Package: stat1080r (via r-universe)

November 2, 2026

Type Package

Title Functions and datasets for STAT 1080: Foundations of Data Science

Version 0.1.0

Author Brennan Bean

Maintainer Brennan Bean <brennan.bean@usu.edu>

Description This package provides companion materials for the course STAT 1080: Foundations of Data Science. Materials include datasets for homework assignments, as well as helpful functions and datasets for in-class assignments. Users should expect regular updates to this package and should re-install it prior to completing any in-class or homework assignment.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Depends R (>= 3.5)

RoxygenNote 7.3.2

Repository https://beanb2.r-universe.dev

Date/Publication 2025-09-27 22:37:00 UTC

RemoteUrl https://github.com/beanb2/stat1080r

RemoteRef HEAD

RemoteSha 70058479e5a465d46925bd9afb1a27770bf91ffe

Contents

campus_crime_hawkes 2
cereal 3
cereal_infected 4
colleges 5
covid_case 6
covid_hos 6

covid_raw 7
crime_hawkes 8
election_survey_2016 8
finch_hawkes 9
housing_hawkes 10
mpg_hawkes 11
ncaa_times 12
patient_hawkes 12
pres_speech_df 13
rt_load_ut 13
satisfy_hawkes 14
swe 15
toluca 15

Index **17**

`campus_crime_hawkes` *Data regarding campus crime.*

Description

Data regarding campus crime.

Usage

```
campus_crime_hawkes
```

Format

A data.frame with 20 rows and 4 columns.

crimes The number of crimes as applicable to the school

police Number of police officers at the school

enrolled Total number of students enrolled at the school

private Private school (1), and others (0).

Source

https://www.hawkeslearning.com/Statistics/dataset-descriptions.html#campus-crime

`cereal` *Nutrition Information for Cereals*

Description

A dataset that provides nutritional information regarding cereals.

Usage

```
cereal
```

Format

A data.frame with 77 rows and 15 variables.

Shelf One of Top, Bottom, or Middle, describing the position of the cereal on the shelf at the grocery store.

Name The name of the cereal.

Manufacturer The manufacturer of the cereal, including American home food products (A), General Mills (G), Kelloggs (K), Nabisco (N), Post Brands (P), Quaker Oats (Q), and Ralston (R).

Type One of cold (C) or hot (H).

Calories The calories in one serving.

Protein The number of grams of protein in one serving.

Fat The number of grams of fat in one serving.

Sodium The number of milligrams (mg) of sodium in one serving.

Fiber The number of grams of fiber in one serving.

Carbohydrates The number of grams of carbohydrates in one serving.

Sugars The numebr of grams of sugar in one serving.

Vitamins vitamins and minerals - 0, 25, or 100, indicating the typical percentage of FDA

Weight The weight of one serving size of cereal.

Cup number of cups in one serving

Source

https://www.hawkeslearning.com/Statistics/beg3/datasets.html

https://www.kaggle.com/jeandsantos/breakfast-cereals-data-analysis-and-clustering

`cereal_infected` *Nutrition Information for Cereals*

Description

A dataset that provides nutritional information regarding cereals, but includes a nefarious twist.

Usage

```
cereal_infected
```

Format

A data.frame with 77 rows and 15 variables.

Shelf One of Top, Bottom, or Middle, describing the position of the cereal on the shelf at the grocery store.

Name The name of the cereal.

Manufacturer The manufacturer of the cereal, including American home food products (A), General Mills (G), Kelloggs (K), Nabisco (N), Post Brands (P), Quaker Oats (Q), and Ralston (R).

Type One of cold (C) or hot (H).

Calories The calories in one serving.

Protein The number of grams of protein in one serving.

Fat The number of grams of fat in one serving.

Sodium The number of milligrams (mg) of sodium in one serving.

Fiber The number of grams of fiber in one serving.

Carbohydrates The number of grams of carbohydrates in one serving.

Sugars The numebr of grams of sugar in one serving.

Vitamins vitamins and minerals - 0, 25, or 100, indicating the typical percentage of FDA

Weight The weight of one serving size of cereal.

