

Paquete rECH

Caja de herramientas para procesar la Encuesta Continua de Hogares (ECH) de Uruguay

Gabriela Mathieu y Richard Detomasi

SIESTA, 22 de octubre de 2019



Introducción



- La Encuesta Continua de Hogares (ECH), releva información de los **hogares particulares uruguayos**, realizada anualmente por el Instituto Nacional de Estadística (INE).
- Objetivos: son obtener **indicadores** para monitorear la evolución del **mercado de trabajo**, el **ingreso** de los hogares y las condiciones de vida de la población.
- Es similar a la EPH de Argentina que se puede analizar con el paquete **eph** (Argentina's Permanent Household Survey Data and Manipulation Utilities)

Objetivos

¿Por qué surge la necesidad de crear un paquete?

- Sistematizar procesamientos para el cálculo de indicadores
- Brindar un manejo fácil de los microdatos de la ECH, al permitir la obtención directa de indicadores (estimaciones) por parte de **personas no expertas**.
- Acercar a R a personas usuarias de otros programas
- Contribuye a la **democratización** de la información pública y el acceso a los datos abiertos.

Características

- rECH permite la conexión a los microdatos de la encuesta desde **2010 hasta 2018** y a los valores del Índice de Precios al Consumo (IPC) de cada año.
- Para un manejo más accesible, los microdatos se **estandarizan** de manera que una misma variable tenga el mismo nombre y formato en todos los años.
- Es posible usar los microdatos con otros nombres de variables pero deberá indicarse en la función que corresponda a que nombre estándar corresponde.
- Algunos **paquetes** de R que se utilizan en este paquete son, **srvyr**, **survey** [`@survey; @survey1`] para las estimaciones e **ineq** [`@ineq`] para los indicadores de distribución del ingreso, **ggplot2** [`@ggplot`] para gráficos, **sf** [`@sf`] para mapas.
- Programado en R 3.6.1 [`@r`]. Quedará accesible en **github** lo antes posible.

Funciones

- El cálculo de un conjunto amplio de **indicadores** (mercado laboral, educación, ingresos, pobreza, salud, etc.) se obtiene ejecutando funciones específicas.
- Las funciones crean **nuevas variables** en la base, cuando es necesario, y permiten obtener **estimaciones puntuales** y por **intervalo** a nivel de hogares o personas -según corresponda-, tablas, gráficos y mapas,
- Permitiendo la apertura por diferentes variables relevantes (departamento, ascendencia étnico-racial, sexo, etc.), siempre que la cantidad de casos sea suficiente.

Classes

```
class "ech"  
slots:  
  Name: h  
  class: data.frame data.frame p metadata  
  list
```

```
class "indicator"  
slots:  
  Name: table  
  class: data.frame plot gg map metadata  
  character
```

Opiniones

- ¿Son necesarias clases específicas para este paquete?
 - Alternativas entre bases de H y P por separado o juntas
 - Formatos uniformes del objeto de input
 - Coherencia entre bases y metadata
- ¿Las clases planteadas son las adecuadas?
 - La clase ``indicator`` debería ser solo la estimación
- ¿Hay otras clases que puedan interesar?

Carga de datos

- Se plantea consumir las bases directo de la web del INE
 - *Actualización automática de las correcciones oficiales*
- El formato como el INE disponibiliza los microdatos no es homogéneo a lo largo del tiempo
 - Archivos comprimidos en diferentes formatos: .zip, .rar
 - ¿Cómo descomprimir un .rar desde R? ¿Hacerlo desde python?
- Independencia de usar bases propias
 - *Necesidad de mantener lógica de nombres de variables*
- ¿Por qué el INE no publica microdatos en formatos R?
- ¿Es preferible que trabajar con .Rds o .RData?

Variables y Valores

- Diccionario acumulado compatible

descripcion	var06	var07	var08	var09	var10	var11	var12	var13	var14	var15
SEXO	e26	e27	e27	e26	e26	e26	e26	e26	e26	e26
EDAD	e27	e28	e28	e27	e27	e27	e27	e27	e27	e27
RELACIÓN DE PARENTESCO	e31	e32	e32	e30	e30	e30	e30	e30	e30	e30

- Repositorio de definición de compatibilización
- Indicadores (habidos y por haber...)

