

PAM

Melissa Ortega Galarza

2022-06-02

1. Replica

Librerías

```
install.packages("cluster")  
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"
```

Método 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.3<-pam(X.s,3)
```

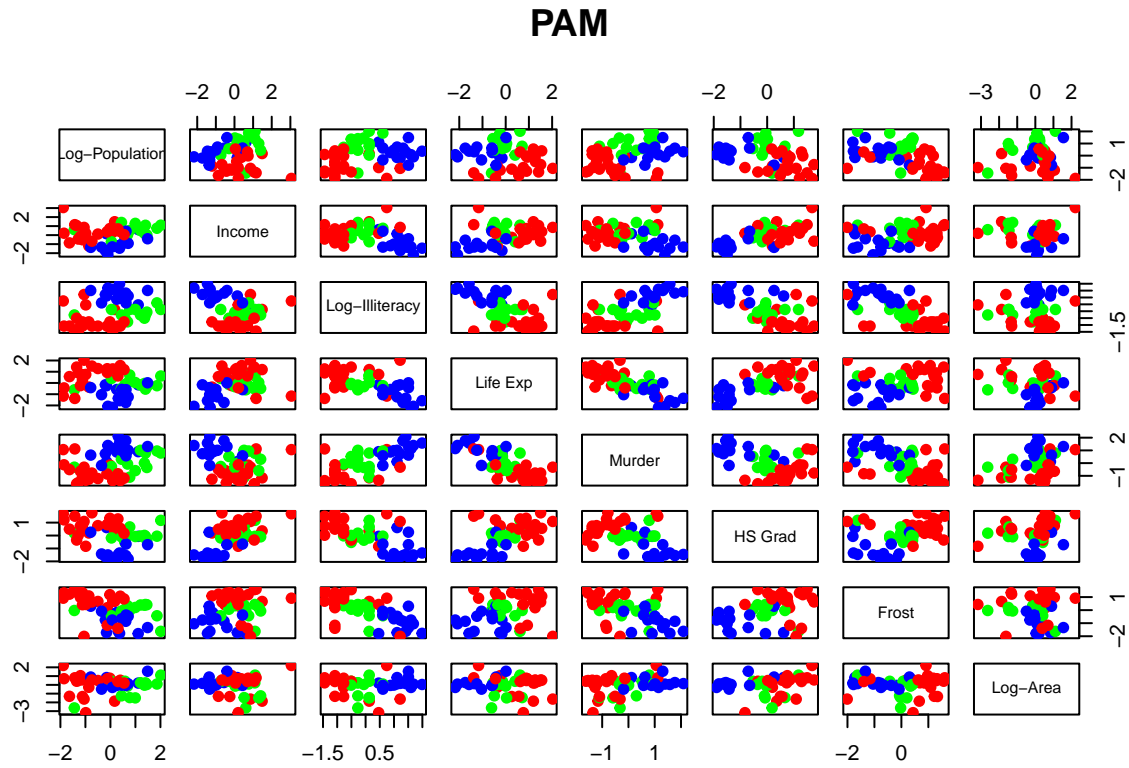
4.- Clusters

