Taller de R: Estadística y Programación

2024-02-20

En este problem set se evalúan los temas vistos en las clases 4 a 8 del curso. Lea atentamente las instrucciones.

Instrucciones

- Este taller pesa el **25%** de la nota total del curso y podrá responderlo de manera individual o en grupo de hasta 3 personas.
- Debe crear un script en el que almacene las respuestas del problem-set. Asigné su código al nombre del archivo. Por ejemplo `201725852.R`.
- En las primeras líneas del script debe escribir su nombre, código y la versión de R sobre la que está trabajando. Además, al inicio del código debe llamar/instalar las librerías que va a usar en la sesión. Por ejemplo: `pacman`, `rio`, `data.table` y `tidyverse` (a lo menos).
- Asegúrate de descargar las bases de datos del repositorio https://github.com/taller-r-202401/problem-sets y crear un nuevo repositorio en su cuenta de GitHub. Si va a trabajar en grupo, solo 1 de los integrantes del grupo debe crear el repositorio y compartir el acceso a los demás integrantes. El repositorio debe ser público para que se pueda acceder desde cualquier cuenta de GitHub. Adicionalmente, este repositorio debe incluir a lo menos tres carpetas: `input` (datos originales), `output` (datos procesados) y `code` (script con la respuesta del taller).
- Por favor sea lo más organizado posible y comente paso a paso cada línea de código, pero recuerden **NO** usar ningún acento o carácter especial dentro del código para evitar problemas al abrir los scripts en los diferentes sistemas operativos.
- **Cada integrante** del grupo deberá colgar el link al repositorio de GitHub que contiene del problem-set en la actividad **problem-set-2** de Bloque Neón antes de las 23:59 horas del **20 de marzo de 2024**.
- No seguir las instrucciones tiene una penalización del **20%** de la nota final.

Problem-set

Los siguientes puntos se realizarán utilizando la Encuesta de Micronegocios 2022, que se centra en empresas con un máximo de 9 empleados. Puedes encontrar el diccionario de datos en el portal del DANE o descargarlo en formato PDF.

1. Importar/exportar bases de datos

- 1.1 Importar

Importe las bases de datos **Módulo de sitio o ubicación** en un objeto llamdo `location` y **Módulo de identificación** en un objeto llamado `identification`.

- 1.2 Exportar

Exporte a la carpeta output los objetos cargados en el punto anterior, guárdelos como **location.rds** y **identification.rds**.

2. Generar variables

- 2.1 Usando la variable **grupos 4**, se debe generar una nueva variable llamada `bussiness_type`, que tomará los siguientes valores:
  - **Agricultura** cuando `grupos 4` sea igual a 1.
  - **Industria manufacturera** cuando `grupos 4` sea igual a 2.
  - **Comercio** cuando `grupos 4` sea igual a 3.
  - **Servicios** cuando `grupos 4` sea igual a 4.
- 2.2 Se debe crear una variable llamada `grupo_etario` que divida a los propietarios de micronegocios en cuatro grupos etarios. Los rangos de edades seleccionados deben ser justificados.
- 2.3 Sobre el objeto `location`, genere una variable llamada `ambulante`, que sera igual a 1 si la variable `P3053` es igual a 3, 4 o 5.

3. Eliminar filas/columnas de un conjunto de datos

- 3.1 Almacene en un objeto llamado `identification_sub` las variables `DIRECTORIO`, `SECUENCIA_P`, `SECUENCIA_ENCUESTA`, `grupo_etario`, `ambulante`, `COD_DEPTO` y `F_EXP`.
- 3.2 Del objeto `location` seleccione solo las variables `DIRECTORIO`, `SECUENCIA_P`, `SECUENCIA_ENCUESTA`, `ambulante P3054`, `P469`, `COD_DEPTO`, `F_EXP` y guárdelo en nuevo objeto llamado `location_sub`.

4. Combinar bases de datos

- 4.1 Use las variables `DIRECTORIO`, `SECUENCIA_P` y `SECUENCIA_ENCUESTA` para unir en una única base de datos, los objetos `location_sub` y `identification_sub`.

5. Descriptivas

- 5.1 Utilizando funciones como `skim` o `summary`, cree breves estadísticas descriptivas de la base de datos creada previamente. `(HINT: Observaciones en NA, conteo de variables únicas)`
- 5.2. Use las funciones `group_by` y `summarise` para extraer variables descriptivas, como la cantidad de asociados por departamento, grupo etario, entre otros. Además, cree un pequeño párrafo con los hallazgos que encuentre.