

# Package ‘coursekata’

May 8, 2026

**Title** Packages and Functions for 'CourseKata' Courses

**Version** 0.19.2

**Date** 2026-03-09

**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://www.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 (>= 3.5.2), glue (>= 1.6.2), lifecycle (>= 1.0.3), 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), knitr (>= 1.40), lubridate (>= 1.8.0), MASS, mockery (>= 0.4.3), mockr (>= 0.1), readr (>= 2.1.2), readxl (>= 1.4.0), rmarkdown (>= 2.17), 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)

**VignetteBuilder** knitr

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

**Maintainer** Adam Blake <adam@coursekata.org>

**Repository** CRAN

**Date/Publication** 2026-03-10 17:10:13 UTC

## Contents

Ames	3
class_data	5
coursekata_attach	5
coursekata_install	6
coursekata_load_theme	6
coursekata_packages	7
coursekata_palette	7
coursekata_palette_provider	8
coursekata_repos	8
coursekata_unload_theme	9
er	9
estimate_extraction	12
fevdata	13
Fingers	14
FingersMessy	15
fit_stats	16
game_data	17
gf_model	17
gf_resid	18
gf_resid_fun	19
gf_sd_ruler	20
gf_squareplot	21
gf_square_resid	22
gf_square_resid_fun	23
middle	24
penguins	25
scale_discrete_coursekata	25
show_cutoffs	26
Smallville	27
split_data	27
Survey	28
Tables	28
theme_coursekata	29
TipExperiment	29
tip_exp	30
World	30

**Index**

**32**

---

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 state 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	<i>Generated "class data" for exploring pairwise tests</i>
------------	--

---

**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	<i>Attach the CourseKata course packages</i>
-------------------	--

---

**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 <code>lm</code> object, or <a href="#">formula</a> .
data	If object is a formula, the data to fit the formula to as a <a href="#">data.frame</a> .
all	If TRUE, return a named list of all related terms (e.g. all <i>F</i> -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 <i>F</i> -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 <a href="#">generate_models()</a> ). Defaults to the widely used Type III SS.

### Details

- $b_0$ : The intercept from the full model.
- $b_1$ : The slope  $b_1$  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. <https://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:

Gender Gender 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.

Sex Sex of participant.

---

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:

Gender Gender 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.

Sex Sex of participant.

---

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 <code>lm</code> 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	<i>Simulated math game data.</i>
-----------	----------------------------------

---

### 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	<i>Add a model to a plot</i>
----------	------------------------------

---

### 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 <code>lm()</code> or <code>aov()</code> .
...	Additional arguments. Typically these are (a) ggplot2 aesthetics to be set with <code>attribute = value</code> , (b) ggplot2 aesthetics to be mapped with <code>attribute = ~expression</code> , or (c) attributes of the layer as a whole, which are set with <code>attribute = value</code> .

**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.

---

`gf_resid`*Add Residual Lines to a Plot*

---

**Description**

This function adds vertical lines representing residuals from a linear model to a ggformula plot. The residuals are drawn from the observed data points to the predicted values from the model.

**Usage**

```
gf_resid(plot, model, linewidth = 0.2, ...)
```

**Arguments**

<code>plot</code>	A ggformula plot object, typically created with <code>gf_point()</code> .
<code>model</code>	A fitted linear model object created using <code>lm()</code> .
<code>linewidth</code>	A numeric value specifying the width of the residual lines. Default is <code>0.2</code> .
<code>...</code>	Additional aesthetics passed to <code>geom_segment()</code> , such as <code>color</code> , <code>alpha</code> , <code>linetype</code> .

**Value**

A ggplot object with residual lines added.

