

# PruebaExcel

Juan Francisco Pallardó Latorre

2024-05-16

```
dataEx = read_excel("rank_prim.xlsx")
```

He eliminado variables como Nombre, Identificador, Latitud, Longitud, barrio, Direccion, NotaReviews, NumReviews y Mixto.

```
eliminar_v = c(1,2,4,5,55,68:70,73)
prueba = dataEx[, -eliminar_v]
```

Elimino las variables constantes: Comedor, Webcam, Internado mixto, Internado especial, Internado femenino, Internado masculino, Chino, Italiano, Portugués, Árabe, Japonés, Griego, Inglés a distancia, Ruso, Euskera, ADU, EP, PEV, PIP, ZC, PPEV, PPEC y PEPLI.

```
eliminar_v = c(6,19:23, 29:38, 40, 44,45,47:50)
prueba = prueba[, -eliminar_v]
```

Arreglo de las variables Inst+Gimnasio, Gimnas+Piscina y Inst+Gimnas+Piscina.

```
for (i in 1:length(prueba$`Inst + Gimnasio`)) {
  if (prueba$`Inst + Gimnasio`[i] == 1) {
    prueba$`Instalaciones deportivas`[i] <- 1
  }
}

for (i in 1:length(prueba$`Inst + Gimnasio`)) {
  if (prueba$`Gimnas + Piscina`[i] == 1) {
    prueba$Piscina[i] <- 1
    prueba$`Inst + Gimnasio`[i] <- 1
  }
}
prueba = subset(prueba, select = - `Gimnas + Piscina`)

for (i in 1:length(prueba$Piscina)) {
  if (prueba$`Inst + Gimnas + Piscina`[i] == 1) {
    prueba$Piscina[i] <- 1
    prueba$`Instalaciones deportivas`[i] <- 1
    prueba$`Inst + Gimnasio`[i] = 1
  }
}
prueba = subset(prueba, select = - `Inst + Gimnas + Piscina`)
names(prueba)[9] = "Gimnasio"
```

Ahora creamos las variables dummy de las variables Tipo, index\_gl\_1, Religion, Precio y Horario. (One-Hot encoding)

```

var_dummy = as.data.frame(model.matrix(~ Tipo - 1, data=prueba))
prueba = cbind(prueba[-which(names(prueba) == "Tipo")], var_dummy)

var_dummy = as.data.frame(model.matrix(~ index_gl_1 - 1, data=prueba))
prueba = cbind(prueba[-which(names(prueba) == "index_gl_1")], var_dummy)

var_dummy = as.data.frame(model.matrix(~ Religion - 1, data=prueba))
prueba = cbind(prueba[-which(names(prueba) == "Religion")], var_dummy)

var_dummy = as.data.frame(model.matrix(~ Precio - 1, data=prueba))
prueba = cbind(prueba[-which(names(prueba) == "Precio")], var_dummy)

var_dummy = as.data.frame(model.matrix(~ Horario - 1, data=prueba))
prueba = cbind(prueba[-which(names(prueba) == "Horario")], var_dummy)

```

Elimino algunas variables dummy que son redundantes como: TipoPúblico, ReligiosoLaico y PrecioEntre 100 y 300€. Además de la variable Huerto Escolar que se repite su información en Huerto.

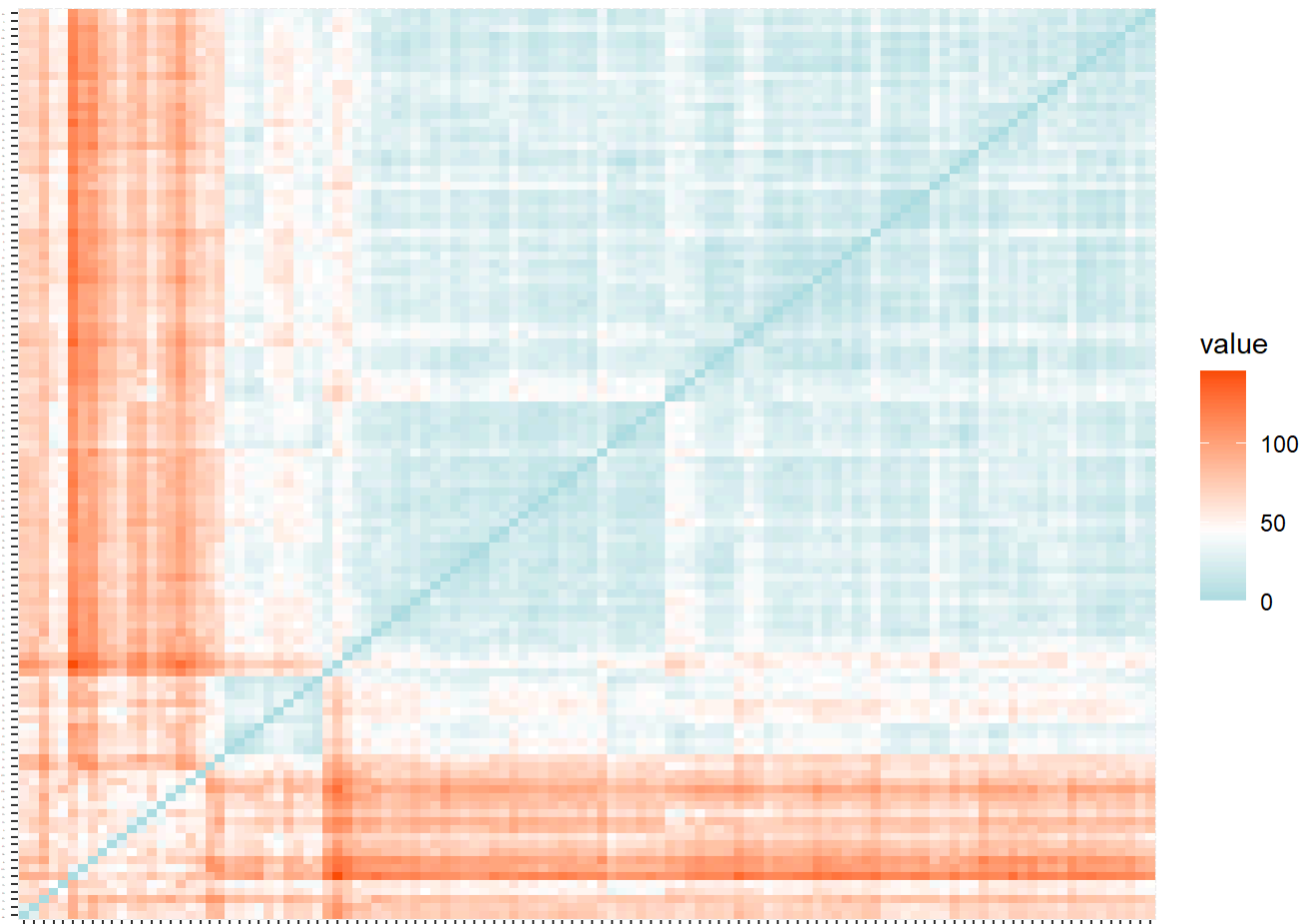
```
prueba_bin = prueba[, -c(12, 59, 63, 65)]
```

