

Nama : M. Mahtubillah
 NIM : 222011569
 No. Absen : 16
 Kelas : 3511
 Mata Kuliah : Analisis Pembah Banda

Tugas Praktikum 11 : Cluster Analysis

1) Kerjakan latihan dari buku Johnson & Wichern (2002)

(12.8) Using the distance in example 12.4, Cluster the items using the average hierarchical procedure, draw dendrogram. Compare result with 12.4 & 12.6.

> Jarak antara pasangan dari 5 objek

$$D = \{d_k\} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & & & & \\ 9 & 0 & & & \\ 3 & 7 & 0 & & \\ 6 & 5 & 9 & 0 & \\ 11 & 10 & 2 & 8 & 0 \end{bmatrix} \end{matrix}$$

$$d_{(35)_1} = \text{avg} \{d_{31}, d_{51}\} = \text{avg} \{3, 11\} = 7$$

$$d_{(35)_2} = \text{avg} \{d_{32}, d_{52}\} = \text{avg} \{7, 10\} = 8.5$$

$$d_{(35)_4} = \text{avg} \{d_{34}, d_{54}\} = \text{avg} \{9, 8\} = 8.5$$

$$D = \{d_k\} = \begin{matrix} & \begin{matrix} (35) & 1 & 2 & 4 \end{matrix} \\ \begin{matrix} (35) \\ 1 \\ 2 \\ 4 \end{matrix} & \begin{bmatrix} 0 & & & \\ 7 & 0 & & \\ 8.5 & 9 & 0 & \\ 8.5 & 6 & 5 & 0 \end{bmatrix} \end{matrix}$$

$$d_{(24)(35)} = \text{avg} \{d_2(35), d_4(35)\} = \text{avg} \{8.5, 8.5\} = 8.5$$

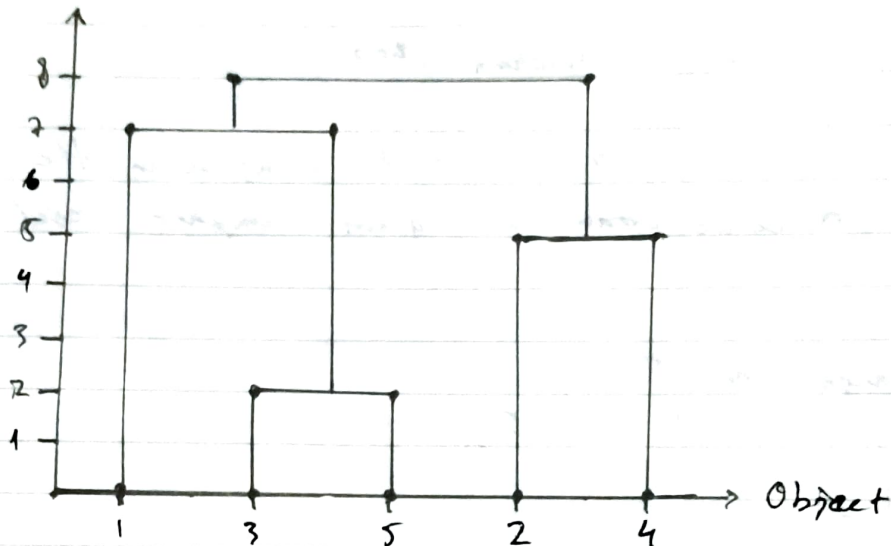
$$d_{(24)_1} = \text{avg} \{d_{21}, d_{41}\} = \text{avg} \{9, 6\} = 7.5$$

$$D = \{d_k\} = \begin{matrix} & \begin{matrix} (35) & 1 & (24) \end{matrix} \\ \begin{matrix} (35) \\ 1 \\ (24) \end{matrix} & \begin{bmatrix} 0 & & \\ 7 & 0 & \\ 8.5 & 7.5 & 0 \end{bmatrix} \end{matrix}$$

$$d_{(135)(24)} = \text{avg} (d_{(124)}, d_{(35)(24)}) = \text{avg} \{7.5, 8.5\} = 8$$

$$D = \{d_{hk}\} = \begin{matrix} & \begin{matrix} (135) & (24) \end{matrix} \\ \begin{matrix} (135) \\ (24) \end{matrix} & \begin{bmatrix} 0 & 8 \\ 8 & 0 \end{bmatrix} \end{matrix}$$

> Dendrogram



> Compare Result with 12.4 & 12.6

12.4 → ... - Single Linkage Method

12.6 → Complete Linkage Method

12.8 → Average Linkage Method

↳ Jika diambil 2 kluster maka metode single and average memiliki pengelompokan sama yaitu 1, 3, dan 5 (kelompok 1) dan 2 & 4 (kelompok 2). Sedangkan, complete 1, 2, dan 4 (kelompok 1) dan 3 & 5 (kelompok 2).

↳ Secara umum, ketiga metode sama-sama baik karena tidak ada kelompok beranggota tunggal.

↳ Jika diambil 3 kluster maka masing-masing metode memiliki kelompok tunggal:

Single = (1, 3, 5), (2), (4) → 2 Tunggal

Complete = (1), (2, 4), (3, 5) → 1 Tunggal

Average = (1), (2, 4), (3, 5) → 1 Tunggal

↳ Metode ... Complete dan average lebih baik pada 3 kluster.

