

# Aplicación de las técnicas Multivariantes

Jeimy Merary Romero Sandoval

2024-05-02

## Table of contents

<b>Aplicacion de las Tecnicas</b>	<b>2</b>
Ejercicio 1: . . . . .	2
Ejercicio 2: . . . . .	10
Ejercicio 3: . . . . .	22
Ejercicio 4: . . . . .	32

# Aplicacion de las Tecnicas

## Ejercicio 1:

Considera que eres un consultor de negocios especializado en el análisis de

pequeñas empresas. Estás investigando diversas características operativas y de gestión en 20 pequeñas empresas para ayudarles a optimizar su rendimiento. Las empresas han sido evaluadas según

las siguientes características: Innovación: Capacidad para introducir nuevos productos o servicios. Comunicación: Efectividad en la comunicación interna y externa. Eficiencia: Uso efectivo de los recursos. Responsabilidad: Adherencia a las regulaciones y compromisos sociales. Atención al Cliente: Calidad en el trato y servicio a clientes. Comunidad: Participación y contribución a la comunidad local. En tu esfuerzo por entender cómo varias características de las empresas pueden influir en su éxito general, te interesa descubrir si algunas de estas características tienden a presentarse juntas o si se influyen mutuamente en el desempeño empresarial.

**¿Qué enfoque utilizarías para evaluar si ciertas características tienden a agruparse en ciertos perfiles de empresas?**

haciendo los siguientes pasos:

```
library(readxl)
dato <- read_excel("C:/Users/DELL/Documents/seminario/ejercicio 1.xlsx")
dato
```

# A tibble: 20 x 6

	Innovación	Comunicación	Eficiencia	Responsabilidad	Atención al Cliente
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	4	5	4	5	4
2	1	2	5	4	1
3	1	2	2	3	4
4	4	5	1	2	2
5	5	5	5	3	4
6	1	2	2	3	3
7	3	3	4	4	5
8	2	2	5	4	1
9	2	1	1	1	1
10	1	4	4	4	4
11	2	3	5	4	2
12	4	5	1	2	3
13	2	3	2	3	4

```

14          3          2          3          2          4
15          2          2          3          5          2
16          2          4          5          3          5
17          4          2          3          2          1
18          3          2          5          1          2
19          5          4          1          5          2
20          2          1          1          1          5
# i 1 more variable: Comunidad <dbl>

```

```

datas <- scale(dato,center = TRUE,scale = TRUE)
head(datas)

```

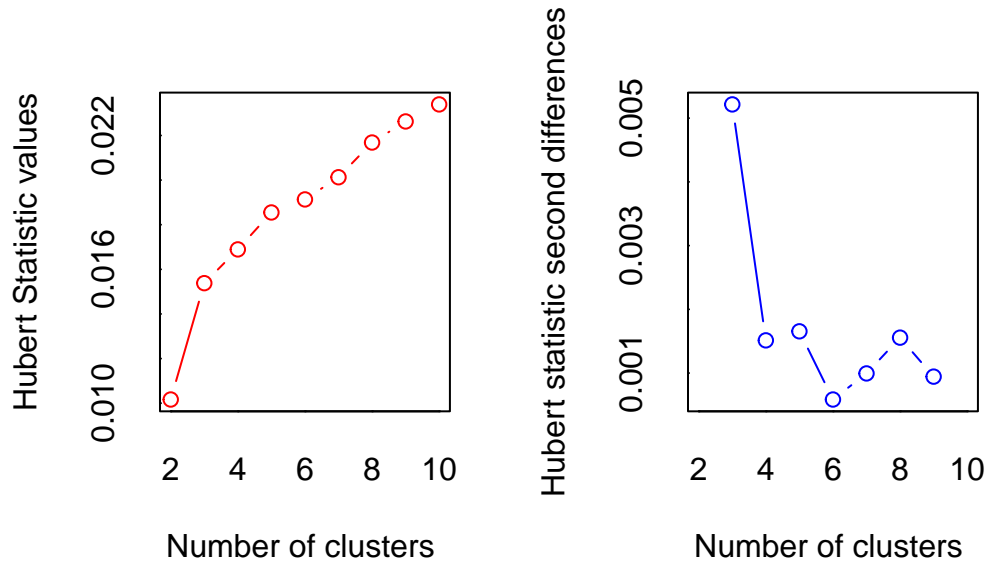
```

      Innovación Comunicación Eficiencia Responsabilidad Atención\r\nal Cliente
[1,]  1.031419    1.5114962  0.5559102    1.48075669    0.73335181
[2,] -1.260623   -0.7004495  1.1735882    0.72139428   -1.36193908
[3,] -1.260623   -0.7004495 -0.6794458   -0.03796812    0.73335181
[4,]  1.031419    1.5114962 -1.2971238   -0.79733052   -0.66350878
[5,]  1.795433    1.5114962  1.1735882   -0.03796812    0.73335181
[6,] -1.260623   -0.7004495 -0.6794458   -0.03796812    0.03492151
      Comunidad
[1,]  0.1296694
[2,]  0.7780165
[3,] -0.5186777
[4,] -1.1670248
[5,]  0.1296694
[6,]  0.1296694

```

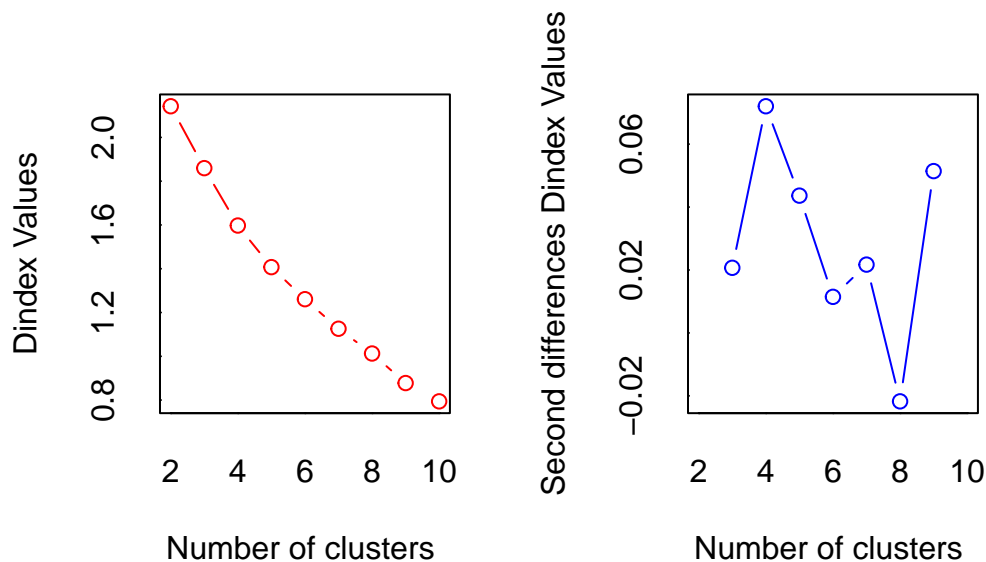
- **Innovación:** Los valores en esta columna representan la capacidad de las empresas para introducir nuevos productos o servicios. Las puntuaciones más altas indican una mayor capacidad de innovación en comparación con el promedio de las empresas.
- **Comunicación:** Aquí se evalúa la efectividad en la comunicación interna y externa. Puntuaciones más altas indican una mejor comunicación en general.
- **Eficiencia:** Esta variable refleja el uso efectivo de los recursos por parte de las empresas. Puntuaciones más altas sugieren una mayor eficiencia en la gestión de recursos.
- **Responsabilidad:** Evalúa la adherencia a las regulaciones y compromisos sociales. Puntuaciones más altas indican una mayor responsabilidad empresarial.
- **Atención al Cliente:** Refleja la calidad en el trato y servicio a los clientes. Valores más altos representan una mejor atención al cliente.
- **Comunidad:** Evalúa la participación y contribución de las empresas a la comunidad local. Puntuaciones más altas indican una mayor implicación comunitaria.

```
#install.packages("NbClust")
library(NbClust)
set.seed(1234)
res.nbclust <- NbClust(datas, distance = "euclidean",
  min.nc = 2, max.nc = 10,
  method = "complete", index = "all")
```



\*\*\* : The Hubert index is a graphical method of determining the number of clusters.

In the plot of Hubert index, we seek a significant knee that corresponds to a significant increase of the value of the measure i.e the significant peak in index second differences plot.



\*\*\* : The D index is a graphical method of determining the number of clusters.

In the plot of D index, we seek a significant knee (the significant peak in the second differences plot) that corresponds to a significant increase of the value of the measure.

