R Notebook

COVID PHU-Level Clustering

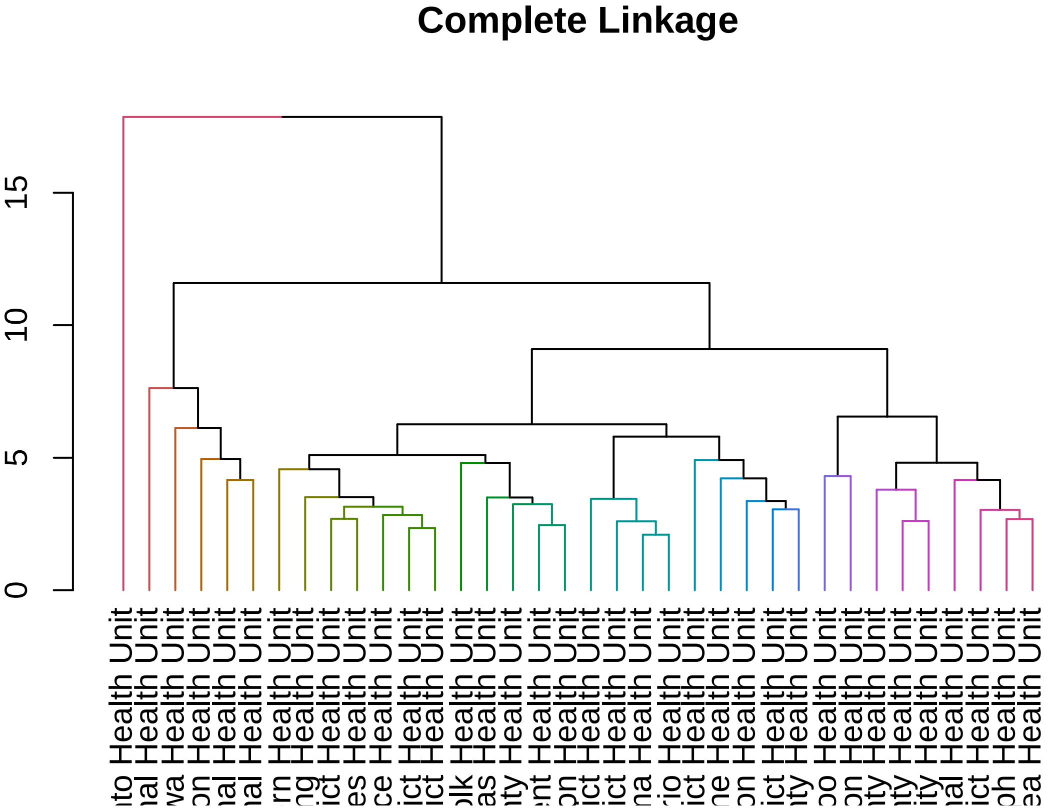
Hierarchical Clustering Methods

Agglomerative Clustering:

Data is first scaled then the euclidean distance is computed between each PHU. Clusters are formed based on least dissimilarities between observations.

