

The R logo is a large, stylized graphic in the background, composed of several blue and yellow rectangular blocks arranged in a 3D-like structure. It is positioned behind the main title text.

ANÁLISIS DE DATOS CON R



contáctenos: enei@inei.gob.pe / 433-3127



Pasaje Hernán Velarde 285 Lima.

Entre la cuadra 01 y 02 de la Av. Arequipa.

Correo: enei@inei.gob.pe / campusvirtual@inei.gob.pe

Teléfonos: 433-3127 - 332-4650

Centro Andino de Formación y Capacitación en Estadística

Cursos Especializados en Estadística e Informática

CURSO BASICO DE “R”



Algunas recomendaciones para escribir en lenguaje R

- a. Al escribir instrucciones podemos hacer comentarios, utilizando el símbolo # al comienzo de la oración.
 - b. El nombre de un objeto debe empezar con una letra y puede incluir letras y números.
 - c. R discrimina entre mayúsculas y minúsculas.
 - d. Las funciones se escriben seguidas de paréntesis.
 - e. Los corchetes se usan para referir posiciones.
 - f. Para programar en R, necesitamos conocer el uso de sus funciones, el tipo objeto al que puede aplicárseles, sus argumentos, así como los valores que requieren ser definidos.
- Podemos acceder a toda esta información a través de diversos procedimientos de solicitud de ayuda en R .

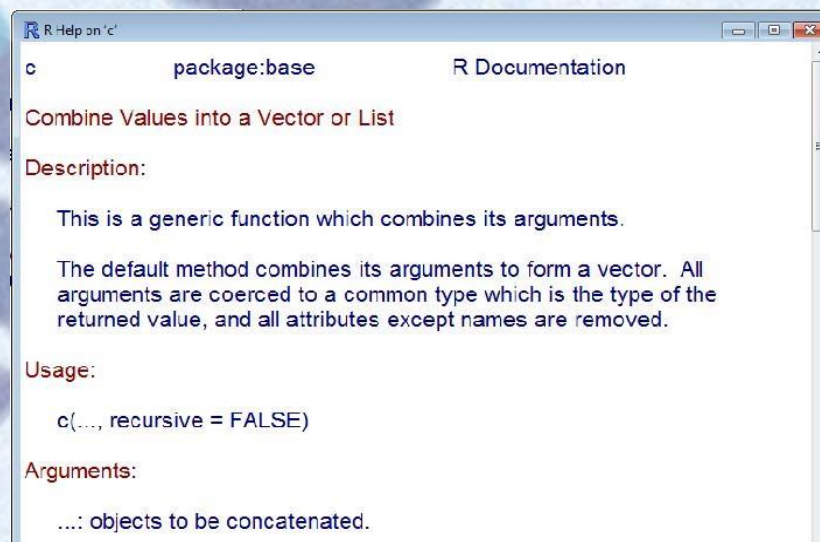
Solicitar ayuda

Hay diversas maneras de solicitar ayuda para escribir en lenguaje R, aquí describiremos tres:

- Si conocemos el nombre de la función

Escribimos en la consola el nombre de esta, precedido por el signo de cierre de interrogación (?). Como resultado R nos mostrará una descripción de la función, la forma de su uso, los argumentos, detalles, valores, funciones asociadas y ejemplos.

```
> ?c  
> |
```



Solicitar ejemplos de la manera en que se emplea una función

```
> example(mean)

mean> x <- c(0:10, 50)

mean> xm <- mean(x)

mean> c(xm, mean(x, trim = 0.10))
[1] 8.75 5.50
```


IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

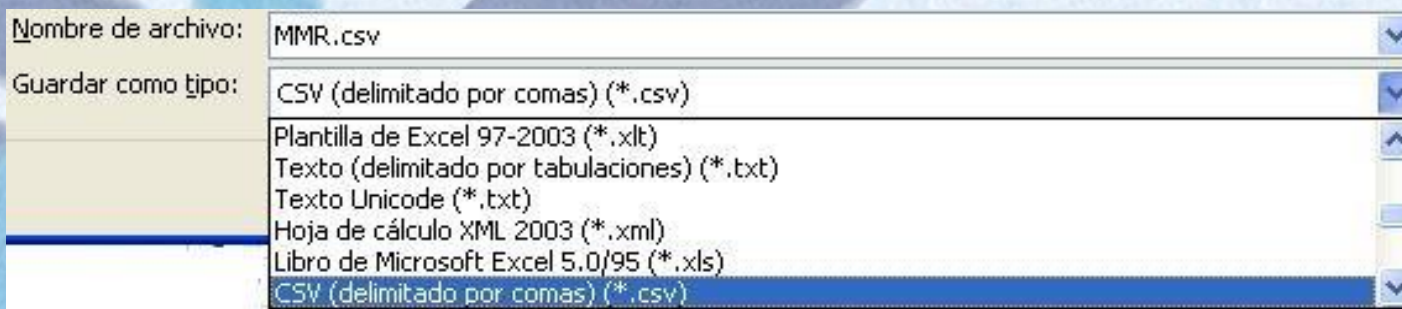
R puede leer archivos externos como marcos de datos (data frame). En esta parte leeremos archivos con extensión *.txt, *.csv, y *.dat, usando el paquete básico instalado en R.

Las bases de datos se encuentran archivadas en formato SPSS, STATA o Excel, lo primero que haremos será preparar estos archivos para que puedan ser leídos por el paquete básico de R. Aunque debemos mencionar que R cuenta con paquetes que se pueden descargar para leer directamente los archivos con extensión *.sav y *.xls o *.xlsx.

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

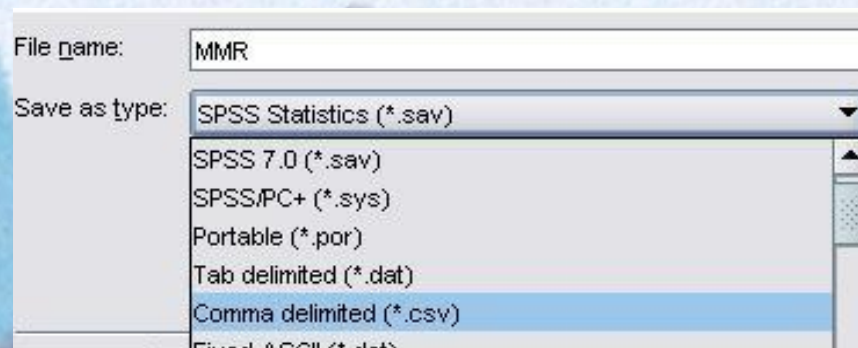
Preparar archivos externos que puedan ser leídos por el paquete básico de R

Para transportar archivos creados en Excel o en SPSS que puedan ser leídos por el paquete básico de R, primero seleccionamos la opción **Guardar como/Save as**; luego abrimos la pestaña **Guardar como tipo/Save as type**; y finalmente seleccionamos un formato. Los formatos disponibles desde Excel son: *.csv o *.txt:



IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Y los formatos disponibles desde SPSS son: *.csv o *.dat.



Los archivos resultantes serán:



IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Leer archivos desde R

Leer archivos con la función `read.table()` Leer

archivos delimitados por comas (*.csv)

```
> Salud=read.table(file="Enaho2013400p0.csv",header=TRUE,sep=",")
> salud
Error: objeto 'salud' no encontrado
> Salud
```

Donde:

Salud es un objeto de tipo marco de datos para R.

