

Package: tigerData (via r-universe)

November 24, 2024

Type Package

Title GC Statistics Datasets

Version 0.2.1

Date 2021-02-07

Author Homer White <hwhite0@georgetowncollege.edu>

Maintainer Homer White <hwhite0@georgetowncollege.edu>

Description A small, informal collection of datasets useful in undergraduate statistics courses.

Depends R (>= 3.5.0)

Encoding UTF-8

License GPL (>=3)

LazyLoad yes

LazyData true

URL <https://github.com/homerhanumat/tigerData>

RoxygenNote 7.3.1

Repository <https://homerhanumat.r-universe.dev>

RemoteUrl <https://github.com/homerhanumat/tigerData>

RemoteRef HEAD

RemoteSha 89d32a98581112bcb07b51d5c95ef34c0a3b77b2

Contents

cars2017	2
diabetes	3
donation	4
event	4
eventParticipation	5
firesetting	5
haliotis	6
lgbtplus	7

nhanes18	8
parkExp	9
parking	10
participant	11
payphones	11
reviews	12
shrooms	13
uniformsNFL	14
whately_2015	14
Index	17

cars2017

Fuel Economy Data on Cars in 2017

Description

2017 version of data set for textbook Modern Data Science With R (see Chapter 9).

Format

A data frame with 1103 observations on the following 7 variables.

make manufacturer of car

model car model

displacement Engine Displacement (Liters)

cylinders number of cylinders

city_mpg fuel economy (mpg), city

hwy_mpg fuel economy (mpg), highway

gears number of gears

Source

<https://www.fueleconomy.gov>

diabetes

*Diabetes Risk***Description**

Subset of survey data collected by the US National Center for Health Statistics (NCHS). The original data was based on home interviews of about 5,000 people per years, from 1999-2004.

Format

A data frame with 9096 observations on the following 23 variables.

sex "male" or "female"

age age of subject in years

pregnant "yes" or "no"

ethnicity Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, or Other/Multi

smoker "yes" or "no"

diabetic "yes" or "no"

height height (meters)

weight weight (kilograms)

waist waist circumference (meters)

wci the proposed body shape index

bmi body mass index

ptfat percent trunk fat

tfat mass of trunk fat

lfat limb fat

llean limb lean tissue

lbmi lean-tissue only BMI

fbmi fat-only BMI

bbmi bone BMI

pfat percent fat

bmd bone mineral density

fmhm_other Framingham risk score

hdl HDL cholesterol

chol cholesterol (LDL?)

bps systolic blood pressure, mmHg

bpd diastolic blood pressure, mmHg

income ratio of family income to poverty threshold. (5 stands for a ratio greater than or equal to 5)

Source

Modified from NCHS in package DataComputing. The original data is from NHANES, the National Health and Nutrition Survey. See http://www.cdc.gov/nchs/nhanes/search/nhanes03_04.aspx# for more information.

donation	<i>Zen Center Donations</i>
----------	-----------------------------

Description

Donations made to a fictional Zen Center. In a family with [participant](#), [event](#) and [eventParticipation](#).

Format

A data frame with 66 observations on the following 5 variables.

Don_ID Donation ID

Part_ID Participant ID

Don_Amount Amount of donation, in dollars.

Don_Date Date of donation.

Don_Form A character vector with two values: cash and check.

Source

Hypothetical data.

event	<i>Zen Center Events</i>
-------	--------------------------

Description

Events at a fictional Zen Center. In a family with [participant](#), [donation](#) and [eventParticipation](#).

Format

A data frame with 15 observations on the following 7 variables.

Event_ID Event ID

Event_Type Type of event. A factor with levels potluck, retreat and sangha.

Event_Location Location of the event.

Event_Start Start-time of the event.

Event_End Ending time of the event.

Event_Cost Nominal cost of the event, in dollars.

Part_ID ID of the participant who organizes the event.

Source

Hypothetical data.

eventParticipation *Zen Center Participants*

Description

Participation of persons in events associated with a fictional Zen Center. In a family with [participant](#), [donation](#) and [event](#).

