

Librería para R: Datos Abiertos CEP

Herramienta para facilitar el acceso a los
Datos Abiertos del CEP XXI

Junio 2023



Ministerio de Economía
Argentina

Secretaría de Industria
y Desarrollo Productivo

CEPXXI
CENTRO DE ESTUDIOS
PARA LA PRODUCCIÓN



DESCRIPCIÓN

La librería para R DatosAbiertosCEP tiene por objetivo colaborar en el acceso y el procesamiento de los datos puestos a disposición desde el Centro de Estudios para la Producción XXI, en la página de Datos Abiertos de la Secretaría de Industria y Desarrollo Productivo del Ministerio de Economía.

El paquete actualmente cuenta con las siguientes funciones:

DA_cruces_existentes(): permite ver el listado de las diferentes bases disponibles junto con su valor *index* (ya que cada serie tiene un valor único identificatorio) y los niveles de desagregación existentes;

descarga_DA(): permite la descarga de cada una de las bases disponibles. Se puede hacerlo indicando su valor correspondiente de *index* o especificando los parámetros puntuales, como:

- **tipo:** define la variable de interés.
valores: Percentil Salario, Salario Mediano, Salario prom por depto y sector, Puestos provincia y sector, Salario mediano por provincia y sector, Var. mediana facturación, Cant empleadoras, Puestos depto, Salario promedio, Salario mediano por depto y sector, Salario prom por provincia y sector, % Muj provincia y sector, Puestos por sector, % Muj por sector.
- **género:** indica si se quiere la base desagregada por sexo biológico o no.
valores: SI, NO.
- **jurisdicción:** indica los diferentes niveles de desagregación geográfica que tiene el dato.
valores: NO, Departamento vivienda, Provincia trabajo, Departamento fiscal, Provincia vivienda.
- **universo:** refiere al universo de empleadores con el que se quiere trabajar.
valores: NO, Privado, Total empresas, Público, Total empleo.
- **sector:** indica el nivel de desagregación de la actividad económica.
valores: NO, Letra, CLAE2, CLAE3, CLAE6.
- **show_info_DA:** permite el acceso a la metodología del dataset.
por default es verdadero, pero puede desactivarse.

deflactar_DA(): permite llevar a precios constantes los valores monetarios nominales de las bases, mediante la utilización del Índice de Precios al Consumidor (IPC) de INDEC. Este comando permite elegir el mes base (en el que el valor del índice es = 100).



indexar_DA(): permite indexar las variables numéricas pudiendo elegir como base de referencia el valor máximo o mínimo de la serie o bien una fecha escogida manualmente, y así poder comparar respecto de ese valor elegido.

diccionario_sectores(): permite sumar niveles de agregación sectorial a los datos descargados. Por ejemplo, en caso de descargar información a tres dígitos (CLAE3), con esta función se puede agregar el CLAE2 y la Letra a la que pertenece ese sector, junto con las descripciones de cada desagregación sectorial.

dicc_depto_prov(): permite agregar fácilmente el nombre del departamento y la provincia a la que pertenece cada dato, en caso de estar desagregado a nivel departamental.

tablero_provincial(): genera un tablero en la aplicación *shiny* en el que se puede analizar información provincial.

A su vez, el paquete permite acceder rápidamente al diccionario de departamentos y de sectores. Para ello se deberá escribir `dicc_depto` y `dicc_sector`, respectivamente.



AUTORES

Este paquete surge de un trabajo colectivo por parte del equipo de datos del Centro de Estudios para la Producción XXI. El mismo está conformado por Nicolás Sidicaro, Gisella Pascuariello, Gisel Trebotic, Ignacio Paola y Guido Sánchez. A su vez, se agradecen los comentarios y el acompañamiento de Pablo Sonzogni (coordinador del área de datos) y Florencia Asef Horno (directora del CEP XXI).



INSTALACIÓN

El paquete actualmente se encuentra en desarrollo, por lo que se debe instalar mediante github.

- `install.packages('devtools')`
- si no tiene instalado devtools
`devtools::install_github("CEProduccionXXI/DatosAbiertosCEP")`



¿CÓMO USARLO?