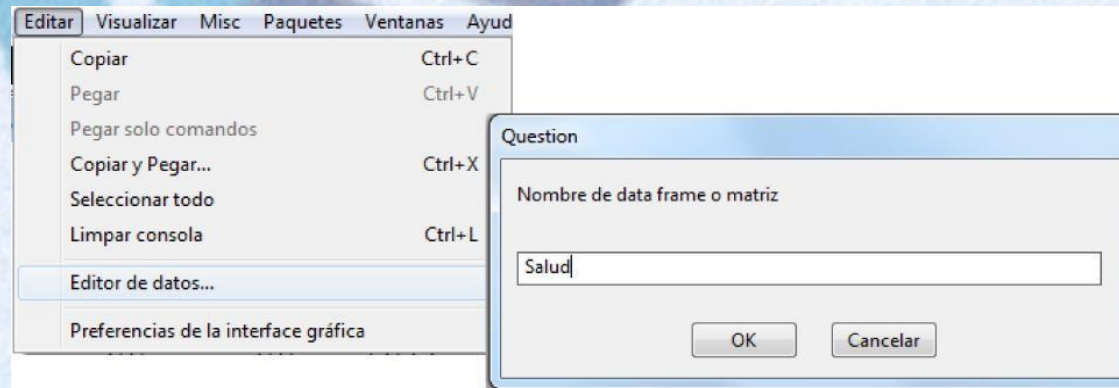
`read.table` es la función que nos permite leer el archivo `Enaho2013400p0.csv` desde R

Header es un argumento de la función `read table` que lee la primera fila del archivo `Enaho2013400p0.csv`, como una fila que contiene los nombres de las variables de la base de datos

Sep indica el elemento que actúa como separador de los datos en este caso, la coma

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

```
> Salud
  CONGLOME VIVIENDA HOGAR CODPERSONA UBIGEO DOMINIO ESTRATO CODINFOR P400N P400I
1     2063      46    11      3 150136      8      1      2      3      2
2     6062       1    11      2  70106      8      1      1      2      1
3     6062     111    11      3  70106      8      1      2      3      2
4     2039      41    11      8 150135      8      1      2      8      2
5     6045     134    11      8  70106      8      1      4      8      4
6     7559      10    11      6 150110      8      1      6      6      6
7     1736      34    11      5  70106      8      1      2      5      2
8     1738      21    11      5  70106      8      1      5      5      5
9     1741      50    11      4  70106      8      1      4      4      4
10    1848      87    11      3 150111      8      1      2      3      2
11    1932       1    11      6 150123      8      1      3      6      3
12    1979      68    11      3 150132      8      1      3      3      3
13    3552      41    11      4  70106      8      1      2      4      2
14    7565     171    11      5 150110      8      1      5      5      5
15    7565     171    11      6 150110      8      1      2      6      2
16    7624      80    11      8 150132      8      1      7      8      7
17    7628      51    11      4 150132      8      1      1      4      1
18    7803      85    11      8 150115      8      1      1      8      1
19    7926      57    11      3 150142      8      1      3      3      3
20    1772      10    11      6 150103      8      1      6      6      6
21    1848      87    11      4 150111      8      1      2      4      2
```



IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

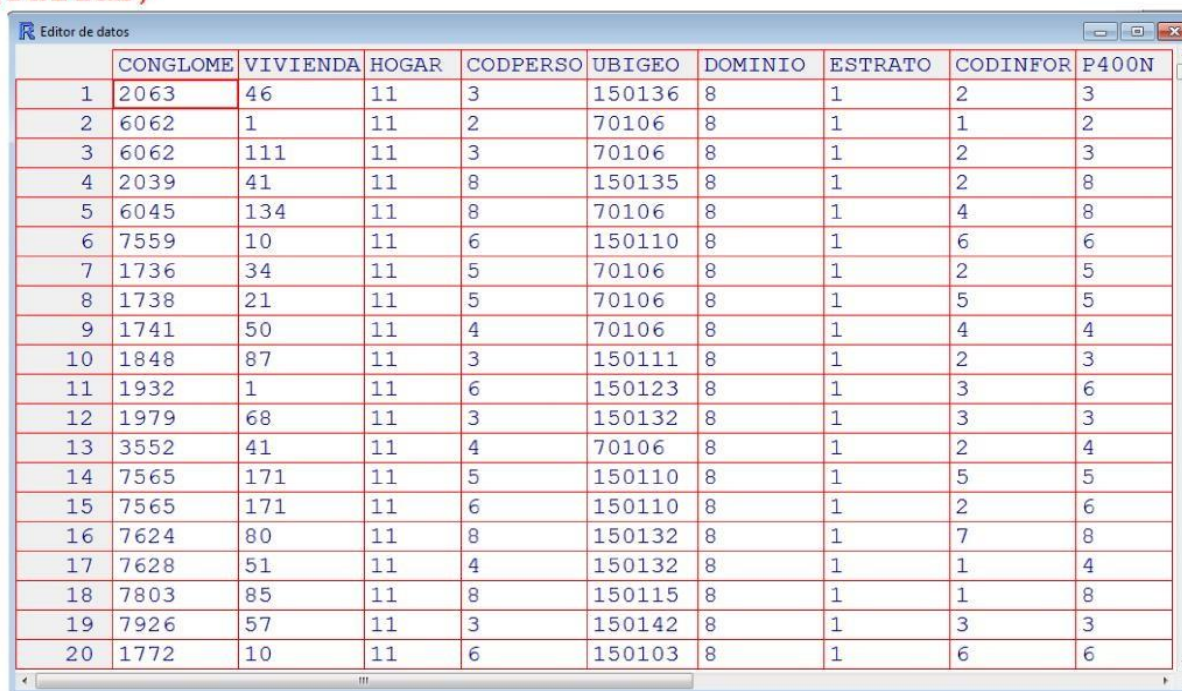
R Editor de datos

	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	UBIGEO	DOMINIO	ESTRATO	CODINFOR	P400N
1	2063	46	11	3	150136	8	1	2	3
2	6062	1	11	2	70106	8	1	1	2
3	6062	111	11	3	70106	8	1	2	3
4	2039	41	11	8	150135	8	1	2	8
5	6045	134	11	8	70106	8	1	4	8
6	7559	10	11	6	150110	8	1	6	6
7	1736	34	11	5	70106	8	1	2	5
8	1738	21	11	5	70106	8	1	5	5
9	1741	50	11	4	70106	8	1	4	4
10	1848	87	11	3	150111	8	1	2	3
11	1932	1	11	6	150123	8	1	3	6
12	1979	68	11	3	150132	8	1	3	3
13	3552	41	11	4	70106	8	1	2	4
14	7565	171	11	5	150110	8	1	5	5
15	7565	171	11	6	150110	8	1	2	6
16	7624	80	11	8	150132	8	1	7	8
17	7628	51	11	4	150132	8	1	1	4
18	7803	85	11	8	150115	8	1	1	8
19	7926	57	11	3	150142	8	1	3	3
20	1772	10	11	6	150103	8	1	6	6

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Leer archivos delimitados por tabulaciones (*.txt ó *.dat)

```
> Salud2=read.table(file="Enaho2013400d.dat",header=TRUE,sep="\t")
> fix(Salud2)
> |
```



