

Curso-taller en línea: “R for Data Science: Fundamentos y Aplicaciones”

Perfil: Profesionales de áreas generadoras de datos (académicas, económicas, administración pública o privada, industria, etc.) y estudiantes de licenciatura (últimos semestres), maestría y doctorado que cumplan con los requisitos que se listan en el siguiente párrafo.

Profesor titular: Dr. Cristóbal Fresno Rodríguez y Dr. Guillermo de Anda Jáuregui, Instituto Nacional de Medicina Genómica.

Fecha: Del 01 al 28 de octubre de 2020.

Horario: jueves (1, 8 y 15) y miércoles (1 y 28) de octubre de 10:00 a 14:00 horas.

REQUISITOS

- R base: instalación de paquetes, manejo de *data.frames*, funciones, ciclos y gráficas básicas.
- Conocimientos básicos de estadística descriptiva.
- Notebook: 4 GB RAM con *Rstudio* y *R* versión 3.6 o superior.
- Disponibilidad de tiempo para trabajo práctico complementario.

OBJETIVO GENERAL DEL CURSO

Adquirir y desarrollar habilidades computacionales para la exploración y análisis de bases de datos utilizando las herramientas disponibles en el lenguaje de programación R.

TEMARIO

Parte 1: Introducción al curso

- 1.1 ¿Qué es la ciencia de datos?
- 1.2 ¿Qué aplicaciones tiene?
- 1.3 Datos limpios: El paradigma *tidyverse*

Parte 2: Exploración de datos: lectura, limpieza, y descripción

- 2.1 Formatos comunes de datos
- 2.2 Lectura de datos: *read.table*, *readr*, *data.table*, *vroom*
- 2.3 Estructuras de datos: listas, *data.frame*, *tibbles*
- 2.4 Estadística descriptiva: medidas de tendencia central, medidas de dispersión, representación gráfica.
- 2.5 Graficación base de R: diagramas de dispersión, cajas, histogramas, barras, etc.

Parte 3: Manipulación de datos: selecciones, transformaciones y rearrreglos

- 3.1 Rearreglo de datos: *reshape2*, *tidyR*
- 3.2 Filtrado de datos con *r-base* y *dplyr*
- 3.3 Unión de conjuntos de datos: *joins*, *binds*
- 3.4 Agregación de datos: *summarize*, *tally* ...
- 3.5 Transformaciones comunes de variables:
 - 3.5.1 variables categóricas (*forcats*)
 - 3.5.2 manipulación de texto (*stringr*)
 - 3.5.3 manipulación de fechas y horas (*lubridate*, *hms*)
- 3.6 Visualización: *ggplot2* (primitivas y capas de visualización)

Parte 4: Presentación de Resultados: gráficas en paneles y generación de reportes

- 4.1 Construye tus propios paneles de gráficas con *cowplot*
- 4.2 Reportes programáticos y reproducibles con *Rmarkdown*

Parte 5: Examen

- 5.1 Evaluación práctica de los temas del curso.

Bibliografía:

R para Ciencia de Datos <https://es.r4ds.hadley.nz/>

R for Data Science: Import, Tidy, Transform, Visualize and Model Data. 1st edition, O'Reilly Media, Inc. 20178, ISBN:1491910399 9781491910399

R Graphics Cookbook: Practical Recipes for Visualizing Data. O'Reilly Media, Inc. 20178. ISBN:1449316956