

INTRODUCCIÓN A R, RSTUDIO Y RCOMMANDER

José Manuel Sánchez Santos

Prof. Titular de Universidad. Dpto. de Estadística. Universidad de Salamanca

Palabras clave: Software R, RStudio, RCommander, R-Project.

Objetivos

El principal objetivo del curso es guiar al alumno en una primera toma de contacto con el software R, uno de los más utilizados en el campo Biosanitario, pero que de entrada puede resultar difícil de manejar.

Otro objetivo no menos importante, será el de mostrar la necesidad de utilizar una interfaz gráfica como RStudio para el control del software R, que simplifique de manera notable el manejo y la programación con el lenguaje R.

Finalmente se pretende mostrar R como una potente herramienta para realizar cálculos estadísticos. Debido a la heterogeneidad de conocimientos estadísticos de los alumnos del curso, se ha optado por mostrarles el uso del paquete RCommander como vía de aprendizaje para realizar en R estadísticas y gráficos básicos.

Contenidos

R es un lenguaje y un entorno de programación para realizar análisis estadísticos y gráficos. Es un proyecto de software libre, surgido de la implementación GNU del lenguaje S. R es probablemente, uno de los lenguajes más utilizados en investigación por la comunidad estadística, sobre todo en los campos de la investigación biomédica, la bioinformática y la matemática financiera. La posibilidad de cargar diferentes bibliotecas o paquetes con finalidades específicas de cálculo o gráfico, ha contribuido enormemente al auge de este lenguaje. Otro importante punto a su favor, es el hecho de que R es un lenguaje orientado a objetos lo que facilita enormemente su uso incluso para investigadores con pocas nociones de programación informática. R se distribuye bajo la licencia GNU GPL y está disponible para los sistemas operativos Windows, Macintosh, Unix y GNU/Linux. Toda la información relativa a este software está disponible en www.r-project.org.

Durante el curso se mostrará con diferentes ejemplos y problemas, la necesidad de utilizar una interfaz de usuario que mejore las prestaciones de R. Esto nos conducirá a la instalación y manejo del programa RStudio, disponible en www.rstudio.com.

En el tramo final del curso se hará hincapié en el uso de paquetes y bibliotecas de R que la comunidad científica pone a disposición de los usuarios, para mostrar cómo gracias a ellos, este software estadístico ha crecido exponencialmente tanto en su utilización como en la creación de potentes herramientas estadísticas. Para realizar una primera toma de contacto con algunas herramientas que incorpora R, se instala y utiliza el paquete "Rcmdr" abreviatura de RCommander, www.rcommander.com, que es un

paquete que crea una interfaz gráfica con menús y botones para realizar cálculos estadísticos y gráficos de una manera visual, al estilo de aplicaciones como SPSS, Minitab, SAS...

Metodología

El curso comienza con la descarga e instalación del software R. A continuación se hace un recorrido por todos sus menús y herramientas, con el propósito de explicar su funcionamiento basado en funciones/comandos incluidos en paquetes (figura 1). Estos paquetes se clasifican en dos grandes grupos, los que se incluyen con la instalación básica y los que el usuario puede descargar de repositorios (CRAN, BioConductor, etc.)

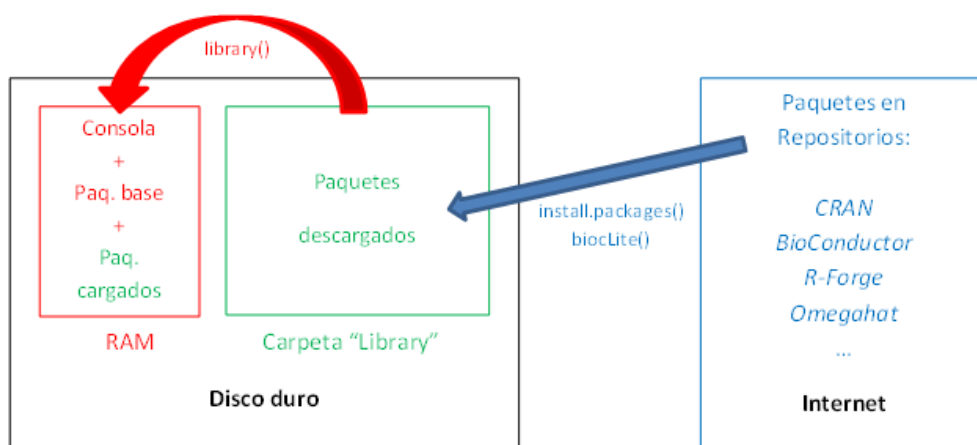


Figura 1: Estructura de R (consola, paquetes base y descargados, repositorios).

Para entender este software y especialmente su lenguaje de programación, se hace especial hincapié en las funciones de ayuda del programa, así como en la necesidad de utilizar los foros especializados en los que miles de especialistas participan de manera altruista. Se recomienda la atenta lectura de algunos manuales sobre R, como por ejemplo, "R para principiantes" de Emmanuel Paradis.

