Package: usdatasets (via r-universe)

November 8, 2024

Type Package

Title A Comprehensive Collection of U.S. Datasets

Version 0.1.0

Maintainer Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

Description Provides a diverse collection of U.S. datasets
encompassing various fields such as crime, economics,
education, finance, energy, healthcare, and more. It serves as
a valuable resource for researchers and analysts seeking to
perform in-depth analyses and derive insights from
U.S.-specific data.

License GPL-3

URL https://github.com/lightbluetitan/usdatasets

Encoding UTF-8

LazyData true

Suggests ggplot2, dplyr, knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

RoxygenNote 7.3.2

VignetteBuilder knitr

Repository https://lightbluetitan.r-universe.dev

RemoteUrl https://github.com/lightbluetitan/usdatasets

RemoteRef HEAD

RemoteSha 495129fdfb9386b71735d2155e4db6dc7c4c9c7a

Contents

---

acs12_tbl_df *American Community Survey 2012*

Description

The dataset name has been changed to 'acs12_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble data frame, helping to differentiate it from other datasets within the package. The original content of the dataset has not been modified in any way.

Usage

```
data(acs12_tbl_df)
```

Format

A tibble with 2,000 observations and 13 variables:

income Income of individuals (integer).

employment Employment status (factor with 3 levels).

hrs_work Number of hours worked per week (integer).

race Race of individuals (factor with 4 levels).

age Age of individuals (integer).

gender Gender of individuals (factor with 2 levels: "male", "female").

citizen Citizenship status (factor with 2 levels: "no", "yes").

time_to_work Time taken to travel to work in minutes (integer).

lang Primary language spoken at home (factor with 2 levels: "english", "other").

married Marital status (factor with 2 levels: "no", "yes").

edu Educational attainment (factor with 3 levels).

disability Disability status (factor with 2 levels).

birth_qrtr Birth quarter of individuals (factor with 4 levels).

Source

American Community Survey, 2012.

---

  age_at_mar_tbl_df   *Age at first marriage of 5,534 US women.*

Description

The dataset name has been changed to 'age_at_mar_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble data frame, helping to differentiate it from other datasets within the package. The original content of the dataset has not been modified in any way.

Usage

```
data(age_at_mar_tbl_df)
```

Format

A tibble with 5,534 observations and 1 variable:

age Age at first marriage (integer).

Source

United States Census Data.

---

  airlines_tbl_df   *Airline names - U.S. Airlines Carrier Codes and Names*

Description

The dataset name has been changed to 'airlines_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble, helping to differentiate it from other datasets within the package. The original content of the dataset has not been modified in any way.

Usage

```
data(airlines_tbl_df)
```

Format

A tibble with 16 observations and 2 variables:

carrier Carrier code (character) representing the airline.

name Name of the airline (character).

Source

U.S. Department of Transportation.

---

airports_tbl_df *Airport metadata - U.S. Airports Information*

Description

The dataset name has been changed to 'airports_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble, helping to differentiate it from other datasets within the package. The original content of the dataset has not been modified in any way.

Usage

```
data(airports_tbl_df)
```

Format

A tibble with 1,458 observations and 8 variables:

faa FAA airport code (character).

name Name of the airport (character).

lat Latitude of the airport (numeric).

lon Longitude of the airport (numeric).

alt Altitude of the airport (numeric).

tz Time zone (numeric).

dst Daylight saving time flag (character).

tzone Time zone name (character).

Source

U.S. Federal Aviation Administration (FAA).

---

airquality_df   *New York Air Quality Measurements*

Description

The dataset name has been changed to 'airquality_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'df' identifies the dataset as a data frame, helping to differentiate it from other datasets within the package. The original content of the dataset has not been modified in any way.

Usage

```
data(airquality_df)
```

Format

A data frame with 153 observations and 6 variables:

Ozone Ozone concentration (parts per billion) from 1 to 331.

Solar.R Solar radiation (watts per square meter).

Wind Wind speed (miles per hour).

Temp Temperature (degrees Fahrenheit).

Month Month of the observation (integer from 5 to 9).

Day Day of the observation (integer from 1 to 31).

Source

United States Environmental Protection Agency (EPA).

---

ames_tbl_df   *Housing prices in Ames, Iowa*

Description

The dataset name has been changed to 'ames_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(ames_tbl_df)
```

Format

A tibble with 2,930 observations and 82 variables:

Order Row number in the dataset.

PID Parcel Identifier.

area Total house area in square feet.

price Sale price of the house.

MS.SubClass Building class type.

MS.Zoning Zoning classification of the property.

Lot.Frontage Lot frontage length in feet.

Lot.Area Total lot area in square feet.

Street Street type access to the property.

Alley Alley type access.

Lot.Shape Shape of the lot.

Land.Contour Land contour around the property.

Utilities Availability of utilities.

Lot.Config Lot configuration.

Land.Slope Slope of the land.

Neighborhood Neighborhood in Ames.

Condition.1 Proximity to main conditions like railroads.

Condition.2 Proximity to secondary conditions.

Bldg.Type Type of building.

House.Style Architectural style of the house.

Overall.Qual Overall quality of the materials and finish.

Overall.Cond Overall condition of the house.

Year.Built Year the house was built.

Year.Remod.Add Year of the last remodel or addition.

Roof.Style Roof style.

Roof.Matl Roof material.

Exterior.1st Primary exterior material.

Exterior.2nd Secondary exterior material.

Mas.Vnr.Type Masonry veneer type.

Mas.Vnr.Area Masonry veneer area in square feet.

Exter.Qual Exterior material quality.

Exter.Cond Condition of the exterior material.

Foundation Type of foundation.

Bsmt.Qual Basement quality.

Bsmt.Cond Basement condition.

Bsmt.Exposure Basement exposure to the outside.

BsmtFin.Type.1 Type 1 of finished basement.

BsmtFin.SF.1 Square feet of finished basement type 1.

BsmtFin.Type.2 Type 2 of finished basement.

BsmtFin.SF.2 Square feet of finished basement type 2.

Bsmt.Unf.SF Unfinished basement area in square feet.

Total.Bsmt.SF Total basement area in square feet.

Heating Type of heating system.

Heating.QC Heating system quality.

Central.Air Presence of central air conditioning.

Electrical Type of electrical system.

X1st.Flr.SF First floor area in square feet.

X2nd.Flr.SF Second floor area in square feet.

Low.Qual.Fin.SF Low-quality finished area in square feet.

Bsmt.Full.Bath Number of full bathrooms in the basement.

Bsmt.Half.Bath Number of half bathrooms in the basement.

Full.Bath Number of full bathrooms above ground.

Half.Bath Number of half bathrooms above ground.

Bedroom.AbvGr Number of bedrooms above ground.

Kitchen.AbvGr Number of kitchens above ground.

Kitchen.Qual Kitchen quality.

TotRms.AbvGrd Total number of rooms above ground.

Functional Functionality of the house.

Fireplaces Number of fireplaces.

Fireplace.Qu Fireplace quality.

Garage.Type Type of garage.

Garage.Yr.Blt Year the garage was built.

Garage.Finish Garage finish type.

Garage.Cars Number of cars the garage can accommodate.

Garage.Area Garage area in square feet.

Garage.Qual Garage quality.

Garage.Cond Garage condition.

Paved.Drive Indicates whether the driveway is paved.

Wood.Deck.SF Wood deck area in square feet.

Open.Porch.SF Open porch area in square feet.

Enclosed.Porch Enclosed porch area in square feet.

X3Ssn.Porch Three-season porch area in square feet.

Screen.Porch Screened porch area in square feet.

Pool.Area Pool area in square feet.

Pool.QC Pool quality.

Fence Type of fence.

Misc.Feature Miscellaneous features of the property.

Misc.Val Value of miscellaneous features.

Mo.Sold Month the house was sold.

Yr.Sold Year the house was sold.

Sale.Type Type of sale.

Sale.Condition Condition of the sale.

Source

Ames Housing Dataset, provided by Dean De Cock

---

births14_tbl_df *US Births 2014*

Description

The dataset name has been changed to 'births14_tbl_df' to avoid confusion with other packages in
the R ecosystem from which datasets have been sourced. This naming convention helps distinguish
this dataset as part of the 'usdatasets' package and assists users in identifying its specific character-
istics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has
not been modified in any way.

Usage

