

Package: academyDatasets (via r-universe)

June 6, 2026

Title Datasets for RStudio Academy

Version 0.4.1

Description A set of datasets for use with RStudio Academy exercises, tutorials, and recipes.

License file LICENSE

URL <https://github.com/rstudio/academyDatasets>

BugReports <https://github.com/rstudio/academyDatasets/issues>

Depends R (>= 2.10)

Imports tibble

Suggests docthis (>= 0.1.1), dplyr, knitr, skimr, tsibble, utils, vctrs

Remotes mdlincoln/docthis

Config/Needs/actions cynkra/fledge, rmarkdown, knitr, skimr

Encoding UTF-8

LazyData true

LazyDataCompression xz

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

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

Date/Publication 2023-05-31 16:39:50 UTC

RemoteUrl <https://github.com/rstudio/academyDatasets>

RemoteRef HEAD

RemoteSha 7cf749c0ff00dddfde87352b047fef97392683

Contents

aus_electricity	3
bitcoin	4
car_parts	5
covid	6
covid_state_pop	7
dmd	8
dominick	9
dominick_oatmeal	11
dominick_soap	12
elec_demand	13
electricity_weekly	14
encounters	15
fda_adverse_daily	16
fda_pt_drugs	17
flu	18
fred_md	19
hospital	20
mdrd	21
mdrd_supplemental	23
medications	24
nhanes_dermatology	25
nhanes_sleep	26
ny_air	27
oikolab_weather	28
pedestrian_counts	29
rideshare	30
riverflow	31
sdtm_adverse_events	32
sdtm_concomitant_meds	33
sdtm_demographics	34
sdtm_lab_results	35
sdtm_subject_visits	36
sdtm_vital_signs	37
solar	38
sunspots	39
tourists	40
us_births	41
vaccines	42

Index

43

aus_electricity	<i>Australian electricity demand data</i>
-----------------	---

Description

This dataset records half-hourly electricity demand for five states in Australia: Victoria, New South Wales, Queensland, Tasmania, and South Australia.

Usage

```
aus_electricity
aus_electricity_dictionary
```

Format

A tibble with 1155264 rows and 3 variables:

```
date [dtm] Starting date-time of demand reading
state [chr] State abbreviation: Victoria (VIC), New South Wales (NSW), Queensland (QLD),
Tasmania (TAS), and South Australia (SA)
demand [dbl] Half-hourly electricity demand in MW
```

An object of class `list` of length 3.

Functions

- `aus_electricity_dictionary`: The `aus_electricity` data dictionary

Source

<https://zenodo.org/record/4659727>

Examples

```
# Convert to `tsibble`
library(tsibble)
aus_electricity %>%
  as_tsibble(key = state, index = date)
```

bitcoin	<i>bitcoin</i>
---------	----------------

Description

This dataset contains the potential influencers of the bitcoin price. There are a total of 18 daily time-series including hash rate, block size, mining difficulty etc. It also encompasses public opinion in the form of tweets and google searches mentioning the keyword bitcoin. The data is scraped from the interactive web-graphs available at <https://bitinfocharts.com>.

For more details, please refer to BitInfoCharts, 2021. Cryptocurrency statistics. URL <https://bitinfocharts.com>

Usage

bitcoin

bitcoin_dictionary

Format

A tibble with 4581 rows and 19 variables:

timestamp [date] Day
 price [dbl] Price in US dollars (USD)
 difficulty [dbl] Mining difficulty
 sent_addresses [int] Number of addresses that sent Bitcoin
 send_usd [dbl] Amount of Bitcoin sent, in USD
 market_cap [dbl] Market value of all existing Bitcoin in USD
 confirmation_time [dbl] Time to record a transaction in block chain
 transactions [int] Number of blockchain transactions
 median_transaction_size [dbl] Median transaction size
 mining_profitability [dbl] Profit in USD/Day for 1 THash/s
 fee_reward [dbl] Average fee percentage in total block reward
 top_100_percent [dbl] Percent of Bitcoin owned by the top 100 richest addresses
 median_transaction_value [dbl] Median transaction value in USD
 av_transaction_value [dbl] Average transaction value in USD
 block_size [dbl] Average (mined?) Bitcoin block size in kilobytes
 hashrate [dbl] Bitcoin hashrate in Ehash/s
 active_addresses [int] Number of unique (from or to) addresses per day
 google_trends [dbl] Google trends interest score for Bitcoin
 tweets [int] Tweets per day with the tag #Bitcoin

An object of class list of length 3.

Functions

- bitcoin_dictionary: The bitcoin data dictionary

Source

<https://forecastingdata.org/>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
bitcoin %>%  
  as_tsibble()
```

car_parts

Monthly Car Part Sales

Description

This dataset contains 2674 intermittent monthly time series that represent car parts sales from January 1998 to March 2002. It was extracted from R expsmooth package.

Usage

```
car_parts
```

```
car_parts_dictionary
```

Format

A tibble with 136374 rows and 3 variables:

part_num [chr] ID of the car part

date [date] Start date of the month

qty [int] Number of parts sold that month

An object of class list of length 3.

Functions

- car_parts_dictionary: The car_parts data dictionary

Source

<https://zenodo.org/record/4656022#.YSzxsNNKj60>

Examples

```
# Convert to `tsibble`
library(tsibble)
car_parts %>%
  as_tsibble(key = part_num, index = date)
```

covid

COVID-19 US Historical Data by State

Description

From [The COVID Tracking Project](#):

We collect, cross-check, and publish COVID-19 **data** from 56 US states and territories in three main areas: **testing, hospitalization, and patient outcomes, racial and ethnic demographic information** via [The COVID Racial Data Tracker](#), and **long-term-care facilities** via the [Long-Term-Care tracker](#). We compile these numbers to provide the most complete picture we can assemble of the US COVID-19 testing effort and the outbreak's effects on the people and communities it strikes.

If you'd like to use the data, whether it's for a specialized project or just to better understand COVID-19 in the US, here are a few things you should know right away.

- We update **the full dataset** each day between about 5:30pm and 7pm Eastern time, with limited additional updates as new information arrives.
- All our data comes from state and territory public health authorities or official statements from state officials. Not all states report all data, which means we can't, either. You can [read more about our data sources here](#).

Usage

covid

covid_dictionary

Format

A tibble with 20780 rows and 6 variables:

date [date] Date on which data was collected by The COVID Tracking Project.

state [fct] Two-letter abbreviation for the state or territory.

tests [dbl] Daily increase in totalTestResults, calculated from the previous day's value. (Original: totalTestResultsIncrease)

cases [dbl] The daily increase in API field positive, which measures Cases (confirmed plus probable) calculated based on the previous day's value. (Original: positiveIncrease)

hospitalizations [dbl] Daily increase in hospitalizedCumulative, calculated from the previous day's value. (Original: hospitalizedIncrease)

deaths [dbl] Daily increase in death, calculated from the previous day's value. (Original: deathIncrease)

An object of class `list` of length 3.

Functions

- covid_dictionary: The covid data dictionary

Source

<https://covidtracking.com/data/download/all-states-history.csv>

See Also

Other COVID-19 datasets: [covid_state_pop](#)

covid_state_pop	<i>COVID-19 US State Populations</i>
-----------------	--------------------------------------

Description

State populations as reported by the [The COVID Tracking Project](#).

Usage

