EXPERIMENT NO:10

AIM: Implement R program on Hierarchical Clustering on mtcars Dataset.

DESCRIPTION:

Hierarchical clustering in R Programming Language is an Unsupervised non-linear algorithm in which clusters are created such that they have a hierarchy(or a pre-determined ordering). For example, consider a family of up to three generations. A grandfather and mother have their children that become father and mother of their children. So, they all are grouped together to the same family i.e they form a hierarchy.

Hierarchical clustering is of two types:

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

In hierarchical clustering, Objects are categorized into a hierarchy similar to a tree-shaped structure which is used to interpret hierarchical clustering models. The algorithm is as follows:

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.
Thumb Rule: Largest vertical distance which doesn't cut any horizontal line defines the optimal number of clusters.

mtcars(motor trend car road test) comprise fuel consumption, performance, and 10 aspects of automobile design for 32 automobiles. It comes pre-installed with dplyr package in R.

Performing Hierarchical clustering on Dataset
Using Hierarchical Clustering algorithm on the dataset using **hclust()** which is pre-installed in stats package when R is installed.

- The values are shown as per the distance matrix calculation with the method as euclidean.
- In the model, the cluster method is average, distance is euclidean and no. of objects are 32.
- The plot dendrogram is shown with x-axis as distance matrix and y-axis as height.
- So, Tree is cut where k = 3 and each category represents its number of clusters.
- The plot denotes dendrogram after being cut. The green lines show the number of clusters as per the thumb rule.

CODE:

```
#loading packages

library(dplyr)

library(ggplot2)


#summary of the dataset in the package

head(mtcars)
```

OUTPUT:

```
> #summary of the dataset in the package
> head(mtcars)
                   mpg cyl disp  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
```

CODE:

