

**Ex.No.7****Hierarchical Clustering Algorithm****Date:** 12-09-23**Aim**

To implement Hierarchical Clustering Algorithm under unsupervised machine learning through R programming.

**Procedure**

1. To do programming in R, first install “RStudio” and “R” in the system. RStudio is an integrated development environment [IDE] for R and python.
2. Select the File in taskbar → open New file → R script or use shortcut “ctrl+shift+N”
3. Install the ‘dplyr’ package and load it in R.
4. Import the built-in dataset ‘mtcars’
5. Apply the Hierarchical Clustering Algorithm on ‘mtcars’ dataset.
6. Write the program in the script and save it using the extension R.
7. Run the program by clicking Run option or use the shortcut “ctrl+enter”.
8. See the output in the console tab.

**Concepts Involved**

- Applying the Hierarchical Clustering Algorithm on a Dataset.

**Hierarchical Clustering Algorithm**

Clustering is the most common form of unsupervised learning, a type of machine learning algorithm used to draw inferences from unlabeled data.

**R - Hierarchical Clustering**

This is of two types:

- **Agglomerative Hierarchical clustering:** It starts at individual leaves and successfully merges clusters together. It is a Bottom-up approach.
- **Divisive Hierarchical clustering:** It starts at the root and recursively split the clusters. It's a top-down approach.

**Algorithm**

1. Make each data point in a single point cluster that forms **N** clusters.
2. Take the two closest data points and make them one cluster that forms **N- 1** clusters.
3. Take the two closest clusters and make them one cluster that forms **N- 2** clusters.
4. Repeat steps 3 until there is only one cluster.

Dendrogram is a hierarchy of clusters in which distances are converted into heights. It clusters **n** units or objects each with **p** feature into smaller groups. Units in the same cluster are joined by a horizontal line.

The leaves at the bottom represent individual units. It provides a visual representation of clusters.

### Script

```
# Installing the package
install.packages("dplyr")

# Loading package
library(dplyr)

# Summary of dataset in package
head(mtcars)
```

### Output

	mpg	cyl	dis	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

### Script

```
# Finding distance matrix

distance_mat <- dist(mtcars, method = 'euclidean')

distance_mat#representing some part for ease of view
```

### Output

Mazda RX4	Mazda RX4 Wag	Datsun 710	Hornet 4 Drive	Hornet Sportabout
Mazda RX4 Wag	0.6153251			
Datsun 710	54.9086059	54.8915169		
Hornet 4 Drive	98.1125212	98.0958939	150.9935191	
Hornet Sportabout	210.3374396	210.3358546	265.0831615	121.0297564
valiant	65.4717710	65.4392224	117.7547018	33.5508692
152.1241352				

### Script

```
# Fitting Hierarchical clustering Model

# to training dataset

set.seed(240) # Setting seed

Hierar_cl <- hclust(distance_mat, method = "average")

Hierar_cl
```