Cup number of cups in one serving

Source

https://www.hawkeslearning.com/Statistics/beg3/datasets.html

https://www.kaggle.com/jeandsantos/breakfast-cereals-data-analysis-and-clustering

`colleges` *University Statistics*

Description

A subset of data from the U.S. Department of Education's College Score card. Contains information about institutions across the US. Institutions are only retained if they offer 4-year programs, have more than 2k students, have an endowment larger than 10 million dollars, and have non-missing information for all the variables included in the dataset. The data represent the most recent statistics that were compiled in May of 2025.

Usage

```
colleges
```

Format

A data frame with 831 rows and 10 columns.

name Name of the college or university

endowment_begin Size of the school endowment at the beginning of the fiscal year, in billions of dollars

endowment_end Size of the school endowment at the end of the fiscal year, in billions of dollars

enrollment Number of enrolled undergraduates

enrollment_men_prop Proportion of male undergraduates

enrollment_women_prop Proportion of female undergraduates

public Is the school a public institution?

attendance_cost Average cost of attendance, in thousands of dollars

income_median Median family income, in thousands of dollars

earnings Median earnings 10 years after enrollment, in thousands of dollars

Source

<https://collegescorecard.ed.gov/data/>

covid_case *Covid Case Data - Country Level Aggregates*

Description

A snapshot in time of COVID case level aggregates by country, as prepared by the World Health Organization.

Usage

```
covid_case
```

Format

A data frame with 241 rows and 6 columns.

name Country name

region World Health Organization region

case_total Cumulative number of cases

case_pop Cumulative number of cases per 100,000 people

death_total Cumulative number of deaths

death_pop Cumulative number of deaths per 100,000 people

Source

<https://data.who.int/dashboards/covid19/>

covid_hos *Covid Case Data - Country Level Time Series Data*

Description

Weekly covid case statistics by country, as reported by the World Health Organization.

Usage

```
covid_hos
```

Format

A data frame with 241 rows and 6 columns.

date Date of report, in yyyy-mm-dd format

country Country name

region World Health Organization region

new_hos Number of new hospitalizations reported that week.

Source

<https://data.who.int/dashboards/covid19/>

`covid_raw` *COVID-19 Case Count Data*

Description

COVID case count data taken from the world health organization. Data were scraped on 11-17-2023 and the updated tables from the website are subject to change.

Usage

```
covid_raw
```

Format

A data.frame with 333,459 rows rows and 8 variables.

Date_reported The date in y-m-d format.

Country_code Two letter country code.

Country Name of the country.

WHO_region World Health Organization region.

New_cases new COVID-19 cases reported on that day.

Cumulative_cases total number of cases reported since Jan 2020.

New_cases new COVID-19 deaths reported on that day.

Cumulative_cases total number of deaths reported since Jan 2020.

Source

https://covid19.who.int/WHO-COVID-19-global-data.csv

`crime_hawkes` *Data regarding high school completion and crime rates across U.S. states in 2014.*

Description

Data regarding high school completion and crime rates across U.S. states in 2014.

Usage

```
crime_hawkes
```

Format

A data.frame with 50 rows and 5 columns.

state U.S. State

completion The 4-year ACGR is the number of students who graduate in 4 years with a regular high school diploma divided by the number of students who form the adjusted cohort for the graduating class." This number has been rounded to the nearest whole number.

crime Rate of violent and property crimes per 100,000 people

violent Rate of violent crimes (murder, rape, robbery, aggravated assault) per 100,000 people

property Rate of property crimes per 100,000 people

Source

https://www.hawkeslearning.com/Statistics/dataset-descriptions.html#high-school-completion-and-crim

`election_survey_2016` *Post election survey of 7606 voters in the 2016 US Presidential Election.*

Description

Post election survey of 7606 voters in the 2016 US Presidential Election.

Usage

```
election_survey_2016
```

Format

A data.frame with 7606 rows and 6 columns.

id voter id

weight a weight based on the estimated representative-ness of the voter in relation to the overall population

vote_method either by mail or in person

vote either Clinton, Trump, or "Other" for all other candidates

income voter's self-identified income bracket

news voter's self-reported interest in the daily news

Source

https://www.kaggle.com/datasets/democracy-fund/2016-voter-survey?resource=download

---

`finch_hawkes` *A sample of data collected on Darwin's finches on Daphe Major island.*

Description

A sample of data collected on Darwin's finches on Daphe Major island.