\*\*\*\*\*

\* Among all indices:

- \* 4 proposed 2 as the best number of clusters
- \* 3 proposed 3 as the best number of clusters
- \* 4 proposed 4 as the best number of clusters
- \* 4 proposed 5 as the best number of clusters
- \* 1 proposed 6 as the best number of clusters
- \* 3 proposed 9 as the best number of clusters
- \* 4 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 2

\*\*\*\*\*

se observa en la siguiente grafia que se puede dividir los resultados en 2 clusters , esto significa

que los datos pueden agruparse de manera más efectiva en dos grupos distintos en lugar de más o menos clusters.

```
#install.packages("clValid")  
library(clValid)
```

Cargando paquete requerido: cluster

```
validclus <- clValid(datas, nClust = 2:5,  
                    clMethods = c("hierarchical", "kmeans", "diana", "fanny", "pam", "clara", "agnes"),  
                    validation = "internal")
```

```
Warning in fanny(Dist, nc, ...): the memberships are all very close to 1/k.  
Maybe decrease 'memb.exp' ?  
Warning in fanny(Dist, nc, ...): the memberships are all very close to 1/k.  
Maybe decrease 'memb.exp' ?
```

```
Warning in vClusters(mat, clMethods[i], nClust, validation = validation, :  
fanny unable to find 3 clusters, returning NA for these validation measures
```

```
Warning in fanny(Dist, nc, ...): the memberships are all very close to 1/k.  
Maybe decrease 'memb.exp' ?
```

```
Warning in vClusters(mat, clMethods[i], nClust, validation = validation, :  
fanny unable to find 4 clusters, returning NA for these validation measures
```

```
Warning in fanny(Dist, nc, ...): the memberships are all very close to 1/k.  
Maybe decrease 'memb.exp' ?
```

```
Warning in vClusters(mat, clMethods[i], nClust, validation = validation, :  
fanny unable to find 5 clusters, returning NA for these validation measures
```

```
Warning in clValid(datas, nClust = 2:5, clMethods = c("hierarchical", "kmeans",  
: rownames for data not specified, using 1:nrow(data)
```

```
summary(validclus)
```

Clustering Methods:

hierarchical kmeans diana fanny pam clara agnes

Cluster sizes:

2 3 4 5

Validation Measures:

		2	3	4	5
hierarchical	Connectivity	8.0960	16.6563	21.7567	24.0663
	Dunn	0.4569	0.3847	0.4901	0.4901
	Silhouette	0.2235	0.2211	0.2399	0.2557
kmeans	Connectivity	10.7623	16.6563	21.7567	25.0329
	Dunn	0.3692	0.3847	0.4901	0.5439
	Silhouette	0.2130	0.2211	0.2399	0.2403
diana	Connectivity	10.7623	19.6849	21.1349	22.5294
	Dunn	0.3692	0.3952	0.4465	0.4901
	Silhouette	0.2130	0.2058	0.1897	0.1271
fanny	Connectivity	14.6663	NA	NA	NA
	Dunn	0.4060	NA	NA	NA
	Silhouette	0.1425	NA	NA	NA
pam	Connectivity	10.7623	20.4710	25.6500	26.7635
	Dunn	0.3692	0.4127	0.5089	0.5089
	Silhouette	0.2130	0.1980	0.2176	0.2348
clara	Connectivity	10.7623	20.4710	25.6500	26.7635
	Dunn	0.3692	0.4127	0.5089	0.5089
	Silhouette	0.2130	0.1980	0.2176	0.2348
agnes	Connectivity	8.0960	16.6563	21.7567	24.0663
	Dunn	0.4569	0.3847	0.4901	0.4901
	Silhouette	0.2235	0.2211	0.2399	0.2557

Optimal Scores:

	Score	Method	Clusters
Connectivity	8.0960	hierarchical	2
Dunn	0.5439	kmeans	5
Silhouette	0.2557	hierarchical	5

Observando esos valores dice que hay que tomar 2 clusters, lo cual se puede representar en la siguiente grafica.

```
library(factoextra)
```

Cargando paquete requerido: ggplot2

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

```
dist <- dist(datas, method = "euclidean")

modelo <- hclust(dist, method = "complete")

modelo <- hclust(dist(dist))
fviz_dend(modelo, cex = 0.5, k=2,
           rect = TRUE,
           k_colors = "jco",
           rect_border = "jco",
           rect_fill = TRUE,
           horiz = TRUE,
           ggtheme = theme_bw())
```

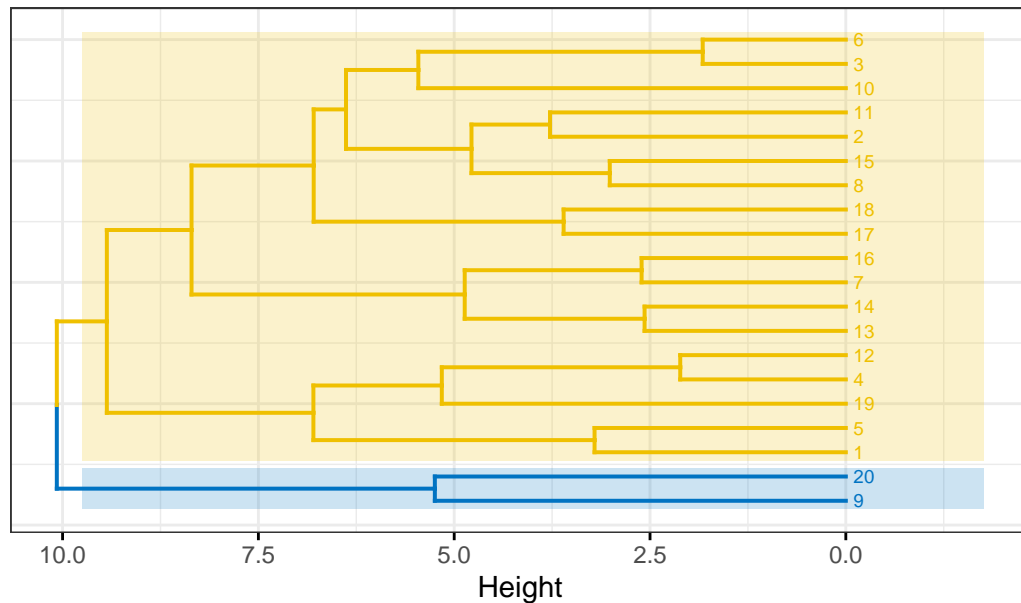
Warning: The ``<scale>`` argument of ``guides()`` cannot be ``FALSE``. Use "none" instead as of ggplot2 3.3.4.

i The deprecated feature was likely used in the factoextra package.

Please report the issue at <https://github.com/kassambara/factoextra/issues>.



## Cluster Dendrogram



observando la grafica anterior se puede concluir que si hay que dividir en 2 cluster, tambien se puede decir que las empresas que estan dentro de esos cluster tienen mucha relacion hablando sobre las calificaciones.

### **¿Cómo podrías organizar estos datos para identificar patrones de rendimiento entre estas empresas?**

-Es posible que las empresas en el primer cluster compartan características similares entre sí que las distinguan claramente del segundo cluster. Podrían tener estrategias de marketing similares, enfoques de gestión comunes o perfiles de clientes específicos.

-Cuando se agrupan las empresas en dos grupos basándose en ciertas características, como por ejemplo cómo se manejan, cómo se promocionan o qué tipo de clientes tienen, se encuentra que algunas de estas características tienden a ir de la mano en cada grupo. Esto significa que ciertas cosas que hacen estas empresas, como su forma de trabajar o cómo se presentan al público, parecen tener un impacto conjunto en su éxito. Es como si estas características se ayudaran entre sí para que las empresas funcionen bien en su conjunto.

- Al agrupar las empresas en función de ciertas características que están relacionadas con el éxito empresarial, como la eficiencia operativa, la satisfacción del cliente o la innovación, el hecho de que se formen dos grupos distintos indica que hay combinaciones de estas características que tienden a estar juntas y que pueden influenciarse de manera positiva entre sí para mejorar el desempeño general de las empresas en cada grupo.

-Si las características están relacionadas con el mercado al que se dirige la empresa o su posición competitiva, el cluster en 2 grupos podría indicar diferentes segmentos de mercado o niveles de competencia.

-Al identificar dos grupos distintos, se puede evaluar qué características están presentes en cada cluster y cómo estas podrían estar contribuyendo al éxito o al desempeño deficiente. Esto podría proporcionar sobre áreas de mejora o estrategias exitosas que podrían aplicarse en otros contextos.

## Ejercicio 2:

En el estudio de la enología, el análisis de las propiedades químicas de los vinos es fundamental para entender sus características y diferencias. Se dispone de un conjunto de datos que incluye 13 concentraciones químicas diferentes de muestras de vino obtenidas de tres tipos de cultivares distintos en el dataset "wine" de la librería "rattle". El conjunto de datos contiene mediciones para 13 variables químicas diferentes (como el alcohol, el magnesio, los taninos, entre otros) para cada muestra de vino. Dado que se conocen los cultivares de origen para cada muestra, ¿cómo podrías utilizar las mediciones químicas disponibles para identificar y describir las diferencias claras entre estos grupos?

```
library(rattle)
```

Cargando paquete requerido: tibble

Cargando paquete requerido: bitops

Rattle: A free graphical interface for data science with R.  
Versión 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.  
Escriba 'rattle()' para agitar, sacudir y rotar sus datos.

```
library(psych)
```

Adjuntando el paquete: 'psych'

The following objects are masked from 'package:ggplot2':

%+%, alpha

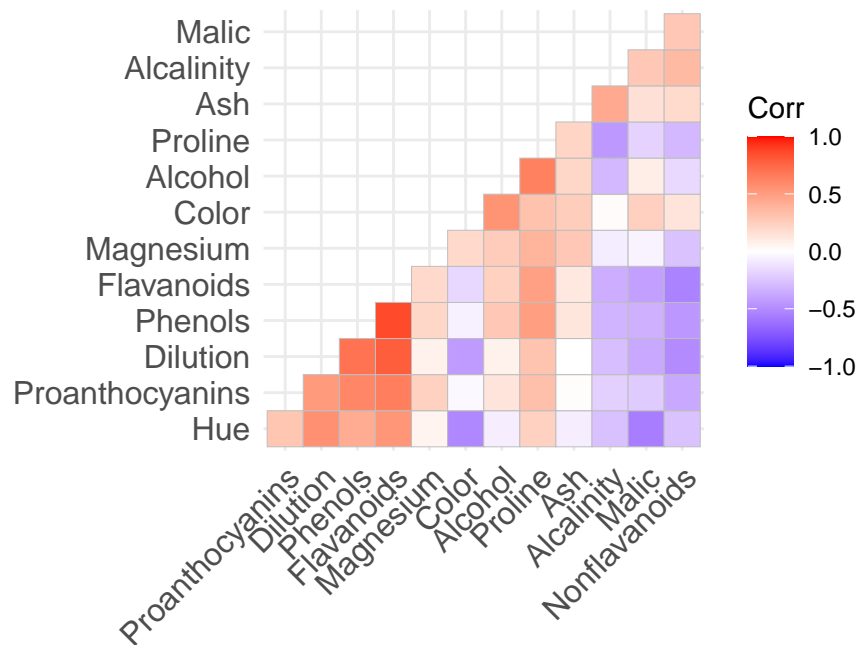
```
library(polycor)
```

Adjuntando el paquete: 'polycor'

