Escuela Nacional de Estadística e Informática







ANÁLISIS DE DATO CON R

Instituto Nacional de Estadística e Informática

Escuela Nacional de Estadística e Informática



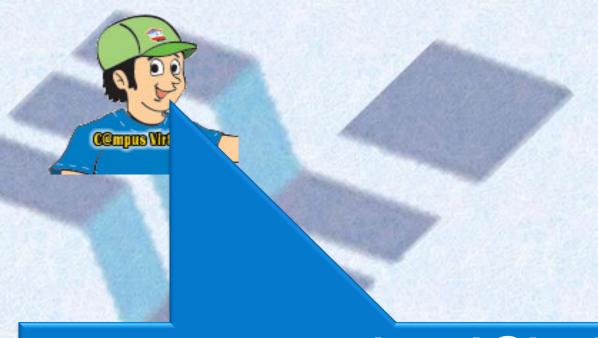


Centro Andino de Formación y Capacitación en Estadística

Cursos Especializados en Estadística e Informática



Recuerda siempre nuestro correo para cualquier consulta



campusvirtual@inei.gob.pe



CURSO BASICO DE "R"





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

recode() es una función que sirve para recodificar vectores numéricos, de caracter 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 seguir los siguientes pasos:

- Instalar el paquete de algún MIRROR cercano a nuestro país, por ejemplo Colombia(Bogota). (No viene en el paquete básico)
- Una vez instalado, hay que cargar el **paquete car**, de la siguiente manera:

Desplegamos el menú Paquetes → Cargar paquete...

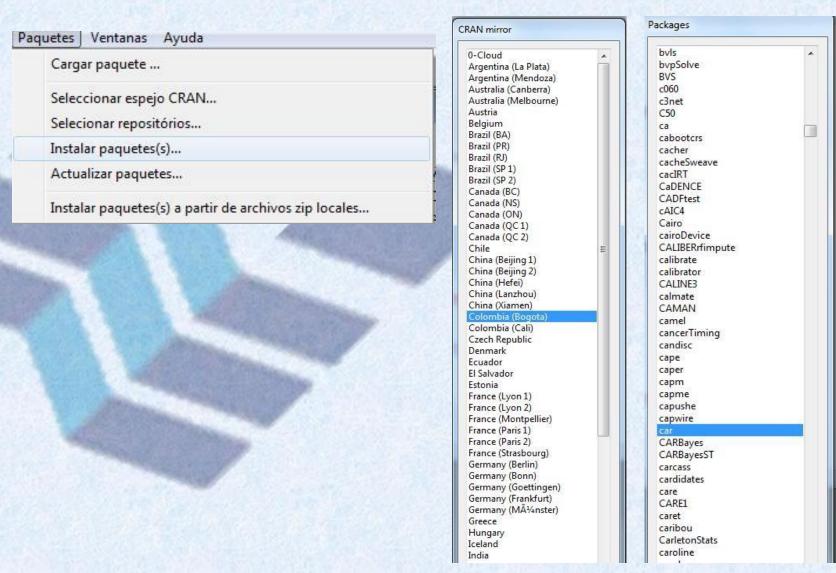
En la ventana emergente "Select one", seleccionamos car, Finalmente hacemos clic sobre el botón OK

Instituto Nacional de Estadística e Informática

Escuela Nacional de Estadística e Informática



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





¿Cómo se pide ayuda en R?

- help(),?
- help.search("")
- help.start()

