Package: pulso (via r-universe)

November 28, 2026

Title Load Microdata from Colombia's 'GEIH' ('DANE')

Version 0.2.1

Description Programmatic access to microdata from Colombia's Gran Encuesta Integrada de Hogares ('GEIH'), published by 'DANE'. Provides a tidy interface to download, parse, and harmonize labor market surveys from 2007 to present. R companion to the 'pulso-co' 'Python' package.

License MIT + file LICENSE

URL https://github.com/Stebandido77/pulso

BugReports https://github.com/Stebandido77/pulso/issues

Encoding UTF-8

Depends R (>= 4.1.0)

Imports httr2 (>= 1.0.0), jsonlite (>= 1.8.0), tibble (>= 3.2.0), rlang (>= 1.1.0), cli (>= 3.6.0), fs (>= 1.6.0), digest (>= 0.6.0), data.table (>= 1.14.0), dplyr (>= 1.0.0), readxl (>= 1.4.0), xml2 (>= 1.3.0)

Suggests haven (>= 2.5.0), vctrs (>= 0.6.0), tidyr, testthat (>= 3.2.0), withr, knitr, rmarkdown

VignetteBuilder knitr

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Esteban Labastidas [aut, cre]

Maintainer Esteban Labastidas <estebanlabastidas123@gmail.com>

Repository https://cran.r-universe.dev

Date/Publication 2026-11-28 04:30:02 UTC

RemoteUrl https://github.com/cran/pulso

RemoteRef HEAD

RemoteSha 53d1c20f9d804a91adf6d287d62c330518efd5a8

Contents

pulso_describe . . . 2
pulso_describe_column . . . 3
pulso_describe_variable . . . 3
pulso_list_columns_metadata . . . 4
pulso_list_validated_range . . . 5
pulso_list_variables . . . 5
pulso_load . . . 6
pulso_load_merged . . . 7
pulso_tpm . . . 9
pulso_update_catalog . . . 10
pulso_validation_status . . . 10

Index **12**

`pulso_describe` *Describe a GEIH module*

Description

Returns a human-readable summary of a GEIH survey module: its survey level, available epochs, and harmonized canonical variables.

Usage

```
pulso_describe(module)
```

Arguments

`module`	Character. Module name (e.g., "ocupados"). Must be one of the modules registered in sources.json.

Value

A multi-line character string.

Examples

```
pulso_describe("ocupados")
```

pulso_describe_column *Describe metadata for a single column*

Description

Pretty-prints metadata for a column in a tibble loaded with `metadata = TRUE`. Output format mirrors Python's `pulso.describe_column()`.

Usage

```
pulso_describe_column(df, column)
```

Arguments

`df`	A tibble loaded with `pulso_load(..., metadata = TRUE)`.
`column`	Character. Column name to describe.

Value

A multi-line character string.

Examples

```
if (interactive()) {
  df <- pulso_load(2024, 6, "ocupados", metadata = TRUE)
  cat(pulso_describe_column(df, "P6430"))
}
```

pulso_describe_variable *Describe a harmonized variable*

Description

Returns a human-readable summary of a canonical variable defined in variable_map.json: its module, description, source mappings per epoch, and any comparability warning.

Usage

```
pulso_describe_variable(variable_name)
```

Arguments

`variable_name`	Character. Canonical variable name (e.g., "sexo"). The lookup is case-insensitive.

Value

A single multi-line character string.

Examples

```
pulso_describe_variable("sexo")
pulso_describe_variable("edad")
```

`pulso_list_columns_metadata` *List metadata for all columns of a tibble*

Description

Returns a tibble summary of all columns with their metadata.

Usage

```
pulso_list_columns_metadata(df)
```

Arguments

`df`	A tibble loaded with `pulso_load(..., metadata = TRUE)`.

Value

A tibble with columns: column, label, type, module, source, has_categories.

