Package 'coursekata'

August 16, 2024

Title Packages and Functions for 'CourseKata' Courses

Version 0.18.0

Date 2024-08-16

Description Easily install and load all packages and functions used in
'CourseKata' courses. Aid teaching with helper functions and augment
generic functions to provide cohesion between the network of packages.
Learn more about 'CourseKata' at <https://coursekata.org>.

License AGPL (>= 3)

URL https://github.com/coursekata/coursekata-r

BugReports https://github.com/coursekata/coursekata-r/issues

Depends R (>= 3.6)

Imports cli (>= 3.2.0), dslabs (>= 0.7.4), ggformula (>= 0.10.1),
ggplot2 (>= 3.5.0), glue (>= 1.6.2), lsr (>= 0.5.2), Metrics,
mosaic (>= 1.8.3), palmerpenguins, purrr (>= 0.3.4), remotes,
rlang (>= 1.0.2), supernova (>= 2.5.1), vctrs (>= 0.4.1),
viridisLite

Suggests fivethirtyeight (>= 0.6.2), lubridate (>= 1.8.0), MASS,
mockery (>= 0.4.3), mockr (>= 0.1), readr (>= 2.1.2), readxl
(>= 1.4.0), usethis (>= 2.1.6), simstudy (>= 0.5.0), testthat
(>= 3.1.2), tibble(>= 3.1.7), tidyr (>= 1.2.0), vdiffr (>=
1.0.2), withr (>= 2.5.0)

Config/testthat/edition 3

Config/testthat/parallel true

Language en-US

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

NeedsCompilation no

Author Adam Blake [cre, aut] (<https://orcid.org/0000-0001-7881-8652>),
Ji Son [aut] (<https://orcid.org/0000-0002-4258-4791>),
Jim Stigler [aut] (<https://orcid.org/0000-0001-6107-7827>)

Maintainer Adam Blake <adam@coursekata.org>

Repository CRAN

Date/Publication 2024-08-16 20:20:02 UTC

Contents

Ames   *Ames, Iowa housing data*

Description

Data describing all residential home sales in Ames, Iowa from the years 2006–2010 as reported by the Ames City Assessor's Office and compiled by De Cock (2011). Ames is located about 30 miles north of Des Moines (the stats capitol) and is home to Iowa State University (the largest university in the state). Each row represents the latest sale of a home (one row per home in the dataset). Columns represent home features and sale prices (outcome). The original dataset includes a uniquely detailed (81 features per home) and comprehensive look at the housing market. The data included here are only a subset used for examples in CourseKata course material. See the references and data source for the full dataset.

Pedagogical Modifications:

To simplify the dataset for instructional purposes, the data were filtered to include only single family homes, residential zoning, 1-2 story homes, homes with brick, cinder block, or concrete foundations, and average to excellent kitchen qualities. Further, the descriptive variables were reduced to the subset described in the format section.

Usage

```
Ames
```

Format

A data frame with 2930 observations on the following 80 variables:

`YearBuilt` Year home was built (YYYY).

`YearSold` Year of home sale (YYYY). Note: all home sales in this dataset occurred between 2006 - 2010. If a home was sold more than once between 2006 - 2010, only its latest sale is included in dataset.

`Neighborhood` One of two neighborhoods in Ames county:

- College Creek (`CollegeCreek`), a neighborhood located adjacent to Iowa State University (the largest University in the state).
- Old Town (`OldTown`), a nationally designated historic district in Ames. The old neighborhood is located just north of the central business district.

`HomeSizeR` Raw above-ground area of home, measured in square feet.

`HomeSizeK` Above-ground area of home, measured in thousands of square feet.

`LotSizeR` Raw total property lot size, measured in square feet.

`LotSizeK` Total property lot size, in thousands of square feet.

`Floors` Number of above-ground floors (1 story or 2 story).

`BuildQuality` Assessor's rating of overall material and finish of the house.

- 10: Very Excellent
- 9: Excellent
- 8: Very Good
- 7: Good
- 6: Above Average
- 5: Average

- 4: Below Average
- 3: Fair
- 2: Poor
- 1: Very Poor

`Foundation` Type of foundation (ground material underneath the house).

