

Lab-3

Q-1 k-mean, k-median, expectation maximization, hierarchical algorithm with R.

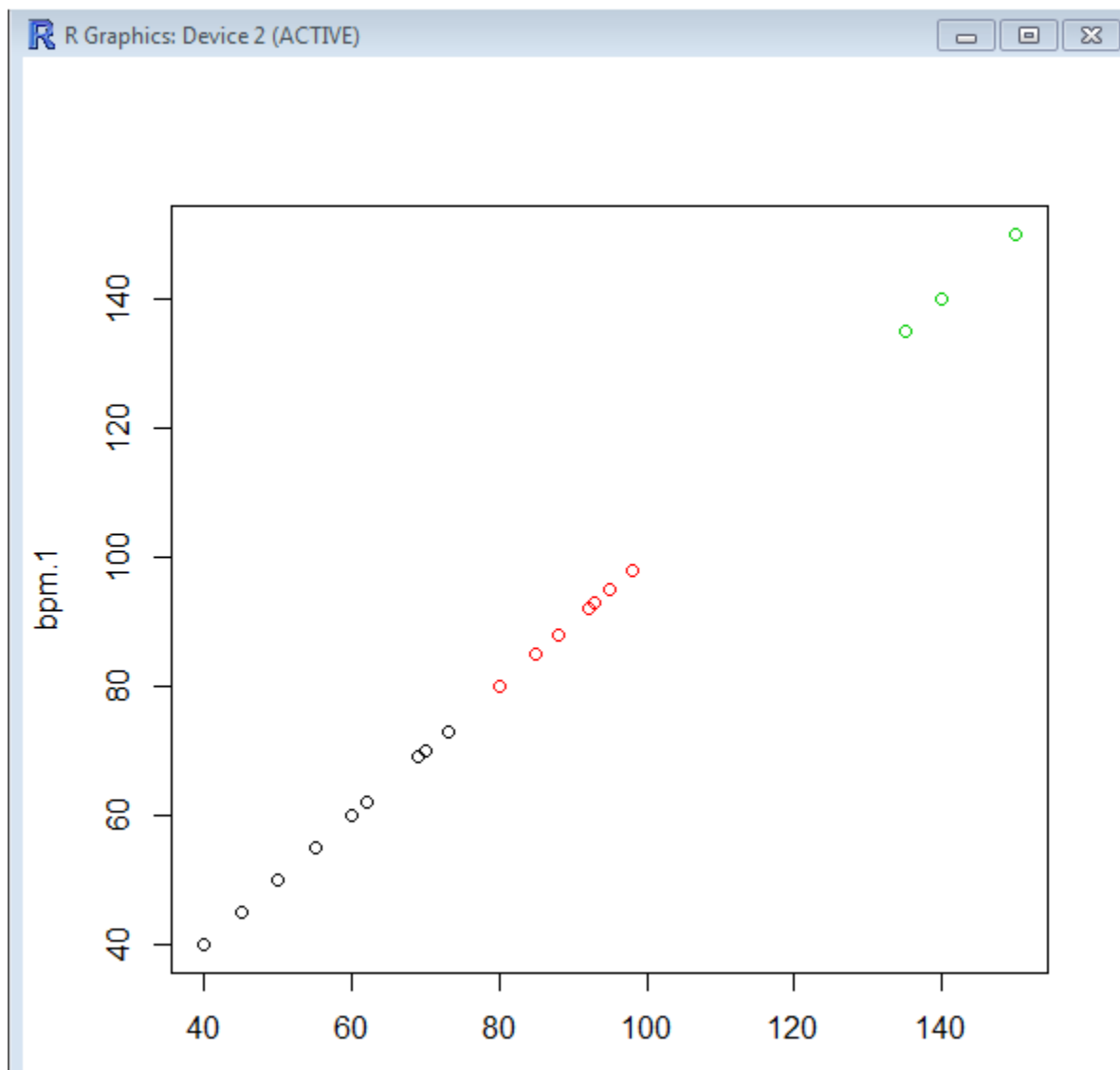
K-means :- We are collecting the heart bit rate from the wear, so we are using k-means algorithm to find the cluster of BPM with our set of collected sample data. Below are sample data file, source code and output in R for the algorithm.