Hago el escalado y centrado de los datos, usare la distancia de Manhattan (ver diferencias con Euclídea).  
Printeo una matriz de todas las instancias y su distancia.

```

schools_cluster = scale(prueba_bin, center=TRUE, scale=TRUE)
midist <- get_dist(schools_cluster, stand = FALSE, method = "manhattan")
fviz_dist(midist, show_labels = TRUE, lab_size = 0.3,
          gradient = list(low = "#00AFBB", mid = "white", high = "#FC4E07"))

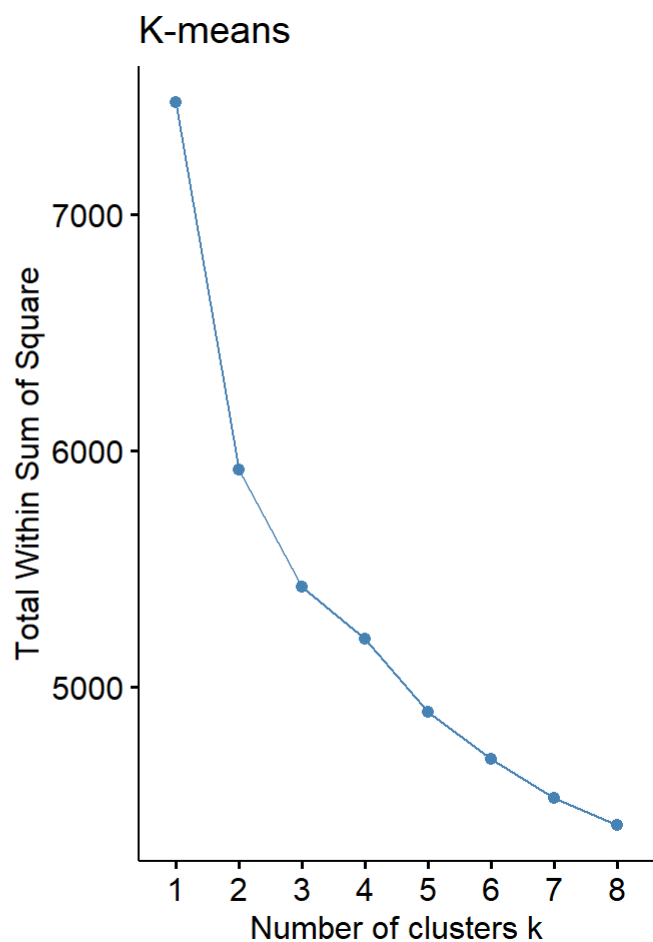
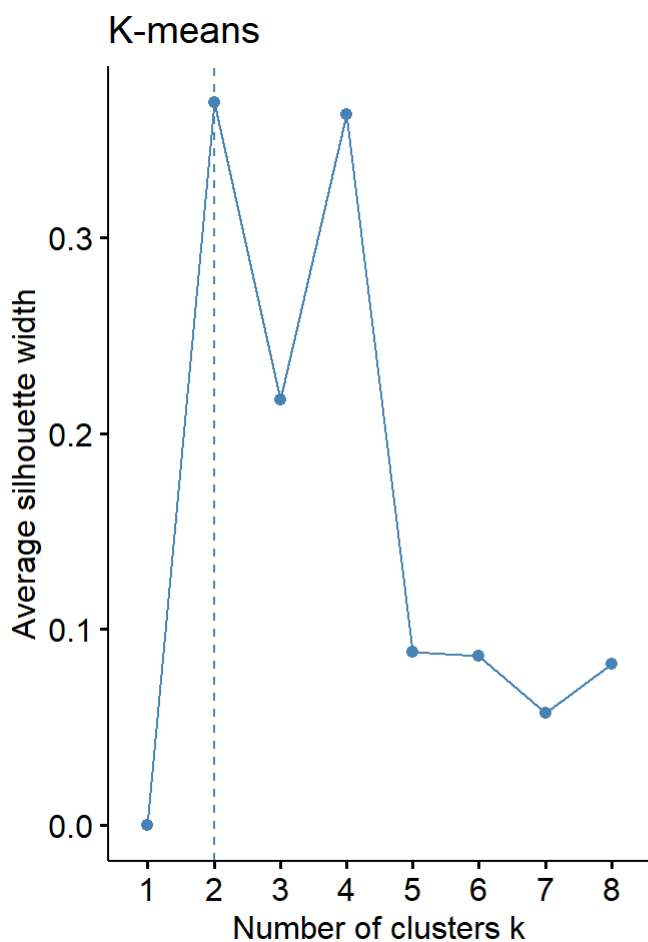
```



En el gráfico obtenido a partir de las distancias podemos observar posibles agrupaciones entre centros que serían los cuadrados azules que se forman a través de la diagonal principal de la matriz. El siguiente paso será realizar un método de partición, en concreto realizaremos el algoritmo **k-medias**. Sin embargo, previo a realizarlo hay que determinar el número óptimo de clusters. Para ello nos basaremos en el coeficiente de Silhouette (A mayor más relación entre clusters) y en la Suma de Cuadrados Residual (A menor mejor).

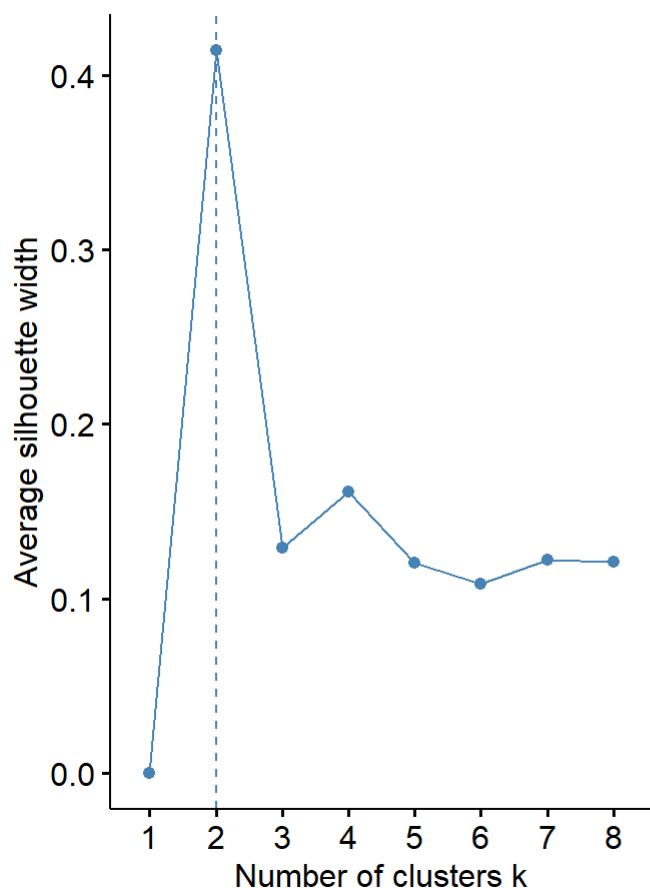
```
p1 = fviz_nbclust(x = schools_cluster, FUNcluster = kmeans, method = "silhouette",
                  k.max = 8, verbose = FALSE) +
  labs(title = "K-means")
p2 = fviz_nbclust(x = schools_cluster, FUNcluster = kmeans, method = "wss",
                  k.max = 8, verbose = FALSE) +
  labs(title = "K-means")
p3 = fviz_nbclust(x = schools_cluster, FUNcluster = pam, method = "silhouette",
                  k.max = 8, verbose = FALSE, diss = midist) +
  labs(title = "K-Medoides - Silhouette")
p4 = fviz_nbclust(x = schools_cluster, FUNcluster = pam, method = "wss",
                  k.max = 8, verbose = FALSE, diss = midist) +
  labs(title = "K-Medoides - Sum")

grid.arrange(p1,p2,nrow = 1)
```

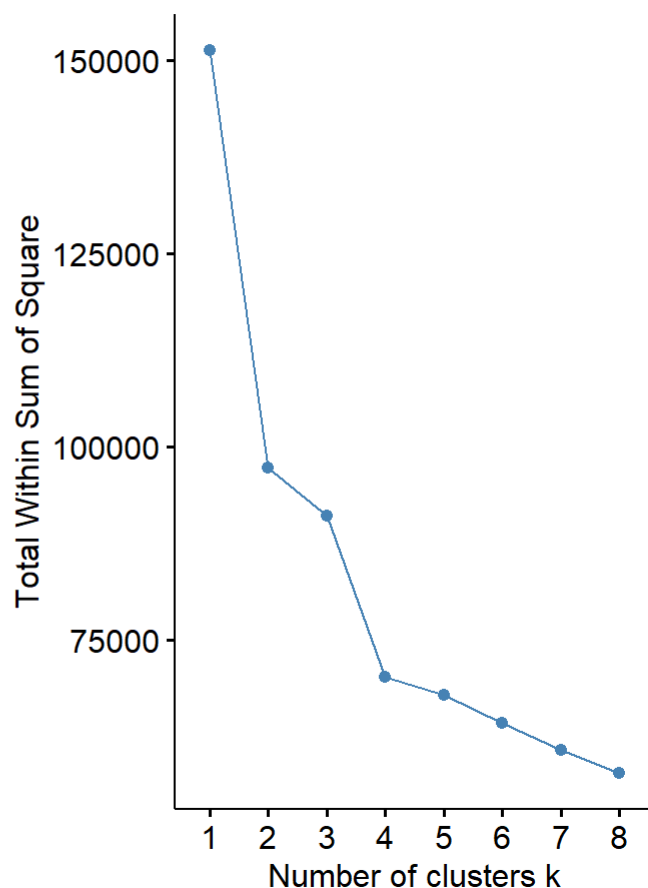


```
grid.arrange(p3,p4,nrow = 1)
```

K-Medoides - Silhouette



K-Medoides - Sum

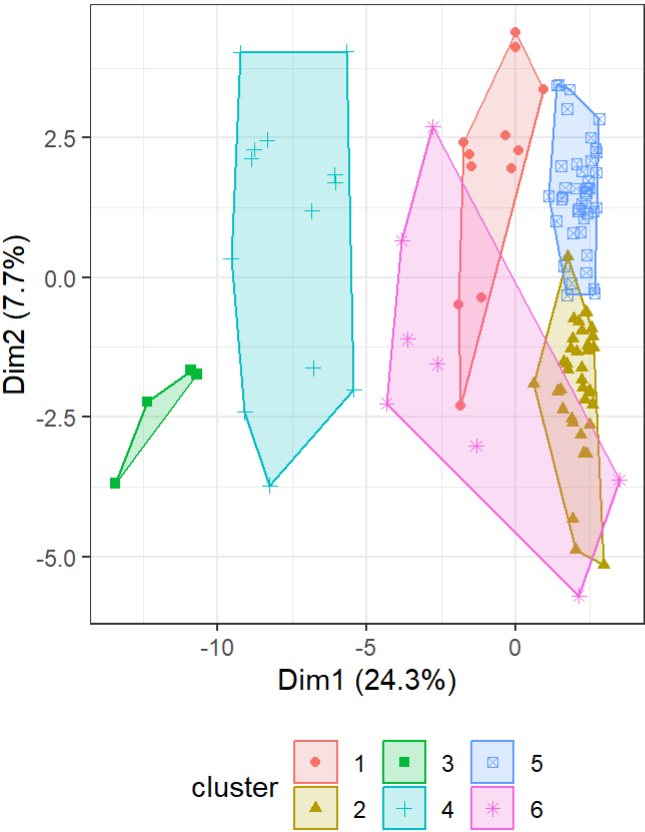


```
set.seed(100)
clust3 <- kmeans(midist, centers = 6, nstart = 20)
table(clust3$cluster)
```