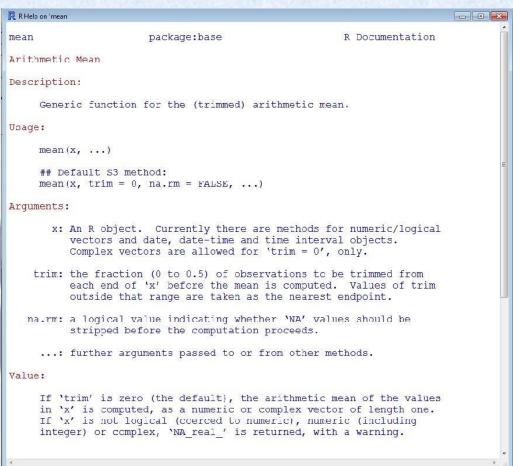


> help()

```
R Help on 'help'
                                                                          - - X
                       package:utils
help
                                                        R Documentation
Documentation
Description:
     'help' is the primary interface to the help systems.
Usage:
     help(topic, package = NULL, lib.loc = NULL,
          verbose = getOption("verbose"),
          try.all.packages = getOption("help.try.all.packages"),
          help type = getOption("help type"))
Arguments:
   topic: usually, a name or character string specifying the topic for
          which help is sought. A character string (enclosed in
          explicit single or double quotes) is always taken as naming a
          topic.
          If the value of 'topic' is a length-one character vector the
          topic is taken to be the value of the only element.
          Otherwise 'topic' must be a name or a reserved word (if
```











```
Principal Components Analysis
     'princomp' performs a principal components analysis on the given
     numeric data matrix and returns the results as an object of class
    princomp(x, ...)
     ## S3 method for class 'formula'
     princomp(formula, data = NULL, subset, na.action, ...)
     ## Default S3 method:
     princomp(x, cor = FALSE, scores = TRUE, covmat = NULL,
              subset = rep len(TRUE, nrow(as.matrix(x))), ...)
     ## S3 method for class 'princomp'
     predict (object, newdata, ...)
 formula: a formula with no response variable, referring only to
          numeric variables.
    data: an optional data frame (or similar: see 'model.frame')
          containing the variables in the formula 'formula'. By
          default the variables are taken from 'environment(formula)'.
 subset: an optional vector used to select rows (observations) of the
          data matrix 'x'.
na.action: a function which indicates what should happen when the data
          contain 'NA's. The default is set by the 'na.action' setting
```



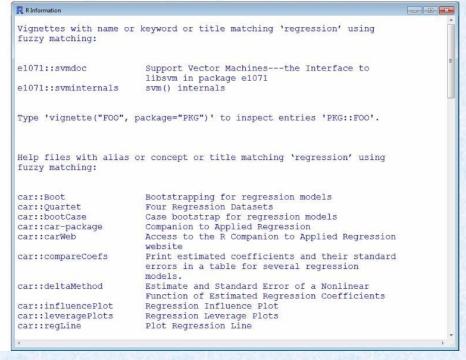
Buscar ayuda sobre un concepto estadístico

> help(regression)

No documentation for 'regression' in specified packages and libraries:

you could try '??regression'

> help.search("regression")





Equivalencias

- > ?mean
- > ??regression
- > help(mean)
- > help.search("regression")



SCRIPTS

¿Qué son los scripts?

- Los scripts son la forma más recomendable de interactuar con R
 - Guardamos registro de todo lo que hacemos
 - Podemos reutilizarlo más adelante
- Los scripts facilitan la colaboración entre investigadores: basta con enviar una base de datos y un script
- Para escribir un script podemos utilizar cualquier editor de texto

Escuela Nacional de Estadística e Informática



SCRIPTS

¿Cómo trabajar con scripts?

- Mediante el comando source()
- Mediante "copia-pega"

Atención

Para que los scripts sean útiles es vital "comentarlos" (#)



ÁREA DE TRABAJO (WORPSPACE)

Todos los objetos que se crean en R se almacenan en el área de trabajo (workspace)

- Podemos guardar el workspace: save.image()
- Podemos recuperar un workspace previamente guardado: load()
- Podemos ver qué objetos se encuentran en el workspace: 1s()
- Podemos eliminar objetos del workspace: rm()



AREA DE TRABAJO (WORPSPACE)

[Previously saved workspace restored]

```
> x<-c(1,2,3,4,5)
> #Vemos los objetos creados en el área de trabajo
> 1s()
               "test.csv" "test.csv1" "test.missing" "x"
[1] "test"
[6] "y"
> X
[1] 1 2 3 4 5
> y
[1] 0 3 6 9 12 15
> test
  prgtype gender id ses schtyp level
```



