

Introducción al lenguaje R

Breve Guía de Instalación y Uso

Ester Bernadó

1. ¿Qué es R?

R es un lenguaje y un entorno de programación. El lenguaje deriva de una implementación de un lenguaje previo denominado lenguaje S y sus posteriores implementaciones denominadas S-Plus, que fueron desarrolladas en Bell Laboratories por Rick Becker, John Chambers y Allan Wilks

R proporciona un conjunto de técnicas estadísticas y técnicas de visualización para el análisis de datos. Además, contiene numerosas librerías para la minería de datos y su funcionalidad es fácilmente extensible gracias a múltiples paquetes que se pueden instalar en el entorno. Su uso está muy extendido en ámbitos de estadística y ciencia de los datos en general.

El sitio web del proyecto R es:

<https://www.r-project.org/about.html>

R está disponible como un software de libre distribución bajo la licencia [Free Software Foundation](#) [GNU General Public License](#).

2. Características de R

Tal como lo describe la *R Foundation*, R incluye las siguientes características:

- funcionalidades para el tratamiento de datos y almacenamiento,
- un conjunto de operadores para cálculos con vectores y matrices,
- una colección de herramientas para el análisis de datos,
- funcionalidades gráficas para la visualización y análisis de datos,
- un lenguaje de programación que incluye condicionales, bucles y funciones que puede definir el usuario.

3. Descargar e instalar R

3.1. Descargar R

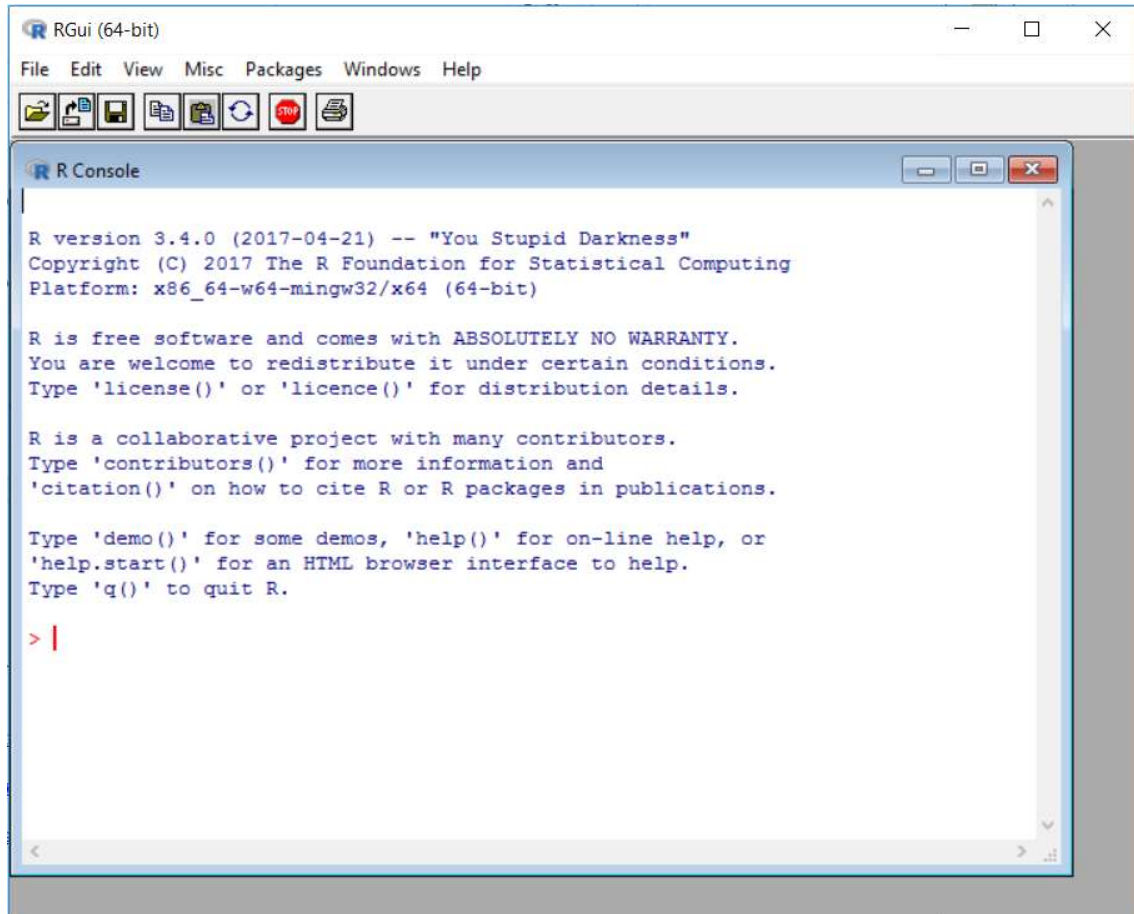
Para descargar el lenguaje R, se debe acceder a la web del proyecto R: <https://www.r-project.org/> y seleccionar *Download R*. En el apartado de descargas se nos redirigirá a la página CRAN (Comprehensive R Archive Network) para escoger el servidor de descargas R más cercano. CRAN es una red de servidores replicados (*mirror servers*) que facilita los procesos de descarga del software R y de otros paquetes que lo complementan. En el sitio web, también se encuentra abundante información, como tutoriales del lenguaje R e información sobre la instalación del entorno.