(12.10) Use Ward's Method to Cluster

Item	x	a.) each item is a cluster
1	2	$\{1\} = ESS = \sum (x_j - \bar{x})^2 = (2-2)^2 = 0$
2	1	$\{2\} = ESS = \sum (x_j - \bar{x})^2 = (1-1)^2 = 0$
3	5	$\{3\} = ESS = \sum (x_j - \bar{x})^2 = (5-5)^2 = 0$
4	8	$\{4\} = ESS = \sum (x_j - \bar{x})^2 = (8-8)^2 = 0$

$$\{1\}\{2\}\{3\}\{4\} = ESS = 0 + 0 + 0 + 0 = 0$$

↳ Karena hanya ada 1 anggota, maka \bar{x} adalah x itu sendiri. Sehingga, ESS bernilai 0.

b.) > Cluster $\{1,2\}\{3\}\{4\}$
 $\{1,2\} = ESS = \sum (x_j - \bar{x})^2 = (2-1.5)^2 + (1-1.5)^2 = 0.5$
 $\{1,2\}\{3\}\{4\} = ESS = 0.5 + 0 + 0 = 0.5$
 increase in ESS = $0.5 - 0 = 0.5$

> Cluster $\{1,3\}\{2\}\{4\}$
 $\{1,3\} = ESS = \sum (x_j - \bar{x})^2 = (2-3.5)^2 + (5-3.5)^2 = 4.5$
 $\{1,3\}\{2\}\{4\} = ESS = 4.5 + 0 + 0 = 4.5$
 increase in ESS = $4.5 - 0 = 4.5$

> Cluster $\{1\}\{2,3\}\{4\}$
 $\{2,3\} = ESS = \sum (x_j - \bar{x})^2 = (1-3)^2 + (5-3)^2 = 8$
 $\{2,3\}\{1\}\{4\} = ESS = 8 + 0 + 0 = 8$
 increase in ESS = $8 - 0 = 8$

> Cluster $\{1\}\{2,4\}\{3\}$
 $\{2,4\} = ESS = \sum (x_j - \bar{x})^2 = (1-4.5)^2 + (8-4.5)^2 = 24.5$
 $\{2,4\}\{1\}\{3\} = ESS = 24.5 + 0 + 0 = 24.5$
 increase in ESS = $24.5 - 0 = 24.5$

> Cluster $\{1\}\{2\}\{3,4\}$
 $\{3,4\} = ESS = \sum (x_j - \bar{x})^2 = (5-6.5)^2 + (8-6.5)^2 = 4.5$
 $\{3,4\}\{1\}\{2\} = ESS = 4.5 + 0 + 0 = 4.5$
 increase in ESS = $4.5 - 0 = 4.5$

> Cluster $\{1,4\} \{2\} \{3\}$
 $\{1,4\} = \text{ESS} = \sum (Y_i - \bar{X})^2 = (2-5)^2 + (8-5)^2 = 18$
 $\{1,4\} \{2\} \{3\} = \text{ESS} = 18 + 0 + 0 = 18$
 increase in ESS = $18 - 0 = 18$

Clusters	ESS Increased
$\{1,2\} \{3\} \{4\}$	0.5 \rightarrow ESS
$\{1,3\} \{2\} \{4\}$	4.5
$\{1,4\} \{3\} \{2\}$	18
$\{1\} \{2,3\} \{4\}$	8
$\{1\} \{2,4\} \{3\}$	24.5
$\{1\} \{2\} \{3,4\}$	4.5

c) Complete step

> Cluster $\{1,2,3\} \{4\}$
 $\{1,2,3\} = \bar{X} = (2+1+5)/3 = 2.67$
 $\text{ESS} = (2-2.67)^2 + (1-2.67)^2 + (5-2.67)^2$
 $= 8.67$

$\{1,2,3\} \{4\} = \text{ESS} = 8.67 + 0 = 8.67$

increase in ESS = $8.67 - 0.5 = 8.17$

> $\{1,2,4\} \{3\} = \bar{X} = (2+1+8)/3 = 3.67$
 $\text{ESS} = (2-3.67)^2 + (1-3.67)^2 + (8-3.67)^2$
 $= 28.67$

$\{1,2,4\} \{3\} = \text{ESS} = 28.67 + 0 = 28.67$

increase in ESS = $28.67 - 0.5 = 28.17$

> $\{1,2\} \{3,4\} = \text{ESS} = 0.5 + 4.5 = 5$

increase in ESS = $5 - 0.5 = 4.5$

Clusters	increase in ESS
$\{1,2,3\} \{4\}$	8.17
$\{1,2,4\} \{3\}$	28.17
$\{1,2\} \{3,4\}$	4.5 \rightarrow ESS

> Cluster {1234}

$$\{1234\} = \bar{x} = (2+1+5+8)/4 = 4$$

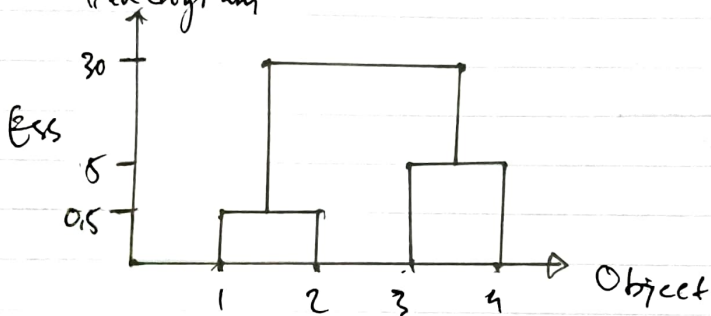
$$ESS = (2-4)^2 + (1-4)^2 + (5-4)^2 + (8-4)^2 = 30$$

$$\text{Increase in ESS} = 30 - 4.5 = 25.5$$

> Tabel ESS

Clusters	ESS
{1} {2} {3} {4}	0
{12} {3} {4}	0.5
{123} {4}	5
{1234}	30

> Pseudogram



(P.11) Use the Kmeans clustering to divide items into $k=2$, initial group AB & CD

> Coordinate of Centroid

$$AB \rightarrow \bar{x}_1 = 2(5+1)/2 = 3$$

$$\bar{x}_2 = (4-2)/2 = 1$$

$$CD \rightarrow \bar{x}_3 = (-1+3)/2 = 1$$

$$\bar{x}_4 = (1+1)/2 = 1$$

Clusters	Coordinate of Centroid	
	\bar{x}_1	\bar{x}_2
AB	3	1
CD	1	1

> Compute Euclidean Distance

$$d^2(A, (AB)) = (5-3)^2 + (4-1)^2 = 13 \rightarrow \text{lebih dekat}$$

$$d^2(A, (CD)) = (5-1)^2 + (4-1)^2 = 25$$

$$d^2(B, (AB)) = (1-3)^2 + (-2-1)^2 = 13$$

$$d^2(B, (CD)) = (1-1)^2 + (-2-1)^2 = 9 \rightarrow \text{lebih dekat, Cluster berubah}$$