#finding the distance matrix

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

distance_mat

OUTPUT:

```
Merc 280C           227.8813169  64.8898713  39.3868519   1.5231546
Merc 450SE          106.4084264 175.1620073 159.8179555 122.3642489 122.3461050
Merc 450SL          106.4320572 175.1189767 159.7760899 122.3443771 122.3355492   0.9826495
Merc 450SLC         106.4010305 175.2118218 159.8495837 122.3934970 122.3586862   1.3726252   2.1383405
Cadillac Fleetwood  119.0239068 355.6627498 349.2832611 315.3904859 315.3557081 197.8842803 197.9154476
Lincoln Continental 104.5112999 348.9901277 341.3154316 306.6760719 306.6406187 187.5997191 187.6330806
Chrysler Imperial    81.4297699 338.1959373 328.4335161 292.7146896 292.6989332 171.6600758 171.6743028
Fiat 128            333.9792070  68.6105903  69.3127910 106.5053149 106.6829794 228.3247948 228.2592340
Honda Civic         344.0518316  72.0014488  78.5387212 116.7280991 116.8711475 238.0141824 237.9588183
Toyota Corolla      341.0218232  76.2806458  76.7731674 113.6290721 113.8118009 235.5183809 235.4481971
Toyota Corona       282.0508820  44.0850975  21.0962017  54.3641713  54.4258314 176.6020527 176.5727477
Dodge Challenger    103.9023864 192.8617917 185.8331870 152.8929263 152.8722437  51.8008639  51.8242520
AMC Javelin         110.3084921 180.5479760 172.5312555 139.1457974 139.1181977  41.2080044  41.2411618
Camaro Z28           10.0761203 273.8367985 257.7469734 219.5520854 219.5276434  98.7203049  98.7566899
Pontiac Firebird     80.8057339 277.4606884 271.3871978 238.1726099 238.1806292 124.3368538 124.3204160
Fiat X1-9           333.4843231  67.9163981  68.5564864 105.7412910 105.8560373 227.7627676 227.7173075
Porsche 914-2       285.1986201  39.4469276  22.1180967  57.6458160  57.8473863 179.5034108 179.4550855
Lotus Europa        296.4572287  72.8971106  50.1094030  74.1443580  74.3824296 193.3074449 193.2407697
Ford Pantera L       21.2655990 287.5238795 269.9772035 231.4081306 231.4024263 112.8181834 112.8296774
Ferrari Dino        226.2036333 113.3023005  80.6550953  56.8365103  56.8987601 131.0272205 131.0077635
Maserati Bora       107.7224977 313.8633093 288.8755628 250.5874125 250.5774357 157.1633256 157.1768956
Volvo 142E          275.1353516  53.6823481  24.6913548  48.8053450  48.8884618 170.4500681 170.4225164
                    Merc 450SLC Cadillac Fleetwood Lincoln Continental Chrysler Imperial    Fiat 128
Mazda RX4 Wag
Datsun 710
Hornet 4 Drive
Hornet Sportabout
Valiant
Duster 360
Merc 240D
Merc 230
Merc 280
Merc 280C
Merc 450SE
Merc 450SL
Merc 450SLC
Cadillac Fleetwood  197.8526242
Lincoln Continental 187.5671081          15.6224446
Chrysler Imperial   171.6557637          40.8399636          25.3714237
Fiat 128            228.4051825         417.7687579         410.0206984        397.2276375
Honda Civic         238.0828999         425.3271621         417.9679574        405.8152201  14.5590942
Toyota Corolla      235.6024098         425.3446517         417.5429986        404.6335386   7.8324789
Toyota Corona       176.6305359         368.3195488         360.0267515        346.5724649  52.8798281
Dodge Challenger     51.8012606         163.6314881         156.2805020        145.9194779 254.2367888
AMC Javelin          41.1929050         176.8610896         169.0925457        157.8097554 241.1203621
Camaro Z28           98.7035830         128.4587210         114.0932078         91.2880886 325.6636235
Pontiac Firebird    124.3726128          78.5385347          72.6947903         68.2030747 339.5857659
> distance_mat
                    Mazda RX4 Mazda RX4 Wag  Datsun 710 Hornet 4 Drive Hornet Sportabout     Valiant
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
Duster 360         241.4076490   241.4088680 294.4790230    169.4299647        70.1767262 194.6094525
Merc 240D           50.1532711    50.1146059  49.6584796    121.2739722       241.5069657  89.5911056
Merc 230            25.4683117    25.3284509  33.1803843    118.2433145       233.4924012  85.0079649
Merc 280            15.3641921    15.2956865  66.9363534     91.4224033       199.3344960  60.2909811
Merc 280C           15.6724727    15.5837744  67.0261397     91.4612914       199.3406564  60.2655656
Merc 450SE         135.4307018   135.4254826 189.1954941     72.4964325        84.3888482  90.6970264
Merc 450SL         135.4014424   135.3960351 189.1631745     72.4313532        84.3683999  90.6769728
Merc 450SLC        135.4794674   135.4723157 189.2345426     72.5718466        84.4332423  90.7092989
Cadillac Fleetwood 326.3395903   326.3355070 381.0926242    234.4403876       116.2804201 266.6280942
Lincoln Continental 318.0469808  318.0429333 372.8012090    227.9726091       108.0624299 259.6304391
Chrysler Imperial  304.7203408   304.7169175 359.3014906    218.1548299        97.2049146 248.7713290
Fiat 128            93.2679950    93.2530993  40.9933763    184.9689734       302.0377212 152.1153263
Honda Civic        102.8307567   102.8238713  52.7704607    191.5518700       310.0324645 158.9615769
Toyota Corolla     100.6040368   100.5887588  47.6535017    192.6714187       309.5581776 159.8302995
Toyota Corona       42.3075233    42.2659224  12.9654743    138.5304725       252.3331988 105.2876428
Dodge Challenger   163.1150750   163.1134210 217.7795805     72.4403915        48.9838851 103.4310693
AMC Javelin        149.6047203   149.6014522 204.3188913     61.3601899        61.4274240  91.0444349
Camaro Z28         233.2228758   233.2248748 286.0049209    163.6632641        70.9665308 187.8463771
Pontiac Firebird   248.6780270   248.6762035 303.3583889    156.2240346        40.0052475 188.5272116
Fiat X1-9           92.5048389    92.4940020  39.8815148    184.4471198       301.5669483 151.4379425
Porsche 914-2       44.4033659    44.4073589  13.1357109    139.1579524       254.1452553 106.0585767
Lotus Europa        65.7328377    65.7362635  25.0948550    163.2367437       272.3582423 130.8248192
Ford Pantera L     245.4247064   245.4293785 297.2940489    180.1140339        89.5934049 203.0177926
Ferrari Dino        66.7661029    66.7764167  90.2415509    130.5523007       215.0673853 106.5694802
Maserati Bora      265.6454248   265.6491465 309.7718171    229.3419352       170.7094473 242.4393015
Volvo 142E          39.1894029    39.1626037  20.6939436    137.0363299       248.0063378 104.1863681
                    Duster 360   Merc 240D    Merc 230    Merc 280   Merc 280C  Merc 450SE  Merc 450SL
Mazda RX4 Wag
Datsun 710
Hornet 4 Drive
Hornet Sportabout
Valiant
Duster 360
Merc 240D          281.2962502
Merc 230           265.8823313  33.6873047
Merc 280           227.8998521  64.7754228  39.2994160
```