- `Brick&Tile`: Brick and Tile
- `CinderBlock`: Cinder Blocks
- `PouredConcrete`: Poured Concrete

`HasCentralAir` Indicator if home contains central air conditioning (0 = No, 1 = Yes).

`Bathrooms` Number of full above-ground bathrooms.

`Bedrooms` Number of full above-ground bedrooms.

`TotalRooms` Number of above-ground rooms in home, excluding bathrooms.

`KitchenQuality` Assessor's rating of kitchen material quality.

- `Excellent`
- `Good`
- `Average`

`HasFireplace` Indicator if home contains at least one fireplace (0 = No, 1 = Yes).

`GarageType` Type of garage.

- `Attached`: includes attached, built-in, basement, and dual-type garages
- `Detached`: includes detached and carport garages
- `None`: home does not have a garage or carport

`GarageCars` Number of cars that can fit in garage.

`PriceR` Sale price of home, in raw USD ($)

`PriceK` Sale price of home, in thousands of USD ($)

`TinySet` (Ignore) Whether or not this row is in `ames_tiny.csv`

Source

https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data

References

De Cock, Dean, (2011). Ames, Iowa: Alternative to the Boston Housing Data as an end of semester regression project, *Journal of Statistics Education, 19*(3). doi:10.1080/10691898.2011.11889627

---

class_data *Generated "class data" for exploring pairwise tests*

Description

These data were generated as outcomes for "students" for three different "instructors" named A, B, and C. The outcome have means such that C > B > A, but the difference is only clearly significant for C > A, and borderline for the others.

Usage

```
class_data
```

Format

An object of class tbl_df (inherits from tbl, data.frame) with 105 rows and 2 columns.

Details

outcome A hypothetical, numerical outcome of an intervention.

teacher Either "A", "B", or "C", associating the outcome to a teacher.

---

coursekata_attach *Attach the CourseKata course packages*

Description

Attach the CourseKata course packages

Usage

```
coursekata_attach(do_not_ask = FALSE, quietly = FALSE)
```

Arguments

do_not_ask   Prevent asking the user to install missing packages (they are skipped).

quietly   Whether to suppress messages.

Value

A named logical vector indicating which packages were attached.

Examples

```
coursekata_attach()
```

---

coursekata_install   *Install or update all CourseKata packages.*

Description

Install or update all CourseKata packages.

Usage

```
coursekata_install(...)

coursekata_update(...)
```

Arguments

...	Arguments passed on to `remotes::install_cran` or `remotes::install_github` depending on whether the package appears to be from CRAN or GitHub.

Value

The state of all the packages after any updates have been performed.

---

coursekata_load_theme   *Utility function for loading all themes.*

Description

This function is called at package start-up and should rarely be needed by the user. The exception is when the user has called `coursekata_unload_theme()` and wants to go back to the CourseKata look and feel. When run, this function sets the CourseKata color palettes `coursekata_palette()`, sets the default theme to `theme_coursekata()`, and tweaks some default settings for specific plots. To restore the original ggplot2 settings, run `coursekata_unload_theme()`.

Usage

```
coursekata_load_theme()
```

Value

No return value, called to adjust the global state of `ggplot2`.

See Also

coursekata_palette theme_coursekata scale_discrete_coursekata coursekata_unload_theme

---

coursekata_packages   *List all CourseKata course packages*

Description

List all CourseKata course packages

Usage

```
coursekata_packages(check_remote_version = FALSE)
```

Arguments

check_remote_version

> Should the remote version number be checked? Requires internet, and will take longer.

Value

A data frame with three variables: the name of the package package, the version, and whether it is currently attached.

Examples

```
coursekata_packages()
```

---

coursekata_palette   *The color palettes used in our theme system*

Description

The color palettes used in our theme system

Usage

```
coursekata_palette(indices = integer(0))
```

Arguments

indices   The indices of the colors to pull (or all colors if no indices are given).

Value

A named list of the requested colors in the palette.

---

coursekata_palette_provider

Create a function that provides a colorblind palette.

Description

Create a function that provides a colorblind palette.

Usage

```
coursekata_palette_provider()
```

Value

A function that accepts one argument n, which is the number of colors you want to use in the plot. This function is used by scales like `scale_color_discrete` to provide colorblind- safe palettes. Where possible, the function will use the hand-picked colors from `coursekata_palette()`, and when more colors are needed than are available, it will use the `viridisLite::viridis()` palette.

