Practical No. 04

AIM - Practical of Clustering.

Source Code -

```
"k-means clustering "
data("iris")
names(iris)
new_data<-subset(iris,select = c(-Species))
new_data
cl <- kmeans(new_data,3)
cl
```

```
data <- new_data
wss <- sapply(1:15, function(k){kmeans(data, k )$tot.withinss})
wss
```

```
plot(1:15, wss, type="b", pch=19, frame=FALSE, xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")
install.packages("cluster")
library(cluster)
clusplot(new_data, cl$cluster, color=TRUE, shade=TRUE, labels=2, lines=0)
cl$cluster
cl$centers
```

```
"agglomerative clustering"
clusters <- hclust(dist(iris[, 3:4]))
plot(clusters)
clusterCut <- cutree(clusters, 3)
table(clusterCut, iris$Species)
install.packages("ggplot2")
library("ggplot2")
ggplot(iris, aes(Petal.Length, Petal.Width, color = Species)) +
  geom_point(alpha = 0.4, size = 3.5) + geom_point(col = clusterCut) +
  scale_color_manual(values = c('black', 'red', 'green'))
```

```
clusters <- hclust(dist(iris[, 3:4]), method='average')
clusterCut1 <- cutree(clusters, 3)
table(clusterCut1, iris$Species)
plot(clusters)
ggplot(iris, aes(Petal.Length, Petal.Width, color=Species)) +
  geom_point(alpha = 0.4, size = 3.5) + geom_point(col=clusterCut1) +
  scale_color_manual(values = c('black', 'red', 'green'))
```

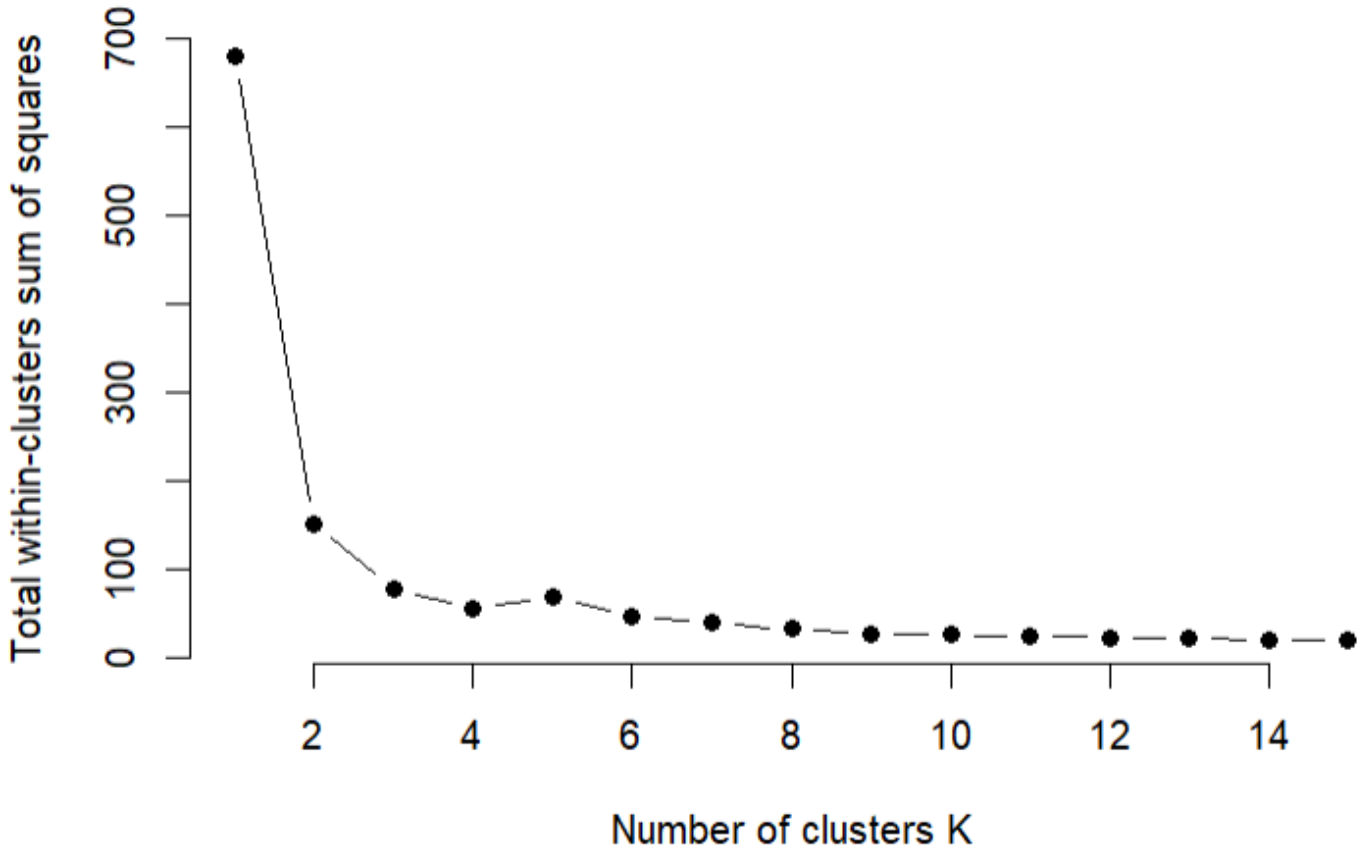
OUTPUT -

```
> "k-means clustering "
[1] "k-means clustering "
> data("iris")
> names(iris)
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
> new_data<-subset(iris,select = c(-Species))
> new_data
```

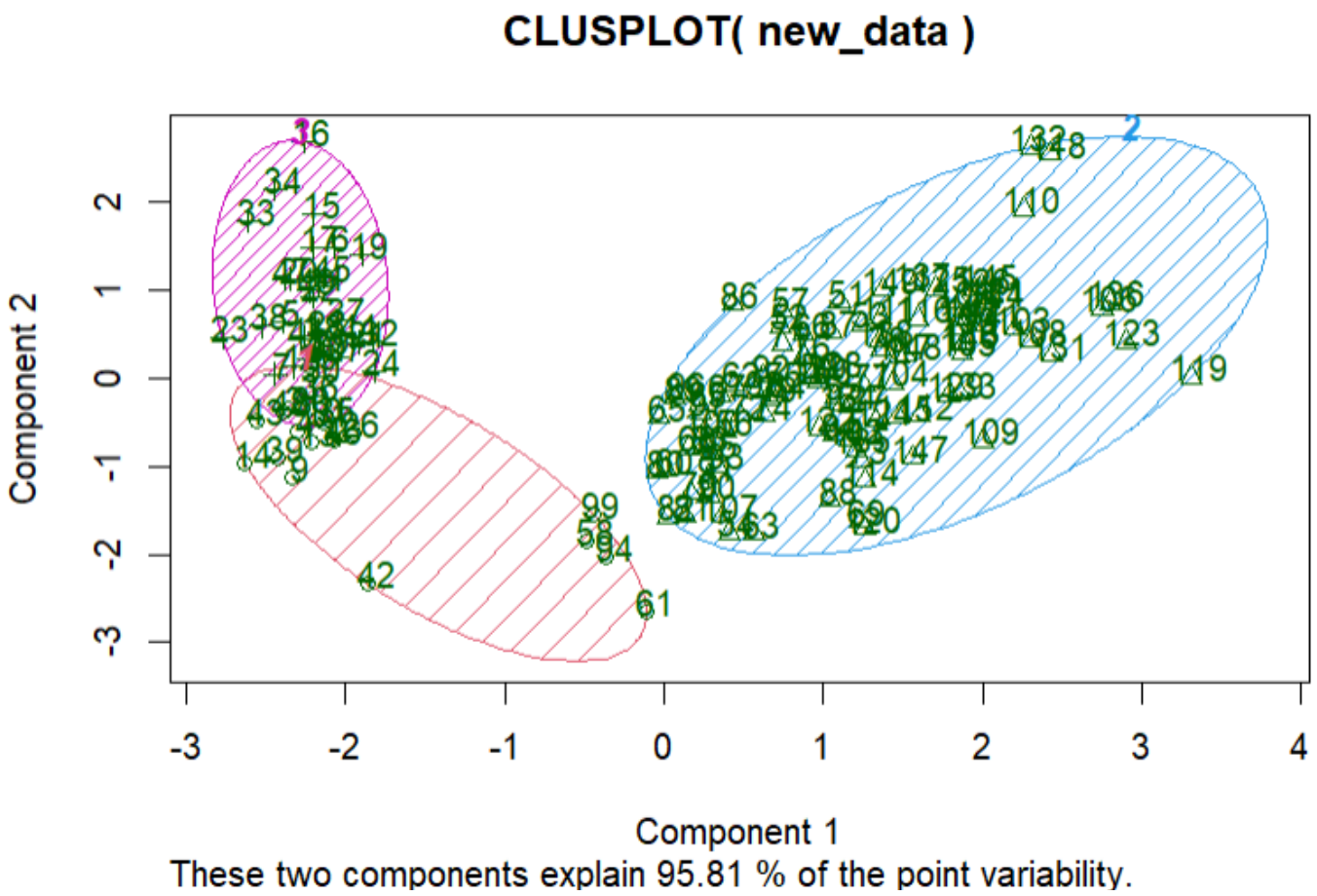
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1