El proceso de instalación es muy sencillo: se descarga el fichero .exe y se ejecuta en el ordenador local. En el proceso de instalación nos pregunta la posibilidad de escoger el directorio donde se almacenarán el resto de paquetes que se instalen en R.

3.2. Entorno R

La instalación del software R viene por defecto con un entorno de programación sencillo que consiste básicamente en una pantalla o consola desde la cual se puede interactuar con el lenguaje R. Se denomina RGui (R Graphical User Interface). R es un lenguaje interpretado. Es decir, se escriben las órdenes y el entorno las procesa una por una y devuelve el resultado.

El entorno RGui tiene un aspecto como el que se ilustra en la siguiente figura:



El símbolo “>” indica que la consola está lista para recibir instrucciones. La salida de la instrucción se ilustra con un número dentro de las llaves “[]”. El número indica el número de elemento que se está visualizando.

Si ejecutamos el típico programa “Hello World” en R, obtenemos:

```
> print("hello world")
[1] "hello world"
> |
```

En el siguiente ejemplo creamos un vector con la secuencia de números de 1 a 100 y visualizamos el resultado:

```
> a<-seq(1:100)
> a
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18
[19] 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
[37] 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
[55] 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
[73] 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
[91] 91 92 93 94 95 96 97 98 99 100
> |
```

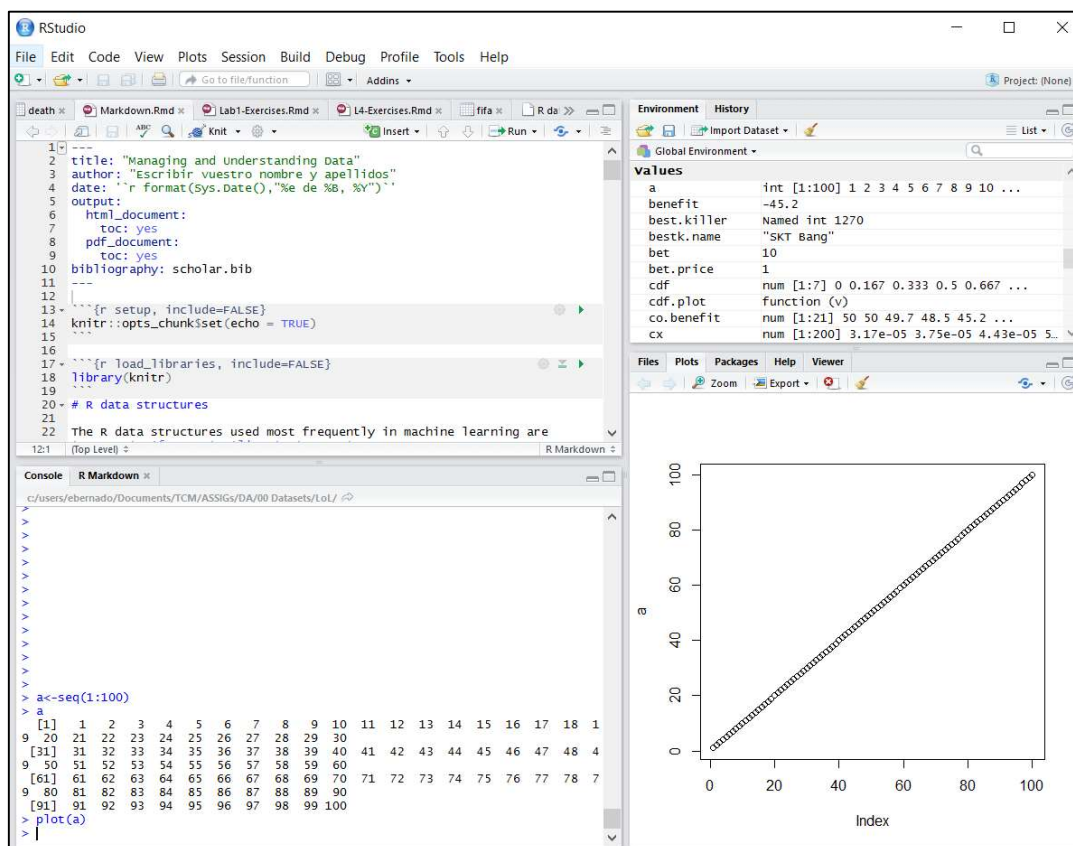
La primera instrucción, crea el vector denominado “a” a partir de la función “seq” que genera una secuencia de números según se especifica en sus parámetros. En este ejemplo, se genera la secuencia de 1 a 100. A continuación, la instrucción “a” visualiza el contenido de la variable.

