**PRACTICAL – 10**

**AIM:**

Implementation and analysis of Clustering Algorithms Like K Means Agglomerative.

**THEORY:**

**K-Means Clustering Algorithm**:

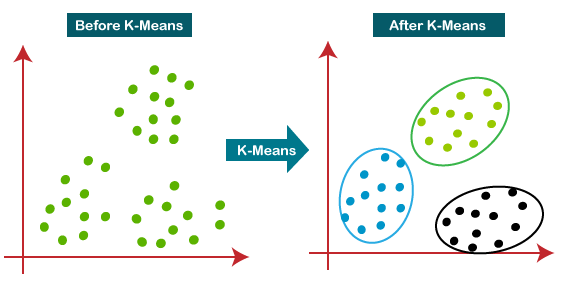
K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science.

**What is K-Means Algorithm?**

* K-Means Clustering is an [Unsupervised Learning algorithm](https://www.javatpoint.com/unsupervised-machine-learning), which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.
* It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.
* It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.
* It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.
* The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.
* The k-means [clustering](https://www.javatpoint.com/clustering-in-machine-learning) algorithm mainly performs two tasks:

1. Determines the best value for K center points or centroids by an iterative process.
2. Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

* Hence each cluster has datapoints with some commonalities, and it is away from other clusters.
* The below diagram explains the working of the K-means Clustering Algorithm:



**How does the K-Means Algorithm Work?**

The working of the K-Means algorithm is explained in the below steps:

**Step-1:** Select the number K to decide the number of clusters.

**Step-2:** Select random K points or centroids. (It can be other from the input dataset).

**Step-3:** Assign each data point to their closest centroid, which will form the predefined K clusters.

**Step-4:** Calculate the variance and place a new centroid of each cluster.

**Step-5:** Repeat the third steps, which means reassign each datapoint to the new closest centroid of

each cluster.