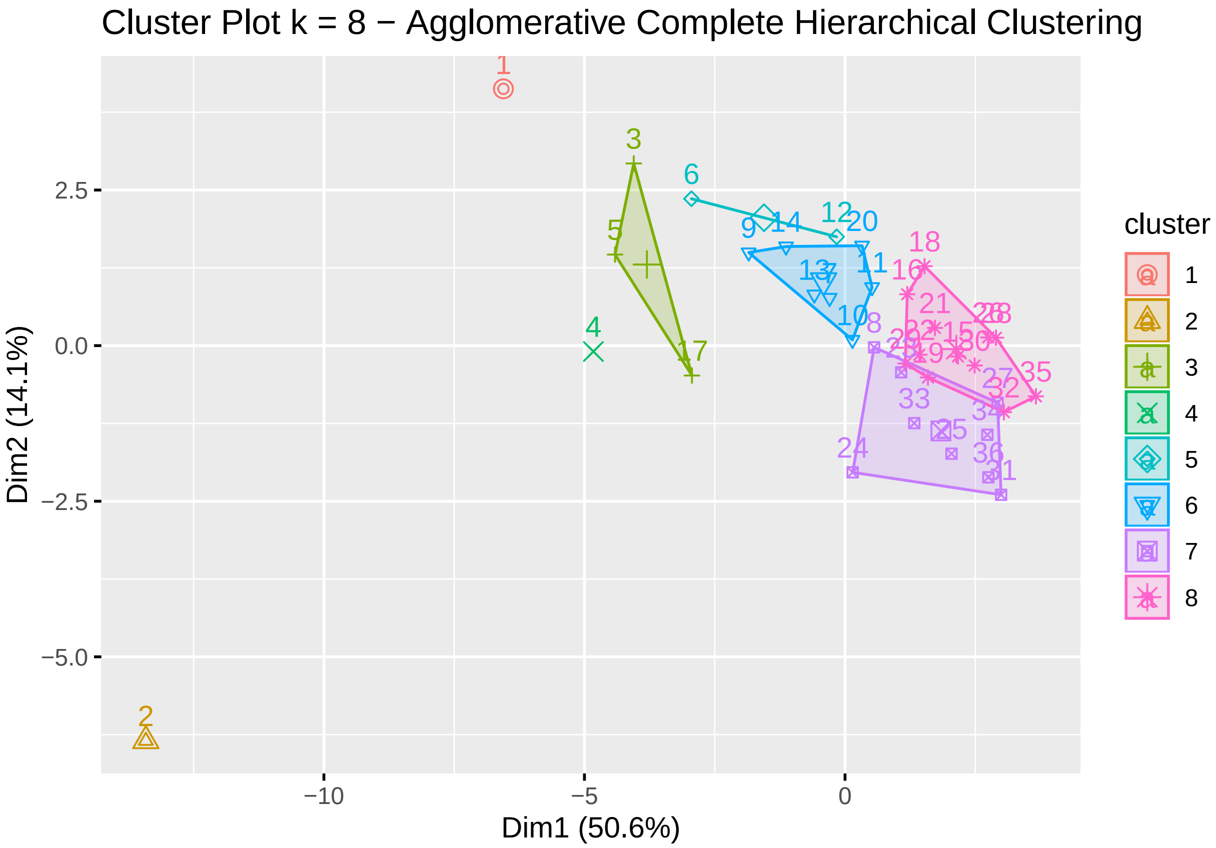
Group average clustering: attempts to produce relatively compact that are relatively far apart. *Complete clustering:* all union observations for two groups are least dissimilar - can produce clusters that violate "closeness" property (one observation may be more similar to observations in other groups). *Single clustering:* grouped based on one dissimilarity measure common between two observations being very small - prone to chaining.

Single linkage appears to have chaining while average linkage groups are not as easily discernable. Complete linkage was chosen for further analysis.