3.3. Descargar R Studio

RStudio es un entorno de programación más completo y usable. Permite trabajar en modo consola, al igual que en RGui, pero además incluye la posibilidad de trabajar con ficheros de instrucciones R (que se guardan con extensión .R). Estos ficheros se pueden guardar e importar posteriormente para su uso posterior. El entorno incluye también la posibilidad de visualizar la salida en una ventana adicional, visualizar el contenido de las variables del entorno, etcétera. El entorno RStudio se puede descargar de:

<https://www.rstudio.com/products/rstudio/download>

A continuación se muestra el aspecto que presenta el entorno RStudio:



En este entorno, se pueden localizar cuatro ventanas principales:

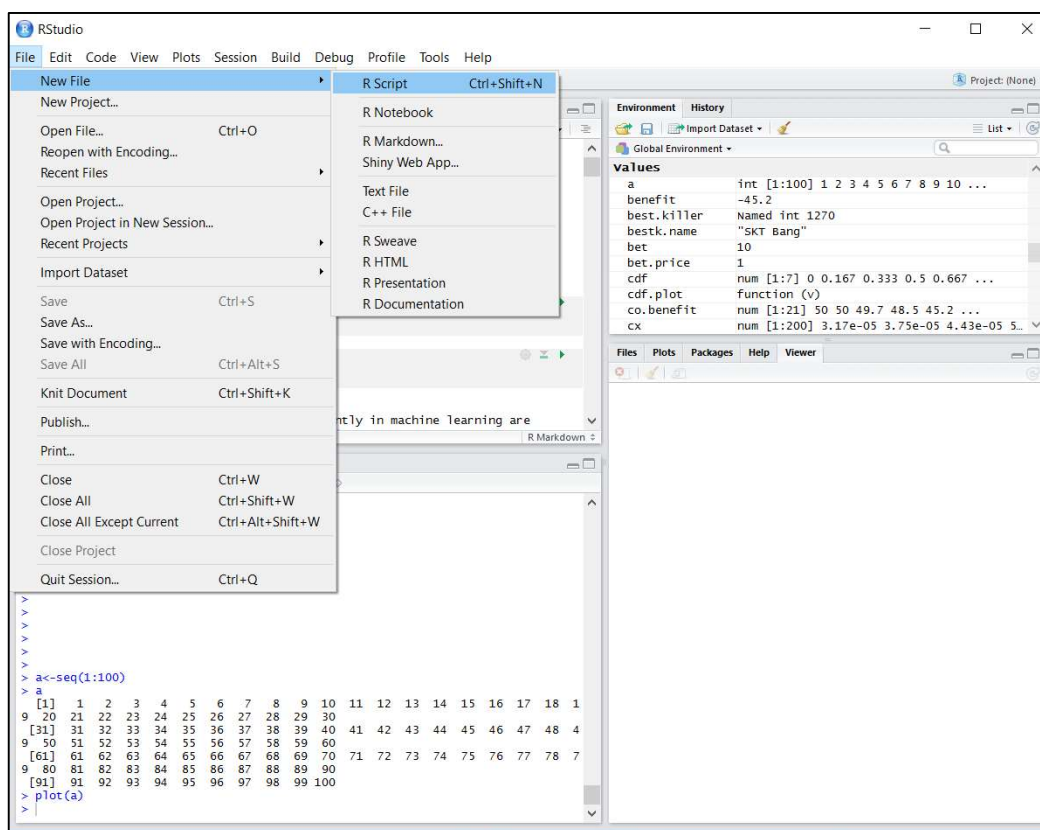
- **Editor** (esquina superior izquierda): en esta ventana, se pueden editar ficheros R. Si hay varios ficheros R abiertos en el entorno, se puede navegar por ellos a través de las pestañas que hay en el borde superior.

