

Package: DataScienceExercises (via r-universe)

October 13, 2026

Type Package

Title Exercises for introductions to R

Version 0.0.1

Description This package contains exercises that help users to practice concepts that are part of most introductory courses to R.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

URL <https://github.com/teaching-datascience/DataScienceExercises>

BugReports <https://github.com/teaching-datascience/DataScienceExercises/issues>

Imports data.table (>= 1.15.0), dplyr (>= 1.1.4), ggcheck (>= 0.0.5), ggplot2 (>= 3.5.0), gradethis (>= 0.2.14), here (>= 1.0.1), knitr (>= 1.45), learnr (>= 0.11.5), purrr (>= 1.0.2), scales (>= 1.3.0), stringr (>= 1.5.1), tibble (>= 3.2.1), tidyr (>= 1.3.0)

Suggests anyflights, gapminder, haven, htmltools, readxl, testthat (>= 3.0.0), usethis, WDI (>= 2.7.8)

Depends R (>= 4.3)

RoxygenNote 7.3.3

Remotes rstudio/ggcheck, rstudio/gradethis

Config/testthat/edition 3

Config/pak/sysreqs cmake make libicu-dev libuv1-dev zlib1g-dev

Repository <https://teaching-datascience.r-universe.dev>

Date/Publication 2026-10-05 19:41:08 UTC

RemoteUrl <https://github.com/teaching-datascience/DataScienceExercises>

RemoteRef main

RemoteSha 99d000fa506c292af7b5e1bf6dea6ba3f7c78786

Contents

aggGDPlifexp	2
beer	3
check_setup	3
econjournals	4
EUFstudents	4
EUFstudentsemesters	5
gdplifexp2007	6
Growth	6
list_exercises	7
nycflights21_small	7
pop_var	8
product_complexity	8
regression_table	9
run_exercise	10
unemp_pop	10
wine2dine	11
Index	12

aggGDPlifexp	<i>Aggregated on GDP, life expectancy, and population over time.</i>
--------------	--

Description

The gapminder data on population size, life expectancy, and GDP per capita levels, aggregated over continents.

Usage

aggGDPlifexp

Format

A data frame with 60 rows and 5 variables:

continent Continent name

year Year of the observation

lifeExp Average life expectancy at birth, in years

pop Average population, in nb of people

gdpPercap Average GDP per capita, in 2005-international dollars

Source

<https://github.com/jennybc/gapminder>

beer	<i>Data on beer consumption</i>
------	---------------------------------

Description

A data set on beer consumption; the data was obtained for a single household

Usage

beer

Format

A data frame with 30 rows and 5 variables

consumption liters of beer consumed

price Price of beer (\$)

price_liquor price of other liquor (\$)

price_other price of other goods and services (an index)

income income (\$)

Source

<http://www.principlesofeconometrics.com/poe4/poe4stata.htm>

check_setup	<i>Check whether your setup is complete</i>
-------------	---

Description

Reports your ‘R’ version, the version of this package, and whether all packages needed by the exercises are installed. If something does not work, run this function and send its output to your lecturer.

Usage

check_setup()

Value

A logical value, invisibly: ‘TRUE’ if the setup is complete.

Examples

```
## Not run:  
check_setup()  
  
## End(Not run)
```

econjournals

Data on economics journals subscription fees

Description

A data set on how much libraries need to pay for subscription plans of journals in the field of economics.

Usage

econjournals

Format

A data frame with 327 rows and 10 variables

title Title of the journal

field The subfield of economics the journal is devoted to, e.g. ‘General’, ‘Finance’ or ‘Urban and Regional’.

publisher The name of the publisher

publisher_type Type of publisher: ‘profit’ or ‘nonprofit’

first_year First year in which journal was published in English language.

sub_price Price to university libraries for print and online access in the United States in 2004. Measured in nominal dollars.

pages_py Pages published in 2003

impact Journal impact factor in 2002

citations Total cites in 1999

papers Papers published in 2002

Source

<https://www.econ.ucsb.edu/~tedb/Journals/pricing.html>

EUFstudents

Artificial data on heights of EUF students in 2021

Description

A an artificial data set about the height of students from the Europa University Flensburg. The composition of the population was set according to official information about the student population in 2021, the heights were created artificially by drawing from a random distribution, which was calibrated for mean and standard deviation parameters of the German population.

