

Package ‘learningr’

October 2, 2019

Type Package

Title Data and Functions to Accompany the Book “Learning R”

Version 0.29.1

Date 2013-10-20

Author Richie Cotton

Maintainer Richie Cotton <richierocks@gmail.com>

Description Crabs in the English channel, deer skulls, English monarchs, half-caste Manga characters, Jamaican cities, Shakespeare's The Tempest, drugged up cyclists and sexually transmitted diseases.

Imports plyr

License Unlimited

LazyData TRUE

Collate 'learningr-package.R' 'buggy_count.R' 'hypotenuse.R'
'fix_filenames.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2019-10-02 15:27:55 UTC

Depends R (>= 2.10)

R topics documented:

alpe_d_huez	2
buggy_count	2
crab_tag	3
deer_endocranial_volume	5
english_monarchs	6
fix_filenames	7
gonorrhoea	7
hafu	8
hypotenuse	9
learningr	10
obama_vs_mccain	10

Index**12**

alpe_d_huez	<i>Alpe d'Huez</i>
-------------	--------------------

Description

Fastest times for the Alpe d'Huez stage of the Tour de France cycle race, plus some contextual information on year and drug use allegations.

Format

A data frame with the following columns.

Time Character time of ride in the form M' S".

NumericTime Numeric time of ride in minutes.

Name Name of rider.

Year Year of race.

Nationality Nationality of rider.

DrugUse Have allegations of drug use been made against the rider? In `alpe_d_huez` the values are "Y" or "N"; in `alpe_d_huez2` this is a logical vector.

Note

The data was kindly compiled by William Hogan.

The dataset is not guaranteed to be error free. Please double check the data if you want to use it for something serious.

References

William Hogan <william.hogan@amd.com> compiled the data from http://en.wikipedia.org/wiki/Alpe_d%27Huez. Richard Cotton <richierocks@gmail.com> made some modifications while importing it into R.

buggy_count	<i>A buggy version of plyr's count function</i>
-------------	---

Description

An old version of the `plyr` package's count function that fails when you pass it a factor input.

Usage

```
buggy_count(df, vars = NULL, wt_var = NULL)
```

Arguments

`df` A data frame or an atomic input.
`vars` Variables in `df` to count unique values of
`wt_var` Optional variable to weight by. See [count](#).

Value

A data frame with label and freq columns.

Note

In case the “buggy” part of the name didn’t give it away, this is not suitable for real world usage!

See Also

[count](#) and [table](#)

Examples

```
## Not run:
buggy_count(factor()) #oops!

## End(Not run)
```

crab_tag	<i>Crab tag</i>
----------	-----------------

Description

Depth and temperature data from a sensor tag attached to an edible crab (*Cancer Pangurus*) in the North Sea in 2008 and 2009.

Format

A list with 5 elements, as follows. `id_block` is a list with 2 elements

Firmware Version No Version of the firmware used in the crab tag.

Firmware Build Level Build level of the firmware used in the crab tag.

`tag_notebook` is a list with 5 elements.

Mission Day Number of days of data.

Last Deployment Date Date and time that the tag was released into the sea.

Deployed by Host Version UNKNOWN

Downloaded by Host Version UNKNOWN

Last Clock Set Date UNKNOWN

lifetime_notebook is a list with 3 elements.

Tag ID The unique ID of the tag.

Pressure Range UNKNOWN

No of sensors The number of sensors on the tag.

deployment_notebook is a data frame with X columns.

Start Start date and time of recording.

Stop Stop date and time of recording.

Logging Rate UNKNOWN

Pointer UNKNOWN

PA inc UNKNOWN

sensors UNKNOWN

Flags UNKNOWN

Resolution UNKNOWN

Fast Rate UNKNOWN

V1 UNKNOWN

V2 UNKNOWN

daylog is a data frame with X columns.

Mission.Day Integer number of days since the start of recording.

Date Date of record.

Max.Temp Maximum temperature (Celcius) recorded that day.

Min.Temp Minimum temperature (Celcius) recorded that day.

Max.Depth Maximum depth (m) recorded that day.

Min.Depth Minimum depth (m) recorded that day.

Batt.Volts Voltage of tag battery.

Note

This data was kindly supplied by Ewan Hunter of CEFAS. It is part of a larger dataset consisting of many crabs.

References

Ewan Hunter <ewan.hunter@cefas.gov.uk> ran the project where the data was collected. A full analysis is in Hunter, E, Eaton, D, Stewart, C, Lawler, A & Smith, M. 2013. Edible crabs "go west": migrations and incubation cycle revealed by electronic tags. <https://www.ncbi.nlm.nih.gov/pubmed/23734180>. Richard Cotton <richierocks@gmail.com> made some modifications while importing it into R.

deer_endocranial_volume

Deer Endocranial Volume

Description

The dataset contains the endocranial volume of 33 red deer (*Cervus elaphus*), using four different methods of measurement: cardiac tomography, filling the skull with glass beads (yes, the skulls are from dead deer), simply measuring the length, width and height and multiplying the numbers, and using Finarelli's equation. "Endocranial volume" is a proxy for brain size.

