

Package: gbcrosswalk (via r-universe)

July 12, 2026

Title Crosswalk Chinese GB Industry Classifications Across Years

Version 0.0.9

Description Clean, build, and compose crosswalks between Chinese GB/T (Guobiao/Tuijian, the recommended national standard prefix) 4754 industry classification vintages. The package starts with the historical workflow for 1986, 1994, 2002, 2011, and 2017 and is designed so later source-document vintages can be added as additional adjacent pairwise crosswalks. It also includes GB/T 4754-2002 to International Standard Industrial Classification of All Economic Activities (ISIC) Revision 3 links and Harmonized System (HS) Combined product code to ISIC Revision 3 links.

License MIT + file LICENSE

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

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

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Depends R (>= 4.1)

Imports utils

Enhances readxl

Repository <https://robin621.r-universe.dev>

Date/Publication 2026-07-12 11:13:24 UTC

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

RemoteRef HEAD

RemoteSha 58869bfdb4c834f9c928b3df2a05c72e34131c31

Contents

gbcrosswalk-package	2
Index	6

Description

Utilities for cleaning, building, and composing Chinese GB/T (Guobiao/Tuijian, the recommended national standard prefix) 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"))
normalize_isic_code(x, level = c("S", "M", "L"))
normalize_hs_level(level = c("HS6", "HS4", "HS2"))
normalize_hs_code(x, level = c("HS6", "HS4", "HS2"))
detect_hs_level(codes)
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_)
convert_gb_to_isic(codes, gb_year = NULL, gb_level = NULL,
  isic_level = c("S", "M", "L"), isic_revision = "3",
  crosswalk = load_gb_isic_crosswalk(),
  pairs = load_gb_crosswalks(),
```

```

        years = c(1986, 1994, 2002, 2011, 2017),
        unmatched = NA_character_)
convert_gb_to_hs(codes, gb_year = NULL, gb_level = NULL,
  hs_level = c("HS6", "HS4", "HS2"),
  hs_system = "HS Combined", isic_revision = "3",
  isic_level = c("S", "M", "L"),
  gb_isic = load_gb_isic_crosswalk(),
  hs_isic = load_hs_isic_crosswalk(),
  pairs = load_gb_crosswalks(),
  years = c(1986, 1994, 2002, 2011, 2017),
  unmatched = NA_character_)
convert_hs_to_gb(codes, hs_level = NULL, gb_year = 2002,
  gb_level = c("M", "S", "L"),
  hs_system = "HS Combined", isic_revision = "3",
  isic_level = c("S", "M", "L"),
  hs_isic = load_hs_isic_crosswalk(),
  gb_isic = load_gb_isic_crosswalk(),
  pairs = load_gb_crosswalks(),
  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)
read_gb_isic_crosswalk(path)
read_hs_isic_crosswalk(path)
gb_crosswalk_path(file = "gb_all_pairs.csv")
gb_isic_crosswalk_path(file = "gb_2002_isic3.csv")
hs_isic_crosswalk_path(file = "hs_combined_isic3.csv")
gb_raw_crosswalk_path(file = NULL)
load_gb_crosswalks(path = gb_crosswalk_path())
load_gb_isic_crosswalk(path = gb_isic_crosswalk_path())
load_hs_isic_crosswalk(path = hs_isic_crosswalk_path())
build_gb_1986_1994(path, sheet = 1)
build_gb_1994_2002(path)
build_gb_2002_2011(path, sheet = 1)
build_gb_2002_isic(path)
build_hs_combined_isic3(path)
build_gb_2011_2017(path, sheet = 1, old_col = NULL, new_col = NULL)

```

Arguments

<code>x, code, codes</code>	Character or numeric GB, ISIC, or HS codes, depending on the function.
<code>width</code>	Target code width for left-zero padding.
<code>level, gb_level, isic_level</code>	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> and <code>convert_gb_to_isic()</code> to infer the GB level from the input vector.
<code>hs_level</code>	HS output/input level. One of "HS6", "HS4", or "HS2". In <code>convert_hs_to_gb()</code> ,

	NULL infers the level from code length.
hs_system	HS classification system to use. The bundled HS file is specifically "HS Combined", from the source column HS - Combined Product Code; it is not a single HS1996, HS2002, HS2007, HS2012, or HS2017 vintage.
isic_revision	ISIC revision to use. The bundled GB-ISIC and HS-ISIC links are specifically for ISIC Revision "3".
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_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. In <code>convert_gb_to_isic()</code> and <code>convert_gb_to_hs()</code> , <code>gb_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> and is used to route non-2002 GB vintages through GB2002 in ISIC/HS helpers.
crosswalk, gb_isic, hs_isic	GB-ISIC or HS-ISIC crosswalk tables. The defaults use package-shipped CSVs.
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.
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)
convert_gb_to_isic(c("0111", "0112"), gb_year = 2002, gb_level = "S")
convert_gb_to_hs("152", gb_year = 2002, gb_level = "M", hs_level = "HS2", isic_level = "M")
convert_hs_to_gb(c("010110", "010111"), gb_year = 2002, gb_level = "M")
```

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_2002_isic
 (gbcrosswalk-package), 2
build_gb_2011_2017
 (gbcrosswalk-package), 2
build_hs_combined_isic3
 (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
convert_gb_to_hs (gbcrosswalk-package), 2
convert_gb_to_isic
 (gbcrosswalk-package), 2
convert_hs_to_gb (gbcrosswalk-package), 2
crosswalk_codes (gbcrosswalk-package), 2
derive_gb_levels (gbcrosswalk-package), 2
detect_gb_year (gbcrosswalk-package), 2
detect_hs_level (gbcrosswalk-package), 2
dominant_prefix (gbcrosswalk-package), 2
fill_down_missing
 (gbcrosswalk-package), 2
gb_crosswalk_path
 (gbcrosswalk-package), 2
gb_isic_crosswalk_path
 (gbcrosswalk-package), 2
gb_raw_crosswalk_path
 (gbcrosswalk-package), 2
gbcrosswalk (gbcrosswalk-package), 2
gbcrosswalk-package, 2
hs_isic_crosswalk_path
 (gbcrosswalk-package), 2
load_gb_crosswalks
 (gbcrosswalk-package), 2
load_gb_isic_crosswalk
 (gbcrosswalk-package), 2
load_hs_isic_crosswalk
 (gbcrosswalk-package), 2
normalize_gb_code
 (gbcrosswalk-package), 2
normalize_gb_for_level
 (gbcrosswalk-package), 2
normalize_hs_code
 (gbcrosswalk-package), 2
normalize_hs_level
 (gbcrosswalk-package), 2
normalize_isic_code
 (gbcrosswalk-package), 2
read_gb_crosswalk_csvs
 (gbcrosswalk-package), 2
read_gb_isic_crosswalk
 (gbcrosswalk-package), 2
read_hs_isic_crosswalk
 (gbcrosswalk-package), 2
standardize_gb_pairs
 (gbcrosswalk-package), 2
write_gb_crosswalk_csvs
 (gbcrosswalk-package), 2