Data Visualization Lab(L13+L14)

Name: Alpanshu Kataria

Reg. No.: 18BCE1267

Faculty: Dr.Parvathi.R

Lab 2

Task:

KMeans_Animation on wheat seeds dataset.

	Area	Perimeter	Compactness	Kernel.Length	Kernel.Width	Asymmetry.Coeff	Kernel.Groove	Type
1	15.26	14.84	0.8710	5.763	3.312	2.2210	5.220	1
2	14.88	14.57	0.8811	5.554	3.333	1.0180	4.956	1
3	14.29	14.09	0.9050	5.291	3.337	2.6990	4.825	1
4	13.84	13.94	0.8955	5.324	3.379	2.2590	4.805	1
5	16.14	14.99	0.9034	5.658	3.562	1.3550	5.175	1
6	14.38	14.21	0.8951	5.386	3.312	2.4620	4.956	1
7	14.69	14.49	0.8799	5.563	3.259	3.5860	5.219	1
8	16.63	15.46	0.8747	6.053	3.465	2.0400	5.877	1
9	16.44	15.25	0.8880	5.884	3.505	1.9690	5.533	1
10	15.26	14.85	0.8696	5.714	3.242	4.5430	5.314	1
11	14.03	14.16	0.8796	5.438	3.201	1.7170	5.001	1
12	13.89	14.02	0.8880	5.439	3.199	3.9860	4.738	1
13	13.78	14.06	0.8759	5.479	3.156	3.1360	4.872	1
14	13.74	14.05	0.8744	5.482	3.114	2.9320	4.825	1
15	14.59	14.28	0.8993	5.351	3.333	4.1850	4.781	1
16	13.99	13.83	0.9183	5.119	3.383	5.2340	4.781	1
17	15.69	14.75	0.9058	5.527	3.514	1.5990	5.046	1
18	14.70	14.21	0.9153	5.205	3.466	1.7670	4.649	1
19	12.72	13.57	0.8686	5.226	3.049	4.1020	4.914	1
20	14.16	14.40	0.8584	5.658	3.129	3.0720	5.176	1
21	14.11	14.26	0.8722	5.520	3.168	2.6880	5.219	1

Code:

```
#require("datasets")

#data("seeds")

seeds <- read.csv("D:/ViT/Sem 6/Data viz/Lab/Lab2/seeds.csv")

str(seeds)

summary(seeds)

head(seeds)

seeds.new<- seeds[,c(1,2,3,4,5,6,7)]

seeds.class<- seeds[,"Type"]

head(seeds.new)

head(seeds.class)

normalize <- function(x){

  return ((x-min(x))/(max(x)-min(x)))

}

seeds.new$Area<- normalize(seeds.new$Area)

seeds.new$Perimeter<- normalize(seeds.new$Perimeter)

seeds.new$Compactness<- normalize(seeds.new$Compactness)

seeds.new$Kernal.Length<- normalize(seeds.new$Kernel.Length)

seeds.new$Kernal.Width<- normalize(seeds.new$Kernel.Width)

seeds.new$Asymmetry.Coeff<- normalize(seeds.new$Asymmetry.Coeff)

seeds.new$Kernal.Groove<- normalize(seeds.new$Kernel.Groove)

seeds.new<- seeds.new[,c(1,2,3,6,8,9,10)]

head(seeds.new)
```

```
head(seeds.class)
result<- kmeans(seeds.new,3)
result$size # gives no. of records in each cluster
result$centers # gives value of cluster center datapoint value(3 centers for k=3)
result$cluster
par(mfrow=c(2,2), mar=c(5,4,2,2))
plot(seeds.new[c(1,2)], col=result$cluster)
plot(seeds.new[c(1,2)], col=seeds.class)
plot(seeds.new[c(3,4)], col=result$cluster)
plot(seeds.new[c(3,4)], col=seeds.class)
plot(seeds.new[c(5,6)], col=result$cluster)
plot(seeds.new[c(5,6)], col=seeds.class)
result$cluster <- as.factor(result$cluster)
library(ggplot2)
ggplot(seeds.new, aes(Kernal.Length, Kernal.Width, color = result$cluster)) + geom_point()
plot(seeds.new[,], col=result$cluster)
table(result$cluster,seeds.class)
library(animation)
km1<-kmeans.ani(seeds.new,3)
library(factoextra)
fviz_cluster(result, data = seeds.new)
k2 <- kmeans(seeds.new, centers = 2, nstart = 25)
k3 <- kmeans(seeds.new, centers = 3, nstart = 25)
```

```
k4 <- kmeans(seeds.new, centers = 4, nstart = 25)

k5 <- kmeans(seeds.new, centers = 5, nstart = 25)

p1 <- fviz_cluster(k2, geom = "point", data = seeds.new) + ggtitle("k = 2")

p2 <- fviz_cluster(k3, geom = "point", data = seeds.new) + ggtitle("k = 3")

p3 <- fviz_cluster(k4, geom = "point", data = seeds.new) + ggtitle("k = 4")

p4 <- fviz_cluster(k5, geom = "point", data = seeds.new) + ggtitle("k = 5")

library(gridExtra)

grid.arrange(p1, p2, p3, p4, nrow = 2)
```

Console Output:

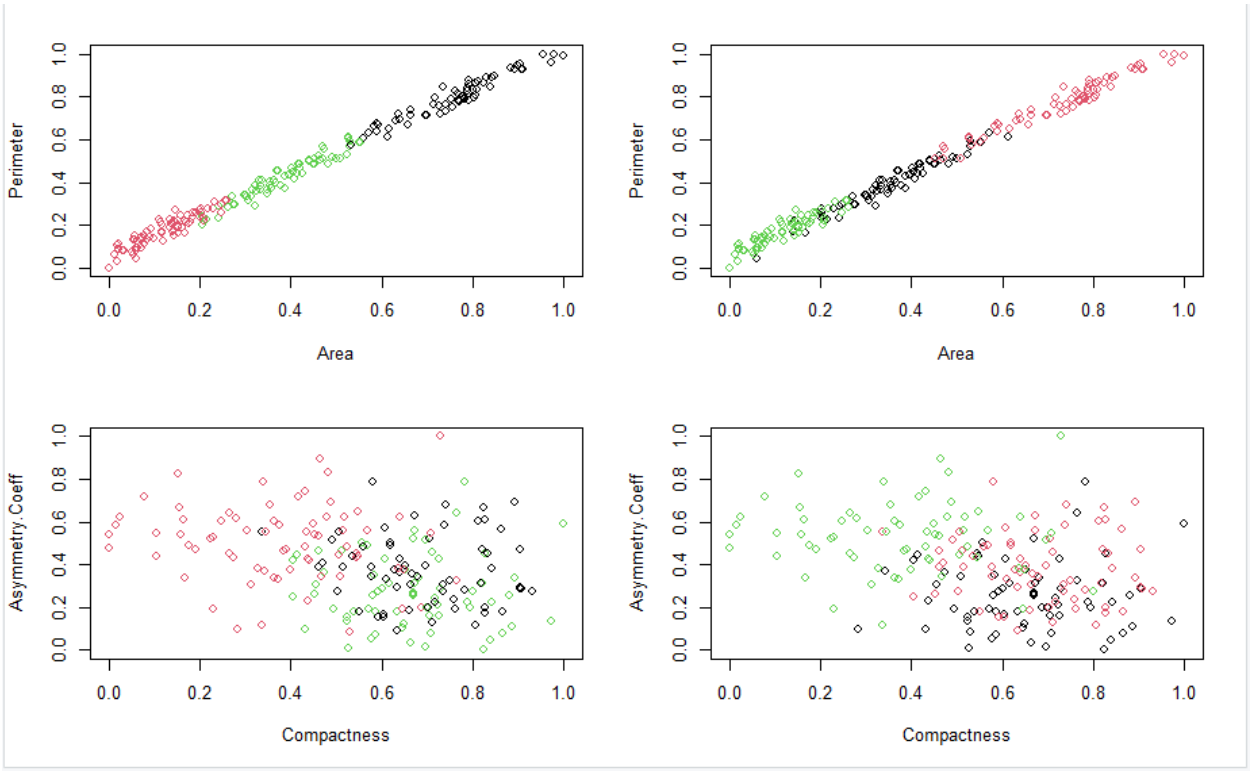
```
> #require("datasets")
> #data("seeds")
> seeds <- read.csv("D:/ViT/Sem 6/Data viz/Lab/Lab2/seeds.csv")
> str(seeds)
'data.frame':   199 obs. of  8 variables:
 $ Area           : num  15.3 14.9 14.3 13.8 16.1 ...
 $ Perimeter      : num  14.8 14.6 14.1 13.9 15 ...
 $ Compactness    : num  0.871 0.881 0.905 0.895 0.903 ...
 $ Kernel.Length  : num  5.76 5.55 5.29 5.32 5.66 ...
 $ Kernel.Width   : num  3.31 3.33 3.34 3.38 3.56 ...
 $ Asymmetry.Coeff: num  2.22 1.02 2.7 2.26 1.35 ...
 $ Kernel.Groove  : num  5.22 4.96 4.83 4.8 5.17 ...
 $ Type           : int  1 1 1 1 1 1 1 1 1 1 ...
> summary(seeds)
      Area          Perimeter      Compactness     Kernel.Length    Kernel.Width
 Min.   :10.59   Min.   :12.41   Min.   :0.8081   Min.   :4.899   Min.   :2.630
 1st Qu.:12.33   1st Qu.:13.47   1st Qu.:0.8571   1st Qu.:5.267   1st Qu.:2.954
 Median :14.43   Median :14.37   Median :0.8734   Median :5.541   Median :3.245
 Mean   :14.92   Mean   :14.60   Mean   :0.8708   Mean   :5.643   Mean   :3.266
 3rd Qu.:17.45   3rd Qu.:15.80   3rd Qu.:0.8868   3rd Qu.:6.002   3rd Qu.:3.564
 Max.   :21.18   Max.   :17.25   Max.   :0.9183   Max.   :6.675   Max.   :4.033
 Asymmetry.Coeff  Kernel.Groove        Type
 Min.   :0.7651   Min.   :4.519   Min.   :1.000
 1st Qu.:2.5700   1st Qu.:5.046   1st Qu.:1.000
 Median :3.6310   Median :5.228   Median :2.000
 Mean   :3.6992   Mean   :5.421   Mean   :1.995
 3rd Qu.:4.7990   3rd Qu.:5.879   3rd Qu.:3.000
 Max.   :8.3150   Max.   :6.550   Max.   :3.000
> head(seeds)
   Area Perimeter Compactness Kernel.Length Kernel.Width Asymmetry.Coeff Kernel.Groove
1 15.26     14.84      0.8710         5.763        3.312           2.221         5.220
2 14.88     14.57      0.8811         5.554        3.333           1.018         4.956
3 14.29     14.09      0.9050         5.291        3.337           2.699         4.825
4 13.84     13.94      0.8955         5.324        3.379           2.259         4.805
5 16.14     14.99      0.9034         5.658        3.562           1.355         5.175
6 14.38     14.21      0.8951         5.386        3.312           2.462         4.956
  Type
```

