

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:

CASE 1: tempus type
The ID is the t parameter, **50902**

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

CASE 2: pc- axis
The ID is the concatenation of path and file, **t20/e245/p08/l0/01001.px**

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

CASE 3: tpx
The ID is the tpx parameter, **33387**

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

CASE 4: series

- Browse a table of interest
- Filter the selected values
- Click on the corresponding value cell
- Take the **code** that appears on the plot

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 periodicities 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 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.

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 the operation for the specified table.

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 <code>get_metadata_operations()</code> .
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.
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.

Setting `metanames/metacodes = TRUE` is useful to get the metadata and therefore see how the data has been filtered, the measures,...

3a 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()` to get all the values at once. There are different approaches to build the filter depending on the table type:

- tempus:** The filter is based on ids, with format: `list(id_variable1 = id_value1, id_variable2 = id_value2)`
- pc-axis:** The filter is based on codes, with format: `list(cod_variable1 = cod_value1, cod_variable2 = cod_value2)`.
- tpx:** The filter is based on codes, with format: `list(cod_variable1 = cod_value1, cod_variable2 = cod_value2)`

For SERIES (operation)

The format is `list(id_variable1 = id_value1, id_variable2 = id_value2)`
The id's of the variables related to an operation can be seen with `get_metadata_variables()` and `get_metadata_values()`

Additional comments

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

Continued 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(shortcut_variable1=name1, shortcut_variable2=name2)
```

See `get_filter_shorcuts()` to get all the shorcuts at once. For a better performance is recommended to use numeric ids for tempus tables and alphanumeric codes for px and tpx tables 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 illustrative examples.

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

- Shortcut **sexo** and **edad** are used for the filter. This shortcut 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.

3b 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

3c 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
1.693519e+12	1	9	2023	113.348	FALSE
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

3c metanames / metacodes

It extracts the metadata from the corresponding dataframe and it adds a column for each variable, taking 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".