11	5.4	3.7	1.5	0.2
12	4.8	3.4	1.6	0.2
13	4.8	3.0	1.4	0.1
14	4.3	3.0	1.1	0.1
15	5.8	4.0	1.2	0.2
16	5.7	4.4	1.5	0.4
17	5.4	3.9	1.3	0.4
18	5.1	3.5	1.4	0.3
19	5.7	3.8	1.7	0.3
20	5.1	3.8	1.5	0.3
21	5.4	3.4	1.7	0.2
22	5.1	3.7	1.5	0.4
23	4.6	3.6	1.0	0.2
24	5.1	3.3	1.7	0.5
25	4.8	3.4	1.9	0.2
26	5.0	3.0	1.6	0.2
27	5.0	3.4	1.6	0.4
28	5.2	3.5	1.5	0.2
29	5.2	3.4	1.4	0.2
30	4.7	3.2	1.6	0.2
31	4.8	3.1	1.6	0.2
32	5.4	3.4	1.5	0.4
33	5.2	4.1	1.5	0.1
34	5.5	4.2	1.4	0.2
35	4.9	3.1	1.5	0.2
36	5.0	3.2	1.2	0.2
37	5.5	3.5	1.3	0.2
38	4.9	3.6	1.4	0.1
39	4.4	3.0	1.3	0.2
40	5.1	3.4	1.5	0.2
41	5.0	3.5	1.3	0.3
42	4.5	2.3	1.3	0.3
43	4.4	3.2	1.3	0.2
44	5.0	3.5	1.6	0.6
45	5.1	3.8	1.9	0.4
46	4.8	3.0	1.4	0.3
47	5.1	3.8	1.6	0.2
48	4.6	3.2	1.4	0.2
49	5.3	3.7	1.5	0.2
50	5.0	3.3	1.4	0.2
51	7.0	3.2	4.7	1.4
52	6.4	3.2	4.5	1.5
53	6.9	3.1	4.9	1.5
54	5.5	2.3	4.0	1.3
55	6.5	2.8	4.6	1.5
56	5.7	2.8	4.5	1.3
57	6.3	3.3	4.7	1.6
58	4.9	2.4	3.3	1.0
59	6.6	2.9	4.6	1.3
60	5.2	2.7	3.9	1.4
61	5.0	2.0	3.5	1.0
62	5.9	3.0	4.2	1.5
63	6.0	2.2	4.0	1.0
64	6.1	2.9	4.7	1.4
65	5.6	2.9	3.6	1.3
66	6.7	3.1	4.4	1.4
67	5.6	3.0	4.5	1.5
68	5.8	2.7	4.1	1.0
69	6.2	2.2	4.5	1.5
70	5.6	2.5	3.9	1.1
71	5.9	3.2	4.8	1.8
72	6.1	2.8	4.0	1.3
73	6.3	2.5	4.9	1.5
74	6.1	2.8	4.7	1.2
75	6.4	2.9	4.3	1.3
76	6.6	3.0	4.4	1.4
77	6.8	2.8	4.8	1.4
78	6.7	3.0	5.0	1.7
79	6.0	2.9	4.5	1.5
80	5.7	2.6	3.5	1.0
81	5.5	2.4	3.8	1.1
82	5.5	2.4	3.7	1.0
83	5.8	2.7	3.9	1.2
84	6.0	2.7	5.1	1.6
85	5.4	3.0	4.5	1.5
86	6.0	3.4	4.5	1.6

```
87          6.7         3.1          4.7         1.5  121         6.9         3.2          5.7         2.3
88          6.3         2.3          4.4         1.3  122         5.6         2.8          4.9         2.0
