

GUIÓN 0 de PRÁCTICAS con Rcmdr: Primeros pasos

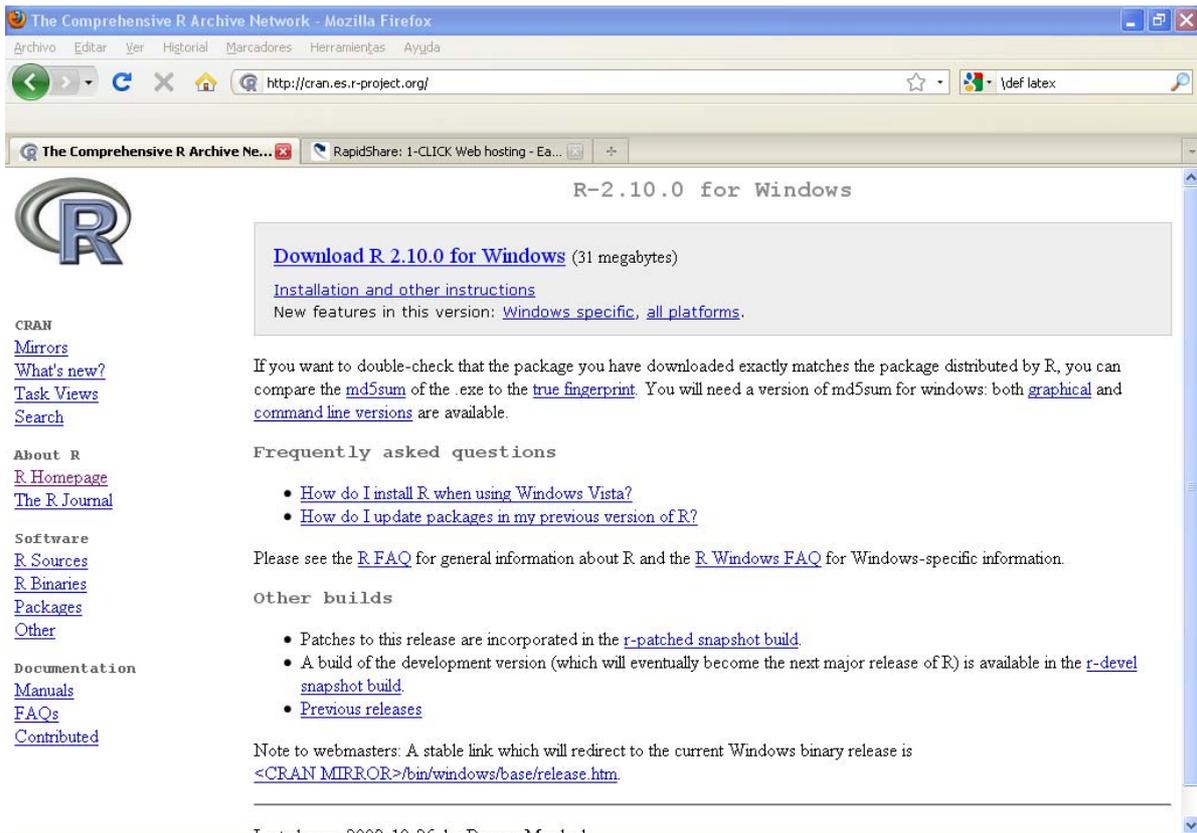
Instalación del programa: Ir a la página principal de R:

<http://www.R-project.org>

The screenshot shows the R Project website interface. On the left, there is a navigation menu with links for 'About R', 'What is R?', 'Contributors', 'Screenshots', 'What's new?', 'Download, Packages', 'CRAN', 'R Project', 'Foundation', 'Members & Donors', 'Mailing Lists', 'Bug Tracking', 'Developer Page', 'Conferences', and 'Search'. Below this is a 'Documentation' section with links for 'Manuals', 'FAQs', 'The R Journal', 'Wiki', and 'Books'. The main content area features the R logo and the title 'The R Project for Statistical Computing'. It displays several statistical plots: a PCA plot titled 'PCA 5 vars' with a biplot showing variables like Fertility, Catholic, Examination, Education, and Agriculture; a bar chart showing the distribution of groups; a dendrogram for clustering into 4 groups; and two histograms showing the distribution of Factor 1 (41%) and Factor 3 (19%). At the bottom, there is a 'Getting Started' section with two bullet points: 'R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).' and '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.'

