Practical No 8

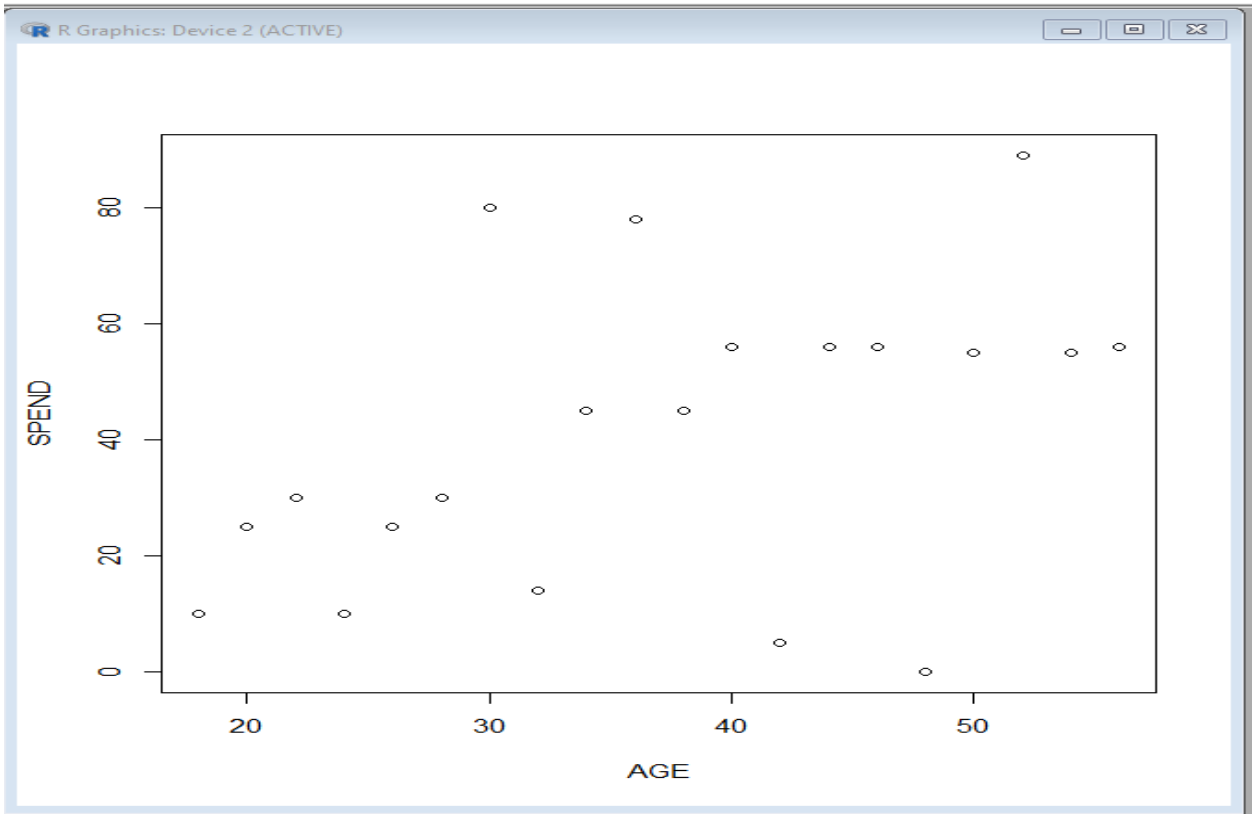
Aim: Demonstration of Clustering

Code:

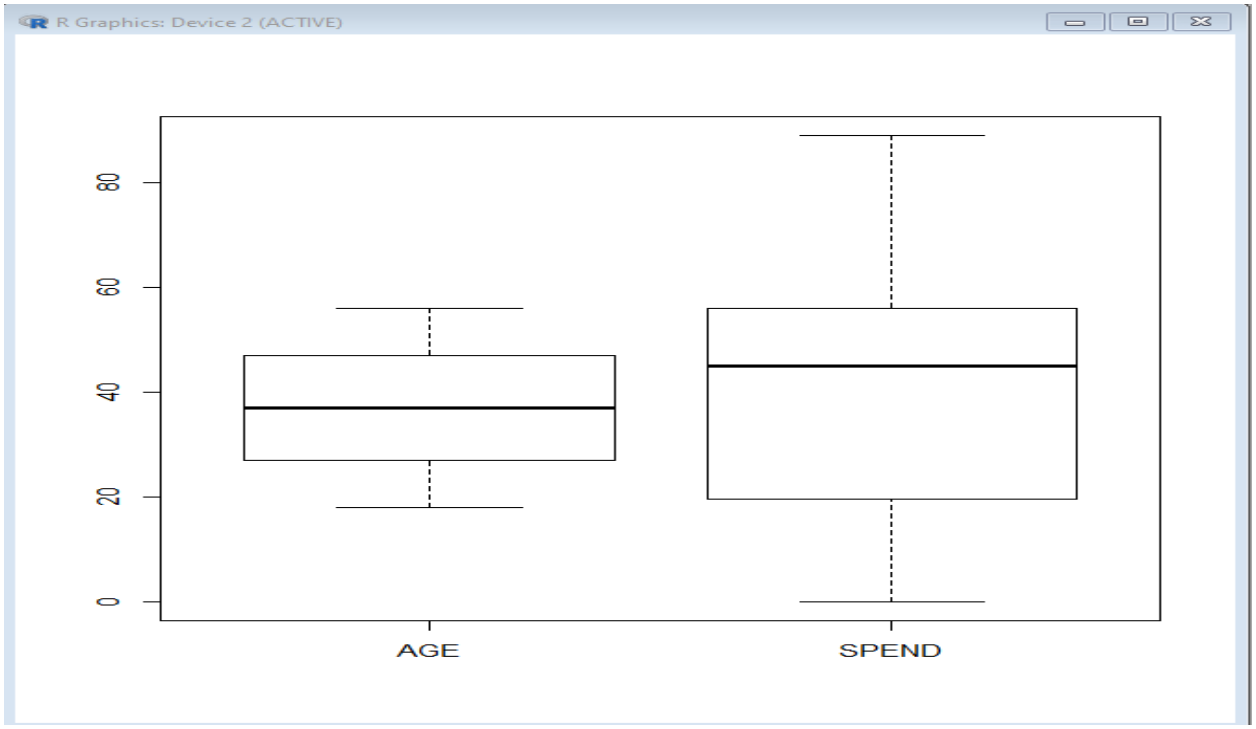
```
df=read.csv("C:/Users/admin/Documents/AGE.csv")
df
```

```
> df=read.csv("C:/Users/admin/Documents/AGE.csv")
> df
   AGE SPEND
1   18    10
2   20    25
3   22    30
4   24    10
5   26    25
6   28    30
7   30    80
8   32    14
9   34    45
10  36    78
11  38    45
12  40    56
13  42     5
14  44    56
15  46    56
16  48     0
17  50    55
18  52    89
19  54    55
20  56    56
```

plot(df)



boxplot(df)



Make the cluster

>set.seed(20)

> c1=kmeans(df[,1:2],3)

> c1

```
> set.seed(20)
> cl=kmeans(df[,1:2],3)
> cl
K-means clustering with 3 clusters of sizes 3, 8, 9

Cluster means:
       AGE    SPEND
1 39.33333 82.33333
2 45.25000 53.00000
3 28.88889 16.55556

Clustering vector:
 [1] 3 3 3 3 3 3 1 3 2 1 2 2 3 2 2 3 2 1 2 2

Within cluster sum of squares by cluster:
[1]  327.3333  595.5000 1829.1111
 (between_SS / total_SS =  82.3 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

#SHOW THE IRIS DATA SET

>iris