89          5.6         3.0          4.1         1.3  123         7.7         2.8          6.7         2.0
90          5.5         2.5          4.0         1.3  124         6.3         2.7          4.9         1.8
91          5.5         2.6          4.4         1.2  125         6.7         3.3          5.7         2.1
92          6.1         3.0          4.6         1.4  126         7.2         3.2          6.0         1.8
93          5.8         2.6          4.0         1.2  127         6.2         2.8          4.8         1.8
94          5.0         2.3          3.3         1.0  128         6.1         3.0          4.9         1.8
95          5.6         2.7          4.2         1.3  129         6.4         2.8          5.6         2.1
96          5.7         3.0          4.2         1.2  130         7.2         3.0          5.8         1.6
97          5.7         2.9          4.2         1.3  131         7.4         2.8          6.1         1.9
98          6.2         2.9          4.3         1.3  132         7.9         3.8          6.4         2.0
99          5.1         2.5          3.0         1.1  133         6.4         2.8          5.6         2.2
100         5.7         2.8          4.1         1.3  134         6.3         2.8          5.1         1.5
101         6.3         3.3          6.0         2.5  135         6.1         2.6          5.6         1.4
102         5.8         2.7          5.1         1.9  136         7.7         3.0          6.1         2.3
103         7.1         3.0          5.9         2.1  137         6.3         3.4          5.6         2.4
