

# Package ‘ChineseNames’

June 21, 2021

**Title** Chinese Name Database 1930-2008

**Version** 1.1.0

**Date** 2021-06-21

**Author** Han-Wu-Shuang Bao [aut, cre]

**Maintainer** Han-Wu-Shuang Bao <baohws@foxmail.com>

**Description** A database of Chinese surnames and Chinese given names (1930-2008).

This database contains nationwide frequency statistics of 1,806 Chinese surnames and 2,614 Chinese characters used in given names, covering about 1.2 billion Han Chinese population (96.8% of the Han Chinese household-registered population born from 1930 to 2008 and still alive in 2008).

This package also contains a function for computing multiple features of Chinese surnames and Chinese given names for scientific research (e.g., name uniqueness, name gender, name valence, and name warmth/competence).

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**URL** <https://github.com/psychbruce/ChineseNames>

**BugReports** <https://github.com/psychbruce/ChineseNames/issues>

**Depends** R (>= 3.6.0)

**Imports** bruceR, data.table

**Suggests** babynames, dplyr

**RoxygenNote** 7.1.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-06-21 16:40:02 UTC

## R topics documented:

ChineseNames . . . . .	2
compute_name_index . . . . .	3
familyname . . . . .	5
givenname . . . . .	5
population . . . . .	6
top1000name.prov . . . . .	7
top100name.year . . . . .	7
top50char.year . . . . .	7
<b>Index</b>	<b>8</b>

---

ChineseNames	<i>ChineseNames: Chinese Name Database 1930-2008</i>
--------------	--

---

## Description

A database of Chinese surnames and Chinese given names (1930-2008). This database contains nationwide frequency statistics of 1,806 Chinese surnames and 2,614 Chinese characters used in given names, covering about 1.2 billion Han Chinese population (96.8% of the Han Chinese household-registered population born from 1930 to 2008 and still alive in 2008). This package also contains a function for computing multiple features of Chinese surnames and Chinese given names for scientific research (e.g., name uniqueness, name gender, name valence, and name warmth/competence).

## Details

Details are described in <https://github.com/psychbruce/ChineseNames>

## Citation

Bao, H.-W.-S. (2021). ChineseNames: Chinese Name Database 1930-2008. R package version 1.1.0. <https://CRAN.R-project.org/package=ChineseNames>

## Note

This database does not contain any individual-level information (so it does not leak personal privacy). All data are at the name level or character level. Extremely rare characters are not included.

## Source

This database was provided by Beijing Meiming Science and Technology Company (in collaboration) and originally obtained from the National Citizen Identity Information Center (NCIIC) of China in 2008.

---

compute\_name\_index      *Compute multiple features of surnames and given names.*

---

## Description

Compute all available name features (indices) based on `familyname` and `givenname`. You can either input a data frame with a variable of Chinese full names (and a variable of birth years, if necessary) or just input a vector of full names (and a vector of birth years, if necessary).

- Usage 1: Input a single value or a vector of name [and birth, if necessary].
- Usage 2: Input a data frame of data and the variable name of `var.fullname` (or `var.surname` and/or `var.givenname`) [and `var.birthyear`, if necessary].

*Caution.* Name-character uniqueness (NU) of cases with birth year  $\geq 2010$  would not be accurately computed due to the year limitation of this name database.

## Usage

```
compute_name_index(
  data = NULL,
  var.fullname = NULL,
  var.surname = NULL,
  var.givenname = NULL,
  var.birthyear = NULL,
  name = NA,
  birth = NA,
  index = c("NLen", "SNU", "SNI", "NU", "CCU", "NG", "NV", "NW", "NC"),
  NU.approx = TRUE,
  digits = 4,
  return.namechar = TRUE,
  return.all = FALSE
)
```

## Arguments

<code>data</code>	Data frame.
<code>var.fullname</code>	Variable name of Chinese full names (e.g., "name").
<code>var.surname</code>	Variable name of Chinese surnames (e.g., "surname").
<code>var.givenname</code>	Variable name of Chinese given names (e.g., "givenname").
<code>var.birthyear</code>	Variable name of birth year (e.g., "birth").
<code>name</code>	<b>If no</b> data, you can just input a vector of full name(s).
<code>birth</code>	<b>If no</b> data, you can just input a vector of birth year(s).
<code>index</code>	Which indices to compute? By default, it computes all available name indices:

- NLen: full-name length (2~4).
- SNU: surname uniqueness (1~6).
- SNI: surname initial (1~26).
- NU: name-character uniqueness (1~6).
- CCU: character-corpus uniqueness (1~6).
- NG: name gender (-1~1).
- NV: name valence (1~5).
- NW: name warmth (1~5).
- NC: name competence (1~5).

