HDBSCAN

With Log-Transformed Data

PMS

18 May, 2023

Assumptions of the Alogrithm

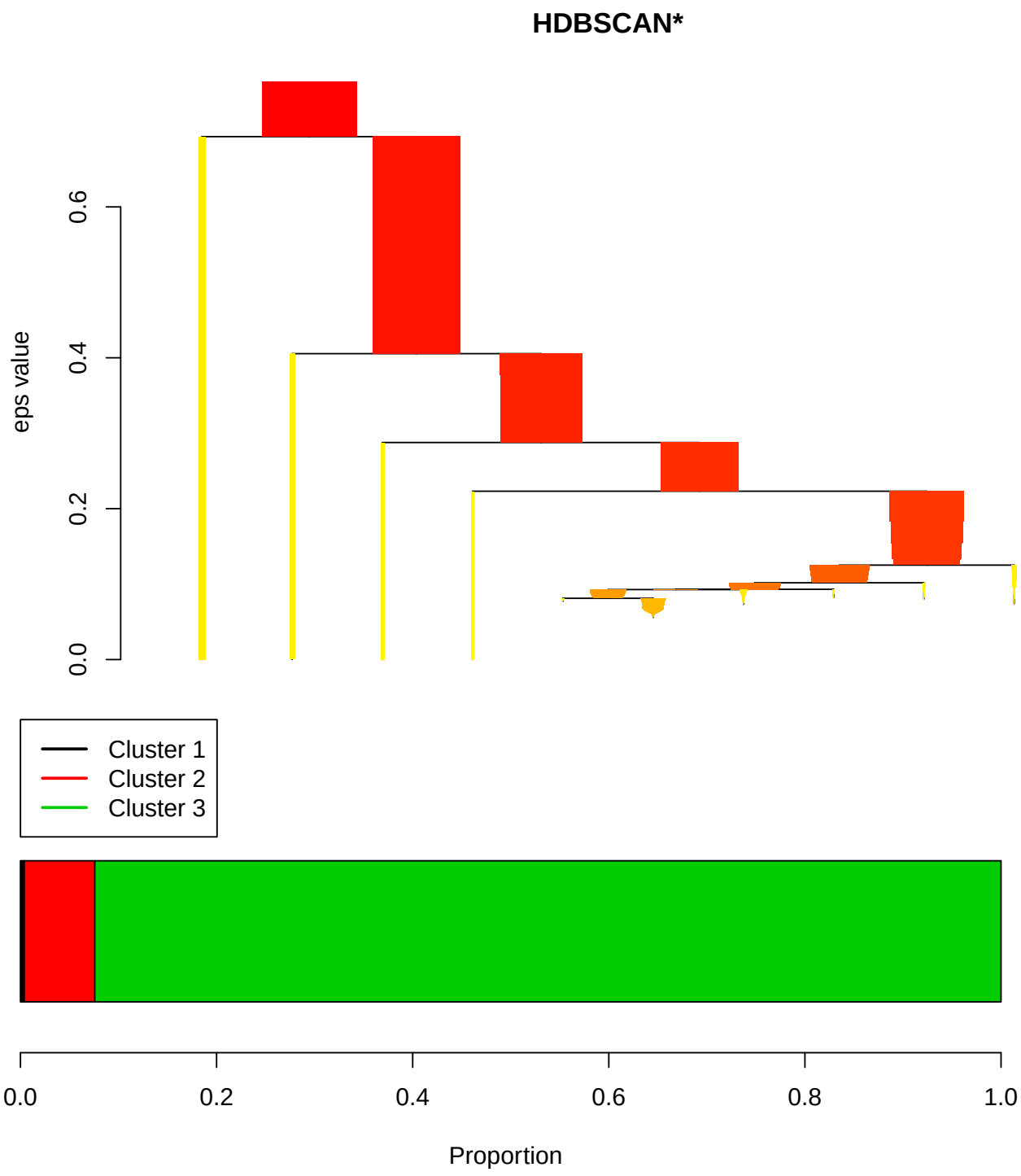
text

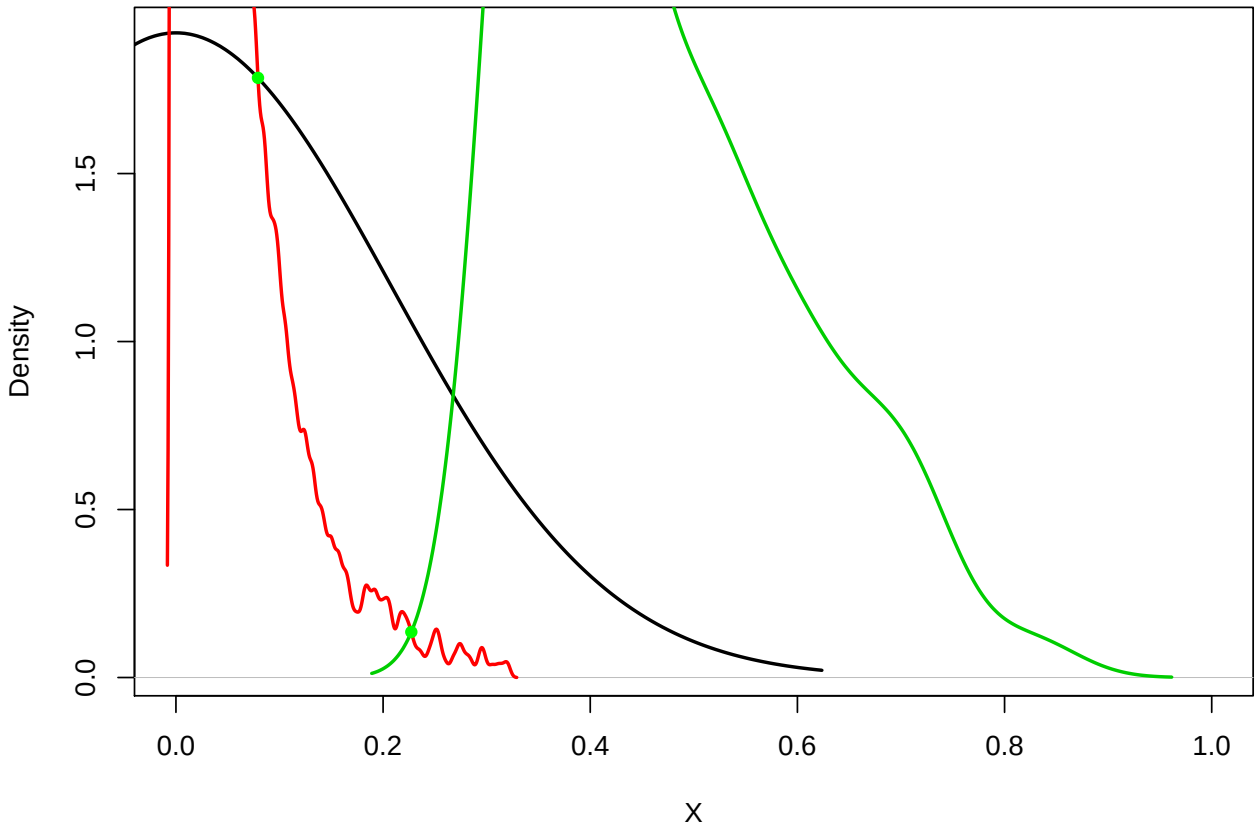
Amenities

Employment

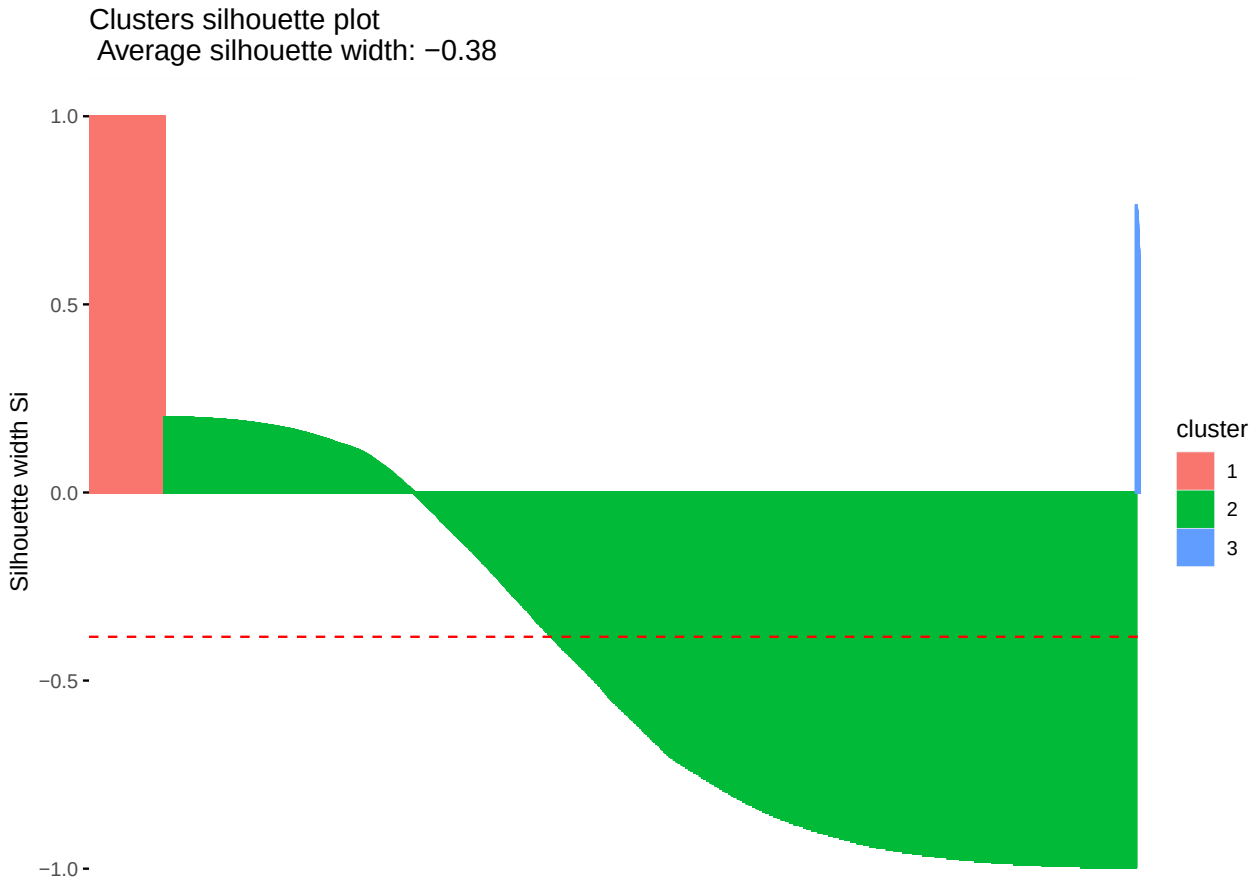
```
## [1] "Silhouette coefficient: 0.669838211728378"
## [1] "Xie Beni coefficient: 27759.4905301196"
```

```
## Warning in if (is.na(res)) {: the condition has length > 1 and only the first
## element will be used
```






```
## [1] "Segment cutoff values:"
## [1] 0.07922764
## [1] 0.2272408
##   cluster  size ave.sil.width
## 1       1  1521          1.00
## 2       2 19546         -0.50
## 3       3    81          0.69
```



