



RRS – R PER A RECERCA SOCIAL

PROJECTE D'ADAPTACIÓ DE MATERIALS DOCENTS

Versió 0.6
Abril 2014

Coordinadors: M. Ajenjo, P. López, F. Miguel
Becaris: L. Parcerisa, A. Montes

Index

Index.....	2
1: Introducció i dades.....	4
Què és R? Versions i actualització.....	4
Per què utilitzar R?.....	4
Instal·lació de R.....	5
Paquets addicionals.....	5
El CRAN (The Comprehensive R Archive Network).....	10
Paquets útils per a Sociologia.....	11
Instal·lació, càrrega i utilització d'un paquet.....	13
.....	13
1) Instal·lació.....	14
2) Càrrega.....	15
3) Utilització d'un paquet.....	16
.....	18
Què és Deducer?.....	18
Consola, Menus i Visor de dades (DEDUCER).....	20
Què és R-studio?.....	26
Instal·lació de R-Studio.....	27
2. Tipus de sortides de resultats: Consola, Gràfiques, Exportació.....	33
3. Estructura de dades en R: Dades i variables.....	35
Tipus de dades.....	35
Tipus de variables.....	37
4. Introducció de dades: Teclat, Arxius de text (CSV), Excel, SPSS, altres. 43	
Teclat.....	43
Importació de dades format Text.....	44
.....	46
Importació de dades format Excel.....	46
Importació de dades format SPSS.....	46

5. Manipulació bàsica de dades.....	51
“Missing values” a R.....	51
Manipulació de VARIABLES.....	51
Recodificar una variable.....	51
Transformar variables.....	53
Conversió de tipus.....	55
Creació de tipologies.....	55
Manipulació de CONJUNTS DE DADES.....	55
Ordenar dades.....	56
Seleccionar subconjunts de casos.....	56
Ampliar/Unir conjunt de dades.....	57
6. Emmagatzematge.....	58

1: Introducció i dades.

Què és R? Versions i actualització.

El programa R és un llenguatge de programació potent (del tipus “orientat a objectes”) i que serveix per a realitzar anàlisi estadística i representació de dades. Una definició alternativa de R és que “*és un paquet estadístic d’última generació al mateix temps que un llenguatge de programació*” (Maurandi *et al.*, 2013: 1). Cal destacar que R és un *software* lliure i obert, per tant, la seva utilització és totalment gratuïta i lliure. Actualment, aquest programa ha adquirit molta rellevància dins la comunitat científica a nivell internacional, fet pel qual és un clar candidat a esdevenir “*lingua franca*” en matèria d’anàlisi de dades i no només això, sinó que també cal destacar que s’ha implantant en moltes universitats d’arreu del món. Per altra banda, R també s’està expandint ràpidament pel món empresarial.

Per què utilitzar R?

Avui en dia podem fer estadística de moltes maneres, que van des de la utilització dels múltiples i diversos paquets estadístics¹ a la utilització de les clàssiques fulles de càlcul (MS-EXCEL, OpenOffice Calc,...).

Segons (Maurandi *et al.*, 2013) els principals avantatges de R són:

- És lliure i la seva distribució es duu a terme sota la llicència GNU, de manera que hom pot utilitzar-lo i, a més a més, modificar-lo i millorar-lo.
- Aquest és un programa multiplataforma i, per tant, existeixen versions per Mac, iPhone, Linux, Windows, etc.
- A través del programa R, és possible analitzar qualsevol tipus de dades.
- Té una capacitat gràfica molt bona en relació als altres paquets estadístics.
- Destaca per la seva compatibilitat amb “*tots els formats de dades*” (txt, dat, csv, xls, sav, sas...), per tant es pot utilitzar per analitzar diverses bases de dades.
- És ampliable mitjançant *extensions* (hi existeixen anàlisis específics per a infinitat de disciplines).
- Actualment ja existeixen milers de tècniques estadístiques que s’han implementat en R, i segueixen augmentant dia rere dia.

Per què R i no un altre programa?

El programa R, al igual que el programa SAS, es un programa basat en comandaments, que permet a l'usuari un major control dels procediments executats, en comparació al programa SPSS. A més del clar avantatge del cost zero de R versus els altres programes, les sortides de processos que ofereix R són concises i deixen a l'usuari l'opció de sol·licitar un major nivell de detall, afavorint una millor pràctica en l'ús de l'estadística i evitant la temptació de tractar d'interpretar tots els estadístics que apareixen en la sortida.

En resum, la gran versatilitat dels procediments estadístics disponibles, la capacitat de produir gràfics de qualitat i l'àmplia documentació gratuïta, entre altres aspectes, fan de R un excel·lent programa estadístic per ser usat tant en docència com en recerca. La gratuïtat de R a més, permet accedir lliurement a un programa d'alta qualitat, així com la transparència en la construcció de R permet un major control del procés de generació de coneixement per part dels usuaris (Salas, 2008)

¹ http://en.wikipedia.org/wiki/List_of_statistical_packages

Table 1. Comparison of general features of the statistical software SPSS, SAS, and R.

Aspecto	Programa estadístico		
	SPSS	SAS	R
Amigabilidad con el usuario	Excelente	Baja-Regular	Baja-Regular
Manipulación de datos	Baja	Buena	Buena
Calidad de gráficos	Regular	Buena-Excelente	Excelente
Control de procesos	Baja	Excelente	Excelente
Costo	US\$ 1500	US\$ 7200	Gratis
Código fuente disponible	No	No	Sí
Variedad análisis estadísticos	Buena	Buena-Excelente	Excelente
Documentación	Excelente	Buena	Buena-Excelente
Soporte técnico	Buena	Buena	Baja
Sistema operativo	Windows*	Windows*	Windows*
		Macintosh*	Macintosh*
		Linux	Linux

Salas, C (2008) ¿Por qué comprar un programa estadístico si existe R?

Instal·lació de R

La instal·lació difereix segons l'entorn o sistema operatiu utilitzat. En Windows es un arxiu d'instal·lació executable, en Linux es tracta d'executar una sèrie de comandaments en consola, etc... Totes les opcions de descarrega es poden trobar a la pàgina **web de CRAN** (The **C**omprehensive **R** Archive **N**etwork): <http://cran.es.r-project.org/>

A partir d'aquí:

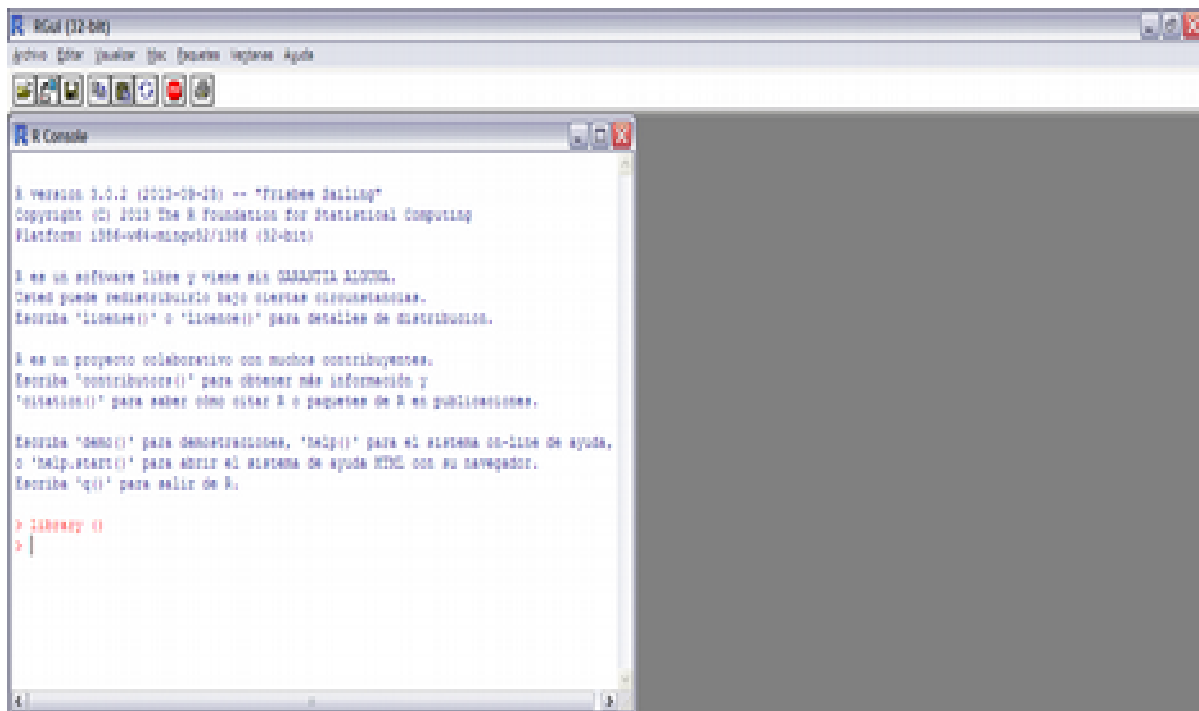
- Per a Windows: "Download R for Windows" i seleccionar l'enllaç "base". "Download R 3.0.2 for Windows (52 megabytes, 32/64 bit)" és un executable, el baixem i l'instal·lem. La versió actual (abril 2014) es 3.1.0, però aquest enllaç de descarrega sempre apuntarà a la versió més actual de R.
- Linux: `sudo apt-get update; sudo apt-get install r-base` (en més detall es pot llegir: <http://numerorojo.wordpress.com/2008/04/27/instalar-r-en-ubuntu/>)
- En Mac OS: Consulteu <http://cran.es.r-project.org/bin/macosx/>

Paquets addicionals

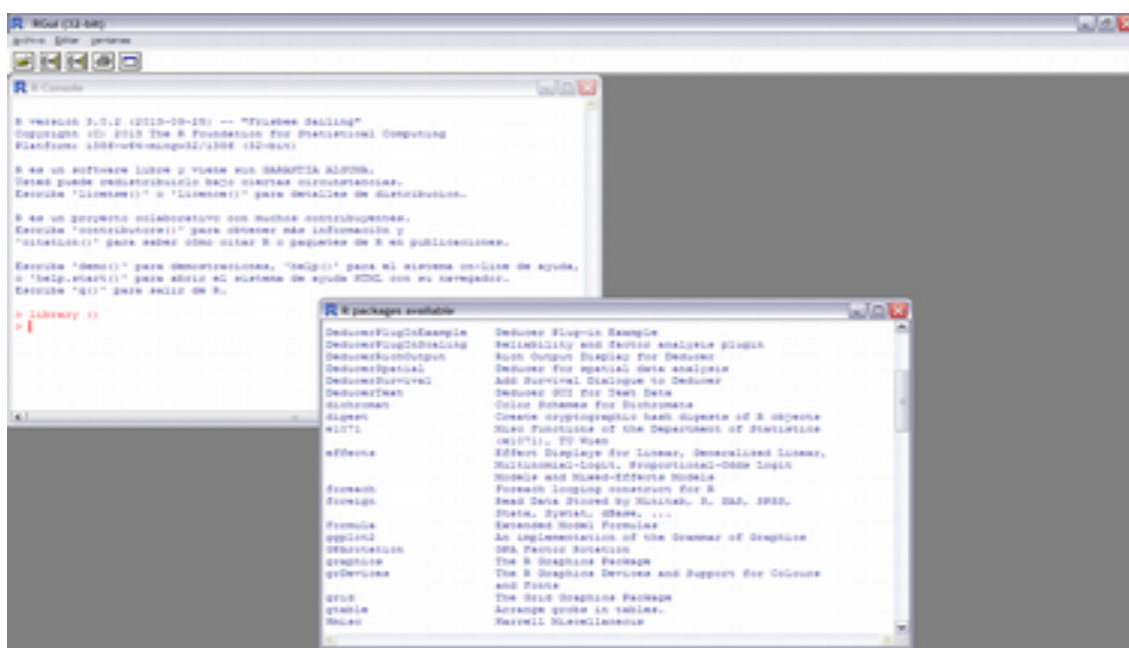
Si bé és cert que amb la simple instal·lació de R hom ja disposa de moltes possibilitats, cal advertir de l'existència d'un gran nombre de mòduls de caràcter opcional, que s'anomenen paquets (*packages*). Els paquets "són col·leccions de funcions i dades", i s'emmagatzemen a un directori del nostre PC que s'anomena *library*.

Paquet = noves funcions per R + dades d'exemple

Per tal de visualitzar els diferents paquets que hom ha instal·lat cal utilitzar la instrucció: **library ()**. En aquest sentit, serà necessari introduir la instrucció **library ()** a dins la consola i pulsar la [intro]. Una vegada haguem fet això se'ns obrirà una finestra anomenada "R packages available" on podrem visualitzar els diferents paquets.



1 Vista Consola de R



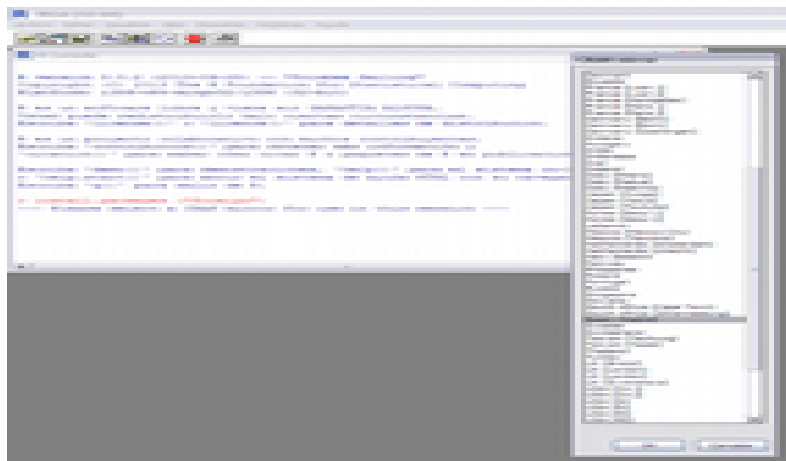
2 finestra R packages available

En aquest sentit, cal remarcar la distinció entre instal·lar i carregar un paquet, ja que no és el mateix. Els paquets instal·lats s'han descarregat des de CRAN al disc dur del nostre PC i estan disponibles per poder ser utilitzats. Els paquets carregats són el subconjunt de paquets instal·lats que s'han “activat”, tot carregant-los a la memòria RAM del PC per a la seva utilització immediata.

Instal·lar paquet = WWW(CRAN) -> PC Carregar paquet = PC -> RAM
--

Instal·lem un paquet amb la instrucció: `install.packages("nom_paquet")` si és el cas que en coneixem el nom, altrament fem servir `install.packages()` per accedir a un llistat de paquets disponibles². Aleshores se'ns obrirà el CRAN mirror.

Quan s'obre la pantalla CRAN mirror, el que està demanant és que hom esculli el servidor repositori a través del qual es descarregarà el paquet en qüestió. En aquest sentit, és bastant indiferent el repositori de R que s'esculli. Una vegada s'hagi seleccionat el repositori cal clicar **OK** i el programa R iniciarà la descarrega del paquet. Si no ens surt cap missatge d'error, significa que s'ha descarregat correctament.

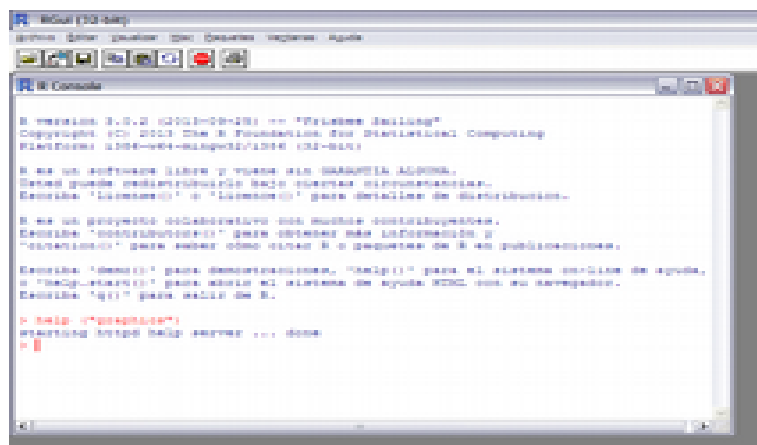


3 CRAN mirror

En cas que ens surti un error és possible que no s'hagi escrit correctament el nom de la funció "install.packages" o el del paquet. No obstant, si hem escrit la instrucció correctament i ens surt que no està disponible, és possible que haguem d'actualitzar la versió de R instal·lada al ordinador: Alguns paquets deixen de funcionar quan s'actualitza d'una versió de R a la posterior, i molts paquets tenen dependències d'altres paquets.

Per tal de saber quins paquets hem instal·lat al PC **i, a més, estan carregats al sistema**, a punt per ser utilitzats, existeix la instrucció: `search ()`

² CRAN s'actualitza sovint: apareixen (i desapareixen) paquets constantment, donat que un gran nombre de desenvolupadors treballen per a crear o millorar els paquets per a R. El desenvolupament "obert" del sistema R implica un grau elevat de heterogeneïtat en la qualitat i una evolució "caòtica" d'aquest sistema descentralitzat, com es habitual a l'àmbit del FOSS (*free and open-source software*).



6 Consola R - cerca d'informació de paquets

Una vegada introduïda la instrucció, si tenim connexió a WWW, automàticament se'n obrirà el navegador WWW per defecte al nostre PC i una pestanya amb la descripció del paquet i múltiples informacions complementàries.



7Exemple descripció de paquet

S'ha de tenir en consideració que la instal·lació de paquets només es pot realitzar amb connexió a internet. Un cop instal·lats, aquests sí que es podran seguir utilitzant sense connexió.

El CRAN (The Comprehensive R Archive Network)

Els paquets que hi ha per a R els podem trobar a <http://cran.r-project.org/>. CRAN és un repositori oficial de paquets disponibles per a R, on actualment s'hi troben més de 5.000 paquets, que els diversos usuaris generen de forma oberta i col·laborativa, i que engloben una gran diversitat de temàtiques.



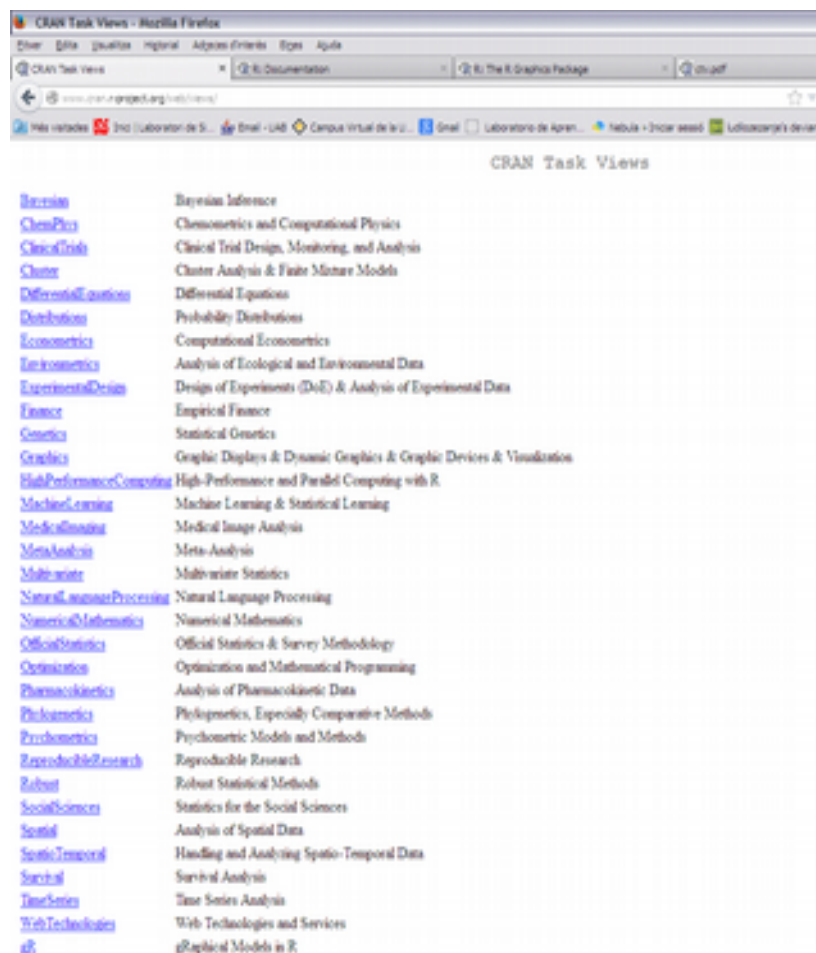
8. Pàgina d'inici CRAN-R-Project

CRAN està “replicat” en múltiples servidors mirall arreu del món, per a distribuir la càrrega de treball. En demanar la descàrrega/instal·lació d'un paquet, R pot demanar seleccionar des d'una llista de servidors: la preferència és el que estigui més proper (Spain, Madrid).

