

Análisis de Datos con el Sistema Estadístico R

Lic. Patricia Vásquez Sotero



Sesión 1

Generalidades Acerca de R

CONTENIDO

- ☐ Objetivos y metodología del curso
- ☐ Un poco de historia sobre R
- ☐ Motivos para su uso
- ☐ Instalación de R
- ☐ Inicio de las sesiones en R
- ☐ R como calculadora científica
- ☐ La ayuda en R

GENERALIDADES

Objetivos

CONOCIMIENTOS Y CAPACIDADES

Proporcionar a los participantes conocimientos y herramientas para familiarizarlos en el análisis de datos que conforman la estadística, a fin de aplicarlos en la solución de problemas propios del quehacer científico en el ejercicio de su carrera profesional y la investigación.

GENERALIDADES

Metodología

- Integración de teoría y práctica con ejemplos y casos.
- Sesiones participativas y trabajos prácticos individuales.
- Requisitos: Estadística Descriptiva. Recomendable conocimientos sobre técnicas de programación.
- Clases: Sábado y domingo de 2pm - 6pm.
- Total horas lectivas: 30 horas. La primera clase es de 2 horas.
- 1 práctica calificada: 15 de septiembre de 2016.
- Examen final: 23 de septiembre de 2016.

SISTEMA ESTADÍSTICO 2

Introducción

EL PAPEL DE LA ESTADÍSTICA EN NUESTROS DÍAS

- ☐ Estadística o *ciencia del estado (en su origen)*
 - *Recolección – Análisis – Interpretación de datos*
- ☐ Su papel hoy en día en
 - Salud
 - Educación
 - Economía, etc... y, cómo no, en INVESTIGACIÓN
- ☐ Utilidad en el entorno global en el que nos encontramos a nivel individual

SISTEMA ESTADÍSTICO R

Introducción

PROGRAMAS DE CÁLCULO ESTADÍSTICO

	Potencia	Dificultad de uso	Precio	Uso prioritario
SPSS	Alta	Baja*	Muy alto	Docencia / profesional
STATA	Muy alta	Media	Medio/Alto	Investigación aplicada
R	Máxima	Alta	Gratuito	Investigación básica y aplicada

UN POCO DE HISTORIA

- R es un **lenguaje de programación** especialmente indicado **para el análisis estadístico**.
- R fue inicialmente diseñado por Robert Gentleman y Ross Ihaka, miembros del Dpto. de Estadística de la Universidad de Auckland, en Nueva Zelanda. Sin embargo, una de las grandes ventajas de R es que hoy en día es, en realidad, **fruto del esfuerzo de miles de personas en todo el mundo** que colaboran en su desarrollo.

¿QUÉ ES R?

- R es un lenguaje de programación para análisis estadístico de datos y gráficos
- Toda la información sobre R se encuentra en <http://www.R-project.org>
- El sistema R contiene dos componentes principales:
 1. Sistema base: contiene el software de lenguaje R y los paquetes complementarios de alta prioridad enumerados en la página 13.
 2. Paquetes complementarios aportados por el usuario

¿QUÉ ES R?

- El paquete estadístico R es uno de los más **flexibles, potentes y profesionales** que existen actualmente para realizar **tareas estadísticas** de todo tipo, desde las más elementales, hasta las más avanzadas.
- Cuenta, además, con la **ventaja de ser gratuito y de descarga e instalación sencillas**. El curso introduce al participante al manejo y uso a nivel básico del sistema estadístico R.
- R incorpora tareas sencillas, pero relevantes, relacionadas con el **manejo y análisis de datos en el contexto estadístico**. Entre ellas se incluye el manejo de datos estadísticos, la representación gráfica de datos univariantes y multivariantes, y la programación de tareas estadísticas con el lenguaje de programación incluido en R.



¿QUIÉNES USAN R?

- Todos los científicos especialmente aquellos que trabajan en países en desarrollo
 - Permite el libre acceso universal a las herramientas más avanzadas para el análisis de datos estadísticos
 - El más utilizado para la enseñanza de las estadísticas de estudiantes de pre y postgrado y pueden usarlo sin costo

MOTIVOS PARA EL USO DE R

RAZONES PARA UTILIZAR (Y NO) UTILIZAR

Pros

- Es libre
- Flexibilidad
- Procedimientos disponibles
- Se aprende estadística
- Los gráficos

Contras

- Al principio suele ser árido
- Cuesta un poco si se está habituado a trabajar con otro tipo de programas

INSTALACIÓN DEL SISTEMA R-BASE

1. Para instalar R hay que bajar un fichero ejecutable de la página web del proyecto R:
<https://cran.r-project.org/mirrors.html>
2. Elija su computadora de la lista *Spain* y luego el sistema operativo (*Linux, MacOS X o Windows*).
3. Haga clic en Base (*Base o Contrib: información sobre el software de terceros disponible para los servicios de CRAN Windows y el entorno correspondiente y variables de creación*).
4. Haga clic en R-3.5.1-win.exe
5. Guardar R

INSTALACIÓN DEL SISTEMA R-BASE



CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

The Comprehensive R Archive Network

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.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2018-07-02, Feather Spray) [R-3.5.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

● Descargar e instalar R

INSTALACIÓN DEL SISTEMA R-BASE

- Tras la descarga habrá que ejecutar el archivo con el nombre
R-3.5.1-win.exe.
- Una vez bajado este ejecutable, hay que ejecutarlo y seguir las instrucciones del programa de instalación.