**Examples**

```
Height_model <- lm(Thumb ~ Height, data = Fingers)
gf_point(Thumb ~ Height, data = Fingers) %>%
  gf_model(Height_model) %>%
  gf_resid(Height_model, color = "red", alpha = 0.5)
```

---

gf_resid_fun	<i>Add Residual Lines from a Function to a Plot</i>
--------------	---

---

## Description

**[Experimental]**

## Usage

```
gf_resid_fun(plot, fun, linewidth = 0.2, ...)
```

## Arguments

plot	A ggformula/ggplot object, typically created with <code>gf_point()</code> .
fun	A function that takes a numeric vector <code>x</code> and returns predicted <code>y</code> .
linewidth	Numeric width of the residual lines. Default 0.2.
...	Additional aesthetics passed to <code>ggplot2::geom_segment()</code> , e.g., <code>color</code> , <code>alpha</code> , <code>linetype</code> .

## Details

Draws vertical residual lines from observed points to predicted values computed by a user-supplied function of `x` (e.g., the function plotted with `gf_function()`).

## Value

A ggplot object with residual segments added.

## Examples

```
set.seed(1)
df <- data.frame(X = 1:10, Y = 2 + 3 * (1:10) + rnorm(10))
my_fun <- function(x) 2 + 3 * x

gf_point(Y ~ X, data = df) %>%
  gf_function(my_fun) %>%
  gf_resid_fun(my_fun, color = "red", alpha = 0.5)
```

gf\_sd\_ruler

*Add a Standard Deviation Ruler to a Plot***Description****[Experimental]****Usage**

```
gf_sd_ruler(
  p,
  y = NULL,
  data = NULL,
  x = NULL,
  where = c("middle", "mean", "median"),
  color = "red",
  size = 0.8,
  ...
)
```

**Arguments**

p	A ggplot object (typically from <code>gf_point()</code> or <code>gf_jitter()</code> ).
y	The y-variable (bare name or string). Defaults to the plot's mapped y aesthetic if omitted.
data	Dataset. Defaults to <code>p\$data</code> .
x	The x-variable for placement. Defaults to the plot's mapped x.
where	Where on the x-axis to place the ruler: "middle" (midpoint of x range), "mean", or "median".
color	Segment color. Default "red".
size	Segment linewidth. Default 0.8.
...	Additional arguments passed to <code>ggplot2::geom_segment()</code> .

**Details**

Adds a vertical segment showing one standard deviation of a variable, placed at a specified x position. Works for both numeric x (scatter) and categorical x (jitter) plots.

**Value**

A ggplot object with the SD ruler segment added.

**Examples**

```
gf_jitter(Thumb ~ Height, data = Fingers) %>%
  gf_model(lm(Thumb ~ NULL, data = Fingers)) %>%
  gf_sd_ruler()
```

---

gf\_squareplot

*Countable-Rectangle Histogram*


---

## Description

**[Experimental]**

## Usage

```
gf_squareplot(
  x,
  data = NULL,
  binwidth = NULL,
  origin = NULL,
  boundary = NULL,
  fill = "#7fcecc",
  color = "black",
  alpha = 1,
  na.rm = TRUE,
  mincount = NULL,
  bars = c("none", "outline", "solid"),
  xbreaks = NULL,
  xrange = NULL,
  show_dgp = FALSE,
  show_mean = FALSE,
  auto_subdivide = FALSE
)
```

## Arguments

x	Formula (~variable) or numeric vector.
data	Data frame (required if x is a formula).
binwidth	Width of histogram bins. Auto-calculated if NULL.
origin	Starting position for bins.
boundary	Alias for origin.
fill	Rectangle fill color. Default "#7fcecc".
color	Rectangle border color. Default "black".
alpha	Transparency. Default 1.
na.rm	Remove NA values. Default TRUE.
mincount	Minimum y-axis height for consistent scaling.
bars	Display style: "none" (squares only), "outline", or "solid".
xbreaks	Number of x-axis breaks or vector of specific positions.
xrange	X-axis limits as c(min, max).

show\_dgp Show DGP annotation overlay. Default FALSE.  
 show\_mean Show dashed mean line. Default FALSE.  
 auto\_subdivide Split bins with >75 observations into sub-columns. Default FALSE.