```
data(births14_tbl_df)
```

Format

A tibble with 1,000 observations and 13 variables:

fage Age of the father (in years).

mage Age of the mother (in years).

mature Indicates if the mother is mature (yes/no).

weeks Number of weeks of pregnancy.

premie Indicates if the baby is a premature birth (yes/no).

visits Number of prenatal visits.

gained Weight gained by the mother during pregnancy (in pounds).

weight Birth weight of the baby (in grams).

lowbirthweight Indicates if the baby is of low birth weight (yes/no).

sex Sex of the baby (male/female).

habit Maternal smoking habits (yes/no).

marital Marital status of the mother (married/single).

whitemom Indicates if the mother is white (yes/no).

Source

National Vital Statistics Reports

---

births_tbl_df *North Carolina births, 100 cases*

Description

The dataset name has been changed to 'births_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(births_tbl_df)
```

Format

A tibble with 150 observations and 9 variables:

f_age Age of the father (in years).

m_age Age of the mother (in years).

weeks Number of weeks of pregnancy.

premature Indicates if the baby is premature (factor: yes/no).

visits Number of prenatal visits.

gained Weight gained by the mother during pregnancy (in pounds).

weight Birth weight of the baby (in grams).

sex_baby Sex of the baby (factor: male/female).

smoke Indicates if the mother smoked during pregnancy (factor: yes/no).

Source

National Vital Statistics Reports

Boston_df *Housing Values in Suburbs of Boston*

Description

The dataset name has been changed to 'Boston_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix '_df' identifies the dataset as a data frame. The original content of the dataset has not been modified in any way.

Usage

```
data(Boston_df)
```

Format

A data frame with 506 observations and 14 variables:

crim Per capita crime rate by town.

zn Proportion of residential land zoned for lots over 25,000 sq. ft.

indus Proportion of non-retail business acres per town.

chas Charles River dummy variable (1 if tract bounds river; 0 otherwise).

nox Nitric oxides concentration (parts per 10 million).

rm Average number of rooms per dwelling.

age Proportion of owner-occupied units built prior to 1940.

dis Weighted distances to five Boston employment centers.

rad Index of accessibility to radial highways.

tax Full-value property tax rate per $10,000.

ptratio Pupil-teacher ratio by town.

black 1000(Bk - 0.63)^2 where Bk is the proportion of Black residents by town.

lstat Percentage of lower status of the population.

medv Median value of owner-occupied homes in $1000s.

Source

Boston Housing Data

Cars93_df   *Data from 93 Cars on Sale in the USA in 1993*

Description

The dataset name has been changed to 'Cars93_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix '_df' identifies the dataset as a data frame. The original content of the dataset has not been modified in any way.

Usage

```
data(Cars93_df)
```

Format

A data frame with 54 observations and 6 variables:

type Type of the car (factor with 3 levels).

price Price of the car (in US dollars).

mpg_city Miles per gallon in the city.

drive_train Drive train type (factor with 3 levels).

passengers Number of passengers the car can accommodate.

weight Weight of the car (in pounds).

Source

1993 Cars Data

---

census_tbl_df   *Random sample of 2000 U.S. Census Data*

Description

The dataset name has been changed to 'census_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(census_tbl_df)
```

Format

A tibble with 500 observations and 8 variables:

census_year Year of the census (in integer).

state_fips_code FIPS code for the state (factor with 47 levels).

total_family_income Total family income (in US dollars).

age Age of the individual (in years).

sex Sex of the individual (factor: male/female).

race_general General race category (factor with 8 levels).

marital_status Marital status of the individual (factor with 6 levels).

total_personal_income Total personal income (in US dollars).

Source

US Census Bureau

---

cia_factbook_tbl_df *CIA Factbook Details on Countries*

Description

The dataset name has been changed to 'cia_factbook_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(cia_factbook_tbl_df)
```

Format

A tibble with 259 observations and 11 variables:

country Name of the country (factor with 259 levels).

area Total area of the country (in square kilometers).

birth_rate Birth rate (number of live births per 1,000 people).

death_rate Death rate (number of deaths per 1,000 people).

infant_mortality_rate Infant mortality rate (number of deaths of infants under one year old per 1,000 live births).

internet_users Number of internet users (in millions).

life_exp_at_birth Life expectancy at birth (in years).

maternal_mortality_rate Maternal mortality rate (number of maternal deaths per 100,000 live births).

net_migration_rate Net migration rate (number of migrants per 1,000 people).

population Total population of the country.

population_growth_rate Population growth rate (percentage).