The following object is masked from 'package:psych':

polyserial

```
library(ggcorrplot)
wine_s <- wine[1:178,2:14] # subconjunto de datos
mat_cor <- hetcor(wine_s)$correlations #matriz de correlación policorica
ggcorrplot(mat_cor,type="lower",hc.order = T)
```



```
library(psych)
cortest.bartlett(mat_cor)->p_esf
```

Warning in cortest.bartlett(mat\_cor): n not specified, 100 used

```
p_esf$p
```

```
[1] 1.836773e-104
```

Dado que el p-valor es mucho menor que cualquier nivel de significancia, se rechaza la hipótesis nula. Esto significa que la matriz de correlación no es una matriz identidad, sugiriendo fuertemente que las variables de estudio están correlacionadas entre sí. lo que significa que se puede aplicar métodos de análisis como el análisis factorial o análisis de componentes principales. Esto implica que se podrá aplicar estos métodos para reducir la dimensiones en los datos.

```
KMO(mat_cor)
```

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = mat\_cor)

Overall MSA = 0.78

MSA for each item =

Alcohol	Malic	Ash	Alcalinity	Magnesium
0.73	0.80	0.44	0.68	0.68
Phenols	Flavanoids	Nonflavanoids	Proanthocyanins	Color
0.87	0.81	0.83	0.85	0.62
Hue	Dilution	Proline		
0.79	0.87	0.82		

El resultado es 0.78 lo que dice que se puede continuar con el análisis Factorial.

```
### prueba de dos modelos con tres factores
modelo1<-fa(mat_cor,
  nfactors = 3,
  rotate = "none",
  fm="mle") # modelo máxima verosimilitud

modelo2<-fa(mat_cor,
  nfactors = 3,
  rotate = "none",
  fm="minres") # modelo minimo residuo
#####comparando las comunialidades
sort(modelo1$communality,decreasing = T)->c1
sort(modelo2$communality,decreasing = T)->c2
head(cbind(c1,c2))
```

	c1	c2
Flavanoids	0.9310641	0.9172700
Alcalinity	0.9271472	0.7510924
Phenols	0.8013574	0.7497843
Color	0.7538577	0.7446974
Dilution	0.7481250	0.7017408
Proline	0.6159081	0.6797750

ambos métodos son capaces de extraer estructuras de factor significativas de los datos,

```
####comparacion de las unicidades
sort(modelo1$uniquenesses,decreasing = T)->u1
sort(modelo2$uniquenesses,decreasing = T)->u2
head(cbind(u1,u2))
```

	u1	u2
Magnesium	0.8372182	0.8003787
Malic	0.7265320	0.6945271
Nonflavanoids	0.6577298	0.6413651
Proanthocyanins	0.5551403	0.6008837
Ash	0.5216287	0.4885421
Hue	0.5025422	0.3734974

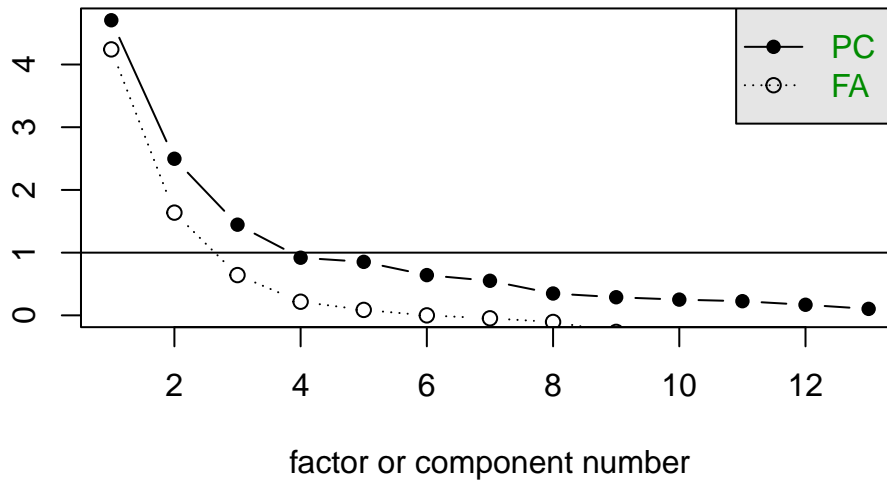
La comparación muestra que la unicidad tiende a ser generalmente más alta en el modelo 1 que en el modelo 2 para la mayoría de las variables, aunque hay excepciones como en Proanthocyanins y Ash, donde el modelo 2 muestra una unicidad más alta o más baja, respectivamente, esto puede reflejar la sensibilidad de cada método a la estructura específica de los datos y a las relaciones entre variables., las unicidades altas en ciertas variables como Magnesium y Malic sugieren que estos elementos pueden tener efectos o características que son distintos y no se explican bien a través de los factores comunes utilizados en estos modelos.

Esto puede ser crucial para la interpretación en contextos donde estas variables son críticas. En resumen, mientras que las comunalidades reflejan cuánto contribuyen los factores comunes a explicar cada variable, las unicidades destacan la varianza que estos factores no logran captar, resaltando la singularidad de cada variable.

```
scree(mat_cor)
```

Eigen values of factors and components

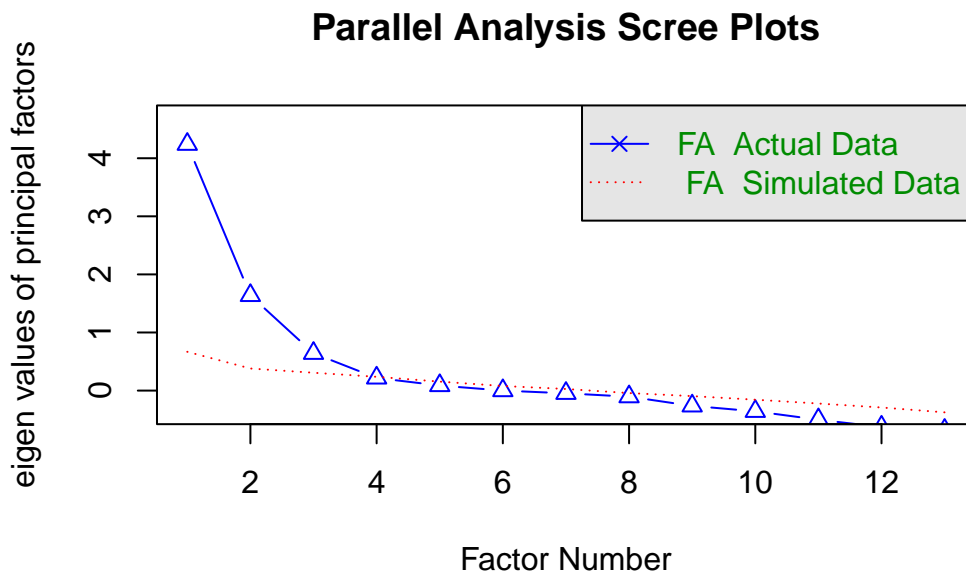
## Scree plot



Al comparar las líneas de PC y FA, es común observar que los autovalores de PCA son generalmente más altos que los de FA para los primeros componentes/factores. Esto se debe a que PCA captura toda la varianza, mientras que el análisis factorial intenta identificar factores latentes que expliquen las correlaciones entre las variables observadas, los factores iniciales suelen ser más influyentes, es decir, explican una mayor proporción de la varianza total compartida entre las variables

```
fa.parallel(mat_cor,n.obs=178,fa="fa",fm="minres")
```

Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :  
The estimated weights for the factor scores are probably incorrect. Try a  
different factor score estimation method.



Parallel analysis suggests that the number of factors = 3 and the number of components = 1

Al observar la grafica anterior se puede dar la interpretación que ambas partes de la gráfica sugiere que el número óptimo de factores a retener podría ser alrededor de cuatro, esto se basa en que la varianza explicada por los factores se estabiliza o disminuye gradualmente después de cuatro factores, lo que indica que agregar más factores no mejora sustancialmente la explicación de la varianza en los datos.

```
#Rotaciones
library(GPArotation)
```

Adjuntando el paquete: 'GPArotation'

The following objects are masked from 'package:psych':

```
equamax, varimin
```

```
names(wine)
```

```

[1] "Type"           "Alcohol"         "Malic"           "Ash"
[5] "Alcalinity"     "Magnesium"       "Phenols"         "Flavanoids"
[9] "Nonflavanoids"  "Proanthocyanins" "Color"           "Hue"
[13] "Dilution"      "Proline"

```

```
str(wine_s)
```

```

'data.frame':  178 obs. of  13 variables:
 $ Alcohol      : num  14.2 13.2 13.2 14.4 13.2 ...
 $ Malic        : num  1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ...
 $ Ash          : num  2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 2.17 2.27 ...
 $ Alcalinity   : num  15.6 11.2 18.6 16.8 21 15.2 14.6 17.6 14 16 ...
 $ Magnesium    : int   127 100 101 113 118 112 96 121 97 98 ...
 $ Phenols      : num  2.8 2.65 2.8 3.85 2.8 3.27 2.5 2.6 2.8 2.98 ...
 $ Flavanoids   : num  3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...
 $ Nonflavanoids : num  0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...
 $ Proanthocyanins: num  2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...
 $ Color        : num  5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...
 $ Hue          : num  1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...
 $ Dilution    : num  3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...
 $ Proline      : int   1065 1050 1185 1480 735 1450 1290 1295 1045 1045 ...