### Details

Creates histograms where individual data points are visible as stacked unit rectangles, making counts easy to visualize. Designed for teaching statistical concepts, particularly sampling distributions.

### Value

A ggplot object with S3 class `c("gf_squareplot", "gg", "ggplot")`.

### Examples

```
gf_squareplot(~Thumb, data = Fingers)
gf_squareplot(~Thumb, data = Fingers, bars = "outline")
```

---

<code>gf_square_resid</code>	<i>Add Squared Residual Visualization to a Plot</i>
------------------------------	---

---

### Description

#### [Experimental]

`gf_squaresid()` was renamed to `gf_square_resid()` for naming consistency and is now deprecated.

### Usage

```
gf_square_resid(plot, model, aspect = 4/6, alpha = 0.1, ...)
gf_squaresid(plot, model, aspect = 4/6, alpha = 0.1, ...)
```

### Arguments

`plot` A ggformula plot object, typically created with `gf_point()`.  
`model` A fitted linear model object created using `lm()`.  
`aspect` A numeric value controlling the square's aspect ratio. Default is 4/6.  
`alpha` A numeric value specifying the transparency of the square's fill. Default is 0.1.  
`...` Additional aesthetics passed to `geom_polygon()`, such as `color` and `fill`.

### Details

This function adds squared residual representations to a ggformula plot, illustrating squared error as a polygon. The function dynamically adjusts the aspect ratio to ensure proper scaling of squares.

**Value**

A ggplot object with squared residuals added.

**Examples**

```
Height_model <- lm(Thumb ~ Height, data = Fingers)
gf_point(Thumb ~ Height, data = Fingers) %>%
  gf_model(Height_model) %>%
  gf_square_resid(Height_model, color = "blue", alpha = 0.5)
```

---

gf\_square\_resid\_fun    *Add Squared Residual Visualization from a Function to a Plot*

---

**Description**

[Experimental]

**Usage**

```
gf_square_resid_fun(plot, fun, aspect = 4/6, alpha = 0.1, ...)
```

**Arguments**

plot	A ggformula/ggplot object, typically created with <code>gf_point()</code> .
fun	A function that takes a numeric vector <code>x</code> and returns predicted <code>y</code> .
aspect	A numeric value controlling the square's aspect ratio. Default is 4/6.
alpha	Transparency of the filled squares. Default 0.1.
...	Additional aesthetics passed to <code>ggplot2::geom_polygon()</code> , e.g., <code>color</code> , <code>fill</code> , <code>linetype</code> .

**Details**

Draws squared residual polygons between observed points and predicted values computed by a user-supplied function of `x`.

**Value**

A ggplot object with squared residual polygons added.

**Examples**

```
set.seed(1)
df <- data.frame(X = 1:10, Y = 2 + 3 * (1:10) + rnorm(10))
my_fun <- function(x) 2 + 3 * x

gf_point(Y ~ X, data = df) %>%
  gf_function(my_fun) %>%
  gf_square_resid_fun(my_fun, color = "red", alpha = 0.3)
```

---

`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.

#### [Experimental]

`outer()` marks values in both outer tails of a distribution. It is the complement of `middle()`: `outer(x, prop)` is equivalent to `tails(x, 1 - prop)`.

### Usage

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

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

```
outer(x, prop)
```

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

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

### Arguments

<code>x</code>	The distribution of values to check.
<code>prop</code>	The total proportion in both tails combined, must be in (0, 1).
<code>greedy</code>	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

---

 show\_cutoffs

*Add Cutoff Markers to a Histogram*


---

**Description**

**[Experimental]**

**Usage**

```
show_cutoffs(plot, color = "#1e3a8a", size = 4, labels = FALSE)
```

**Arguments**

plot	A ggplot histogram with fill mapped to a distribution part function, e.g., fill = ~middle(Thumb, .95).
color	Marker/line color. Default "#1e3a8a".
size	Marker size. Default 4.
labels	Whether to add text annotations explaining the cutoffs. Default FALSE.

