ETC3250 Lab 12

Di Cook

20 October 2015

Unsupervised Classification

THERE IS NOTHING TO TURN IN FOR THIS WEEK

Task 1

We are going to take the proximity output from a random forest fit to the paintings training data. This is a matrix of information. What are the dimensions of the matrix?

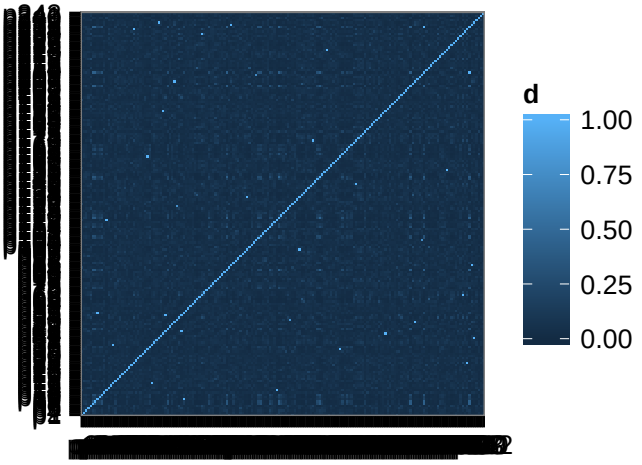
Task 2

Write a paragraph describing how the proximity matrix is calculated. Can it be considered to be a distance metric?

Task 3

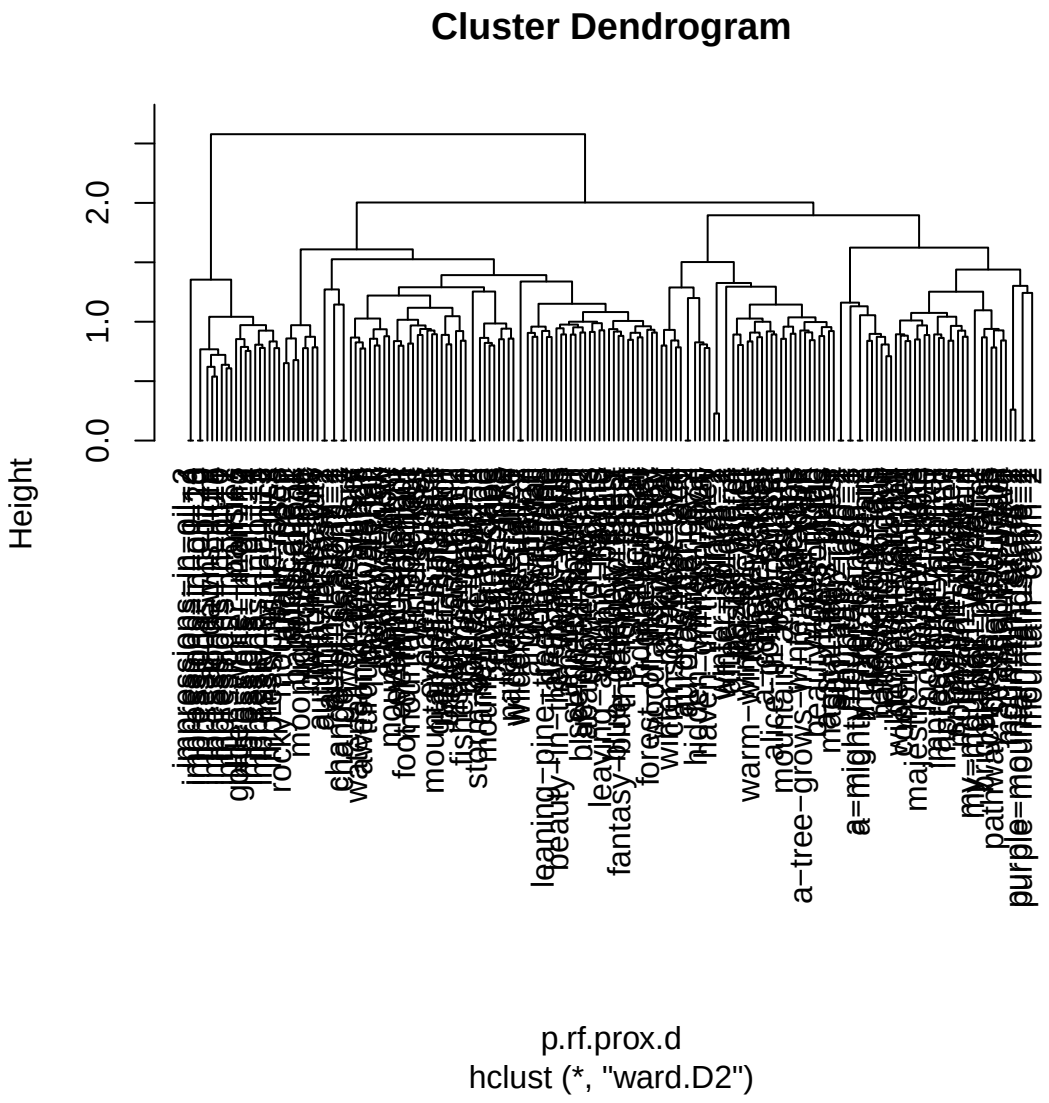
Read in the proximity matrix. Check the dimensions.