Examples

```
if (interactive()) {
  df <- pulso_load(2024, 6, "ocupados", metadata = TRUE)
  summary <- pulso_list_columns_metadata(df)
  print(summary)
}
```

pulso_list_validated_range *List validated GEIH periods*

Description

Returns the registry entries for which the downloaded data has been manually validated against DANE published figures.

Usage

```
pulso_list_validated_range()
```

Value

A tibble with one row per validated (year, month) pair, with columns: `year` (integer), `month` (integer), `epoch` (character), `validated` (logical, always TRUE), `validated_at` (character, ISO-8601 or NA), `num_modules` (integer). Sorted by year, month ascending.

Examples

```
pulso_list_validated_range()
```

pulso_list_variables *List harmonized variables from the variable map*

Description

Returns a tibble summarizing all canonical variables defined in variable_map.json, optionally filtered by module.

Usage

```
pulso_list_variables(module = NULL)
```

Arguments

`module`	Character or NULL. When non-NULL, restricts output to variables whose `module` field equals this value (exact match, case-sensitive). Available modules in the current variable map: `caracteristicas_generales`, `desocupados`, `inactivos`, `ocupados`, `otros_ingresos`, `vivienda_hogares`.

Value

A tibble with one row per variable and columns:

canonical_name chr. Key used throughout the pulso package.

module chr. Survey module the variable belongs to.

description_es chr. Spanish description, or NA if absent.

comparability chr. Always NA in this version (see Note).

has_warning lgl. TRUE when a comparability_warning is present and non-empty for the variable.

num_epochs int. Number of epoch mappings defined.

Rows are sorted ascending by canonical_name.

Note

The comparability column is always NA_character_ in the current version because the comparability field (expected values: "high" / "limited") has not yet been added to variable_map.json. Only comparability_warning (free text) is present in the data.

Examples

```
# All variables
pulso_list_variables()

# Subset by module
pulso_list_variables(module = "ocupados")
```

---

pulso_load *Load GEIH microdata for a single year-month-module*

Description

Downloads and parses microdata from Colombia's Gran Encuesta Integrada de Hogares (GEIH), published by DANE.

Usage

```
pulso_load(
  year,
  month,
  module,
  area = NULL,
  harmonize = TRUE,
  cache = TRUE,
  metadata = FALSE,
  strict = FALSE,
  allow_unvalidated = NULL
)
```

Arguments

year	Integer. Year (2007 to current year).
month	Integer. Month (1-12).
module	Character. Module name (e.g., "ocupados").
area	Character or NULL. Optional area filter. NOT IMPLEMENTED in v0.1.0.
harmonize	Logical. Whether to apply harmonization. Default TRUE.
cache	Logical. Whether to cache downloads. Default TRUE.
metadata	Logical. Whether to attach DANE metadata to result. Default FALSE for Python parity. When TRUE, attaches metadata via `attr(df, "pulso_metadata")` and triggers lazy download of codebook on first call.
strict	Logical. When FALSE (default, same as 'pulso-co' in 'Python'), periods not yet manually validated against DANE published figures are loaded with a one-time warning per period. When TRUE, such periods raise `pulso_data_not_validated`.
allow_unvalidated	Deprecated. Use `strict` instead (`allow_unvalidated = FALSE` is `strict = TRUE`).

Value

A tibble with the microdata. If metadata = TRUE, the tibble has an attribute "pulso_metadata" with structured column info.

Examples

```
if (interactive()) {
  df <- pulso_load(year = 2024, month = 6, module = "ocupados",
                   metadata = TRUE)
  cat(pulso_describe_column(df, "P6430"))
}
```

---

pulso_load_merged *Load and merge GEIH microdata across multiple modules*

Description

Downloads, parses, and joins microdata from multiple GEIH modules for a single year-month. All modules must be at the same survey level (persona or hogar).