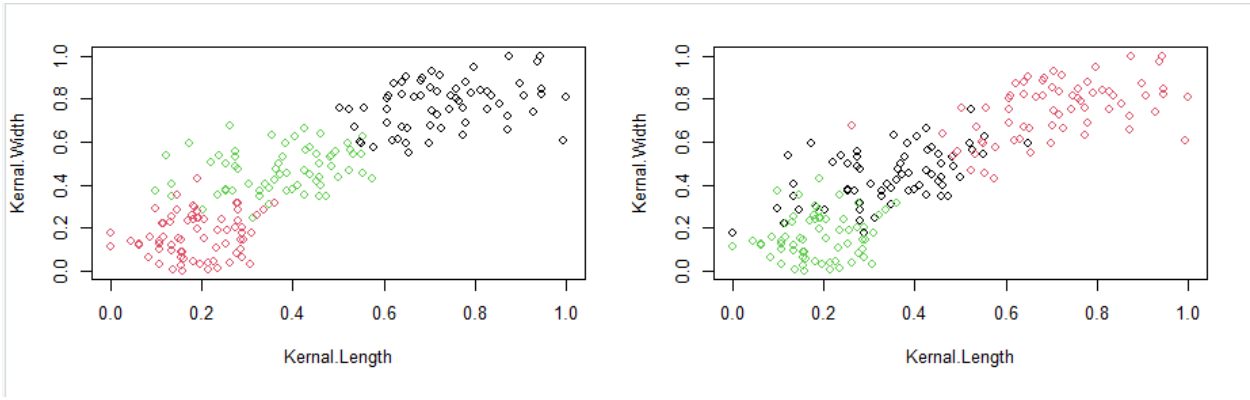
```
  Type
1    1
2    1
3    1
4    1
5    1
6    1
> seeds.new<- seeds[,c(1,2,3,4,5,6,7)]
> seeds.class<- seeds[,"Type"]
> head(seeds.new)
   Area Perimeter Compactness Kernel.Length Kernel.Width Asymmetry.Coeff Kernel.Groove
1 15.26     14.84      0.8710         5.763        3.312           2.221         5.220
2 14.88     14.57      0.8811         5.554        3.333           1.018         4.956
3 14.29     14.09      0.9050         5.291        3.337           2.699         4.825
4 13.84     13.94      0.8955         5.324        3.379           2.259         4.805
5 16.14     14.99      0.9034         5.658        3.562           1.355         5.175
6 14.38     14.21      0.8951         5.386        3.312           2.462         4.956
> head(seeds.class)
[1] 1 1 1 1 1 1
> normalize <- function(x){
+   return ((x-min(x))/(max(x)-min(x)))
+ }
> seeds.new$Area<- normalize(seeds.new$Area)
> seeds.new$Perimeter<- normalize(seeds.new$Perimeter)
> seeds.new$Compactness<- normalize(seeds.new$Compactness)
> seeds.new$Kernal.Length<- normalize(seeds.new$Kernel.Length)
> seeds.new$Kernal.Width<- normalize(seeds.new$Kernel.Width)
> seeds.new$Asymmetry.Coeff<- normalize(seeds.new$Asymmetry.Coeff)
> seeds.new$Kernal.Groove<- normalize(seeds.new$Kernel.Groove)
> seeds.new<- seeds.new[,c(1,2,3,6,8,9,10)]
> head(seeds.new)
       Area Perimeter Compactness Asymmetry.Coeff Kernal.Length Kernal.Width
1 0.4409821 0.5020661   0.5707804      0.19283699     0.4864865    0.4861012
2 0.4050992 0.4462810   0.6624319      0.03349713     0.3688063    0.5010691
3 0.3493862 0.3471074   0.8793103      0.25614909     0.2207207    0.5039202
4 0.3068933 0.3161157   0.7931034      0.19787017     0.2393018    0.5338560
5 0.5240793 0.5330579   0.8647913      0.07813349     0.4273649    0.6642908
6 0.3578848 0.3719008   0.7894737      0.22475794     0.2742117    0.4861012
```