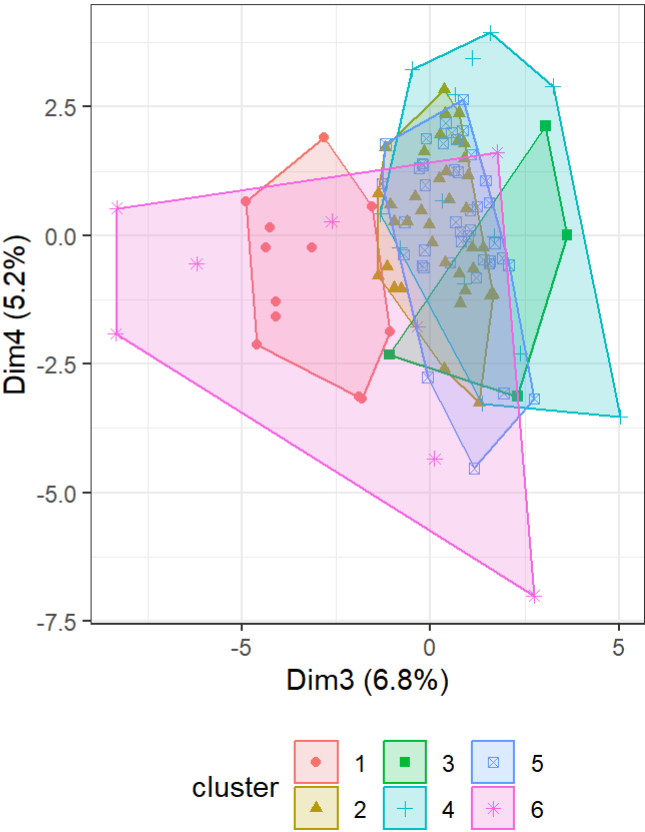
```
##
##  1  2  3  4  5  6
## 12 37  4 13 42  8
```

```
p1 = fviz_cluster(object = list(data=schools_cluster, cluster=clust3$cluster), stand = FALSE,
  ellipse.type = "convex", geom = "point", show.clust.cent = FALSE,
  labelsize = 8) +
  labs(title = "K-MEDIAS + Proyeccion PCA",
    subtitle = "Dist manhattan, K=4") +
  theme_bw() +
  theme(legend.position = "bottom")
p2 = fviz_cluster(object = list(data=schools_cluster, cluster=clust3$cluster), stand = FALSE,
  ellipse.type = "convex", geom = "point", show.clust.cent = FALSE,
  labelsize = 8, axes = 3:4) +
  labs(title = "K-MEDIAS + Proyeccion PCA",
    subtitle = "Dist manhattan, K=4") +
  theme_bw() +
  theme(legend.position = "bottom")
grid.arrange(p1, p2, nrow = 1)
```

K-MEDIAS + Proyeccion PCA  
Dist manhattan, K=4

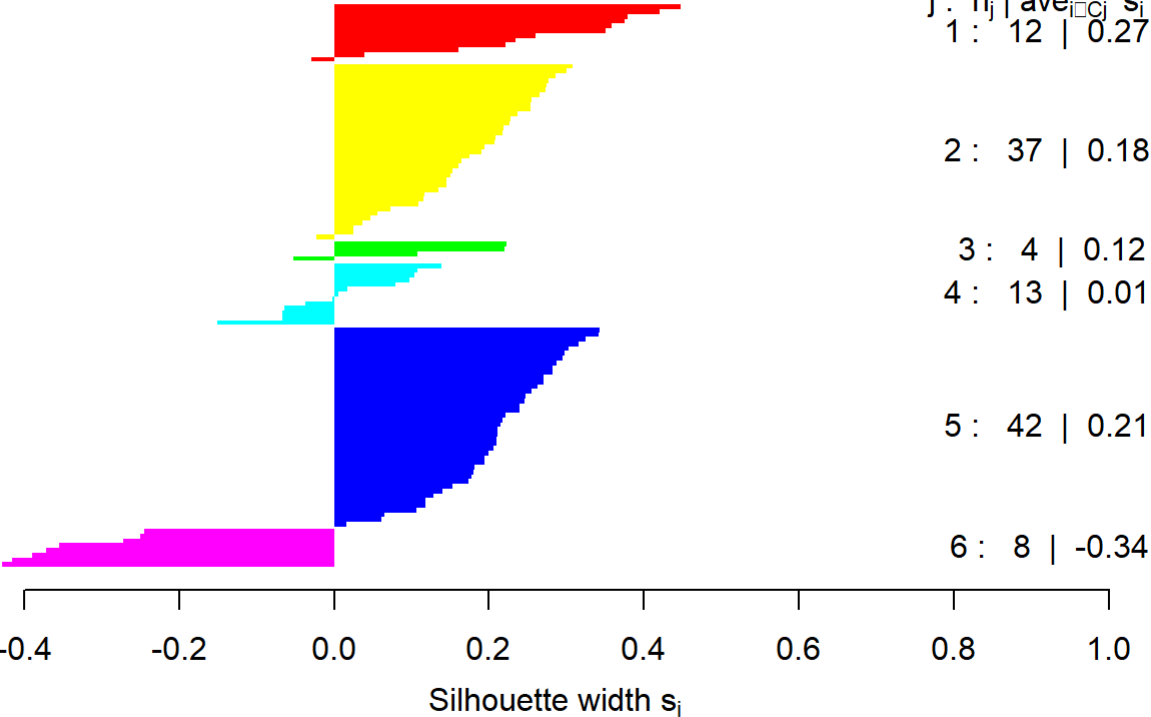


K-MEDIAS + Proyeccion PCA  
Dist manhattan, K=4



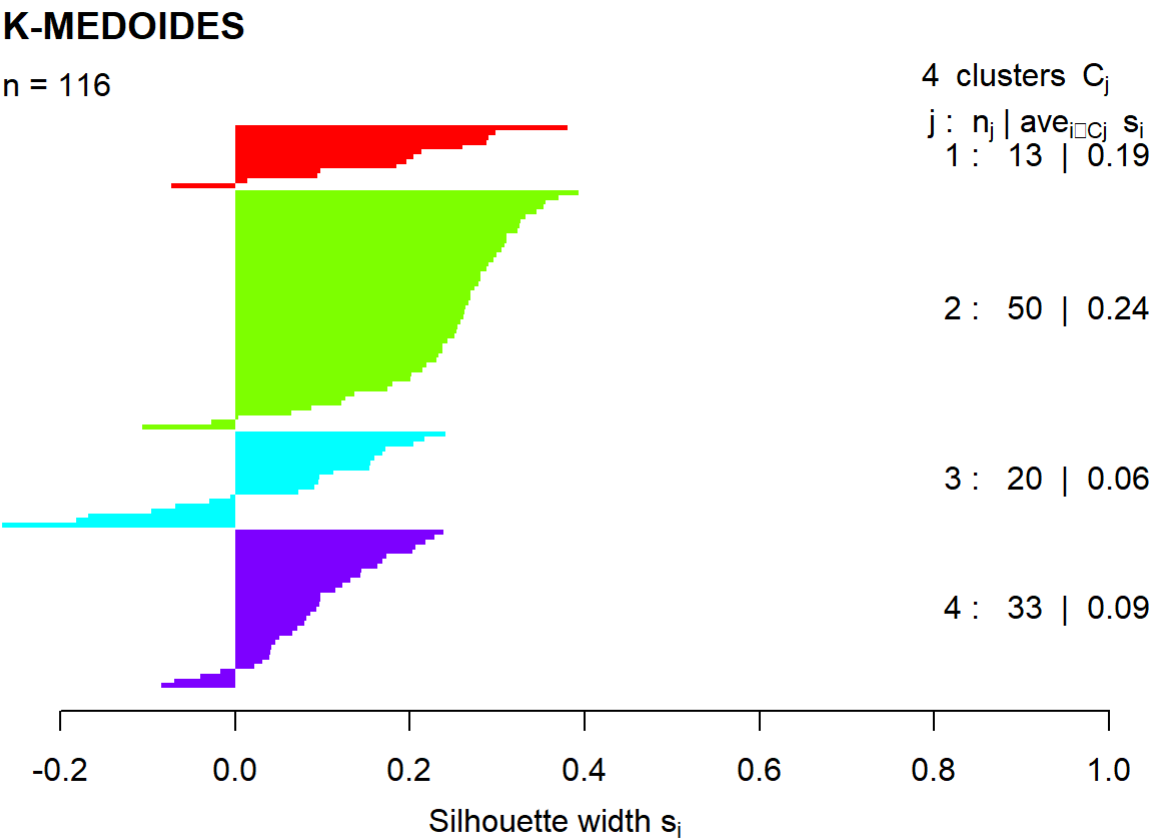
```
plot(silhouette(clust3$cluster, midist), col=rainbow(6), border=NA, main = "K-MEDIAS")
```