```

```
str(wine)
```

```

'data.frame':  178 obs. of  14 variables:
 $ Type        : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
 $ Alcohol     : num  14.2 13.2 13.2 14.4 13.2 ...
 $ Malic       : num  1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ...
 $ Ash         : num  2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 2.17 2.27 ...
 $ Alcalinity  : num  15.6 11.2 18.6 16.8 21 15.2 14.6 17.6 14 16 ...
 $ Magnesium   : int   127 100 101 113 118 112 96 121 97 98 ...
 $ Phenols     : num  2.8 2.65 2.8 3.85 2.8 3.27 2.5 2.6 2.8 2.98 ...
 $ Flavanoids  : num  3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...
 $ Nonflavanoids : num  0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...
 $ Proanthocyanins: num  2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...
 $ Color       : num  5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...
 $ Hue         : num  1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...
 $ Dilution   : num  3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...
 $ Proline     : int   1065 1050 1185 1480 735 1450 1290 1295 1045 1045 ...

```



```
# Convertir a numérico si es necesario
wine_s <- as.data.frame(lapply(wine_s, as.numeric))
wine <- as.data.frame(lapply(wine, as.numeric))

# Verificar nuevamente tipos de datos y valores faltantes
str(wine_s)
```

```
'data.frame': 178 obs. of 13 variables:
 $ Alcohol      : num 14.2 13.2 13.2 14.4 13.2 ...
 $ Malic        : num 1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ...
 $ Ash          : num 2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 2.17 2.27 ...
 $ Alcalinity   : num 15.6 11.2 18.6 16.8 21 15.2 14.6 17.6 14 16 ...
 $ Magnesium    : num 127 100 101 113 118 112 96 121 97 98 ...
 $ Phenols      : num 2.8 2.65 2.8 3.85 2.8 3.27 2.5 2.6 2.8 2.98 ...
 $ Flavanoids   : num 3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...
 $ Nonflavanoids : num 0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...
 $ Proanthocyanins: num 2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...
 $ Color        : num 5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...
 $ Hue          : num 1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...
 $ Dilution     : num 3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...
 $ Proline      : num 1065 1050 1185 1480 735 ...
```

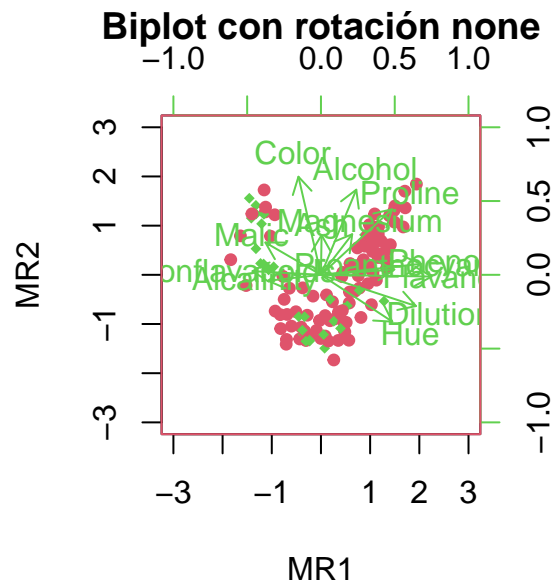
```
str(wine)
```

```
'data.frame': 178 obs. of 14 variables:
 $ Type         : num 1 1 1 1 1 1 1 1 1 1 ...
 $ Alcohol      : num 14.2 13.2 13.2 14.4 13.2 ...
 $ Malic        : num 1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ...
 $ Ash          : num 2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 2.17 2.27 ...
 $ Alcalinity   : num 15.6 11.2 18.6 16.8 21 15.2 14.6 17.6 14 16 ...
 $ Magnesium    : num 127 100 101 113 118 112 96 121 97 98 ...
 $ Phenols      : num 2.8 2.65 2.8 3.85 2.8 3.27 2.5 2.6 2.8 2.98 ...
 $ Flavanoids   : num 3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...
 $ Nonflavanoids : num 0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...
 $ Proanthocyanins: num 2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...
 $ Color        : num 5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...
 $ Hue          : num 1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...
 $ Dilution     : num 3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...
 $ Proline      : num 1065 1050 1185 1480 735 ...
```

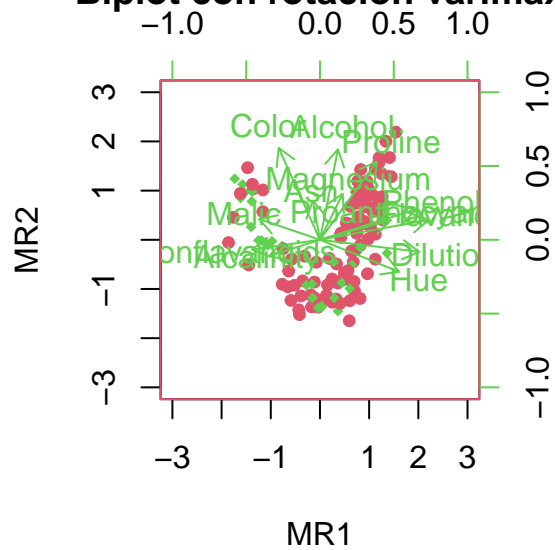
```
# Define los métodos de rotación
rot <- c("none", "varimax", "quartimax", "Promax")

# Define la función bi_mod
bi_mod <- function(tipo, wine_s, wine) {
  biplot.psych(fa(wine_s, nfactors = 2, fm = "minres", rotate = tipo),
    main = paste("Biplot con rotación", tipo),
    col = c(2, 3, 4), pch = c(21, 18), group = wine[, "Malic"])
}

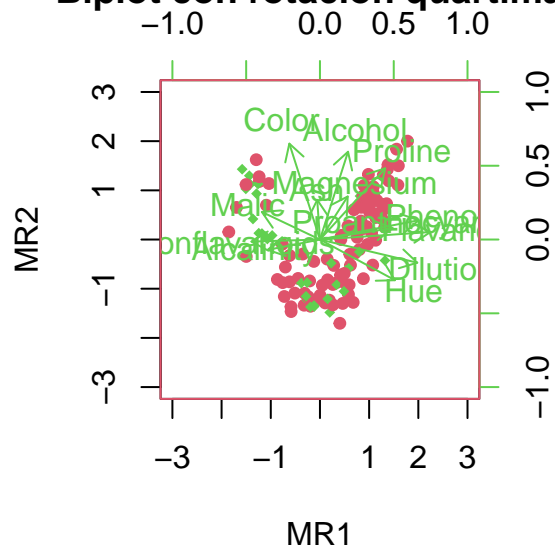
# Aplicar bi_mod corregido
results <- sapply(rot, function(tipo) bi_mod(tipo, wine_s, wine))
```



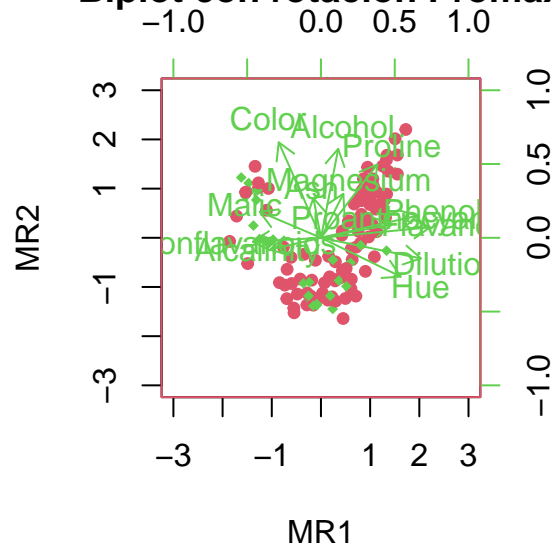
**Biplot con rotación varimax**



**Biplot con rotación quartimax**

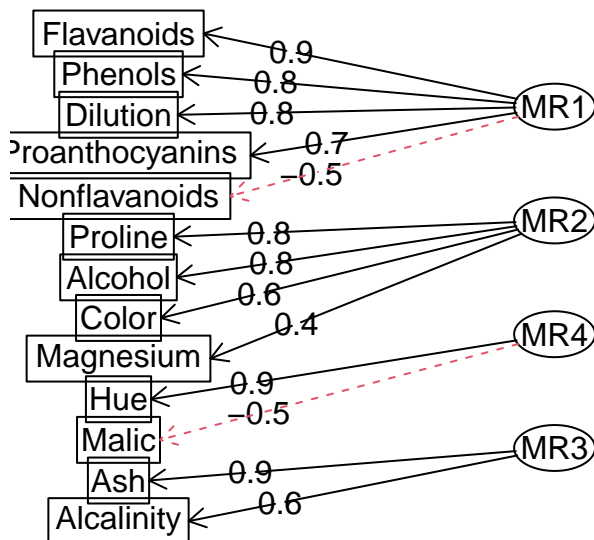


## Biplot con rotación Promax



```
modelo_varimax<-fa(mat_cor,nfactors = 4,rotate = "varimax",
                   fa="minres")
fa.diagram(modelo_varimax)
```

## Factor Analysis



MR1: Este factor está asociado con altas cargas para las variables flavanoids , phenols , proanthocyanins y Nonflavanoids . Estas variables podrían estar relacionadas en términos de características químicas o propiedades de los vinos que afectan a estas medidas. La línea punteada indica la contribución de este factor a estas variables.