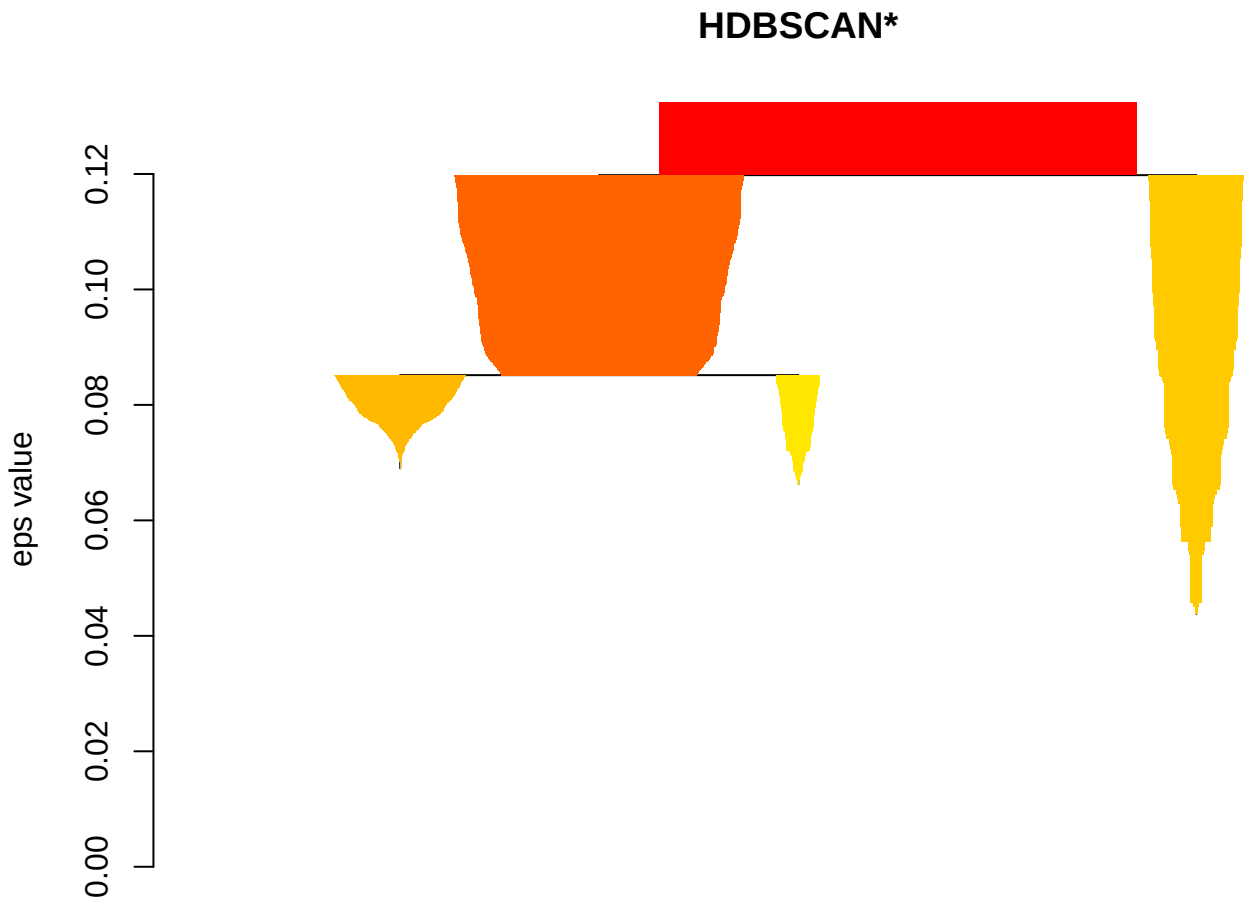

```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##  Cluster 1 Cluster 2 Cluster 3
##       1777     22610        96
##
##
##
##  DB Population:
##  Cluster 1 Cluster 2 Cluster 3
##       70.6      71.6      60.5
##
##
##
##  CSD Population:
##  Cluster 1 Cluster 2 Cluster 3
##   232673.2  240884.7  260239.5
##
##
##
##  CMA Type:
##   Cluster 1 Cluster 2 Cluster 3
##         799      9783        45
## B       741      9532        39
## D       182      2467        11
## K        55       828         1
```

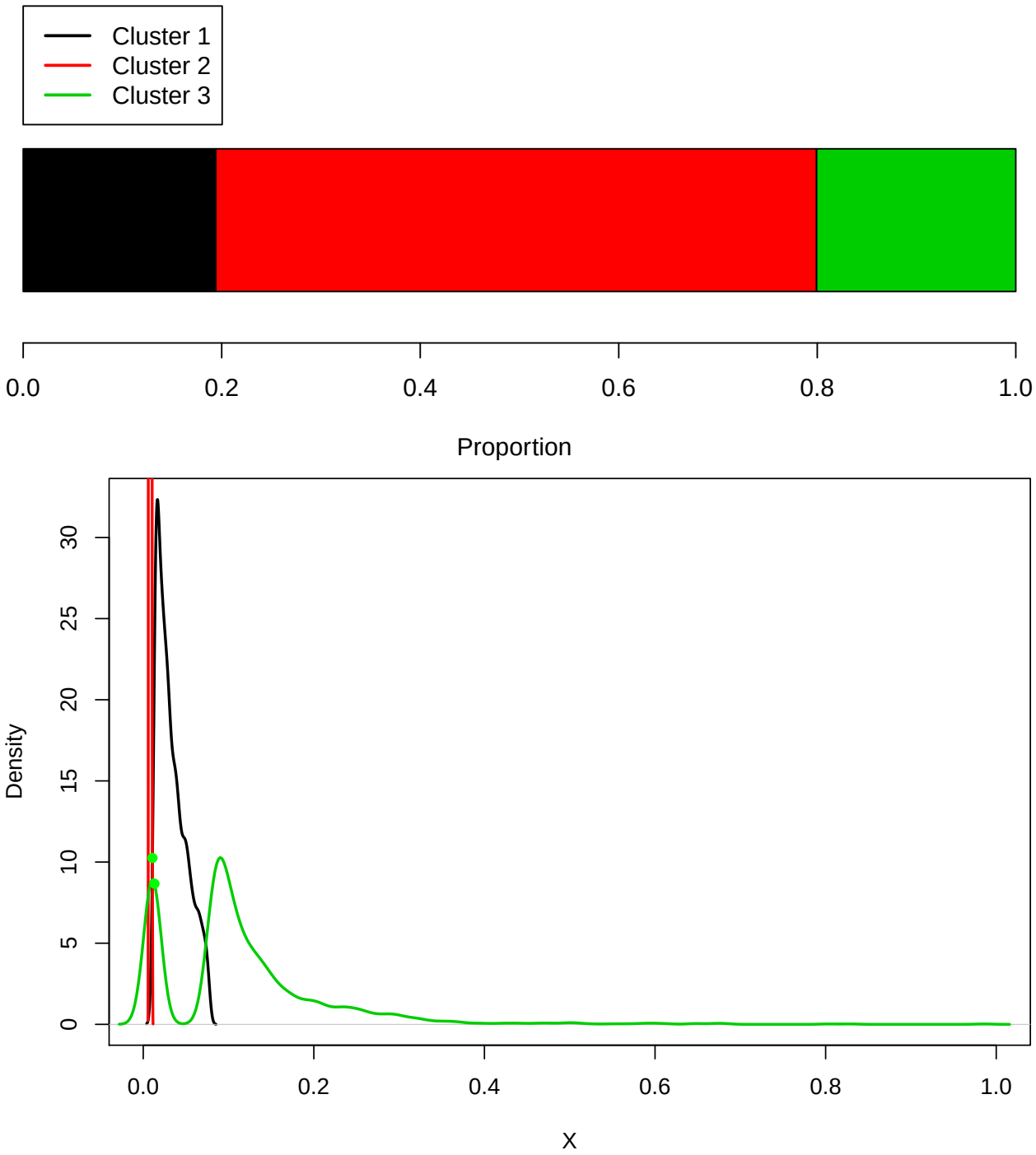
```
##
##
##
##  Index of Remoteness:
##  Cluster 1 Cluster 2 Cluster 3
##      0.224     0.226     0.259
##
##
##
##  Provinces:
##                       Cluster 1 Cluster 2 Cluster 3
## Alberta                      58       701         4
## BritishColumbia              84      1113         2
## NewBrunswick                 11       194         2
## NorthwestTerritories          3        13         0
## NovaScotia                   62       714         3
## Ontario                     276      3568        14
## Quebec                       91      1281         6
## Saskatchewan                 13       108         0
## NA's                       1179     14918        65
##
##
##
##  Amenity dense:
##   Cluster 1 Cluster 2 Cluster 3
## 0      1611     20419        88
## 1       116      1692         7
## 2        25       242         0
## F        25       257         1
##
##
##
##  PMS_prox_idx_emp :
##  Cluster 1 Cluster 2 Cluster 3
##     0.0258   0.02562    0.0273
```

Pharmacy

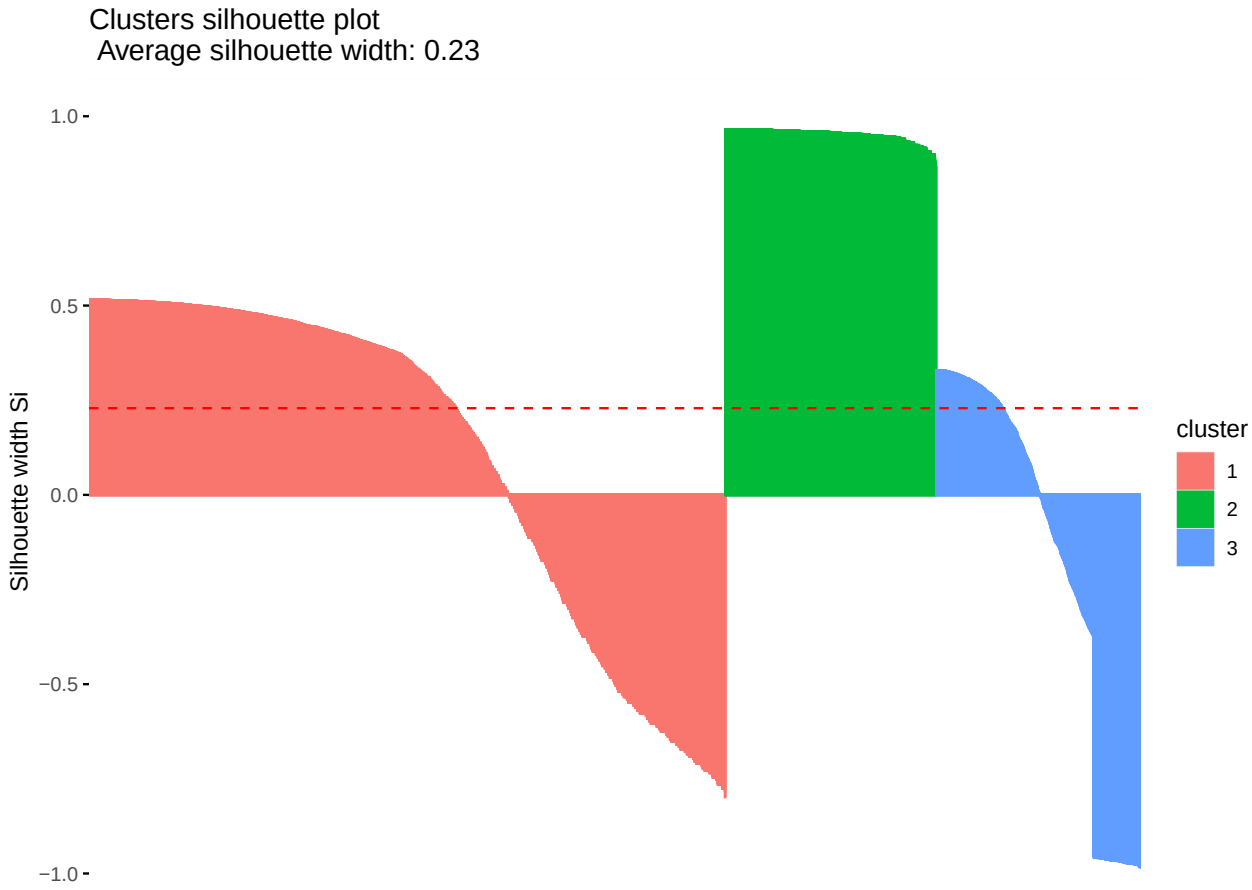
```
## [1] "Silhouette coefficient: 0.443449260743095"
## [1] "Xie Beni coefficient: Inf"
```

```
## Warning in if (is.na(res)) {: the condition has length > 1 and only the first
## element will be used
```






```
## [1] "Segment cutoff values:"
## [1] 0.01078839
## [1] 0.01319184
##   cluster size ave.sil.width
## 1       1 5366          0.11
## 2       2 1779          0.95
## 3       3 1720         -0.15
```



