

PAM

Medel Colorado Yoselin Merari

2022-06-05

PARTITION AROUND MEDOIDS (PAM)

Cargar libreria

```
library(cluster)
```

Cargar la matriz de datos.

```
X<-as.data.frame(state.x77)
colnames(X)
```

```
## [1] "Population" "Income"      "Illiteracy" "Life Exp"   "Murder"
## [6] "HS Grad"    "Frost"       "Area"
```

```
#----- # Transformacion de datos #-----
```

```
#1.- Transformación de las variables x1,x3 y x8 # con la función de logaritmo.
```

```
X[,1]<-log(X[,1])
colnames(X)[1]<-"Log-Population"
```

```
X[,3]<-log(X[,3])
colnames(X)[3]<-"Log-Illiteracy"
```

```
X[,8]<-log(X[,8])
colnames(X)[8]<-"Log-Area"
```

```
#----- # Metodo PAM #-----
```

```
#1.- Separación de filas y columnas.
```

```
dim(X)
```

```
## [1] 50 8
```

```
n<-dim(X)[1]
```

```
p<-dim(X)[2]
```

2.- Estandarización univariante.

```
X.s<-scale(X)
```

3.- Aplicación del algoritmo

```
pam.5<-pam(X.s,5)
```

4.- Clusters

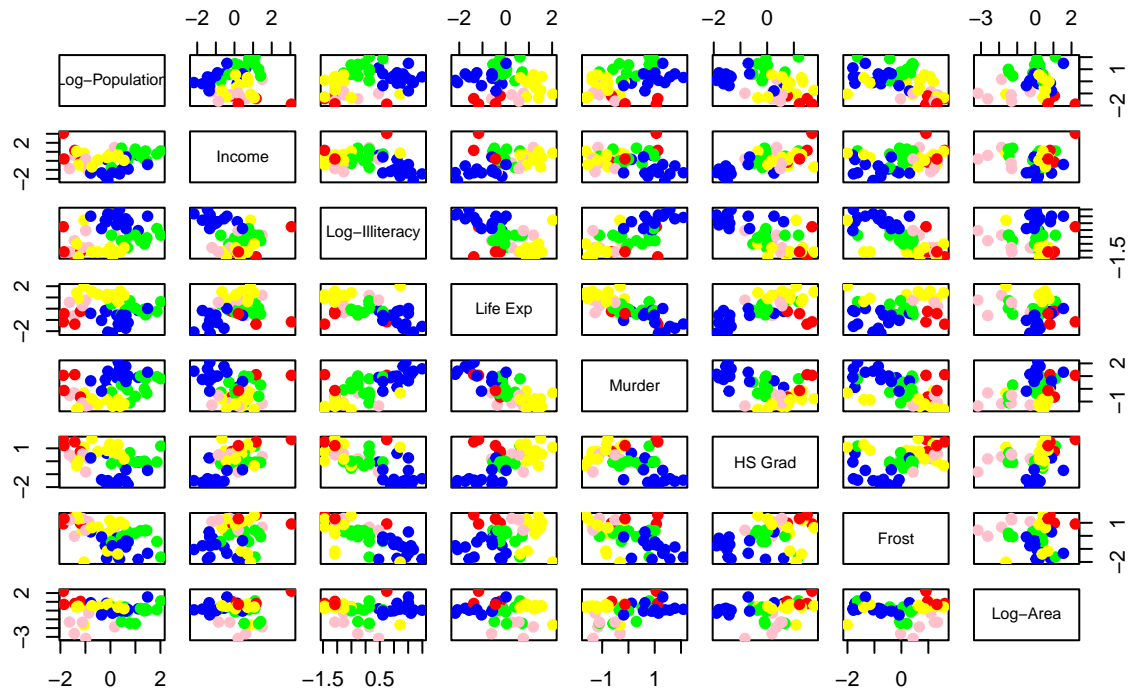
```
cl.pam<-pam.5$clustering  
cl.pam
```

##	Alabama	Alaska	Arizona	Arkansas	California
##	1	2	1	1	3
##	Colorado	Connecticut	Delaware	Florida	Georgia
##	4	5	5	1	1
##	Hawaii	Idaho	Illinois	Indiana	Iowa
##	4	4	3	3	4
##	Kansas	Kentucky	Louisiana	Maine	Maryland
##	4	1	1	5	3
##	Massachusetts	Michigan	Minnesota	Mississippi	Missouri
##	3	3	4	1	3
##	Montana	Nebraska	Nevada	New Hampshire	New Jersey
##	2	4	2	5	3
##	New Mexico	New York	North Carolina	North Dakota	Ohio
##	1	3	1	4	3
##	Oklahoma	Oregon	Pennsylvania	Rhode Island	South Carolina
##	3	4	3	5	1
##	South Dakota	Tennessee	Texas	Utah	Vermont
##	4	1	1	4	5
##	Virginia	Washington	West Virginia	Wisconsin	Wyoming
##	1	4	1	4	2