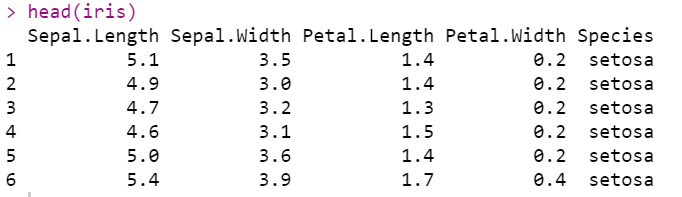
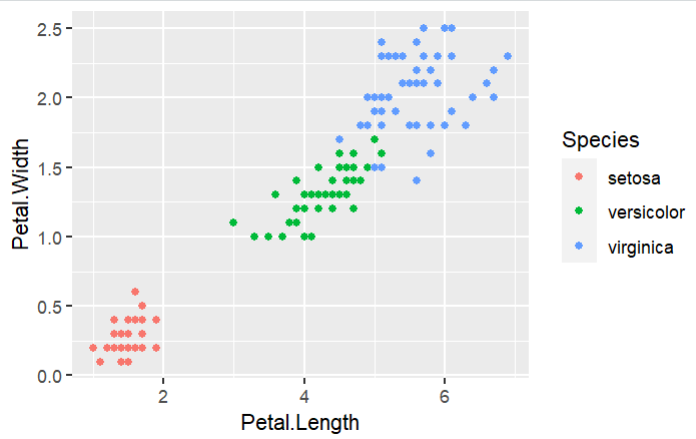
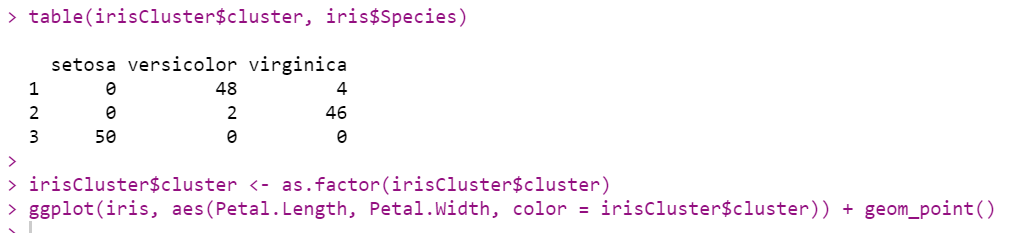
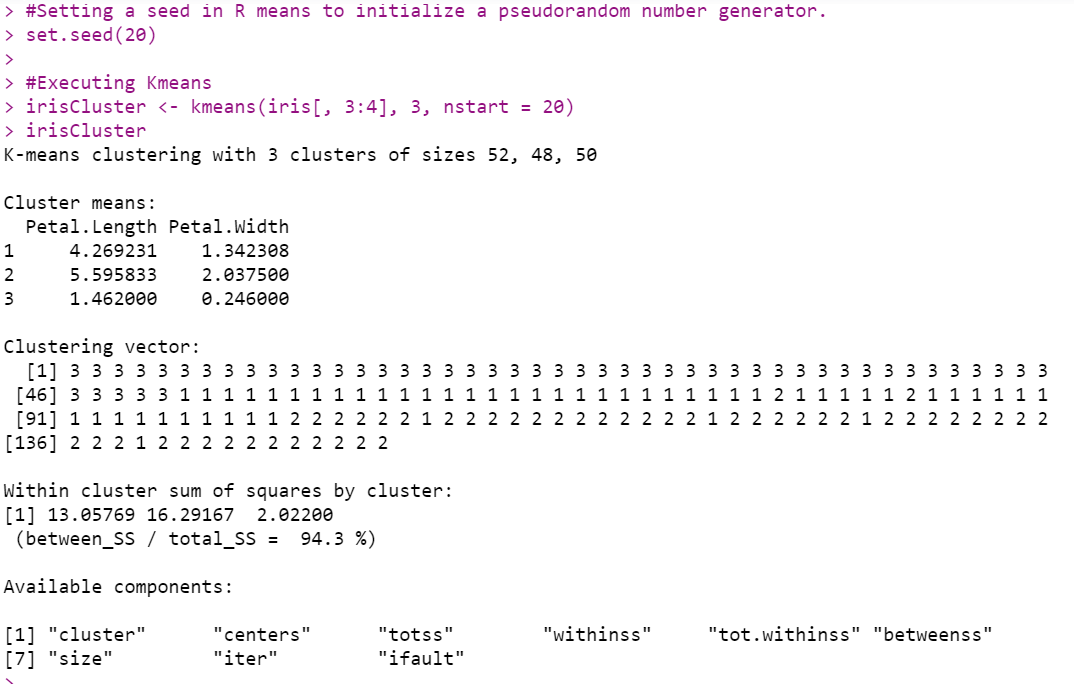
**Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.

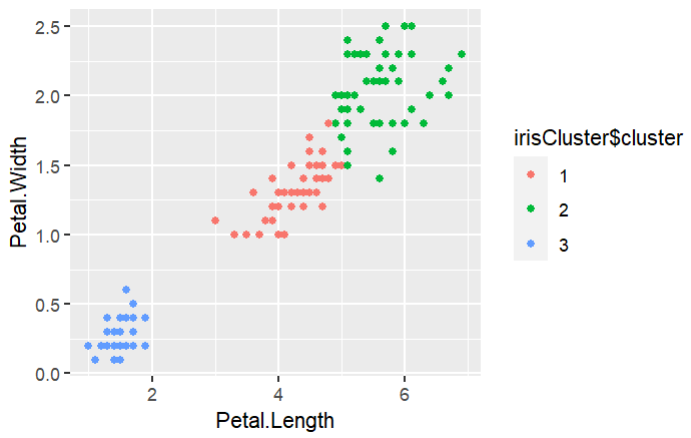
**Step-7**: The model is ready.

**FUNCTIONS USED IN R:**

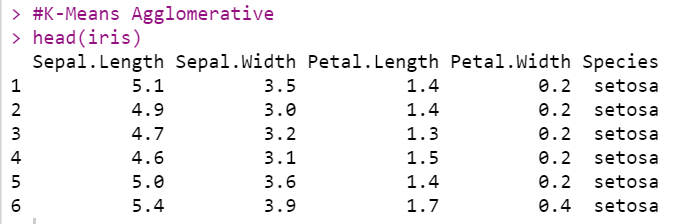
* hclust: Performs hierarchical clustering on a distance or similarity structure.
* cutree: Returns a vector of group numbers for the observations that were clustered. Specify either the number of groups desired or a clustering height.

