

HDBSCAN

With Log-Transformed Data

PMS

24 May, 2023

Introduction

The following analysis is an implementation of the [HDBSCAN](#) algorithm on individual, log-scaled proximity measures (no auxiliary variables included).

Please note that for the following plots, HDBSCAN considers cluster 1 to be “noise points” and are thus not part of an actual cluster.

Assumptions of the Algorithm

[This](#) fast implementation of HDBSCAN (Campello et al., 2013) computes the hierarchical cluster tree representing density estimates along with the stability-based flat cluster extraction. HDBSCAN essentially computes the hierarchy of all DBSCAN* clusterings, and then uses a stability-based extraction method to find optimal cuts in the hierarchy, thus producing a flat solution.

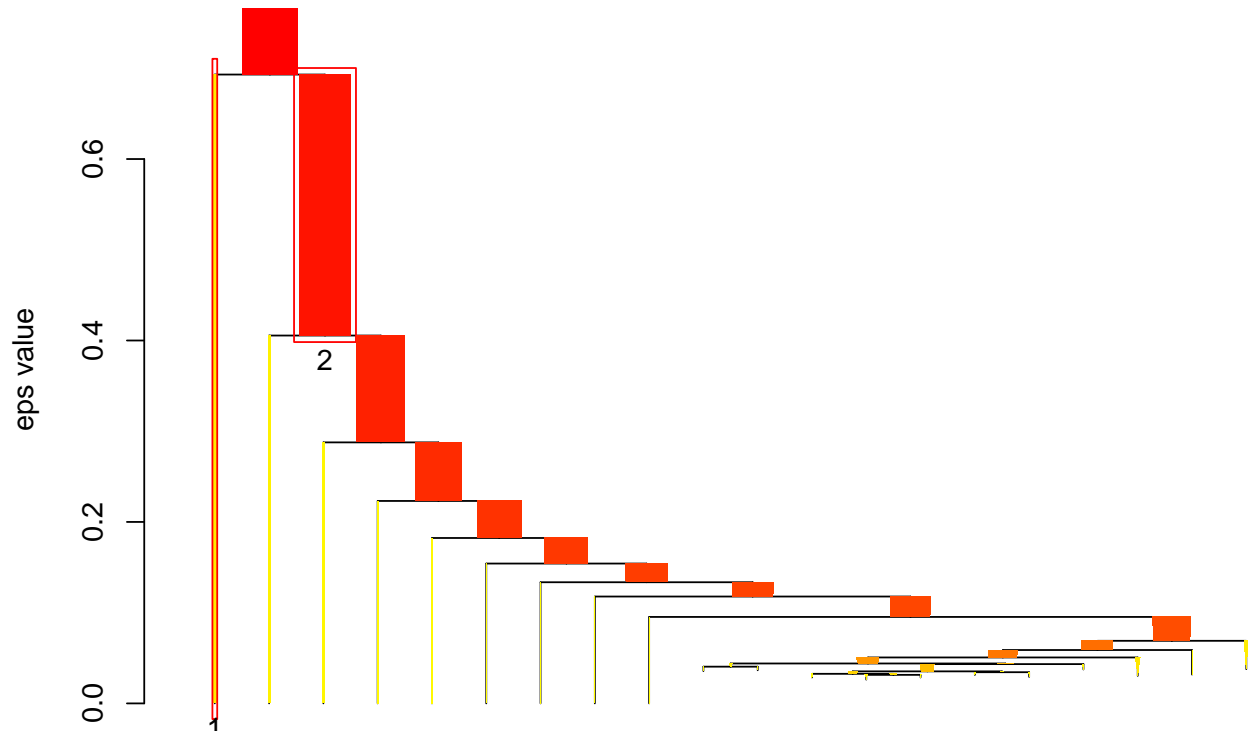
HDBSCAN performs the following steps:

- Compute mutual reachability distance mrd between points (based on distances and core distances).
 - Use mdr as a distance measure to construct a minimum spanning tree.
 - Prune the tree using stability.
 - Extract the clusters.
-

Amenities

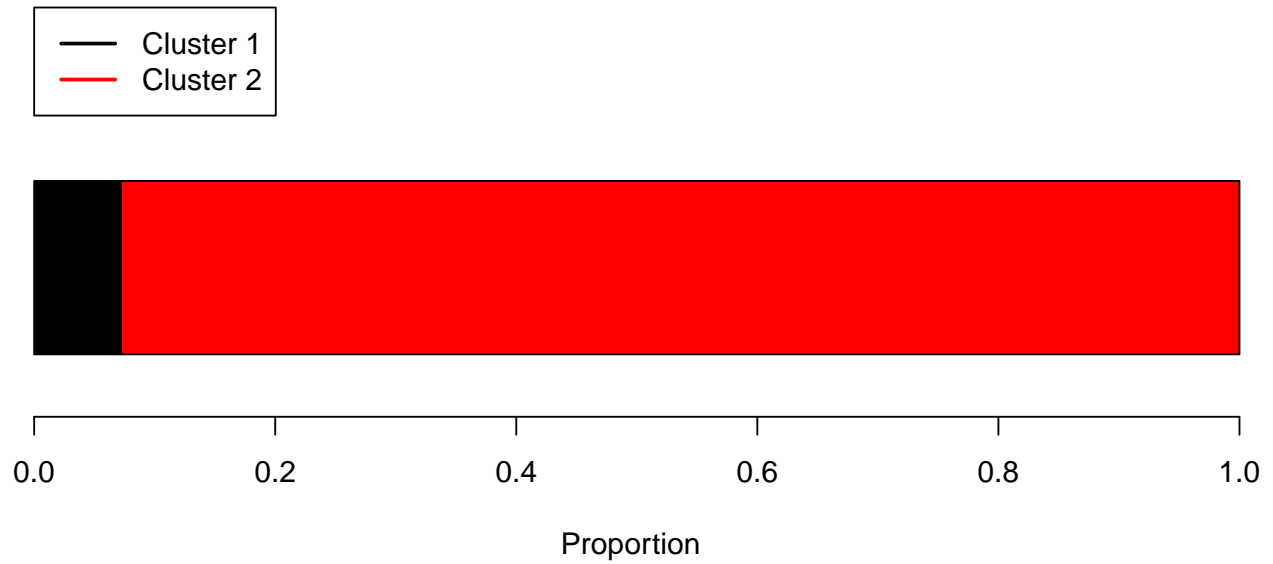
Employment

HDBSCAN*



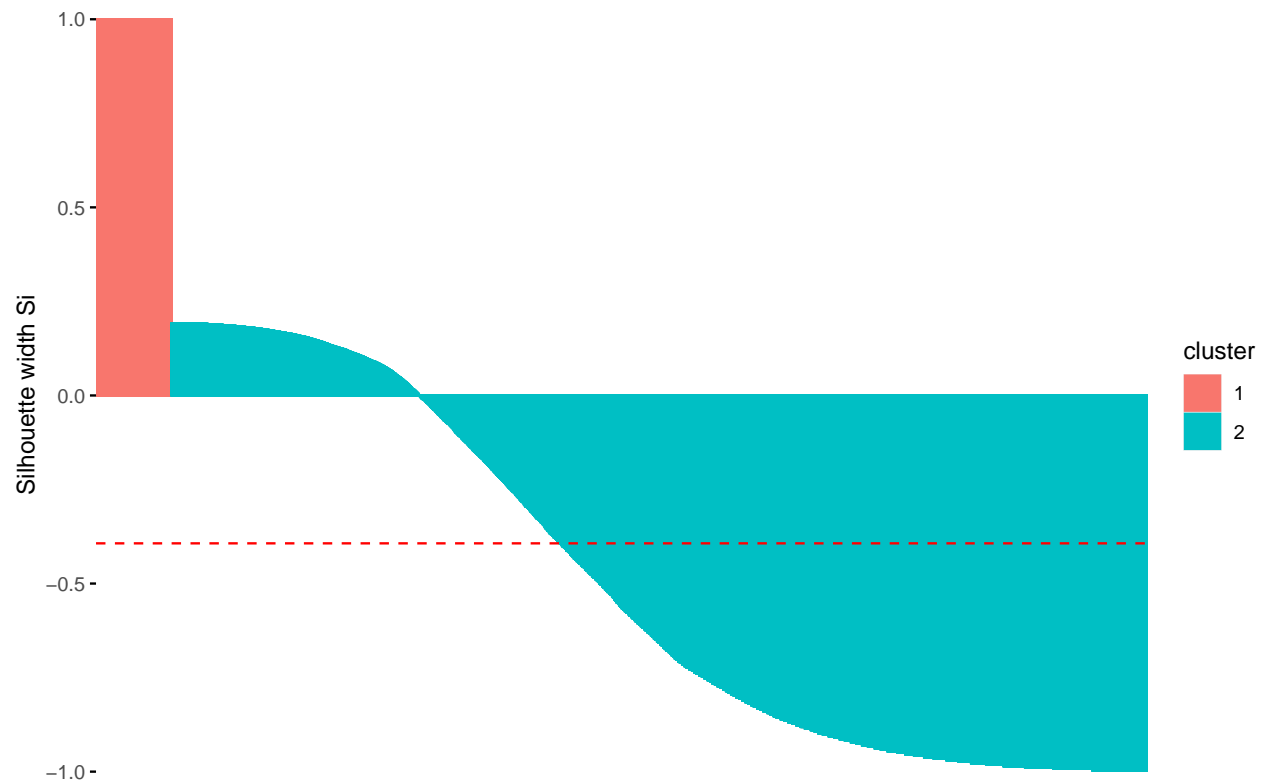
```
## [1] "Silhouette coefficient: 0.687337430388787"  
## [1] "Xie Beni coefficient: 18775.3001891401"  
## [1] "Davies Bouldin coefficient: 0.401367542912651"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.21475  
##   cluster  size ave.sil.width  
## 1         1  1521          1.0  
## 2         2 19594         -0.5
```