**Details**

Adds downward-pointing triangle markers at the empirical quantile cutoffs on a histogram that uses a distribution part function (middle(), tails(), upper(), lower(), or outer()) in its fill aesthetic.

**Value**

A ggplot object with cutoff markers and optional labels.

**Examples**

```
gf_histogram(~Thumb, data = Fingers, fill = ~middle(Thumb, .95)) %>%
  show_cutoffs(labels = TRUE)
```

---

Smallville	<i>Simulated housing data</i>
------------	-------------------------------

---

**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	<i>Split data into train and test sets.</i>
------------	---

---

**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	<i>Students at a university were asked to enter a random number between 1-20 into a survey.</i>
--------	---

---

**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	<i>Tables data</i>
--------	--------------------

---

**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	<i>A simple theme built on top of <code>ggplot2::theme_bw</code></i>
------------------	--

---

### 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	<i>Data from an experiment about smiley faces and tips</i>
---------------	--

---

### 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	<i>Simulated data for an experiment about smiley faces and tips</i>
---------	---

---

### 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	<i>Data on countries from the Happy Planet Index project.</i>
-------	---

---

### 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

## \* datasets

- Ames, 3
- class\_data, 5
- er, 9
- fevdata, 13
- Fingers, 14
- FingersMessy, 15
- game\_data, 17
- penguins, 25
- Smallville, 27
- Survey, 28
- Tables, 28
- tip\_exp, 30
- TipExperiment, 29
- World, 30

Ames, 3

aov(), 17

b (estimate\_extraction), 12

b0 (estimate\_extraction), 12

b1 (estimate\_extraction), 12

class\_data, 5

coursekata\_attach, 5

coursekata\_install, 6

coursekata\_load\_theme, 6

coursekata\_load\_theme(), 9

coursekata\_packages, 7

coursekata\_palette, 7

coursekata\_palette(), 6, 8, 25

coursekata\_palette\_provider, 8

coursekata\_repos, 8

coursekata\_unload\_theme, 9, 29

coursekata\_unload\_theme(), 6

coursekata\_update (coursekata\_install), 6

data.frame, 12

er, 9

estimate\_extraction, 12

f (estimate\_extraction), 12

fevdata, 13

Fingers, 14

FingersMessy, 15

fit\_stats, 16

fitstats (fit\_stats), 16

formula, 12

fVal (estimate\_extraction), 12

game\_data, 17

generate\_models(), 12

gf\_model, 17

gf\_resid, 18

gf\_resid\_fun, 19

gf\_sd\_ruler, 20

gf\_square\_resid, 22

gf\_square\_resid(), 22

gf\_square\_resid\_fun, 23

gf\_squareplot, 21

gf\_squaresid (gf\_square\_resid), 22

ggformula::gf\_hline(), 17

ggformula::gf\_lm(), 17

ggformula::gf\_segment(), 17

ggplot2::geom\_polygon(), 23

ggplot2::geom\_segment(), 19, 20

ggplot2::theme\_bw, 29

lm, 12, 16

lm(), 17

lower (middle), 24

middle, 24

middle(), 24

outer (middle), 24

p (estimate\_extraction), 12

palmerpenguins::penguins, 25

penguins, 25

PRE (estimate\_extraction), 12  
pre (estimate\_extraction), 12

remotes::install\_cran, 6  
remotes::install\_github, 6

scale\_discrete\_coursekata, 25  
show\_cutoffs, 26  
Smallville, 27  
split\_data, 27  
Survey, 28

Tables, 28  
tails (middle), 24  
theme\_coursekata, 29  
theme\_coursekata(), 6  
tip\_exp, 30  
TipExperiment, 29

upper (middle), 24

viridisLite::viridis(), 8

World, 30