**TUGAS MINGGU KE-13**

**STATISTIKA DESKRIPTIF**



**NAMA : MUKHAMAD IKHSANUDIN**

**NIM : 082011633086**

**PROGRAM STUDI S1 SISTEM INFORMASI**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS AIRLANGGA**

**2021**

Tugas pertemuan 26 🡪 dikumpulkan hari ini, tgl. 11-06-2021 jam 23.59 🡪 di upload ke Aula dan kirim ke email eto-w@fst.unair .ac.id dengan subject : Clustering

Code dan outputnya jadikan satu di notebook R-nya

1. Carilah **3 dataset** yang sesuai untuk Clustering kemudian gunakan metode :
2. 5 metode agglomerative (centroid, single-linkage, complete-linkage, average-linkage dan ward)
3. Tentukan nilai k (banyaknya klaster) berdasarkan dendogramnya
4. Tentukan nilai k (banyaknya klaster) berdasarkan nilai BIC atau yang lain
5. Buatlah table rekapitulasi

|  |  |  |  |
| --- | --- | --- | --- |
| No | Metode | Nomer klaster | Anggota klaster |
| 1 | Centroid | 1 |  |
|  |  | 2 |  |
|  |  |  |  |
|  |  | k |  |
| 2 | single-linkage |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. Gunakan no 1 untuk metode-metode berikut :
2. K-means dan buatlah table rekapitulasi seperti no 1d.
3. K-medoids dan buatlah table rekapitulasi seperti no 1d.
4. K-medians dan buatlah table rekapitulasi seperti no 1d.

==============================================================================

Code ditaruh diantara tanda berikut :

```{R}

Syntax di sini

```

Carilah **3 dataset** yang sesuai untuk Clustering kemudian gunakan metode :

1. **5 metode agglomerative (centroid, single-linkage, complete-linkage, average-linkage dan ward)**

library(tidyverse)

library(cluster.datasets)

library(factoextra)

```{r}

# Dataset 1

data("nutrient")

DataClus1 <- nutrient

summary(DataClus1)

DataClus1\_Fix <- scale(DataClus1)

DataClus1\_Fix

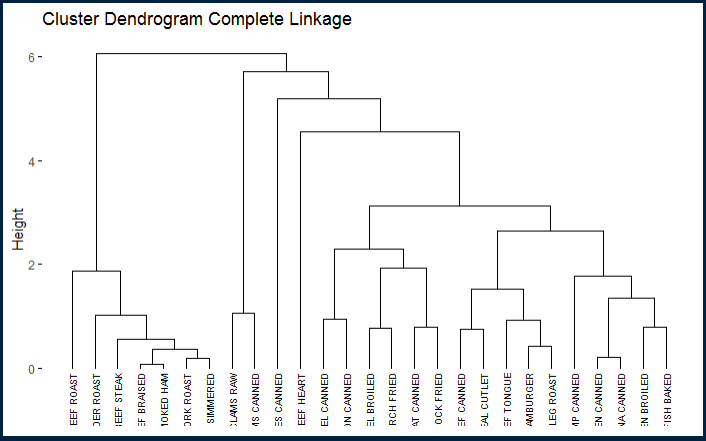
Clus1Dist <- dist(x = DataClus1\_Fix, method = "euclidean")

# Complete Linkage

Clus1\_Com <- hclust(d = Clus1Dist, method = "complete")

Clus1\_Com

fviz\_dend(Clus1\_Com, cex = 0.5, main = "Cluster Dendrogram Complete Linkage")

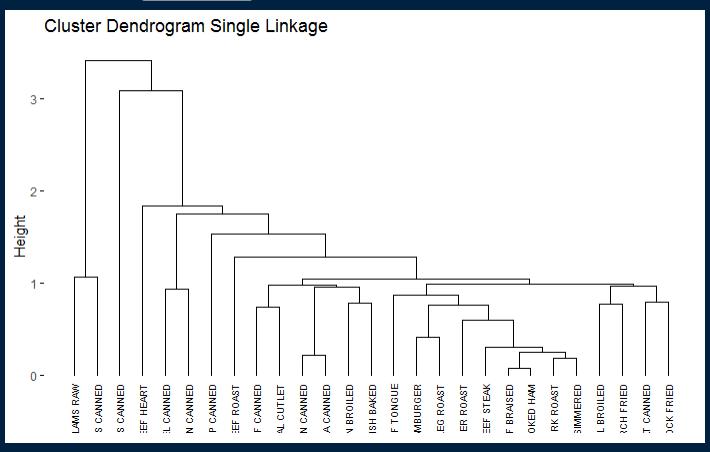


# Single Linkage

Clus1\_Sin <- hclust(d = Clus1Dist, method = "single")

Clus1\_Sin

fviz\_dend(Clus1\_Sin, cex = 0.5, main = "Cluster Dendrogram Single Linkage")

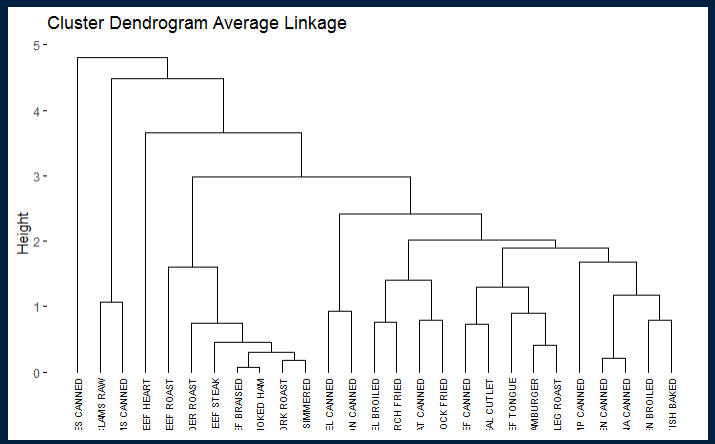


# Average Linkage

Clus1\_Ave <- hclust(d = Clus1Dist, method = "average")

Clus1\_Ave

fviz\_dend(Clus1\_Ave, cex = 0.5, main = "Cluster Dendrogram Average Linkage")

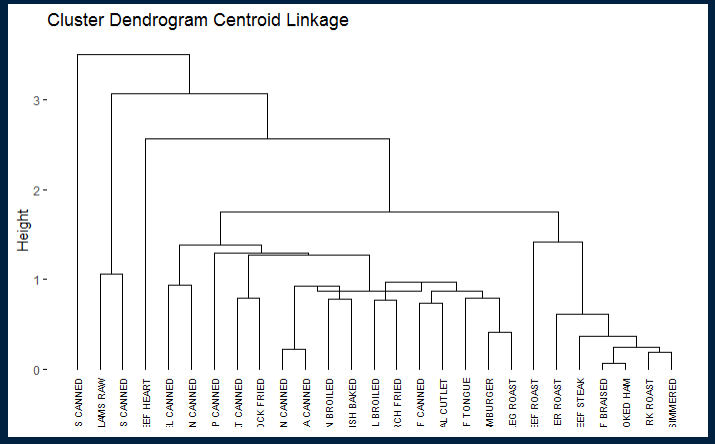


# Centroid Linkage

Clus1\_Cen <- hclust(d = Clus1Dist, method = "centroid")

Clus1\_Cen

fviz\_dend(Clus1\_Cen, cex = 0.5, main = "Cluster Dendrogram Centroid Linkage")

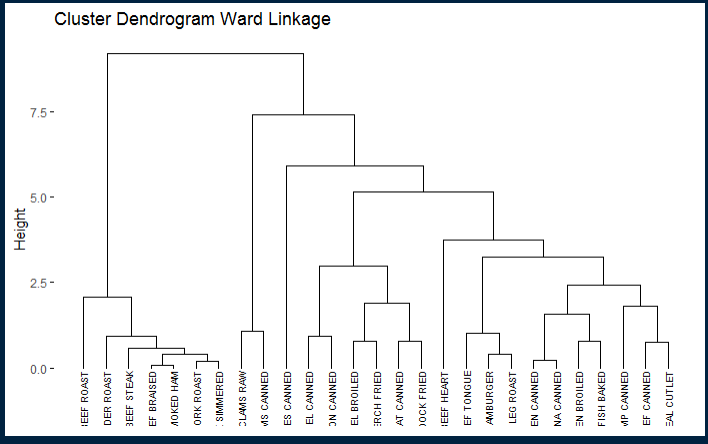


# Ward Linkage

Clus1\_War <- hclust(d = Clus1Dist, method = "ward.D2")

Clus1\_War

fviz\_dend(Clus1\_War, cex = 0.5, main = "Cluster Dendrogram Ward Linkage")