ÁREA DE TRABAJO (WORPSPACE)

```
> #Si quiero borrar algún objeto
> rm("test.csv")
> rm("test.csv1")
> 1s()
[1] "test"
                   "test.missing" "x"
> #Si quiero quardar mi área de trabajo
> save.image("Areal")
> #Si quiero visualizar el directorio de trabajo
> dir()
                               "auto.dta"
                                                         "Clase2.R"
 [1] "Areal"
 [4] "CostaNorte.RData"
                               "Enaho01-2013-200.dta"
                                                         "Enaho2013400d.dat"
 [7] "Enaho2013400p.csv"
                               "Enaho2013400p.sav"
                                                         "Enaho2013400p.xlsx"
[10] "Enaho2013400p0.csv"
                               "hogares.csv"
                                                         "hogares.dat"
[13] "hogares.sav"
                               "hogares.xls"
                                                         "hogares.xlsx"
[16] "Importar.R"
                                                         "mujeres.dta"
                               "ImportarSTATA.R"
                               "SumariaM.csv"
[19] "scan.txt"
                                                         "SumariaM.sav"
[22] "test.csv"
                                                         "test missing.txt"
                               "test.txt"
[25] "test missing comma.txt"
> rm("test")
> rm("test.missing")
> rm("x")
> ls()
[1] "y"
> rm("y")
> 1s()
character (0)
> #Llamo a area de trabajo guardada
> load("Area1")
> ls()
                  "test.missing" "x"
[1] "test"
> X
[1] 1 2 3 4 5
```



ARCHIVO HISTÓRICO

- Todos los comandos que vamos ejecutando en R (también los que producen errores) se van guardando en un archivo histórico
 - Podemos desplazarnos entre los comandos anteriores con flecha_arriba y flecha_abajo
 - Podemos guardar el archivo histórico: savehistory(file='"')
 - Podemos recuperar archivos históricos: loadhistory(file='"')
- El archivo histórico es muy útil como punto de partida para la elaboración de un script



ARCHIVO HISTÓRICO

```
> X
[1] 1 2 3 4 5
> #Hallo el promedio
> mean(x)
[1] 3
> #Hallo la medianan
> median(x)
[1] 3
> #Hallo la desviación estandar
> std(x)
Error: no se pudo encontrar la función "std"
> sd(x)
[1] 1.581139
> #Verifico el directorio de trabajo
> getwd()
[1] "D:/School/R/Escuela/Datos"
> #Guardo mi archivo histórico"
> savehistory("Historia4")
> dir()
 [1] "Areal"
                               "auto.dta"
                                                        "Clase2.R"
 [4] "CostaNorte.RData"
                               "Enaho01-2013-200.dta"
                                                        "Enaho2013400d.dat"
 [7] "Enaho2013400p.csv"
                               "Enaho2013400p.sav"
                                                        "Enaho2013400p.xlsx"
[10] "Enaho2013400p0.csv"
                               "Historia4"
                                                        "hogares.csv"
[13] "hogares.dat"
                                                        "hogares.xls"
                              "hogares.sav"
                                                        "ImportarSTATA.R"
[16] "hogares.xlsx"
                               "Importar.R"
[19] "mujeres.dta"
                              "scan.txt"
                                                        "SumariaM.csv"
[22] "SumariaM.sav"
                              "test.csv"
                                                        "test.txt"
                               "test missing comma.txt"
[25] "test missing.txt"
```



ARCHIVO HISTÓRICO

Veamos el archivo histórico con el explorador de archivos

```
Ilistoria4: Bloc de notas
 Archivo Edición Formato Ver Ayuda
 help()
 help(mean)
 help(princomp)
 help(regression)
 help.search("regression")
 ??regression
 x < -c(1, 2, 3, 4, 5)
 #Vemos los objetos creados en el área de trabajo
1s()
X
 #Si quiero borrar algún objeto
 rm("test.csv")
 rm("test.csv1")
 #Si quiero guardar mi área de trabajo
 save.image("Areal")
 #Si quiero visualizar el directorio de trabajo
 dir()
 rm ("test")
 rm ("test.missing")
 rm ("x")
 13()
 rm ("y")
1s()
 #Llamo a area de trabajo quardada
 load("Areal")
1s()
X
X < C(1,2,3,4,5)
 #Hallo el promedio
 mean(x)
 #Hallo la medianan
 median(x)
Hallo la desviación estandar
```