Clusters	CoC	
	\bar{x}_1	\bar{x}_2
A	5	4
(BCD)	1	0

> Square Distance

$$d^2(A, A) = (5-5)^2 + (4-4)^2 = 0$$

$$d^2(A, (BCD)) = (5-1)^2 + (4-0)^2 = 32$$

$$d^2(B, A) = (1-5)^2 + (-2-4)^2 = 52$$

$$d^2(C, A) = (-1-5)^2 + (1-4)^2 = 45$$

$$d^2(D, (BCD)) = (-1-1)^2 + (1-0)^2 = 2$$

$$d^2(D, A) = (3-5)^2 + (1-4)^2 = 13$$

$$d^2(B, (BCD)) = (1-1)^2 + (-2-0)^2 = 4$$

Cluster	Item			
	A	B	C	D
A	0	52	45	13
(BCD)	32	4	5	2

C(2,2)

> CoC

$$AC \rightarrow \bar{x}_1 = (5+1)/2 = 3$$

$$\bar{x}_2 = (3-2)/2 = 0,5$$

$$BD \rightarrow \bar{x}_1 = (-1-3)/2 = -2$$

$$\bar{x}_2 = (1-2)/2 = -0,5$$

Cluster	Observation	
	\bar{x}_1	\bar{x}_2
(AC)	3	0,5
(BD)	-2	-0,5

> Compute Euclidean Distance

$$d^2(A, (AC)) = (5-3)^2 + (3-0,5)^2 = 6,25 \rightarrow \text{lebih dekat}$$

$$d^2(A, (BD)) = (5+2)^2 + (3+0,5)^2 = 61,25$$

$$d^2(C, (AC)) = (1-3)^2 + (-2-0,5)^2 = 10,25 \rightarrow \text{lebih dekat}$$

$$d^2(C, (BD)) = (1+2)^2 + (-2+0,5)^2 = 11,25$$

> Square Distance

$$d^2(A, (AC)) = 6,25$$

$$d^2(A, (BD)) = 61,25$$

$$d^2(C, (AC)) = 10,25$$

$$d^2(C, (BD)) = 11,25$$

$$d^2(B, (AC)) = (-1-3)^2 + (-1-0,5)^2 = 16,25$$

$$d^2(B, (BD)) = (-1+2)^2 + (1+0,5)^2 = 3,25$$

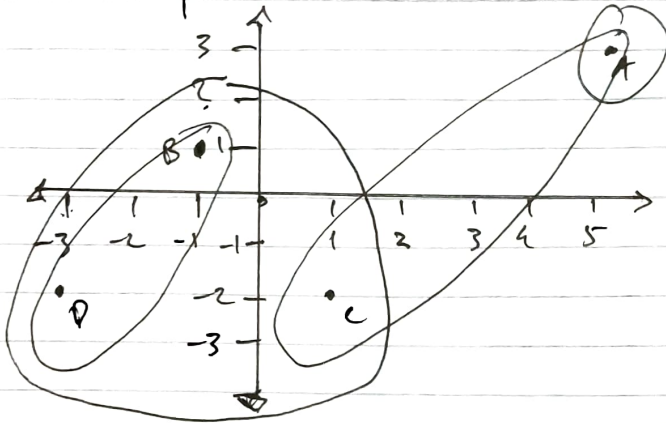
$$d^2(D, (AC)) = (-3-3)^2 + (-2-0,5)^2 = 42,25$$

$$d^2(D, (BD)) = (-3+2)^2 + (-2+0,5)^2 = 3,25$$

Cluster	Item			
	A	B	C	D
(AC)	10,25	16,25	10,25	42,25
(BD)	61,25	3,25	11,25	3,15

•> Compare Solution
 → Dengan Scheme, k-mean, dengan awalan berbeda memberikan hasil berbeda yaitu kluster AC & BD.

→ Graph



→ Solusi terbentuk adalah kluster (AC) dan (BD).
 → Garis terbentuk cukup baik karena tidak berumpuk dan dapat dipisahkan dengan garis lurus.
 → Item C lebih dekat dengan BD.

Penugasan Praktikum 11

M.Mahbubillah

2022-11-24

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(DT)

## Warning: package 'DT' was built under R version 4.2.2

library(readxl)
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.2.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.2.2

## Welcome! Want to learn more? See two factoextra-related books at
## https://goo.gl/ve3WBa

library(psych)

## Warning: package 'psych' was built under R version 4.2.2

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##   %+%, alpha

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 4.2.2
```



```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine
```

Impor Dataset

```
dataPenugasan1 <- read_excel("penugasanP10.xlsx", sheet=1)
kabkot_papua<-as.factor(dataPenugasan1$Kab)
data_papua<-as.matrix(dataPenugasan1[,2:5])
rownames(data_papua)<-kabkot_papua
data_papua<-as.data.frame(data_papua)
head(data_papua)
```

```
##           Miskin   IPM   TPT   Gini
## Merauke      10.03 70.09  2.61 0.380
## Jayawijaya   37.22 58.03  2.39 0.342
## Jayapura     12.44 71.69  9.68 0.432
## Nabire       24.15 68.83  6.31 0.349
## Kepulauan Yapen 26.30 67.66  5.78 0.400
## Biak Numfor   24.57 72.19 10.42 0.401
```

Analisis Deskriptif

