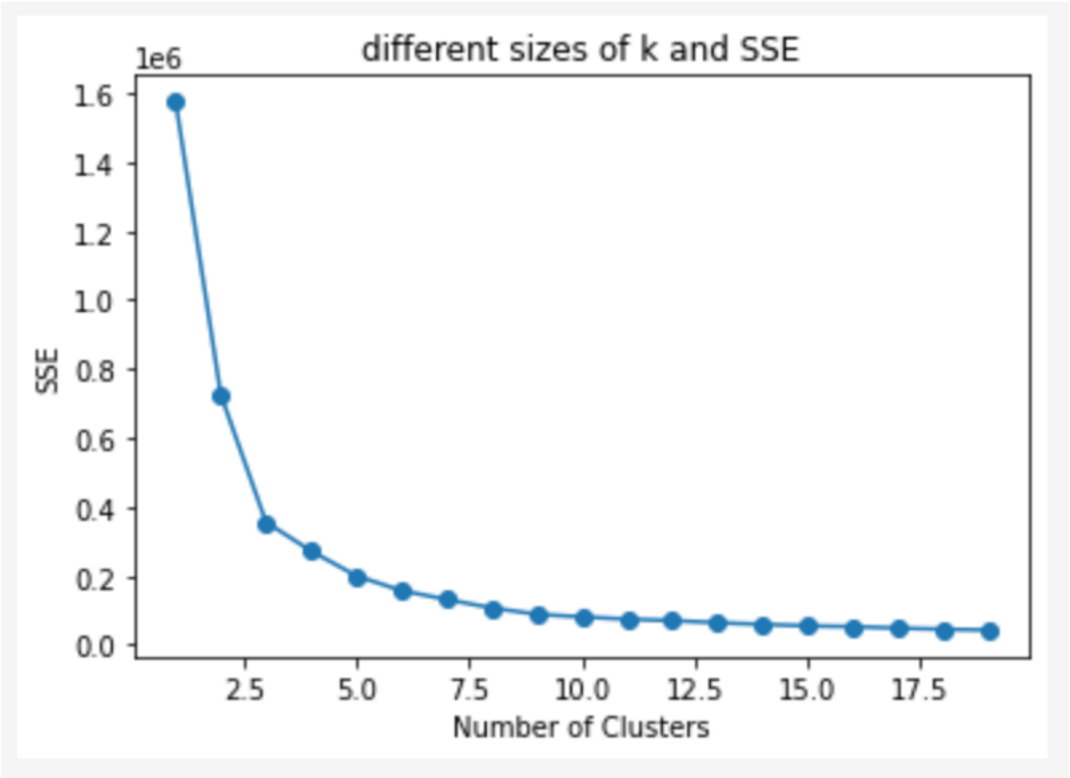
1)

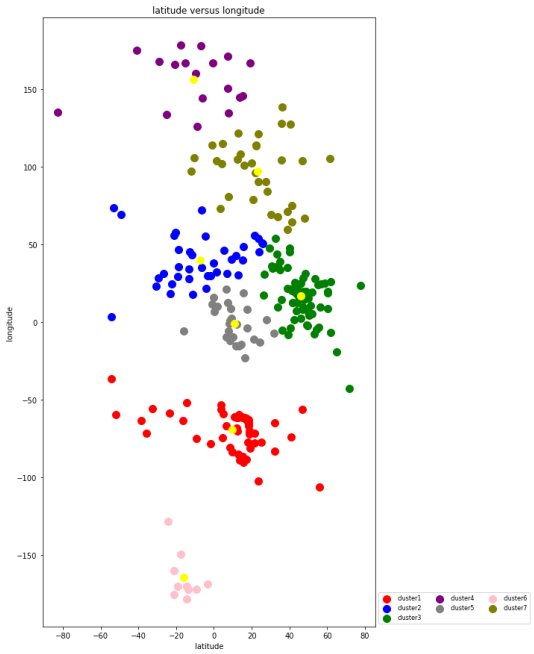
a. Experiment with some different sizes of k and observe the range of the Sum of Squares Error (SSE)