**CODE & OUTPUT:**

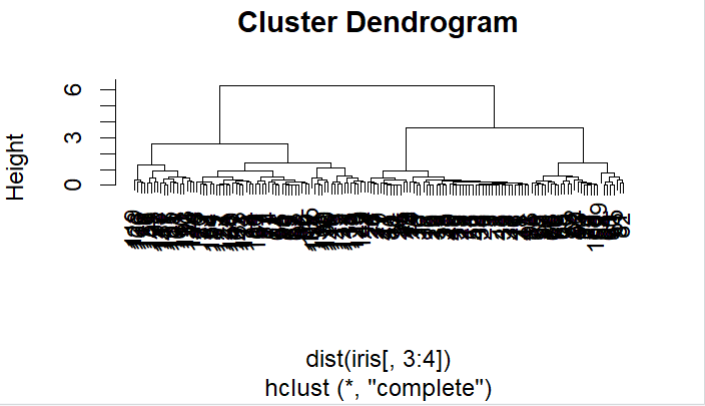
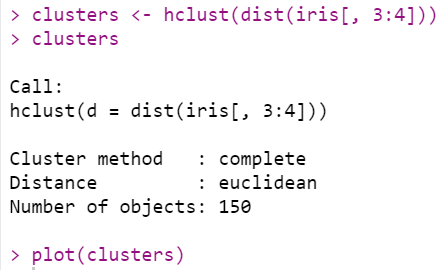
1. **IMPLEMENT K-MEANS CLUSTERING:-**
2. **LOADING DATASET:**
3. **LOADING LIBRARY FOR PLOTING GRAPH:**
4. **EXECUTING K-MENAS AND PLOTTING GRAPH:**