```
covid_state_pop
```

Format

A tibble with 50 rows and 2 variables:

state [factor] Two-letter abbreviation for the state or territory.

population [int] The state population, as reported by the COVID-19 Tracking Project.

Source

<https://api.covidtracking.com/v2/states.json>

See Also

Other COVID-19 datasets: [covid](#)

`dmd`*Duchenne Muscular Dystrophy Dataset*

Description

From the data source page: "This dataset is from M. Percy, listed in Table 38 of DF Andrews and AM Herzberg: Data, New York: Springer-Verlag, 1985 and also available on StatLib. The 209 observations correspond to blood samples on 192 patients (17 patients have two samples in the dataset) collected in a project to develop a screening program for female relatives of boys with DMD. The program's goal was to inform a woman of her chances of being a carrier based on serum markers as well as her family pedigree. Another question of interest is whether age and season should be taken into account. Enzyme levels were measured in known carriers (75 samples) and in a group of non-carriers (134 samples). Note that the original observation numbers (within subject) on this dataset do not agree with replicates of hospital IDs, so they have been recomputed here. Another anomaly of the dataset is that 16 out of 17 subjects having two blood samples drawn had differing carrier status for the two observations. The first two serum markers, creatine kinase and hemopexin, are inexpensive to obtain, while the last two, pyruvate kinase and lactate dehydrogenase, are more expensive. It is of interest to measure how much pk and ld add toward predicting the carrier status. The importance of age and sample date is also of interest. Percy noted that the water supply for the lab changed during the study."

Usage

`dmd``dmd_dictionary`

Format

A tibble with 192 rows and 8 variables:

`hospid` [dbl] Hospital ID`age` [dbl] Age in Years`creatine_kinase` [dbl] Creatine Kinase`hemopexin` [dbl] Hemopexin`pyruvate_kinase` [dbl] Pyruvate Kinase`lactate_dehydrogenase` [dbl] Lactate Dehydrogenase`carrier` [dbl] Carrier of Duchenne Muscular Dystrophy`date` [date] Date of Study

An object of class `list` of length 3.

Functions

- `dmd_dictionary`: The `dmd` data dictionary

Source

<https://biostat.app.vumc.org/wiki/pub/Main/DataSets/dmd.html>

dominick

Dominick's Finer Foods department sales data

Description

From the [Chicago Booth Kilts Center for Marketing](#):

From 1989 to 1994, Chicago Booth and Dominick's Finer Foods entered into a partnership for store-level research into shelf management and pricing. Randomized experiments were conducted in more than 25 different categories throughout all stores in this 100-store chain. As a by-product of this research cooperation, approximately nine years of store-level data on the sales of more than 3,500 UPCs is available in this dataset. This data is unique for the breadth of its coverage and for the information available on retail margins.

The customer count file includes information about in-store traffic. The data is store specific and on a daily basis. The customer count data refers to the number of customers visiting the store and purchasing something. Also in the customer count file is a total dollar sales and total coupons redeemed figure, by DFF defined department. These figures are compiled daily from the register/scanner receipts.

Usage

dominick

dominick_dictionary

Format

A tibble with 279519 rows and 25 variables:

store [int] Store number
date [date] Date
custcoun [int] Number of customers
grocery [dbl] Non-specialty grocery sales in dollars
gm [dbl] General merchandising sales in dollars
dairy [dbl] Dairy sales in dollars
frozen [dbl] Frozen food sales in dollars
meat [dbl] Meat sales in dollars
fish [dbl] Fish sales in dollars
produce [dbl] Produce sales in dollars
saladbar [dbl] Salad bar sales in dollars
floral [dbl] Floral sales in dollars

deli [dbl] Deli sales in dollars
cheese [dbl] Cheese case sales in dollars
bakery [dbl] Bakery sales in dollars
pharmacy [dbl] Pharmacy sales in dollars
jewelry [dbl] Jewelry sales in dollars
cosmetic [dbl]
haba [dbl] Health and beauty aids sales in dollars
camera [dbl] Camera sales in dollars
photofin [dbl] Photo development sales in dollars
video [dbl] Video sales in dollars
beer [dbl] Beer sales in dollars
wine [dbl] Wine sales in dollars
spirits [dbl] Alcoholic spirits sales in dollars

An object of class `list` of length 3.

Functions

- `dominick_dictionary`: The dominick data dictionary

Source

<https://www.chicagobooth.edu/research/kilts/datasets/dominicks>

See Also

Other Dominick's datasets: [dominick_oatmeal](#), [dominick_soap](#)

Examples

```
# Convert to `tsibble`  
library(tsibble)  
dominick %>%  
  as_tsibble(key = c("store"), index = date)
```

`dominick_oatmeal`*Dominick's Finer Foods oatmeal sales data*

Description

From the [Chicago Booth Kilts Center for Marketing](#):

From 1989 to 1994, Chicago Booth and Dominick's Finer Foods entered into a partnership for store-level research into shelf management and pricing. Randomized experiments were conducted in more than 25 different categories throughout all stores in this 100-store chain. As a by-product of this research cooperation, approximately nine years of store-level data on the sales of more than 3,500 UPCs is available in this dataset. This data is unique for the breadth of its coverage and for the information available on retail margins.

The data contain a description of each UPC in a category and sales information at the store level for each UPC in a category. The information is stored on a weekly basis.

Note: This is historical data and the products are not for sale.

Usage

`dominick_oatmeal``dominick_oatmeal_dictionary`

Format

A tibble with 974069 rows and 7 variables:

`week` [date] Start date of the week

`store` [int] Store number

`product` [chr] Abbreviated product name

`size` [chr] Product size

`price` [dbl] Price per unit in dollars

`profit` [dbl] Profit per unit in dollars

`move` [int] Number of units sold during the week

An object of class `list` of length 3.

Functions

- `dominick_oatmeal_dictionary`: The `dominick_oatmeal` data dictionary

Source

<https://www.chicagobooth.edu/research/kilts/datasets/dominicks>

See Also

Other Dominick's datasets: [dominick_soap](#), [dominick](#)

Examples

```
# Convert to `tsibble`  
library(tsibble)  
dominick_oatmeal %>%  
  as_tsibble(key = c("store", "product", "size", "price"), index = week)
```

dominick_soap

Dominick's Finer Foods bath soap sales data

Description

From the [Chicago Booth Kilts Center for Marketing](#):

From 1989 to 1994, Chicago Booth and Dominick's Finer Foods entered into a partnership for store-level research into shelf management and pricing. Randomized experiments were conducted in more than 25 different categories throughout all stores in this 100-store chain. As a by-product of this research cooperation, approximately nine years of store-level data on the sales of more than 3,500 UPCs is available in this dataset. This data is unique for the breadth of its coverage and for the information available on retail margins.

The data contain a description of each UPC in a category and sales information at the store level for each UPC in a category. The information is stored on a weekly basis.

Note: This is historical data and the products are not for sale.

Usage

dominick_soap

dominick_soap_dictionary

Format

A tibble with 415833 rows and 7 variables:

week [date] Start date of the week
store [int] Store number
product [chr] Abbreviated product name
size [chr] Product size
price [dbl] Price per unit in dollars
profit [dbl] Profit per unit in dollars
move [int] Number of units sold during the week

An object of class `list` of length 3.

Functions

- dominick_soap_dictionary: The dominick_soap data dictionary

Source

<https://www.chicagobooth.edu/research/kilts/datasets/dominicks>

See Also

Other Dominick's datasets: [dominick_oatmeal](#), [dominick](#)

Examples