```
> iris
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1            5.1         3.5          1.4         0.2     setosa
2            4.9         3.0          1.4         0.2     setosa
3            4.7         3.2          1.3         0.2     setosa
4            4.6         3.1          1.5         0.2     setosa
5            5.0         3.6          1.4         0.2     setosa
6            5.4         3.9          1.7         0.4     setosa
7            4.6         3.4          1.4         0.3     setosa
8            5.0         3.4          1.5         0.2     setosa
9            4.4         2.9          1.4         0.2     setosa
10           4.9         3.1          1.5         0.1     setosa
11           5.4         3.7          1.5         0.2     setosa
12           4.8         3.4          1.6         0.2     setosa
13           4.8         3.0          1.4         0.1     setosa
```

#View(iris)

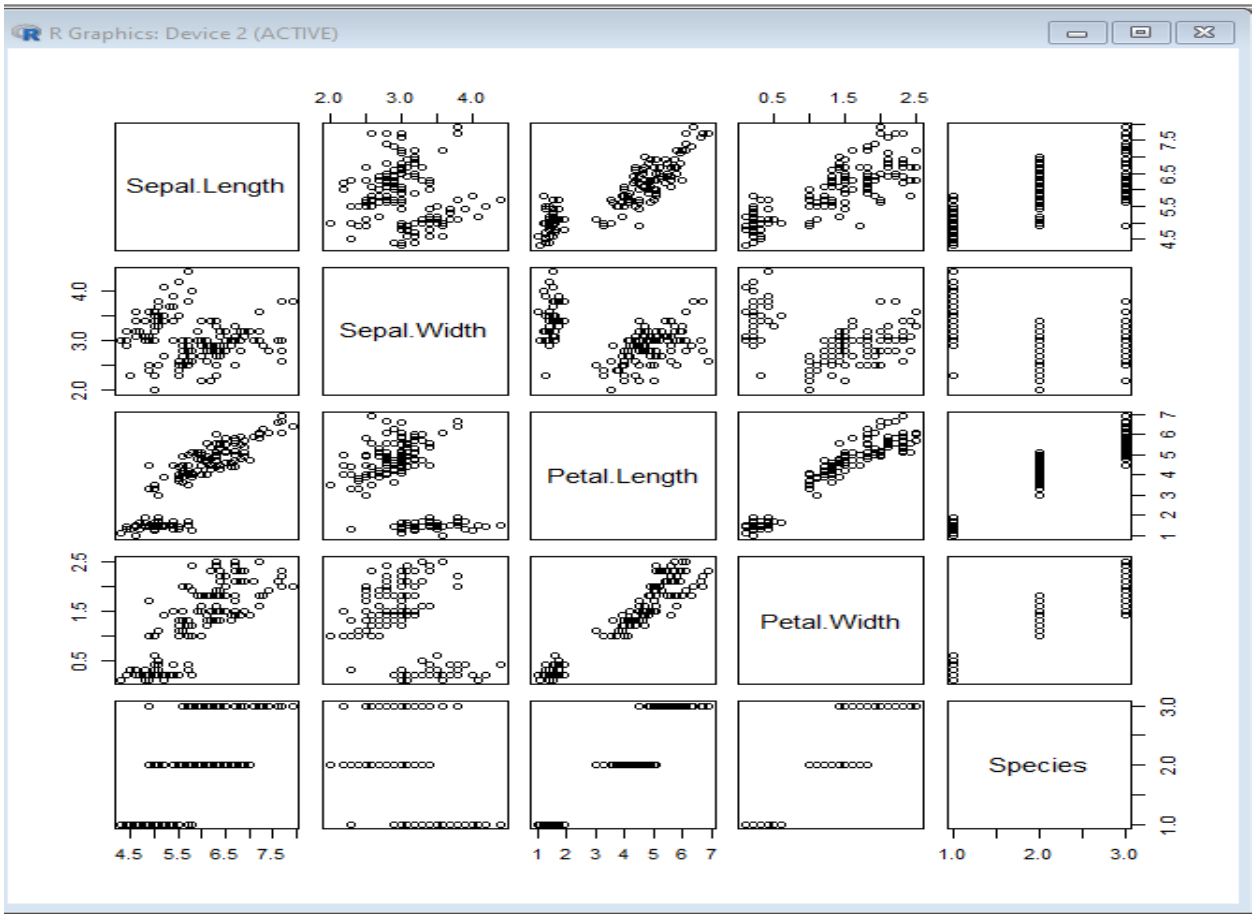
Data: iris

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa
17	5.4	3.9	1.3	0.4	setosa
18	5.1	3.5	1.4	0.3	setosa
19	5.7	3.8	1.7	0.3	setosa

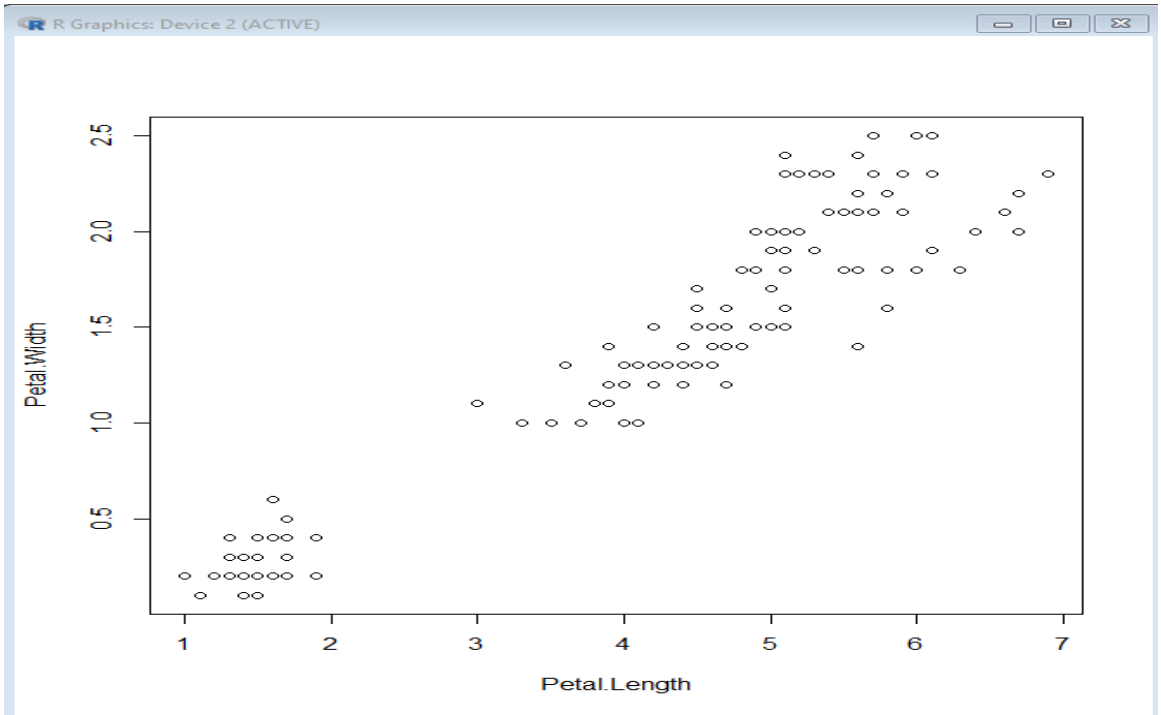
head(iris)
summary(iris)