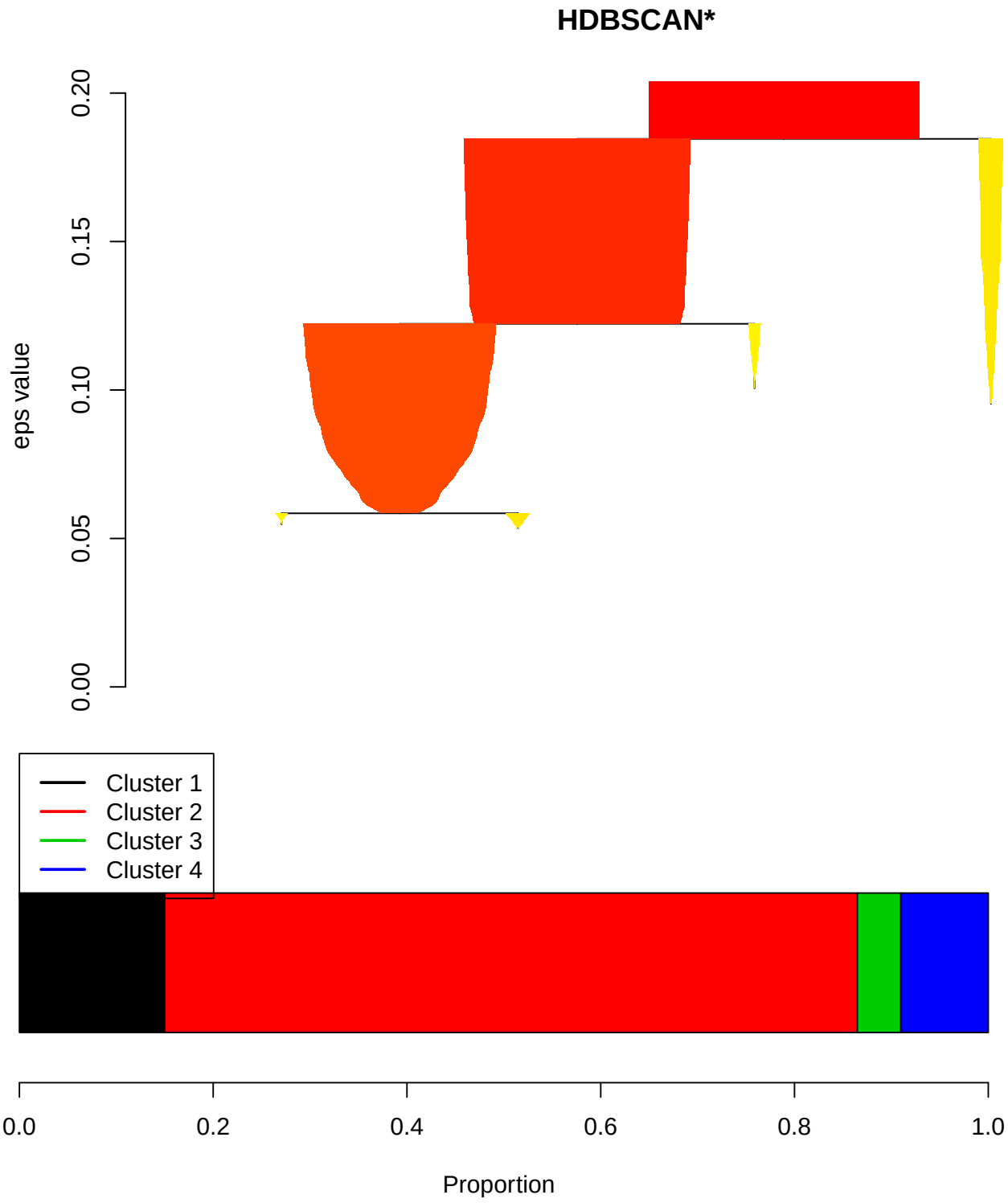

```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##  Cluster 1 Cluster 2 Cluster 3
##      14837      4920      4726
##
##
##
##  DB Population:
##  Cluster 1 Cluster 2 Cluster 3
##         72      70.6      70.7
##
##
##
##  CSD Population:
##  Cluster 1 Cluster 2 Cluster 3
##   240052.5  238405.1  243388.1
##
##
##
##  CMA Type:
##   Cluster 1 Cluster 2 Cluster 3
##        6432      2146      2049
## B      6253      2065      1994
## D      1604       533       523
## K       548       176       160
```

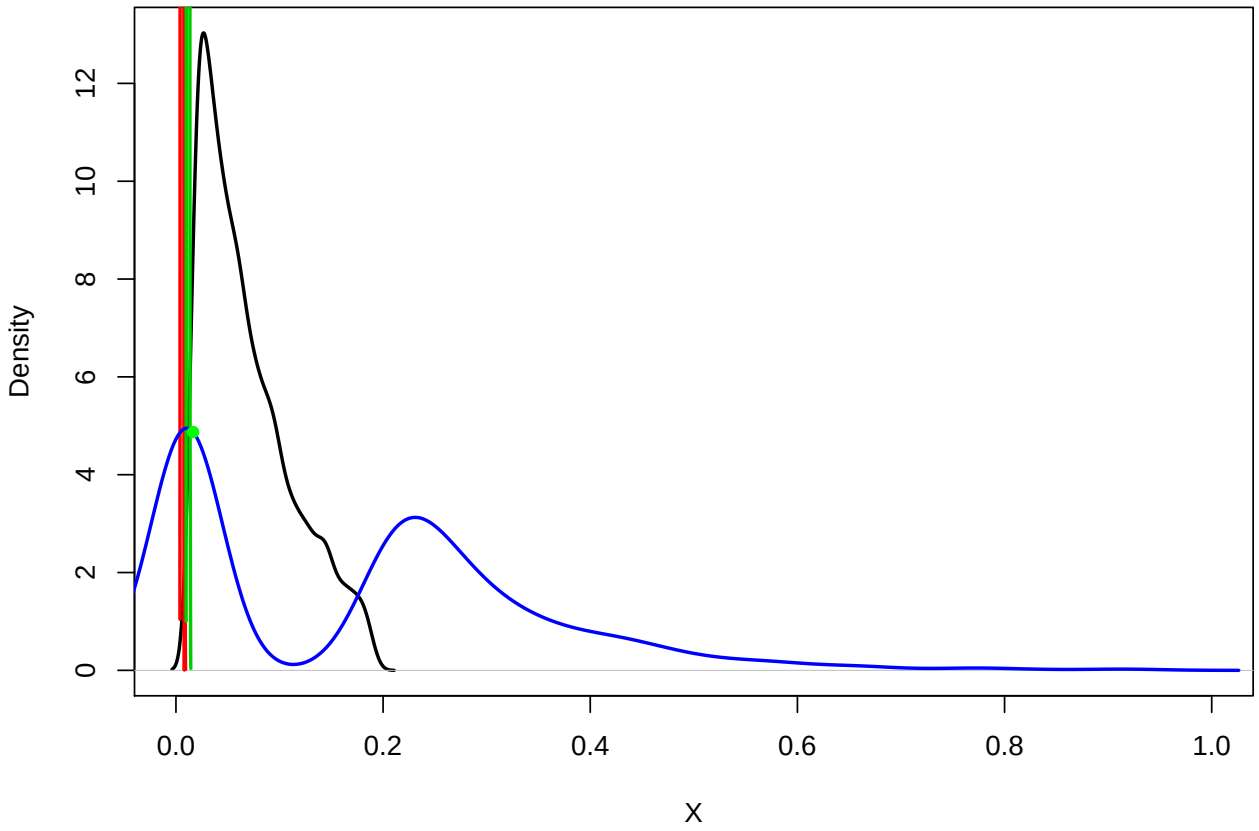
```
##
##
##
##  Index of Remoteness:
##  Cluster 1 Cluster 2 Cluster 3
##      0.226     0.228     0.226
##
##
##
##  Provinces:
##                       Cluster 1 Cluster 2 Cluster 3
## Alberta                     446       165       152
## BritishColumbia             743       246       210
## NewBrunswick                119        48        40
## NorthwestTerritories         12         1         3
## NovaScotia                  474       160       145
## Ontario                    2294       781       783
## Quebec                      865       267       246
## Saskatchewan                 65        26        30
## NA's                       9819      3226      3117
##
##
##
##  Amenity dense:
##   Cluster 1 Cluster 2 Cluster 3
## 0     13388      4473      4257
## 1      1098       345       372
## 2       163        56        48
## F       188        46        49
##
##
##
##  PMS_prox_idx_pharma :
##  Cluster 1 Cluster 2 Cluster 3
##    0.04526   0.04385   0.04276
```

Childcare

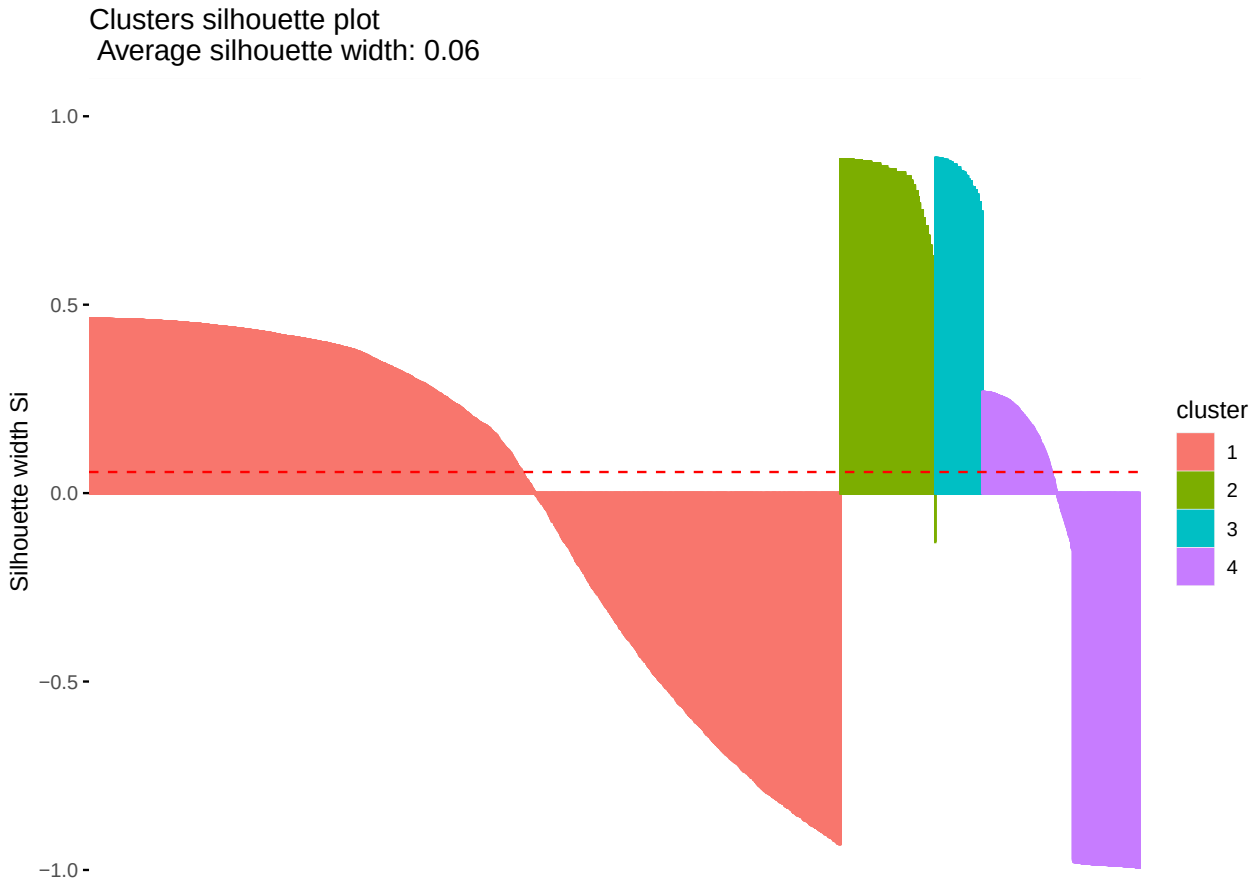
```
## [1] "Silhouette coefficient: 0.400812993352801"
## [1] "Xie Beni coefficient: Inf"
```

```
## Warning in if (is.na(res)) {: the condition has length > 1 and only the first
## element will be used
```






```
## [1] "Segment cutoff values:"
## [1] NA
## [1] NA
## [1] 0.01663626
##   cluster size ave.sil.width
## 1       1 8677         -0.01
## 2       2 1096          0.83
## 3       3  544          0.85
## 4       4 1820         -0.34
```



