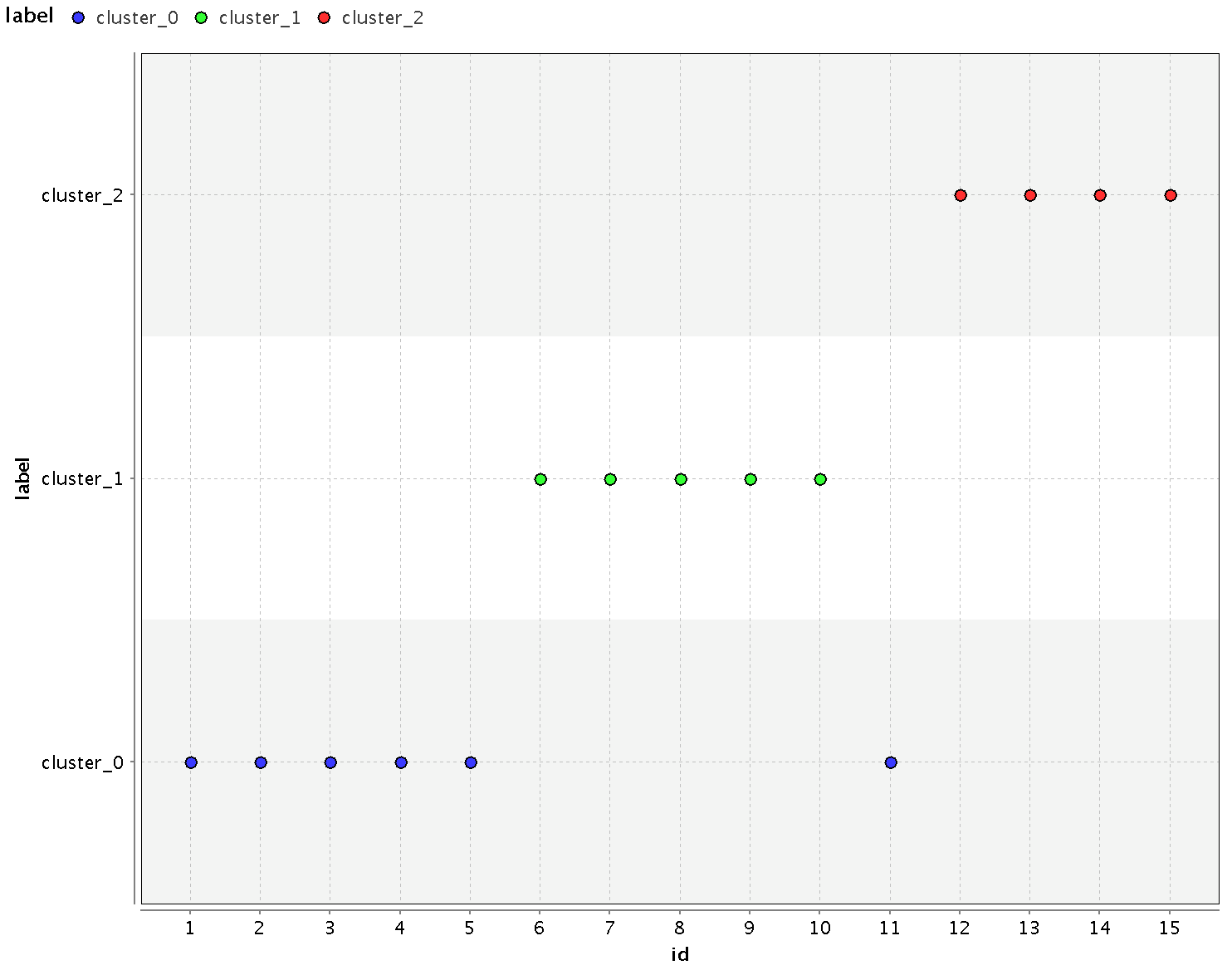
Lab 4 Clustering

B00060572 David Kelly

1.1 Did it identify the three categories of documents? (view scatter plot of cluster V label)

a) k-means correctly identified 14 of the 15 documents



1.2 For documents assigned to the wrong cluster, hover over their point on the scatter plot to get the row ID. Looking back at the document vector, can you see why these documents were allocated to the wrong cluster?