Clusters silhouette plot
Average silhouette width: -0.39



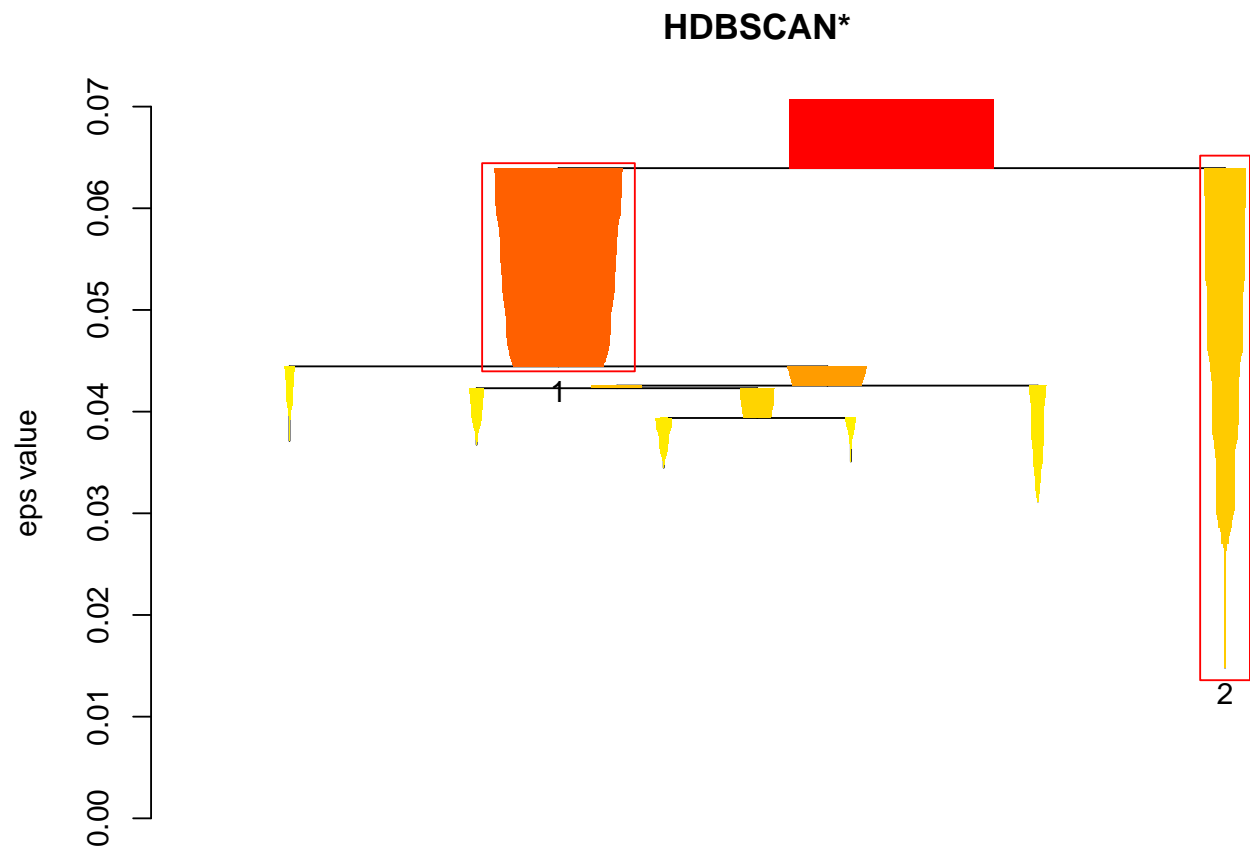
```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
## Cluster 1 Cluster 2
##      1778      22705
##
##
##
## DB Population:
## Cluster 1 Cluster 2
##      72.8      71.4
##
##
##
## CSD Population:
## Cluster 1 Cluster 2
## 235089.4 240777.8
##
##
##
## CMA Type:
## Cluster 1 Cluster 2
##      784      9843
## B      752      9560
## D      176      2484
## K       66       818
```

```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.226      0.226
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              55      708
## BritishColumbia      87     1112
## NewBrunswick          22      185
## NorthwestTerritories   1       15
## NovaScotia            60      719
## Ontario              283     3575
## Quebec                102     1276
## Saskatchewan           7      114
## NA's                 1161    15001
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      1610    20508
## 1       128     1687
## 2        21     246
## F        19     264
##
##
##
## PMS_prox_idx_emp :
## Cluster 1 Cluster 2
##      0.02622   0.02559

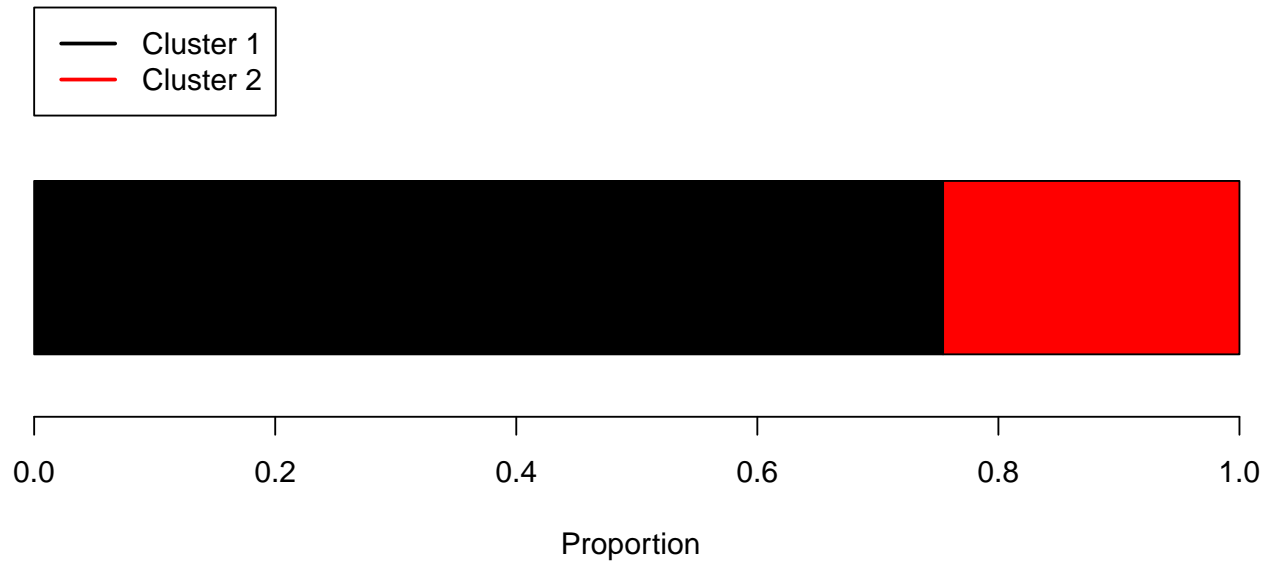
```

Pharmacy



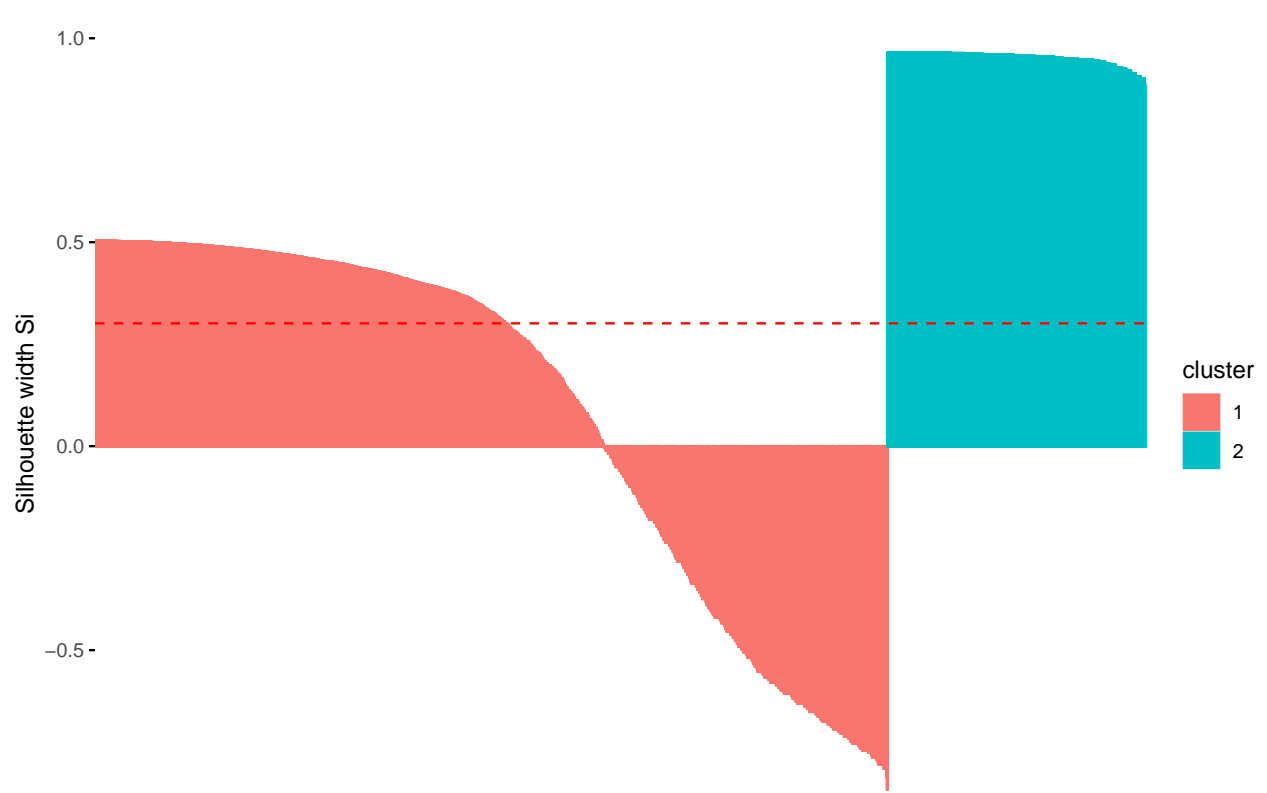
```
## [1] "Silhouette coefficient: 0.434971414389449"  
## [1] "Xie Beni coefficient: Inf"  
## [1] "Davies Bouldin coefficient: 1.20826229805656"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.0114  
##   cluster size ave.sil.width  
## 1         1 5527         0.09  
## 2         2 1802         0.95
```

Clusters silhouette plot
Average silhouette width: 0.3



```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##   Cluster 1 Cluster 2
##      18481      6002
##
##
##
## DB Population:
##   Cluster 1 Cluster 2
##        72      69.8
##
##
##
## CSD Population:
##   Cluster 1 Cluster 2
##  240832.9  238924.7
##
##
##
## CMA Type:
##   Cluster 1 Cluster 2
##        8013      2614
## B        7776      2536
## D        2032       628
## K         660       224
```



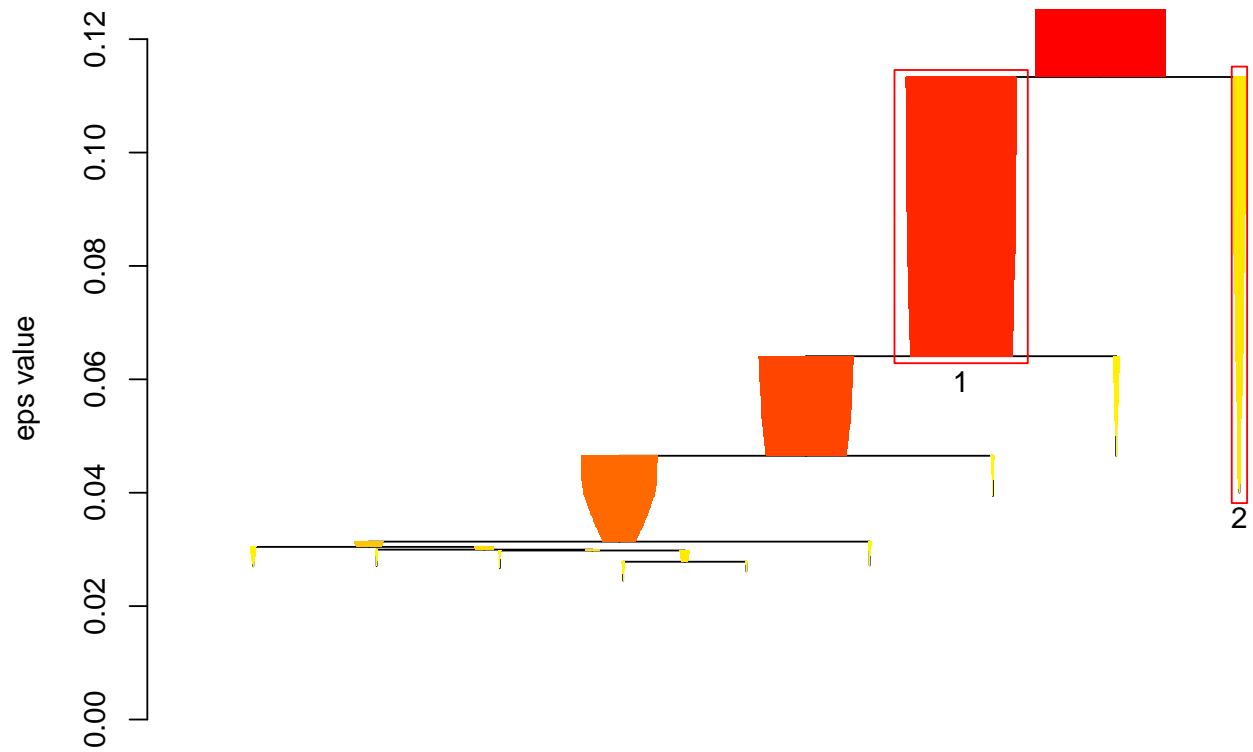
```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.226      0.226
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              565      198
## BritishColumbia      903      296
## NewBrunswick          154       53
## NorthwestTerritories   11        5
## NovaScotia            565      214
## Ontario              2907      951
## Quebec               1057      321
## Saskatchewan           83       38
## NA's                 12236     3926
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      16668      5450
## 1       1377       438
## 2        215        52
## F        221        62
##
##
##
## PMS_prox_idx_pharma :
## Cluster 1 Cluster 2
##      0.0445      0.04448