```

```{R}

#Dataset 2

data("achieve")

DataClus2 <- achieve

summary(DataClus2)

DataClus2\_Fix <- scale(DataClus2)

DataClus2\_Fix

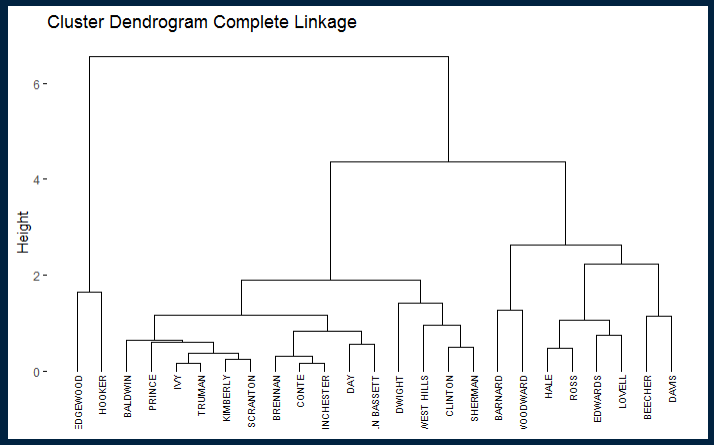
Clus2Dist <- dist(x = DataClus2\_Fix, method = "euclidean")

# Complete Linkage

Clus2\_Com <- hclust(d = Clus2Dist, method = "complete")

Clus2\_Com

fviz\_dend(Clus2\_Com, cex = 0.5, main = "Cluster Dendrogram Complete Linkage")

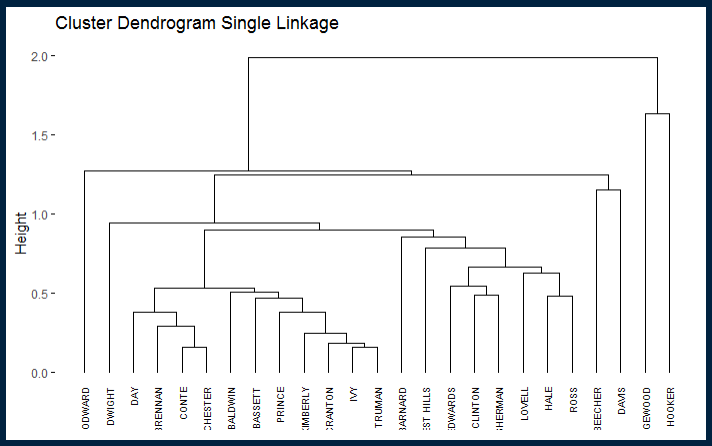


# Single Linkage

Clus2\_Sin <- hclust(d = Clus2Dist, method = "single")

Clus2\_Sin

fviz\_dend(Clus2\_Sin, cex = 0.5, main = "Cluster Dendrogram Single Linkage")

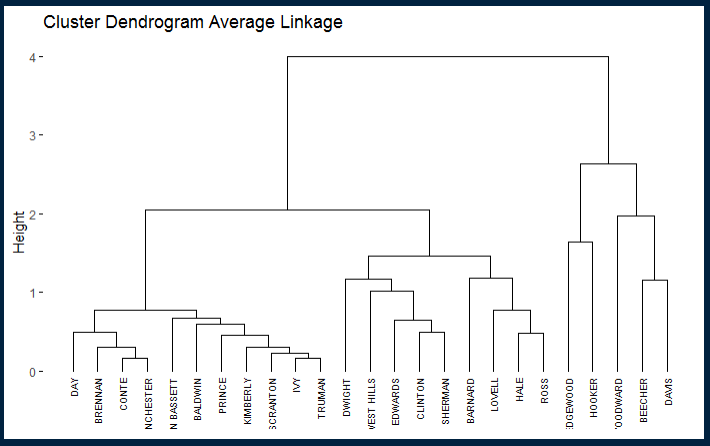


# Average Linkage

Clus2\_Ave <- hclust(d = Clus2Dist, method = "average")

Clus2\_Ave

fviz\_dend(Clus2\_Ave, cex = 0.5, main = "Cluster Dendrogram Average Linkage")

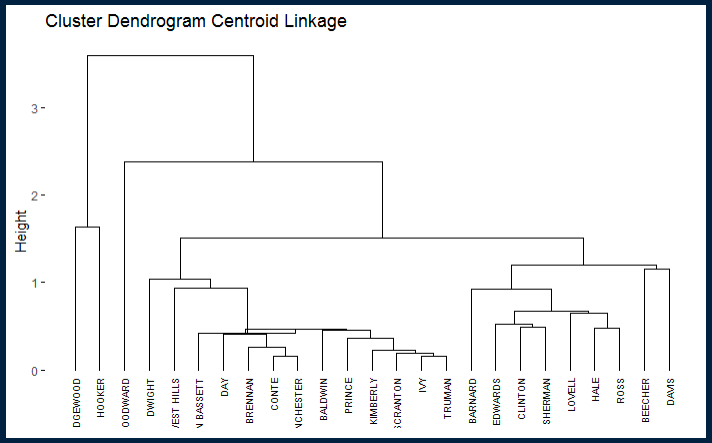


# Centroid Linkage

Clus2\_Cen <- hclust(d = Clus2Dist, method = "centroid")

Clus2\_Cen

fviz\_dend(Clus2\_Cen, cex = 0.5, main = "Cluster Dendrogram Centroid Linkage")

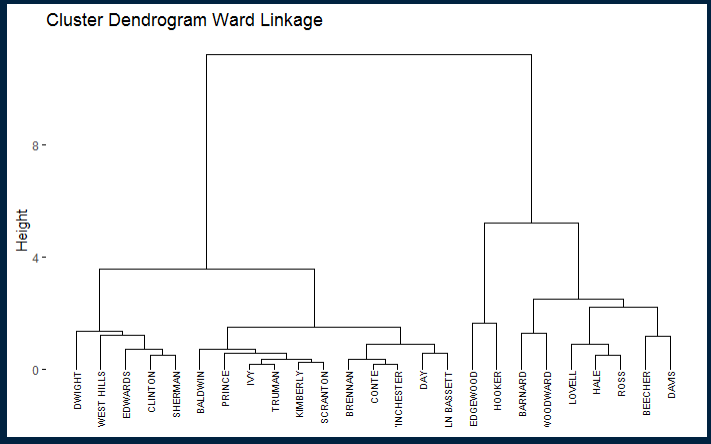


# Ward Linkage

Clus2\_War <- hclust(d = Clus2Dist, method = "ward.D2")

Clus2\_War

fviz\_dend(Clus2\_War, cex = 0.5, main = "Cluster Dendrogram Ward Linkage")