OUTPUT:

```
Fiat X1-9           227.8176554        417.2490481         409.4998363       396.7597522   5.1473415
Porsche 914-2       179.5720446        370.0956775         362.0145494       348.8466861  49.0644372
Lotus Europa        193.3969216        388.5350012         379.4716659       364.5994326  49.9112509
Ford Pantera L      112.8332602        134.8119464         119.7236456        95.3805385 337.1639236
Ferrari Dino        131.0704490        328.5441628         317.7063117       300.1640703 128.3950054
Maserati Bora       157.1683970        214.9366858         199.3420611       174.2936864 349.5338830
Volvo 142E          170.4843735        364.1000930         355.4009443       341.2896659  61.3301247
                    Honda Civic Toyota Corolla Toyota Corona Dodge Challenger AMC Javelin  Camaro Z28
Mazda RX4 Wag
Datsun 710
Hornet 4 Drive
Hornet Sportabout
Valiant
Duster 360
Merc 240D
Merc 230
Merc 280
Merc 280C
Merc 450SE
Merc 450SL
Merc 450SLC
Cadillac Fleetwood
Lincoln Continental
Chrysler Imperial
Fiat 128
Honda Civic
Toyota Corolla       14.3480626
Toyota Corona        63.8985563     59.8451285
Dodge Challenger    261.8498815    261.8345312   205.0347927
AMC Javelin         248.9636504    248.6917065   191.5580526       14.0154995
Camaro Z28          335.8883188    332.6589699   273.6316895      100.3046106 105.6062618
Pontiac Firebird    347.0655360    347.1667643   290.6240706       85.8075196  99.2836114  86.2665759
Fiat X1-9            14.7807070     10.3922856    51.8411748      253.6624046 240.5266823 325.1490914
Porsche 914-2        59.4588768     56.3243031     8.6535903      206.6452569 193.3080584 276.8924414
Lotus Europa         64.0495153     53.8846563    31.2536926      226.5004836 212.7568765 287.6179004
Ford Pantera L      347.8337714    343.9920962   285.1287911      118.7516779 123.3832044  19.3589023
Ferrari Dino        141.7044478    133.4707617    82.2355734      174.9280395 161.1060307 216.7489910
Maserati Bora       362.1620777    355.2601619   299.1865216      185.9059273 185.1553411 102.5946154
Volvo 142E           73.3766041     67.7189421    12.2505275      201.3682522 187.6978440 266.5277736
                    Pontiac Firebird   Fiat X1-9 Porsche 914-2 Lotus Europa Ford Pantera L Ferrari Dino
Mazda RX4 Wag
Datsun 710
Hornet 4 Drive
Hornet Sportabout
Valiant
Duster 360
Merc 240D
Merc 230
Merc 280
Merc 280C
Merc 450SE
Merc 450SL
Merc 450SLC
Cadillac Fleetwood
Lincoln Continental
Chrysler Imperial
Fiat 128
Honda Civic
Toyota Corolla
Toyota Corona
Dodge Challenger
AMC Javelin
Camaro Z28
Pontiac Firebird
Fiat X1-9               339.1396182
Porsche 914-2           292.1646488  48.3775209
Lotus Europa            311.3862342  49.8406880    33.7678653
Ford Pantera L          101.7389686 336.7018783   288.5852993  297.5376920
Ferrari Dino            255.0570519 127.8210813    87.9105966   80.4553451    224.4587490
Maserati Bora           188.3240020 349.1199576   303.9222549  303.2796468     86.9383253  223.5342175
Volvo 142E              286.7497823  60.4120429    18.7555858   27.8104457    277.4803312   70.4751034
                    Maserati Bora
Mazda RX4 Wag
Datsun 710
Hornet 4 Drive
Hornet Sportabout
Valiant
Duster 360
Merc 240D
Merc 230
Merc 280
Merc 280C
Merc 450SE
Merc 450SL
Merc 450SLC
Cadillac Fleetwood
Lincoln Continental
Chrysler Imperial
Fiat 128
Honda Civic
Toyota Corolla
Toyota Corona
Dodge Challenger
AMC Javelin
```

