



Análisis y visualización de datos con el lenguaje R

Capacitador (es):

- **Dr. Mario Hernández-Guzmán**
Doctor en Ciencias con la especialidad en Biotecnología (Posdoctorante CICESE)
- **Dr. Valentín Pérez Hernández**
Doctor en Ciencias en Alimentos y Biotecnología (Posdoctorante CINVESTAV-IPN)

Resumen del curso:

Este curso brindará una capacitación, de básica e intermedia, sobre el uso del software R y RStudio (Posit™) para analizar, manipular y visualizar datos. Los participantes aprenderán a realizar análisis y limpieza de datos, y a utilizar paqueterías para la manipulación y su visualización, empleando paqueterías como dplyr (tidyverse) y ggplot2. Además, identificarán repositorios de paqueterías e instalación de estas, y a configurar y guardar ambientes de trabajo en R.

Objetivo del curso:

Proporcionar a los participantes las habilidades necesarias para llevar a cabo análisis básico de datos utilizando el software R.

Registro

El **REGISTRO** se hará a través de un formulario en *Microsoft® Forms* disponible en:

<https://forms.office.com/r/BvUqSs1vgj>

Duración del curso:

La duración del curso es de 2 semanas, 2 horas diarias: una hora de enseñanza, y una hora de práctica y resolución de dudas. Quedarse a ***la hora práctica es opcional***.

Modalidad y ubicación	Fechas	Horario
Presencial – Biblioteca CICESE	Inicio: 26/06/23 Fin: 07/07/23	10:00 a 12:00 horas.



Aprendizaje esperado:

Al finalizar el curso, los participantes habrán adquirido conocimiento sobre el uso del software R (y RStudio (Posit™)) para análisis y visualización de datos. Los participantes estarán familiarizados con el software R y tendrán la capacidad de realizar diversos análisis.

Formato del curso:

El curso se impartirá en de forma presencial en la Biblioteca CICESE, en un horario de 10:00 a 12:00 h. Se usarán diapositivas para presentar la información del curso por lecciones y se realizarán ejercicios prácticos en el programa RStudio con el lenguaje R. Los participantes trabajarán en su computadora personal. Durante el curso también se brindará apoyo y comentarios de parte del (los) instructor (es) y demás participantes. Se contará con sesiones de preguntas y respuestas. Los comandos, diapositivas y material general estará disponible en un repositorio de GitHub con acceso público para todos los integrantes e interesados.

Requisitos previos:

Se requieren conocimientos básicos de manejo de software y estar familiarizados con la consola de comandos (CLI); aunque no es enteramente limitante.

A quien va dirigido:

A estudiantes universitarios, de posgrado, investigadores y al público en general que estén interesados en realizar investigación relacionada con el análisis de datos (en general).

Contenido del Curso

1. Introducción general
 - a. Introducción general de R (R Project)
 - b. Instalación de R y RStudio (Posit™)
 - c. Uso y configuración de R y RStudio
 - d. Instalación de paquetes en R y RStudio, y repositorios de paquetes
 - e. Guardar scripts y sesiones de trabajo del ambiente R
2. Manejo de bases de datos
 - a. Importación de datos al ambiente R
 - b. Bases de datos en texto plano (txt, tsv, csv)
 - c. Manipulación e inter-conversión de bases de datos en texto plano
 - d. Bases de datos en Microsoft® Excel (paquetería xlsx:.)
 - e. Formatos de tabla de datos: “long” & “wide” formats



3. Manipulación de bases de datos
 - a. Estadísticos con uno o múltiples factores.
 - b. Uso *tidyverse* para manipulación de bases de datos
 - i. Funciones `arrange()`, `mutate()` y `summarise()`
 - ii. Manejo de regex (expresiones regulares)
4. Gráficos en R
 - a. Entendiendo la sintaxis de `ggplot2`
 - b. Scatter/Line plots
 - c. Box-plot
 - d. Bar-plot
 - e. Otros
5. Aplicación de conocimiento adquirido: Trae tus datos y lo analizamos/graficamos en R.
 - a. Manipulación y edición de base de datos; visualización con `ggplot2`
 - b. ¡Publication-ready plots!