Uno de los principales escollos del manejo de R, radica en el hecho de que su interfaz de usuario se basa fundamentalmente en el uso de una consola en la que se ejecutan los comandos en una única línea (figura 2). Los usuarios pueden utilizar "scripts" (ficheros de sintaxis) en los que escribir varias líneas de código para ejecutar en la consola. Las ventanas de gráficos, ayuda, etc. se superponen lo que dificulta la visualización completa de los elementos que se están manejando.

Existen hoy en día otras interfaces de usuario que facilitan el manejo del entorno R, orientadas tanto a mejorar los ficheros de sintaxis como a gestionar todo lo relacionado con gráficos, paquetes, ficheros, ayuda, visualización de objetos, etc. Podemos destacar TinnR, Emacs, notepad++, WinEdt, gedit, Vim, Bluefish, ConTEXT... Algunos de ellos son complicados de configurar y otros son de pago.

Sin duda, una de las interfaces más utilizadas es RStudio (figura 3) cuya página web, www.rstudio.com, nos ofrece versiones para cualquier sistema operativo, versiones de escritorio o de servidor y sobre todo de manera gratuita, ya que RStudio ofrece versiones con licencia AGPLv3 para usuarios particulares.

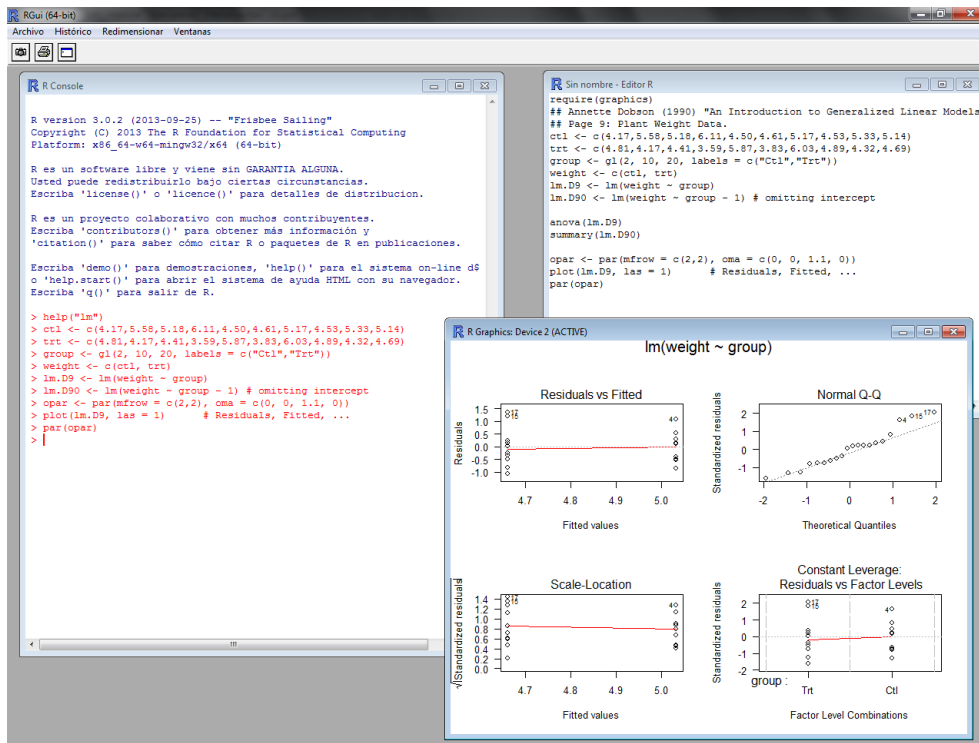


Figura 2: Interfaz por defecto de R con consola, guión y gráfico.