- **Consola** (esquina inferior izquierda): esta ventana funciona como consola, de forma análoga a la del entorno RGui. En ella, se ejecutan instrucciones que son interpretadas. En la figura, se muestra la creación de un vector “a” con números del 1 al 100 y a continuación se visualiza el contenido del vector y se representa gráficamente con la instrucción *plot*.
- **Entorno** (esquina superior derecha): se visualizan las variables del entorno. Hay una lista de variables con su identificador y el tipo de variable. A través de las pestañas superiores se puede consultar el histórico de instrucciones ejecutadas.
- **Visor** (esquina inferior derecha): muestra la salida gráfica. En el ejemplo que se muestra en la figura, se muestra la salida de la función *plot* que se ha ejecutado en la consola. En esta ventana, también se pueden seleccionar otras funcionalidades (a partir de las pestañas del borde superior): Files (ficheros del sistema), Plots (ventana para mostrar los resultados gráficos que se muestra activa en esta figura), Packages (listado de paquetes disponibles), Help (ayuda sobre funciones de R) y Viewer (salida de las instrucciones de R).

3.4. Familiarizarse con el entorno RStudio

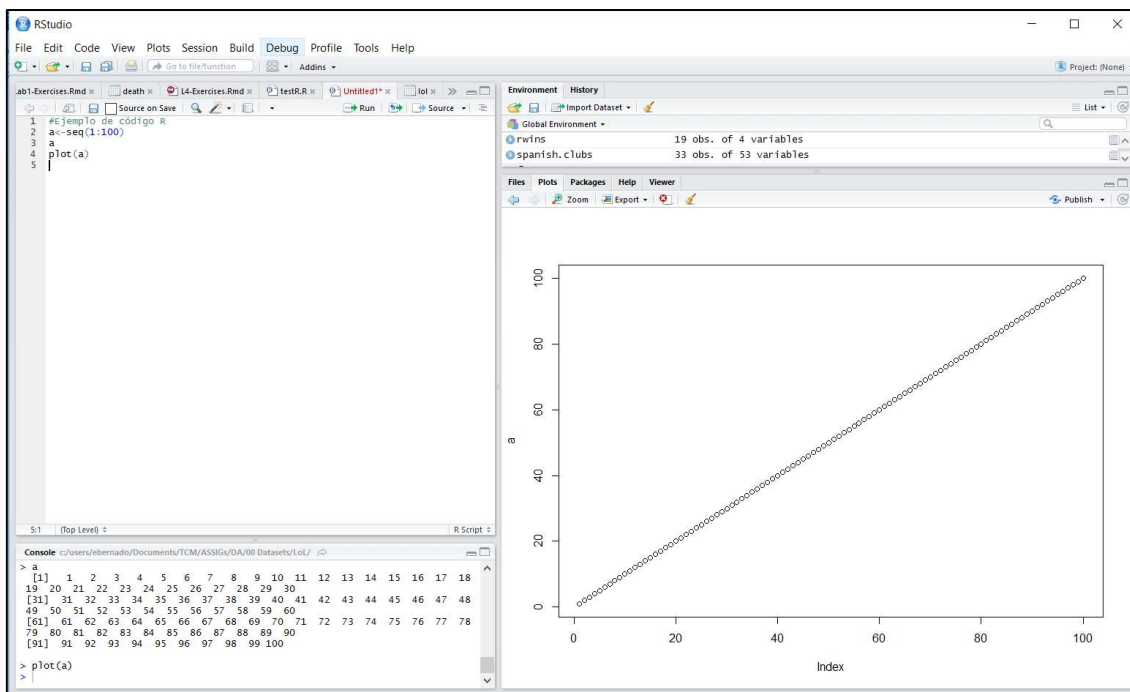
Es recomendable familiarizarse con el entorno RStudio para aprovechar su máxima potencialidad. Una de las formas básicas de trabajar con RStudio es a través de la consola, introduciendo instrucciones que se ejecutan interactivamente, de forma análoga al entorno RGui.