Degut a la diversitat creixent de temàtiques i paquets el CRAN s'organitza per “vistes” o temes, agrupant així els paquets segons la seva funció o naturalesa. D'aquesta forma es poden descarregar/instal·lar col·leccions completes de paquets d'utilitat per a diverses disciplines.

Existeixen col·leccions de paquets per desenvolupar treballs concrets o per àrees de treball específiques que s'anomenen *Task Views*. Com ens recorden (Maurandi et al., 2013: 9), “*instal·lant una Task View automàticament conjunts sencers de paquets s'instal·len sense haver d'anar instal·lant paquet a paquet*”.

Es pot consultar el llistat de Task Views a < <http://cran.es.r-project.org/web/views/>>



9. Task Views, al CRAN

Per tal d'instal·lar de manera automàtica aquestes *Views*, s'ha de procedir a la instal·lació prèvia del paquet **ctv**, per exemple, a través dels procediments de (1) descarrega: **install.packages ("ctv")** i, a continuació, (2) instal·lació a RAM: **library ("ctv")**. Una vegada haguem fet això, les *Views* es poden instal·lar a través del procediment **install.views** o bé mitjançant **update.views**.

Paquets útils per a Sociologia.

Dins del *Task View* de *Social Sciences* disponible al CRAN, destaquem aquests paquets, d'entre molts altres existents, com a paquets útils per a la pràctica sociològica. Si bé, aquests paquets s'actualitzen de forma molt ràpida i diversa, podent quedar obsolets en poc temps.

- **Brglm**: aquest paquet conté diverses eines per ajustar els models lineals amb resposta binominal.
- **Class**: aquest paquet conté diverses funcions de classificació.
- **Dynlm**: aquest paquet possibilita realitzar models lineals i de regressió.
- **Effects**: aquest paquet permet realitzar efectes i visualitzacions en gràfiques i tabulacions.
- **Gam**: aquest paquet permet desenvolupar models additius.
- **Geepack**: aquest paquet permet realitzar equacions d'estimació generalitzades i estructures de correlació.
- **Gnm**: aquest paquet permet realitzar funcions i models no lineals.

- ****Hmisc**: aquest paquet permet realitzar múltiples funcions útils per l'anàlisi de dades, gràfics, càlcul de mida de mostra, importació de conjunts, etc.
- **Lattice**: aquest paquet permet desenvolupar un sistema de visualització de dades d'alt nivell amb un èmfasi en les dades multivariades.
- **Lmtest**: aquest paquet conté una col·lecció de proves, conjunts de dades i exemples per a la comprovació de diagnòstic en els models de regressió lineal.
- ****Mitools**: aquest paquet és una eina d'anàlisis simples i múltiples.
- **Mvnmle**: aquest paquet permet cercar l'estimació de versemblança i la matriu de variància-covariància amb valors perduts.
- **Network**: aquest paquet permet realitzar anàlisi de xarxes.
- **Optmatch**: aquest paquet conté eines per desenvolupar funcions de matching.
- **Pls**: aquest paquet permet realitzar regressions multivariades i anàlisis de components principals.
- **Rgl**: aquest paquet permet realitzar gràfics 3D.
- **Robustbase**: aquest paquet permet realitzar càlculs de la "robustesa".
- ****Survey**: aquest paquet permet realitzar estadística descriptiva, models lineals i acumulatius, log-lineal, gràfics, etc.
- **Survival**: aquest paquet permet realitzar estadístics descriptius, etc.
- ****Zelig**: aquest paquet fa estimacions i ajuda a interpretar els resultats d'una ampla gamma de models estadístics. Els propis autors el defineixen com "everyone's statistical software".

Nota:

(**) Paquets que poden ser especialment interessants per a científics socials.

Això no vol dir que exclusivament els paquets disponibles dins de la *Task View* de *Social Sciences* siguin útils, així com que alguns dels que surten però no hem citat aquí tampoc ho puguin ser. Aquí tenim tota una sèrie de paquets útils disponibles fora de la *Task View Social Sciences*.

A la *Task View* anomenada *Cluster* trobem tota un conjunt de paquets de càlcul i gràfics. A continuació es presenten dos exemples dels diferents paquets que podem trobar dins d'aquesta *Task View*:

- **Amap**: aquest paquet ens proporciona implementacions alternatives de k-means i agrupacions jeràrquiques d'aglomeració.
- **Factoclass**: el paquet factoclass realitza una combinació de mètodes factorials i anàlisi de conglomerats.

`survey`: analysis of complex survey samples **Nom paquet**

Summary statistics, two-sample tests, generalised linear models, cumulative link models, unequally weighted survey samples. Variances by Taylor series linearisation or replicate direct standardization. PPS sampling without replacement. Principal components, factor

Descripció

Version: 3.29-5 **Versió**

Depends: R (> 2.10.0)

Suggests: [foreign](#), [survival](#), [MASS](#), [splines](#), [KernSmooth](#), [hexbin](#), [miscTools](#), [lattice](#)

Enhances: [edfWindex](#), [survey](#)

Published: 2013-06-12

Author: Thomas Lumley **Autor**

Maintainer: Thomas Lumley <tlumley at u.washington.edu>

License: [GPL-2](#) | [GPL-3](#)

URL: <http://faculty.washington.edu/tlumley/survey/>

NeedsCompilation: no

Citation: [survey citation info](#)

Materials: [NEWS](#)

In views: [OfficialStatistics](#), [SocialSciences](#), [Survival](#)

CRAN checks: [survey results](#)

Downloads: **Enllaços de descarrega**

Reference manual: [survey.pdf](#)

Vignettes: [Estimates in subpopulations](#)
[Two-phase designs in epidemiology](#)
[Obsolete formulas for two-phase variances](#)
[Analysing PPS designs](#)
[A survey analysis example](#)

Package source: [survey_3.29-5.tar.gz](#)

MacOS X binary: [survey_3.29-5.tgz](#)

Windows binary: [survey_3.29-5.zip](#)

Old sources: [survey archive](#)

10 Exemple paquet Survey

També hi han *Task View* d'utilitat, com ara la de **Distributions** (on trobem paquets per calcular Chi, Eta, i altres índex d'associació), o d'altres per la realització de gràfics (destacant els paquets de **ggplot**, **rgraphics**, etc.).

El desenvolupament del repositori CRAN segons la lògica FOSS fa complicat conèixer la totalitat de paquets disponibles i d'utilitat. La utilització quotidiana de R implica la consulta de fonts diverses de coneixement expert respecte de les tècniques concretes.

És pràctica comuna la cerca a buscadors WWW per paraules clau (e.g. "r package logistic regression")

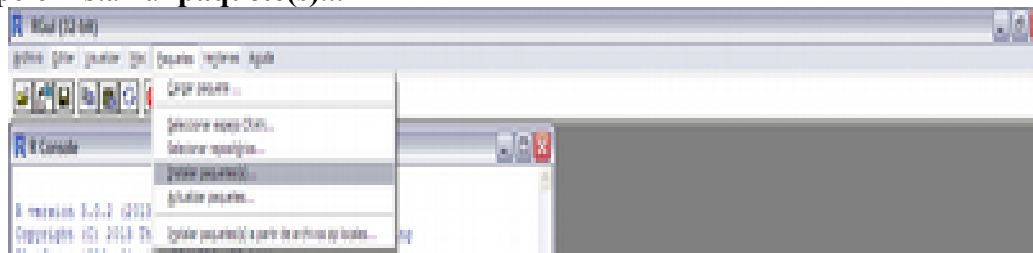
Instal·lació, càrrega i utilització d'un paquet.

A banda de les instruccions escrites directament (tipus llenguatge de programació com s'ha vist fins ara), a la part superior de la consola d'execució de R podem trobar un sistema de menús que ajuda a gestionar el programa i a seleccionar les instruccions pertinents per a diferents tasques.

A continuació, s'utilitzarà aquest sistema de menús per a la instal·lació, càrrega i utilització de paquets.

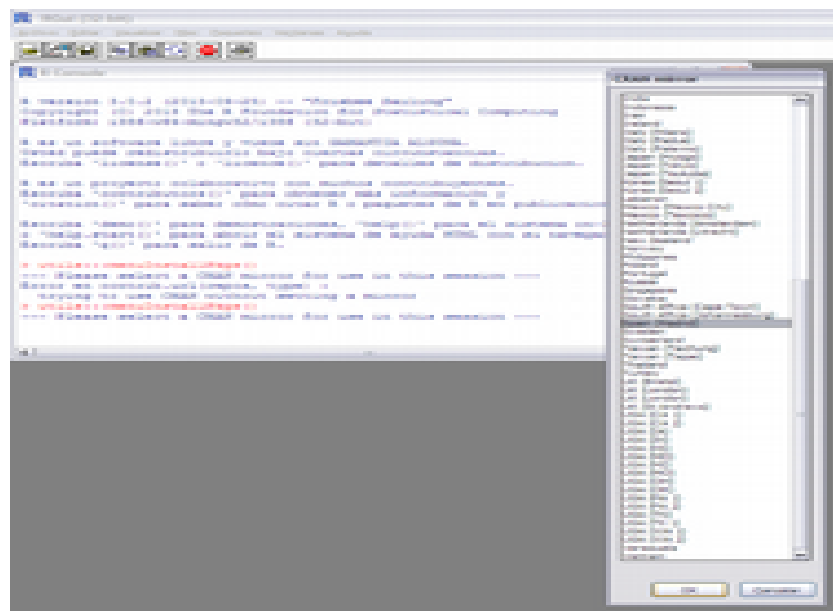
1) Instal·lació

Un cop determinat el paquet que volem utilitzar el primer de tot és descarregar-lo i instal·lar-lo. Per fer-ho ens adreçarem a la finestra “**Paquets**” i clicarem damunt de l’opció instal·lar **paquete(s)**...



11. Pas 1-Instal·lar paquets

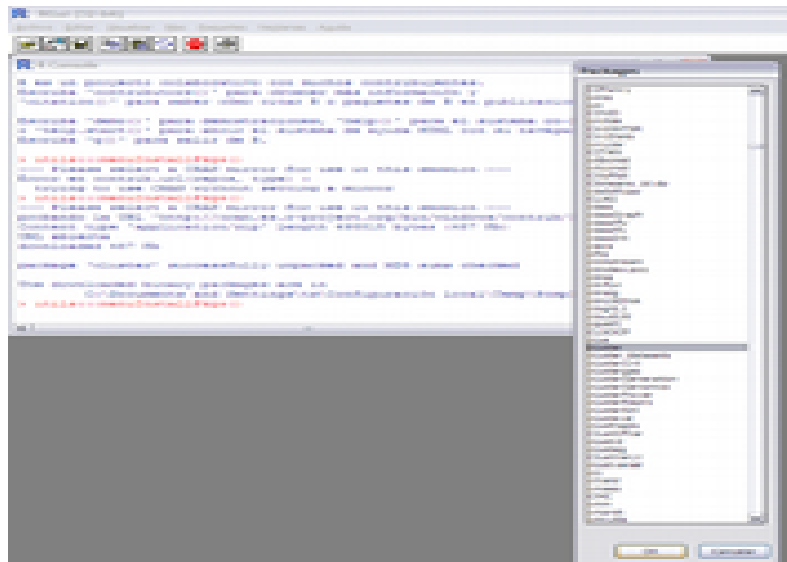
Seguidament, com ja s’ha explicat anteriorment, s’obrirà la finestra CRAN mirror. En aquest cas, haurem de clicar el repositori més pròxim, que en el nostre cas és el de **Spain (Madrid)** i clicarem **OK**.



12. Pas 2 - Instal·lar paquets >Cran mirror

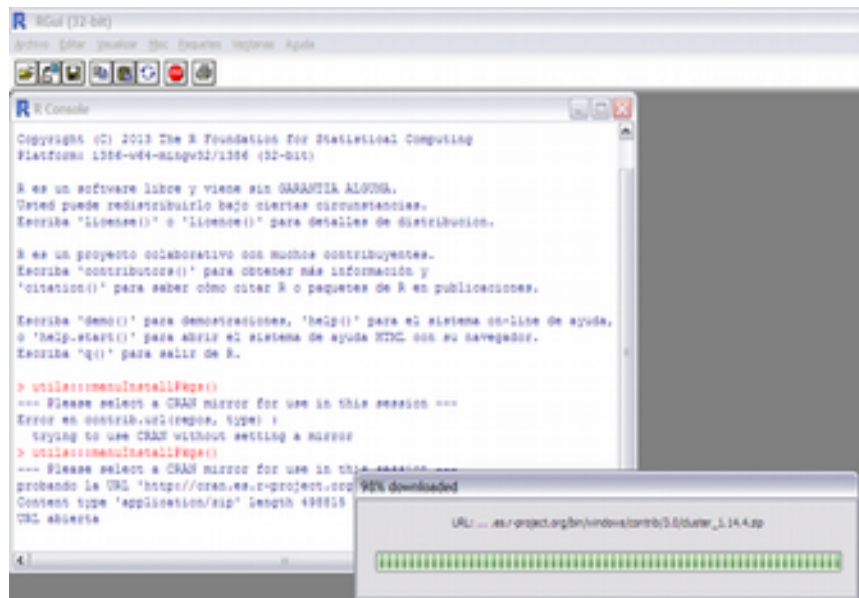
Quan haguem finalitzat el pas anterior, se’ns obrirà una nova finestra anomenada **Packages**. En aquesta finestra es poden visualitzar tots els paquets disponibles per R. Aleshores es tractarà de seleccionar el que ens interressi i clicar **OK**.

Per exemple, provem de descarregar des del CRAN al nostre disc dur el paquet **cluster**.



13. Pas 3 - instal·lar paquets > Cran Mirror > Packages

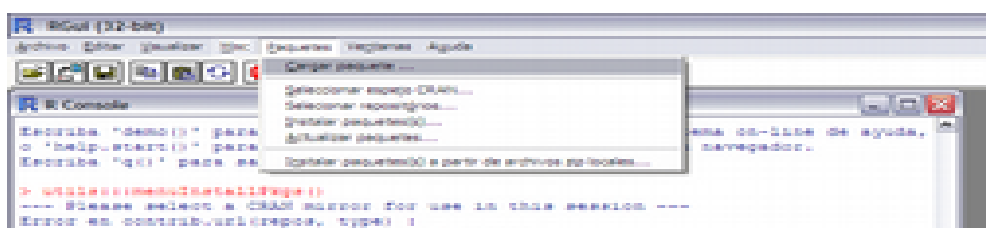
Posteriorment apareixerà una barra de descàrrega.



14. Exemple descàrrega paquet

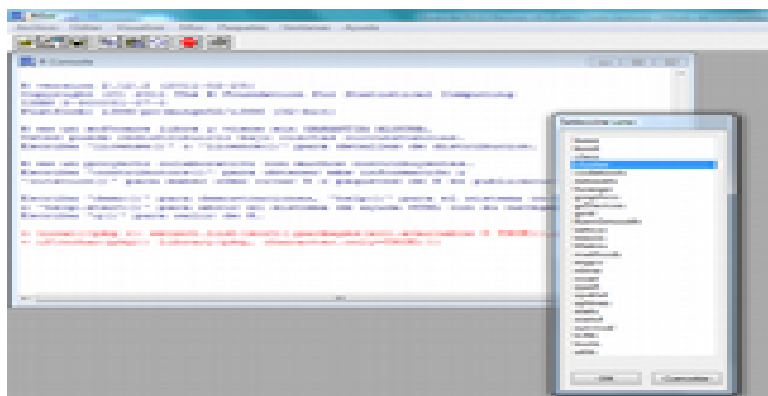
2) Càrrega

Quan haguem fet això, caldrà que carreguem el paquet a la memòria RAM, per poder-lo utilitzar. Per fer-ho ens adreçarem a la pestanya **Paquetes** i clicarem damunt de **Cargar paquete**.



15. Pas 1 - Carregar un paquet: Paquetes > Cargar paquete

Després de clicar a carregar paquet se'ns obrirà una nova finestra, on visualitzarem tots els paquets que tenim instal·lats i caldrà que seleccionem el que volem utilitzar i, seguidament cliquem **OK**.



16 Pas 2 - Carregar un paquet: Paquetes > Cargar paquete > Seleccionar paquete

3) Utilització d'un paquet.

Una vegada haguem instal·lat i carregat els paquets ja podrem començar a fer ús dels complements que hem afegit amb els paquets al programa.

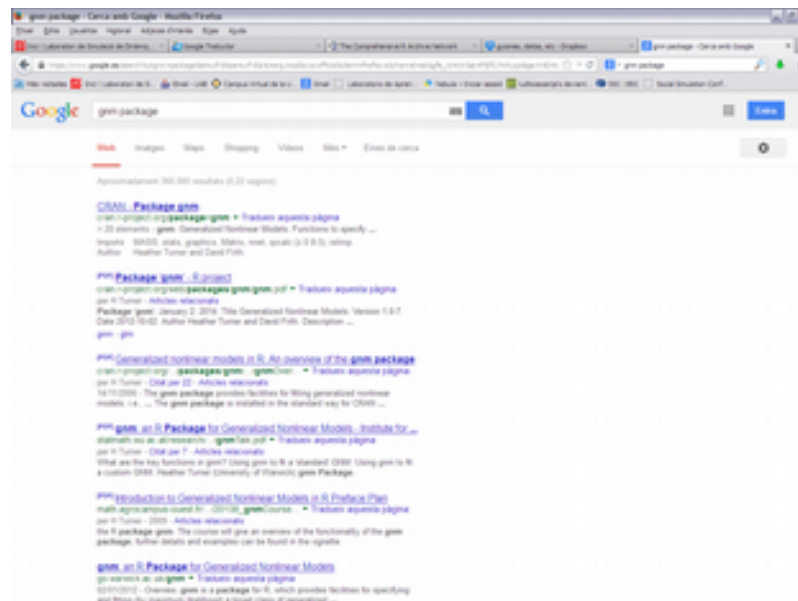
Molts paquets s'instal·len amb dades “de mostra” proporcionades pels desenvolupadors del paquet. Per exemple, el paquet *cluster* instal·la, entre d'altres, una base de dades amb el percentatge de vot republicà a cadascun dels Estats a eleccions USA des de 1856 fins a 1976 (es pot visualitzar un conjunt de dades si s'escriu a la consola en seu nom, en aquest cas **votes.repub**).

Per a veure un exemple d'utilització d'alguna de les funcions que ha afegit el paquet *cluster* al sistema R es pot introduir la següent seqüència d'ordres a la consola:

```
data(votes.repub)
agn1 <- agnes(votes.repub, metric = "manhattan", stand =
TRUE)
agn1
plot(agn1)
```

- La primera línia especifica que es treballarà amb les dades emmagatzemades.
- La següent crea un objecte anomenat “agn1” i li dona com a contingut el resultat d'una anàlisi AGNES (Agglomerative Nesting Hierarchical Clustering) de les dades triades, amb la opció distancia Manhattan.
- La següent mostra el contingut d'aquest objecte anomenat “agn1” (text).
- La inferior mostra gràficament els mateixos resultats (gràfics).

Per tal de conèixer més bé el ventall de possibilitats que ens ofereix un paquet determinat, o bé conèixer les seves característiques podem buscar-lo a través d'un cercador.



17. Exemple cerca d'informació d'un paquet

Per a treballar amb el programa R, és imprescindible saber buscar la informació sobre els paquets. En aquest sentit, és altament recomanable consultar els paquets que t'interessin en funció de la teva especialitat i les seves funcions a través de les diferents Task Views. Per saber més sobre cada paquet, es pot buscar el nom del paquet a través d'algun cercador (com per exemple, google). D'aquesta manera podrem accedir a manuals elaborats pels autors del paquet o per experts que l'hagin utilitzat i n'expliquin el funcionament i les principals característiques.



Què és Deducer?

Deducer és un complement al programa R: un paquet desenvolupat amb l'objectiu de convertir al sistema R en una alternativa lliure a altres programaris de pagament d'anàlisi de dades tals com SAS, SPSS, JMP i Minitab. Compta amb un sistema de menús per gestionar i manipular bases de dades i analitzar-les, així com un editor de dades tipus Excel per veure i editar bases de dades.