104         6.3         2.9          5.6         1.8  138         6.4         3.1          5.5         1.8
105         6.5         3.0          5.8         2.2  139         6.0         3.0          4.8         1.8
106         7.6         3.0          6.6         2.1  140         6.9         3.1          5.4         2.1
107         4.9         2.5          4.5         1.7  141         6.7         3.1          5.6         2.4
108         7.3         2.9          6.3         1.8  142         6.9         3.1          5.1         2.3
109         6.7         2.5          5.8         1.8  143         5.8         2.7          5.1         1.9
110         7.2         3.6          6.1         2.5  144         6.8         3.2          5.9         2.3
111         6.5         3.2          5.1         2.0  145         6.7         3.3          5.7         2.5
112         6.4         2.7          5.3         1.9  146         6.7         3.0          5.2         2.3
113         6.8         3.0          5.5         2.1  147         6.3         2.5          5.0         1.9
114         5.7         2.5          5.0         2.0  148         6.5         3.0          5.2         2.0
115         5.8         2.8          5.1         2.4  149         6.2         3.4          5.4         2.3
116         6.4         3.2          5.3         2.3  150         5.9         3.0          5.1         1.8
117         6.5         3.0          5.5         1.8
118         7.7         3.8          6.7         2.2
119         7.7         2.6          6.9         2.3
120         6.0         2.2          5.0         1.5
> cl <- kmeans(new_data,3)