```
> View(iris)
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> summary(iris)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
 Median :5.800   Median :3.000   Median :4.350   Median :1.300
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
       Species
 setosa    :50
 versicolor:50
 virginica :50
```

plot(iris)


R Graphics: Device 2 (ACTIVE)
Sepal.Length
Sepal.Width
Petal.Length
Petal.Width
Species

plot(iris[,3:4])


R Graphics: Device 2 (ACTIVE)
Petal.Width
Petal.Length

kmeansc1=kmeans(iris[,3:4],3)

kmeansc1

```
> kmeanscl=kmeans(iris[,3:4],3)
> kmeanscl
K-means clustering with 3 clusters of sizes 50, 46, 54

Cluster means:
  Petal.Length Petal.Width
1     1.462000    0.246000
2     5.626087    2.047826
3     4.292593    1.359259

Clustering vector:
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
 [75] 3 3 3 2 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 3 2 2 2 2 2
[112] 2 2 2 2 2 2 2 2 3 2 2 2 3 2 2 3 3 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2
[149] 2 2

Within cluster sum of squares by cluster:
[1]  2.02200 15.16348 14.22741
 (between_SS / total_SS =  94.3 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
> |
```

PRINT CONFUSION MATRIX

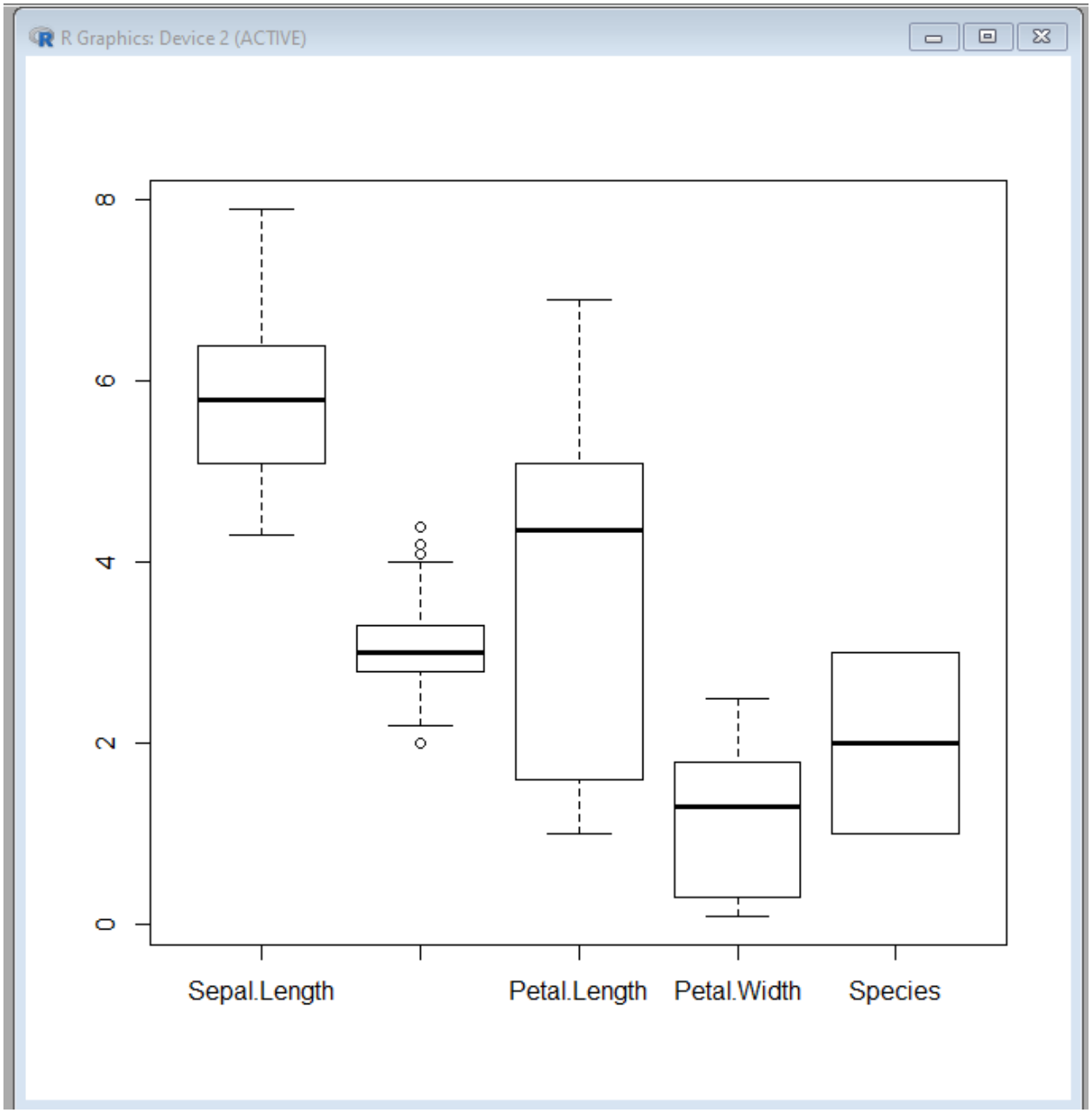
>table(kmeansc1$cluster,iris$Species)

```
> table(kmeanscl$cluster,iris$Species)

   setosa versicolor virginica
1      50          0         0
2       0          2        44
3       0         48         6
> |
```

CALCULATION OF ACCURACY 94.6%

boxplot(iris)


R Graphics: Device 2 (ACTIVE)
0
2
4
6
8
Sepal.Length
Petal.Length
Petal.Width
Species