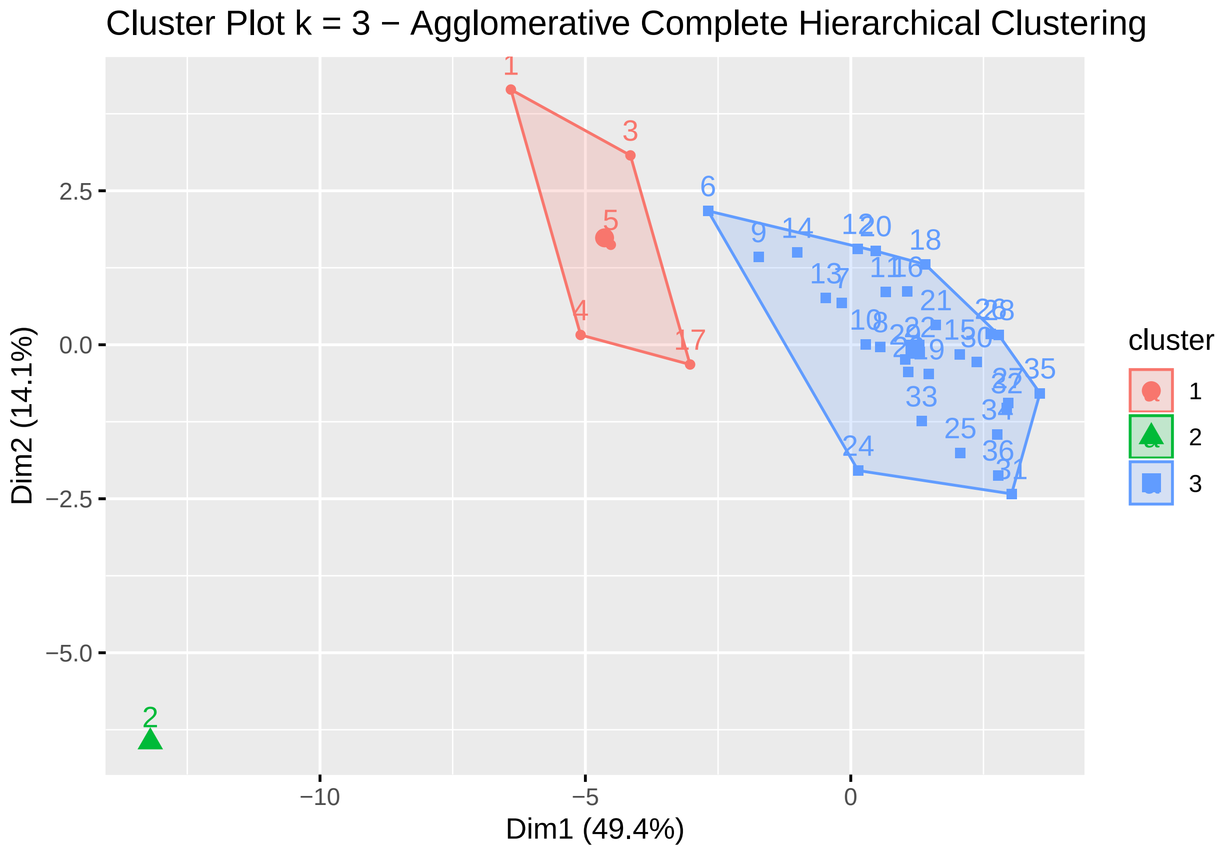
With 8 groups it seems as though there is too much separation…

