Package: summarytabl (via r-universe)

May 26, 2026

Type Package

Title Generate Summary Tables for Categorical, Ordinal, and Continuous Data

Version 0.2.1.9000

Maintainer Ama Nyame-Mensah <ama@anyamemensah.com>

URL https://anyamemensah.github.io/summarytabl/, https://github.com/anyamemensah/summarytabl

BugReports https://github.com/anyamemensah/summarytabl/issues

Description Provides functions for tabulating and summarizing categorical, multiple response, ordinal, and continuous variables in R data frames. Makes it easy to create clear, structured summary tables, so you spend less time wrangling data and more time interpreting it.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports cli, dplyr (>= 1.1.4), purrr (>= 1.1.0), rlang, stats, tibble, tidyr,

RoxygenNote 7.3.3

Suggests knitr, rmarkdown, testthat (>= 3.0.0), tidyselect

VignetteBuilder knitr

Config/testthat/edition 3

Depends R (>= 4.1.0)

Roxygen list(markdown = TRUE)

Config/pak/sysreqs libicu-dev

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

Date/Publication 2026-01-26 01:59:38 UTC

RemoteUrl https://github.com/anyamemensah/summarytabl

RemoteRef HEAD

RemoteSha 34d5534b8e9827dc584423c4fe022e6835b240d8

Contents

`cat_group_tbl` *Summarize two categorical variables*

Description

`cat_group_tbl()` summarizes nominal or categorical variables by a grouping variable, returning frequency counts and percentages.

Usage

```
cat_group_tbl(
  data,
  row_var,
  col_var,
  margins = "all",
  na.rm.row_var = FALSE,
  na.rm.col_var = FALSE,
  pivot = "longer",
  only = NULL,
  ignore = NULL
)
```

Arguments

`data`	A data frame.
`row_var`	A character string of the name of a variable in `data` containing categorical data. This is the primary categorical variable.
`col_var`	A character string of the name of a variable in `data` containing categorical data. This is the secondary categorical variable.

margins	A character string that determines how percentage values are calculated; whether they sum to one across rows, columns, or the entire table (i.e., all). Defaults to all, but can also be set to rows or columns.
na.rm.row_var	A logical value indicating whether missing values for row_var should be removed before calculations. Default is FALSE.
na.rm.col_var	A logical value indicating whether missing values for col_var should be removed before calculations. Default is FALSE.
pivot	A character string that determines the format of the table. By default, longer returns the data in the long format. To return the data in the wide format, specify wider.
only	A character string or vector of character strings of the types of summary data to return. Default is NULL, which returns both counts and percentages. To return only counts or percentages, use count or percent, respectively.
ignore	An optional named vector or list that defines values to exclude from row_var and col_var. If set to NULL (default), all values are retained. To exclude multiple values from row_var or col_var, provide them as a named list.

Value

A tibble showing the count and percentage of each category in row_var by each category in col_var.

Author(s)

Ama Nyame-Mensah

Examples

```
cat_group_tbl(data = nlsy,
              row_var = "gender",
              col_var = "bthwht",
              pivot = "wider",
              only = "count")

cat_group_tbl(data = nlsy,
              row_var = "birthord",
              col_var = "breastfed",
              pivot = "longer")
```

---

cat_tbl *Summarize a categorical variable*

Description

cat_tbl() summarizes nominal or categorical variables, returning frequency counts and percentages.

Usage

```
cat_tbl(data, var, na.rm = FALSE, only = NULL, ignore = NULL)
```

Arguments

data	A data frame.
var	A character string of the name of a variable in data containing categorical data.
na.rm	A logical value indicating whether missing values should be removed before calculations. Default is FALSE.
only	A character string or vector of character strings of the types of summary data to return. Default is NULL, which returns both counts and percentages. To return only counts or percentages, use count or percent, respectively.
ignore	An optional vector that contains values to exclude from var. Default is NULL, which retains all values.

Value

A tibble showing the count and percentage of each category in var

Author(s)

Ama Nyame-Mensah

Examples

```
cat_tbl(data = nlsy, var = "gender")

cat_tbl(data = nlsy, var = "race", only = "count")

cat_tbl(data = nlsy,
        var = "race",
        ignore = "Hispanic",
        only = "percent",
        na.rm = TRUE)
```

---

check_named_vctr *Check a named vector*

Description

This function checks whether named lists and vectors contain invalid values (like NULL or NA), have invalid names (such as missing or empty names), ensures the number of valid names matches the number of supplied values, and confirms that valid names from the object correspond to the provided names. If any of these checks fail, the function returns the default value.

