

Taller de R: Estadística y Programación

Problem set 4

2024-04-20

En este taller se evalúan los temas vistos en las clases 11 y 13 del curso. Lea atentamente las instrucciones del taller.

Instrucciones

- Este taller representa el **25%** de la nota total del curso y puede ser realizado de manera individual o en grupos de hasta 3 personas. En las primeras líneas del script, escriba su nombre, código y la versión de R que está utilizando. Además, al inicio del código, debe incluir las librerías que utilizará en la sesión, por ejemplo: `pacman`, `rio`, `data.table`, `tidyverse`, `sf`, y `rvest`.
- Asegúrese de descargar las bases de datos desde el repositorio <https://github.com/taller-r-202403/problem-sets> y de crear un nuevo repositorio en su cuenta de GitHub. Si trabaja en grupo, solo un integrante debe crear el repositorio y compartir el acceso con los demás. El repositorio debe ser público para permitir el acceso desde cualquier cuenta de GitHub. Incluya al menos tres carpetas en el repositorio: `input` (datos originales), `output` (datos procesados), y `code` (script con la respuesta del taller).
- Todos los integrantes del grupo deben publicar el enlace al repositorio de GitHub en la actividad **Problem-set-3** del Bloque Neón antes de las 23:59 horas del 26 de abril de 2024.
- Por favor, organice su trabajo cuidadosamente y comente paso a paso cada línea de código. Recuerde **NO** usar acentos ni caracteres especiales dentro del código para evitar problemas al abrir los scripts en diferentes sistemas operativos.
- No seguir estas instrucciones resultará en una penalización del **20%** en la nota final.

Solucionar:

Se debe emplear un bucle o una función para importar los archivos de la Gran Encuesta Integrada de Hogares que se encuentran en la carpeta `input`. Además, si es necesario, se puede acceder al [diccionario](#) o descargar el documento [técnico](#) de los datos.

1. Bucle (35%)

- **1.1 Lista de archivos input**

Cree un objeto que almacene el vector de nombres de los archivos dentro de la carpeta `input`. Asegúrese de que cada archivo contenga la ruta con la ubicación de cada archivo.

Hint: Para este punto, puede usar la función `list.files()` con el argumento `recursive = T`.

- **1.2 Importar archivos:**

Usa el objeto creado en el punto anterior como insumo de una función que permita importar los archivos de **Fuerza de trabajo**, **No ocupados** y **Ocupados** para todos los meses.

Hint: Para este punto, puedes crear una función que importe un archivo y combinarla con la función `lapply`. O puedes crear un bucle y almacenar los dataframes en un objeto tipo lista.

- **1.3 Combinar conjuntos de datos**

Combina todos los `data.frame` que importaste en el punto anterior tres `data.frame`.

Hint: Para este punto, puede usar la función `rbindlist` de la librería `data.table`.

2. Preparación (35%)

- **2.1 Creación d bases de datos**

Cree tres bases de datos diferentes:

1. Usando la base de datos llamada **fuerza de trabajo**, suma el número de individuos que hacen parte de la fuerza laboral (**ft == 1**) y aquellos que hacen parte de la población en edad de trabajar (**pet == 1**) por mes, asegurándote de tener en cuenta el factor de expansión.
2. Usando la base de datos llamada **Ocupados**, suma el número de individuos que se encuentren empleados (**ft == 1**) por mes, asegurándote de tener en cuenta el factor de expansión.
3. Usando la base de datos llamada **No ocupados**, suma el número de individuos desempleados (**dsi == 1**) por mes, asegurándote de tener en cuenta el factor de expansión.”

- **2.2 Colapsar datos a nivel mensual**

Unifica todas las bases de datos creadas en el punto anterior en una única base llamada **Output**, que debe contener al menos cinco columnas: **Población en edad de trabajar**, **fuerza laboral**, **ocupados**, **desempleados** y el **mes** correspondiente.

- **2.3 Tasas de desempleo y ocupación.**

Divida el número de individuos **desempleados** por la **fuerza laboral** para obtener la tasa de desempleo, y los **ocupados** por la **población en edad de trabajar** para obtener la tasa de ocupación.

3. GGplot2 (30%)

Grafique las tasas de desempleo y ocupación para cada mes utilizando la función `geom_line`. (**Hint:** Realice un pivot wider a las tasas de modo que los valores estén en una sola columna)