Usage

EUFstudents

Format

A data frame with 6402 rows and 2 variables

Gender Gender of the person

Height Height of the person

Source

Official statistics on the student population of the Europa-Universität Flensburg, <https://www.uni-flensburg.de/die-universitaet/profil-der-euf/zahlen-und-fakten>

EUFstudentsemesters	<i>Study semester of the EUF students in 2021</i>
---------------------	---

Description

A data set on the distribution of study semesters for all students of the Europa-University Flensburg in the spring semester 2021.

Usage

EUFstudentsemesters

Format

A data frame with 5875 rows and 2 variables

Semester The study semester of the student

ID An ID for the student

Source

Official statistics on the student population of the Europa-Universität Flensburg, <https://www.uni-flensburg.de/die-universitaet/profil-der-euf/zahlen-und-fakten>

 gdplifexp2007

Data on GDP, life expectancy, and population size in 2007.

Description

A subset of the gapminder data on population size, life expectancy, and GDP per capita levels in 2007.

Usage

```
gdplifexp2007
```

Format

A data frame with 142 rows and 5 variables:

country Country name

continent Continent name

lifeExp Life expectancy at birth, in years

pop Population in 2007, in nb of people

gdpPercap GDP per capita, in 2005-international dollars

Source

<https://github.com/jennybc/gapminder>

 Growth

Annual GDP growth of Germany, 2015-2020

Description

A small data set on the annual growth rate of German real GDP. It is used in the exercises on data import, where it is also provided as the csv file ‘growth_data_1.csv’.

Usage

```
Growth
```

Format

A data frame with 6 rows and 3 variables

country Country name, here always ‘Germany’

Growth Annual growth rate of real GDP, in percent

year Year of the observation

Source

World Bank World Development Indicators, indicator ‘NY.GDP.MKTP.KD.ZG’, obtained via the ‘WDI’ package; see ‘data-raw/Growth.R’.

list_exercises	<i>List the available exercises</i>
----------------	-------------------------------------

Description

Lists all exercise collections that ship with this package.

Usage

```
list_exercises()
```

Value

A character vector of exercise names, invisibly. The names and their titles are printed to the console.

Examples

```
## Not run:
list_exercises()

## End(Not run)
```

nycflights21_small	<i>A reduced data set on flights from NYC in 2021</i>
--------------------	---

Description

A random sample of 50k flights (out of 336776) from NYC in 2021.

Usage

```
nycflights21_small
```

Format

A data frame with 50000 rows and 5 variables:

arr_delay Arrival delay, in minutes. Negative times represent early arrivals.

dep_delay Departure delay, in minutes. Negative times represent early departures.

month Month of departure.

carrier Two letter carrier abbreviation.

distance Distance between airports, in miles.

Source

<https://github.com/simonpcouch/anyflights>

pop_var

Compute the population variance of a vector

Description

Computes the population variance as defined by

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

where μ is the population mean ($\frac{1}{N} \sum_{i=1}^N x_i$) and N is the size of the population

Usage

```
pop_var(x)
```

Arguments

x A numerical vector

Value

The population variance of this vector

Examples

```
pop_var(c(2, 3, 4, 5))
pop_var(rnorm(20, 1, 0))
```

product_complexity

Product complexity and export shares of periphery countries

Description

A data set on the complexity of traded products and the share of these products in the exports of periphery countries. It is used in the exercises on data import, where it is also provided as the csv file ‘product_complexity.csv’.