Usage

```
check_named_vctr(x, names, default)
```

Arguments

x	A named vector.
names	A character vector or list of character vectors of length one specifying the names to be matched.
default	Default value to return

Value

Either the original object, x, or the default value.

Author(s)

Ama Nyame-Mensah

Examples

```
# returns NULL
check_named_vctr(x = c(one = 1, two = 2, 3),
                 names = c("one", "two", "three"),
                 default = NULL)

# returns x
check_named_vctr(x = list(one = 1, two = 2, three = 3),
                 names = list("one", "two", "three"),
                 default = NULL)

# also returns x
check_named_vctr(x = c(baako = 1, mmienu = 2, mmiensa = 3),
                 names = list("baako", "mmienu", "mmiensa"),
                 default = NULL)
```

---

depressive *Depressive Symptoms Data*

Description

Subset of data from the National Longitudinal Survey of Youth (NLSY) 1979 Children and Young Adults. This dataset includes survey responses about feelings and behaviors linked to depressive symptoms in children and young adults. For more information about the National Longitudinal Survey of Youth, visit: https://www.nlsinfo.org/.

Usage

```
depressive
```

Format

A data frame with 11,551 rows and 12 columns:

cid Child identification number)

race race of child (1 = Hispanic, 2 = Black, 3 = Non-Black,Non-Hispanic)

sex sex of child (1 = male, 2 = female)

yob year of child's bith

dep_1 how often child feels sad and blue (1 = often, 2 = sometimes, 3 = hardly ever)

dep_2 how often child feels nervous, tense, or on edge (1 = often, 2 = sometimes, 3 = hardly ever)

dep_3 how often child feels happy (1 = often, 2 = sometimes, 3 = hardly ever)

dep_4 how often child feels bored (1 = often, 2 = sometimes, 3 = hardly ever)

dep_5 how often child feels lonely (1 = often, 2 = sometimes, 3 = hardly ever)

dep_6 how often child feels tired or worn out (1 = often, 2 = sometimes, 3 = hardly ever)

dep_7 how often child feels excited about something (1 = often, 2 = sometimes, 3 = hardly ever)

dep_8 how often child feels too busy to get everything (1 = often, 2 = sometimes, 3 = hardly ever)

`mean_group_tbl` *Summarize multiple response variables by group or pattern*

Description

`mean_group_tbl()` calculates summary statistics (i.e., mean, median, standard deviation, minimum, maximum, and count of non-missing values) for continuous (i.e., interval and ratio-level) variables, grouped either by another variable in your dataset or by a matched pattern in the variable names.

Usage

```
mean_group_tbl(
  data,
  var_stem,
  group,
  var_input = "stem",
  regex_stem = FALSE,
  ignore_stem_case = FALSE,
  group_type = "variable",
  group_name = NULL,
  regex_group = FALSE,
  ignore_group_case = FALSE,
  remove_group_non_alnum = TRUE,
  na_removal = "listwise",
  only = NULL,
  var_labels = NULL,
  ignore = NULL
)
```

Arguments

data	A data frame.
var_stem	A character vector with one or more elements, where each represents either a variable stem or the complete name of a variable present in data. A variable 'stem' refers to a common naming pattern shared among related variables, typically reflecting repeated measures of the same idea or a group of items assessing a single concept.
group	A character string representing a variable name or a pattern used to search for variables in data.
var_input	A character string specifying whether the values supplied to var_stem should be treated as variable stems (stem) or as complete variable names (name). By default, this is set to stem, so the function searches for variables that begin with each stem provided. Setting this argument to name directs the function to look for variables that exactly match the provided names.
regex_stem	A logical value indicating whether to use Perl-compatible regular expressions when searching for variable stems. Default is FALSE.
ignore_stem_case	A logical value indicating whether the search for columns matching the supplied var_stem is case-insensitive. Default is FALSE.
group_type	A character string that defines how the group argument should be interpreted. Should be one of pattern or variable. Defaults to variable, which searches for a matching variable name in data.
group_name	An optional character string used to rename the group column in the final table When group_type is set to variable, the column name defaults to the matched variable name from data. When set to pattern, the default column name is group.
regex_group	A logical value indicating whether to use Perl-compatible regular expressions when searching for group variables or matching variable name patterns. Default is FALSE.
ignore_group_case	A logical value specifying whether the search for a grouping variable (if group_type is variable) or for variables matching a pattern (if group_type is pattern) should be case-insensitive. Default is FALSE. Set to TRUE to ignore case.
remove_group_non_alnum	A logical value indicating whether to remove all non-alphanumeric characters (i.e., anything that is not a letter or number) from group. Default is TRUE.
na_removal	A character string specifying how missing values are handled. Must be one of listwise or pairwise. Defaults to listwise. • listwise: Removes any row that has at least one missing value across all variables returned or analyzed. (Effectively uses complete cases only.) • pairwise: Handles missing values per variable or per pair of variables, using all available data, even if other variables in the row have missing values.