From the graph, we know that SEE drop sharply before 7 clusters, so 7 is the best cluster

b.

```
corrdinates of cluster centers:
[[   9.62003573  -69.14907131]
 [  -7.29818399   39.86657559]
 [  46.09751427   16.72002277]
 [ -10.81353477  156.11725861]
 [  10.75386428   -1.10167986]
 [ -15.84210807 -164.35100116]
 [  23.31902544   97.03025315]]
Sum of squared distances to centroids: 132052.61386621272
Number of iterations run: 8
```



I think this is the best clustering because the data points are equally divided into 7 clusters and SSE is very low

c.

Cluster	Centroid	Continent
0	9.62003573 -69.14907131	North America
1	-7.29818399 39.86657559	Africa
2	46.09751427 16.72002277	Europe
3	-10.81353477 156.11725861	Oceania
4	10.75386428 -1.10167986	Africa
5	-15.84210807 -164.35100116	Oceania
6	23.31902544 97.03025315	Asia

2)

a

we compare four different method

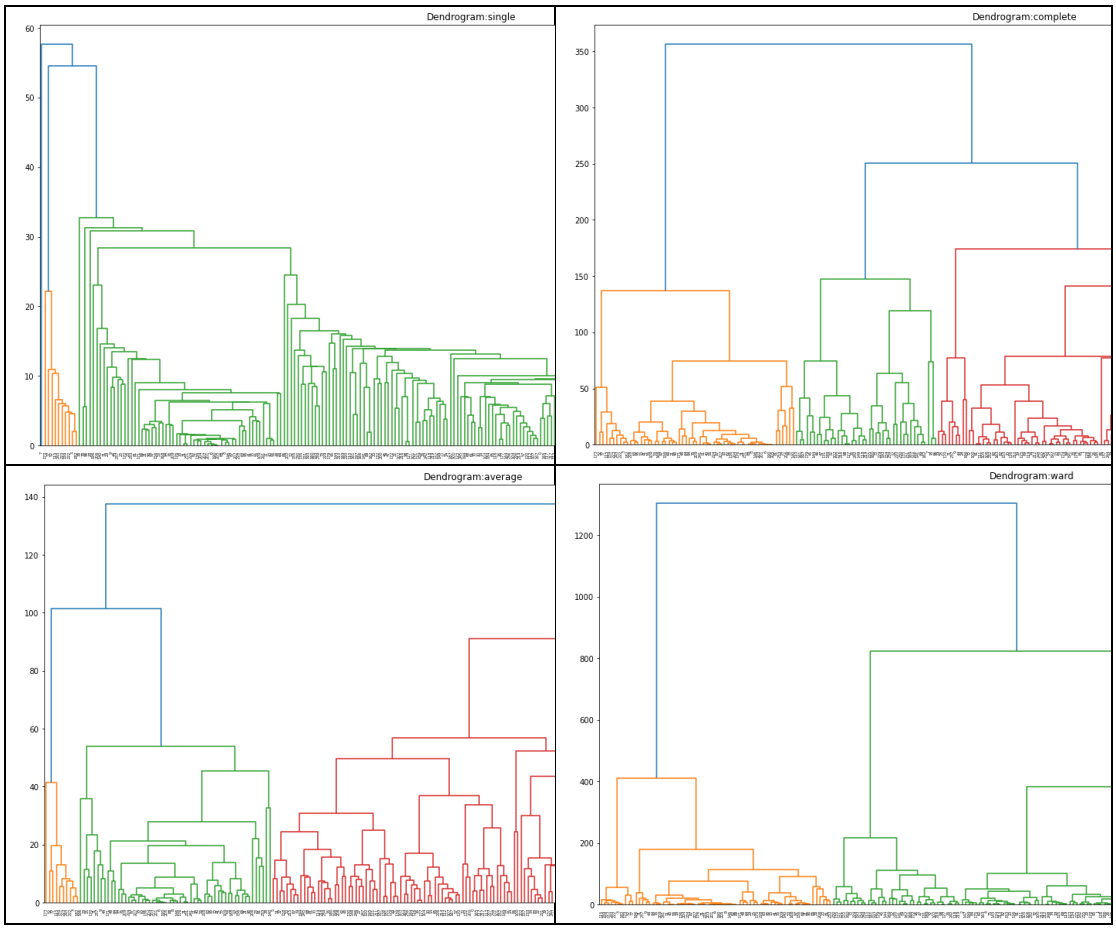
'single' uses the minimum of the distances between all observations of the two sets.