OUTPUT:

```
Camaro Z28
Pontiac Firebird
Fiat X1-9
Porsche 914-2
Lotus Europa
Ford Pantera L
Ferrari Dino
Maserati Bora
Volvo 142E              289.1157363
```

CODE:

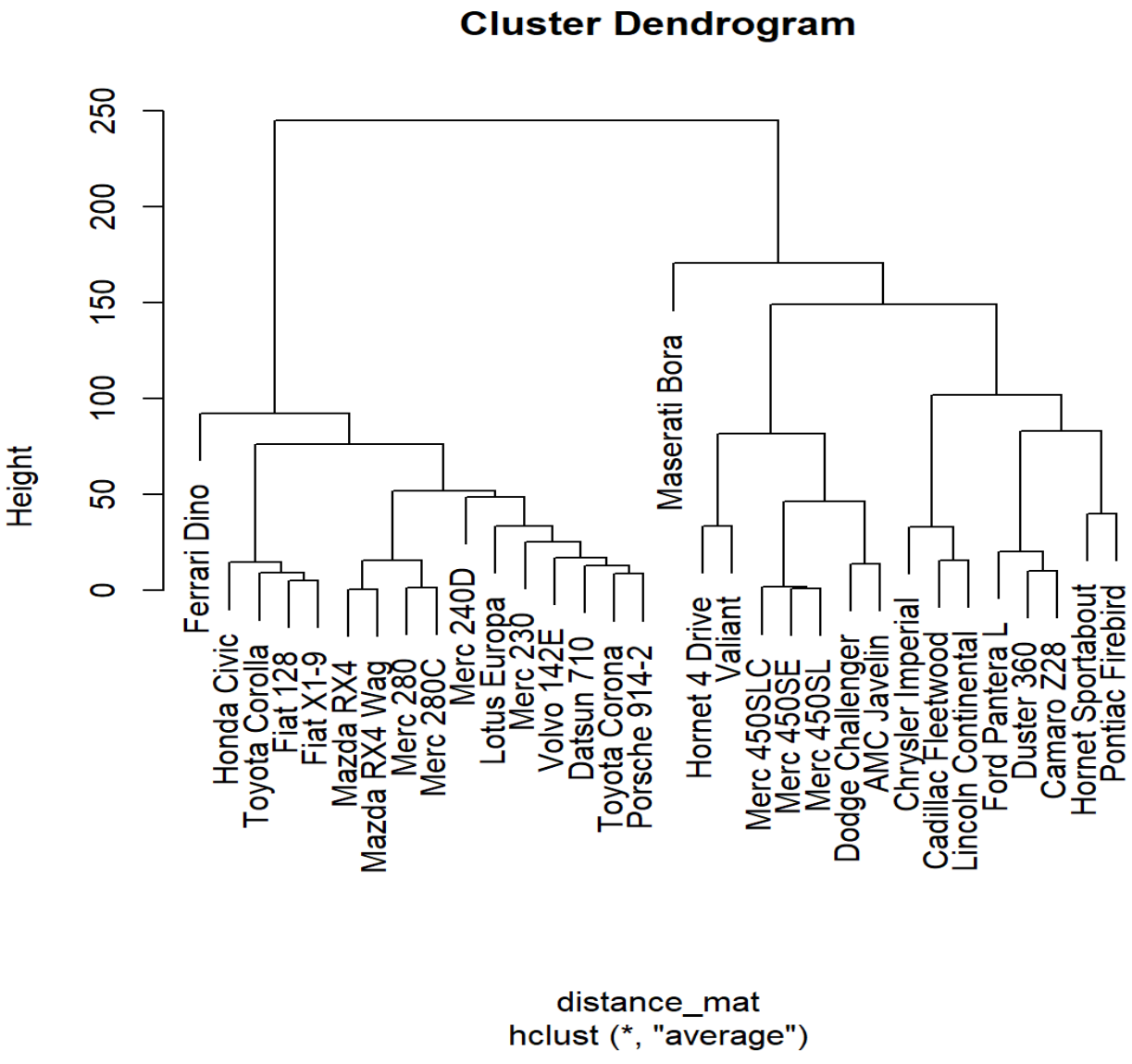
#fitting the hierarchical clustering model to the training data