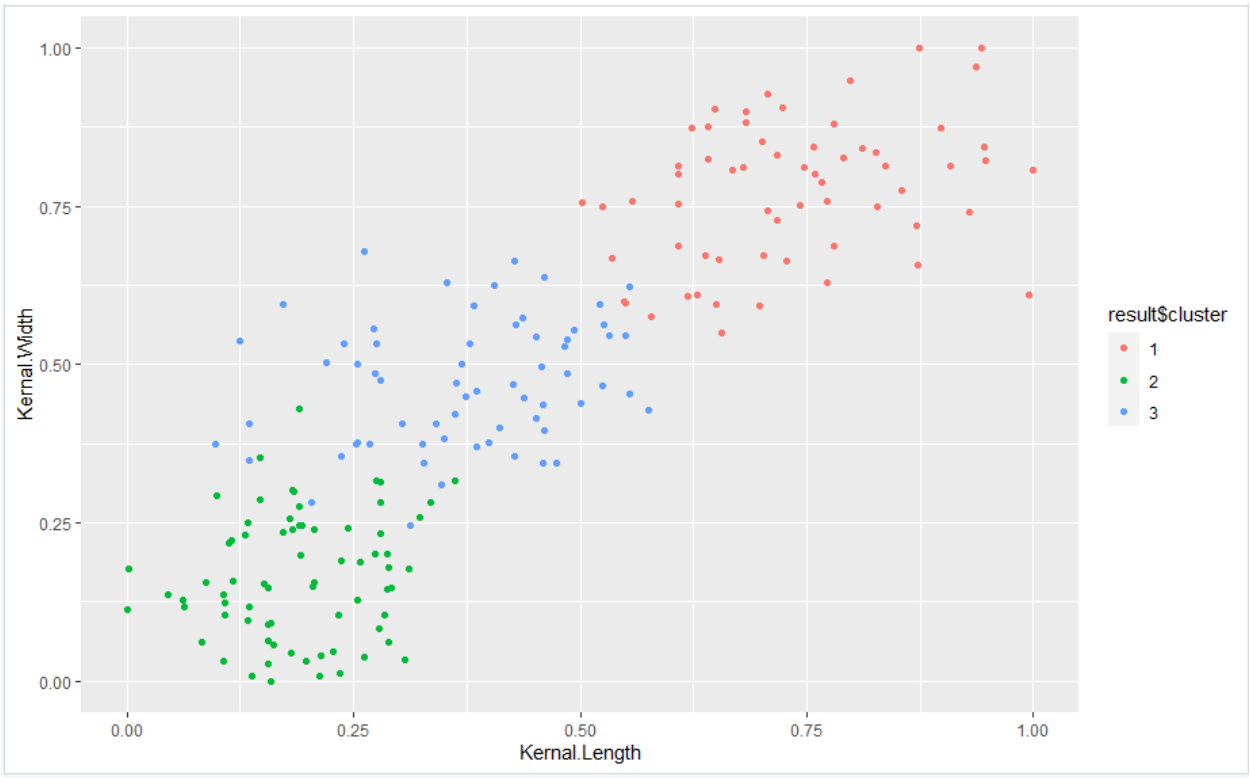
```
  Kernal.Groove
1     0.3451502
2     0.2151649
3     0.1506647
4     0.1408173
5     0.3229936
6     0.2151649
> head(seeds.class)
[1] 1 1 1 1 1 1
> result<- kmeans(seeds.new,3)
> result$size # gives no. of records in each cluster
[1] 62 72 65
> result$centers # gives value of cluster center datapoint value(3 centers for k=3)
       Area Perimeter Compactness Asymmetry.Coeff Kernal.Length Kernal.Width
1 0.7606994 0.7970874   0.6940753       0.3746018     0.7336167    0.7720093
2 0.1257344 0.1781451   0.3777349       0.5028705     0.1902997    0.1642116
3 0.3865621 0.4250159   0.6617479       0.2754671     0.3722453    0.4685454
  Kernal.Groove
1     0.7580328
2     0.2832075
3     0.3224028
> result$cluster
  [1] 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 2 3 3 3 2 3 3 2 2 3 3 3 3 3 3 3 1 3 3 3 3 3
 [42] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 2 3 3 3 3 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [83] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 3 1 1 1
[124] 1 1 1 1 3 3 3 1 3 3 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[165] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 3 2 2 2 2 2 2 2 2
> par(mfrow=c(2,2), mar=c(5,4,2,2))
> plot(seeds.new[c(1,2)], col=result$cluster)
> plot(seeds.new[c(1,2)], col=seeds.class)
> plot(seeds.new[c(3,4)], col=result$cluster)
> plot(seeds.new[c(3,4)], col=seeds.class)
> plot(seeds.new[c(5,6)], col=result$cluster)
> plot(seeds.new[c(5,6)], col=seeds.class)
> result$cluster <- as.factor(result$cluster)
> library(ggplot2)
> ggplot(seeds.new, aes(Kernal.Length, Kernal.Width, color = result$cluster)) + geom_point()
```

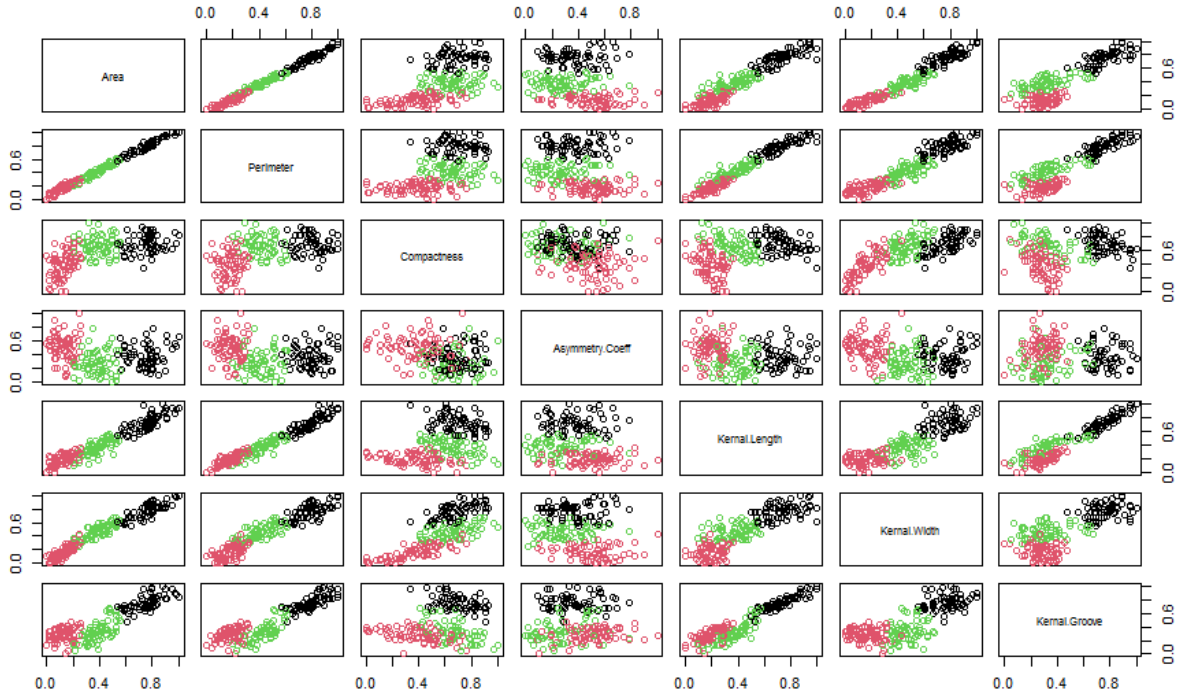
```
> ggplot(seeds.new, aes(Kernal.Length, Kernal.Width, color = result$cluster)) + geom_point()
> plot(seeds.new[,], col=result$cluster)
> table(result$cluster,seeds.class)
   seeds.class
     1  2  3
  1  2 60  0
  2  9  0 63
  3 55  8  2
> library(animation)
> km1<-kmeans.ani(seeds.new,3)
> library(factoextra)
Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
> fviz_cluster(result, data = seeds.new)
> k2 <- kmeans(seeds.new, centers = 2, nstart = 25)
> k3 <- kmeans(seeds.new, centers = 3, nstart = 25)
> k4 <- kmeans(seeds.new, centers = 4, nstart = 25)
> k5 <- kmeans(seeds.new, centers = 5, nstart = 25)
> p1 <- fviz_cluster(k2, geom = "point", data = seeds.new) + ggtitle("k = 2")
> p2 <- fviz_cluster(k3, geom = "point", data = seeds.new) + ggtitle("k = 3")
> p3 <- fviz_cluster(k4, geom = "point", data = seeds.new) + ggtitle("k = 4")
> p4 <- fviz_cluster(k5, geom = "point", data = seeds.new) + ggtitle("k = 5")
> library(gridExtra)
> grid.arrange(p1, p2, p3, p4, nrow = 2)
>
```