Es recomendable elegir siempre la última versión disponible.

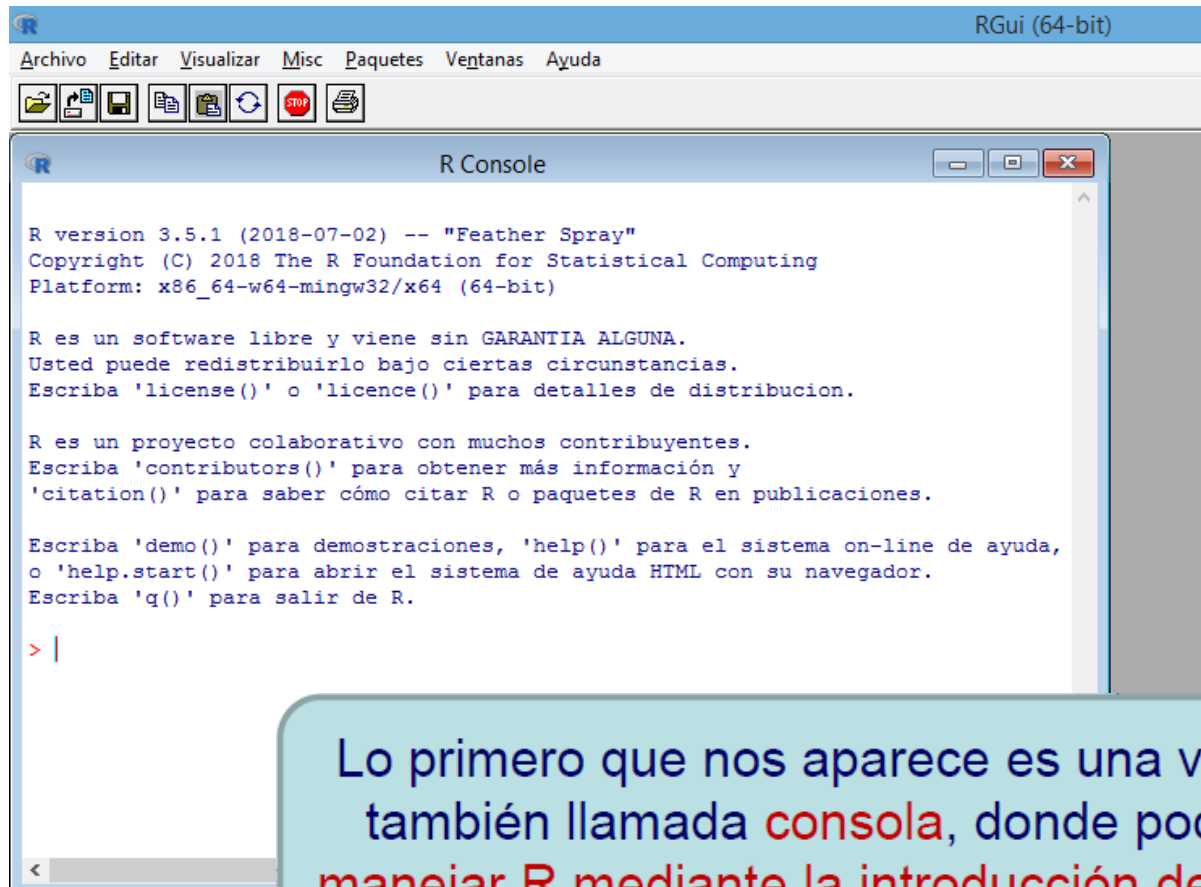
Tarea

Instalar la última versión de R tras desinstalar cualquier versión previa que pudiera haber en el ordenador de trabajo.

INICIO DE LAS SESIONES EN R

- Una vez instalado, podemos ya ejecutar el programa. Para ello, sólo hace falta hacer click en el icono de **R** o seleccionarlo desde el menú de programas.
- En el modo por defecto, abriendo **R** aparece una sola ventana, **la consola o ventana de comandos de R** en la cual introduciremos los comandos y será dónde se verán los resultados de los análisis.
- En la cabecera esta ventana aparece la versión que estamos utilizando. Conviene **visitar habitualmente la página del proyecto** para ver si hay una nueva versión.

