Package: hellodatascience (via r-universe)

June 12, 2026

Type Package

Title Datasets from the Hello Data Science Book

Version 0.1.1

Description Provides datasets used for analysis and visualizations in the open-access Hello Data Science book.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Depends R (>= 3.5)

URL https://hellodata-science.github.io/hellodatascience/

BugReports https://github.com/hellodata-science/hellodatascience/issues

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Mine Dogucu [aut, cre] (ORCID: <https://orcid.org/0000-0002-8007-934X>), Catalina Medina [aut] (ORCID: <https://orcid.org/0000-0003-2847-8180>), Alma Castro [aut]

Maintainer Mine Dogucu <mdogucu@gmail.com>

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

Date/Publication 2026-06-12 19:25:51 UTC

RemoteUrl https://github.com/cran/hellodatascience

RemoteRef HEAD

RemoteSha 64d70125548d4dac0b268439c520d3a7c6af4a75

Contents

`atus_college` *ATUS (American Time Use Survey) College/University Student Data*

Description

The 2024 data was downloaded from U.S. Bureau of Labor Statistics' website https://www.bls.gov/tus/data/datafiles-2024.htm and subset to include only respondents who are enrolled in college or university. This dataset is used only for educational purposes. Those conducting real research should download the data from its original source. BLS.gov cannot vouch for the data or analyses derived from these data after the data have been retrieved from BLS.gov.

Usage

```
atus_college
```

Format

A data frame with 312 rows and 5 variables. Each row represents a college student.

employment full time or part time employment status of respondent

age age

enrollment are you enrolled as a full-time or part-time student?

weekly_earnings weekly earnings at main job

household_size number of people living in respondent's household

time_alone total nonwork-related time respondent spent alone (in minutes)

sleep_time time spent sleeping

work_time time spent working at main job

degree_class_time time spent taking class for degree, certification, or licensure

shopping_time time spent shopping (store, telephone, internet)

lunch_break_time time spent taking a lunch break

sports_time time spent participating in sports, exercise, or recreation

religious_time time spent attending or participating in religious services

Source

U.S. Bureau of Labor Statistics (2025). https://nssdc.gsfc.nasa.gov/planetary/factsheet/index.html.

confederations *FIFA Confederations Data*

Description

The data was obtained from FIFA website https://inside.fifa.com/associations/ and contains information on the FIFA Member Associations (MAs), also known are confederations, which are responsible for the development and governance of football/soccer within their region

Usage

```
confederations
```

Format

A data frame with 7 rows and 2 variables. Each row represents a FIFA confederation, including FIFA itself.

confederation name of the FIFA member association

region region or continent overseed by the confederation

Source

(2026). https://inside.fifa.com/associations/.

country_capital *FIFA World Cup 2026 Country and Capital Data*

Description

The data was obtained from The World Data website https://theworlddata.com/world-population-by-country/ which contains information on the Men's FIFA World Cup 2026 qualifying teams and their 2025 population, and from the WorldData.info website https://www.worlddata.info/capital-cities.php which contains information on the capital cities of all countries.

Usage

```
country_capital
```

Format

A data frame with 5 rows and 3 variables. Each row represents a country men's soccer team that qualified for the FIFA World Cup 2026.

country name of the country

capital name of the country's capital city

population population size in millions based on the United Nations Population Division estimates for 2025

Source

(2026). https://theworlddata.com/world-population-by-country/.

(2025). https://www.worlddata.info/capital-cities.php.

`country_rank` *FIFA World Cup 2026 Country and Ranking Data*

Description

The data was scraped from Whereig website https://www.whereig.com/football/fifa-world-rankings.html/ and contains information on the Men's FIFA World Cup 2026 qualifying teams and ranking data as of 01 April 2026

Usage

```
country_rank
```

Format

A data frame with 7 rows and 3 variables. Each row represents a country men's soccer team that qualified for the FIFA World Cup 2026.

country name of the country soccer team

fifa_rank FIFA world ranking as of April, 2026

confederation region/continent association affiliated to FIFA

Source

Whereig editors (2026). https://www.whereig.com/football/fifa-world-rankings.html/.