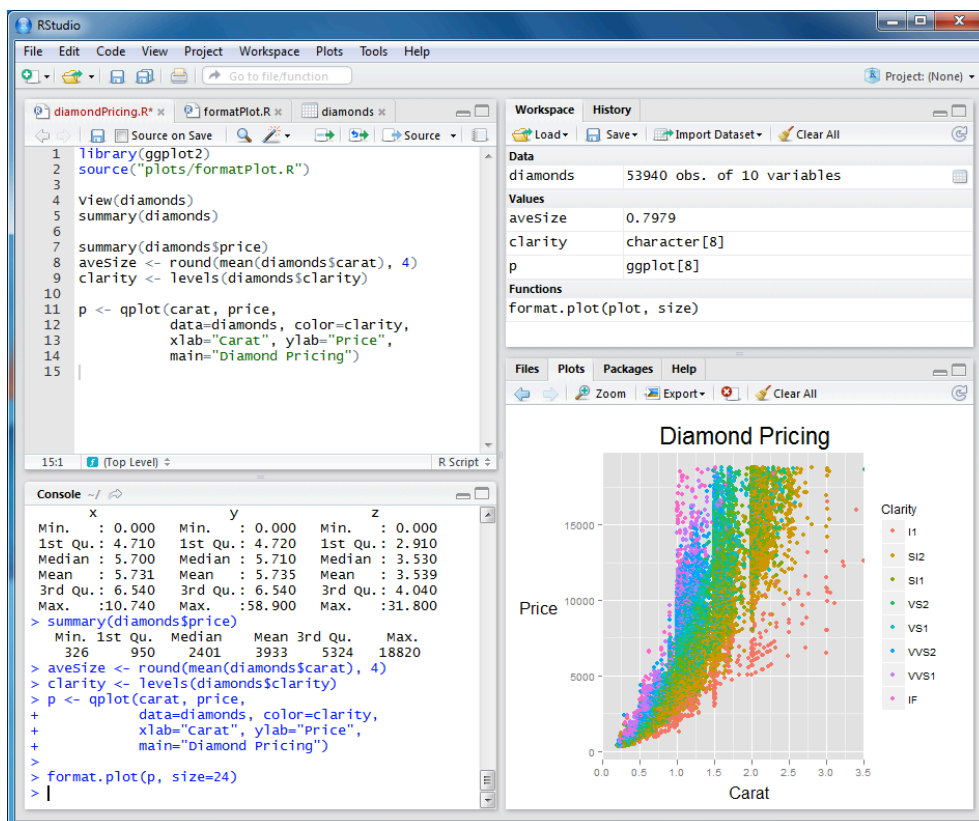


Figura 3: RStudio, interfaz para el manejo de R.

RStudio es un entorno de desarrollo integrado para R que incluye la consola, un editor de sintaxis que resalta comandos, funciones, texto, etc., herramientas para gráficos, historial de objetos, manejo y actualización de paquetes, navegador para las páginas de

ayuda, explorador de archivos... Y todo ello en una única ventana con múltiples pestañas completamente configurable.

Una vez instalado RStudio se realiza un pequeño tutorial para mostrar sus virtudes, y mediante un guión ("script") proporcionado a los alumnos, se comienza un exhaustivo recorrido por los principales tipos de objetos y normas elementales de programación que proporciona el lenguaje R. Mediante multitud de ejemplos se definen y utilizan objetos como *vectores*, *factores*, *matrices*, *arrays*, *listas*, *data frames*, *series temporales*, *funciones* y *expresiones*.

Finalmente se realizan algunos cálculos y gráficos estadísticos pero haciendo uso del paquete "RCmdr" de John Fox. Debido a la heterogeneidad de conocimientos estadísticos de los alumnos del curso no se profundiza mucho en la infinidad de técnicas estadísticas que proporciona R. Se ha optado por presentarles el paquete RCommander para que tomen contacto con la estadística básica de R, ya que este paquete crea una interfaz gráfica con menú y botones (al estilo de otros programas como SPSS o Minitab) que facilita el uso de dichas técnicas (figura 4). Además esta interfaz proporciona la sintaxis correspondiente a las técnicas que se van usando, lo que contribuye al aprendizaje progresivo de comandos y funciones estadísticas para su posterior uso en la consola o en "scripts".

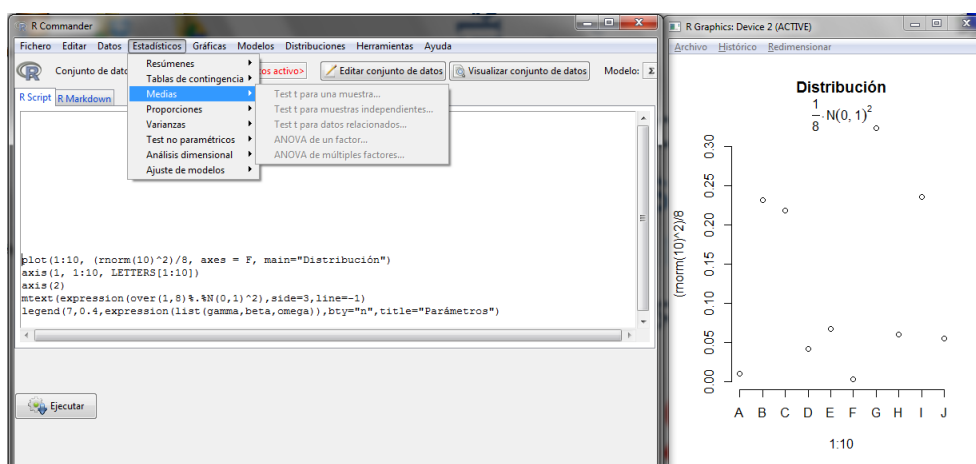


Figura 4: RCommander, interfaz gráfica para estadística básica en R.

Referencias

1. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL www.R-project.org.
2. RStudio (2012). RStudio: Integrated development environment for R (Version 0.98.953) [Computer software]. Boston, MA. URL www.rstudio.org.
3. Fox, J. (2005). The R Commander: A basic statistics graphical user interface to R. Journal of Statistical Software, 14(9): 1-42. URL www.rcommander.com
4. The Comprehensive R Archive Network CRAN. URL www.cran.r-project.org
5. Bioconductor: open source software for Bioinformatics. Fred Hutchinson Cancer Research Center. URL www.bioconductor.org.
6. Emmanuel Paradis (2002): R for beginners [monografía en Internet]. Traducida por Jorge A. Ahumada (2005): R para principiantes. Disponible en: http://cran.r-project.org/doc/contrib/rdebuts_es.pdf.