INICIO DE LAS SESIONES EN R



```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

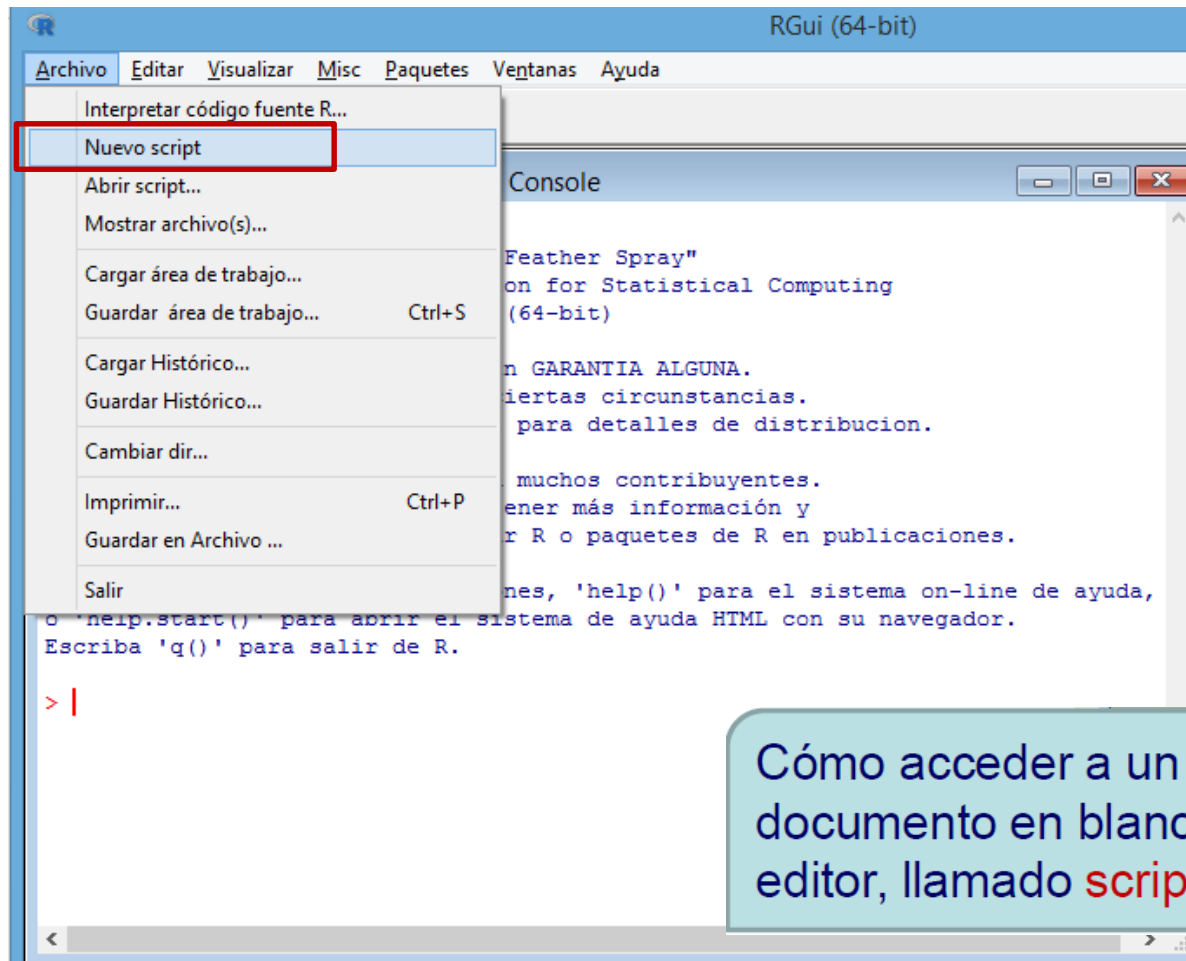
> |
```

Lo primero que nos aparece es una ventana, también llamada **consola**, donde podemos manejar R mediante la introducción de código.