See Also

scale_discrete_coursekata

---

coursekata_repos   *Get repositories for the packages.*

Description

Ensures a default CRAN is set if one is not already set, and adds the repository for fivethirtyeight-data.

Usage

```
coursekata_repos(repos = getOption("repos"))
```

Arguments

repos   Optionally set a repository character vector to augment.

Value

A set of repositories that can be used to install or update the CourseKata packages.

Examples

```
coursekata_repos()
```

---

coursekata_unload_theme

Restore ggplot2 *default settings*

Description

This function will restore all of the tweaks to themes and plotting to the original ggplot2 defaults. If you want to go back to the CourseKata look and feel, run coursekata_load_theme().

Usage

```
coursekata_unload_theme()
```

Value

No return value, called to restore the global state of ggplot2.

See Also

coursekata_load_theme

---

er *Emergency room canine therapy*

Description

Data from: Controlled clinical trial of canine therapy versus usual care to reduce patient anxiety in the emergency department.

Abstract:

Objective:
Test if therapy dogs can reduce anxiety in emergency department (ED) patients.

Methods:
In this controlled clinical trial (NCT03471429), medically stable, adult patients were approached if the physician believed that the patient had "moderate or greater anxiety." Patients were allocated on a 1:1 ratio to either 15 min exposure to a certified therapy dog and handler (dog), or usual care (control). Patient reported anxiety, pain and depression were assessed using a 0-10 scale (10=worst). Primary outcome was change in anxiety from baseline (T0) to 30 min and 90 min after exposure to dog or control (T1 and T2 respectively); secondary outcomes were pain, depression and frequency of pain medication.

Results:
Among 98 patients willing to participate in research, 7 had aversions to dogs, leaving 91 (93%) were willing to see a dog; 40 patients were allocated to each group (dog or control). No data were normally distributed. Median baseline anxiety, pain and depression were similar between groups. With dog exposure, anxiety decreased significantly from T0 to T1: 6 (IQR 4-9.75) to

T1: 2 (0-6) compared with 6 (4-8) to 6 (2.5-8) in controls (P<0.001, for T1, Mann-Whitney U). Dog exposure was associated with significantly lower anxiety at T2 and a significant overall treatment effect on two-way repeated measures ANOVA for anxiety, pain and depression. After exposure, 1/40 in the dog group needed pain medication, versus 7/40 in controls (P=0.056, Fisher's).

Conclusions:
Exposure to therapy dogs plus handlers significantly reduced anxiety in ED patients.

Usage

```
er
```

Format

A data frame with 84 observations on the following 53 variables:

id Subject ID

condition Whether the subject saw a Dog or was in the Control group

age Subject's age in years

gender Subject's self-identified gender

race Subject's self-identified race

veteran Is the subject a veteran?

disabled Is the subject disabled?

dog_name The name of the therapy dog

base_pain Subject's self reported pain before the intervention (T0)

base_depression Subject's self reported depression before the intervention (T0)

base_anxiety Subject's self reported anxiety before the intervention (T0)

base_total The sum of the subject's base_* scores

later_pain Subject's self reported pain after the intervention (T1)

later_depression Subject's self reported depression after the intervention (T1)

later_anxiety Subject's self reported anxiety after the intervention (T1)

later_total The sum of the subject's later_* scores

last_pain Subject's self reported pain after the intervention (T2)

last_depression Subject's self reported depression after the intervention (T2)

last_anxiety Subject's self reported anxiety after the intervention (T2)

last_total The sum of the subject's last_* scores

change_pain The change in subject's pain from before the intervention to after

change_depression The change in subject's depression from before the intervention to after

change_anxiety The change in subject's anxiety from before the intervention to after

change_total The sum of the subject's change_* scores

provider_male Was the health care provider male?

