

Package ‘gbcrosswalk’

July 11, 2026

Title Crosswalk Chinese GB Industry Classifications Across Years

Version 0.0.5

Description Tools for cleaning, building, and composing crosswalks between Chinese GB/T 4754 industry classification vintages. The package starts with the historical workflow for 1986, 1994, 2002, 2011, and 2017 and is designed so later PDF-derived vintages can be added as additional adjacent pairwise crosswalks.

License MIT + file LICENSE

URL <https://github.com/robin621/gbcrosswalk>

BugReports <https://github.com/robin621/gbcrosswalk/issues>

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.1)

Imports utils

Enhances readxl

NeedsCompilation no

Author Yajie Wang [aut, cre] (affiliation: Assistant Professor, Division of Global and Area Studies, Chinese University of Hong Kong (Shenzhen))

Maintainer Yajie Wang <wangyajie621@gmail.com>

Repository CRAN

Date/Publication 2026-07-11 08:50:02 UTC

Contents

gbcrosswalk-package	2
Index	5

gbcrosswalk-package *Crosswalk Chinese GB Industry Classifications Across Years*

Description

Utilities for cleaning, building, and composing Chinese GB/T 4754 industry classification crosswalks. The package standardizes adjacent pairwise crosswalks into a long format and composes them across years.

Usage

```
normalize_gb_code(x, width = 4)
normalize_gb_for_level(x, level = c("S", "M", "L"))
codes_to_gb_level(code, level = c("S", "M", "L"))
derive_gb_levels(data, year, s_col = NULL)
fill_down_missing(x)
dominant_prefix(codes, digits = 3, threshold = 0.5,
                 special_cutoff = NULL, special_suffix = "")
add_collapsed_targets(data, from_col, to_col, to_all_col = NULL,
                      parent_col = NULL, parent_digits = 3,
                      threshold = 0.5, special_cutoff = NULL,
                      special_suffix = "")
as_gb_pair(data, from_year, to_year, from_col = NULL, to_col = NULL,
           levels = c("S", "M", "L"), source_pair = NULL)
compose_gb_crosswalk(pairs, from_year, to_year, level = c("S", "M", "L"),
                    years = c(1986, 1994, 2002, 2011, 2017))
crosswalk_codes(codes, crosswalk, unmatched = NA_character_)
convert_gb_codes(codes, from_year = NULL, to_year = NULL, level = NULL,
                 pairs = load_gb_crosswalks(),
                 years = c(1986, 1994, 2002, 2011, 2017),
                 unmatched = NA_character_)
detect_gb_year(codes, pairs = load_gb_crosswalks(),
               years = c(1986, 1994, 2002, 2011, 2017),
               level = c("auto", "S", "M", "L"),
               ties = c("all", "first"), details = FALSE)
convert_gb_column(data, column, pairs = load_gb_crosswalks(), from_year, to_year,
                  level = c("S", "M", "L"), output_col = NULL,
                  years = c(1986, 1994, 2002, 2011, 2017),
                  unmatched = NA_character_)
standardize_gb_pairs(pairs)
write_gb_crosswalk_csvs(pairs, dir, prefix = "gb", include_combined = TRUE)
read_gb_crosswalk_csvs(path)
gb_crosswalk_path(file = "gb_all_pairs.csv")
gb_raw_crosswalk_path(file = NULL)
load_gb_crosswalks(path = gb_crosswalk_path())
build_gb_1986_1994(path, sheet = 1)
build_gb_1994_2002(path)
```

```
build_gb_2002_2011(path, sheet = 1)
build_gb_2011_2017(path, sheet = 1, old_col = NULL, new_col = NULL)
```

Arguments

x, code, codes	Character or numeric GB codes.
width	Target code width for left-zero padding.
level	One of "S" for 4-digit, "M" for 3-digit, "L" for 2-digit codes, "auto" where supported, or NULL in <code>convert_gb_codes()</code> to infer it from the input vector.
ties	Whether to return all tied best matches or just the first one.
details	Whether to return all candidate-year scores as well as the best match.
data	A data frame containing crosswalk columns or a user data frame.
column	Name of the input code column to convert.
year, from_year, to_year	GB classification years. In <code>convert_gb_codes()</code> , <code>to_year</code> is required and <code>from_year</code> can be NULL to infer it from the input vector.
s_col, from_col, to_col, old_col, new_col	Column names for source and target codes.
to_all_col, parent_col	Names for collapsed target and dominant parent columns.
parent_digits, digits	Number of prefix digits used for parent codes.
threshold	Share required for a dominant parent prefix.
special_cutoff, special_suffix	Optional fallback rule for service-sector-style parent codes.
levels	Levels to include when standardizing a pairwise crosswalk.
source_pair	Optional source label.
pairs	A standard pair table or list of pair tables. Defaults to the package-shipped merged CSV in <code>convert_gb_column()</code> .
file	Package-shipped filename under <code>extdata</code> or <code>raw-crosswalks</code> .
dir	Directory where CSV files should be written.
prefix	Filename prefix for exported CSV files.
include_combined	Whether to also write one combined all-pairs CSV.
output_col	Name of the appended converted-code column.
years	Ordered vector of supported years.
crosswalk	A composed crosswalk with <code>from_code</code> and <code>to_code</code> .
unmatched	Value returned for unmatched input codes.
path	Path to a CSV, XLS, or XLSX crosswalk file, or a directory of CSV files.
sheet	Excel sheet index or name.

Value

Either a character vector, a cleaned data frame, a standard pairwise crosswalk, or a composed crosswalk depending on the function.

Examples

```
normalize_gb_code(c("123", "0123"))
codes_to_gb_level("0123", "M")
convert_gb_codes(c("572", "593", "843"), from_year = 2017, to_year = 2011, level = "M")
convert_gb_codes(c("572", "593", "843"), to_year = 2011)
```

Index

`add_collapsed_targets`
 (gbcrosswalk-package), 2

`as_gb_pair` (gbcrosswalk-package), 2

`build_gb_1986_1994`
 (gbcrosswalk-package), 2

`build_gb_1994_2002`
 (gbcrosswalk-package), 2

`build_gb_2002_2011`
 (gbcrosswalk-package), 2

`build_gb_2011_2017`
 (gbcrosswalk-package), 2

`codes_to_gb_level`
 (gbcrosswalk-package), 2

`compose_gb_crosswalk`
 (gbcrosswalk-package), 2

`convert_gb_codes` (gbcrosswalk-package),
2

`convert_gb_column`
 (gbcrosswalk-package), 2

`crosswalk_codes` (gbcrosswalk-package), 2

`derive_gb_levels` (gbcrosswalk-package),
2

`detect_gb_year` (gbcrosswalk-package), 2

`dominant_prefix` (gbcrosswalk-package), 2

`fill_down_missing`
 (gbcrosswalk-package), 2

`gb_crosswalk_path`
 (gbcrosswalk-package), 2

`gb_raw_crosswalk_path`
 (gbcrosswalk-package), 2

`gbcrosswalk` (gbcrosswalk-package), 2

`gbcrosswalk-package`, 2

`load_gb_crosswalks`
 (gbcrosswalk-package), 2

`normalize_gb_code`
 (gbcrosswalk-package), 2

`normalize_gb_for_level`
 (gbcrosswalk-package), 2

`read_gb_crosswalk_csvs`
 (gbcrosswalk-package), 2

`standardize_gb_pairs`
 (gbcrosswalk-package), 2

`write_gb_crosswalk_csvs`
 (gbcrosswalk-package), 2