```

```{R}

#Dataset 3

data("milk")

DataClus3 <- milk

summary(DataClus3)

DataClus3\_Fix <- scale(DataClus3)

DataClus3\_Fix

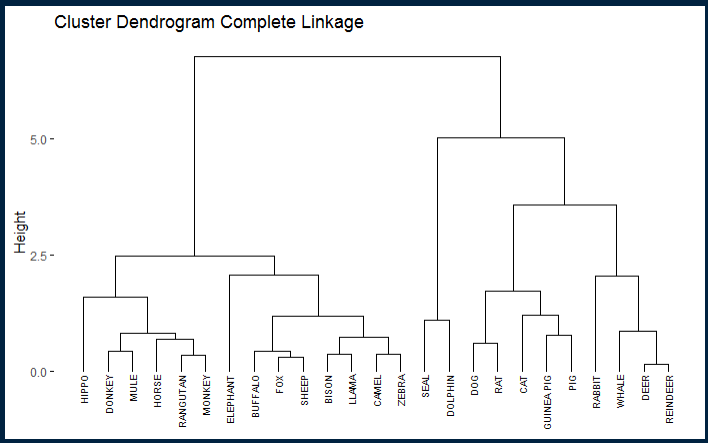
Clus3Dist <- dist(x = DataClus3\_Fix, method = "euclidean")

# Complete Linkage

Clus3\_Com <- hclust(d = Clus3Dist, method = "complete")

Clus3\_Com

fviz\_dend(Clus3\_Com, cex = 0.5, main = "Cluster Dendrogram Complete Linkage")

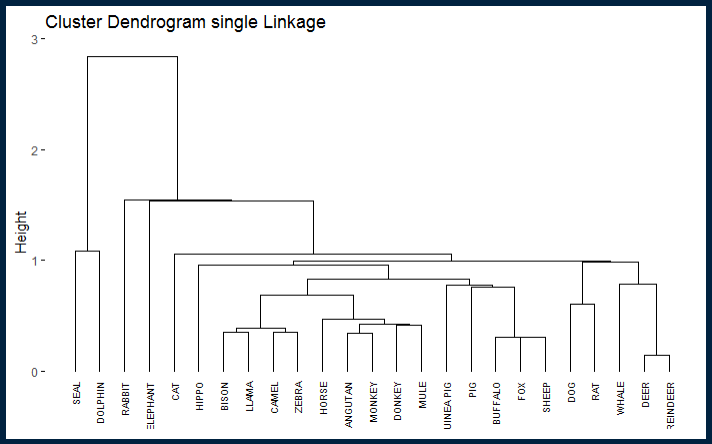


# Single Linkage

Clus3\_Sin <- hclust(d = Clus3Dist, method = "single")

Clus3\_Sin

fviz\_dend(Clus3\_Sin, cex = 0.5, main = "Cluster Dendrogram single Linkage")

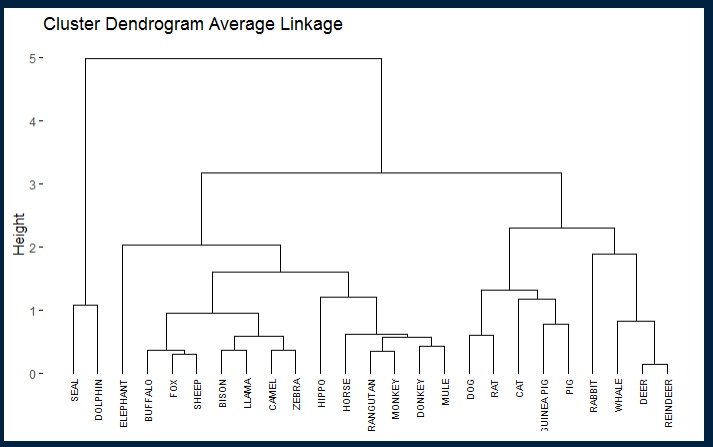


# Average Linkage

Clus3\_Ave <- hclust(d = Clus3Dist, method = "average")

Clus3\_Ave

fviz\_dend(Clus3\_Ave, cex = 0.5, main = "Cluster Dendrogram Average Linkage")

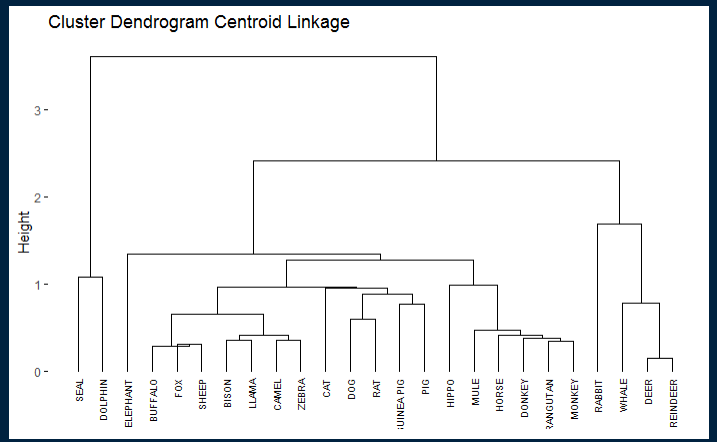


# Centroid Linkage

Clus3\_Cen <- hclust(d = Clus3Dist, method = "centroid")

Clus3\_Cen

fviz\_dend(Clus3\_Cen, cex = 0.5, main = "Cluster Dendrogram Centroid Linkage")

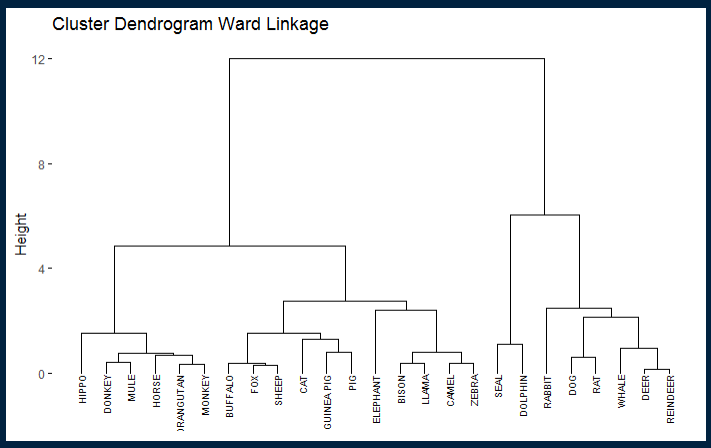


# Ward Linkage

Clus3\_War <- hclust(d = Clus3Dist, method = "ward.D2")

Clus3\_War

fviz\_dend(Clus3\_War, cex = 0.5, main = "Cluster Dendrogram Ward Linkage")



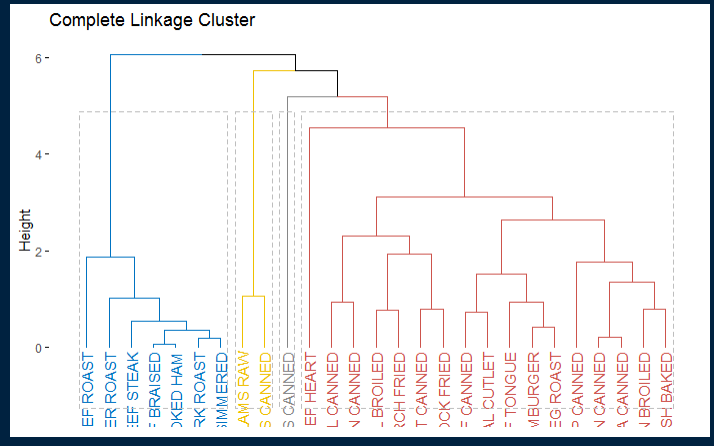
```