MR2: Este factor está asociado principalmente con Proline y Alcohol , y en menor medida con Color y Magnesium . Esto sugiere que estas variables podrían compartir una influencia común o estar relacionadas de alguna manera en el análisis de las características del vino.

MR3: Este factor está asociado principalmente con Ash y Alkalinity . Estas variables podrían tener una relación en términos de las propiedades físicas o químicas del vino relacionadas con la ceniza y la alcalinidad.

MR4: Este factor está asociado principalmente con Hue y Malic . Estas variables podrían tener una relación en términos del color y la acidez del vino.

```
print(modelo_varimax$loadings,cut=0)
```

Loadings:

	MR1	MR2	MR4	MR3
Alcohol	0.124	0.774	-0.140	-0.055
Malic	-0.265	-0.007	-0.509	0.166
Ash	0.040	0.329	-0.031	0.943
Alkalinity	-0.217	-0.349	-0.251	0.598
Magnesium	0.176	0.388	0.018	0.130
Phenols	0.815	0.271	0.207	0.015
Flavanoids	0.892	0.211	0.312	0.007
Nonflavanoids	-0.517	-0.145	-0.167	0.254
Proanthocyanins	0.665	0.139	0.074	-0.024
Color	-0.173	0.614	-0.522	0.065
Hue	0.311	-0.028	0.856	-0.035
Dilution	0.764	-0.027	0.399	-0.028
Proline	0.302	0.811	0.205	-0.052

	MR1	MR2	MR4	MR3
SS loadings	3.137	2.174	1.724	1.368
Proportion Var	0.241	0.167	0.133	0.105
Cumulative Var	0.241	0.409	0.541	0.646

MR1: Este factor estan asociados con las características químicas que contribuyen a la calidad y sabor del vino. Las variables con altos loadings en este factor, como Flavanoids, Phenols y Proanthocyanins, son conocidas por sus efectos en el aroma, sabor y propiedades antioxidantes

del vino. Un vino con altos loadings en este factor podría ser percibido como más complejo en sabor y con un perfil antioxidante más marcado.

MR2: Aquí se encuentran variables como Alcohol, Proline y Color. Estas características pueden estar relacionadas con la fortaleza y el cuerpo del vino, así como con su color y tonalidad. Un vino con altos loadings en este factor podría ser percibido como más robusto en sabor y con una mayor concentración de compuestos relacionados con el color.

MR4: Este factor está asociado con variables como Hue y Ash. La hue se refiere al tono del color del vino, mientras que Ash puede estar relacionado con las propiedades de la ceniza en los vinos. Un vino con altos loadings en este factor podría tener un tono de color particular y propiedades asociadas con la ceniza, que pueden influir en su percepción visual y en boca.

MR3: Este factor involucra variables como Alkalinity y Nonflavanoids. La alcalinidad puede influir en la sensación en boca y el equilibrio del vino, mientras que los compuestos no flavonoides pueden tener efectos específicos en el perfil sensorial y la estructura del vino. Un vino con altos loadings en este factor podría tener una sensación en boca particular y ciertas características sensoriales distintivas

### **Ejercicio 3:**

Un consorcio de investigación ambiental está estudiando el impacto y las prácticas de sostenibilidad en 13 ciudades de tamaño mediano. Cada ciudad ha sido evaluada con base en 8 indicadores clave que reflejan su desempeño y esfuerzos hacia la sostenibilidad:

X1: Tasa de reciclaje (porcentaje del total de desechos reciclados).

X2: Producción per cápita de energía renovable (en kWh).

X3: Consumo per cápita de agua (en metros cúbicos).

X4: Índice de calidad del aire (medido como el promedio anual de partículas PM2.5).

X5: Extensión de áreas verdes por persona (en metros cuadrados).

X6: Inversión en infraestructura sostenible (en millones de dólares).

X7: Participación ciudadana en programas ambientales (en porcentaje de la población).

X8: Eficiencia en la gestión de residuos (toneladas de residuos procesados por día).

Dado el complejo panorama de indicadores para evaluar la sostenibilidad en estas ciudades, es crucial simplificar la información para entender mejor qué ciudades están liderando en prácticas sostenibles y cuáles necesitan mejorar. Considerando este objetivo, ¿cómo podrías consolidar estos múltiples indicadores para facilitar una comparación efectiva entre las ciudades?

##Se utilizara el metodo de PCA

```
library(readxl)
ejercicio_3 <- read_excel("C:/Users/DELL/Documents/seminario/ejercicio 3.xlsx")
ejercicio_3
```

```
# A tibble: 13 x 8
      X1     X2     X3     X4     X5     X6     X7     X8
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  0.13  0.94  2.17  5.79  5.48 11.1   3.98  5.8
2  1.55  4.3   4.33  5.4   5.34 11.0   5.58  8.1
3  1.21  6.9   6.5   4.07  6.89 24.8   7.28 14.4
4  2.48  8.69  8.67  7.41  4.14 31.6  10.3 17.9
5  6.01 11.2  10.8   0.03 10.0  10.2   1.97  7.55
6  6.74 13.1  13     1.55  5.87  6.13   1.06  3.95
7  7.58 14.8  15.2  11.4   2.71  6.96   1.67  4.75
8  8.01 18.2  17.3   6.33  2.56 41.5  12.4 26.2
9  8.12 19.6  19.5   6.51  3.46 31.0  10.2 20.2
10 11.5  21.7  21.7   5.06  5.92 43.3  12.2 26.6
11 10.7  24.1  23.8   3.97  5.6  11.4   5.24  7.33
12 12.0  24.8  26     4.89  4.87  9.71   4.09  6.83
13 14.4  25.4  28.2   7.76  3.52 34.4   9.8  20.4
```

```
data("ejercicio_3")
```

Warning in data("ejercicio\_3"): data set 'ejercicio\_3' not found

```
head(ejercicio_3)
```

```
# A tibble: 6 x 8
      X1     X2     X3     X4     X5     X6     X7     X8
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  0.13  0.94  2.17  5.79  5.48 11.1   3.98  5.8
2  1.55  4.3   4.33  5.4   5.34 11.0   5.58  8.1
3  1.21  6.9   6.5   4.07  6.89 24.8   7.28 14.4
4  2.48  8.69  8.67  7.41  4.14 31.6  10.3 17.9
5  6.01 11.2  10.8   0.03 10.0  10.2   1.97  7.55
6  6.74 13.1  13     1.55  5.87  6.13   1.06  3.95
```

```
str(ejercicio_3)
```

```
tibble [13 x 8] (S3: tbl_df/tbl/data.frame)
 $ X1: num [1:13] 0.13 1.55 1.21 2.48 6.01 ...
 $ X2: num [1:13] 0.94 4.3 6.9 8.69 11.24 ...
 $ X3: num [1:13] 2.17 4.33 6.5 8.67 10.83 ...
 $ X4: num [1:13] 5.79 5.4 4.07 7.41 0.03 1.55 11.4 6.33 6.51 5.06 ...
 $ X5: num [1:13] 5.48 5.34 6.89 4.14 10.04 ...
 $ X6: num [1:13] 11.1 11 24.8 31.6 10.2 ...
 $ X7: num [1:13] 3.98 5.58 7.28 10.34 1.97 ...
 $ X8: num [1:13] 5.8 8.1 14.4 17.87 7.55 ...
```

```
apply(X = ejercicio_3, MARGIN = 2, FUN = mean)
```

	X1	X2	X3	X4	X5	X6	X7	X8
	6.956154	14.914615	15.166923	5.397692	5.107692	21.006154	6.598462	13.072308

El PCA ha reducido la dimensionalidad del conjunto de datos de 8 variables a un conjunto de componentes principales que capturan la mayor parte de la variabilidad en los datos.

```
apply(X = ejercicio_3, MARGIN = 2, FUN = var)
```

	X1	X2	X3	X4	X5	X6	X7
	20.658376	66.657160	71.201490	7.975636	3.970703	189.893976	16.246164
X8							
	68.046053						

Una alta varianza en estas variables indica que los valores de esa variable están más dispersos, lo que podría significar una mayor variabilidad en la situación de cada ciudad en ese aspecto particular de la sostenibilidad y en este caso las variables mas altas son las siguientes:

- X3: Consumo per cápita de agua (en metros cúbicos).
- X2: Producción per cápita de energía renovable (en kWh).
- X6: Inversión en infraestructura sostenible (en millones de dólares).

Por otro lado, una baja varianza indica que los valores tienden a estar más cercanos a la media, lo que podría indicar una mayor uniformidad en ese aspecto entre las ciudades

```
pca <- prcomp(ejercicio_3, scale = TRUE)
names(pca)
```

```
[1] "sdev"      "rotation" "center"    "scale"     "x"
```



```
pca$center
```

	X1	X2	X3	X4	X5	X6	X7	X8
	6.956154	14.914615	15.166923	5.397692	5.107692	21.006154	6.598462	13.072308

- Los números asociados a cada variable representan el promedio de esa variable antes de que se estandarice es decir, se reste la media y se divida por la desviación estándar durante el análisis de PCA.
- El valor 6.956154 para X1 indica que, en promedio, la Tasa de reciclaje en las ciudades evaluadas es de aproximadamente el 6.95% antes de aplicar el escalado durante PCA.
- El valor 14.914615 para X2 indica que, en promedio, la Producción per cápita de energía renovable es de aproximadamente 14.91 kWh antes del escalado.