only — A character string or vector of character strings specifying which summary statistics to return. Defaults to NULL, which includes mean (mean), median (median) standard deviation (sd), minimum (min), maximum (max), and count of non-missing values (nobs).

var_labels — An optional named character vector or list used to assign custom labels to variable names. Each element must be named and correspond to a variable included in the returned table. If var_input is set to stem, and any element is either unnamed or refers to a variable not present in the table, all labels will be ignored and the table will be printed without them.

ignore — An optional named vector or list indicating values to exclude from variables matching specified stems (or names), and, if applicable, from a grouping variable in data. Defaults to NULL, indicating that all values are retained. To specify exclusions for variables identified by var_stem, use the corresponding stems or variable names as names in the vector or list. To exclude multiple values from these variables or a grouping variable, supply them as a named list.

Value

A tibble showing summary statistics for continuous variables, grouped either by a specified variable in the dataset or by matching patterns in variable names.

Author(s)

Ama Nyame-Mensah

Examples

```
sdoh_child_ages_region <-
  dplyr::select(sdoh, c(REGION, ACS_PCT_AGE_0_4, ACS_PCT_AGE_5_9,
                        ACS_PCT_AGE_10_14, ACS_PCT_AGE_15_17))

mean_group_tbl(data = sdoh_child_ages_region,
               var_stem = "ACS_PCT_AGE",
               group = "REGION",
               group_name = "us_region",
               na_removal = "pairwise",
               var_labels = c(
                 ACS_PCT_AGE_0_4 = "% of population between ages 0-4",
                 ACS_PCT_AGE_5_9 = "% of population between ages 5-9",
                 ACS_PCT_AGE_10_14 = "% of population between ages 10-14",
                 ACS_PCT_AGE_15_17 = "% of population between ages 15-17"))

set.seed(0222)
grouped_data <-
  data.frame(
    symptoms.t1 = sample(c(0:10, -999), replace = TRUE, size = 50),
    symptoms.t2 = sample(c(NA, 0:10, -999), replace = TRUE, size = 50)
  )

mean_group_tbl(data = grouped_data,
               var_stem = "symptoms",
```

```
  group = ".t\\d",
  group_type = "pattern",
  na_removal = "listwise",
  ignore = c(symptoms = -999))
```

`mean_tbl` *Summarize continuous variables*

Description

`mean_tbl()` calculates summary statistics (i.e., mean, median, standard deviation, minimum, maximum, and count of non-missing values) for continuous (i.e., interval and ratio-level) variables.

Usage

```
mean_tbl(
  data,
  var_stem,
  var_input = "stem",
  regex_stem = FALSE,
  ignore_stem_case = FALSE,
  na_removal = "listwise",
  only = NULL,
  var_labels = NULL,
  ignore = NULL
)
```

Arguments

`data`	A data frame.
`var_stem`	A character vector with one or more elements, where each represents either a variable stem or the complete name of a variable present in `data`. A variable 'stem' refers to a common naming pattern shared among related variables, typically reflecting repeated measures of the same idea or a group of items assessing a single concept.
`var_input`	A character string specifying whether the values supplied to `var_stem` should be treated as variable stems (`stem`) or as complete variable names (`name`). By default, this is set to `stem`, so the function searches for variables that begin with each stem provided. Setting this argument to `name` directs the function to look for variables that exactly match the provided names.
`regex_stem`	A logical value indicating whether to use Perl-compatible regular expressions when searching for variable stems. Default is `FALSE`.
`ignore_stem_case`	A logical value indicating whether the search for columns matching the supplied `var_stem` is case-insensitive. Default is `FALSE`.

`na_removal` A character string specifying how missing values are handled. Must be one of `listwise` or `pairwise`. Defaults to `listwise`.

- `listwise`: Removes any row that has at least one missing value across all variables returned or analyzed. (Effectively uses complete cases only.)
- `pairwise`: Handles missing values per variable or per pair of variables, using all available data, even if other variables in the row have missing values.

