

**Forma de trabajar:** Incluir para cada pregunta los siguiente:

1. El enunciado de la pregunta.
  2. Los comandos de R utilizados para responder a la pregunta.
  3. Los resultados mostrados por R como respuesta a cada comando.
  4. Copia, en su caso, de las gráficas generadas por R.
- 

## 1. Preparación de datos

- (a) Cargar en R los datos del archivo *Adult-data.csv* en el frame *Frame0*. Los datos proceden de la página <https://archive.ics.uci.edu/ml/datasets.php> (UCI Machine Learning Repository).
- (b) Extraer en el subframe *SubFrame0* las variables '*Edad*', '*Educacion-num años*', '*Raza*' y '*Nivel de ingresos*'.
- (c) Determinar si en *SubFrame0* existen campos no definidos (con contenido ?).
- (d) Obtener el *SubFrame1* eliminando los registros de *SubFrame0* con algún campo no definido.
- (e) Obtener los rangos de variación de las variables '*Edad*' y '*Educacion-num años*'.
- (f) Obtener los valores que pueden tomar las variables '*Raza*' y '*Nivel de ingresos*'.
- (g) Obtener la tabla de frecuencias de la variable '*Raza*'.
- (h) Queremos hacer un clustering de las variables '*Edad*', '*Educacion-num años*' y '*Nivel de ingresos*', pero no queremos perder la información de la variable '*Raza*'. Obtener el *SubFrame2* redefiniendo numéricamente la variable '*Nivel de ingresos*' para que sea representativa en el clustering y la variable '*Raza*' para que apenas influya en el clustering.

## 2. Clustering y análisis de los resultados

- (a) Calcular mediante el criterio elbow el valor adecuado del número de clusters  $k$ .
- (b) Obtener con dicho valor de  $k$  el clustering correspondiente.
- (c) Obtener las gráficas bidimensionales del clustering '*Edad*'-'*Educacion-num años*', '*Edad*'-'*Nivel de ingresos*' y '*Educacion-num años*'-'*Nivel de ingresos*'.
- (d) A la vista de las gráficas, describir las propiedades características de cada uno de los clusters obtenidos.
- (e) Calcular para cada uno de los clusters la tabla de frecuencia de la variable '*Raza*'.
- (f) Comparar las tablas de frecuencia anteriores con la tabla de frecuencias de la variable '*Raza*' obtenida en la sección anterior.