Otra forma muy habitual y muy cómoda es a través de la edición de ficheros .R. Un fichero .R es un fichero de texto que contiene instrucciones R que se pueden ejecutar. La ventaja de trabajar con ficheros es que se pueden guardar y recuperar para su uso posterior. Para crear un fichero R, seleccionar File>New File>R Script tal como se muestra en la figura siguiente.



El fichero por defecto lleva la etiqueta “Untitled1”. Una vez se ha creado el nuevo fichero, se pueden escribir las instrucciones como si se tratara de un fichero de texto. En el ejemplo que sigue se ejecutan tres instrucciones:

```
a<-seq(1:100)
a
plot(a)
```



El código del ejemplo contiene una línea que empieza con el símbolo “#”. Este es un comentario que el lenguaje R no procesará como instrucción en el momento en que se interprete el código.

Para interpretar el código, hay dos opciones: interpretar instrucción por instrucción o todo el código a la vez. Para interpretar instrucción por instrucción se debe seleccionar la opción “Run” disponible en el borde superior derecho de la ventana de edición. Esta ejecuta la instrucción donde se encuentra el cursor, a la vez que pasa el cursor a la instrucción siguiente para su posterior interpretación. La salida de las instrucciones sale a consola, si es de tipo texto o en el visor (Viewer) si es de tipo gráfico. Alternativamente a la instrucción Run, se puede ejecutar con las teclas Ctrl+Return.

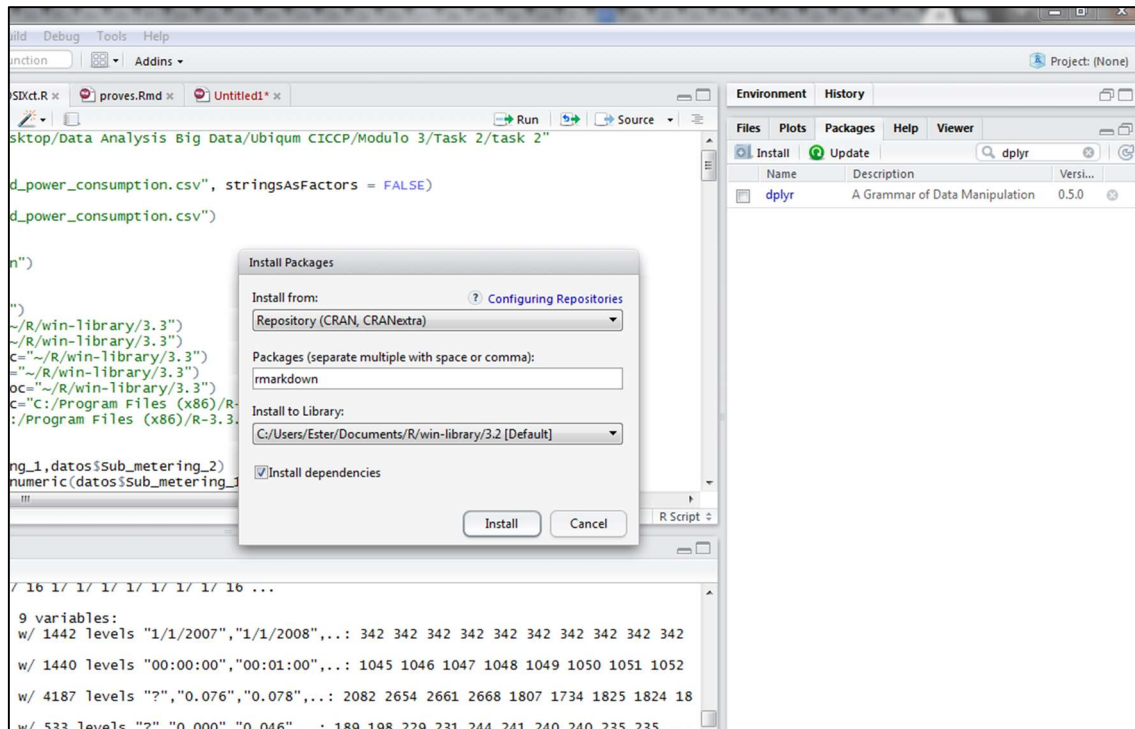
Para ejecutar el código entero, se debe seleccionar la opción “Source” que está en el borde superior del Editor, a la derecha de la opción Run.