For details, see <https://github.com/psychbruce/ChineseNames>

NU.approx	Whether to <i>approximately</i> compute name-character uniqueness (NU) using <i>the nearest two birth cohorts with relative weights</i> (which would be more precise than just using a single birth cohort). Default is TRUE.
digits	Number of decimal places. Default is 4.
return.namechar	Whether to return separate name characters. Default is TRUE.
return.all	Whether to return all temporary variables in the computation of the final variables. Default is FALSE.

### Value

A new data frame (`data.table`) with name indices appended. Full names are split into `name0` (surnames, with compound surnames automatically detected), `name1`, `name2`, and `name3` (given-name characters).

### Note

For details and examples, see <https://github.com/psychbruce/ChineseNames>

### Examples

```
## Prepare
sn=familyname$surname[1:12]
gn=c(top100name.year$name.all.1960[1:6],
      top100name.year$name.all.2000[1:6],
      top100name.year$name.all.1960[95:100],
      top100name.year$name.all.2000[95:100])
demodata=data.frame(name=paste0(sn, gn),
                    birth=c(1960:1965, 2000:2005,
                           1960:1965, 2000:2005))

## Compute
newdata=compute_name_index(demodata,
                           var.fullname="name",
                           var.birthyear="birth")

# use View(newdata) to see the results
```

---

familyname	<i>1,806 Chinese surnames and nationwide frequency.</i>
------------	---

---

**Description**

1,806 Chinese surnames and nationwide frequency.

**Usage**

```
data(familyname)
```

**Format**

A data frame with 7 variables:

surname surname (in Chinese)

compound 0 = single surname, 1 = compound surname

initial initial letter (a-z)

initial.rank initial order (1-26)

n.1930\_2008 total counts in the database

ppm.1930\_2008 proportion in population (ppm = parts per million)

surname.uniqueness surname uniqueness

**Details**

<https://github.com/psychbruce/ChineseNames>

---

givenname	<i>2,614 Chinese characters used in given names and nationwide frequency.</i>
-----------	---

---

**Description**

2,614 Chinese characters used in given names and nationwide frequency.

**Usage**

```
data(givenname)
```

**Format**

A data frame with 25 variables:

character character used in given names (in Chinese)

pinyin pinyin (pronunciation)

bihua number of strokes in a character

n.male total counts in male

n.female total counts in female

name.gender difference in proportions of a character used by male vs. female

n.1930\_1959, n.1960\_1969, n.1970\_1979, n.1980\_1989, n.1990\_1999, n.2000\_2008 total counts in a birth cohort

ppm.1930\_1959, ppm.1960\_1969, ppm.1970\_1979, ppm.1980\_1989, ppm.1990\_1999, ppm.2000\_2008 proportion (parts per million) in a birth cohort

name.ppm average ppm (parts per million) across all cohorts

name.uniqueness name-character uniqueness (in naming practices)

corpus.ppm proportion (parts per million) in contemporary Chinese corpus

corpus.uniqueness character-corpus uniqueness (in contemporary Chinese corpus)

name.valence name valence (positivity of character meaning) (based on subjective ratings from 16 raters, ICC = 0.921)

name.warmth name warmth/morality (based on subjective ratings from 10 raters, ICC = 0.774)

name.competence name competence/assertiveness (based on subjective ratings from 10 raters, ICC = 0.712)

**Details**

<https://github.com/psychbruce/ChineseNames>

---

population

*Population statistics for the Chinese name database.*

---

**Description**

Population statistics for the Chinese name database.

**Usage**

data(population)

**Details**

<https://github.com/psychbruce/ChineseNames>

---

top1000name.prov      *Top 1,000 given names in 31 Chinese mainland provinces.*

---

**Description**

Top 1,000 given names in 31 Chinese mainland provinces.

**Usage**

data(top1000name.prov)

**Details**

<https://github.com/psychbruce/ChineseNames>

---

top100name.year      *Top 100 given names in 6 birth cohorts.*

---

**Description**

Top 100 given names in 6 birth cohorts.

**Usage**

data(top100name.year)

**Details**

<https://github.com/psychbruce/ChineseNames>

---

top50char.year      *Top 50 given-name characters in 6 birth cohorts.*

---

**Description**

Top 50 given-name characters in 6 birth cohorts.

**Usage**

data(top50char.year)

**Details**

<https://github.com/psychbruce/ChineseNames>

# Index

ChineseNames, [2](#)  
compute\_name\_index, [3](#)  
  
familyname, [3, 5](#)  
  
givenname, [3, 5](#)  
  
population, [6](#)  
  
top1000name.prov, [7](#)  
top100name.year, [7](#)  
top50char.year, [7](#)