Hay tres niveles de funciones en R

- Las que están accesibles por defecto: ejemplos mean(), var(), cor()
- Las que están instaladas pero no inmediatamente accesibles
- · Las que no están instaladas: hay miles disponibles en internet



```
> x
[1] 1 2 3 4 5
> ?source
> mean(x)
[1] 3
> y<-c(2,3,4,6,7)
> cor(x,y)
[1] 0.9912407
> #Quiero crear una distribución normal multivariante
> #Tengo información que el comando es myrnorm
> ?myrnorm
No documentation for 'myrnorm' in specified packages and libraries:
you could try '??myrnorm'
> ??myrnorm
```

```
Help files with alias or concept or title matching 'mvrnorm' using fuzzy matching:

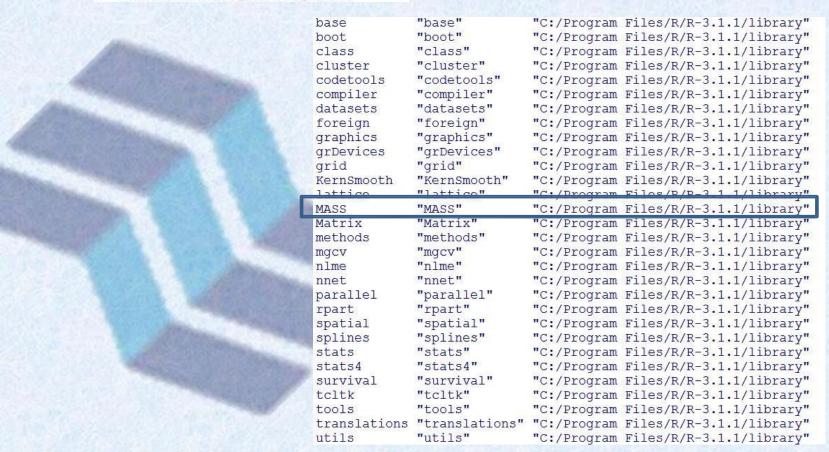
MASS::mvrnorm

Simulate from a Multivariate Normal Distribution

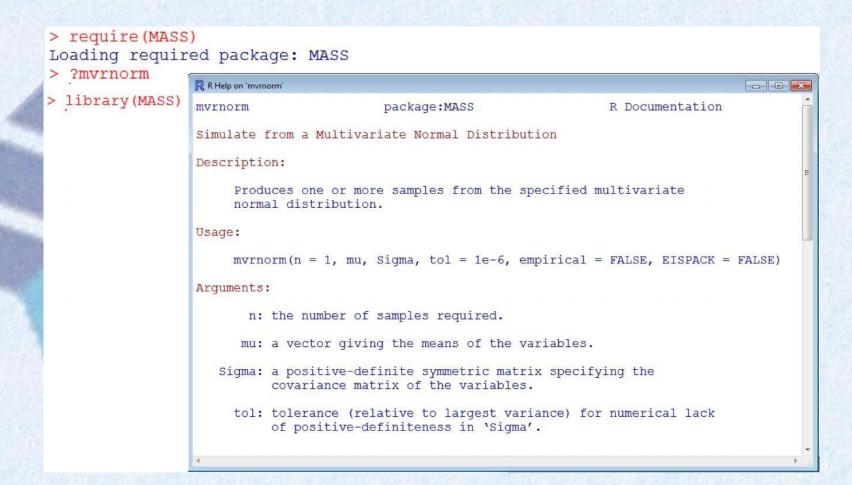
Type '?PKG::FOO' to inspect entries 'PKG::FOO', or 'TYPE?PKG::FOO' for entries like 'PKG::FOO-TYPE'.
```



> installed.packages()









<u>Funciones a Instalar en R</u>

```
> #Imaginemos que queremos utilizar el método de componentes principales
> #Sobre los vectores x e y
> x
[1] 1 2 3 4 5
> y
[1] 2 3 4 6 7
> #Solicitamos a R ayuda sobre PCA (Analisis de componentes principales)
> ?PCA
No documentation for 'PCA' in specified packages and libraries:
you could try '??PCA'
> install.packages("FactoMiner")
Installing package into 'C:/Users/lulloa/Documents/R/win-library/3.1'
(as 'lib' is unspecified)
---- Please select a CRAN mirror for use in this session ---
```