K-MEDIAS  
n = 116



Average silhouette width : 0.14

```
clust4 <- pam(schools_cluster, k = 4)
plot(silhouette(clust4$clustering, midist), col=rainbow(4), border=NA, main = "K-MEDOIDES")
```



```
sil = data.frame(silhouette(clust3$cluster, midist))

mal = sil$sil_width < -0.045
malClasifiS = sil[mal,]
```

```
misclust = factor(clust3$cluster)

mediasCluster = aggregate(schools_cluster, by = list("cluster" = misclust), mean)[-1]
rownames(mediasCluster) = paste0("c",1:6)
kable(t(round(mediasCluster,2)))
```

	c1	c2	c3	c4	c5	c6
Biblioteca	0.13	-0.07	0.13	0.13	0.13	-0.82
Laboratorio / taller	-0.02	-0.01	0.29	0.00	0.11	-0.64
Aula de informática	0.68	-0.18	0.68	0.68	-0.39	0.41
Aula de música	0.50	-0.36	0.68	0.52	-0.13	0.41
Instalaciones deportivas	0.16	0.16	0.16	0.16	0.01	-1.41
Transporte	0.29	0.04	1.70	0.24	-0.35	-0.06
Cafetería	-0.19	-0.19	1.18	0.23	-0.19	1.18

	<b>c1</b>	<b>c2</b>	<b>c3</b>	<b>c4</b>	<b>c5</b>	<b>c6</b>
Gimnasio	0.74	-0.35	1.25	0.62	-0.31	0.48
Salón de actos	0.74	-0.51	1.45	1.29	-0.48	0.92
Pilota	-0.13	-0.13	1.78	-0.13	-0.13	0.82
Cursos de verano	-0.09	-0.09	-0.09	-0.09	-0.09	1.25
Aula de idiomas	-0.01	-0.11	1.47	0.51	-0.22	0.14
Piscina	-0.09	-0.09	-0.09	-0.09	-0.09	1.25
Castellano	-0.21	0.05	1.01	0.17	-0.21	0.40
Valenciano	-0.23	0.01	0.89	0.46	-0.23	0.33
Inglés	0.09	0.09	0.09	0.09	0.09	-1.25
Francés	0.56	-0.76	0.56	0.56	0.28	-0.02
Alemán	0.40	-0.37	1.16	1.04	-0.37	0.78
EI	-0.36	0.16	-1.41	0.16	0.01	0.16
BACH	0.96	-0.50	2.00	1.04	-0.50	0.75
FP	0.44	-0.25	-0.25	0.71	-0.25	0.79
SEC	1.40	-0.63	1.58	1.41	-0.58	0.75
PIL	-0.09	-0.09	-0.09	-0.09	-0.09	1.25
zonas_verdes	-0.16	0.17	0.08	-0.13	-0.06	-0.09
densidad_poblacion	0.30	-0.53	-0.10	-0.26	0.50	-0.13
turismos_e	-0.17	0.28	-0.21	-0.16	-0.08	-0.26
paro	0.41	-0.61	-0.34	0.01	0.50	-0.29
renta_media	-0.16	0.47	0.92	-0.18	-0.54	0.69
riesgo_pobreza	-0.02	-0.24	-0.26	-0.01	0.25	-0.07
index_equi	0.13	-0.06	0.62	0.42	-0.20	0.14
index_soci	-0.33	0.64	0.74	-0.39	-0.59	0.89
index_glob	-0.39	0.68	0.46	-0.24	-0.64	0.95
bus_count	0.30	-0.21	-0.25	0.02	0.14	-0.11
metro_count	0.45	-0.13	-0.19	0.20	0.04	-0.53
Futbol	-0.38	-0.38	2.58	2.13	-0.38	-0.38
Baloncesto	-0.37	-0.37	2.69	1.98	-0.37	-0.37
Baile	-0.41	-0.41	2.40	1.97	-0.41	0.29
Ajedrez	-0.34	-0.34	2.93	1.42	-0.34	0.07
Patinaje	-0.34	-0.34	2.93	1.42	-0.34	0.07

	c1	c2	c3	c4	c5	c6
Guitarra	-0.32	-0.32	3.08	1.51	-0.32	-0.32
Judo	-0.32	-0.32	2.23	1.77	-0.32	-0.32
Teatro	-0.27	-0.27	2.68	0.94	-0.27	0.22
Manualidades	-0.27	-0.27	2.68	0.94	-0.27	0.22
Programacion	-0.21	-0.21	2.24	0.92	-0.21	-0.21
ComedorCocinapropia	1.23	-0.64	1.23	1.23	-0.42	0.72
ComedorCateringExterno	-1.23	0.64	-1.23	-1.23	0.42	-0.72
Becas	-0.46	0.07	1.00	0.03	-0.02	-0.09
Excursiones_Convivencias	-0.43	-0.43	2.32	1.90	-0.43	0.60
Enfermeria	-0.07	-0.16	2.11	0.42	-0.18	0.07
Huerto	-0.31	-0.31	3.24	1.06	-0.31	0.14
PatioInfantil	-0.41	-0.41	2.40	1.97	-0.41	0.29
CentroTecnologico	-0.51	-0.11	0.72	1.00	-0.10	-0.20
GoogleClassroom	-0.37	-0.37	1.16	2.22	-0.37	0.01
Teams	-0.27	-0.27	2.68	1.24	-0.27	-0.27
HerramientasPropiasTecnologicas	-0.25	-0.25	2.88	1.03	-0.25	-0.25
NecesidadesEspecialesAlumnos	-0.45	-0.45	2.18	2.18	-0.45	0.53
TipoPrivado	1.51	-0.65	1.51	1.51	-0.60	0.70
index_gl_1No Vulnerable	-0.36	0.87	1.14	-0.25	-0.87	0.89
index_gl_1Pot. Vulnerable	0.49	-0.26	-0.52	0.04	0.23	-0.52
index_gl_1Vulnerable	-0.04	-0.68	-0.74	0.23	0.70	-0.48
ReligionReligioso	1.38	-0.56	1.77	1.23	-0.56	0.60
PrecioGratis o <100€	0.09	0.09	-2.60	0.09	0.09	0.09
HorarioHorario ampliado	1.28	-0.16	-0.15	-0.35	-0.22	0.65
HorarioHorario ampliado de mañana	-0.21	-0.08	-0.21	0.92	-0.09	-0.21
HorarioHorario ampliado de mañana y tarde	-1.15	0.19	0.23	-0.05	0.26	-0.54