LA CONSOLA Y EL EDITOR DE R

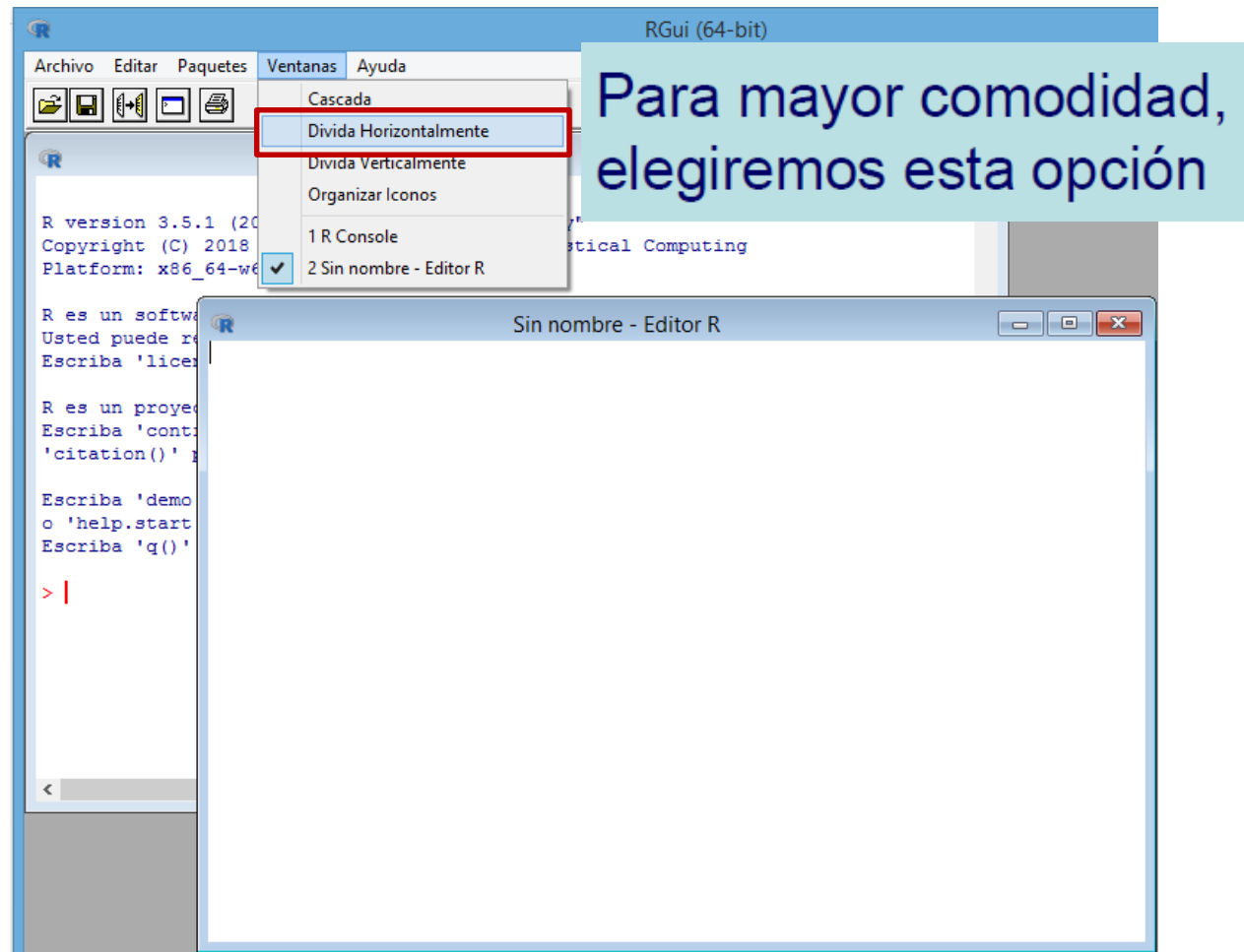
- Sin embargo, ésta no es la manera más eficiente de trabajar en R.
- Cuando estemos realizando un trabajo de mediana complejidad, será muy útil manejar todas las entradas de código que solicitemos a R en un entorno donde podamos corregirlas, retocarlas, repetirlas, guardarlas para continuar el trabajo en otro momento, etc.
- Ésta es la función del editor de R.