'complete' or 'maximum' linkage uses the maximum distances between all observations of the two sets.

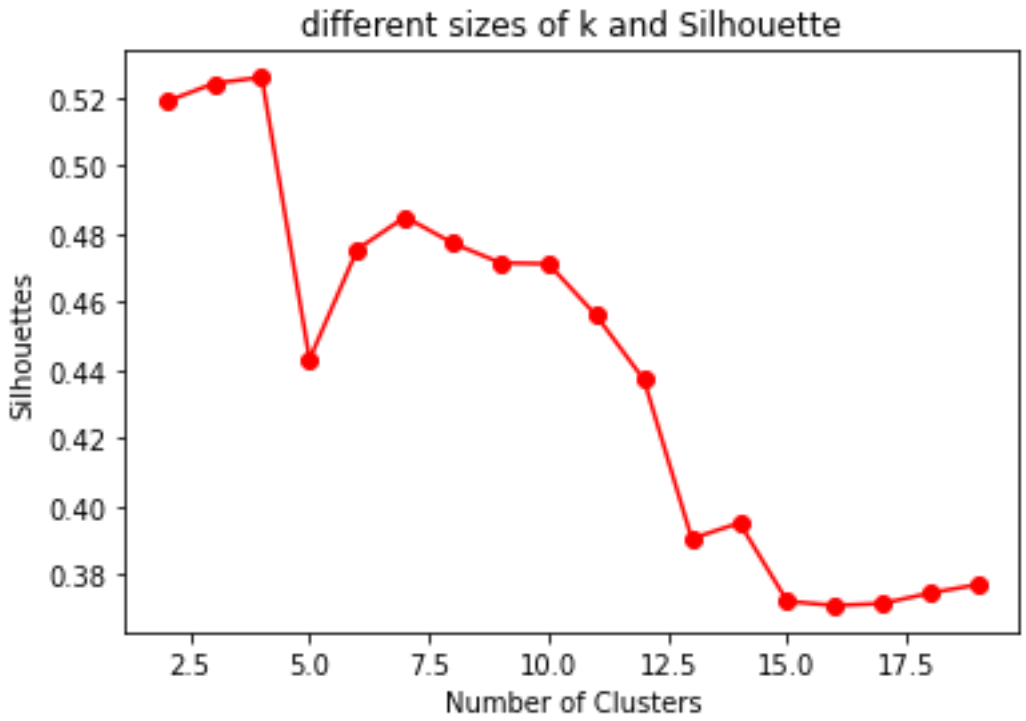
'average' uses the average of the distances of each observation of the two sets.

'ward' minimizes the variance of the clusters being merged.

