

Laboratorio 0: Introducción a R y RCommander

Instalación

Primero es necesario instalar el programa R que se puede descargar libremente. En Madrid, el *mirror* más cercano es:

<http://cran.es.r-project.org/bin/windows/base>

Se baja el programa *.exe correspondiente (ejemplo: R-2.8.1-win32.exe) y se instala el programa aceptando las opciones que pide por defecto.

Después se instala el entorno de ventanas de R, denominado RCommander (Rcmdr).

Se puede hacer de dos maneras posibles:

1. 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)
```

2. Desde el menú superior de ventanas de R: **Paquetes** → **Instalar paquetes** → **Rcmdr**

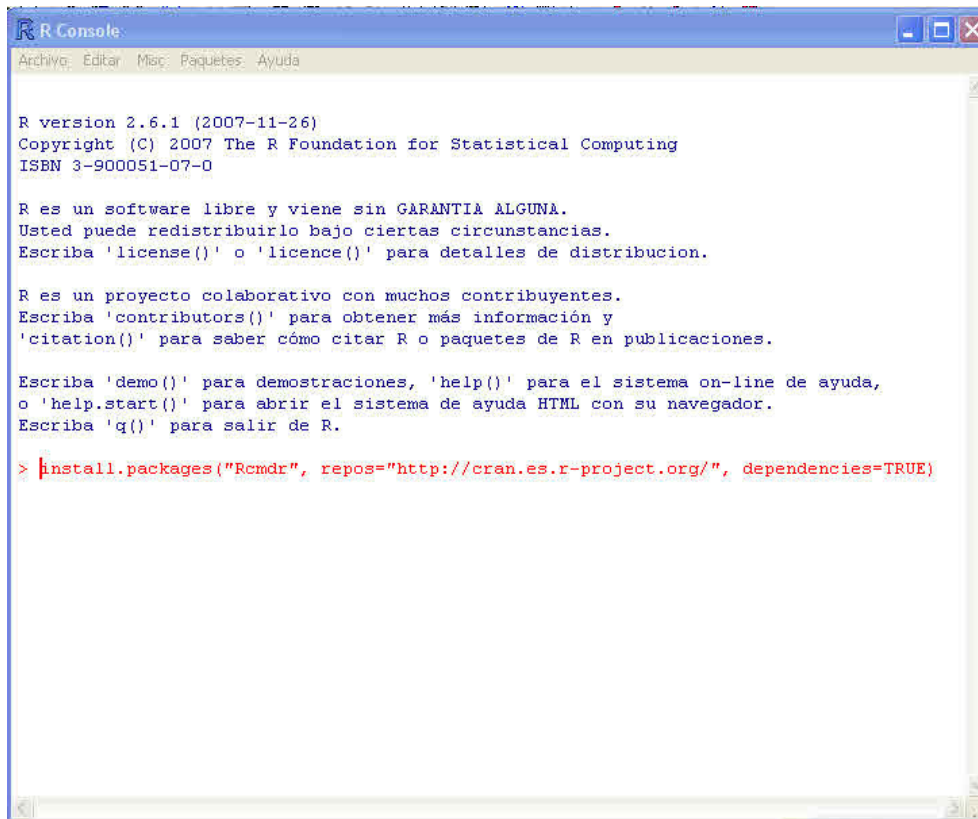


Figura 1: Ventana principal de R

Trabajo con RCommander

Una vez que ha arrancado R, para arrancar RCommander se *carga* primero la librería Rcmdr tecleando:

```
library(Rcmdr)
```

Alternativamente se puede hacer desde el menú superior de ventanas de R:

Paquetes → **Cargar paquete** → **Rcmdr**

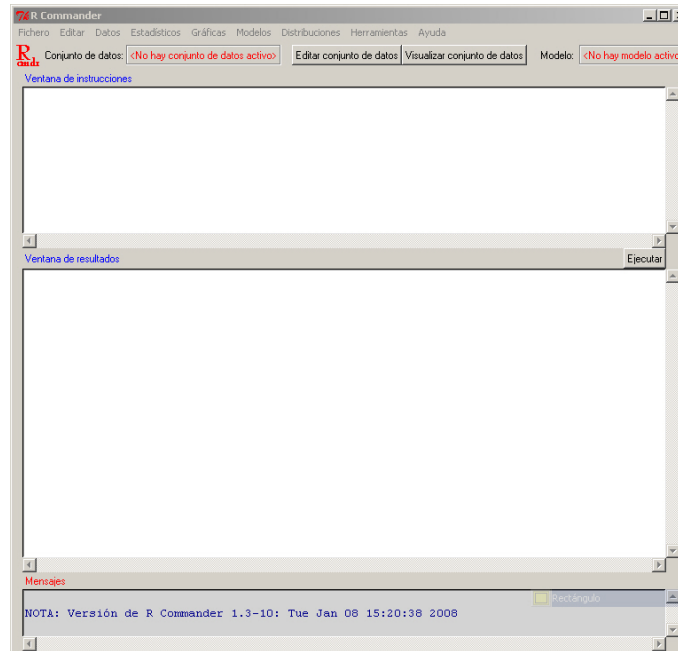
Aparece el *R Commander graphical user interface* (GUI).

Se obtiene un entorno de ventanas con tres partes:

Script Window

Output Window

Messages Window



Los menús principales de RCommander son:

- **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*, y cuando comienza el programa no se asume **ninguno** por defecto por lo que lo primero que hay que hacer es cargar un conjunto de datos para poder trabajar.

Si hay varios conjuntos de datos en memoria, sólo se puede seleccionar uno al mismo tiempo para trabajar con él.

Formas de introducir los datos:

- Directamente escribiéndolos en el programa (para pequeños conjuntos de datos) usando

Datos → Nuevo Conjunto de Datos

- Importando los datos desde un fichero ASCII o desde el *portapapeles*.
- Importando los datos desde un fichero de otros programas: Minitab, SPSS, Stata, Excel, Access ó dBase.
- Importar los datos incluidos en algún paquete o librería propia de R.

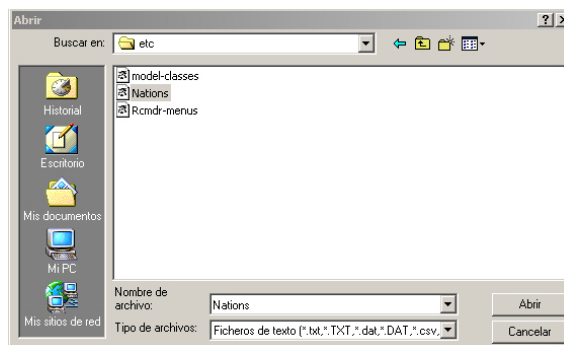
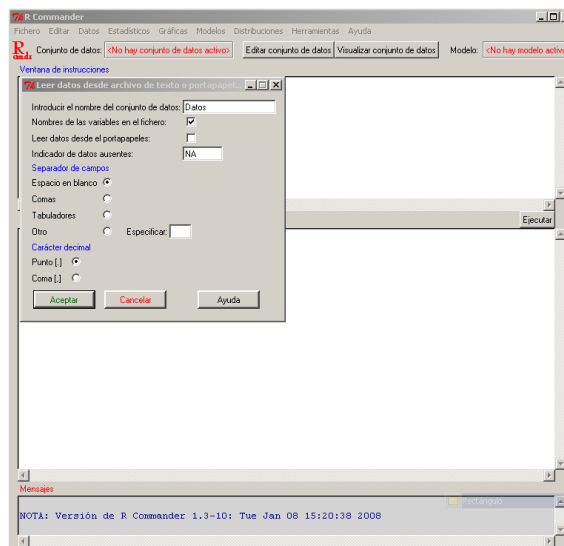
Ejemplo

Leer, por ejemplo, los datos del fichero `Nations.txt`.

Datos → **Importar Datos** → **Desde archivo de texto o portapapeles** → **Directorio** (donde esté guardado el fichero `Nations.txt`)

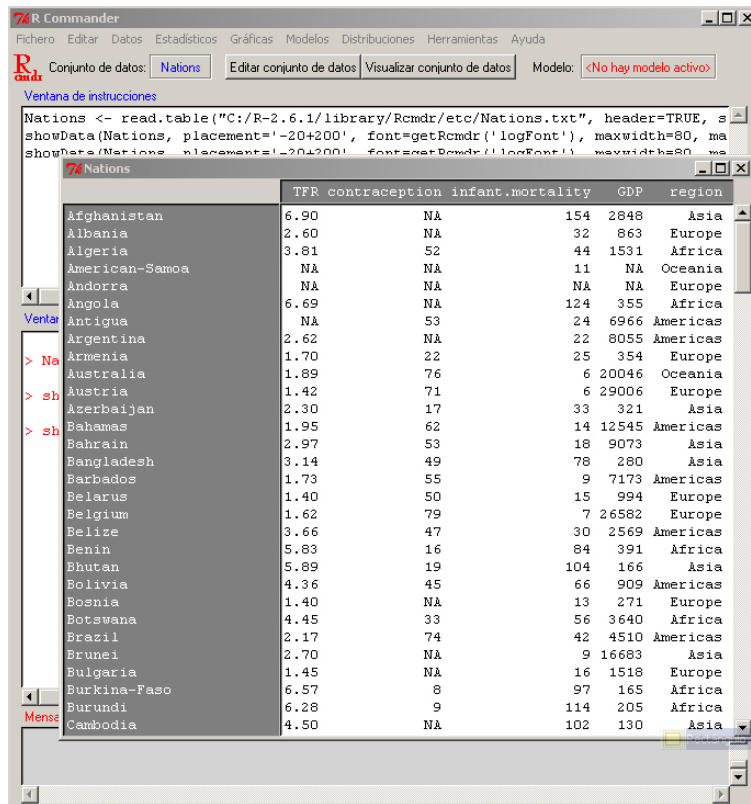
Se puede descargar de la web de la asignatura, o encontrar en el propio directorio de Rcmdr:

```
c:\R-2.8.1\library\Rcmdr\etc\Nations.txt
```



Para ver los datos que se acaban de importar, se pueden usar los botones

Editar Conjunto de Datos o **Visualizar Conjunto de Datos**.



```
Nations <- read.table("C:/R-2.6.1/library/Rcmdr/etc/Nations.txt", header=TRUE, showData(Nations, placement="-20+200", font=getRcmdr('logFont'), maxwidth=80, showData(Nations, placement="-20+200", font=getRcmdr('logFont'), maxwidth=80, ma
```

	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 contraceptivos 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.

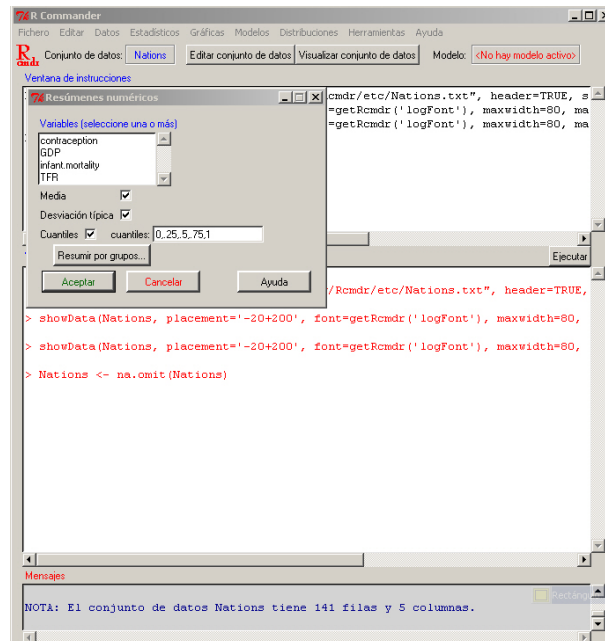
Las variables **TFR**, **contraception**, **infant.mortality** y **GDP** son numéricas (variables cuantitativas); a su vez, **region** es nominal y contiene los nombres de las regiones.

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

Ejemplo

Para eliminar los datos missing del fichero de datos activo, se hace:

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

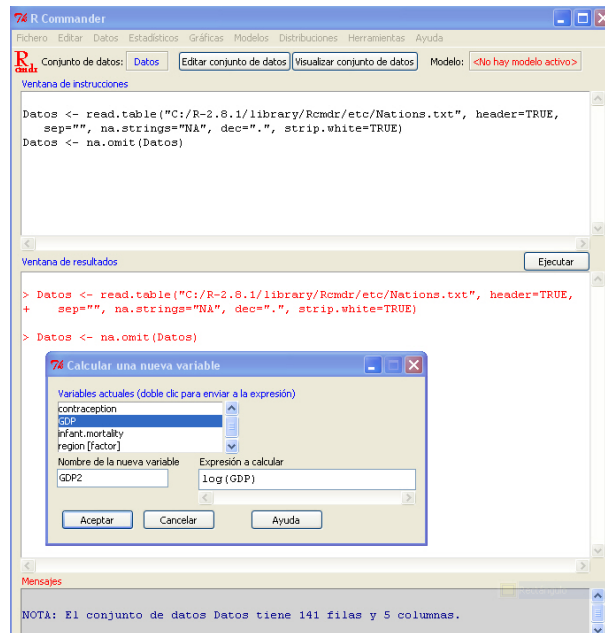


Ejemplo

Calcular nuevas variables a partir de las existentes:

Datos → Modificar variables del conjunto activo → Calcular una nueva variable

Calculamos, por ejemplo, el logaritmo de la variable GPD:



Páginas sobre R y aplicaciones en RCommander

<http://socserv.mcmaster.ca/jfox/Misc/Rcmdr>

<http://knuth.uca.es/repos/R-contribuciones/Commander-es.pdf>

<http://cran.es.r-project.org/other-docs.html>