```
cl.pam<-pam.3$clustering  
cl.pam
```

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

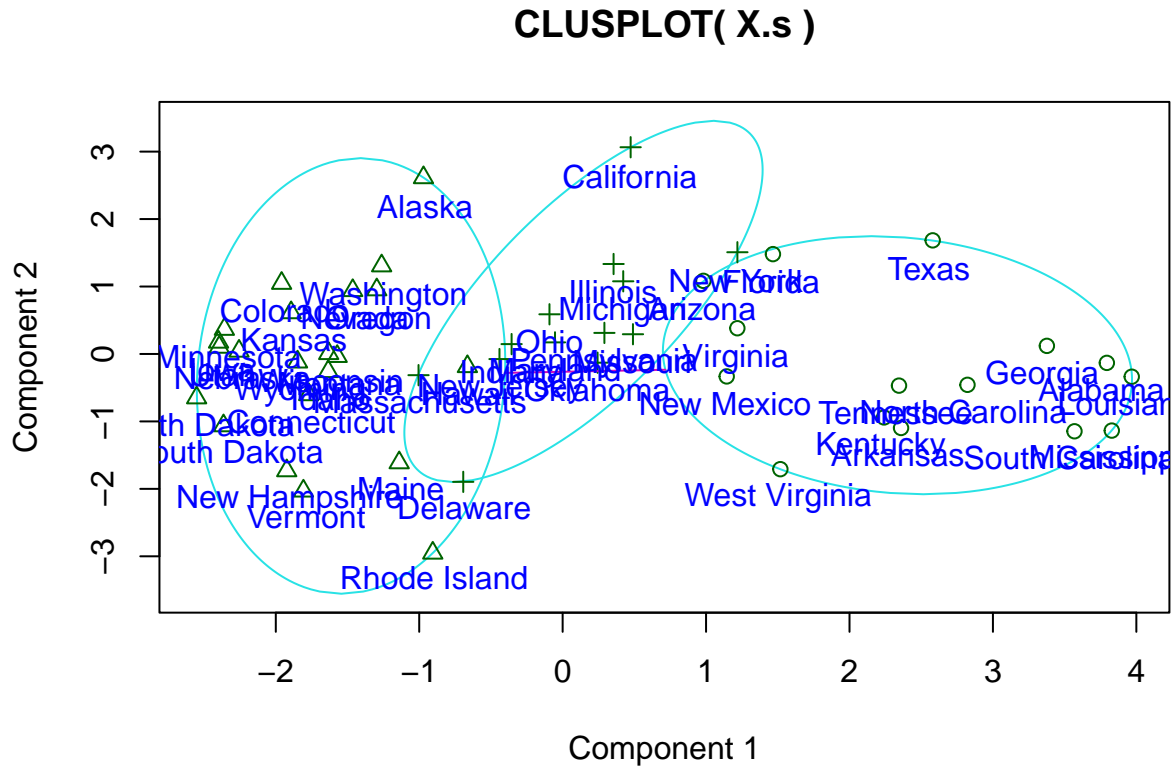
5.- Scatter plot de la matriz con los grupos

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



Visualización 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 grafica de la eficacia de clasificacion de una observación dentro de un grupo.

1.- Generacion de los calculos

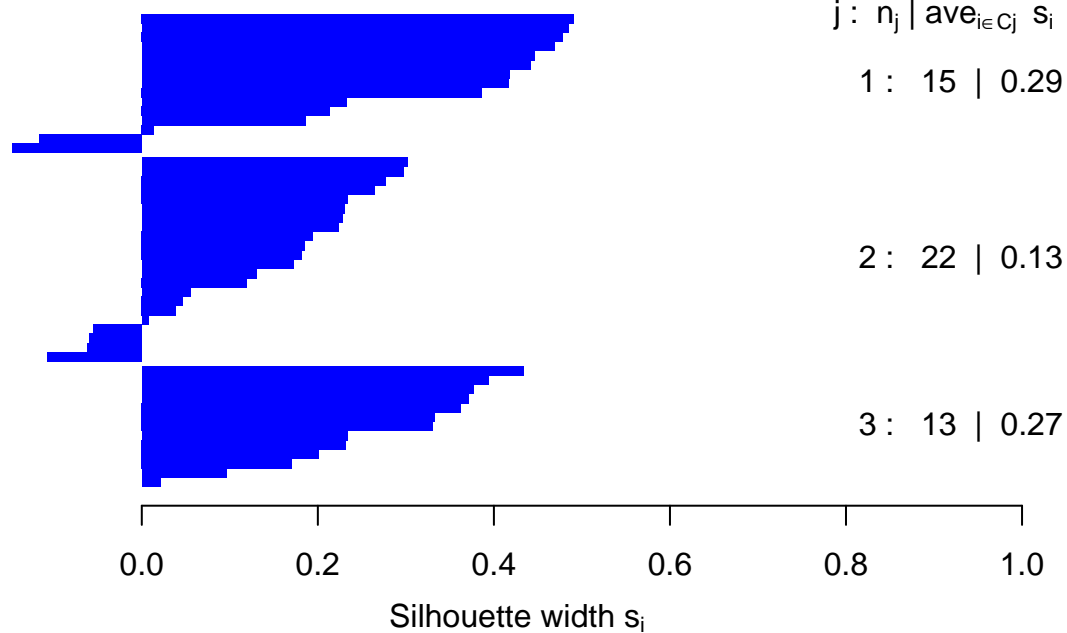
```
dist.Euc<-dist(X.s, method = "euclidean")  
Sil.pam<-silhouette(cl.pam, dist.Euc)
```

2.- Generación del gráfico

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

Silhouette for PAM

n = 50



Average silhouette width : 0.22

2. Cambiar el Número de Clusters.

Librerías

```
install.packages("cluster")
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"
```

Método 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.6<-pam(X.s,6)
```