provider The health care provider's status: either an `Advanced Practitioner`, `Resident` physician, or `Attending` physician

heart_rate The subject's heart rate at baseline (T0)

resp_rate The subject's respiratory rate at baseline (T0)

sp_o2 The subject's SpO2 at baseline (T0)

bp_syst The subject's systolic blood pressure at baseline (T0)

bp_diast The subject's diastolic blood pressure at baseline (T0)

med_given Was the subject given medication prior to the study? (T0)

mh_none None of the other medical history items were indicated

mh_asthma Medical history: asthma

mh_smoker Medical history: smoker

mh_cad Medical history: coronary artery disease

mh_diabetes Medical history: diabetes mellitus

mh_hypertension Medical history: hypertension

mh_stroke Medical history: prior stroke

mh_chronic_kidney Medical history: chronic kidney disease

mh_copd Medical history: chronic obstructive pulmonary disease

mh_hyperlipidemia Medical history: hyperlipidemia

mh_hiv Medical history: HIV

mh_other Medical history: other (write-in)

ph_adhd Psychiatric history: attention-deficit/hyperactivity disorder

ph_anxiety Psychiatric history: anxiety

ph_bipolar Psychiatric history: bipolar

ph_borderline Psychiatric history: borderline personality disorder

ph_depression Psychiatric history: depression

ph_schizophrenia Psychiatric history: schizophrenia

ph_ptsd Psychiatric history: PTSD

ph_none None of the other psychiatric history items were indicated

ph_other Psychiatric history: other (write-in)

References

Kline, J. A., Fisher, M. A., Pettit, K. L., Linville, C. T., & Beck, A. M. (2019). Controlled clinical trial of canine therapy versus usual care to reduce patient anxiety in the emergency department. *PloS One, 14*(1), e0209232. doi:10.1371/journal.pone.0209232

---

estimate_extraction   *Extract estimates/statistics from a model*

Description

This collection of functions is useful for extracting estimates and statistics from a fitted model. They
are particularly useful when estimating many models, like when bootstrapping confidence intervals.
Each function can be used with an already fitted model as an [lm](#) object, or a formula and associated
data can be passed to it. **All of these assume the comparison is the empty model.**

Usage

```
b0(object, data = NULL)

b1(object, data = NULL)

b(object, data = NULL, all = FALSE, predictor = character())

f(object, data = NULL, all = FALSE, predictor = character(), type = 3)

pre(object, data = NULL, all = FALSE, predictor = character(), type = 3)

p(object, data = NULL, all = FALSE, predictor = character(), type = 3)

fVal(object, data = NULL, all = FALSE, predictor = character(), type = 3)

PRE(object, data = NULL, all = FALSE, predictor = character(), type = 3)
```

Arguments

object	A [lm](#) object, or [formula](#).
data	If object is a formula, the data to fit the formula to as a [data.frame](#).
all	If TRUE, return a named list of all related terms (e.g. all *F*-values).The name for the full model value is the name of the function (e.g. "f"), and the names for the constituent terms are the term names prefixed by the function name (e.g. "f_a:b" for the *F*-value of the a:b interaction term).
predictor	Filter the output down to just the statistics for these terms (e.g. "hp" to just get the statistics for that term in the model). This argument is flexible: you can pass a character vector of terms (c("hp", "hp:cyl")), a one-sided formula (~hp), or a list of formulae (c(~hp, ~hp:cyl)).
type	The type of sums of squares to calculate (see [generate_models()](#)). Defaults to the widely used Type III SS.

Details

- `b0`: The intercept from the full model.
- `b1`: The slope b1 from the full model.
- `b`: The coefficients from the full model.
- `f`: The F value from the full model.
- `pre`: The Proportional Reduction in Error for the full model.
- `p`: The *p*-value from the full model.
- `sse`: The SS Error (SS Residual) from the model.
- `ssm`: The SS Model (SS Regression) for the full model.
- `ssr`: Alias for SSM.

Value

The value of the estimate as a single number.

References

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data Analysis: A Model Comparison Approach to Regression, ANOVA, and Beyond* (3rd ed.). New York: Routledge. ISBN:879-1138819832

Examples

```
supernova(lm(mpg ~ disp, data = mtcars))

change_p_decimals <- supernova(lm(mpg ~ disp, data = mtcars))
print(change_p_decimals, pcut = 8)
```

---

fevdata	*Forced Expiratory Volume (FEV) Data*

Description

Data from: Fundamentals of Biostatistics Notes from: Kahn, M.

Abstract:

Sample of 654 youths, aged 3 to 19, in the area of East Boston during middle to late 1970's. Interest concerns the relationship between smoking and FEV. Since the study is necessarily observational, statistical adjustment via regression models clarifies the relationship.