Usage

```
finch_hawkes
```

Format

A data.frame with 100 rows and 11 columns.

band Refers to an individual's identity, more specifically, the number on a metal leg band it was given

species Species name

sex Male, female, or unknown. The reason for the "unknown" category is that males start their lives looking like females. After one or more years they molt into a plumage with some black feathering that indicates they are males

first_adult_year The year after the individual hatched from an egg

last_year The last year of that individual's life

weight Weight, in grams

wing Wing length, in millimeters

tarsus Tarsus length (part of leg), in millimeters

beak_length Beak length, in millimeters

beak_depth Beak depth, in millimeters

beak_width Beak width, in millimeters

Source

https://www.hawkeslearning.com/Statistics/dataset-descriptions.html#finch-data

---

housing_hawkes *Mount Pleasant Real estate.*

Description

Information about properties for sale in three subdivisions of Mount Pleasant, South Carolina, in the year 2017.

Usage

```
housing_hawkes
```

Format

A data.frame with 245 rows and 24 columns.

id Property ID number

price Asking price in dollars

duplex Is the property a duplex?

bedrooms Number of bedrooms

baths_total baths_full + baths_half

baths_full Number of full bathrooms

baths_half Number of half bathrooms

stories Number of stories

Subdivision Name of the subdivision the property is located in

sqrt Estimated floor area inside the house

year The year the house was constructed

acres The size of the lot in acres

new Has the house been lived in previously?

style The style of the house

garage Number of covered parking spots on the property

misc Miscellaneous exterior features (text string)

pool Does the home have a private pool?

dock Does the home have a private dock?

fence Does the home have a fenced yard?

porch Does the home have a screened porch?

Amenities Text screen of other amenities associated with the property

golf_course Is the home located on a golf course?

fireplace Does the home have a fireplace?

num_fireplace Number of fireplaces

Source

https://www.hawkeslearning.com/Statistics/dataset-descriptions.html#mount-pleasant-real-estate

---

`mpg_hawkes` *Miles Per Gallon*

Description

Data on vehicle fuel economy for model years 1984—2019

Usage

```
mpg_hawkes
```

Format

A data.frame with 38,693 rows and 11 columns.

city city miles per gallon

cylinders number of cylinders in the vehicle engine

displ engine displacement (liters)

drive drive axle type (front-wheel, four-wheel, etc.)

fuelType type of fuel (regular, premium, diesel, etc.)

highway highway miles per gallon

make manufacturer

model model name

trans type of transmission

VClass vehicle size class

year model year

Source

https://www.seattlecentral.edu/qelp/sets/057/057.html

https://www.hawkeslearning.com/Statistics/dataset-descriptions.html#miles-per-gallon

ncaa_times *NCAA Women's Cross Country Championship Run Times*

Description

Contains the run times for women who recorded valid run times in both the 2021 and 2022 NCAA National Championship Meets.

Usage

```
ncaa_times
```

Format

A data.frame with 98 rows and 3 columns.

name The name of the athlete

final_2021 The 6k run time in seconds for the 2021 meet

final_2022 The 6k run time in seconds for the 2022 meet

Source

https://xc.tfrrs.org/results_search.html

patient_hawkes *Medical statistics for a sample of 100 patients from a hospital.*

Description

Medical statistics for a sample of 100 patients from a hospital.

Usage

```
patient_hawkes
```

Format

A data.frame with 100 rows and 8 columns

id patient id number

gender patient gender (dichotomous)

marital_status one of married (M), widowed (W), divorced (D) or single (S)

age patient age

weight patient weight in pounds

cholesterol patient total cholesterol

blood_pressure patient systolic blood pressure in mmHg

cohort patients separated by decade of age

Source

https://www.hawkeslearning.com/Statistics/beg3/dataset-descriptions.html#decreases-in-cholesterol-l

---

`pres_speech_df` *Presidential inaugural addresses*

Description

This is a web scraped dataset of all presidential inauguration speeches. It does not include second or third inaugural addresses, nor does it include presidential addresses when the vice president replaced a dead president mid term.

Usage

```
pres_speech_df
```

Format

A data.frame with 53 rows rows and 2 variables.

year The year in which the inaugural address was given.

speech The full text of the inaugural speech as a single character strring.

