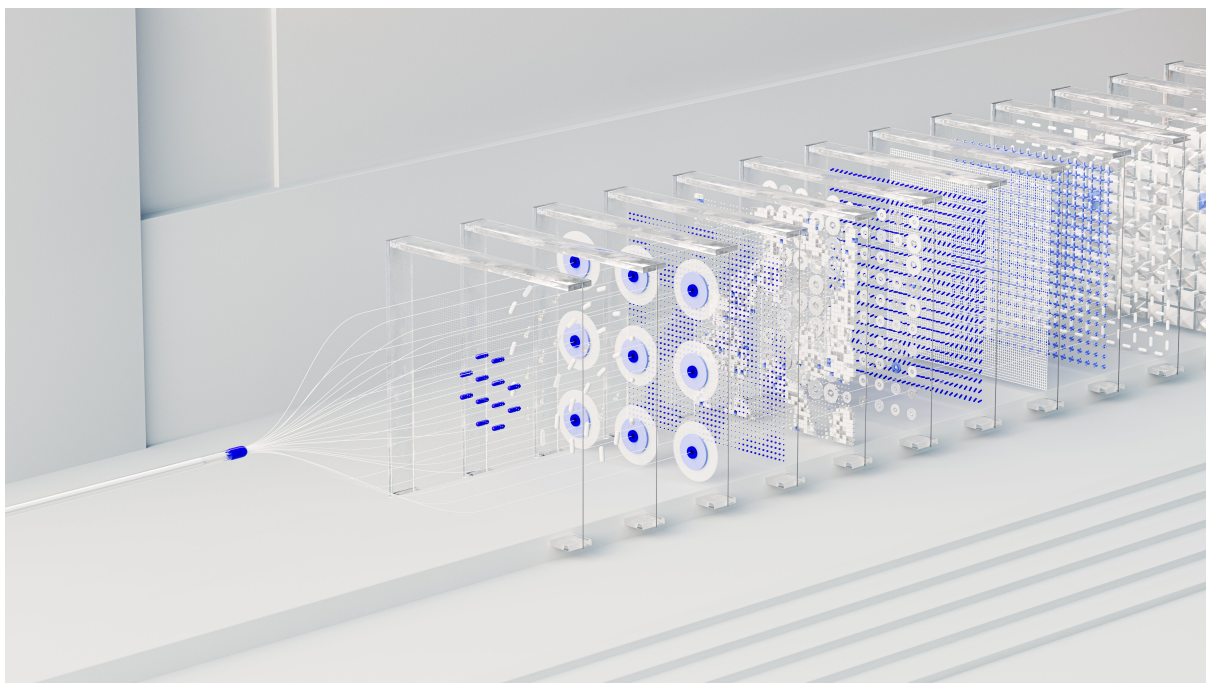


## CASO PRÁCTICO 101 HERRAMIENTAS PARA EL CIENTÍFICO DE DATOS



### R COMO HERRAMIENTA DEL CIENTÍFICO DE DATOS CIENCIA DE DATOS PARA NEGOCIOS

Fernando Aleisy González  
19 de enero de 2024

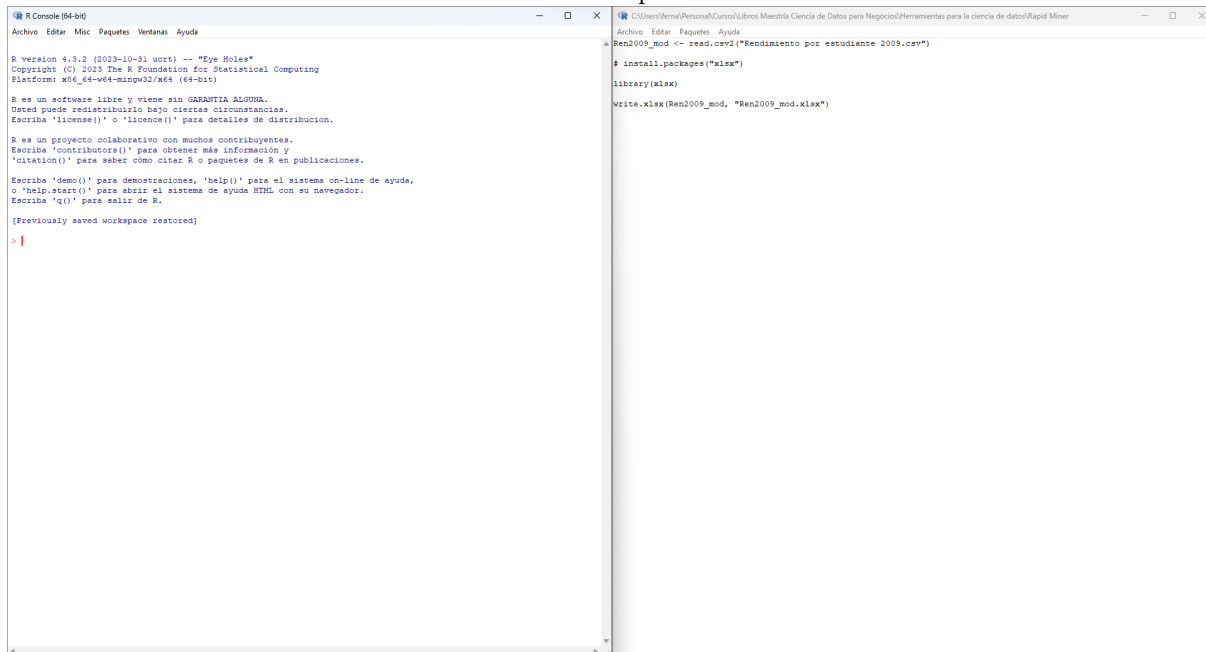
## Índice

<b>Descripción de la Herramienta</b>	<b>3</b>
Por qué usar R . . . . .	3
RStudio, el IDE más usado . . . . .	4
<b>Funciones más destacadas</b>	<b>6</b>
<b>Aplicaciones</b>	<b>7</b>
<b>Ejemplo de aplicación</b>	<b>8</b>
<b>Referencias</b>	<b>9</b>

## Descripción de la Herramienta

R es un lenguaje que fue desarrollado por estadísticos como un entorno para el análisis de datos, lo que implica que no es un lenguaje para el desarrollo de software. La interactividad que nos ofrece R es fundamental para el éxito en el campo del análisis de los datos ya que esta permite una rápida exploración de los datos (Irizarry, 2019).

Figura 1: Entorno de R con ventanas separadas. En la ventana de la izquierda se muestra la consola y en la ventana de la derecha el area donde se escribe el script



## Por qué usar R

R es un entorno de software libre para computación estadística y gráficos. Compila y se ejecuta en una amplia variedad de plataformas UNIX, Windows y MacOS lo que permite que el desarrollo se de más rápido que en los softwares de uso comercial debido a que los usuarios hacen desarrollos, los documentan y los suben al CRAN de R de manera cotidiana (Galán et al., 2016). Por los anterior, existen algunas razones de peso para decidirse por el uso de R en el análisis de datos:

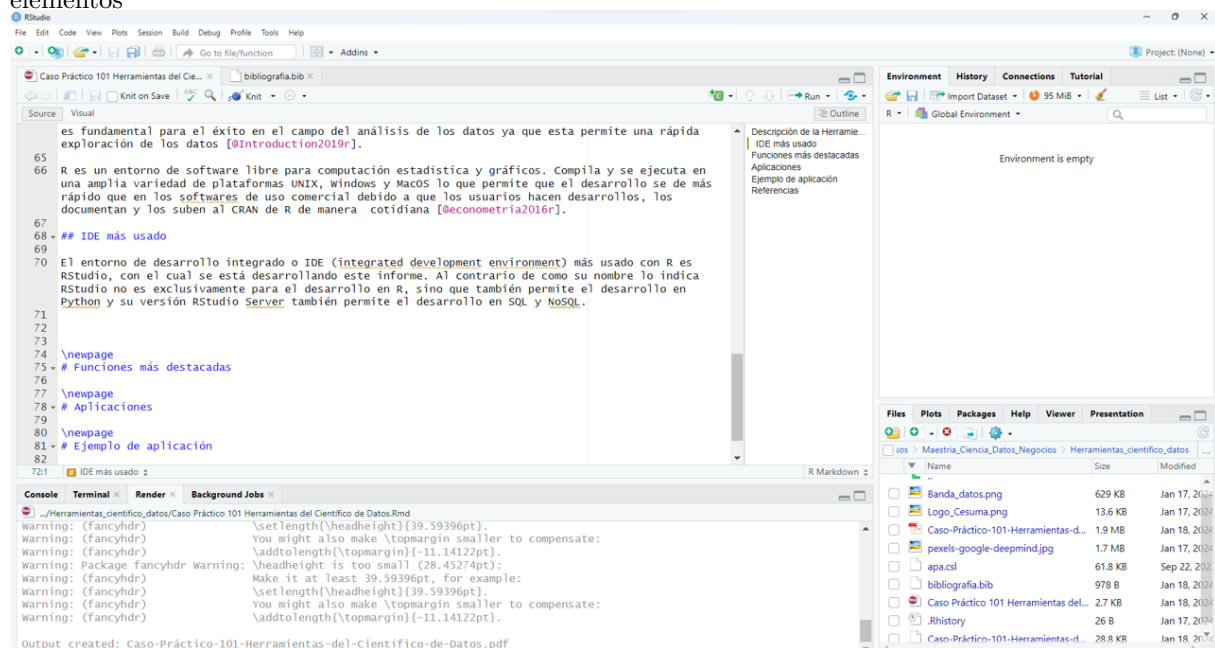
- R es ampliamente utilizado en la comunidad estadística y científica para realizar análisis estadísticos complejos, por lo que en Internet se puede encontrar respuestas a casi todos los inconvenientes que se presenten a la hora de realizar un análisis de datos.
- R es ampliamente utilizado en entornos académicos y de investigación, lo que hace que sea más fácil colaborar y compartir código y resultados con colegas.
- R tiene una amplia gama de funciones y paquetes específicos para estadísticas descriptivas, inferencia estadística, regresión, análisis multivariado y más.
- R proporciona herramientas poderosas para crear visualizaciones de datos de alta calidad. La librería **ggplot2** es especialmente popular por su flexibilidad y capacidad para producir gráficos estéticos y personalizables.
- R está diseñado para trabajar con conjuntos de datos de manera eficiente. Su sintaxis es especialmente adecuada para manipular y transformar datos, lo que facilita la limpieza y preparación de datos para análisis.
- R cuenta con una comunidad activa de usuarios y desarrolladores. Existen numerosos paquetes que amplían las funcionalidades de R para abordar diversas necesidades analíticas. Esta riqueza de paquetes permite a los usuarios acceder a una amplia gama de herramientas especializadas.