```

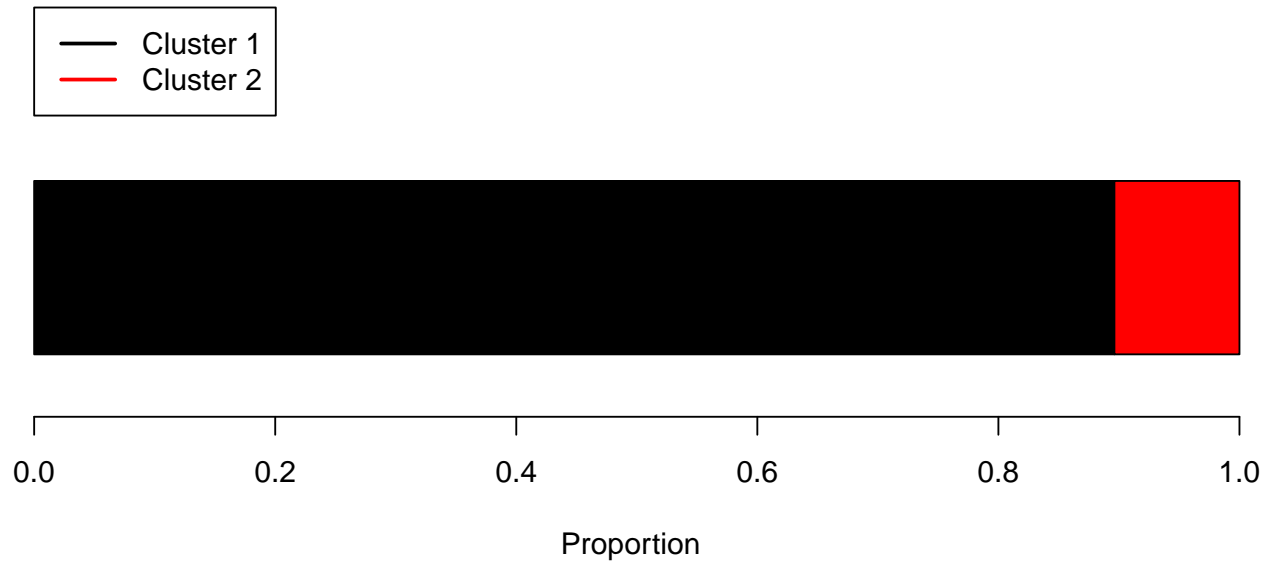
Childcare

HDBSCAN*



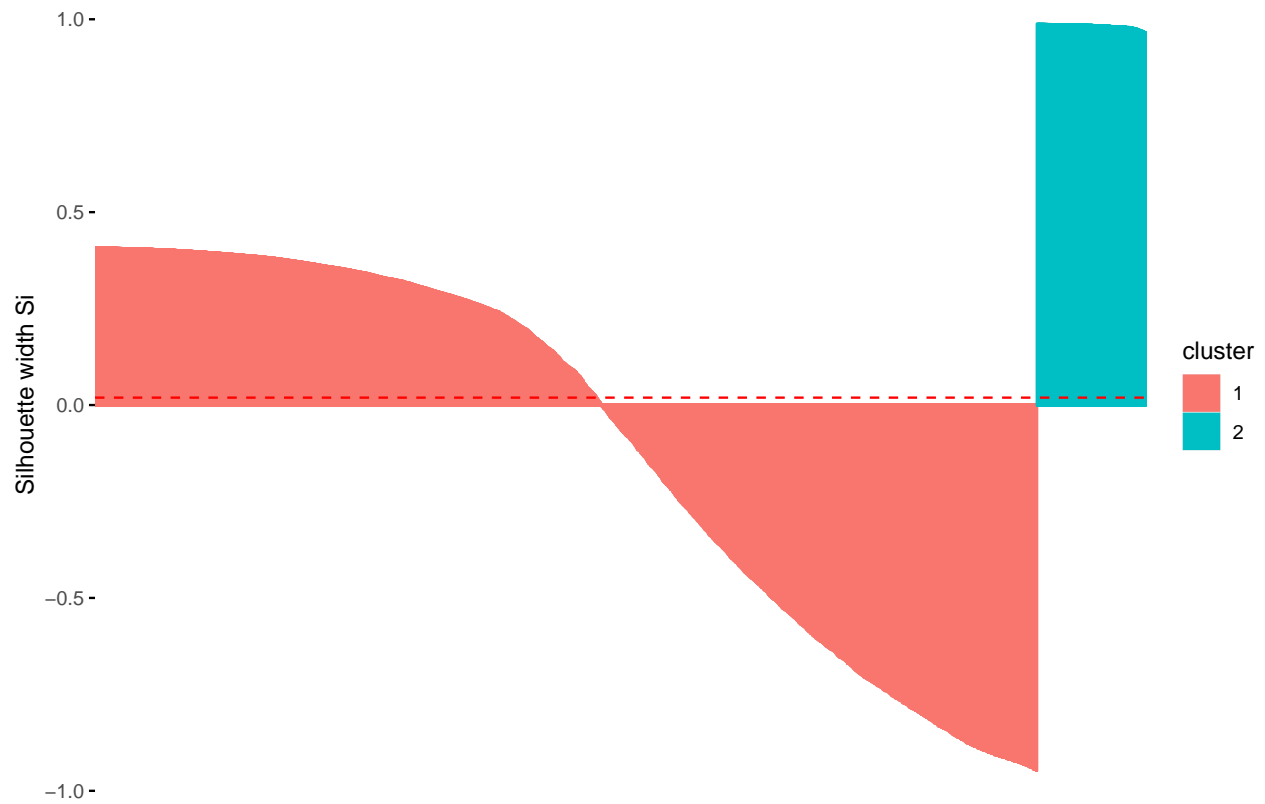
```
## [1] "Silhouette coefficient: 0.410851184885891"  
## [1] "Xie Beni coefficient: Inf"  
## [1] "Davies Bouldin coefficient: 2.15897203245473"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.0087  
##   cluster  size ave.sil.width  
## 1       1 10330        -0.09  
## 2       2  1196         0.98
```

Clusters silhouette plot
Average silhouette width: 0.02



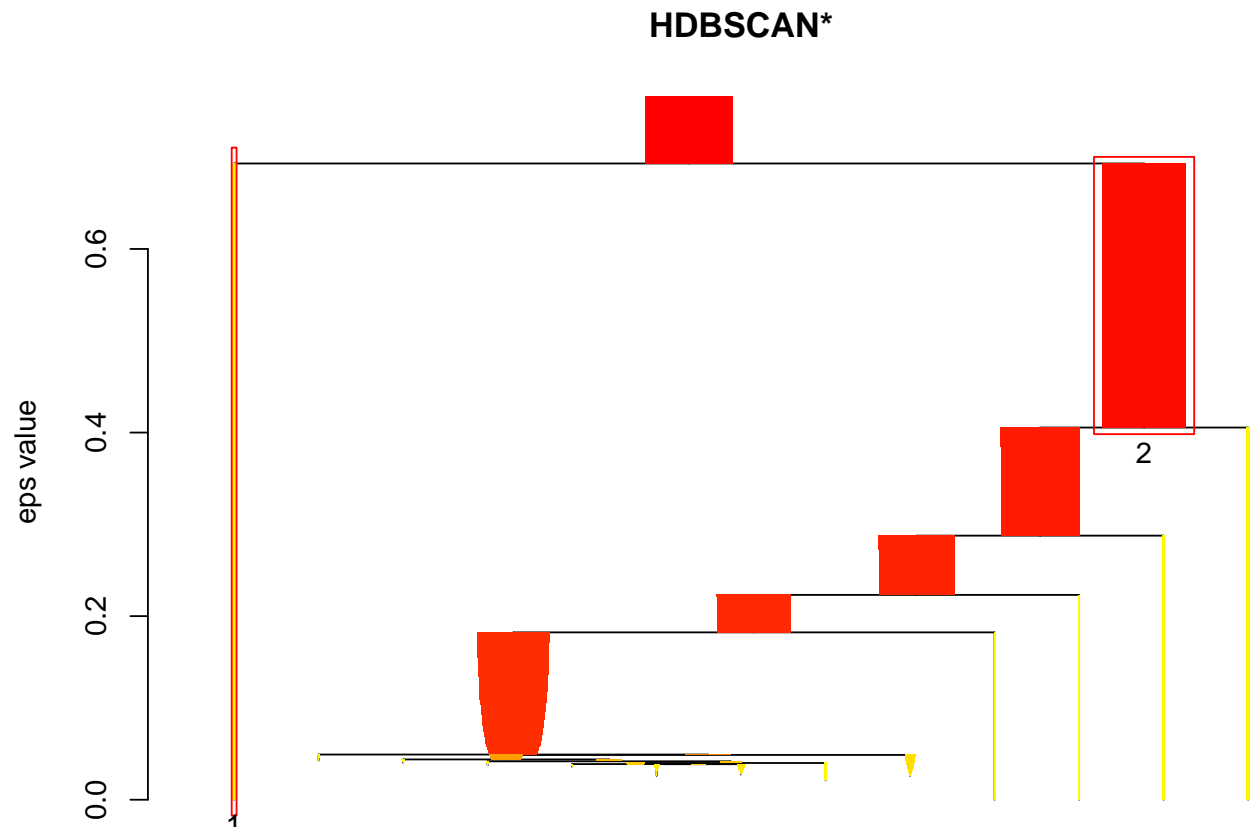
```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##   Cluster 1 Cluster 2
##      21952      2531
##
##
##
## DB Population:
##   Cluster 1 Cluster 2
##       71.3      72.8
##
##
##
## CSD Population:
##   Cluster 1 Cluster 2
##  240439.2  239721.1
##
##
##
## CMA Type:
##   Cluster 1 Cluster 2
##      9521      1106
## B      9248      1064
## D      2404       256
## K       779       105
```

```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.226      0.225
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              685       78
## BritishColumbia      1094      105
## NewBrunswick          186       21
## NorthwestTerritories   13        3
## NovaScotia            700       79
## Ontario              3448      410
## Quebec               1241      137
## Saskatchewan          109       12
## NA's                 14476     1686
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      19833      2285
## 1       1625       190
## 2        239        28
## F        255        28
##
##
##
## PMS_prox_idx_childcare :
## Cluster 1 Cluster 2
##      0.07675    0.07711

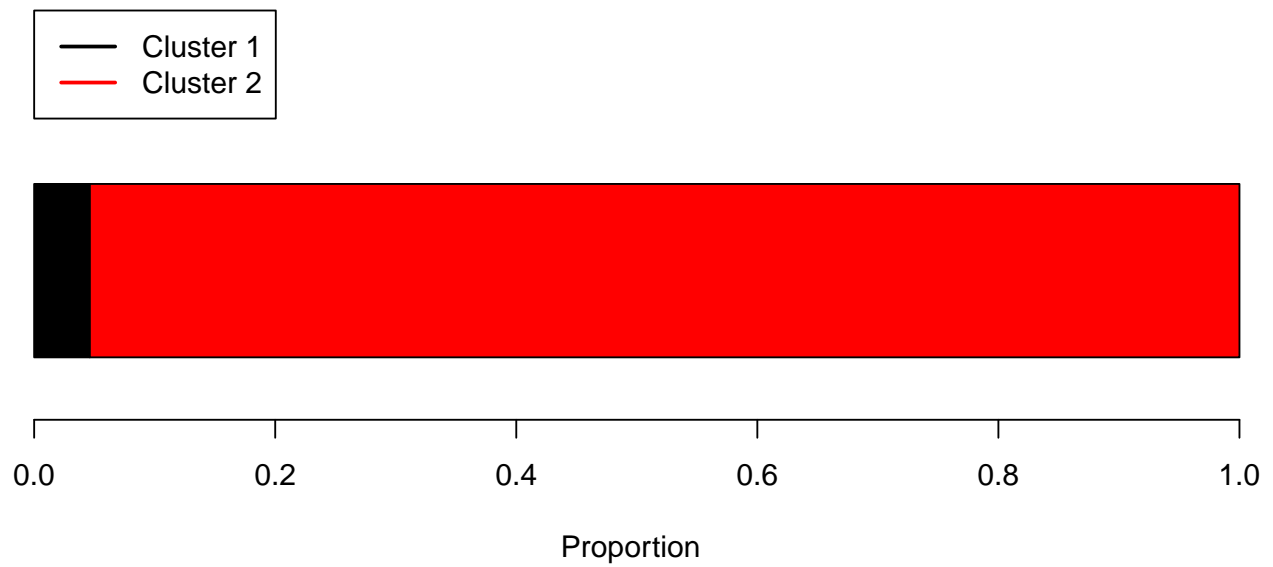
```

Health



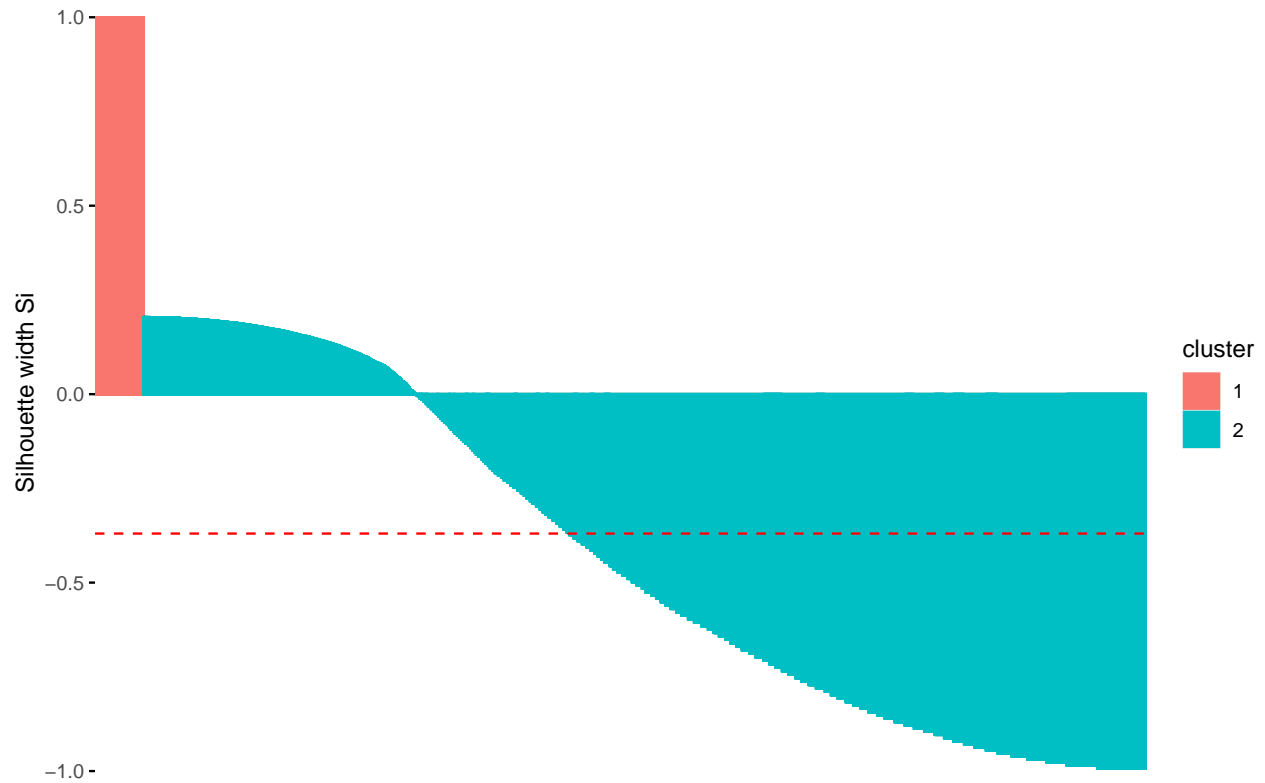
```
## [1] "Silhouette coefficient: 0.726800535711154"
## [1] "Xie Beni coefficient: 21709.6233018117"
## [1] "Davies Bouldin coefficient: 0.355478247608528"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.09555  
##   cluster  size ave.sil.width  
## 1         1   687          1.00  
## 2         2 14181         -0.44
```

