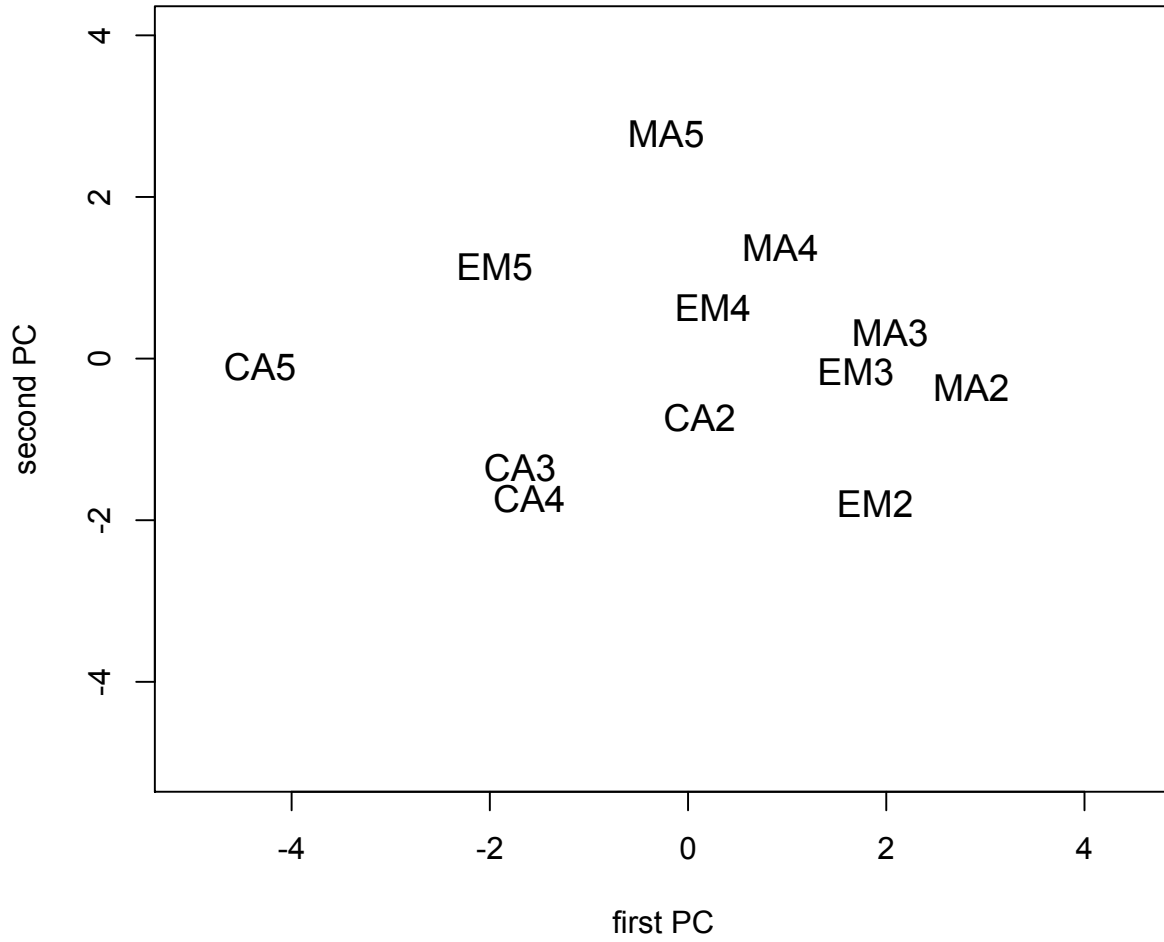


La segunda componente depende de los contenidos gastados en pan, vino y leche.

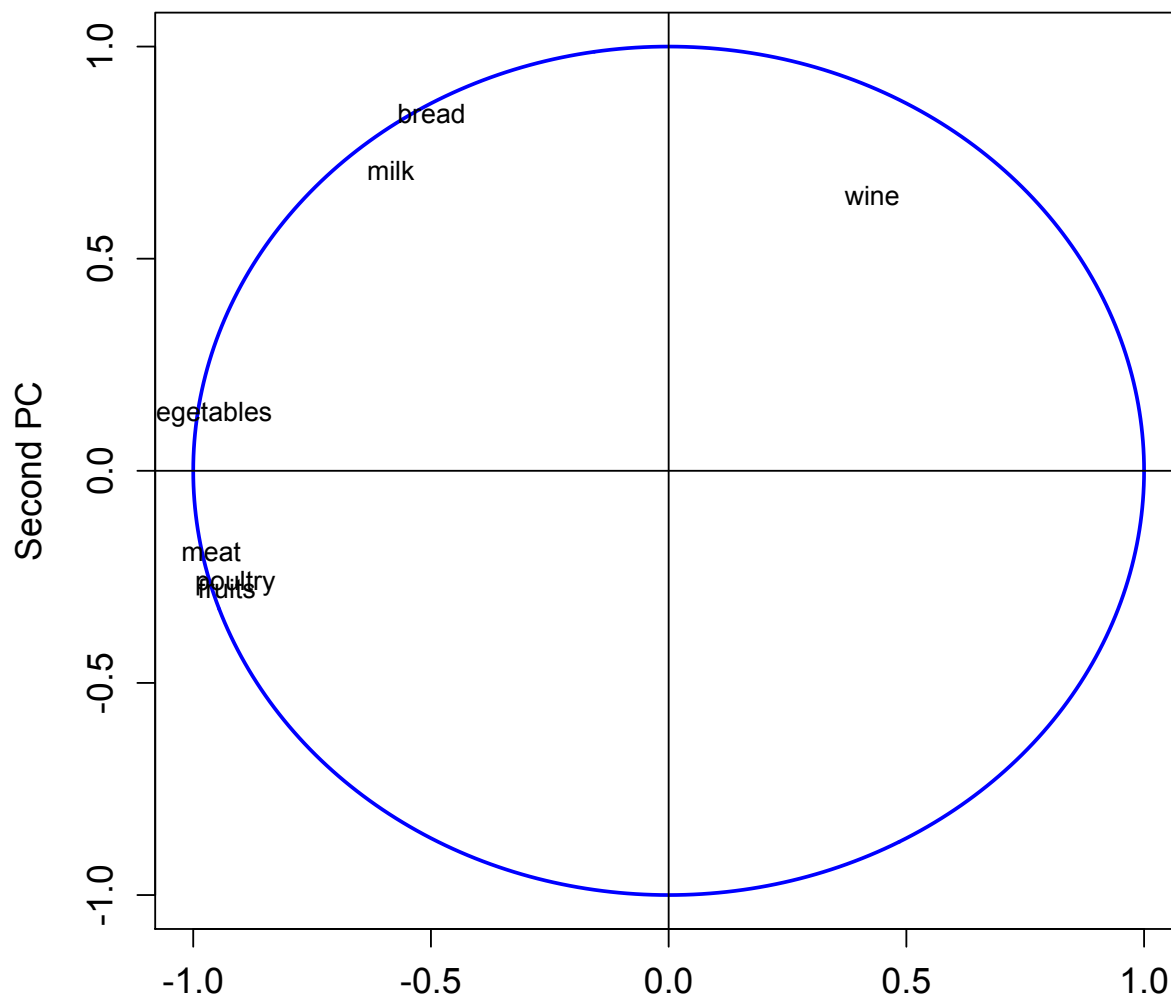
Si se usa la figura correspondiente a las dos primeras componentes principales (scatterplot) así como la información de las correlaciones, se puede inferir que las familias con mayor número de hijos quedan a la izquierda en la figura (los gastos en vegetales, fruta, carne y aves son mayores). Al mismo tiempo, las familias de los gerentes también ejercen mayor gasto en estos rubros y las familias de los trabajadores manuales ejercen menos gasto en esos mismos rubros, pero gastan más en pan, leche y vino (en forma similar para las familias de empleados).

Para llevar a cabo un análisis de agrupamientos, se consideran los datos estandarizados ó equivalentemente se usa  $d_{ij}^2 = (x_{gi} - x_{gj})' W (x_{gi} - x_{gj})$  con  $W = \text{diag} \left( \frac{1}{\widehat{\text{VAR}}(x_1)}, \dots, \frac{1}{\widehat{\text{VAR}}(x_7)} \right)$

# French Food data



## French Food data





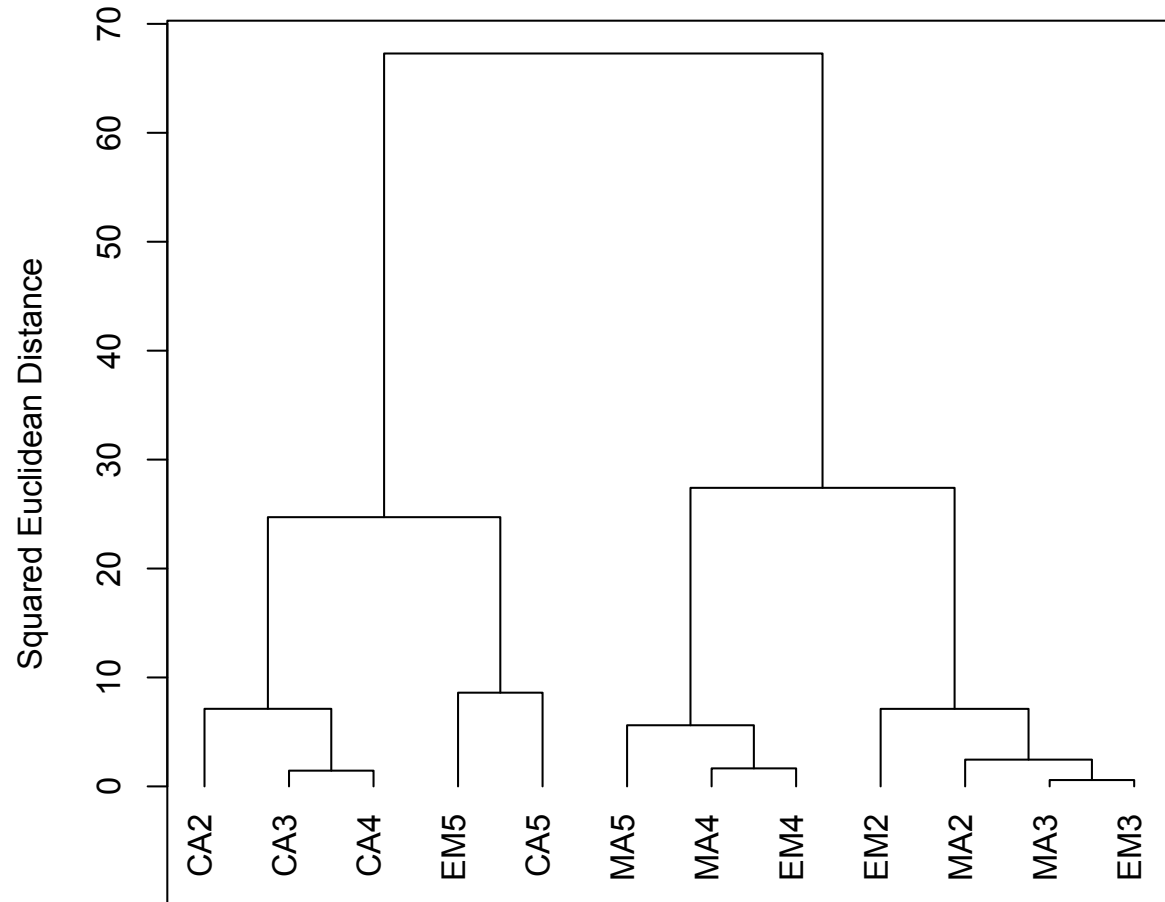
Al calcular el algoritmo jerárquico aglomerativo de Ward se obtiene el dendograma de la figura D1. Si se corta a una distancia de 30 hay dos grupos  $\{CA2, CA3, CA4, EM5, CA5\}$  y  $\{MA5, MA4, EM4, EM2, MA2, MA3, EM3\}$ , el primer grupo sería el de familias cuyo gasto en comida primordialmente se da en frutas, vegetales, carne y aves. En el segundo grupo de familias se gasta más en pan, leche y vino. Si se corta a una distancia de 20, se obtienen 4 grupos  $\{CA2, CA3, CA4\}$ ,  $\{EM2, MA2, EM3, MA3\}$ ,  $\{EM4, MA4, MA5\}$  y  $\{EM5, CA5\}$  los factores que determinan cada grupo son el nivel social (social-profesional) y el tamaño de las familias

### Ejemplo Billetes del Banco Suizo

Se seleccionan al azar 20 billetes de la muestra  $X_1, \dots, X_{200}$  de billetes del banco suizo con la

Figura D1

Ward Dendrogram for French Food



cual ya habíamos trabajado. Al aplicar el algoritmo jerárquico aglomerativo de Ward obtenemos el dendograma en la figura D2.

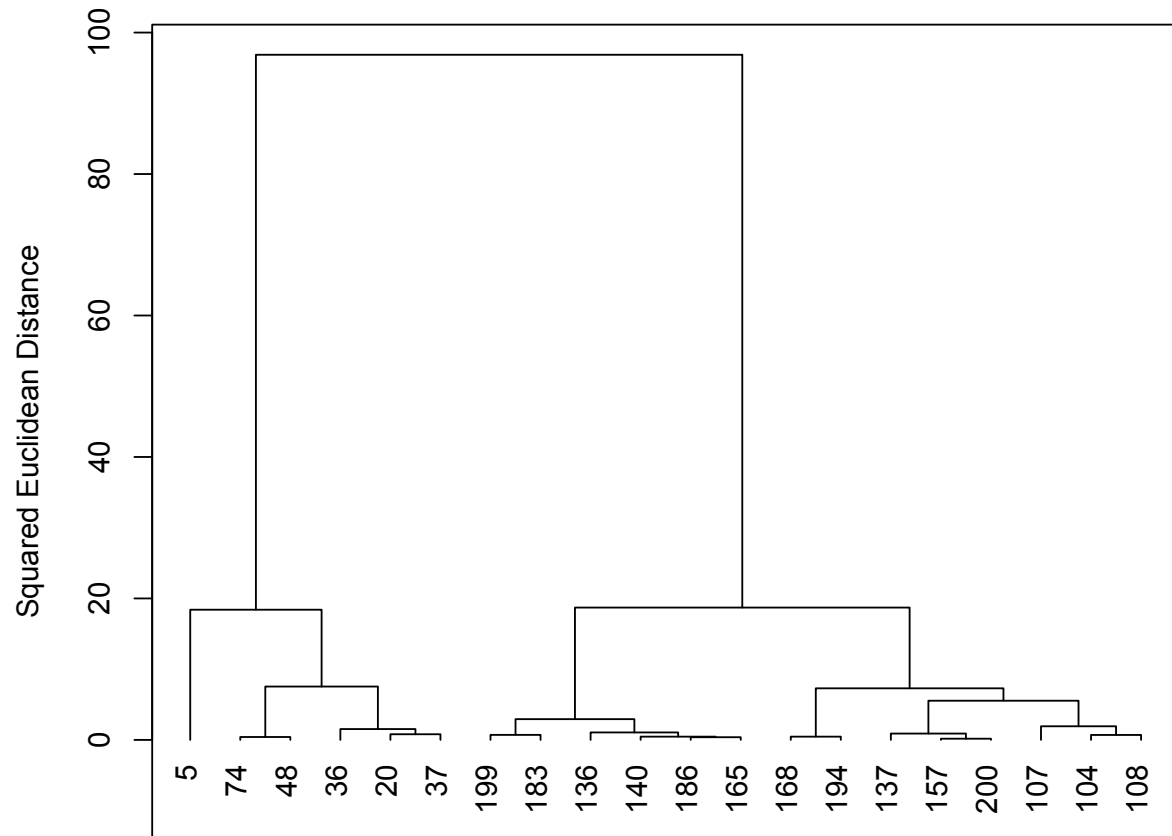
Si se corta en una distancia de 40 se obtienen dos grupos, los billetes en uno de los grupos tienen números entre 1 y 100 y los billetes del otro grupo, números entre 101 y 200.

Recordemos que se tenía información referente a que los billetes numerados 1, ..., 100 son verdaderos mientras que los numerados 101, ..., 200 son falsos.



Figura D2

Dendrogram for 20 Swiss bank notes



Ward algorithm