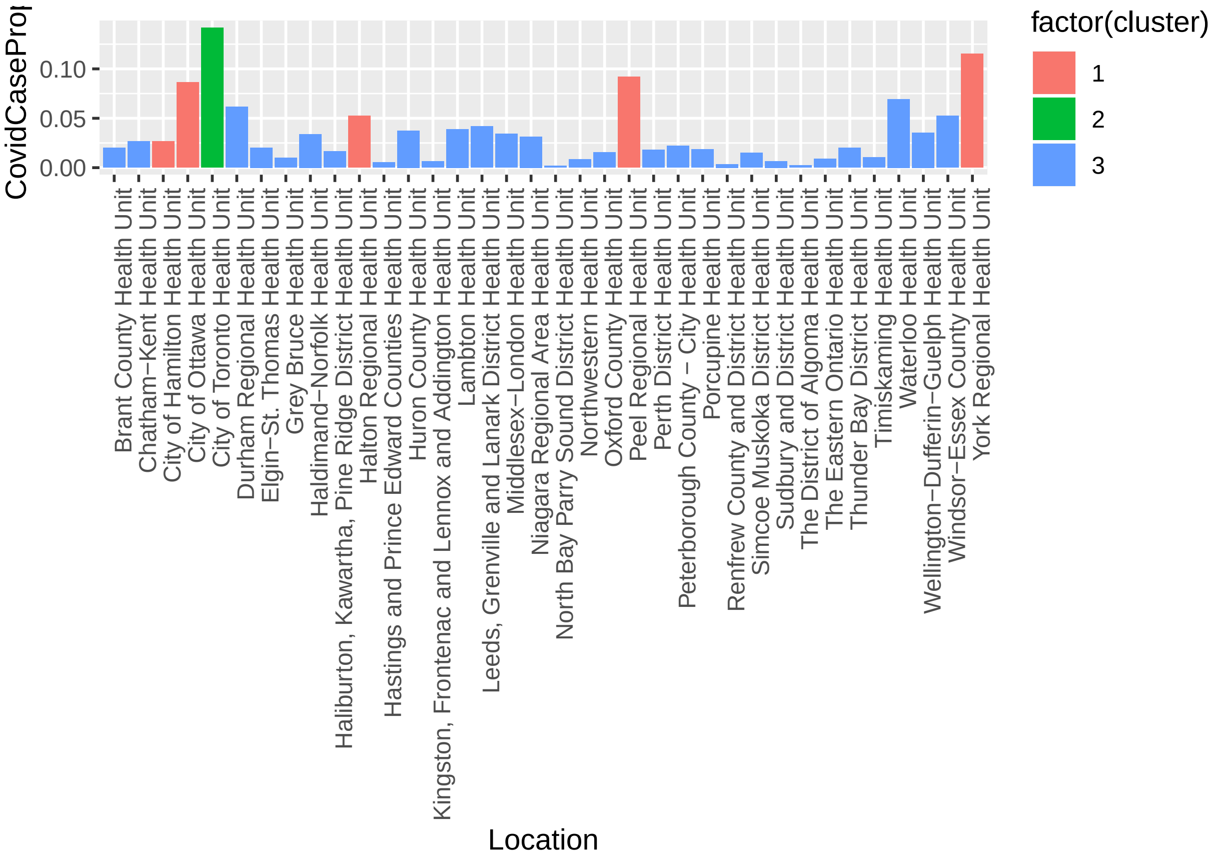
```
## # A tibble: 8 x 2
##   cluster     n
##     <int> <int>
## 1       1     1
## 2       2     1
## 3       3     3
## 4       4     1
## 5       5     2
## 6       6     7
## 7       7     9
## 8       8    12
```



With 3 groups it appears as though there is not enough separation…

```
## # A tibble: 3 x 2
##   cluster     n
##     <int> <int>
## 1       1     5
## 2       2     1
## 3       3    30
```


Cluster Plot k = 3 − Agglomerative Complete Hierarchical Clustering
Dim2 (14.1%)
Dim1 (49.4%)
2.5
0.0
−2.5
−5.0
−10
−5
0
1
3
5
4
17
2
6
9
14
12
20
18
13
7
11
16
21
10
8
23
22
15
28
30
19
35
27
32
33
34
25
24
36
31
cluster
1
2
3