`only` A character string or vector of character strings specifying which summary statistics to return. Defaults to `NULL`, which includes mean (mean), median (median) standard deviation (sd), minimum (min), maximum (max), and count of non-missing values (nobs).

`var_labels` An optional named character vector or list used to assign custom labels to variable names. Each element must be named and correspond to a variable included in the returned table. If `var_input` is set to `stem`, and any element is either unnamed or refers to a variable not present in the table, all labels will be ignored and the table will be printed without them.

`ignore` An optional named vector or list indicating values to exclude from variables matching specified stems (or names). Defaults to `NULL`, indicating that all values are retained. To specify exclusions for variables identified by `var_stem`, use the corresponding stems or variable names as names in the vector or list. To exclude multiple values from these variables, supply them as a named list.

Value

A tibble showing summary statistics for continuous variables.

Author(s)

Ama Nyame-Mensah

Examples

```
sdoh_child_ages <-
  dplyr::select(sdoh, c(ACS_PCT_AGE_0_4, ACS_PCT_AGE_5_9,
                        ACS_PCT_AGE_10_14, ACS_PCT_AGE_15_17))

mean_tbl(data = sdoh_child_ages, var_stem = "ACS_PCT_AGE")

mean_tbl(data = sdoh_child_ages,
         var_stem = "ACS_PCT_AGE",
         na_removal = "pairwise",
         var_labels = c(
           ACS_PCT_AGE_0_4 = "% of population between ages 0-4",
           ACS_PCT_AGE_5_9 = "% of population between ages 5-9",
           ACS_PCT_AGE_10_14 = "% of population between ages 10-14",
           ACS_PCT_AGE_15_17 = "% of population between ages 15-17"))
```

nlsy *National Longitudinal Survey of Youth (NLSY) Data*

Description

These data are a subset from the National Longitudinal Survey of Youth (NLSY) 1979 Children and Young Adults.The data contains 2,976 observations and 10 variables.

For more information about the National Longitudinal Survey of Youth, visit https://www.nlsinfo.org/.

Usage

```
nlsy
```

Format

A tibble with 2,976 rows and 11 columns:

CID Child identification number)

race race of child (Hispanic, Black, Non-Black,Non-Hispanic)

gender gender of child (1 = male, 0 = female)

birthord birth order of child

magebirth Age of mother at birth of child

bthwht whether child was born low birth weight (1 = yes, 0 = no)

breastfed whether child was breastfed (1 = yes, 0 = no)

medu Highest grade completed by child's mother

math PIAT Math Standard Score

read PIAT Reading Recognition Standard Score

hhnum Number of household members in household

sdoh *2020 Social Determinants of Health (SDOH) Data*

Description

Subset of data from the 2020 Social Determinants of Health (SDOH) Database. For more information about the 2020 SDOH Database, visit: https://www.ahrq.gov/sdoh/index.html.

Usage

```
sdoh
```

Format

A tibble with 3,229 rows and 29 columns:

YEAR SDOH file year

COUNTYFIPS State-county FIPS Code (5-digit)

STATEFIPS State FIPS Code (2-digit)

STATE State name

COUNTY County name

REGION Census region name

TERRITORY Territory indicator (1= U.S. Territory, 0= U.S. State or DC)

ACS_PCT_AGE_0_4 Percentage of population between ages 0-4

ACS_PCT_AGE_5_9 Percentage of population between ages 5-9

ACS_PCT_AGE_10_14 Percentage of population between ages 10-14

ACS_PCT_AGE_15_17 Percentage of population between ages 15-17

NOAAC_PRECIPITATION_JAN Monthly (January) precipitation (Inches)

NOAAC_PRECIPITATION_FEB Monthly (February) precipitation (Inches)

NOAAC_PRECIPITATION_MAR Monthly (March) precipitation (Inches)

NOAAC_PRECIPITATION_APR Monthly (April) precipitation (Inches)

NOAAC_PRECIPITATION_MAY Monthly (May) precipitation (Inches)

NOAAC_PRECIPITATION_JUN Monthly (June) precipitation (Inches)

NOAAC_PRECIPITATION_JUL Monthly (July) precipitation (Inches)

NOAAC_PRECIPITATION_AUG Monthly (August) precipitation (Inches)

NOAAC_PRECIPITATION_SEP Monthly (September) precipitation (Inches)

NOAAC_PRECIPITATION_OCT Monthly (October) precipitation (Inches)

NOAAC_PRECIPITATION_NOV Monthly (November) precipitation (Inches)