Clusters silhouette plot
Average silhouette width: -0.37



```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
## Cluster 1 Cluster 2
##      1138      23345
##
##
##
## DB Population:
## Cluster 1 Cluster 2
##       73       71.4
##
##
##
## CSD Population:
## Cluster 1 Cluster 2
##    249166  239936.5
##
##
##
## CMA Type:
## Cluster 1 Cluster 2
##       505      10122
## B       470      9842
## D       118      2542
## K        45       839
```

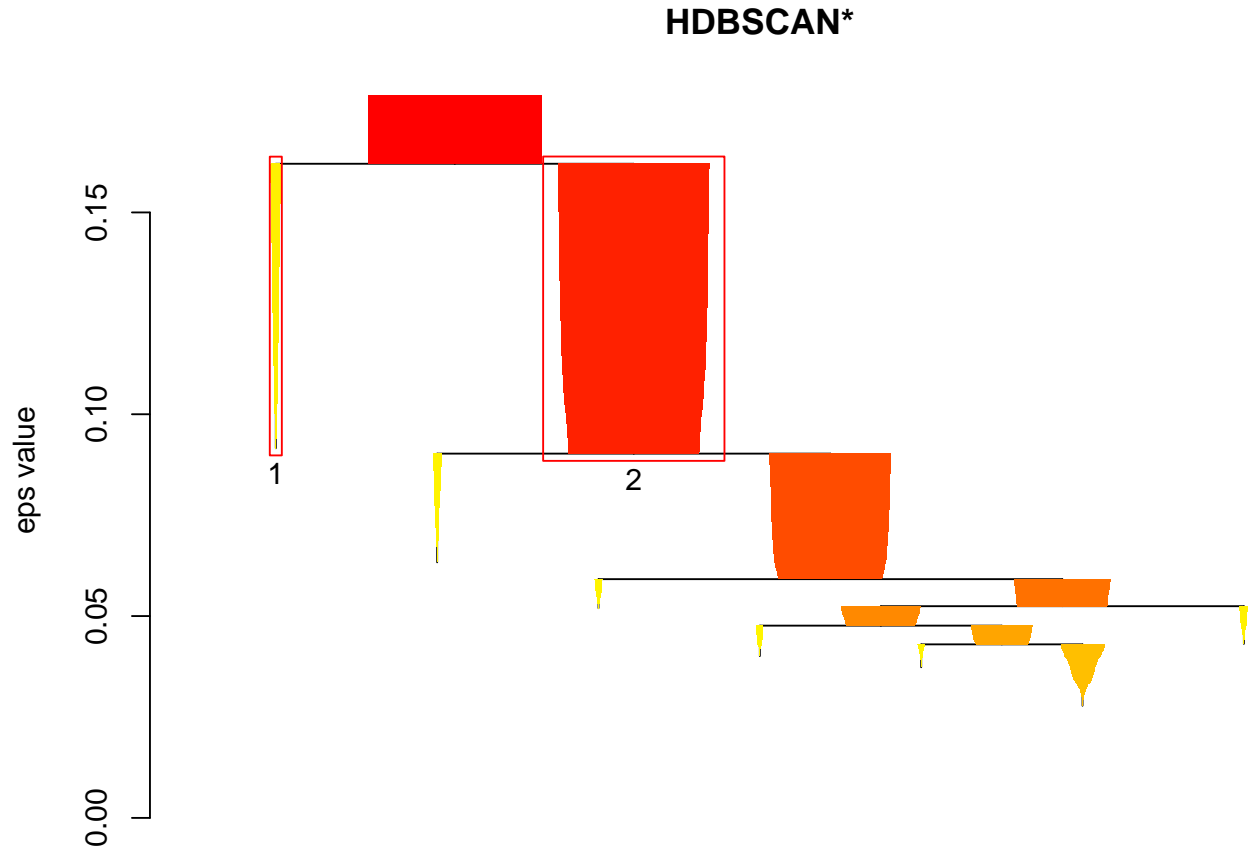


```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.23      0.226
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              38      725
## BritishColumbia      47     1152
## NewBrunswick          8      199
## NorthwestTerritories  1       15
## NovaScotia           35      744
## Ontario              206     3652
## Quebec               57     1321
## Saskatchewan          3      118
## NA's                 743     15419
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      1034     21084
## 1        83     1732
## 2        18      249
## F         3      280
##
##
##
## PMS_prox_idx_health :
## Cluster 1 Cluster 2
##      0.01494  0.01365

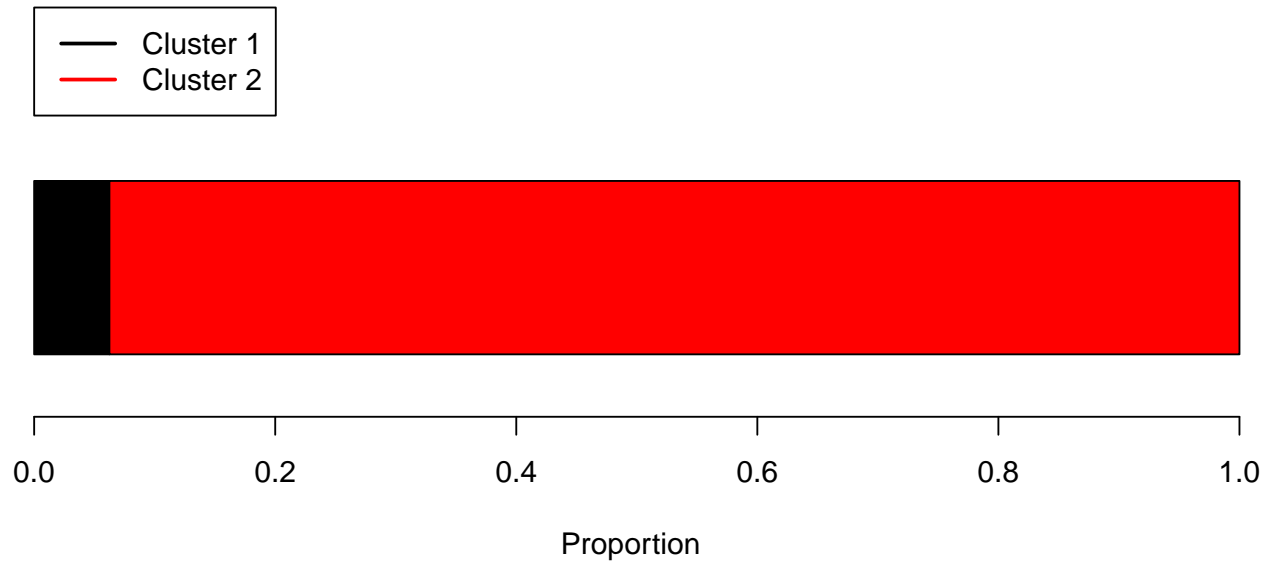
```

Grocery



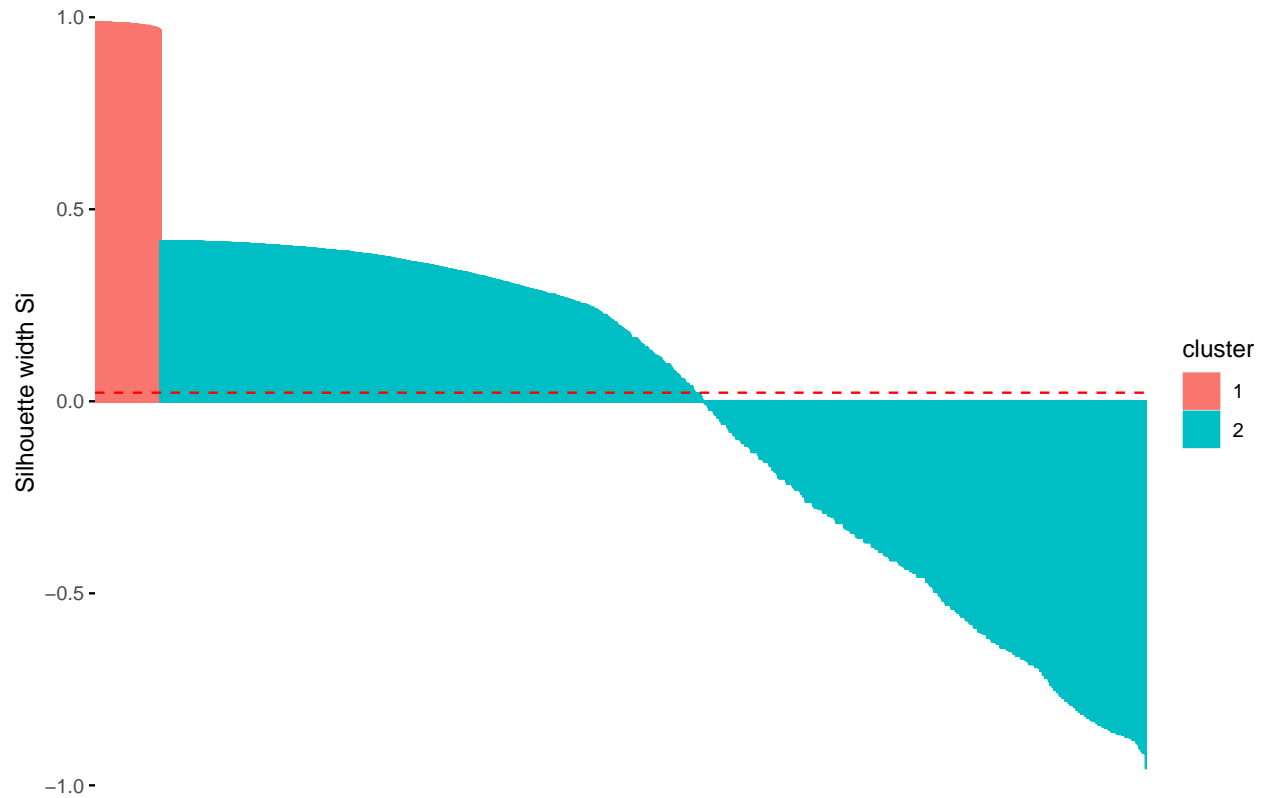
```
## [1] "Silhouette coefficient: 0.405722984464385"  
## [1] "Xie Beni coefficient: Inf"  
## [1] "Davies Bouldin coefficient: 2.06783585991517"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.1146  
##   cluster size ave.sil.width  
## 1         1  408          0.98  
## 2         2 6126         -0.04
```