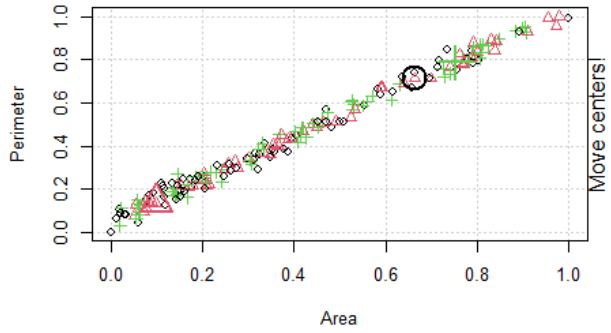
Plot Output:

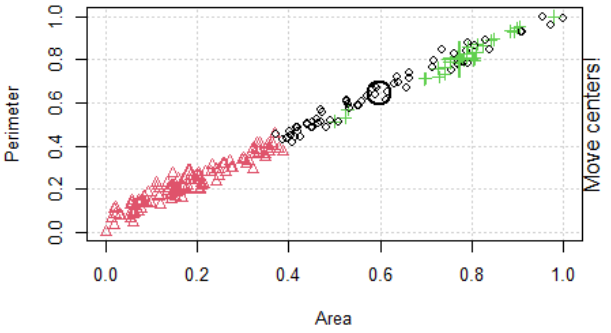

Perimeter
Area
Perimeter
Area
Asymmetry.Coeff
Compactness
Asymmetry.Coeff
Compactness

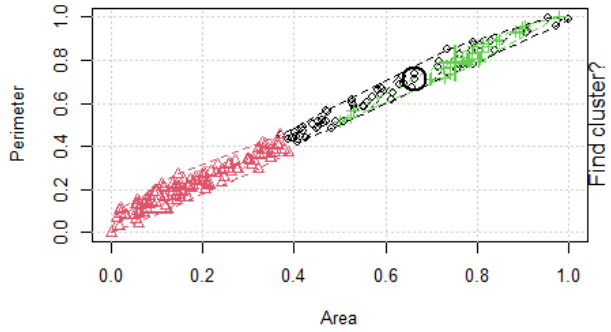

Kernal.Width
Kernal.Length
Kernal.Width
Kernal.Length

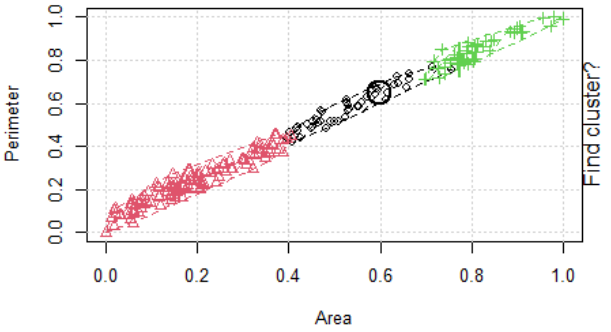

1.00
0.75
0.50
0.25
0.00
Kernal.Width
0.00
0.25
0.50
0.75
1.00
Kernal.Length
result$cluster
1
2
3

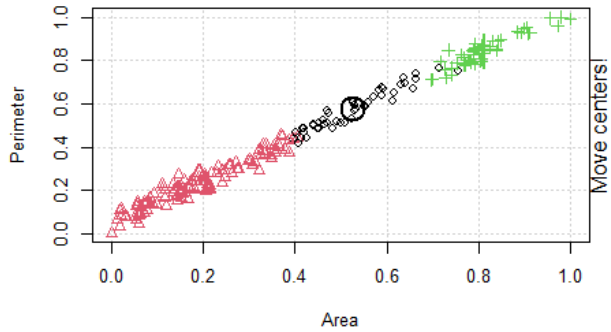

Area
Perimeter
Compactness
Asymmetry.Coeff
Kernal.Length
Kernal.Width
Kernal.Groove
0.0
0.4
0.8
0.0
0.6

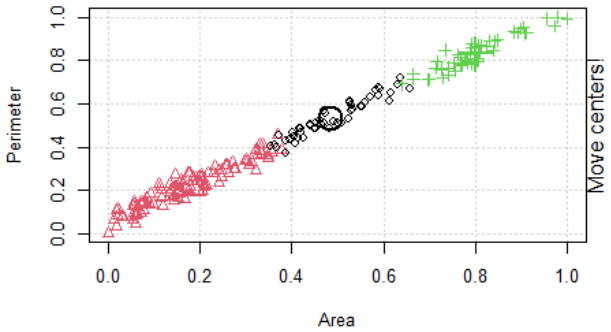

Perimeter
Area
Move centers!