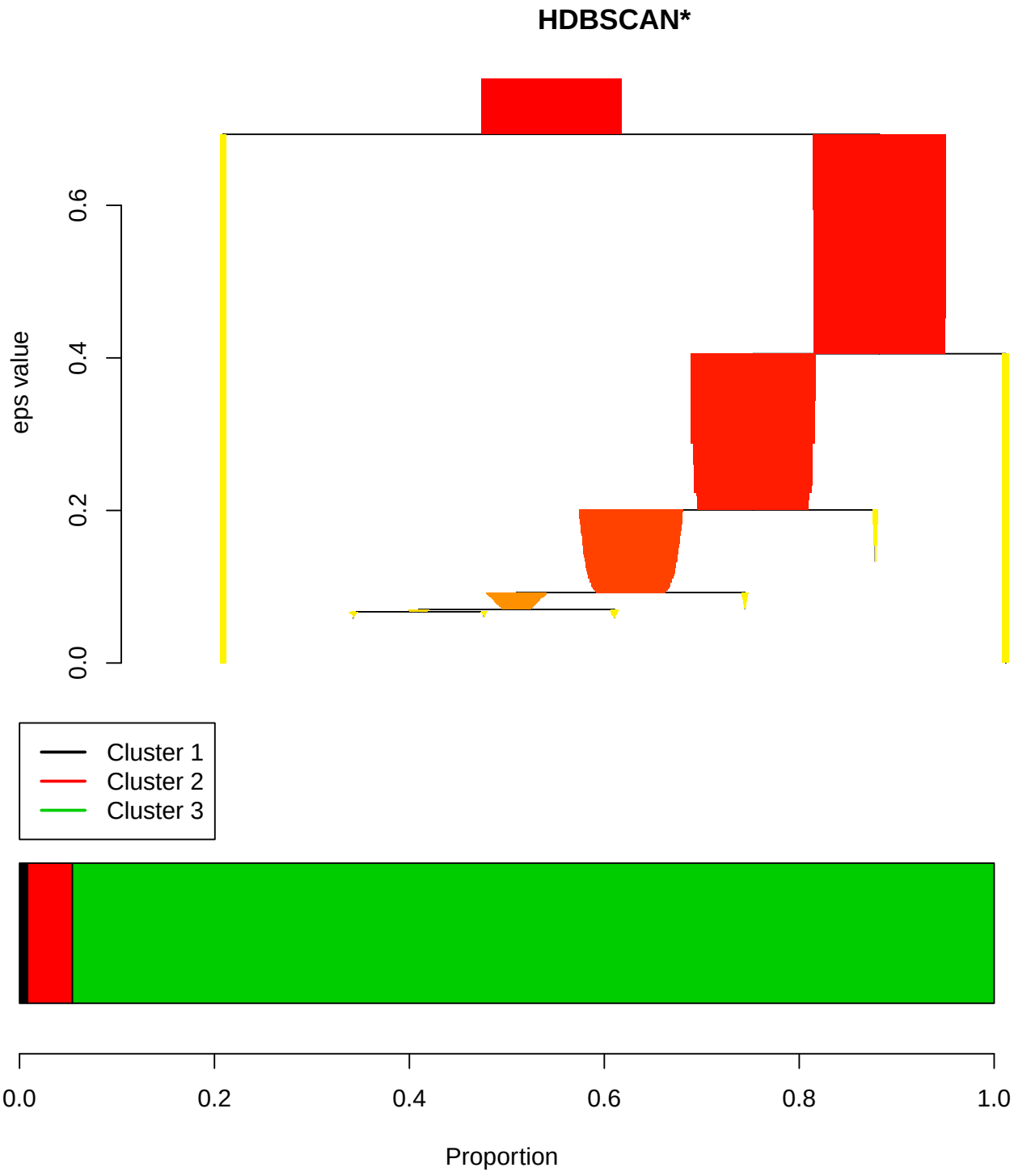

```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##  Cluster 1 Cluster 2 Cluster 3 Cluster 4
##      17505      2211      1100      3667
##
##
##
##  DB Population:
##  Cluster 1 Cluster 2 Cluster 3 Cluster 4
##       71.5        72        68      72.3
##
##
##
##  CSD Population:
##  Cluster 1 Cluster 2 Cluster 3 Cluster 4
##   242930.7  248623.4  224980.6  227749.3
##
##
##
##  CMA Type:
##   Cluster 1 Cluster 2 Cluster 3 Cluster 4
##        7566       972       474      1615
## B      7417       930       469      1496
## D      1910       230       125       395
## K       612        79        32       161
```

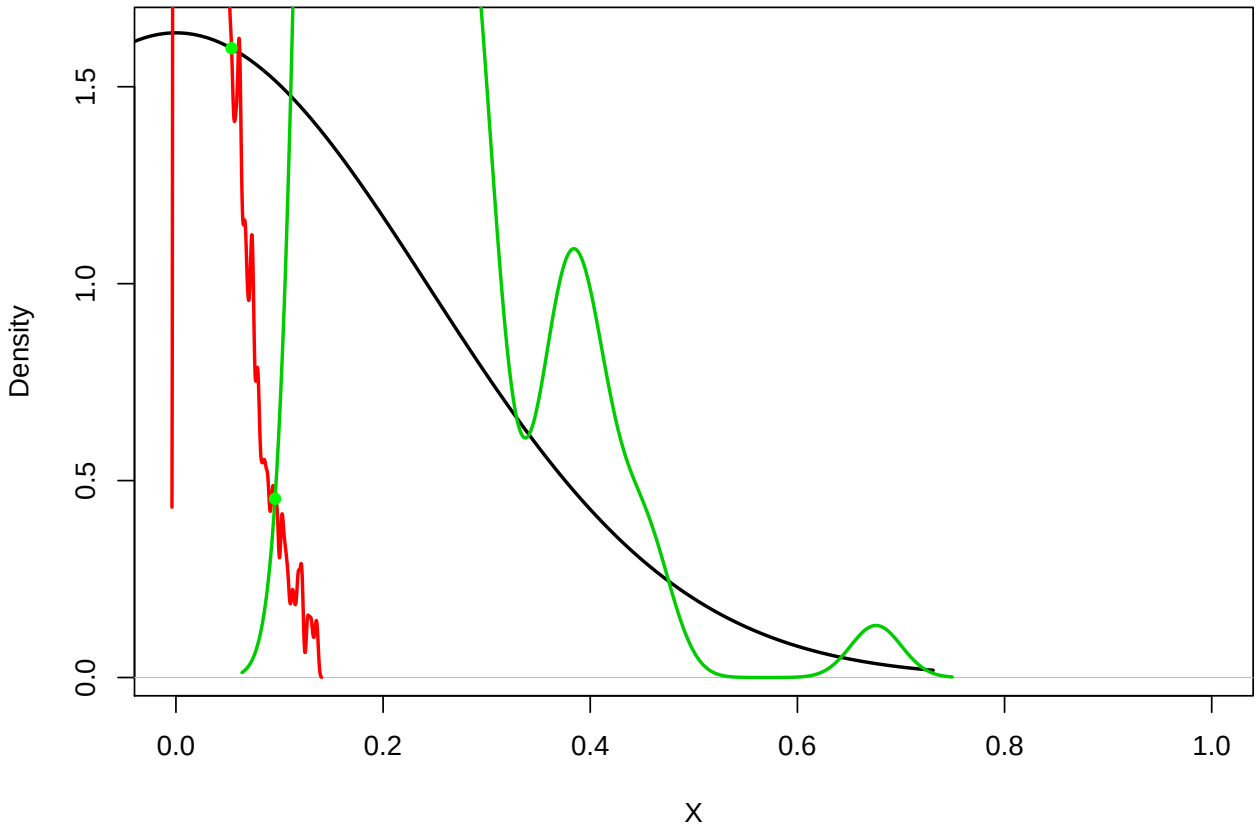
```
##
##
##
##  Index of Remoteness:
##  Cluster 1 Cluster 2 Cluster 3 Cluster 4
##      0.225     0.227     0.229     0.231
##
##
##
##  Provinces:
##                       Cluster 1 Cluster 2 Cluster 3 Cluster 4
## Alberta                     527        73        38       125
## BritishColumbia             859        98        58       184
## NewBrunswick                145        28         9        25
## NorthwestTerritories         14         1         0         1
## NovaScotia                  567        71        34       107
## Ontario                    2774       345       174       565
## Quebec                      997       116        69       196
## Saskatchewan                 82        10         7        22
## NA's                      11540      1469       711      2442
##
##
##
##  Amenity dense:
##   Cluster 1 Cluster 2 Cluster 3 Cluster 4
## 0     15841      2002       988      3287
## 1      1286       150        88       291
## 2       178        33        13        43
## F       200        26        11        46
##
##
##
##  PMS_prox_idx_childcare :
##  Cluster 1 Cluster 2 Cluster 3 Cluster 4
##    0.07676   0.07547   0.07958   0.07688
```