Usage

```
product_complexity
```

Format

A data frame with 10 rows and 3 variables

commoditycode The code identifying the product according to the Harmonized System (HS)

product_complexity The complexity of the product, measured as its product complexity index (PCI)

exp_share The share of this product in total exports, in percent

Source

Own computations based on the BACI trade data of the CEPII; see ‘data-raw/product_complexity.R’.

regression_table	<i>Summarize a linear regression model in a concise table</i>
------------------	---

Description

Returns the estimated coefficients of a linear regression model together with their standard errors, test statistics, p-values and the bounds of the 95

Usage

```
regression_table(model, digits = 3, level = 0.95)
```

Arguments

model	A fitted model object, usually the result of a call to [stats::lm()].
digits	The number of digits to which the results are rounded.
level	The confidence level of the reported interval.

Details

This is a convenience function: the same information is available from ‘summary(model)’ and [stats::confint()], which you should also know how to use.

Value

A ‘data.frame’ with one row per estimated coefficient and the columns ‘term’, ‘estimate’, ‘std_error’, ‘statistic’, ‘p_value’, ‘lower_ci’ and ‘upper_ci’.

Examples

```
linmod <- lm(consumption ~ price_liquor, data = beer)
regression_table(linmod)
```

run_exercise	<i>Start an exercise</i>
--------------	--------------------------

Description

Starts one of the interactive exercise collections of this package in your web browser. This is a convenience wrapper around `[learnr::run_tutorial()]`: it fills in the package name and makes sure the exercise opens in a browser window rather than the RStudio viewer pane.

Usage

```
run_exercise(name, launch_browser = TRUE, ...)
```

Arguments

name	The name of the exercise, e.g. "Basics". Case does not matter. Call <code>[list_exercises()]</code> to see the available names.
launch_browser	Should the exercise open in your default web browser? 'TRUE' by default, which is usually what you want.
...	Further arguments passed on to <code>[learnr::run_tutorial()]</code> .

Details

To stop an exercise, press the red stop button in RStudio, or press 'Esc' in the console.

Value

Called for its side effect of starting the exercise.

Examples

```
## Not run:
run_exercise("Basics")

## End(Not run)
```

unemp_pop	<i>Population and labor force of Germany, 2015-2020</i>
-----------	---

Description

A small data set on the size of the German population and its labor force. It is used in the exercises on data import, where it is also provided as the csv file 'unemp_pop.csv'.

Usage

unemp_pop

Format

A data frame with 6 rows and 4 variables

country Country name, here always ‘Germany’

year Year of the observation

population Total population, in nb of people

labor_force Total labor force, in nb of people

Source

World Bank World Development Indicators, indicators ‘SP.POP.TOTL’ and ‘SL.TLF.TOTL.IN’, obtained via the ‘WDI’ package; see ‘data-raw/unemp_pop.R’.

wine2dine	<i>Data on wine properties</i>
-----------	--------------------------------

Description

A data set containing information about the alcohol and residual sugar contained in white and red wines, as well as their quality

Usage

wine2dine

Format

A data frame with 6497 rows and 4 variables:

alcohol The alcohol content of the wine

quality An quality score between 0 and 10, based on sensory data

residual_sugar The amount of sugar left in the wine

kind The kind of wine, i.e. red or white

Source

<https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/>

Index

* datasets

- aggGDPlifexp, [2](#)
- beer, [3](#)
- econjournals, [4](#)
- EUFstudents, [4](#)
- EUFstudentsemesters, [5](#)
- gdplifexp2007, [6](#)
- Growth, [6](#)
- nycflights21_small, [7](#)
- product_complexity, [8](#)
- unemp_pop, [10](#)
- wine2dine, [11](#)

aggGDPlifexp, [2](#)

beer, [3](#)

check_setup, [3](#)

econjournals, [4](#)

EUFstudents, [4](#)

EUFstudentsemesters, [5](#)

gdplifexp2007, [6](#)

Growth, [6](#)

list_exercises, [7](#)

nycflights21_small, [7](#)

pop_var, [8](#)

product_complexity, [8](#)

regression_table, [9](#)

run_exercise, [10](#)

unemp_pop, [10](#)

wine2dine, [11](#)