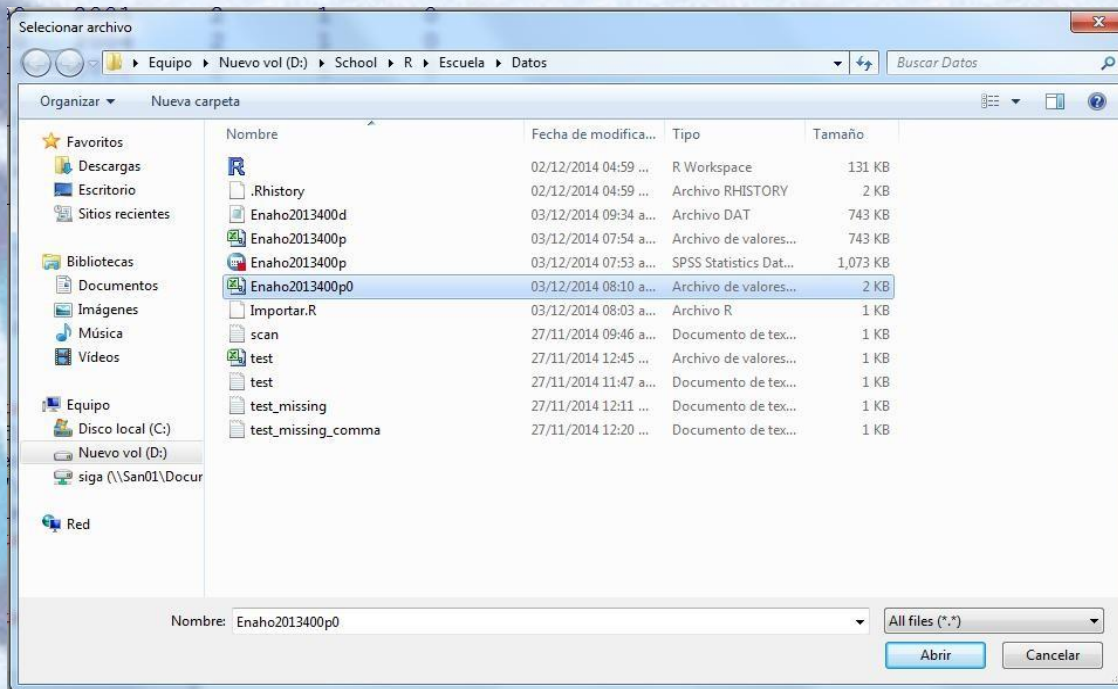
	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	UBIGEO	DOMINIO	ESTRATO	CODINFOR	P400N
1	2063	46	11	3	150136	8	1	2	3
2	6062	1	11	2	70106	8	1	1	2
3	6062	111	11	3	70106	8	1	2	3
4	2039	41	11	8	150135	8	1	2	8
5	6045	134	11	8	70106	8	1	4	8
6	7559	10	11	6	150110	8	1	6	6
7	1736	34	11	5	70106	8	1	2	5
8	1738	21	11	5	70106	8	1	5	5
9	1741	50	11	4	70106	8	1	4	4
10	1848	87	11	3	150111	8	1	2	3
11	1932	1	11	6	150123	8	1	3	6
12	1979	68	11	3	150132	8	1	3	3
13	3552	41	11	4	70106	8	1	2	4
14	7565	171	11	5	150110	8	1	5	5
15	7565	171	11	6	150110	8	1	2	6
16	7624	80	11	8	150132	8	1	7	8
17	7628	51	11	4	150132	8	1	1	4
18	7803	85	11	8	150115	8	1	1	8
19	7926	57	11	3	150142	8	1	3	3
20	1772	10	11	6	150103	8	1	6	6

En este caso sep="\t", indica que las tabulaciones son los separadores de datos.

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Leer archivo seleccionándolos desde una carpeta

```
> Salud4=read.table(file.choose(),header=TRUE,sep=",")
```



IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

```
> fix(Salud4)
```

	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	UBIGEO	DOMINIO	ESTRATO	CODINFOR	P400N
1	2063	46	11	3	150136	8	1	2	3
2	6062	1	11	2	70106	8	1	1	2
3	6062	111	11	3	70106	8	1	2	3
4	2039	41	11	8	150135	8	1	2	8
5	6045	134	11	8	70106	8	1	4	8
6	7559	10	11	6	150110	8	1	6	6
7	1736	34	11	5	70106	8	1	2	5
8	1738	21	11	5	70106	8	1	5	5
9	1741	50	11	4	70106	8	1	4	4
10	1848	87	11	3	150111	8	1	2	3
11	1932	1	11	6	150123	8	1	3	6
12	1979	68	11	3	150132	8	1	3	3
13	3552	41	11	4	70106	8	1	2	4
14	7565	171	11	5	150110	8	1	5	5
15	7565	171	11	6	150110	8	1	2	6
16	7624	80	11	8	150132	8	1	7	8
17	7628	51	11	4	150132	8	1	1	4
18	7803	85	11	8	150115	8	1	1	8
19	7926	57	11	3	150142	8	1	3	3
20	1772	10	11	6	150103	8	1	6	6

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Leer un segmento de archivo

Leer un segmento de archivo de Excel

Primero seleccionamos en Excel, la hoja o el segmento de hoja que queremos leer, y activamos la opción copiar.

	A	B	C	D	E
1	CASEID	MMIDX	MM1	MM2	MM3
2	000804001 03	7	2	0	
3	001009101 02	2	2	0	
4	001202101 02	1	2		
5	001607501 03	3	2		
6	002106101 02	2	2		
7	005700501 02	3	2		
8	007703301 02	3	2		
9	008101201 02	2	2		

Luego escribimos en la Consola lo siguiente:

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Luego escribimos en la Consola lo siguiente:

```
> mater5 = read.table("clipboard")
> mater5
```

	CASEID	MMIDX	MM1	MM2
000804001	3	7	2	0
001009101	2	2	2	0
001202101	2	1	2	0
001607501	3	3	2	0
002106101	2	2	2	0
005700501	2	3	2	0

En caso que seleccionemos un segmento que no incluya la primera fila, la cual por lo general, contiene el nombre de las variables, R les asignará los siguientes nombres: V1, V2, V3...Vn.

	A	B	C	D	E
1	CASEID	MMIDX	MM1	MM2	MM3
2	000804001 03	7	2	0	
3	001009101 02	2	2	0	
4	001202101 02	1	2		
5	001607501 03	3	2		
6	002106101 02	2	2		
7	005700501 02	3	2		
8	007703301 02	3	2		
9	008101201 02	2	2		
10	008106601 02	1	2		

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

```
> mater6 = read.table("clipboard")
> mater6
```

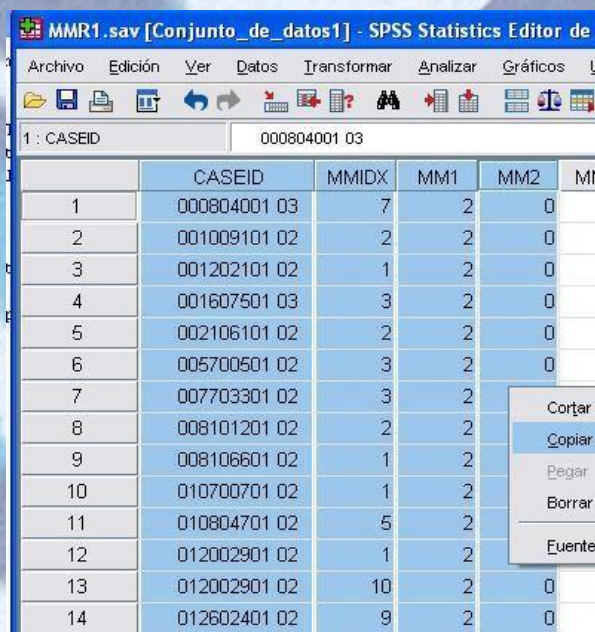
	V1	V2	V3	V4	V5
1	1202101	2	1	2	0
2	1607501	3	3	2	0
3	2106101	2	2	2	0
4	5700501	2	3	2	0

Sin embargo, atención con este procedimiento. Como se puede observar, R ha dividido la variable CASEID en V1 y V2, y ha anulado los ceros a la izquierda que identifican a la variable CASEID como una variable de tipo carácter, y no de tipo numérico.

