

Ejercicio Propuesto 1 (Resuelto)

Una empresa de comercio electrónico desea medir el volumen de gasto mensual en euros en función de las reseñas de los clientes, los cupones y el número de visitas a la web de la empresa. Vamos a utilizar el archivo [GASTOS1.RData](#) que contiene el gasto mensual de los clientes en euros, números de reseñas positivas realizadas por el cliente, los cupones promocionales disponibles para el cliente y el número de visitas realizadas a la página web.

Se desea clasificar a los 24 clientes encuestados por el número de visitas realizadas a la página web (Visitas), el gasto mensual de los clientes en euros (Gasto) y los cupones promocionales disponibles para el cliente (Cupones)

Se pide:

1. Realizar un diagrama de dispersión 3-D para mostrar la distribución de los datos y estudiar los posibles grupos que se pueden hacer.
2. Utilizar un análisis clúster jerárquico. (Etiquetar los casos mediante Identificación personal, id)
 - I. Método: Vecino más lejano; Medida: Intervalo- Distancia euclídea al cuadrado; Transformar valores: Estandarizar las variables (puntuaciones Z)
 - II. Obtener la Matriz de distancia y el Dendograma
 - III. Analizar las tablas obtenidas y sacar conclusiones

Solución

Considerando las variables de interés gastos y número de visitas a la página web, repetir el análisis hecho en el ejercicio propuesto 3. Para realizar el apartado 1.

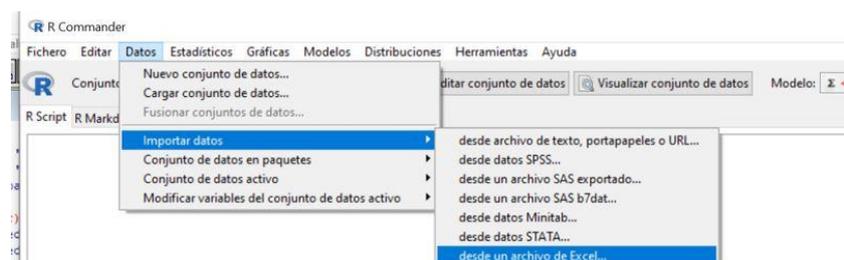
- Seleccionar en el menú principal Gráficas/Gráficos 3D/Diagrama de dispersión en 3D
- Variables explicativas: visitas y cupones; Variable explicada: gasto.
- Pestaña Opciones: Identificar Observaciones Interactivamente con el ratón.

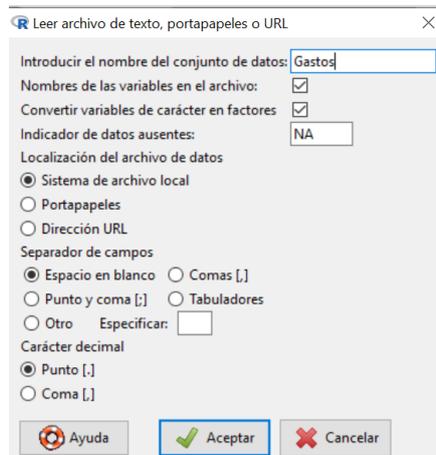
En primer lugar:

```
> library(Rcmdr)
```

1. Realizar un diagrama de dispersión 3-D para mostrar la distribución de los datos y estudiar los posibles grupos que se pueden hacer.

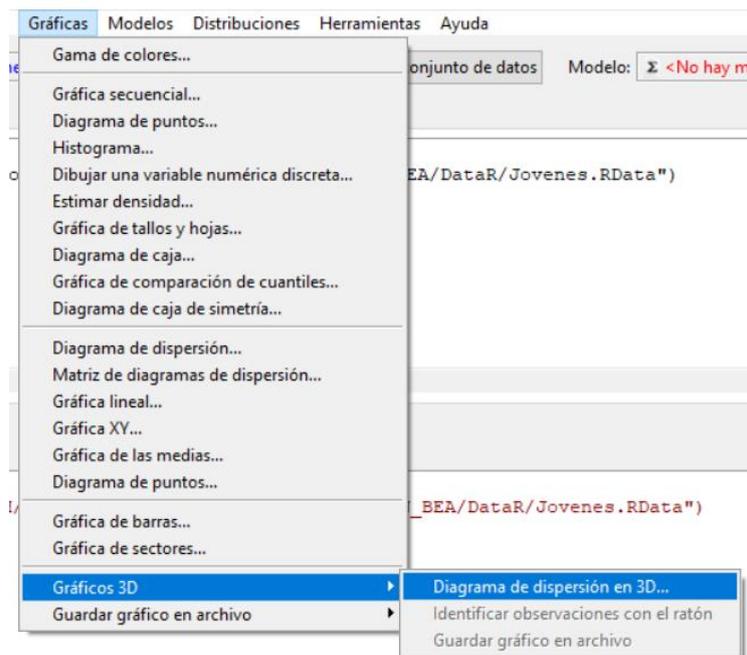
En primer lugar tenemos que cargar los datos



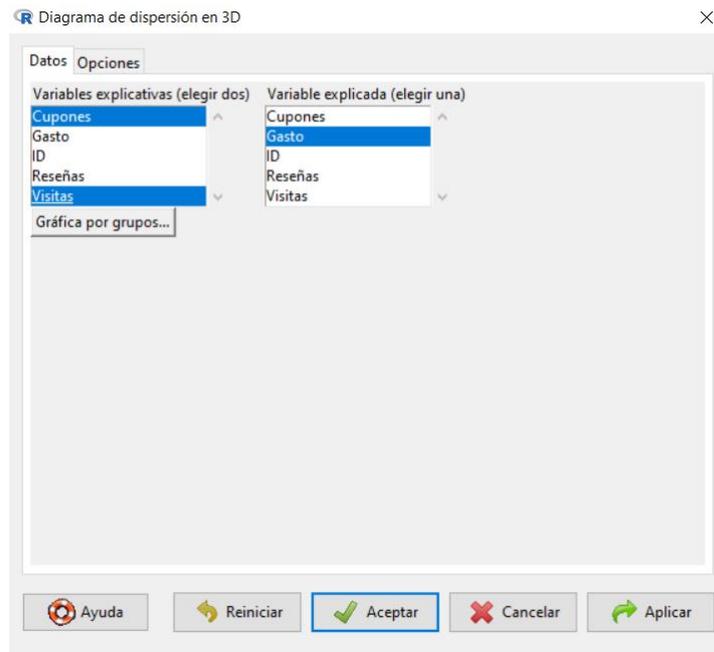


Pulsamos **Aceptar** y elegimos el dichero: [Gastos.xlsx](#) Gastos.xlsx

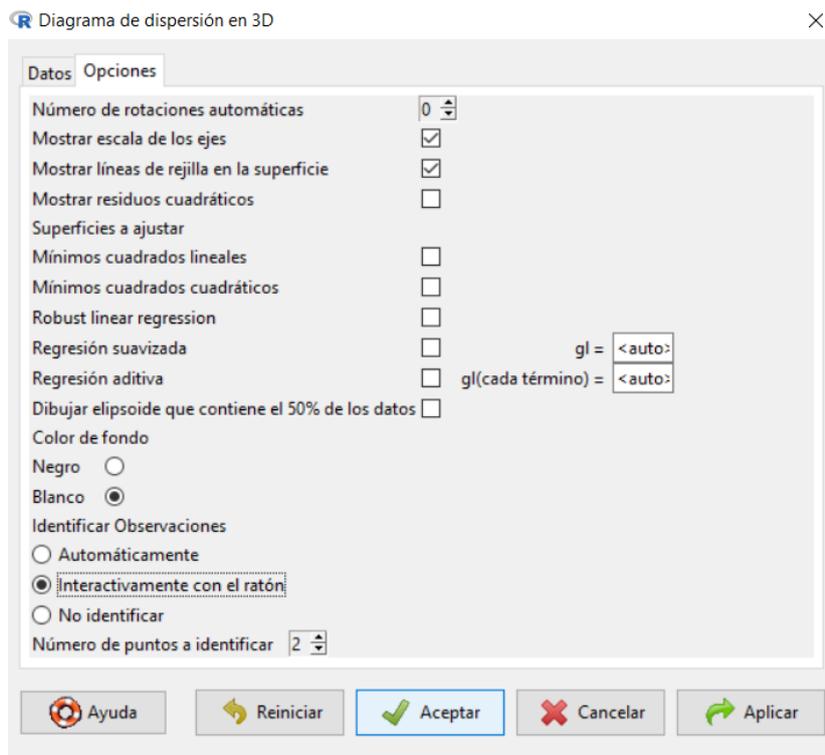
Para elaborar el gráfico, tenemos que ir a la opción *Gráficas*, seleccionar *Gráficos 3D*, *Diagrama de dispersión en 3D*. Ver la siguiente imagen



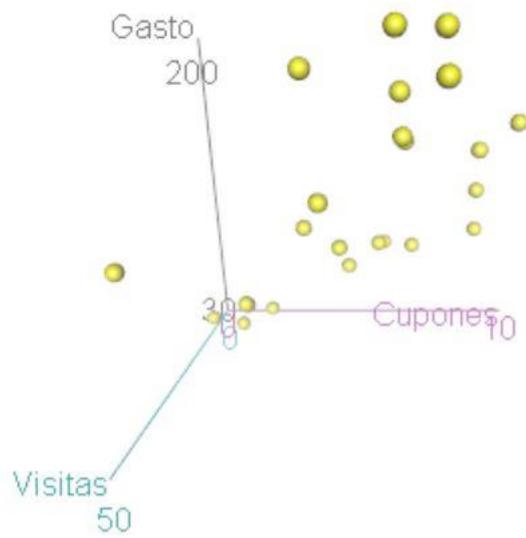
Se abrirá una ventana donde se muestra dos cuadros para seleccionar las *Variables explicativas* (se selecciona las variables cupones y visitas) y en *Variable explicada*, se selecciona la variable gasto.



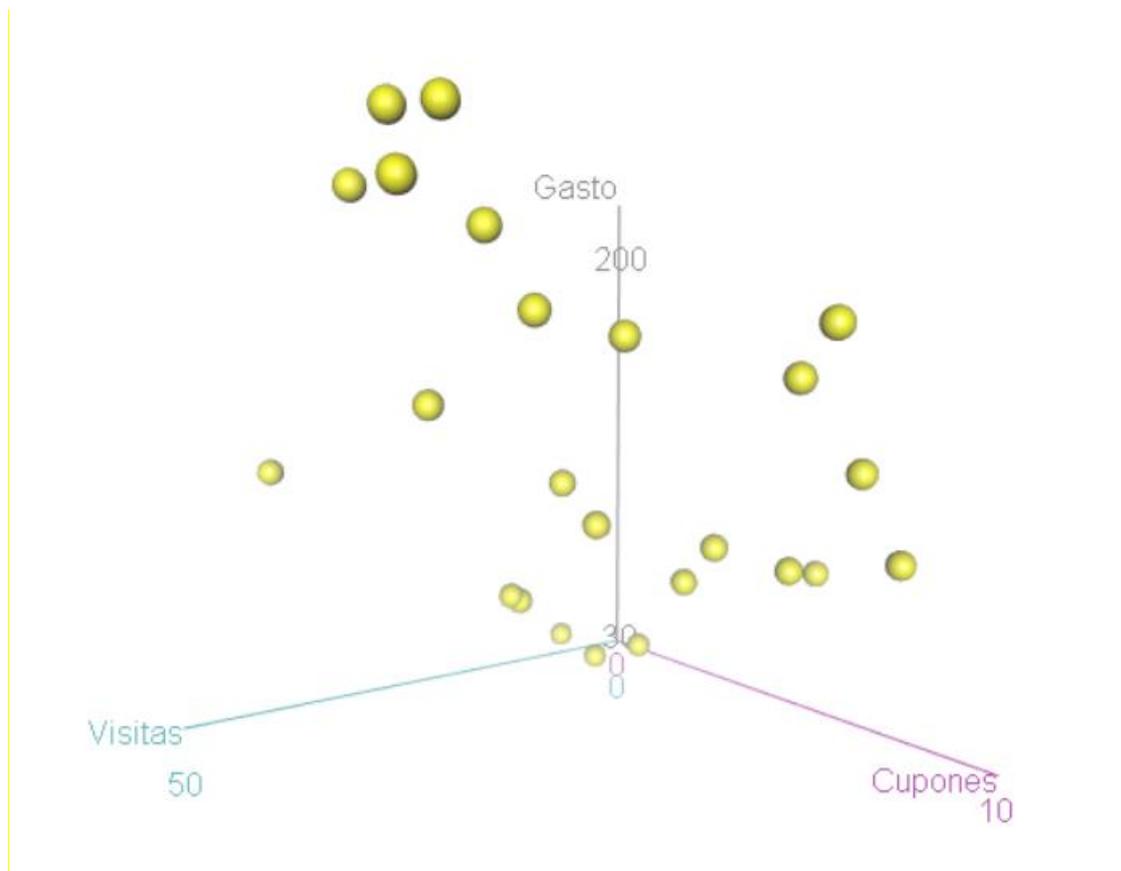
Por último, en *Opciones*, seleccionamos en *identificar Observaciones*, Interactivamente con el ratón. Terminamos con *Aceptar*.



Se obtiene el siguiente gráfico



Cuando se obtiene el gráfico, presiona el lado derecho del ratón y con el cursor podrás rotar el gráfico.

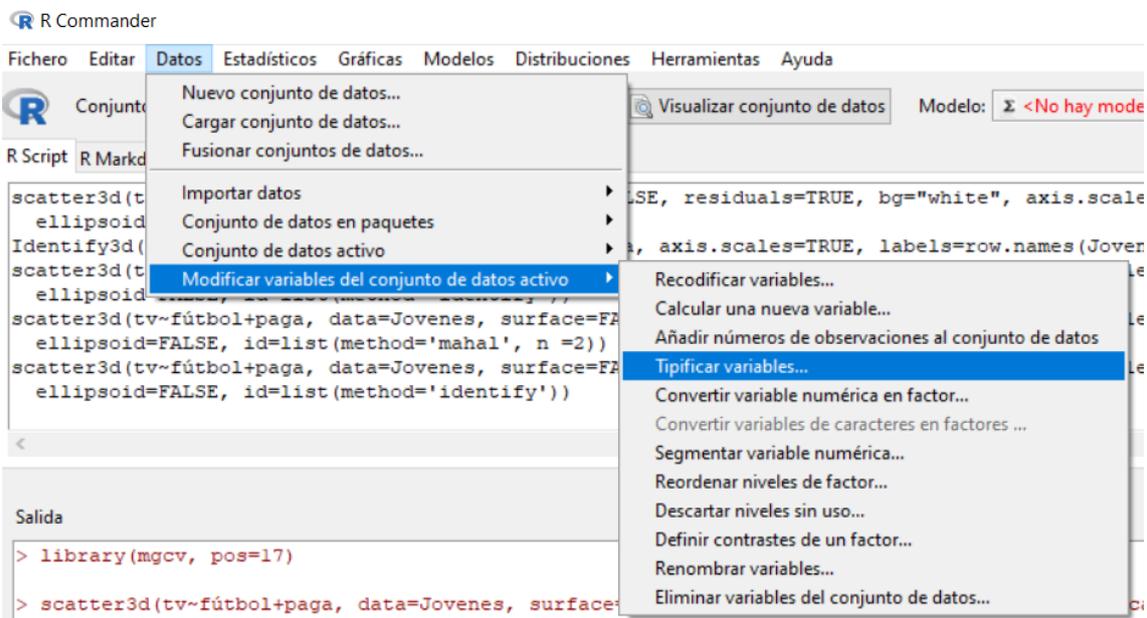


Observando el gráfico podemos observar 3 grupos de sujetos considerando estas variables.

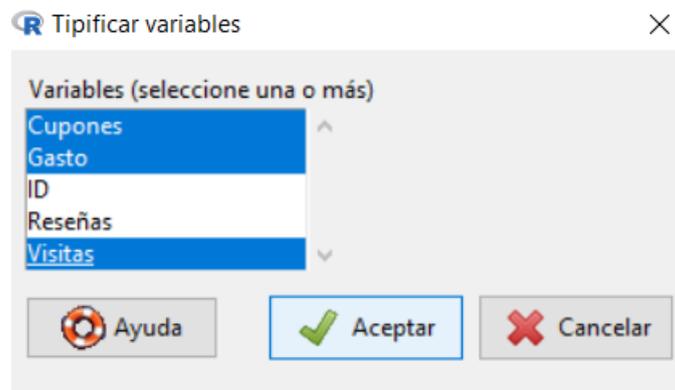
2. Utilizar un análisis clúster jerárquico. (Etiquetar los casos mediante Identificación personal, id)

- IV. Método: Vecino más lejano; Medida: Intervalo- Distancia euclídea al cuadrado; Transformar valores: Estandarizar las variables (puntuaciones Z)
- V. Obtener la Matriz de distancia y el Dendograma
- VI. Analizar las tablas obtenidas y sacar conclusiones

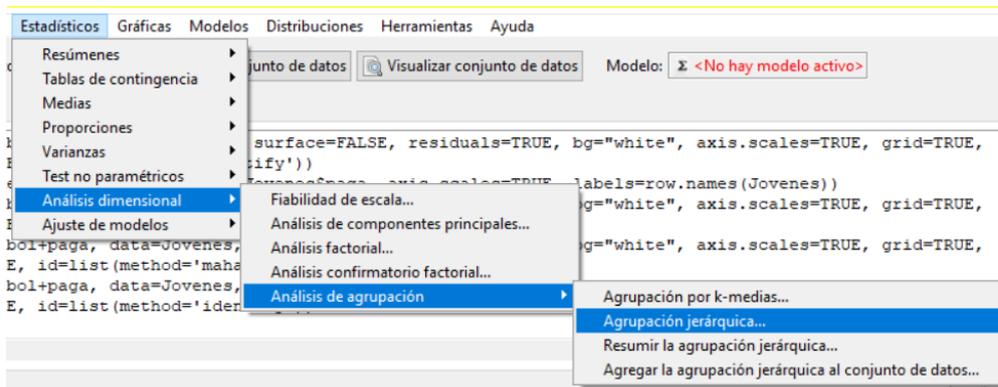
Primero, vamos a estandarizar las variables con las que vamos a trabajar: Cupones, Visitas, Gasto, y trabajamos con las variables estandarizadas. Para estandarizar las variables, vamos a la opción **Datos, Modificar variables del conjunto de datos activo, Tipificar variables**.



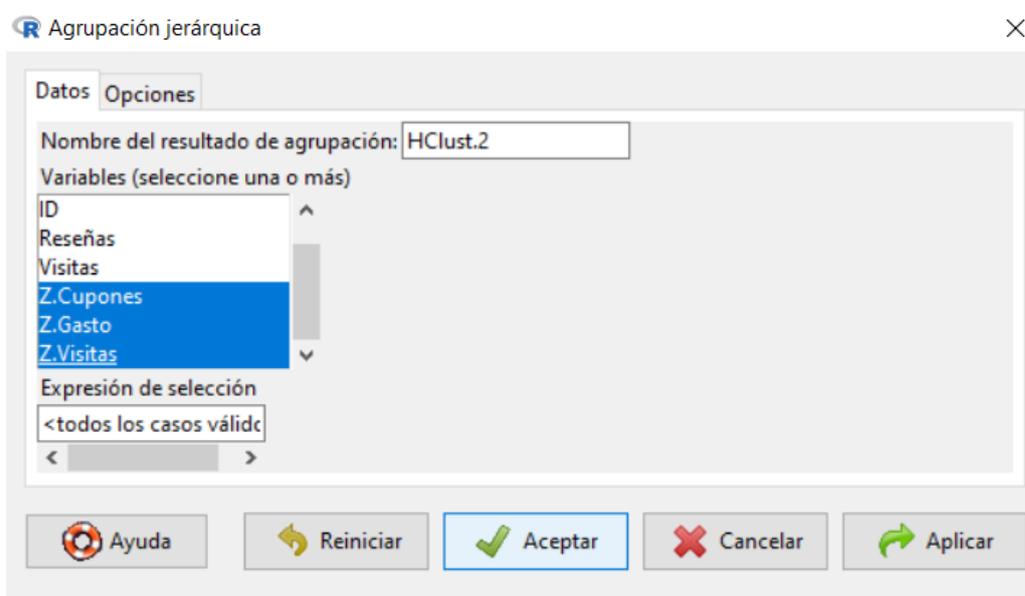
Seleccionamos las variables que se va a estandarizar. Luego damos en **Aceptar** y se guardarán las variables estandarizadas en la base de datos.



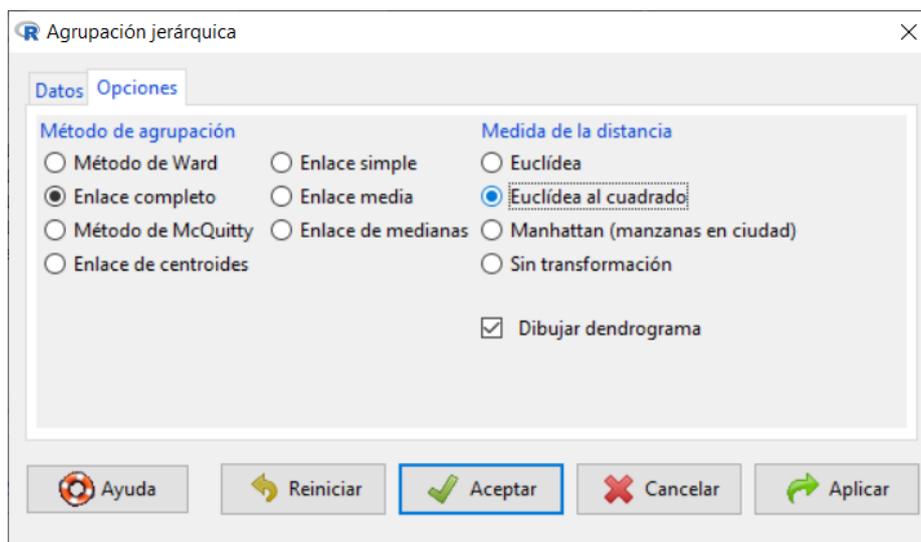
En segundo lugar, para agrupar o clúster los datos, vamos a seleccionar la opción **Estadísticos, Análisis dimensional, Análisis de agrupación, Agrupación jerárquico**



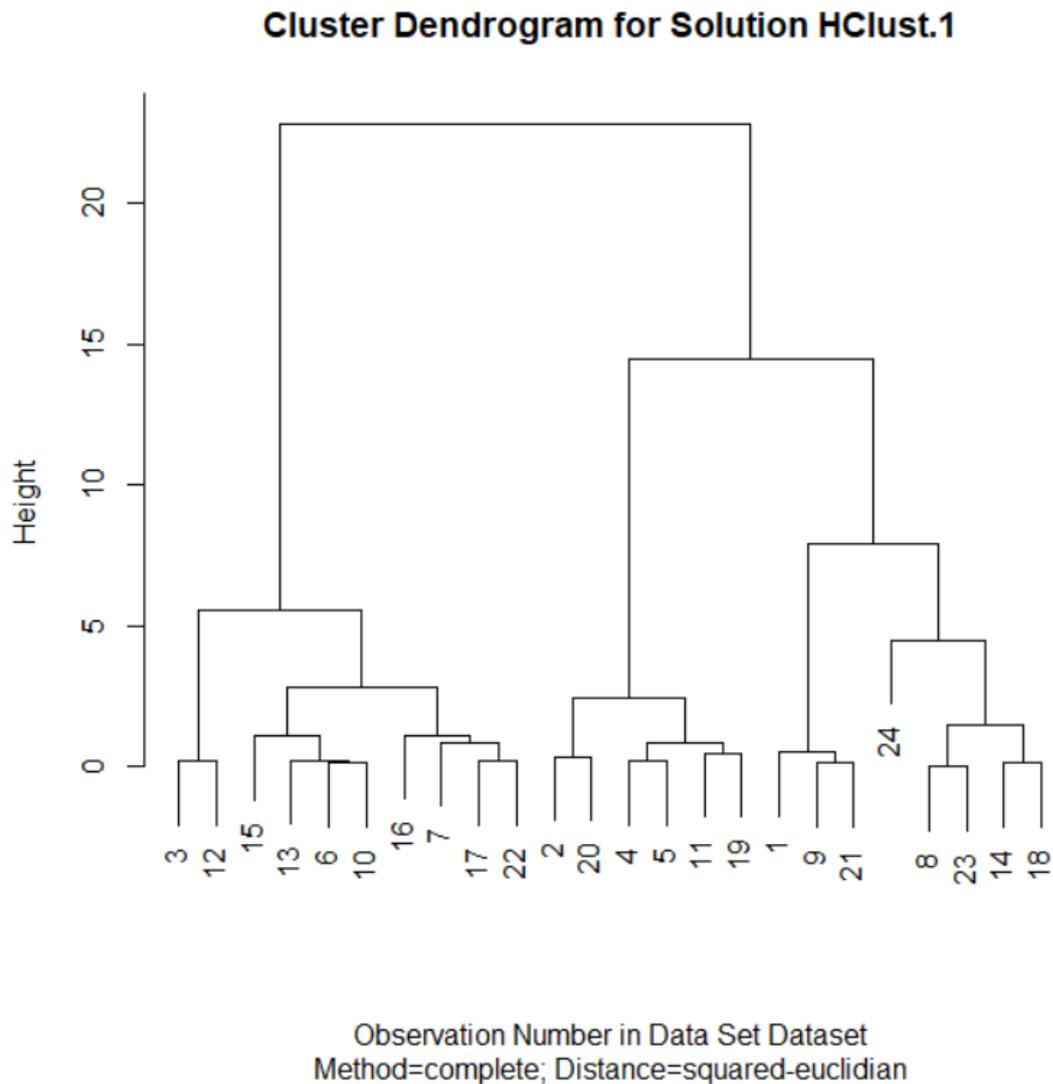
En la pestaña **Datos**, seleccionamos las variables estandarizadas (ZCupones, ZGasto, ZVisitas).



En *Opciones*, seleccionamos en **Método de agrupación**: *Enlace completo* y en **Medida de la distancia**: *Euclídea al cuadrado*. Para terminar, damos en *Aceptar*.



Se pulsa aceptar y se muestra el siguiente Dendrograma.



Teniendo en cuenta este dendrograma y el diagrama de dispersión obtenido podemos pensar en una solución de tres clústeres.

- Clúster 1: (3, 12, 15, 13, 6, 10, 16, 7, 17, 22)
- Clúster 2: (2, 20, 4, 5, 11, 19,)
- Clúster 3: (1, 9, 21, 8, 23, 14, 18, 24)

Por último, tenemos la matriz de distancias:

```
> d <- dist(model.matrix(~-1 + Z.Cupones+Z.Gasto+Z.Visitas, Gastos))^2
```

```
> d
```