Ejemplo de función: edu.years()

Descripción	Variable	Códigos
Asistencia alguna vez a establecimiento de enseñanza	e49	1 = Sí / 2 = No
Años aprobados posgrado	e51_11	Nº
Años aprobados Univ. o similar	e51_9	Nº
Años aprobados Terciario no univ.	e51_10	Nº
Años aprobados Magisterio o Prof.	e51_8	Nº
Años aprobados Educ. Técnica	e51_7	Nº
Años aprobados Bachillerato Tecno.	e51_6	Nº
Años aprobados Bachillerato	e51_5	Nº
Años aprobados Ciclo Básico	e51_4	Nº
Años aprobados Primaria Común	e51_2	Nº
Años aprobados Primaria Especial	e51_3	Nº
Asistencia educ. preescolar	e193	1 Asiste act.; 2 Sí, asistió; 3 No asistió

Ejemplo de función: edu.years()

descripcion	var06	var07	var08	var09	var10	var11	var12
asistencia alguna vez a un establecimiento de enseñanza	e51	e53	e51	e50	e50	e50	e49
asistencia a educación preescolar						e193	e193
años aprobados en primaria común	e52_1_1	e54_1_1	e54_1_1_V	e51_2	e51_2	e51_2	e51_2
años aprobados en primaria especial				e51_3	e51_3	e51_3	e51_3
años aprobados en ciclo básico				e51_4	e51_4	e51_4	e51_4
años aprobados en bachillerato	e52_2_1	e54_2_1	e54_2_1_V	e51_5	e51_5	e51_5	e51_5
años aprobados en bachillerato tecnológico				e51_6	e51_6	e51_6	e51_6
años aprobados en educación técnica	e52_3_1	e54_3_1	e54_3_1_V	e51_7	e51_7	e51_7	e51_7
años aprobados en magisterio o profesorado	e52_4_1	e54_4_1	e54_4_1_V	e51_8	e51_8	e51_8	e51_8
años aprobados en universidad o similar	e52_5_1	e54_5_1	e54_5_1_V	e51_9	e51_9	e51_9	e51_9
años aprobados en terciario no universitario	e52_6_1	e54_6_1	e54_6_1_V	e51_10	e51_10	e51_10	e51_10
años aprobados en postgrado	e52_7_1	e54_7_1	e54_7_1_V	e51_11	e51_11	e51_11	e51_11

Ejemplo de función: `edu.years()`

This function allows you to estimate the years of education.

Usage

```
edu.years(data = ech2017, output = "var", geo.unit = "uy",  
by = NULL, weights = ~pesoano, sf = FALSE)
```