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Leer un segmento de archivo de SPSS

Los procedimientos son los mismos que cuando se trabaja con un segmento de Excel



	CASEID	MMIDX	MM1	MM2	MM3
1	000804001 03	7	2	0	
2	001009101 02	2	2	0	
3	001202101 02	1	2	0	
4	001607501 03	3	2	0	
5	002106101 02	2	2	0	
6	005700501 02	3	2	0	
7	007703301 02	3	2		
8	008101201 02	2	2		
9	008106601 02	1	2		
10	010700701 02	1	2		
11	010804701 02	5	2		
12	012002901 02	1	2		
13	012002901 02	10	2	0	
14	012602401 02	9	2	0	

```
> mater7 = read.table("clipboard")
> mater7
```

	V1	V2	V3	V4	V5
1	804001	3	7	2	0
2	1009101	2	2	2	0
3	1202101	2	1	2	0
4	1607501	3	3	2	0
5	2106101	2	2	2	0
6	5700501	2	3	2	0
7	7703301	2	3	2	0
8	8101201	2	2	2	0
9	8106601	2	1	2	0
10	10700701	2	1	2	0
11	10804701	2	5	2	0

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

La diferencia con Excel es que, al copiar un segmento de archivo SPSS no se trasladan los nombres originales de las variables, incluso si las incluimos en la selección. SPSS les asigna nuevos nombres. Y al igual que en Excel cuando copiábamos un segmento de archivo que no incluía las filas con los nombres de las variables, se debe tener cuidado con la variable CASEID, ya que R la descompone en dos variables (V1 y V2).

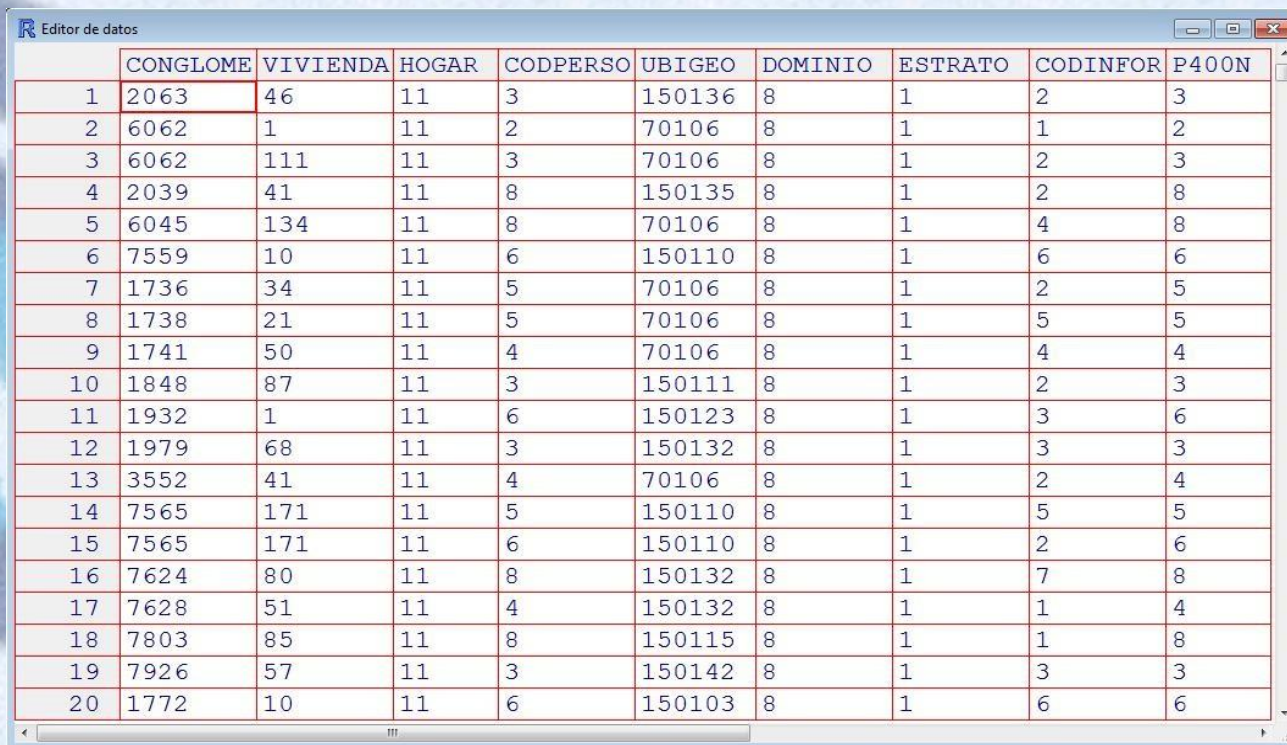
Leer archivos con la función `read.csv()`

```
> mater8 = read.csv(file="MATERMAT.csv")
> mater8
```

	CASEID	MMIDX	MM1	MM2	MM3	MM4	MM5	MM6	MM7	MM8	MM9	MM10	
1	000804001	03	7	2	0	NA	NA	NA	14	25	NA	2	NA
2	001009101	02	2	2	0	NA	NA	NA	6	30	NA	2	NA
3	001202101	02	1	2	0	NA	NA	NA	NA	38	NA	2	NA

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

```
> Data9=read.csv(file="Enaho2013400p.csv")
> fix(Data9)
```



	CONGLOME	VIVIENDA	HOGAR	CODPERSONO	UBIGEO	DOMINIO	ESTRATO	CODINFOR	P400N
1	2063	46	11	3	150136	8	1	2	3
2	6062	1	11	2	70106	8	1	1	2
3	6062	111	11	3	70106	8	1	2	3
4	2039	41	11	8	150135	8	1	2	8
5	6045	134	11	8	70106	8	1	4	8
6	7559	10	11	6	150110	8	1	6	6
7	1736	34	11	5	70106	8	1	2	5
8	1738	21	11	5	70106	8	1	5	5
9	1741	50	11	4	70106	8	1	4	4
10	1848	87	11	3	150111	8	1	2	3
11	1932	1	11	6	150123	8	1	3	6
12	1979	68	11	3	150132	8	1	3	3
13	3552	41	11	4	70106	8	1	2	4
14	7565	171	11	5	150110	8	1	5	5
15	7565	171	11	6	150110	8	1	2	6
16	7624	80	11	8	150132	8	1	7	8
17	7628	51	11	4	150132	8	1	1	4
18	7803	85	11	8	150115	8	1	1	8
19	7926	57	11	3	150142	8	1	3	3
20	1772	10	11	6	150103	8	1	6	6

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Explorar el contenido de un archivo

Ver el nombre de todas las variables de un archivo usando la función `names()`

```
> names(mater1)
```

[1]	"CASEID"	"MMIDX"	"MM1"	"MM2"	"MM3"	"MM4"	"MM5"
[8]	"MM6"	"MM7"	"MM8"	"MM9"	"MM10"	"MM11"	"MM12"
[15]	"MM13"	"MM14"	"MM15"	"MM16"	"MMC1"	"MMC2"	"filter_."
[22]	"V001"	"V002"	"V003"	"V101"	"V102"	"V103"	"V190"

```
> names(Data9)
[1] "CONGLOME" "VIVIENDA" "HOGAR"      "CODPERSONA" "UBIGEO"      "DOMINIO"      "ESTRATO"      "CODINFOR"
[9] "P400N"      "P400I"      "P400A1"     "P400A2"     "P400A3"     "P401C"      "P401D1"      "P401D2"
```

Listar todos los casos de una variable

```
> Data10=read.csv(file="Enaho2013400p0.csv")
> fix(Data10)
> Data10$DOMINIO
[1] 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
```


IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Listar el nombre de las filas [1] (casos) y el de las columnas [2] (variables) usando la **función labels**

```
> labels(mater1)
[[1]]
 [1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12"
[13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22" "23" "24"
[25] "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] "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"
[61] "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] "85" "86" "87" "88" "89" "90" "91" "92" "93" "94" "95" "96"
[97] "97" "98" "99" "100" "101" "102" "103" "104" "105" "106" "107" "108"
[109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
[121] "121" "122" "123" "124" "125" "126"

[[2]]
 [1] "CASEID"  "MMIDX"   "MM1"     "MM2"     "MM3"     "MM4"
 [7] "MM5"     "MM6"     "MM7"     "MM8"     "MM9"     "MM10"
[13] "MM11"    "MM12"    "MM13"    "MM14"    "MM15"    "MM16"
[19] "MMC1"    "MMC2"    "filter_." "V001"     "V002"     "V003"
[25] "V101"    "V102"    "V103"    "V190"
```

