

El paquete MicroDatosES

Carlos J. Gil Bellosta
cgb@datanalytics.com

Datanalytics

Noviembre de 2012

¿Qué son los microdatos?

Microdatos

En el ámbito de las encuestas y los censos, los microdatos son los ficheros de información al nivel mínimo de agregación (individuo).

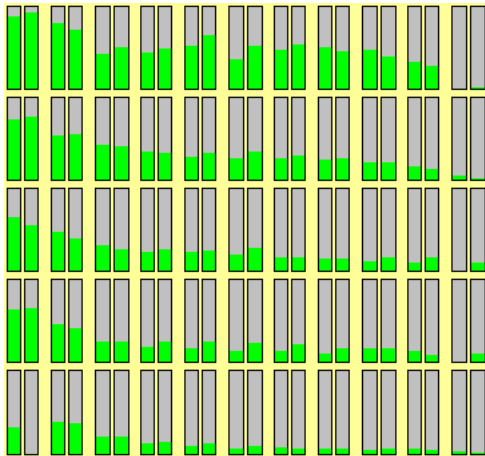
- El INE proporciona microdatos sobre desempleo (EPA), nacimientos, defunciones, etc.
- El CIS, de encuestas de opinión
- El Ayuntamiento de Madrid, de calidad del aire
- Etc.

¿A quién interesan microdatos? ¿Por qué?

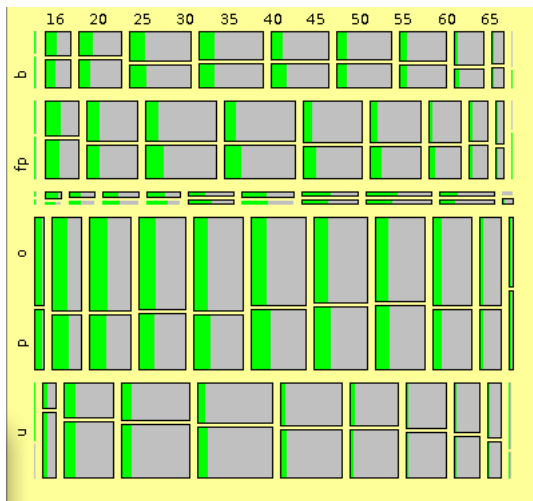
- Institutos estadísticos (INE, etc.)
- Científicos sociales (sociólogos, políticos, médicos, etc.)
- Periodistas de datos
- ¿Público en general?

Los microdatos permiten responder a preguntas distintas (¿y más interesantes?) que las que responden quienes crean la información agregada.

Lo que no te cuenta la prensa (I)



Lo que no te cuenta la prensa (II)



¿Cómo son los metadatos? (EPA, INE, columnas)

ENCUESTA DE POBLACION ACTIVA 2005-2012				
REGISTRO:				NIVEL = 1 LONGITUD = 176
DESCRIPCIÓN: DISEÑO DE REGISTRO ANONIMIZADO GRATUITO INEBASE				
CAMPO	COMIENZO	FIN	LONGITUD	DESCRIPCIÓN DEL CAMPO
CONTROL	1		0	AGRUPACIÓN DE CAMPOS
CICLO	1	3	3	Período de referencia
IDEN	4		0	AGRUPACIÓN DE CAMPOS
CCAA	4	5	2	<u>Comunidad autónoma</u>
PROV	6	7	2	<u>Provincia</u>
NVIVI	8	12	5	<u>Número de orden de la vivienda</u>
NIVEL	13	13	1	Variable que indica el nivel del registro en el fichero
NPERS	14	15	2	<u>Número de la persona</u>
A	16		0	AGRUPACIÓN DE CAMPOS
EDAD5	16	17	2	<u>Edad, grupos quinquenales de años cumplidos</u>
RELPP1	18	18	1	<u>Relación con la persona de referencia</u>
SEXO1	19	19	1	<u>Sexo</u>
NCONY	20	21	2	<u>Número de persona del cónyuge</u>
NPADRE	22	23	2	<u>Número de persona del padre</u>
NMADRE	24	25	2	<u>Número de persona de la madre</u>
RELLMILI	26	26	1	<u>Relleno antigua variable MILI</u>
ECIV1	27	27	1	<u>Estado civil legal</u>

¿Cómo son los metadatos? (EPA, INE, códigos)

Tipo de contrato temporal		
Filtro:	Asalariados con contrato temporal	
Variable:	DUCON3	
Valores:	01	Por circunstancias de la producción
	02	De aprendizaje, formación o práctica
	03	Estacional o de temporada
	04	Cubre un periodo de prueba
	05	Cubre la ausencia total o parcial de otro trabajador
	06	Para obra o servicio determinado
	07	Verbal no incluido en las opciones anteriores
	08	Otro tipo
	00	No sabe

¿Por qué el paquete MicroDatosEs?

El paquete MicroDatosES automatiza la carga de (ciertos) ficheros de microdatos. Sus objetivos son:

- Reducción de los costes sociales
- Facilidad de acceso a información relevante
- Fomento del uso de R

Ejemplo de uso: carga de los microdatos de la EPA

```
library(MicroDatosEs)

raw.epa <- epa2005("EPAWEBT0312")
class(raw.epa) # data.set (del paquete "survey")

dat.epa <- as.data.frame(raw.epa)
```

Ejemplo de uso: cálculo de la tasa de paro

```
# se eliminan los inactivos
tasa.paro <- dat.epa[ grep("Inactivos",
  dat.epa$aoi, invert=T), ]
tasa.paro$aoi <- factor(tasa.paro$aoi)
tmp <- sum(tasa.paro$factorel) / 100
xtabs( factorel ~ aoi, data = tasa.paro ) / tmp
#Ocupados subempleados por insuficiencia de horas
#                               10.199275
#                               Resto de ocupados
#                               64.785718
#                               Parados que buscan primer empleo
#                               2.386597
#                               Parados que han trabajado antes
#                               22.628410
```

Ámbito del paquete

Idealmente, el paquete debería comprender:

- Microdatos facilitados por administraciones públicas españolas
- Que tengan continuidad en el tiempo (con el mismo formato)

Funciones públicas (a día de hoy)

Las funciones

- `epa2005(epa.file)`, para ficheros de la EPA (post 2005)
- `defun2011(file)`, para el fichero de defunciones (post 2011)

cargan ficheros utilizando ciertos ficheros de metadatos incluidos en el paquete.

La función `test.metadata(file, md.1, md.2, md.3,...)` permite cargar ficheros usando ficheros de metadatos `md.` [1-3] (documentados) construidos por el usuario.

Funciones públicas (en el futuro)

La idea es que:

- usuarios de R interesados en determinados ficheros de microdatos creen sus ficheros de metadatos `md.` [1–3],
- verifiquen que funcionan correctamente usando `test.metadata`,
- tengan la gentileza de contribuir con ellos al desarrollo del paquete.