1. **Tentukan nilai k (banyaknya klaster) berdasarkan dendogramnya**

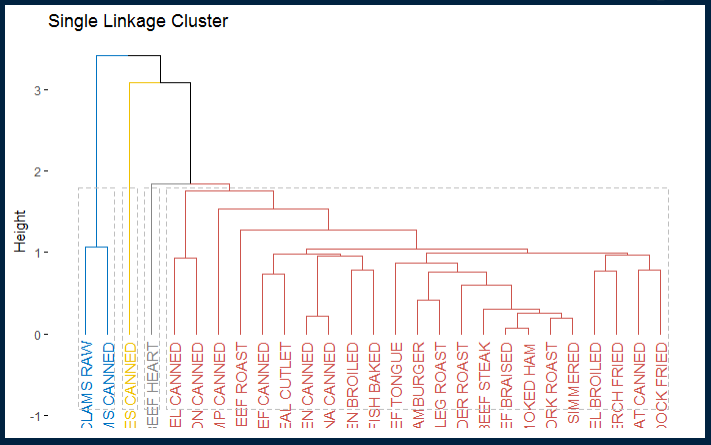
```{R}

#Dataset 1

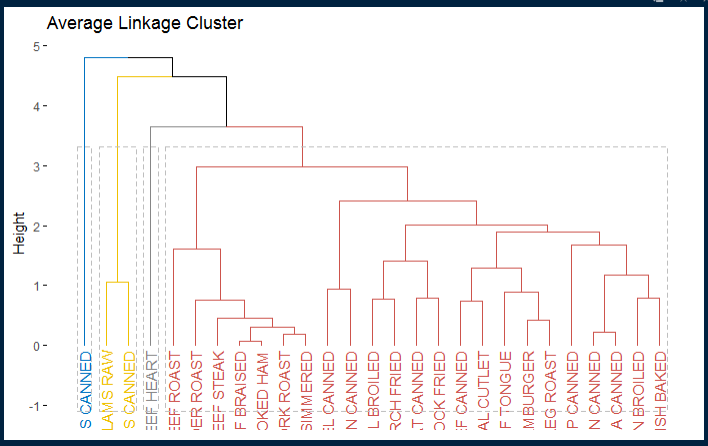
fviz\_dend(Clus1\_Com, k = 4, k\_colors = "jco", rect = T, main = "Complete Linkage Cluster")



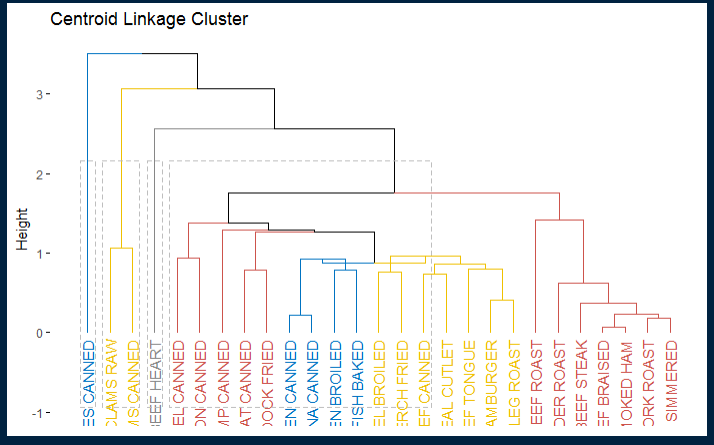
fviz\_dend(Clus1\_Sin, k = 4, k\_colors = "jco", rect = T, main = "Single Linkage Cluster")



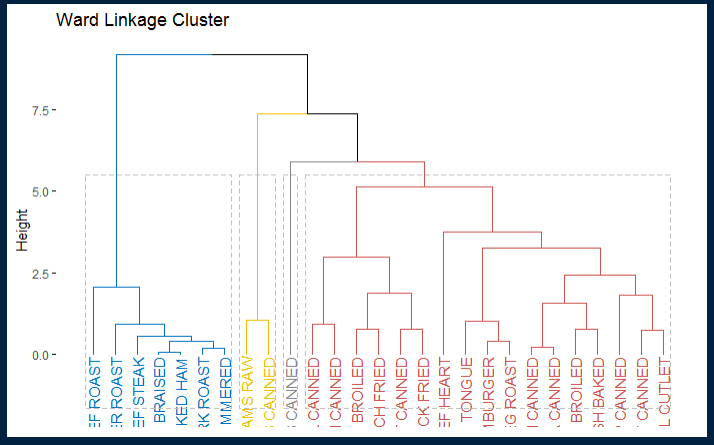
fviz\_dend(Clus1\_Ave, k = 4, k\_colors = "jco", rect = T, main = "Average Linkage Cluster")



fviz\_dend(Clus1\_Cen, k = 4, k\_colors = "jco", rect = T, main = "Centroid Linkage Cluster")



fviz\_dend(Clus1\_War, k = 4, k\_colors = "jco", rect = T, main = "Ward Linkage Cluster")

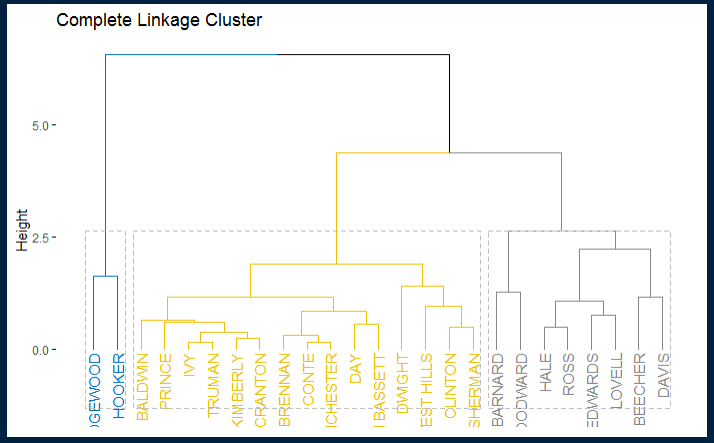