Format

A data frame with the following columns.

SkullID A unique identifier for each red deer.

VolCT Endocranial volume calculated by cardiac tomography.

VolBead Endocranial volume calculated by glass beads.

VolLWH Endocranial volume calculated by length*width*height.

VolFinarelli Endocranial volume calculated by Finarelli's equation.

VolCT2 A second measurement via cardiac tomography.

VolBead2 A second measurement via glass beads.

VolLWH2 A second measurement via l*w*h.

Note

The data was kindly provided by Corina Logan. Second measurements are provided for several of the deer. Finarelli's equation is used for estimating the brain volume of non-bovid ruminant Artiodactylid species (say that 10 times fast).

$$\ln(\text{volume}(\text{skull})) = 2.6616 * \ln(\text{width}(\text{skull})) - 6.2722$$

References

The dataset was collected by Corina Logan <itsme@corinalogan.com>. It is stored in the Dryad Digital Repository, doi: [10.5061/dryad.4t7h2](https://doi.org/10.5061/dryad.4t7h2). A more full analysis is given in the paper Logan CJ, Clutton-Brock TH. 2013. Validating methods for estimating endocranial volume in individual red deer (*Cervus elaphus*). Behavioural Processes 92:143-146. doi:10.1016/j.beproc.2012.10.015. <http://www.sciencedirect.com/science/article/pii/S037663571200232X> Richard Cotton <richierocks@gmail.com> made some modifications while importing it into R.

english_monarchs	<i>English Monarchs</i>
------------------	-------------------------

Description

Names, dates and houses of English kings and queens from post-Roman rule (the fifth century) until England invaded Ireland in the Early 13th century.

Format

A data frame with the following columns.

name Name of monarch(s).

house Royal house of monarch(s).

start.of.reign Year they rose to power.

end.of.reign Year they left power.

domain Region of England ruled over.

Note

This dataset is a bit messy and ambiguous in places, because history is like that. In fact, the messy parts of the dataset are in general a good indicator that something interesting was happening at the time. (See, for example, missing or multiple rulers, starts and ends of reigns in the same year, and rulers that appear several times with different territories.) Even defining a monarch of England is tricky. Most of the monarchs in this dataset were around before England existed (it consisted of seven territories called the heptarchy). The data stops before John I (the bad guy from the Robin Hood stories) because he proclaimed himself King of Ireland, although some people consider monarchs up to Anne, five hundred years later, to be English Monarchs even though they ruled over Ireland, Wales and Scotland to varying degrees.

The heptarchy consisted of East Anglia, Essex, Kent, Mercia, Northumbria, Sussex and Wessex. Northumbria was originally divided into Deira and Bernicia. There are also periods of Norse and Danish rule.

The dataset was compiled from Wikipedia and thus is not guaranteed to be error free. Please double check the data if you want to use it for something serious.

References

Richard Cotton <richierocks@gmail.com> compiled the dataset from various Wikipedia pages.
http://en.wikipedia.org/wiki/Kings_of_england
http://en.wikipedia.org/wiki/Kings_of_East_Anglia
http://en.wikipedia.org/wiki/Kings_of_Essex
http://en.wikipedia.org/wiki/Kings_of_Kent
http://en.wikipedia.org/wiki/Kings_of_Mercia
http://en.wikipedia.org/wiki/Kings_of_Northumbria
http://en.wikipedia.org/wiki/Kings_of_Sussex
http://en.wikipedia.org/wiki/Kings_of_Wessex

fix_filenames	<i>Convert filenames to match those in the book</i>
---------------	---

Description

Some filenames have been altered in order to comply with portability requirements on CRAN. This function converts the filenames between the CRAN forms and the book forms.

Usage

```
fix_filenames(x = c("tobook", "tocran"),
             dir = system.file("extdata", packages = "learningr"))
```

Arguments

x	Either "tocran" or "tobook".
dir	Directory containing the files.

Value

A logical vector of length 4, TRUE for each file whose name was changed.

Examples

```
## Not run:
#To convert the files to the book form, use:
fix_filenames("tobook")
#The files were converted to CRAN form using:
fix_filenames("tocran", "learningr/inst/extdata")

## End(Not run)
```

gonorrhoea	<i>Gonorrhoea</i>
------------	-------------------

Description

Rates of gonorrhoea infection in the US by year, with contextual information about age, ethnicity and gender.

Format

A data frame with the following columns.

Year Year that infected people visited the clinic.

Age.Group Age group of person infected.

Ethnicity Ethnicity of person infected.

Gender Gender of person infected.

Rate Number of infections per 100000 people.

References

Compiled by Richard Cotton <richierocks@gmail.com> from <http://www.cdc.gov/std/stats11/tables/22b.htm>

hafu

Hafu

Description

Half-caste manga characters.

Format

Both data frames have the following columns.

Year Integer year that the manga was made.