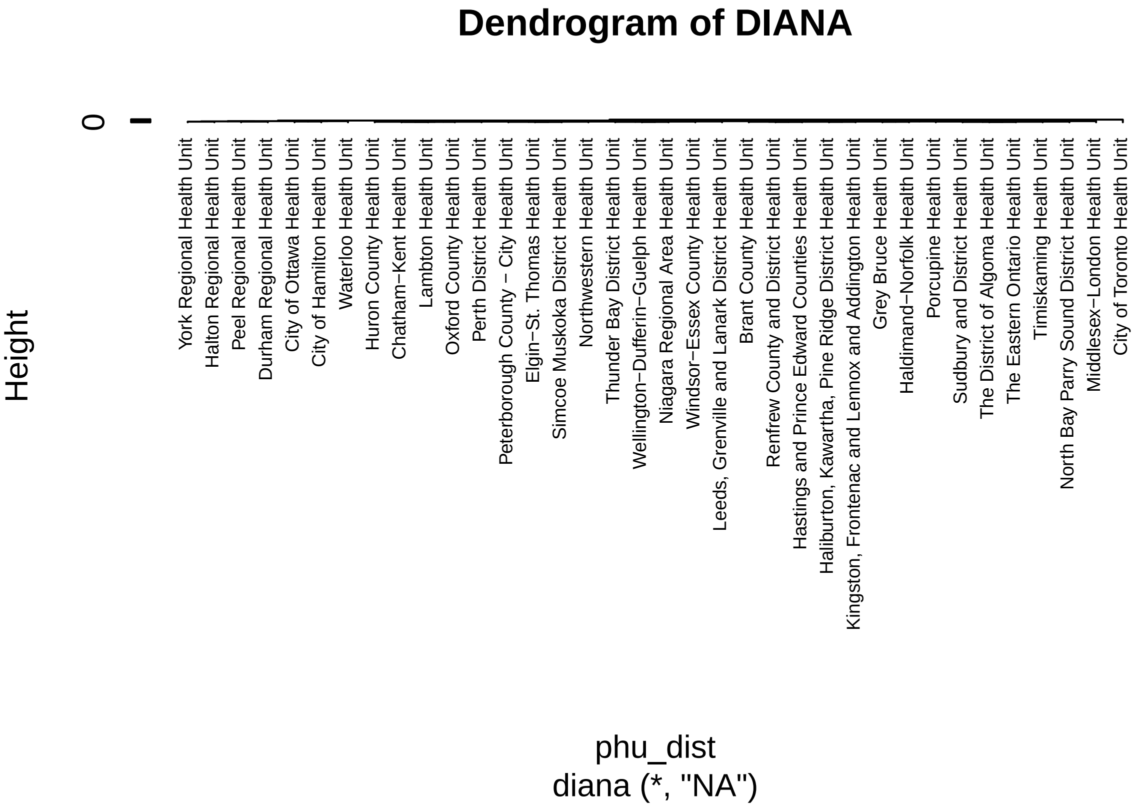

CovidCaseProp
0.10
0.05
0.00
Brant County Health Unit
Chatham−Kent Health Unit
City of Hamilton Health Unit
City of Ottawa Health Unit
City of Toronto Health Unit
Durham Regional Health Unit
Elgin−St. Thomas Health Unit
Grey Bruce Health Unit
Haldimand−Norfolk Health Unit
Haliburton, Kawartha, Pine Ridge District Health Unit
Halton Regional Health Unit
Hastings and Prince Edward Counties Health Unit
Huron County Health Unit
Kingston, Frontenac and Lennox and Addington Health Unit
Lambton Health Unit
Leeds, Grenville and Lanark District Health Unit
Middlesex−London Health Unit
Niagara Regional Area Health Unit
North Bay Parry Sound District Health Unit
Northwestern Health Unit
Oxford County Health Unit
Peel Regional Health Unit
Perth District Health Unit
Peterborough County − City Health Unit
Porcupine Health Unit
Renfrew County and District Health Unit
Simcoe Muskoka District Health Unit
Sudbury and District Health Unit
The District of Algoma Health Unit
The Eastern Ontario Health Unit
Thunder Bay District Health Unit
Timiskaming Health Unit
Waterloo Health Unit
Wellington−Dufferin−Guelph Health Unit
Windsor−Essex County Health Unit
York Regional Health Unit
Location
factor(cluster)
1
2
3