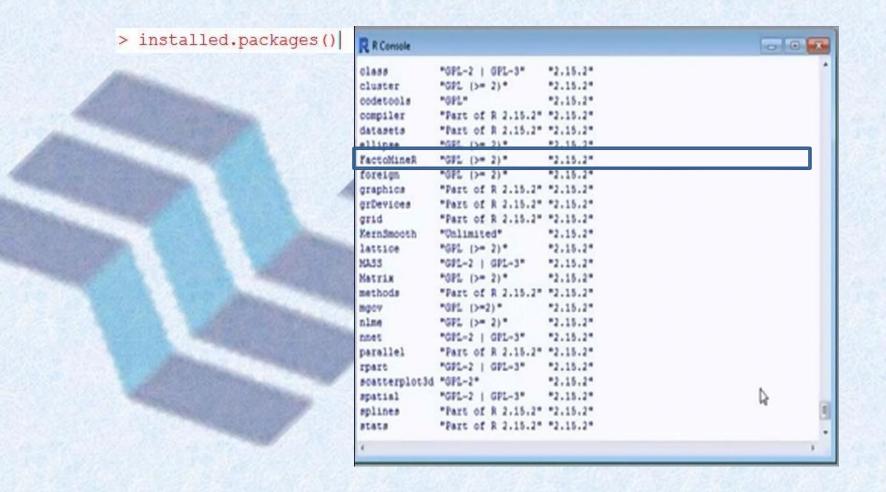
URL: ... t.r-project.org/bin/windows/contrib/2.15/FactoMineR_1.20.zip



CRAN mirror

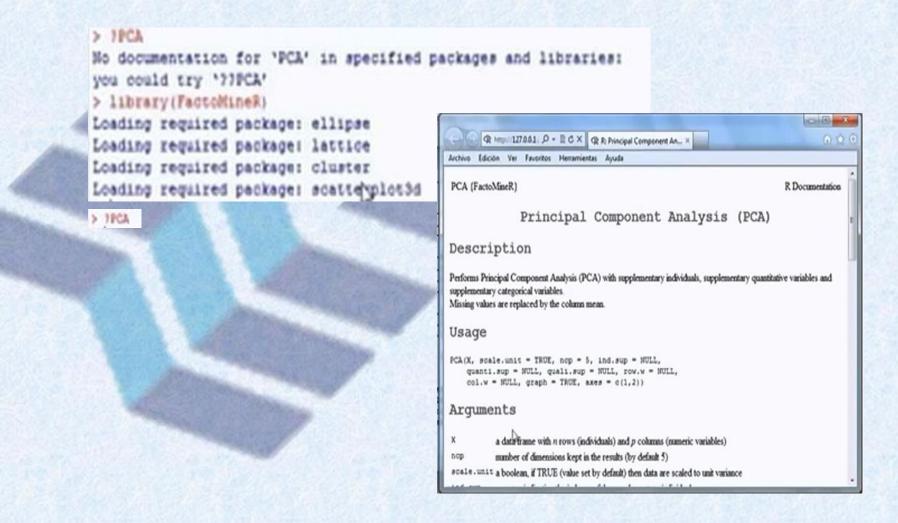


<u>Funciones a Instalar en R</u>





<u>Funciones a Instalar en R</u>





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

recode() es una función que sirve para recodificar vectores numéricos, de caracter 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 seguir los siguientes pasos:

- Instalar el paquete de algún MIRROR cercano a nuestro país, por ejemplo Colombia(Bogota). (No viene en el paquete básico)
- Una vez instalado, hay que cargar el **paquete car**, de la siguiente manera:

Desplegamos el menú Paquetes → Cargar paquete...