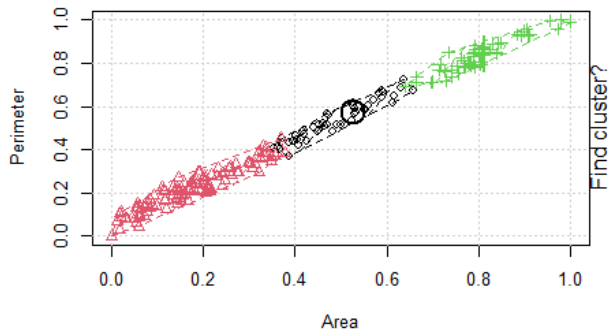

Perimeter
Area
Move centers!

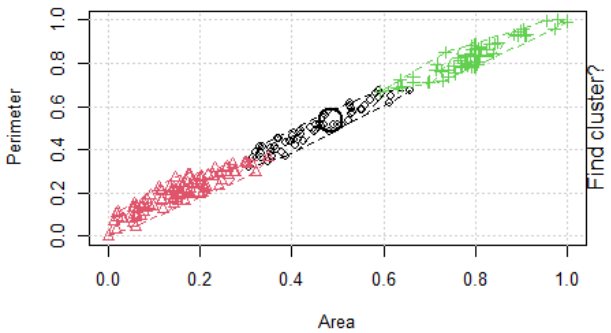

Perimeter
Area
Find cluster?

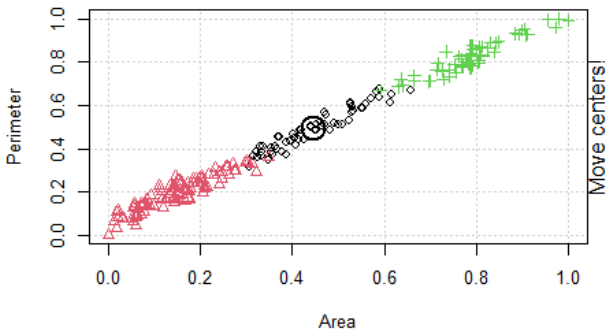

Perimeter
Area
Find cluster?

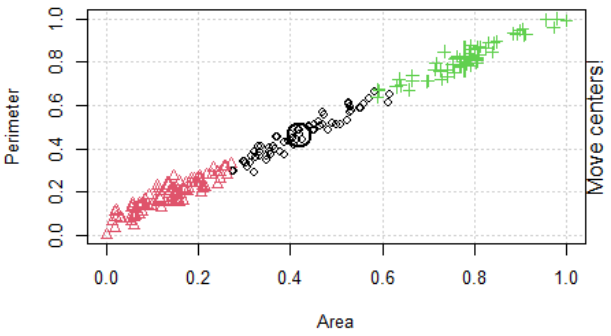

Perimeter
Area
Move centers!

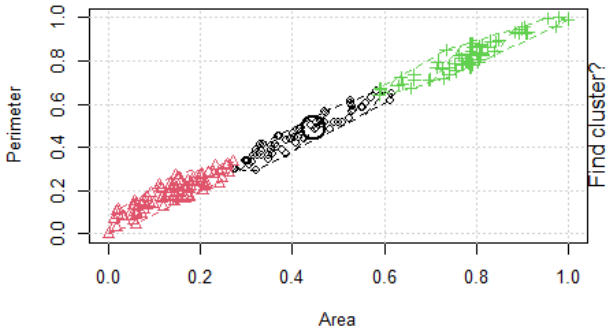

Perimeter
Area
Move centers!

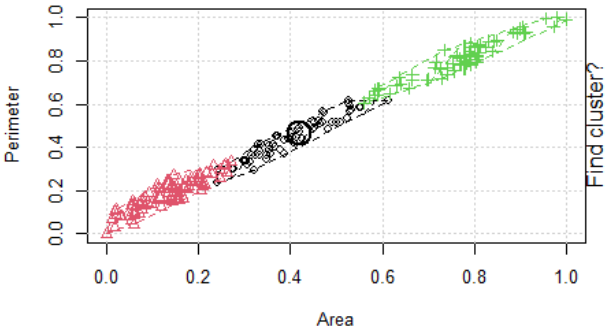

Perimeter
Area
Find cluster?

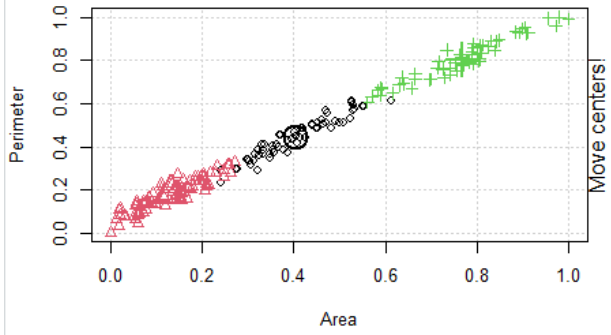

Perimeter
Area
Find cluster?