Bajar primero el programa *R* desde un *mirror* cercano (por ejemplo desde Madrid)

<http://cran.es.r-project.org/bin/windows/base/R-2.10.1-win32.exe>



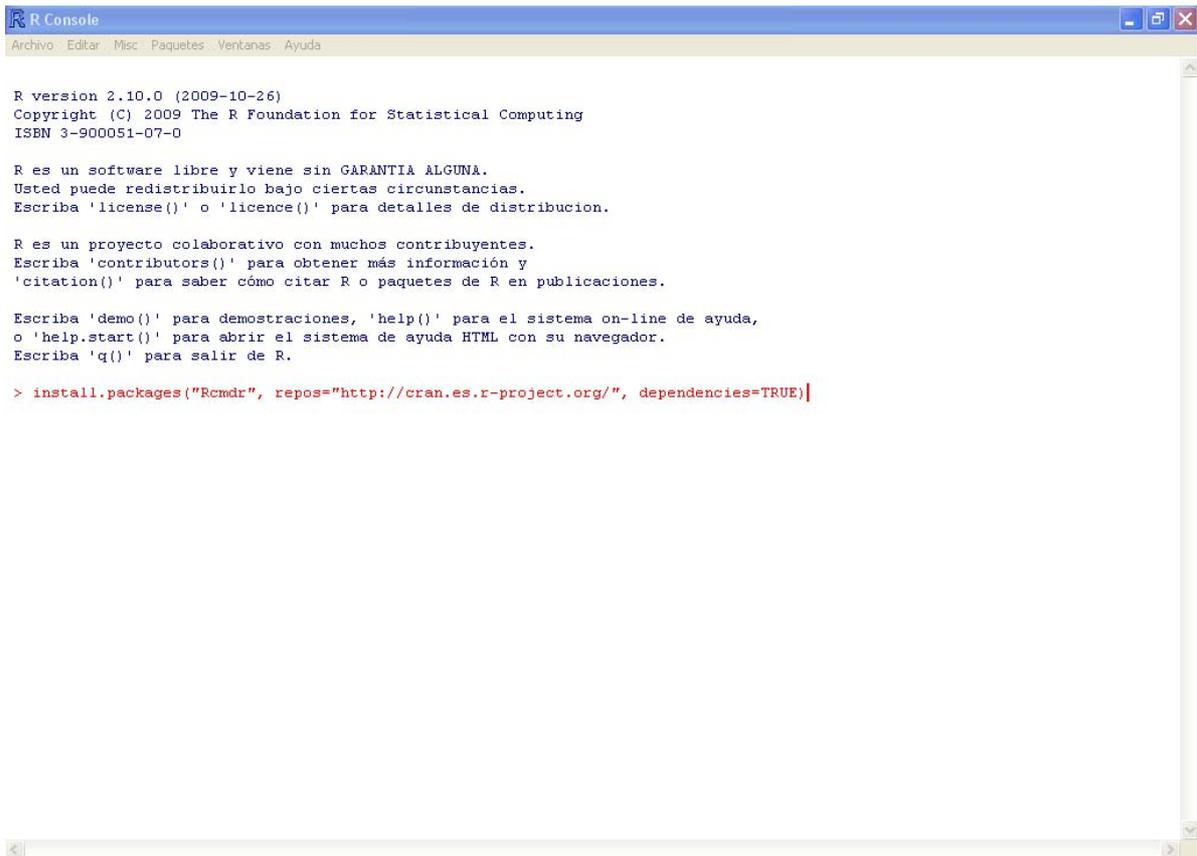
Instalar el programa *R* (haciendo click en el fichero descargado).

Desde la línea de comandos de *R*, instalar la librería *Rcmdr* con todas las librerías asociadas (dependencias):

```
install.packages("Rcmdr", repos="http://cran.es.r-project.org/",  
dependencies=TRUE)
```

O bien, a partir del menú de ventanas

Paquetes → Instalar Paquete(s)



```
R Console
Archivo  Editor  Misc  Paquetes  Ventanas  Ayuda

R version 2.10.0 (2009-10-26)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> install.packages("Rcmdr", repos="http://cran.es.r-project.org/", dependencies=TRUE)|
```

Una vez que ha arrancado *R*, se carga la librería *Rcmdr* tecleando:

library(Rcmdr)