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Conocer el número de variables que contiene un marco de datos usando la función **length**

```
> length(Data10)
[1] 16
```

Listar un intervalo de casos de una variable del marco de datos haciendo referencia a su nombre

Ejemplo: listar los 10 primeros casos de la variable UBIGEO. Es decir, las diez primeras filas de la variable UBIGEO.

```
> Data10$UBIGEO[1:10]
[1] 150136 70106 70106 150135 70106 150110 70106 70106 70106 150111
```


IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Segmentar archivos

La variable DOMINIO, del marco de datos SumariaM, contiene información del dominio geográfico donde se encuentran los hogares entrevistados

```
> Sumaria$DOMINIO
[1] 4 4 7 7 7 5 5 6 6 6 6 6 7 6 6 5 7 2 2 2 2 2 5 5 5 4 1 1 8 8 8 8 8 8 5 7 7 3 6 1 1 6 7 7 3
[46] 1 1 5 6 8 5 1 1 1 1 8 1 7 7 7 7 6 8 6 6 5 5 5 7 5 1 1 1 8 8 8 8 8 2 7 7 7 3 5 1 1 1 6 7 7
[91] 7 3 1
```

Donde: 1= Costa Norte, 2= Costa Centro, 3=Costa Sur, 4=Sierra Norte, 5=Sierra Centro
6=Sierra Sur 7=Selva 8=Lima Metropolitana.

Por ejemplo si quisiéramos trabajar solo con los hogares de la Selva, que brindaron información., entonces debemos segmentar el marco de datos.

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Para ello usaremos la función **subset()** de la siguiente manera:

```
> CNorte <- subset(Sumaria, subset=DOMINIO==7)
> fix(CNorte)
```

Donde:

- CNorte Nombre del nuevo marco de datos que contendrá los elementos seleccionados.
- Subset Función que selecciona todos los elementos del objeto (en este caso del marco de datos Sumaria) que deben conservarse, en este caso, todos los elementos de DOMINIO igual al código 7.

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

La opción **labels** nos permite conocer los casos (filas) que han sido seleccionadas para integrar el nuevo marco de datos, y además nos lista los nombres de las variables consideradas en él.

```
> labels(CNorte)
[[1]]
[1] "3" "4" "5" "13" "17" "36" "37" "43" "44" "58" "59" "60" "61" "69" "80" "81" "82" "89"
[19] "90" "91"

[[2]]
[1] "AÑO" "MES" "CONGLOME" "VIVIENDA" "HOGAR" "UBIGEO" "DOMINIO" "ESTRATO"
[9] "PERCEPHO" "GRU52HD2" "INGPEIHD" "MIEPERHO" "TOTMIEHO" "INGMO1HD" "INGMO2HD" "INGHOG1D"
[17] "INGHOG2D" "GASHOG1D" "GASHOG2D" "LINPE" "LINEA" "ld" "POBREZA" "FACTOR07"
```

El nuevo marco de datos al que hemos llamado: “CNorte”, contiene información sobre 20 casos y 24 variables.

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

A continuación podemos verificar, que ahora, sólo los casos que registraron códigos 7 forman parte de variable DOMINIO.

```
> CNorte$DOMINIO  
[1] 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
```

Guardar y recargar marcos de datos

Guardar usando `save()`

En el punto anterior se creó el marco de datos “CNorte”, si queremos usarlo en posteriores sesiones de trabajo, necesitamos guardarlo en un archivo. Para ello utilizamos la función **`save()`** de la siguiente manera:

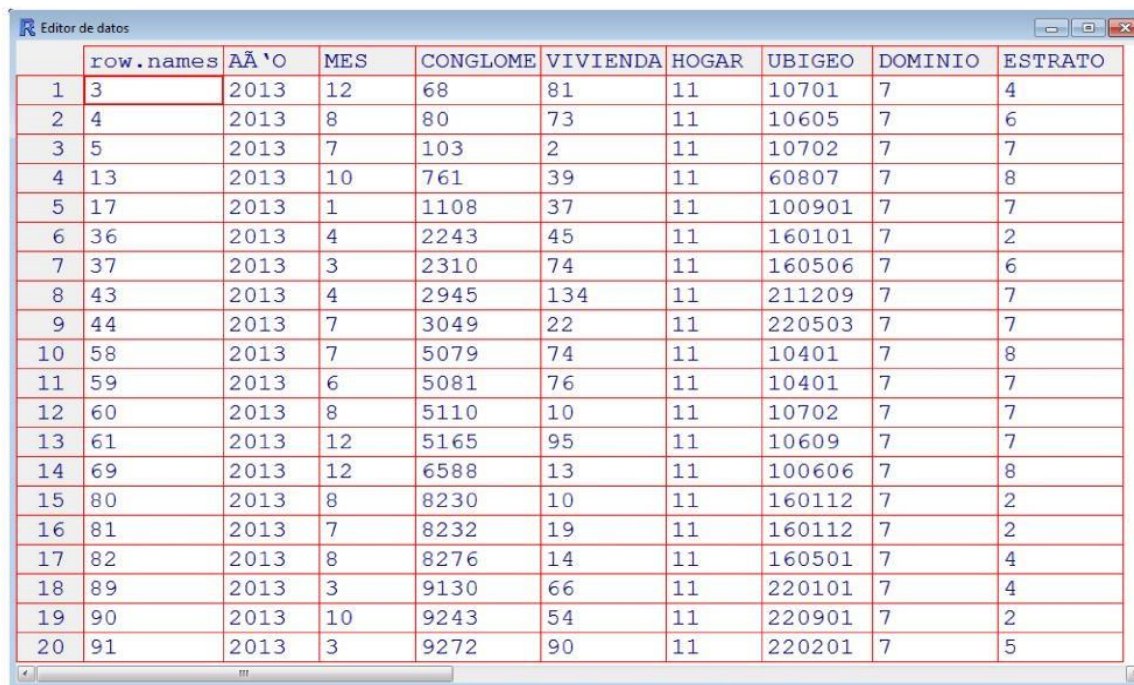
```
> save(CNorte, file="D:/School/R/Escuela/Datos/CostaNorte.RData")
```

El archivo “CostaNorte.RData” guarda el objeto: marco de datos “CNorte”.

IV. TRATAMIENTO Y EXPLORACIÓN DE ARCHIVOS

Recargar un archivo de extensión ***.RData** en una nueva sesión de trabajo usando **load()**

```
> load("D:/School/R/Escuela/Datos/CostaNorte.RData")
> CNorte
> fix(CNorte)
```



	row.names	AÃ'O	MES	CONGLOME	VIVIENDA	HOGAR	UBIGEO	DOMINIO	ESTRATO
1	3	2013	12	68	81	11	10701	7	4
2	4	2013	8	80	73	11	10605	7	6
3	5	2013	7	103	2	11	10702	7	7
4	13	2013	10	761	39	11	60807	7	8
5	17	2013	1	1108	37	11	100901	7	7
6	36	2013	4	2243	45	11	160101	7	2
7	37	2013	3	2310	74	11	160506	7	6
8	43	2013	4	2945	134	11	211209	7	7
9	44	2013	7	3049	22	11	220503	7	7
10	58	2013	7	5079	74	11	10401	7	8
11	59	2013	6	5081	76	11	10401	7	7
12	60	2013	8	5110	10	11	10702	7	7
13	61	2013	12	5165	95	11	10609	7	7
14	69	2013	12	6588	13	11	100606	7	8
15	80	2013	8	8230	10	11	160112	7	2
16	81	2013	7	8232	19	11	160112	7	2
17	82	2013	8	8276	14	11	160501	7	4
18	89	2013	3	9130	66	11	220101	7	4
19	90	2013	10	9243	54	11	220901	7	2
20	91	2013	3	9272	90	11	220201	7	5

V. TRATAMIENTO DE VARIABLES

Convertir vectores en factores o crear variables cualitativas

Para crear variables cualitativas en R, debemos convertir los vectores numéricos (variables cuantitativas) de un marco de datos, en factores (variables cualitativas).