VER CATÁLOGO DE DATASETS →

DA_cruces_existentes()

Con la función `DA_cruces_existentes()` se accede a la información que puede descargarse mediante el paquete. Allí se encuentran la mayoría de las bases de datos puestas a disposición en sus distintas desagregaciones, así como un indicador único por base que permite descargarlas más fácilmente (*index*).

- Para ver los posibles cruces:
`View(DA_cruces_existentes())`
- Si se quieren ver los posibles cruces de algún dato. Por ejemplo, los datos disponibles de salario mediano para el universo de empresas privadas:
`View(DatosAbiertosCEP::DA_cruces_existentes(tipo = 'Salario mediano', universo = 'Privado'))`

DESCARGAR DATASETS → `descarga_DA()`

Una vez seleccionada la base con la que se quiere trabajar, se puede descargar con `descarga_DA()`. Hay dos opciones:

OPCIÓN 1

Una primera opción para descargar el dato es mediante el `index` de la base:

```
datos2 <- descarga_DA(index_base = 29)
```

OPCIÓN 2

Se deberán indicar los cruces buscados. Se deben utilizar los cinco parámetros:

- `tipo`: refiere al dato que se quiere descargar;
- `genero`: indica si se quiere la base desagregada por sexo biológico o no;
- `sector`: indica el nivel de desagregación de actividad económica;
- `jurisdicción`: indica los diferentes niveles de desagregación geográfica que tiene el dato;
- `universo`: refiere al universo de empleadores con el que se quiere trabajar. Se puede escoger por empresas privadas, empresas públicas y privadas, empleadores públicos, total del empleo y NO (para los casos en los que no se indica el universo).



Por ejemplo:

```
datos <- descarga_DA(tipo = 'Salario mediano', universo  
= 'Privado', sector = 'CLAE6', jurisdiccion =  
'NO', genero='NO')
```

- Los datos obtenidos con cualquiera de estas opciones serán iguales.
- La función cuenta con el parámetro `show_info_DA` para facilitar el acceso a la metodología de la base. Este parámetro por default es verdadero, pero eventualmente puede desactivarse.

INFORMACIÓN AUXILIAR → `diccionario_sectores()`

Asimismo, a los datos descargados se les podrá añadir información referida a la descripción del sector productivo. Por un lado, este comando permite añadir los códigos de los distintos niveles de agregación del Clasificador de Actividad Económica (CLAE), que son: 6, 3, 2 dígitos y Letra (por default se añaden todas las categorías). Por otro lado, se pueden agregar las descripciones de los sectores (por default aparecen pero es posible desactivarlas).

- Añadir sector
`datos <- diccionario_sectores(datos,`
Nombre de la base a la que se quiere añadir la información

`agregacion = 'letra'`

Nivel de agregación del que se quiere sumar datos,
opciones: "clae6", "clae3", "clae2", "letra"

`descripciones = T)`

T si se quiere sumar la descripción / F si no

INFORMACIÓN AUXILIAR → `dicc_depto_prov()`

Asimismo, a los datos descargados se les podrá añadir información sobre la desagregación geográfica, siempre que corresponda.

- Añadir departamento
`test <- descarga_DA(index_base = 19)`
Para descargar alguna base que tenga departamento, bajo el nombre de "test"

`test <- dicc_depto_prov(test)`

Nombre de la base a la que se quiere añadir la información



OPERACIONES AUXILIARES → deflactar_DA()

También es posible convertir a precios constantes los valores monetarios corrientes, escogiendo las variables que se quieren deflactar y el mes base sobre el que se desea operar (en el que el valor índice es = 100). Como deflactor se utiliza una serie propia elaborada en base al Índice de Precios al Consumidor (IPC) de INDEC y, para ciertos períodos puntuales, el relevamiento de precios de otras Direcciones Provinciales de Estadística. Cabe mencionar, que el número índice empalmado que se construyó fue en base a diferentes índices en los cuáles existen diferencias de representatividad. El código permite o bien generar nuevas variables con los valores constantes, o sobrescribir las variables originales.