```
summary(data_papua)
```

```
##           Miskin           IPM           TPT           Gini
## Min.      :10.03   Min.      :31.55   Min.      : 0.000   Min.      :0.187
## 1st Qu.:24.15   1st Qu.:48.37   1st Qu.: 0.710   1st Qu.:0.326
## Median :29.54   Median :56.31   Median : 2.390   Median :0.362
## Mean     :28.21   Mean     :57.29   Mean     : 3.279   Mean     :0.355
## 3rd Qu.:36.72   3rd Qu.:66.40   3rd Qu.: 4.680   3rd Qu.:0.400
## Max.     :41.76   Max.     :79.94   Max.     :12.370   Max.     :0.448
```

Matriks Korelasi

```
cor(dataPenugasan1[,2:5])
```

```
##           Miskin           IPM           TPT           Gini
## Miskin   1.0000000 -0.7698378 -0.6846585 -0.3597086
## IPM      -0.7698378  1.0000000  0.8270919  0.3354901
## TPT      -0.6846585  0.8270919  1.0000000  0.1730048
## Gini     -0.3597086  0.3354901  0.1730048  1.0000000
```

Berdasarkan matriks korelasi di atas, dapat diketahui bahwa terdapat hubungan yang cukup signifikan antar variabel dalam data sehingga hasil profiling yang akan dilakukan setelah cluster terbentuk nanti adalah menggunakan metode PCA (Principal Component Analysis).

Matriks Kovarian

```
cov(dataPenugasan1[,2:5])
```

```
##           Miskin           IPM           TPT           Gini
## Miskin  92.5370030 -83.5670691 -21.86649951 -0.226079286
## IPM     -83.5670691 127.3375261  30.98700099  0.247348929
## TPT     -21.8664995  30.9870010  11.02288103  0.037528214
## Gini    -0.2260793  0.2473489  0.03752821  0.004268786
```

Berdasarkan matriks kovarian di atas, dapat dilihat bahwa terdapat sebuah variabel yang sangat homogen. Hal ini dapat dilihat pada varians Gini Ratio sebesar 0,004 yang dibandingkan dengan varians IPM sebesar 127,337. Hal ini dapat terjadi karena data belum standar sehingga dalam analisis kali ini, data yang digunakan adalah data standar dari data papua.

Standardisasi

```
z_penugasan <- round(scale(data_papua),4)
```

```
# datatable(z_penugasan, caption = "Data Hasil Standardisasi")
```

```
z_penugasan
```

```
##           Miskin           IPM           TPT           Gini
## Merauke      -1.8903    1.1345   -0.2015    0.3826
## Jayawijaya    0.9362    0.0658   -0.2678   -0.1990
## Jayapura     -1.6398    1.2763    1.9280    1.1785
## Nabire       -0.4224    1.0229    0.9129   -0.0918
## Kepulauan Yapen -0.1989    0.9192    0.7533    0.6887
## Biak Numfor  -0.3788    1.3206    2.1509    0.7041
## Paniai        0.8832   -0.0866   -0.7888    0.3826
## Puncak Jaya   0.6784   -0.7903   -0.4515    0.3214
## Mimika       -1.4506    1.4979    1.2744   -0.2449
## Boven Digoel  -0.9152    0.3760   -0.0599    1.4234
## Mappi        -0.3299    0.0764    0.3708    1.1020
## Asmat        -0.2831   -0.5971   -0.6683    0.7347
## Yahukimo      0.9487   -0.7016   -0.3822    0.5969
## Pegunungan Bintang 0.2013   -1.0499   -0.4967    1.4234
## Tolikara      0.3978   -0.6901   -0.5961    0.5051
## Sarmi        -1.4911    0.5621    0.5967    1.0561
## Keerom       -1.2364    0.8075   -0.1141    0.6275
## Waropen       0.1379    0.6781    0.0124   -0.1377
## Supiori       0.9040    0.4442    0.4220   -0.5510
## Mamberamo Raya 0.0173   -0.4881   -0.0840   -2.2040
## Nduga         0.8843   -2.2808   -0.6774   -2.5713
## Lanny Jaya    1.0308   -0.8355   -0.9515   -1.1326
## Mamberamo Tengah 0.8520   -0.8612   -0.7738   -0.3826
## Yalimo        0.4788   -0.7929   -0.7949    0.1071
## Puncak        0.9092   -1.2626   -0.9876   -0.8112
## Dogiyai       0.0422   -0.2169   -0.9545   -0.4439
## Intan Jaya    1.2990   -0.8417   -0.9876   -1.3163
## Deiyai        1.4082   -0.6937   -0.9214    0.0306
## Kota Jayapura -1.7728    2.0074    2.7382   -1.1785
## attr(,"scaled:center")
```

```
##      Miskin      IPM      TPT      Gini
## 28.213793 57.287586  3.278966  0.355000
## attr(,"scaled:scale")
##      Miskin      IPM      TPT      Gini
##  9.61961553 11.28439303  3.32007244  0.06533595
```

Analisis Kluster Metode Hirarki