dendrogram



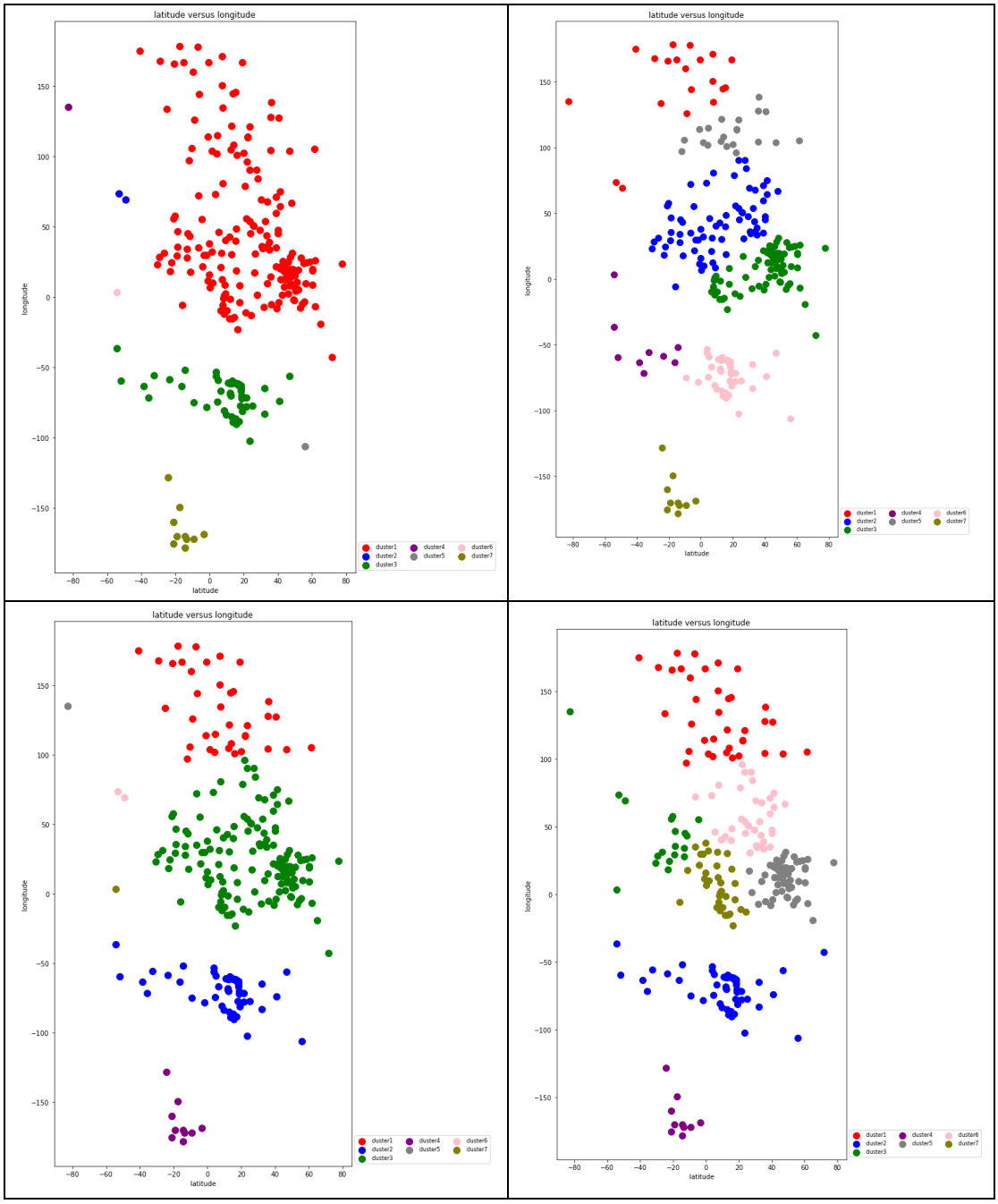
From the graph, we know that `linkage = 'ward'` works better. we use silhouette coefficient to help us experiment with different number of clusters



The higher silhouette coefficient the better.

1. From 2 to 4, 4 clusters has the best silhouette coefficient, but it is has higher Sum of Squares Error if we consider k-Means algorithm
2. From 5 to 19, **7 clusters** has the best silhouette coefficient, so 7 cluster is the best.

We use scatter plot to help visualize the clusters



From the graph we know that the best parameters: n_clusters = 7, affinity = 'euclidean', linkage = 'ward'

b

Cluster	Continent
0	Asia
1	North America
2	Africa
3	Oceania
4	Europe
5	Asia

6	Africa

3)

AgglomerativeClustering is the most accurate grouping of countries into the correct continent

a

Cluster	Continent
0	Asia
1	North America
2	Africa
3	Oceania
4	Europe
5	Asia
6	Africa