Deducer està orientat cap a usuaris de SPSS amb nivell bàsic d'estadística.

Deducer proveeix una interfície gràfica per a usuaris de R (JGR), encoratjant als usuaris no tècnics a aprendre i realitzar anàlisis sense necessitat de conèixer el llenguatge de programació de R, alhora, que permet l'ús del mateix sense limitar les opcions als usuaris ja experts en la matèria.

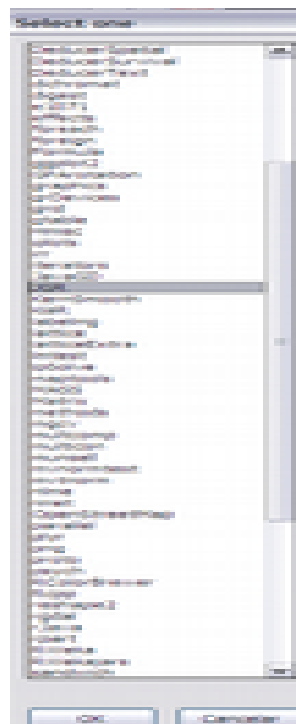
R = Programació estadística d'anàlisi
R + JGR() + Deducer = Anàlisi estadística amb R mitjançant menús

Instal·lació: El complement *Deducer* es pot descarregar gratuïtament des de:

1. el repositori CRAN, com a un paquet complementari a R, o
2. la seva pàgina web: <http://www.deducer.org>, on cal seleccionar el sistema operatiu (Windows, MacOS X o Linux) i seguir les instruccions corresponents.

Instal·lació Deducer

El paquet *Deducer* requereix la instal·lació prèvia del paquet **JGR ()**. Podem utilitzar el procediment que hem descrit anteriorment [**Paquetes > instal·lar paquets > Cran mirror: Madrid (Spain) > seleccionar JGR () > OK**]. La instal·lació de JGR() implica la instal·lació d'alguns paquets addicionals ('png', 'rJava', 'JavaGD' i 'iplots'); aquest procediment és automàtic.

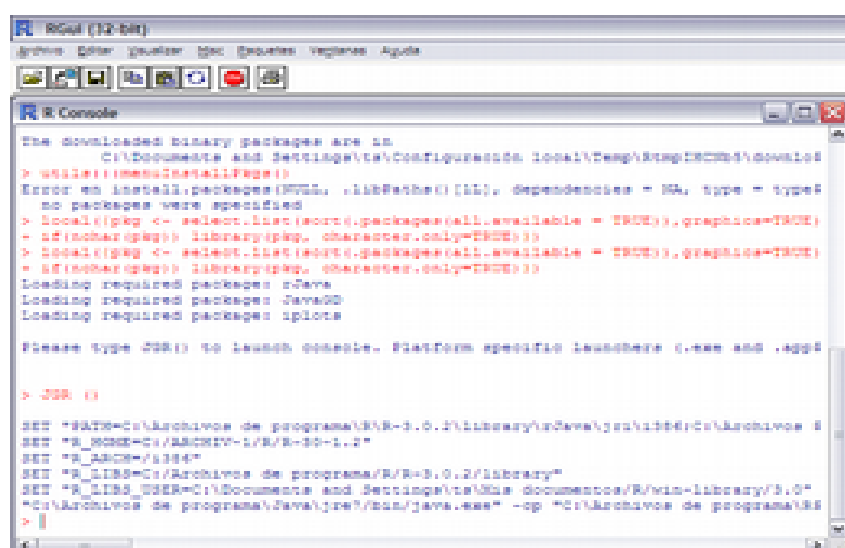


18. Exemple instal·lació JGR ()

Per a descarregar propiament Deducer (i, automàticament, les seves dependències) es pot utilitzar el procediment que hem descrit anteriorment [**Paquetes > instal·lar paquets > Cran mirror: Madrid (Spain) > seleccionar Deducer > OK**]. La instal·lació de Deducer implica la instal·lació d'alguns paquets addicionals.

Una vegada haguem descarregat Deducer, haurem d'instal·lar-lo a la RAM per al seu ús immediat [**Paquetes > Cargar paquete > Cran mirror: Madrid (Spain) > seleccionar Deducer > OK**].

Finalment, el procediment per activar la nova interfície gràfica i els nous menús implica escriure a la consola: **JGR ()** i clicant intro.



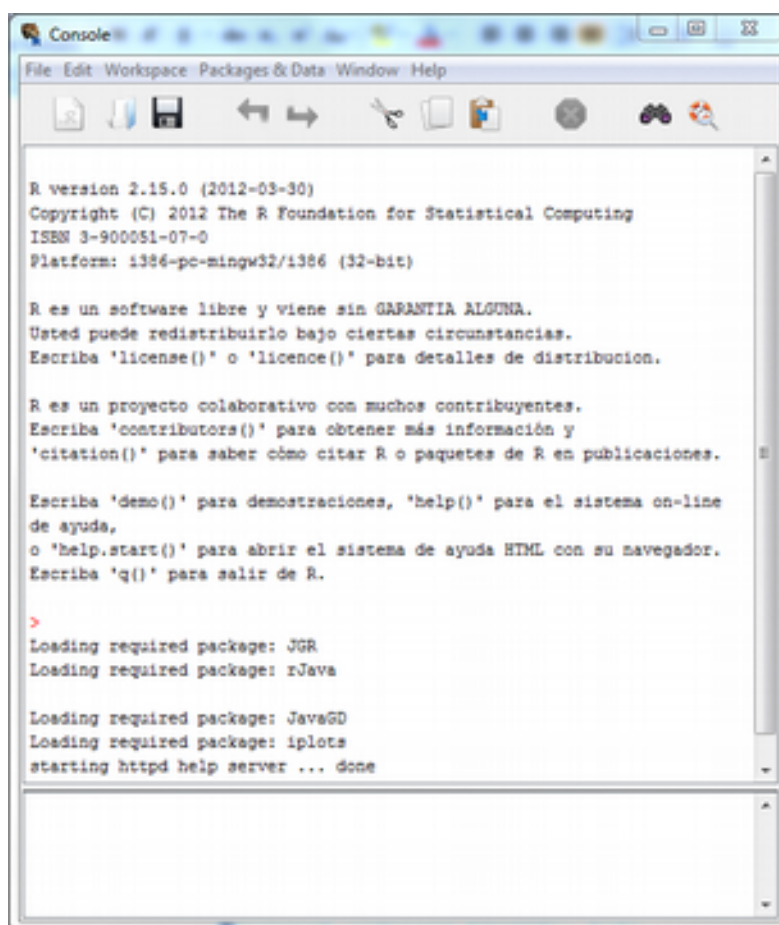
19. Exemple càrrega JGR ()

Se'ns obrirà la consola del Deducer (diferent de la de R), la qual explicarem de manera més detallada tot seguit. A més a la consola original de R s'han afegit un conjunt de nous menús, a la dreta de la opció "Ayuda": **Deducer, Data, Analysis, Plots**.

Consola, Menus i Visor de dades (DEDUCER)

Deducer s'estructura en dues finestres diferents: la consola i el visor de dades.

1. La consola recull tots els menús (part superior de la consola de Deducer, o de la consola de R -segons la versió-) per treballar amb les dades i realitzar les anàlisis estadístiques, alhora que també és per on es poden introduir els comandaments manualment (part inferior). A més és on es veuran els resultats de les anàlisis (part central).



20. Consola pròpia del Deducer



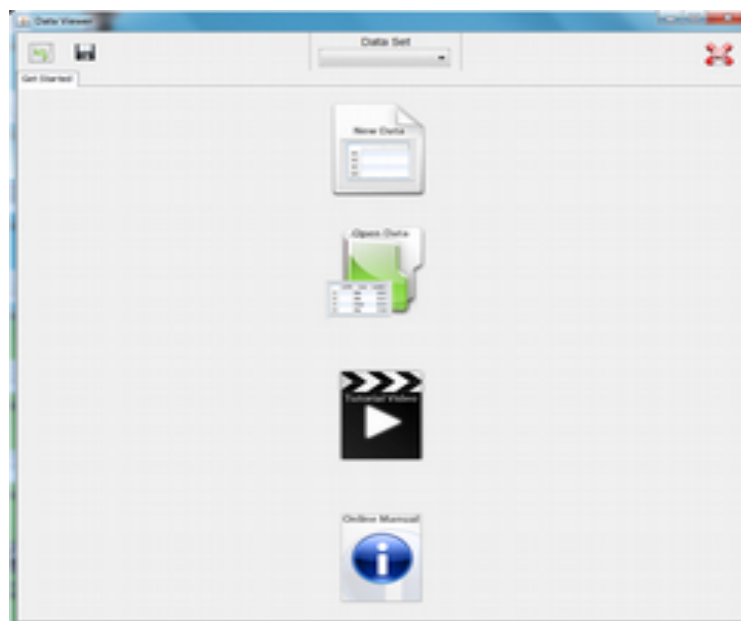
21. Consola de R (amb Deducer carregat)

En el menú de la consola és on tenim les funcions de treball (DEDUCER: Workspace / Packages & Data, R-CONSOLE: Deducer / Data / Analysis / Plots) que son les que

utilitzarem per realitzar les modificacions pertinents a l'espai de treball, modificar les dades prèviament als anàlisis, realitzar els diversos càlculs i funcions, generar gràfics i/o carregar els paquets que creiem pertinents.

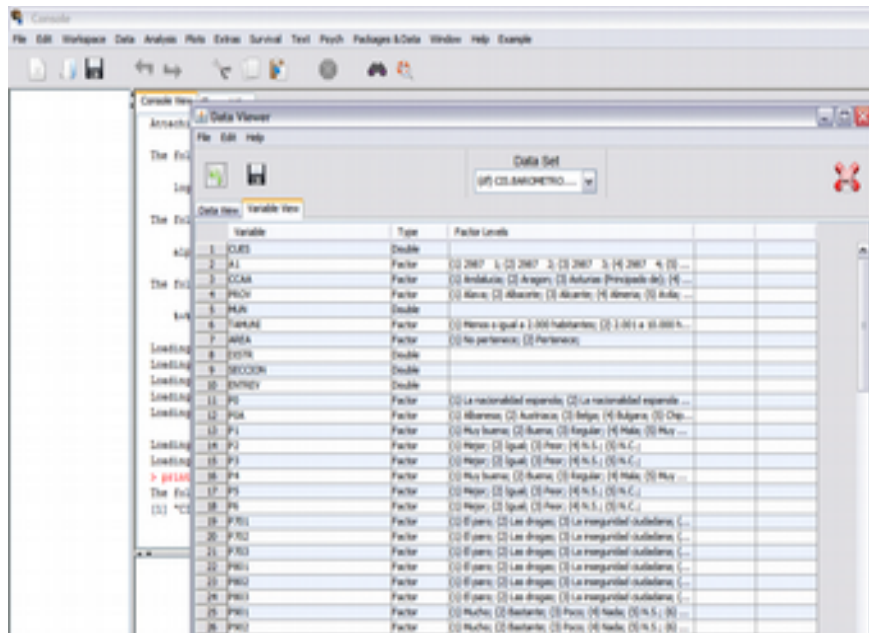
És important tenir en compte que els missatges d'avís surten reflectits exclusivament a la Consola (en forma de sintaxis) i que per tant s'ha d'estar atent per què no passin desapercebuts missatges rellevants, com per exemple, la invalidació d'algun procediment.

2. El visor de dades, molt semblat a un full de càlcul d'Excel, és un segon espai on es poden veure i editar les dades i les variables amb les quals anirem a treballar. Prèviament a l'obertura d'aquest sorgeix una primera interfície on podrem escollir entre obrir una nova base de dades o carregar una ja construïda. Des d'aquí es on obrirem la nostra matriu principal amb la que treballarem.



22. Data Viewer (crear matriu o carregar bases de dades).

Un cop carregada una base de dades s'obre el Visor de Dades (vista de dades i vista de variables) amb la següent forma:



23. Exemple Vista de variables

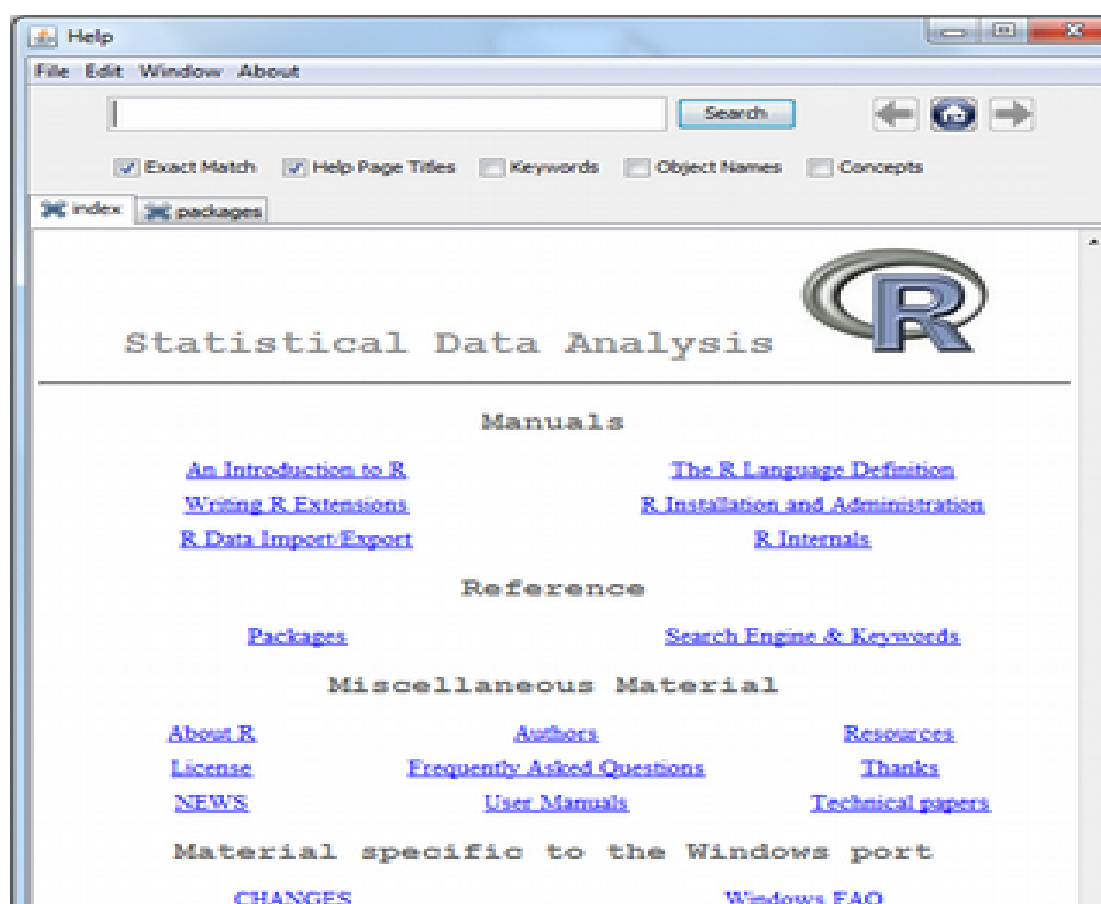
Exemple: Si hem carregat prèviament el paquet “cluster” disposem de la base de dades “votes.repub” amb el % de vot republicà, per a cada Estat, a diverses eleccions. Si volem visualitzar-lo al visor haurèm de

- (1) carregar-ho a l'espai de dades: `data("votes.repub")`, i
- (2) obrir el visor de dades amb el menú Deducer / Data viewer

La interfície *DEDUCER* afegeix a R funcionalitats per a la realització de les següents tasques:

- Carregar dades des d'arxius originals en diversos formats (txt, CSV, SPSS, etc.).
- Visualitzar les dades i els tipus de variables en el visor de dades per separat.
- Realitzar transformacions de les dades (recodificació, editar funcions, transformacions, traslladar, fusionar).
- Anàlisi estadística (comparació de mitjanes, taules de contingència, anàlisi de regressió).
- Una interfície gràfica d'usuari per a la creació de gràfics utilitzant el paquet **ggplot2**.

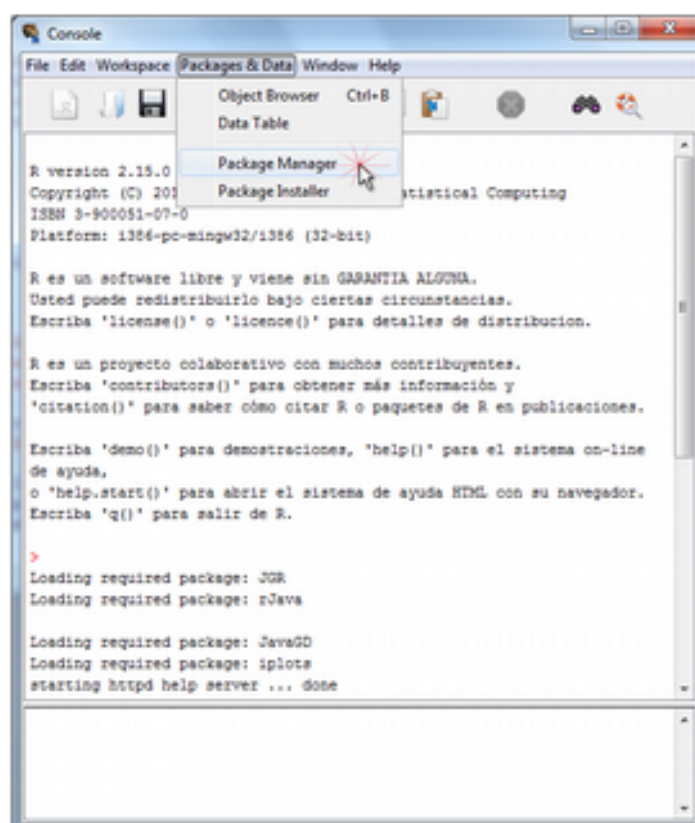
Deducer incorpora una pestanya d'ajuda (*Help*) que obre en una nova pestanya un repositori de materials d'ajuda i guia en el desenvolupament de les diverses tasques de R. Aquest, a més d'informació en torn a les tasques pròpies de R també afegeix informació corresponent a l'ús del programa *Deducer* i dels diversos paquets instal·lats en el programari.



24. Vista de recursos de la pestanya d'ajuda.

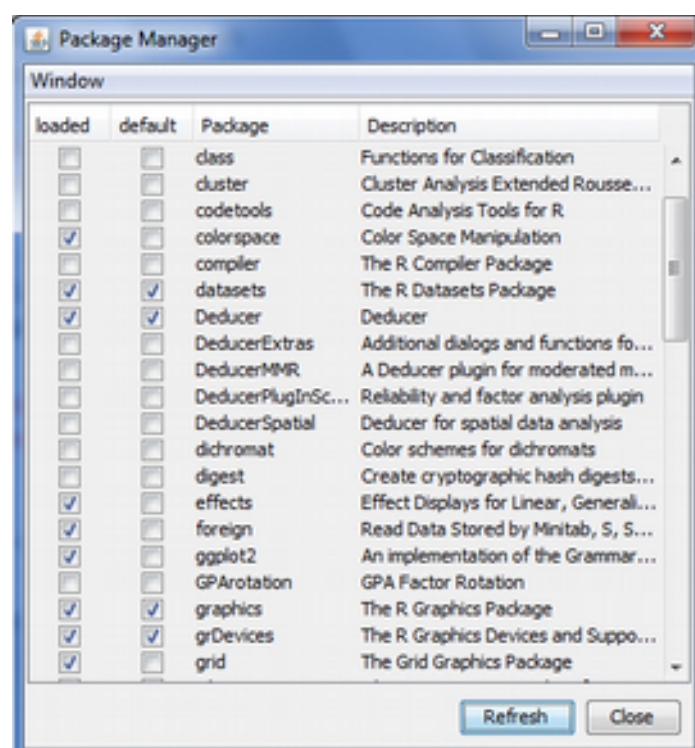
Gestió de paquets amb Deducer

Per tal de gestionar i carregar paquets amb Deducer hem d'anar a **Packages & Data** i clicar damunt de **Package Manager**.



25. Exemple càrrega Package Manager

Seguidament ens apareixerà la finestra del Package Manager, a través de la qual podrem carregar i gestionar els paquets que tenim descarregats per a R.



26. Package Manager



Què és R-studio?

RStudio és una GUI (Graphical user interface) per a R, multi-plataforma (Windows, Linux i Mac). Aquesta aglutina en una mateixa pantalla tots els entorns de treball de R, i manté la filosofia de les instruccions i la programació tot aportant algunes ‘ajudes’ que milloren la eficiència i la usabilitat de R³.