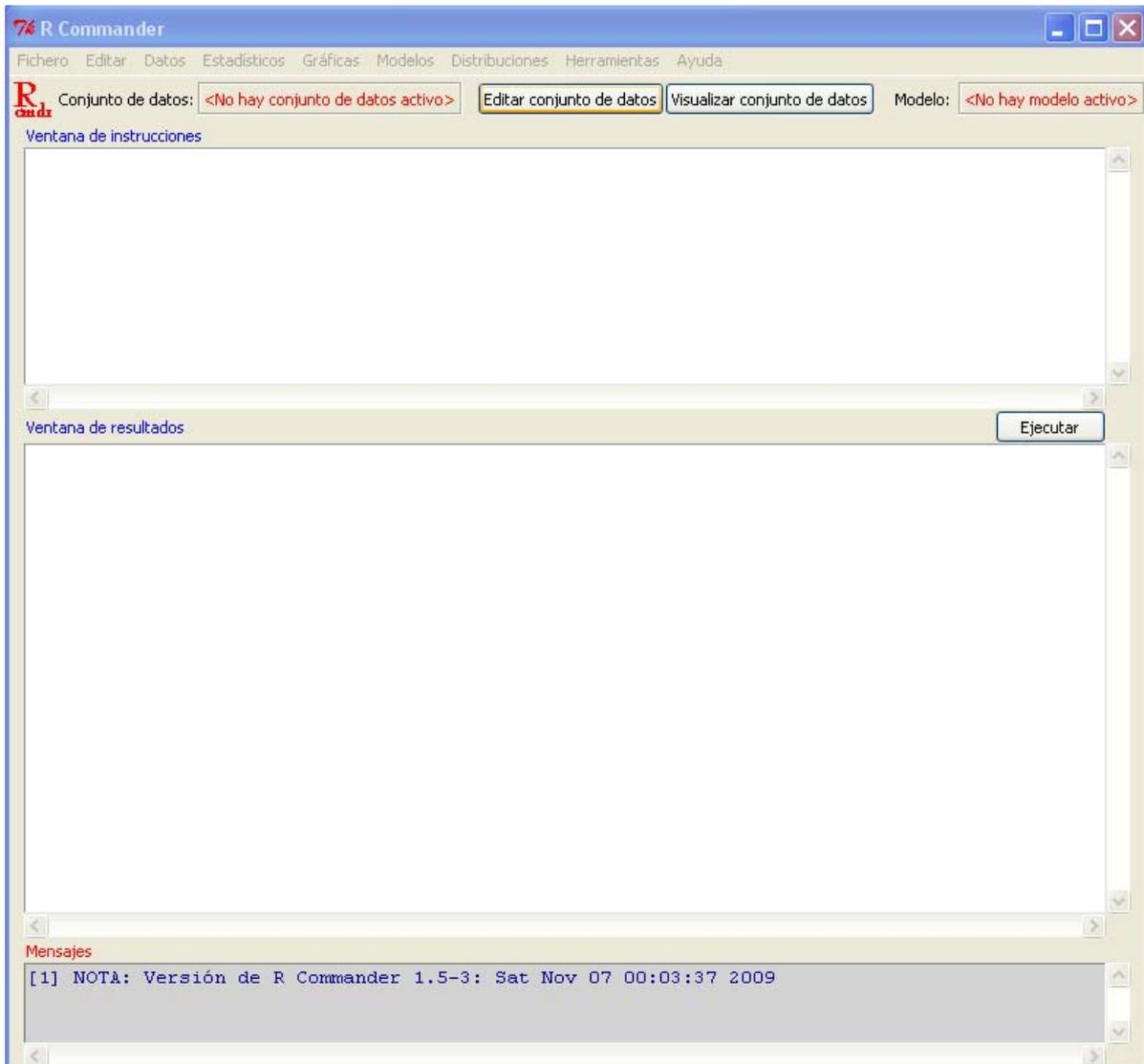
O bien, se puede hacer desde el menú superior de la ventana de *R*:

Paquetes → Cargar paquete → Rcmdr

Aparece el *R Commander graphical user interface*.

NOTA:

Si se cierra accidentalmente la ventana de *RCommander*, basta escribir en la consola (pantalla en blanco) de *R* la orden **Commander ()** y pulsar **OK**.



Ventanas Principales

- ✓ Ventana de instrucciones
- ✓ Ventana de resultados
- ✓ Mensajes

Menús Principales de *Rcmdr* :

Fichero: Menú para cargar y grabar ficheros script, salidas y el *R workspace*.

Editar: Submenús para editar los contenidos de los scripts y salidas. Pulsando con el botón derecho encima de las ventanas script y output aparece también un menú contextual.

Datos: Submenús para lectura y manipulación de datos.

Estadísticos: Submenús para realizar numerosos análisis estadísticos.

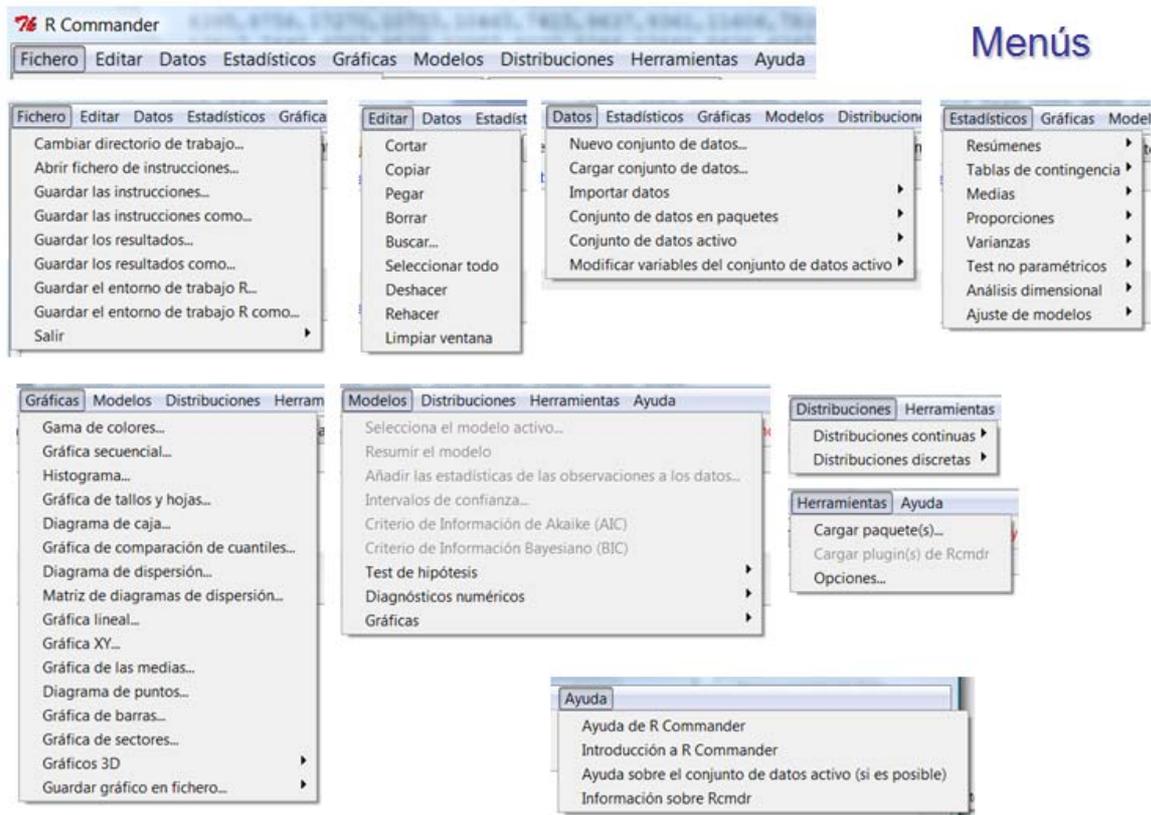
Gráficas: Submenús para realizar gráficos estadísticos simples.