Format

A data frame with 107 observations on the following 5 variables.

Part_ID Participant ID

Event_ID ID of the participant.

EP_Amount_Paid Amount actually paid by the participant.

EP_Notes Miscellaneous comments

EP_Attended Whether or not the participant actually attended.

Source

Hypothetical data.

firesetting *Fire-setting Among Teenagers*

Description

Modified from a dataset obtained in the course of a study on factors that are associated with fire-setting among at-risk youth. The data comes from national surveys of at-risk teenagers.

Format

A data frame with 975 observations on the following 6 variables.

age Child's age in years.

sex Sex of the child.

race Child's race.

school.attitude A measure of child's perceptions about school, combined from surveys given to child and to his/her parents. Higher scores indicate poorer attitudes.

academic A measure of the child's academic performance. Higher scores indicate more academic problems.

adhd Scaled scores from a test for ADHD. Higher scores indicates more problems with ADHD.

fires Whether or not the child sets fires (0 = does not, 1 = does).

Source

Doctroal dissertation by Carrie H. Bowling, University of Kentucky, 2013.. Further details in [../doc/firesetting_phd_proposal.pdf](#).

haliotis

Age of Sea Snails?

Description

Subset of data from a study on sea snails of the genus *Haliotis*. The age of a such a snail (in years) is quite close to the number of rings it has on its shell plus 1.5. The idea is to predict the age of an individual sea snail from other characteristics. The rest of the data is withheld for evaluation purposes.

Format

A data frame with 2923 observations on the following 9 variables.

Type male, femal or infant

LongSh length of longest shell

Diam diameter

Height height

WhWt weight when whole

ShuckWt weight when shucked

ViscWt weight of the snail's viscera

ShellWt weight of the snails's shell

Age number of rings, plus 1.5

Details

From the ICU website: "The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope – a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age."

Source

Dataset derived from the UCI Machine Learning Repository. See <http://archive.ics.uci.edu/ml/datasets/Abalone> for more information, including a citation of the original research article.

lgbtplus

Homonegativity in Greek Life

Description

Georgetown College students surveyed their peers on attitudes about LBGT issues and persons.

Format

A data frame with 75 observations on the following 22 variables.

Gender

LGBTPlus

Age

Race

School.System

LGBTPlusFamily

LGBTPlusFamily.Proximity

LGBTPlusFriends

College.Year

Sports

Greek

Q1G

Q2G

Q3G

Q4E

Q5V

Q6R

Q7V

Q8V

Q9E

Q10V

Q11E

Note

From Hunter Gatewood (email): "It was coded mainly based upon the numbers stated within the questions. For question 7 (on the front), if they answered no for question 6 (front), we coded it as 5. For question 11, we had to separate it between 'for a short time' (1-3 semesters) and 'for a longer time (4+ semesters) with coding (1,2)."

Source

Hunter Gatewood, Molly Dixon, Almond Bailey. Class: PSY 311, Fall 2014. Instructor: Dr. Regan Lookadoo. For survey form see [../doc/lbgt_likert.pdf](#).

 nhanes18

National Health and Nutrition Examination Survey

Description

Results of the Survey as of 2018.

Format

A data frame with 9842 observations on the following 18 variables.

id An ID number for the subject

gender "male" or "female"

age age of subject in years

arthritis Was the subject ever told he/she had arthritis? (yes, no)

edu level of education attained

married marital status

income income level

cholesterol blood cholesterol (mmol/L)

glucose glucose, refridgerated serum (mmol/L)

iron iron (umol/L)

sodium sodium (mmol/L)

weight weight (kg)

systolic systolic blood pressure, mm/Hg

diastolic diastolic blood pressure, mm/Hg

asthma Does the subject have asthma? (yes, no)

heartattack Was the subject ever told he/she had a heart attack? (yes, no)

liver Was the subject ever told he/she had a liver condition? (yes, no)

cancer Was the subject ever told he/she had a cancer or malignancy? (yes, no)

Source

NHANES, the National Health and Nutrition Survey. See http://www.cdc.gov/nchs/nhanes/search/nhanes03_04.aspx# for more information.