```

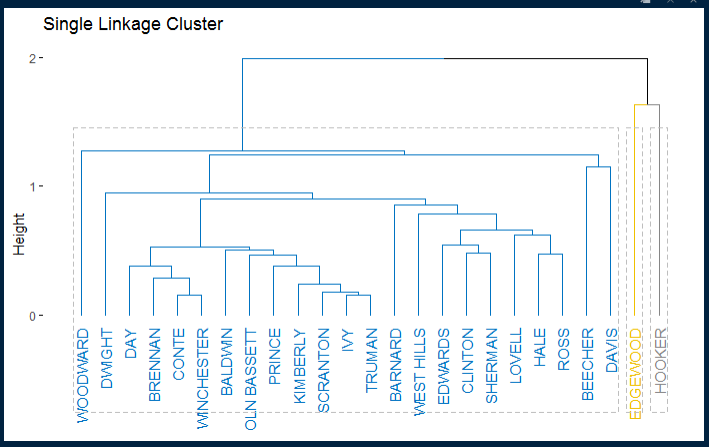
```{R}

#Dataset 2

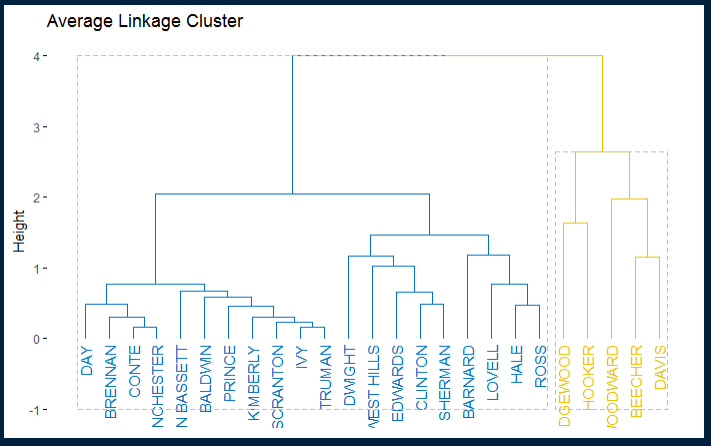
fviz\_dend(Clus2\_Com, k = 3, k\_colors = "jco", rect = T, main = "Complete Linkage Cluster")



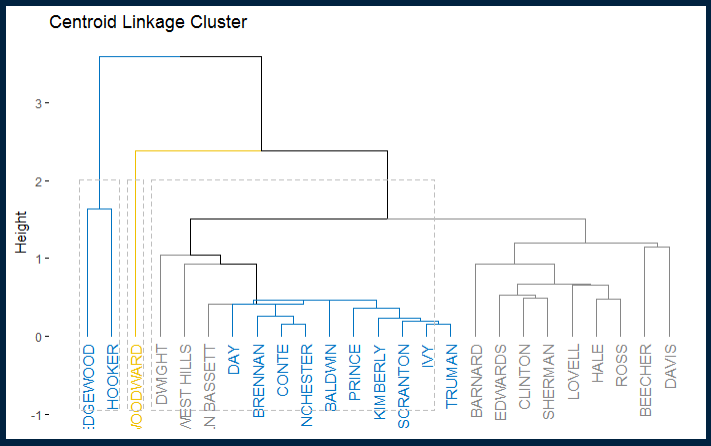
fviz\_dend(Clus2\_Sin, k = 3, k\_colors = "jco", rect = T, main = "Single Linkage Cluster")



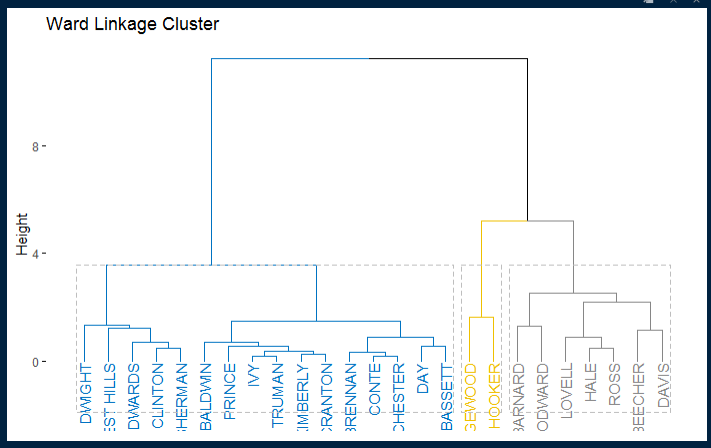
fviz\_dend(Clus2\_Ave, k = 2, k\_colors = "jco", rect = T, main = "Average Linkage Cluster")



fviz\_dend(Clus2\_Cen, k = 3, k\_colors = "jco", rect = T, main = "Centroid Linkage Cluster")



fviz\_dend(Clus2\_War, k = 3, k\_colors = "jco", rect = T, main = "Ward Linkage Cluster")

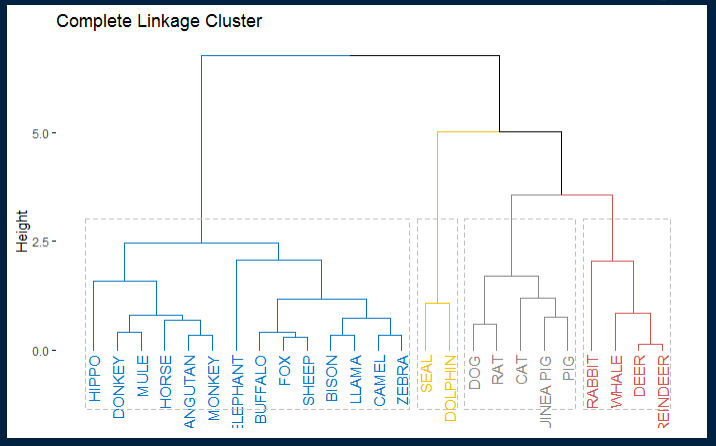