Source_Kmeans.txt



k-mean-bpm_data.txt

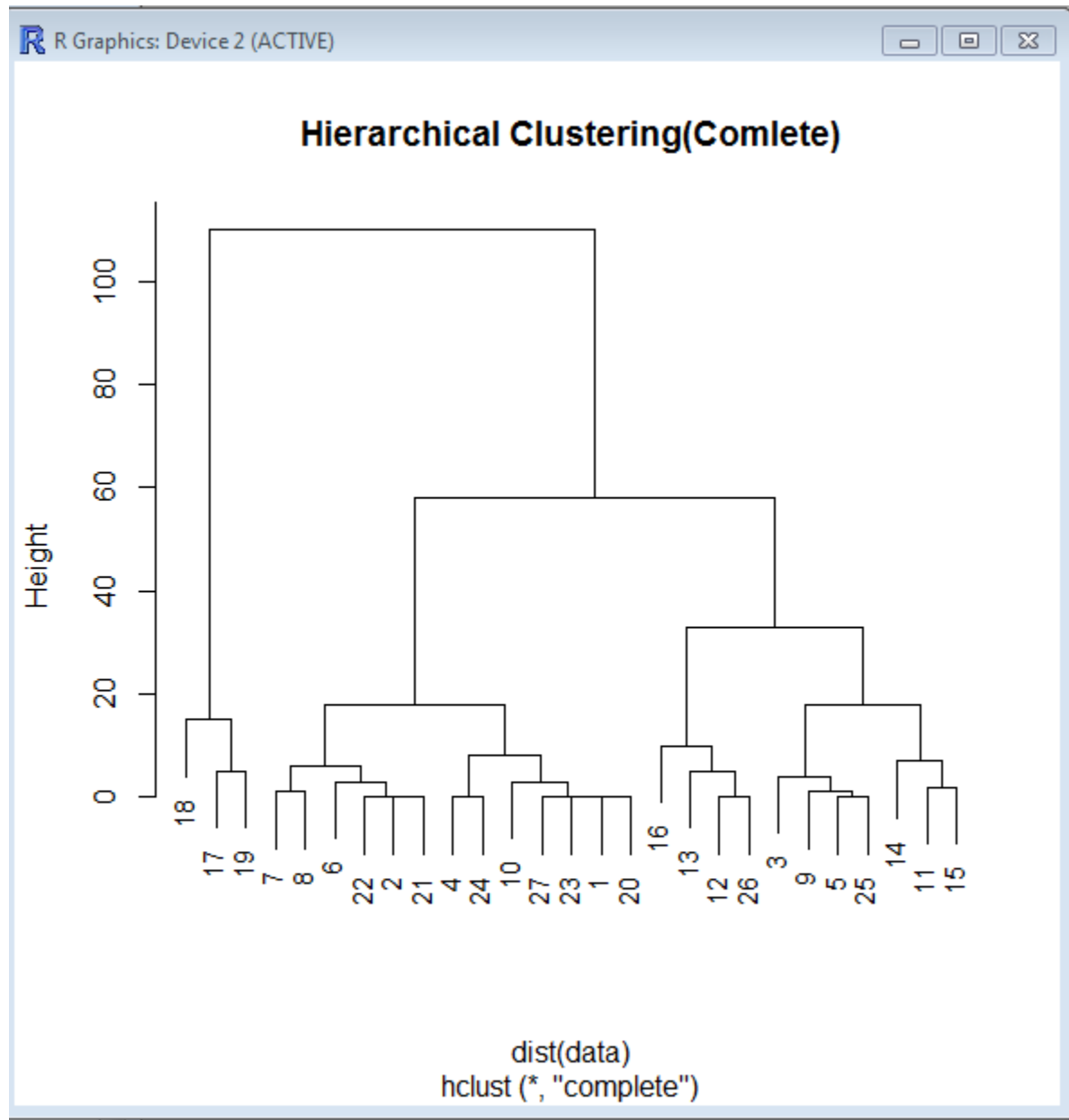


Hierarchical Clustering: Hierarchical clustering is clustering algorithm for combining data into one cluster or dividing them up to individual cluster. It doesn't have cluster known on prior basis. We don't define cluster like k-means here. The output of this is Dendrogram.

With the same above data set we performed hierarchical clustering and below are the source and output of it.



