

## Desafío - Gráficos

- Para realizar este desafío debes haber estudiado previamente todo el material disponibilizado correspondiente a la unidad.
- Una vez terminado el desafío, comprime la carpeta que contiene el desarrollo de los requerimientos solicitados y sube el `.zip` en el LMS.
- Desarrollo desafío:
  - El desafío se debe desarrollar de manera Individual.

### 1. Genere una submuestra de casos

- Utilice los últimos 4 dígitos de su rut como semilla pseudoaleatoria.
- Seleccione el 50% de los casos.
- Cada base generada debe contener los siguientes elementos:
  - El índice de desarrollo humano (`undp_hdi`).
  - El nombre del país (`ccodealp`).
  - La región a la que pertenece (`ht_region`).
  - El PIB per capita. (`gle_cgdpc`).
  - El total de la población (`imf_pop`).
- Si su apellido está entre la A y la M, escoja las siguientes variables del módulo Educación:
  - `ffp_hf`: Human Flight and Brain Drain.
  - `wef_qes`: Quality of the educational system.
  - `wdi_expedu`: Government expenditure on education, total (% of GDP).
  - `wdi_ners`: School enrollment, secondary (% net).
- Si su apellido está entre la N y la Z, escoja las siguientes variables del módulo Salud:
  - `wef_imort`: Infant mortality, deaths/1000 live births.
  - `who_alc2000`: Alcohol consumption per capita (2000-).
  - `who_tobt`: Current smoking of any tobacco product (Total).
  - `wdi_exph`: Government expenditure on health, total (% of GDP).
  - Guarde esta tabla procesada en un nuevo objeto.
  - Renombre las categorías de la variable `ht_region` de números a regiones.

## 2. Genere una función que ingrese su objeto y devuelva:

- Por cada variable existente en su objeto, calcule las medidas descriptivas para los casos continuos.
- Para cada variable discreta, que calcule la frecuencia.
- Reporte las estadísticas descriptivas para `gle_cgdpc`, `undp_hdi`, `imf_pop`. Compare las estadísticas con algún compañero. ¿Ve alguna diferencia sustancial en alguna de ellas?

## 3. Genere una función que liste las observaciones perdidas de una variable

- La función debe contener los siguientes argumentos:
  - `dataframe`: La función debe ingresar un objeto DataFrame.
  - `var`: Variable a inspeccionar.
  - `print_list`: Opción para imprimir la lista de observaciones perdidas en la variable. Debe ser False por defecto.
- La función debe retornar la cantidad de casos perdidos y el porcentaje correspondiente.
- Cuando `print_list = True`, debe retornar la lista de casos.
- Analice todas las variables y sus casos perdidos.  
Para las 3 variables con un mayor porcentaje de casos perdidos, solicite la lista de países con ausencia de datos.

## 4. Grafique los histogramas

- Genere una función que grafique un histograma en conjunto y señale las medias.
- La función debe incluir los siguientes argumentos:
  - `dataframe`: La base de datos donde se encuentran los datos específicos.
  - `var`: La variable a graficar.
  - `sample_mean`: Booleano. Si es verdadero, debe generar una recta vertical indicando la media de la variable en la selección muestral. Por defecto debe ser `False`.
  - `true_mean`: Booleano. Si es verdadero, debe generar una recta vertical indicando la media de variable en la base de datos completa.
- Implemente las funciones para las 4 variables seleccionadas según su grupo.  
¿En qué variables la media de la submuestra es mayor a la de la muestra completa?

## 5. Genere una función que devuelva un dotplot

- La función debe contener los siguientes argumentos como argumentos:
  - `dataframe`: La tabla de datos donde buscar las variables.
  - `plot_var`: La variable a analizar y extraer las medias.
  - `plot_by`: La variable agrupadora.
  - `global_stat`: Booleano. Si es `True` debe graficar la media global de la variable. Por defecto debe ser `False`.
  - `statistic`: Debe presentar dos opciones. `mean` para la media y `median` para la mediana. Por defecto debe ser `mean`.
- Implemente la función en las 3 variables con una menor cantidad de datos perdidos.

## 6. Guarde la base de datos

- La submuestra creada tiene un método llamado `to_csv`.  
Acceda a este y guarde la base de datos con la siguiente nomenclatura:  
`subsample_<iniciales>_demo.csv`.  
(Subela a la plataforma, junto al desafío)