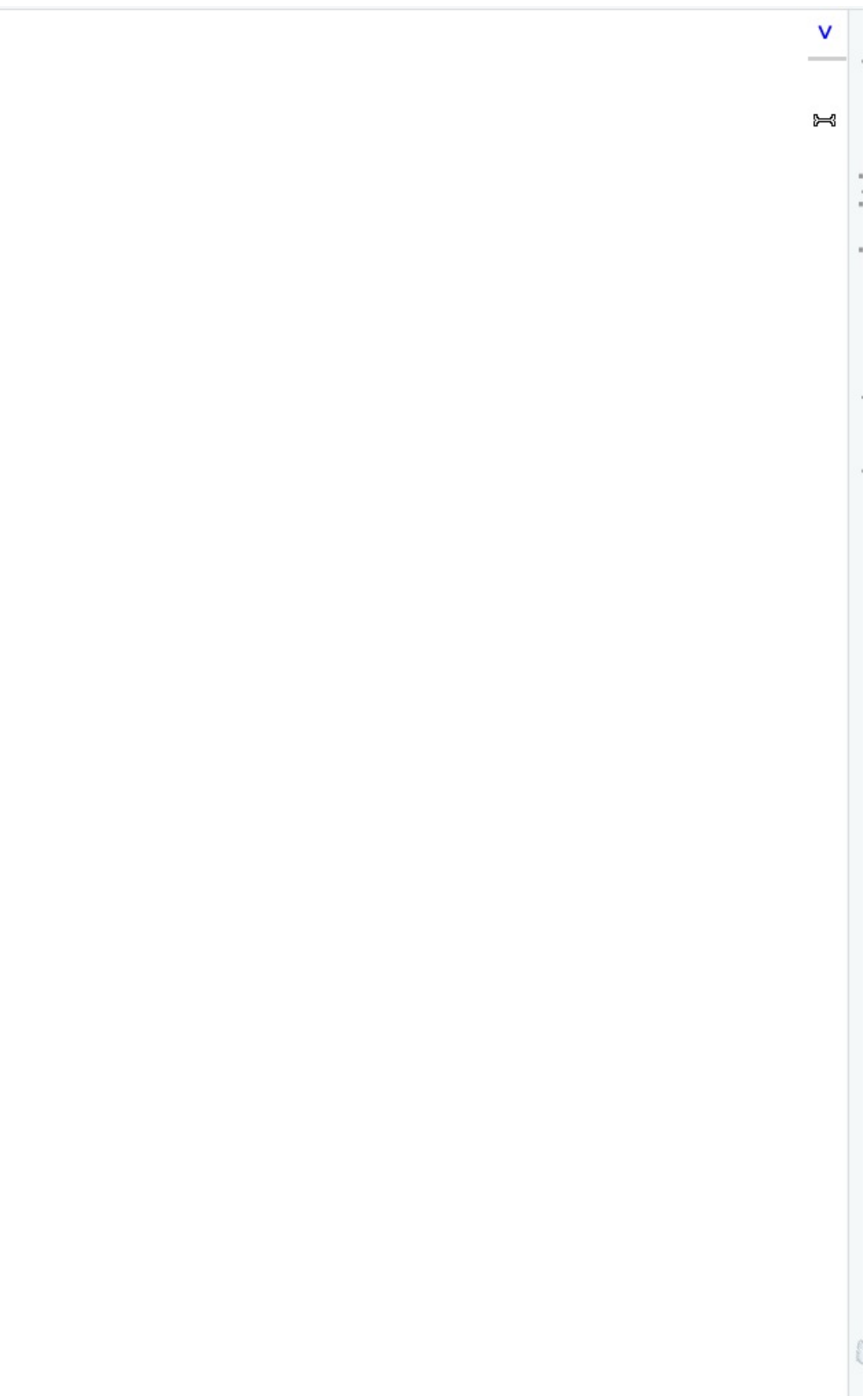
Arguments

<code>data</code>	Dataframe with necessary variables Defaults to ech.
<code>output</code>	May be "var" (default) or "ind", and determines that the output is the ECH with the new variables generated or the corresponding indicator.
<code>geo.unit</code>	The corresponding administrative unit to make the indicator; NULL (default and correspond to all country), other options: "region_3", "region_4", "dpto", "secc" (Only works with 'output="ind"')
<code>by</code>	A vector at most 2 variables of the ECH object (Only works with 'output="ind"')
<code>weights</code>	Default is NULL and correspond to "pesoano" (year estimation), but have others options: "pesosem" to six-monthly cut, "pesotri" to quarterly cut, and "pesomen" to monthly cut.
<code>sf</code>	Default is FALSE, but if TRUE return the 'Sf' object of the indicator

Objetos de salida

- `ech` con nuevas variables
 - Se crean las variables necesarias para el cálculo del indicador correspondiente
- `indicator` con la estructura correspondiente a la combinación de `geo.unit` y `by`
 - En `'geo.unit'` se indica la unidad espacial
 - En `'by'` se indica(n) la(s) variable(s) de agrupación
- `sf` el indicador con la geometría correspondiente
 - A nivel de departamento y sección

`edu.years(output = "ind")`




`edu.years(output = "ind", geo.unit = "dpto")`



```
edu.years(output = "ind", geo.unit = "dpto", by =  
"e26")
```



!  GRACIAS!