Clusters silhouette plot
Average silhouette width: 0.02



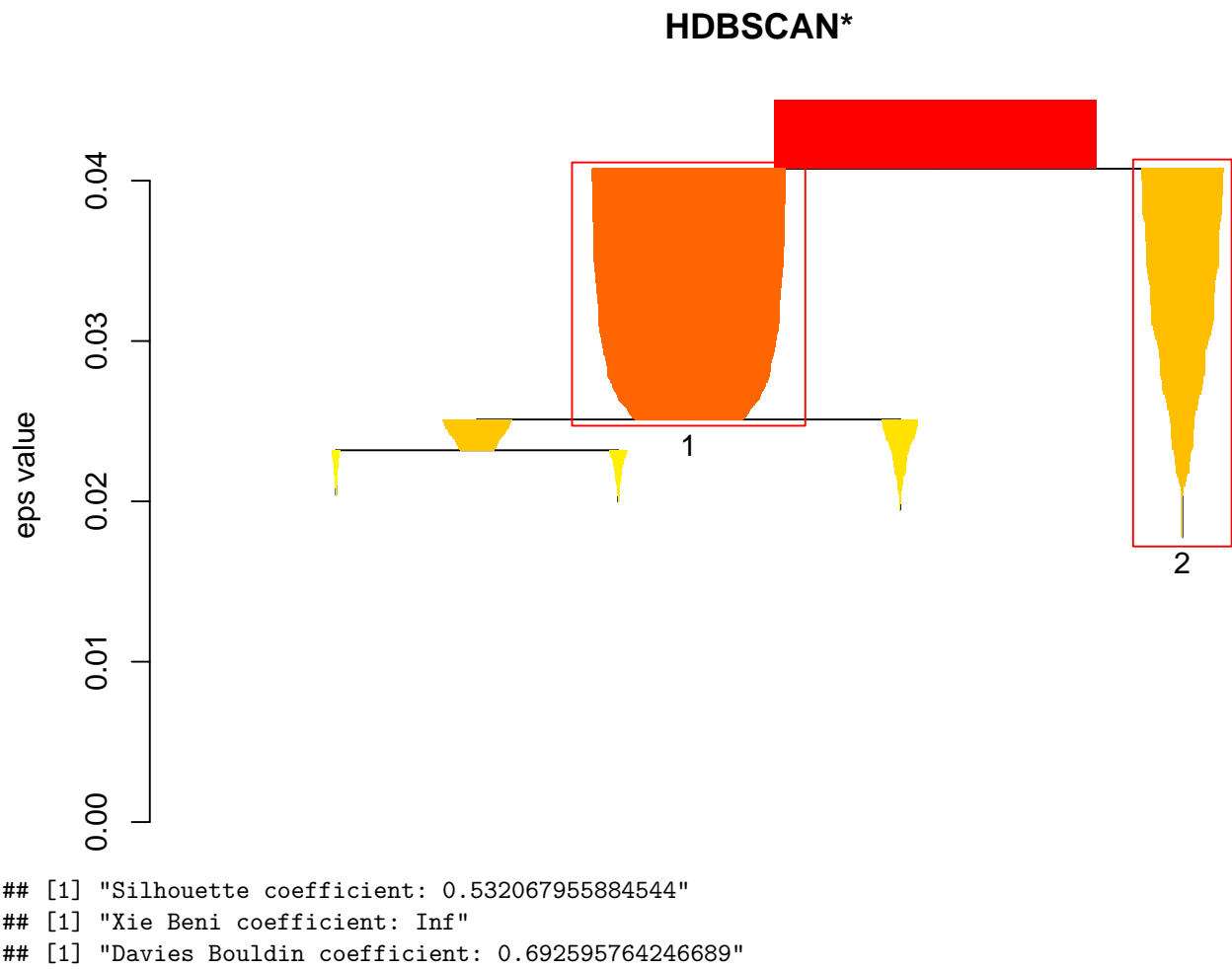
```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
## Cluster 1 Cluster 2
##      1528      22955
##
##
##
## DB Population:
## Cluster 1 Cluster 2
##      73.3      71.4
##
##
##
## CSD Population:
## Cluster 1 Cluster 2
## 215871.6 241996.3
##
##
##
## CMA Type:
## Cluster 1 Cluster 2
##      669      9958
## B      634      9678
## D      159      2501
## K       66       818
```

```

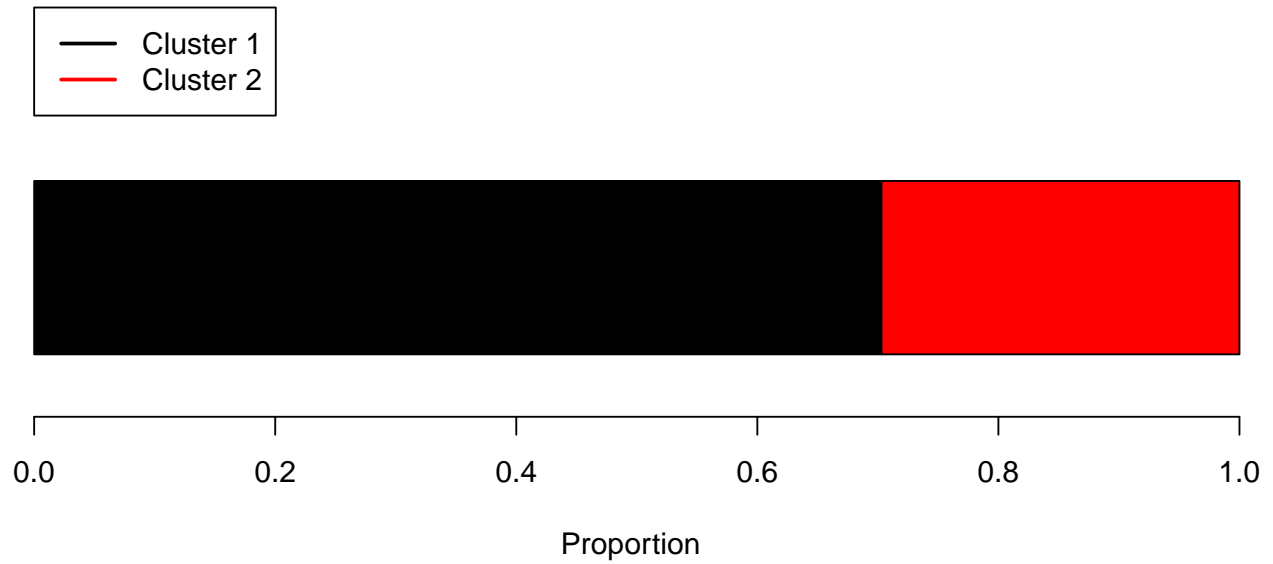
##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.225      0.226
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              41      722
## BritishColumbia      73     1126
## NewBrunswick         20      187
## NorthwestTerritories  2       14
## NovaScotia           33      746
## Ontario              237     3621
## Quebec               80     1298
## Saskatchewan          5      116
## NA's                 1037    15125
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      1386     20732
## 1       109     1706
## 2        11      256
## F        22      261
##
##
##
## PMS_prox_idx_grocery :
## Cluster 1 Cluster 2
##      0.072    0.07073

```

Primary Education

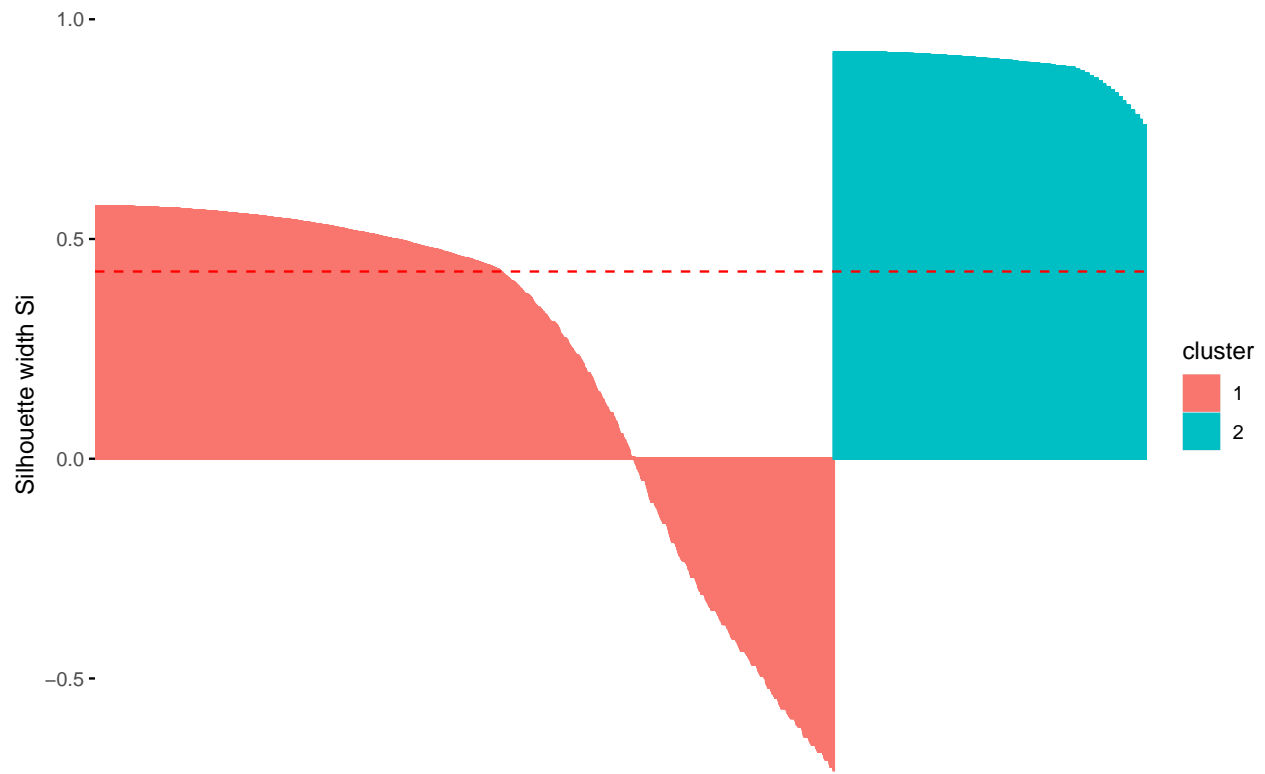


Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.05295  
##   cluster size ave.sil.width  
## 1       1 6774         0.23  
## 2       2 2862         0.89
```

Clusters silhouette plot
Average silhouette width: 0.43



```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##   Cluster 1 Cluster 2
##      17176      7307
##
##
##
## DB Population:
##   Cluster 1 Cluster 2
##      70.7      73.4
##
##
##
## CSD Population:
##   Cluster 1 Cluster 2
##  239548.1  242285.8
##
##
##
## CMA Type:
##   Cluster 1 Cluster 2
##      7479      3148
## B      7173      3139
## D      1893       767
## K        631       253
```

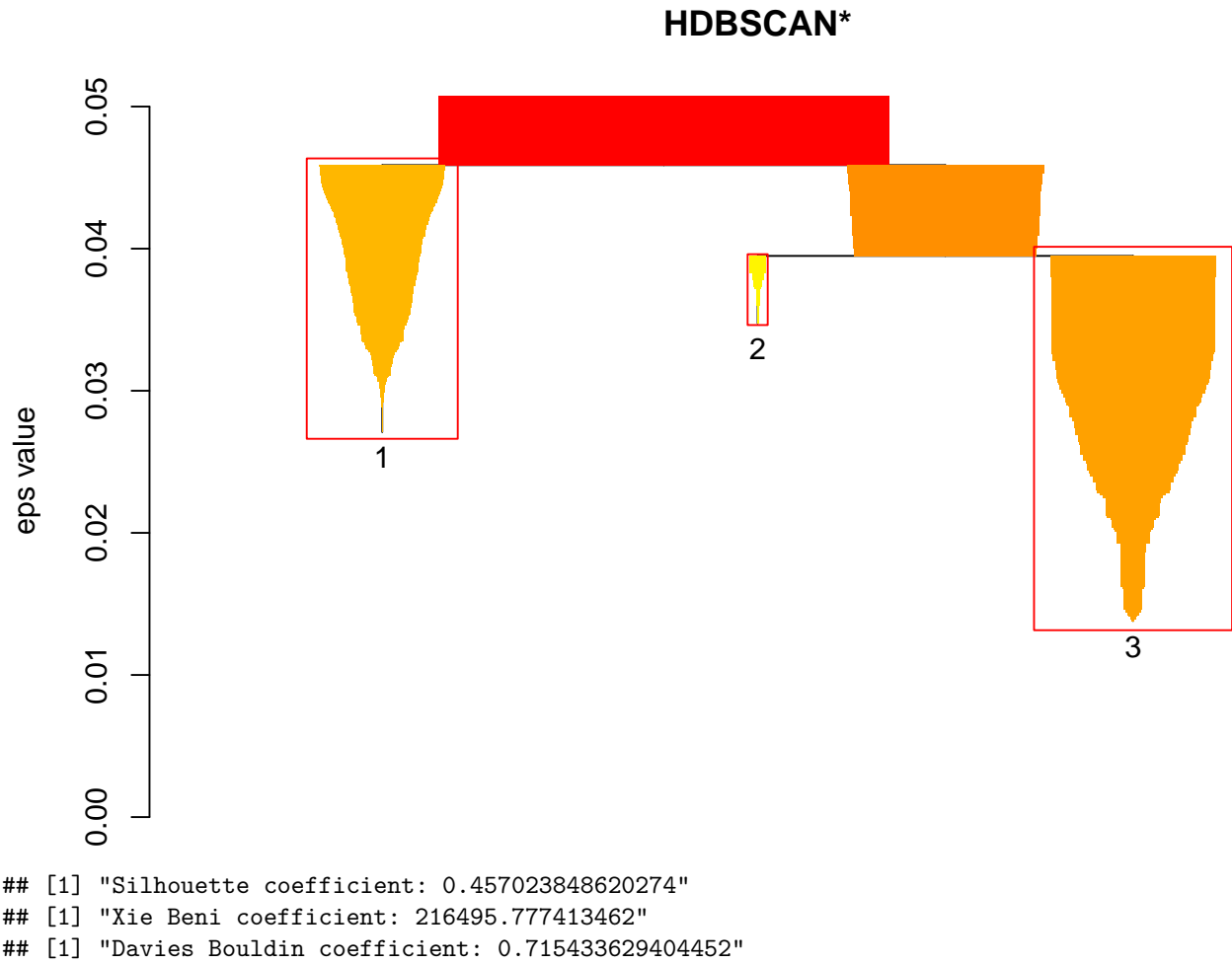


