

Package ‘regioncode’

May 9, 2026

Title Convert Region Names and Division Codes of China Over Years

Version 0.2.1

Author Yue Hu [aut, cre],
Xinyi Ye [aut],
Qian Qiu [aut],
Yufei Sun [aut],
Wenquan Wu [aut]

Maintainer Yue Hu <yuehu@tsinghua.edu.cn>

Description A tool to conquer the difficulties to convert various region names and administration division codes of Chinese regions. The current version enables seamlessly converting Chinese regions' formal names, common-used names, and codes between each other at the city level from 1986 to 2019.

License MIT + file LICENSE

Encoding UTF-8

BugReports <https://github.com/sammo3182/regioncode/issues>

Depends R(>= 3.6.0)

Suggests dplyr, knitr, rmarkdown

LazyData true

RoxygenNote 7.3.3

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2026-04-07 12:10:02 UTC

Contents

corruption	2
regioncode	2
Index	5

 corruption

China's Corruption Investigations Dataset

Description

A dataset containing information on almost 20,000 officials who were investigated during Xi Jinping's anti-corruption campaign.

Usage

corruption

Format

A data frame with 6 variables:

province 2-digit province number

prefecture Prefecture name in Chinese

county County name in Chinese

province_id 6-digit province number

prefecture_id 6-digit province number

county_id 6-digit province number

Source

[doi:10.7910/DVN/9QZRAD](https://doi.org/10.7910/DVN/9QZRAD)

 regioncode

regioncode is developed to conquer the difficulties to convert various region names and administration division codes of Chinese regions. In the current version, regioncode enables seamlessly converting Chinese regions' formal names, common-used names, and geocodes between each other at the prefectural level from 1986 to 2019.

Description

regioncode is developed to conquer the difficulties to convert various region names and administration division codes of Chinese regions. In the current version, regioncode enables seamlessly converting Chinese regions' formal names, common-used names, and geocodes between each other at the prefectural level from 1986 to 2019.

Usage

```

regioncode(
  data_input,
  year_from = 1999,
  year_to = 2015,
  convert_to = "code",
  incomplete_name = "none",
  zhixiashi = FALSE,
  to_dialect = "none",
  to_pinyin = FALSE,
  province = FALSE
)

```

Arguments

data_input	A character vector for names or a six-digit integer vector for division codes to convert.
year_from	A integer to define the year of the input. The default value is 1999.
year_to	A integer to define the year to convert. The default value is 2015.
convert_to	A character indicating the converting methods. At the prefectural level, valid methods include converting between codes in different years, from codes to city ranks, from codes to region names, from region names to city ranks, from region names to division codes, from region names or division codes to sociopolitical area names, and between names in different years. The current version automatically detect the type of the input. Users only need to choose the output to be codes (code), names (name) , area (area) or city ranks (rank). The default option is code. When province is TRUE, one can also choose abbre, abbreToCode, abbreToName, and abbreToArea to convert between names/codes and abbreviations of provinces.
incomplete_name	A character string indicating how incomplete (short) names are handled. One of "none" (default), "from", "to", or "both". See "Details".
zhixiashi	A logic string to indicate whether treating division codes and names of municipality directly under the central government (Only makes a difference for prefectural-level conversion). The default value is FALSE.
to_dialect	A character indicating the language transformation. At the prefectural level, valid transformation include dia_group, dia_sub_group. At the province level, valid transformation is dia_super. The default value is "none". When province is TRUE, one can also choose dia_super to get the language zone of provinces.
to_pinyin	A logic string to indicate whether the output is in pinyin spelling instead of Chinese characters. The default is FALSE.
province	A logic string to indicate the level of converting. The default is FALSE.

Details

In many national and regional data in China studies, the source applies incomplete names instead of the official, full name of a given region. A typical case is that "Xinjiang" is used much more often

than "Xinjiang Weiwuer Zizhiqu" (the Xinjiang Uygur Autonomous Region) for the name of the province. In other cases the "Shi" (City) is often omitted to refer to a prefectural city. `regioncode` accounts this issue by offering the argument `incomplete_name`.

- "none": no short name will be used for either input or output;
- "from": input data is short names instead of the full, official ones;
- "to": output results will be short names;
- "both": both input and output are using short names.

The argument makes a difference only when code or name are chose in `convert_to`. Users can use this argument together with `name` to convert between names and incomplete names.

Value

The function returns a character or numeric vector depending on what method is specified.

Examples

```
## Not run:  
# The example can be run well but CRAN does not like Chinese characters, so here just "dontrun" it.  
  
library(regioncode)  
  
regioncode(  
  data_input = corruption$prefecture_id,  
  year_from = 2016,  
  year_to = 2017  
)  
  
## End(Not run)
```

Index

* **datasets**

corruption, [2](#)

corruption, [2](#)

regioncode, [2](#)