Series Name of series.

Character Name of character.

Gender Gender of character.

Father Nationality of character's father.

Mother Nationality of character's mother.

Eyes Character's eye colour.

Hair Character's hair colour.

Notes Notes on data collection or ambiguity.

hafu2 has these additional columns.

FathersNationalityIsUncertain

MothersNationalityIsUncertain

Note

The dataset was kindly provided by Gwern Branwen.

hafu2 is a lightly cleaned up version of hafu.

Gwern's notes: The following list includes manga, light novel, anime, and video game characters (there being little point in keeping the mediums separate). It also includes characters who are not hafu themselves but a quarter-foreign inasmuch as they imply a hafu at some point. Characters are treated separately even if they are in the same work (eg. siblings). Classification is based on in-universe or out-of-universe information, since appearance can be highly misleading in anime (blue eyes may indicate heroic status, rather than being Caucasian; hair color may be chosen for contrast against other characters or signal stereotypes like red hair indicating a fiery personality), and different groups will identify the same anime character as belonging to their own race (Lu 2009), perhaps due to minimalistic drawings intended to save money or enable viewers to project themselves onto a character.

References

The dataset was compiled by Gwern Branwen <gwern0@gmail.com>. The original is available from <http://www.gwern.net/hafu#list>.

hypotenuse

Dumb hypotenuse function.

Description

Calculate the (Pythagorean) hypotenuse of two numeric vectors using the obvious algorithm.

Usage

```
hypotenuse(x, y)
```

Arguments

x	A numeric vector.
y	A numeric vector.

Value

A numeric vector of the hypotenuse of the inputs.

Note

This algorithm fails when the inputs are very large or very small, making it unsuitable for real-world use.

References

Cleve Moler (MATLAB creator and discoverer of the Moler-Morrison algorithm for calculating hypotenuses) discusses the pro and cons of several algorithms here. <http://blogs.mathworks.com/cleve/2012/07/30/pythagorean-addition>

See Also

[hypot](#)

Examples

```
hypotenuse(5, 12)           #okay
hypotenuse(1e-300, 1e-300) #fails
hypotenuse(1e300, 1e300)   #fails
```

learningr

Datasets to accompany the book Learning R

Description

learningr contains datasets that are used in examples in the book “Learning R”.

Author(s)

Richard Cotton <richierocks@gmail.com>

References

Cotton, R (2013) *Learning R* O Reilly. ISBN 978-1-4493-5710-8.

obama_vs_mccain

Obama vs. McCain

Description

State-by-state voting information in the 2008 US presidential election, along with contextual information on income, unemployment, ethnicity and religion.

Format

A data frame with 52 observations (one for each US state) and the following columns.

State The name of the US state.

Region The US Federal region.

Obama Percentage of voters who voted for Barack Obama in the 2008 presidential election.

McCain Percentage of voters who voted for John McCain in the 2008 presidential election.

Turnout Percentage of people who voted in the 2008 presidential election.

Unemployment Percentage of people who are unemployed.

Income Mean annual income in US dollars.

Population Number of people living in the state.

Catholic Percentage of people identifying as Catholic.

Protestant Percentage of people identifying as Protestant.

Other Percentage of people identifying as religious, but not Catholic or Protestant.

Non.religious Percentage of people identifying as non-religious.

Black Percentage of people identifying as black.

Latino Percentage of people identifying as Latino.

Urbanization Percentage of people living in an urban area.

Note

Religious identification data are not available for Alaska and Hawaii. The totals of these columns is generally less than 100, since some people didn't give an answer. The District of Columbia is included, even though it isn't a state. The dataset is not guaranteed to be error free. Please double check the data if you want to use it for something serious.

References

This dataset was kindly compiled and provided by Edwin Thoen <edwinthoen@hotmail.com>.

The voting information came from <http://www.uselectionatlas.org/>, extracted on 2011-12-09.

The ethnicity, income and urbanisation information came from <http://quickfacts.census.gov>, extracted on 2011-12-09.

The unemployment information came from <http://data.bls.gov/timeseries/LNS14000000>, extracted 2011-12-09.

The religious information came from Table 12 of the American Religious Identification Survey 2008. http://commons.trincoll.edu/aris/files/2011/08/ARIS_Report_2008.pdf.

Index

*Topic **datasets**

- alpe_d_huez, 2
- crab_tag, 3
- deer_endocranial_volume, 5
- english_monarchs, 6
- gonorrhoea, 7
- hafu, 8
- obama_vs_mccain, 10

- alpe_d_huez, 2
- alpe_d_huez2 (alpe_d_huez), 2

- buggy_count, 2

- count, 3
- crab_tag, 3

- deer_endocranial_volume, 5

- english_monarchs, 6

- fix_filenames, 7

- gonorrhoea, 7

- hafu, 8
- hafu2 (hafu), 8
- hypot, 10
- hypotenuse, 9

- learningr, 10
- learningr-package (learningr), 10

- obama_vs_mccain, 10

- table, 3