

¿A quién está dirigido el curso?



Al personal docente y alumnos interesados en introducirse al mundo del análisis estadístico y la ciencia de datos, para sus proyectos de investigación y trabajos de tesis.

Software utilizado en el curso



CENTRO INTERDISCIPLINARIO DE  
CIENCIAS MARINAS

**Mayores informes:**

Unidad de Tecnología Educativa y Campus  
Virtual

lperezr@ipn.mx

Teléfonos: (01)(612)1225344 y 1225366

Extensión: 81558

Segunda planta del Edificio 71 (Edif. de Plancton)

Av. Instituto Politécnico Nacional, S/N

Col. Playa Palo de Santa Rita

Visita nuestra página web:

<https://www.cicimar.ipn.mx>



Centro Interdisciplinario de Ciencias Marinas

## Introducción al análisis de datos con R

MSC. Alberto González Espinoza

14 de octubre - 4 de noviembre, 2019



Curso introductorio al análisis estadístico y la ciencia de datos con la plataforma R, enfocado al análisis cuantitativo y cualitativo de datos para proyectos de tesis e investigación.

- 30 horas de duración con valor curricular.
- Registro oficial ante de la Dirección de Educación Superior del Instituto Politécnico Nacional.

## ¿Qué es R?

R, es una plataforma de análisis estadístico avanzado de datos, que se compone del lenguaje de programación con el mismo nombre, su intérprete de comandos y una amplia biblioteca de funciones. Este sistema se distribuye gratuitamente debido a que se encuentra bajo la Licencia Pública General (GPL por sus siglas en inglés) del proyecto colaborativo de software libre GNU.

## ¿Qué podemos hacer con R?

Además de sus herramientas base para análisis estadístico, por ser un proyecto de código abierto, R incorpora una gran variedad de paquetes modulares, que son complementos especializados para temas específicos, desarrollados por especialistas en alguna parte del mundo. Estos paquetes se basan a menudo en métodos muy innovadores y permiten utilizar R para una amplia gama de problemas: existen paquetes para procesar datos tan diversos como censos, series de tiempo, secuencias genéticas o información económica, así como implementar una gran variedad de métodos estadísticos. Existen también, varios paquetes para la elaboración de gráficas y otros para hacer interactuar código de R y procesadores de texto como  $\text{\LaTeX}$  para la elaboración de reportes.

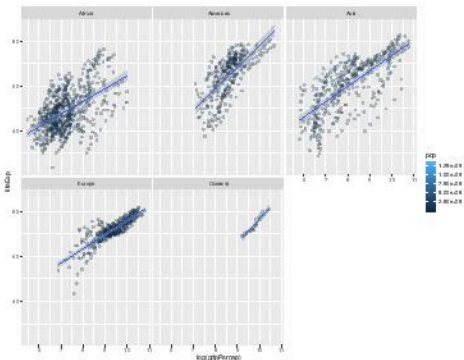


Figura 1: Ejemplo de un gráfico creado en R con la biblioteca de funciones ggplot2.

## ¿Por qué usar R?

R es una herramienta ampliamente utilizada en el campo de la investigación científica gracias a las características antes mencionadas, permitiendo realizar cálculos y análisis complejos sobre conjuntos de datos de gran tamaño, y obtener resultados precisos y confiables rápidamente, sin invertir en costosas licencias de software.

### Objetivo del curso

Al final del curso, los alumnos deberán ser capaces de crear scripts en el lenguaje R, que les permitan analizar estadísticamente distintos conjuntos de datos, obteniendo tablas con datos agregados, tablas de frecuencias, modelos lineales predictivos y gráficas con múltiples variables y facetas.

### Temario

1. Bases del lenguaje R
  - ¿Qué es R?
  - La consola de comandos.
  - Operaciones aritméticas.
  - Variables.
  - Tipos de datos.
  - Vectores.
  - Secuencias.
  - Tablas de frecuencias.
  - Cálculos estadísticos.
  - Números aleatorios.
  - Muestreo aleatorio.
2. Manejo de conjuntos de datos
  - Creación de dataframes.
  - Estructura de los dataframes.
  - Filtrar datos.
  - Conjuntos de datos (datasets).
  - Lectura de archivos CSV.
  - Gráficas con ggplot2.

- Exportar gráficos.
3. Análisis de los datos
    - Agregación de datos con dplyr.
    - Graficar múltiples variables.
    - Cálculo de correlación.
    - Modelo lineal.
    - Rangos en tablas de frecuencias.
    - Tabla de probabilidades.
    - Facetas en los gráficos.

### Evaluación



Al final de cada unidad de competencia, los alumnos elaborarán un ejercicio de evaluación para resolver una problemática con aplicación práctica sobre un conjunto de datos proporcionado.

### Detalles del curso

Imparte: MSC. Alberto González Espinoza  
Correo electrónico: [algonzalez@ipn.mx](mailto:algonzalez@ipn.mx)  
Curso de nivel licenciatura.  
Clave: DES/CICIMAR/C/53/2019  
Fechas: Del 14 de octubre al 4 de noviembre, 2019.  
Horario: 12:00 p.m. – 14:00 p.m.  
Sede: Sala multimedia del CICIMAR-IPN.  
Costo: \$657.72 MN

### Referencias

- [1] Vikram Dayal. *An Introduction to R for Quantitative Economics Graphing, Simulating and Computing*. Springer, 2018.
- [2] Christian Heumann and Michael Schomaker Shalabh. *Introduction to Statistics and Data Analysis*. Springer, 2016.
- [3] Dan E. Kelley. *Oceanographic Analysis with R*. Springer, 2018.
- [4] Manas A. Pathak. *Beginning Data Science with R*. Springer, 2014.