```
# Convert to `tsibble`  
library(tsibble)  
dominick_soap %>%  
  as_tsibble(key = c("store", "product", "size", "price"), index = week)
```

elec_demand

Victoria, Australia Electricity Demand Data

Description

Single time series representing the half hourly electricity demand (in gigawatts) for Victoria, Australia in 2014.

Usage

elec_demand

elec_demand_dictionary

Format

A tibble with 17520 rows and 2 variables:

timestamp [dtm] Datetime of observation

demand [dbl] Electricity demand (GW)

An object of class list of length 3.

Functions

- elec_demand_dictionary: The elec_demand data dictionary

Source

<https://zenodo.org/record/4656069#.YSe0vdNKiM8>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
elec_demand %>%  
  as_tsibble(index = timestamp)
```

electricity_weekly	<i>Weekly Electricity Consumption</i>
--------------------	---------------------------------------

Description

This dataset contains weekly aggregated electricity consumption for 321 clients in Portugal from 2012 to 2014.

Usage

```
electricity_weekly  
  
electricity_weekly_dictionary
```

Format

A tibble with 50076 rows and 3 variables:

```
client [chr] ID of the electric company client  
date [date] Date  
power [int] Weekly electricity consumption, in kilowatts (kW)
```

An object of class `list` of length 3.

Functions

- `electricity_weekly_dictionary`: The `electricity_weekly` data dictionary

Source

<https://zenodo.org/record/4656141#.YSkxU9NKiM8>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
electricity_weekly %>%  
  as_tsibble(key = client, index = date)
```

encounters

Synthea Synthetic Encounters data

Description

encounters describes simulated visits of a patient population with different characteristics of their visits

Usage

encounters

encounters_dictionary

Format

A `spec_tbl_df`:

`id` [chr] Primary Key. Unique Identifier of the encounter.

`start` [dtm] The date and time the encounter started.

`stop` [dtm] The date and time the encounter concluded.

`patient` [chr] Foreign key to the Patient.

`organization` [chr] Foreign key to the Organization.

`provider` [chr] Foreign key to the Provider.

`payer` [chr] Foreign key to the Payer.

`encounterclass` [chr] The class of the encounter, such as ambulatory, emergency, inpatient, wellness, or urgentcare

`code` [dbl] Encounter code from SNOMED-CT

`description` [chr] Description of the type of encounter.

`base_encounter_cost` [dbl] The base cost of the encounter, not including any line item costs related to medications, immunizations, procedures, or other services.

`total_claim_cost` [dbl] The total cost of the encounter, including all line items.

`payer_coverage` [dbl] The amount of cost covered by the Payer.

`reasoncode` [dbl] Diagnosis code from SNOMED-CT, only if this encounter targeted a specific condition.

`reasondescription` [chr] Description of the reason code.

An object of class `list` of length 3.

Functions

- `encounters_dictionary`: The encounters data dictionary

Source

<https://synthea.mitre.org/downloads>

See Also

Other Synthea Synthetic Patient Population data: [medications](#)

fda_adverse_daily *Daily counts of FDA drug adverse event reports*

Description

This data set contains daily counts of reports received by the Food and Drug Administration (FDA) from 2004 to 2020 regarding adverse events associated with the administration of drugs in medical settings. This data was collected from the [FDA Adverse Event Reporting System \(FAERS\)](#), and has been made available through [openFDA](#).

According to openFDA: "an adverse event is submitted to the FDA to report any undesirable experience associated with the use of a medical product in a patient. For drugs, this includes serious drug side effects, product use errors, product quality problems, and therapeutic failures for prescription or over-the-counter medicines and medicines administered to hospital patients or at outpatient infusion centers.

Reporting of adverse events by healthcare professionals and consumers is voluntary in the United States. FDA receives some adverse event reports directly from healthcare professionals (such as physicians, pharmacists, nurses and others) and consumers (such as patients, family members, lawyers and others). Healthcare professionals and consumers may also report adverse events to the products' manufacturers. If a manufacturer receives an adverse event report, it is normally required to send the report to FDA."

Usage

fda_adverse_daily

fda_adverse_daily_dictionary

Format

A tibble with 5968 rows and 3 variables:

receive_date [date] Date that the report was first received by FDA.

public [dbl] Number of reports that were submitted directly by a member of the public.

manufacturer [dbl] Number of reports that were submitted through a drug manufacturer.

An object of class list of length 3.

Functions

- fda_adverse_daily_dictionary: The fda_adverse_daily data dictionary

Source

<https://open.fda.gov/apis/drug/event/>

See Also

Other openFDA datasets: [fda_pt_drugs](#)

fda_pt_drugs

Patient and drug information for FDA drug adverse events

Description

This data set contains selected variables from a subset of FDA drug adverse event reports received by the FDA in January 2019.

Usage

fda_pt_drugs

fda_pt_drugs_dictionary

Format

A tibble with 5765 rows and 17 variables:

report_id [chr] The 8-digit Safety Report ID number, also known as the case report number or case ID. Can be used to identify or find a specific adverse event report.

receive_date [date] Date that the report was first received by FDA.

receipt_date [date] Date that the most recent information in the report was received by FDA.

country [chr] The name of the country where the adverse event occurred.

reporter [chr] Category of individual who submitted the report: physician, pharmacist, other health professional, lawyer or consumer/non-health professional.

age [dbl] Age of the patient when the adverse event first occurred.

sex [chr] The sex of the patient.

weight [dbl] The patient weight, in kilograms (kg).

drug [chr] Drug name. This may be the valid trade name of the product (e.g. "advil" or "aleve") or the generic name (e.g. "ibuprofen").

dosage [dbl] The number portion of a dosage; when combined with dosage_unit the complete dosage information is represented.

dosage_unit [chr] The drug dosage unit: kilograms (kg), grams (g), milligrams (mg) or micrograms (ug).

indication [chr] Indication for the drug's use.

drug_start_date [date] Date the patient began taking the drug.

drug_end_date [date] Date the patient stopped taking the drug.

serious [lg] A logical value indicating whether or not the adverse event was serious, i.e. resulted in death, a life threatening condition, hospitalization, disability, congenital anomaly, or some other serious condition.

reaction [chr] Patient reaction, as a term from the Medical Dictionary for Regulatory Activities, encoded in British English.

outcome [chr] Outcome of the patient reaction at the time of last observation: recovered, recovering, not recovered, recovered with sequelae (consequent health issues), fatal or unknown.

An object of class list of length 3.

Functions

- `fda_pt_drugs_dictionary`: The `fda_pt_drugs` data dictionary

Source

<https://open.fda.gov/apis/drug/event/>

See Also

Other openFDA datasets: [fda_adverse_daily](#)

flu

Protein Sequences of Influenza B Virus Strains

Description

A data set of proteins, their sequences, and other attributes of 15,091 unique strains of Influenza B.

Usage

`flu`

`flu_dictionary`

Format

A tibble with 130560 rows and 16 variables:

`protein` [chr] Abbreviation of viral protein

`sequence_accession` [chr] Unique identifier given to the protein sequence record to allow for tracking of different versions of that sequence.

`complete_genome` [chr] Is the viral strain's complete genome known?

`complete_sequence` [chr] Is the complete sequence of this viral protein known?

`segment` [dbl] One of eight single-stranded RNA segments that encodes the viral protein

segment_length [dbl] Number of RNA nucleotides in segment
 collection_date [date] Date of sample collection
 host_species [chr] Species that the viral strain infects
 country [chr] Country of strain origin
 state_province [chr] State or province of origin if applicable
 geographic_grouping [chr] Geographic origin of viral strain
 flu_season [chr] For geographic regions in the northern hemisphere, the two digit year for the fall and winter season when the strain was recorded.
 strain_name [chr] Name of viral strain
 sequence [chr] Protein sequence in amino acid
 submission_date [date] Date of entry submission
 passage_history [chr] An indicator of what cell line was used for culturing the virus. Nomenclature for passage history is notoriously unstandardized. See <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6599686/>
 An object of class list of length 3.

Functions

- flu_dictionary: The flu data dictionary

Source

<https://FAIRsharing.org>: IRD; Influenza Research Database; DOI: <https://doi.org/10.25504/FAIRsharing.ws7cgw>; Last edited: March 30, 2020, 1:15 p.m.; Last accessed: May 27 2021 4:25 p.m.

References

Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S, Larsen CN, Lee AJ, Li X, Macken C, Mahaffey C, Pickett BE, Reardon B, Smith T, Stewart L, Suloway C, Sun G, Tong L, Vincent AL, Walters B, Zaremba S, Zhao H, Zhou L, Zmasek C, Klem EB, Scheuermann RH; Nucleic Acids Res ; 2016; 10.1093/nar/gkw857 ;

fred_md

U.S. macro-economic indicators from the FRED-MD database.

Description

A dataset containing six macro-economic indicators tracked by the Federal Reserve Bank, extracted from the FRED-MD database.

For more information on a variable, look it up by name here: <https://fred.stlouisfed.org/>

Usage

fred_md

fred_md_dictionary

Format

A tibble with 751 rows and 7 variables:

date [date] Date

rpi [dbl] Real personal income, in billions of dollars

hwi [int] Help-wanted index: the number of help-wanted ads in major newspapers

unrate [dbl] Civilian unemployment rate (percent)

ce16ov [dbl] Thousands of employed civilians

houst [int] Total number of new privately owned houses

cpiaucs1 [dbl] Consumer price index (all items)

An object of class `list` of length 3.**Functions**

- `fred_md_dictionary`: The `fred_md` data dictionary

Source

<https://research.stlouisfed.org/econ/mccracken/fred-databases/>

Examples