Source

CIA World Factbook

---

cle_sac_tbl_df   *Cleveland and Sacramento Demographic and Income Data (2000)*

Description

The dataset name has been changed to 'cle_sac_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(cle_sac_tbl_df)
```

Format

A tibble with 500 observations and 8 variables:

year Year of the observation (integer).

state State of the observation (factor with 2 levels).

city City of the observation (character).

age Age of the individual (integer).

sex Sex of the individual (factor with 2 levels).

race Race of the individual (character).

marital_status Marital status of the individual (character).

personal_income Personal income of the individual (integer).

Source

Cleveland Study

---

county_tbl_df *United States Counties*

Description

The dataset name has been changed to 'county_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(county_tbl_df)
```

Format

A tibble with 3,142 observations and 15 variables:

name Name of the county.

state State in which the county is located (factor with 51 levels).

pop2000 Population of the county in the year 2000.

pop2010 Population of the county in the year 2010.

pop2017 Population of the county in the year 2017.

pop_change Change in population over the years.

poverty Poverty rate in the county.

homeownership Rate of homeownership in the county.

multi_unit Percentage of multi-unit housing.

unemployment_rate Unemployment rate in the county.

metro Indicates if the county is in a metropolitan area (factor with 2 levels).

median_edu Median education level in the county (factor with 4 levels).

per_capita_income Per capita income in the county.

median_hh_income Median household income in the county.

smoking_ban Indicates if there is a smoking ban in place (factor with 3 levels).

Source

United States Census Bureau

---

env_regulation_tbl_df   *American Adults on Regulation and Renewable Energy*

Description

The dataset name has been changed to 'env_regulation_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(env_regulation_tbl_df)
```

Format

A tibble with 705 observations and 1 variable:

statement Environmental regulation statement (character).

Source

Environmental Regulation Study

---

fcid_tbl_df   *Summary of male heights from USDA Food Commodity Intake Database*

Description

The dataset name has been changed to 'fcid_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(fcid_tbl_df)
```

Format

A tibble with 100 observations and 2 variables:

height Height of the individual (numeric).

num_of_adults Number of adults in the household (integer).

Source

Family Characteristics and Income Study

---

goog_tbl_df *Google stock data*

Description

The dataset name has been changed to 'goog_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(goog_tbl_df)
```

Format

A tibble with 98 observations and 7 variables:

date Date of the stock price observation (factor with 98 levels).

open Opening price of the stock (numeric).

high Highest price during the trading session (numeric).

low Lowest price during the trading session (numeric).

close Closing price of the stock (numeric).

volume Number of shares traded (integer).

adj_close Adjusted closing price of the stock (numeric).

Source

Google Stock Market Data

| govrace10_tbl_df | *Election results for 2010 Governor races in the U.S.* |

Description

The dataset name has been changed to 'govrace10_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(govrace10_tbl_df)
```

Format

A tibble with 37 observations and 23 variables:

id Identification number (numeric).

state State name (character).

abbr State abbreviation (character).

name1 Name of the first candidate (character).

perc1 Percentage of votes for the first candidate (numeric).

party1 Political party of the first candidate (character).

votes1 Number of votes for the first candidate (numeric).

name2 Name of the second candidate (character).

perc2 Percentage of votes for the second candidate (numeric).

party2 Political party of the second candidate (character).

votes2 Number of votes for the second candidate (numeric).

name3 Name of the third candidate (character).

perc3 Percentage of votes for the third candidate (numeric).

party3 Political party of the third candidate (character).

votes3 Number of votes for the third candidate (numeric).

name4 Name of the fourth candidate (character).

perc4 Percentage of votes for the fourth candidate (numeric).

party4 Political party of the fourth candidate (character).

votes4 Number of votes for the fourth candidate (numeric).

name5 Name of the fifth candidate (character).

perc5 Percentage of votes for the fifth candidate (numeric).

party5 Political party of the fifth candidate (character).

votes5 Number of votes for the fifth candidate (numeric).

Source

2010 Gubernatorial Races

---

homicides15_tbl_df   *Homicides in nine cities in 2015*

Description

The dataset name has been changed to 'homicides15_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(homicides15_tbl_df)
```

Format

A tibble with 1922 observations and 15 variables:

uid Unique identifier (integer).

city_name City name where the homicide occurred (character).

offense_code Offense code (character).

offense_type Type of offense (character).

date_single Date of the homicide (POSIXct).

address Location address of the homicide (character).

longitude Longitude of the homicide location (numeric).

latitude Latitude of the homicide location (numeric).

location_type Type of location where the homicide occurred (character).

location_category Category of the location (character).

fips_state FIPS code of the state (integer).

fips_county FIPS code of the county (character).

tract Census tract where the homicide occurred (character).

block_group Block group number (integer).

block Block number (integer).

Source

2015 Homicides Data

houserace10_tbl_df	*Election results for the 2010 U.S. House of Represenatives races*

Description

The dataset name has been changed to 'houserace10_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(houserace10_tbl_df)
```

Format

A tibble with 435 observations and 24 variables:

id Unique race identifier (numeric).

state Name of the state (character).

abbr State abbreviation (character).

num District number (numeric).

name1 Name of the first candidate (character).

perc1 Percentage of votes for the first candidate (numeric).

party1 Party affiliation of the first candidate (character).

votes1 Number of votes for the first candidate (numeric).

name2 Name of the second candidate (character).

perc2 Percentage of votes for the second candidate (numeric).

party2 Party affiliation of the second candidate (character).

votes2 Number of votes for the second candidate (numeric).

name3 Name of the third candidate (character).

perc3 Percentage of votes for the third candidate (numeric).

party3 Party affiliation of the third candidate (character).

votes3 Number of votes for the third candidate (numeric).

name4 Name of the fourth candidate (character).

perc4 Percentage of votes for the fourth candidate (numeric).

party4 Party affiliation of the fourth candidate (character).

votes4 Number of votes for the fourth candidate (numeric).

name5 Name of the fifth candidate (character).

perc5 Percentage of votes for the fifth candidate (numeric).

party5 Party affiliation of the fifth candidate (character).

votes5 Number of votes for the fifth candidate (numeric).

Source

2010 U.S. House of Representatives Election Data

---

house_tbl_df *United States House of Representatives historical make-up*

Description

The dataset name has been changed to 'house_tbl_df' to avoid confusion with other packages in the R ecosystem from which datasets have been sourced. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and assists users in identifying its specific characteristics. The suffix 'tbl_df' identifies the dataset as a tibble. The original content of the dataset has not been modified in any way.

Usage

```
data(house_tbl_df)
```

Format

A tibble with 116 observations and 12 variables:

congress Congress number (numeric).

year_start Starting year of the congress (numeric).

year_end Ending year of the congress (numeric).

seats Total number of seats in the House of Representatives (numeric).

p1 Abbreviation of the first party (character).

np1 Number of seats for the first party (numeric).

p2 Abbreviation of the second party (character).

np2 Number of seats for the second party (numeric).

other Number of seats for other parties (numeric).

vac Number of vacant seats (numeric).

del Number of delegate seats (numeric).

res Number of resident commissioner seats (numeric).

Source

Historical House of Representatives Data

---

immigration_tbl_df   *Poll on illegal workers in the US*

Description

The dataset name has been changed to 'immigration_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(immigration_tbl_df)
```

Format

A tibble with 910 observations and 2 variables:

response Factor indicating the response to immigration-related questions, with 4 levels.

political Factor indicating the political alignment associated with the responses, with 3 levels.

Source

Data from surveys on immigration attitudes

---

leg_mari_tbl_df   *Legalization of Marijuana Support in 2010 California Survey*

Description

The dataset name has been changed to 'leg_mari_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(leg_mari_tbl_df)
```

Format

A tibble with 119 observations and 1 variable:

response Factor indicating responses related to legal marijuana, with 2 levels.

Source

Data from surveys on attitudes towards legal marijuana

---

marathon_tbl_df *New York City Marathon Times (outdated)*

Description

The dataset name has been changed to 'marathon_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(marathon_tbl_df)
```

Format

A tibble with 59 observations and 3 variables:

year Integer indicating the year of the marathon event.

gender Factor indicating the gender of the participants, with 2 levels.

time Numeric value representing the marathon completion time in hours.

Source

Data from marathon event results

---

military_tbl_df *US Military Demographics*

Description

The dataset name has been changed to 'military_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(military_tbl_df)
```

Format

A tibble with an unspecified number of observations and 6 variables:

grade Factor indicating the military grade, with 3 levels.

branch Factor indicating the branch of the military, with 4 levels.

gender Factor indicating the gender of the participants, with 2 levels.

race Factor indicating the race of the participants, with 7 levels.

hisp Logical indicating whether the participants identify as Hispanic.

rank Integer representing the rank of the participants.

Source

Data from military personnel demographics

---

minn38_df *Minnesota High School Graduates of 1938*

Description

The dataset name has been changed to 'minn38_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a data frame. The original content of the dataset has not been modified.

Usage

```
data(minn38_df)
```

Format

A data frame with 168 observations and 5 variables:

hs Factor indicating the high school status, with 3 levels.

phs Factor indicating the post-high school status, with 4 levels.

fol Factor indicating the field of study, with 7 levels.

sex Factor indicating the gender of the participants, with 2 levels.

f Integer representing the associated numerical value for the participants.

Source

Data from the Minnesota 1938 study

---

mlb_players_18_tbl_df *Batter Statistics for 2018 Major League Baseball (MLB) Season*

Description

The dataset name has been changed to 'mlb_players_18_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(mlb_players_18_tbl_df)
```

Format

A tibble with 1270 observations and 19 variables:

name Character string representing the name of the player.

team Character string indicating the team the player belongs to.

position Character string indicating the position played by the player.

games Integer representing the number of games played.

AB Integer indicating the number of at-bats.

R Integer representing the number of runs scored.

H Integer representing the number of hits.

doubles Integer indicating the number of doubles hit.

triples Integer indicating the number of triples hit.

HR Integer representing the number of home runs hit.

RBI Integer indicating the number of runs batted in.

walks Integer indicating the number of walks received.

strike_outs Integer indicating the number of strikeouts.

stolen_bases Integer representing the number of stolen bases.

caught_stealing_base Integer indicating the number of times caught stealing.

AVG Numeric representing the batting average.

OBP Numeric representing the on-base percentage.

SLG Numeric representing the slugging percentage.

OPS Numeric representing the on-base plus slugging percentage.

Source

Data from Major League Baseball (MLB) player statistics for the 2018 season

---

mn_police_use_of_force_df

Minneapolis police use of force data.

Description

The dataset name has been changed to 'mn_police_use_of_force_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a data frame. The original content of the dataset has not been modified.

Usage

```
data(mn_police_use_of_force_df)
```

Format

A data frame with 12925 observations and 13 variables:

response_datetime Character string representing the date and time of the response.

problem Character string describing the nature of the problem.

is_911_call Character string indicating whether the incident was initiated by a 911 call.

primary_offense Character string indicating the primary offense involved in the incident.

subject_injury Character string describing the injuries sustained by the subject, if any.

force_type Character string describing the type of force used by the police.

force_type_action Character string describing the specific actions related to the use of force.

race Character string indicating the race of the subject involved in the incident.

sex Character string indicating the sex of the subject.

age Integer representing the age of the subject.

type_resistance Character string describing the type of resistance offered by the subject.

precinct Character string indicating the precinct in which the incident occurred.

neighborhood Character string representing the neighborhood where the incident occurred.

Source

Data from police use of force reports in Minnesota

---

nba_players_19_tbl_df *NBA Players for the 2018-2019 season*

Description

The dataset name has been changed to 'nba_players_19_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'us-datasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(nba_players_19_tbl_df)
```

Format

A tibble with 494 observations and 7 variables:

first_name Character string representing the player's first name.

last_name Character string representing the player's last name.

team Character string indicating the name of the team.

team_abbr Character string representing the team's abbreviation.

position Character string indicating the player's position on the team.

number Character string representing the player's jersey number.

height Numeric value representing the player's height.

Source

Data from NBA players' statistics in 2019

---

ncbirths_tbl_df *North Carolina births, 1000 cases*

Description

The dataset name has been changed to 'ncbirths_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(ncbirths_tbl_df)
```

Format

A tibble with 1000 observations and 13 variables:

fage Integer representing the father's age.

mage Integer representing the mother's age.

mature Factor with 2 levels indicating if the mother is mature (>=35 years).

weeks Integer representing the number of gestation weeks.

premie Factor with 2 levels indicating if the baby was born prematurely.

visits Integer representing the number of prenatal visits.

marital Factor with 2 levels indicating the marital status of the mother.

gained Integer representing the mother's weight gain during pregnancy (in pounds).

weight Numeric value representing the baby's birth weight (in grams).

lowbirthweight Factor with 2 levels indicating if the baby was born with low birth weight.

gender Factor with 2 levels indicating the baby's gender.

habit Factor with 2 levels indicating if the mother has a smoking habit.

whitemom Factor with 2 levels indicating if the mother is white.

Source

Data from birth records in North Carolina

`nycvehiclethefts_tbl_df`

Thefts of motor vehicles 2014 to 2017

Description

The dataset name has been changed to 'nycvehiclethefts_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(nycvehiclethefts_tbl_df)
```

Format

A tibble with 35,746 observations and 9 variables:

uid Integer value representing a unique identifier for each vehicle theft incident.

date_single Character value representing the single date of the theft incident.

date_start Character value representing the start date of the theft incident.

date_end Character value representing the end date of the theft incident.

longitude Numeric value indicating the longitude where the incident occurred.

latitude Numeric value indicating the latitude where the incident occurred.

location_type Character value representing the type of location where the theft took place.

location_category Character value indicating the category of the location.

census_block Character value indicating the census block where the incident took place.

Source

Data from the New York City Vehicle Thefts records

`nyc_marathon_tbl_df`   *New York City Marathon Times*

Description

The dataset name has been changed to 'nyc_marathon_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(nyc_marathon_tbl_df)
```

Format

A tibble with 102 observations and 7 variables:

year Numeric value representing the year the marathon took place.

name Character value representing the name of the runner.

country Character value indicating the country of origin of the runner.

time Time variable in 'hms' format representing the finish time of the runner.

time_hrs Numeric value representing the finish time of the runner in hours.

division Character value indicating the division (category) the runner participated in.

note Character value containing additional notes, if any, about the runner or the race.

Source

Data from the New York City Marathon records

```
offshore_drilling_tbl_df
```
California poll on drilling off the California coast

Description

The dataset name has been changed to 'offshore_drilling_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(offshore_drilling_tbl_df)
```

Format

A tibble with 828 observations and 2 variables:

v1 Factor with 4 levels, representing different responses or categories related to offshore drilling.

v2 Factor with 3 levels, representing secondary categories or classifications related to the responses in v1.

Source

Data related to offshore drilling opinions or classifications

---

orings_tbl_df *1986 Challenger disaster and O-rings*

Description

The dataset name has been changed to 'orings_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(orings_tbl_df)
```

Format

A tibble with 23 observations and 4 variables:

mission Integer representing the mission number.

temperature Integer representing the launch temperature in Fahrenheit.

damaged Integer representing the number of damaged O-rings in the mission.

undamaged Numeric representing the number of undamaged O-rings in the mission.

Source

Data from NASA missions related to O-ring performance.

---

oscars_tbl_df *Oscar winners, 1929 to 2018*

Description

The dataset name has been changed to 'oscars_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(oscars_tbl_df)
```

Format

A tibble with 184 observations and 11 variables:

oscar_no Numeric indicating the Oscar number.

oscar_yr Numeric representing the year the Oscar was awarded.

award Character string indicating the category of the award.

name Character string with the name of the recipient.

movie Character string indicating the movie for which the award was given.

age Numeric indicating the age of the recipient at the time of the award.

birth_pl Character string indicating the birthplace of the recipient.

birth_date Date representing the birthdate of the recipient.

birth_mo Numeric indicating the birth month.

birth_d Numeric indicating the birth day.

birth_y Numeric indicating the birth year.

Source

Data from historical Oscar award records.

---

piracy_tbl_df   *Piracy and PIPA/SOPA*

Description

The dataset name has been changed to 'piracy_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(piracy_tbl_df)
```

Format

A tibble with 534 observations and 8 variables:

name Character string indicating the name of the politician.

party Factor with 3 levels representing the politician's party affiliation.

state Factor with 50 levels indicating the U.S. state the politician represents.

money_pro Numeric representing the amount of pro-piracy funding received.

money_con Numeric representing the amount of anti-piracy funding received.

years Integer indicating the number of years in office.

stance Factor with 5 levels indicating the politician's stance on piracy.

chamber Factor with 2 levels indicating the chamber of the U.S. Congress (House or Senate).

Source

Data on political stances and funding related to piracy.

---

precip_numeric   *Annual Precipitation in US Cities*

Description

The dataset name has been changed to 'precip_numeric' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a numeric vector. The original content of the dataset has not been modified.

Usage

```
data(precip_numeric)
```

Format

A numeric vector with 70 observations representing average annual precipitation (in inches) for various cities in the United States.

Mobile Numeric value representing the average annual precipitation in Mobile.

Juneau Numeric value representing the average annual precipitation in Juneau.

Phoenix Numeric value representing the average annual precipitation in Phoenix.

Los Angeles Numeric value representing the average annual precipitation in Los Angeles.

... Additional cities included in the dataset.

Source

Data on precipitation for various U.S. cities.

---

presidents_ts   *Quarterly Approval Ratings of US Presidents*

Description

The dataset name has been changed to 'presidents_ts' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a time series object. The original content of the dataset has not been modified.

Usage

```
data(presidents_ts)
```

Format

A time series object with 120 observations, covering quarterly data from 1945 to 1975. Each observation represents the number of presidents' approval ratings during a given quarter. The data is structured as follows:

Qtr1 Numeric values representing the approval ratings for the first quarter.

Qtr2 Numeric values representing the approval ratings for the second quarter.

Qtr3 Numeric values representing the approval ratings for the third quarter.

Qtr4 Numeric values representing the approval ratings for the fourth quarter.

Source

Data on presidential approval ratings from 1945 to 1975.

---

prrace08_tbl_df   *Election results for the 2008 U.S. Presidential race*

Description

The dataset name has been changed to 'prrace08_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(prrace08_tbl_df)
```

Format

A tibble with 51 observations and 7 variables:

state Factor indicating the U.S. state (including Washington D.C.) where the election took place.

state_full Factor providing the full name of the U.S. state corresponding to the abbreviation.

n_obama Integer representing the number of votes received by Barack Obama in the state.

p_obama Numeric representing the percentage of total votes received by Barack Obama in the state.

n_mc_cain Integer representing the number of votes received by John McCain in the state.

p_mc_cain Numeric representing the percentage of total votes received by John McCain in the state.

el_votes Integer indicating the number of electoral votes allocated to the state.

Source

Data on the 2008 U.S. presidential race results by state.

---

road_df   *Road Accident Deaths in US States*

Description

The dataset name has been changed to 'road_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a data frame. The original content of the dataset has not been modified.

Usage

```
data(road_df)
```

Format

A data frame with 26 observations and 6 variables:

deaths Integer indicating the number of road deaths.

drivers Integer representing the number of licensed drivers.

popden Numeric indicating the population density (people per square mile).

rural Numeric indicating the percentage of rural roads.

temp Integer representing the average temperature (in degrees Fahrenheit).

fuel Numeric indicating the fuel consumption per capita (in gallons).

Source

Data on road safety statistics, including deaths, drivers, population density, and environmental factors.

---

senaterace10_tbl_df   *Election results for the 2010 U.S. Senate races*

Description

The dataset name has been changed to 'senaterace10_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(senaterace10_tbl_df)
```

Format

A tibble with 38 observations and 23 variables:

id Numeric identifier for the election race.

state Character string indicating the U.S. state where the election took place.

abbr Character string representing the state abbreviation.

name1 Character string indicating the name of the first candidate.

perc1 Numeric indicating the percentage of votes received by the first candidate.

party1 Character string indicating the party affiliation of the first candidate.

votes1 Numeric indicating the total votes received by the first candidate.

name2 Character string indicating the name of the second candidate.

perc2 Numeric indicating the percentage of votes received by the second candidate.

party2 Character string indicating the party affiliation of the second candidate.

votes2 Numeric indicating the total votes received by the second candidate.

name3 Character string indicating the name of the third candidate.

perc3 Numeric indicating the percentage of votes received by the third candidate.

party3 Character string indicating the party affiliation of the third candidate.

votes3 Numeric indicating the total votes received by the third candidate.

name4 Character string indicating the name of the fourth candidate.

perc4 Numeric indicating the percentage of votes received by the fourth candidate.

party4 Character string indicating the party affiliation of the fourth candidate.

votes4 Numeric indicating the total votes received by the fourth candidate.

name5 Character string indicating the name of the fifth candidate.

perc5 Numeric indicating the percentage of votes received by the fifth candidate.

party5 Character string indicating the party affiliation of the fifth candidate.

votes5 Numeric indicating the total votes received by the fifth candidate.

Source

Data on U.S. Senate races held in 2010, including candidates' names, vote percentages, and party affiliations.

```
sp500_1950_2018_tbl_df
```
                                   Daily observations for the S&P 500 - Historical Data (1950-2018)

Description

The dataset name has been changed to 'sp500_1950_2018_tbl_df' to avoid confusion with other
packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the
'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been
modified.

Usage

```
data(sp500_1950_2018_tbl_df)
```

Format

A tibble with 17346 observations and 7 variables:

Date Factor indicating the date of the recorded stock prices.

Open Numeric representing the opening price of the stock.

High Numeric representing the highest price of the stock during the day.

Low Numeric representing the lowest price of the stock during the day.

Close Numeric representing the closing price of the stock.

Adj.Close Numeric representing the adjusted closing price of the stock.

Volume Numeric representing the trading volume of the stock.

Source

Historical data on S&P 500 stock prices from 1950 to 2018.

```
sp500_tbl_df
```
                        Financial information for 50 S&P 500 companies

---

Description

The dataset name has been changed to 'sp500_tbl_df' to avoid confusion with other packages in
the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets'
package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(sp500_tbl_df)
```

Format

A tibble with 50 observations and 12 variables:

stock Factor indicating the stock ticker symbol of the company.

market_cap Numeric representing the market capitalization of the company.

ent_value Numeric representing the enterprise value of the company.

trail_pe Numeric representing the trailing price-to-earnings ratio.

forward_pe Numeric representing the forward price-to-earnings ratio.

ev_over_rev Numeric representing the enterprise value to revenue ratio.

profit_margin Numeric representing the profit margin of the company.

revenue Numeric representing the total revenue generated by the company.

growth Numeric representing the growth rate of the company.

earn_before Numeric representing the earnings before interest and taxes (EBIT).

cash Numeric representing the cash holdings of the company.

debt Numeric representing the total debt of the company.

Source

Data on S&P 500 companies, including financial metrics and ratios.

---

state_abb_character   *US State Facts and Figures - U.S. State Abbreviations*

Description

The dataset name has been changed to 'state_abb_character' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a character vector. The original content of the dataset has not been modified.

Usage

```
data(state_abb_character)
```

Format

A character vector with 50 elements representing U.S. state abbreviations:

state_abb Character vector of state abbreviations, e.g., "AL" for Alabama, "CA" for California.

Source

U.S. state abbreviations.

---

state_area_numeric *US State Facts and Figures - US State Areas*

Description

The dataset name has been changed to 'state_area_numeric' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a numeric dataset. The original content of the dataset has not been modified.

Usage

```
data(state_area_numeric)
```

Format

A numeric dataset with 50 elements representing the area of U.S. states in square kilometers:

state_area Numeric values indicating the area of each state, measured in square kilometers.

Source

U.S. state areas.

---

state_center_list *US State Facts and Figures - US State Centers*

Description

The dataset name has been changed to 'state_center_list' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a list. The original content of the dataset has not been modified.

Usage

```
data(state_center_list)
```

Format

A list with 2 elements, each containing numeric values representing the geographical center coordinates of U.S. states:

x Numeric vector of length 50 representing the x-coordinates (longitude) of the state centers.

y Numeric vector of length 50 representing the y-coordinates (latitude) of the state centers.

Source

Geographical data for U.S. state centers.

---

state_division_factor   *US State Facts and Figures - US State Divisions*

Description

The dataset name has been changed to 'state_division_factor' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a factor. The original content of the dataset has not been modified.

Usage

```
data(state_division_factor)
```

Format

A factor with 50 observations representing the divisions of U.S. states. It contains 9 levels:

East South Central Region including Alabama, Kentucky, Mississippi, and Tennessee.

Pacific Region including California, Oregon, and Washington.

Mountain Region including Colorado, Idaho, Montana, Nevada, Utah, and Wyoming.

West South Central Region including Arkansas, Louisiana, Oklahoma, and Texas.

New England Region including Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.

South Atlantic Region including Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, Washington, D.C., and West Virginia.

East North Central Region including Illinois, Indiana, Michigan, Ohio, and Wisconsin.

West North Central Region including Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.

Middle Atlantic Region including New Jersey, New York, and Pennsylvania.

Source

U.S. Census Bureau regional divisions.

state_name_character   *US State Facts and Figures - US State Names*

Description

The dataset name has been changed to 'state_name_character' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a character vector. The original content of the dataset has not been modified.

Usage

```
data(state_name_character)
```

Format

A character vector with 50 observations representing the names of U.S. states.

"Alabama" Name of the state.

"Alaska" Name of the state.

"Arizona" Name of the state.

"Arkansas" Name of the state.

"California" Name of the state.

"Colorado" Name of the state.

"Connecticut" Name of the state.

"Delaware" Name of the state.

"Florida" Name of the state.

"Georgia" Name of the state.

"Hawaii" Name of the state.

"Idaho" Name of the state.

"Illinois" Name of the state.

"Indiana" Name of the state.

"Iowa" Name of the state.

"Kansas" Name of the state.

"Kentucky" Name of the state.

"Louisiana" Name of the state.

"Maine" Name of the state.

"Maryland" Name of the state.

"Massachusetts" Name of the state.

"Michigan" Name of the state.

"Minnesota" Name of the state.

"Mississippi" Name of the state.

"Missouri" Name of the state.

"Montana" Name of the state.

"Nebraska" Name of the state.

"Nevada" Name of the state.

"New Hampshire" Name of the state.

"New Jersey" Name of the state.

"New Mexico" Name of the state.

"New York" Name of the state.

"North Carolina" Name of the state.

"North Dakota" Name of the state.

"Ohio" Name of the state.

"Oklahoma" Name of the state.

"Oregon" Name of the state.

"Pennsylvania" Name of the state.

"Rhode Island" Name of the state.

"South Carolina" Name of the state.

"South Dakota" Name of the state.

"Tennessee" Name of the state.

"Texas" Name of the state.

"Utah" Name of the state.

"Vermont" Name of the state.

"Virginia" Name of the state.

"Washington" Name of the state.

"West Virginia" Name of the state.

"Wisconsin" Name of the state.

"Wyoming" Name of the state.

Source

U.S. Census Bureau.

---

state_region_factor   *US State Facts and Figures - US State Regions*

Description

The dataset name has been changed to 'state_region_factor' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a factor variable representing U.S. state regions.

Usage

```
data(state_region_factor)
```

Format

A factor variable with 50 observations, representing the region of each U.S. state. The regions are classified into four levels:

"Northeast" States located in the Northeast region.

"South" States located in the Southern region.

"North Central" States located in the North Central region.

"West" States located in the Western region.

Source

U.S. Census Bureau.

---

state_x77_matrix   *US State Facts and Figures - US State Demographics and Statistics (1977)*

Description

The dataset name has been changed to 'state_x77_matrix' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a matrix variable representing various demographic and statistical attributes of U.S. states in 1977.

Usage

```
data(state_x77_matrix)
```

Format

A matrix with 50 rows and 8 columns representing various demographic and statistical characteristics of U.S. states. The columns include:

Population Population of the state.

Income Median income of the state's residents.

Illiteracy Illiteracy rate (percentage).

Life Exp Life expectancy (in years).

Murder Murder rate (per 100,000 inhabitants).

HS Grad High school graduation rate (percentage).

Frost Number of days with frost.

Area Total area of the state (in square miles).

Source

U.S. Census Bureau (1977).

UCBAdmissions_table   *Student Admissions at UC Berkeley*

Description

The dataset name has been changed to 'UCBAdmissions_table' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a table object. The original content of the dataset has not been modified.

Usage

```
data(UCBAdmissions_table)
```

Format

A table object with 24 entries representing the admissions data at U.C. Berkeley:

Admit A factor with levels "Admitted" and "Rejected".

Gender A factor with levels "Male" and "Female".

Dept A factor representing the department with levels "A", "B", "C", "D", "E", and "F".

values Numeric counts of admissions based on gender and department.

Source

U.C. Berkeley admissions data from 1973.

USAccDeaths_ts   *Accidental Deaths in the US 1973-1978*

Description

The dataset name has been changed to 'USAccDeaths_ts' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a time series object. The original content of the dataset has not been modified.

Usage

```
data(USAccDeaths_ts)
```

Format

A time series object with 72 observations representing monthly accidental deaths in the U.S. from 1973 to 1979:

years A numeric vector representing the years from 1973 to 1979.

months A character vector representing the months from January to December.

deaths Numeric values representing the number of accidental deaths for each month.

Source

U.S. accidental deaths data.

USArrests_df   *Violent Crime Rates by US State*

Description

The dataset name has been changed to 'USArrests_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a data frame. The original content of the dataset has not been modified.

Usage

```
data(USArrests_df)
```

Format

A data frame with 50 observations and 4 variables representing the rates of arrests in the U.S.:

Murder Numeric vector representing the murder rates per 100,000 residents.

Assault Integer vector representing the assault rates per 100,000 residents.

UrbanPop Integer vector representing the percentage of the population living in urban areas.

Rape Numeric vector representing the rape rates per 100,000 residents.

Source

U.S. arrests data from 1973.

UScitiesD_dist *Distances Between European Cities and Between US Cities*

Description

The dataset name has been changed to 'UScitiesD_dist' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a distance object. The original content of the dataset has not been modified.

Usage

```
data(UScitiesD_dist)
```

Format

A distance object containing the distances (in miles) between selected U.S. cities:

Atlanta Distance from Atlanta to other cities.

Chicago Distance from Chicago to other cities.

Denver Distance from Denver to other cities.

Houston Distance from Houston to other cities.

LosAngeles Distance from Los Angeles to other cities.

Miami Distance from Miami to other cities.

NewYork Distance from New York to other cities.

SanFrancisco Distance from San Francisco to other cities.

Seattle Distance from Seattle to other cities.

Washington.DC Distance from Washington D.C. to other cities.

Source

U.S. cities distance data.

---

usdatasets *usdatasets: A Comprehensive Collection of U.S. Datasets*

Description

This package provides a wide variety of datasets related to crime, economy, society, politics, and sports within the United States for testing, learning, and research purposes.

Details

usdatasets: A Comprehensive Collection of U.S. Datasets

A Comprehensive Collection of U.S. Datasets.

Author(s)

Maintainer: Renzo Cáceres Rossi <arenzocaceresrossi@gmail.com>

See Also

Useful links:

- <https://github.com/lightbluetitan/usdatasets>

USJudgeRatings_df *Lawyers' Ratings of State Judges in the US Superior Court*

Description

The dataset name has been changed to 'USJudgeRatings_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a data frame. The original content of the dataset has not been modified.

Usage

```
data(USJudgeRatings_df)
```

Format

A data frame with 43 observations and 12 variables representing ratings for U.S. judges:

CONT Numeric vector representing the judges' ratings on control.

INTG Numeric vector representing the judges' ratings on integrity.

DMNR Numeric vector representing the judges' ratings on demeanor.

DILG Numeric vector representing the judges' ratings on diligence.

CFMG Numeric vector representing the judges' ratings on communications with clients.

DECI Numeric vector representing the judges' ratings on decisiveness.

PREP Numeric vector representing the judges' ratings on preparation.

FAMI Numeric vector representing the judges' ratings on family law expertise.

ORAL Numeric vector representing the judges' ratings on oral communications.

WRIT Numeric vector representing the judges' ratings on written communications.

PHYS Numeric vector representing the judges' ratings on physical appearance.

RTEN Numeric vector representing the judges' ratings on overall rating.

Source

U.S. judge ratings data.

USPersonalExpenditure_matrix

Personal Expenditure Data

Description

The dataset name has been changed to 'USPersonalExpenditure_matrix' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a matrix. The original content of the dataset has not been modified.

Usage

```
data(USPersonalExpenditure_matrix)
```

Format

A matrix with 5 rows and 5 columns representing U.S. personal expenditures in different categories over selected years:

Food and Tobacco Numeric values representing expenditures on food and tobacco for the years 1940, 1945, 1950, 1955, and 1960.

Household Operation Numeric values representing expenditures on household operations for the years 1940, 1945, 1950, 1955, and 1960.

Medical and Health Numeric values representing expenditures on medical and health services for the years 1940, 1945, 1950, 1955, and 1960.

Personal Care Numeric values representing expenditures on personal care for the years 1940, 1945, 1950, 1955, and 1960.

Private Education Numeric values representing expenditures on private education for the years 1940, 1945, 1950, 1955, and 1960.

Source

U.S. personal expenditure data.

---

uspop_ts   *Populations Recorded by the US Census*

Description

The dataset name has been changed to 'uspop_ts' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a time series object. The original content of the dataset has not been modified.

Usage

```
data(uspop_ts)
```

Format

A time series object with 19 observations representing the U.S. population from 1790 to 1970:

values Numeric vector containing the population values in millions.

Source

U.S. Census Bureau.

---

us_crime_rates_spec_tbl_df
                         US Crime Rates

Description

The dataset 'us_crime_rates_spec_tbl_df' contains crime statistics for the United States, including various types of crimes and population data for each year. This dataset is structured as a tibble for ease of use within the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package.

Usage

```
data(us_crime_rates_spec_tbl_df)
```

Format

A tibble with 60 rows and 12 columns:

year Numeric year of the recorded data, e.g., 2000, 2001.

population Numeric population total for the respective year.

total Numeric total number of crimes reported.

violent Numeric total number of violent crimes.

property Numeric total number of property crimes.

murder Numeric total number of murders.

forcible_rape Numeric total number of forcible rapes.

robbery Numeric total number of robberies.

aggravated_assault Numeric total number of aggravated assaults.

burglary Numeric total number of burglaries.

larceny_theft Numeric total number of larcenies.

vehicle_theft Numeric total number of vehicle thefts.

Source

Federal Bureau of Investigation (FBI) Uniform Crime Reporting (UCR) Program.

---

us_temp_tbl_df *US Temperature Data*

Description

The dataset 'us_temp_tbl_df' contains temperature records from various weather stations across the United States, providing both maximum and minimum temperature readings. This dataset is structured as a tibble for ease of use within the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package.

Usage

```
data(us_temp_tbl_df)
```

Format

A tibble with 10,118 rows and 9 columns:

station Character string representing the weather station identifier.

name Character string for the name of the weather station.

latitude Numeric value for the latitude of the weather station.

longitude Numeric value for the longitude of the weather station.

elevation Numeric value for the elevation of the weather station in meters.

date Date of the recorded temperature data.

tmax Numeric value for the maximum temperature recorded (in degrees Celsius).

tmin Numeric value for the minimum temperature recorded (in degrees Celsius).

year Factor representing the year of the recorded data.

Source

National Oceanic and Atmospheric Administration (NOAA).

---

us_time_survey_tbl_df   *American Time Survey 2009 - 2019*

Description

The dataset name has been changed to 'us_time_survey_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a tibble. The original content of the dataset has not been modified.

Usage

```
data(us_time_survey_tbl_df)
```

Format

A tibble with 11 observations and 8 variables representing time use in various activities:

year Numeric value representing the year of the survey.

household_activities Numeric value representing time spent on household activities (in hours).

eating_and_drinking Numeric value representing time spent on eating and drinking (in hours).

leisure_and_sports Numeric value representing time spent on leisure and sports activities (in hours).

sleeping Numeric value representing time spent sleeping (in hours).

caring_children Numeric value representing time spent caring for children (in hours).

working_employed Numeric value representing time spent working while employed (in hours).

working_employed_days_worked Numeric value representing the number of days worked while employed.

Source

U.S. Bureau of Labor Statistics.

VADeaths_matrix   *Death Rates in Virginia (1940)*

Description

The dataset name has been changed to 'VADeaths_matrix' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a matrix. The original content of the dataset has not been modified.

Usage

```
data(VADeaths_matrix)
```

Format

A matrix containing mortality rates (per 1000) for different demographic groups in Virginia:

Rural Male Mortality rates for rural males by age group.

Rural Female Mortality rates for rural females by age group.

Urban Male Mortality rates for urban males by age group.

Urban Female Mortality rates for urban females by age group.

Source

Virginia mortality data.

```
voter_count_spec_tbl_df
```
US Voter Turnout Data.

Description

The dataset name has been changed to 'voter_count_spec_tbl_df' to avoid confusion with other packages in the R ecosystem. This naming convention helps distinguish this dataset as part of the 'usdatasets' package and identifies it as a special tibble. The original content of the dataset has not been modified.

Usage

```
data(voter_count_spec_tbl_df)
```

Format

A special tibble containing voting statistics across different years and regions:

year Year of the election.

region Region of the voters.

voting_eligible_population Total population eligible to vote.

total_ballots_counted Total number of ballots counted.

highest_office Total votes for the highest office.

percent_total_ballots_counted Percentage of total ballots counted.

percent_highest_office Percentage of votes for the highest office.

Source

Election data from various sources.

---

women_df   *Average Heights and Weights for American Women*

Description

The dataset name has been kept as 'women_df' to maintain consistency with other datasets in the R ecosystem. This naming convention helps clearly identify this dataset within the context of its application. The original content of the dataset has not been modified.

Usage

```
data(women_df)
```

Format

A data frame containing measurements of women's height and weight:

height Height of women in inches.

weight Weight of women in pounds.

Source

Based on statistical data for women's height and weight.

Index