Y así sucesivamente para las demás variables.

```
pca$scale
```

	X1	X2	X3	X4	X5	X6	X7	X8
	4.545149	8.164384	8.438098	2.824117	1.992662	13.780202	4.030653	8.249003

El valor 4.545149 para X1 indica que la Tasa de reciclaje tiene una desviación estándar de aproximadamente 4.55% antes del escalado durante PCA.

El valor 8.164384 para X2 indica que la Producción per cápita de energía renovable tiene una desviación estándar de aproximadamente 8.16 kWh antes del escalado. Y

así sucesivamente para las demás variables.

```
pca$rotation
```

	PC1	PC2	PC3	PC4	PC5	PC6
X1	0.3551761	-0.4594807	0.03418166	-0.175477344	0.4262555	0.60639707
X2	0.3769254	-0.4286792	0.03801483	0.070933220	-0.2656430	-0.57860281
X3	0.3763377	-0.4334478	0.00620033	-0.005478072	-0.2364483	-0.01686902
X4	0.2140581	0.2014459	-0.67199135	-0.664424455	-0.1253127	-0.05404693
X5	-0.2942276	-0.1413771	0.59850961	-0.699072953	-0.2124654	-0.02748401
X6	0.3894736	0.3446872	0.26238969	-0.118028764	0.2985259	-0.16069794
X7	0.3858970	0.3672919	0.20563146	0.123468724	-0.6503276	0.45248395
X8	0.3964218	0.3208648	0.27645845	-0.069339783	0.3448821	-0.25094786
	PC7	PC8				

```

X1 -0.18198549  0.219628939
X2 -0.26313489  0.439428491
X3  0.45614646 -0.637308817
X4 -0.04376550  0.005167487
X5 -0.02476905 -0.003932809
X6  0.63954673  0.350371723
X7 -0.13749489  0.109903482
X8 -0.50903627 -0.466499638

```

en PC1, las variables con los mayores valores absolutos mayores en magnitud son X6, X7, y X8. Esto significa que estas tres variables tienen una contribución significativa al primer componente principal. En PC2, las variables más influyentes son X1, X2, y X3, mientras que en PC3, las variables más influyentes son X4, X5, y X6.

```
head(pca$x)
```

	PC1	PC2	PC3	PC4	PC5	PC6
[1,]	-2.6636458	1.323754	-0.6734126	-0.006578704	0.02543303	0.13579392
[2,]	-2.0489412	1.107953	-0.4377562	0.137707486	-0.14295528	0.20289829
[3,]	-1.3342320	1.433398	0.9150931	-0.177507386	-0.13951493	-0.18814527
[4,]	0.2564063	2.117565	-0.2836960	-0.027359029	-0.19512574	-0.01758377
[5,]	-2.5869249	-1.129072	2.1037315	-0.462653490	0.14578153	-0.04768086
[6,]	-1.9892208	-1.333303	0.2614460	0.666490276	0.37790932	-0.00566878

	PC7	PC8
[1,]	0.08875470	-0.01239567
[2,]	-0.15253052	-0.01524319
[3,]	0.08732463	-0.02144284
[4,]	0.07698271	0.04477051
[5,]	-0.06050534	-0.02491833
[6,]	0.06063114	0.03525300

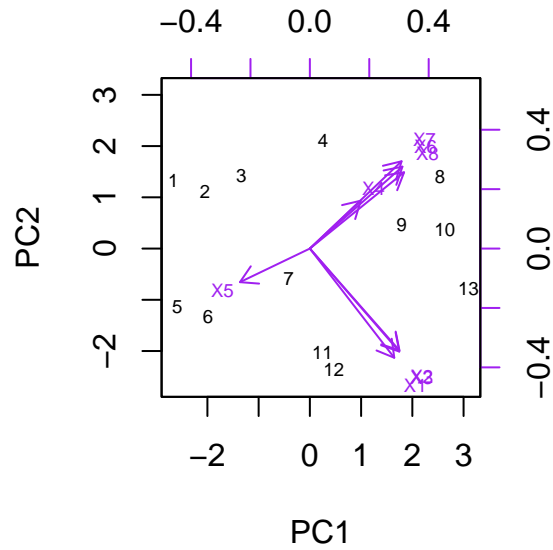
Cada valor en las celdas de `pc$x` representa la proyección de una observación en el espacio definido por los componentes principales. Por ejemplo, en la primera fila de `pca$x`, los números -2.66, 1.32, -0.67, -0.01, 0.03, 0.14, 0.09 y -0.01 indican la posición relativa de esa observación a lo largo de cada uno de los ocho componentes principales. Cuanto más cercanas estén dos observaciones en este espacio, mayor será su similitud en términos de las características originales que se han reducido a través del análisis de componentes principales.

```
dim(pca$x)
```

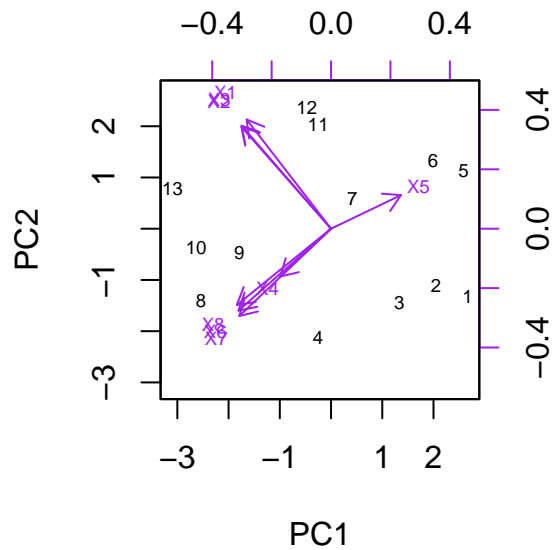
```
[1] 13  8
```

significa que hay 13 observaciones en los datos originales y 8 componentes principales resultantes del análisis de componentes principales (PCA)

```
biplot(x = pca, scale = 0, cex = 0.6, col = c("black", "purple"))
```



```
pca$rotation <- -pca$rotation  
pca$x <- -pca$x  
biplot(x = pca, scale = 0, cex = 0.6, col = c("black", "purple"))
```



```
library(ggplot2)
pca$sdev^2
```

```
[1] 4.202467342 2.129335069 1.452848665 0.145329749 0.047957456 0.013948110
[7] 0.006509750 0.001603858
```

El primer componente principal tiene una varianza explicada de aproximadamente 4.20. El segundo componente principal tiene una varianza explicada de aproximadamente 2.13. El tercer componente principal tiene una varianza explicada de aproximadamente 1.45. Y así sucesivamente para los demás componentes principales.

```
library(ggplot2)
pca$sdev^2
```

```
[1] 4.202467342 2.129335069 1.452848665 0.145329749 0.047957456 0.013948110
[7] 0.006509750 0.001603858
```

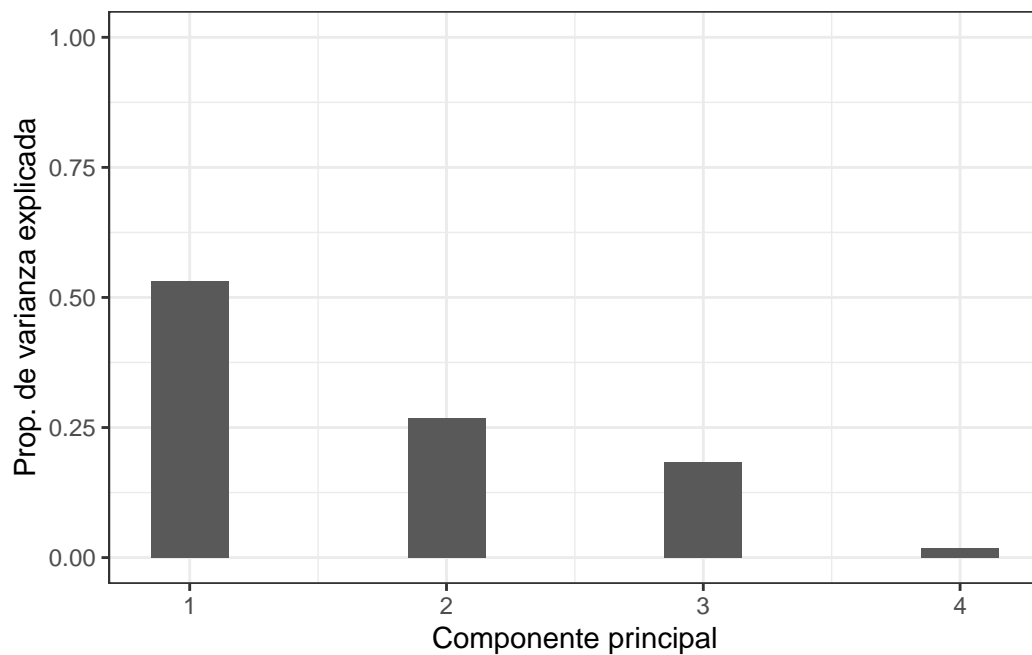
cada valor en la lista explica la varianza asociada con cada componente principal en orden descendente de importancia.

```
prop_varianza <- pca$sdev^2 / sum(pca$sdev^2)
prop_varianza
```

```
[1] 0.5253084178 0.2661668837 0.1816060831 0.0181662186 0.0059946820
[6] 0.0017435138 0.0008137188 0.0002004823
```

cada valor en la lista indica la proporción de varianza explicada por el componente principal correspondiente en relación con la varianza total de los datos.

```
ggplot(data = data.frame(prop_varianza, pc = 1:4),
       aes(x = pc, y = prop_varianza)) +
  geom_col(width = 0.3) +
  scale_y_continuous(limits = c(0,1)) +
  theme_bw() +
  labs(x = "Componente principal",
       y = "Prop. de varianza explicada")
```