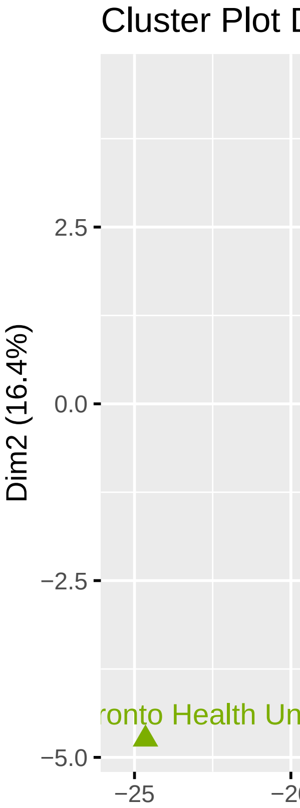
Divisive Hierarchical Clustering - DIANA

All observations begin as a single cluster and are divisible separated based on largest average dissimilarity between an observation and all others.

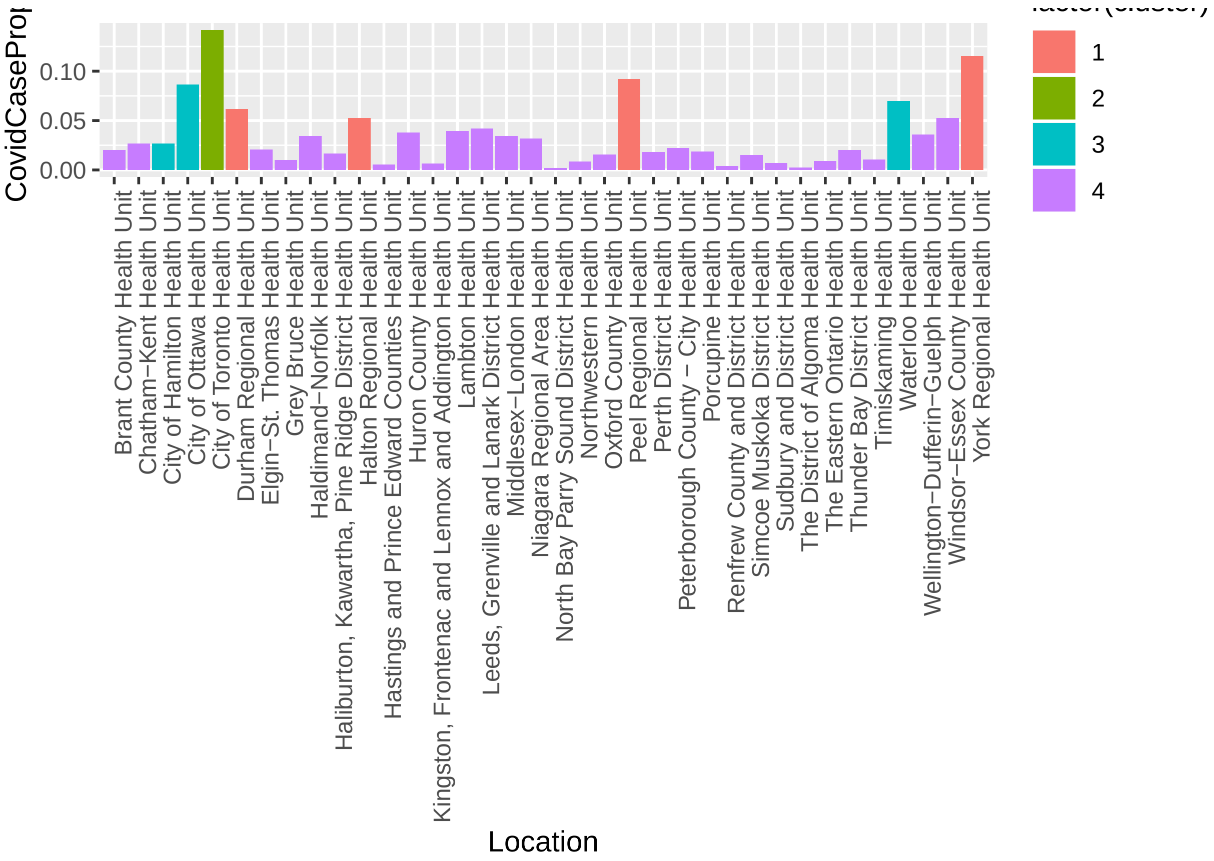
Hierarchical clustering will always cluster even if there is no grouping in the data. The divisive coefficient is on a 0-1 scale, representing the extent to which the hierarchical structure produced by a dendrogram actually represents the data (closer to 1 = better representation).

```
## [1] "Divisive Coefficient:  0.769955649859812"
```