RStudio està orientat cap a usuaris programadors de codi informàtic amb nivell expert en anàlisi estadística.

Entre d’altres coses, RStudio permet:

- Obrir diversos “scripts” alhora.
- Executar porcions del codi amb només marcar-ho en els script.
- Mostrar el “workspace”.
- Mostrar els objectes del “workspace”.
- Integrar l’ajuda.
- Integrar la “llibreria”.

Notes:

Script: un script és un fitxer amb instruccions (tipus la SINTAXI a l'SPSS). Aquest concepte es tradueix literalment per guió (o també arxiu d’ordres o arxiu de processament per lots), i es refereix a un programa simple, que generalment s’emmagatzema en un arxiu de text pla. Així doncs, podem entendre els scripts com “un programa que dona instruccions a un altre programa més avançat”⁴. A través de R-Studio podem obrir diferents script alhora per treballar amb diverses línies o aproximacions d’anàlisi alhora.

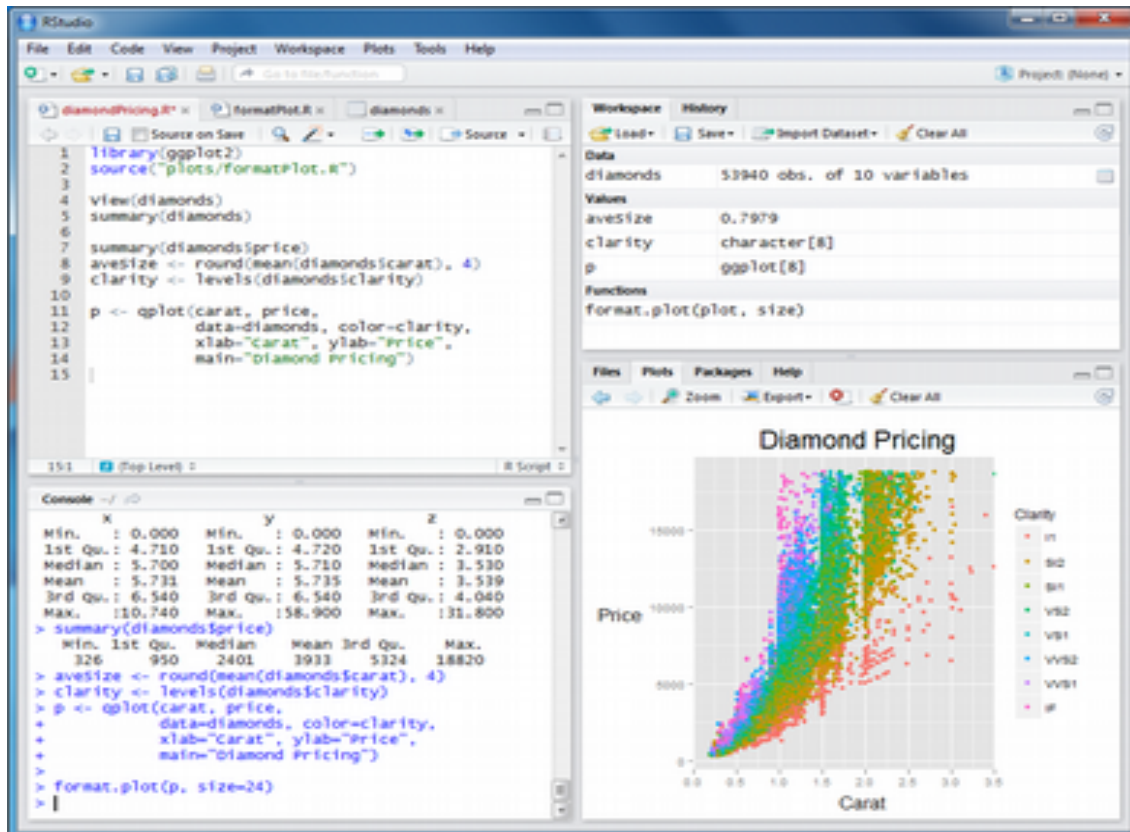
Workspace⁵: aquest concepte fa referència a l’espai de treball on generalment s’inclouen el conjunt d’objectes que han estat prèviament definits per l’usuari (més endavant s’explicarà què són els objectes. Per ara podem dir que aquests inclouen elements com les variables, els *dataframes*, els vectors, etc.) i que són emmagatzemats en una mena de memòria intermitja, mentre l’usuari està treballant amb R.

Llibreria: podem entendre la llibreria com un espai en el qual s’emmagatzemen funcions. Quan iniciem una sessió amb R s’acostumen a carregar les llibreries bàsiques. No obstant, sovint esdevé necessari carregar llibreries específiques que disposin de la implementació de funcions concretes.

³ Informació extreta de <http://www.um.es/ae/FEIR/feir-1.html>

⁴ <http://programaciondesarrollo.es/scripts-que-son-y-para-que-sirven/>

⁵ Informació extreta de <http://www.um.es/ae/FEIR/feir-1.html>



27. GUI de RStudio

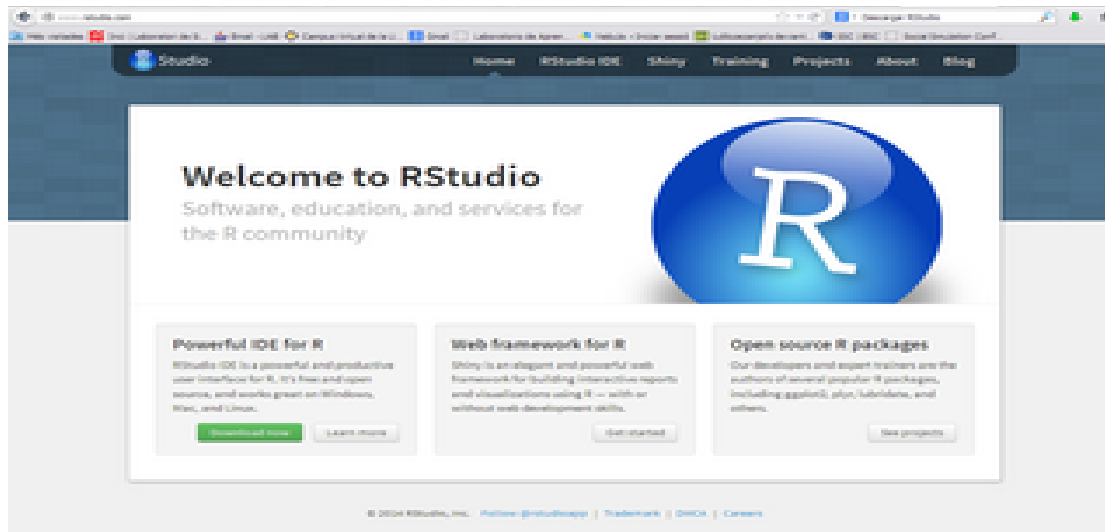
Rstudio s'organitza en quatre zones de treball diferenciades⁶, tal i com podem apreciar a la figura anterior.

1. A la zona superior esquerra, podem obrir i editar fitxers amb **codi R**, anomenats *scripts* (així com també amb altres llenguatges acceptats per R).
2. A la zona inferior esquerra podem observar-hi la consola de R, a través de la qual podem executar els **comandaments** "individuals" de R, i consultar **resultats**.
3. A la zona superior dreta hi trobem dues pestanyes:
 - Workspace: on podem observar-hi la llista dels **objectes** que s'han creat a la memòria.
 - History: que conté l'històric amb les línies de codi que s'han **executat** en R.
4. A la zona inferior dreta, hi trobem quatre pestanyes:
 - Files: aquestes donen accés a l'arbre de directoris, així com també als fitxers del **disc dur**.
 - Plots: és on apareixen els **gràfics** que hem creat en la consola.
 - Packages: aquesta pestanya facilita la gestió i/o administració dels **paquets** de R que hem instal·lat.
 - Help: és molt útil i permet obrir les pàgines **d'ajuda**.

Instal·lació de R-Studio

⁶ Informació extreta de <http://r-es.org/Introducci%C3%B3n+a+RStudio>

Per tal de descarregar-nos Rstudio hem d'anar a la web <http://rstudio.org/>. Una vegada estiguem dins la web oficial de Rstudio, hem de clicar damunt la pestanya **Rstudio IDE**.



28. Pàgina d'inici de la web de RStudio

Seguidament clicarem damunt de la pestanya **Download Rstudio**. I una vegada estiguem a la segona pantalla, clicarem, **Download Rstudio Desktop**.

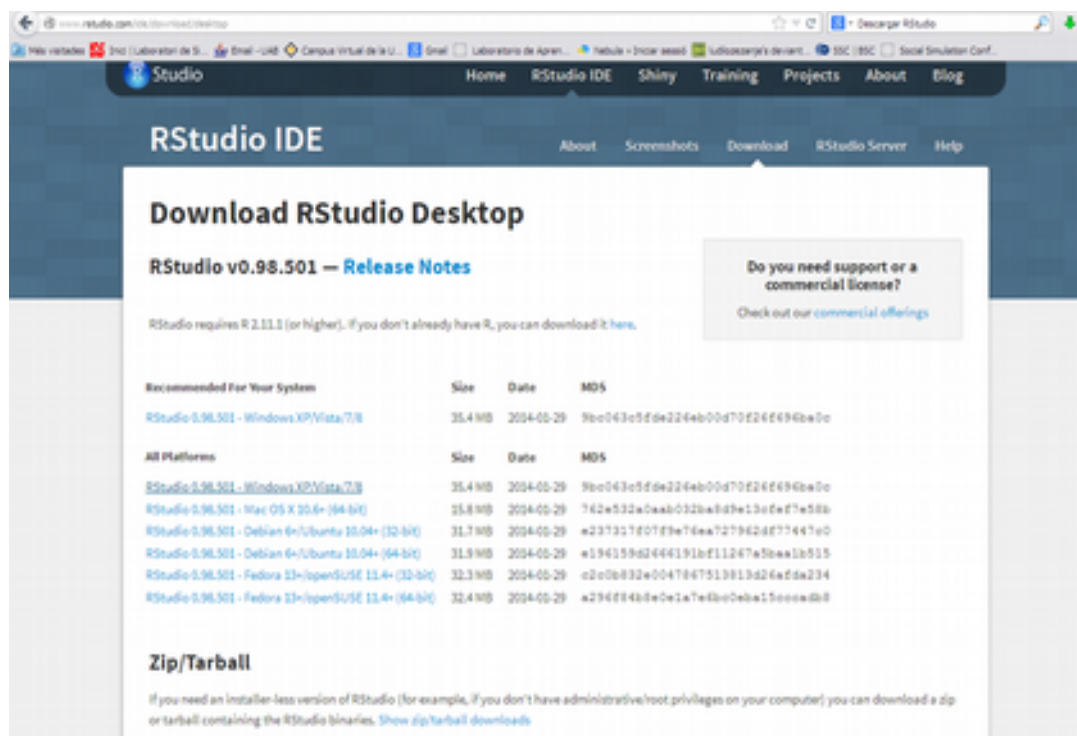


29. Pàgina de descàrrega de RStudio -1



30. Pàgina de descàrrega de RStudio -2

Quan haguem realitzat els passos anteriors, se'ns obrirà una nova pantalla des de la qual podrem seleccionar la plataforma per a la qual volem descarregar RStudio.



31. Download RStudio Desktop

Després de seleccionar la plataforma se'ns obrirà una pestanya per desar el fitxer. Clicarem damunt l'opció **Desa el fitxer** i el desarem a l'escriptori.

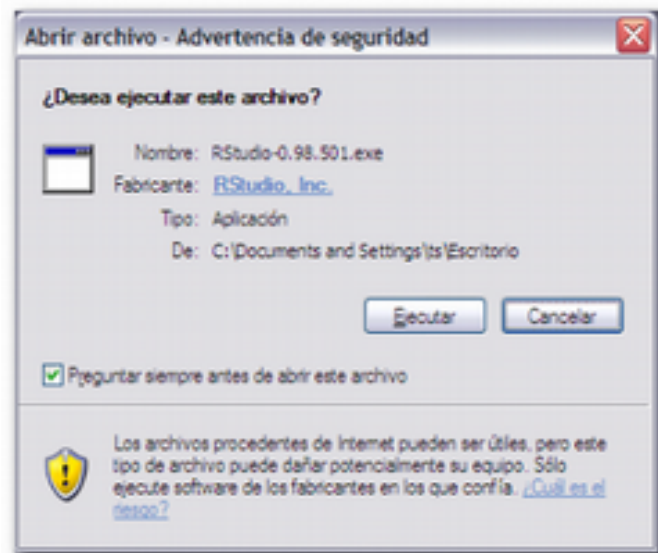


32. Pas 1- Exemple de descàrrega RStudio [Desar el fitxer]



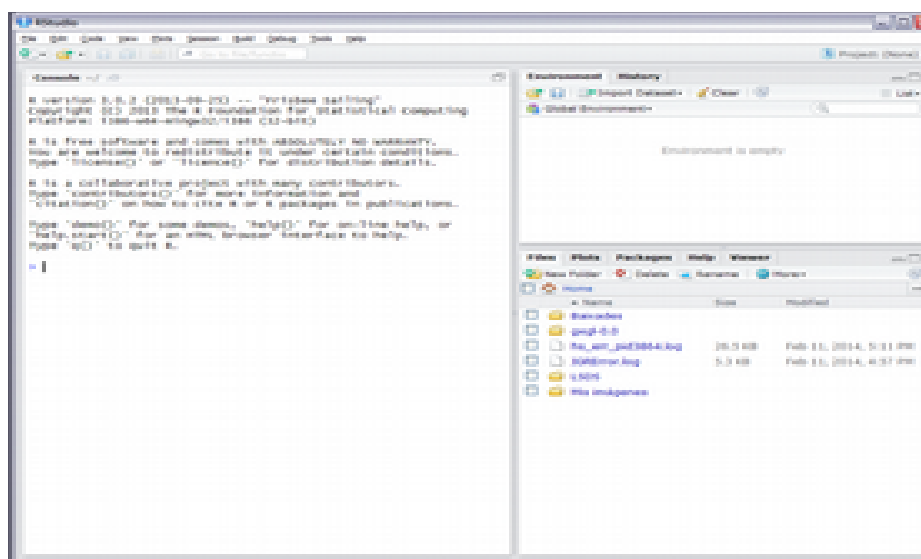
33. Pas 2- Exemple de descàrrega RStudio [Guardar fitxer]

Una vegada l'haguem descarregat, ens apareixerà el *logo* de *RStudio* a l'escriptori, farem doble clic a polsarem **Executar**. Seguidament, procedirem a la instal·lació.



34. Pas 3- Exemple de descàrrega RStudio [Executar arxiu]

Una vegada l'haguem instal·lat, ho executarem: per exemple, des de Windows, anirem a la pestanya de **Inici > Tots els programes > Rstudio > Rstudio**. Després de seguir aquest procediment se'ns obrirà una nova pantalla i ja podrem començar a treballar.



35. vista inicial - GUI de RStudio

A la segona part d'aquest manual, dedicada a la preparació de dades, s'utilitzarà el sistema de menús *Deducer*.

No obstant això, de manera gradual s'aniran introduint les instruccions de programació a consola (als requadres com ara aquest).

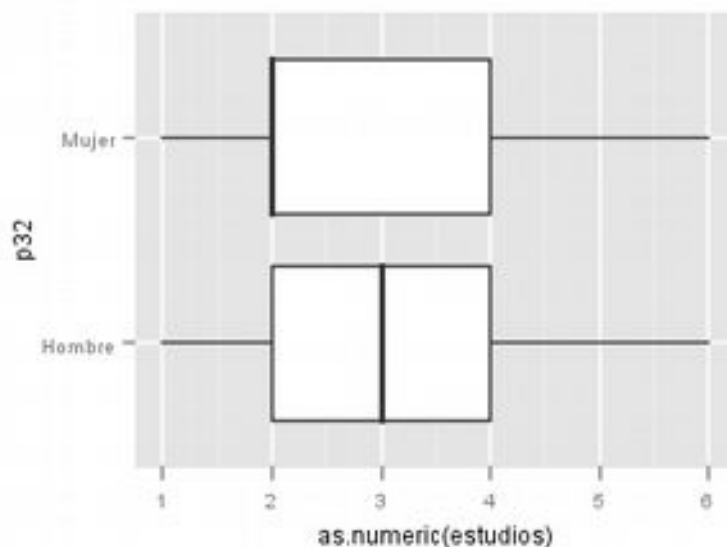
I a la tercera part, dedicada a estadística descriptiva, s'utilitzarà extensament l'entorn de programació *R-Studio*.

2. Tipus de sortides de resultats:

Consola, Gràfiques, Exportació.

En l'entorn de R, després d'escriure una instrucció i prémer la tecla [Intro], o amb *Deducer* després d'acceptar executar un procés dissenyat per menús, els resultats apareixen:

1. en la finestra de gràfics, en format d'imatge, o
2. en la consola, en format de text.



36. Exemple de resultat gràfic amb R.

A la finestra gràfica, el menú “File” exporta el gràfic en diversos possibles formats d'imatge (PDF, EPS, PNG, JPEG, BMP, TIFF) i el menú “Edit” permet copiar-lo al portapapers per a enganxar-ho a altres aplicacions.

A la finestra de terminal, o consola, el format dels resultats és text simple (conegut com a ASCII) i els resultats extensos, per exemple taules amb moltes files i columnes, queden tallades. És fàcilment exportable a qualsevol aplicació, però només es visualitza bé amb tipus de lletra monoespaiat, altrament les columnes queden trencades, etc...

Exemples de resultat de text en R/Deducer

Amb tipus estàndard (Times New) Columnes no ajustades	<pre>> frequencias(CISjunio2012[c("estudios","ocupa","condicion","estatus")], r.digits = 1) \$estudios ----- -- Frequencies -- -- -- Value # of Cases % Cumulative % 1 Sin estudios 194 7.8 7.8 2 Primaria 1057 42.6 50.4 3 Secundaria 305 12.3 62.7 4 F.P. 418 16.9 79.6 5 Medios universitarios 202 8.1 87.7 6 Superiores 304 12.3 100.0 -- -- Case Summary -- -- -- Valid Missing Total # of cases 2480 2 2482 -- -- -----</pre>
Amb tipus monoespaiat	<pre>> frequencias(CISjunio2012[c("estudios","ocupa","condicion","estatus")], r.digits = 1) \$estudios ----- -- Frequencies --</pre>

(Courier New) Columnes ajustades	--	Value	# of Cases	% Cumulative	%	--
	1	Sin estudios	194	7.8	7.8	
	2	Primaria	1057	42.6	50.4	
	3	Secundaria	305	12.3	62.7	
	4	F.P.	418	16.9	79.6	
	5	Medios universitarios	202	8.1	87.7	
	6	Superiores	304	12.3	100.0	
	--					--
	--	Case Summary				--
	--					--
		Valid	Missing	Total		
	# of cases	2480	2	2482		
	--					--
	--					--

Es recomanable instal·lar addicionalment el paquet *Deducer Richoutput* que genera resultats d'alguns procediments en format HTML (concretament, `add.cross.strata.test`, `add.test`, `contin.tests.to.table`, `contingency.tables`, `cor.matrix`, `descriptive.table`, `frequencies`, `k.sample.test`, `one.sample.test`, `print.contin.table`, `print.contin.tests`, `print.contingency.tables`, `print.cor.matrix`, `print.freq.table`, `print.multi.test`, i `two.sample.test`).

Per a instal·lar aquest paquet, que a data 29/04/2014 encara no està al CRAN sinó a un repositori "personal" del desenvolupador, cal executar, a la consola, el següent codi:

```
install.packages("DeducerRichOutput", repos="http://R-Forge.R-project.org")
```

Exemple de resultat de text en R, amb *Deducer Richoutput*:

Frequencies (catlab)

	Value	# of Cases	%	Cumulative %
1	Administrativo	363	76.60	76.60
2	Seguridad	27	5.70	82.30
3	Directivo	84	17.70	100.00

Case Summary (catlab)

	Valid	Missing	Total	% Missing
1	474.00	0.00	474.00	0.00

3. Estructura de dades en R: Dades i variables.

El sistema R emmagatzema a una memòria de treball totes les dades que son carregades, totes les instruccions, i tots els gràfics en forma “d’objectes”. Tota aquesta informació es coneix com a *Workspace*, està disponible directament per a la sessió actual. El workspace és a la memòria temporal (es perd en sortir de R), tot i que abans de sortir el programa demana si es vol desar l’espai de treball al disc dur de forma que en arrencar una nova sessió es començaria exactament on es va deixar actualment.