Modelos: Submenús para obtener resúmenes numéricos, intervalos de confianza, tests de hipótesis y gráficos complejos.

Distribuciones: Submenús de distribuciones de probabilidad estándar.

Herramientas: Submenús para cargar otros paquetes o datos de R.

Ayuda: menú con ayuda para *Rcmdr* y un manual del mismo.



En *Rcmdr* se asume que existe un conjunto de datos activo, aunque cuando se comienza a trabajar no hay ninguno por defecto. Así, hay que cargar previamente **siempre** un conjunto de datos.

Si hay varios conjuntos de datos sólo se puede seleccionar uno de ellos.

Formas de introducir los datos

- Directamente (para pequeños conjuntos de datos) desde

Datos → Nuevo Conjunto de Datos

- Importar los datos desde un fichero ASCII o desde el portapapeles.

- Importar los datos desde un fichero tipo *Minitab*, *SPSS*, *Stata*, *Excel*, *Access* o *dBase*.

- Importar los datos incluidos en algún paquete o librería de *R*.

Leer por ejemplos los datos del fichero [Nations.txt](#)

Datos → **Importar Datos** → **Desde archivo de texto o portapapeles** → **Aceptar**

Ir al subdirectorio:

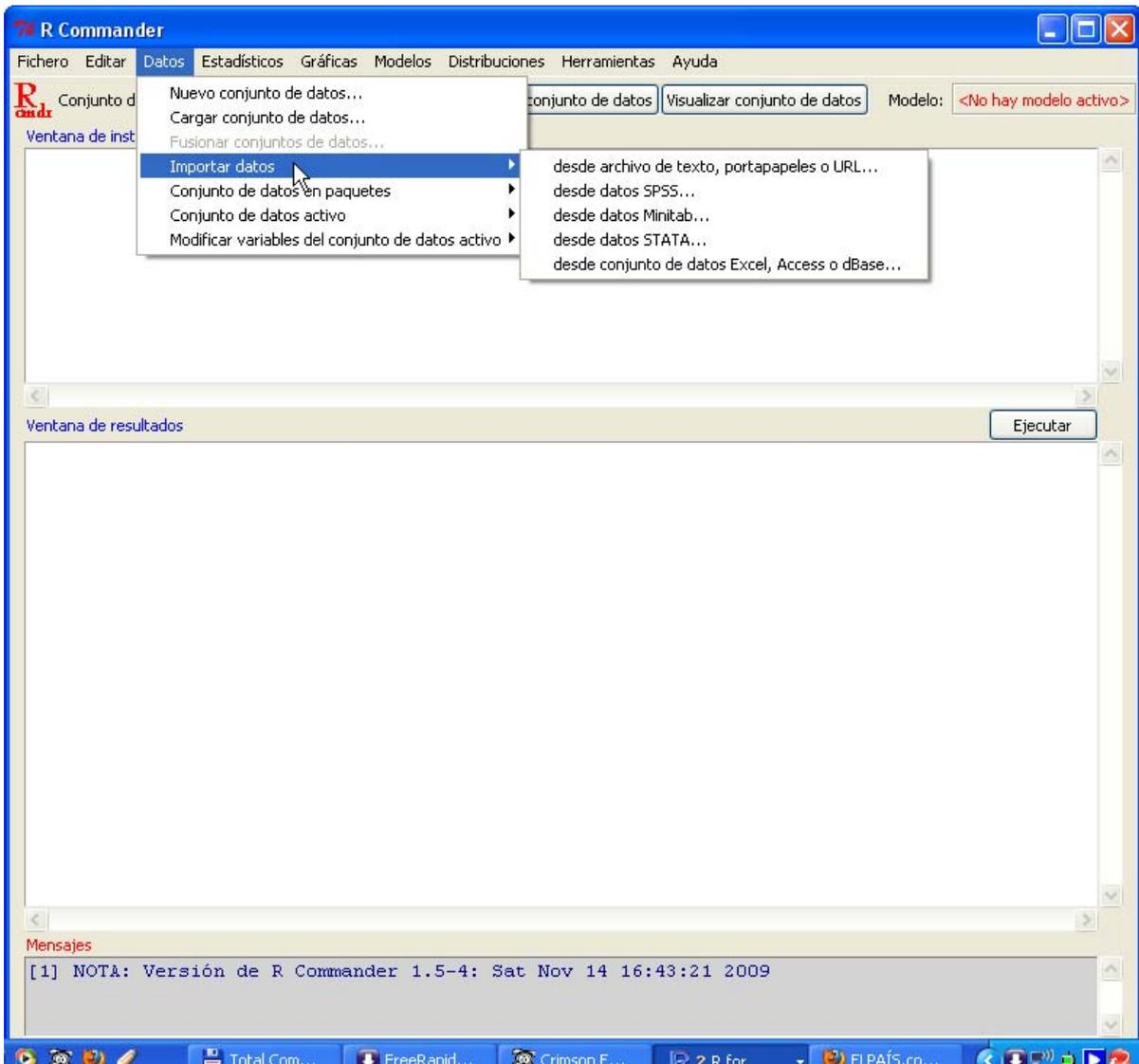
`C:/R-2.10.1/library/Rcmdr/etc/`

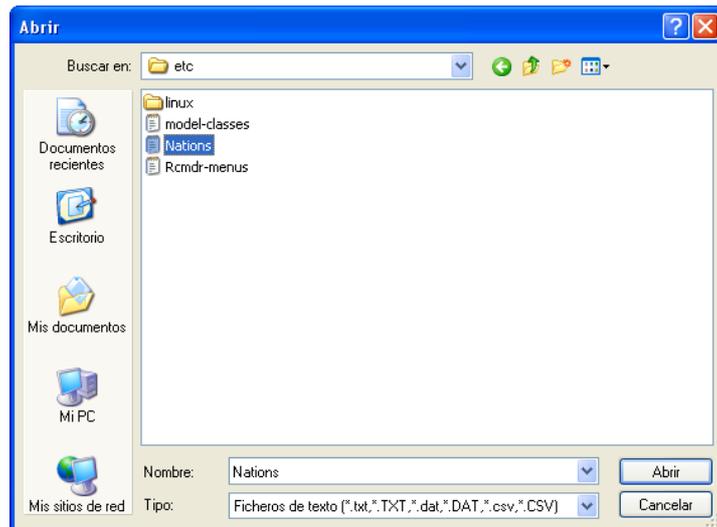
y buscar el fichero [Nations.txt](#)

NOTA:

El procedimiento anterior equivale a usar el comando básico de *R* (que se observa al mismo tiempo en la *ventana de instrucciones*):

```
Datos = read.table("C:/R-2.10.1/library/Rcmdr/etc/Nations.txt", header=TRUE,  
sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
```





Se pueden usar los botones **Editar Conjunto de Datos** o **Visualizar Conjunto de Datos**

