Package 'randomNames'

March 7, 2019

Version 1.4-0.0

Date 2019-3-7

Title Generate Random Given and Surnames

Depends R (>= 3.3)

Suggests knitr, rmarkdown

Imports crayon, data.table (>= 1.12.0), toOrdinal (>= 1.1)

Maintainer Damian W. Betebenner <dbetebenner@nciea.org>

Description Function for generating random gender and ethnicity correct first and/or last names. Names are chosen proportionally based upon their probability of appearing in a large scale data base of real names.

URL https://CenterForAssessment.github.io/randomNames, https://github.com/CenterForAssessment/randomNames, https://cran.r-project.org/package=randomNames

BugReports https://github.com/CenterForAssessment/randomNames/issues

VignetteBuilder knitr

LazyLoad Yes

LazyData Yes

LazyDataCompression xz

ByteCompile TRUE

License GPL-3

NeedsCompilation no

Author Damian W. Betebenner [aut, cre]

Repository CRAN

Date/Publication 2019-03-07 16:50:03 UTC

R topics documented:

`randomNames-package` *Generate Random Given and Surnames*

Description

Functions for generating random gender and ethnicity correct first and/or last names where names are proportionally sampled based upon their frequency in a large scale database.

Details

Package:	randomNames
Type:	Package
Version:	1.4-0.0
Date:	2019-3-7
License:	GPL-3
LazyLoad:	yes

Author(s)

Damian W. Betebenner <DBetebenner@nciea.org>

`randomNames` *Random Names Function*

Description

Function to generate random gender and ethnicity correct first and/or last names. Names are chosen proportionally based upon their probability of appearing in a large scale database of real names. Function to generate random gender and ethnicity correct first and/or/last names. The function probabilistically samples from the embedded randomNamesData dataset to provide a realistic list of first and/or last names.

Usage

```
randomNames(n,
            gender,
            ethnicity,
            which.names="both",
            name.order="last.first",
            name.sep=", ",
            sample.with.replacement=TRUE,
            return.complete.data=FALSE)
```

Arguments

`n`	OPTIONAL. Positive integer indicating how many name to produce. Best to use when no gender or ethnicity data is provided and one simply wants `n` random first and/or last names.
`gender`	OPTIONAL. A vector indicating the genders for the names to be calculated. The maximum of `n`, the length of `gender` and the length of `ethnicity` is the number of random names returned. Note that the gender vector is only employed for deriving first names. If no gender vector is provided, the function randomly samples from both genders to produce a sample of names. Current gender codes are `0: Male` and `1: Female`. See examples for various use cases.
`ethnicity`	OPTIONAL. A vector indicating the ethnicities for the names to be calculated. The maximum of `n`, the length of `gender` and the length of `ethnicity` is the number of random names returned. Note that the ethnicity vector is employed for both deriving first and last names. If no ethnicity vector is provided the function samples from all ethnicity to produce a sample of names. Current ethnicity codes are: **1** American Indian or Native Alaskan **2** Asian or Pacific Islander **3** Black (not Hispanic) **4** Hispanic **5** White (not Hispanic) **6** Middle-Eastern, Arabic
`which.names`	OPTIONAL. One of `"both"` (the default), `"first"`, or `"last"`, `"complete.data"` indicating what names to produce. `"complete.data"` provides a data.table with `ethnicity` and `gender` codes.
`name.order`	OPTIONAL. If `which.names` is `"both"`, then names can be returned as either `"last.first"` (the default) or `"first.last"`.
`name.sep`	OPTIONAL. If `which.names` is `"both"`, then names are separated by the `name.sep` string provided. Defaults to comma-space separated.
`sample.with.replacement`	Boolean argument (defaults to TRUE) indicating whether sampling is done with replacement.
`return.complete.data`	Boolean argument (defaults to FALSE) indicating whether to return data including gender and ethnicity codes used for name construct. If set to TRUE, data is returned as a data.frame/data.table.

Details

Typical use of the function is to submit a vector of genders and ethnicities to derived a gender and ethnicity representative vector of first and/or last names.

Value

Function returns a character vector containing first and/or last names.

Author(s)

Damian W. Betebenner <dbetebenner@nciea.org>

See Also

randomNamesData

Examples

```
randomNames()  ## Returns a single name in "last, first" format

randomNames(5, which.names="first") ## Returns 5 first names

randomNames(5, return.complete.data=TRUE) ## Returns entire data.table

test.df <- data.frame(GENDER=sample(0:1, 100, replace=TRUE),
ETHNICITY=sample(1:6, 100, replace=TRUE))

test.names <- randomNames(gender=test.df$GENDER,
ethnicity=test.df$ETHNICITY)

head(test.names)


ethnicities <- c("African American", "Hispanic", "Asian", "White", "Native American")
genders <- c("Female", "Male")

test.df <- data.frame(GENDER=sample(genders, 100, replace=TRUE),
ETHNICITY=sample(ethnicities, 100, replace=TRUE))

test.names <- randomNames(gender=test.df$GENDER,
ethnicity=test.df$ETHNICITY)

head(test.names)
```

---

randomNamesData *First names (by gender and ethnicity) and last names (by ethnicity) for randomNames function*

Description

A dataset compiled from a large scale data set of names providing first names by gender and ethnicity and last names by ethnicity. The dataset provides frequency probabilities from the original data set so that sampling provides a reasonable set of random names. The dataset is for use with the randomNames function to quickly generate random names that can be used, for example, to anonymize results.

Format

An environment of first and last names by gender and ethnicity (first names) and ethnicity (last names) Gender is coded 0 for male and 1 for female, ethnicity is coded 1 for American Indian or Native Alaskan, 2 for Asian or Pacific Islander, 3 for Black (not Hispanic), 4 for Hispanic, and 5 for White (not Hispanic). For example, the array `first_names_e1_g0` in `randomNamesData` provides first and associated frequency probabilities for male, American Indians or Native Alaskans. To view names of all environment arrays type `ls(randomNamesData)` after loading the package.

Source

Large scale state data

See Also

`randomNames`

Index