Els objectes més rellevants al sistema d’anàlisi R son les dades, que s’emmagatzemen en diferents formats interns.

Tipus de dades

Vectors: Són matrius d'una dimensió, que solament poden contenir valors numèrics, alfanumèrics i valors lògics.

Matriu: Una matriu és un vector amb un atribut addicional (dim), que al seu torn és un vector numèric de longitud 2 que conté la definició del nombre de files i columnes.

Dataframe: Un *dataframe* (de vegades es tradueix com a ‘base de dades’) és una generalització de les matrius on cada columna pot contenir tipus de dades diferents a la resta de columnes, mantenint la mateixa longitud. És el que més s’assembla a una matriu de dades de SPSS o de qualsevol paquet estadístic estàndard.

Crear un nou objecte vector: `a4 <- c(1,3,5,7)`

Crear un nou objecte matriu:

`A <- matrix(c(2,4,3,1,5,7),nrow=2,ncol=3,byrow = TRUE)`

Crear un nou objecte Base de Dades:

```
n = c(2, 3, 5)
s = c("aa", "bb", "cc")
b = c(TRUE, FALSE, TRUE)
df = data.frame(n, s, b)
```

Amb el sistema *Deducer* instal·lat, com a mecanisme simplificador, el Visor de Dades estructura les dades en dues pestanyes anomenades: “vista de dades” (Data View) i “vista de variables” (Variable View).

Carregarem dades per a mostrar exemples de visualització. Al sistema de menús de *Deducer*, cliquem a Data / Data Viewer / Open Data Frame i seleccionem l’arxiu “BAROMETRO_CIS_JUNIO_2012[1].rda” que acompanya aquest Manual.

Aquestes pestanyes tenen les següents característiques:

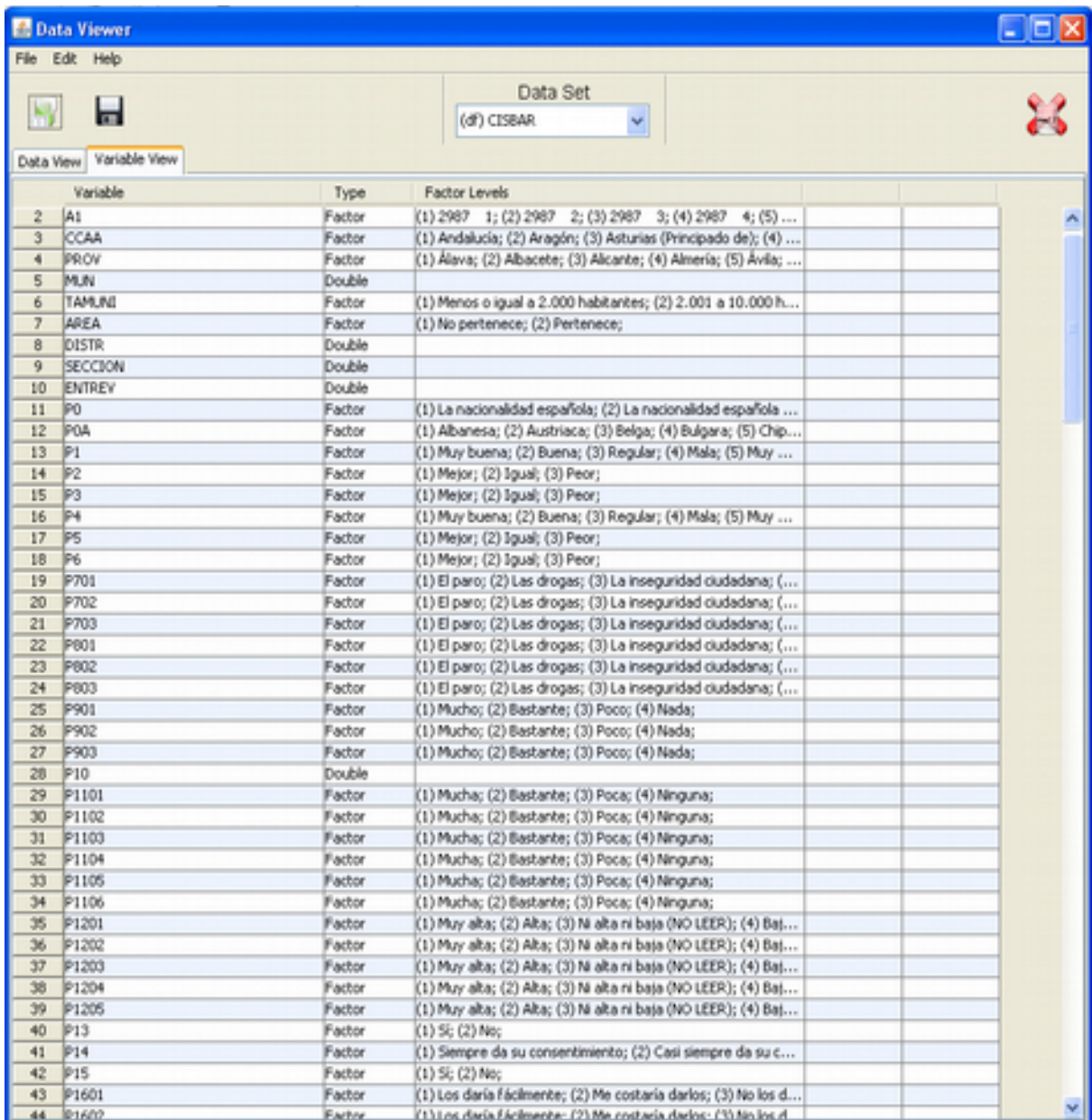
Vista de dades: està dividida en columnes i files donant lloc a cel·les o caselles on es recullen les dades. Cada columna té assignat un nom de variable, ja sigui especificat per

l'usuari o bé pel propi programa. Les files, al seu torn, estan numerades de forma correlativa i representen la informació dels diferents casos estadístics.

	CUES	A1	CCAA	PROV	MUN	TAMUNI	AREA	DISTR	SECCION	ENTREV	PO	PGF
414	867.0/2987	867	Cataluña	Girona	79.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
415	868.0/2987	868	Cataluña	Girona	79.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
416	869.0/2987	869	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
417	870.0/2987	870	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
418	871.0/2987	871	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
419	872.0/2987	872	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
420	873.0/2987	873	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
421	874.0/2987	874	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
422	875.0/2987	875	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
423	876.0/2987	876	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
424	877.0/2987	877	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
425	878.0/2987	878	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
426	879.0/2987	879	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
427	880.0/2987	880	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
428	881.0/2987	881	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
429	882.0/2987	882	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
430	883.0/2987	883	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
431	884.0/2987	884	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
432	885.0/2987	885	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
433	886.0/2987	886	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
434	887.0/2987	887	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
435	888.0/2987	888	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
436	889.0/2987	889	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
437	890.0/2987	890	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
438	891.0/2987	891	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
439	892.0/2987	892	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
440	893.0/2987	893	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
441	894.0/2987	894	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
442	895.0/2987	895	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
443	896.0/2987	896	Cataluña	Girona	0.0/10.001 a 50...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
444	897.0/2987	897	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
445	898.0/2987	898	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
446	899.0/2987	899	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
447	900.0/2987	900	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
448	901.0/2987	901	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
449	902.0/2987	902	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
450	903.0/2987	903	Cataluña	Girona	0.0/2.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
451	1499.0/2987	1499	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
452	1500.0/2987	1500	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
453	1501.0/2987	1501	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
454	1502.0/2987	1502	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
455	1503.0/2987	1503	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
456	1504.0/2987	1504	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
457	1505.0/2987	1505	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
458	1506.0/2987	1506	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
459	1507.0/2987	1507	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...
460	1508.0/2987	1508	Andalucía	Málaga	0.0/50.001 a 10...	No pertenece	0.0	0.0	0.0	0.0	0.0	La nacional...

37. Vista de dades, en Deducer (R-GUI)

Vista de variables: recull les característiques (columnes) de cadascuna de les variables (files). Informa sobre el Nom de la variable, el Tipus (Numèric, Cadena, Data,...), i introdueix “etiquetes” per als valors de les variables categòriques (“Factor Levels”). Aquestes variables defineixen les columnes de la vista de dades i caracteritzen els individus (files).



The screenshot shows the 'Data Viewer' window in R. The 'Data Set' dropdown is set to '(df) CESBAR'. The 'Variable View' tab is selected. The table below lists the variables, their types, and their factor levels.

Variable	Type	Factor Levels
2 A1	Factor	(1) 2987 1; (2) 2987 2; (3) 2987 3; (4) 2987 4; (5) ...
3 CCAA	Factor	(1) Andalucía; (2) Aragón; (3) Asturias (Principado de); (4) ...
4 PROV	Factor	(1) Álava; (2) Albacete; (3) Alicante; (4) Almería; (5) Ávila; ...
5 MUN	Double	
6 TAMUNI	Factor	(1) Menos o igual a 2.000 habitantes; (2) 2.001 a 10.000 h...
7 AREA	Factor	(1) No pertenece; (2) Pertenece;
8 DISTR	Double	
9 SECCION	Double	
10 ENTREV	Double	
11 P0	Factor	(1) La nacionalidad española; (2) La nacionalidad española ...
12 P0A	Factor	(1) Albanesa; (2) Austriaca; (3) Belga; (4) Bulgara; (5) Chip...
13 P1	Factor	(1) Muy buena; (2) Buena; (3) Regular; (4) Mala; (5) Muy ...
14 P2	Factor	(1) Mejor; (2) Igual; (3) Peor;
15 P3	Factor	(1) Mejor; (2) Igual; (3) Peor;
16 P4	Factor	(1) Muy buena; (2) Buena; (3) Regular; (4) Mala; (5) Muy ...
17 P5	Factor	(1) Mejor; (2) Igual; (3) Peor;
18 P6	Factor	(1) Mejor; (2) Igual; (3) Peor;
19 P701	Factor	(1) El paro; (2) Las drogas; (3) La inseguridad ciudadana; (...)
20 P702	Factor	(1) El paro; (2) Las drogas; (3) La inseguridad ciudadana; (...)
21 P703	Factor	(1) El paro; (2) Las drogas; (3) La inseguridad ciudadana; (...)
22 P801	Factor	(1) El paro; (2) Las drogas; (3) La inseguridad ciudadana; (...)
23 P802	Factor	(1) El paro; (2) Las drogas; (3) La inseguridad ciudadana; (...)
24 P803	Factor	(1) El paro; (2) Las drogas; (3) La inseguridad ciudadana; (...)
25 P901	Factor	(1) Mucho; (2) Bastante; (3) Poco; (4) Nada;
26 P902	Factor	(1) Mucho; (2) Bastante; (3) Poco; (4) Nada;
27 P903	Factor	(1) Mucho; (2) Bastante; (3) Poco; (4) Nada;
28 P10	Double	
29 P1101	Factor	(1) Mucha; (2) Bastante; (3) Poca; (4) Ninguna;
30 P1102	Factor	(1) Mucha; (2) Bastante; (3) Poca; (4) Ninguna;
31 P1103	Factor	(1) Mucha; (2) Bastante; (3) Poca; (4) Ninguna;
32 P1104	Factor	(1) Mucha; (2) Bastante; (3) Poca; (4) Ninguna;
33 P1105	Factor	(1) Mucha; (2) Bastante; (3) Poca; (4) Ninguna;
34 P1106	Factor	(1) Mucha; (2) Bastante; (3) Poca; (4) Ninguna;
35 P1201	Factor	(1) Muy alta; (2) Alta; (3) Ni alta ni baja (NO LEER); (4) Baj...
36 P1202	Factor	(1) Muy alta; (2) Alta; (3) Ni alta ni baja (NO LEER); (4) Baj...
37 P1203	Factor	(1) Muy alta; (2) Alta; (3) Ni alta ni baja (NO LEER); (4) Baj...
38 P1204	Factor	(1) Muy alta; (2) Alta; (3) Ni alta ni baja (NO LEER); (4) Baj...
39 P1205	Factor	(1) Muy alta; (2) Alta; (3) Ni alta ni baja (NO LEER); (4) Baj...
40 P13	Factor	(1) Sí; (2) No;
41 P14	Factor	(1) Siempre da su consentimiento; (2) Casi siempre da su c...
42 P15	Factor	(1) Sí; (2) No;
43 P1601	Factor	(1) Los daría fácilmente; (2) Me costaría darlos; (3) No los d...
44 P1602	Factor	(1) Los daría fácilmente; (2) Me costaría darlos; (3) No los d...

38. Vista de variables en Deducer (R-GUI)

El nom de les variables (primera columna) no poden tenir accents, ni "ñ", ni "ç", ni espais en blanc, ni cap caràcter fora de l'estàndard anglès.

Tipus de variables

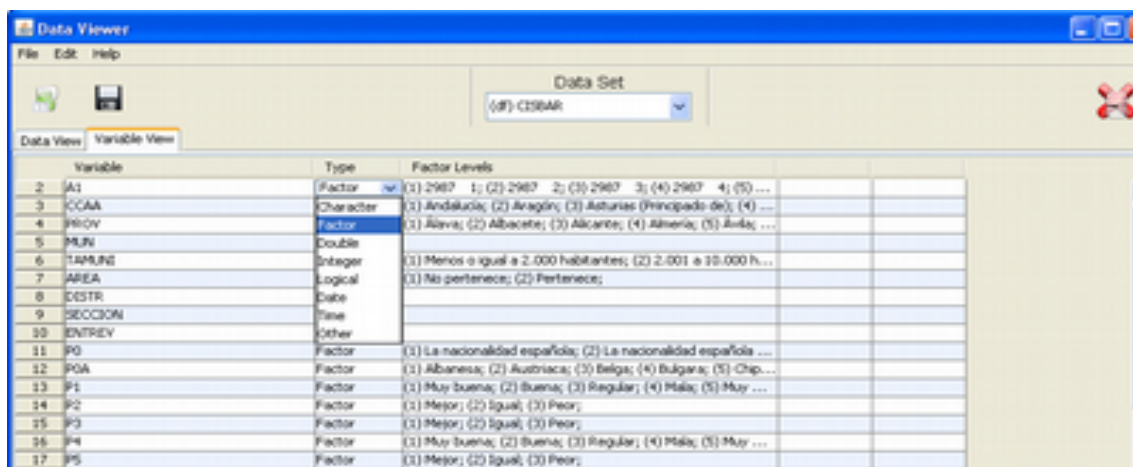
Les variables d'un *dataframe* de R poden ser e diferent tipus:

- **Qualitatius o Categòrics:** etiqueta (numèrica o textual) que representa el grup o categoria a la qual pertany el cas. Es poden diferenciar entre nominals (per exemple el sexe) i ordinals (nivell d'estudis). En R s'anomenen FACTORS, i en el cas de ser de nivell ordinal ORDERED FACTORS.
- **Quantitatius:** valors numèrics pels quals té sentit realitzar aritmètica. Es poden diferenciar entre contínues (índex de massa corporal) i discretes (nombre de

fills). En R s'anomenen DOUBLE si tenen decimals i INTEGER si representen dades discretes.

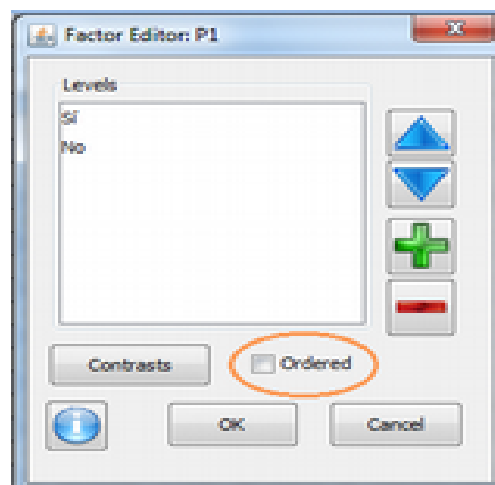
A la pestanya “Vista de variables”, *Deducer* permet la classificació de les variables en:

- Character: variables de cadena (text)
- Factor: variables categòriques, amb etiqueta (nominals o ordinals)
- Double: variables quantitatives contínues
- Integer: variables quantitatives discretes
- Logical: variables lògiques (o dicotòmiques Sí/No)
- Date: variables de data
- Time: variables de temps



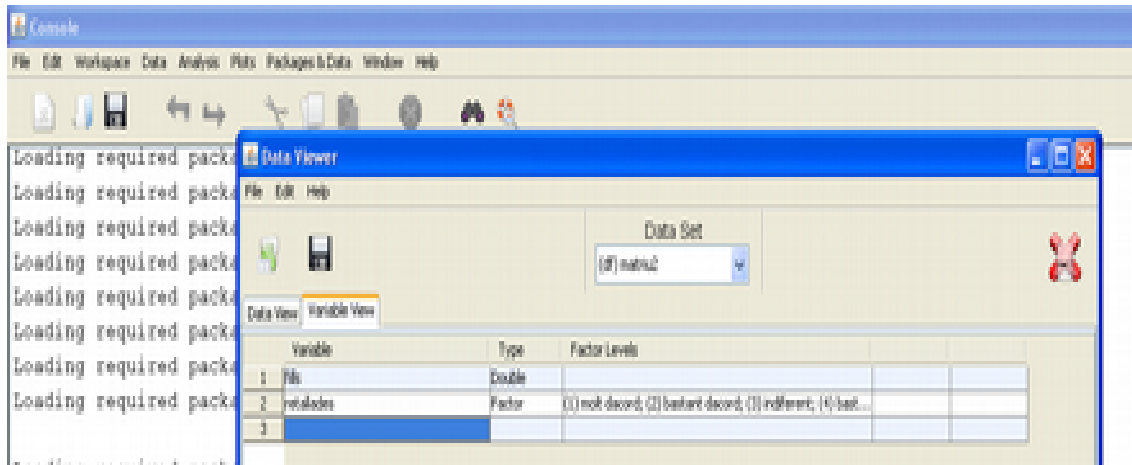
39. Vista de variables [pestanya tipus de variables]

Els nivells, o “*etiquetes dels valors*”, de les variables categòriques (factors) es mostraran en la columna “Factor Levels” i es podran editar a la pròpia cel·la. Quan la categoria de la variable (**levels**) pugui prendre diversos valors ordenables seguint una escala preestablerta (variable ordinal) marcarem la casella “**Ordered**”.



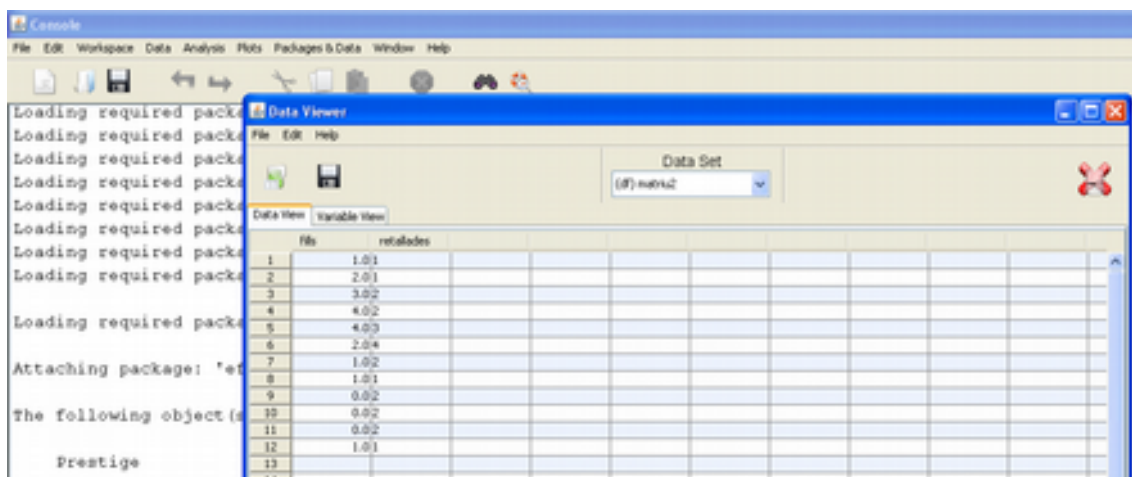
40. Finestra Factor Levels

Per crear una nova variable contínua amb *Deducer*, primer de tot ens haurem d'adreçar a la vista de variables, i seguidament introduïrem el nom de la variable al primer espai disponible de la primera columna. Per exemple, podem introduir la variable número de fills posant: **fills** (Compte! No acceptarà “Nombre de Fills”, perquè té espais en blanc). Una vegada haguem introduït el nom, anirem a la casella del costat “Type” i seleccionarem l'opció “Integer”.