```

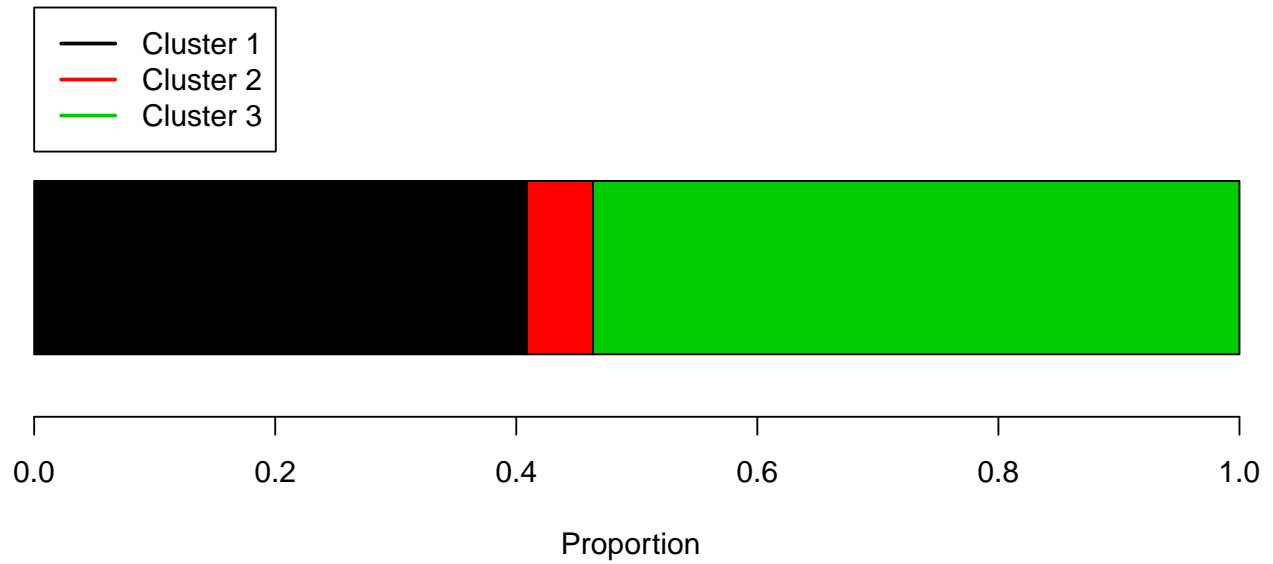
##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.226      0.226
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              519      244
## BritishColumbia      843      356
## NewBrunswick          147       60
## NorthwestTerritories    11       5
## NovaScotia            551      228
## Ontario              2676     1182
## Quebec               976      402
## Saskatchewan          85       36
## NA's                 11368     4794
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      15488      6630
## 1       1303       512
## 2        187       80
## F        198       85
##
##
##
## PMS_prox_idx_educpri :
## Cluster 1 Cluster 2
##      0.11746    0.11734

```

Secondary Education

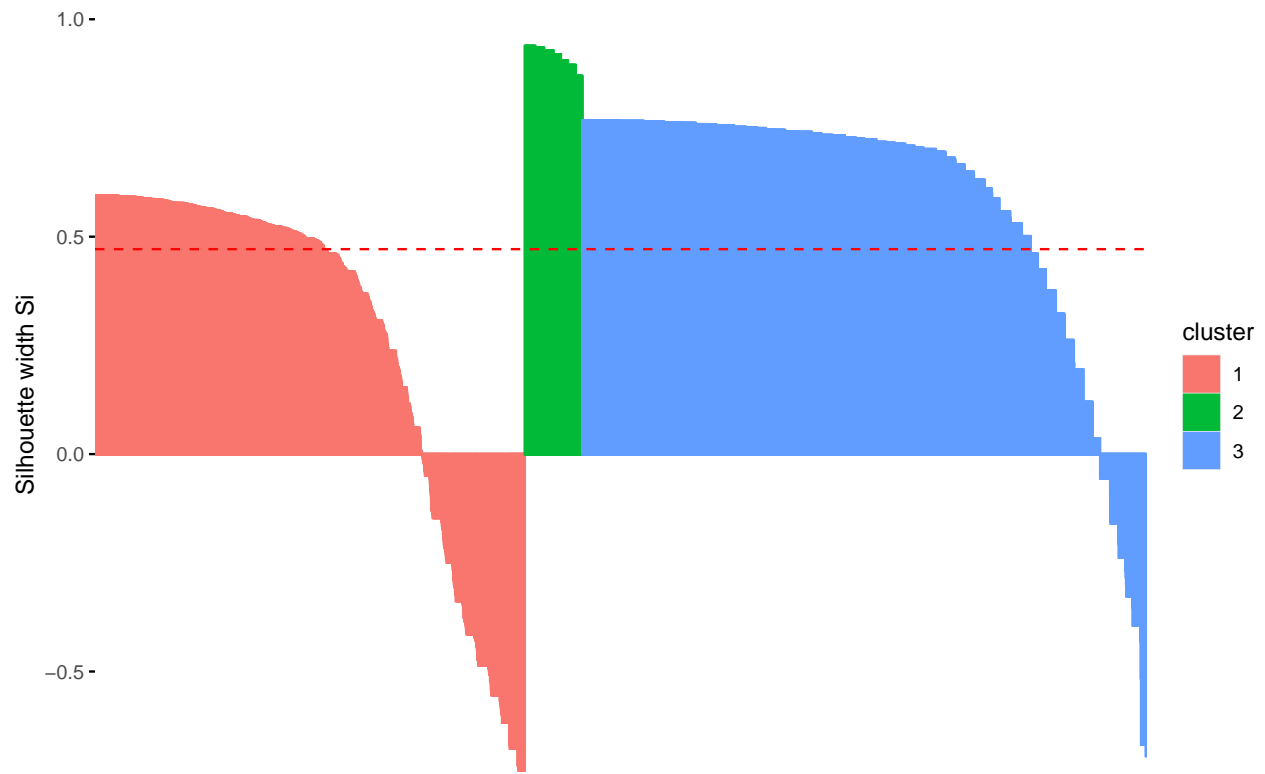


Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.0625  
## [1] 0.096  
##   cluster size ave.sil.width  
## 1         1 1974          0.27  
## 2         2  264          0.92  
## 3         3 2589          0.58
```

Clusters silhouette plot
Average silhouette width: 0.47



```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##   Cluster 1 Cluster 2 Cluster 3
##      10007      1334      13142
##
##
##
## DB Population:
##   Cluster 1 Cluster 2 Cluster 3
##      71.1      69.7      72
##
##
##
## CSD Population:
##   Cluster 1 Cluster 2 Cluster 3
##  241033.3  214198.7  242513.1
##
##
##
## CMA Type:
##   Cluster 1 Cluster 2 Cluster 3
##      4300      585      5742
## B      4235      539      5538
## D      1100      157      1403
## K       372       53      459
```

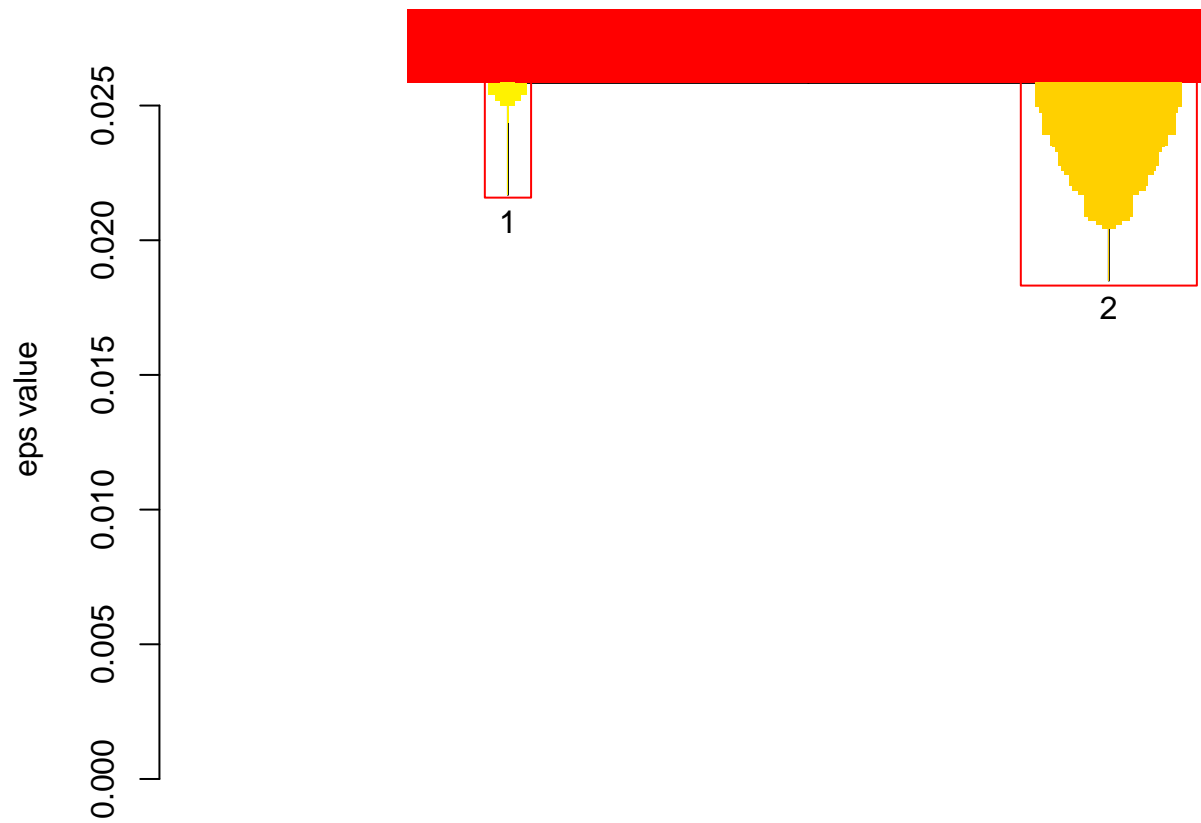
```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2 Cluster 3
##      0.225      0.23      0.227
##
##
##
## Provinces:
##              Cluster 1 Cluster 2 Cluster 3
## Alberta              309      33      421
## BritishColumbia      509      77      613
## NewBrunswick          85      13      109
## NorthwestTerritories   6       1       9
## NovaScotia           343      44      392
## Ontario              1588     194     2076
## Quebec               588      72      718
## Saskatchewan          50       7       64
## NA's                 6529     893     8740
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2 Cluster 3
## 0      9019      1209      11890
## 1       769       97       949
## 2       103       15       149
## F       116       13       154
##
##
##
## PMS_prox_idx_educsec :
## Cluster 1 Cluster 2 Cluster 3
##      0.10137  0.11699  0.10406

