

Módulo 1: Tratamiento de datos en R

1.1. Introducción a R

Rafael Zambrano

rafazamb@gmail.com

R: Un lenguaje para la estadística

- R fue creado en 1992 en Nueva Zelanda por Ross Ihaka y Robert Gentleman
- Es un conjunto de programas integrados para el manejo de datos, simulaciones, cálculos y realización de gráficos
- Es una implementación libre, independiente y *opensource* del lenguaje de programación S, que fue desarrollado a mediados de los años 70 en Bell Labs

¿Por qué R?

- Permite manipular los datos rápidamente y de forma precisa
- Fácil automatización
- Lee prácticamente cualquier tipo de datos
- Compatible con grandes conjuntos de datos
- Gratuito
- Capacidades avanzadas de gráficos
- Se ejecuta en muchas plataformas
- Mejora su funcionalidad constantemente

Instalación de R y RStudio

- El software de R puede encontrarse para los distintos sistemas operativos en <https://cran.rstudio.com>

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

- Después de instalar R, a continuación puede instalarse un entorno de desarrollo amigable llamado RStudio, disponible en www.rstudio.com/products/rstudio/download



Instalación de R y RStudio

The screenshot displays the RStudio application window with four main panels, each highlighted by a red box and a label:

- 1- Code Editor:** The top-left panel shows an R script file named 'diamondPricing.R'. The code includes library loading, data viewing, summary generation, and a ggplot2 plot of carat vs price, colored by clarity.
- 2- R Console:** The bottom-left panel shows the output of the R script, including summary statistics for the 'diamonds' dataset and the execution of the plot command.
- 3- Workspace and History:** The top-right panel shows the 'Workspace' tab with the 'diamonds' dataset loaded. It also displays the 'History' tab with a list of executed commands.
- 4- Plots and files:** The bottom-right panel shows the 'Plots' tab with a scatter plot titled 'Diamond Pricing'. The plot shows 'Price' on the y-axis (ranging from 5000 to 10000) and 'Carat' on the x-axis (ranging from 1 to 3). The data points are colored by clarity, with a legend on the right showing 'VS1', 'VS2', 'VVS1', and 'IF'.

¡Gracias!

Contacto: Rafael Zambrano

rafazamb@gmail.com