Package ‘genero’

October 13, 2022

Title Estimate Gender from Names in Spanish and Portuguese
Version 0.1.0
Maintainer Juan Pablo Marin Diaz <jpmarindiaz@gmail.com>
Description Estimate gender from names in Spanish and Portuguese. Works with vectors and dataframes. The estimation works not only for first names but also full names. The package relies on a compilation of common names with its most frequent associated gender in both languages which are used as look up tables for gender inference.

URL https://github.com/datasketch/genero
License MIT + file LICENSE
Encoding UTF-8
LazyData true

BugReports https://github.com/datasketch/genero/issues
Depends R (>= 3.1)
Imports
Suggests testthat (>= 2.1.0), knitr, rmarkdown
VignetteBuilder knitr
RoxygenNote 7.0.2
NeedsCompilation no
Author Juan Pablo Marin Diaz [aut, cre]
Repository CRAN
Date/Publication 2020-03-09 10:10:02 UTC

R topics documented:

genero ................................................................. 2
names_gender_es .................................................. 3
names_gender_pt .................................................. 3

Index 4
Panel component for shiny panels layout

Usage

genero(
  names,
  result_as = c(male = "male", female = "female"),
  lang = "es",
  col = NULL,
  na = NA,
  rev_weights = FALSE
)

Arguments

names A vector or data.frame with names or full names
result_as A named vector with names c("male", "female") values can be used to override the results
lang Use "es" for Spanish (default), "pt" for Portuguese.
col The name of the column with the names or full names when the input is a data frame.
na String to be used when there is not match for gender
rev_weights Boolean to indicate if weights should be reversed when input names have the format Last Name First Name

Value

A vector of data frame with the estimated gender for the input. When the input is data.frame a column is attached next to the column used for the input names with the result.

Examples

genero(c("Juan", "Pablo", "Camila", "Mariana"))
names_gender_es

Names with gender in Spanish

Description
These data was collected and organized manually from multiple sources. It consists of more than 9810 names in Spanish and its corresponding associated gender accounting for name variations.

Usage
names_gender_es

Format
Data frame with two columns: name and gender.

Examples
names_gender_es

names_gender_pt

Names with gender in Portuguese

Description
These data is created and derived from https://brasil.io/dataset/genero-nomes/nomes it consists of more than 50,000 names in Portuguese and its corresponding associated gender.

Usage
names_gender_pt

Format
Data frame with two columns: name and gender.

Examples
names_gender_pt
Index

* datasets
  names_gender_es, 3
  names_gender_pt, 3

genero, 2

names_gender_es, 3
names_gender_pt, 3