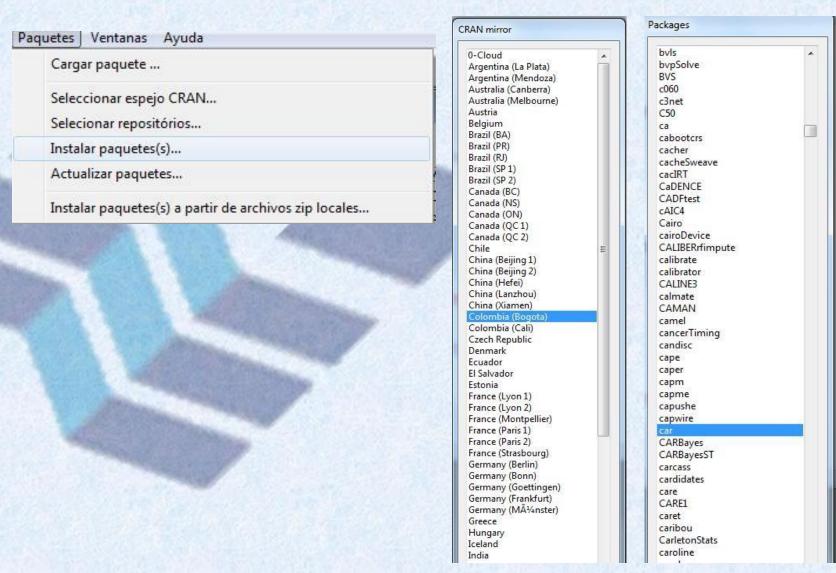
En la ventana emergente "Select one", seleccionamos car, Finalmente hacemos clic sobre el botón OK

Instituto Nacional de Estadística e Informática

Escuela Nacional de Estadística e Informática



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





Todo en R son objetos; los datos también, pero...

¿qué tipos de objetos existen en R? : función class()

```
> x
[1] 1 2 3 4 5
> class(x)
[1] "numeric"
> v<-c("Saul", "Rosa", "Franklin", "Fidel")
> class(v)
[1] "character"
> is.numeric(x)
[1] TRUE
> is.numeric(v)
[1] FALSE
> is.character(v)
[1] TRUE
```



Existen muchas clases de datos en R. Veremos:

- números (numeric)
- cadenas (character)
- factores (factor)
- valores lógicos (logical)

Existen una serie de funciones para trabajar con clases.

Funciones para clases:

- o class()
- is.className(): is.numeric(), is.character()
- as.className(): as.numeric(), as.character()



```
Ejemplo
> a<-"string"
> is.character(a)
> a<-5
> is.character(a)
> is.numeric(a)
> b <-as.character(a)
> b
> is.character(b)
> z<-as.character(x)
> is.numeric(z)
[1] FALSE
> is.character(z)
[1] TRUE
 Z.
    "1" "2" "3" "4" "5"
```



Operaciones con caracteres

- ¿De qué clase es?: class()
- ¿Es de la clase carácter?: is.character()
- Convertir en carácter: as.character()
- Concatenación de cadenas: paste(a,b,sep='"')
- Extracción de subcadenas: substr(string, comienzo, longitud)

```
> c1<-"Café"
> c2<-"con"
> c3<-"leche"
> paste(c1,c2,c3)
[1] "Café con leche"
> #Si quiero quitar el separador
> paste(c1,c2,c3, sep="")
[1] "Caféconleche"
> paste(c1,c2,c3, sep="%")
[1] "Café%con%leche"
```



```
> d
[1] "Café%con%leche"
> #queremos extaer una porcion de la cadena
> substr(d, 6, 8)
[1] "con"
```