41. Exemple creació d'una variable

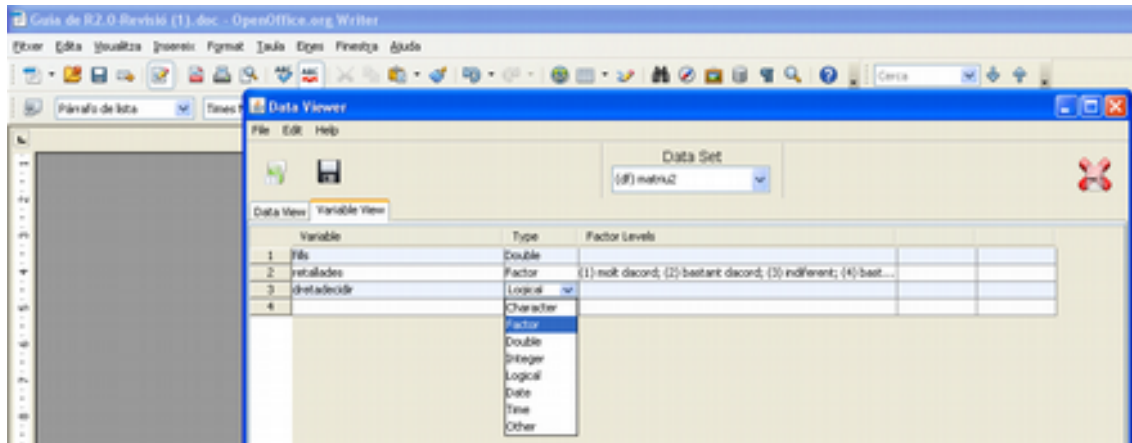
Una vegada haguem fet això, ja ens podem dirigir a la pestanya “Data view” i introduir les respostes. En aquest cas posarem el número de fills per cada cas.



42. Exemple introducció de dades

Al moment d'introduir les dades al “Data View”, *Deducer* interpreta el tipus de variable segons la informació proporcionada, i fins i tot canvia (sense avis) el tipus. Això pot generar problemes: si definim una variable com a INTEGER però introduïm un número amb decimals “2.0” la converteix a DOUBLE, si introduïm un número decimal amb coma “2,3” per comptes de punt “2.3” la converteix a CHARACTER.

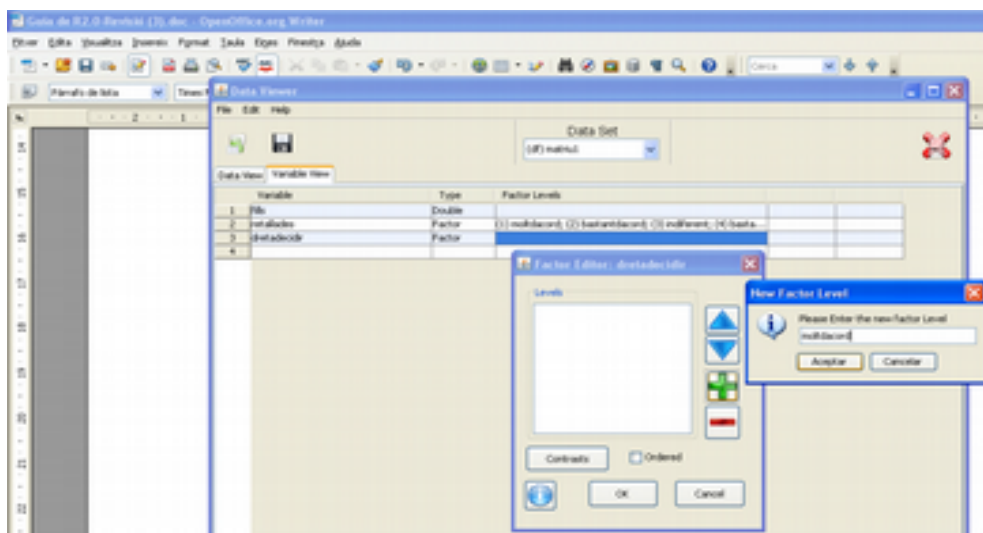
Per a crear una nova variable categòrica amb *Deducer*, en primer clicarem a “**Variable view**”. En segon lloc introduïrem el nom a la columna de “**Variable**” i posteriorment clicarem amb el botó esquerra en la casella de ubicada a la columna “**Type**” i seleccionarem l'opció “**Factor**”.



43. Exemple creació d'una variable

Una vegada haguem realitzat el pas anterior, ens dirigirem farem un clic amb el botó esquerra del ratolí a la casella corresponent de la columna **Factor Levels**.

Una vegada haguem fet això haurem de prémer el símbol + de color verd (**Add**) i introduir l'etiqueta de la resposta (per exemple, molt d'acord, bastant d'acord, etc.).

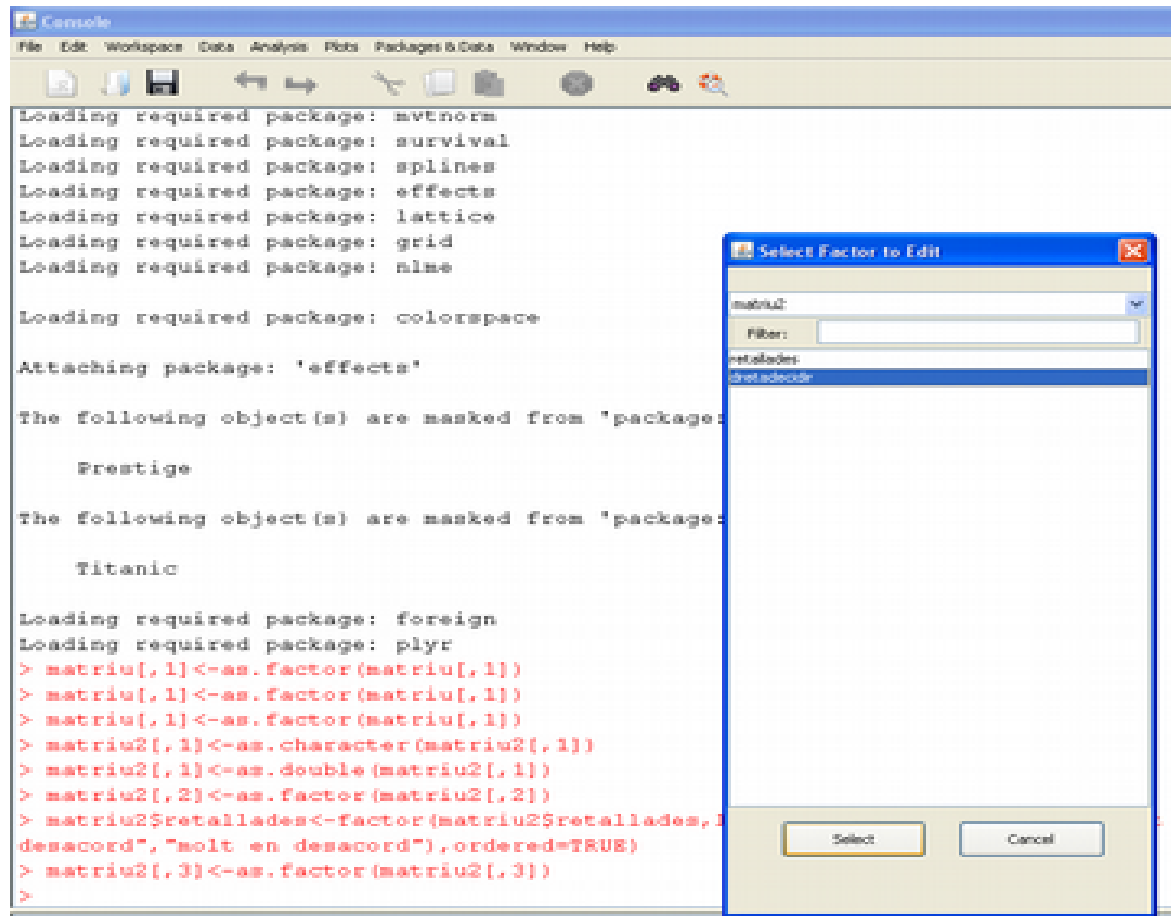


44. Exemple introducció nom etiqueta de resposta.

Quan haguem introduït les etiquetes de les respostes, haurem en les variables categòriques ordinals haurem d'ordenar-les amb les fletxes de color blau i clicar la casella ☐ **Ordered**. Després només caldrà que fem un clic a Ok i ja podrem introduir les respostes a la pestanya **Data View**.

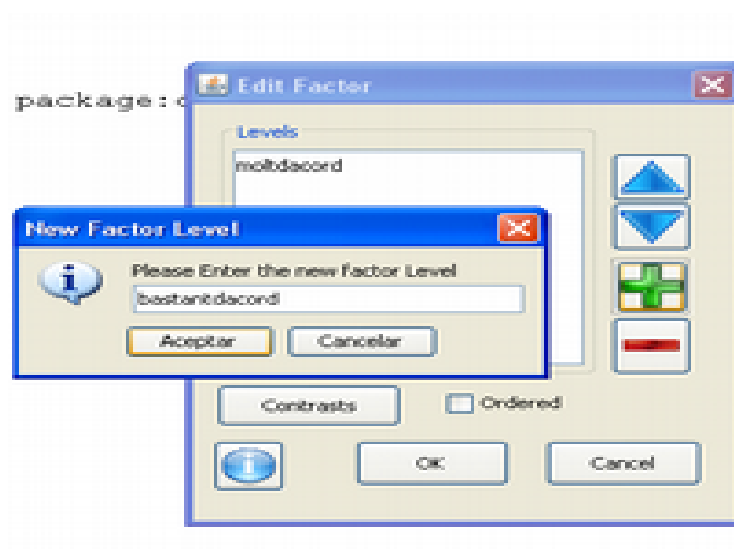
Alternativament també podem dirigir-nos a la consola, i clicar damunt de “**Data**” > “**Edit factor**”.

Una vegada haguem clicat damunt Edit Factor seleccionarem la variable que volem codificar. En aquest cas, agafarem la variable referent al grau d'acord amb el dret a decidir, anomenada **dretadecidir**, i clicarem “**Select**”.



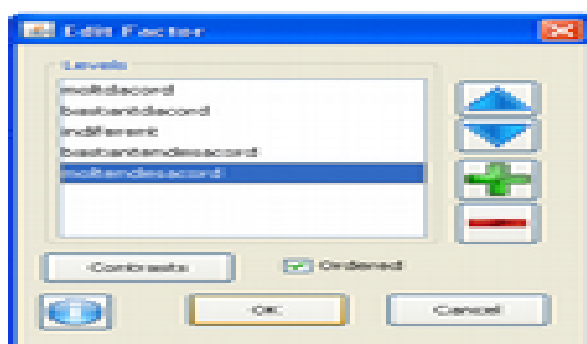
45. Exemple creació de variable amb menús consola, Deducer (R-GUI)

Per introduir les diferents opcions de resposta, clicarem el botó **verd + (Add)** i introduïrem el nom. Per exemple, molt d'acord, bastant d'acord, etc.



46. Exemple creació de variable amb menú de consola Deducer (R-GUI)

Quan haguem fet el pas anterior, clicarem a **Ordered**. I amb les fletxes blaves, ordenarem les respostes i clicarem **OK**.



47. Exemple creació de variable amb menú de consola Deducer (R-GUI)

4. Introducció de dades: Teclat, Arxius de text (CSV), Excel, SPSS, altres.

R es un programa altament compatible amb pràcticament qualsevol format de dades. Amb *Deducer* podem obrir una base de dades utilitzant el menú File / Open Data. Amb aquesta opció podem carregar dades que es trobin en format de R, en format text o altres tipus de format com per exemple Excel o SPSS.

Per que aquest procediment no doni error, la ruta del fitxer NO pot contenir accents, p.e., donaria error si provem d'obrir l'arxiu que es troba a un dispositiu USB de memòria amb una ruta
D:/Recerca/Immigrats/DadesMataró/dades-immi2006.sav).

Teclat

Si tenim un full de dades (“dataframe”) i volem editar-ho a ma, amb el teclat, utilitzarem la funció **edit()** e introduïrem els comandaments que volem realitzar en la pantalla.

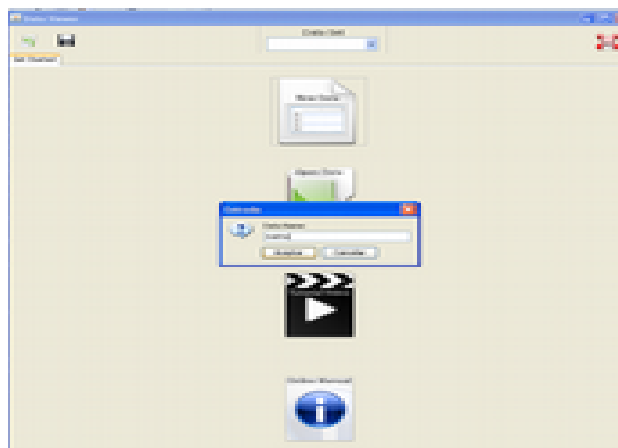
En el cas de no tenir un “dataframe” creat hem d'introduir nosaltres les dades. En primer lloc generem una matriu nova des del visor de dades del *Deducer* (New Data) i introduïrem un nom. A partir d'aquí, en les vistes de dades i variables afegirem les nostres dades.

Obrir i crear fitxers

Per tal de dur a terme l'anàlisi de les dades, el primer pas que s'ha de realitzar consisteix en procedir a l'obertura d'un arxiu de treball, o bé a crear-lo de nou. D'aquesta manera, podem obrir un arxiu de R que ja existeix, importar alguna base de dades que estigui disponible en algun altre programa (SPSS, Excel, etc.) o bé introduir les dades de la matriu manualment.

Com podem crear una nova base de dades?

Quan hom vulgui introduir les dades directament a través de la consola, cal seleccionar l'opció “**New Data**” i introduir el nom que es vol donar a la nova base de dades.



48. Exemple creació d'una nova matriu

Per introduir les dades podem recórrer a diferents procediment:

- Crear noves files i columnes i introduir dades manualment.
- Copiar dades d'altres aplicacions i enganxar-les a la taula.
- Importar dades d'altres aplicacions (*veure més endavant*).

En aquells casos en els quals tinguem diverses bases de dades que estan carregades a la sessió que hem iniciat en R, podem visualitzar-les procedint a la seva selecció a partir de la llista anomenada **“Data Set”**. Per a carregar les dades a la sessió de R podem clicar damunt de l'opció **“Open Data”**. Finalment, també podem procedir a guardar-les a través del botó **“Save Data”** o bé tancar-ho mitjançant l'opció **“Remove from Workspace”**.

Quan creem una matriu nova, primer de tot ens adreçarem al visor de variables i introduïrem la informació corresponent a les variables que hem utilitzat per enquestar així com les característiques d'aquestes variables. Aquestes s'organitzaran en columnes. Posteriorment anirem, anirem al visor de dades i introduïrem la informació corresponent a les respostes dels individus que hem enquestat.

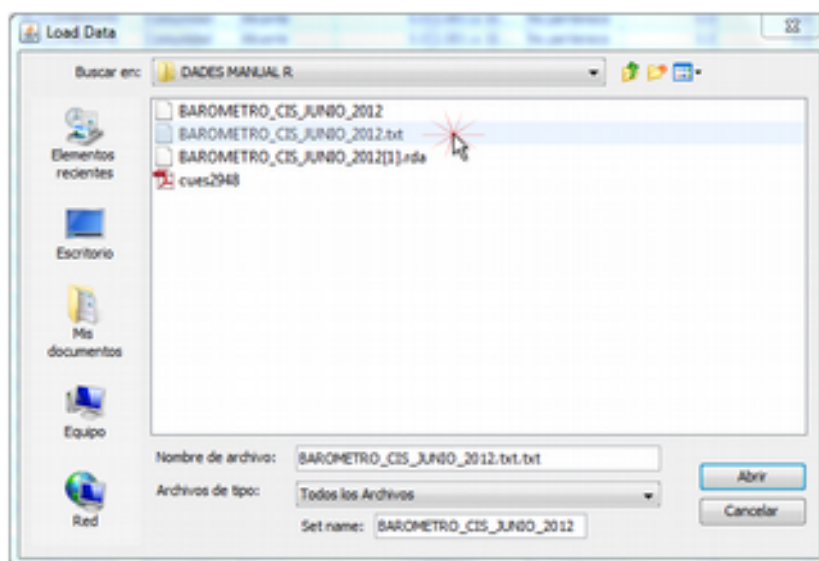
En R, i per tant en *Deducer*, el separador de decimals es el punt, i no la coma. Una dada que contingui una coma no és tractada com a numèrica, sinó com a text.

Importació de dades format Text

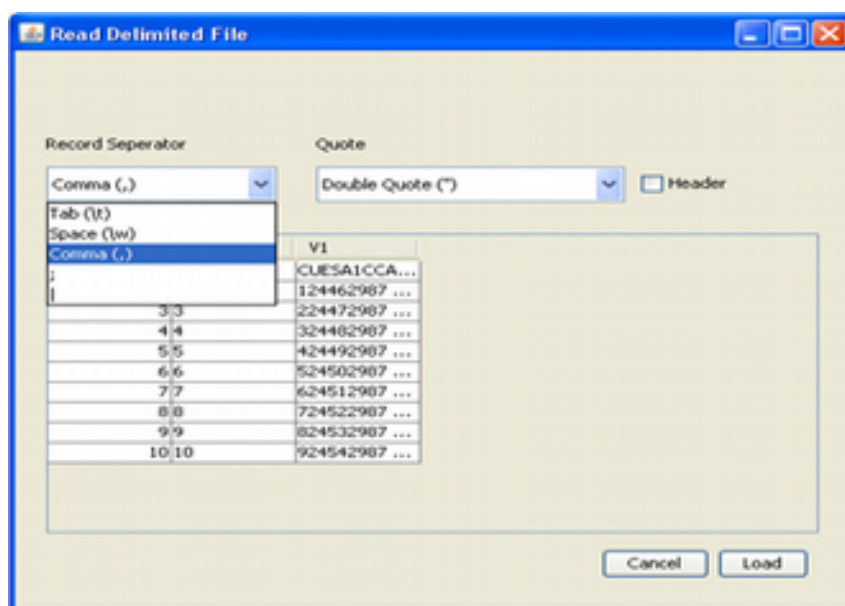
Moltes bases de dades es disposen organitzades en registres separats per espais en blanc o altres indicadors (comes, dos-punts,...), amb o sense una primera línia amb els encapçalaments o identificadors de les variables.

Per a incorporar a un workspace de R dades en aquest format, a *Deducer* cal seleccionar el menú “Packages & Data” i l'opció “Data Viewer”. Fer clic a File / Open Data per cercar la ubicació de l'arxiu de dades. A continuació es mostra un exemple en el qual carreguem la base de dades del Baròmetre del CIS amb format txt.

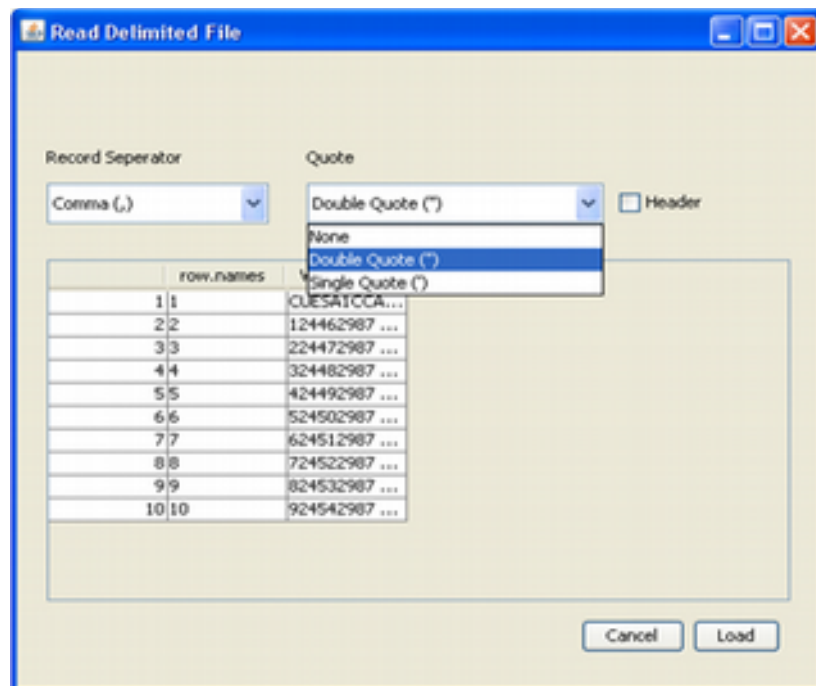
En seleccionar-lo amb l'opció Text file (.txt y .csv) apareix una finestra on podem especificar quin caràcter separa les variables (tabulador, espai, coma...), si hi ha un delimitador específic per a les variables de cadena (“Quote”) i si el fitxer inclou els noms de les variables (“Header”).