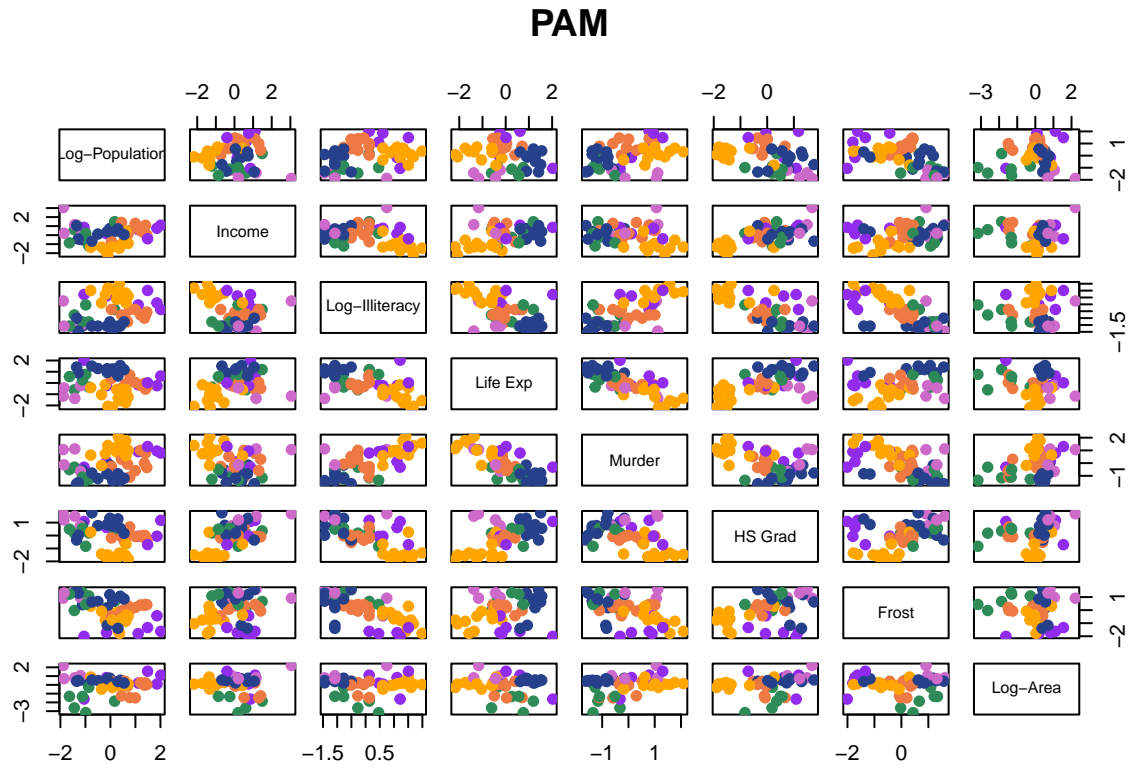
4.- Clusters

```
cl.pam<-pam.6$clustering
cl.pam
```

##	Alabama	Alaska	Arizona	Arkansas	California
##	1	2	3	1	3
##	Colorado	Connecticut	Delaware	Florida	Georgia
##	4	5	5	3	1
##	Hawaii	Idaho	Illinois	Indiana	Iowa
##	3	4	6	6	4
##	Kansas	Kentucky	Louisiana	Maine	Maryland
##	4	1	1	5	6
##	Massachusetts	Michigan	Minnesota	Mississippi	Missouri
##	6	6	4	1	6
##	Montana	Nebraska	Nevada	New Hampshire	New Jersey
##	2	4	2	5	6
##	New Mexico	New York	North Carolina	North Dakota	Ohio
##	1	3	1	4	6
##	Oklahoma	Oregon	Pennsylvania	Rhode Island	South Carolina
##	6	4	6	5	1
##	South Dakota	Tennessee	Texas	Utah	Vermont
##	4	1	3	4	5
##	Virginia	Washington	West Virginia	Wisconsin	Wyoming
##	1	4	1	4	2

