

Análisis de Datos Categóricos

Ayudantía 1: Tutorial Descarga R

Roberto Velázquez

¿Qué es R?

- Un software de acceso libre y gratuito creado por **Robert Gentleman** y **Ross Ihaka (R)**, que tuvo su origen en 1993 como resultado del desarrollo del lenguaje de programación estadística S.
- El software fue desarrollado sobre la base del lenguaje homónimo, el cual se encuentra enfocado al **manejo y análisis estadístico, minería de datos, visualización de información**.
- En términos de experiencia de usuario, R es definido como un lenguaje de programación **orientado a objetos**, más que bases de datos (como SPSS o Stata).

¿Qué es R?

- Para el uso de todo el potencial analítico de esta herramienta, el trabajo de manipulación y estimación de datos y representaciones gráficas se realiza en base a **paquetes o librerías** gratuitas y comunitarias descargables desde el mismo programa.
- Actualmente existen más de **3000 librerías disponibles**, muchas de ellas creadas por destacados académicos de las ciencias sociales.

Ventajas de R

- Software libre y disponible en cualquier plataforma.
- Funciona en una variedad de sistemas operativos (Linux, MacOS, Windows, entre otros).
- Permite trabajar con datos que poseen diferentes formatos (SPSS, Stata, SAS, Excel, etc.).
- Algunas técnicas avanzadas y robustas solo pueden realizarse con este software.

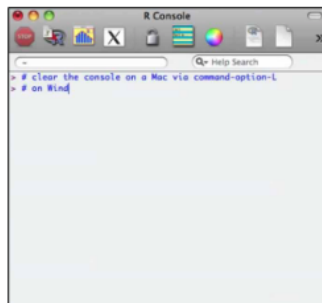
Ventajas de R

- Crea gráficos de alta calidad.
- Crea flujos de trabajo (informes reproducibles dinámicos).
- Contiene un módulo base de herramientas estadísticas que permiten a R interactuar con otros lenguajes: Python, Pearl, C, C++, etc.
- Capacidad gráfica de alto rendimiento. Funciona muy bien con LATEX.
- R forma parte de un proyecto colaborativo y abierto, en cuya base hay una comunidad académica activa.

¿Qué es RStudio?

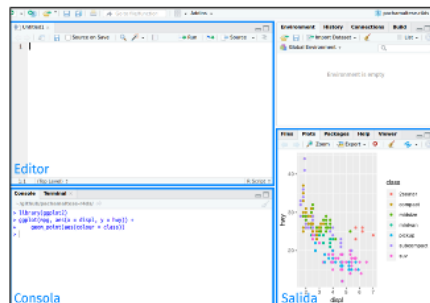
- La consola inicial de trabajo para el lenguaje R es poco intuitiva, por lo que se recomienda utilizar una interface gráfica.
- La interface más utilizada es **RStudio**, y es la que estaremos usando en el curso. Existen otras interfaces gráficas como Anaconda o R Commander, etc.
- El diseño de **RStudio** es muy fluido e interactivo debido a la visualización simultánea de varios elementos: **código, gráfico, consola y objetos**.

¿Qué es RStudio?



R

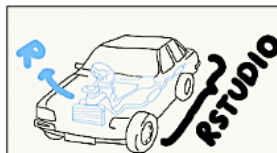
Lenguaje



RStudio

Salida

Editor



Instalación

PASO 1 - INSTALACIÓN R (lenguaje)

Para Windows:

- 1 Descargar la última versión desde <https://cran.r-project.org/bin/windows/base/>
- 2 Abrir el archivo ejecutable (.exe) y seguir las instrucciones para instalar el software.

Para Mac:

- 1 Descargar la última versión desde <https://cran.r-project.org/bin/macosx/>
- 2 Abrir el archivo ejecutable (.pkg) y seguir las instrucciones para instalar el software.

PASO 2 - INSTALACIÓN RStudio (editor)

- 1 Descargar la última versión del software desde <https://www.rstudio.com/products/rstudio/download/>
- 2 Seguir las instrucciones para instalar el software.
- 3 A partir de ahora utilice esta interfaz para usar R.

Explorando R Studio

The screenshot displays the R Studio interface with four main panels, each labeled with a number in a purple box:

- 1 – Editor/Source:** Contains the R script code for loading data and creating a plot.
- 2 – Environment:** Shows the workspace with the 'diamonds' dataset (53940 obs. of 10 variables) and the 'p' variable (ggplot object).
- 3 – Console:** Displays the output of the R commands, including summary statistics for the 'diamonds' dataset and the execution of the plot creation commands.
- 4 – Assistance:** Shows a preview of the 'Diamond Pricing' scatter plot, which is a faceted plot of Price vs. Carat, colored by Clarity.