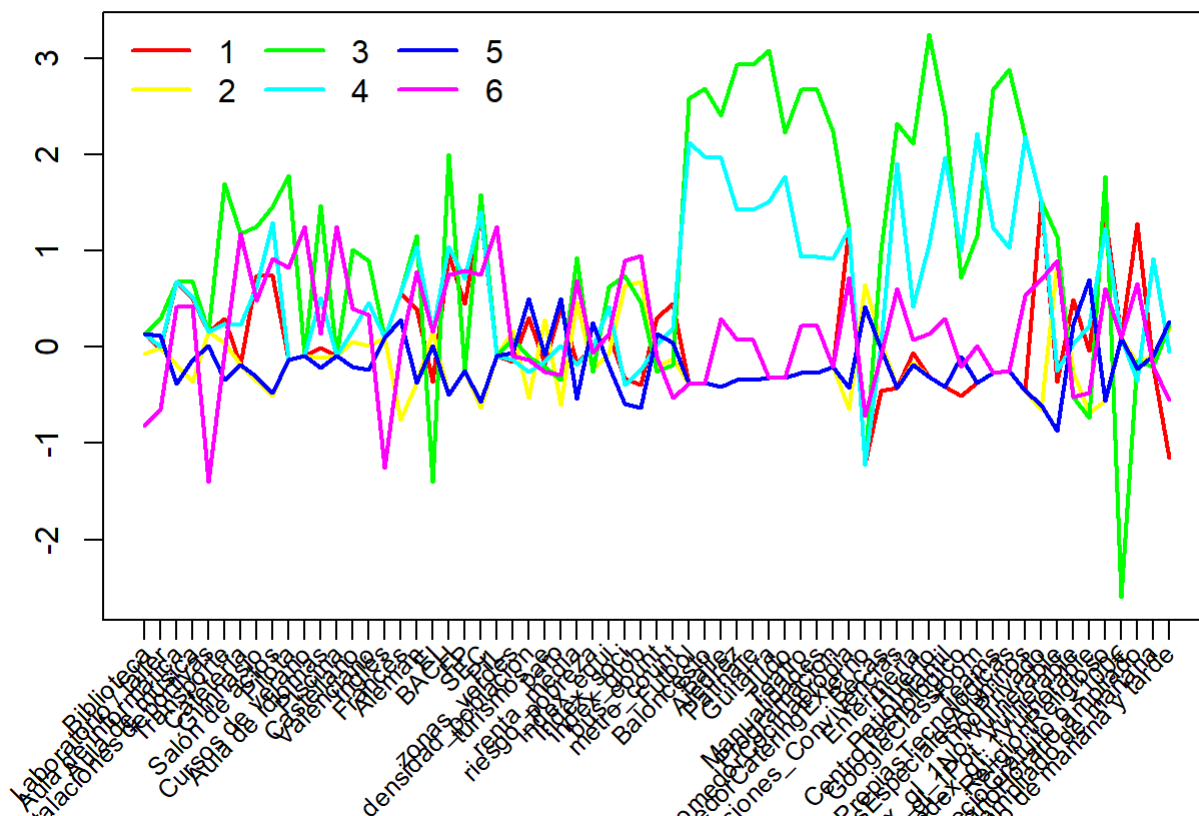
```

par(mar = c(5, 4, 4, 2) + 0.1)
matplot(t(mediasCluster), type = "l", col = rainbow(6), ylab = "", xlab = "", lwd = 2,
        lty = 1, main = "Perfil medio de los clusters", xaxt = "n")
axis(side = 1, at = 1:ncol(schools_cluster), labels = FALSE)
text(x = 1:ncol(schools_cluster), y = par("usr")[3] - 0.3,
     labels = colnames(schools_cluster), srt = 45, adj = 1, xpd = TRUE, cex = 0.8)
legend("topleft", as.character(1:6), col = rainbow(6), lwd = 2, ncol = 3, bty = "n")

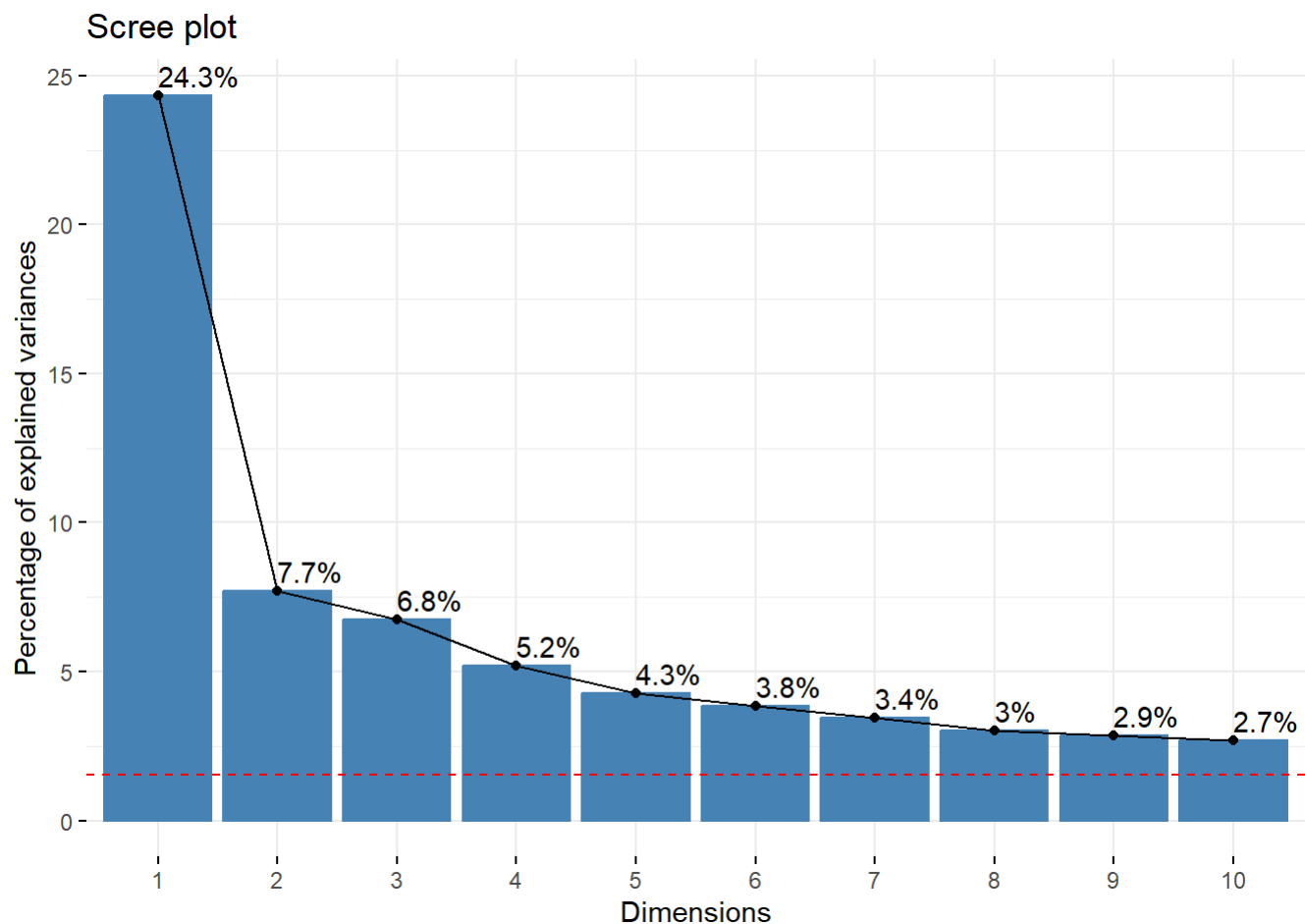
```



## Perfil medio de los clusters



```
res.pca = PCA(prueba_bin, scale.unit = TRUE, graph = FALSE, ncp = 10)
eig.val <- get_eigenvalue(res.pca)
VPmedio = 100 * (1/nrow(eig.val))
fviz_eig(res.pca, addlabels = TRUE) +
  geom_hline(yintercept=VPmedio, linetype=2, color="red")
```



```
kable(eig.val[1:6,])
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	15.825838	24.347443	24.34744
Dim.2	5.004237	7.698827	32.04627
Dim.3	4.387578	6.750121	38.79639
Dim.4	3.387804	5.212006	44.00840
Dim.5	2.775554	4.270082	48.27848
Dim.6	2.497510	3.842322	52.12080

K = 4

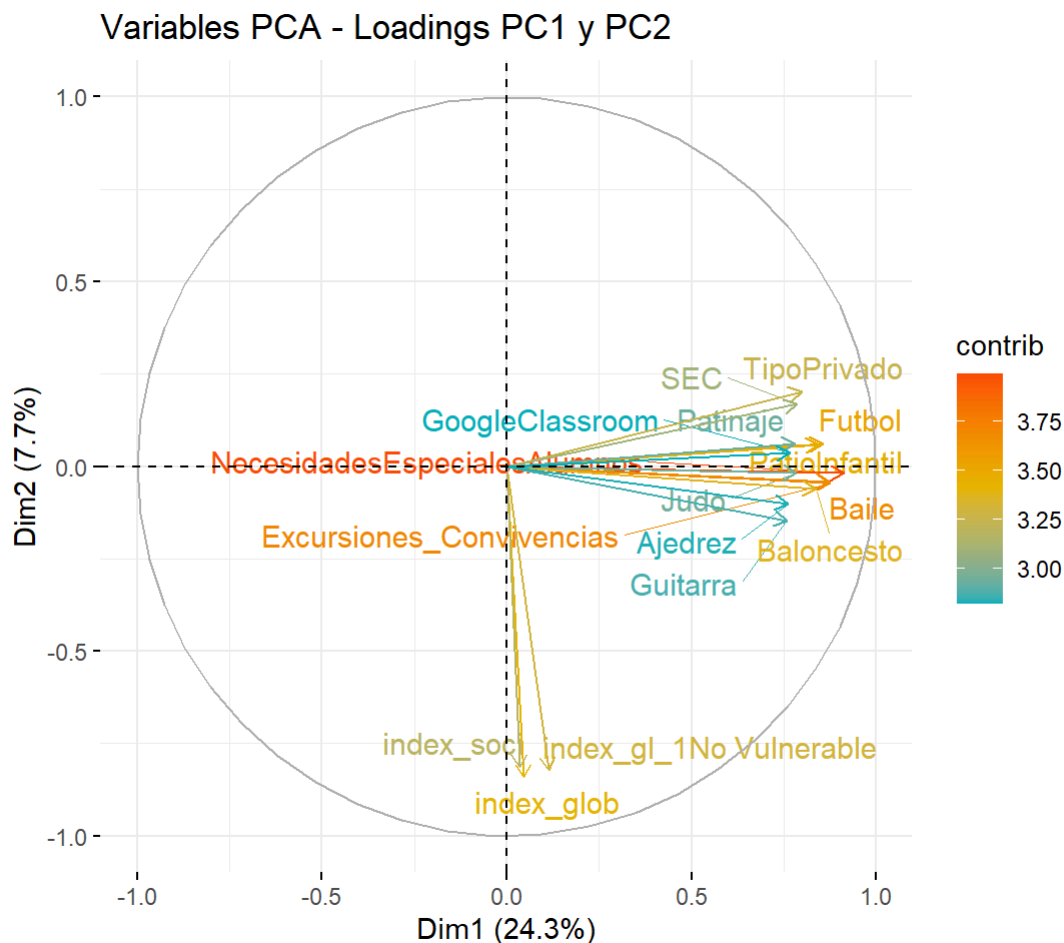
```
res.pca = PCA(prueba_bin, scale.unit = TRUE, graph = FALSE, ncp = K)
```

