INE data extraction with ineapir : : **CHEAT SHEET**

The **ineapir** package allows to extract open data and metadata published by the INE (Spain). The data is obtained using calls to the INE API JSON service which access via URL requests to the data required by introducing the ID of the serie/tabla desired.

1 How to obtain ID's

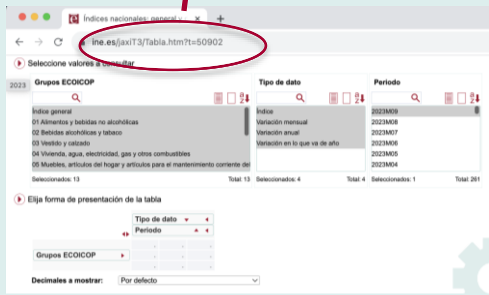
Go to INE website and find a table/ series with the desired data (no need to be filtered). Depending the type of data hosted on the INE database we can distinguish several types of URL's:

FROM A TABLE

CASE 1: tempus type

The **idTable** is the t parameter, **50902**

ine.es/jaxiT3/Tabla.htm?t=50902

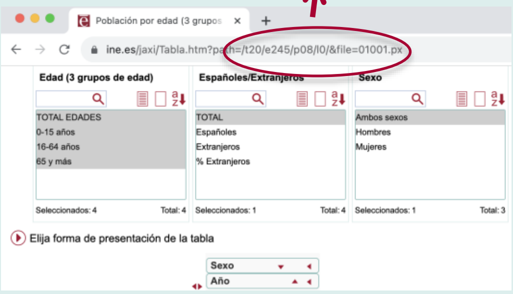

Índices nacionales: general
ine.es/jaxiT3/Tabla.htm?t=50902
Seleccione valores a consultar
Grupos ECOICOP
Índice general
01 Alimentos y bebidas no alcohólicas
02 Bebidas alcohólicas y tabaco
03 Vestido y calzado
04 Vivienda, agua, electricidad, gas y otros combustibles
05 Muebles, artículos del hogar y artículos para el mantenimiento corriente del
Tipo de dato
Índice
Variación mensual
Variación anual
Variación en lo que va de año
Periodo
2023M09
2023M08
2023M07
2023M06
2023M05
2023M04
Elija forma de presentación de la tabla
Tipo de dato
Periodo
Grupos ECOICOP
Decimales a mostrar:
Por defecto

`idTable = "50902"`

CASE 2: pc- axis

The **idTable** is the concatenation of the path and file, **t20/e245/p08/l0/01001.px**

ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=01001.px

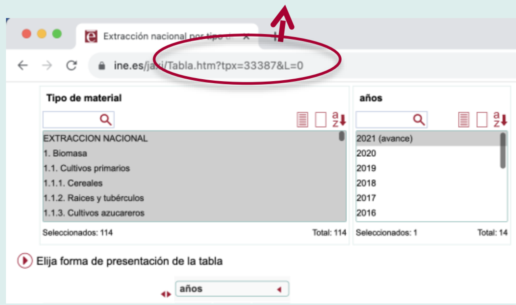

Población por edad (3 grupos
ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=01001.px
Edad (3 grupos de edad)
TOTAL EDADES
0-15 años
16-64 años
65 y más
Seleccionados: 4 Total: 4
Españoles/Extranjeros
TOTAL
Españoles
Extranjeros
% Extranjeros
Seleccionados: 1 Total: 4
Sexo
Ambos sexos
Hombres
Mujeres
Seleccionados: 1 Total: 3
Elija forma de presentación de la tabla
Sexo
Año

`idTable ="t20/e245/p08/l0/01001.px"`

CASE 3: tpx

The **idTable** is the tpx parameter, **33387**

https://www.ine.es/jaxi/Tabla.htm?tpx=33387&L=0

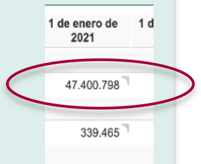

Extracción nacional por tipo
ine.es/jaxi/Tabla.htm?tpx=33387&L=0
Tipo de material
EXTRACCION NACIONAL
1. Biomasa
1.1. Cultivos primarios
1.1.1. Cereales
1.1.2. Raices y tubérculos
1.1.3. Cultivos azucareros
Seleccionados: 114 Total: 114
años
2021 (avance)
2020
2019
2018
2017
2016
Seleccionados: 1 Total: 14
Elija forma de presentación de la tabla
años

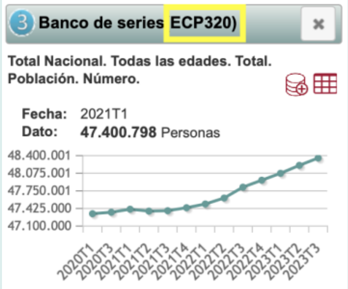
`idTable="33387"`

FROM A SERIES

CASE 4: series

1. Browse a table of interest
2. Filter the selected values
3. Click con the corresponding value cell
4. Take the codSeries that appears on the plot