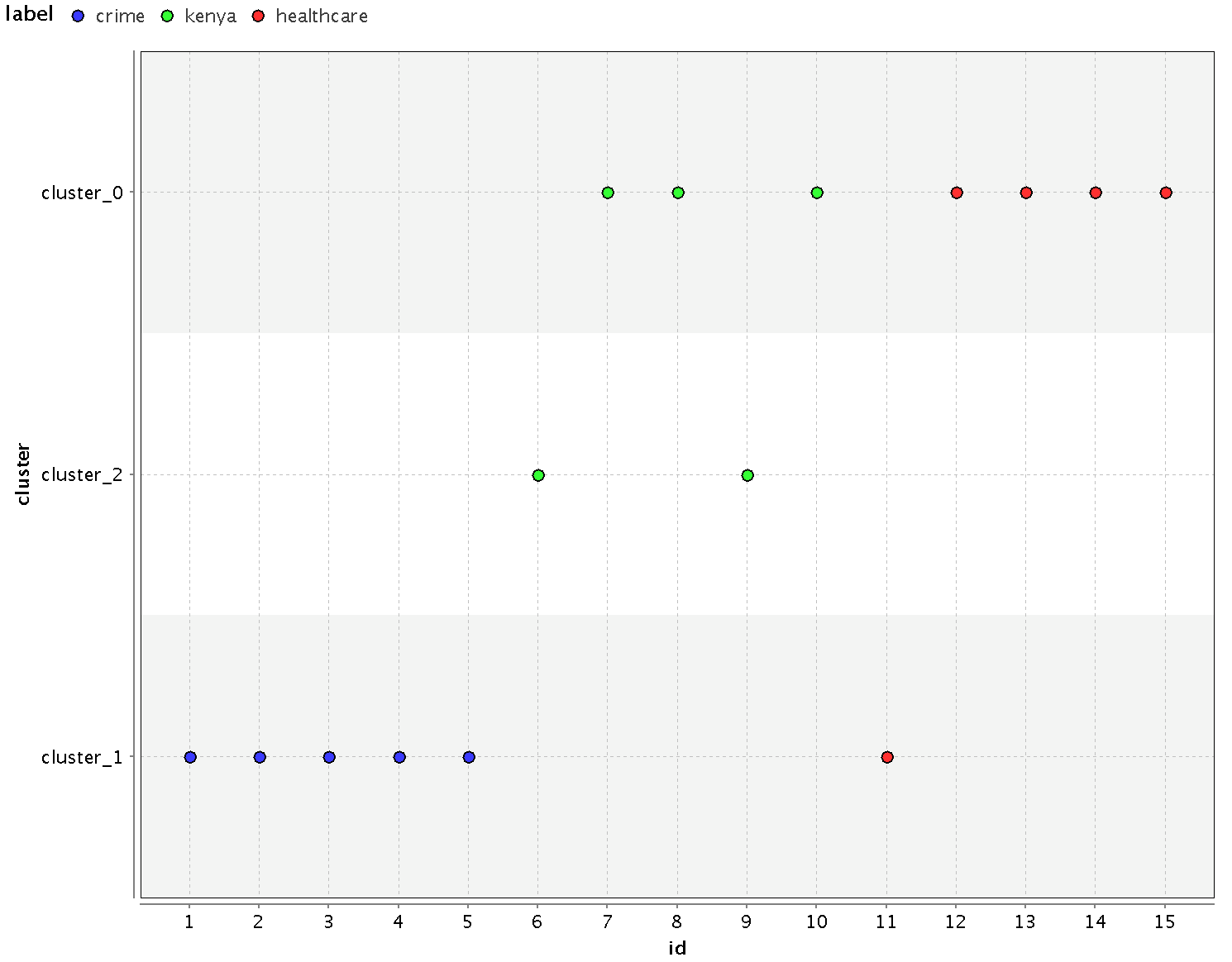
a) Yes, looking back at the document vector I can see that from the word list generated in preprocessing I have no terms that match in the incorrectly clustered document (H1.txt)

2.1 Did agglomerative clustering do better or worse the k-means?

a) Using measure as numeric and changing numerical measure to cosine had no affect on on clustering the document. The same document was incorrectly clustered.