Como ya vimos anteriormente, la variable DOMINIO da cuenta de los niveles geográficos de los hogares entrevistados. Donde: 1= Costa Norte, 2= Costa Centro, 3=Costa Sur, 4=Sierra Norte, 5=Sierra Centro 6=Sierra Sur 7=Selva 8=Lima Metropolitana.

```
> Sumaria$DOMINIO
 [1] 4 4 7 7 7 5 5 6 6 6 6 6 7 6 6 5 7 2 2 2 2 2 5 5 5 4 1 1 8 8 8 8 8 8 5 7 7 3 6 1 1 6 7 7 3
[46] 1 1 5 6 8 5 1 1 1 1 8 1 7 7 7 7 6 8 6 6 5 5 5 7 5 1 1 1 8 8 8 8 8 2 7 7 7 3 5 1 1 1 6 7 7
[91] 7 3 1
```

Como vemos, la variable DOMINIO es un vector numérico. Para convertirlo en un factor (variable cualitativa) con el mismo nombre DOMINIO, escribimos las siguientes instrucciones:

V. TRATAMIENTO DE VARIABLES

R Editor de datos

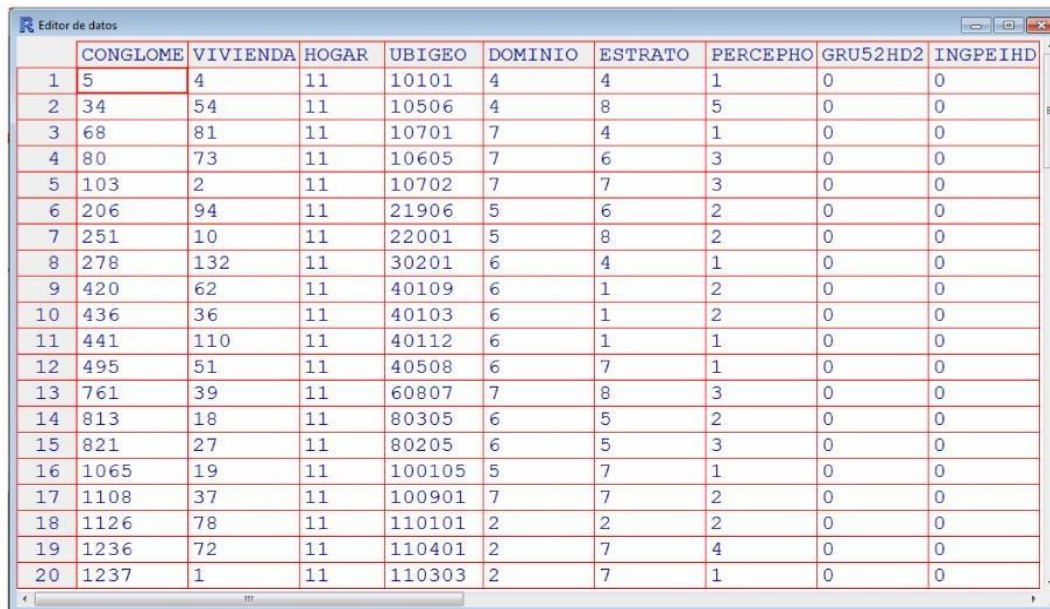
	AÑO	MES	CONGLOME	VIVIENDA	HOGAR	UBIGEO	DOMINIO	ESTRATO	PERCEPHO
1	2013	11	5	4	11	10101	S.Norte	4	1
2	2013	8	34	54	11	10506	S.Norte	8	5
3	2013	12	68	81	11	10701	Selva	4	1
4	2013	8	80	73	11	10605	Selva	6	3
5	2013	7	103	2	11	10702	Selva	7	3
6	2013	9	206	94	11	21906	S.Centro	6	2
7	2013	1	251	10	11	22001	S.Centro	8	2
8	2013	10	278	132	11	30201	S.Sur	4	1
9	2013	4	420	62	11	40109	S.Sur	1	2
10	2013	8	436	36	11	40103	S.Sur	1	2
11	2013	1	441	110	11	40112	S.Sur	1	1
12	2013	5	495	51	11	40508	S.Sur	7	1
13	2013	10	761	39	11	60807	Selva	8	3
14	2013	5	813	18	11	80305	S.Sur	5	2
15	2013	1	821	27	11	80205	S.Sur	5	3
16	2013	10	1065	19	11	100105	S.Centro	7	1
17	2013	1	1108	37	11	100901	Selva	7	2
18	2013	12	1126	78	11	110101	C.Centro	2	2
19	2013	2	1236	72	11	110401	C.Centro	7	4
20	2013	3	1237	1	11	110303	C.Centro	7	1

V. TRATAMIENTO DE VARIABLES

Eliminar variables de un marco de datos

Eliminamos de Sumaria, 2 variables : AÑO, MES Para ello, escribimos las siguientes instrucciones:

```
> Sumaria$AÑO <-NULL
> Sumaria$MES <-NULL
> length(Sumaria)
[1] 22
> fix(Sumaria)
```



	CONGLOME	VIVIENDA	HOGAR	UBIGEO	DOMINIO	ESTRATO	PERCEPHO	GRU52HD2	INGPEIHD
1	5	4	11	10101	4	4	1	0	0
2	34	54	11	10506	4	8	5	0	0
3	68	81	11	10701	7	4	1	0	0
4	80	73	11	10605	7	6	3	0	0
5	103	2	11	10702	7	7	3	0	0
6	206	94	11	21906	5	6	2	0	0
7	251	10	11	22001	5	8	2	0	0
8	278	132	11	30201	6	4	1	0	0
9	420	62	11	40109	6	1	2	0	0
10	436	36	11	40103	6	1	2	0	0
11	441	110	11	40112	6	1	1	0	0
12	495	51	11	40508	6	7	1	0	0
13	761	39	11	60807	7	8	3	0	0
14	813	18	11	80305	6	5	2	0	0
15	821	27	11	80205	6	5	3	0	0
16	1065	19	11	100105	5	7	1	0	0
17	1108	37	11	100901	7	7	2	0	0
18	1126	78	11	110101	2	2	2	0	0
19	1236	72	11	110401	2	7	4	0	0
20	1237	1	11	110303	2	7	1	0	0

V. TRATAMIENTO DE VARIABLES

Renombrar variables

Para renombrar variables utilizamos la función **names()**.