`codSeries="ECP320"`

2 Main functions

OBTAINING DATA

- ***get_data_table(****idTable, filter, nlast, det, tip, lang, validate , verbose, unnest, metanames, metacodes****)***
  It returns the data of the idTable specified according to the filter and the other arguments.
- ***get_data_series(****codSeries,nlast, dateStart, dateEnd, det, tip, lang, validate , verbose, unnest****)***
  It returns the data of the codSeries specified according to the parameters.
- ***get_data_series_filter(****operation, filter, periodicity, nlast, det,…****)***
  It returns the data of the operation specified according to the filter and the other parameters.

Auxiliar functions for Operations

- ***get_metadata_operations****(operation, ..)*
  It returns all available operations if no argument is passed.
- ***get_metadata_periodicity****(idem.)*
  It returns the periocities for the specified operation.
- ***get_metadata_publications(****operation, ..****)***
  It returns all available publications for the specified operation.
- ***get_metadata_series_operation(****operation, ..****)***
  It returns all series involved in the specified operation.
- ***get_metadata_variables / values(****operation,(variable) ..****)***
  It returns all available variables for a specific operation. In case of values it returns the values for a specific operation/variable.
- ***get_metadata_series_varval(****operation, ..****)***
  It returns all variables/values for the specified operation.
- ***get_metadata_tables_operation*** *(operation, ..****)***
  Get all tables for a given operation

Auxiliar functions for Tables

- ***get_metadata_table_groups/values(****idTable, (idGroup),...****)***
  It returns all available groups and values for the specified table.
- ***get_metadata_table_ varval(****idTable, ..****)***
  Get metadata information about the variables and values for a given table
- ***get_metadata_series_table****(idTable, filter ..****)***
  Get all the series for a given table
- ***get_metadata_operation_table(****idTable, ..****)***
  It returns all operations for the specified table.

Auxiliar functions for Series

- ***get_metadata_series****(codSeries, tip..****)***
Get information for the given series.
- ***get_metadata_series_values****(codSeries, tip..****)***
Get all the values (and variables) for the given series.

3 Arguments

OPTION	TYPE	DEFAULT	EFFECTS
idTable	int		Id of the table
codSeries	string		Code of the series.
operation	string		Code of the operation. To obtain a list of available operations see get_metadata_operations().
filter	list		Filter variables for the specified values. For more information see Section 3a (filter)
nlast	int	NULL / 1	Number of periods to retrieve. By default is set to all periods for tables and 1 for series.
dateStart	yyyy/mm/dd		Initial date of the requested data. If dateStart and dateEnd are equal, the specified dates are retrieved
dateEnd	yyyy/mm/dd		End date of the requested data. If no end date is entered, all dates will be queried, from the corresponding start date to the last available period.(*) (*)
det	int	NULL	Level of detail. Valid values: 0, 1 or 2.
tip	string	NULL	'A' for friendly output (e.g. readable dates), set to 'M' to include metadata or set to 'AM' for both.
lang	string	"ES"	Language for the data. 'ES' for Spanish or 'EN' for English.
validate	logical	TRUE	Validate the input parameters. A FALSE value means fewer API calls.
verbose	logical	FALSE	Print additional information, including the URL to call the API service.
unnest	logical	FALSE	Set to TRUE to obtain a single data frame of data.
metanames	logical	FALSE	Set to TRUE to extract the name of the values that defined the table. (**)
metacodes	logical	FALSE	Set to TRUE to extract the codes and ids of the values that defined the table. (**)

(*) It can be a vector of dates, where each date represents the end date of individual ranges where the initial/final date should be found at the same position in the dateEnd/dateStart vector. The length of the vector must be ≤ to the length of the dateStart vector.

(**) Setting *metanames/metacodes = TRUE* is useful to get the metadata and therefore see how the data has been filtered, the measures,…

3 a filter argument

Data from **tables** and **operations** can be filtered with a list according to the variables/values they contain. Let's see how to construct the filter:

FOR TABLES

See get_metadata_table_varval(idTable) to get all the values involved in the filter at once. There are different approaches to build the filter depending on the table type:

1. **tempus**: The return of get_metadata_table_varval(idTable) is like:

```
#>    Id            Fk_Variable             Nombre              Codigo
#> 1 304092         762                     Índice general       00
```

   so the filter can be constructed as:

   filter<- list(Fk_Variable1 = Id1 , …..)

