

Package ‘learningtower’

August 25, 2026

Title OECD PISA Datasets from 2000-2022 in an Easy-to-Use Format

Version 2.0.0

Description The Programme for International Student Assessment (PISA) is a global study conducted by the Organization for Economic Cooperation and Development (OECD) in member and non-member countries to assess educational systems by assessing 15-year-old school students academic performance in mathematics, science, and reading. This datasets contains information on their scores and other socioeconomic characteristics, information about their school and its infrastructure, as well as the countries that are taking part in the program.

Depends R ($\geq$ 3.5.0)

Encoding UTF-8

RoxygenNote 7.3.3

VignetteBuilder knitr

License MIT + file LICENSE

URL <https://kevinwang09.github.io/learningtower/>,
<https://github.com/kevinwang09/learningtower>

BugReports <https://github.com/kevinwang09/learningtower/issues>

Imports tibble, dplyr

Suggests testthat ($\geq$ 3.0.0), knitr, rmarkdown, ggplot2, forcats,
scales, digest, jsonlite, curl

Config/testthat/edition 3

LazyData true

LazyDataCompression xz

NeedsCompilation no

Author Kevin Wang [aut, cre],
Paul Yacobellis [aut],
Erika Siregar [aut],
Sarah Romanes [aut],
Kim Fitter [aut],
Giulio Valentino Dalla Riva [aut],

Dianne Cook [aut],
Nick Tierney [aut],
Priya Dingorkar [aut],
Shabarish Sai Subramanian [aut],
Guan Ru Chen [aut]

Maintainer Kevin Wang <kevinwangstats@gmail.com>

Repository CRAN

Date/Publication 2026-08-25 10:50:08 UTC

Contents

countrycode	2
load_student	3
school	3
student	4

Index	6
--------------	----------

countrycode	<i>Country iso3c and name mapping for PISA OECD countries participants.</i>
-------------	---

Description

A dataset containing mapping of the country ISO code to the country names. More information on participating countries can be found at <https://www.oecd.org/en/about/programmes/pisa/pisa-participants.html>.

Format

A tibble of the following variables

- country: Country 3 character code. Note that some regions/territories are coded as country for ease of input. Character.
- country_name: Country name. Note that some regions/territories are coded as country for ease of input. Character.

load_student	<i>load_student() function allows the user to extract the PISA student scores for any desired year from 2000-2022</i>
--------------	---

Description

load_student() function was created to extract the data of student's scores in any years from 2000-2022, the function requires any of the year as its argument or a string "all" that will return all the PISA scores of the students from the years 2000-2022.

Usage

```
load_student(year = "2000")
```

Arguments

year	is the required parameter for the function to display the dataset the user wants to view the PISA scores for the selected year else the entire student data will be available to the user
------	---

Value

A dataset of PISA scores of students that took the test in the selected year as per user from the years 2000-2018

Examples

```
## Not run:
library(learningtower)
student_all <- load_student("all")
student_2000 <- load_student("2000")

## End(Not run)
```

school	<i>Subset of the School data available for the years 2000-2022 from the PISA OECD database</i>
--------	--

Description

A subset data containing school weight and other information from the triennial testing of 15 year olds around the globe. Original data available from <https://www.oecd.org/en/about/programmes/pisa/pisa-data.html>.

Format

A tibble of the following variables

- year: Year of the PISA data. Integer.
- country: Country 3 character code. Note that some regions/territories are coded as country for ease of input. Character.
- school_id: The school identification number, unique for each country and year combination. Character.
- fund_gov: Percentage of total funding for school year from government. Numeric.
- fund_fees: Percentage of total funding for school year from student fees or school charges paid by parents. Numeric.
- fund_donation: Percentage of total funding for school year from benefactors, donations, bequests, sponsorship, parent fundraising. Numeric.
- enrol_boys: Number of boys in the school. Numeric.
- enrol_girls: Number of girls in the school. Numeric.
- stratio: Student-Teacher ratio. Numeric.
- public_private: Is the school a public or private school. Factor.
- staff_shortage: Shortage of staff. Numeric.
- sch_wgt: The final survey weight score for the schools. Numeric.
- school_size: The school size. Numeric.

student

Processed and Sampled PISA Student Data (2000-2022)

Description

This dataset provides a clean and processed subset of the OECD PISA student data for the years 2000-2022. The original data is sourced from <https://www.oecd.org/en/about/programmes/pisa/pisa-data.html> and has been prepared for analysis. A sampling of 50 students per country (for OECD countries) has been included for each year. The data curation and sampling process are documented in https://github.com/kevinwang09/learningtower_masonry/blob/master/Code/student_bind_rows.Rmd

Format

A tibble of the following variables

- year: Year of the PISA data. Integer.
- country: Country 3 character code. Note that some regions/territories are coded as "country" for ease of input. Factor.
- school_id: Unique school identifier for each country and year. Character.
- student_id: Unique student identifier within each school. Integer.

- `mother_educ`: Mother's highest level of education, from "less than ISCED1" to "ISCED 3A". Factor.
- `father_educ`: Father's highest level of education, from "less than ISCED1" to "ISCED 3A". Factor.
- `gender`: Gender of the student. Only "male" and "female" are recorded. Factor. Note that we call this variable `gender` and not `sex` as this term was used in the OECD PISA database.
- `computer`: Possession of computer. Only "yes" and "no" are recorded. Factor.
- `internet`: Access to internet. Only "yes" and "no" are recorded. Factor.
- `math`: Simulated score in mathematics. Numeric.
- `read`: Simulated score in reading. Numeric.
- `science`: Simulated score in science. Numeric.
- `stu_wgt`: The final survey weight score for the student score. Numeric.
- `desk`: Possession of desk to study at. Only "yes" and "no" are recorded. Factor.
- `room`: Possession of a room of your own. Only "yes" and "no" are recorded. Factor.
- `dishwasher`: Possession of a dishwasher. Only "yes" and "no" are recorded. Factor. Note that in 2015 and 2018, all entries are missing.
- `television`: Number of televisions. "0", "1", "2" are code for no, one and two TVs in the house. "3+" codes for three or more TVs. Factor. Note that in 2003, all entries are missing.
- `computer_n`: Number of computers. "0", "1", "2" are code for no, one and two computers in the house. "3+" codes for three or more computers. Factor. Note that in 2003, all entries are missing.
- `car`: Number of cars. "0", "1", "2" are code for no, one and two cars in the house. "3+" codes for three or more cars. Factor. Note that in 2003, all entries are missing.
- `book`: Number of books. Factor. Note that encoding is different in the years 2000 and 2003 compared to all other years. Factor. Evaluate `table(student$book, student$year)` for a demo.
- `wealth`: Index of family wealth. Numeric. Note that in 2003, all entries are missing.
- `escs`: Index of economic, social and cultural status. Numeric.

Examples

```
library(dplyr)
data(student_subset_2000)
data(student_subset_2003)
dplyr::bind_rows(
  student_subset_2000,
  student_subset_2003
)
```

Index

countrycode, [2](#)

load_student, [3](#)

school, [3](#)

student, [4](#)

student_subset_2000 (student), [4](#)

student_subset_2003 (student), [4](#)

student_subset_2006 (student), [4](#)

student_subset_2009 (student), [4](#)

student_subset_2012 (student), [4](#)

student_subset_2015 (student), [4](#)

student_subset_2018 (student), [4](#)

student_subset_2022 (student), [4](#)