Health

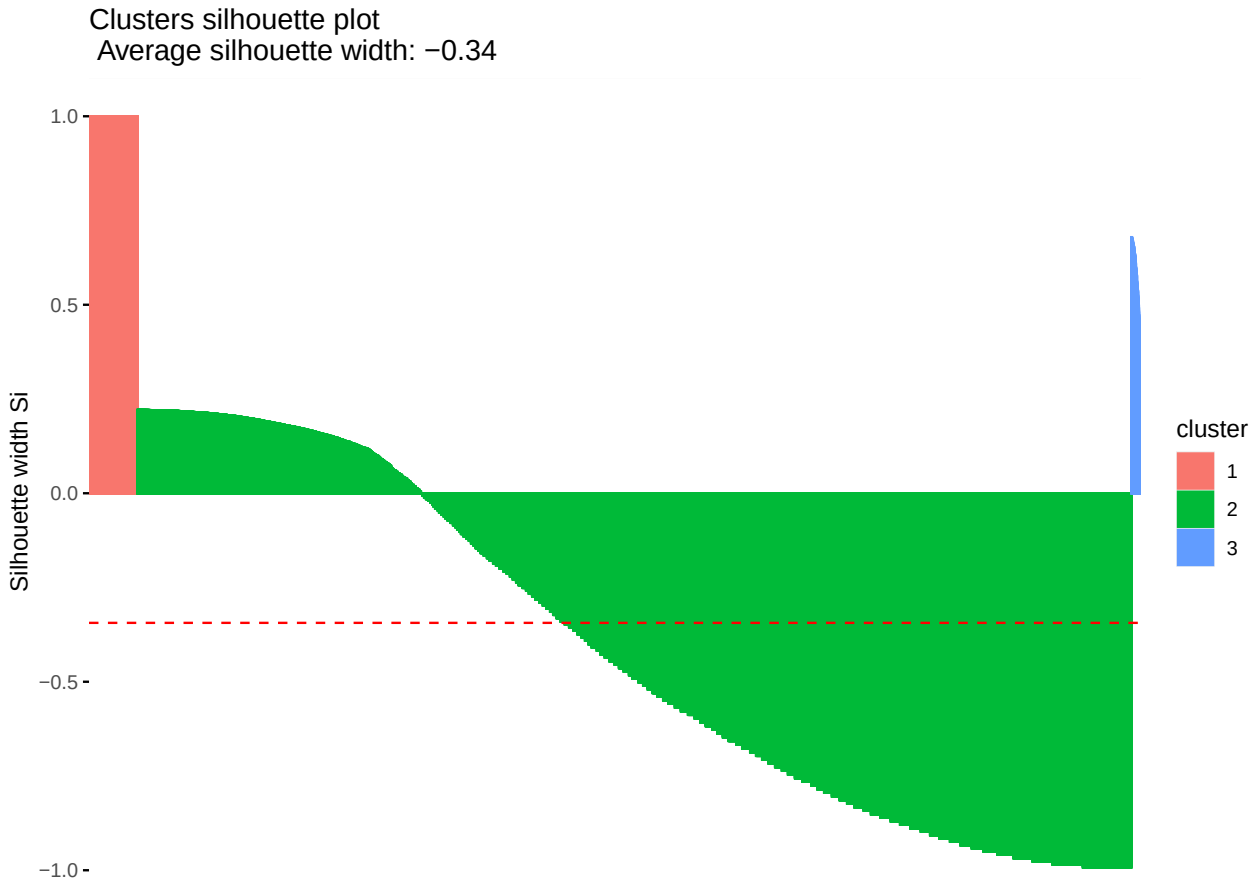
```
## [1] "Silhouette coefficient: 0.71150342405053"
## [1] "Xie Beni coefficient: 119651.993317046"
```

```
## Warning in if (is.na(res)) {: the condition has length > 1 and only the first
## element will be used
```






```
## [1] "Segment cutoff values:"
## [1] 0.05372428
## [1] 0.09568582
##   cluster  size ave.sil.width
## 1       1   687          1.00
## 2       2 14109         -0.42
## 3       3   123          0.59
```




```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##  Cluster 1 Cluster 2 Cluster 3
##       1135     23140       208
##
##
##
##  DB Population:
##  Cluster 1 Cluster 2 Cluster 3
##       73.1      71.5        64
##
##
##
##  CSD Population:
##  Cluster 1 Cluster 2 Cluster 3
##   241247.6  240424.7    228912
##
##
##
##  CMA Type:
##   Cluster 1 Cluster 2 Cluster 3
##         489     10052        86
## B       477      9754        81
## D       124      2507        29
## K        45       827        12
```

```
##
##
##
##  Index of Remoteness:
##  Cluster 1 Cluster 2 Cluster 3
##      0.227     0.226     0.242
##
##
##
##  Provinces:
##                       Cluster 1 Cluster 2 Cluster 3
## Alberta                      26       730         7
## BritishColumbia              49      1135        15
## NewBrunswick                  7       198         2
## NorthwestTerritories          0        16         0
## NovaScotia                   34       737         8
## Ontario                     190      3647        21
## Quebec                       60      1303        15
## Saskatchewan                  7       113         1
## NA's                        762     15261       139
##
##
##
##  Amenity dense:
##   Cluster 1 Cluster 2 Cluster 3
## 0      1030     20896       192
## 1        81      1724        10
## 2        14       248         5
## F        10       272         1
##
##
##
##  PMS_prox_idx_health :
##  Cluster 1 Cluster 2 Cluster 3
##    0.01254   0.01372   0.01896
```