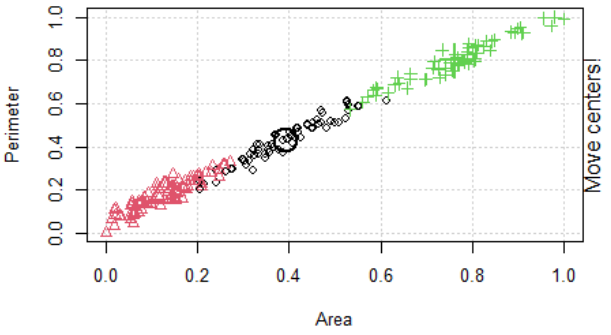

Perimeter
Area
Move centers!
0.0
0.2
0.4
0.6
0.8
1.0

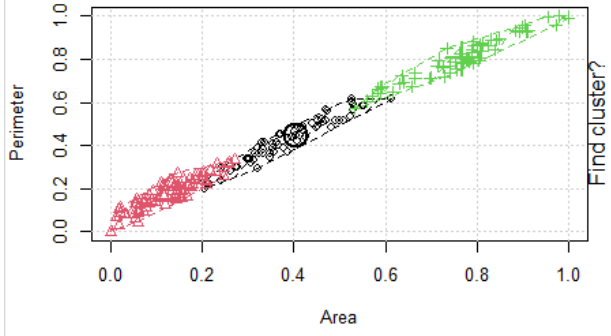

Perimeter
Area
Move centers!
0.0
0.2
0.4
0.6
0.8
1.0

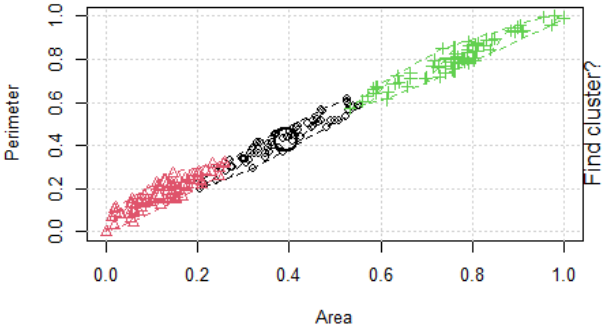

Perimeter
Area
Find cluster?
0.0
0.2
0.4
0.6
0.8
1.0

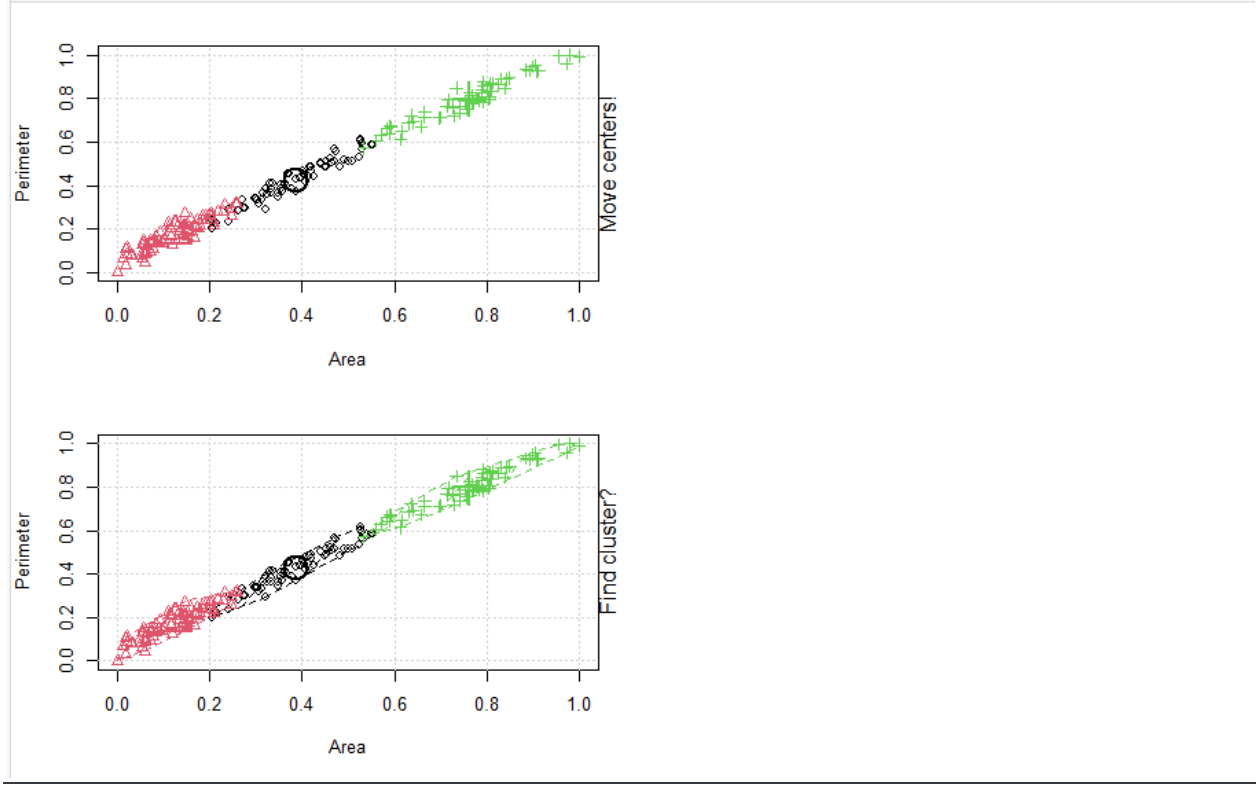

Perimeter
Area
Find cluster?
0.0
0.2
0.4
0.6
0.8
1.0

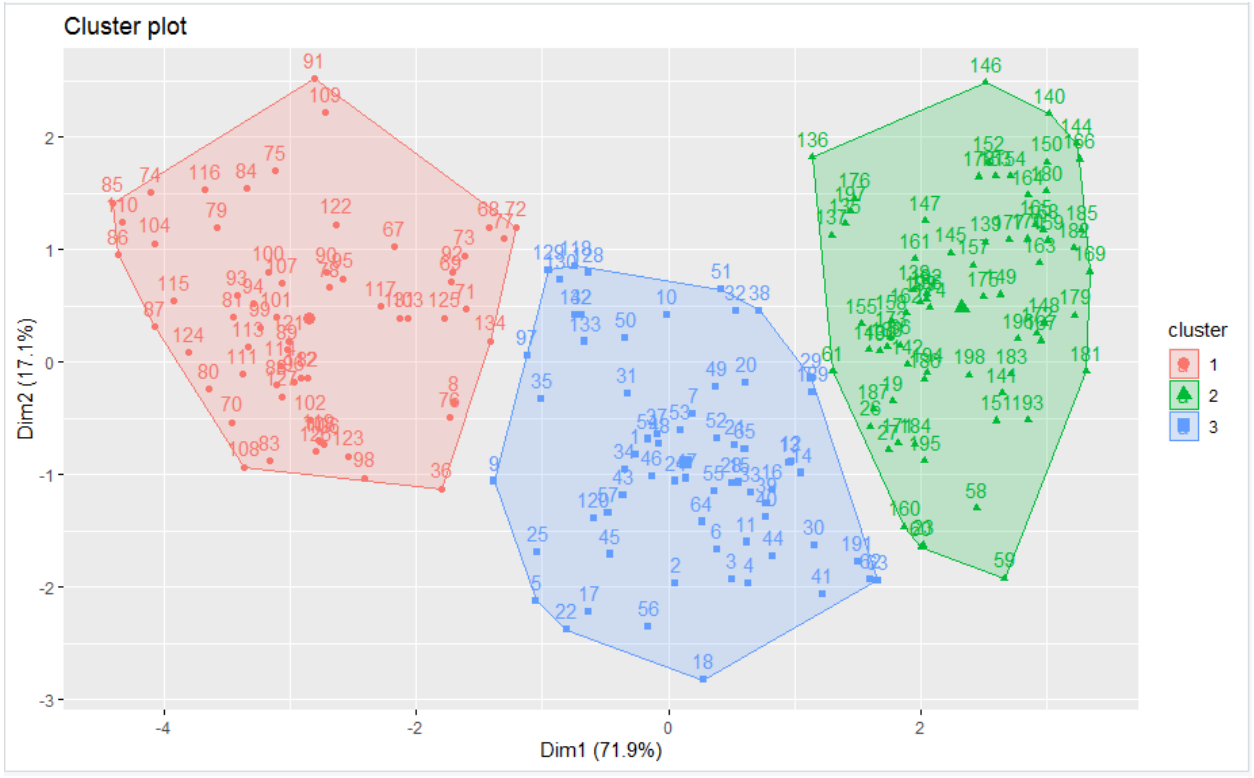