Description

A subset of the parking data frame, giving only the subject involved in the experiment. In the experiment, parked cars were approached by either an expensive car or a cheap one. The approaching car waited for the spot, and while waiting either honked once or did not honk at all.

Format

A data frame with 237 observations on the following 12 variables.

confcar The type of car that was waiting for the parking spot (or that just drove by). Either a Nissan Maxima or an Infinity Q45. The car is "confronting" the parked car, hence the name of the variable.

sex Sex of the driver of the parked car.

race Race of the driver of the parked car.

num Number of people in the parked car (including the driver).

horn The waiting car either honked the horn once, or did not honk at all.

carval Book value of the parked car, in dollars.

month Month in which the incident occurred.

day Day of the week on which the incident occurred.

miltime Time at which the incident occurred, in military units. For example, 1130 denotes 11:30AM, while 1350 denotes for 1:50PM.

time Time in seconds for the parked car to depart the parking spot.

ccstatus Status of the waiting "confronting" car. The Maxima is considered a low-status car, whereas the Infinity Q45 is an expensive, "high-status" car.

valuediff Difference in value between the confronting car and the parked car, in dollars. The values of the confronting cars were as follows: Maxima: 5200, Infinity Q45: 57000.

Note

This is almost the original data. B. Ruback indicates (personal communication) that several observations are missing and cannot be recovered at the present time.

Source

"Territorial Defense in Parking Lots: Retaliation Against Waiting Drivers", B. Ruback and D. Juieng, *Journal of Applied Social Psychology*, Volume 27, Issue 9, May 1997, pp. 821-834. Provided by B. Ruback.

parking

*Territoriality in Parking***Description**

A study of how long it takes a driver to vacate his/her spot in a parking lot.

Format

A data frame with 237 observations on the following 12 variables.

confcar The type of car that was waiting for the parking spot (or that just drove by). Either a Nissan Maxima, a Lexus or an Infinity Q45. The car is "confronting" the parked car, hence the name of the variable.

sex Sex of the driver of the parked car.

race Race of the driver of the parked car.

num Number of people in the parked car (including the driver).

horn The waiting car either did not intrude on the parked car at all, intruded only slightly by driving by, or stopped near the parking spot and waited. In that case the waiting car either honked the horn once, or did not honk at all.

carval Book value of the parked car, in dollars.

month Month in which the incident occurred.

day Day of the week on which the incident occurred.

miltime Time at which the incident occurred, in military units. For example, 1130 denotes 11:30AM, while 1350 denotes for 1:50PM.

time Time in seconds for the parked car to depart the parking spot.

ccstatus Status of the waiting "confronting" car. The Maxima is considered a low-status car, whereas the Lexus and Infinity Q45 are expensive, "high-status" cars.

valuediff Difference in value between the confronting car and the parked car, in dollars. The values of the confronting cars were as follows: Maxima: 5200, Lexus: 43000, Infinity Q45: 57000.

Note

This is almost the original data. B. Ruback indicates (personal communication) that three observations are missing and cannot be recovered at the present time.

Source

"Territorial Defense in Parking Lots: Retaliation Against Waiting Drivers", B. Ruback and D. Juieng, *Journal of Applied Social Psychology*, Volume 27, Issue 9, May 1997, pp. 821-834. Provided by B. Ruback.

participant	<i>Zen Center Participants</i>
-------------	--------------------------------

Description

Persons associated with fictional Zen Center. In a family with [event](#), [donation](#) and [eventParticipation](#).

Format

A data frame with 11 observations on the following 11 variables.

Part_ID Participant ID

Part_Class Type of the participant. A factor with levels member and visitor.

Part_Fname First name of participant.

Part_Mname Middle name of participant.

Part_Lname Last name of participant.

Part_Address Street address of participant.

Part_City City of participant.

Part_State State of participant.

Part_Zip Postal code of participant.

Part_Email Email address of participant.

Part_Phone Phone number of participant.

Source

Hypothetical data.

payphones	<i>Territoriality in Pay-Phones</i>
-----------	-------------------------------------

Description

Will people using a public pay-phone talk longer if someone is waiting to use their phone? In this experiment, conducted in 1989, "the investigators measured the length of time (in seconds) that subjects spent on the telephone under one of three conditions: when alone (A), when one person was using an adjacent telephone (B), or when one person was using an adjacent telephone and another person was waiting to use one of the two telephones. The study was conducted in an alcove of a shopping mall, an area that contained only the two adjacent telephones." (Quotation from Business Statistics, 6th. ed. 1992, by W. Daniel and J. Terrell.)

Format

A data frame with 56 observations on the following 3 variables.

sex Sex of the subject.

treatment Which condition the subject was put into (A, B or C as described above) by the researchers.

time Time in seconds that the subject spent on the phone.

Source

R.B. Ruback, K.D. Poe, and P.Doriat, "Waiting on a Phone: Intrusion on Callers Leads to Territorial Defense" *Social Psychology Quarterly*, 52:232-241. Gender data provided by R.B. Ruback (personal communication).

reviews

Amazon.com Book Reviews

Description

Amazon.com reader-reviews of several popular books.

Format

A data frame with 243,269 observations on the following 5 variables.

book The book under review. Values along with book-titles are as follows:

- hunger: "The Hunger Games"
- shades: "Fifty Shades of Gray"
- fault: "The Fault in our Stars"
- martian: "The Martian"
- unbroken: "Unbroken"
- gonegirl: "The Gone Girl"
- traingirl: "Girl on a Train"
- goldfinch: "The Goldfinch"

rating rating assigned (1-5)

URL_fragment Prepend "https://www.amazon.com/" to get the full URL of the review.

review_title Title of the review; usually a concise judgment of the book.

content HTML of the review text.

Source

This data frame is a compilation of the data sets in "Amazon Book Reviews", in the UC-Irvine Machine Learning Repository. See <https://archive.ics.uci.edu/ml/datasets/Amazon+book+reviews> for more information.

shrooms

Can You Eat This Mushroom?

Description

Subset of data from a study on edibility of mushrooms. The individual mushrooms come from 23 species of gilled mushrooms in the Agaricus and Lepiota Family. The aim is to come up with a rule for predicting, on the basis of an individual mushroom's characteristics, whether or not the mushroom is edible. Remaining data is held back for evaluation of proposed rules.

Format

A data frame with 5891 observations on the following 23 variables.

class Whether the mushroom is edible or poisonous.

cap.shape

cap.surface

cap.color

bruises Whether or not the mushroom is bruised.

odor

gill.attachment

gill.spacing

gill.size

gill.color

stalk.shape

stalk.root

stalk.surface.above.ring

stalk.surface.below.ring

stalk.color.above.ring

stalk.color.below.ring

veil.type

veil.color

ring.number

ring.type

spore.print.color

population

habitat

Source

A sample from of mushroom records drawn from The Audubon Society Field Guide to North American Mushrooms (1981). G. H. Lincoff (Pres.), New York: Alfred A. Knopf, Original data contributed by Jeffrey Schlimmer to the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml>, Irvine, CA: University of California, School of Information and Computer Science. See <http://archive.ics.uci.edu/ml/datasets/Mushroom>.

 uniformsNFL

Malevolence of NFL Uniforms and Penalty Yardage

Description

28 NFL teams from the 1980's. Team uniforms were rated for their "malevolence", and the average penalty yardage for each team was also recorded.

Format

A data frame with 28 observations on the following 3 variables.

team Name of NFL team.

malevolence Rating of "malevolence" accorded to the team uniform. High scores indicate more malevolence.

z_pen_yards Mean penalty yardage per game for the team, expressed as a z-score.

Source

The Dark Side of Self- and Social Perception: Black Uniforms and Aggression in Professional Sports, Frank and Gilovich, Journal of Personality and Social Psychology, 1988, Vol. 54, No. 1, 74-85.

 whately_2015

Weather data from Macleish Field Stations

Description

Weather data collected at the Macleish Field Station in Whately, MA during 2015. This is a copy of the whately_2015 data from package 'macleish': <<https://github.com/beanumber/macleish>>.