LA CONSOLA Y EL EDITOR DE R

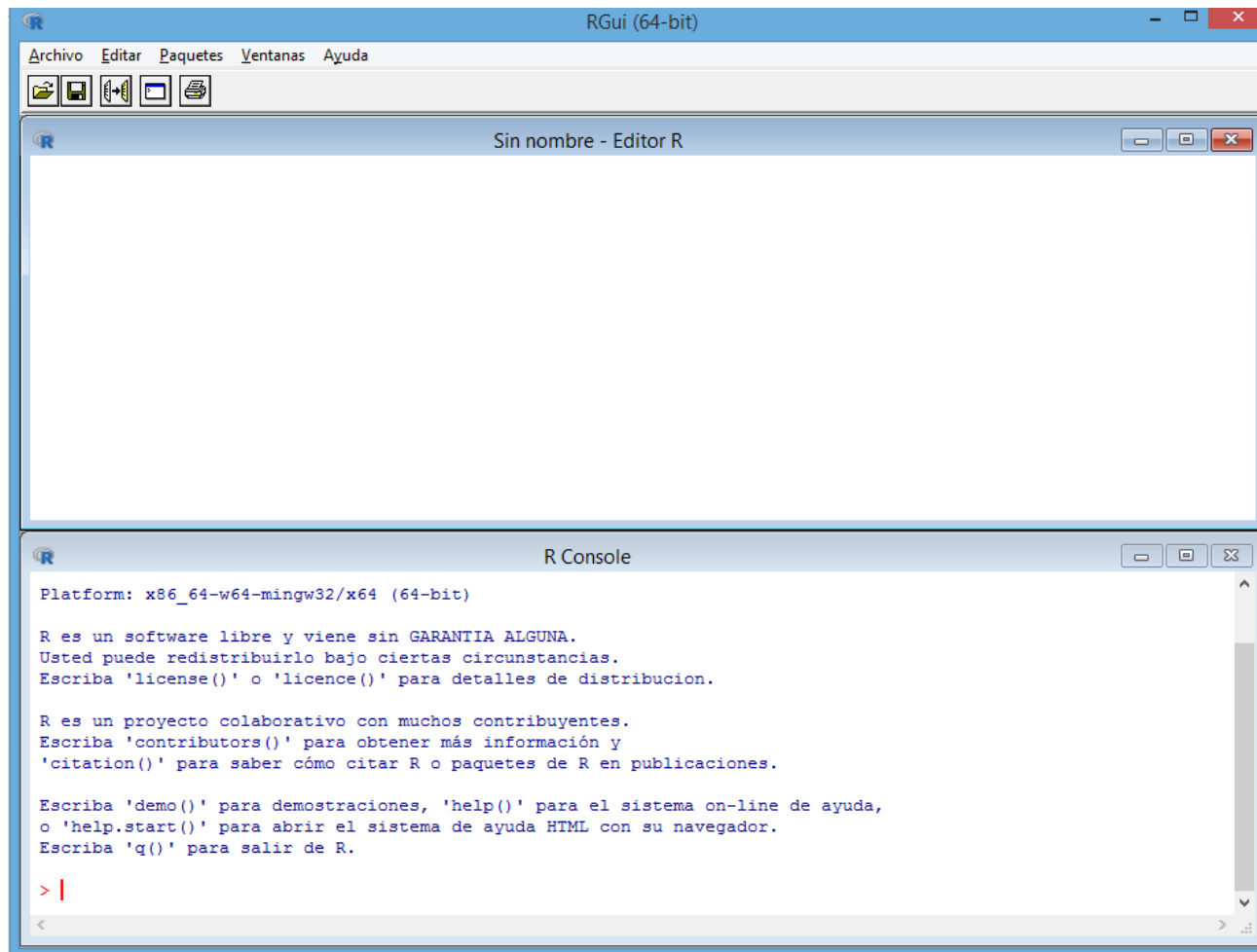


Cómo acceder a un documento en blanco del editor, llamado **script**.