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

#plotting the dendrogram

```
plot(Hierar_cl)
```

OUTPUT:



CODE:

#choosing the number of clusters (cut by the number of clustering)

num_clusters<-3 **#change this to the desired number of clusters**

fit<-cutree(Hierar_cl,k=num_clusters)

fit

#display the count of datapoints in each cluster

table(fit)

OUTPUT:

```
> fit
          Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive   Hornet Sportabout
                  1                   1                   1                   2                   2
            Valiant          Duster 360           Merc 240D            Merc 230            Merc 280
                  2                   2                   1                   1                   1
          Merc 280C          Merc 450SE          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
                  1                   2                   2                   2                   2
Lincoln Continental   Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
                  2                   2                   1                   1                   1
      Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28    Pontiac Firebird
                  1                   2                   2                   2                   2
          Fiat X1-9       Porsche 914-2        Lotus Europa      Ford Pantera L         Ferrari Dino
                  1                   1                   1                   2                   1
      Maserati Bora          Volvo 142E
                  3                   1
>
> #display the count of datapoints in each cluster
> table(fit)
fit
 1  2  3
16 15  1
```

CODE:

#highlight clusters on the dendrogram

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

#create a dataframe with cluster assignments

clustered_data<-data.frame(mtcars,Cluster=factor(fit))

#define the custom cluster colors

custom_colors<-c("red","pink","lavender")