> cl
K-means clustering with 3 clusters of sizes 21, 96, 33

Cluster means:
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1     4.738095    2.904762     1.790476   0.3523810
2     6.314583    2.895833     4.973958   1.7031250
3     5.175758    3.624242     1.472727   0.2727273

Clustering vector:
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
  3   1   1   1   3   3   3   3   1   1   3   3   1   1   3   3   3   3   3   3
 21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
  3   3   3   3   1   1   3   3   3   1   1   3   3   3   1   3   3   3   1   3
 41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
  3   1   1   3   3   1   3   1   3   3   2   2   2   2   2   2   2   1   2   2
 61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80
  1   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
 81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100
  2   2   2   2   2   2   2   2   2   2   2   2   2   1   2   2   2   2   1   2
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
  2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
  2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
141 142 143 144 145 146 147 148 149 150
  2   2   2   2   2   2   2   2   2   2
```

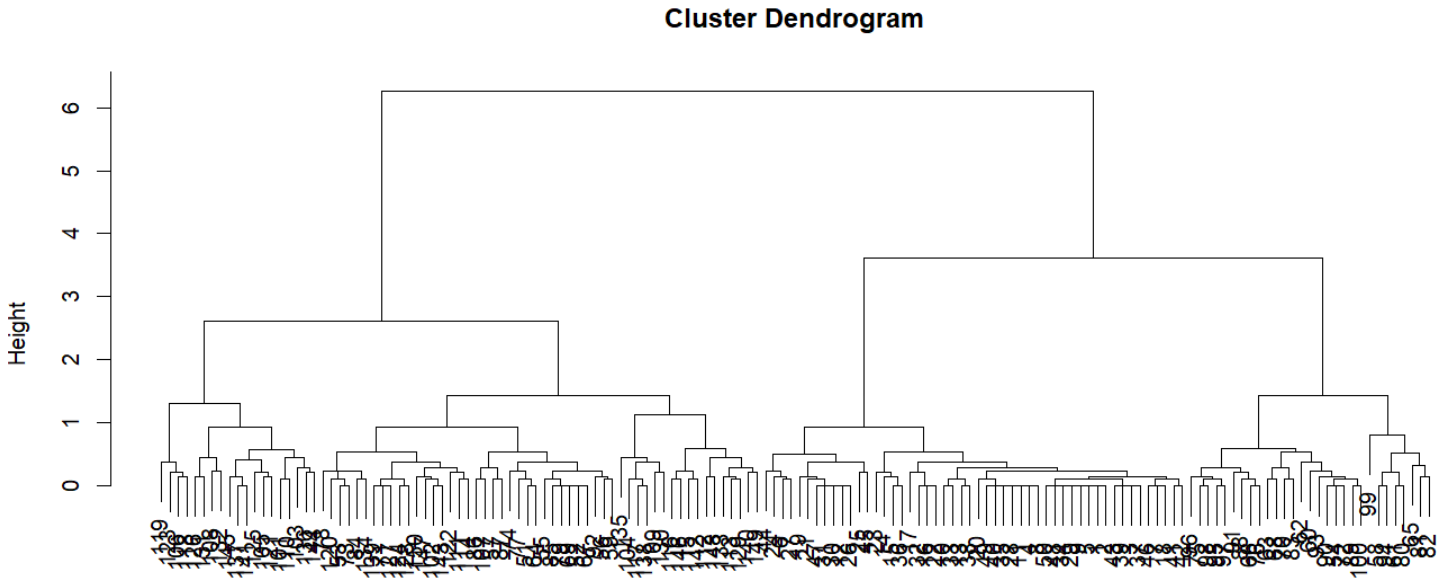
```
Within cluster sum of squares by cluster:
[1]  17.669524 118.651875   6.432121
 (between_SS / total_SS =  79.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
> data <- new_data
> wss <- sapply(1:15, function(k){kmeans(data, k )$tot.withinss})
> wss
 [1] 681.37060 152.34795  78.85144  57.22847  69.24240  47.61943  39.98885