Pedagogical Notes::

This is a versatile dataset that can be used throughout an introductory statistics course as well as an introductory modeling course. It includes many issues from statistical adjustment in observational studies, to subgroup analysis, quadratic regression and analysis of covariance.

Usage

```
fevdata
```

Format

A data frame with 654 observations on the following 5 variables:

AGE Age, in years

FEV Forced expiratory volume, in liters

HEIGHT Height, in inches

SEX 0 = Female, 1 = Male

SMOKE 0 = Non-smoker, 1 = Smoker

References

Kahn,M. (2003). Data Sleuth, STATS, 37, 24. <http://jse.amstat.org/datasets/fev.txt> Rosner, B. (1999). Fundamentals of Biostatistics, Pacific Grove, CA: Duxbury

---

Fingers	*Data from introductory statistics students at a university.*

Description

Students at a university taking an introductory statistics course were asked to complete this survey as part of their homework.

Usage

```
Fingers
```

Format

A data frame with 157 observations on the following 16 variables:

Sex Sex of participant.

RaceEthnic Racial or ethnic background.

FamilyMembers Members of immediate family (excluding self).

SSLast Last digit of social security number (NA if no SSN).

Year Year in school: 1=First, 2=Second, 3=Third, 4=Fourth, 5=Other

Job Current employment status: 1=Not Working, 2=Part-time Job, 3=Full-time Job

MathAnxious Agreement with the statement "In general I tend to feel very anxious about mathematics": 1=Strongly Disagree, 2=Disagree, 3=Neither Agree nor Disagree, 4=Agree, 5=Strongly Agree

Interest Interest in statistics and the course: 1=No Interest, 2=Somewhat Interested, 3=Very Interested

GradePredict Numeric prediction for final grade in the course. The value is converted from the student's letter grade prediction. 4.0=A, 3.7=A-, 3.3=B+, 3.0=B, 2.7=B-, 2.3=C+, 2.0=C, 1.7=C-, 1.3=Below C-

Thumb Length in mm from tip of thumb to the crease between the thumb and palm.

Index Length in mm from tip of index finger to the crease between the index finger and palm.

Middle Length in mm from tip of middle finger to the crease between the middle finger and palm.

Ring Length in mm from tip of ring finger to the crease between the middle finger and palm.

Pinkie Length in mm from tip of pinkie finger to the crease between the pinkie finger and palm.

Height Height in inches.

Weight Weight in pounds.

FingersMessy *Raw data from introductory statistics students at a university.*

Description

This is the Fingers dataset before it was cleaned. In the cleaning process, we converted the values from numbers to appropriate types (where applicable), removed outliers that suggested data was input incorrectly, and we removed incomplete cases. The description for the dataset is: Students at a university taking an introductory statistics course were asked to complete this survey as part of their homework. (This is the same data set as the Fingers data)

Usage

FingersMessy

Format

A data frame with 157 observations on the following 16 variables:

Sex Sex of participant.

RaceEthnic Racial or ethnic background.

FamilyMembers Members of immediate family (excluding self).

SSLast Last digit of social security number (NA if no SSN).

Year Year in school: 1=First, 2=Second, 3=Third, 4=Fourth, 5=Other

Job Current employment status: 1=Not Working, 2=Part-time Job, 3=Full-time Job

MathAnxious Agreement with the statement "In general I tend to feel very anxious about mathematics": 1=Strongly Disagree, 2=Disagree, 3=Neither Agree nor Disagree, 4=Agree, 5=Strongly Agree

Interest Interest in statistics and the course: 1=No Interest, 2=Somewhat Interested, 3=Very Interested

GradePredict Numeric prediction for final grade in the course. The value is converted from the student's letter grade prediction. 4.0=A, 3.7=A-, 3.3=B+, 3.0=B, 2.7=B-, 2.3=C+, 2.0=C, 1.7=C-, 1.3=Below C-

Thumb Length in mm from tip of thumb to the crease between the thumb and palm.

Index Length in mm from tip of index finger to the crease between the index finger and palm.

Middle Length in mm from tip of middle finger to the crease between the middle finger and palm.

Ring Length in mm from tip of ring finger to the crease between the middle finger and palm.

Pinkie Length in mm from tip of pinkie finger to the crease between the pinkie finger and palm.