```
fviz_pca_var(res.pca, axes = c(1,2), repel = TRUE, col.var = "contrib",
```

```
  select.var = list(contrib = 16) ,
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
```

```
  labelsiz = 4,
```

```
  title = 'Variables PCA - Loadings PC1 y PC2')
```

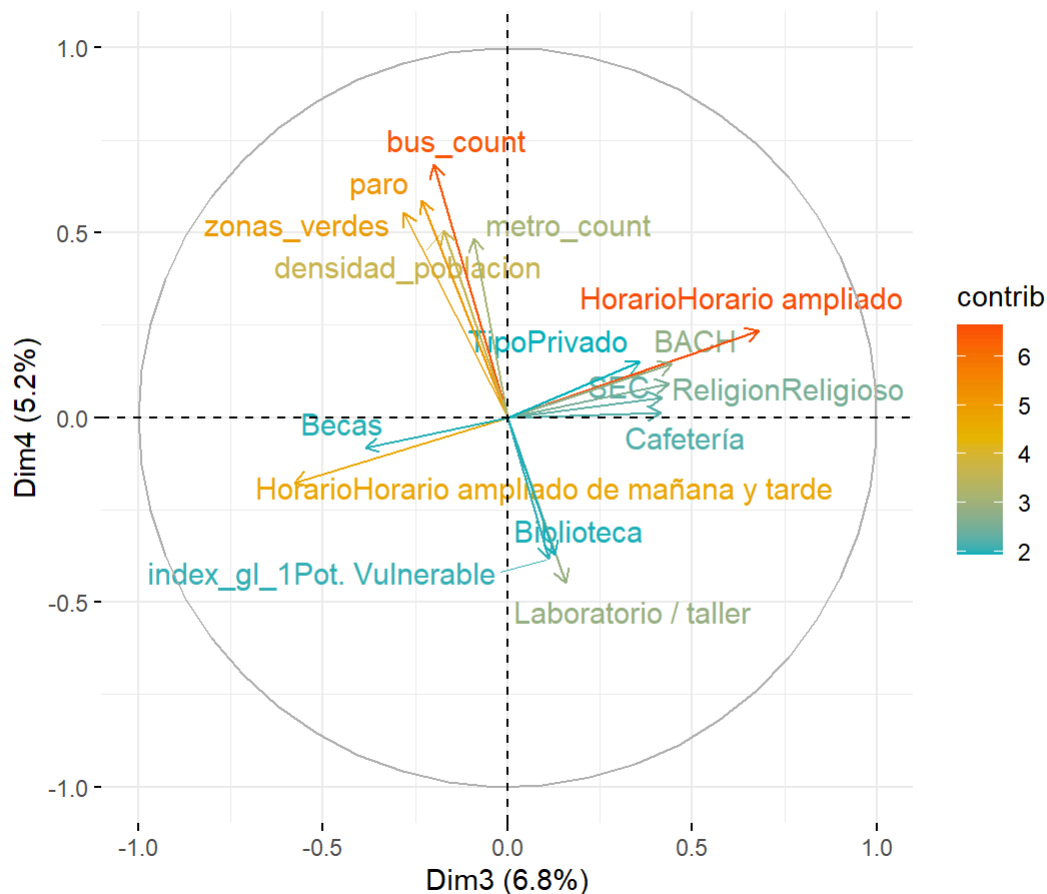


```
fviz_pca_var(res.pca, axes = c(3,4), repel = TRUE, col.var = "contrib",
  select.var = list(contrib=16),
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

  labelsiz = 4,

  title = 'Variables PCA - Loadings PC3 y PC4')
```

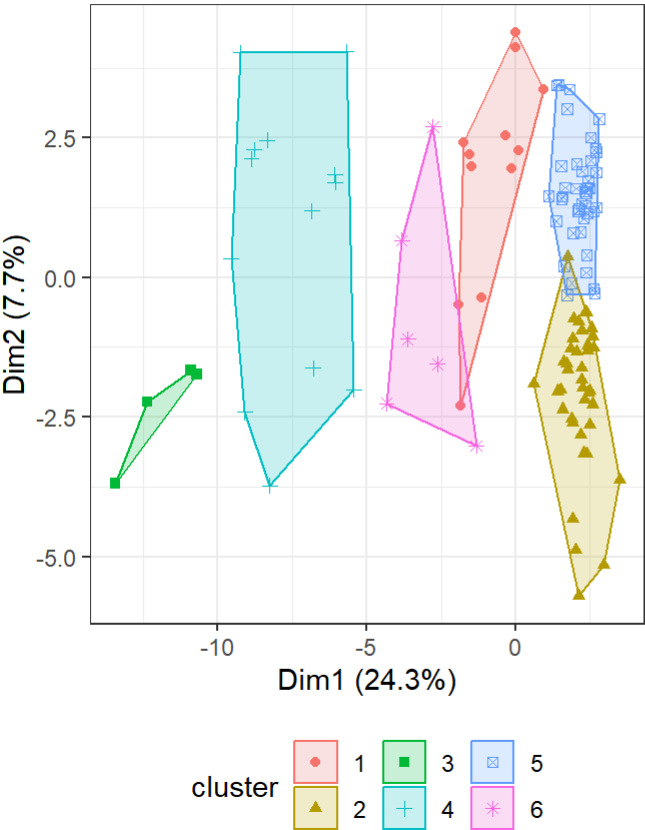
## Variables PCA - Loadings PC3 y PC4



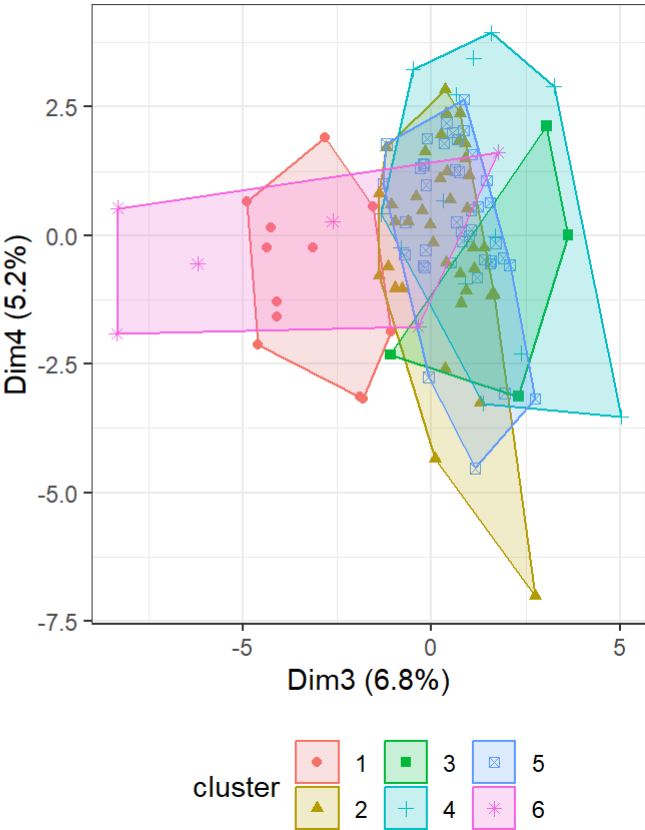
```
n_clusters = as.data.frame(clust3$cluster)
```

```
clust3$cluster[89] = 2
clust3$cluster[91] = 2
p1 = fviz_cluster(object = list(data=schools_cluster, cluster=clust3$cluster), stand = FALSE,
  ellipse.type = "convex", geom = "point", show.clust.cent = FALSE,
  labelsize = 8) +
  labs(title = "K-MEDIAS + Proyeccion PCA",
    subtitle = "Dist manhattan, K=4") +
  theme_bw() +
  theme(legend.position = "bottom")
p2 = fviz_cluster(object = list(data=schools_cluster, cluster=clust3$cluster), stand = FALSE,
  ellipse.type = "convex", geom = "point", show.clust.cent = FALSE,
  labelsize = 8, axes = 3:4) +
  labs(title = "K-MEDIAS + Proyeccion PCA",
    subtitle = "Dist manhattan, K=4") +
  theme_bw() +
  theme(legend.position = "bottom")
grid.arrange(p1, p2, nrow = 1)
```

K-MEDIAS + Proyeccion PCA  
Dist manhattan, K=4

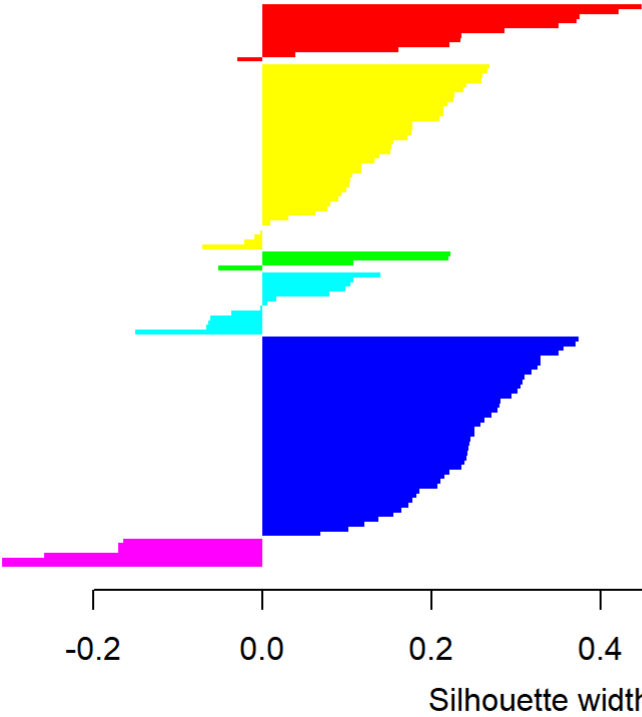


K-MEDIAS + Proyeccion PCA  
Dist manhattan, K=4