2. **pc-axis**: **The return of** get_metadata_table_varval(idTable) is like:

```
#>      Nombre    Codigo     Variable.Nombre        Variable.Codigo
#> 1    Hombres   hombres    Sexo                   sexo
```

   so the filter can be constructed as:

   filter<- list(Variable.Codigo = Codigo , …..)

3. **tpx**: **The return of** get_metadata_table_varval(idTable) is like:

```
         Codigo                    Variable.Nombre        Variable.Codigo
 #1      extraccionnacional        tipo de material        tipodematerial
```

   so the filter can be constructed as:

   filter<- list(Variable.Codigo=Codigo , …..)

Additional comments

1. **Several values**: For letting a variable take several values, use concat. operator, ie,
   list(id_variable1 = c(id_value1a,id_value1b))
2. **All values**: For letting a variable take all posible values, left the equal empty,ie,
   list(id_variable1 = "")

Continues on next page

Shorcuts

Instead of using Id's /codes for filtering variables, shorcuts can be used. They are standardized expressions. Their format is:

filter= list(Shorcut= value1, Shorcut= value2)

See get_filter_shorcuts() to get all the shorcuts at once. For a better performance is recommended to use numeric codes instead of shortcuts.

Shortcut	Variable ID	Comment
nac	349	National
prov	115	Provinces
isla	20	Islands
edad	355, 356, 357, 360	Age wrapper

Remark: There exists a wrapper which detects the variable being used. Its format is the following:

filter= list(values = c(id_value1, id_value2))

Regular Expressions

When **shorcuts** are used for filtering, regular expressions can be used in the filter. Let's show some illlustrative examples.
Example:

```
filter <- list(sexo = "total", edad = "2[0-5]+")
```

- Shorcut **sexo** and **edad** are used for the filter. This shorcut sexo linked with variable 18 and the edad is a wrapper between variables 355, 356, 357, 360
- For filter the **edad** a regular expression is used, edad = "2[0-5]+" . It filters all ages between 20 and 25 years old.

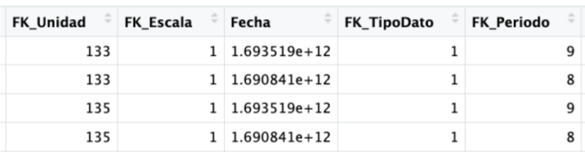
OPTION	TYPE	DEFAULT	EFFECTS
idTable	int		Id of the table
codSeries	string		Code of the series.
operation	string		Code of the operation. To obtain a list of available operations see get_metadata_operations().
filter	list		Filter variables for the specified values. For more information see
nlast	int	NULL / 1	Number of periods to retrieve. By default is set to all periods for tables and 1 for series.
det	int	NULL	Level of detail. Valid values: 0, 1 or 2.
tip	string	NULL	'A' for friendly output (e.g. readable dates), set to 'M' to include metadata or set to 'AM' for both.
lang	string	"ES"	Language for the data. 'ES' for Spanish or 'EN' for English.
validate	logical	TRUE	Validate the input parameters. A FALSE value means fewer API calls.
verbose	logical	FALSE	Print additional information, including the URL to call the API service.
unnest	logical	FALSE	Set to TRUE to obtain a single data frame of data.
metanames	logical	FALSE	Set to TRUE to extract the name of the values that defined the table.
metacodes	logical	FALSE	Set to TRUE to extract the codes and ids of the values that defined the table.

3 b tip

Argument that deals with the layout of the output and with the metadata.

tip = NULL

No readable dates and no metadata.

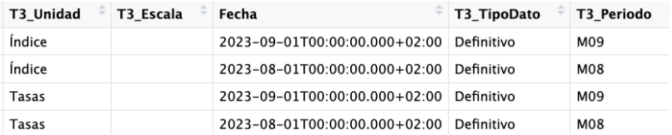


FK_Unidad	FK_Escala	Fecha	FK_TipoDato	FK_Periodo
133	1	1.693519e+12	1	9
133	1	1.690841e+12	1	8
135	1	1.693519e+12	1	9
135	1	1.690841e+12	1	8

tip = 'A'

Friendly output:

Dates in yyyy/mm/dd format and variables/values in names instead of codes



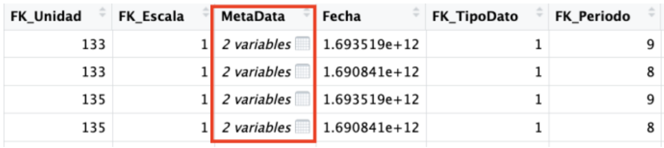
T3_Unidad	T3_Escala	Fecha	T3_TipoDato	T3_Periodo
Índice		2023-09-01T00:00:00.000+02:00	Definitivo	M09
Índice		2023-08-01T00:00:00.000+02:00	Definitivo	M08
Tasas		2023-09-01T00:00:00.000+02:00	Definitivo	M09
Tasas		2023-08-01T00:00:00.000+02:00	Definitivo	M08

Notice the difference with the previous image

tip = 'M'

Metadata included as a dataframe.