Source

https://www.presidency.ucsb.edu/documents/app-categories/spoken-addresses-and-remarks/presidential/inaugural-addresses

---

`rt_load_ut` *Utah Design Snow Loads*

Description

Reliability-targeted design snow loads for the state of Utah. A design snow load is the pressure exerted by accumulated snow that a building needs to be able to withstand in order to prevent collapse.

Usage

```
rt_load_ut
```

Format

A data.frame with 257 rows and 6 columns

name name of the weather station at which the measurements were recorded

latitude decimal degrees latitude (geographic location)

longitude decimal degrees longitude

elevation altitude of the station, measured in feet

eco_region the level I ecological region as defined by the U.S. Environmental Protection Agency

load the design snow load, measured in pounds per square foot (psf)

Source

https://doi.org/10.26077/200k-pr86

---

satisfy_hawkes *Data regarding campus crime.*

Description

Data regarding campus crime.

Usage

```
satisfy_hawkes
```

Format

A data.frame with 15,000 rows and 12 columns.

employee_id The ID number associated with the individual employee

satisfaction_level How satisfied the employee is in their position (scale of 0 to 1)

last_evaluation_score How management rated employee performance during the last evaluation (scale of 0 to 1)

number_of_projects The number of projects an employee is currently working on

average_monthly_hours he average number of hours the employee works in a month

years_spent_at_company The number of years the employee has worked at the company

work_accident A binary variable that indicates whether the employee experienced an accident at work

left_company A binary variable that indicates whether an employee left the company

promotion_in_last_5_years A binary variable that indicates whether an employee received a promotion in the last 5 years

department The department that the employee works in

salary The level of the employee' s salary(low, medium, high)

salary_range The dollar range for the salary levels

Source

https://www.hawkeslearning.com/Statistics/dataset-descriptions.html#employee-satisfaction

https://www.kaggle.com/ludobenistant/hr-analytics

`swe` *Utah SNOTEL station SWE measurements.*

Description

This is data obtained from the National Water and Climate Center report generator. It contains snow water equivalent (SWE) measurements from two snowpack telemetery (SNOTEL) stations in Utah. One at Tony Grove, Lake up Logan Canyon, and another at Ben Lomond Peak, east of Ogden.

Usage

```
swe
```

Format

A data.frame with 32,972 rows rows and 5 variables.

date The date of the measurement in y-m-d format.

id The SNOTEL measurement ID.

name The SNOTEL station name.

swe The measured snow water equiavlent, in 10ths of millimeters.

qc An NWCC quality control flag where "S" stands for suspect measurements.

Source

https://wcc.sc.egov.usda.gov/reportGenerator/

`toluca` *Comparison of lot size vs work hours at the Toluca Company*

Description

Comparison of lot size vs work hours at the Toluca Company

Usage

```
toluca
```

Format

A data.frame with 25 rows and 2 columns

size lot size (no units originally provided, but assumed to be thousands of square feet)

hours average hours it takes to make single part

Source

http://www.cnachtsheim-text.csom.umn.edu/Kutner/Chapter%20%201%20Data%20Sets/CH01PR20.txt

Index

∗ **datasets**
  campus_crime_hawkes, 2
  cereal, 3
  cereal_infected, 4
  colleges, 5
  covid_case, 6
  covid_hos, 6
  covid_raw, 7
  crime_hawkes, 8
  election_survey_2016, 8
  finch_hawkes, 9
  housing_hawkes, 10
  mpg_hawkes, 11
  ncaa_times, 12
  patient_hawkes, 12
  pres_speech_df, 13
  rt_load_ut, 13
  satisfy_hawkes, 14
  swe, 15
  toluca, 15

campus_crime_hawkes, 2
cereal, 3
cereal_infected, 4
colleges, 5
covid_case, 6
covid_hos, 6
covid_raw, 7
crime_hawkes, 8

election_survey_2016, 8

finch_hawkes, 9

housing_hawkes, 10

mpg_hawkes, 11

ncaa_times, 12

patient_hawkes, 12

pres_speech_df, 13

rt_load_ut, 13

satisfy_hawkes, 14
swe, 15

toluca, 15