# Tarea 2 de Introducción Análisis de Datos

Estudiantes:

Sofia Elena Arias Juárez

Alisson Steller Alfaro

Luis Diego Chavarría Chacón

Andrés González Romero

## Paso 1: Lectura del archivo de datos CSV

Después de instalar RStudio, procedemos a configurar el directorio de trabajo utilizando la siguiente instrucción:

A screenshot of a computer

Description automatically generated

Este es directorio donde leeremos el CSV.

A screenshot of a computer

Description automatically generated

Ahora en RStudio, empleamos la siguiente función para leer el archivo de datos, asegurándonos de especificar si el archivo tiene nombres en cada columna utilizando el parámetro **header = TRUE**, y **sep** para definir la separación decimal. Finalmente, cargamos los datos en una variable:

A screen shot of a computer code

Description automatically generated

## Paso 2: Impresión del archivo de datos

Para mostrar los datos, se utiliza la función **View** para su impresión.

A screenshot of a video game

Description automatically generated

A screenshot of a computer

Description automatically generated

## Paso 3: Creación de gráficos

Del link proporcionado, seleccionamos dos gráficos: el histograma y de barras.

Para el histograma se utiliza la siguiente instrucción:

A black background with white text

Description automatically generated

**hist():** Indica la creación de un histograma en R.

**datos$Height:** Selecciona como datos la altura de las personas y serán utilizados en la creación del histograma.

**main:** Indica el título del histograma.

**xlab:** Indica el título del eje x.

**ylab:** Indica el título del eje y.

**col:** Es el color el cual será pintado el histograma.

Nos daría el siguiente resultado:

**A graph of blue bars

Description automatically generated with medium confidence**

Ahora decidimos demostrar la media genero para el grafico de barras. Por lo que primero se hizo un cálculo previo a la generación de este.

A screen shot of a computer

Description automatically generated

**tapply():** Calcula la media de las alturas de los dos géneros, toma las dos columnas **datos$Weight, datos$Gender**, y la función que se va aplicar, en este caso **mean**.

**barplot():** Crea un gráfico de barras.

**mean\_height:** Es el vector de medias de altura.

**main:** Es el título del gráfico.

**xlab:** Indica el título del eje x.

**ylab:** Indica el título del eje y.

**col:** Son los colores de las barras.

A screenshot of a computer

Description automatically generated