

# PAM

Edgar Ortiz Mota

2022-05-29

```
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"
```

## Transformación 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.- Separacion de filas y columnas.

```
dim(X)
```

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

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

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

2.- Estandarizacion univariante.

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

3.- Aplicacion del algoritmo

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

4.- Clusters

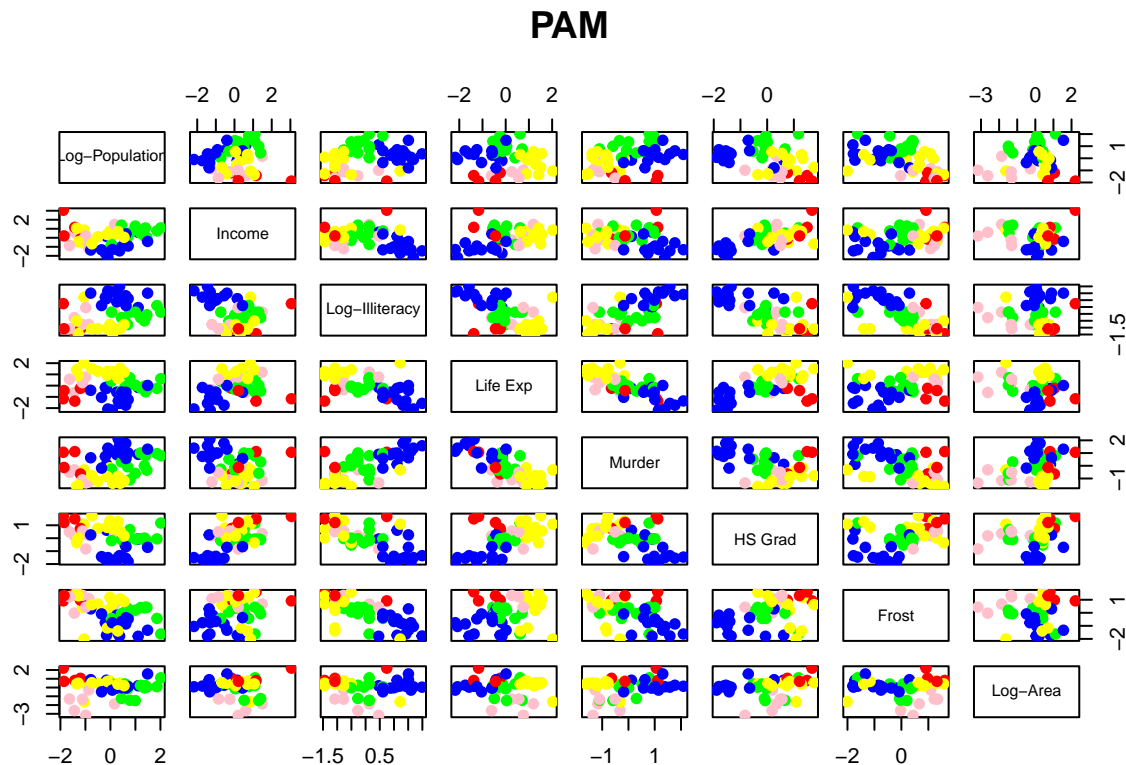
```
cl.pam<-pam.5$clustering
cl.pam
```

```
##           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

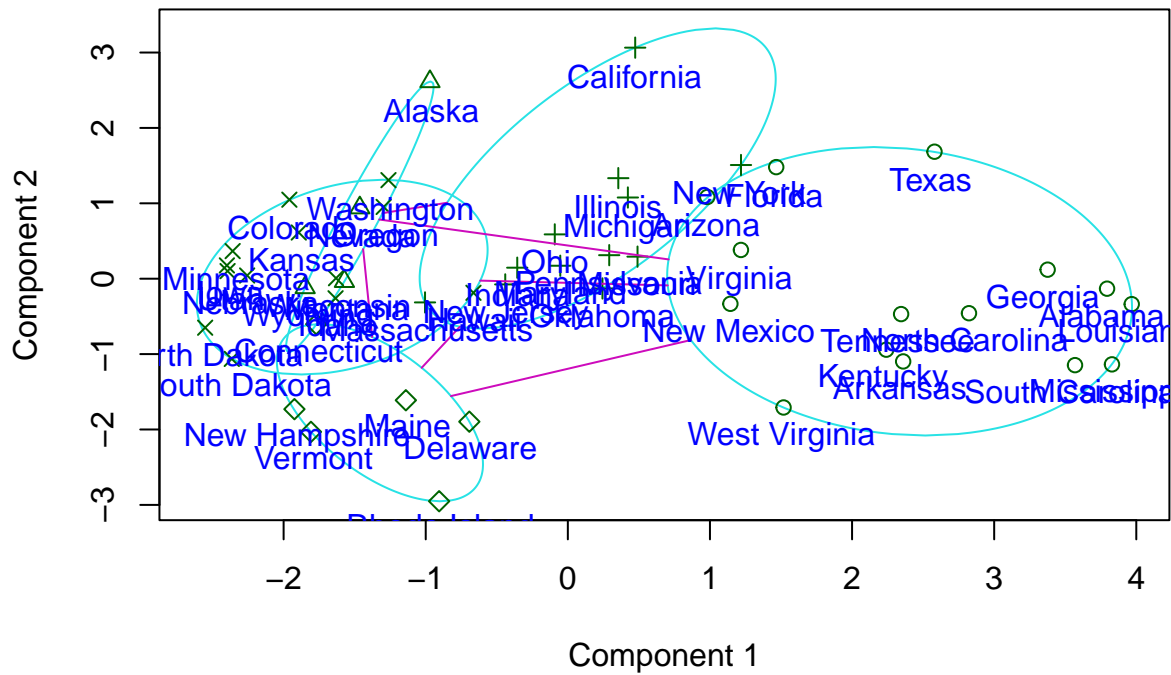
```
col.cluster<-c("blue","red","green","yellow","pink")[cl.pam]