```
# Convert to `tsibble`
library(dplyr)
library(tsibble)
fred_md %>%
  mutate(date = yearmonth(date)) %>%
  as_tsibble(index = date)
```

 hospital

Monthly counts of patients use of medical products

Description

Monthly patient count for products that are related to medical problems. There are 767 time series that had a mean count of at least 10 and no zeros.

Usage

hospital

hospital_dictionary

Format

A tibble with 64428 rows and 5 variables:

sku [chr] Hospital stock-keeping unit (SKU) code, representing a specific medical product

entity_code [chr] Code related to medical product for use with medical billing and insurance purposes

month [int] Month of interest

year [int] Year of interest

patient_counts [int] Number of patients who received the medical product

An object of class list of length 3.

Functions

- hospital_dictionary: The hospital data dictionary

Source

<https://robjhyndman.com/expsmooth/>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
library(dplyr)  
hospital %>%  
  mutate(  
    date = yearmonth(paste(year, month, sep = "-")),  
    .keep = "unused"  
  ) %>%  
  as_tsibble(key = c(sku, entity_code), index = date)
```

Description

From **NIDDK**:

The Modification of Diet in Renal Disease (MDRD) study consisted of two randomized clinical trials that investigated whether protein restriction and control of blood pressure had an effect on the progression of chronic kidney disease (CKD). The study tested two hypotheses—that (1) a reduction in dietary protein and phosphorous intake and (2) the maintenance of blood pressure at a level below that usually recommended safely and effectively delays the progression of CKD.

Our data is from Study 2, which included patients with relatively advanced renal disease (GFR between 13 and 24 ml/min). From **NIDDK**:

In study 2, 255 patients with GFR of 13 to 24 ml/min/1.73 m² were randomly assigned to the low-protein diet (0.58 g per kilogram per day) or a very-low-protein diet (0.28 g per kilogram per day) with a keto acid-amino acid supplement, and a usual- or a low-blood-pressure group (same values as those in study 1). The length of follow-up varied from 18-to-45-months, with monthly evaluations of the patients. The primary outcome was the change in GFR rate over time.

Dropout:

Many patients dropped out of the study before completion. Whether or not a patient dropped out is captured in the dropout variable. Reasons for dropout included dialysis, kidney transplant, death, and other medical reasons.

Usage

```
mdrd
```

```
mdrd_dictionary
```

Format

A tibble with 1988 rows and 10 variables:

ptid [dbl] Patient identifier

gfr [dbl] Glomerular filtration rate in milliliters per minute. A measure of how much blood the kidneys filter per minute.

months [dbl] Number of months after the start of the study that the measurement was taken.

diet1_normbp [dbl] Was the participant assigned to the low-protein, normal-blood pressure diet? (0 = No, 1 = Yes)

diet1_lowbp [dbl] Was the participant assigned to the low-protein, low-blood pressure diet? (0 = No, 1 = Yes)

dietk_normbp [dbl] Was the participant assigned to the very low-protein, normal-blood pressure diet? (0 = No, 1 = Yes)

dietk_lowbp [dbl] Was the participant assigned to the very low-protein, low-blood pressure diet? (0 = No, 1 = Yes)

log_protein [dbl] Logarithm of the grams of protein consumed per day.

followupmonths [dbl] Number of months until patient follow-up.

dropout [dbl] Did the patient drop out of the study? (0 = No, 1 = Yes)

An object of class list of length 3.

Functions

- `mdrd_dictionary`: The mdrd data dictionary

Source

<https://repository.niddk.nih.gov/studies/mdrd/>

See Also

Other MDRD datasets: [mdrd_supplemental](#)

mdrd_supplemental	<i>Modification of Diet In Renal Disease - Supplemental Data</i>
-------------------	--

Description

Supplemental data for the [mdrd](#) dataset. Note: this data is simulated and is not from the original MDRD study.

Usage

```
mdrd_supplemental
```

```
mdrd_supplemental_dictionary
```

Format

A tibble with 255 rows and 5 variables:

```
ptid [dbl] Patient identifier
```

```
sex [chr] Sex
```

```
age [dbl] Age (years)
```

```
height [dbl] Height (meters)
```

```
weight [dbl] Weight (kilograms)
```

An object of class `list` of length 3.

Functions

- `mdrd_supplemental_dictionary`: The mdrd_supplemental data dictionary

Source

<https://repository.niddk.nih.gov/studies/mdrd/>

See Also

Other MDRD datasets: [mdrd](#)

medications

Synthea Synthetic Medications data

Description

medications describes simulated medication history of the patient population

Usage

medications

medications_dictionary

Format

A `spec_tbl_df`:

`start` [dtm] The date and time the medication was prescribed.

