

# Introducción al análisis de datos con R

**Manuel Mejías Leiva**

Universidad de Valladolid

[manuel.mejias@uva.es](mailto:manuel.mejias@uva.es)

5 Junio – 9 Junio 2023 (11:00-13:00)

## **Descripción del curso:**

El objetivo de este curso es que los/as alumnos/as se familiaricen con el funcionamiento de R, RStudio y los fundamentos del análisis y la manipulación de datos. Al finalizarlo, los/as alumnos/as estarán preparados para trabajar con R tanto para los cursos de métodos cuantitativos como para sus propios análisis.

Durante todas las sesiones, se trabajarán con datos de encuestas que abordan temáticas relacionadas con las desigualdades educativas y sociales. Esto permitirá a los/as estudiantes aplicar los conocimientos adquiridos a situaciones reales y relevantes, lo que les ayudará a comprender mejor la complejidad de estos problemas y a desarrollar habilidades prácticas en el manejo de datos.

## **Estructura del curso:**

El curso constará de cuatro bloques principales, seguidos de una sesión final de recapitulación, y se estructurará del siguiente modo:

### **Día 1: Conceptos básicos de R y RStudio**

El primer día aprenderemos qué es R y cómo trabajar con él utilizando y configurando RStudio. Cubriremos los aspectos básicos del lenguaje R, como tipos y estructuras de datos, cómo importar datos y funciones y herramientas básicas para conocer nuestros datos.

### **Día 2: Introducción a la limpieza y manipulación de datos**

Después de familiarizarnos con R, el segundo día se centrará en la limpieza y manipulación de datos, principalmente utilizando los paquetes **dplyr** y **tidyr**. Los temas cubrirán cómo filtrar, seleccionar, resumir y reorganizar variables, así como cambiar la estructura de los datos cuando sea necesario. También aprenderemos a crear nuevas variables y a trabajar con grupos o categorías.

### **Día 3: Introducción a la visualización de datos y flujos de trabajo**

En la primera parte de esta lección, cubriremos los fundamentos de la manipulación de datos y flujos de trabajo y cómo configurar un entorno de trabajo fluido en R que sea reproducible, portable y autónomo. La segunda parte del día 3 consistirá en una introducción a **ggplot2**, un paquete de visualización de datos, para generar nuestros primeros gráficos en R.

### **Día 4: Relaciones entre variables, regresiones y presentación de resultados**

Esta sesión se centrará en cómo visualizar y entender la relación entre dos o más variables utilizando R. Además, se cubrirá una variedad de herramientas, paquetes y directrices que serán útiles para aquellos que deseen familiarizarse con el análisis de regresión en R. Por último, se trabajará con una serie de herramientas para exportar los resultados de las regresiones en forma de gráficos y tablas.

### **Día 5: Caso práctico de síntesis**

El último día del curso se realizará un caso práctico utilizando datos del Programa para la Evaluación Internacional de Alumnos (**PISA** por sus siglas en inglés), donde los/as alumnos/as repasarán y pondrán en práctica los diferentes temas tratados a lo largo del curso.

### **Antes del curso:**

Antes de comenzar el curso, se recomienda que los/as estudiantes instalen R y RStudio en sus ordenadores personales y se familiaricen con el entorno de trabajo. A continuación, se presentan los enlaces a la web oficial para la descarga de R y RStudio:

- Descarga de R: <https://cran.r-project.org/>
- Descarga de RStudio Desktop: <https://rstudio.com/products/rstudio/download/>

Es importante tener en cuenta que R debe ser instalado primero antes de instalar RStudio, ya que RStudio es una interfaz gráfica para interactuar con R.

### **Recursos bibliográficos:**

- "R for Data Science" de Hadley Wickham y Garrett Grolemund: Este libro es una introducción completa al lenguaje de programación R, desde los conceptos básicos hasta la manipulación de datos, visualización y modelado estadístico.
- "An Introduction to R" de Venables, Smith y R Development Core Team: Este libro es una guía completa y práctica para el uso de R en la estadística y la ciencia de datos.
- "Advanced R" de Hadley Wickham: Este libro se centra en temas más avanzados de programación en R, como el manejo de funciones, el uso de paquetes y la programación orientada a objetos.
- "Data Manipulation with R" de Phil Spector: Este libro se centra en la manipulación de datos en R, incluyendo la limpieza, la transformación y el subconjunto de datos.
- "ggplot2: Elegant Graphics for Data Analysis" de Hadley Wickham: Este libro se centra en la creación de gráficos elegantes y efectivos en R utilizando el paquete ggplot2.