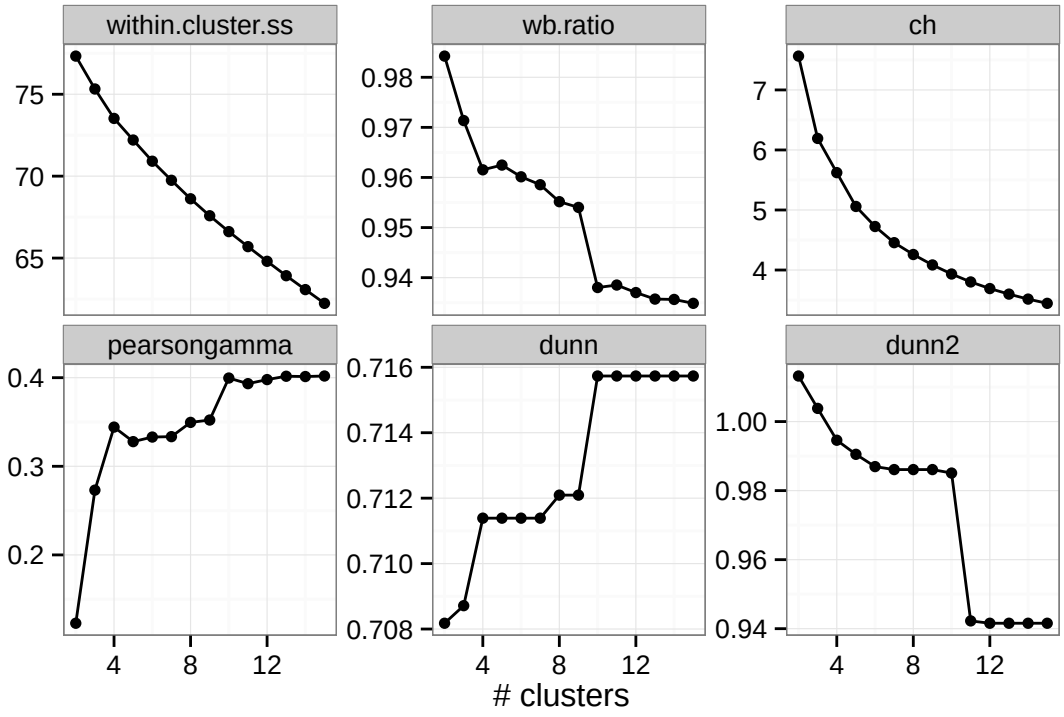
```
## [1] 178 178
```



Task 4

Run hierarchical clustering using `ward.D2` linkage, and examine results for solutions ranging from k=2 to 15. Compute the statistics to determine the ideal number of clusters.





Task 5

Explore one of the cluster solutions that is suggested by the stats, and one that seems reasonable to you based on your knowledge of the paintings. How well does the solution match the class labels provided with the data?

```
##
##                1  2  3  4  5  6  7  8
##   cold         2  4  0  2  4 11  0  0
##   dusk         4  2  1  1 14  5  0  3
##   flowers      0  3  0  2 10  1  6  0
##   impressions  0  2  0 15  0  0  0  0
##   oval         4  0  0  0  0  1  0  0
##   scene        1 13  2  0  8  2  0  2
##   trees        0  0  0  0 16  2  0  0
##   water        0  6  9  0 13  4  0  3
```