

# Iniciación al análisis de datos con R

David García Callejas

david.garcia.callejas@gmail.com



**Junta de Andalucía**

Consejería de la Presidencia,  
Administración Pública e Interior

Centro de Estudios Andaluces

# Organización

5 sesiones:

- Introducción. Tipos de datos, proyectos en Rstudio, buenas prácticas: 2:30h
- Visualización de datos con ggplot2: 2:30h
- Gestión y tratamiento de datos. Fundamentos de tidyverse: 2:30h
- Programación y funciones: 2:30h
- Análisis estadísticos univariantes: 2:00h



# Organización

Temporización:

- Sesiones de 2-2:30h
- Tarea de aprovechamiento: 1 mes

# Sesión 1: Introducción

## Objetivos:

- Conocer el entorno de trabajo de R y Rstudio
- Instalar paquetes externos
- Trabajar con proyectos autocontenidos en Rstudio
- Repasar los tipos de datos básicos en R: Vectores, matrices, listas, dataframes

## Sesión 1: Introducción

### Ejercicios:

- Crea una serie de matrices cuyos elementos sean números aleatorios entre 0 y 1, almacénalas en una lista y guarda dicha lista en disco.
- Crea una tabla de 1000 filas en el que la primera columna sea una columna numérica que siga una distribución normal con media 0 y desviación típica 1. La segunda columna debe ser una distribución aleatoria de las letras del alfabeto. Pista: usa las funciones “rnorm” y “sample”, y el vector “letters”, que viene por defecto en R

## Sesión 2: Visualización de datos con ggplot2

### Objetivos:

- Entender la sintaxis de ggplot para visualizar diferentes tipos de datos
- Generar gráficos de puntos, barras, cajas, etc.
- Modificar la apariencia y la estética de las figuras generadas.

## Sesión 2: Visualización de datos con ggplot2

### Ejercicios:

- Usando el dataset “iris”:
- Selecciona dos variables y crea una gráfica con su relación.
- Hay al menos dos maneras de diferenciar los puntos relativos a cada especie: generando diferentes formas/colores en un mismo panel, o bien separando cada especie en paneles distintos. Crea una figura con cada manera.
- Utiliza el paquete “patchwork” para generar una sola figura uniendo las dos figuras anteriores.



## Sesión 3: Gestión y tratamiento de datos

### Objetivos:

- Seleccionar, reemplazar, unir y agregar datos en formato dataframe.
- Entender los fundamentos del conjunto de paquetes “tidyverse”

## Sesión 3: Gestión y tratamiento de datos

### Ejercicios:

- Usando los conjuntos de datos “nations” y “nations\_co2”, crea una tabla con la siguiente información por país:
  - último año con datos de PIB (gdp) disponibles
  - PIB total de dicho año ( $\text{gdp per capita} * \text{población total}$ )
  - Emisiones de CO2 totales de dicho año ( $\text{co2 per capita} * \text{población total}$ )

## Sesión 4: Programación y funciones

### Objetivos:

- Familiarizarse con las estructuras de control, bucles, y vectorización.
- Crear e importar funciones

## Sesión 4: Programación y funciones

### Ejercicios:

- Crea una función que muestre por pantalla la media y la desviación típica de un vector numérico. Opcionalmente, también la mediana.
- Si el vector no es numérico, la función debe mostrar por pantalla un mensaje de error escrito por ti.

## Sesión 5: Análisis estadísticos univariantes

### Objetivos:

- Familiarizarse con la distribución de datos gaussiana, y con las medidas descriptivas básicas: media, desviación típica, mediana, intervalos de confianza.
- Familiarizarse con análisis estadísticos básicos en R

## Sesión 5: Análisis estadísticos univariantes

### Ejercicios:

- Usando el conjunto de datos “starwars\_info\_personajes.csv”, muestra si hay diferencias estadísticas entre el peso medio de los humanos y del resto de personajes
- Usando el mismo conjunto de datos, ¿podemos predecir la altura de un personaje en función de su peso? ¿por qué? ¿hay alguna característica peculiar en los datos?



**Junta de Andalucía**