LA CONSOLA Y EL EDITOR DE R



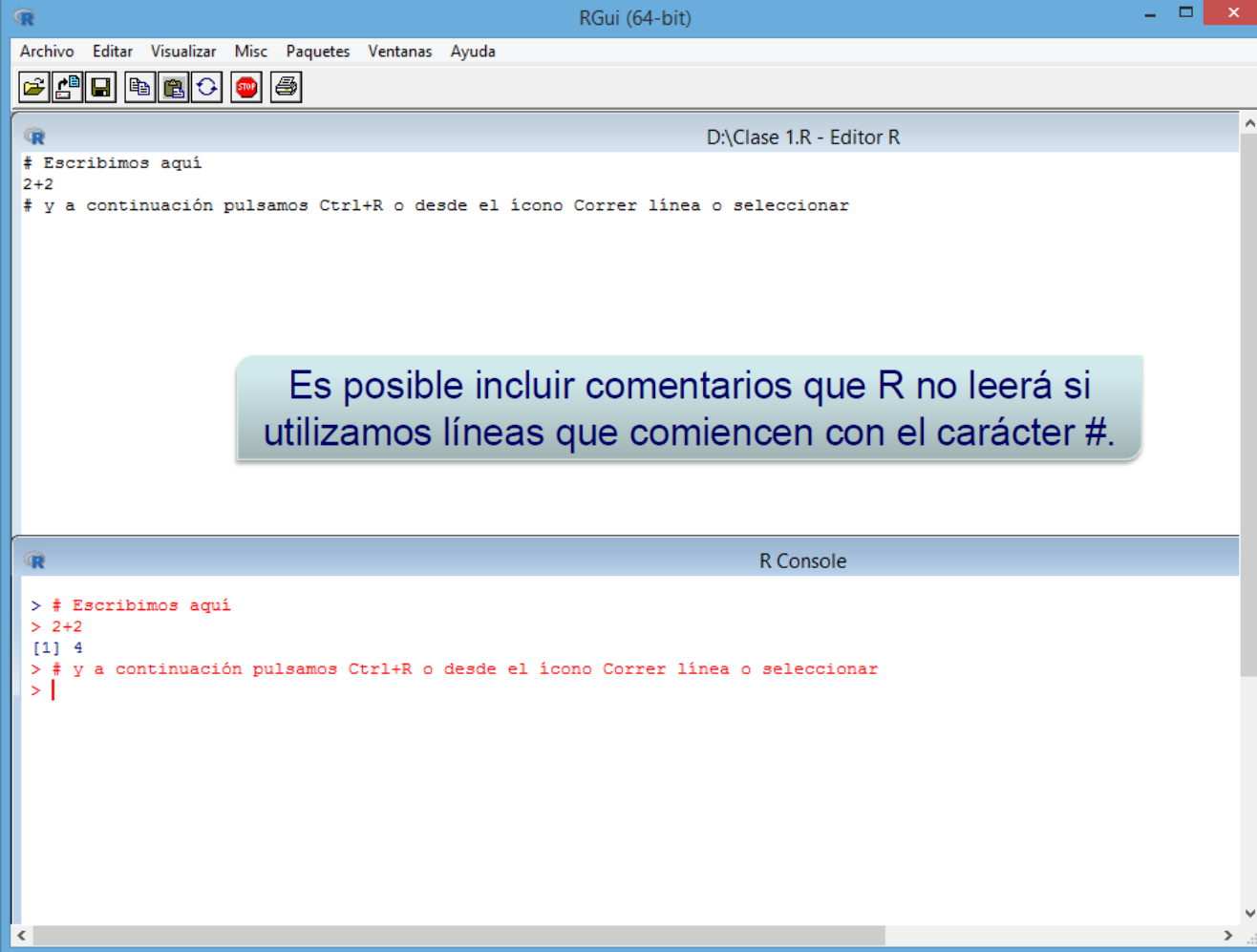
LA CONSOLA Y EL EDITOR DE R



LA CONSOLA Y EL EDITOR DE R

- Justo después de la cabecera, aparece una línea en blanco con el símbolo `>` en el margen izquierdo: **Prompt**. A partir de aquí R espera que escribamos **Comandos** e instrucciones para comenzar a trabajar.
- Para ejecutar un comando escrito utilizamos **Ctrl+R**.
- El signo **#** indica la introducción de un comentario.
- Las órdenes elementales consisten en expresiones o en asignaciones:
 - ▶ Si una orden consiste en una expresión, se evalúa, se imprime y su valor se pierde.
 - ▶ Una asignación, por el contrario, evalúa una expresión, no la imprime y guarda su valor en una variable
- El comando **demo()** permite realizar demostraciones del uso de R.
- Mediante **citation()** podemos saber cómo citar R o paquetes de R en publicaciones y así dar el crédito a la enorme cantidad de personas que desarrollan este proyecto.

LA CONSOLA Y EL EDITOR DE R



The screenshot displays the RGui (64-bit) application window. The top menu bar includes 'Archivo', 'Editar', 'Visualizar', 'Misc', 'Paquetes', 'Ventanas', and 'Ayuda'. Below the menu is a toolbar with icons for file operations and execution. The main window is titled 'D:\Clase 1.R - Editor R' and contains the following text:

```
# Escribimos aquí  
2+2  
# y a continuación pulsamos Ctrl+R o desde el icono Correr línea o seleccionar
```

Below the editor is the 'R Console' window, which shows the execution of the code from the editor:

```
> # Escribimos aquí  
> 2+2  
[1] 4  
> # y a continuación pulsamos Ctrl+R o desde el icono Correr línea o seleccionar  
> |
```

Es posible incluir comentarios que R no leerá si utilizamos líneas que comiencen con el carácter #.

LA CONSOLA Y EL EDITOR DE R

- En cada línea sólo se tiene un máximo de 128 caracteres: si queremos escribir más, podemos utilizar otra línea que comienza con el símbolo **+** en el lado izquierdo de la pantalla. Esto también puede ocurrir cuando el comando que terminamos de escribir no está completo (si falta un paréntesis, etc.).
- Si encontramos que un error es la falta de algo, podemos volver a teclearlo, y al pulsar **Ctrl+R**, ejecutará lo que hayamos solicitado.
- Si lo que ocurre es que nos hemos equivocado, al pulsar la tecla **Esc** el **prompt >** volverá a aparecer.
- Con la tecla **↑** recuperamos las instrucciones utilizadas en la sesión y con las teclas **→** **←** podemos corregir errores.
- Para separar expresiones se emplea **;** (punto y coma).
- Con el ícono **STOP** se interrumpe la edición o ejecución en curso.
- **q()** es el comando para salir de R.

R COMO CALCULADORA

- El uso más básico de R sería utilizarlo como una **calculadora científica**. Así:

3+4

Sumará 3 y 4 y devolverá el resultado.

- Cualquier **función matemática** que se nos ocurra está en R:

log(x), exp(x), log(x,n), log10(x), sqrt(x),
factorial(x), choose(n,x), gamma(x), lgamma(x),
floor(x), ceiling(x), trunc(x), round(x,digits=0),
signif(x,digits=6), cos(x), sin(x), tan(x), acos(x),
asin(x), atan(x), acosh(x), asinh(x), atanh(x), abs(x)

y podemos aplicarla sobre cualquier número.

R COMO CALCULADORA

- Sin embargo, lo que convierte a R en una potente herramienta de trabajo es que está diseñado de forma que la mayoría de operaciones (p.e. `+`, `-`, `*`, `/`, etc.) y de funciones (p.e. `log()`) están definidas con **carácter vectorial, es decir para operar componente a componente**.
- Es conveniente pues, en la medida de lo posible, explotar dicha posibilidad ya que agiliza en mucho el tiempo de computación.
- La función principal para definir un vector es a través de sus componentes, con la función `c()`, mediante el comando más importante en R que es `<-` el de la asignación.

R COMO CALCULADORA

- Los paréntesis **()** se emplean para los argumentos de las funciones y para agrupar expresiones algebraicas.
- Los corchetes **[]** o dobles corchetes **[[]]** para seleccionar partes de un objeto así como el **\$**. Las llaves **{ }** para agrupar expresiones.
- Para referirnos a la componente *n*ésima del vector *v* escribimos **v[n]**.