Covid Case Total Proportions (per PHU) data and DIANA cluster assignment are added to `phu_prop` to assess groupings by covid output.



Mixture Model Clustering Methods

EM Algorithm

Soft version of K-means clustering that is based on probabilities rather than deterministic assignments of points to groups. Used package `Mclust`.

For all parameters:

Between PHU and proportion of cases:

```r
# EM view classes ----
classes <- mutate(as.data.frame(mcovid$classification), loc = phu$Location[mcovid$data[,2]])
colnames(classes)[1] <- 'class'
classes <- classes[order(classes),][1:36,]
classes
```

```
##    class                                                     loc
## 1      1              North Bay Parry Sound District Health Unit
## 2      1                               Peel Regional Health Unit
## 4      1                              City of Ottawa Health Unit
## 5      1                                     Lambton Health Unit
## 6      1 Kingston, Frontenac and Lennox and Addington Health Unit
## 9      1                                    Waterloo Health Unit
## 13     1                         The Eastern Ontario Health Unit
## 14     1                                 Timiskaming Health Unit
## 3      2                  Peterborough County - City Health Unit
## 7      2                  Wellington-Dufferin-Guelph Health Unit
## 12     2                            City of Hamilton Health Unit
## 16     2                              City of Toronto Health Unit
## 17     2                              Halton Regional Health Unit
## 18     2                                 Huron County Health Unit
## 20     2                             Elgin-St. Thomas Health Unit
## 22     2                                 Northwestern Health Unit
## 23     2                                York Regional Health Unit
## 24     2                                 Chatham-Kent Health Unit
## 28     2                               Perth District Health Unit
## 29     2                              Durham Regional Health Unit
## 30     2                         Thunder Bay District Health Unit
## 32     2                             Middlesex-London Health Unit
## 33     2                         Windsor-Essex County Health Unit
## 8      3                                 Brant County Health Unit
## 10     3                       The District of Algoma Health Unit
## 11     3         Leeds, Grenville and Lanark District Health Unit
## 15     3                        Niagara Regional Area Health Unit
## 19     3                      Simcoe Muskoka District Health Unit
## 21     3                         Sudbury and District Health Unit
## 25     3                                    Porcupine Health Unit
## 26     3                  Renfrew County and District Health Unit
## 27     3                                   Grey Bruce Health Unit
## 31     3                            Haldimand-Norfolk Health Unit
## 34     3    Haliburton, Kawartha, Pine Ridge District Health Unit
## 35     3          Hastings and Prince Edward Counties Health Unit
## 36     3                                Oxford County Health Unit
```