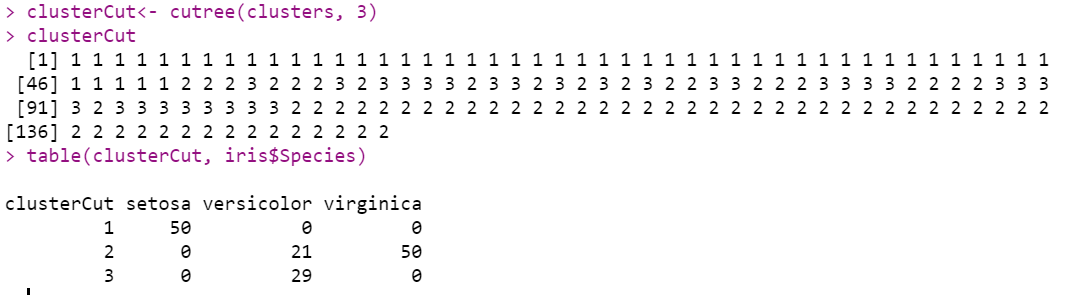
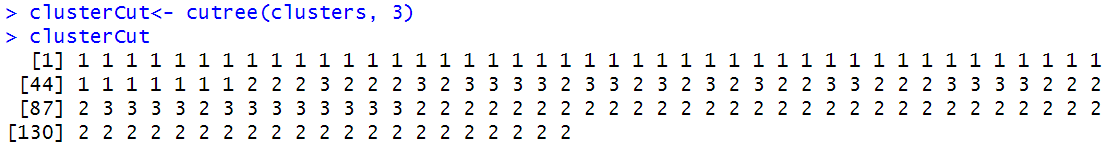
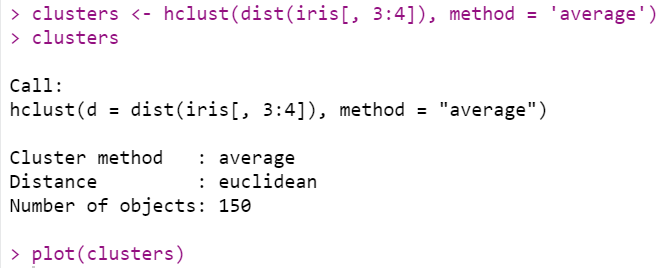
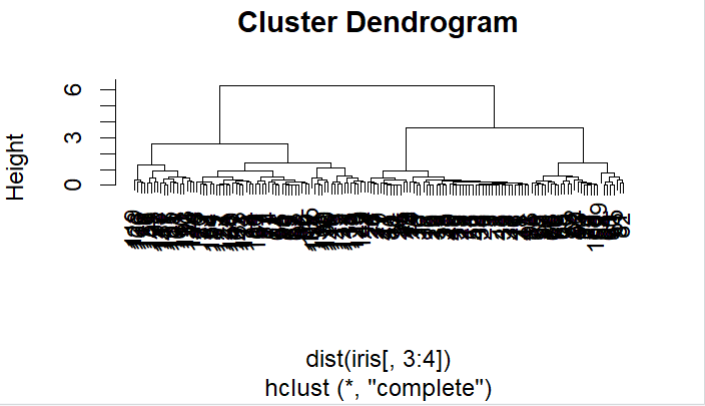
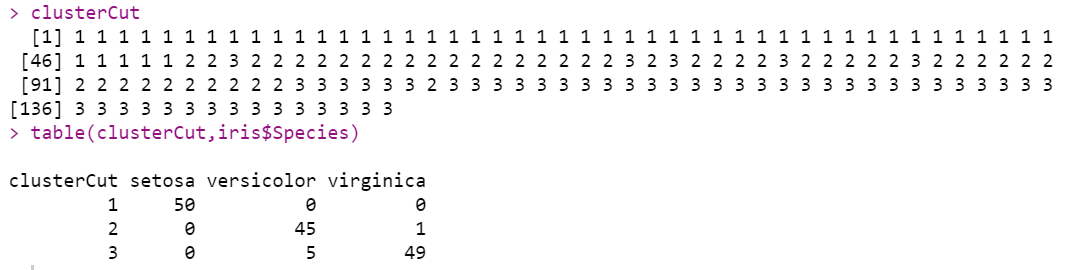


1. **K-MEANS AGGOLOMERATIVE:-**
2. **LOADING DATASET:**

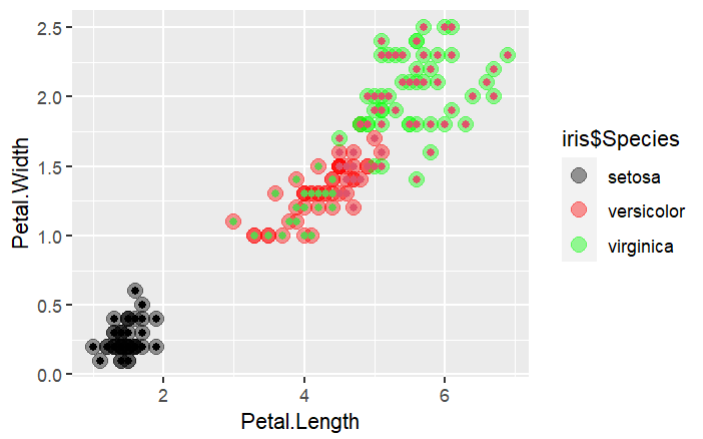


1. **CREATING CLUSTERS:**



1. **CREATING GROUP OF CLUSTERS:**
2. **CREATING AVERAGE LINKAGE CLUSTERS:**
3. **CREATING GROUP CLUSTER FOR AVERGE LINKAGE:**
4. **PLOTTING CLUSTERD GRAPH:**

ggplot(iris, aes(Petal.Length, Petal.Width, color = iris$Species)) + geom\_point(alpha = 0.4, size = 3.5) + geom\_point(col = clusterCut) + scale\_color\_manual(values = c('black', 'red', 'green'))



**CONCLUSION:**

In this practical I learnt the implementation of Clustering Algorithms Like K Means Agglomerative.