pairs(X.s, col=col.cluster, main="PAM", pch=19)
```



Visualizacion con Componentes Principales

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

## CLUSPLOT( X.s )



## Silhouette

Representación grafica de la eficacia de clasificación de una observación dentro de un grupo.

1.- Generacion de los calculos

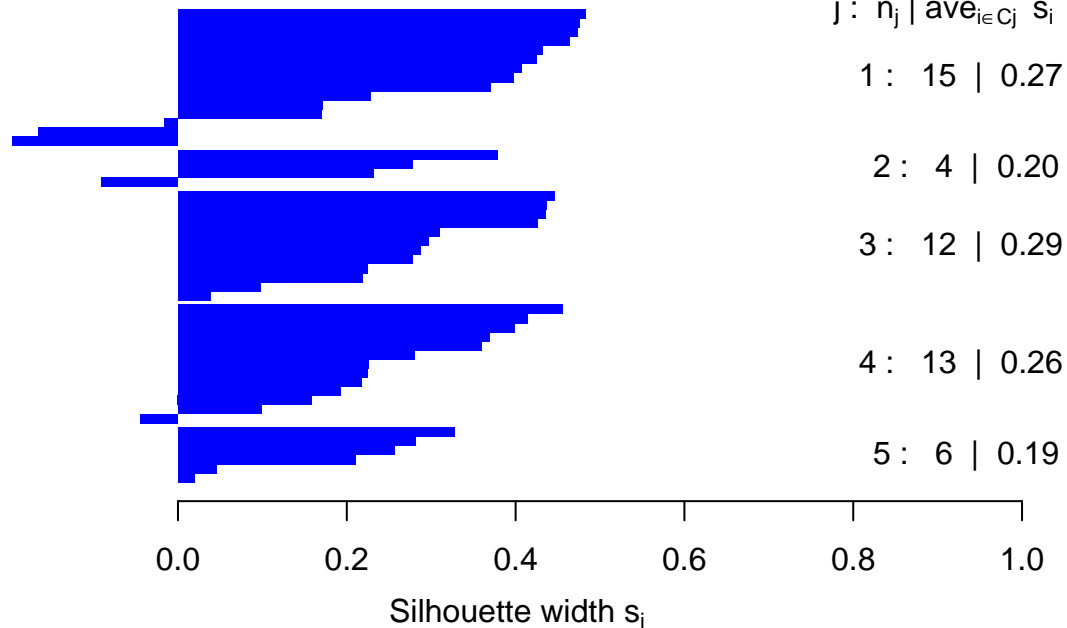
```
dist.Euc<-dist(X.s, method = "euclidean")
Sil.pam<-silhouette(cl.pam, dist.Euc)
```

2.- Generacion del grafico

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

## Silhouette for PAM

n = 50



Average silhouette width : 0.26

Como se puede observar, al aumentar el numero de clusters la capacidad de agrupamiento de Silhouette fue aumentando es por eso que se decidio cambiar de 3 a 6 el numero de clusters ya que asi se amplia la capacidad de agrupamiento, que paso de un .23 con 3 clusters a un .26 con 6 clusters.