Usage

```
pulso_load_merged(
  year,
  month,
  modules,
  harmonize = TRUE,
  cache = TRUE,
  metadata = FALSE,
  strict = FALSE,
  allow_unvalidated = NULL
)
```

Arguments

year	Integer. Year (2007 to current year).
month	Integer. Month (1-12).
modules	Character vector. Module names to merge (length >= 2). Use pulso_load for a single module.
harmonize	Logical. Whether to lowercase column names. Default TRUE.
cache	Logical. Whether to cache downloads. Default TRUE.
metadata	Logical. Whether to attach DANE metadata. Default FALSE. When TRUE, metadata is composed from the first module's column index (multi-module metadata composition is planned for a future release).
strict	Logical. Passed through to each pulso_load(). Default FALSE: unvalidated periods load with a warning.
allow_unvalidated	Deprecated; see pulso_load.

Value

A tibble with columns from all requested modules, joined on shared identifier keys (directorio, secuencia_p, orden for persona-level modules). The join is an outer join by default so that modules covering different person subsets (e.g., ocupados vs desocupados) are combined correctly.

Note

Mixed-level merges (persona + hogar modules in the same call) are deferred to v0.2.0. If you need a hogar-level module, merge the result manually after separate calls.

Examples

```
if (interactive()) {
  df <- pulso_load_merged(2024, 6, c("ocupados", "caracteristicas_generales"))
  nrow(df)
}
```

`pulso_tpm` *Fetch Tasa de Politica Monetaria (TPM) from Banco de la Republica*

Description

Returns Colombia's monetary policy rate as a tibble. Data source: Banco de la Republica SDMX API (DF_CBR_DAILY_HIST). When the API is unavailable, falls back to a bundled snapshot extending to 2026-04-21.

Usage

```
pulso_tpm(start = NULL, end = NULL, use_fixture = NULL)
```

Arguments

`start`	Optional start date as character "YYYY-MM-DD" or Date.
`end`	Optional end date as character "YYYY-MM-DD" or Date.
`use_fixture`	Logical or NULL. If NULL (default), auto-detect: tries API first, falls back to snapshot if unavailable. If TRUE, always use bundled snapshot. If FALSE, force API call.

Value

A tibble with columns:

fecha Date. Observation date.

valor numeric. TPM rate in percentage points.

serie character. Always "tpm".

Examples

```
if (interactive()) {
  # Recent TPM
  tpm_2024 <- pulso_tpm(start = "2024-01-01", end = "2024-12-31")
  # Full history (uses fixture if API down)
  tpm_all <- pulso_tpm()
}
```

pulso_update_catalog *Update the catalog of DANE releases*

Description

DANE publishes a new GEIH month roughly every four weeks, faster than package releases reach CRAN. This downloads the catalog maintained in the pulso repository (refreshed monthly) into the user cache, so new months become loadable without upgrading the package. The catalog is only replaced when the download is valid and covers a later period than the one in use.

Usage

```
pulso_update_catalog(
  url = "https://raw.githubusercontent.com/Stebandido77/pulso/main/data/sources.json"
)
```

Arguments

`url`	Character. Catalog URL. Defaults to the `main` branch of the pulso repository on GitHub.

Value

Invisibly, a list with `previous_end`, `latest_end`, `n_periods` and `path` (where the catalog was saved, or `NULL` if it was not newer).

Examples

```
if (interactive()) {
  pulso_update_catalog()
}
```

pulso_validation_status *Validation status for a GEIH period*

Description

Returns structured metadata about a specific year-month entry in the pulso registry, including whether it has been manually validated against published DANE figures.

Usage

```
pulso_validation_status(year, month)
```

Arguments

year	Integer. Year (2007 to current year).
month	Integer. Month (1-12).

Value

A one-row tibble with columns: year, month, epoch, validated, validated_by, validated_at, source_url, file_size_mb, modules_available, checksum_sha256.

Examples

```
pulso_validation_status(2024, 6)
```

Index

pulso_describe, 2
pulso_describe_column, 3
pulso_describe_variable, 3
pulso_list_columns_metadata, 4
pulso_list_validated_range, 5
pulso_list_variables, 5
pulso_load, 6, *8*
pulso_load_merged, 7
pulso_tpm, 9
pulso_update_catalog, 10
pulso_validation_status, 10