Format

For both, a data frame (`[dplyr::tbl_df()]`) with roughly 52,560 rows and 8 or 9 variables.

The following variables are values that are found in either the ‘whately_2015’ or ‘orchard_2015’ data tables.

All variables are averaged over the 10 minute interval unless otherwise noted.

when Timestamp for each measurement set in Eastern Standard Time.

temperature average temperature, in Celsius

wind_speed Wind speed, in meters per second

wind_dir Wind direction, in degrees

rel_humidity How much water there is in the air, in millimeters

pressure Atmospheric pressure, in millibars

rainfall Total rainfall, in millimeters

solar_radiation Amount of radiation coming from the sun, in Watts/meters². Solar measurement for Whately

par_density Photosynthetically Active Radiation (sunlight between 400 and 700 nm), in average density of Watts/meters². One of two solar measurements for Orchard

par_total Photosynthetically Active Radiation (sunlight between 400 and 700 nm), in average total over measurement period of Watts/meters². One of two solar measurements for Orchard

Details

The Macleish Field Station is a remote outpost owned by Smith College and used for field research. There are two weather stations on the premises. One is called ‘WhatelyMet’ and the other is ‘OrchardMet’.

The ‘WhatelyMet’ station is located at (42.448470, -72.680553) and the ‘OrchardMet’ station is at (42.449653, -72.680315).

‘WhatelyMet’ is located at the end of Poplar Hill Road in Whately, Massachusetts, USA. The meteorological instruments of ‘WhatelyMet’ (except the rain gauge) are mounted at the top of a tower 25.3 m tall, well above the surrounding forest canopy. The tower is located on a local ridge at an elevation 250.75m above sea level.

‘OrchardMet’ is located about 250 m north of the first tower in an open field next to an apple orchard. Full canopy trees (~20 m tall) are within 30 m of this station. This station has a standard instrument configuration with temperature, relative humidity, solar radiation, and barometric pressure measured between 1.5 and 2.0 m above the ground. Wind speed and direction are measured on a 10 m tall tower and precipitation is measured on the ground. Ground temperature is measured at 15 and 30 cm below the ground surface 2 m south of the tower. The tower is located 258.1 m above sea level. Data collection at OrchardMet began on June 27th, 2014.

The variables shown above are weather data collected at ‘WhatelyMet’ and ‘OrchardMet’ during 2015. Solar radiation is measured in two different ways: see ‘SlrW_Avg’ or the ‘PAR’ variables for Photosynthetic Active Radiation.

Note that a loose wire resulted in erroneous temperature reading at OrchardMet in late November, 2015.

Source

These data are recorded at <<https://www.smith.edu/about-smith/sustainable-smith/ceeds>>

Index

* datasets

- [cars2017, 2](#)
 - [diabetes, 3](#)
 - [donation, 4](#)
 - [event, 4](#)
 - [eventParticipation, 5](#)
 - [firesetting, 5](#)
 - [haliotis, 6](#)
 - [lgbtplus, 7](#)
 - [nhanes18, 8](#)
 - [parkExp, 9](#)
 - [parking, 10](#)
 - [participant, 11](#)
 - [payphones, 11](#)
 - [reviews, 12](#)
 - [shrooms, 13](#)
 - [uniformsNFL, 14](#)
 - [whately_2015, 14](#)
-
- [cars2017, 2](#)
 - [diabetes, 3](#)
 - [donation, 4, 4, 5, 11](#)
 - [event, 4, 4, 5, 11](#)
 - [eventParticipation, 4, 5, 11](#)
 - [firesetting, 5](#)
 - [haliotis, 6](#)
 - [lgbtplus, 7](#)
 - [nhanes18, 8](#)
 - [parkExp, 9](#)
 - [parking, 10](#)
 - [participant, 4, 5, 11](#)
 - [payphones, 11](#)
 - [reviews, 12](#)