#create a scatterplot with custom colors

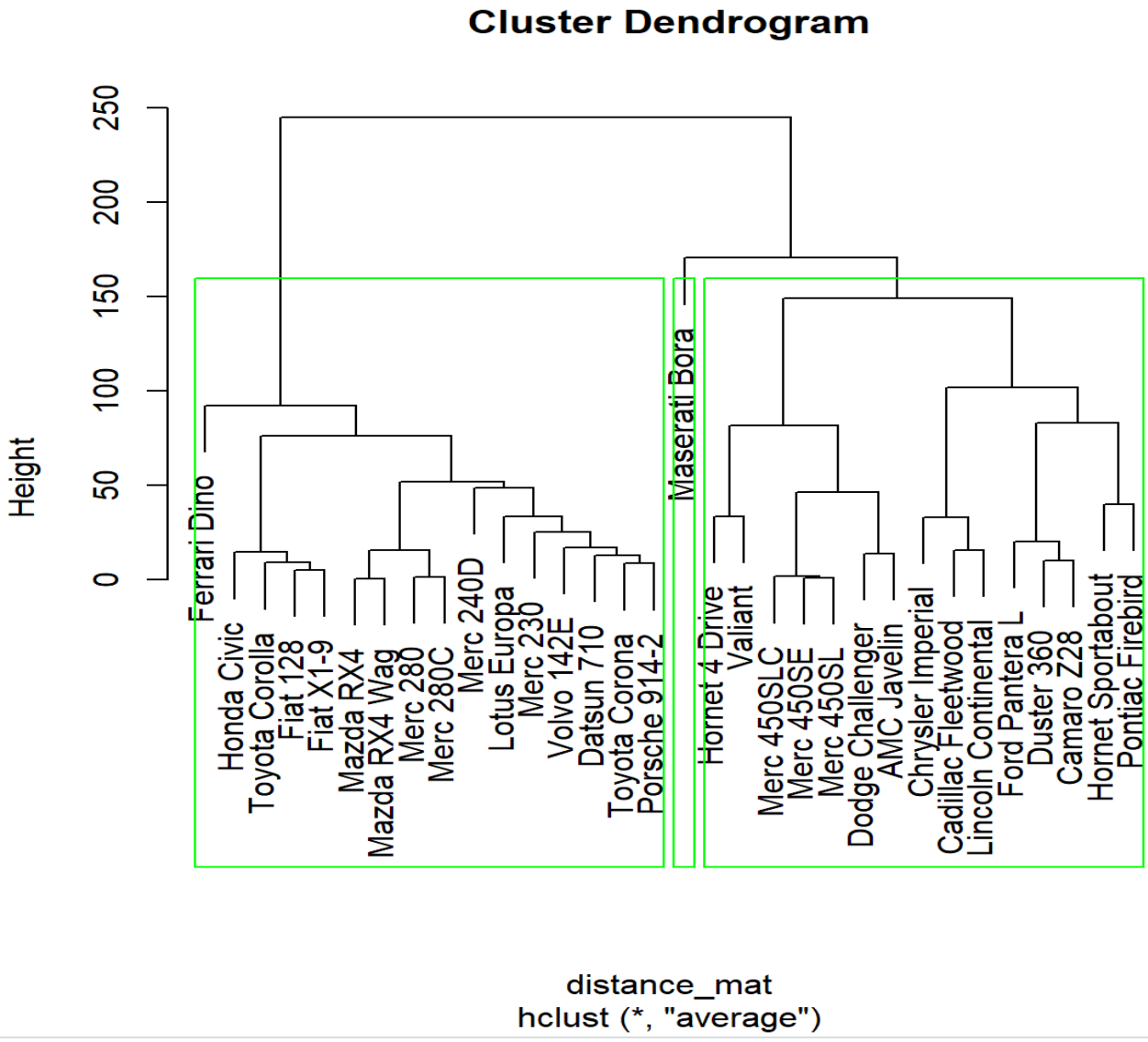
scatterplot<-ggplot(clustered_data,aes(x=mpg,

