

# Ejercicio 2

Importar datos y funciones de `dplyr` y `tidyr`

Unidad de Estadística      Servicio de Gestión Estratégica  
Departamento de Desarrollo Sostenible e Inteligente

18 Noviembre, 2023

## Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Importar datos: Encuesta Continua de Hogares</b>	<b>2</b>
<b>3. Crear nuevas variables</b>	<b>2</b>
3.1. Resumen de datos . . . . .	2
3.2. Uso de lista y filtrado . . . . .	2
<b>4. Uso de joins y funciones de <code>tidyr</code></b>	<b>2</b>

## 1. Introducción

En este ejercicio aplicaremos algunas de las cosas aprendidas en el Módulo 3. Trabajaremos con los datos de la **ECH**<sup>1</sup> que están dentro de la carpeta **Bases**. La idea es que importen los datos, realicen algunos procesamiento utilizando las funciones de la librería **dplyr** y **tidyr**. También vamos a guardar algunos objetos generados en el ejercicio en formato **RData**

- Recordar cargar los paquetes o librerías, utilizando la función `library()`
- Solucionar los conflictos resultantes

## 2. Importar datos: Encuesta Continua de Hogares

Importar los datos de la ECH **baseHog.csv** que están dentro de la carpeta **Bases**. Cargar la base en el objeto **dat**. Se puede usar cualquier función para importar datos vista en clase.

Sobre los datos, chequear además:

- Que tengan la dimensión adecuada
- Que las variables tengan la clase que corresponde

## 3. Crear nuevas variables

Al objeto **dat** que contiene los datos le vamos a agregar las siguientes variables:

- **ypc\_svl**: Ingreso del hogar sin valor locativo per cápita.  $(ht11 - ht13)/ht19$
- **ind\_VmedAdec**: Variable que indique si la vivienda es **Inadecuada** o **Medianamente Adecuada**
- **ht13\_rec**: Recodificar la variable valor locativo en 3 categorías (**Menos de 10000**, **Entre 10000 y 20000** y **Más de 20000**).
- **ht11\_cent**: Restarle el promedio a la variable **ht11**.
- **ht11\_centMun**: Hacer lo mismo que en la variable anterior pero dentro de cada municipio, es decir, restarle el promedio a la variable **ht11**.

### 3.1. Resumen de datos

Resumir la información de las variables creadas utilizando las funciones **group\_by()** y la función de resumen **summarise()**

### 3.2. Uso de lista y filtrado

Crear una lista llamada **mi\_lista** y guardar:

- En la primer componente los resultados **res01**
- En la segunda componente guardar los resultados de **res01** solo para los municipios (A,C,D)
- En la tercer componente guardar los resultados de **res01** solo para los municipios (A,C,D) y las columnas **mujer\_por** y **ht11**.

## 4. Uso de joins y funciones de tidyr

- Utilizar la función **left\_join** para juntar datos de hogares de la ECH con datos de *personas* (**basePer.csv**).
- Elegir solamente las columnas “correlativ”, “desocupado”, “ocupados”, “subempleo”, “pobpcoac” y “TotEduc” de personas y considerar los mismos **NAs** que para la base de hogares.

---

<sup>1</sup>Más información sobre la ECH en <https://cfervirtual.montevideo.gub.uy/moodle/mod/folder/view.php?id=73598>

- Juntar datos de personas ECH:
  - Levantar los datos personas ECH
  - Elegir solamente las columnas “correlativ”, “desocupado”, “ocupados”, “subempleo”, “pobpcoac” y “TotEduc” de personas
- Mezclar datos personas y hogares considerando los mismos NAs que para hogares.

La variable **pobpcoac** de la base de hogares cuantifica la *condición de actividad económica* de una persona. Presentar en formato de tabla la distribución de personas inactivas que realizan quehaceres en el hogar (pobpcoac==6), en cuyas filas figuren los municipios, en las columnas figure el sexo y en cada una de las celdas de la tabla se muestre el promedio de edad según sexo, para cada municipio de Montevideo.

Finalmente, guardar los datos obtenidos en archivo `misRes.RData`, dentro de una carpeta nueva llamada **Resultados**.