Package ‘testDriveR’

February 6, 2019

Type Package

Title Teaching Data for Statistics and Data Science

Version 0.5.1

Description Provides data sets for teaching statistics and data science courses. It includes a sample of data from John Edmund Kerrich’s famous coinflip experiment. These are data that I use for teaching SOC 4015 / SOC 5050 at Saint Louis University (SLU). The package also contains an R Markdown template with the required formatting for assignments in my courses SOC 4015, SOC 4650, SOC 5050, and SOC 5650 at SLU.

License GPL-3

URL https://github.com/chris-prener/testDriveR

BugReports https://github.com/chris-prener/testDriveR/issues

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Suggests dplyr, ggplot2, knitr, rmarkdown, testthat

NeedsCompilation no

Author Christopher Prener [aut, cre] (<https://orcid.org/0000-0002-4310-9888>)

Maintainer Christopher Prener <chris.prener@slu.edu>

Repository CRAN

Date/Publication 2019-02-06 13:30:02 UTC

R topics documented:

auto17 ................................................................. 2
childMortality ..................................................... 3
gss14 ................................................................. 4
gss14_simple ....................................................... 5
kerrich ................................................................. 6
testDriveR ............................................................ 7

Index 8
Description

A data set containing model year 2017 vehicles for sale in the United States.

Usage

data(auto17)

Format

A data frame with 1216 rows and 21 variables:

- **id**: DOT vehicle ID number
- **mfr**: vehicle manufacturer
- **mfrDivision**: vehicle brand
- **carLine**: vehicle name
- **carClass**: vehicle type, numeric
- **carClassStr**: vehicle type, string
- **cityFE**: fuel economy, city
- **hwyFE**: fuel economy, highway
- **combFE**: fuel economy, combined
- **guzzlerStr**: poor fuel economy
- **fuelStr**: fuel, abbrev.
- **fuelStr2**: fuel, full
- **fuelCost**: estimated fuel cost
- **displ**: engine displacement
- **transStr**: transmission, full
- **transStr2**: transmission, abbrev.
- **gears**: number of gears
- **cyl**: number of cylinders
- **airAsp**: air aspiration method
- **driveStr**: vehicle drive type, abbrev.
- **driveStr2**: vehicle drive type, full

Source

https://www.fueleconomy.gov/feg/download.shtml
Examples

str(auto17)
head(auto17)

tabular(childMortality)  

UNICEF Childhood Mortality Data

Description

A data set containing time series data by country for estimated under-5, infant, and neonatal mortality rates.

Usage

data(childMortality)

Format

A data frame with 28982 rows and 6 variables:

countryISO  two-letter country code

countryName  full name of country

continent  name of continent

category  type of mortality rate - infant_MR, child_MR, or under5_MR

year  year of estimate

estimate  estimated mortality rate

Source

http://childmortality.org

Examples

str(childMortality)
Description

A data set containing data on work, salary, and education from the 2014 General Social Survey. Missing data are explicitly identified with NAs and all data are represented as factors when appropriate.

Usage

data(gss14)

Format

A data frame with 2538 rows and 19 variables:

YEAR  GSS year for this respondent
INCOME06  Total family income (2006 version)
INCOM16  Rs family income when 16 yrs old
REG16  Region of residence, age 16
RACE  Race of respondent
SEX  Respondents sex
SPDEG  Spouses highest degree
MADEG  Mothers highest degree
PADEG  Fathers highest degree
DEGREE  Rs highest degree
CHILDS  Number of children
SPWRKSLF  Spouse self-emp. or works for somebody
SPHRS1  Number of hrs spouse worked last week
MARITAL  Marital status
WRKSLF  R self-emp or works for somebody
HRS1  Number of hours worked last week
WRKSTAT  Labor force status
ID_  Respondent id number
BALLOT  Ballot used for interview

Source

https://gssdataexplorer.norc.org
Examples

str(gss14)
head(gss14)

gss14_simple 2014 General Social Survey (Simplified)

Description

A data set containing data on work, salary, and education from the 2014 General Social Survey. Missing data are not explicitly identified with NAs and all data are represented numerically instead of as factors when appropriate.

Usage

data(gss14_simple)

Format

A data frame with 2538 rows and 19 variables:

YEAR    GSS year for this respondent
INCOME06 Total family income (2006 version)
INCOM16  Rs family income when 16 yrs old
REG16    Region of residence, age 16
RACE     Race of respondent
SEX      Respondents sex
SPDEG    Spouses highest degree
MADEG    Mothers highest degree
PADEG    Fathers highest degree
DEGREE   Rs highest degree
CHILDS   Number of children
SPWRKSLF Spouse self-emp. or works for somebody
SPHRS1   Number of hrs spouse worked last week
MARITAL  Marital status
WRKSLF   R self-emp or works for somebody
HRS1     Number of hours worked last week
WRKSTAT  Labor force status
ID_      Respondent id number
BALLOT   Ballot used for interview
Source

https://gssdataexplorer.norc.org

Examples

str(gss14_simple)
head(gss14_simple)

kerrich Kerrich Coin Toss Trial Outcomes

Description

A data set containing 2,000 trials of coin flips from statistician John Edmund Kerrich’s 1940s experiments while imprisoned by the Nazis during World War Two.

Usage

data(kerrich)

Format

A data frame with 1216 rows and 21 variables:

- **id** trial
- **outcome** outcome of each trial; TRUE = heads, FALSE = tails
- **average** cumulative mean of outcomes

Source

https://stats.stackexchange.com/questions/76663/john-kerrich-coin-flip-data/77044#77044
https://books.google.com/books/about/An_experimental_introduction_to_the_theo.html?id=JBTvAAAAMAAJ&hl=en

References


Examples

str(kerrich)

```
if (require("ggplot2")) {
  ggplot(data = kerrich) +
  geom_hline(mapping = aes(yintercept = .5, color = "p(heads)") +
  geom_line(mapping = aes(x = id, y = average)) +
  ylim(0,1)
}
```
Description

The goal of testDriveR is to provide data sets for teaching statistics and data science courses. This package includes a sample of data from John Edmund Kerrich’s famous coinflip experiment. These are data that I use for teaching SOC 4015 / SOC 5050 at Saint Louis University.

Details

There are currently five data sets that are included in the package:

- auto17 - A data set containing model year 2017 vehicles for sale in the United States
- childMortality - A data set containing childhood mortality time series data by country from UNICEF
- gss14 - A data set containing a selection of variables related to work and education from the 2014 General Social Survey
- gss14_simple - A simple version of gss14 without factors created and without missing data explicitly declared
- kerrich - A data set containing 2000 trials of coin flips by John Edmund Kerrich
Index

*Topic datasets
  auto17, 2
  childMortality, 3
  gss14, 4
  gss14_simple, 5
  kerrich, 6

auto17, 2, 7
childMortality, 3, 7
gss14, 4, 7
gss14_simple, 5, 7
kerrich, 6, 7
testDriveR, 7
testDriveR-package (testDriveR), 7