

# Taller de R: Estadística y Programación

16/08/2023

En este taller se evalúan los temas vistos en las clases 4 a 7 del curso. Lea atentamente las instrucciones del taller.

## Instrucciones

- Este taller constituye el **20%** de la nota total del curso y puede ser completado de forma individual o en grupos de hasta 3 personas.
- Deberá crear un repositorio en GitHub y, posteriormente, un proyecto en R que esté vinculado a dicho repositorio. Este proyecto debe contener, al menos, tres carpetas: **input** para los datos originales, **output** para los datos procesados y **code** para el script con las respuestas al taller. En las primeras líneas del script, escriba su nombre, código y la versión de R que está utilizando. Además, al comienzo del código, asegúrese de llamar o instalar las bibliotecas que utilizará, como **pacman**, **rio**, **data.table** y **tidyverse**, como mínimo.
- Por favor, sea lo más organizado posible y comente cada línea de código de forma detallada. Sin embargo, recuerde **NO** utilizar acentos ni caracteres especiales en el código para evitar problemas de compatibilidad al abrir los scripts en diferentes sistemas operativos.
- Si opta por trabajar en grupo, asegúrese de que se cree un único repositorio en GitHub para todo el equipo. No obstante, **cada integrante** del grupo deberá subir el enlace al repositorio de GitHub en Bloque Neón antes de las 23:59 horas del **28 de septiembre de 2023**.
- El incumplimiento de estas instrucciones resultará en una penalización del **20%** en la nota final.

## Problem-set

### 1. Importar/exportar bases de datos

Los siguientes puntos se realizarán utilizando la Encuesta de Micronegocios 2021, 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.

#### 2.1 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**.

#### 2.2 Exportar

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

### 3. Generar variables

- 3.1. Sobre el objeto `identification`, genere una variable que será igual a: **Agricultura** cuando sea **igual a 1**, **Industria manufacturera** cuando sea **igual a 2**, **comercio** cuando sea **igual a 3** y **servicios** cuando sea **igual a 4**. a esta variable le llamará `bussiness_type`.
- 3.2. Sobre el objeto `location`, genere una variable llamada `local`, que sea igual a 1 si la variable P3053 es igual a 6 o 7.

### 4. Eliminar filas/columnas de un conjunto de datos

- 4.1 Almacene en un objeto llamado `identification_sub` el subconjunto de observaciones del objeto `identification` que pertenezcan a la industria manufacturera.
- 4.2 Del objeto `location` seleccione solo las variables `DIRECTORIO`, `SECUENCIA_P`, `SECUENCIA_ENCUESTA`, `P3054`, `P469`, `COD_DEPTO`, `F_EXP` y guardelo en nuevo objeto llamado `location_sub`.
- 4.3 Del objeto `identification_sub`, seleccione las variables `DIRECTORIO`, `SECUENCIA_P`, `SECUENCIA_ENCUESTA`, `P35`, `P241`, `P3032_1`, `P3032_2`, `P3032_3`, `P3033` y `P3034` y sobre escriba el objeto `identification_sub`.

### 5. Combinar bases de datos

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