49. Exemple importa dades text amb Deducer 1/3



50. Exemple importa dades text amb Deducer 2/3



51. Exemple importa dades text amb Deducer 1/3

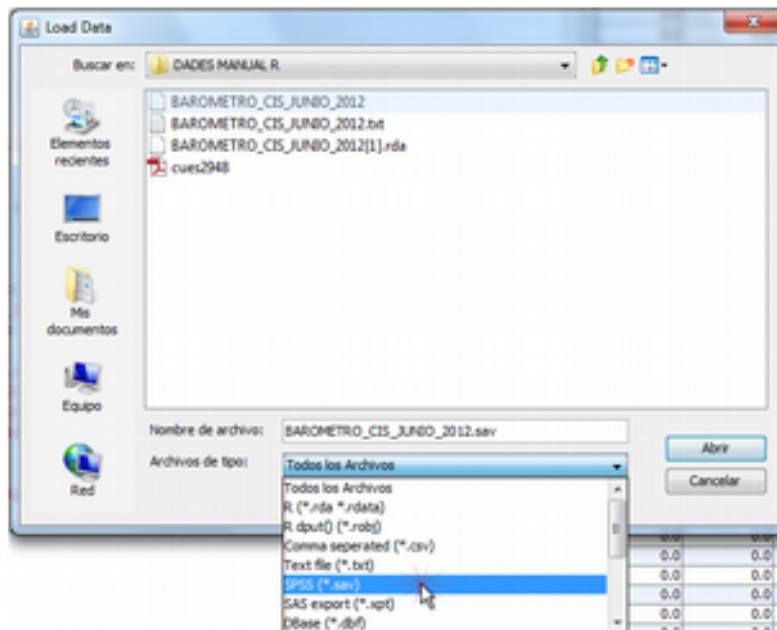
Importació de dades format Excel

Per importar directament arxius d'Excel hem de tenir instal·lat el paquet XLConnect. *Deducer* ho instal·la i carrega automàticament la primera vegada que volem importar un arxiu d'aquest tipus.

Com a solució alternativa hem d'obrir l'arxiu en Excel i després usar "Guardar com" per crear un arxiu de tipus CSV (delimitat per comes), que pot ser fàcilment importable des d'R amb la anterior opció de TEXT.

Importació de dades format SPSS

Amb *Deducer*, en seleccionar un fitxer de SPSS (.sav) la base de dades s'obre automàticament.



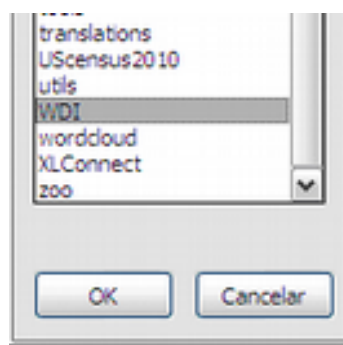
52. Exemple importar a R base de dades des de SPSS

Les variables que tenen etiquetes definides al SPSS es guarden a R com a **factors**, i per tant es perden els valors numèrics del codis.

Descarrega de bases de dades des de paquets R (CRAN)

Hi ha una funció per carregar des del propi programari R dades disponibles a bases de dades online ubicades a servidor organismes internacionals. Ambdues opcions es realitzen a partir d'instruccions de consola, amb la mateixa fórmula.

Primer es carregar el paquet que es vol utilitzar per accedir a les dades *online*. En el nostre cas, i com a exemple, utilitzarem el paquet WDI que conté dades del World Bank's World Development Indicators, desagregades per països.



53. Carrega del paquet

Un cop carregat aquest paquet, que el trobarem al CRAN, mitjançant el procediment estàndard que ja hem explicat prèviament, hem de seleccionar la informació que volem extreure, ja que el programari, no ens permet carregar la matriu sencera amb tots els seus indicadors. Per això, introduïm la següent fórmula:

Exemple programació:

Nom.matriu <- Nom del paquet(Country="països que volem seleccionar", indicators="indicadors que volem seleccionar", Start=any d'inici, End=any de final)

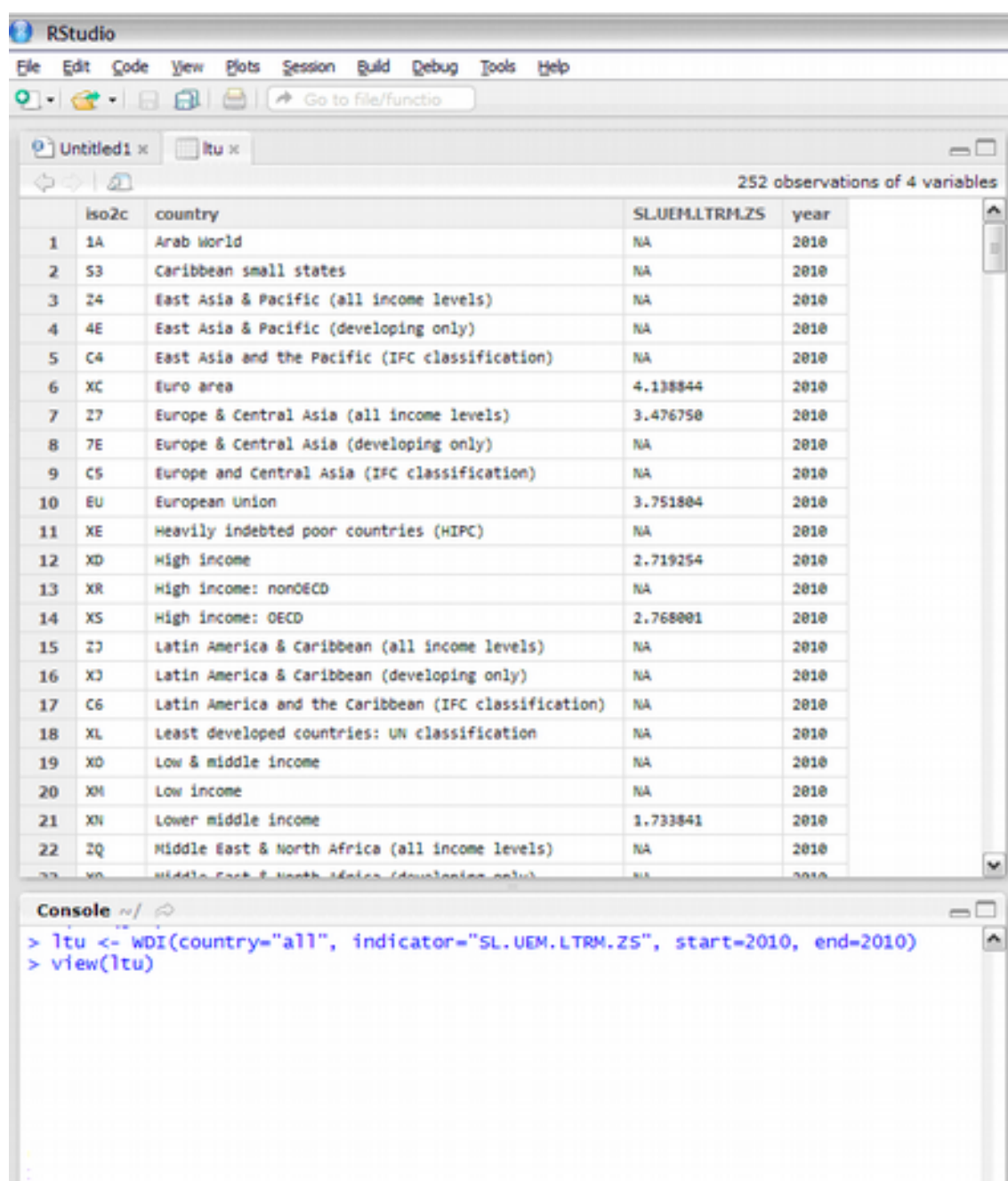
En el nostre cas això quedaria:

```
ltu <- WDI (Country="all", indicators ="SL.UEM.LTRM.ZS", start=2008, end=2010)
```

```
> ltu <- WDI(country="all", indicator="SL.UEM.LTRM.ZS", start=2010, end=2010)
```

Un cop carregada la informació, podrem realitzar les operacions que vulguem amb aquest conjunt de dades, però es recomanable realitzar una primera visualització per assegurar-nos que les hem carregades bé. Us expliquem les opcions per a cada cas:

a) Rstudio: Introduïrem la funció *view* (nom de la nova matriu). Amb això, tindrem l'opció de carregar la matriu en la seva funció clàssica.



	iso2c	country	SLUEM.LTRM.ZS	year
1	1A	Arab World	NA	2010
2	S3	Caribbean small states	NA	2010
3	Z4	East Asia & Pacific (all income levels)	NA	2010
4	4E	East Asia & Pacific (developing only)	NA	2010
5	C4	East Asia and the Pacific (IPC classification)	NA	2010
6	XC	Euro area	4.138844	2010
7	Z7	Europe & Central Asia (all income levels)	3.476750	2010
8	7E	Europe & Central Asia (developing only)	NA	2010
9	C5	Europe and Central Asia (IPC classification)	NA	2010
10	EU	European Union	3.751804	2010
11	XE	Heavily indebted poor countries (HIPC)	NA	2010
12	XD	High income	2.719254	2010
13	XR	High income: nonOECD	NA	2010
14	XS	High income: OECD	2.768001	2010
15	Z3	Latin America & Caribbean (all income levels)	NA	2010
16	X3	Latin America & Caribbean (developing only)	NA	2010
17	C6	Latin America and the Caribbean (IPC classification)	NA	2010
18	XL	Least developed countries: UN classification	NA	2010
19	XD	Low & middle income	NA	2010
20	XN	Low income	NA	2010
21	XN	Lower middle income	1.733841	2010
22	ZQ	Middle East & North Africa (all income levels)	NA	2010
23	XN	Middle East & North Africa (developing only)	NA	2010

```

> ltu <- WDI(country="all", indicator="SL.UEM.LTRM.ZS", start=2010, end=2010)
> view(ltu)

```

54. Matriu carregada en RStudio i comandament de carrega

b) Deducer: Donarem a l'opció data viewer per tal de que se'ns carregui.

```

> ltu <- WDI(country="all", indicator="SL.UEM.LTRM.ZS", start=2010, end=2010)
> deducer("Data viewer")
>

```

55. Comandaments per carregar matriu en Deducer

Data Viewer

File Edit Help

Data Set
(df) Its

Data View Variable View

	iso2c	country	SLUEML...	year				
1	1A	Arab World	...	2010				
2	S3	Caribbean small states	...	2010				
3	Z4	East Asia & Pacific (all income levels)...	...	2010				
4	4E	East Asia & Pacific (developing only) ...	...	2010				
5	C4	East Asia and the Pacific (IFC classifi...	...	2010				
6	Xc	Euro area	4.138844	2010				
7	Z7	Europe & Central Asia (all income lev...	3.476750	2010				
8	7E	Europe & Central Asia (developing o...	...	2010				
9	C5	Europe and Central Asia (IFC classifi...	...	2010				
10	EU	European Union	3.751804	2010				
11	XE	Heavily indebted poor countries (HDP...	...	2010				
12	XD	High income	2.719254	2010				
13	XR	High income: nonOECD	...	2010				
14	XS	High income: OECD	2.768001	2010				
15	ZJ	Latin America & Caribbean (all incom...	...	2010				
16	XJ	Latin America & Caribbean (developi...	...	2010				
17	C6	Latin America and the Caribbean (IF...	...	2010				
18	XL	Least developed countries: UN classi...	...	2010				
19	XO	Low & middle income	...	2010				
20	XM	Low income	...	2010				
21	XN	Lower middle income	1.733841	2010				
22	ZQ	Middle East & North Africa (all incom...	...	2010				
23	XQ	Middle East & North Africa (developi...	...	2010				
24	C7	Middle East and North Africa (IFC da...	...	2010				
25	XP	Middle income	...	2010				
26	XU	North America	2.595562	2010				
27	XY	Not classified	...	2010				
28	OE	OECD members	2.590387	2010				
29	S4	Other small states	...	2010				
30	S2	Pacific island small states	...	2010				
31	S1	Small states	...	2010				
32	85	South Asia	1.383646	2010				
33	C8	South Asia (IFC classification) ...	...	2010				
34	ZG	Sub-Saharan Africa (all income levels...	...	2010				
35	ZF	Sub-Saharan Africa (developing only...	...	2010				
36	C9	Sub-Saharan Africa (IFC classificatio...	...	2010				
37	XT	Upper middle income	...	2010				

56. Matriu carregada en Deducer

5. Manipulació bàsica de dades

“Missing values” a R

En R tots els valors codificats com “no sap”, “no contesta”, s’han de codificar com a valors perduts (NA). Saber identificar i tractar aquests casos es fonamental per a poder realitzar de forma correcta les anàlisis i la seva interpretació. En general la no disponibilitat de dades es representa a través dels valors NA (=Not Available).

Al “Data View” de *Deducer*, qualsevol cel·la que en quedi en blanc es considera per defecte com un valor no disponible NA. El sistema R no permet, a diferència d’altres com SPSS, definir com a *perduts* determinat valors que realment han estat introduïts a la base de dades.

En R, una forma genèrica de tractar de forma diferencial valors NS o NC com a perduts es (1) codificar-los dins de la base de dades, (2) crear una còpia de la variable original en la qual els valors corresponents estiguin en blanc (NA) i (3) fer les anàlisis corresponents seleccionant la versió de la variable que interressi en cada cas, amb o sense NA.

Manipulació de VARIABLES

En la manipulació de les bases de dades a vegades es necessari transformar la informació que tenim disponible per aconseguir noves variables a partir de les disponibles. El llenguatge de programació R proporciona potents eines per fer qualsevol tipus de transformacions, tot i que els GUIs actuals (*Deducer*, *RStudio*, *R-Commander*, etc...) cobreixen directament una petita part d’aquestes possibilitats.

Aquesta secció repassa les transformacions més habituals en el tractament quantitatiu de dades socials per a recerca, tot indicant el procediment per realitzar-les amb *Deducer* quan sigui possible, o amb programació directa per consola.

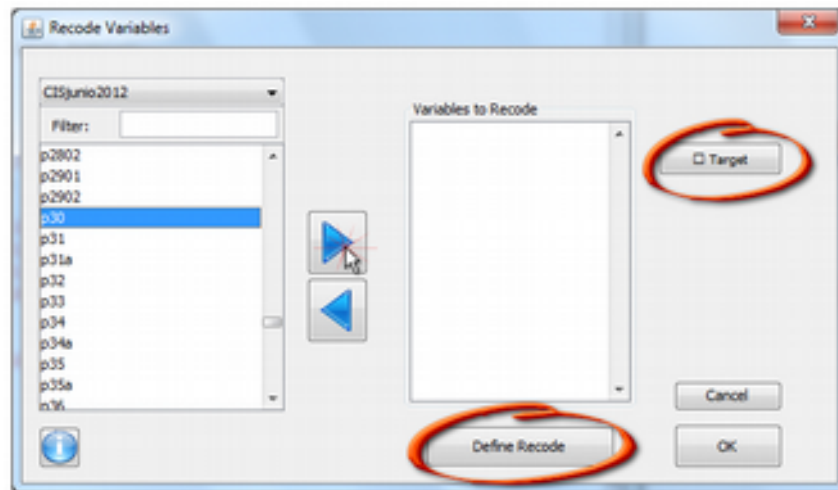
Per a tots els procediments següents s'utilitzarà l'arxiu en format R “BAROMETRO_CIS_JUNIO_2012[1].rda” adjunt a aquest Manual.

Recodificar una variable

Recodificar una variable consisteix en assignar nous codis o valors a partir dels seus valors originals, o agrupar rangs de valors existents en nous valors, de manera que es modifiqui la seva codificació original.

Amb *Deducer* les variables es recodifiquen des del menú Data / Recode Variables. Es poden recodificar en les mateixes variables (**no recomanable**) o en variables de nova creació (anomenades “target”).

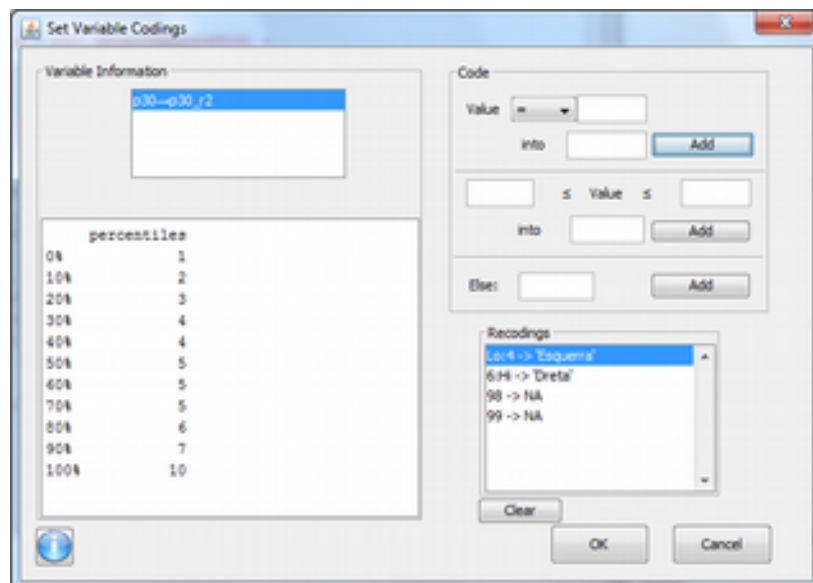
Per a recodificar la variable P30 del Baròmetre CIS Junio 2012 (una escala de 10 punts d'autoposicionament polític, amb extrems: “Esquerra”=1, “Dreta”=10) de forma que es generi la nova variable dicotòmica P30_r2: “Esquerra”=1,2,3,4,5 i “Dreta”=6,7,8,9,10 cal utilitzar el menú de *Deducer* Data / Recode Variables.



57. Finestra recodificar variables

En la pestanya “*Define Recode*” podem definir com volem fer la recodificació:

1. El panell de l'esquerra mostra informació sobre les variables que poden ser útils per la recodificació. Per a les variables numèriques es mostra una taula de percentils i per a les variables categòriques una taula de freqüències.
2. En el panell de la dreta (“Code”) s'especifica la recodificació, els seus valors i les seves etiquetes.
3. Es pot recodificar valors com a *no available* (“valors perduts”, o *missing values* per falta de dades, o per resposta impossible) indicant NA com a valor.
4. Quan procedim a fer una recodificació, en el cas dels valors antics que no es codifiquen queden igual que en la matriu antiga i, per tant, aquests valors no s'inclouran dins de les noves categories. Els valors antics que no es codifiquin com a valors perduts, el programa R no és capaç d'interpretar que són valors perduts i per tant els comptabilitza a l'hora de realitzar l'anàlisi. Com es veu a la imatge els valors recodificats iguals o menors de quatre queden codificats com a esquerra, mentre que els valors recodificats iguals o majors de 6 queden codificats com a dreta (es refereix a l'eix ideològic esquerra - dreta). No obstant, el valor antic 5, al no haver estat recodificat es manté igual que en la matriu antiga.