```
head(dataPenugasan1)
```

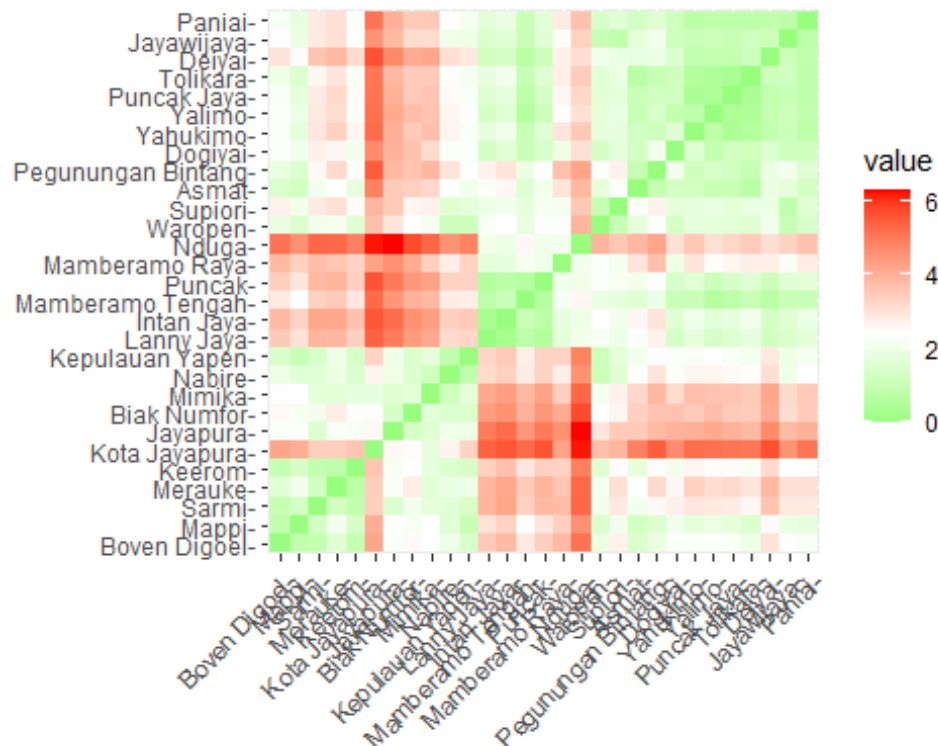
```
## # A tibble: 6 × 5
##   Kab      Miskin    IPM    TPT    Gini
##   <chr>      <dbl> <dbl> <dbl> <dbl>
## 1 Merauke      10.0  70.1  2.61  0.38
## 2 Jayawijaya   37.2  58.0  2.39  0.342
## 3 Jayapura     12.4  71.7  9.68  0.432
## 4 Nabire       24.2  68.8  6.31  0.349
## 5 Kepulauan Yapen 26.3  67.7  5.78  0.4
## 6 Biak Numfor   24.6  72.2 10.4  0.401
```

```
AKU <- princomp(dataPenugasan1[,2:5], cor = T)
summary(AKU, loadings = TRUE)
```

```
## Importance of components:
##              Comp.1      Comp.2      Comp.3      Comp.4
## Standard deviation  1.6350236  0.9392869  0.54452075  0.38462323
## Proportion of Variance 0.6683256  0.2205650  0.07412571  0.03698376
## Cumulative Proportion 0.6683256  0.8888905  0.96301624  1.00000000
##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4
## Miskin  0.546      0.815  0.194
## IPM    -0.575 -0.129  0.202  0.782
## TPT    -0.537 -0.340  0.507 -0.582
## Gini   -0.288  0.931  0.195 -0.108
```

Mengukur jarak masing-masing observasi

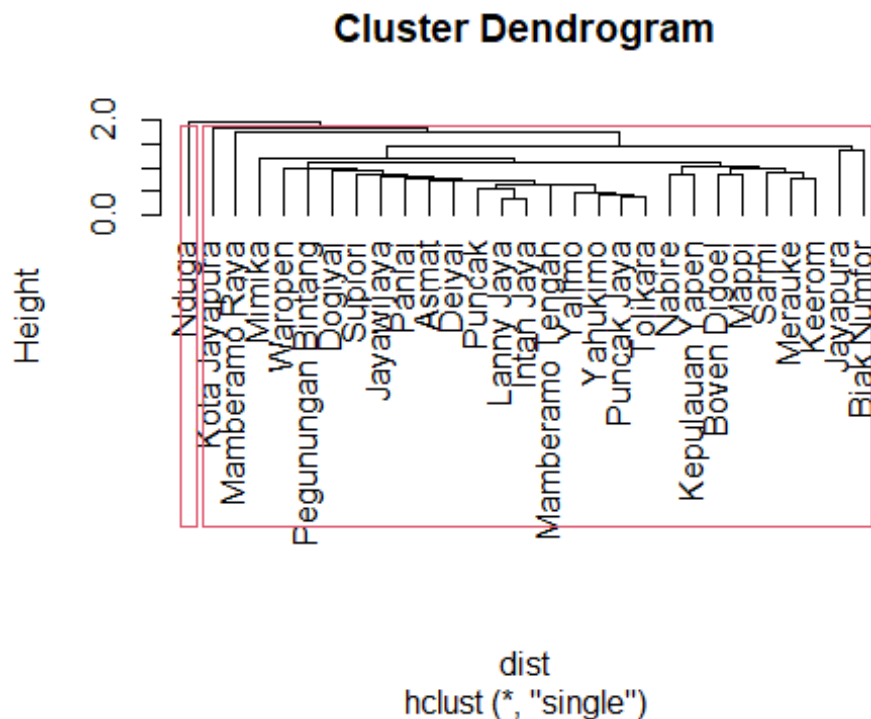
```
dist<-dist(z_penugasan)
fviz_dist(dist, gradient=list(low="green", mid="white", high="red"))
```



Berdasarkan hasil di atas, dapat dilihat bahwa hubungan yang paling dekat berdasarkan jarak pada masing-masing kabupaten/kota di Provinsi Papua ditunjukkan dengan warna kehijauan. Sedangkan jarak paling jauh disimbolkan dengan warna merah dan putih menunjukkan ukuran kedekatan berdasarkan jarak menengah.

Single Linkage