Code in Editor/Source:

```
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 view(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- ggplot(carat, price,
12             data=diamonds, color=clarity,
13             xlab="Carat", ylab="Price",
14             main="diamond pricing")
15
```

Output in Console:

```
Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
1st Qu.: 4.710   1st Qu.: 4.720   1st Qu.: 2.910
Median : 5.700   Median : 5.710   Median : 3.530
Mean   : 5.731   Mean   : 5.735   Mean   : 3.539
3rd Qu.: 6.540   3rd Qu.: 6.540   3rd Qu.: 4.040
Max.   :10.740   Max.   :18.900   Max.   :31.800

> summary(diamonds$price)
   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
   326     950     2401     3933     5324    18820

> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- ggplot(carat, price,
+             data=diamonds, color=clarity,
+             xlab="Carat", ylab="Price",
+             main="diamond pricing")
> format.plot(p, size=
>
```

Environment Panel:

Object	Class	Attributes
diamonds	data.frame	53940 obs. of 10 variables
aveSize	numeric	0.7979
clarity	character	[8]
p	ggplot	[8]

Assistance Panel:

Diamond Pricing

Price

Clarity

- I1
- I2
- S11
- VSI2
- VSI1
- VVS2
- VVS1
- IF

Explorando R Studio

La consola de RStudio está dividida en cuatro ventanas o paneles que se describen a continuación:

- ➊ **Editor/Source:** para escribir, editar y guardar los scripts (códigos computacionales) que son creados por el usuario.
- ➋ **Environment:** en donde aparecen todos los objetos que son creados en la sesión junto con su estructura.
- ➌ **Console:** donde se encuentran los resultados ejecutados desde el panel Editor/Source. También es posible hacer cálculos directamente en esta consola.
- ➍ **Assistance:** en donde se visualizan los gráficos, las bases de datos, los directorios del sistema, así como la ayuda de las funciones de las librerías.

Directorio de trabajo en RStudio

Para editar y ejecutar un nuevo código desde RStudio siga las siguientes instrucciones:

- ➊ Crear un nuevo archivo de R: File -> New File -> R Script.
- ➋ Escribir los comandos en el panel de código.
- ➌ Ubicar el cursor en la línea que desea compilar (seleccionar las líneas) y ejecutar con los comandos Ctrl + Enter (command + Enter en mac).
- ➍ Guardar el archivo en la carpeta deseada: File -> Save

Resumen gráfico de interface en RStudio

Write Code

Navigate tabs
Open in new window
Save
Find and replace
Compile as notebook
Run selected code

R Support

Import data with wizard
History of past commands to run/copy
Display .Rpres slideshows
File > New File > R Presentation

The screenshot shows the RStudio IDE with the following components and labels:

- Source Editor:**
 - Line numbers 1-22.
 - Code:

```
# Good Start...  
"P0030001"  
"P0030002"  
"P0030003"  
"P0030004"  
  
get_digit <- function() {  
  ("num" %>% row(10 ^ n))  
  %>% K(10 ^ (n - 1))  
}
```
 - Labels: "Cursors of shared users", "Re-run previous code", "Source with or without Echo", "Show file outline", "Multiple cursors/column selection with Alt + mouse drag", "Code diagnostics that appear in the margin. Hover over diagnostic symbols for details.", "Syntax highlighting based on your file's extension", "Tab completion to finish function names, file paths, arguments, and more.", "Multi-language code snippets to quickly use common blocks of code.", "Jump to function in file", "Change file type", "Working Directory", "Press ↑ to see command history", "Maximize, minimize panes", "Drag pane boundaries".
- Environment/History/Build/Kit/Presentation:**
 - Labels: "Load workspace", "Save workspace", "Delete all saved objects", "Search inside environment", "Choose environment to display from list of parent environments", "Display objects as list or grid".
- Data:**
 - Variables: iris (150 obs. of 5 variables), Values, Functions, Foo (function (x)).
 - Labels: "Displays saved objects by type with short description", "View in data viewer", "View function source code".
- Files:**
 - Labels: "Create folder", "Upload file", "Delete file", "Rename file", "Change directory", "Path to displayed directory".
- Console:**
 - Code:

```
> foo(1)  
[1] 2  
> foo <- function(x) x + 1  
> foo(2)  
foo(2)  
> foo(3)
```
 - Labels: "A File browser keyed to your working directory. Click on file or directory name to open."

Breve tarea

- 1 Instalar R y RStudio.
- 2 Abrir y guardar un directorio de trabajo para ejercitar con la interface.
- 3 Explorar los distintos paneles de la interface.

Referencias

- Garrett Golemund (2014). *Hands-On Programming with R*. O'Reilly. Media, Inc.,s.
- Hadley Wickham (2015). *Advanced R*, CRC Press, Taylor & Francis Group, Boca Raton, FL. En <https://adv-r.hadley.nz/>
- Hadley Wickham and Garrett Golemund (2017). *R for Data Science*. Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media, Inc.. En <https://r4ds.had.co.nz/>
- R Project: <https://www.r-project.org/>
- RStudio: <https://rstudio.com/>