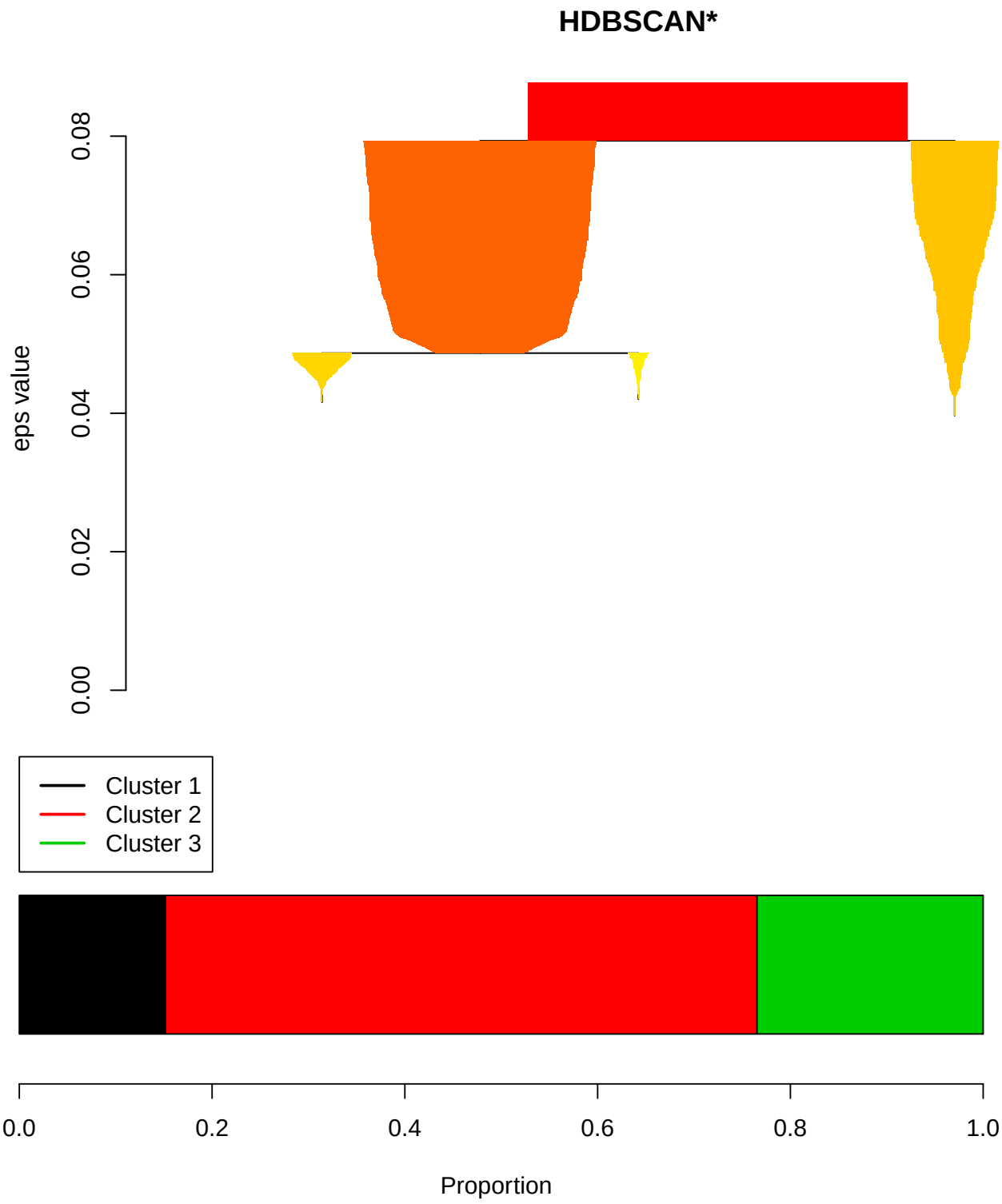
Grocery

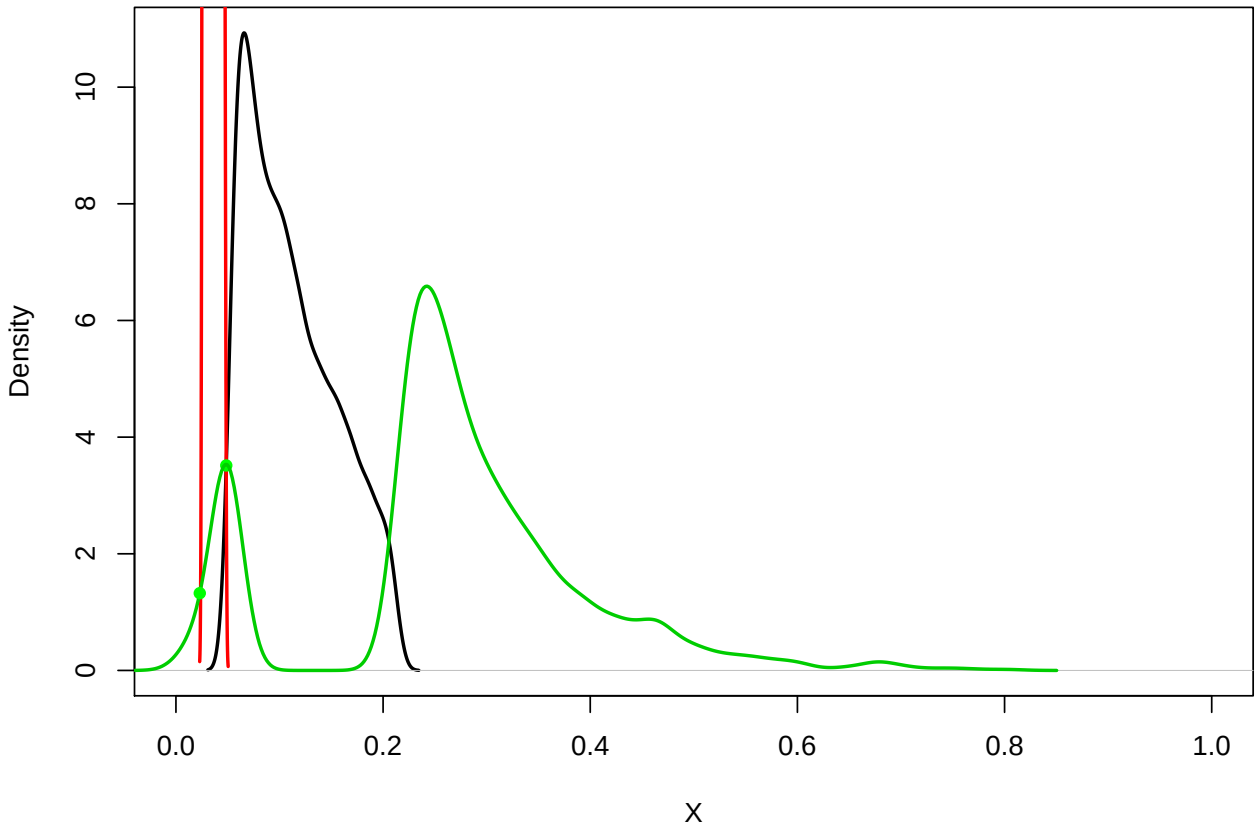
```
## [1] "NO CLUSTERS DETECTED"
```

Primary Education

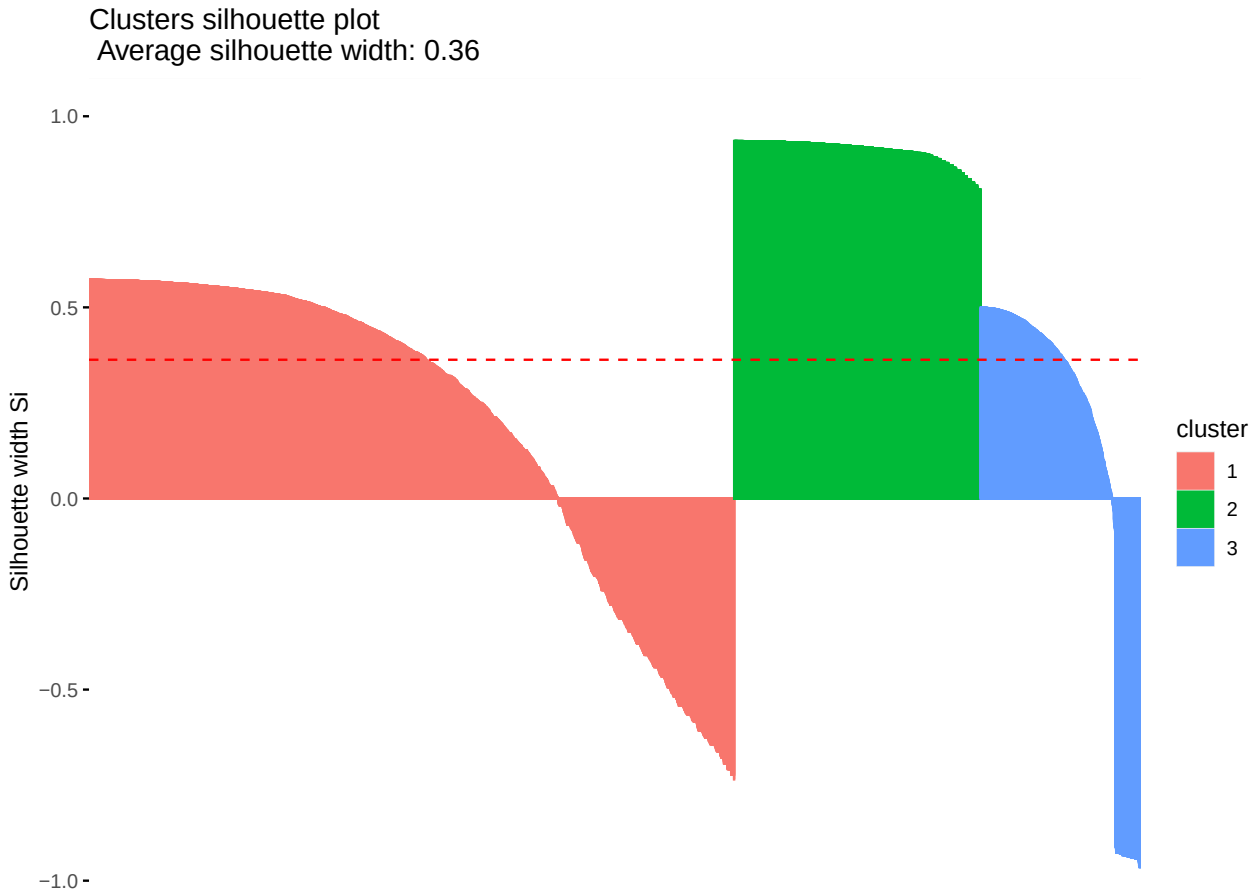
```
## [1] "Silhouette coefficient: 0.491369087890728"
## [1] "Xie Beni coefficient: 855312.566930832"
```

```
## Warning in if (is.na(res)) {: the condition has length > 1 and only the first
## element will be used
```






```
## [1] "Segment cutoff values:"
## [1] 0.04859196
## [1] 0.02301243
##   cluster size ave.sil.width
## 1       1 6896          0.20
## 2       2 2636          0.91
## 3       3 1703          0.16
```



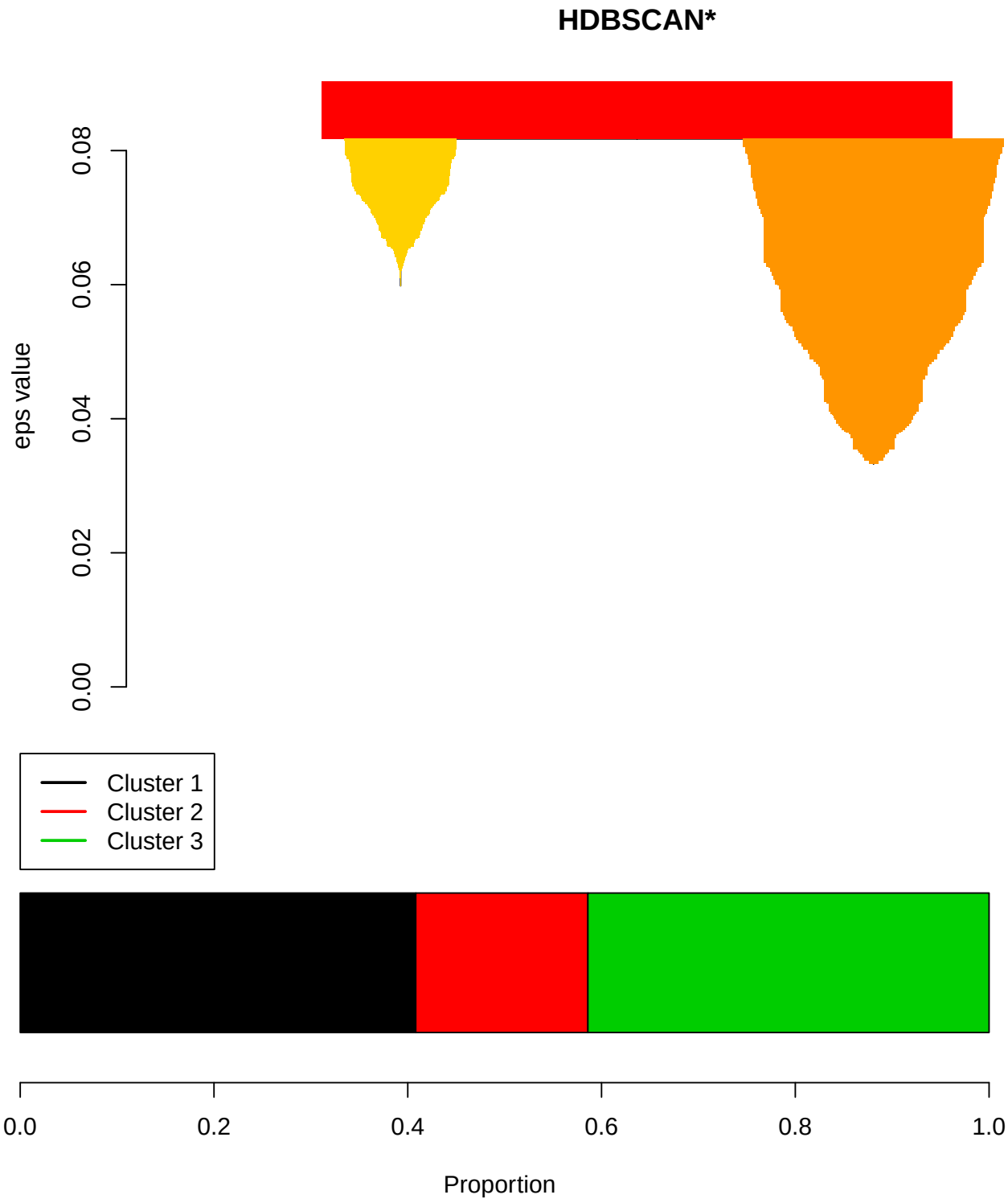

```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##  Cluster 1 Cluster 2 Cluster 3
##      15014      5753      3716
##
##
##
##  DB Population:
##  Cluster 1 Cluster 2 Cluster 3
##       71.6      69.6      73.8
##
##
##
##  CSD Population:
##  Cluster 1 Cluster 2 Cluster 3
##   240761.7  242149.5  235998.2
##
##
##
##  CMA Type:
##   Cluster 1 Cluster 2 Cluster 3
##        6511      2508      1608
## B      6342      2396      1574
## D      1629       634       397
## K       532       215       137
```

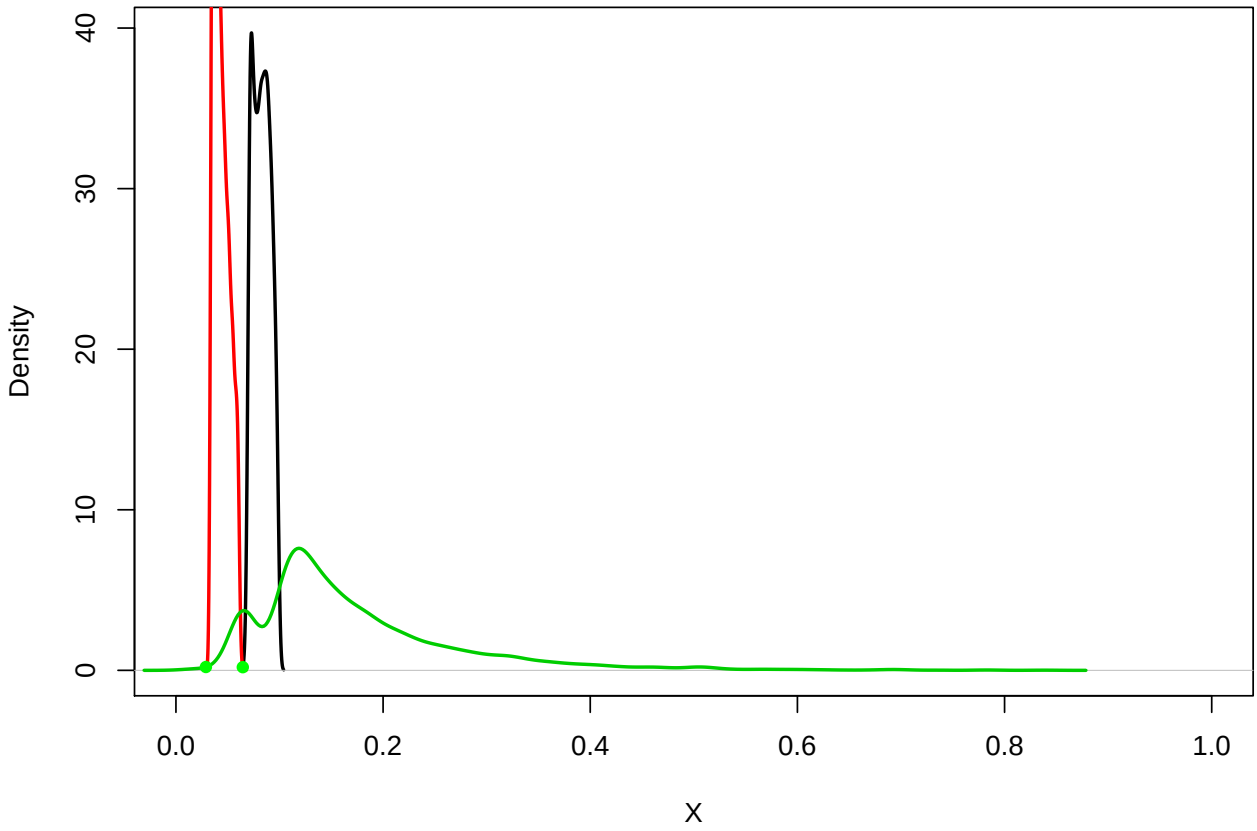
```
##
##
##
##  Index of Remoteness:
##  Cluster 1 Cluster 2 Cluster 3
##      0.227     0.224     0.226
##
##
##
##  Provinces:
##                       Cluster 1 Cluster 2 Cluster 3
## Alberta                     461       169       133
## BritishColumbia             732       266       201
## NewBrunswick                116        53        38
## NorthwestTerritories         14         0         2
## NovaScotia                  503       162       114
## Ontario                    2338       947       573
## Quebec                      854       316       208
## Saskatchewan                 64        41        16
## NA's                       9932      3799      2431
##
##
##
##  Amenity dense:
##   Cluster 1 Cluster 2 Cluster 3
## 0     13563      5194      3361
## 1      1098       442       275
## 2       170        66        31
## F       183        51        49
##
##
##
##  PMS_prox_idx_educpri :
##  Cluster 1 Cluster 2 Cluster 3
##    0.11763   0.11848   0.11501
```