Source_Heirarchical.txt

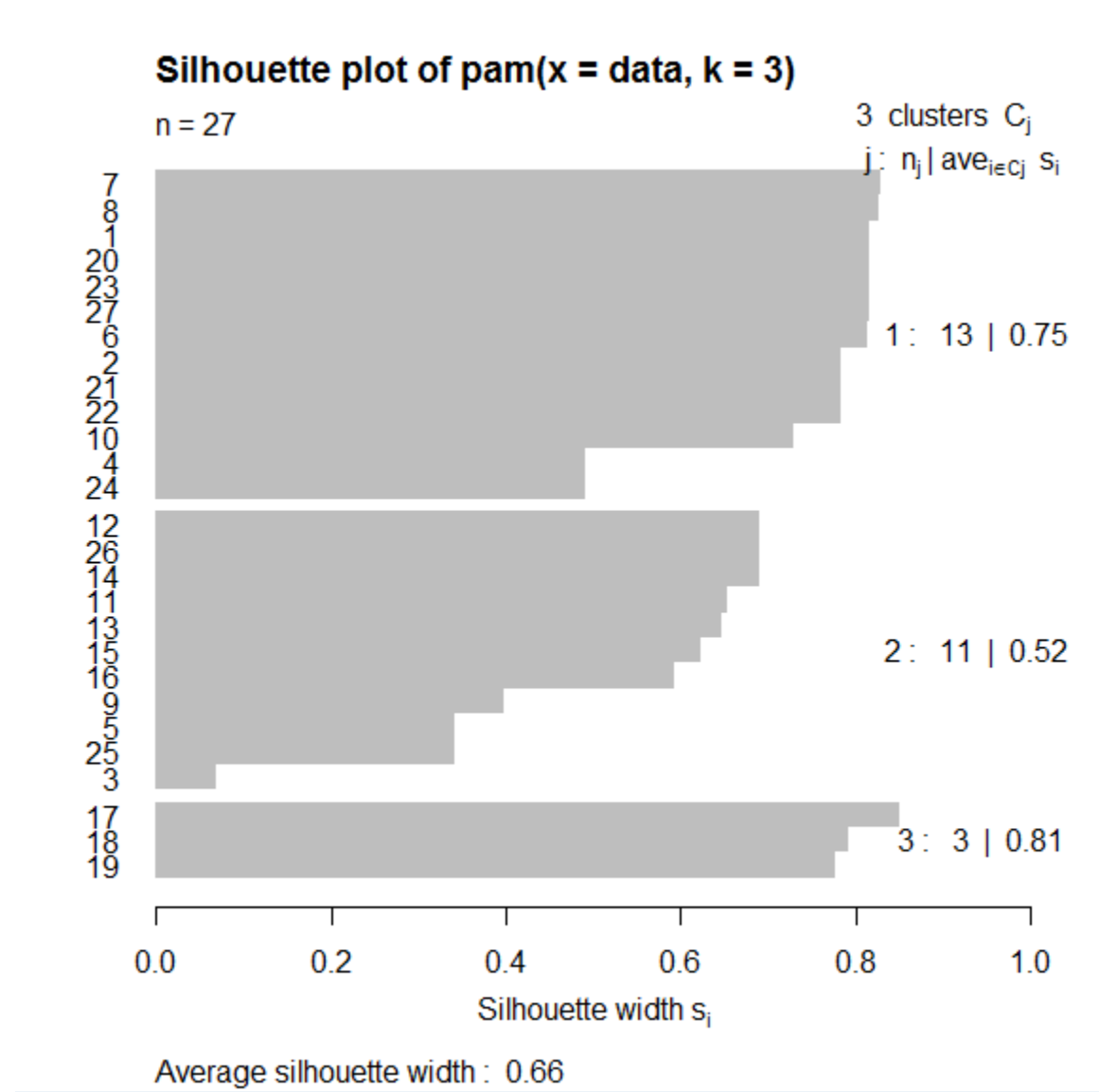


K-Medoids:- Like k-means clustering here also we need to define number of clusters. Medoids are similar in concept to means or centroids, but medoids are always members of the data set. Medoids are most commonly used on data when a mean or centroid cannot be defined such as 3-D trajectories or in the gene expression context.

Below are the source and output for 3 cluster and interpretation right below it.