`stop` [dtm] The date and time the prescription ended, if applicable.

`patient` [chr] Foreign key to the Patient.

`payer` [chr] Foreign key to the Payer.

`encounter` [chr] Foreign key to the Encounter where the medication was prescribed.

`code` [dbl] Medication code from RxNorm.

`description` [chr] Description of the medication.

`base_cost` [dbl] The line item cost of the medication.

`payer_coverage` [dbl] The amount covered or reimbursed by the Payer.

`dispenses` [dbl] The number of times the prescription was filled.

`totalcost` [dbl] The total cost of the prescription, including all dispenses.

`reasoncode` [dbl] Diagnosis code from SNOMED-CT specifying why this medication was prescribed.

`reasondescription` [chr] Description of the reason code.

An object of class `list` of length 3.

Functions

- `medications_dictionary`: The medications data dictionary

Source

<https://synthea.mitre.org/downloads>

See Also

Other Synthea Synthetic Patient Population data: [encounters](#)

nhanes_dermatology	<i>National Health and Nutrition Examination Survey, Dermatology, 2017-2018</i>
--------------------	---

Description

From the **NCHS**:

The National Center for Health Statistics (NCHS), Division of Health and Nutrition Examination Surveys (DHANES), part of the Centers for Disease Control and Prevention (CDC), has conducted a series of health and nutrition surveys since the early 1960's. The National Health and Nutrition Examination Surveys (NHANES) were conducted on a periodic basis from 1971 to 1994. In 1999, NHANES became continuous. Every year, approximately 5,000 individuals of all ages are interviewed in their homes and complete the health examination component of the survey. The health examination is conducted in a mobile examination center (MEC); the MEC provides an ideal setting for the collection of high quality data in a standardized environment.

The dermatology questionnaire section provides personal interview data on sun exposure and sun protective behavior.

Usage

nhanes_dermatology

nhanes_dermatology_dictionary

Format

A tibble with 3419 rows and 8 variables:

seq_no [dbl] Respondent sequence number

sun_reaction [fct] If after several months of not being in the sun, you then went out in the sun without sunscreen or protective clothing for a half hour, which one of these would happen to your skin?

shade [fct] When you go outside on a very sunny day, for more than one hour, how often do you stay in the shade?

long_sleeves [fct] When you go outside on a very sunny day, for more than one hour, how often do you wear a long sleeved shirt?

sunscreen [fct] When you go outside on a very sunny day, for more than one hour, how often do you use sunscreen?

sunburns [dbl] How many times in the past year have you had a sunburn?

time_outdoors_workday [dbl] During the past 30 days, how much time did you usually spend outdoors between 9 in the morning and 5 in the afternoon on the days that you worked or went to school?

time_outdoors_weekend [dbl] During the past 30 days, how much time did you usually spend outdoors between 9 in the morning and 5 in the afternoon on the days when you were not working or going to school?

An object of class list of length 3.

Functions

- `nhanes_dermatology_dictionary`: The `nhanes_dermatology` data dictionary

Source

<https://www.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire&Cycle=2017-2018>

See Also

Other NHANES datasets: [nhanes_sleep](#)

nhanes_sleep	<i>National Health and Nutrition Examination Survey, Sleep Disorders, 2017-2018</i>
--------------	---

Description

From the **NCHS**:

The National Center for Health Statistics (NCHS), Division of Health and Nutrition Examination Surveys (DHANES), part of the Centers for Disease Control and Prevention (CDC), has conducted a series of health and nutrition surveys since the early 1960's. The National Health and Nutrition Examination Surveys (NHANES) were conducted on a periodic basis from 1971 to 1994. In 1999, NHANES became continuous. Every year, approximately 5,000 individuals of all ages are interviewed in their homes and complete the health examination component of the survey. The health examination is conducted in a mobile examination center (MEC); the MEC provides an ideal setting for the collection of high quality data in a standardized environment.

The sleep disorders (variable name prefix SLQ) data set has questions on sleep habits and disorders.

Usage

`nhanes_sleep`

`nhanes_sleep_dictionary`

Format

A tibble with 6161 rows and 11 variables:

`seq_no` [dbl] Respondent sequence number

`sleep_time_workday` [time] What time do you usually fall asleep on weekdays or workdays?

`wake_time_workday` [time] What time do you usually wake up on weekdays or workdays?

`sleep_length_workday` [dbl] Number of hours usually sleep on weekdays or workdays.

`sleep_time_weekend` [time] What time do you usually fall asleep on weekends or non-workdays?

`wake_time_weekend` [time] What time do you usually wake up on weekends or non-workdays?

sleep_length_weekend [dbl] Number of hours usually sleep on weekends or non-workdays.
 snore [fct] In the past 12 months, how often did you snore while you were sleeping?
 stop_breathing [fct] In the past 12 months, how often did you snort, gasp, or stop breathing while you were asleep?
 told_doctor [lgl] Have you ever told a doctor or other health professional that you have trouble sleeping?
 overly_sleepy [fct] In the past month, how often did you feel excessively or overly sleepy during the day?

An object of class `list` of length 3.

Functions

- `nhanes_sleep_dictionary`: The `nhanes_sleep` data dictionary

Source

<https://www.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire&Cycle=2017-2018>

See Also

Other NHANES datasets: [nhanes_dermatology](#)

ny_air

Daily predicted New York air quality

Description

Daily predictions of PM2.5 (inhalable particulate matter) concentrations for New York state in 2016, by county.

Usage

`ny_air`

`ny_air_dictionary`

Format

A tibble with 22692 rows and 4 variables:

`date` [date] Date

`county` [dbl] County

`pm25_max` [dbl] Maximum predicted value of PM2.5 concentration

`pm25_median` [dbl] Median predicted value of PM2.5 concentration

An object of class `list` of length 3.

Functions

- `ny_air_dictionary`: The `ny_air` data dictionary

Source

<https://data.cdc.gov/Environmental-Health-Toxicology/Daily-PM2-5-Concentrations-All-County-2001-2020-7vdq-ztk9>

`oikolab_weather`

Monash University Weather Data

Description

Eight time series representing the hourly climate data nearby Monash University, Clayton, Victoria, Australia from 2010-01-01 to 2021-05-31.

Usage

`oikolab_weather`

`oikolab_weather_dictionary`

Format

A tibble with 100057 rows and 9 variables:

`timestamp` [dtm] Datetime of observation

`temperature` [dbl] temperature (C)

`dewpoint_temperature` [dbl] dewpoint temperature (C)

`wind_speed` [dbl] wind speed (m/s)

`mean_sea_level_pressure` [dbl] mean sea level pressure (Pa)

`relative_humidity` [dbl] relative humidity (0-1)

`surface_solar_radiation` [dbl] surface solar radiation (W/m²)

`surface_thermal_radiation` [dbl] surface thermal radiation (W/m²)

`total_cloud_cover` [dbl] total cloud cover (0-1)

An object of class `list` of length 3.

