

SOC4001 Procesamiento avanzado de bases de datos en R

Tarea 2

Ponderación: 12% de la nota final del curso

Entrega: Desde el momento de entrega, los estudiantes tienen hasta el día domingo 18 de Octubre a las 13:59pm para completar esta tarea y subirla a Classroom.

Formato: Entregar esta tarea en un archivo comprimido (zip) que contenga la carpeta **workflow** y todas las subcarpetas y archivos.

Instrucciones:

- 1) Descargar los datos usados en la Tarea 1 (T1) desde:

https://raw.githubusercontent.com/mebucca/dar_soc4001/master/slides/class_4/Duncan.csv

(botón derecho sobre los datos y luego, “guardar como”)

- 2) Descargar y descomprimir el template de “workflow” visto en clase desde:

https://github.com/mebucca/dar_soc4001/blob/master/slides/class_4/workflow.zip

- 3) Leer archivo **README.md**.

- 4) Mover **Duncan.csv** a la carpeta **data**.

- 5) En **1_masterfile.R** establecer directorio correspondiente a la ubicación de los archivos en tu computador.

- 6) En **2_exploration.R** implementa las siguientes instrucciones de la tarea 1:

- 6.1 (T1-4) Muestra las primeras y las últimas 6 observaciones de la base de datos en la consola.

- 7) En **3_recoding.R** implementa las siguientes instrucciones de la tarea 1:

- 7.1 (T1-5) Crea una base de datos que contenga sólo las variables “type” and “prestige”. Llama tal objeto “subdatos_duncan”.
- 7.2 (T1-8) Crea una nueva variable llamada “educacion_agg”. Asigna valor 1 a “educacion_agg” para aquellas observaciones en las cuales la variable “education” toma un valor por sobre la media. Asigna valor 0 a “educacion_agg” para aquellas observaciones en las cuales la variable “education” toma un valor igual o menor a la media.

- 8) En **4_analyses.R** implementa las siguientes instrucciones:

- 8.1 Usa un loop para calcular la media y la desviación estándar de la variable “prestige” para las observaciones en cada uno de los niveles de la variable “educacion_agg”. Almacena los resultados en una matriz, donde las filas correspondan a cada nivel de “educacion_agg”. En la primera columna almacena la media y en la segunda la desviación estándar. Usa nombres informativos para filas y columnas.
- 8.2 Usa la función `write.table()` para almacenar la tabla en la carpeta **results**. Guarda esta tabla con el nombre “mi_table.txt”. Documentación del comando en:

<https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/write.table>

- 8.3 Crea un scatterplot de las variables “prestige” and “education” y almacénalo en la carpeta **results** con el nombre “mi_grafico.jpeg”.
- 9) Ejecuta `1_masterfile.R` y asegúrate que la totalidad del código corra en “un click”. Comprueba que los archivos con resultados hayan sido creados exitosamente.