```

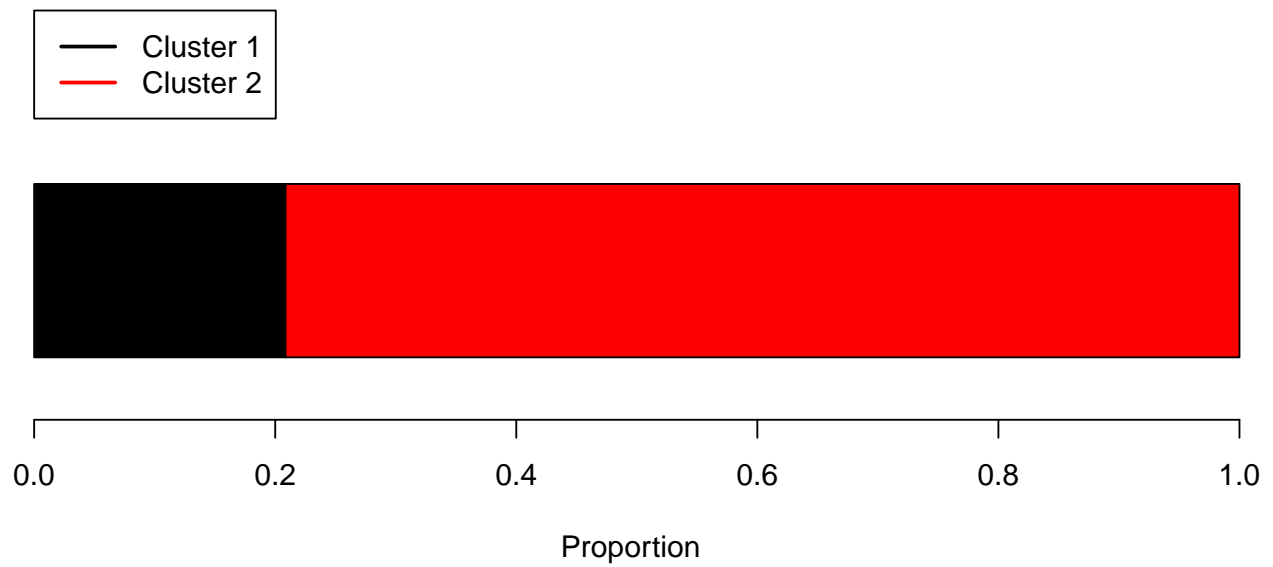
Libraries

HDBSCAN*



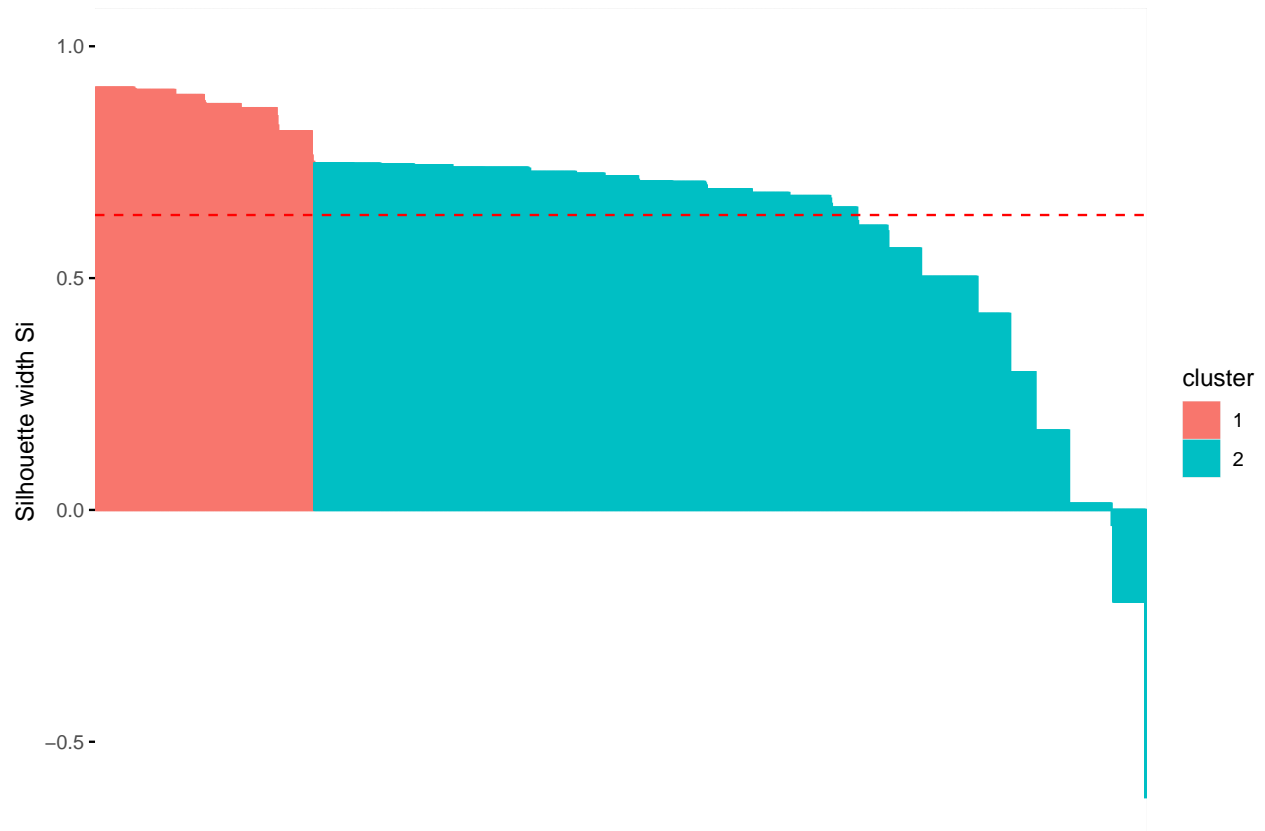
```
## [1] "Silhouette coefficient: 0.454816989136992"  
## [1] "Xie Beni coefficient: Inf"  
## [1] "Davies Bouldin coefficient: 0.669349169759769"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.0577  
##   cluster size ave.sil.width  
## 1         1  268          0.88  
## 2         2 1017          0.57
```

Clusters silhouette plot
Average silhouette width: 0.64



```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
## Cluster 1 Cluster 2
##      5108      19375
##
##
##
## DB Population:
## Cluster 1 Cluster 2
##      72.4      71.2
##
##
##
## CSD Population:
## Cluster 1 Cluster 2
## 254096.3 236743.3
##
##
##
## CMA Type:
## Cluster 1 Cluster 2
##      2199      8428
## B      2150      8162
## D      556      2104
## K      203      681
```

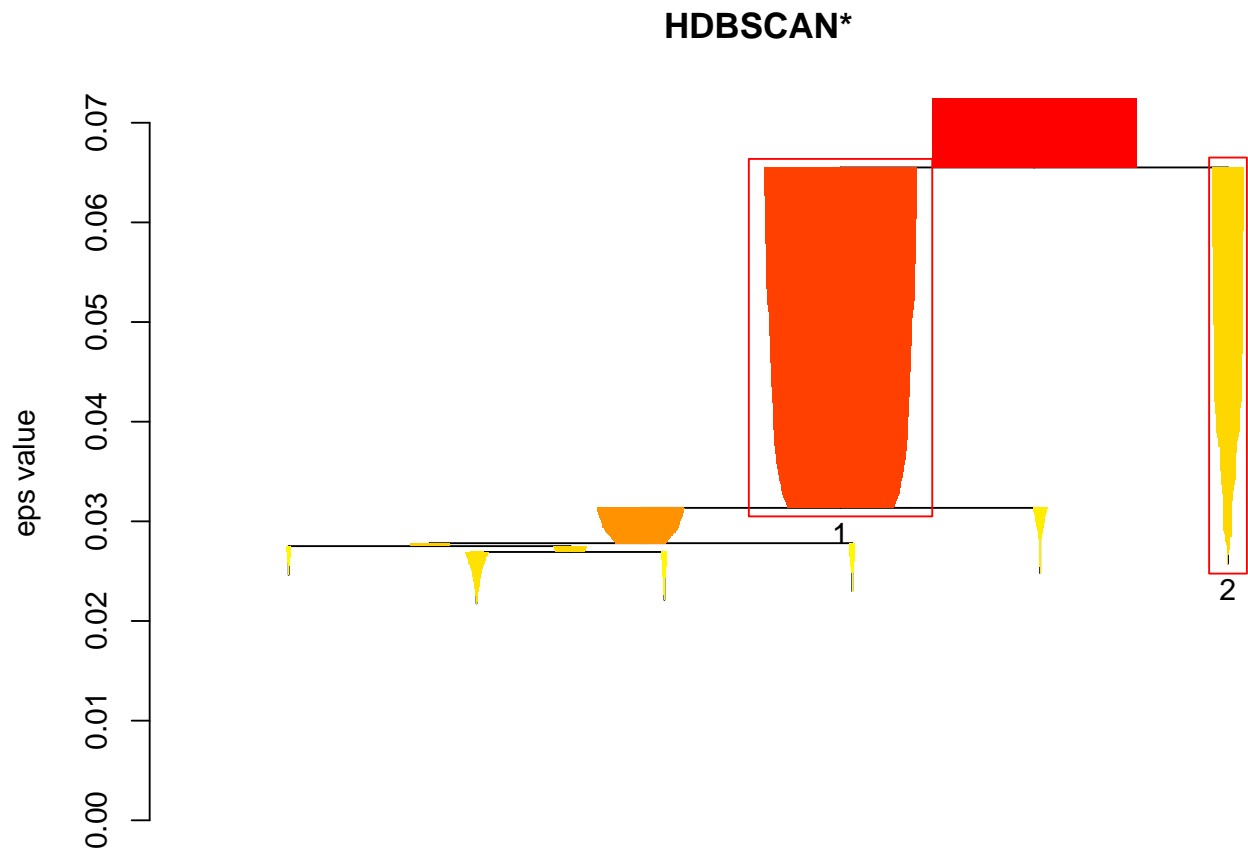


```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.225      0.227
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              168      595
## BritishColumbia      261      938
## NewBrunswick          49      158
## NorthwestTerritories   2       14
## NovaScotia            154      625
## Ontario               846     3012
## Quebec               286     1092
## Saskatchewan           17      104
## NA's                 3325     12837
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      4598     17520
## 1       386     1429
## 2        56      211
## F        68      215
##
##
##
## PMS_prox_idx_lib :
## Cluster 1 Cluster 2
##      0.11642   0.11329

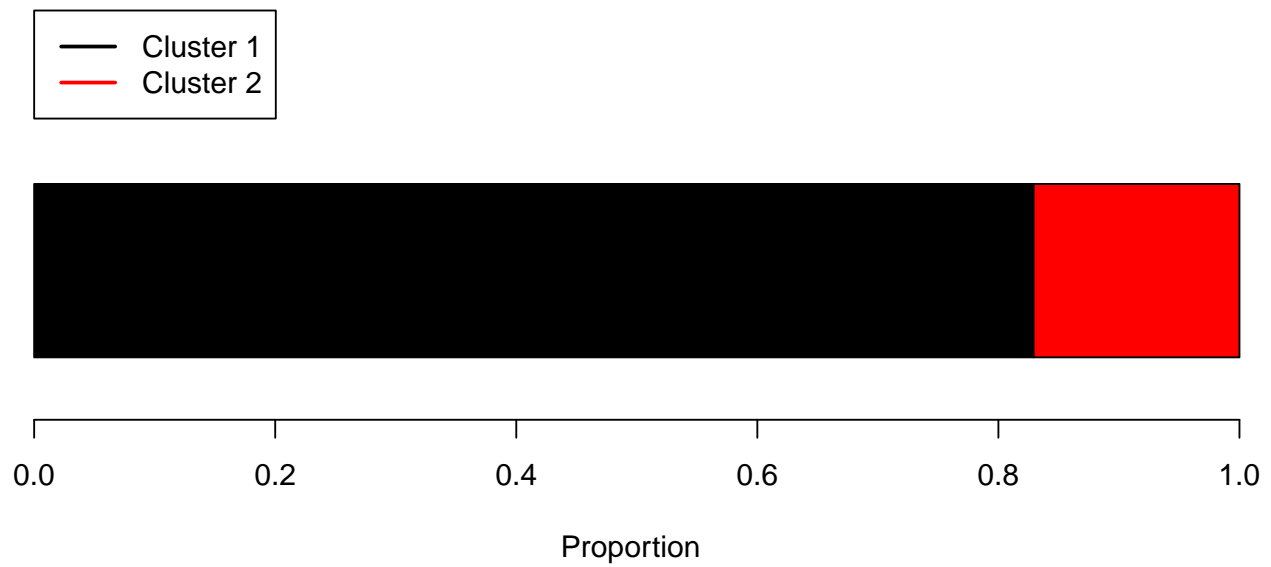
```

Parks



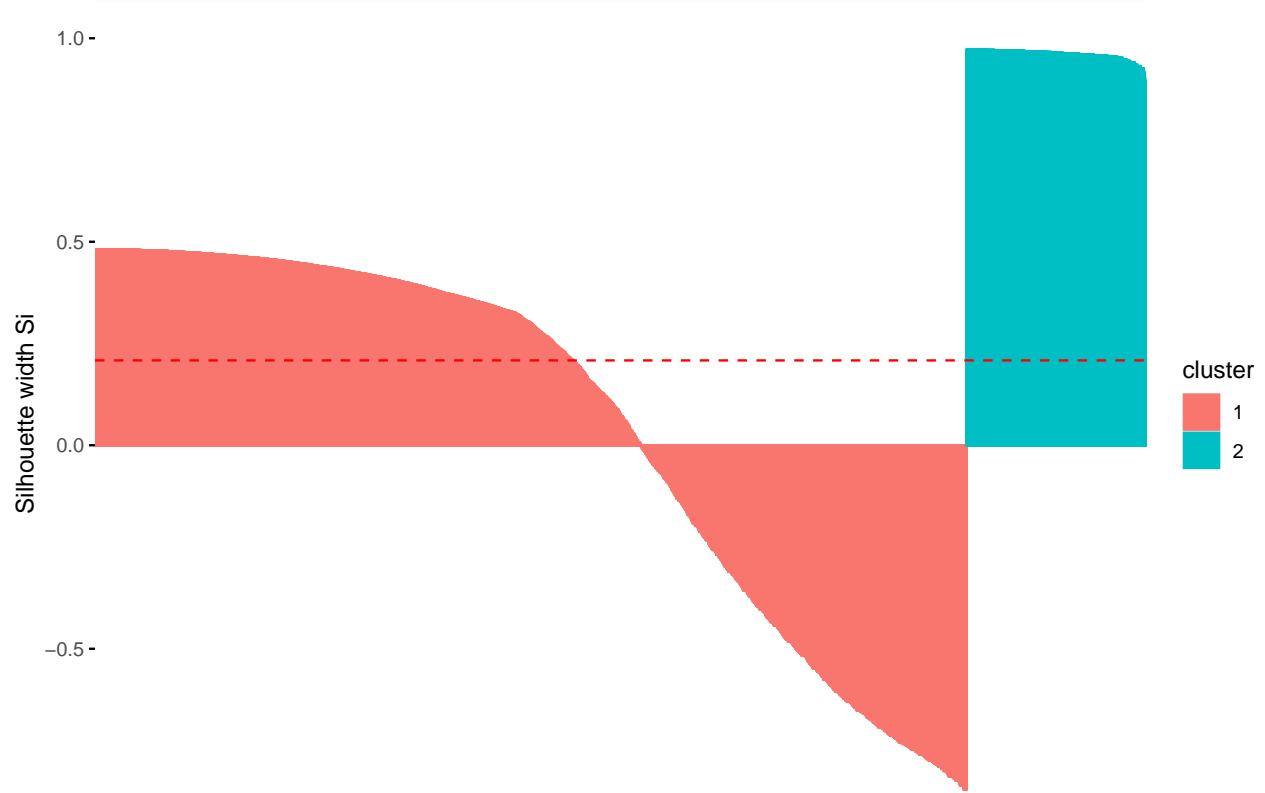
```
## [1] "Silhouette coefficient: 0.332151764469907"  
## [1] "Xie Beni coefficient: 431856.623546124"  
## [1] "Davies Bouldin coefficient: 6.54391888894235"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.01955  
##   cluster size ave.sil.width  
## 1      1 8725      0.05  
## 2      2 1797      0.96
```