R COMO CALCULADORA

Tarea

Ejecutar los siguientes comandos:

```
log(((3+2)*5)+6)
```

Equivale a $\log(\{(3+2)*5\}+6)$ o $\log(\{\{3+2\}*5\}+6)$

```
v<-c(2,1,3,4)
```

crea un vector llamado v con cuatro valores

```
v
```

nos muestra el valor de dicho vector

```
w<-c(0,2,-2,1)
```

crea un vector llamado w

```
w[3]
```

nos muestra la tercera componente de w

R COMO CALCULADORA

Tarea

Calcula las siguientes operaciones, analiza los resultados obtenidos y observa los mensajes de advertencia:

$$2*v-3*w+2$$

multiplica componente a componente

$$w/v$$

divide componente a componente

$$v^3$$

eleva al cubo cada componente

$$v^w$$

eleva cada componente de v a su correspondiente componente de w

R COMO CALCULADORA

Tarea

`sqrt(w)`

produce un “warning” por pedir que haga la raíz de un negativo

`vw<-c(v,w)`

une los dos vectores y los almacena en uno nuevo vw

`vw`

muestra el resultado de vw

`vwa<-c(vw,6)`

une otra componente

`vwa/v`

da un “warning” pues ambos vectores no tienen el mismo número de componentes. Pero, ¿qué hace?

SALVAR E INICIAR SESIONES ANTERIORES

- Desde el menú **Archivo>guardar área de trabajo** se puede guardar una imagen del espacio de trabajo realizado hasta ese momento. La extensión para estos archivos es **.RData**.
- Para salir de R podemos seleccionar **Salir** del menú **Archivo**, ejecutar la orden **q()** o cerrar directamente la ventana del programa (o la consola). En ese momento, R siempre pregunta si el usuario quiere guardar el actual espacio de trabajo (si se acaba de hacer ya no es necesario).
- Podemos retomar el espacio de trabajo de otro día abriendo desde el menú **Archivo>cargar área de trabajo** la imagen anteriormente guardada, para restaurar los objetos creados en otras sesiones; también es posible haciendo doble click sobre el fichero **.RData**.
- También es posible guardar todos los comandos de la sesión desde el menú **Archivo>Guardar Histórico**. La extensión de estos cheros es **.RHistory**. Se pueden editar: lo veremos en scripts.

LA AYUDA EN R

- R dispone de una ayuda muy completa sobre todas las funciones, procedimientos y elementos que configuran el lenguaje.
- Además de las opciones de menú propias de R, desde la ventana de comandos se puede acceder a información específica sobre las funciones de R con el comando **help** o mediante **?**.

help(log)

Nos abre la ventana de ayuda sobre la función **log**.

? ls

Hace lo mismo para la función **ls**.

Tarea

Consultar la ayuda sobre los comandos **apropos** y **find**.

Consultar la ayuda sobre **if** (observar la diferencia entre **?if**, **help("if")** y **help('if')**).

LA AYUDA EN R

- Cuando no sepamos el nombre exacto del comando o de la función que necesitamos, pero sí que sepamos el tema sobre el que queremos ayuda (en inglés), podemos utilizar la función `help.search`.
- Por ejemplo, si queremos introducir datos, y no sabemos que comandos nos pueden ayudar:

`help.search("data input")`

Con un poco de suerte aparecerá información sobre el nombre de funciones asociadas a esta búsqueda.

`?read.table`

Nos aportará dicha info.

- Es posible acceder a manuales vía el menú Ayuda-Manuales (en PDF) o vía CRAN:

<http://cran.r-project.org/>

LA AYUDA EN R

Algunos de los manuales disponibles son:

- **An Introduction to R** introduce el lenguaje, a la vez que presenta cómo utilizar R como herramienta para el análisis estadístico y gráfico. Existe una traducción de éste en español.
- **R Language Definition** documenta el lenguaje per se, es decir los objetos en los que se basa y los detalles del proceso de evaluación. De gran utilidad cuando se va a programar funciones.
- **Writing R Extensions** cubre cómo crear nuestras propias librerías, cómo escribir los archivos de ayuda y cómo utilizar y conectar con lenguajes externos (C, C++, Fortran, etc.).
- **R Data Import/Export** describe las utilidades para importar y exportar datos en R incluyendo librerías disponibles en CRAN.

MÁS SOBRE LA AYUDA EN R

- Desde el propio programa, a través del menú Ayuda podemos ir a la página web de las FAQ. Hay dos, una más genérica **R FAQ** y otra para windows **R for Windows FAQ**.
- Si nos vemos muy apurados podemos intentar preguntar a alguien que ya haya pasado por los mismos problemas anteriormente.
- Para ello existen varias listas de correo (foros) donde preguntar dudas, aunque antes de plantear una pregunta conviene revisar antes lo que ya se ha preguntado y que por tanto ya aparece en las FAQ.
- Por supuesto, pertenecer a una lista permite recibir todas las dudas que allí en el foro se plantean (aunque a veces puede llegar a ser cansino recibir tantos correos).
- El acceso a estas listas se puede realizar desde la página web del proyecto R.