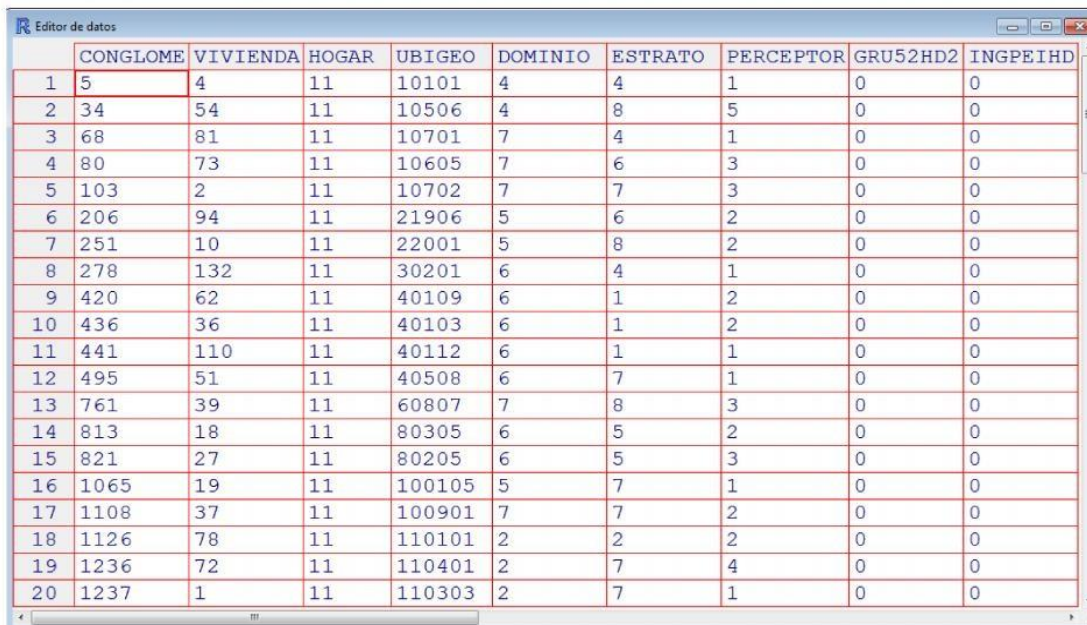
Ejemplo: renombrar la variable V101 (región de residencia) como DPTO.

```
> names(Sumaria)[c(7)] <- c("PERCEPTOR")
> fix(Sumaria)
```

Donde:

[c(7)] es la posición de la variable PERCEPHO, en el marco de datos Sumaria.

c("PERCEPTOR")
Nuevo nombre a asignar a la posición 7 del marco de datos Sumaria.



	CONGLOME	VIVIENDA	HOGAR	UBIGEO	DOMINIO	ESTRATO	PERCEPTOR	GRU52HD2	INGPEIHD
1	5	4	11	10101	4	4	1	0	0
2	34	54	11	10506	4	8	5	0	0
3	68	81	11	10701	7	4	1	0	0
4	80	73	11	10605	7	6	3	0	0
5	103	2	11	10702	7	7	3	0	0
6	206	94	11	21906	5	6	2	0	0
7	251	10	11	22001	5	8	2	0	0
8	278	132	11	30201	6	4	1	0	0
9	420	62	11	40109	6	1	2	0	0
10	436	36	11	40103	6	1	2	0	0
11	441	110	11	40112	6	1	1	0	0
12	495	51	11	40508	6	7	1	0	0
13	761	39	11	60807	7	8	3	0	0
14	813	18	11	80305	6	5	2	0	0
15	821	27	11	80205	6	5	3	0	0
16	1065	19	11	100105	5	7	1	0	0
17	1108	37	11	100901	7	7	2	0	0
18	1126	78	11	110101	2	2	2	0	0
19	1236	72	11	110401	2	7	4	0	0
20	1237	1	11	110303	2	7	1	0	0

V. TRATAMIENTO DE VARIABLES

Crear nuevas variables a partir de otras existentes haciendo cálculos

Crearemos nuevas variables, a partir de cálculos matemáticos que efectuaremos sobre la variable GASHOG2D (Gasto del hogar).

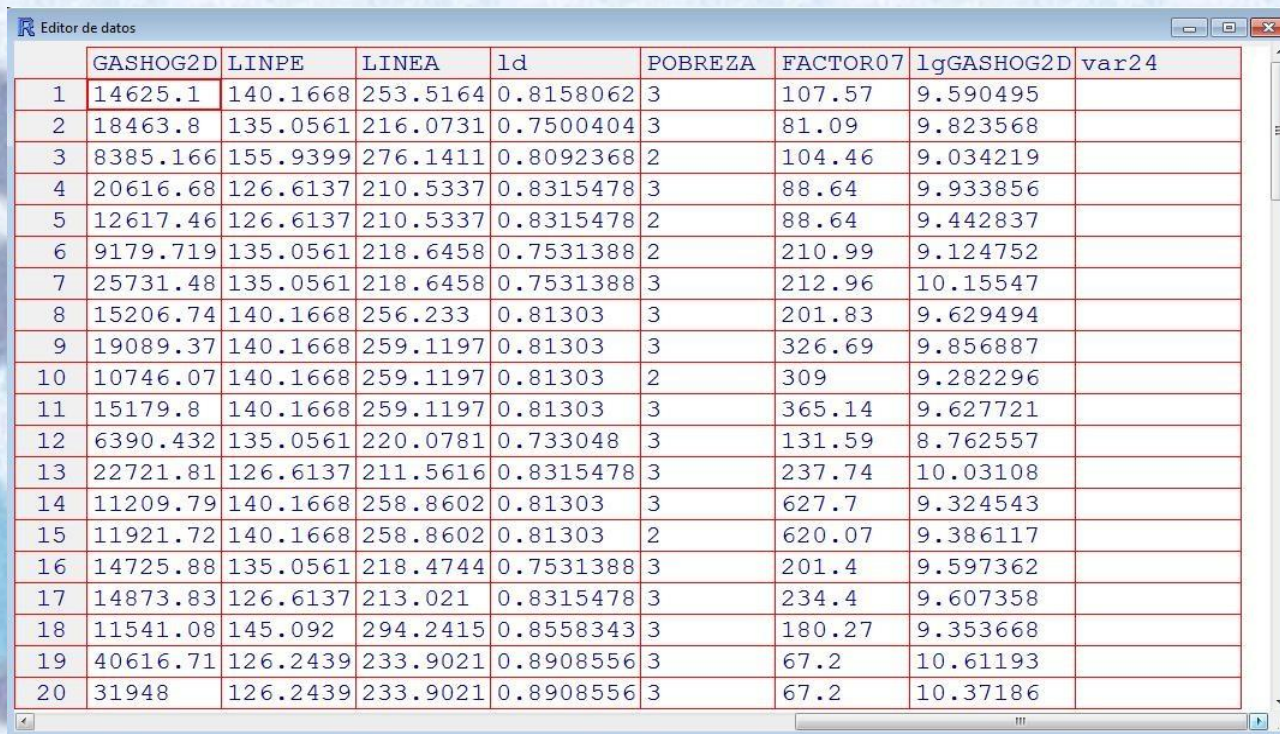
Logaritmo natural de una variable

```
> Sumaria$lgGASHOG2D <- with(Sumaria, log(GASHOG2D))  
> fix(Sumaria)
```

Donde:

- | | |
|-------------------|--|
| lgGASHOG2D | es el nombre de la variable a crear en el marco de datos Sumaria |
| with() | Es la función que permite la creación de la nueva variable luego de realizar la función log() en la base de datos. |
| log() | función que permite calcular el logaritmo natural de una variable cuantitativa. |