```
res1_par <- hclust(dist,method = "single")
plot(res1_par,hang=-2,cex=1)
rect.hclust(res1_par,k=2)
```

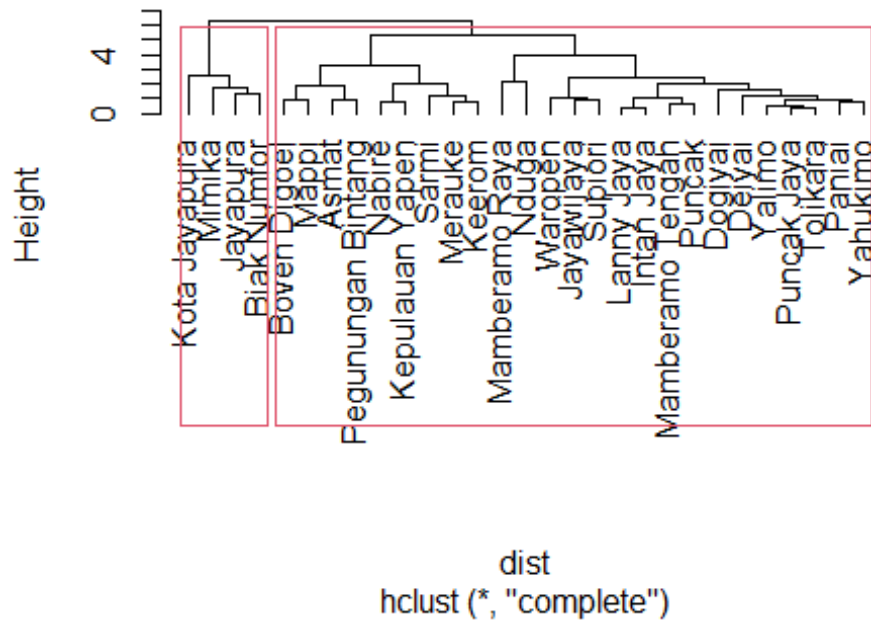



Berdasarkan single linkage yang memperhitungkan jarak paling dekat sebagai dasar pembentukan cluster, dapat dilihat bahwa Nduga menjadi salah satu kabupaten/kota yang diindikasikan sebagai pencilon pada data papua karena pada level clustering sebanyak 2 cluster, Nduga menjadi satu-satunya anggota pada cluster 1 sedangkan kabupaten/kota lain berada di cluster 2. Dikarenakan masih terdapat cluster yang terdiri dari satu observasi saja, maka metode single linkage masih kurang bagus.

Complete Linkage

```
res1_par <- hclust(dist,method = "complete")
plot(res1_par,hang=-2,cex=1)
rect.hclust(res1_par,k=2)
```

Cluster Dendrogram

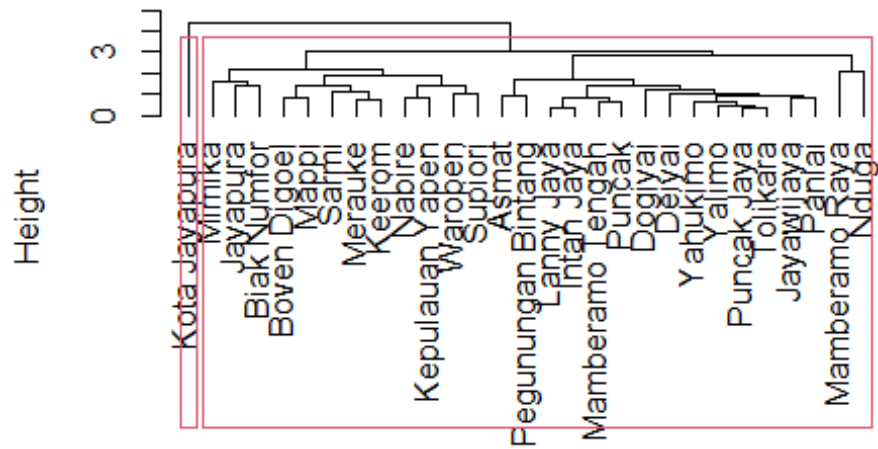


Berdasarkan complete linkage yang memperhitungkan jarak terjauh dalam menentukan cluster, dapat dilihat bahwa cluster 1 terdiri atas 4 kabupaten/kota sedangkan yang lain berada di cluster 2. Dibandingkan dengan single linkage, hasil ini cukup bagus untuk pembentukan cluster. Cluster yang terbentuk dapat dikatakan sudah baik dikarenakan tidak terdapat cluster yang hanya berisi satu observasi.

Average Linkage