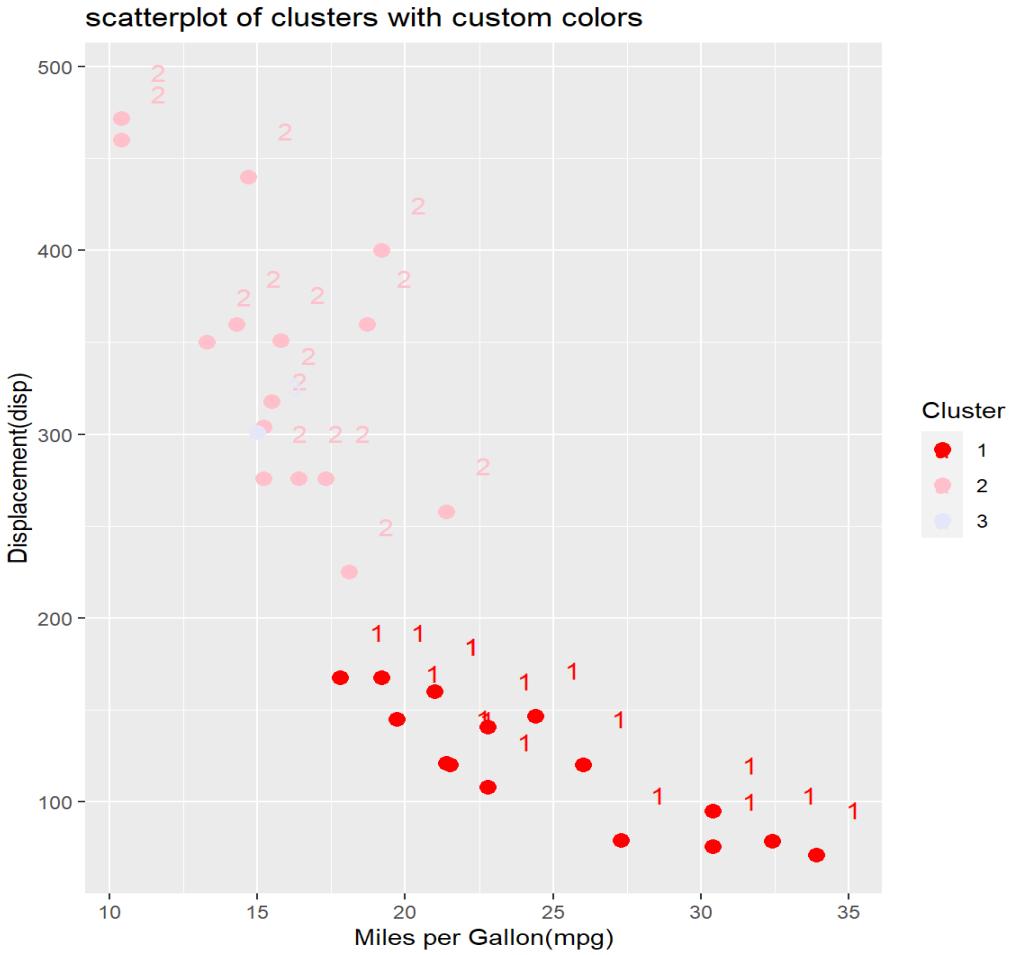
```
                    y=disp,
                    color=Cluster))+
  geom_point(size=3)+
  geom_text(aes(label=Cluster),hjust=0,vjust=0,
            nudge_x=1,nudge_y=20,size=4)+
  labs(x="Miles per Gallon(mpg)",y="Displacement(disp)",
       title="scatterplot of clusters with custom colors")+
  scale_color_manual(values=custom_colors,name='Cluster')
```

#display scatter plot

scatterplot

OUTPUT:





CODE:

 # Calculate the mean mpg for each cluster

```
cluster_means <- clustered_data %>%
  group_by(Cluster) %>%
  summarise(mean_mpg = mean(mpg))
```

 # Create bar graphs for "mpg" variable within each cluster

```
bar_graph <- ggplot(cluster_means, aes(x = Cluster, y = mean_mpg, fill = Cluster)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = round(mean_mpg, 2), y = mean_mpg + 0.5), vjust = -0.5, position =
position_dodge(width = 0.9)) +
  labs(x = "Cluster", y = "Miles per Gallon (mpg)", title = "Bar Graphs of miles per gallon (mpg) by
Cluster") +
  scale_fill_manual(values = custom_colors, name = "Cluster") +
  theme_minimal()
```

#display barplot

print(bar_graph)

OUTPUT: