

Hoja de Trabajo – Clustering

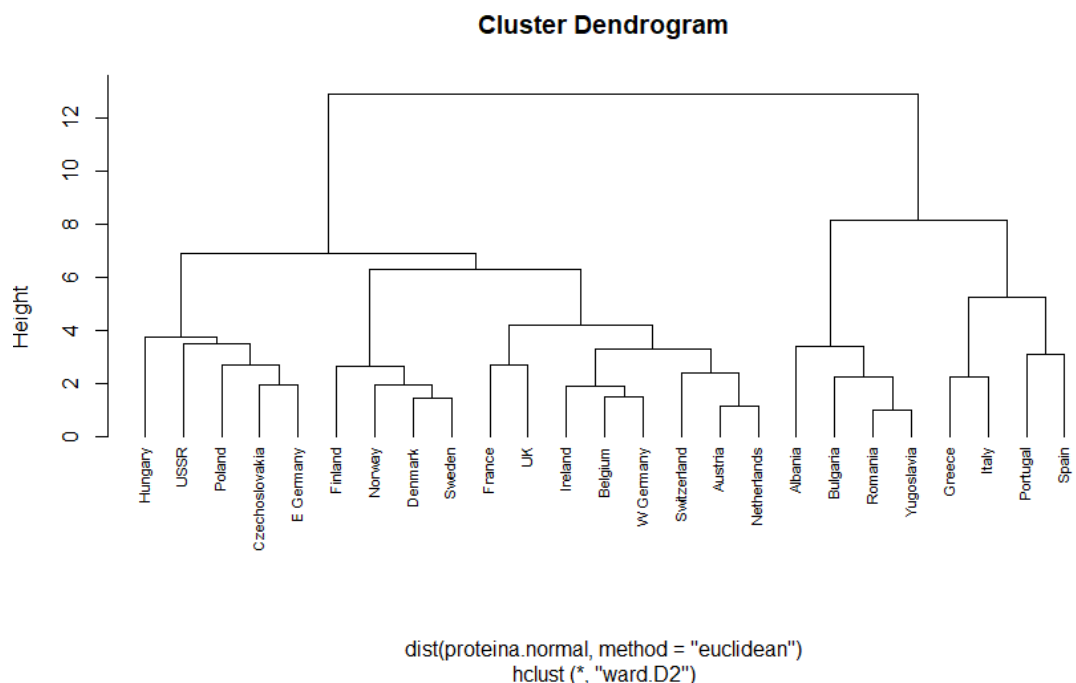
En la siguiente hoja de deberá utilizar las fuentes de datos indicadas para analizar la información usando R

Para cada uno de los ejercicios:

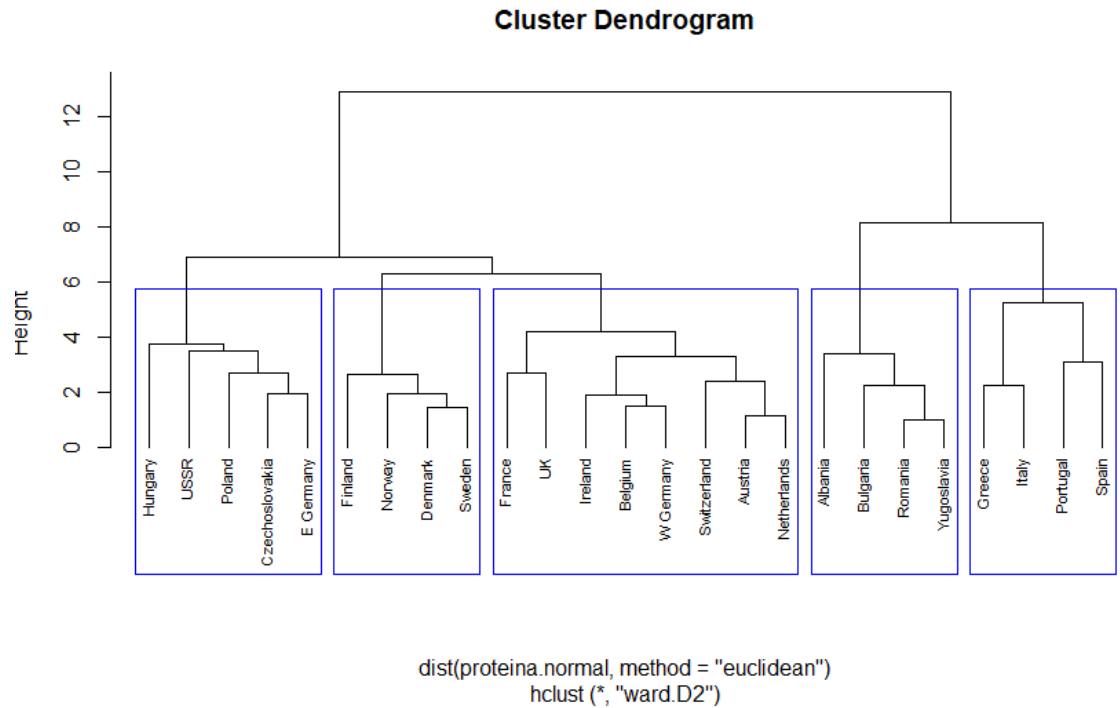
- Suba el archivo .R indicado
- Adjunte los PLOT como imagen en esta hoja de trabajo
- Responda a las preguntas indicadas

1. Ejercicio # 1

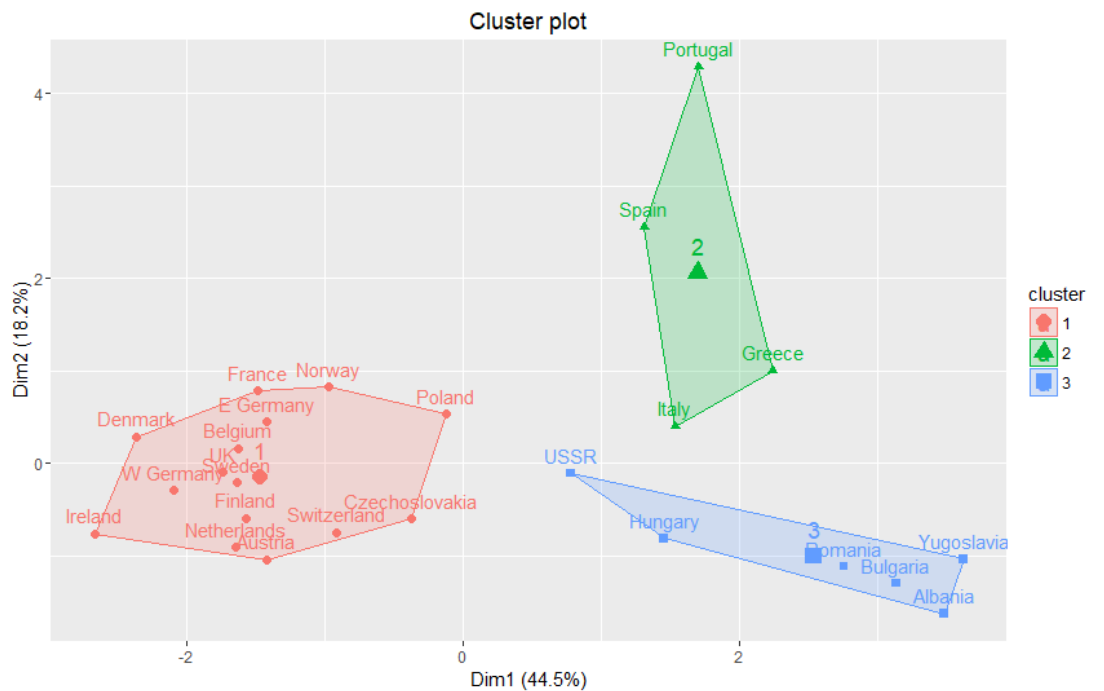
- Cree un nuevo script .R llamado "Clustering"
- Lea el archivo subido al portal llamado "Protein"
 - Este archivo contiene un listado de diferentes observaciones y posee 10 variables que indican el consumo de proteínas en Europa y sus calificaciones según las variables.
- Debe de realizar un script que permita por medio de la técnica de agrupamiento jerárquico (con el método a su elección) y de K-Means realizar los cluster que mejor agrupen la información dada.
- Debe de crear y adjuntar los siguientes plots:
 - Un plot con el dendrograma del modelo jerárquico



- ii. Un plot con los diferentes grupos clusters definidos (cuttree)



- iii. Un plot utilizando factoextra para representar los k-means



e. Debe de responder lo siguiente:

i. ¿Cuál es el mejor número de clusters? ¿Por qué?

3 clústeres, ya que denota una mayor diferencia entre los datos disponibles, separando claramente áreas geográficas con centroides separados.

ii. ¿Cuáles serían las etiquetas para cada cluster?

El algoritmo parece haber detectado las tendencias regionales, por lo tanto, cada etiqueta representa partes de Europa:

- Europa Central (Rojo)
- Europa Oriental (Verde)
- Europa Occidental (Azul)



2. Ejercicio 2

- Comente dentro del script de "Clustering" la elaboración de este ejercicio.
- Obtenga un resumen de las principales variables estadísticas del dataset.
- Calcule la media, moda y desviación estandar de una variable a su elección.