- R facilita la reproducibilidad en la investigación y análisis de datos. Los scripts y notebooks de R pueden documentar de manera efectiva los pasos realizados, lo que facilita compartir y reproducir análisis.
- R puede integrarse con otros lenguajes de programación, como C, C++, Java y Python, en especial cuando se utiliza el IDE RStudio. Esto permite aprovechar bibliotecas específicas de otros lenguajes cuando sea necesario.
- R es un software de código abierto, lo que significa que es gratuito y está disponible para su modificación y distribución. Esto ha contribuido a su popularidad y a la creación de una comunidad activa de usuarios y desarrolladores.

## RStudio, el IDE más usado

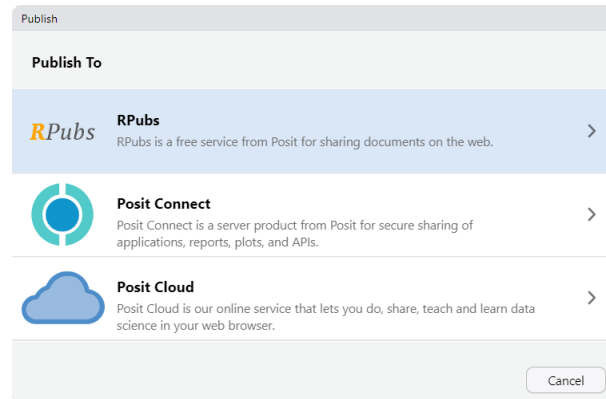
El entorno de desarrollo integrado o IDE (integrated development environment) más usado con R es RStudio, con el cual se está desarrollando este informe. Al contrario de como su nombre lo indica RStudio no es exclusivamente para el desarrollo en R, sino que también permite el desarrollo en Python y su versión RStudio Server también permite el desarrollo en SQL y NoSQL.

Figura 2: Vista general del IDE de RStudio. RStudio se divide en la zona de los scripts, el terminal y un panel lateral donde se muestra las variables creadas, el directorio de trabajo, las gráficas, entre otros elementos



Utilizando el procesador de texto RMarkdown que nos ofrece RStudio se abre la posibilidad de crear documentos en el lenguaje Markdown, el cual es un lenguaje de marcado sencillo que sirve para agregar formato, vínculos e imágenes con facilidad al texto simple, lo que permite crear paginas web de manera sencilla que luego se pueden alojar en RPubS, Posit Connect o Posit Cloud, siendo la primera la más popular.

Figura 3: plataforma de publicación abierta para documentos HTML producidos con RMarkdown desde RStudio



RMarkdown también nos ofrece la posibilidad de exportar nuestros informes a un documento de Word y si instalamos una versión ligera de LaTeX, también podremos exportar los informes en pdf permitiendo utilizar el lenguaje de marcado de LaTeX en RStudio como complemento al lenguaje de marcado Markdown. Cabe recalcar que no se debe mezclar LaTeX con Markdown, por lo que se debe avitar ciertas formalidades si se desea aprovechar el etiquetado de RMarkdown y mostrar tablas (`data.frames`) y gráficas hechas con código R.

En términos sencillos RMarkdown es un procesador de texto que ofrece además la posibilidad de incluir trozos de código desde R u otros formatos como Python o SQL (Boccardo & Ruiz, 2019).

## Funciones más destacadas

## Aplicaciones

## Ejemplo de aplicación



## Referencias

- Boccardo, G., & Ruiz, F. (2019). *RStudio para estadística descriptiva en ciencias sociales*. 23. <https://bookdown.org/gboccardo/manual-ED-UCH/>
- Galán, J., Feregrino, J., Ruíz, L., Quintana, L., Mendoza, M., & Rosales, R. (2016). *Econometría aplicada usando r*. 23.
- Irizarry, R. A. (2019). *Introduction to data science, data analysis and prediction algorithms with r*. <https://rafalab.dfci.harvard.edu/dsbook/>
- The R Foundation. (2002). *Statutes of “the r foundation for statistical computing”*. <https://www.r-project.org/foundation/>