

Taller de R: Estadística y Programación

01/09/2022

En este taller se evalúan los temas vistos en las clases 1 a 6 del curso. Lea atentamente las instrucciones del taller.

Instrucciones

- Este taller pesa el **30%** de la nota total del curso y podrá responderlo de manera individual o en grupo de hasta 3 personas. En las primeras líneas del script debe escribir su nombre, código y la versión de R sobre la que está trabajando.
- Recuerde hacer fork sobre el repositorio <https://github.com/taller-r-202202/pset-1> para que cree una versión del repositorio en su cuenta de GitHub. Si va a trabajar en grupo, solo 1 integrante del grupo debe hacer fork del repositorio y compartir el acceso a los demás integrantes. Este repositorio debe incluir a lo menos tres carpetas: **input** (datos originales), **output** (datos procesados) y **code** (script con la respuesta del taller).
- Deberá colgar en Bloque Neón el link al repositorio de GitHub antes de las 23:59 horas del 18 de septiembre de 2022.
- 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.
- No seguir las instrucciones tiene una penalización del **20%** de la nota final.

Problem set 1

1. Vectores

Cree un vector que contenga los números del 1 al 100, posteriormente cree otro vector que contenga los números impares de 1 a 99. Use el vector de números impares para crear un vector que contenga los números pares del primer vector.

2. Importar/exportar bases de datos

2.1 Importar

Importe las bases de datos **Enero - Cabecera - Características generales (Personas).csv** y **Enero - Cabecera - Ocupados.csv**.

2.2 Exportar

Exporte a la carpeta output los objetos cargados en el punto anterior, guárdelos como **Características generales (Personas).rds** **Ocupados.rds**.

3. Generar variables

- Sobre el objeto que contiene la base `Ocupados`, genere una variable `ocupado` que sea igual a 1.
- Sobre el objeto que contiene las `Características generales`, genere una variable `joven` que sea igual a 1 si la persona tiene entre 18 y 28 años de edad.

4. Eliminar filas/columnas de un conjunto de datos

- Del objeto que contiene las `Características generales`, deje únicamente las observaciones para las personas entre 18 y 70 años de edad.
- Del objeto que contiene las `Características generales`, seleccione las variables `secuencia_p`, `orden`, `hogar`, `directorio`, `P6020`, `P6040`, `P6920`, `DPT0`, `fex_c_2011`, `ESC` y `MES`
- Del objeto que contiene la base `Ocupados`, seleccione las variables `secuencia_p`, `orden`, `hogar`, `directorio`, `ocupado`, `INGLAB0` y `P6050`

5. Combinar bases de datos

Use las variables `secuencia_p`, `orden`, `hogar` y `directorio` para unir en una única base de datos, los objetos del punto anterior.

6. Descriptivas de un conjunto de datos: ¿Cómo varió el número de ocupados del mes de Enero al mes de Febrero?

Use las funciones `ggplot()`, `group_by()` y `summarize()` entre otras, para generar algunas estadísticas descriptivas (gráficos y tablas) numero de ocupados e ingresos laborales promedio. Tenga en cuenta algunas dimensiones como departamento, sexo y edad. Las tablas las puede plotear sobre la consola, pero los gráficos los debe exportar en formato `.jpeg` a la carpeta *output*. Debe generar mínimo 3 gráficos y 3 tablas.