Al observar el grafico anterior se puede ver que los primeros tres componentes principales son los más importantes para comprender la estructura y la variabilidad en los datos.

```
prop_varianza_acum <- cumsum(prop_varianza)
prop_varianza_acum
```

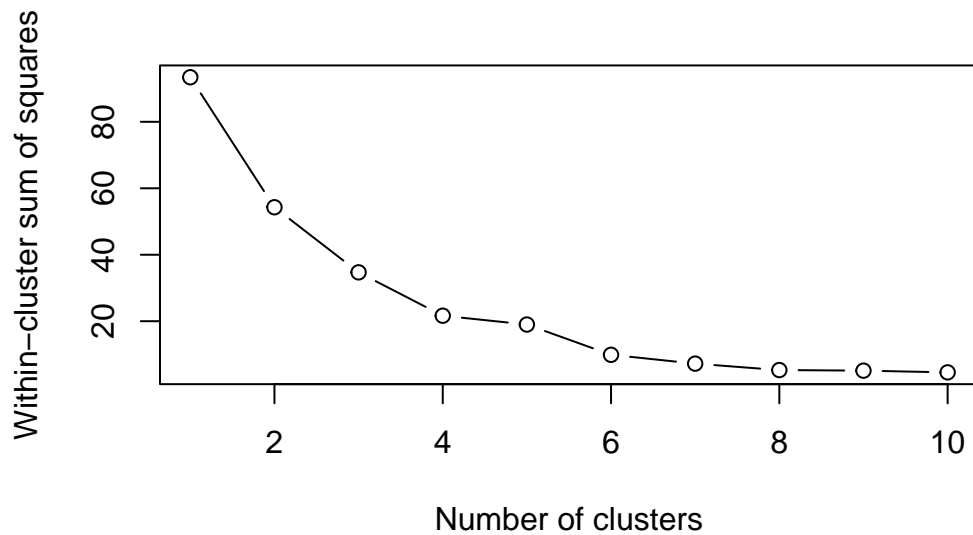
```
[1] 0.5253084 0.7914753 0.9730814 0.9912476 0.9972423 0.9989858 0.9997995
[8] 1.0000000
```

El primer valor, 0.5253084, indica que el primer componente principal explica el 52.53% de la varianza total en los datos. El segundo valor, 0.7914753, muestra que los dos primeros componentes principales juntos explican el 79.15% de la varianza. El tercer valor, 0.9730814, indica que los tres primeros componentes principales explican el 97.31% de la varianza acumulada. Y así sucesivamente, hasta llegar al último valor de 1.0000000, que muestra que todos los componentes principales juntos explican el 100% de la varianza. como se tenia previsto.

os primeros dos o tres componentes principales retienen la mayoría de la información en los datos originales, lo que sugiere que podrían ser suficientes para un análisis más simplificado sin perder mucha información.

```
library(cluster)

# Determinar el número óptimo de clusters
wss <- numeric(10)
for (i in 1:10) {
  wss[i] <- sum(kmeans(pca$x[, 1:3], centers = i)$withinss)
}
plot(1:10, wss, type = "b", xlab = "Number of clusters", ylab = "Within-cluster sum of squares")
```

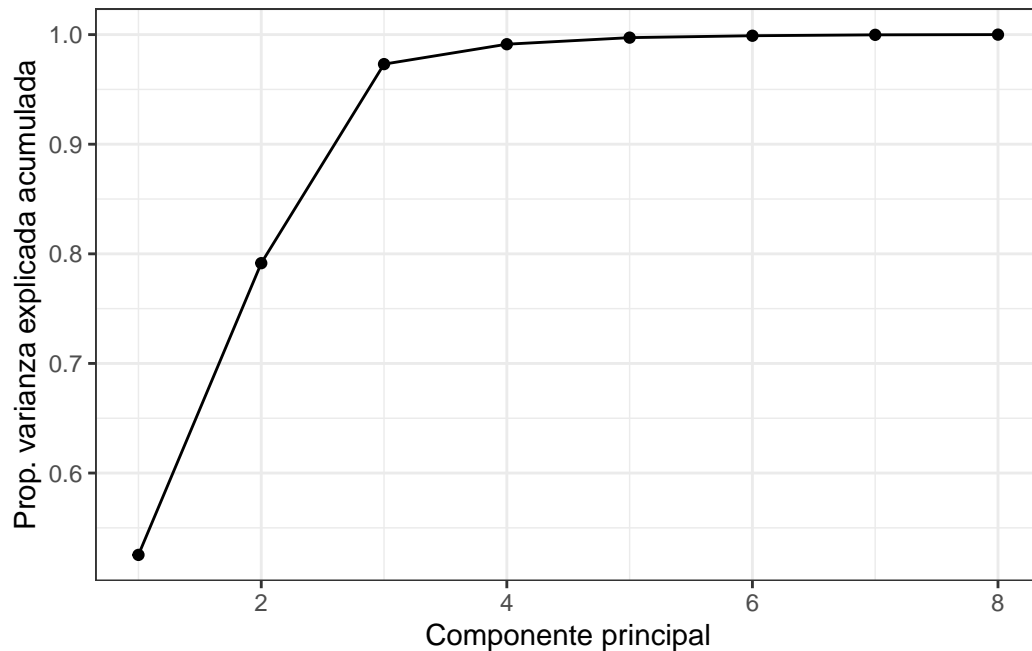


```
# Aplicar k-means con el número óptimo de clusters
k <- which.min(wss)
clusters <- kmeans(pca$x[, 1:3], centers = k)
```

al elegir los numeros de clustes que son tres se puede observar en el grafico si estarian bien tomar esos datos.

en este caso se hara de las variables anteriores que eran 8

```
ggplot(data = data.frame(prop_varianza_acum, pc = 1:8),
       aes(x = pc, y = prop_varianza_acum, group = 1)) +
  geom_point() +
  geom_line() +
  theme_bw() +
  labs(x = "Componente principal",
       y = "Prop. varianza explicada acumulada")
```



**¿cómo podrías consolidar estos múltiples indicadores para facilitar una comparación efectiva entre las ciudades?**

se puede hacer dividiendo las ciudades en tres clusters, donde se puede observar que algunas ciudades están liderando en prácticas sostenibles, representadas por un alto puntaje en sostenibilidad en comparación con otras ciudades que podrían beneficiarse de mejoras en sus iniciativas sostenibles debido a puntajes más bajos en los indicadores.

#### Ejercicio 4:

```
library(readxl)
ejercicio_4 <- read_excel("C:/Users/DELL/Documents/seminario/ejercicio 4.xlsx")
ejercicio_4
```

```
# A tibble: 10 x 3
  Variables Tienda Marca
  <chr>      <dbl> <dbl>
1 A         2     2
2 B         5     9
3 C         3     5
4 D         4     8
5 E         6     7
```



6	F	9	3
7	G	8	4
8	H	7	5
9	I	4	6
10	J	5	7

```
datos<-(ejercicio_4)
k.means.fit <-kmeans(datos[,2:3], 3, nstart = 10)
k.means.fit
```

K-means clustering with 3 clusters of sizes 5, 3, 2

Cluster means:

	Tienda	Marca
1	4.8	7.4
2	8.0	4.0
3	2.5	3.5

Clustering vector:

```
[1] 3 1 3 1 1 2 2 2 1 1
```

Within cluster sum of squares by cluster:

```
[1] 8 4 5
```

(between\_SS / total\_SS = 80.8 %)

Available components:

[1]	"cluster"	"centers"	"totss"	"withinss"	"tot.withinss"
[6]	"betweenss"	"size"	"iter"	"ifault"	

En el resultado anterior se puede observar que indica que el 80.8% de la variabilidad total es explicada por la agrupación en 3 clusters.

```
k.means.fit$centers
```

	Tienda	Marca
1	4.8	7.4
2	8.0	4.0
3	2.5	3.5

Estos valores dan una idea de cómo están distribuidos los datos alrededor los centros en cada cluster como para el cluster 1, las observaciones tienden a tener valores más altos en la variable tienda y valores también altos en la variable marca, mientras que para el cluster 3, las observaciones tienden a tener valores bajos en ambas variables.

```
k.means.fit$ifault
```

```
[1] 0
```

este resultado es que no se ha producido ninguna falla o error durante el proceso de clustering mediante k-means.

```
grupos=k.means.fit$cluster  
table(datos$Variables,grupos)
```

```
grupos  
  1 2 3  
A 0 0 1  
B 1 0 0  
C 0 0 1  
D 1 0 0  
E 1 0 0  
F 0 1 0  
G 0 1 0  
H 0 1 0  
I 1 0 0  
J 1 0 0
```

Cada celda de la tabla indica cuántas observaciones de la variable correspondiente están asignadas a cada grupo.

```
library(dplyr)
```

Adjuntando el paquete: 'dplyr'