```
plot(silhouette(clust3$cluster, midist), col=rainbow(6), border=NA, main = "K-MEDIAS")
```

K-MEDIAS  
n = 116



6 clusters  $C_j$

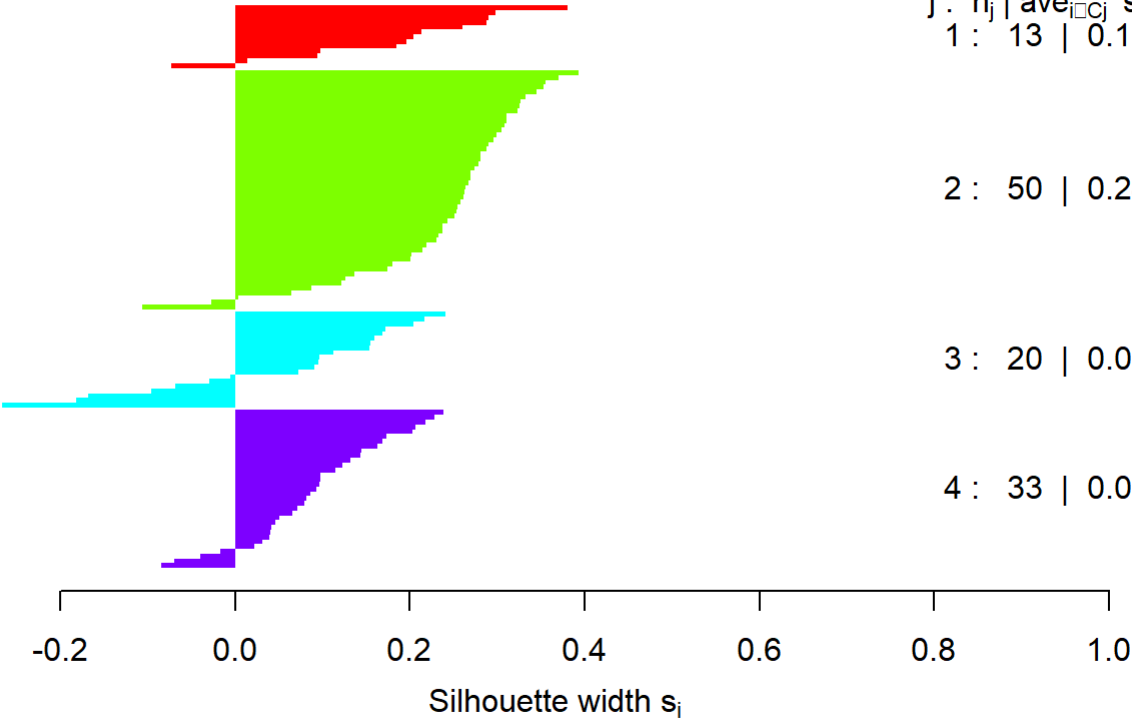
j	$n_j$	ave $s_i$
1	12	0.26
2	39	0.13
3	4	0.12
4	13	0.01
5	42	0.25
6	6	-0.23

Average silhouette width : 0.16

```
clust4 <- pam(schools_cluster, k = 4)
plot(silhouette(clust4$clustering, midist), col=rainbow(4), border=NA, main = "K-MEDOIDES")
```

K-MEDOIDES

n = 116



```
sil = data.frame(silhouette(clust3$cluster, midist))

mal = sil$sil_width < -0.045
malClasifiS = sil[mal,]
```

```
misclust = factor(clust3$cluster)

mediasCluster = aggregate(schools_cluster, by = list("cluster" = misclust), mean)[,-1]
rownames(mediasCluster) = paste0("c",1:6)
kable(t(round(mediasCluster,2)))
```

	c1	c2	c3	c4	c5	c6
Biblioteca	0.13	-0.26	0.13	0.13	0.13	0.13
Laboratorio / taller	-0.02	-0.19	0.29	0.00	0.11	0.29
Aula de informática	0.68	-0.20	0.68	0.68	-0.39	0.68
Aula de música	0.50	-0.36	0.68	0.52	-0.13	0.68
Instalaciones deportivas	0.16	-0.16	0.16	0.16	0.01	0.16
Transporte	0.29	0.02	1.70	0.24	-0.35	0.06
Cafetería	-0.19	-0.19	1.18	0.23	-0.19	1.63

	<b>c1</b>	<b>c2</b>	<b>c3</b>	<b>c4</b>	<b>c5</b>	<b>c6</b>
Gimnasio	0.74	-0.37	1.25	0.62	-0.31	0.91
Salón de actos	0.74	-0.52	1.45	1.29	-0.48	1.45
Pilota	-0.13	-0.13	1.78	-0.13	-0.13	1.14
Cursos de verano	-0.09	-0.09	-0.09	-0.09	-0.09	1.70
Aula de idiomas	-0.01	-0.12	1.47	0.51	-0.22	0.29
Piscina	-0.09	-0.09	-0.09	-0.09	-0.09	1.70
Castellano	-0.21	0.17	1.01	0.17	-0.21	-0.21
Valenciano	-0.23	0.11	0.89	0.46	-0.23	-0.23
Inglés	0.09	-0.18	0.09	0.09	0.09	0.09
Francés	0.56	-0.81	0.56	0.56	0.28	0.56
Alemán	0.40	-0.37	1.16	1.04	-0.37	1.16
EI	-0.36	0.16	-1.41	0.16	0.01	0.16
BACH	0.96	-0.50	2.00	1.04	-0.50	1.17
FP	0.44	-0.25	-0.25	0.71	-0.25	1.14
SEC	1.40	-0.63	1.58	1.41	-0.58	1.21
PIL	-0.09	-0.09	-0.09	-0.09	-0.09	1.70
zonas_verdes	-0.16	0.18	0.08	-0.13	-0.06	-0.24
densidad_poblacion	0.30	-0.48	-0.10	-0.26	0.50	-0.32
turismos_e	-0.17	0.25	-0.21	-0.16	-0.08	-0.24
paro	0.41	-0.57	-0.34	0.01	0.50	-0.43
renta_media	-0.16	0.53	0.92	-0.18	-0.54	0.43
riesgo_pobreza	-0.02	-0.24	-0.26	-0.01	0.25	0.02
index_equi	0.13	-0.07	0.62	0.42	-0.20	0.30
index_soci	-0.33	0.68	0.74	-0.39	-0.59	0.68
index_glob	-0.39	0.71	0.46	-0.24	-0.64	0.85
bus_count	0.30	-0.15	-0.25	0.02	0.14	-0.49
metro_count	0.45	-0.14	-0.19	0.20	0.04	-0.60
Futbol	-0.38	-0.38	2.58	2.13	-0.38	-0.38
Baloncesto	-0.37	-0.37	2.69	1.98	-0.37	-0.37
Baile	-0.41	-0.41	2.40	1.97	-0.41	0.53
Ajedrez	-0.34	-0.34	2.93	1.42	-0.34	0.21
Patinaje	-0.34	-0.34	2.93	1.42	-0.34	0.21

	<b>c1</b>	<b>c2</b>	<b>c3</b>	<b>c4</b>	<b>c5</b>	<b>c6</b>
Guitarra	-0.32	-0.32	3.08	1.51	-0.32	-0.32
Judo	-0.32	-0.32	2.23	1.77	-0.32	-0.32
Teatro	-0.27	-0.27	2.68	0.94	-0.27	0.38
Manualidades	-0.27	-0.27	2.68	0.94	-0.27	0.38
Programacion	-0.21	-0.21	2.24	0.92	-0.21	-0.21
ComedorCocinapropia	1.23	-0.65	1.23	1.23	-0.42	1.23
ComedorCateringExterno	-1.23	0.65	-1.23	-1.23	0.42	-1.23
Becas	-0.46	0.03	1.00	0.03	-0.02	0.09
Excursiones_Convivencias	-0.43	-0.43	2.32	1.90	-0.43	0.95
Enfermeria	-0.07	-0.17	2.11	0.42	-0.18	0.21
Huerto	-0.31	-0.31	3.24	1.06	-0.31	0.29
PatioInfantil	-0.41	-0.41	2.40	1.97	-0.41	0.53
CentroTecnologico	-0.51	-0.13	0.72	1.00	-0.10	-0.10
GoogleClassroom	-0.37	-0.37	1.16	2.22	-0.37	0.14
Teams	-0.27	-0.27	2.68	1.24	-0.27	-0.27
HerramientasPropiasTecnologicas	-0.25	-0.25	2.88	1.03	-0.25	-0.25
NecesidadesEspecialesAlumnos	-0.45	-0.45	2.18	2.18	-0.45	0.86
TipoPrivado	1.51	-0.65	1.51	1.51	-0.60	1.15
index_gl_1No Vulnerable	-0.36	0.89	1.14	-0.25	-0.87	0.81
index_gl_1Pot. Vulnerable	0.49	-0.27	-0.52	0.04	0.23	-0.52
index_gl_1Vulnerable	-0.04	-0.68	-0.74	0.23	0.70	-0.39
ReligionReligioso	1.38	-0.56	1.77	1.23	-0.56	0.99
PrecioGratuito o <100€	0.09	0.09	-2.60	0.09	0.09	0.09
HorarioHorario ampliado	1.28	-0.13	-0.15	-0.35	-0.22	0.74
HorarioHorario ampliado de mañana	-0.21	-0.09	-0.21	0.92	-0.09	-0.21
HorarioHorario ampliado de mañana y tarde	-1.15	0.17	0.23	-0.05	0.26	-0.63

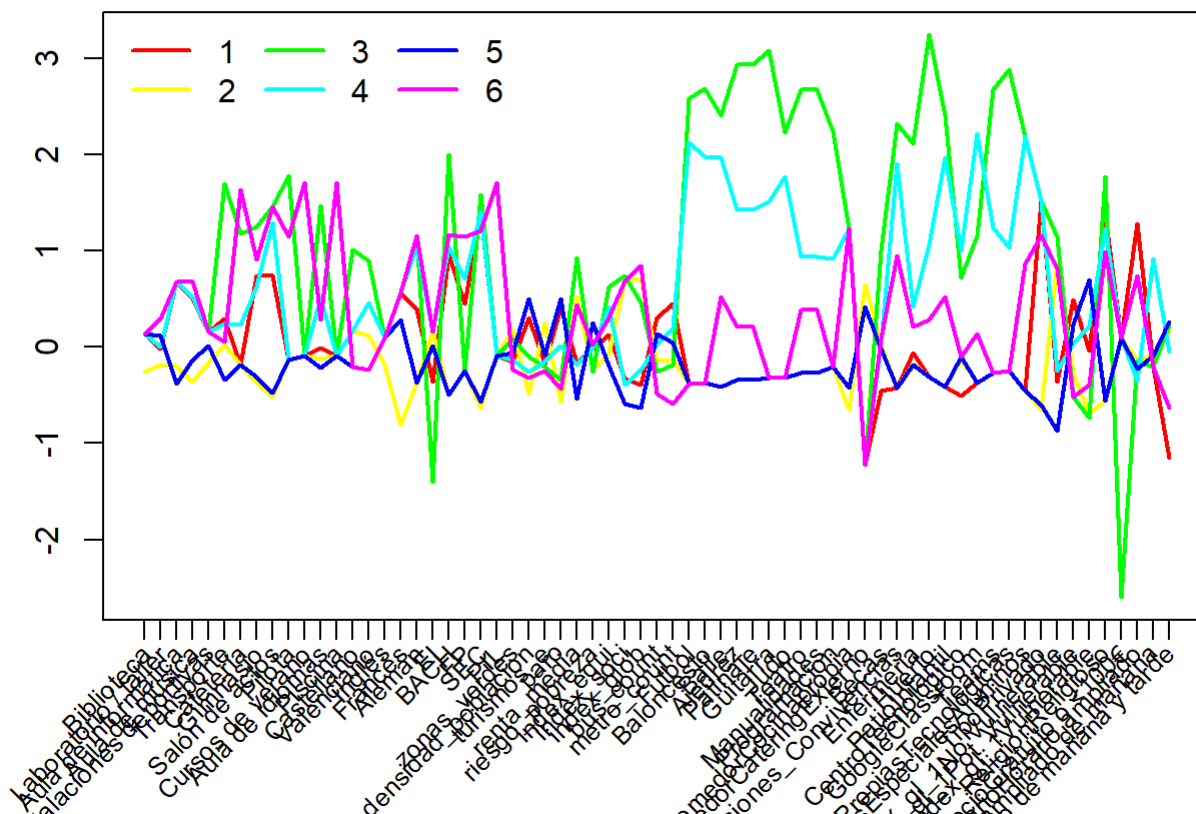
```