Secondary Education

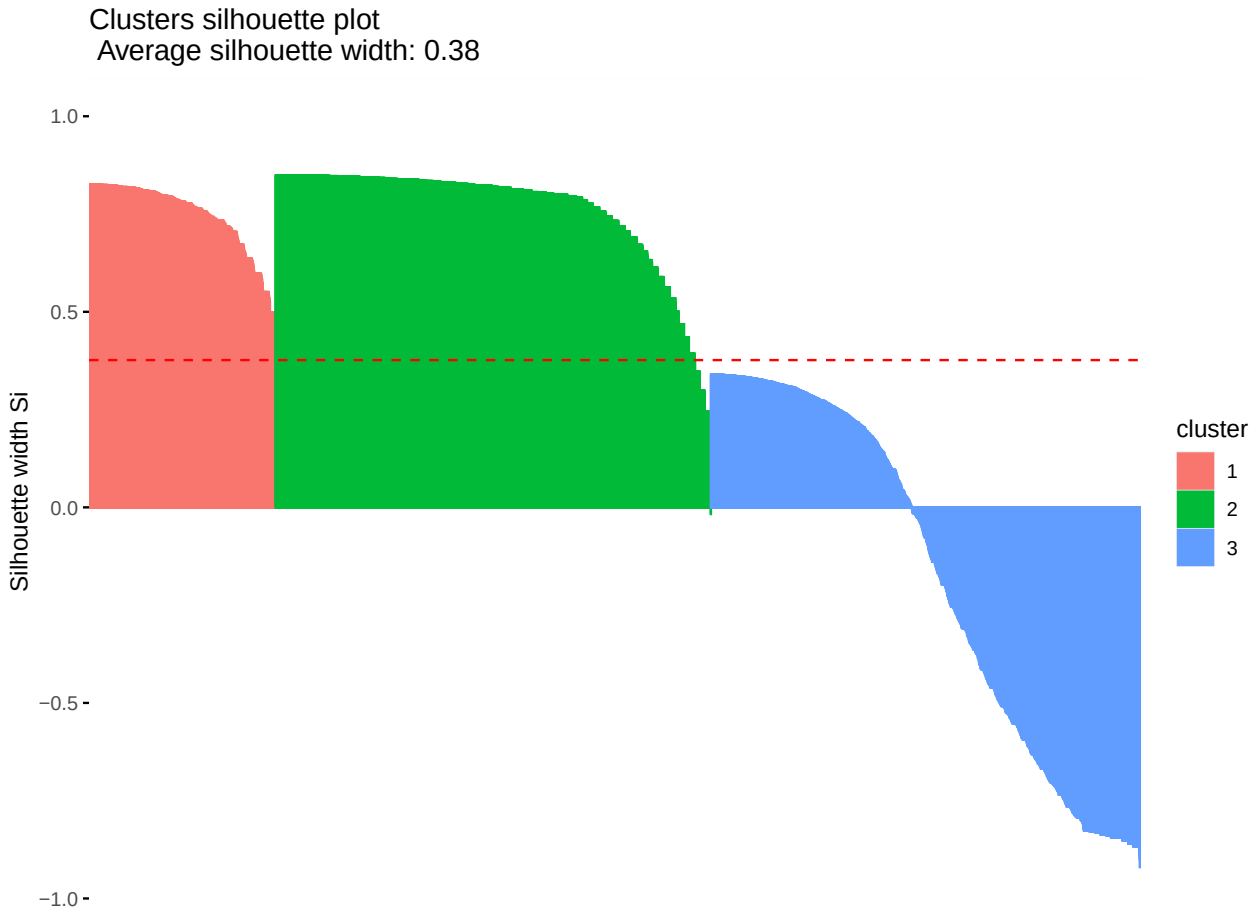
```
## [1] "Silhouette coefficient: 0.502911585364128"
## [1] "Xie Beni coefficient: 121053.51347719"
```

```
## Warning in if (is.na(res)) {: the condition has length > 1 and only the first
## element will be used
```






```
## [1] "Segment cutoff values:"
## [1] 0.06455027
## [1] 0.02893298
##   cluster size ave.sil.width
## 1       1 1255          0.75
## 2       2 2925          0.76
## 3       3 2884         -0.17
```




```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##  Cluster 1 Cluster 2 Cluster 3
##       4341     10167      9975
##
##
##
##  DB Population:
##  Cluster 1 Cluster 2 Cluster 3
##       67.6      71.6        73
##
##
##
##  CSD Population:
##  Cluster 1 Cluster 2 Cluster 3
##     233466  240213.5  243521.5
##
##
##
##  CMA Type:
##    Cluster 1 Cluster 2 Cluster 3
##         1949      4384      4294
## B       1746      4290      4276
## D        480      1105      1075
## K        166       388       330
```

```
##
##
##
##  Index of Remoteness:
##  Cluster 1 Cluster 2 Cluster 3
##      0.231     0.226     0.224
##
##
##
##  Provinces:
##                       Cluster 1 Cluster 2 Cluster 3
## Alberta                     128       326       309
## BritishColumbia             204       491       504
## NewBrunswick                 39        96        72
## NorthwestTerritories          3         6         7
## NovaScotia                  151       312       316
## Ontario                     642      1592      1624
## Quebec                      262       559       557
## Saskatchewan                 20        53        48
## NA's                       2892      6732      6538
##
##
##
##  Amenity dense:
##   Cluster 1 Cluster 2 Cluster 3
## 0      3957      9178      8983
## 1       295       760       760
## 2        43       109       115
## F        46       120       117
##
##
##
##  PMS_prox_idx_educsec :
##  Cluster 1 Cluster 2 Cluster 3
##    0.10759   0.10224   0.10344
```