`mx_us_wc_ranks` *Mexico's and the United States' World Cup Participation Rankings Data*

Description

The data was gathered from The Soccer World Cups website https://www.thesoccerworldcups.com/world_cups.php and contains information about every World Cup played including national teams, standings, and more

Usage

```
mx_us_wc_ranks
```

Format

A data frame with 2 rows and 5 variables. Each row represents Mexico's and the United States' men soccer team and their final participation ranking at the last four world cup tournaments.

country name of the country soccer team

2010 final ranking in the 2010 world cup

2014 final ranking in the 2014 world cup

2018 final ranking in the 2018 world cup

2022 final ranking in the 2022 world cup

Source

(2026). https://www.thesoccerworldcups.com/world_cups.php.

---

penn_world *Penn World Data*

Description

The data was downloaded from https://www.rug.nl/ggdc/productivity/pwt/ and contains information about different economic measures of countries around the world. The dataset has been subset and variable names have been modified for exercise purposes.

Usage

```
penn_world
```

Format

A data frame with 12810 rows and 14 variables. Each row represents a country in a specific year.

Country Code 3-letter ISO country code

Country country name

Currency Unit currency unit

Year year

Real GDP Expenditure expenditure-side real GDP at chained PPPs (in mil. 2017US$)

Real GDP Output output-side real GDP at chained PPPs (in mil. 2017US$)

Population population (in millions)

Emp number of persons engaged (in millions)

Average Hours average annual hours worked by persons engaged

PL Consumption price level of household consumption, price level of USA GDPo in 2017=1

PL Capital Formation price level of capital formation, price level of USA GDPo in 2017=1

PL Gov price level of government consumption, price level of USA GDPo in 2017=1

PL Exports price level of exports, price level of USA GDPo in 2017=1

PL Imports price level of imports, price level of USA GDPo in 2017=1

Source

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150-3182, available for download at http://www.ggdc.net/pwt/.

`planets` *Planets Data*

Description

The data was scraped from NASA's website https://nssdc.gsfc.nasa.gov/planetary/factsheet/index.html and contains information on the planets of our Solar System

Usage

```
planets
```

Format

A data frame with 8 rows and 7 variables. Each row represents a planet.

name name of the planet

mass mass in 10^24 kg

length_of_day length of day in hours

mean_temp whether mean temperature in C is positive or not {negative}{positive}

n_moons number of moons

ring_system whether the planet has set of rings around it {TRUE} {FALSE}

surface_pressure surface pressure in bars

Source

David R. Williams (2024). https://nssdc.gsfc.nasa.gov/planetary/factsheet/index.html.

produce_prices *Fruit and Vegetable Prices*

Description

How much do fruits and vegetables cost? United States Department of Agriculture (USDA) Economic Research Service (ERS), estimated average prices for 153 commonly consumed fresh and processed fruits and vegetables. USDA ERS calculated average prices at retail stores using 2022 retail scanner data from Circana (formerly Information Resources Inc. (IRI)). A selection of retail establishments—grocery stores, supermarkets, supercenters, convenience stores, drug stores, and liquor stores—across the United States provides Circana with weekly retail sales data (revenue and quantity).

Usage

```
produce_prices
```

Format

A data frame with 155 rows and 10 variables:

id ID of item

produce name of produce

form form of produce, either 'Canned', 'Dried', 'Fresh', 'Frozen', or 'Juice'

retail_price average retail price per pound or per pint

retail_price_unit unit for the 'retail_price', either 'per pint' or 'per pound'

cup_equivalent_size For most fruits and vegetables, a cup equivalent is the edible portion that will fit into a 1-cup measuring cup; for raisins and other dried fruit, it is the edible portion that will fit into a 1/2-cup; and for leafy vegetables, 2 cups. An edible cup equivalent is the unit of measurement used by the U.S. Department of Agriculture and the Department of Health and Human Services to report fruit and vegetable consumption recommendations.

cup_equivalent_unit unit for 'cup_equivalent_size'

cup_equivalent_price average retail price per 'cup_equivalent_unit' of produce

type type of produce, either 'fruit' or 'vegetables'

year year # Add more items for each column

Source

U.S. Department of Agriculture, Economic Research Service. (2024). Fruit and vegetable prices. https://www.ers.usda.gov/data-products/fruit-and-vegetable-prices

Index