par(mar = c(5, 4, 4, 2) + 0.1)
matplot(t(mediasCluster), type = "l", col = rainbow(6), ylab = "", xlab = "", lwd = 2,
        lty = 1, main = "Perfil medio de los clusters", xaxt = "n")
axis(side = 1, at = 1:ncol(schools_cluster), labels = FALSE)
text(x = 1:ncol(schools_cluster), y = par("usr")[3] - 0.3,
     labels = colnames(schools_cluster), srt = 45, adj = 1, xpd = TRUE, cex = 0.8)
legend("topleft", as.character(1:6), col = rainbow(6), lwd = 2, ncol = 3, bty = "n")

```



## Perfil medio de los clusters



Explicación de los clúster: -Cluster 1 (rojo): Centros generalmente de primaria (no EI). Barrios bastante poblados y situados en zonas con gran número de paradas de bus y de metro. Comedor con cocina propia. Sin enfermería, ni centro tecnológico. No capacitados para alumnos con necesidades especiales. Generalmente centros privados. Se sitúan en barrios potencialmente vulnerables. Precios entre 100-300€. Horario ampliado.

-Cluster 2 (amarillo): No suelen tener bibliotecas, ni laboratorio/taller, ni aula de informática, ni aula de música, ni instalaciones deportivas. Tampoco presentan transporte privado. No cafetería, gimnasio, salón de actos... No idiomas (tan solo el Valenciano y poco). Barrios con zonas verdes, poca densidad\_poblacion, alto turismo\_e, bajo paro, segunda mejor renta\_media, baja pobreza. Presenta cocina de catering, son centros de tipo público. En cuanto a los índices se observa que se sitúan en barrios no vulnerables. Son laicos. Usualmente gratuitos.

-Cluster 3 (verde): Centros con transporte privado. Tienen cafetería. Pilota. Generalmente centros con aula de idiomas. Presentan generalmente la opción de tener líneas en castellano y en Valenciano. Son centros que no presentan EI por lo general, es decir, se inician a partir de primaria sus alumnos y su educación llega hasta bachiller pasando por secundaria. No presentan FP. Se localizan en barrios donde la renta media es superior al resto, así como menor pobreza. Estos barrios presentan índices de vulnerabilidad socioeconómica y vulnerabilidad con respecto a los equipamientos del barrio altos, lo que hace que sean barrios no vulnerables. Suelen presentar un gran número de extraescolares: Baloncesto, Fútbol, Patinaje, Ajedrez, Guitarra... La comida del comedor es propia. Se realizan convivencias, buenas instalaciones: Pilota, Enfermería, Huerto, Patio Infantil, Aula Tecnológica... Uso de Teams y tecnologías propias. Centros religiosos y de pago entre 100-300€. Usualmente Privados (como C4). En cuanto a los horarios hay de todo de mañana y de mañana y tarde. Capacitados para alumnos con necesidades especiales. Suelen presentar becas.

-Cluster 4 (azul-claro): presentan aula de informática, salón de actos generalmente. No presentan otras instalaciones como Pilota. No realizan cursos de verano. Algunos presentan líneas en valenciano y castellano. Idiomas de francés y alemán. Suelen ser centros desde primaria hasta bachiller, incluso presentan FP algunos. Alto index\_equi pero bajo index\_glob e index\_soci. Presentan gran número de extraescolares.

Comedor con cocina propia. Usualmente presenta gran cantidad de instalaciones. Prefieren usar Google Classroom. Capacidad de atender a niños con necesidades especiales. Religiosos y tipo privado. Horario ampliado de mañana.

-Cluster 5 (azul-oscuro): Pocas instalaciones (sin aula de informática, música, instalaciones deportivas, cafetería, aula de idiomas...). Tampoco tiene transporte privado. Son centros con EI y EP, unicamente. Se sitúan en barrios con una alta densidad de población, alto paro, baja renta\_media, alto riesgo de pobreza, indices bajos lo que hace que sean barrios vulnerables. Vemos que tienen buena conexion de buses. Bajo numero de extraescolares. Cocina con catering externo. Poco uso de tecnologías. Centros laicos y públicos. Horarios predominantemente de mañana y tarde.

-Cluster 6 (rosa): Presentan algunas instalaciones como aula de informática, cafetería, gimnasio, salon de actos y piscina (suelen hacer cursos de verano). Suelen presentar idiomas como Francés y Alemán (más que el C4). Centros desde EP hasta Bachiller. Tienen PIL. Situado en barrios con poca población, renta\_media normal-alta y poco paro. En cuanto a los indices generalmente altos lo que hace que se sitúen en barrios . Pocas conexiones de bus/metro. Pocas extraescolares fuera de los idiomas. Comedor con cocina propia. Centros privados mayoritariamente, siendo alguno religioso. Horario ampliado.

```
#library(xlsx)
#write.xlsx(prueba_bin, "DatosInfPrim.xlsx", row.names = FALSE)
```