5.- Scatter plot de la matriz con los grupos

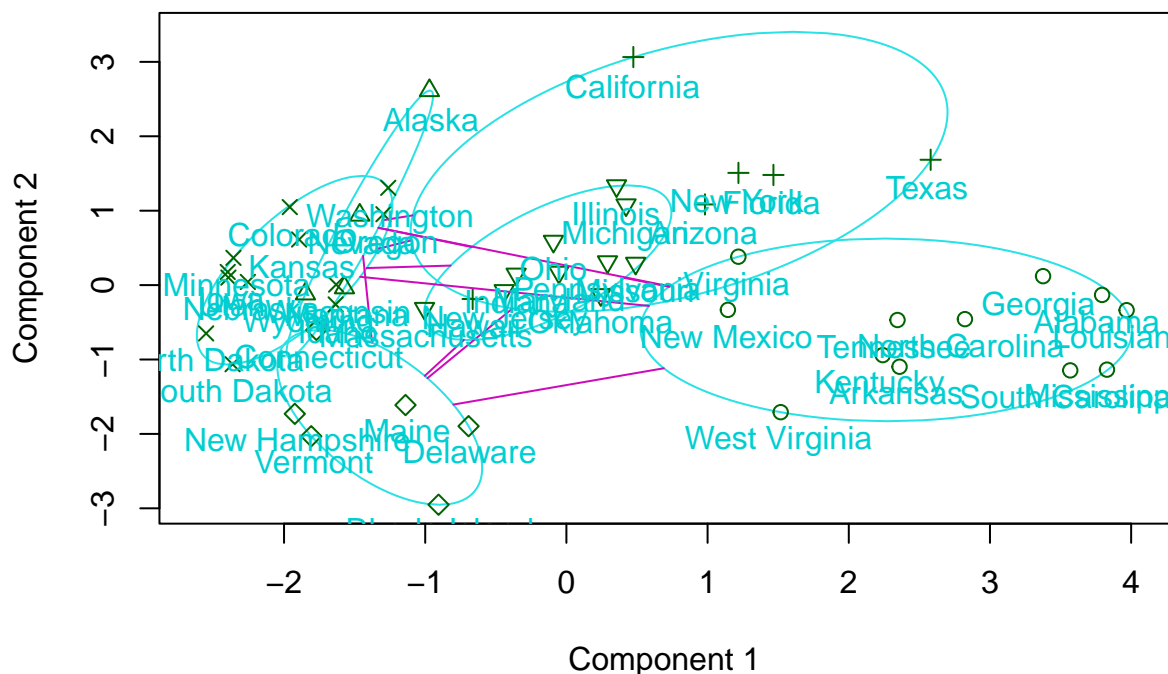
```
col.cluster<-c("orange1","orchid3","purple2","royalblue4","seagreen","sienna2")[cl.pam]
pairs(X.s, col=col.cluster, main="PAM", pch=19)
```



Visualización con Componentes Principales

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

CLUSPLOT(X.s)



These two components explain 62.5 % of the point variability.

Silhouette

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

1.- Generacion de los calculos

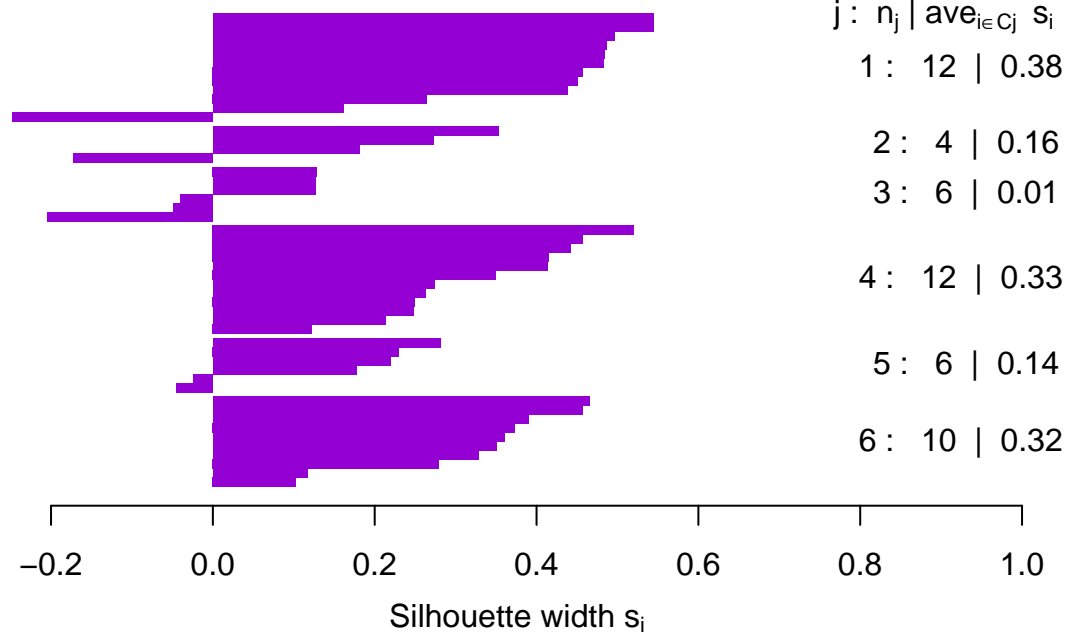
```
dist.Euc<-dist(X.s, method = "euclidean")
Sil.pam<-silhouette(cl.pam, dist.Euc)
```

2.- Generación del gráfico

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

Silhouette for PAM

n = 50



Average silhouette width : 0.27

Análisis

Como podemos observar obtenemos un grafico totalmente si ocupamos 6 clusters.