Análisis de datos en R

```
> library("foreign", lib.loc="C:/Program Files/R/R-3.1.1/library")
> dir()
 [1] "Areal"
                               "auto.dta"
                                                          "Clase2.R"
                               "Enaho01-2013-200.dta"
 [4] "CostaNorte.RData"
                                                          "Enaho2013400d.dat"
     "Enaho2013400p.csv"
                               "Enaho2013400p.sav"
                                                         "Enaho2013400p.xlsx"
[10] "Enaho2013400p0.csv"
                               "Historia4"
                                                         "hogares.csv"
[13] "hogares.dat"
                               "hogares.sav"
                                                         "hogares.xls"
[16] "hogares.xlsx"
                               "Importar.R"
                                                         "ImportarSTATA.R"
[19] "MiScript.txt"
                                                         "scan.txt"
                               "mujeres.dta"
[22] "SumariaM.csv"
                               "SumariaM.sav"
                                                          "test.csv"
[25] "test.txt"
                               "test missing.txt"
                                                         "test missing comma.txt"
> read.dta("auto.dta")
                make price mpg rep78 headroom trunk weight length turn
1
         AMC Concord 4099
                                            2.5
                                                   11
                                                        2930
                                                                 186
                                                                       40
           AMC Pacer 4749 17
                                    3
                                            3.0
                                                   11
                                                        3350
                                                                 173
                                                                       40
          AMC Spirit 3799
                                   NA
                                            3.0
                                                   12
                                                        2640
                                                                 168
                                                                       35
                                                        3250
4
       Buick Century 4816
                                            4.5
                                                                 196
                                                                       40
                                                   16
5
                                            4.0
       Buick Electra 7827
                             15
                                                   20
                                                        4080
                                                                 222
                                                                       43
                                            4.0
       Buick LeSabre
                     5788
                             18
                                                   21
                                                        3670
                                                                 218
                                                                       43
          Buick Opel
                      4453
                                                   10
                                                        2230
                                                                 170
                                                                       34
                                   NA
                                            3.0
                                            2.0
         Buick Regal 5189
                             20
                                                   16
                                                        3280
                                                                 200
                                                                       42
9
       Buick Riviera 10372
                            16
                                            3.5
                                                   17
                                                        3880
                                                                 207
                                                                       43
       Buick Skylark 4082
                                            3.5
                                                                       42
10
                                                   13
                                                        3400
                                                                 200
11
        Cad. Deville 11385
                            14
                                            4.0
                                                        4330
                                                                 221
                                                                       44
12
       Cad. Eldorado 14500
                                            3.5
                                                   16
                                                        3900
                                                                 204
                                                                       43
13
        Cad. Seville 15906
                                            3.0
                                                   13
                                                        4290
                                                                 204
                                                                       45
14
                                            2.5
                                                        2110
      Chev. Chevette 3299
                                                                 163
                                                                       34
```



Análisis de datos en R

- > data1<-read.dta("auto.dta")
- > fix(data1)

	make	price	mpg	rep78	headroom	trunk	weight	length	turn
1	AMC Concord	4099	22	3	2.5	11	2930	186	40
2	AMC Pacer	4749	17	3	3	11	3350	173	40
3	AMC Spirit	3799	22	NA	3	12	2640	168	35
4	Buick Century	4816	20	3	4.5	16	3250	196	40
5	Buick Electra	7827	15	4	4	20	4080	222	43
6	Buick LeSabre	5788	18	3	4	21	3670	218	43
7	Buick Opel	4453	26	NA	3	10	2230	170	34
8	Buick Regal	5189	20	3	2	16	3280	200	42
9	Buick Riviera	10372	16	3	3.5	17	3880	207	43
10	Buick Skylark	4082	19	3	3.5	13	3400	200	42
11	Cad. Deville	11385	14	3	4	20	4330	221	44
12	Cad. Eldorado	14500	14	2	3.5	16	3900	204	43
13	Cad. Seville	15906	21	3	3	13	4290	204	45
14	Chev. Chevette	3299	29	3	2.5	9	2110	163	34
15	Chev. Impala	5705	16	4	4	20	3690	212	43
16	Chev. Malibu	4504	22	3	3.5	17	3180	193	31
17	Chev. Monte Carlo	5104	22	2	2	16	3220	200	41
18	Chev. Monza	3667	24	2	2	7	2750	179	40
19	Chev. Nova	3955	19	3	3.5	13	3430	197	43
20	Dodge Colt	3984	30	5	2	8	2120	163	35



Análisis de datos en R

```
> class(datos[,2])
[1] "integer"
> is.numeric(datos[,2])
[1] TRUE
> table(datos[,2])
> class(datos[,3])
[1] "integer"
> table(datos[,3])
 132 1751
   for (i in 1:dim(datos)[2])
               print(colnames(datos)[i])
print(table(datos[,1]))
```



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)