 [8]  32.96739  28.17230  26.26447  25.65746  23.93863  22.48749  21.04254
[15]  20.19866
> plot(1:15, wss, type="b", pch=19, frame=FALSE,
+      xlab="Number of clusters K",
+      ylab="Total within-clusters sum of squares")
```



```
> install.packages("cluster")
WARNING: Rtools is required to build R packages but is not currently installed. Please download
and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Kunal/AppData/Local/R/win-library/4.2'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/cluster_2.1.4.zip'
Content type 'application/zip' length 585996 bytes (572 KB)
downloaded 572 KB

package 'cluster' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Kunal\AppData\Local\Temp\RtmpmKwA7v\downloaded_packages
> library(cluster)
> clusplot(new_data, cl$cluster, color=TRUE, shade=TRUE, labels=2, lines=0)
```

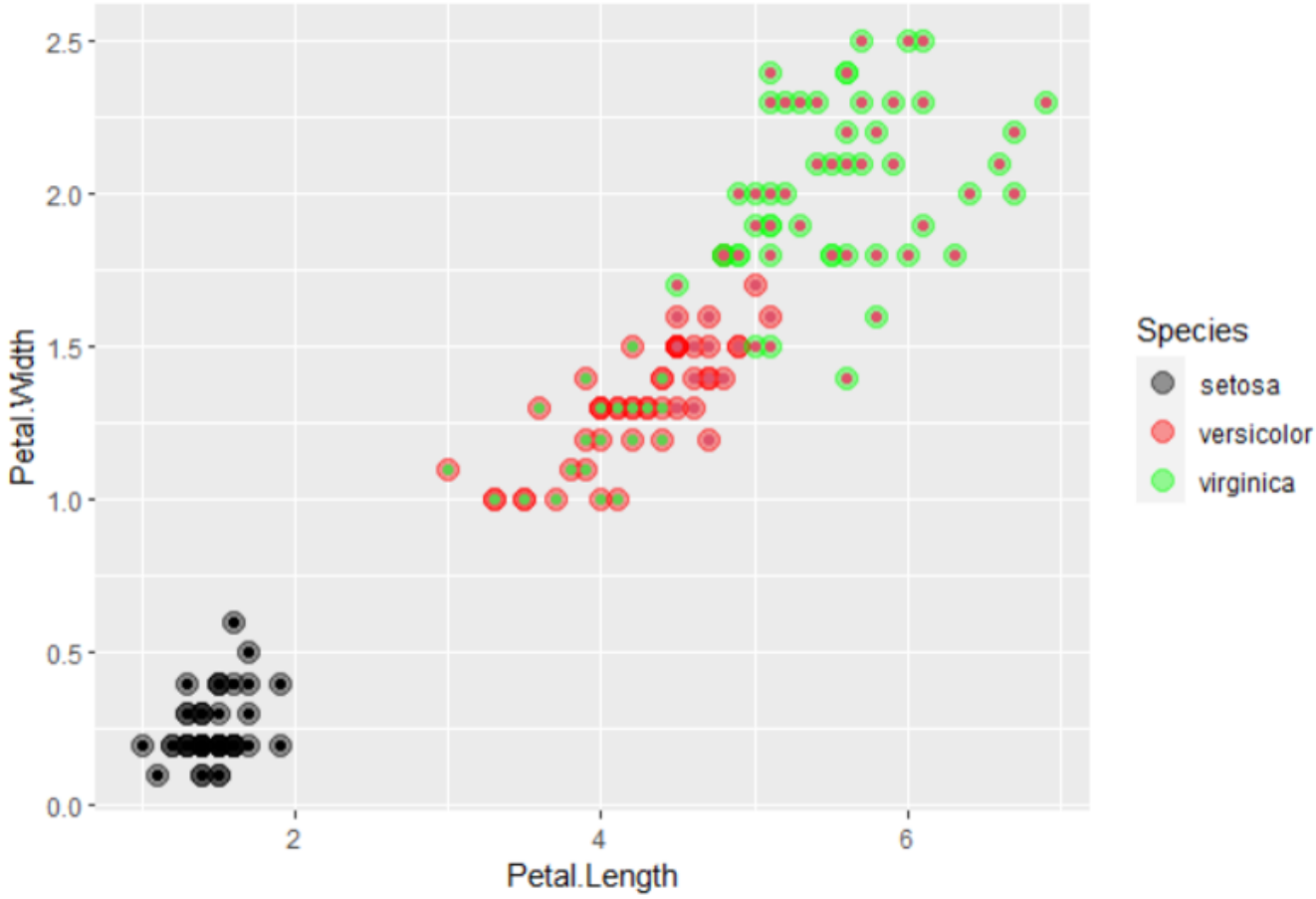


```
> cl$cluster
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25
  3   1   1   1   3   3   3   3   1   1   3   3   1   1   3   3   3   3   3   3   3   3   3   3   1
 26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50
  1   3   3   3   1   1   3   3   3   1   3   3   3   1   3   3   1   1   3   3   1   3   1   3   3
 51  52  53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75
  2   2   2   2   2   2   2   1   2   2   1   2   2   2   2   2   2   2   2   2   2   2   2   2   2
 76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100
  2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   1   2   2   2   2   1   2
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125
  2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