Functions

- `oikolab_weather_dictionary`: The `oikolab_weather` data dictionary

Source

<https://zenodo.org/record/5184708#.YSVNgNNKiM9>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
oikolab_weather %>%  
  as_tsibble(index = timestamp)
```

pedestrian_counts	<i>Melbourne pedestrian count data</i>
-------------------	--

Description

This dataset records hourly pedestrian counts captured from 66 sensors in Melbourne city, starting from May 2009.

Usage

```
pedestrian_counts  
  
pedestrian_counts_dictionary
```

Format

A tibble with 3132346 rows and 3 variables:

```
date [dtm] Date-time of sensor reading  
sensor_id [chr] Sensor ID  
ped_count [int] Hourly count of pedestrians
```

An object of class `list` of length 3.

Functions

- `pedestrian_counts_dictionary`: The `pedestrian_counts` data dictionary

Source

<https://zenodo.org/record/4656626#.YSVj8FNKjUI>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
pedestrian_counts %>%  
  as_tsibble(key = sensor_id, index = date)
```

rideshare

Hourly Summaries of Rideshare Service Data

Description

This dataset contains various hourly time series representations of attributes related to Uber and Lyft rideshare services for various locations in New York between 2018-11-26 and 2018-12-18. For a given starting location, provider and service,

Usage

rideshare

rideshare_dictionary

Format

A tibble with 84396 rows and 19 variables:

source_location [chr] Starting point of the ride

provider_name [chr] Rideshare service provider

provider_service [chr] Provider-specific ride type identifier

timestamp [dtm] Hour

price_min [dbl] Minimum price estimate for rides in USD

price_mean [dbl] Mean price estimate for rides in USD

price_max [dbl] Maximum price estimate for rides in USD

distance_min [dbl] Minimum Distance between source and destination

distance_mean [dbl] Mean Distance between source and destination

distance_max [dbl] Maximum Distance between source and destination

surge_min [dbl] Minimum multiplier by which price was increased, default 1

surge_mean [dbl] Mean multiplier by which price was increased, default 1

surge_max [dbl] Maximum multiplier by which price was increased, default 1

api_calls [int] Number of API calls in the hour

temp [dbl] Temperature (F)

rain [dbl] Rain in the last hour (inches)

humidity [dbl] Humidity (%)

clouds [dbl] Cloud cover (0-1)

wind [dbl] Wind speed (mph)

An object of class `list` of length 3.

Functions

- `rideshare_dictionary`: The rideshare data dictionary

Source

<https://zenodo.org/record/5122114#.YS0--9NKiM9>

Examples

```
# Convert to `tsibble`
library(tsibble)
rideshare %>%
  as_tsibble(key = source_location:provider_service, index = timestamp)
```

riverflow	<i>Daily flow rate of the Saugeen River</i>
-----------	---

Description

This dataset contains a time series representing the daily mean flow in cubic meters per second of the Saugeen River in Ontario, Canada from 01/01/1915 to 12/31/1979.

Usage

```
riverflow

riverflow_dictionary
```

Format

A tibble with 23741 rows and 2 variables:

```
date [date] Date of observation
flow_rate [dbl] Volumetric flow rate, in cubic meters per second
```

An object of class `list` of length 3.

Functions

- `riverflow_dictionary`: The riverflow data dictionary

Source

<https://zenodo.org/record/4656058#.YSY4VtNKjJJ>

Examples

```
# Convert to `tsibble`
library(tsibble)
as_tsibble(riverflow)
```

sdm_adverse_events *SDTM Adverse Events data*

Description

sdm_adverse_events describes adverse event data related to a simulated clinical trial. An adverse event is an undesirable medical occurrence that happens while a subject is enrolled in a clinical trial. These occurrences must be reported and analyzed to see if they are caused by the treatment under study.

Usage

sdm_adverse_events

sdm_adverse_events_dictionary

Format

A spec_tbl_df:

USUBJID [chr] Unique Subject Identifier

AESEQ [dbl] Sequence Number

AESTDT [date] Start date of the adverse event

AESTDY [dbl] The number of days the subject had been enrolled in the study when the adverse event began.

AEENDT [date] End date of the adverse event

AEENDY [dbl] The number of days the subject had been enrolled in the study when the adverse event ended.

AETERM [chr] The reported term for the adverse event, i.e. how the reporter described the adverse event.

AEDECOD [chr] The official term for the adverse event, i.e. the dictionary derived description of the adverse event.

AEBODSYS [chr] Body system or organ class involved in the adverse event.

AESER [chr] Was the adverse event serious? Y = Yes, N = No.

AEONGO [chr] Is the adverse event ongoing? Y = Yes, N = No.

AESEV [chr] Severity of the adverse event.

AEREL [chr] The investigator's opinion as to whether or not the adverse event was related to the study treatment.

AEOUT [chr] Outcome of the adverse event.

An object of class list of length 3.

Functions

- sdm_adverse_events_dictionary: The sdm_adverse_events data dictionary

Source

<https://rhoinc.github.io/data-library/>

See Also

Other Rho SDTM datasets: [sdm_concomitant_meds](#), [sdm_demographics](#), [sdm_lab_results](#), [sdm_subject_visits](#), [sdm_vital_signs](#)

sdm_concomitant_meds *SDTM Concomitant Medication data*

Description

sdm_concomitant_meds describes concomitant medication data related to a simulated clinical trial. Concomitant medications are medications that subjects take during the trial, concomitant with the treatment being studied. Concomitant medications are not part of the treatment being studied but may be confounding variables, or cause interaction effects or adverse events.

Usage

```
sdm_concomitant_meds
```

```
sdm_concomitant_meds_dictionary
```

Format

A tibble with 307 rows and 12 variables:

USUBJID [chr] Unique Subject Identifier

CMSEQ [dbl] Sequence Number

CMSTDT [date] Start date of the concomitant medication

CMSTDY [dbl] The number of days the subject had been enrolled in the study when the concomitant medication began.

CMENDT [date] End date of the concomitant medication

CMENDY [dbl] The number of days the subject had been enrolled in the study when the concomitant medication ended.

CMTRT [chr] Reported name of the concomitant medication, i.e concomitant treatment

PREFTERM [chr] The nonproprietary, i.e. generic, name of the concomitant medication

ATCTEXT2 [chr] Anatomical Therapeutic Chemical (ATC) of the concomitant medication

CMONGO [chr] Is the concomitant medication ongoing? Y = Yes, N = No

CMDOSE [dbl] Amount of concomitant medication taken per administration (dose)

CMROUTE [chr] Route of administration of the concomitant medication

An object of class `list` of length 3.

Functions

- `sdm_concomitant_meds_dictionary`: The `sdm_concomitant_meds` data dictionary

Source

<https://rhoinc.github.io/data-library/>

See Also

Other Rho SDTM datasets: [sdm_adverse_events](#), [sdm_demographics](#), [sdm_lab_results](#), [sdm_subject_visits](#), [sdm_vital_signs](#)

<code>sdm_demographics</code>	<i>SDTM Demographics data</i>
-------------------------------	-------------------------------

Description

`sdm_demographics` describes demographic data related to subjects of a simulated clinical trial.