CURIOSIDADES EN R

- La experiencia de Ross and Robert en el diseño e implementación del lenguaje está descrito en Ihaka and Gentleman (1996), R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*, 5, 299-314.
- La importancia de este lenguaje es tal que existen dos congresos (useR -anual- y DSC -bianual-) enteramente dedicados a su uso.
- Desde 2009 existe una publicación The R Journal enteramente dedicada a artículos sobre el desarrollo y la aplicación de R.

Tarea

Consultar toda la información disponible en la página del proyecto R.

BIBLIOGRAFÍA EN R

- ¡La bibliografía de R es extensísima! Imposible darla toda entera.
- En la página web del proyecto existe disponible mucha *Contributed Documentation* que además es de libre distribución.
- Entre la enorme cantidad de libros (¡posiblemente en la actualidad haya más de 100!) enteramente dedicados a R destacamos únicamente dos:
 - ▶ John M. Chambers (2008), “Software for Data Analysis: Programming with R”. Springer, New York.
 - ▶ Peter Dalgaard (2008), “Introductory Statistics with R”, 2nd edition. Springer.
- Una muy buen referencia es: Michael J. Crawley (2007), The R book. John Wiley and Sons, Ltd.

BIBLIOGRAFÍA EN R



CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

Contributed Documentation

Note: The CRAN area for contributed documentation is frozen and no longer actively maintained.

[English](#) --- [Other Languages](#)

Manuals, tutorials, etc. provided by users of R. The R core team does not take any responsibility for contents, but we appreciate the effort very much and encourage everybody to contribute to this list! To submit, follow the submission instructions on the [CRAN main page](#). All material below is available directly from CRAN, you may also want to look at the list of [other R documentation](#) available on the Internet.

Note: Please use the [directory listing](#) to sort by name, size or date (e.g., to see which documents have been updated lately).

English Documents

Documents with more than 100 pages:

- “**Visual Statistics. Use R!**” by Alexey Shipunov ([PDF](#), 2017-05-15, 388 pages). All book materials are accessible from [Alexey Shipunov's English R page](#).
- “**Using R for Data Analysis and Graphics - Introduction, Examples and Commentary**” by John Maindonald ([PDF](#), data sets and scripts are available at [JM's homepage](#)).
- “**Practical Regression and Anova using R**” by Julian J. Faraway ([PDF](#), data sets and scripts are available at the [book homepage](#)).
- The [Web Appendix](#) to the book “An R Companion to Applied Regression” (second edition) by John Fox and Sanford Weisberg contains information about R to fit a variety of regression models.
- “**An Introduction to S and the Hmisc and Design Libraries**” by Carlos Alzola and Frank E. Harrell, especially of interest to SAS users, users of the Hmisc or Design packages, or R users interested in data manipulation, recoding, etc. ([PDF](#)).
- “**Statistical Computing and Graphics Course Notes**” by Frank E. Harrell, includes material on S, LaTeX, reproducible research, making good graphs, brief overview of computer languages, etc. ([PDF](#)).
- “**An Introduction to R: Software for Statistical Modelling & Computing**” by Petra Kuhnert and Bill Venables ([ZIP 3.8MB](#)): A 360 page PDF document of lecture notes in combination with the data sets and R scripts used in the manuscript.
- “**Introduction to the R Project for Statistical Computing for Use at the ITC**” by David Rossiter ([PDF](#), 2012-08-20, 141 pages).
- “**Analysis of Epidemiological Data Using R and Epicale**” by Virasakdi Chongsuvivatwong ([PDF](#)).
- “**Statistics Using R with Biological Examples**” by Kim Seefeld and Ernst Linder ([PDF](#)).
- “**IcebreakerR**” by Andrew Robinson ([PDF](#), 2016-06-21, 161 pages).
- “**Applied Statistics for Bioinformatics Using R**” by Wim Krijnen ([PDF](#), 2009-11-17, 278 pages).
- “**An Introduction to R**” by Longhow Lam ([PDF](#), 2010-10-28, 212 pages).
- “**R and Data Mining: Examples and Case Studies**” by Yanchang Zhao ([PDF](#), 2013-04-26, 160 pages).
- “**A Student's Guide to R**” by Nicholas J. Horton, Randall Pruim, and Daniel T. Kaplan ([PDF](#), 2015-11-16, 119 pages).
- “**Is R Suitable Enough for Biostatisticians?**” by Adrian Olszewski ([PDF](#), 2015-06-28, 365 pages).

Documents with fewer than 100 pages:

Comunicación constante con la Escuela del INEI

Correo de la Dirección Técnica de la ENEI

Sr. Eduardo Villa Morocho (Eduardo.villa@inei.gob.pe)

Coordinación Académica

Sra. María Elena Quirós Cubillas (Maria.Quiros@inei.gob.pe)

Correo de la Escuela del INEI

enei@inei.gob.pe

Área de Educación Virtual

Sr. Gonzalo Anchante (gonzalo.anchante@inei.gob.pe)