  2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
> cl$centers
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1     4.738095    2.904762     1.790476   0.3523810
2     6.314583    2.895833     4.973958   1.7031250
3     5.175758    3.624242     1.472727   0.2727273
> "agglomarative clustering "
[1] "agglomarative clustering "
> clusters <- hclust(dist(iris[, 3:4]))
> plot(clusters)
```

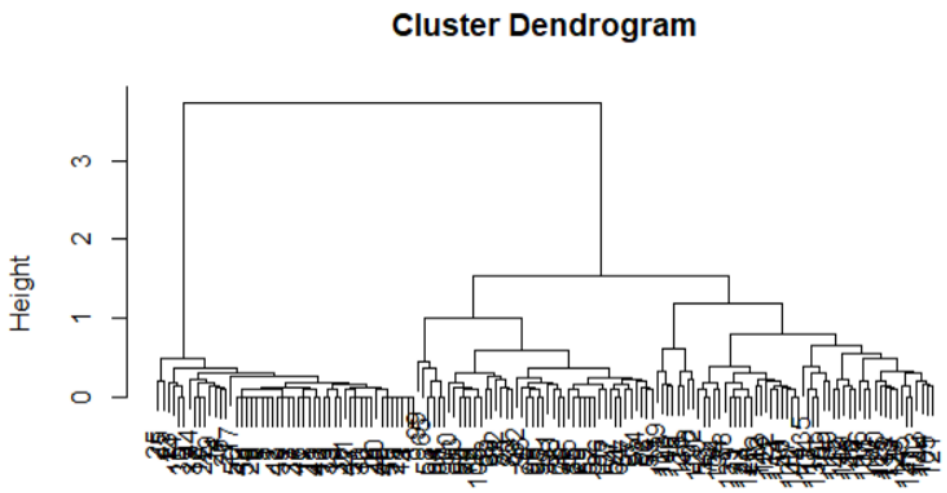



```
> clusterCut <- cutree(clusters, 3)
> table(clusterCut, iris$Species)

clusterCut setosa versicolor virginica
         1     50          0         0
         2      0         21        50
         3      0         29         0
> install.packages("ggplot2")
```

```
package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Kunal\AppData\Local\Temp\RtmpmKwA7v\downloaded_packages
> library("ggplot2")
Need help? Try Stackoverflow: https://stackoverflow.com/tags/ggplot2
> ggplot(iris, aes(Petal.Length, Petal.Width, color = Species)) +
+   geom_point(alpha = 0.4, size = 3.5) + geom_point(col = clusterCut) +
+   scale_color_manual(values = c('black', 'red', 'green'))
```




```
> clusters <- hclust(dist(iris[, 3:4]), method='average')
> clusterCut1 <- cutree(clusters, 3)
> table(clusterCut1, iris$Species)

clusterCut1 setosa versicolor virginica
          1     50          0         0
          2      0         45         1
          3      0          5        49
> plot(clusters)
```




```
> ggplot(iris, aes(Petal.Length, Petal.Width, color=Species)) +
+   geom_point(alpha = 0.4, size = 3.5) + geom_point(col=clusterCut1) +
+   scale_color_manual(values = c('black', 'red', 'green'))
```