

Ejercicio Propuesto 2 (Resuelto)

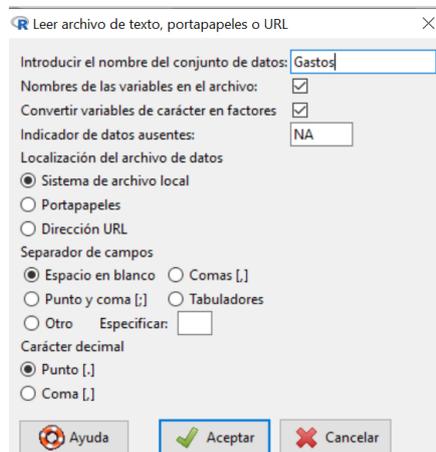
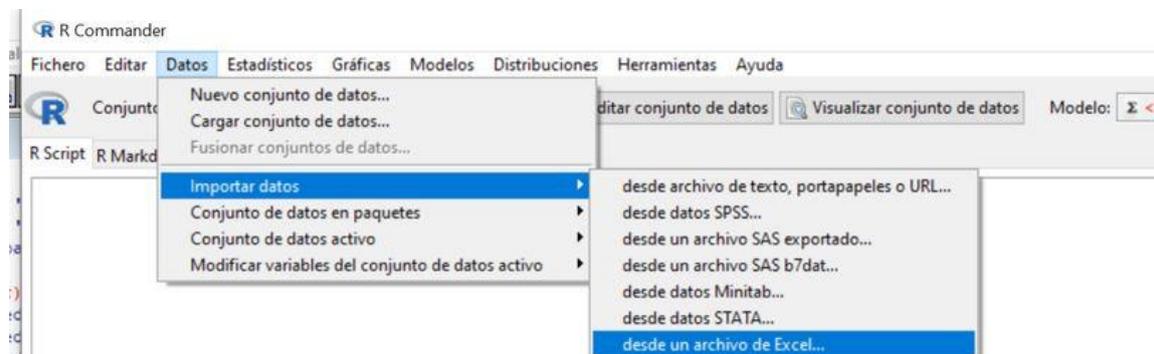
Vamos a utilizar de nuevo el archivo de datos [GASTOS1.RData](#) que contiene información sobre 24 clientes.

Se pide:

1. **Tipificar las variables visitas, cupones y gasto**
2. **Realizar un análisis de conglomerados de k-medias con tres conglomerados según las variables tipificadas visitas, cupones y gasto ($Z_{visitas}$, $Z_{cupones}$ y Z_{gasto}).**
3. **Resumir los resultados obtenidos e interpretar la solución.**

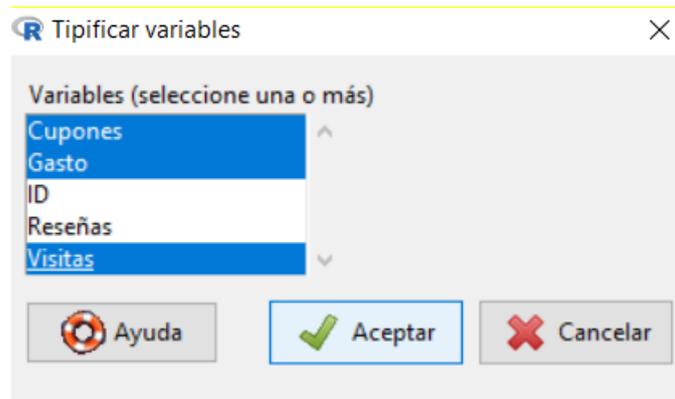
Solución

En primer lugar, vamos a cargar el conjunto de datos

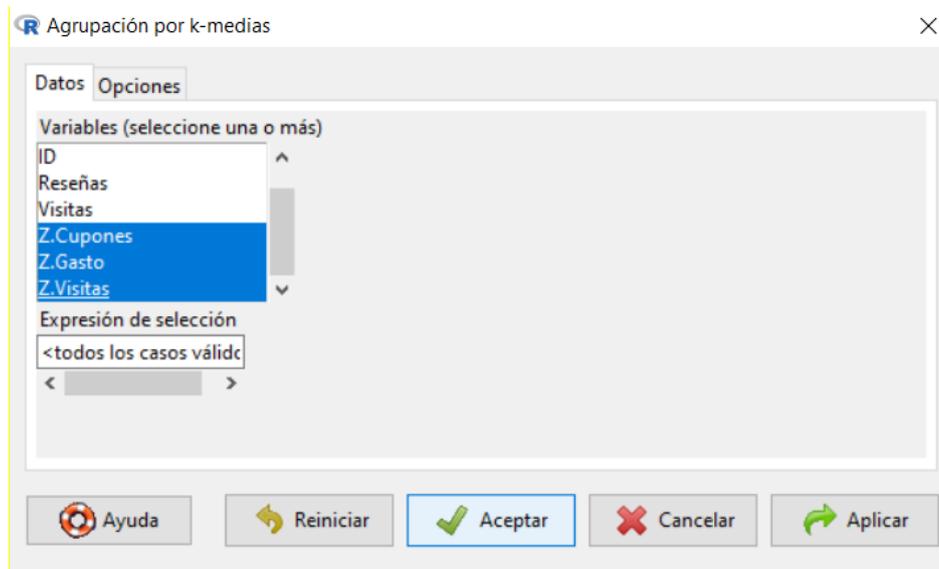


Pulsamos **Aceptar** y elegimos el dichero: [Gastos1.xlsx](#)

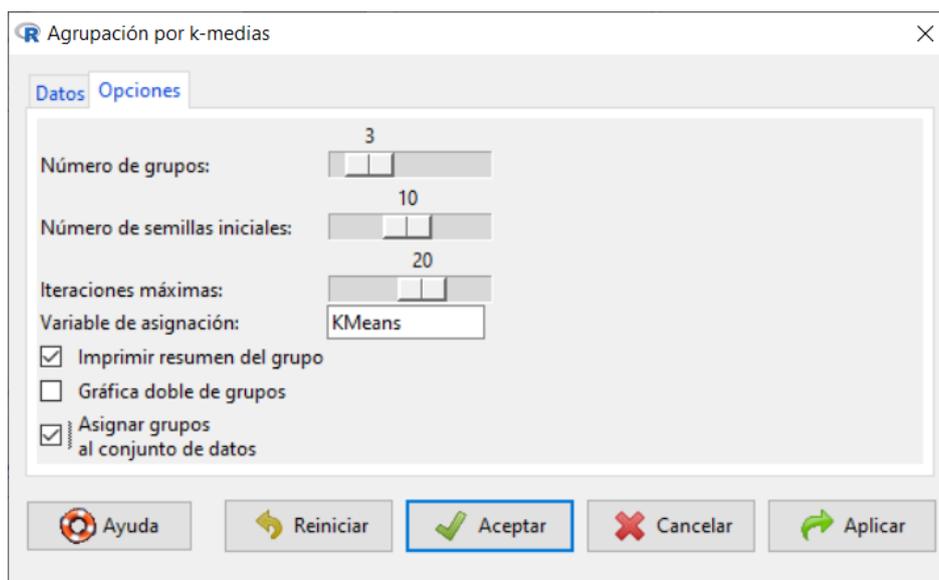
Empezamos tipificando las variables, en el menú **Datos**, **Modificar variables del conjunto de datos activo**, **Tipificar variables**. Seleccionamos las variables que deseamos tipificar: Cupones, Gasto, Visitas. Trabajamos con las variables estandarizadas.



Realizamos el procedimiento **Agrupación k-medias**, y seleccionamos las variables que acabamos de crear en la pestaña **Datos**.



En la pestaña **Opciones**, **Número de grupos: 3** y **Iteraciones máximas: 20**. Pulsamos **Aceptar**.



Las salidas son las siguientes:

```

> .cluster$size # Cluster Sizes
[1] 6 8 10

> .cluster$centers # Cluster Centroids
new.x.Z.Cupones new.x.Z.Gasto new.x.Z.Visitas
1 -1.4483016 -1.0948018 -0.4238294
2 0.4358966 1.1145775 1.1219014
3 0.5202637 -0.2347809 -0.6432235

> .cluster$withinss # Within Cluster Sum of Squares
[1] 6.103121 4.038434 9.080704

> .cluster$tot.withinss # Total Within Sum of Squares
[1] 19.22226

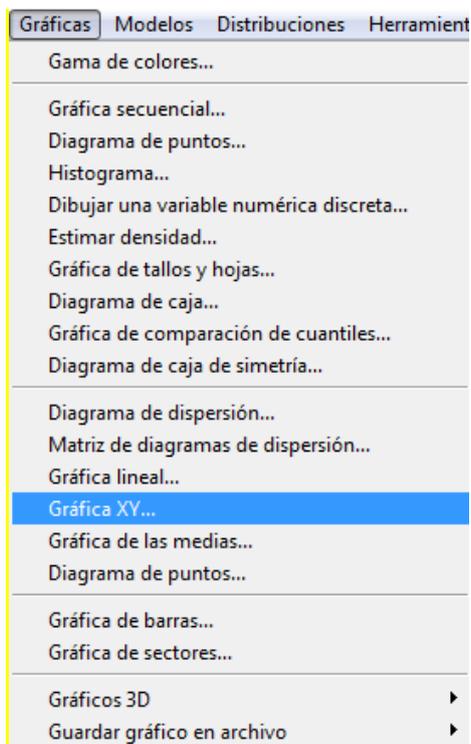
> .cluster$betweenss # Between Cluster Sum of Squares
[1] 49.77774

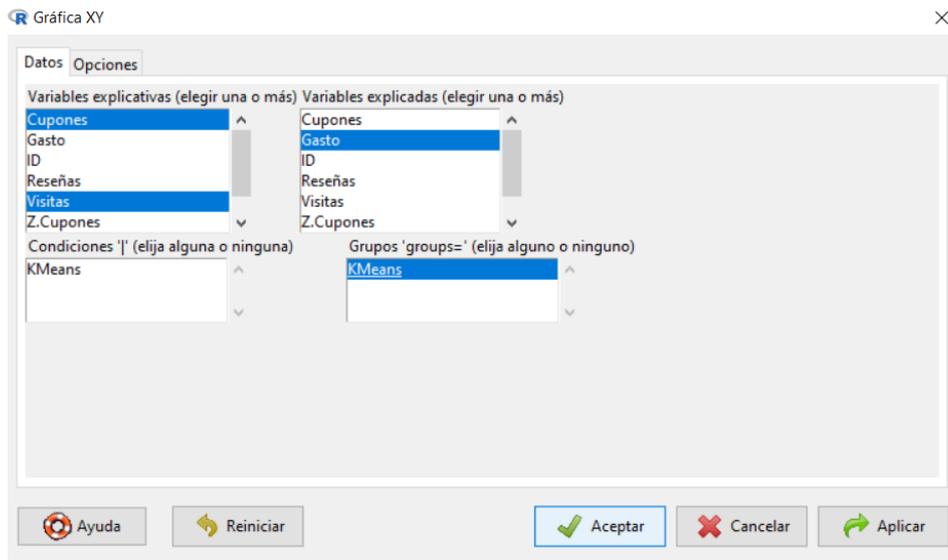
```

Los centros de los conglomerados que se han obtenido corresponden a:

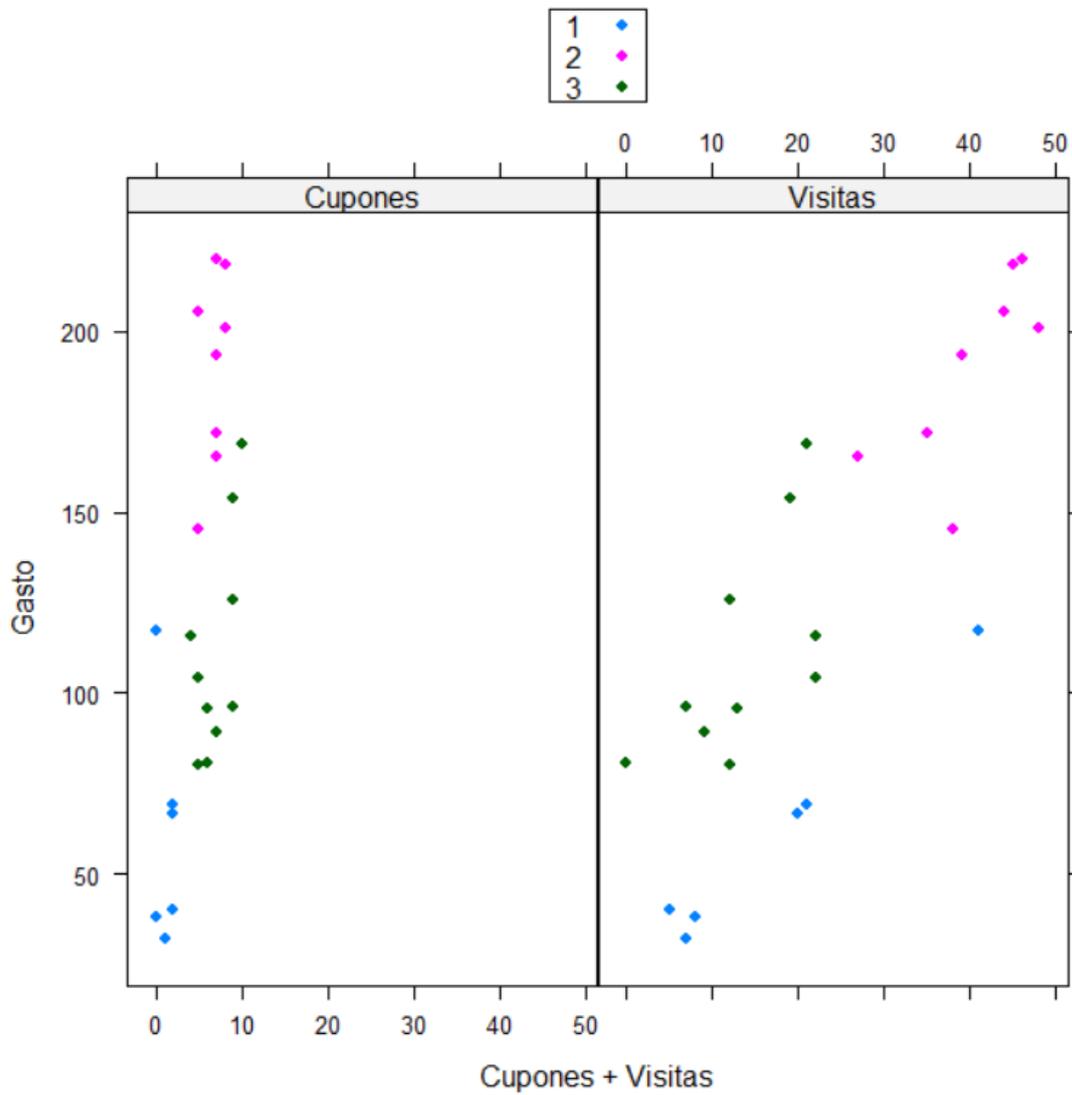
- Clúster 1: contienen 6 clientes.
- Clúster 2: 8 clientes.
- Clúster 3: 10 clientes.

Finalmente, representamos el diagrama de dispersión de la variable Cupones y Visitas en función de Gasto, identificando los conglomerados de pertenencia. Para ello nos vamos al menú **Gráficas, Gráfica XY**. En este menú, pestaña **Datos** seleccionamos como variables explicativas Cupones y Visitas, y como explicada la variable Gasto. En grupos ponemos que identifique los individuos con la variable KMeans.





Pulsamos *Aceptar*. Se obtiene el siguiente gráfico de dispersión



En este gráfico se puede observar que los clientes del clúster 1 tienen Gastos más bajas y que tienen menos de 5 Cupones, el número de Visitas a la página web está entre 10 y 40. Los del clúster 3 los clientes hacen un Gasto entre 130 a 180 euros, tienen entre 5 a 10 Cupones y las Visitas a la página web son menos de 20. Los del clúster 2 son los que más Gasto tienen, es decir, gastan más de 150 euros, usan entre 5 a 15 Cupones y las Visitas a la web lo hacen más de 20 veces.

ID	Gasto	Reseñas	Cupones	Visitas	Z.Cupones	Z.Gasto	Z.Visitas	KMeans
1	40.00000	16	2	5	-1.1670780	-1.44372586	-1.22162601	1
2	96.37964	11	9	7	1.1952004	-0.48262716	-1.08865991	3
3	153.83627	12	9	19	1.1952004	0.49683103	-0.29086334	3
4	80.00000	15	5	12	-0.1546730	-0.76184939	-0.75624467	3
5	95.52166	2	6	13	0.1827953	-0.49725299	-0.68976162	3
6	218.41755	8	8	45	0.8577320	1.59774231	1.43769591	2
7	165.16466	15	7	27	0.5202637	0.68994506	0.24100105	2
8	68.95851	19	2	21	-1.1670780	-0.95007272	-0.15789724	1
9	38.04224	9	0	8	-1.8420147	-1.47709958	-1.02217686	1
10	219.83571	14	7	46	0.5202637	1.62191768	1.50417896	2
11	80.46985	12	6	0	0.1827953	-0.75383984	-1.55404125	3
12	168.63266	1	10	21	1.5326687	0.74906370	-0.15789724	3
13	201.00000	11	8	48	0.8577320	1.30082694	1.63714506	2
14	115.72520	7	4	22	-0.4921413	-0.15284499	-0.09141419	3
15	205.61247	10	5	44	-0.1546730	1.37945534	1.37121287	2
16	145.30944	20	5	38	-0.1546730	0.35147496	0.97231458	2
17	193.23329	14	7	39	0.5202637	1.16842858	1.03879763	2
18	104.00728	14	5	22	-0.1546730	-0.35259934	-0.09141419	3
19	89.02093	1	7	9	0.5202637	-0.60807036	-0.95569382	3
20	125.59358	0	9	12	1.1952004	0.01538033	-0.75624467	3
21	32.27354	16	1	7	-1.5045463	-1.57543813	-1.08865991	1
22	172.02126	18	7	35	0.5202637	0.80682898	0.77286543	2
23	66.51911	11	2	20	-1.1670780	-0.99165687	-0.22438029	1
24	117.01736	4	0	41	-1.8420147	-0.13081766	1.17176372	1