```
res1_par <- hclust(dist,method = "average")
plot(res1_par,hang=-2,cex=1)
rect.hclust(res1_par,k=2)
```

Cluster Dendrogram

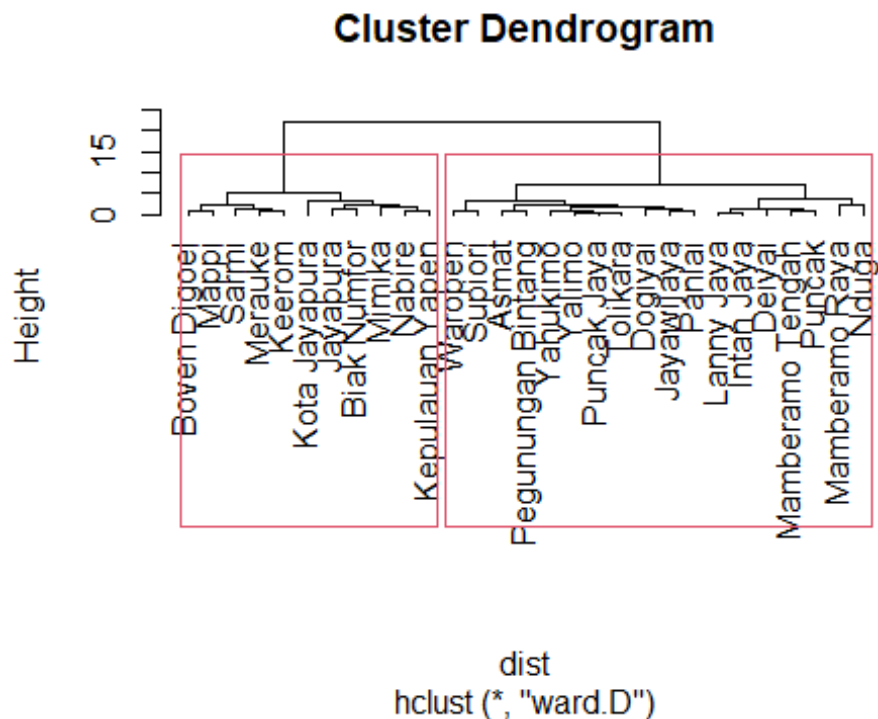


dist
hclust (*, "average")

Berdasarkan

average linkage yang memperhitungkan jarak paling dekat sebagai dasar pembentukan cluster, dapat dilihat bahwa Kota Jayapura menjadi salah satu kabupaten/kota yang diindikasikan sebagai pencilon pada data papua karena pada level clustering sebanyak 2 cluster. Kota Jayapura menjadi satu-satunya anggota pada cluster 1 sedangkan kabupaten/kota lain berada di cluster 2. Dikarenakan masih terdapat cluster yang terdiri dari satu observasi saja, maka metode average linkage masih kurang bagus. ### Ward.D Method

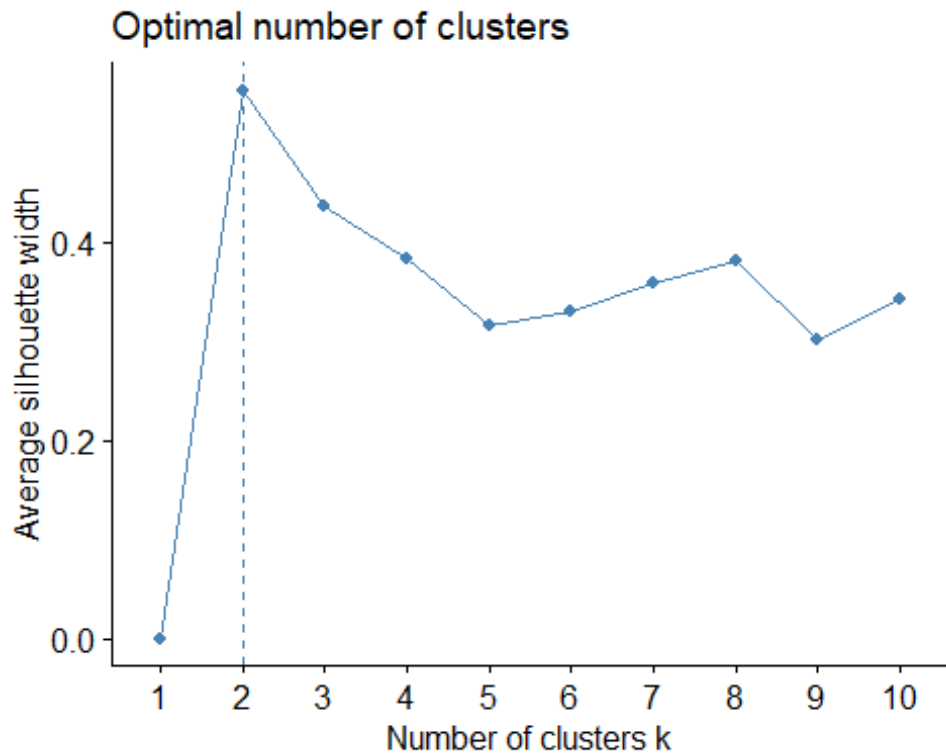
```
res1_par <- hclust(dist,method = "ward.D")
plot(res1_par,hang=-2,cex=1)
rect.hclust(res1_par,k=2)
```



Berdasarkan metode Ward yang memperhitungkan nilai SSE sebagai dasar pembentukan cluster, hasil clustering menunjukkan pengelompokan yang paling berbeda dibandingkan cara-cara yang lain yaitu cluster 1 dengan 11 anggota kabupaten/kota dan cluster 2 dengan 18 anggota kabupaten/kota sehingga cukup seimbang dalam membagi observasi dalam dua cluster. Hasil clustering dapat dikatakan lebih baik daripada ketiga hasil sebelumnya.

Analisis Cluster Non-Hirarki

```
fviz_nbclust(data_papua, kmeans)
```

Berdasarkan Elbow

Plot di atas, dapat dilihat bahwa cluster 2 memiliki perpotongan yang paling tajam dibandingkan cluster lainnya.

##Plot K-means dengan scalling

```
kmeans2 <- kmeans(z_penugasan,centers = 2,nstart=25)
```

```
kmeans3 <- kmeans(z_penugasan,centers = 3,nstart=25)
```

```
kmeans5 <- kmeans(z_penugasan,centers = 5,nstart=25)
```

```
kmeans7 <- kmeans(z_penugasan,centers = 7,nstart=25)
```

#compare plot

```
plot1 <- fviz_cluster(kmeans2,geom=c("point"),data=z_penugasan,  
show.clust.cent = TRUE) + ggtitle("k = 2")
```

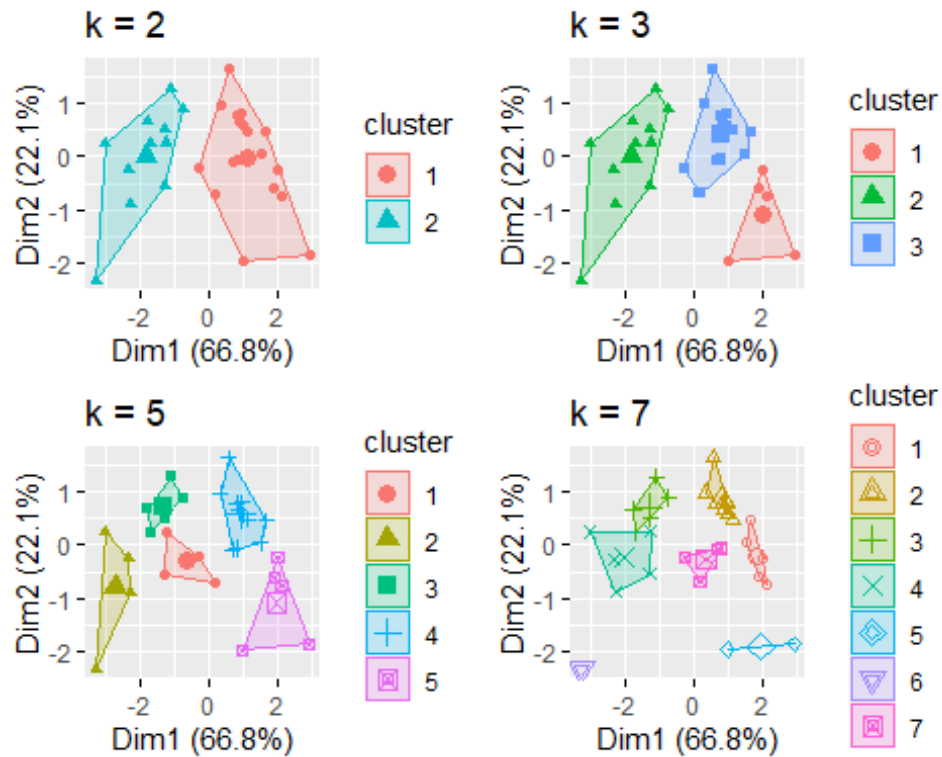
```
plot2 <- fviz_cluster(kmeans3,geom="point",data=z_penugasan) + ggtitle("k =  
3")
```