Clusters silhouette plot
Average silhouette width: 0.21



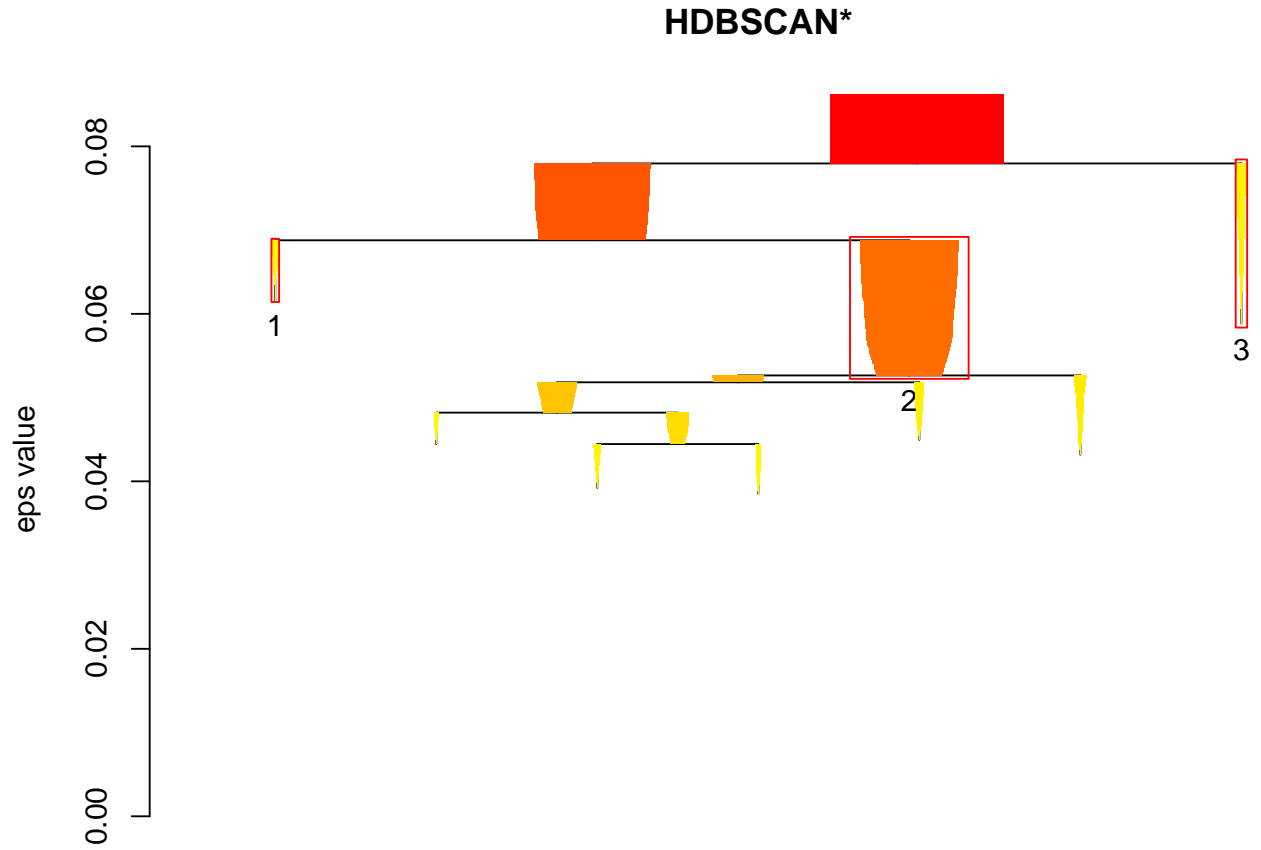
```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##   Cluster 1 Cluster 2
##      20313      4170
##
##
##
## DB Population:
##   Cluster 1 Cluster 2
##      71.9      69.4
##
##
##
## CSD Population:
##   Cluster 1 Cluster 2
##  236544.8  258961.9
##
##
##
## CMA Type:
##   Cluster 1 Cluster 2
##      8824      1803
## B      8519      1793
## D      2243       417
## K       727       157
```

```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2
##      0.227      0.222
##
##
##
## Provinces:
##              Cluster 1 Cluster 2
## Alberta              643      120
## BritishColumbia      984      215
## NewBrunswick          173       34
## NorthwestTerritories   12        4
## NovaScotia            655      124
## Ontario              3151      707
## Quebec               1128      250
## Saskatchewan          103       18
## NA's                 13464     2698
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2
## 0      18376      3742
## 1       1485       330
## 2        215        52
## F        237        46
##
##
##
## PMS_prox_idx_parks :
## Cluster 1 Cluster 2
##      0.06852      0.0691

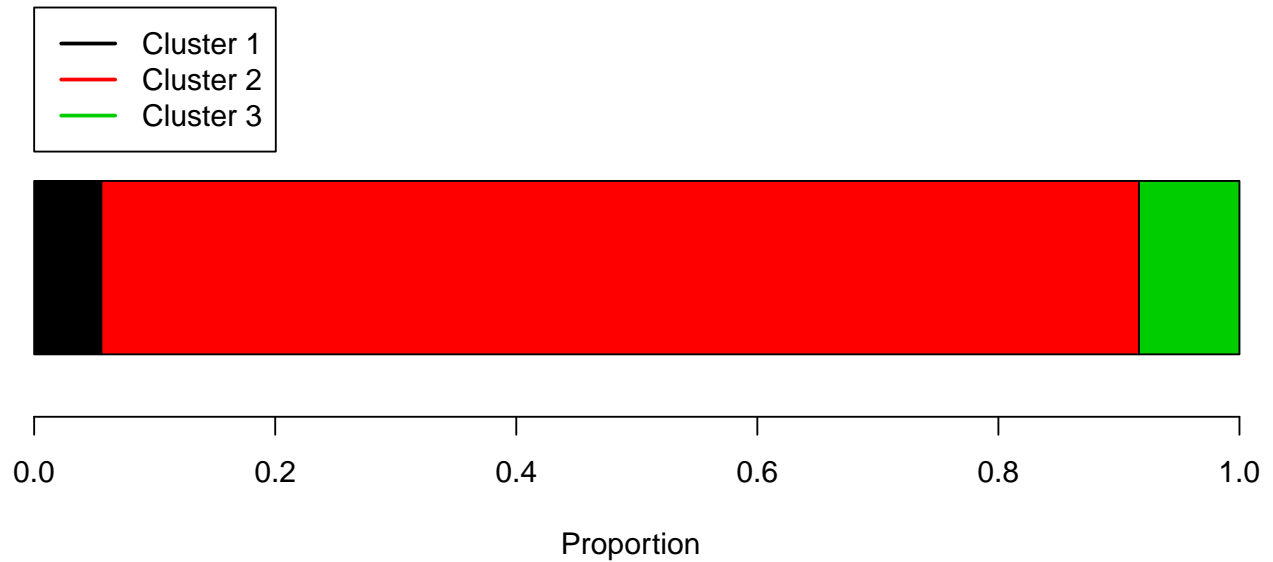
```

Transit

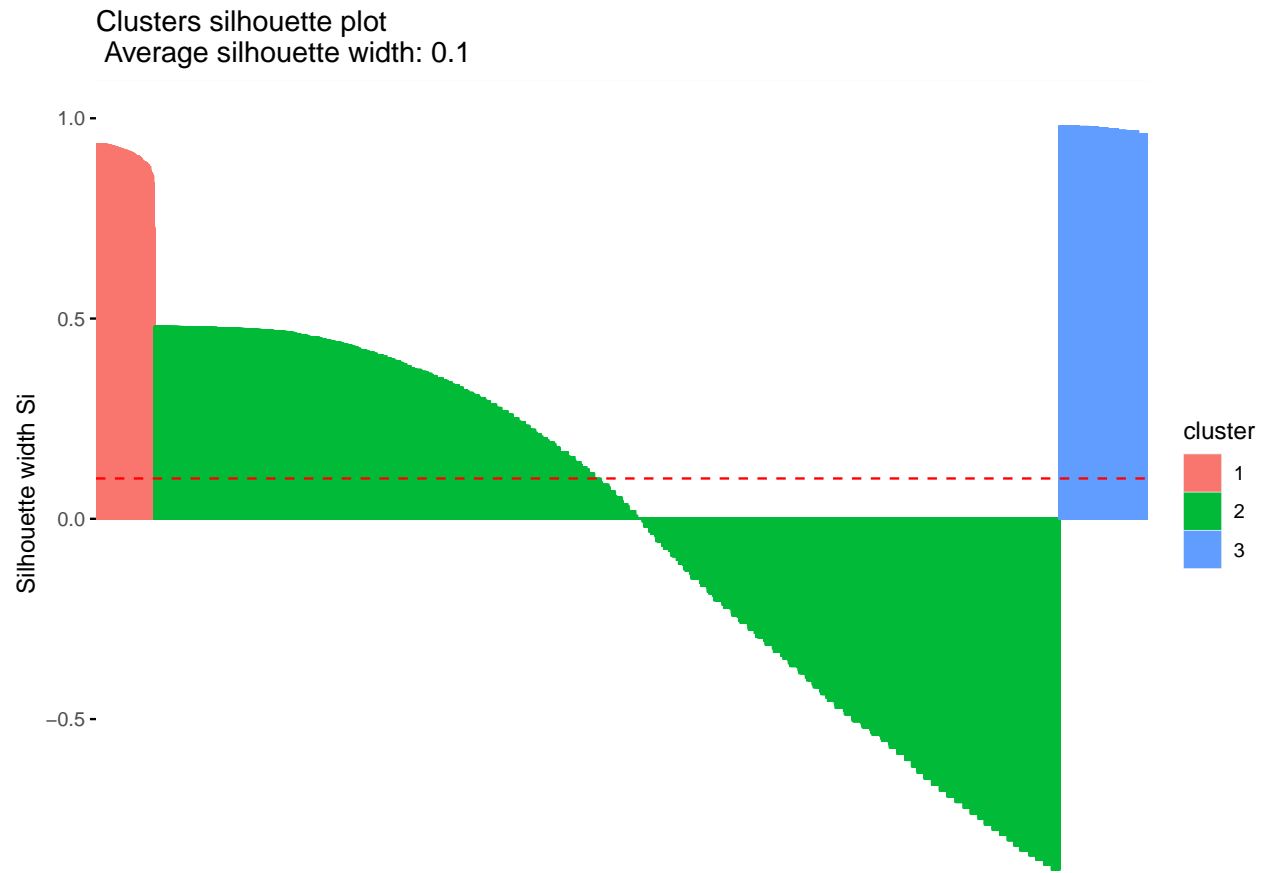


```
## [1] "Silhouette coefficient: 0.340639023025164"  
## [1] "Xie Beni coefficient: Inf"  
## [1] "Davies Bouldin coefficient: 3.19593452462934"
```

Proportion of DBs in each cluster



```
## [1] "Segment cutoff values:"  
## [1] 0.0039  
## [1] 0.0355  
##   cluster size ave.sil.width  
## 1      1 334      0.91  
## 2      2 5154     -0.04  
## 3      3 499      0.97
```



```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##   Cluster 1 Cluster 2 Cluster 3
##       1361      21087      2035
##
##
##
## DB Population:
##   Cluster 1 Cluster 2 Cluster 3
##       70.5      71.3      73.5
##
##
##
## CSD Population:
##   Cluster 1 Cluster 2 Cluster 3
## 240817.1 240423.9 239451.8
##
##
##
## CMA Type:
##   Cluster 1 Cluster 2 Cluster 3
##         620      9115      892
## B         556      8902      854
## D         134      2310      216
## K          51       760       73
```



```

##
##
##
## Index of Remoteness:
## Cluster 1 Cluster 2 Cluster 3
##      0.227      0.226      0.23
##
##
##
## Provinces:
##              Cluster 1 Cluster 2 Cluster 3
## Alberta              31      680      52
## BritishColumbia      69     1036      94
## NewBrunswick         10      183      14
## NorthwestTerritories   3       12       1
## NovaScotia           43      658      78
## Ontario              218     3334     306
## Quebec               75     1197     106
## Saskatchewan          7      105       9
## NA's                 905     13882    1375
##
##
##
## Amenity dense:
## Cluster 1 Cluster 2 Cluster 3
## 0      1224     19048     1846
## 1       94      1571      150
## 2       20      231       16
## F       23      237       23
##
##
##
## PMS_prox_idx_transit :
## Cluster 1 Cluster 2 Cluster 3
##      0.019   0.01806   0.01779

```

Conclusion

text

```
## list(PMS_prox_idx_emp = 0.21475, PMS_prox_idx_pharma = 0.0114,  
##      PMS_prox_idx_childcare = 0.0087, PMS_prox_idx_health = 0.09555,  
##      PMS_prox_idx_grocery = 0.1146, PMS_prox_idx_educpri = 0.05295,  
##      PMS_prox_idx_educsec = c(0.0625, 0.096), PMS_prox_idx_lib = 0.0577,  
##      PMS_prox_idx_parks = 0.01955, PMS_prox_idx_transit = c(0.0039,  
##      0.0355))
```