Source_Medoids.txt



The plot indicates that there is a good structure to the clusters, with most observations seeming to belong to the cluster that they're in. There is a summary measure at the bottom of the plot labeled "Average Silhouette Width". This table shows how to use the value:

Range of SC	Interpretation
0.71-1.0	A strong structure has been found
0.51-0.70	A reasonable structure has been found
0.26-0.50	The structure is weak and could be artificial
< 0.25	No substantial structure has been found

One novel feature of pam is that it finds observations from the original data that are typical of each cluster in the sense that they are closest to the center of the cluster. The indexes of the **medoids** are stored in the id.med component of the pam object, so we can use that component as a subscript into the vector of data BPM to see which ones were selected:

```
> data$bpm[data.pam$id.med]
```

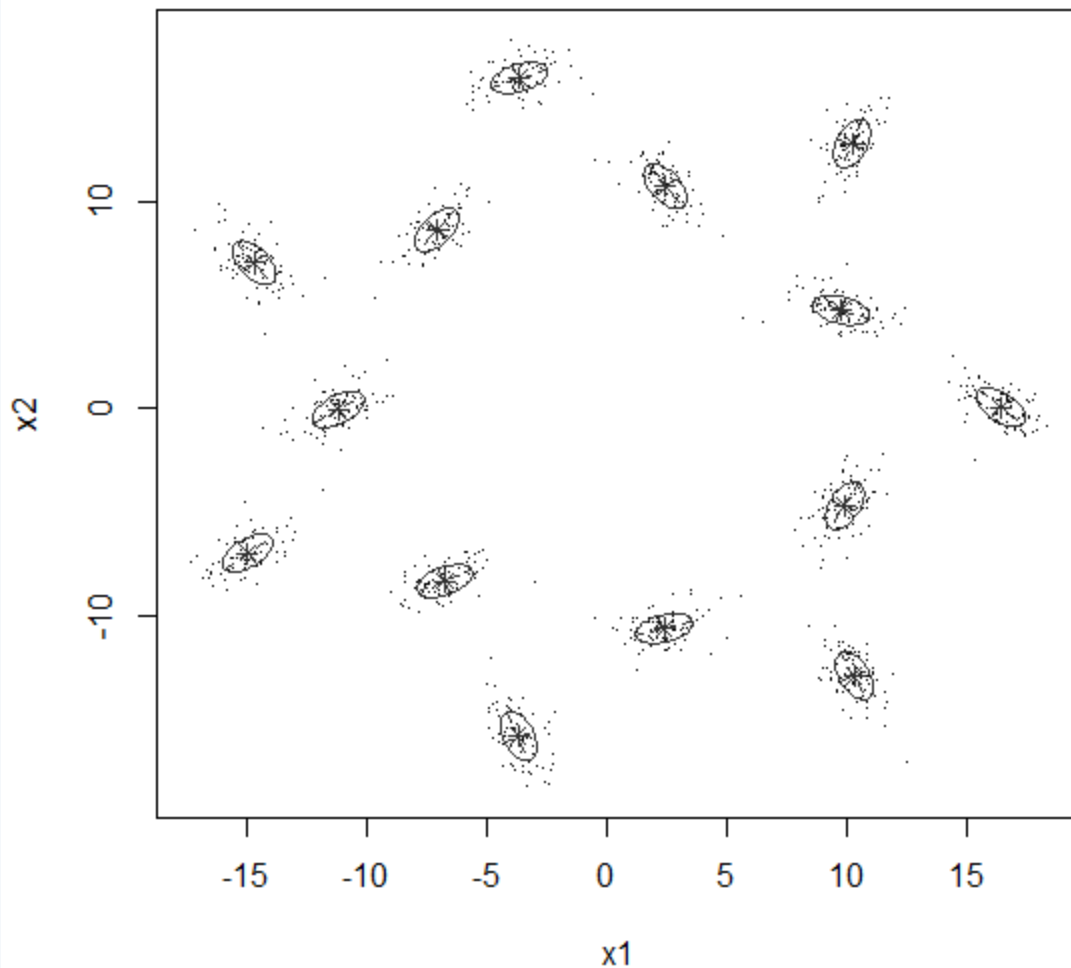
```
[1] 88 60 140
```

Expectation Maximization: In statistics, an expectation–maximization (EM) algorithm is an iterative method for finding maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables.

We generated algorithm with R in-build wreath data set and below are the source code and outcome.



EM.R



We can see that the mixture models created to represent the point are in conformation to the data set. On this case, the groups don't have an intersection between them, so all points were classified to the right group. The cluster orientation allows the method to find a better Ellipsoid to represent those points.

Q-2 Snap shot for data collection of project wear app.



This is the result of heart bit and steps collected from the wear device. This is just static collection and we will make this run on timely basis to collect periodic data for our project.