**Output**

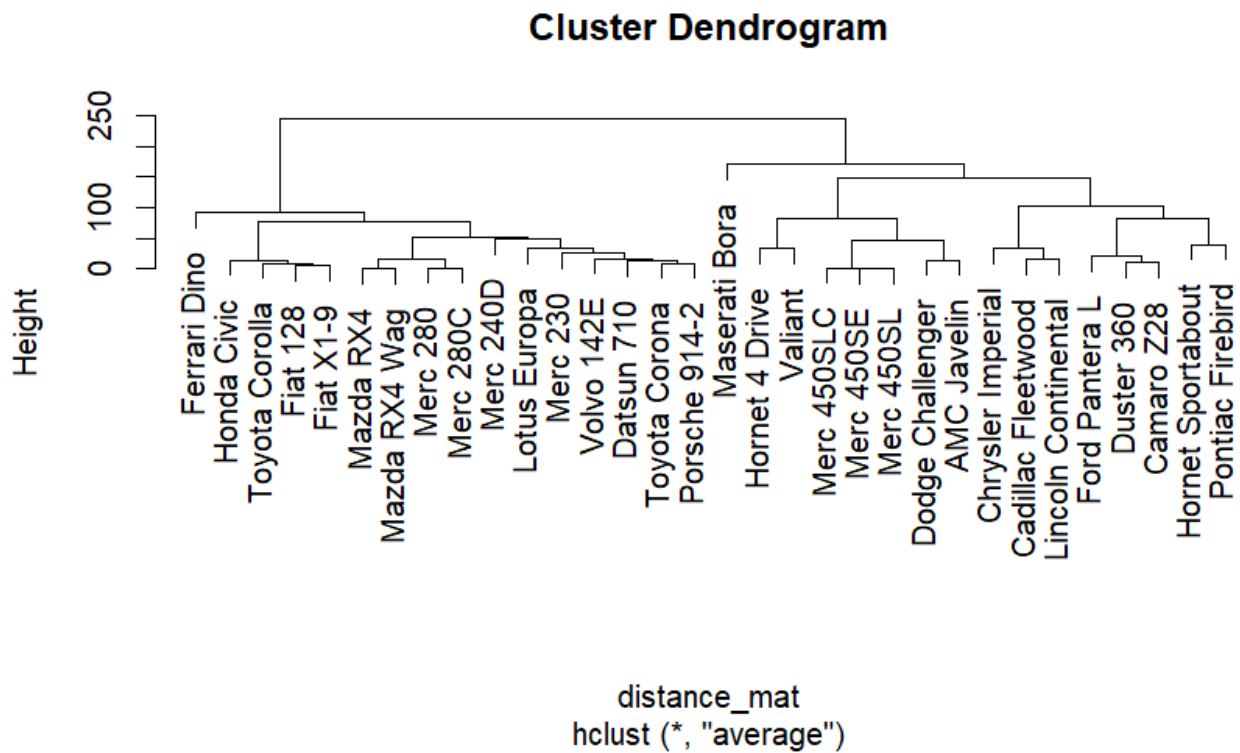
```
call:
hclust(d = distance_mat, method = "average")
```

```
Cluster method   : average
Distance         : euclidean
Number of objects: 32
```

**Script**

```
# Plotting dendrogram
```

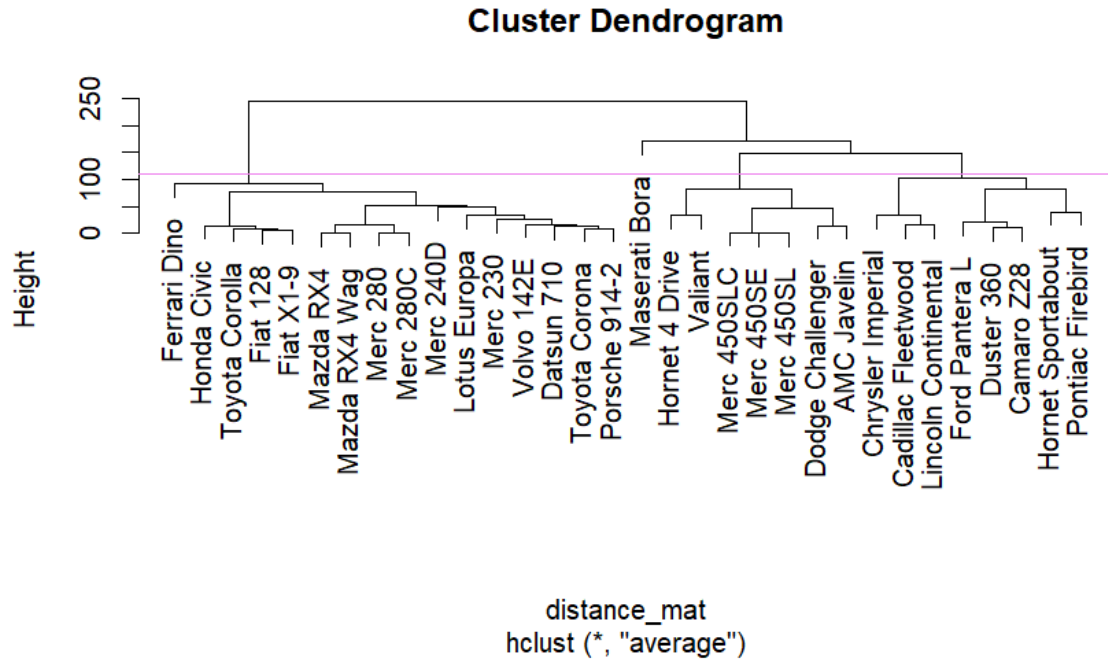
```
plot(Hierar_cl)
```

**Output****Script**

```
# Choosing no. of clusters
```

```
# Cutting tree by height
```

```
abline(h = 110, col = "violet")
```

