcode97215"   
## [43] "zipcode97217" "zipcode97219"   
## [45] "zipcode97220" "zipcode97221"   
## [47] "zipcode97222" "zipcode97223"   
## [49] "zipcode97224" "zipcode97225"   
## [51] "zipcode97227" "zipcode97229"   
## [53] "zipcode97230" "zipcode97232"   
## [55] "zipcode97233" "zipcode97236"   
## [57] "zipcode97239" "zipcode97266"   
## [59] "zipcode97267" "zipcode97321"   
## [61] "zipcode97325" "zipcode97429"   
## [63] "zipcode97527" "zipcode97701"   
## [65] "zipcode97702" "zipcode97703"   
## [67] "zipcode97707" "zipcode97734"   
## [69] "zipcode97738" "zipcode97741"   
## [71] "zipcode97753" "zipcode97754"   
## [73] "zipcode97756" "zipcode97759"   
## [75] "zipcode97760" "zipcode98632"   
## [77] "zipcode98660" "zipcode98683"   
## [79] "zipcode98685" "communitySuburban"   
## [81] "communityRural" "distance"   
## [83] "income$25,001-$49,999" "income$50,000-$79,999"   
## [85] "income$80,000-$119,999" "income$120,000-$149,999"  
## [87] "income$150,000 or more" "internet"

## cluster size ave.sil.width  
## 1 1 249 0.19  
## 2 2 93 0.34  
## 3 3 2 -0.11  
## 4 4 1 0.00

* Hopkins statistic is 0.041
* Analysis identified clusters
* Divisive coefficient is 0.878
* Average silhouette width is 0.231



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| cluster | n | distance\_mean | distance\_median | distance\_range |
| 1 | 249 | 6.7 | 5 | 0.4-30.0 |
| 2 | 93 | 15.1 | 12 | 1.0-100.0 |
| 3 | 2 | 140.0 | 140 | 130.0-150.0 |
| 4 | 1 | 75.0 | 75 | 75.0-75.0 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | zipcodeClass | n | pct |
| 1 | 1 | 249 | 1 |
| 2 | 2 | 93 | 1 |
| 3 | 2 | 2 | 1 |
| 4 | 1 | 1 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | community | n | pct |
| 1 | Urban | 109 | 0.44 |
| 1 | Suburban | 128 | 0.51 |
| 1 | Rural | 12 | 0.05 |
| 2 | Urban | 13 | 0.14 |
| 2 | Suburban | 30 | 0.32 |
| 2 | Rural | 50 | 0.54 |
| 3 | Rural | 2 | 1.00 |
| 4 | Rural | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | income | n | pct |
| 1 | $25,000 or less | 25 | 0.10 |
| 1 | $25,001-$49,999 | 57 | 0.23 |
| 1 | $50,000-$79,999 | 58 | 0.23 |
| 1 | $80,000-$119,999 | 38 | 0.15 |
| 1 | $120,000-$149,999 | 27 | 0.11 |
| 1 | $150,000 or more | 44 | 0.18 |
| 2 | $25,000 or less | 10 | 0.11 |
| 2 | $25,001-$49,999 | 26 | 0.28 |
| 2 | $50,000-$79,999 | 31 | 0.33 |
| 2 | $80,000-$119,999 | 17 | 0.18 |
| 2 | $120,000-$149,999 | 5 | 0.05 |
| 2 | $150,000 or more | 4 | 0.04 |
| 3 | $50,000-$79,999 | 1 | 0.50 |
| 3 | $80,000-$119,999 | 1 | 0.50 |
| 4 | $50,000-$79,999 | 1 | 1.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| cluster | internet | n | pct |
| 1 | 0 | 4 | 0.02 |
| 1 | 1 | 245 | 0.98 |
| 2 | 0 | 4 | 0.04 |
| 2 | 1 | 89 | 0.96 |
| 3 | 1 | 2 | 1.00 |
| 4 | 0 | 1 | 1.00 |

## Save objects

## size isdir mode  
## data/processed/clusterAnalysis.RData 11984348 FALSE 666  
## mtime  
## data/processed/clusterAnalysis.RData 2018-07-10 07:56:24  
## ctime  
## data/processed/clusterAnalysis.RData 2018-07-06 12:14:44  
## atime exe  
## data/processed/clusterAnalysis.RData 2018-07-06 12:14:44 no