```

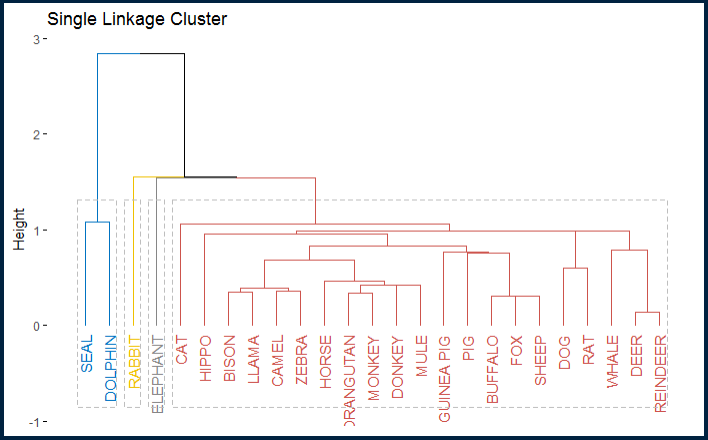
```{R}

#Dataset 3

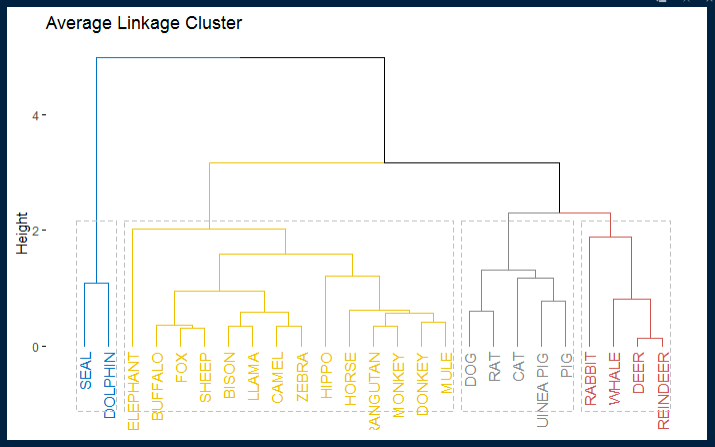
fviz\_dend(Clus3\_Com, k = 4, k\_colors = "jco", rect = T, main = "Complete Linkage Cluster")



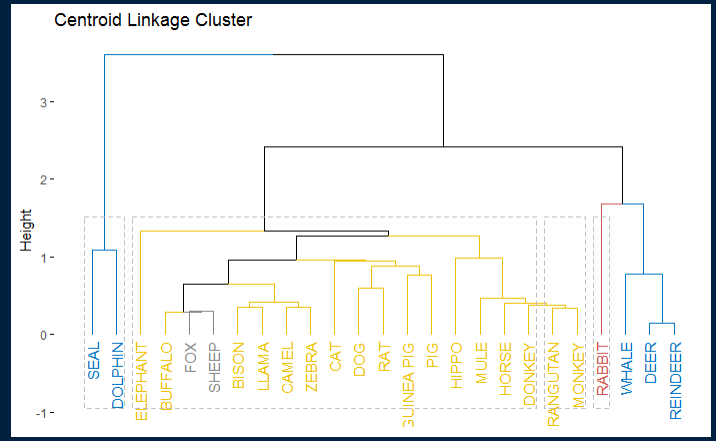
fviz\_dend(Clus3\_Sin, k = 4, k\_colors = "jco", rect = T, main = "Single Linkage Cluster")



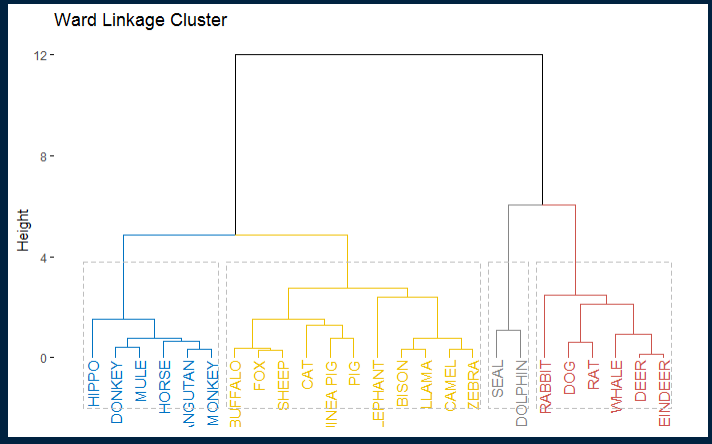
fviz\_dend(Clus3\_Ave, k = 4, k\_colors = "jco", rect = T, main = "Average Linkage Cluster")



fviz\_dend(Clus3\_Cen, k = 4, k\_colors = "jco", rect = T, main = "Centroid Linkage Cluster")



fviz\_dend(Clus3\_War, k = 4, k\_colors = "jco", rect = T, main = "Ward Linkage Cluster")



