

# 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. Además, al inicio del código debe llamar/installar las librerías que va a usar en la sesión. Por ejemplo: `pacman`, `rio` y `tidyverse` (a lo menos).
- Recuerde hacer fork sobre el repositorio <https://github.com/taller-r-202301/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

Los siguientes puntos se llevarán a cabo con la Encuesta de Micronegocios- 2021, la cual se enfoca en las empresas que desarrollan una actividad económica con un máximo de 9 personas ocupadas. El diccionario de datos lo encontrará virtual en el portal del DANE o lo podrá descargar en formato pdf.

## 2.1 Importar

Importe las bases de datos **Módulo de sitio o ubicación** y **Módulo de identificación**.

## 2.2 Exportar

Exporte a la carpeta output los objetos cargados en el punto anterior, guárdelos como **Modulo\_ubicacion.rds** y **Modulo\_identificacion.rds**.

## 3. Generar variables

- Sobre el objeto que contiene la base **Identificación**, genere una variable que sera 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 llamara **bussiness\_type**.
- Sobre el objeto que contiene la **Ubicación**, 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

- Del objeto que contiene las **Identificación**, deje únicamente las observaciones de la industria manufacturera.
- Del objeto que contiene las **ubicación**, seleccione las variables **DIRECTORIO**, **SECUENCIA\_P**, **SECUENCIA\_ENCUESTA**, **P3054**, **P469**, **COD\_DEPTO**, **F\_EXP**.
- Del objeto que contiene la base **Identificación**, seleccione las variables **DIRECTORIO**, **SECUENCIA\_P**, **SECUENCIA\_ENCUESTA**, **P35**, **P241**, **P3032\_1**, **P3032\_2**, **P3032\_3**, **P3033** y **P3034**.

## 5. Combinar bases de datos

Use las variables **DIRECTORIO**, **SECUENCIA\_P** y **SECUENCIA\_ENCUESTA** 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). 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.