#5.- Scatter plot de la matriz con los grupos

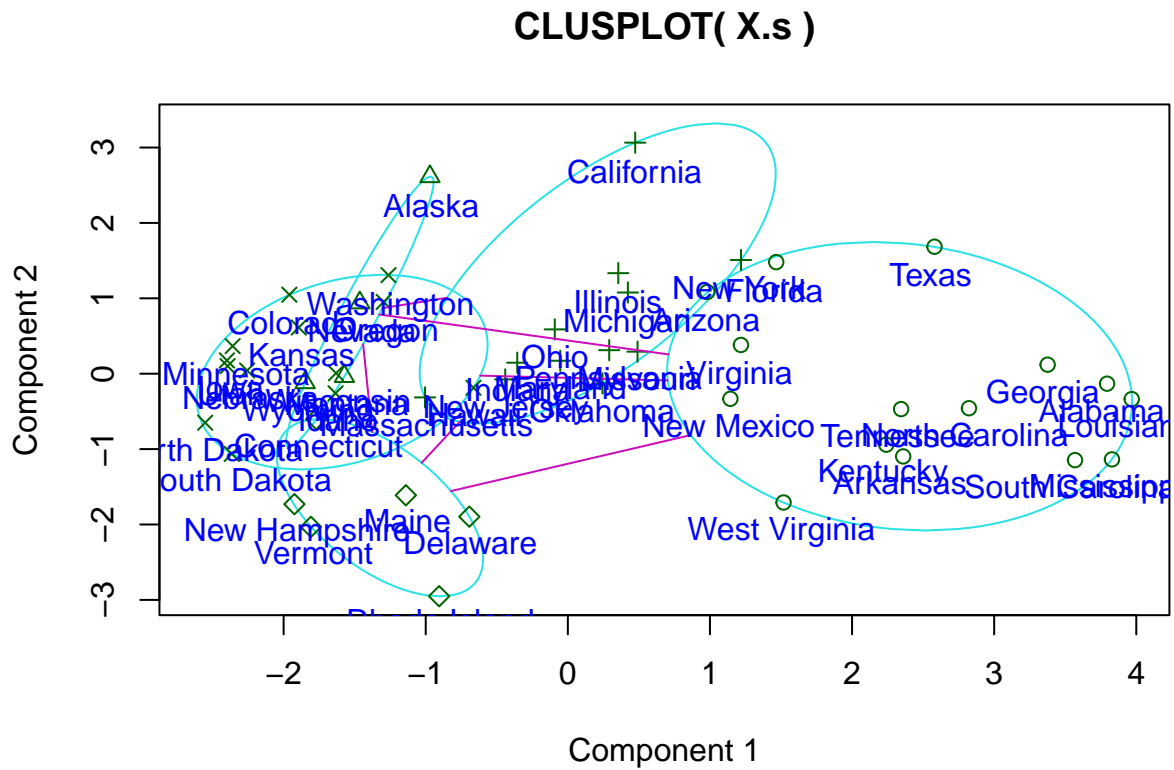
```
col.cluster<-c("blue","red","green","yellow","pink")[cl.pam]  
pairs(X.s, col=col.cluster, main="PAM", pch=19)
```

PAM



#----- # Visualizacion con Componentes Principales #-----

```
clusplot(X.s,cl.pam)
text(princomp(X.s)$scores[,1:2],
     labels=rownames(X.s),pos=1, col="blue")
```



These two components explain 62.5 % of the point variability.

#-----# Silhouette #-----

Representación gráfica de la eficacia de
clasificación de una observacion dentro de un
grupo.

1.- Generación de los calculos

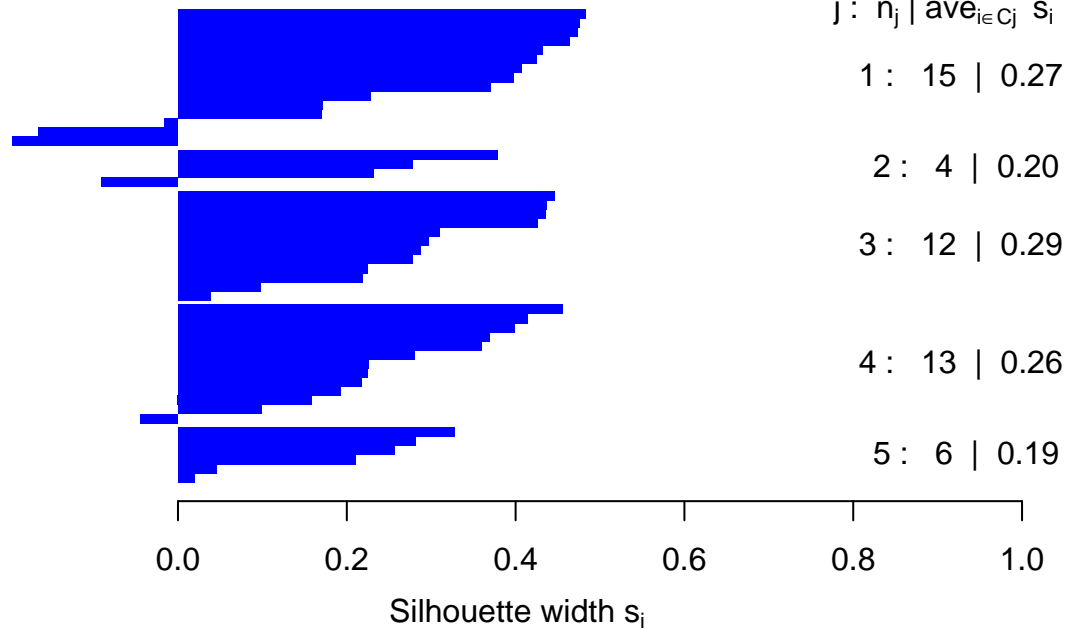
```
dist.Euc<-dist(X.s, method = "euclidean")
Sil.pam<-silhouette(cl.pam, dist.Euc)
```

#2.- Generación del gráfico

```
plot(Sil.pam, main="Silhouette for PAM",
     col="blue")
```

Silhouette for PAM

n = 50



Average silhouette width : 0.26