```

1. **Tentukan nilai k (banyaknya klaster) berdasarkan nilai BIC atau yang lain**

```{R}

#Dataset 1

library (mclust)

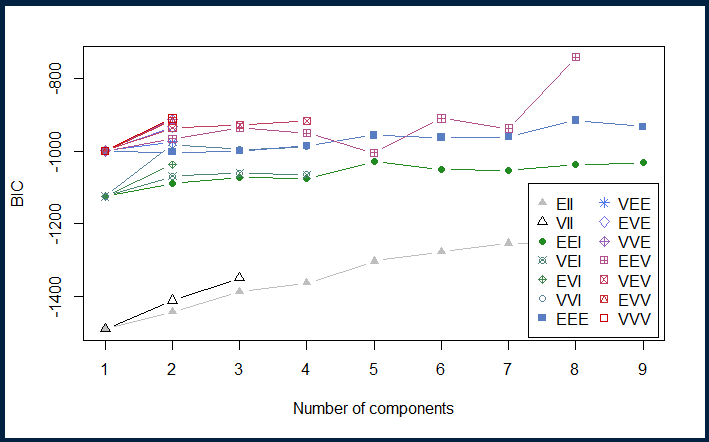
Clust1 <- as.matrix(nutrient)

Cluster1<-Mclust(nutrient)

summary(Cluster1)

plot(Cluster1)

```



```{R}

#Dataset 2

library (mclust)

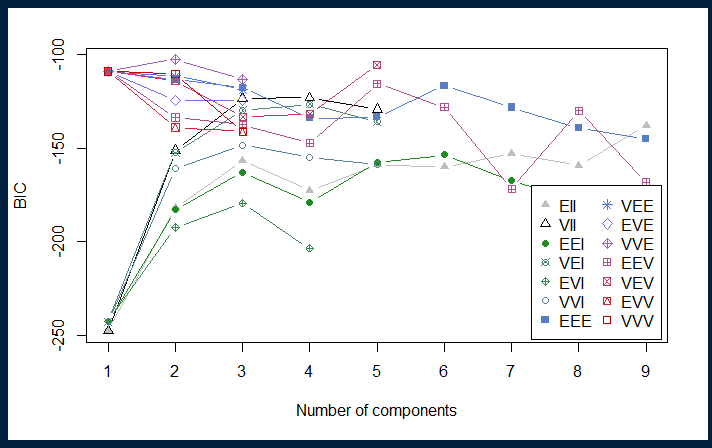
Clust2 <- as.matrix(achieve)

Cluster2<-Mclust(achieve)

summary(Cluster2)

plot(Cluster2)

```



```{R}

#Dataset 3

library (mclust)

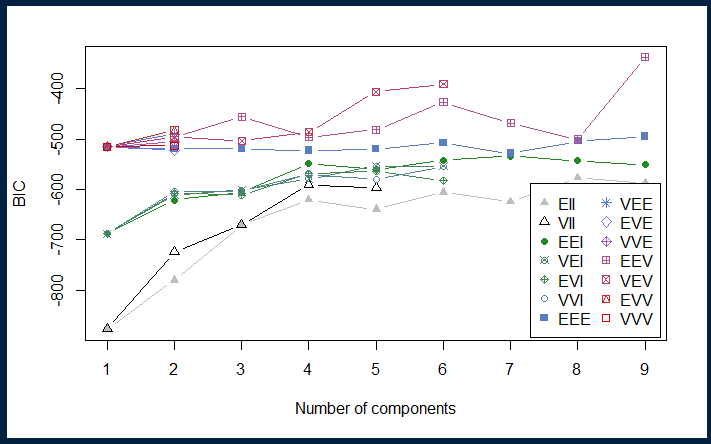
Clust3 <- as.matrix(milk)

Cluster3<-Mclust(milk)

summary(Cluster3)

plot(Cluster3)

```



1. Buatlah table rekapitulasi

Dataset 1 : nutrient

|  |  |  |  |
| --- | --- | --- | --- |
| No | Metode | Nomer klaster | Anggota klaster |
| 1 | Complete-Linkage | 1 | 7 |
|  |  | 2 | 2 |
|  |  | 3 | 1 |
|  |  | 4 | 17 |
| 2 | Single-Linkage | 1 | 2 |
|  |  | 2 | 1 |
|  |  | 3 | 1 |
|  |  | 4 | 23 |
| 3 | Average-Linkage | 1 | 2 |
|  |  | 2 | 1 |
|  |  | 3 | 1 |
|  |  | 4 | 23 |
| 4 | Centroid | 1 | 4 |
|  |  | 2 | 9 |
|  |  | 3 | 1 |
|  |  | 4 | 12 |
| 5 | Ward | 1 | 7 |
|  |  | 2 | 2 |
|  |  | 3 | 1 |
|  |  | 4 | 17 |