Usage

```
sdm_demographics
```

```
sdm_demographics_dictionary
```

Format

A `spec_tbl_df`:

USUBJID [chr] Unique Subject Identifier

SITE [chr] Name of study site

SITEID [chr] Study Site Identifier

AGE [dbl] Age of subject

SEX [chr] Sex of subject

RACE [chr] Race of subject

ARM [chr] Treatment arm that subject is assigned to

ARMCD [chr] Code for treatment arm that subject is assigned to

SBJTSTAT [chr] Subject status in study

RFSTDTC [date] Reference start date. Usually the day the subject begins treatment.

RFENDTC [date] Reference end date. Usually the day the subject takes their last treatment.

RFENDY [dbl] The number of days after the RFSTDTC that the RFENDTC occurred, i.e. the number of days the subject spent in the study/in treatment

SAFFL [chr] Safety population flag. Did the subject actually receive a treatment (including placebo) and should therefore be included in the population of subjects studied to determine the safety of the treatment? Y = Yes and N = No

SAFFN [dbl] Numeric coding of safety population flag. Y = Yes and N = No

An object of class list of length 3.

Functions

- sdm_demographics_dictionary: The sdm_demographics data dictionary

Source

<https://rhoinc.github.io/data-library/>

See Also

Other Rho SDTM datasets: [sdm_adverse_events](#), [sdm_concomitant_meds](#), [sdm_lab_results](#), [sdm_subject_visits](#), [sdm_vital_signs](#)

sdm_lab_results	<i>SDTM Lab Test Results data</i>
-----------------	-----------------------------------

Description

sdm_lab_results describes the results of lab tests performed on subjects in a simulated clinical trial.

Usage

```
sdm_lab_results
```

```
sdm_lab_results_dictionary
```

Format

A spec_tbl_df:

USUBJID [chr] Unique Subject Identifier

VISIT [chr] Protocol defined text description of the visit

VISITNUM [dbl] Visit number

LBDT [date] Date of specimen collection

LB DY [dbl] Study day of specimen collection

LBCAT [chr] Category of the lab test performed

LBTEST [chr] Name of the lab test performed

LBSTRESU [chr] Standard units for lab test result (LBSTRESN)

LBSTRESN [dbl] Lab test result in standard units
LBSTNRLO [dbl] Lower limit of normal range lab test result
LBSTNRHI [dbl] Upper (high) limit of normal range for lab test result

An object of class `list` of length 3.

Functions

- `sdm_lab_results_dictionary`: The `sdm_lab_results` data dictionary

Source

<https://rhoinc.github.io/data-library/>

See Also

Other Rho SDTM datasets: [sdm_adverse_events](#), [sdm_concomitant_meds](#), [sdm_demographics](#), [sdm_subject_visits](#), [sdm_vital_signs](#)

`sdm_subject_visits` *SDTM Subject Visits data*

Description

`sdm_subject_visits` describes the clinical visits of subjects in a simulated clinical trial.

Usage

```
sdm_subject_visits
```

```
sdm_subject_visits_dictionary
```

Format

A `spec_tbl_df`:

USUBJID [chr] Unique Subject Identifier

VISIT [chr] Protocol defined text description of the visit

VISITNUM [dbl] Visit number

SVDT [date] Subject visit date

SVDY [dbl] Study day of subject visit

SVSTATUS [chr] Status of subject visit: Completed, Terminated, Missed, Expected, Overdue, Failed

An object of class `list` of length 3.

Functions

- `sdm_subject_visits_dictionary`: The `sdm_subject_visits` data dictionary

Source

<https://rhoinc.github.io/data-library/>

See Also

Other Rho SDTM datasets: [sdtm_adverse_events](#), [sdtm_concomitant_meds](#), [sdtm_demographics](#), [sdtm_lab_results](#), [sdtm_vital_signs](#)

sdtm_vital_signs	<i>SDTM Vital Signs data</i>
------------------	------------------------------

Description

sdtm_vital_signs describes the vital signs of subjects in a simulated clinical trial. Vital signs include things like heart rate, blood pressure, temperature, and so on.

Usage

```
sdtm_vital_signs
```

```
sdtm_vital_signs_dictionary
```

Format

A `spec_tbl_df`:

USUBJID [chr] Unique Subject Identifier

VISIT [chr] Protocol defined text description of the visit

VISITNUM [dbl] Visit number

VSDT [date] Date vital signs were collected

VSDY [dbl] Day of study when vital signs were collected

VSCAT [chr] Category of vital signs test

VSTEST [chr] Name of vital signs test

VSSTRESU [chr] Standard units of vital signs test result

VSSTRESN [dbl] Result of vital signs test in standard units

VSSTNRLO [dbl] Lower limit of normal range for result of vital signs test

VSSTNRHI [dbl] Upper (high) limit of normal range for result of vital signs test

An object of class `list` of length 3.

Functions

- `sdtm_vital_signs_dictionary`: The `sdtm_vital_signs` data dictionary

Source

<https://rhoinc.github.io/data-library/>

See Also

Other Rho SDTM datasets: [sdtm_adverse_events](#), [sdtm_concomitant_meds](#), [sdtm_demographics](#), [sdtm_lab_results](#), [sdtm_subject_visits](#)

solar

Solar Power

Description

This dataset contains weekly solar power production records from 137 photovoltaic (PV) power plants in Alabama, in 2006. The data were originally collected by the National Renewable Energy Laboratory (NREL): <https://www.nrel.gov/>

Usage

```
solar
```

```
solar_dictionary
```

Format

A tibble with 7124 rows and 3 variables:

plant [chr] ID of the photovoltaic (PV) power plant

date [date] Date

power [dbl] Solar power capacity, in megawatts (MW)

An object of class `list` of length 3.

Functions

- `solar_dictionary`: The solar data dictionary

Source

<https://zenodo.org/record/4656151#.YSktINNKjSU>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
solar %>%  
  as_tsibble(key = plant, index = date)
```

sunspots	<i>Daily historical sunspot data</i>
----------	--------------------------------------

Description

A dataset containing daily sunspot counts, from January 08, 1818 to May 31, 2020.

Usage

```
sunspots
```

```
sunspots_dictionary
```

Format

A tibble with 73924 rows and 2 variables:

```
date [date] Date of observation
```

```
sunspots [int] Observed number of sunspots
```

An object of class `list` of length 3.

Functions

- `sunspots_dictionary`: The sunspots data dictionary

Source

<https://zenodo.org/record/4654773#.YSV1vdNKg8M>

Examples

```
# Convert to `tsibble`  
tsibble::as_tsibble(sunspots, key = NULL, index = date)
```

tourists

Kaggle Tourism forecasting competition time series

Description

From [Athanasopoulos, Hyndman, Song, and Wu \(2010\)](#)

The data include 366 monthly series. They were supplied by both tourism bodies (such as Tourism Australia, the Hong Kong Tourism Board and Tourism New Zealand) and various academics, who had used them in previous tourism forecasting studies.

In order to adhere to all confidentiality agreements with all parties, the data are presented under coded titles.

City and countries were fabricated and assigned to the data. This data does not reflect true tourism trends for these cities.

Usage

tourists

tourists_dictionary

Format

A tibble with 109280 rows and 4 variables:

month [date] Start date of the month

city [chr] Fabricated city

country [chr] Fabricated country

tourists [int] Number of tourists who visited the location in the specified month

An object of class `list` of length 3.

Functions

- `tourists_dictionary`: The tourists data dictionary

Source

<https://zenodo.org/record/4656096>

Examples

