

Package ‘okcupiddata’

August 19, 2016

Title OkCupid Profile Data for Introductory Statistics and Data
Science Courses

Version 0.1.0

Description Cleaned profile data from “OkCupid Profile Data for Introductory
Statistics and Data Science Courses” (Journal of Statistics Education 2015
<<http://www.amstat.org/publications/jse/v23n2/kim.pdf>>).

Depends R (>= 2.10)

License CC0

Encoding UTF-8

LazyData TRUE

URL <http://github.com/rudeboybert/okcupiddata>

BugReports <https://github.com/rudeboybert/okcupiddata/issues>

RoxygenNote 5.0.1

NeedsCompilation no

Author Albert Y. Kim [aut, cre],
Adriana Escobedo-Land [aut]

Maintainer Albert Y. Kim <albert.y.s.kim@gmail.com>

Repository CRAN

Date/Publication 2016-08-19 17:05:06

R topics documented:

okcupiddata	2
profiles	2
Index	4

 okcupiddata

okcupiddata:

Description

R package of cleaned profile data from "OkCupid Profile Data for Introductory Statistics and Data Science Courses": 59,946 OkCupid users who were living within 25 miles of San Francisco, had active profiles on June 26, 2012, were online in the previous year, and had at least one picture in their profile. The original data, publication, code, and codebook can be found at https://github.com/rudeboybert/JSE_OkCupid.

References

Albert Y. Kim, Adriana Escobedo-Land (2015). OkCupid Profile Data for Introductory Statistics and Data Science Courses. *Journal of Statistics Education*, 23(2). <http://www.amstat.org/publications/jse/v23n2/kim.pdf>.

Examples

```
library(okcupiddata)
data(profiles)
```

 profiles

Cleaned OkCupid profile data

Description

Cleaned profile data 59,946 OkCupid users who were living within 25 miles of San Francisco, had active profiles on June 26, 2012, were online in the previous year, and had at least one picture in their profile. The original data and codebook can be found at https://github.com/rudeboybert/JSE_OkCupid.

Usage

```
profiles
```

Format

A data.frame with 59946 rows and 22 variables:

age Age

body_type Body type

diet Dietary habits

drinks Drinking habits

drugs Drug usage habits

education Education level
ethnicity Ethnicity
height Height in inches
income Income
job Job
last_online Date/time of last login to OkCupid
location Location
offspring Number of offspring
orientation Sexual orientation
pets Number of pets
religion Religious affiliation
sex Sex. Note at the time OkCupid only allowed for male/female binary. This has since been relaxed.
sign Astrological sign
smokes Smoking habits
speaks Languages spoken
status Relationship status
essay0 Response to first essay question (my self summary), trimmed to 140 characters

Details

The differences between the cleaned and original version of profiles data are:

Essay Responses Due to file size restrictions, only the first 140 characters of each user's first essay response (my self summary) is included

Missing income values Previously coded as -1, they are now coded as NA

All other missing values Previously coded as "", they are now coded as NA

offspring and sign String instances of "'’" are replaced with apostrophes

last_online Date/time strings are converted to USA/Pacific timezone POSIXct date-time objects using [parse_date_time](#)

Source

https://github.com/rudeboybert/JSE_OkCupid

Examples

```

library(okcupiddata)
data(profiles)
# If using RStudio:
# View(profiles)
summary(profiles$income)

```

Index

*Topic **datasets**

profiles, [2](#)

okcupiddata, [2](#)

okcupiddata-package (okcupiddata), [2](#)

parse_date_time, [3](#)

profiles, [2](#)