Dataset 2 : achieve

|  |  |  |  |
| --- | --- | --- | --- |
| No | Metode | Nomer klaster | Anggota klaster |
| 1 | Complete-Linkage | 1 | 2 |
|  |  | 2 | 15 |
|  |  | 3 | 8 |
| 2 | Single-Linkage | 1 | 23 |
|  |  | 2 | 1 |
|  |  | 3 | 1 |
| 3 | Average-Linkage | 1 | 20 |
|  |  | 2 | 5 |
| 4 | Centroid | 1 | 12 |
|  |  | 2 | 1 |
|  |  | 3 | 12 |
| 5 | Ward | 1 | 16 |
|  |  | 2 | 2 |
|  |  | 3 | 7 |

Dataset 3 : milk

|  |  |  |  |
| --- | --- | --- | --- |
| No | Metode | Nomer klaster | Anggota klaster |
| 1 | Complete-Linkage | 1 | 14 |
|  |  | 2 | 2 |
|  |  | 3 | 5 |
|  |  | 4 | 4 |
| 2 | Single-Linkage | 1 | 2 |
|  |  | 2 | 1 |
|  |  | 3 | 1 |
|  |  | 4 | 21 |
| 3 | Average-Linkage | 1 | 2 |
|  |  | 2 | 14 |
|  |  | 3 | 5 |
|  |  | 4 | 4 |
| 4 | Centroid | 1 | 5 |
|  |  | 2 | 18 |
|  |  | 3 | 2 |
|  |  | 4 | 1 |
| 5 | Ward | 1 | 6 |
|  |  | 2 | 11 |
|  |  | 3 | 2 |
|  |  | 4 | 6 |

1. Gunakan no 1 untuk metode-metode berikut :
2. K-means dan buatlah table rekapitulasi seperti no 1d.
3. K-medoids dan buatlah table rekapitulasi seperti no 1d.
4. K-medians dan buatlah table rekapitulasi seperti no 1d.

```{R}

#datasets 1

fviz\_nbclust(nutrient, kmeans, method = "silhouette")

kmean1 <- kmeans(nutrient, 2)

kmean1

#datasets 2

fviz\_nbclust(achieve, kmeans, method = "silhouette")

kmean2 <- kmeans(achieve, 2)

kmean2

#datasets 3

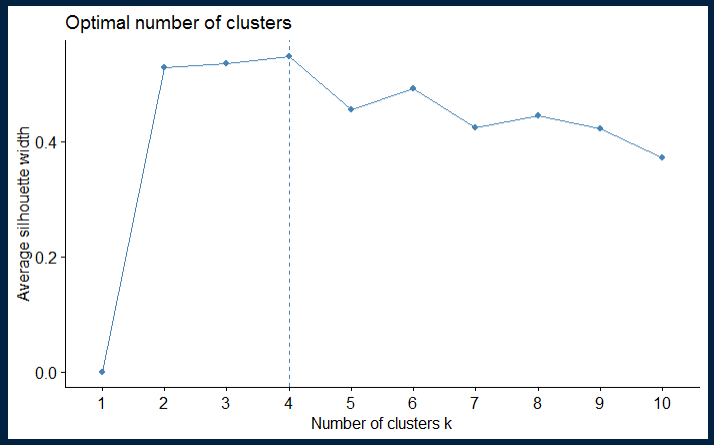
fviz\_nbclust(milk, kmeans, method = "silhouette")

kmean3 <- kmeans(milk, 2)

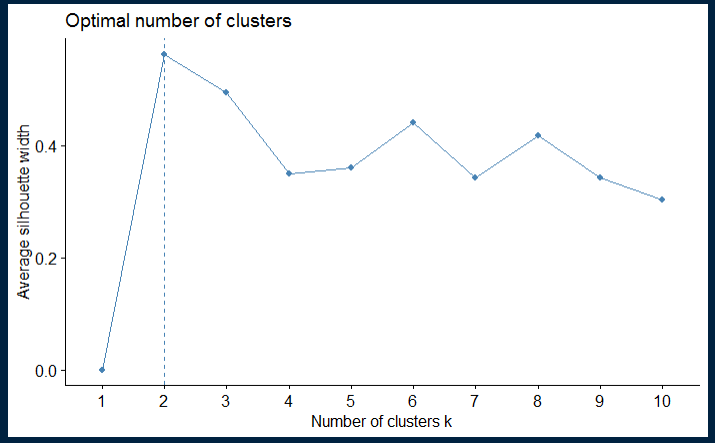
kmean3

```

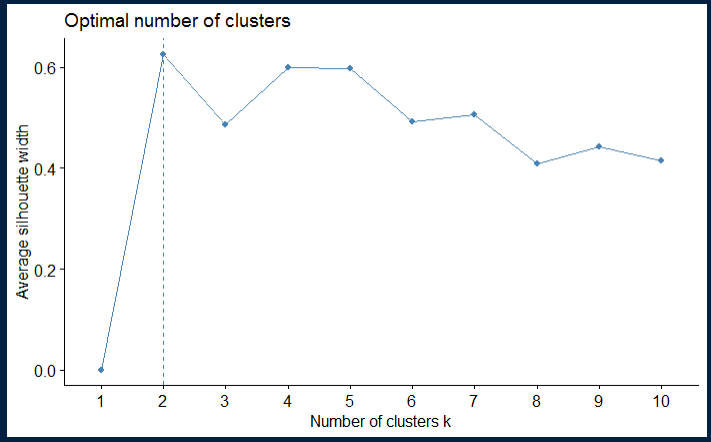
Datasets 1



Datasets 2



Datasets 3



|  |  |  |  |
| --- | --- | --- | --- |
| Datasets | Metode | Nomer klaster | Anggota klaster |
| 1 | K-Means | 1 | 13 |
|  |  | 2 | 12 |
| 2 | K-Means | 1 | 12 |
|  |  | 2 | 14 |
| 3 | K-Means | 1 | 11 |
|  |  | 2 | 14 |