Height Height in inches.

Weight Weight in pounds.

---

fit_stats *Test the fit of a model on a train and test set.*

Description

Test the fit of a model on a train and test set.

Usage

```
fit_stats(model, df_train, df_test)

fitstats(model, df_train, df_test)
```

Arguments

model   An lm model.

df_train   A data frame with the training data.

df_test   A data frame with the test data.

Value

A data frame with the fit statistics.

---

game_data *Simulated math game data.*

Description

The simulated results of a small study comparing the effectiveness of three different computer-based math games in a sample of 105 fifth-grade students. All three games focused on the same topic and had identical learning goals, and none of the students had any prior knowledge of the topic.

Usage

```
game_data
```

Format

A data frame with 105 observations on the following 2 variables:

game The game the student was randomly assigned to, coded as "A", "B", or "C".

outcome Each student's score on the outcome test.

---

gf_model   *Add a model to a plot*

Description

When teaching about regression it can be useful to visualize the data as a point plot with the outcome on the y-axis and the explanatory variable on the x-axis. For regression models, this is most easily achieved by calling ggformula::gf_lm(), with empty models ggformula::gf_hline() using the mean, and a more complicated call to ggformula::gf_segment() for group models. This function simplifies this by making a guess about what kind of model you are plotting (empty/null, regression, group) and then making the appropriate plot layer for it.

Usage

```
gf_model(object, model, ...)
```

Arguments

object	A plot created with the ggformula package.
model	A linear model fit by either lm() or aov().
...	Additional arguments. Typically these are (a) ggplot2 aesthetics to be set with attribute = value, (b) ggplot2 aesthetics to be mapped with attribute = ~ expression, or (c) attributes of the layer as a whole, which are set with attribute = value.

Details

This function only works with models that have a continuous outcome measure.

Value

a gg object (a plot layer) that can be added to a plot.

---

middle   *Find a percentage of a distribution*

Description

Given a distribution, find which values lie in the upper, lower, or middle proportion of the distribution. Useful when you want to do something like shade in the middle 95% of a plot. This is a greedy operation, meaning that if the cutoff point is between two whole numbers the specified region will suck up the extra space. For example, the requesting the upper 30% of the [1 2 3 4] will return [FALSE FALSE TRUE TRUE] because the 30% was greedy.

Usage

```
middle(x, prop = 0.95, greedy = TRUE)

tails(x, prop = 0.95, greedy = TRUE)

lower(x, prop = 0.025, greedy = TRUE)

upper(x, prop = 0.025, greedy = TRUE)
```

Arguments

x	The distribution of values to check.
prop	The proportion of values to find.
greedy	Whether the function should be greedy, as per the description above.

Details

Note that NA values are ignored, i.e. they will always return FALSE.

Value

A logical vector indicating which values are in the specified region.

Examples

```
upper(1:10, .1)
lower(1:10, .2)
middle(1:10, .5)
tails(1:10, .5)

sampling_distribution <- do(1000) * mean(rnorm(100, 5, 10))
sampling_distribution %>%
  gf_histogram(~mean, data = sampling_distribution, fill = ~ middle(mean, .68)) %>%
  gf_refine(scale_fill_manual(values = c("blue", "coral")))
```

penguins *A modified form of the* palmerpenguins::penguins *data set.*

Description

The modifications are to select only a subset of the variables, and convert some of the units.

Usage

```
penguins
```

Format

A data frame with 333 observations on the following 7 variables:

species The species of penguin, coded as "Adelie", "Chinstrap", or "Gentoo".

gentoo Whether the penguin is a Gentoo penguin (1) or not (0).

body_mass_kg The mass of the penguin's body, in kilograms.

flipper_length_m The length of the penguin's flipper, in m.

bill_length_cm The length of the penguin's bill, in cm.

female Whether the penguin is female (1) or not (0).

island The island where the penguin was observed, coded as "Biscoe", "Dream", or "Torgersen".

---

scale_discrete_coursekata

A discrete color scale constructor with colorblind-safe palettes.

Description

See [coursekata_palette()](#) for more information.

Usage

```
scale_discrete_coursekata(...)
```

Arguments

... Additional parameters passed on to the scale type.

Value

A discrete color scale.