Perimeter
Area
Move centers!
0.0
0.2
0.4
0.6
0.8
1.0

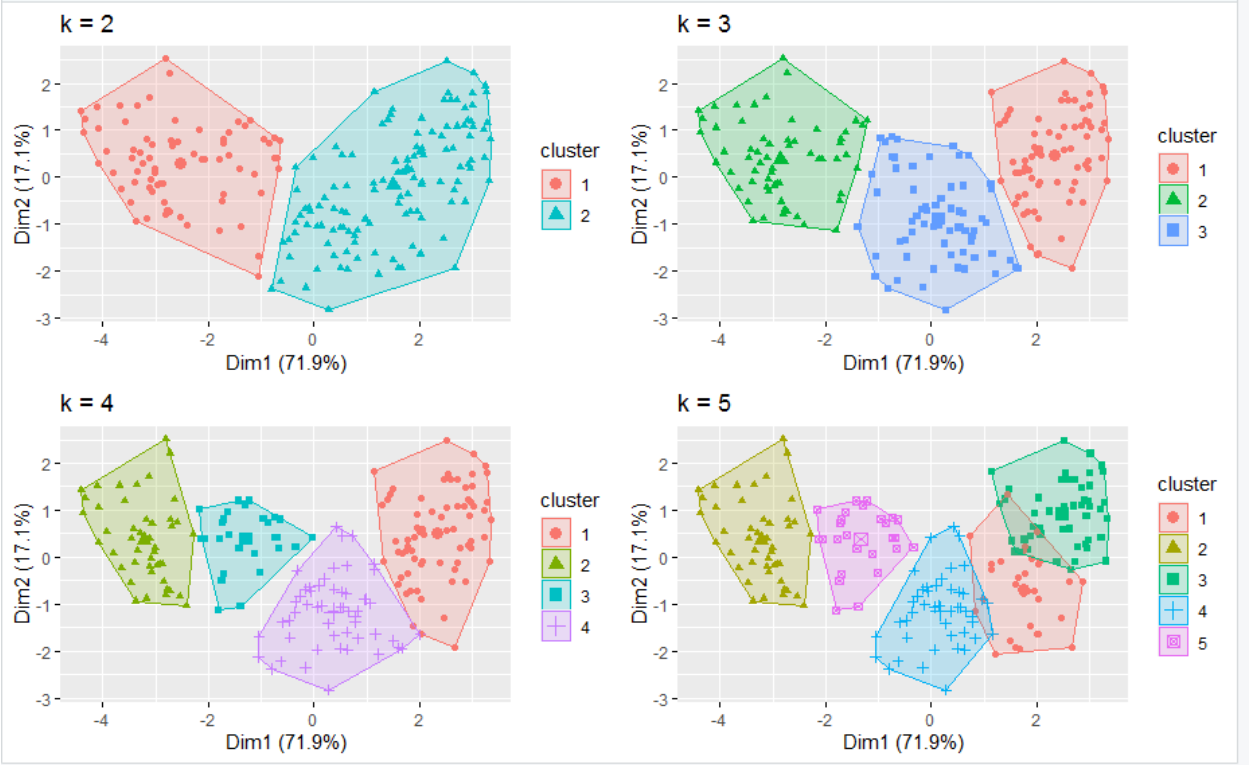

Perimeter
Area
Move centers!
0.0
0.2
0.4
0.6
0.8
1.0


Perimeter
Area
Find cluster?
0.0
0.2
0.4
0.6
0.8
1.0


Perimeter
Area
Find cluster?
0.0
0.2
0.4
0.6
0.8
1.0


Perimeter
Area
Move centers!
Find cluster?


Cluster plot
Dim1 (71.9%)
Dim2 (17.1%)
cluster
1
2
3


k = 2
k = 3
k = 4
k = 5
Dim1 (71.9%)
Dim2 (17.1%)
cluster
1
2
3
4
5