Para guardar el fichero, se debe seleccionar la opción “File>Save As”, desde donde se podrá especificar la ruta y nombre del fichero deseados.

3.5. Descargar RMarkdown

RMarkdown es un lenguaje basado en etiquetas que nos permitirá generar documentos en múltiples formatos como HTML y PDF, donde se combina el texto con formato de presentación, el código R y el resultado de las ejecuciones del código R. RMarkdown permite crear **informes dinámicos**, es decir informes que presentan la salida de ejecutar instrucciones y a la vez con un formato de presentación definitivo. En el siguiente enlace <http://rmarkdown.rstudio.com> se puede encontrar más información.

Para instalar el RMarkdown podéis hacerlo desde la ventana de la derecha: Packages>Install. En la subventana, se indica que se desea instalar “rmarkdown”.

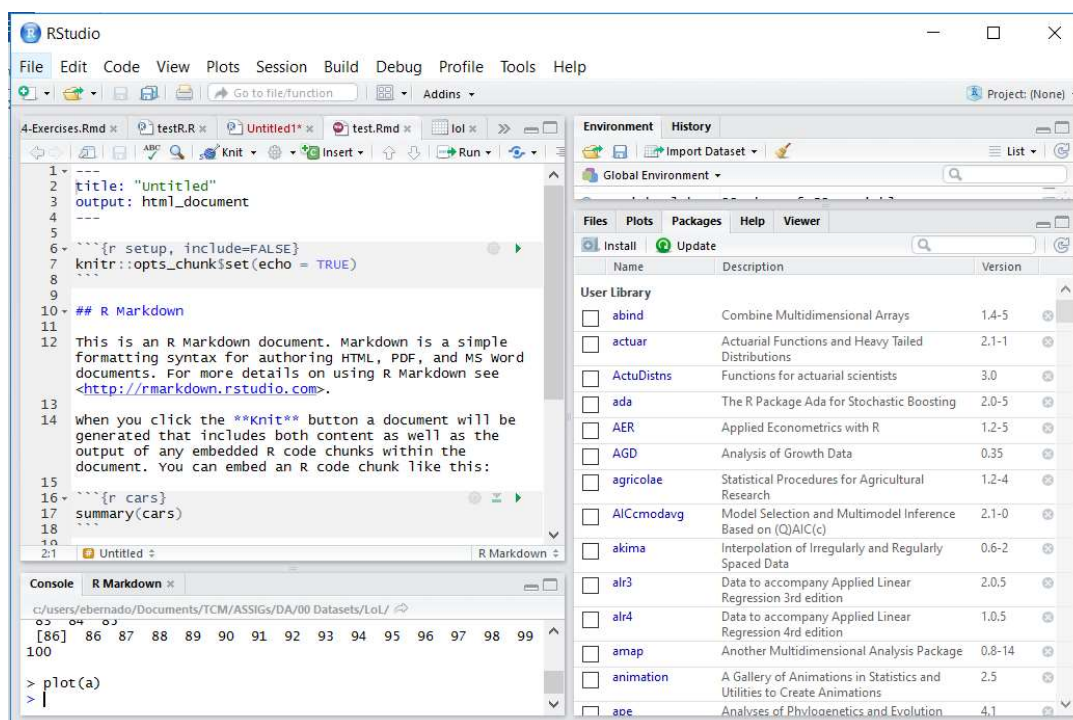


Otra forma es ejecutar la instrucción desde consola:

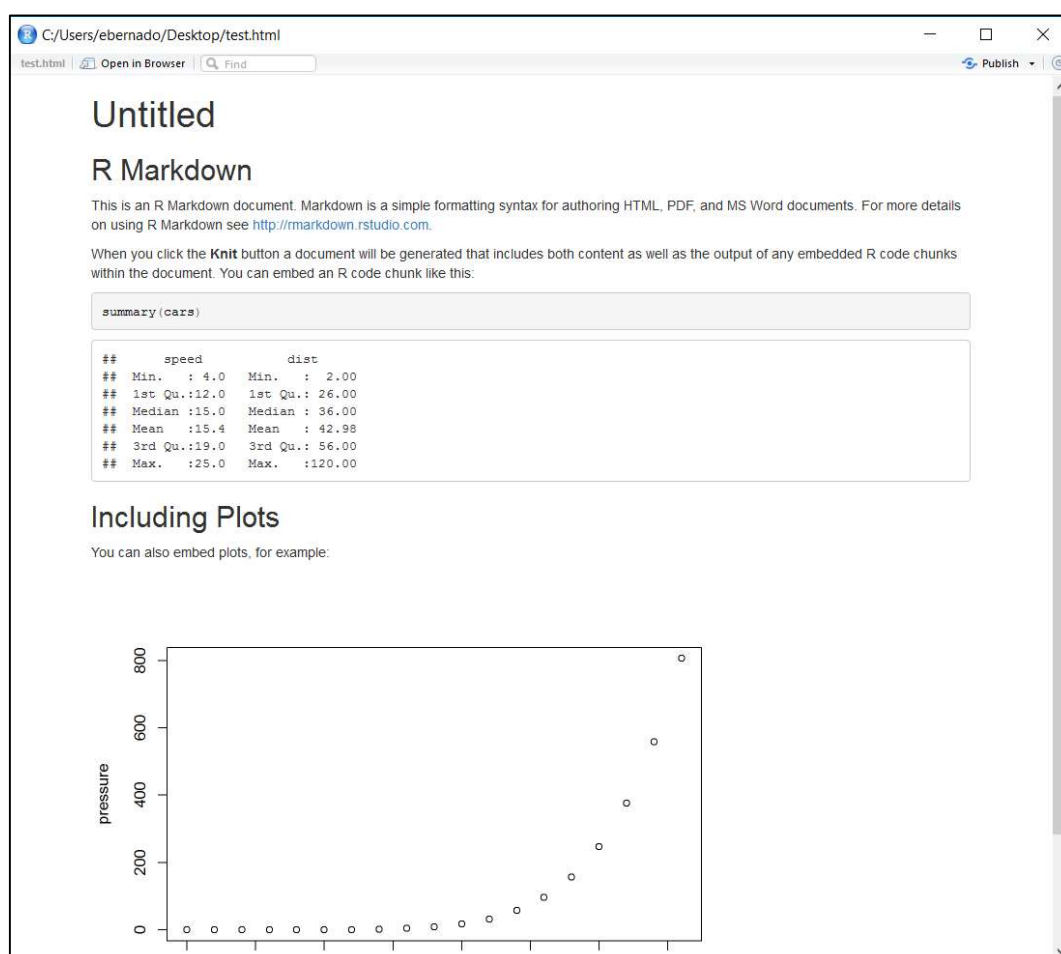
```
install.packages("rmarkdown")
```

Una vez instalado el RMarkdown, éste se integra en el entorno RStudio. Para trabajar con RMarkdown desde RStudio, se debe crear un fichero con extensión .Rmd. Para ello, se debe acceder a “File>New File>R Markdown”. Con ello, se crea un fichero que por defecto se llama Untitledx.Rmd, donde x es un número.

Al crear un fichero .Rmd se genera un contenido por defecto con algunas instrucciones ilustrativas. Para ejecutar un fichero RMarkdown, se debe seleccionar el botón **Knit** que aparece en el borde superior del editor cuando se edita un fichero Rmd, tal como se puede observar en la figura siguiente.



Al ejecutar un código Rmd, se interpreta el código R y se formatea el texto según está especificado en el fichero. Se produce un fichero de salida como la que se muestra a continuación:



4. Otras consideraciones sobre R

Para los que conozcan otros entornos de análisis de datos como SAS y SPSS existe una diferencia fundamental entre estos entornos y R. Los entornos como SAS y SPSS proporcionan un gran volumen de información a partir de la salida de los análisis estadísticos como la regresión o el análisis discriminante, por poner algunos ejemplos. En cambio, R trabaja de forma distinta. La filosofía de R es que el análisis estadístico se realiza de forma iterativa, partir de un conjunto de pasos que están detallados en las instrucciones del lenguaje. La salida de R suele ser mínima, y el resultado completo se guarda en objetos R para su posterior consulta a través de instrucciones específicas sobre estos objetos. Conocer estas diferencias de uso de los distintos softwares facilita su comprensión y su manejo óptimo.

5. Fuentes de información

Existen múltiples fuentes de información sobre R en Internet: libros de introducción a R, tutoriales, cursos y blogs sobre R. Existen también comunidades online que comparten mucha información y que pueden ser muy útiles como consulta y fuente de interacción con otros usuarios de R:

<http://cran.es.r-project.org/doc/manuals/R-intro.pdf>

<https://www.r-bloggers.com/>

<https://stackoverflow.com/questions/tagged/r>