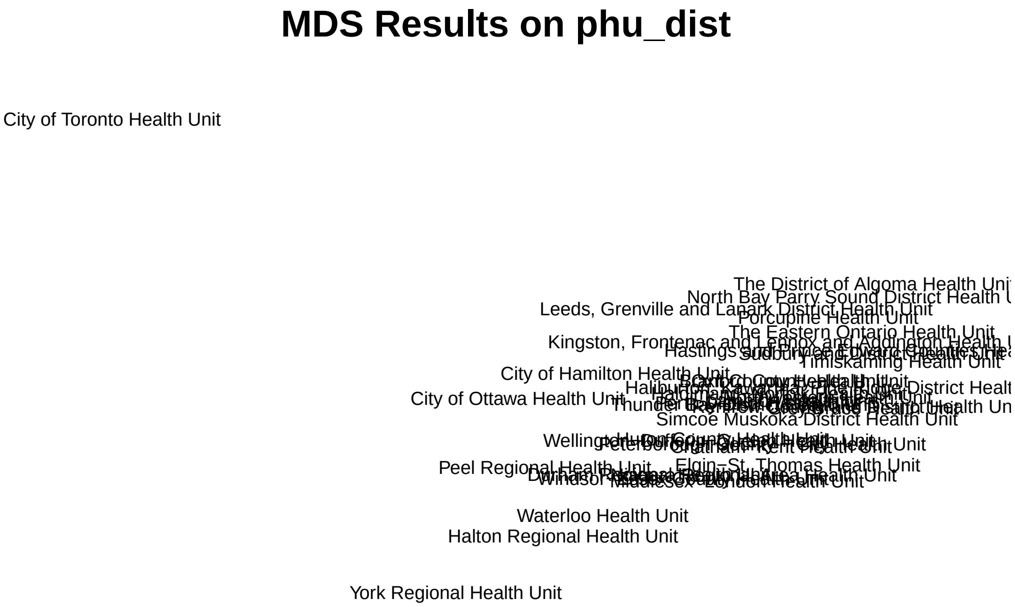
Multidimensional Scaling

Uses euclidean distances of scaled PHU data. Similarly represents hierarchical clustering (Toronto is its own cluster, denser regions such as Peel, Halton, York, Waterloo, Ottawa…etc. separate from smaller districts).

```
labels(phu_dist)
```

```
##  [1] "York Regional Health Unit"
##  [2] "City of Toronto Health Unit"
##  [3] "Halton Regional Health Unit"
##  [4] "City of Ottawa Health Unit"
##  [5] "Peel Regional Health Unit"
##  [6] "Waterloo Health Unit"
##  [7] "Huron County Health Unit"
##  [8] "Perth District Health Unit"
##  [9] "Durham Regional Health Unit"
## [10] "Thunder Bay District Health Unit"
## [11] "Peterborough County - City Health Unit"
## [12] "Middlesex-London Health Unit"
## [13] "Wellington-Dufferin-Guelph Health Unit"
## [14] "Windsor-Essex County Health Unit"
## [15] "Northwestern Health Unit"
## [16] "Chatham-Kent Health Unit"
## [17] "City of Hamilton Health Unit"
## [18] "Elgin-St. Thomas Health Unit"
## [19] "Oxford County Health Unit"
## [20] "Niagara Regional Area Health Unit"
## [21] "Simcoe Muskoka District Health Unit"
## [22] "Lambton Health Unit"
## [23] "Brant County Health Unit"
## [24] "Leeds, Grenville and Lanark District Health Unit"
## [25] "Porcupine Health Unit"
## [26] "Renfrew County and District Health Unit"
## [27] "Sudbury and District Health Unit"
## [28] "Grey Bruce Health Unit"
## [29] "Haldimand-Norfolk Health Unit"
## [30] "Haliburton, Kawartha, Pine Ridge District Health Unit"
## [31] "The District of Algoma Health Unit"
## [32] "Hastings and Prince Edward Counties Health Unit"
## [33] "Kingston, Frontenac and Lennox and Addington Health Unit"
## [34] "The Eastern Ontario Health Unit"
## [35] "Timiskaming Health Unit"
## [36] "North Bay Parry Sound District Health Unit"
```

```
plot(cmdscale(phu_dist)[,1], cmdscale(phu_dist)[,2], type = "n",
     xlab = "", ylab = "",
     asp = 1, axes = FALSE,
     main = "MDS Results on phu_dist")
text(cmdscale(phu_dist)[,1], cmdscale(phu_dist)[,2], rownames(cmdscale(phu_dist)), cex = 0.6)
```



Other unexplored methods:

Combinatorial - works directly with data without assuming underlying probabilistic model.