- Llevar a precios constantes los valores monetarios

```
test <- descarga_DA(index_base = 1)
```

Para descargar alguna base que tenga valores monetarios, bajo el nombre de "test"

```
test <- deflactar_DA(data = test,  
                    mes_base='2022-11-01',
```

Fecha de base sobre la que se quiere tener el valor constante de las variables. Formato: "YYYY-MM-DD"

```
variables_monetarias = "
```

VARIABLES que se desean deflactar. En caso de que no se utilice el parámetro o se lo deje en blanco, se deflactarán todas

```
pisar_datos == T)
```

En caso de ser verdadero se reemplazarán las variables originales por las deflactadas. Por default es falso.

OPERACIONES AUXILIARES → indexar_DA()

La función `indexar_DA()` permite indexar las variables numéricas respecto a algún valor puntual (su respectivo valor máximo, mínimo o alguna fecha en particular que deberá tener formato YYYY-MM-DD). Si la serie llegase a estar desagregada en varias dimensiones, por ejemplo, datos de empleo por letra y por provincia, es posible elegir si se quiere fijar la base de comparación teniendo en cuenta sólo la letra; sólo la provincia o ambas (el default toma todas las desagregaciones).



- Indexar variables seleccionadas

```
test <- descarga_DA(index_base = 7)
```

Para descargar una serie con el nombre "test"

```
test <- indexar_DA(data = test,  
                  base_indice='max')
```

Acá se toma el máximo de la serie como base del índice (también se podrían tomar valores 'min' o una fecha YYYY-MM-DD).

```
variables_datos_abiertos = c('w_mean','p10')
```

Vector con variables a indexar. Por default se indexan todas las variables posibles. En caso de querer alguna específica, indicarlo en la función.

```
variables_agrupar = c('fecha','mujer')
```

Indica las variables que se desean agrupar para tomar base de comparación del índice. Pueden ser "todas",¹ "ninguna" o un vector con variables seleccionadas. Por default, se tienen en cuenta todas las columnas por los que esté desagregada nuestra variable de interés.

```
pisar_datos = F)
```

Indica si se sobreescriben las variables originales. Si es F, entonces se agregan columnas, si es T se pisan las variables originales.

VISUALIZACIÓN → tablero_provincial()

Para construir el tablero es necesario ejecutar la función `tablero_provincial()` que lo abrirá directamente. Si el cuadro abierto queda gris, deberá apretarse en "abrir en navegador" para poder visualizarlo. No se puede agrupar por la variable de fecha, lógicamente, ya que justamente queremos tener una evolución temporal. La opción `variables_agrupar()` tiene sentido solamente si se indexa por máximo o mínimo.

¹ No se puede agrupar por la variable de fecha, lógicamente, ya que justamente queremos tener una evolución temporal. La opción `variables_agrupar()` tiene sentido solamente si se indexa por máximo o mínimo.



OTRAS SERIES DE DATOS DISPONIBLES



APORTES

El paquete también incorpora datos relevados por el equipo de Coyuntura del CEP XXI. Estos datos provienen de diferentes fuentes y pueden encontrarse en indicadoresargentina.produccion.gob.ar.

Es posible corroborar el listado de series disponibles con el comando `DPexistentes()`, donde figura el nombre de la serie, su descripción, la unidad de medida en la que se encuentra, la temporalidad, la fuente y otros datos importantes. Para acceder se puede ejecutar directamente la función o bien guardar el resultado en un elemento:

```
datos_productivos <- DPexistentes()
```

Aclaración sobre la función `DPexistentes()`: los valores en pesos constantes se encuentran deflactados siempre en función del último mes disponible, independientemente de lo que indique la unidad de medida de cada indicador (la variable “um”). Esto se debe a un problema del sistema de visualización de los tableros que está en proceso de ajuste.

Una vez seleccionado el dato que se quiere descargar se puede utilizar la función `descarga_DP()` para acceder a la serie.

```
test <- descarga_DP(index=30
Serie que se quiere descargar, el index proviene de la
variable ID de DPexistentes()
    show_info_data = T)
Muestra información relevante de la serie
(default = T)
```

El paquete actualmente se encuentra en desarrollo, por lo que se irá actualizando con el correr de los meses. Si se encuentra algún error o se desea proponer alguna funcionalidad que no fue tomada en cuenta se agradecerán los comentarios mediante [esta plataforma](#).



Ministerio de Economía
Argentina

Secretaría de Industria
y Desarrollo Productivo

CEPXXI CENTRO DE ESTUDIOS
PARA LA PRODUCCIÓN