The following objects are masked from 'package:stats':

```
filter, lag
```

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
dif=data.frame(datos,grupos)
dif=data.frame(dif) %>%
  mutate(grupos=dplyr::recode(grupos,
                                "3"="A",
                                "1"="B",
                                "3"="C",
                                "1"="D",
                                "1"="E",
                                "2"="F",
                                "2"="G",
                                "2"="H",
                                "1"="I",
                                "1"="J"))
table(dif$grupos,dif$Variable)
```

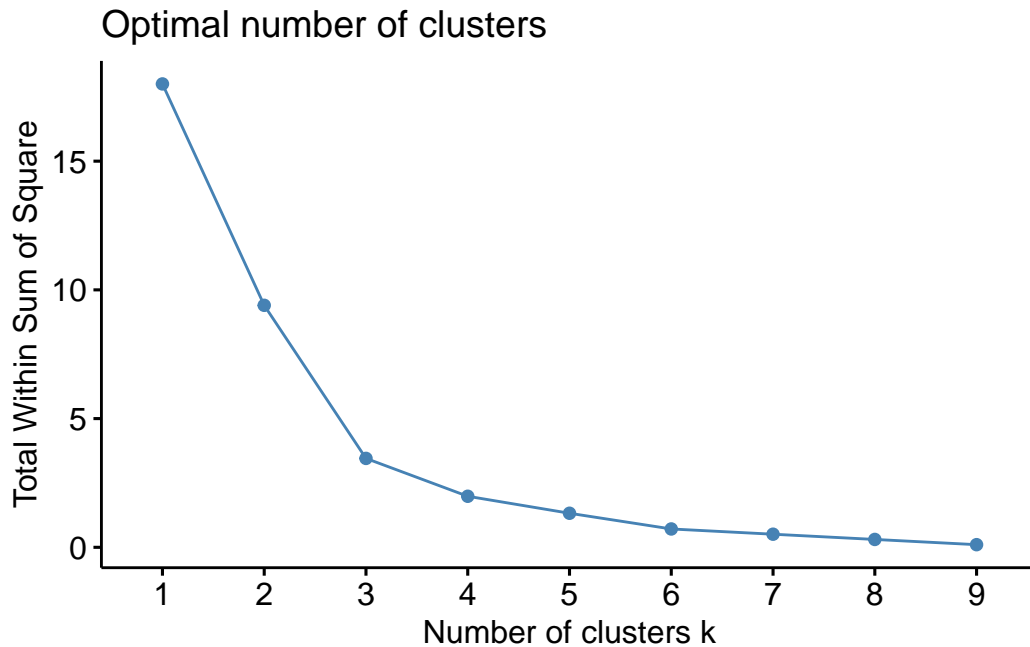
	A	B	C	D	E	F	G	H	I	J
C	1	0	1	0	0	0	0	0	0	0
H	0	0	0	0	0	1	1	1	0	0
J	0	1	0	1	1	0	0	0	1	1

**Diferencia en la Lealtad:** Al observar los puntajes de lealtad hacia las tiendas y las marcas, se puede notar que hay una variación considerable entre los encuestados. Algunos muestran una mayor lealtad hacia las tiendas locales como los encuestados F y G, mientras que otros tienen una mayor lealtad hacia las marcas como encuestados B e I.

**Relación entre Tienda y Marca:** También se puede notar que la lealtad hacia las tiendas y hacia las marcas no siempre está alineada. Por ejemplo, el encuestado F muestra una alta lealtad hacia la tienda puntaje de 9, pero una baja lealtad hacia la marca como puntaje de 3, lo que indica una desconexión entre su preferencia por la tienda local y la marca en sí misma.

Aunque los datos son limitados ya que solo son 10 encuestados, se pueden observar patrones generales en las preferencias de los consumidores. Algunos tienen una lealtad equilibrada entre tienda y marca, mientras que otros muestran preferencias más marcadas hacia una u otra.

```
d2 <- scale(datos[,2:3])
rownames(d2) <- datos$Variables
fviz_nbclust(x = d2, FUNcluster = kmeans, method = "wss", k.max = 9,
              diss = get_dist(d2, method = "euclidean"), nstart = 50)
```



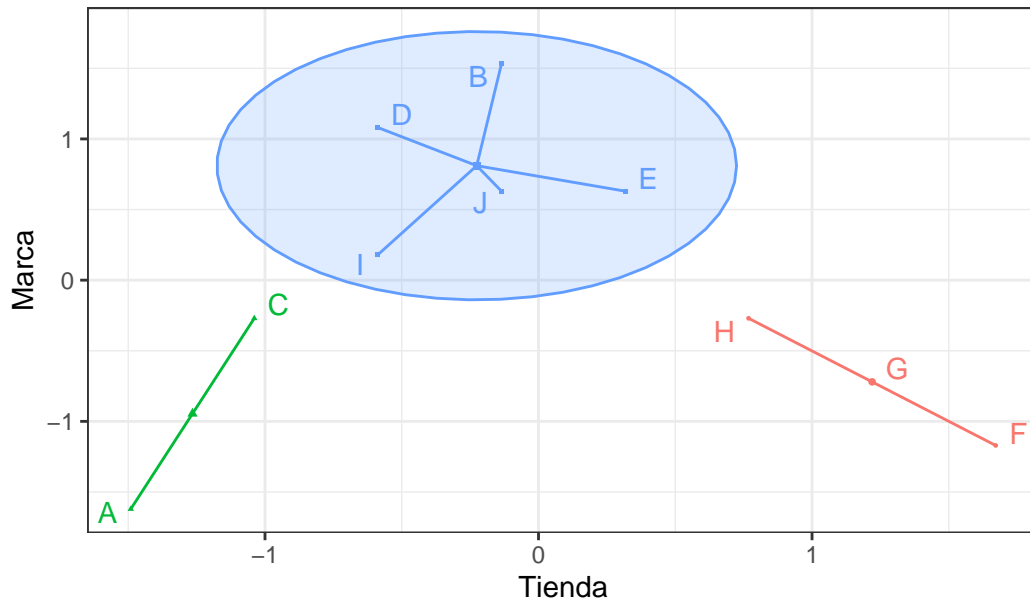
Esta figura de números óptimos de clusters indica que, con los datos actuales, puede hacerse una clasificación de máximo 9 clusters, aunque ya se ha evidenciado que tres clusters pueden ser suficientes.

```
d2f=data.frame(d2)
km_clusters <- kmeans(x = d2f, centers = 3, nstart = 50)

# Las funciones del paquete factoextra emplean el nombre de las filas del
# dataframe que contiene los datos como identificador de las observaciones.
# Esto permite añadir labels a los gráficos.
fviz_cluster(object = km_clusters, data = d2f, show.clust.cent = TRUE,
              ellipse.type = "euclid", star.plot = TRUE, repel = TRUE,
              pointsize=0.5,outlier.color="black") +
  labs(title = "Resultados clustering K-means") +
  theme_bw() + theme(legend.position = "none")
```

Too few points to calculate an ellipse  
 Too few points to calculate an ellipse

## Resultados clustering K-means

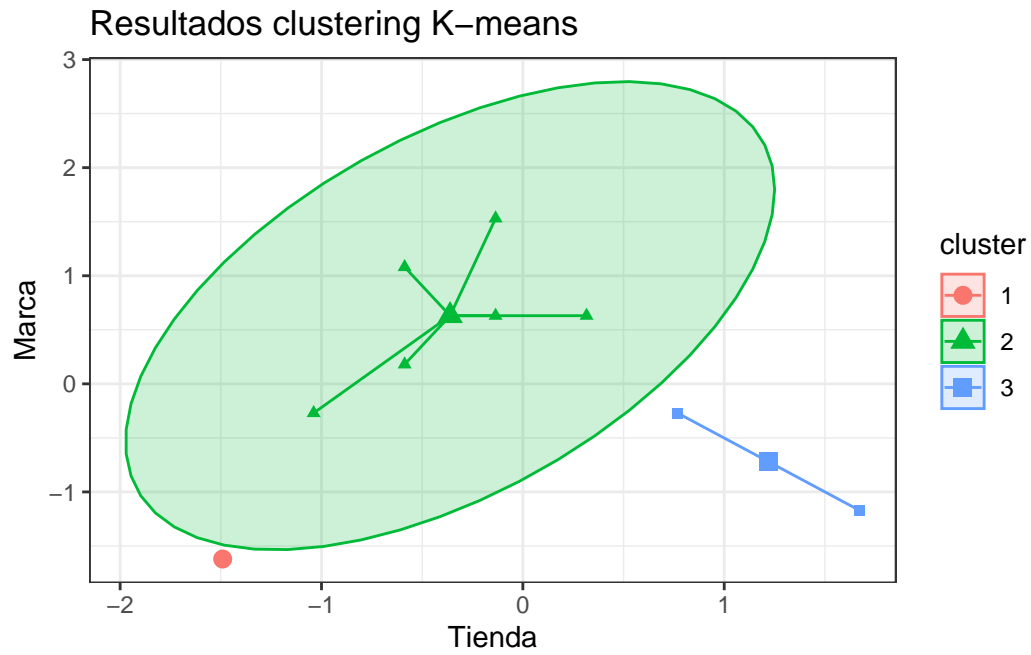


Se puede notar que las preferencias de las personas están influenciadas por su nivel de lealtad hacia la tienda o la marca. Por ejemplo, el encuestado F muestra una alta lealtad hacia la tienda con un nivel de 9, lo que indica que su preferencia se inclina más hacia la tienda que hacia la marca, ya que su lealtad hacia la marca es de 3.

Al observar los clusters, se puede ver que en el grupo de color rojo se encuentran agrupados los encuestados D, B, E, J e I. En el grupo de color azul se encuentran agrupados los encuestados H, G y F. Mientras que en el grupo verde se encuentran agrupados los encuestados C y A.

```
require(cluster)
pam.res <- pam(d2f, 3)
# Visualización
fviz_cluster(pam.res, geom = "point", ellipse.type = "norm",
              show.clust.cent = TRUE, star.plot = TRUE) +
  labs(title = "Resultados clustering K-means") + theme_bw()
```

Too few points to calculate an ellipse  
Too few points to calculate an ellipse



en esta grafica se puede observar mejor como se distribuyen los cluster concentrandose mas las variables en el cluster 2 de color verde.