58. Exemple de recodificació

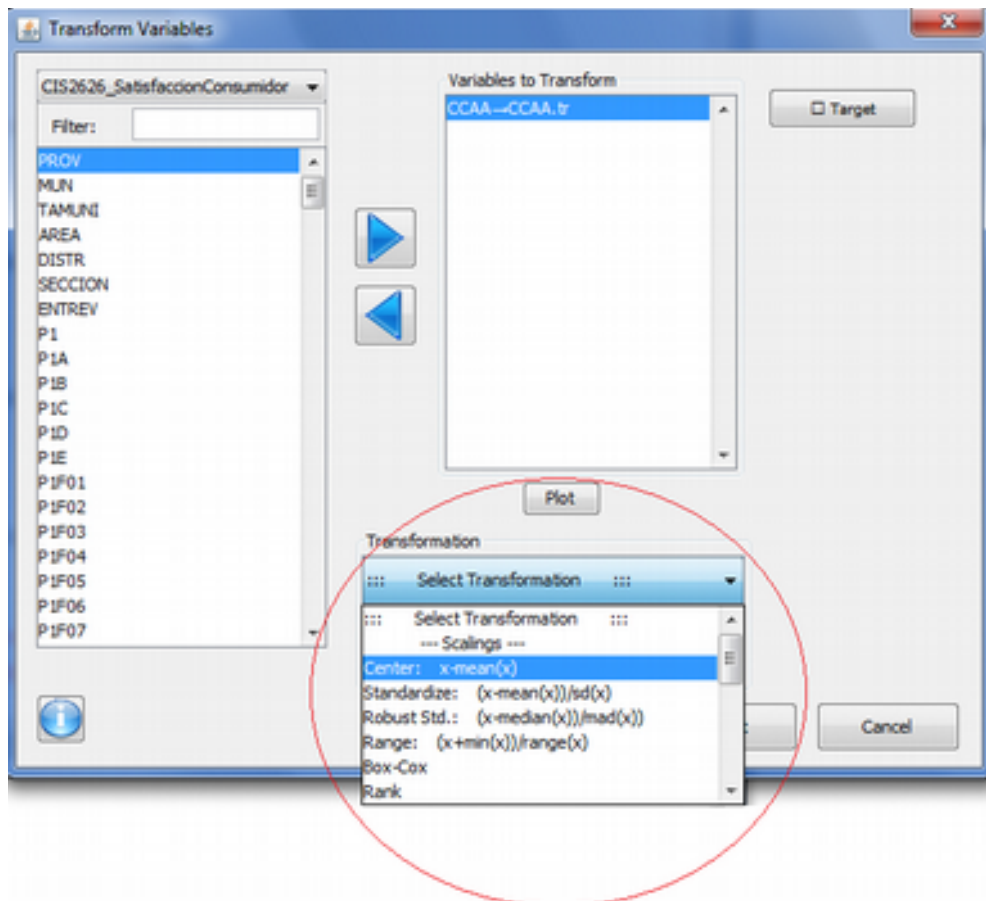
Com es pot veure a la finestra de consola de *Deducer*, la instrucció corresponent per a aquesta codificació és:

```
CISjunio2012[c("p30_r3")] <-
recode.variables(CISjunio2012[c("p30")] , "Lo:4 ->
'Esquerra';6:Hi -> 'Dreta';98 -> NA;99 -> NA;")
```

Transformar variables

Deducer ofereix una opció per a transformar variables mitjançant el menú Data / Transform que proporciona una gran varietat d'opcions per transformar i re-escalar variables. Aquestes opcions són:

- o **Center**: Reescala les variables perquè tinguin mitjana 0.
- o **Standardize**: Reescala les variables perquè tinguin mitjana 0 i desviació estàndard 1.
- o **Robust Standardize**: Reescala les variables perquè tinguin mitjana 0 i desviació absoluta mitjana 1.
- o **Range**: Transforma la variable perquè prengui valors entre 0 i 1.
- o **Box-Cox**: Transforma la variable per intentar obtenir una distribució normal.
- o **Rank**: Reemplaça els valors pel seu rang.
- o **Log**: Retorna el logaritme neperià (per a valors majors que 0).
- o **Square root**: Retorna l'arrel quadrada.
- o **Absolute value**: Retorna el valor absolut.
- o **Quantiles**: Divideix la variable en grups amb el mateix nombre d'observacions.
- o **Equal width**: Divideix la variable en grups amb intervals de la mateixa amplitud.
- o **Custom**: Permet definir transformacions personalitzades.



59. Exemple transformació de variables

Recordeu sempre utilitzar el botó [Target] per a canviar el nom de la variable transformada. Altrament, hi ha perill de sobre-escriure les dades originals, i per tant perdre-les si no hi ha còpies prèvies de la base de dades.

Com a alternativa a les transformacions incorporades al menú de *Deducer*, el llenguatge de programació R permet generar noves variables a partir de les existents amb la màxima flexibilitat. Aquest procediment, però, implica el coneixement del llenguatge de programació.

Com a exemple, calculem l'any de naixement a partir de la edat. Podem visualitzar el contingut de la variable p33 (edat) amb la instrucció `CISjunio2012$p33`. Podem crear una nova variable a la base de dades anomenada `CISjunio2012$p33_naixement`, descomptant de 2012 la edat de cada persona. La instrucció indica a R que generi un nou objecte (`p33_naixement`) a la base de dades corresponent (`CISjunio2012`) i que el contingut d'aquest nou objecte sigui el resultat del càlcul `2012 - CISjunio2012$p33`:

```
CISjunio2012$p33_naixement <- 2012 - CISjunio2012$p33
```

Es pot crear una nova variable a partir de les existents P1201n fins a P1207n (versió numèrica d'escala de valoració de 1 a 4 de la pregunta 12 del qüestionari) referides a la "dependència tecnològica" de la persona entrevistada. Aquesta nova variable (p12n) pot ser la mitjana dels valors de les 7 preguntes per a cada persona. Però prèviament a la suma s'ha de transforma cada variable per tal que el valor 1 indiqui màxima necessitat (=5) i el valor 4 mínima necessitat (=0). La instrucció en R seria:

```
CISjunio2012[c("p12n")] <- ((4 - CISjunio2012$p1201n) +
(4 - CISjunio2012$p1202n) + (4 - CISjunio2012$p1203n) +
(4 - CISjunio2012$p1204n) + (4 - CISjunio2012$p1205n) +
(4 - CISjunio2012$p1206n) + (4 -
CISjunio2012$p1207n) ) / 7
```

Conversió de tipus

En determinades situacions, necessitem que R entengui determinats factors com a números o a números com a caràcters. Aquesta funció s'anomena conversió de tipus. En el format R, quan afegim a un vector numèric algun element no numèric aquest converteix tot el vector de manera automàtica en no numèric.

Lógico	Convierte
is.numeric()	as.numeric()
is.character()	as.character()
is.vector()	as.vector()
is.matrix()	as.matrix()
is.data.frame()	as.data.frame()

Creació de tipologies

Una tipologia es una variable "construïda" a partir del valors combinats d'altres variables. Com a exemple trivial, amb les variables p32 ("Hombres", "Mujeres") i p35 ("Católico/a",...) i es pot construir la tipologia Catòlic / Catòlica / No catòlic / No catòlica mitjançant instruccions **ifelse**.

PENDENT

Manipulació de CONJUNTS DE DADES

Una situació comuna és la de trobar-nos les dades recollides en matrius o bases de dades diferents i que nosaltres vulguem unificar tota aquesta informació en una sola. Aquí es poden donar diverses opcions:

1. Els individus/casos (files) estan en bases diferents

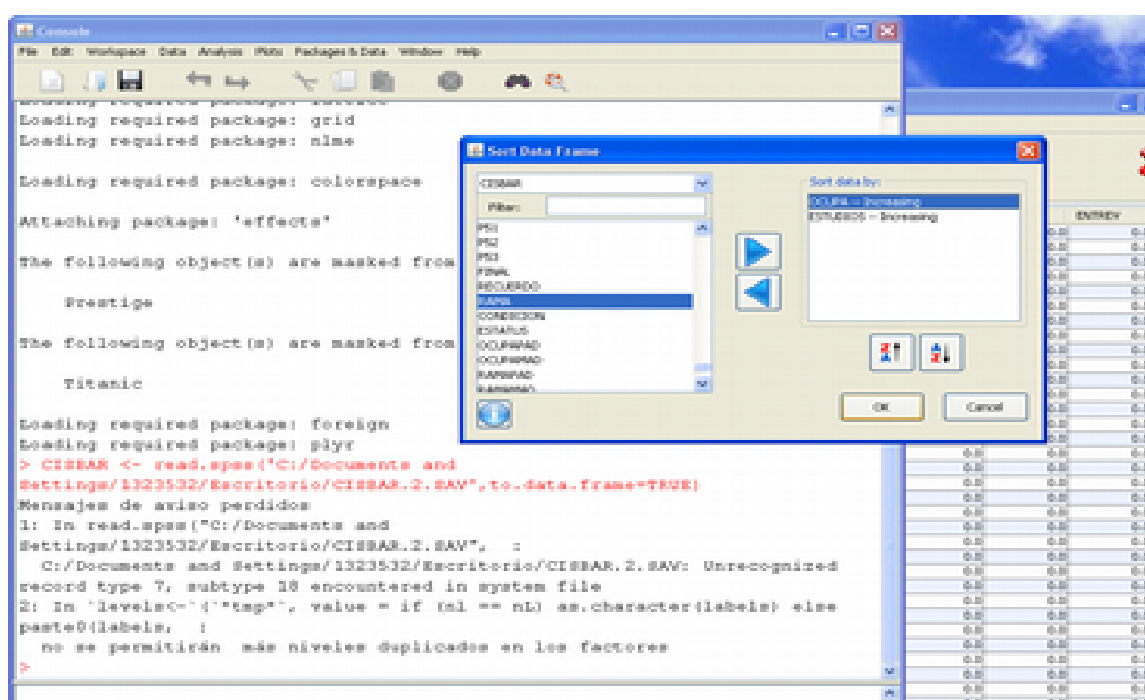
2. Les variables (columnes) estan en bases de dades diferents.

En tots dos casos el que es pretén fer és fusionar els arxius. En el primer cas s'afegiran noves files d'individus (afegir casos) i per a això és necessari que el nou individu tingui les mateixes característiques (variables) que la resta d'individus. En cas contrari s'imputarà un valor perdut en aquelles variables en les quals difereixi. En el segon cas (afegir variables) es crearan noves columnes de dades. Si les noves columnes són de diferent longitud a les ja existents, s'emplenarà els espais en blanc amb valors perduts (*NA: not available*) fins que la matriu coincideixi.

En altres situacions el que interessa es canviar l'ordre dels individus dins de l'arxius, segons diversos criteris, o be treballar amb subconjunt de casos.

Ordenar dades

Amb *Deducer*, seleccionar el menú Data / Sort a la finestra de Consola. Apareixerà la següent finestra:



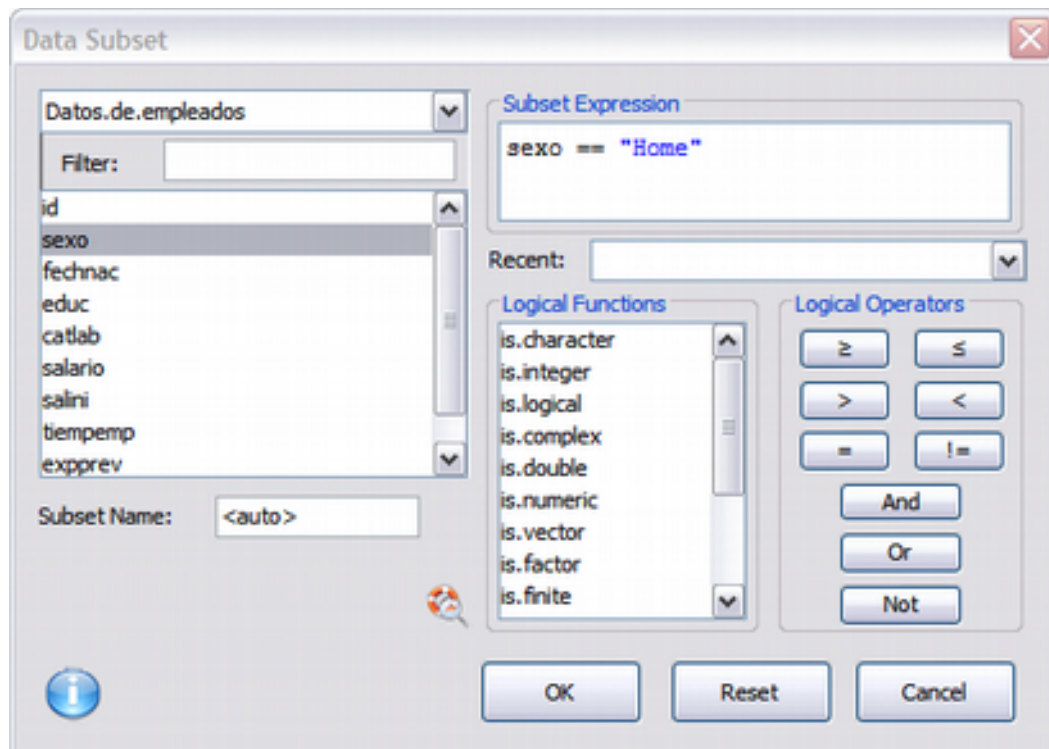
60. Exemple ordenar dades arxius CIS

Selecciona el conjunt de dades que ha de ser ordenat (del menú desplegable) i la variable o variables que s'utilitzaran com a criteris d'ordenació. Triar el tipus d'ordenació per a cada variable i executar l'ordre amb el botó [OK].

Per tal de dur a terme l'ordenació de les files d'un *dataframe* s'acostuma a utilitzar la funció anomenada "order ()". Aquesta per defecte tendeix a ordenar les dades de manera ascendent. No obstant, també podem canviar el sentit de l'ordenació utilitzant el signe "menys".

Seleccionar subconjunts de casos

Amb *Deducer*, seleccionar el menú Data / Subset a la finestra de Consola. Apareixerà la següent finestra:



61. Exemple de seleccionar subconjunt de casos

Enter the expression for the desired subset in the **Subset Expression** space. The subset desired for this example is those intersections where the bike count was greater than or equal to the pedestrian count. Note that to create a subset on non-numeric variables (e.g., *type*) the values of the variable must be enclosed with quotation marks (e.g., *type* == "none"). Include a name for the subset in the **Subset Name** space. Deducer will automatically provide a name if one is not supplied. Click **OK**. The subset will appear as a new entry in the pull down menu of the **Data Viewer**. (The original data remains unchanged.)

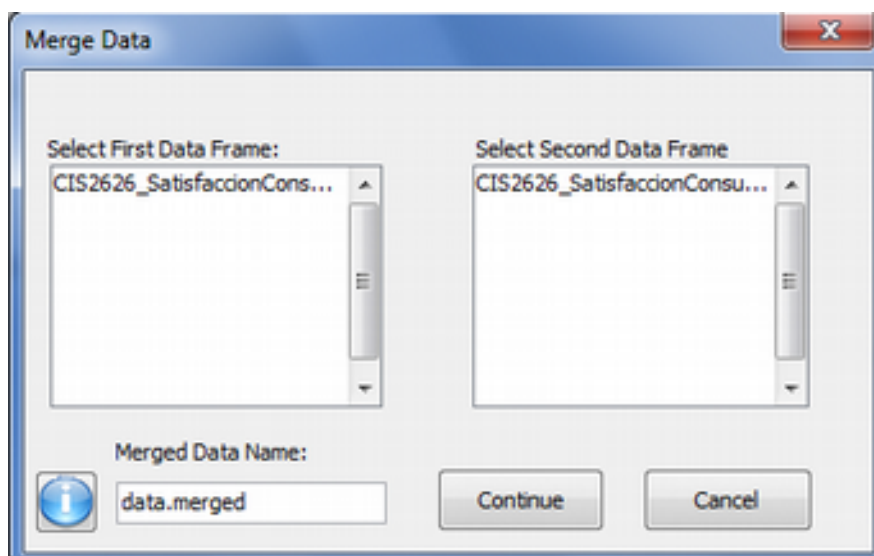
Introduir la expressió que defineix el subconjunt a l'espai "Subset Expression". Compte amb les variables qualitatives, o factors: els valors de la variable s'han d'escriure entre cometes dobles (e.g., "Home"). Es pot indicar igualment un nom pel subconjunt de dades, o *Deducer* assignarà un nom automàticament.

Tot subconjunt apareix al menú desplegable del "Visor de Dades" i altres finestres d'operació. Les dades originals es mantenen sense canvi.

Ampliar/Unir conjunt de dades

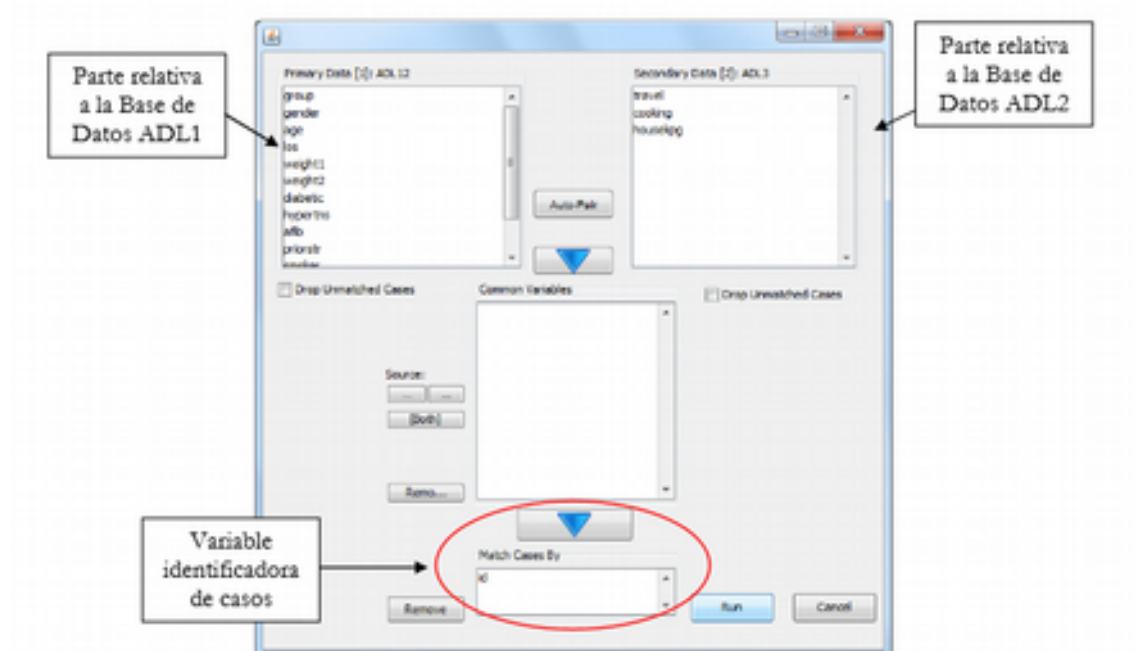
Més enllà d'afegir casos i variables existeix l'opció de fusionar les dues matrius,:

En tots dos casos el que es pretén fer és fusionar els arxius i el seu procediment es realitza a partir del Data → Merge Data.



62. Exemple ampliar/unir conjunt de dades

Per combinar dades de dues matrius utilitzarem l'opció “*Auto-Pair*” que ens permet aparellar variables diferents. Per afegir casos hem de seleccionar totes les variables comunes i afegir l'identificador en el requadre “*Match Cases By*”. Amb les variables la mecànica es exactament la mateixa.



63. Exemple extret de Manual d'introducció a Deducer. Servei d'estadística de la UAB. 2013.

6. Emmagatzematge

Els objectes creats durant una sessió de R poden emmagatzemar-se en un arxiu per al seu ús posterior. Al finalitzar la sessió, R pregunta si desitja fer-ho. En cas afirmatiu tots els objectes s'emmagatzemen en l'arxiu .RData2 en el directori de treball. En la següent ocasió que executi R, es recuperaran els objectes d'aquest arxiu així com l'historial d'ordres.

És recomanable que utilitzis un directori de treball diferent per a cada problema que s'analitzi amb R. És molt comú crear objectes amb noms *standard* o amb noms ja establerts en relació a la matriu amb la que es treballa. Aquests noms tenen sentit dins d'un anàlisis concret, però és molt difícil dilucidar el seu significat quan s'han realitzat diversos anàlisis en el mateix directori. De la mateixa manera, treballar en el mateix directori ens pot fer més difícil la feina si acumulem infinitat d'objectes de diferents anàlisis.

Les bases de dades poden ser guardades en els següents formats:

- R workspace (extensió .rda i .rdata)
- R object (extensió .robj)
- Comma separated (extensió .csv)
- Tab delimited (extensió .txt)
- DBase (extensió .dbf)
- Stata (extensió .dta)
- ARFF (extensió .arff)

També podem guardar totes les bases de dades obertes en un sol arxiu utilitzant el menú Workspace → Save as...