```
plot3 <- fviz_cluster(kmeans5,geom="point",data=z_penugasan) + ggtitle("k =  
5")
```

```
plot4 <- fviz_cluster(kmeans7,geom="point",data=z_penugasan) + ggtitle("k =  
7")
```

#Library(gridExtra)

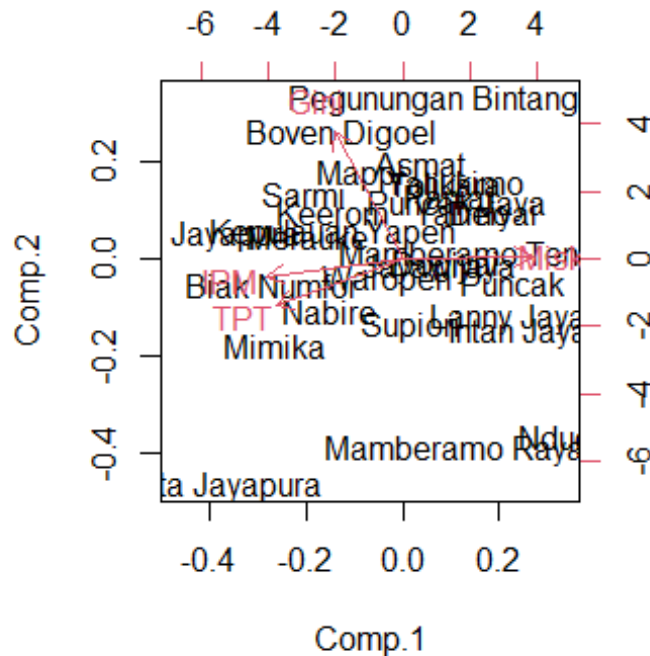
```
grid.arrange(plot1,plot2,plot3,plot4,nrow=2)
```



Berdasarkan

output di atas, dapat disimpulkan bahwa sebenarnya jika cluster dibagi menjadi 7 cluster, cluster yang terbentuk masih dapat digolongkan sebagai cluster yang baik. Namun, pada analisis cluster kali ini hanya akan menggunakan 2 cluster saja.

```
AKU <- princomp(data_papua, cor = T)
biplot(AKU)
```



Jika analisis kluster di atas diinterpretasikan bersama dengan biplot, maka dapat disimpulkan bahwa kluster 1 (Kabupaten Jayawijaya, Paniai, Puncak Jaya, dst) cenderung memiliki persentase penduduk miskin yang lebih tinggi dari kluster 2 (Kabupaten Merauke, Jayapura, Nabire, Kep. Yapen, dst). Sementara itu, kluster 2 cenderung memiliki TPT dan IPM yang lebih tinggi dari kluster 1. Sedangkan untuk gini ratio, antara kluster 1 dan 2 cenderung sama.

```
# append cluster assignment
```

```
mydata <- data.frame(data_papua, kmeans2$cluster)
```

```
mydata
```

	Miskin	IPM	TPT	Gini	kmeans2.cluster
## Merauke	10.03	70.09	2.61	0.380	2
## Jayawijaya	37.22	58.03	2.39	0.342	1
## Jayapura	12.44	71.69	9.68	0.432	2
## Nabire	24.15	68.83	6.31	0.349	2
## Kepulauan Yapen	26.30	67.66	5.78	0.400	2
## Biak Numfor	24.57	72.19	10.42	0.401	2
## Paniai	36.71	56.31	0.66	0.380	1
## Puncak Jaya	34.74	48.37	1.78	0.376	1
## Mimika	14.26	74.19	7.51	0.339	2
## Boven Digoel	19.41	61.53	3.08	0.448	2
## Mappi	25.04	58.15	4.51	0.427	2
## Asmat	25.49	50.55	1.06	0.403	1
## Yahukimo	37.34	49.37	2.01	0.394	1
## Pegunungan Bintang	30.15	45.44	1.63	0.448	1
## Tolikara	32.04	49.50	1.30	0.388	1
## Sarimi	13.87	63.63	5.26	0.424	2

## Keerom	16.32	66.40	2.90	0.396	2
## Waropen	29.54	64.94	3.32	0.346	1
## Supiori	36.91	62.30	4.68	0.319	1
## Mamberamo Raya	28.38	51.78	3.00	0.211	1
## Nduga	36.72	31.55	1.03	0.187	1
## Lanny Jaya	38.13	47.86	0.12	0.281	1
## Mamberamo Tengah	36.41	47.57	0.71	0.330	1
## Yalimo	32.82	48.34	0.64	0.362	1
## Puncak	36.96	43.04	0.00	0.302	1
## Dogiyai	28.62	54.84	0.11	0.326	1
## Intan Jaya	40.71	47.79	0.00	0.269	1
## Deiyai	41.76	49.46	0.22	0.357	1
## Kota Jayapura	11.16	79.94	12.37	0.278	2

```
aggregate(z_penugasan,by=list(kmeans2$cluster),FUN=mean)
```

##	Group.1	Miskin	IPM	TPT	Gini
## 1	1	0.6514556	-0.6111611	-0.5749833	-0.3137667
## 2	2	-1.0660182	1.0000727	0.9408818	0.5134273