**Output****Script**

```
# Cutting tree by no. of clusters
```

```
fit <- cutree(Hierar_cl, k = 3 )
```

```
fit
```

**Output**

Mazda RX4	Mazda RX4 Wag	Datsun 710	Hornet 4 Drive
	1	1	1
2			
Hornet Sportabout		valiant	Duster 360
OD			Merc 24
	2	2	2
1	Merc 230	Merc 280	Merc 280C
SE			Merc 450
	1	1	1
2	Merc 450SL	Merc 450SLC	Cadillac Fleetwood
al			Lincoln Continent
	2	2	2
2	Chrysler Imperial	Fiat 128	Honda Civic
la			Toyota Corol
	2	1	1
1	Toyota Corona	Dodge Challenger	AMC Javelin
28			Camaro Z

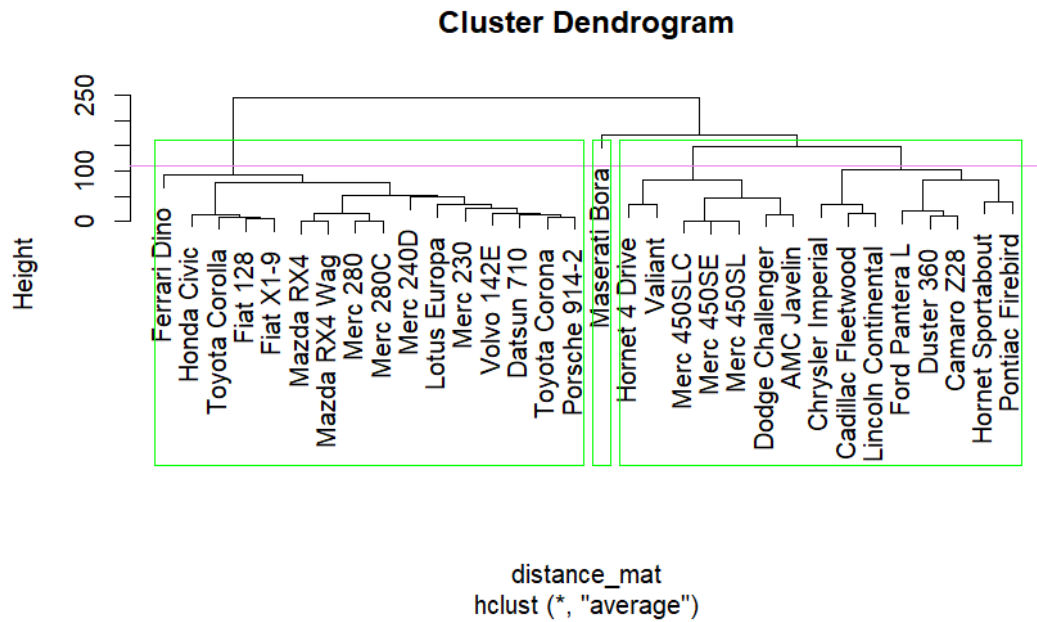
	1	2	2	
2	Pontiac Firebird	Fiat X1-9	Porsche 914-2	Lotus Euro
pa				
1	2	1	1	
2E	Ford Pantera L	Ferrari Dino	Maserati Bora	Volvo 14
1	2	1	3	

## Script

```
table(fit)
```

```
rect.hclust(Hierar_cl, k = 3, border = "green")
```

## Output



## Result

Thus, the cluster dendrogram is plotted using Hierarchical Clustering Algorithm using R programming successfully.