Package 'genderBR'

April 17, 2018

Type Package

Title Predict Gender from Brazilian First Names

Version 1.1.0

Description A method to predict and report gender from Brazilian first names
using the Brazilian Institute of Geography and Statistics' Census data.

License GPL (>= 2)

Depends R (>= 3.1.2)

Imports dplyr (>= 0.5.0), jsonlite, httr

Encoding UTF-8

LazyData true

URL https://github.com/meirelesff/genderBR

BugReports https://github.com/meirelesff/genderBR/issues

RoxygenNote 6.0.1

NeedsCompilation no

Author Fernando Meireles [aut, cre]

Maintainer Fernando Meireles <fmeireles@ufmg.br>

Repository CRAN

Date/Publication 2018-04-17 19:25:05 UTC

R topics documented:

---

get_gender   *Predict gender from Brazilian first names*

Description

`get_gender` uses the IBGE's 2010 Census data to predict gender from Brazilian first names. More specifically, it uses data on the number of females and males with the same name in Brazil, or in a given Brazilian state, and calculates the proportion of females using it. The function classifies a name as male or female only when that proportion is higher than a given threshold (e.g., `female if proportion > 0.9, the default`, or `male if proportion < 0.1`); proportions below this threshold are classified as missing (`NA`). This method is based on the 'gender' package developed by Muellen (2016). gender: Predict Gender from Names Using Historical Data.

Multiple names can be passed to the function call. To speed the calculation process, the package aggregates equal first names to make fewer requests to the IBGE's API. Also, the package contains an internal dataset with all the names reported by the IBGE to make faster classifications – although this option does not support getting results by State.

Usage

```
get_gender(names, state = NULL, prob = FALSE, threshold = 0.9,
  internal = TRUE)
```

Arguments

names	A string specifying a person's first name. Names can also be passed to the function as a full name (e.g., Ana Maria de Souza). `get_gender` is case insensitive.
state	A string with the state of federation abbreviation (e.g., `RJ` for Rio de Janeiro). If state is set to a value different from `NULL`, the `internal` argument is ignored.
prob	Report the proportion of female uses of the name? Defaults to `FALSE`.
threshold	Numeric indicating the threshold used in predictions. Defaults to 0.9.
internal	Use internal data to predict gender? Allowing this option makes the function way faster, but it does not support getting results by State. Defaults to `TRUE`.

Details

Information on the Brazilian first names uses by gender was collect in the 2010 Census (Censo Demografico de 2010, in Portuguese), in July of that year, by the Instituto Brasileiro de Demografia e Estatistica (IBGE). The surveyed population includes 190,8 million Brazilians living in all 27 states. According to the IBGE, there are more than 130,000 unique first names in this population.

Value

`get_gender` may returns three different values: `Female`, if the name provided is female; `Male`, if the name provided is male; or `NA`, if we can not predict gender from the name given the chosen threshold.

If the `prob` argument is set to `TRUE`, then the function returns the proportion of females uses of the provided name.

Note

Names with different spell (e.g., Ana and Anna, or Marcos and Markos) are considered different names. Additionally, only names with more than 20 occurrences, or more than 15 occurrences in a given state, are included in the IBGE's data.

References

For more information on the IBGE's data, please check (in Portuguese): http://censo2010.ibge.gov.br/nomes/

See Also

map_gender

Examples

```
# Use get_gender to predict the gender of a person based on her/his first name
get_gender('MARIA DA SILVA SANTOS')
get_gender('joao')

## Not run:
# It is possible to filter results by state
get_gender('ana', state = 'sp')

# To change the employed threshold
get_gender('ariel', threshold = '0.8')

# Or to get the proportion of females
# with the name provided
get_gender('iris', prob = TRUE)

# Multiple names can be predict at the same time
get_gender(c('joao', 'ana', 'benedita', 'rafael'))

# In different states
get(rep('cris', 3), c('sp', 'am', 'rs'))

## End(Not run)
```

---

get_states	*State's abbreviations*

Description

Use this function to get a data.frame with the full names, abbreviations (acronym), and IBGE codes of all Brazilian states.

Usage

```
get_states()
```

Value

A `tbl_df, tbl, data.frame` with two variables: `state, abb`, and `code`.

---

map_gender   *Map the use of Brazilian first names by gender and by state*

Description

`map_gender` retrieves data on the number of male or female uses of a given first name by state from the Instituto Brasileiro de Geografia e Estatistica's 2010 Census API.

Usage

```
map_gender(name, gender = NULL)
```

Arguments

name   A string with a Brazilian first name. The name can also be passed to the function
              as a full name (e.g., Ana Maria de Souza). `get_gender` is case insensitive.

gender   A string with the gender to look for. Valid inputs are `m`, for males, `f`, for females,
              and `NULL`, in which case the function returns results for all persons with a given
              name.

Details

Information on the gender associated with Brazilian first names was collect in the 2010 Census (Censo Demografico de 2010, in Portuguese), in July of that year, by the Instituto Brasileiro de Demografia e Estatistica (IBGE). The surveyed population includes 190,8 million Brazilians living in all 27 states. According to the IBGE, there are more than 130,000 unique first names in this population.

Value

`get_gender` returns a `tbl_df, tbl, data.frame` with the following variables:

- `nome` State's name.
- `uf` State's abbreviation.
- `freq` Total number of persons with the name provided.
- `populacao` State's total population.
- `sexo` Same as the `sexo` argument provided.
- `prop` Persons with the name and gender provided per 100,000 inhabitants.

Note

Names with different spell (e.g., Ana and Anna, or Marcos and Markos) are considered different names. Additionally, only names with more than 20 occurrences, or more than 15 occurrences in a given state, are considered.

References

For more information on the IBGE's data, please check (in Portuguese): http://censo2010.ibge.gov.br/nomes/

See Also

get_gender

Examples

```
# Map the use of the name 'Joao' for
# males in Brazil by state
map_gender('Joao', gender = 'm')


# Map the use of the name 'Maria'
map_gender('maria')

# The function accepts full names
map_gender('Maria da Silva Santos')

# Or names in uppercase
map_gender('MARIA DA SILVA SANTOS')
```

Index