See Also

coursekata_palette

Smallville   *Simulated housing data*

Description

These data are simulated to be similar to the Ames housing data, but with far fewer variables and much smaller effect sizes.

Usage

```
Smallville
```

Format

A data frame with 32 observations on the following 4 variables:

PriceK Price the home sold for (in thousands of dollars)

Neighborhood The neighborhood the home is in (Eastside, Downtown)

HomeSizeK The size of the home (in thousands of square feet)

HasFireplace Whether the home has a fireplace (0 = no, 1 = yes)

---

split_data   *Split data into train and test sets.*

Description

Split data into train and test sets.

Usage

```
split_data(data, prop = 0.7)
```

Arguments

data	A data frame.
prop	The proportion of rows to assign to the training set.

Value

A list with two data frames, train and test.

Survey	*Students at a university were asked to enter a random number between 1-20 into a survey.*

Description

Students at a university taking an introductory statistics course were asked to complete this survey as part of their homework.

Usage

```
Survey
```

Format

A data frame with 211 observations on the following 1 variable:

Any1_20 The random number between 1 and 20 that a student thought of.

Tables	*Tables data*

Description

Data about tips collected from an experiment with 44 tables at a restaurant.

Usage

```
Tables
```

Format

A data frame with 44 observations on the following 2 variables.

TableID A number assigned to each table.

Tip How much the tip was.

---

theme_coursekata *A simple theme built on top of* `ggplot2::theme_bw`

Description

The `coursekata` package automatically loads this theme when the package is loaded. This is in addition to a number of other plot tweaks and option settings. To just restore the theme to the default, you can run set_theme(theme_grey). If you want to restore all plot related settings and/or prevent them when loading the package, see `coursekata_unload_theme`.

Usage

```
theme_coursekata()
```

Value

A gg theme object

Examples

```
gf_boxplot(Thumb ~ RaceEthnic, data = Fingers, fill = ~RaceEthnic)
```

TipExperiment *Data from an experiment about smiley faces and tips*

Description

Tables were randomly assigned to receive checks that either included or did not include a drawing of a smiley face. Data was collected from 44 tables in an effort to examine whether the added smiley face would cause more generous tipping.

Usage

```
TipExperiment
```

Format

A data frame with 44 observations on the following 3 variables.

TableID A number assigned to each table.

Tip How much the tip was.

Condition Which experimental condition the table was randomly assigned to.

Check (Simulated) The amount of money the table paid for their meal.

FoodQuality (Simulated) The perceived quality of the food.

---

tip_exp *Simulated data for an experiment about smiley faces and tips*

Description

These are simulated data that are similar to the `TipExperiment` data. Hypothetical tables were randomly assigned to receive checks that either included or did not include a drawing of a smiley face, either from a male or a female server.

Usage

```
tip_exp
```

Format

A data frame with 44 observations on the following 3 variables.

gender Whether the server was `female` or `male`

condition Whether the check had a `smiley face` or not (`control`)

tip_percent The size of the tip as a percentage of the price of the meal

---

World *Data on countries from the Happy Planet Index project.*

Description

These data have been updated with some historical height data (from Our World in Data), drinking data (collected by the World Health Organization featured in fivethirtyeight), population and land characteristics, and vaccination data (from March 2023).

Usage

```
World
```

Format

A data frame with 130 observations on the following 14 variables:

Country Name of country

Region One of 5 UN defined regions: Africa, Americas, Asia, Europe, Oceania

Code Three-letter country codes defined by the International Organization for Standardization (ISO) to represent countries in a way that avoids errors since a country's name changes depending on the language being used.

LifeExpectancy Average life expectancy (in years)

GirlsH1900 The average of 18-year-old girls heights in 1900 (in cm)

GirlsH1980 The average of 18-year-old girls heights in 1980 (in cm)

Happiness Score on a 0-10 scale for average level of happiness (10 being happiest)

GDPperCapita Gross Domestic Product (per capita)

FertRate The average number of children that will be born to a woman over her lifetime

PeopleVacc Total number of people vaccinated in the country

PeopleVacc_per100 Total number of people vaccinated in the country (in percent)

Population2010 Population (in millions) in 2010

Population2020 Population (in millions) in 2020

WineServ Average wine consumption per capita for those age 15 and over per week (collected by WHO)

Index