A screenshot of the R Commander interface. The main window shows the command: `Datos <- read.table("C:/R-2.10.0/library/Rcmdr/etc/Nations.txt",`. Below this, a "Datos" window displays a table with 5 columns: TFR, contraception, infant.mortality, GDP, and region. The data is sorted by TFR in descending order. A note at the bottom states: "[2] NOTA: El conjunto de datos Datos tiene 207 filas y 5 columnas."

	TFR	contraception	infant.mortality	GDP	region
Afghanistan	6.90	NA	154	2848	Asia
Albania	2.60	NA	32	863	Europe
Algeria	3.81	52	44	1531	Africa
American-Samoa	NA	NA	11	NA	Oceania
Andorra	NA	NA	NA	NA	Europe
Angola	6.69	NA	124	355	Africa
Antigua	NA	53	24	6966	Americas
Argentina	2.62	NA	22	8055	Americas
Armenia	1.70	22	25	354	Europe
Australia	1.89	76	6	20046	Oceania
Austria	1.42	71	6	29006	Europe
Azerbaijan	2.30	17	33	321	Asia
Bahamas	1.95	62	14	12545	Americas
Bahrain	2.97	53	18	9073	Asia
Bangladesh	3.14	49	78	280	Asia
Barbados	1.73	55	9	7173	Americas
Belarus	1.40	50	15	994	Europe
Belgium	1.62	79	7	26582	Europe
Belize	3.66	47	30	2569	Americas
Benin	5.83	16	84	391	Africa
Bhutan	5.89	19	104	166	Asia
Bolivia	4.36	45	66	909	Americas
Bosnia	1.40	NA	13	271	Europe
Botswana	4.45	33	56	3640	Africa
Brazil	2.17	74	42	4510	Americas
Brunei	2.70	NA	9	16683	Asia
Bulgaria	1.45	NA	16	1518	Europe
Burkina-Faso	6.57	8	97	165	Africa
Burundi	6.28	9	114	205	Africa
Cambodia	4.50	NA	102	130	Asia

La primera línea del fichero contiene los nombres de las variables: *TFR* (tasa de fertilidad total, expresada como el número de descendientes por mujer), *contraception* (porcentaje de uso de anticonceptivos usados entre mujeres enparejadas), *infant.mortality* (tasa de mortalidad infantil por cada 1000 nacimientos), *GDP* (PIB, en dólares) y *region*.