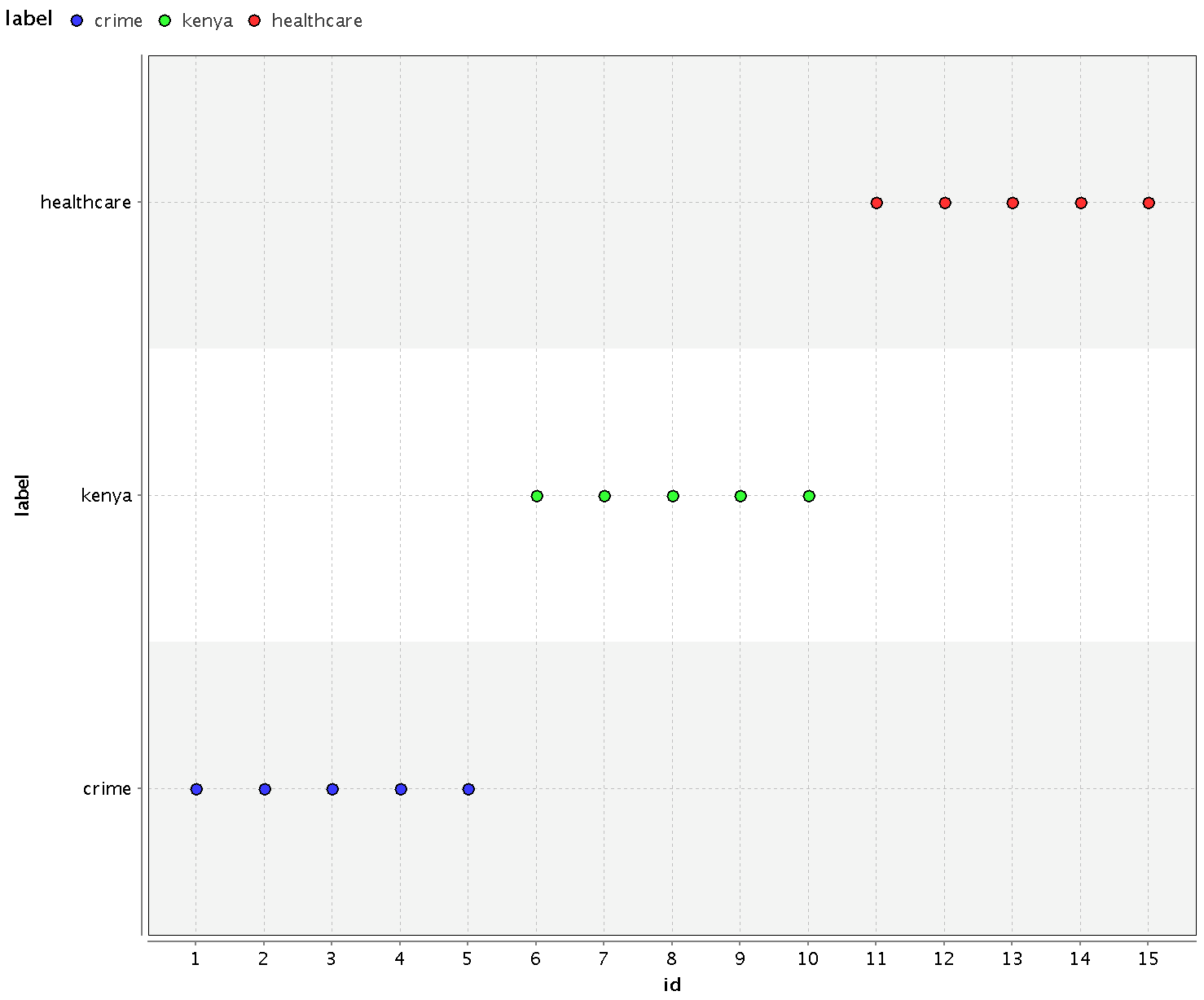
2.2 Change mode to complete linkage. Does that improve the performance?

a) Complete linkage decreased agglomerative clustering’s ability correctly cluster documents



3.1 How did the performance of SOM compare with agglomerative clustering with complete linkage?

a) SOM worked better than k-means and agglomerative clustering. SOM was able to correctly label each document



4.1 What value of K gives the smallest DB value?

a) 2

4.2 Looking at a plot of the iris dataset, and ignoring the class label, does two clusters make sense?

a) Yes, The Iris-setosa a clearly separated from both other labels which are intermingled with each other

4.3 According to a DB evaluation, how many clusters are in the dataset?

a) 9