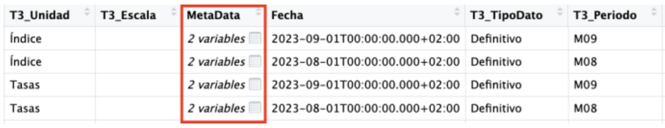
e.g. Variables/values filtered.



FK_Unidad	FK_Escala	MetaData	Fecha	FK_TipoDato	FK_Periodo
133	1	*2 variables*	1.693519e+12	1	9
133	1	*2 variables*	1.690841e+12	1	8
135	1	*2 variables*	1.693519e+12	1	9
135	1	*2 variables*	1.690841e+12	1	8

tip = 'AM'

Friendly output and metadata included.



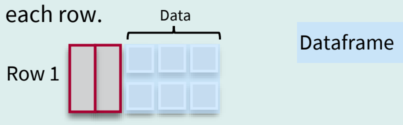
T3_Unidad	T3_Escala	MetaData	Fecha	T3_TipoDato	T3_Periodo
Índice		*2 variables*	2023-09-01T00:00:00.000+02:00	Definitivo	M09
Índice		*2 variables*	2023-08-01T00:00:00.000+02:00	Definitivo	M08
Tasas		*2 variables*	2023-09-01T00:00:00.000+02:00	Definitivo	M09
Tasas		*2 variables*	2023-08-01T00:00:00.000+02:00	Definitivo	M08

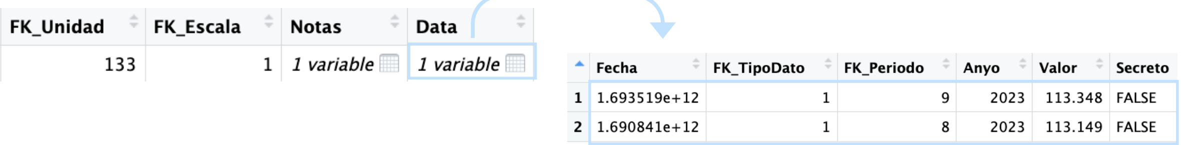
3 c unnest

Parameter that allows to unnest the data. Advisable to set it always to TRUE.

unnest = FALSE

Data is contained in dataframes for each row.



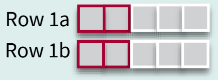


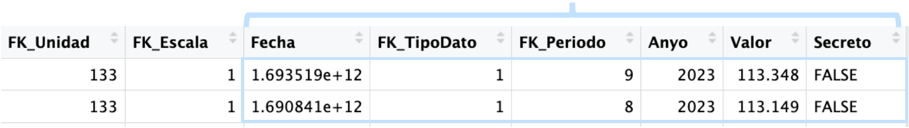
FK_Unidad	FK_Escala	Notas	Data
133	1	*1 variable*	*1 variable*

	Fecha	FK_TipoDato	FK_Periodo	Anyo	Valor	Secreto
1	1.693519e+12	1	9	2023	113.348	FALSE
2	1.690841e+12	1	8	2023	113.149	FALSE

unnest = TRUE

For each row, the dataframes of the data are extracted in columns/rows





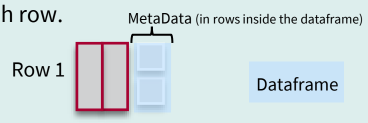
FK_Unidad	FK_Escala	Fecha	FK_TipoDato	FK_Periodo	Anyo	Valor	Secreto
133	1	1.693519e+12	1	9	2023	113.348	FALSE
133	1	1.690841e+12	1	8	2023	113.149	FALSE

3 c metanames / metacodes

It extracts the metadata from the corresponding dataframe and it adds a column for each variable , takin the corresponding value of the variable. Advisable to set it always to TRUE in case that the data wants to be filtered in a posterior analysis by a certain variable/value. In case it is set to FALSE the metanames and metacodes are included in a dataframe called "Metadata" for each row.

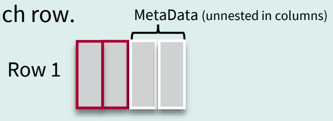
metacodes = FALSE

Data is contained in dataframes for each row.



metacodes = TRUE

Data is contained in dataframes for each row.



Remark:

- For setting metanames/metacodes = TRUE, the argument tip must be set to "M" or "AM".
- It is advised to set them to TRUE.

Continues on next page