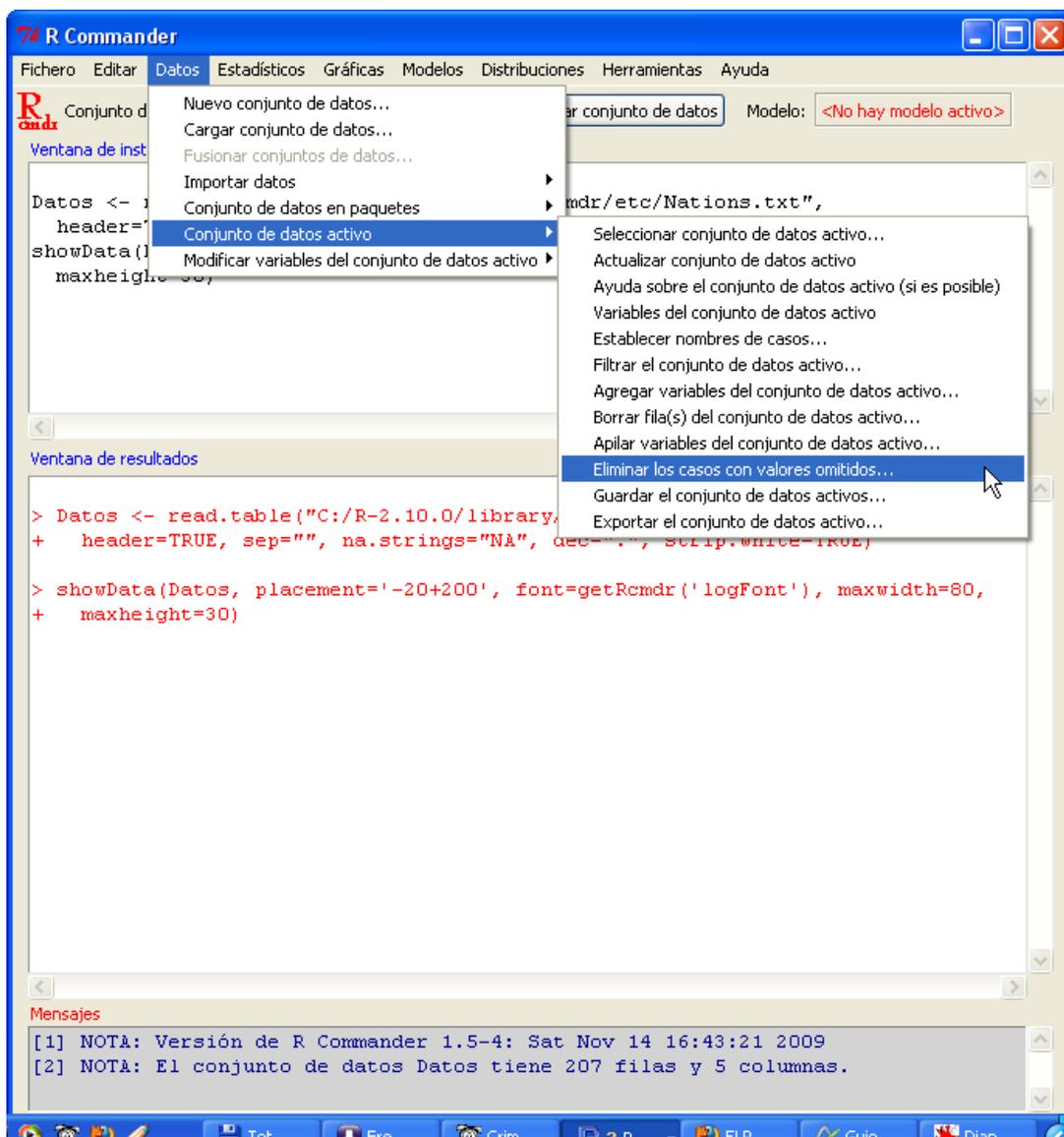
Algunos de los datos son missing. En *R*, es conveniente usar el código **NA** (representando *not available*) para codificar los datos faltantes.

Las variables *TFR*, *contraception*, *infant.mortality* y *GDP* son **numéricas** (variables cuantitativas); a su vez, *region* es textual y contiene los nombres de las regiones.

Cuando se leen los datos, *R* trata la variable *region* como un **factor**, es decir, como una variable categórica.