```
# Convert to `tsibble`  
library(dplyr)  
library(tsibble)  
tourists %>%  
  mutate(month = yearmonth(month)) %>%  
  as_tsibble(key = c("city", "country"), index = month)
```

us_births

us_births

Description

This dataset contains a single very long daily time series representing the number of births in US from 01/01/1969 to 31/12/1988. It was extracted from R mosaicData package. The length of this time series is 7305.

For more details, please refer to Pruim, R., Kaplan, D., Horton, N., 2020. mosaicData: project mosaic data sets. R package version 0.18.0. <https://CRAN.R-project.org/package=mosaicData>

Usage

```
us_births
```

```
us_births_dictionary
```

Format

A tibble with 7305 rows and 2 variables:

```
timestamp [date] Day
```

```
births [int] Number of births in the US
```

An object of class list of length 3.

Functions

- `us_births_dictionary`: The `us_births` data dictionary

Source

<https://forecastingdata.org/>

Examples

```
# Convert to `tsibble`  
library(tsibble)  
us_births %>%  
  as_tsibble()
```

vaccines

COVID-19 Vaccine Distribution Allocations

Description

A dataset containing the number of COVID-19 vaccines allocated in the US by date, manufacturer, and state/territory.

Usage

vaccines

vaccines_dictionary

Format

A tibble with 3591 rows and 5 variables:

jurisdiction [chr] State or territory

vaccine [chr] Vaccine type

week [date] Week the vaccines were allocated

first_dose [dbl] Number of first doses allocated

second_dose [dbl] Number of second doses allocated

An object of class `list` of length 3.

Functions

- `vaccines_dictionary`: The vaccines data dictionary

Source

<https://data.cdc.gov/browse?category=Vaccinations>

Index

- * **COVID-19 datasets**
 - covid, 6
 - covid_state_pop, 7
- * **DMD datasets**
 - dmd, 8
- * **Dominick's datasets**
 - dominick, 9
 - dominick_oatmeal, 11
 - dominick_soap, 12
- * **MDRD datasets**
 - mdrd, 21
 - mdrd_supplemental, 23
- * **NHANES datasets**
 - nhanes_dermatology, 25
 - nhanes_sleep, 26
- * **Rho SDTM datasets**
 - sdtm_adverse_events, 32
 - sdtm_concomitant_meds, 33
 - sdtm_demographics, 34
 - sdtm_lab_results, 35
 - sdtm_subject_visits, 36
 - sdtm_vital_signs, 37
- * **Synthea Synthetic Patient Population data**
 - encounters, 15
 - medications, 24
- * **datasets**
 - aus_electricity, 3
 - bitcoin, 4
 - car_parts, 5
 - covid, 6
 - covid_state_pop, 7
 - dmd, 8
 - dominick, 9
 - dominick_oatmeal, 11
 - dominick_soap, 12
 - elec_demand, 13
 - electricity_weekly, 14
 - encounters, 15
 - fda_adverse_daily, 16
 - fda_pt_drugs, 17
 - flu, 18
 - fred_md, 19
 - hospital, 20
 - mdrd, 21
 - mdrd_supplemental, 23
 - medications, 24
 - nhanes_dermatology, 25
 - nhanes_sleep, 26
 - ny_air, 27
 - oikolab_weather, 28
 - pedestrian_counts, 29
 - rideshare, 30
 - riverflow, 31
 - sdtm_adverse_events, 32
 - sdtm_concomitant_meds, 33
 - sdtm_demographics, 34
 - sdtm_lab_results, 35
 - sdtm_subject_visits, 36
 - sdtm_vital_signs, 37
 - solar, 38
 - sunspots, 39
 - tourists, 40
 - us_births, 41
 - vaccines, 42
- * **flu datasets**
 - flu, 18
- * **hospital timeseries datasets**
 - hospital, 20
- * **openFDA datasets**
 - fda_adverse_daily, 16
 - fda_pt_drugs, 17
- aus_electricity, 3
- aus_electricity_dictionary (aus_electricity), 3
- bitcoin, 4
- bitcoin_dictionary (bitcoin), 4

- car_parts, 5
- car_parts_dictionary (car_parts), 5
- covid, 6, 7
- covid_dictionary (covid), 6
- covid_state_pop, 7, 7

- dmd, 8
- dmd_dictionary (dmd), 8
- dominick, 9, 12, 13
- dominick_dictionary (dominick), 9
- dominick_oatmeal, 10, 11, 13
- dominick_oatmeal_dictionary (dominick_oatmeal), 11
- dominick_soap, 10, 12, 12
- dominick_soap_dictionary (dominick_soap), 12

- elec_demand, 13
- elec_demand_dictionary (elec_demand), 13
- electricity_weekly, 14
- electricity_weekly_dictionary (electricity_weekly), 14
- encounters, 15, 24
- encounters_dictionary (encounters), 15

- fda_adverse_daily, 16, 18
- fda_adverse_daily_dictionary (fda_adverse_daily), 16
- fda_pt_drugs, 17, 17
- fda_pt_drugs_dictionary (fda_pt_drugs), 17
- flu, 18
- flu_dictionary (flu), 18
- fred_md, 19
- fred_md_dictionary (fred_md), 19

- hospital, 20
- hospital_dictionary (hospital), 20

- mdrd, 21, 23
- mdrd_dictionary (mdrd), 21
- mdrd_supplemental, 23, 23
- mdrd_supplemental_dictionary (mdrd_supplemental), 23
- medications, 16, 24
- medications_dictionary (medications), 24

- nhanes_dermatology, 25, 27
- nhanes_dermatology_dictionary (nhanes_dermatology), 25

- nhanes_sleep, 26, 26
- nhanes_sleep_dictionary (nhanes_sleep), 26
- ny_air, 27
- ny_air_dictionary (ny_air), 27

- oikolab_weather, 28
- oikolab_weather_dictionary (oikolab_weather), 28

- pedestrian_counts, 29
- pedestrian_counts_dictionary (pedestrian_counts), 29

- rideshare, 30
- rideshare_dictionary (rideshare), 30
- riverflow, 31
- riverflow_dictionary (riverflow), 31

- sdtm_adverse_events, 32, 34–38
- sdtm_adverse_events_dictionary (sdtm_adverse_events), 32
- sdtm_concomitant_meds, 33, 33, 35–38
- sdtm_concomitant_meds_dictionary (sdtm_concomitant_meds), 33
- sdtm_demographics, 33, 34, 34, 36–38
- sdtm_demographics_dictionary (sdtm_demographics), 34
- sdtm_lab_results, 33, 34, 35, 35, 37, 38
- sdtm_lab_results_dictionary (sdtm_lab_results), 35
- sdtm_subject_visits, 33–35, 36, 36, 38
- sdtm_subject_visits_dictionary (sdtm_subject_visits), 36
- sdtm_vital_signs, 33–36, 37, 37
- sdtm_vital_signs_dictionary (sdtm_vital_signs), 37
- solar, 38
- solar_dictionary (solar), 38
- sunspots, 39
- sunspots_dictionary (sunspots), 39

- tourists, 40
- tourists_dictionary (tourists), 40

- us_births, 41
- us_births_dictionary (us_births), 41

- vaccines, 42
- vaccines_dictionary (vaccines), 42