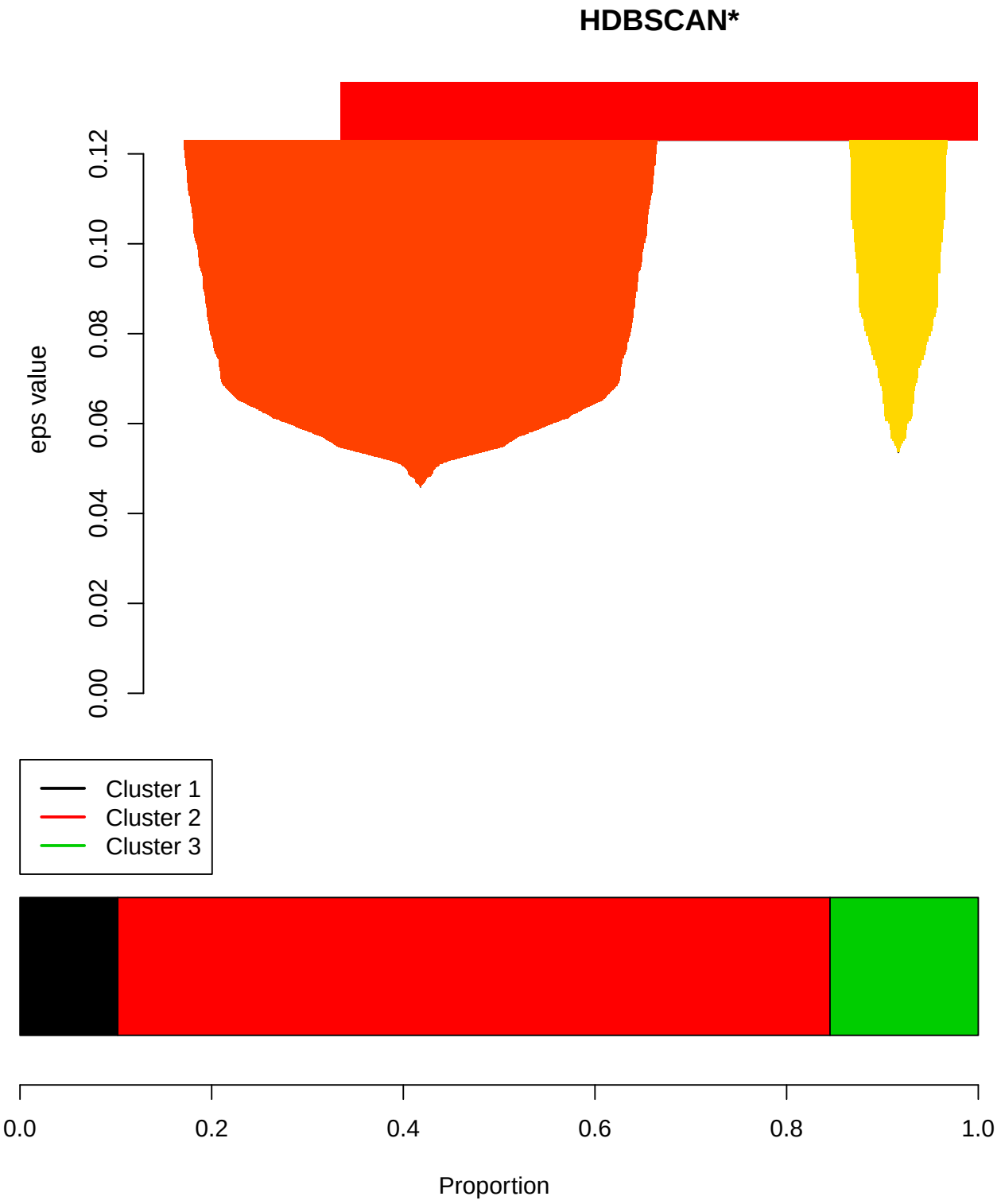
Libraries

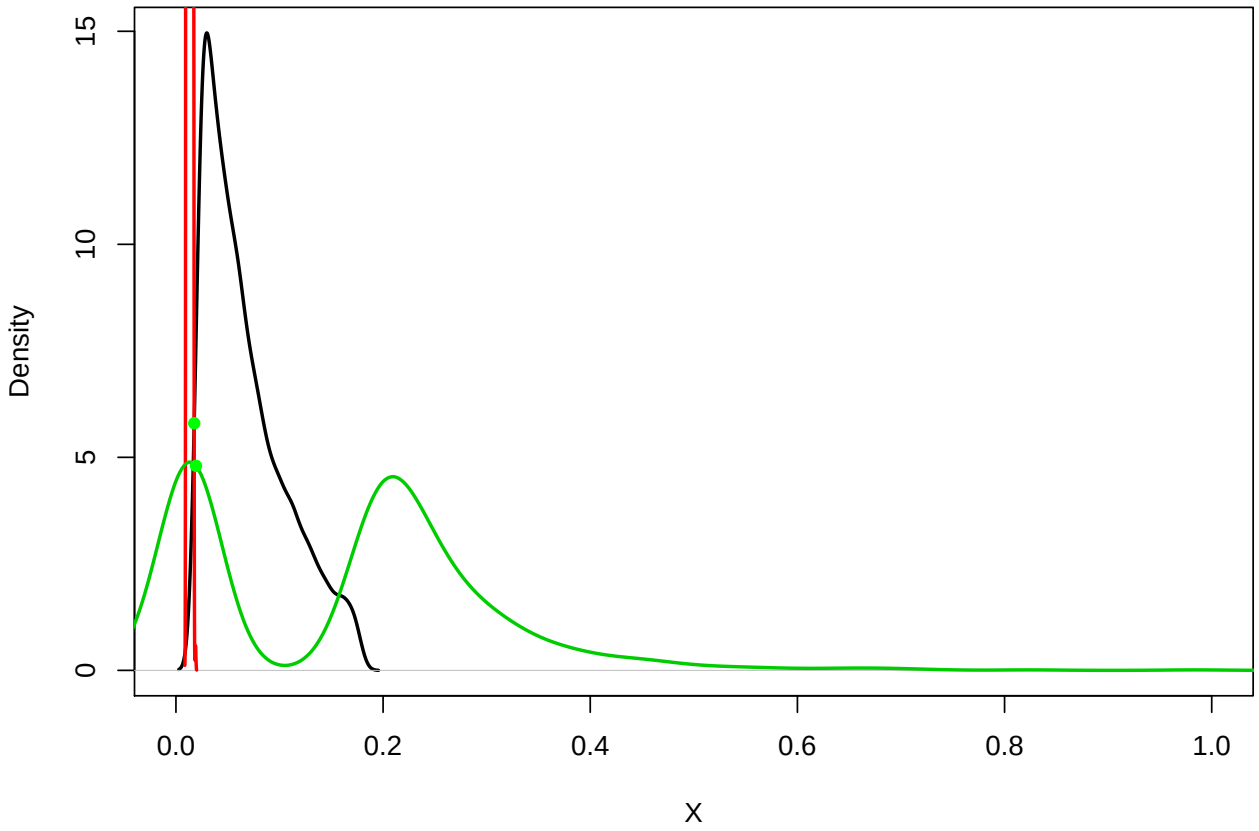
```
## [1] "NO CLUSTERS DETECTED"
```

Parks

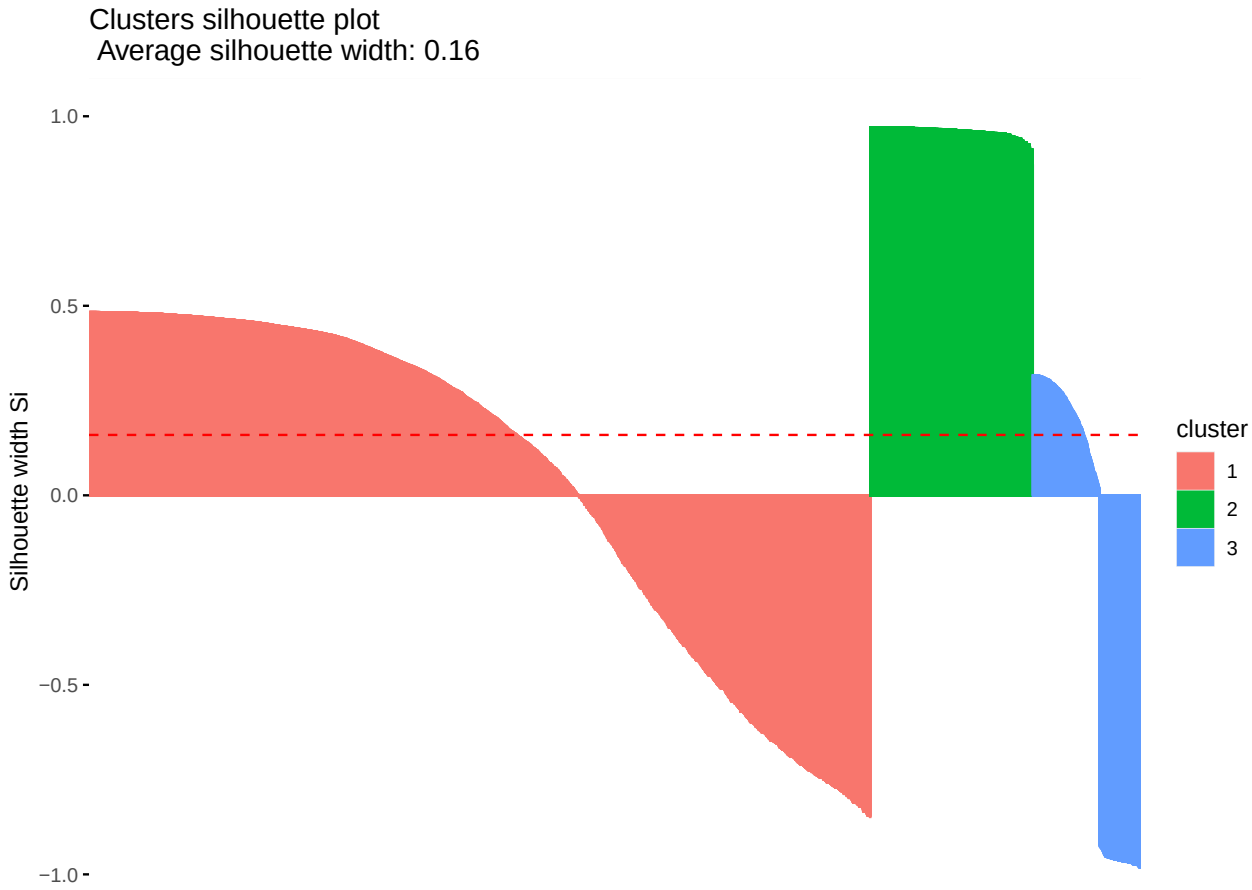
```
## [1] "Silhouette coefficient: 0.343965286339036"
## [1] "Xie Beni coefficient: Inf"
```

```
## Warning in if (is.na(res)) {: the condition has length > 1 and only the first
## element will be used
```






```
## [1] "Segment cutoff values:"
## [1] 0.01765479
## [1] 0.01937567
##   cluster size ave.sil.width
## 1       1 8696          0.04
## 2       2 1810          0.96
## 3       3 1193         -0.22
```




```
## [1] "Cluster profiles:"
## [1] "Num of DBs:"
##  Cluster 1 Cluster 2 Cluster 3
##      18236      3764      2483
##
##
##
##  DB Population:
##  Cluster 1 Cluster 2 Cluster 3
##       71.6      73.5      67.8
##
##
##
##  CSD Population:
##  Cluster 1 Cluster 2 Cluster 3
##   239109.5  239498.1  250897.9
##
##
##
##  CMA Type:
##   Cluster 1 Cluster 2 Cluster 3
##        7889      1643      1095
## B      7671      1599      1042
## D      1991       409       260
## K       685       113        86
```

```
##
##
##
##  Index of Remoteness:
##  Cluster 1 Cluster 2 Cluster 3
##      0.226     0.224     0.229
##
##
##
##  Provinces:
##                       Cluster 1 Cluster 2 Cluster 3
## Alberta                     583       102        78
## BritishColumbia             905       162       132
## NewBrunswick                165        25        17
## NorthwestTerritories         12         3         1
## NovaScotia                  561       139        79
## Ontario                    2890       593       375
## Quebec                     1031       203       144
## Saskatchewan                 92        20         9
## NA's                      11997      2517      1648
##
##
##
##  Amenity dense:
##   Cluster 1 Cluster 2 Cluster 3
## 0     16471      3409      2238
## 1      1365       263       187
## 2       196        41        30
## F       204        51        28
##
##
##
##  PMS_prox_idx_parks :
##  Cluster 1 Cluster 2 Cluster 3
##    0.06832   0.07044   0.06805
```

Transit

```
## [1] "NO CLUSTERS DETECTED"
```

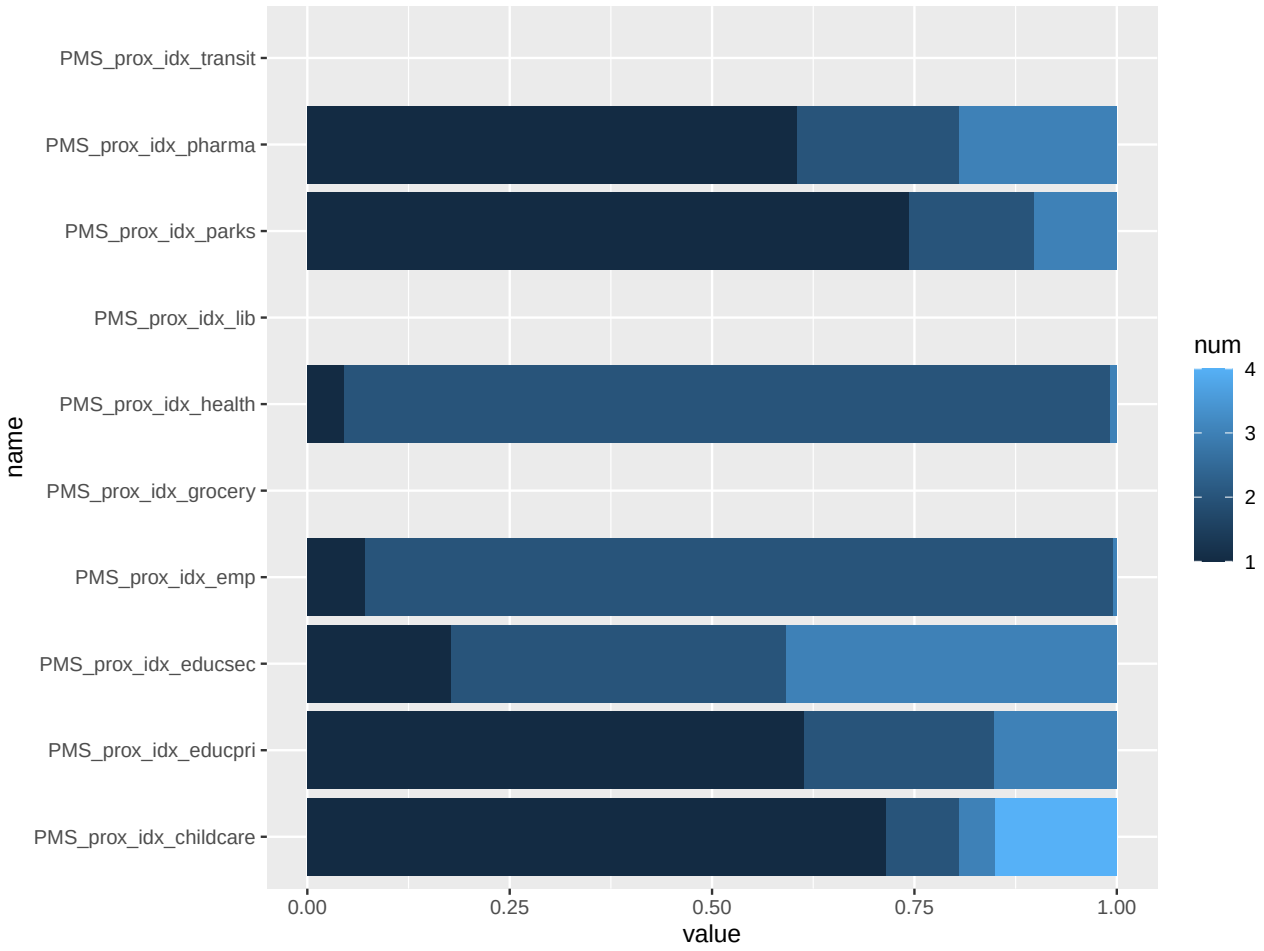
Conclusion

text

```
## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
## i Please use `all_of()` or `any_of()` instead.
##   # Was:
##   data %>% select(amenities)
##
##   # Now:
##   data %>% select(all_of(amenities))
##
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
```

```
## # A tibble: 40 x 3
##    name                     value   num
##    <chr>                    <dbl> <int>
##  1 PMS_prox_idx_emp        0.0719     1
##  2 PMS_prox_idx_pharma     0.605      1
##  3 PMS_prox_idx_childcare  0.715      1
##  4 PMS_prox_idx_health     0.0460     1
##  5 PMS_prox_idx_grocery   NA          1
##  6 PMS_prox_idx_educpri    0.614      1
##  7 PMS_prox_idx_educsec    0.178      1
##  8 PMS_prox_idx_lib       NA          1
##  9 PMS_prox_idx_parks      0.743      1
## 10 PMS_prox_idx_transit   NA          1
## # ... with 30 more rows
```

```
## Warning: Removed 18 rows containing missing values (`position_stack()`).
```


PMS_prox_idx_transit
PMS_prox_idx_pharma
PMS_prox_idx_parks
PMS_prox_idx_lib
PMS_prox_idx_health
PMS_prox_idx_grocery
PMS_prox_idx_emp
PMS_prox_idx_educsec
PMS_prox_idx_educpri
PMS_prox_idx_childcare
name
0.00
0.25
0.50
0.75
1.00
value
num
4
3
2
1