V. TRATAMIENTO DE VARIABLES



R Editor de datos

	GASHOG2D	LINPE	LINEA	ld	POBREZA	FACTOR07	lgGASHOG2D	var24
1	14625.1	140.1668	253.5164	0.8158062	3	107.57	9.590495	
2	18463.8	135.0561	216.0731	0.7500404	3	81.09	9.823568	
3	8385.166	155.9399	276.1411	0.8092368	2	104.46	9.034219	
4	20616.68	126.6137	210.5337	0.8315478	3	88.64	9.933856	
5	12617.46	126.6137	210.5337	0.8315478	2	88.64	9.442837	
6	9179.719	135.0561	218.6458	0.7531388	2	210.99	9.124752	
7	25731.48	135.0561	218.6458	0.7531388	3	212.96	10.15547	
8	15206.74	140.1668	256.233	0.81303	3	201.83	9.629494	
9	19089.37	140.1668	259.1197	0.81303	3	326.69	9.856887	
10	10746.07	140.1668	259.1197	0.81303	2	309	9.282296	
11	15179.8	140.1668	259.1197	0.81303	3	365.14	9.627721	
12	6390.432	135.0561	220.0781	0.733048	3	131.59	8.762557	
13	22721.81	126.6137	211.5616	0.8315478	3	237.74	10.03108	
14	11209.79	140.1668	258.8602	0.81303	3	627.7	9.324543	
15	11921.72	140.1668	258.8602	0.81303	2	620.07	9.386117	
16	14725.88	135.0561	218.4744	0.7531388	3	201.4	9.597362	
17	14873.83	126.6137	213.021	0.8315478	3	234.4	9.607358	
18	11541.08	145.092	294.2415	0.8558343	3	180.27	9.353668	
19	40616.71	126.2439	233.9021	0.8908556	3	67.2	10.61193	
20	31948	126.2439	233.9021	0.8908556	3	67.2	10.37186	

V. TRATAMIENTO DE VARIABLES

Recodificar una variable numérica usando la función `recode()` o `Recode()`

`recode()` es una función que sirve para recodificar vectores numéricos, de carácter o factores. En este caso `Recode` es un alias de la función `recode`, y se utiliza para evitar conflictos con funciones similares entre ciertos paquetes de R. (Fox, 2011 en R Documentation).

Para activar la función `recode()` necesitamos cargar el paquete `car`, de la siguiente manera:

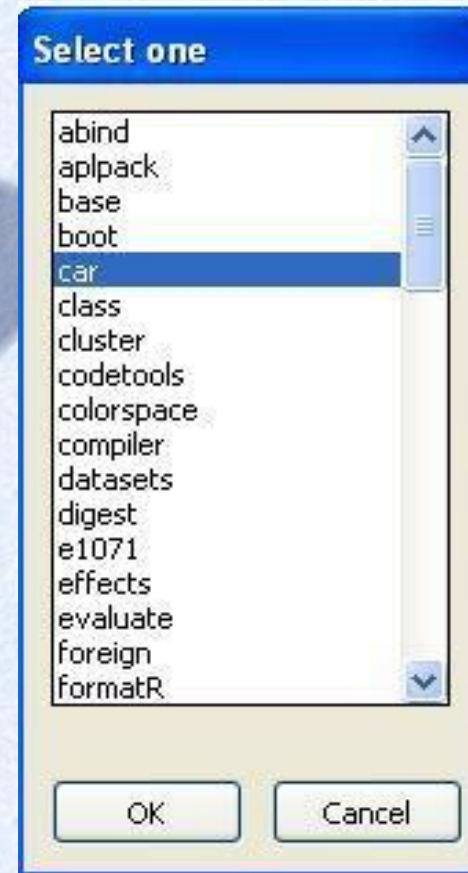
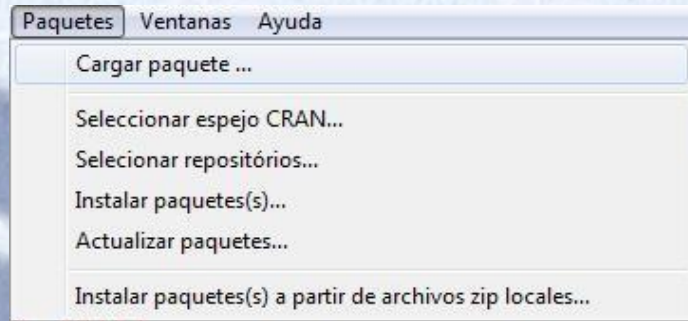
Desplegamos el menú:

Paquetes --> Cargar paquetes...

En la ventana emergente “Seleccione”, seleccionamos `car`,

Finalmente hacemos clic sobre el botón OK

V. TRATAMIENTO DE VARIABLES



V. TRATAMIENTO DE VARIABLES

Ejemplo: Reducir a tres categorías la variable V190 que tiene cinco (ver 5.1), creando la variable NSE (Nivel socioeconómico).

Para ello crearemos un factor numérico usando la función `Recode()`:

```
> mater1$NSE<-Recode(mater1$V190,  
+ '1:2=1; 3=2; 4:5=3',  
+ as.factor.result=TRUE)
```

También podemos crear un factor carácter usando la misma función:

```
> mater1$NSECTG<-Recode(mater1$V190,  
+ '1:2="Pobre"; 3="Medio"; 4:5="Rico"',  
+ as.factor.result=TRUE)
```


V. TRATAMIENTO DE VARIABLES

A continuación usando **table()**, podemos observar los resultados de ambos procedimientos:

```
> table(mater1$NSE)
 1  2  3
87 23 16
> table(mater1$NSECTG)
Medio Pobre Rico
 23    87   16
```

En la tabla de frecuencia de la variable NSE, los resultados se presentan en orden numérico, mientras que en la de la variable NSECTG el orden es alfabético. Para forzar el orden “Pobre-Medio-Rico” en la variable NSECTG, debemos agregar el argumento **levels** a la función **Recode**, de la siguiente manera:

```
> mater1$NSECTG<-Recode(mater1$V190,
+ '1:2="Pobre"; 3="Medio"; 4:5="Rico"',
+ as.factor.result=TRUE, levels=c("Pobre","Medio","Rico"))
```

V. TRATAMIENTO DE VARIABLES

Donde:

levels indica en qué orden deben presentarse las categorías .

Verificamos los resultados

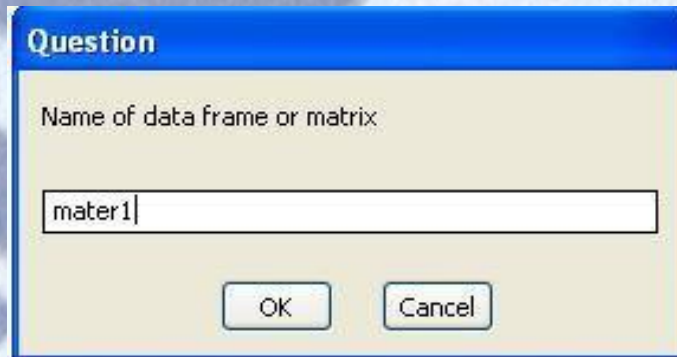
```
> table(mater1$NSECTG)
```

Pobre	Medio	Rico
87	23	16

V. TRATAMIENTO DE VARIABLES

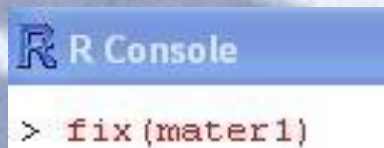
Modificar datos desde el Editor

Ya describimos la ventana de edición de datos. Aquí ilustramos cómo acceder concretamente a ella para editar datos. Para ello desplegamos el menú:
Editar → Editor de Datos...



V. TRATAMIENTO DE VARIABLES

Al abrirse el editor de datos, en la Consola se escribirá automáticamente la siguiente instrucción:



```
> fix(mater1)
```

Una vez abierto el editor, podemos hacer modificaciones directamente sobre la base de datos, como en el siguiente ejemplo: fila 3 de la variable MM6.

	CASEID	MMIDX	MM6
1	000804001 03	7	14
2	001009101 02	2	6
3	001202101 02	1	NA
4	001607501 03	3	11

	CASEID	MMIDX	MM6
1	000804001 03	7	14
2	001009101 02	2	6
3	001202101 02	1	1
4	001607501 03	3	11

**Recuerda siempre nuestro correo
para cualquier consulta**



campusvirtual@inei.gob.pe

Comunicación constante con la Escuela del INEI

Correo de la Escuela del INEI
enei@inei.gob.pe

Área de Educación Virtual
campusvirtual@inei.gob.pe