A continuación eliminamos los casos *missing* que tienen código NA:

Datos → Conjunto de datos activo → Eliminar los casos con valores omitidos



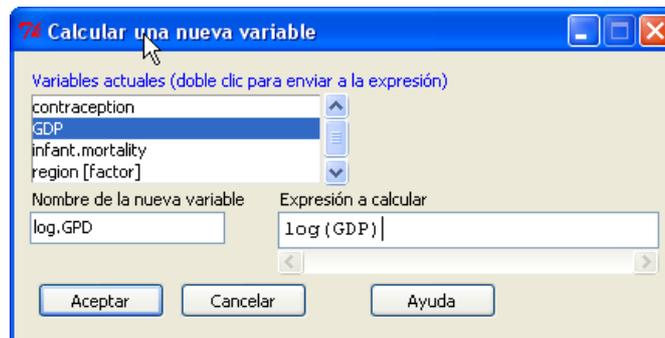
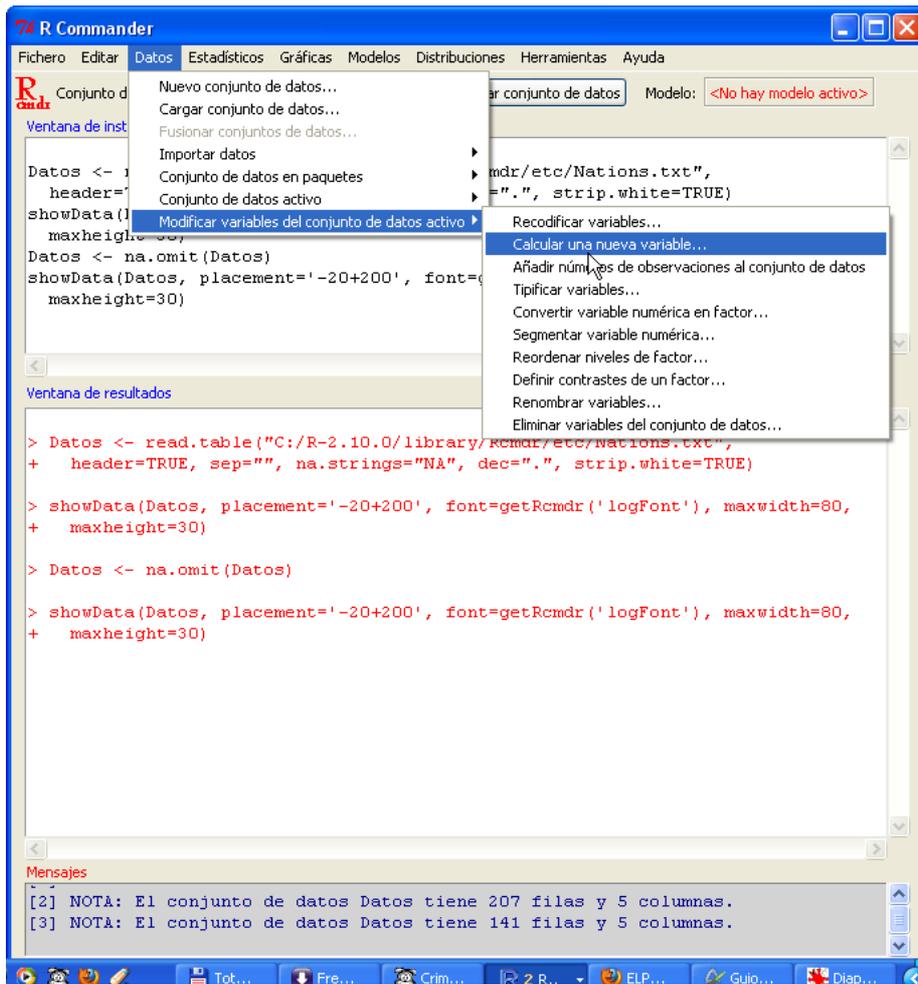


Se obtiene el mismo conjunto de datos sin las filas que tuviesen algún valor NA:

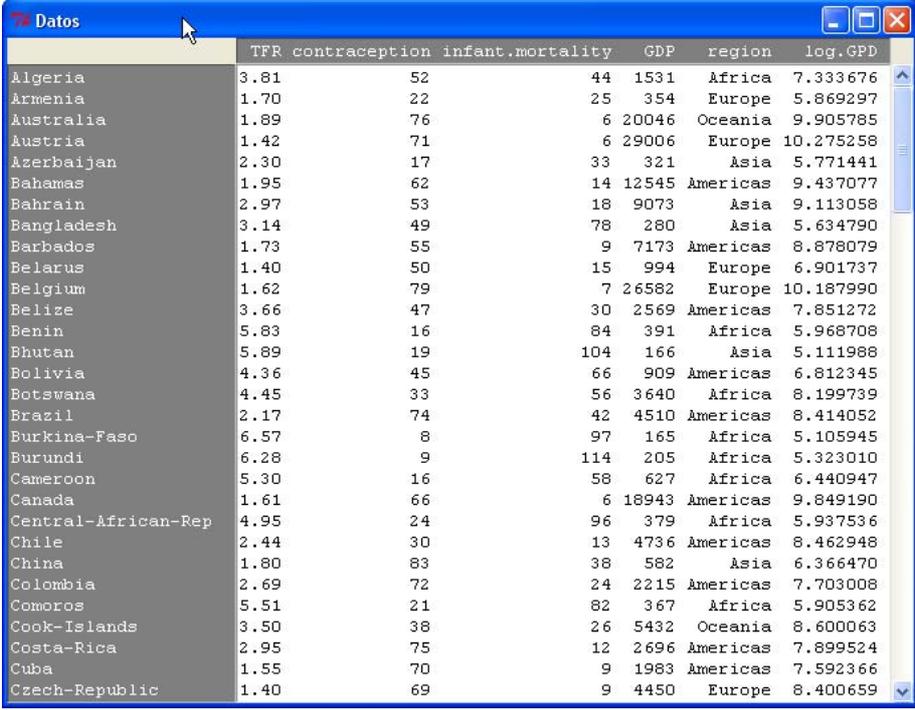
	TFR	contraception	infant.mortality	GDP	region
Algeria	3.81		52	44 1531	Africa
Armenia	1.70		22	25 354	Europe
Australia	1.89		76	6 20046	Oceania
Austria	1.42		71	6 29006	Europe
Azerbaijan	2.30		17	33 321	Asia
Bahamas	1.95		62	14 12545	Americas
Bahrain	2.97		53	18 9073	Asia
Bangladesh	3.14		49	78 280	Asia
Barbados	1.73		55	9 7173	Americas
Belarus	1.40		50	15 994	Europe
Belgium	1.62		79	7 26582	Europe
Belize	3.66		47	30 2569	Americas
Benin	5.83		16	84 391	Africa
Bhutan	5.89		19	104 166	Asia
Bolivia	4.36		45	66 909	Americas
Botswana	4.45		33	56 3640	Africa
Brazil	2.17		74	42 4510	Americas
Burkina-Faso	6.57		8	97 165	Africa
Burundi	6.28		9	114 205	Africa
Cameroon	5.30		16	58 627	Africa
Canada	1.61		66	6 18943	Americas
Central-African-Rep	4.95		24	96 379	Africa
Chile	2.44		30	13 4736	Americas
China	1.80		83	38 582	Asia
Colombia	2.69		72	24 2215	Americas
Comoros	5.51		21	82 367	Africa
Cook-Islands	3.50		38	26 5432	Oceania
Costa-Rica	2.95		75	12 2696	Americas
Cuba	1.55		70	9 1983	Americas
Czech-Republic	1.40		69	9 4450	Europe

Se pueden calcular nuevas variables a partir de las existentes.

Datos → Modificar Variables del Conjunto Activo → Calcular una Nueva Variable



Si visualizamos el nuevo conjunto de datos, se obtiene



	TFR	contraception	infant.mortality	GDP	region	log.GPD
Algeria	3.81		52	44 1531	Africa	7.333676
Armenia	1.70		22	25 354	Europe	5.869297
Australia	1.89		76	6 20046	Oceania	9.905785
Austria	1.42		71	6 29006	Europe	10.275258
Azerbaijan	2.30		17	33 321	Asia	5.771441
Bahamas	1.95		62	14 12545	Americas	9.437077
Bahrain	2.97		53	18 9073	Asia	9.113058
Bangladesh	3.14		49	78 280	Asia	5.634790
Barbados	1.73		55	9 7173	Americas	8.878079
Belarus	1.40		50	15 994	Europe	6.901737
Belgium	1.62		79	7 26582	Europe	10.187990
Belize	3.66		47	30 2569	Americas	7.851272
Benin	5.83		16	84 391	Africa	5.968708
Bhutan	5.89		19	104 166	Asia	5.111988
Bolivia	4.36		45	66 909	Americas	6.812345
Botswana	4.45		33	56 3640	Africa	8.199739
Brazil	2.17		74	42 4510	Americas	8.414052
Burkina-Faso	6.57		8	97 165	Africa	5.105945
Burundi	6.28		9	114 205	Africa	5.323010
Cameroon	5.30		16	58 627	Africa	6.440947
Canada	1.61		66	6 18943	Americas	9.849190
Central-African-Rep	4.95		24	96 379	Africa	5.937536
Chile	2.44		30	13 4736	Americas	8.462948
China	1.80		83	38 582	Asia	6.366470
Colombia	2.69		72	24 2215	Americas	7.703008
Comoros	5.51		21	82 367	Africa	5.905362
Cook-Islands	3.50		38	26 5432	Oceania	8.600063
Costa-Rica	2.95		75	12 2696	Americas	7.899524
Cuba	1.55		70	9 1983	Americas	7.592366
Czech-Republic	1.40		69	9 4450	Europe	8.400659

Introducción manual de nuevos datos en R/R Commander

Ejemplo: Se quiere introducir estos tres vectores:

X: 11, 12, 13, 16, 16, 17, 18, 21

Y: 14, 15, 15, 15, 16, 16, 16, 17

Z: 11, 11, 11, 12, 19, 20, 20, 20

Luego se quiere juntarlos en un solo conjunto (*matriz de datos*), con cada uno de los vectores en cada columna.

Se escribe en la *ventana de instrucciones* lo siguiente:

```
X = c(11,12,13,16,16,17,18,21)
```

```
Y = c(14,15,15,15,16,16,16,17)
```

```
Z = c(11,11,11,12,19,20,20,20)
```

```
data.xyz = data.frame(X,Y,Z)
```

luego se marca todo el párrafo y se pulsa al botón

A rectangular button with a light blue border and the text "Ejecutar" in the center.

NOTA:

La opción de

Datos → Nuevo Conjunto de Datos

No es muy recomendable: a menudo se bloquea por falta de memoria *RAM*.