NOAAC_PRECIPITATION_DEC Monthly (December) precipitation (Inches)

HHC_PCT_HHA_NURSING Percentage of home health agencies offering nursing care services

HHC_PCT_HHA_PHYS_THERAPY Percentage of home health agencies offering physical therapy services

HHC_PCT_HHA_OCC_THERAPY Percentage of home health agencies offering occupational therapy services

HHC_PCT_HHA_SPEECH Percentage of home health agencies offering speech pathology services

HHC_PCT_HHA_MEDICAL Percentage of home health agencies offering medical social services

HHC_PCT_HHA_AIDE Percentage of home health agencies offering home health aide services

`select_group_tbl` *Summarize multiple response variables by group or pattern*

Description

`select_group_tbl()` displays frequency counts and percentages for multiple response variables (e.g., a series of questions where participants answer "Yes" or "No" to each item) as well as ordinal variables (such as Likert or Likert-type items with responses ranging from "Strongly Disagree" to "Strongly Agree", where respondents select one response per statement, question, or item), grouped either by another variable in your dataset or by a matched pattern in the variable names.

Usage

```
select_group_tbl(
  data,
  var_stem,
  group,
  var_input = "stem",
  regex_stem = FALSE,
  ignore_stem_case = FALSE,
  group_type = "variable",
  group_name = NULL,
  margins = "all",
  regex_group = FALSE,
  ignore_group_case = FALSE,
  remove_group_non_alnum = TRUE,
  na_removal = "listwise",
  pivot = "longer",
  only = NULL,
  var_labels = NULL,
  ignore = NULL,
  force_pivot = FALSE
)
```

Arguments

`data`	A data frame.
`var_stem`	A character vector with one or more elements, where each represents either a variable stem or the complete name of a variable present in `data`. A variable 'stem' refers to a common naming pattern shared among related variables, typically reflecting repeated measures of the same idea or a group of items assessing a single concept.
`group`	A character string representing a variable name or a pattern used to search for variables in `data`.
`var_input`	A character string specifying whether the values supplied to `var_stem` should be treated as variable stems (`stem`) or as complete variable names (`name`). By

default, this is set to stem, so the function searches for variables that begin with each stem provided. Setting this argument to name directs the function to look for variables that exactly match the provided names.

regex_stem
: A logical value indicating whether to use Perl-compatible regular expressions when searching for variable stems. Default is FALSE.

ignore_stem_case
: A logical value indicating whether the search for columns matching the supplied var_stem is case-insensitive. Default is FALSE.

group_type
: A character string that defines how the group argument should be interpreted. Should be one of pattern or variable. Defaults to variable, which searches for a matching variable name in data.

group_name
: An optional character string used to rename the group column in the final table When group_type is set to variable, the column name defaults to the matched variable name from data. When set to pattern, the default column name is group.

margins
: A character string that determines how percentage values are calculated; whether they sum to one across rows, columns, or the entire variable (i.e., all). Defaults to all, but can also be set to rows or columns. Note: This argument only affects the final table when group_type is variable.

regex_group
: A logical value indicating whether to use Perl-compatible regular expressions when searching for group variables or matching variable name patterns. Default is FALSE.

ignore_group_case
: A logical value specifying whether the search for a grouping variable (if group_type is variable) or for variables matching a pattern (if group_type is pattern) should be case-insensitive. Default is FALSE. Set to TRUE to ignore case.

remove_group_non_alnum
: A logical value indicating whether to remove all non-alphanumeric characters (i.e., anything that is not a letter or number) from group. Default is TRUE.

na_removal
: A character string specifying how missing values are handled. Must be one of listwise or pairwise. Defaults to listwise.

  - listwise: Removes any row that has at least one missing value across all variables returned or analyzed. (Effectively uses complete cases only.)
  - pairwise: Handles missing values per variable or per pair of variables, using all available data, even if other variables in the row have missing values.

pivot
: A character string that determines the format of the table. By default, longer returns the data in the long format. To return the data in the wide format, specify wider.

only
: A character string or vector of character strings of the types of summary data to return. Default is NULL, which returns both counts and percentages. To return only counts or percentages, use count or percent, respectively.

var_labels
: An optional named character vector or list used to assign custom labels to variable names. Each element must be named and correspond to a variable included in the returned table. If var_input is set to stem, and any element is either

unnamed or refers to a variable not present in the table, all labels will be ignored and the table will be printed without them.

`ignore` An optional named vector or list indicating values to exclude from variables matching specified stems (or names), and, if applicable, from a grouping variable in `data`. Defaults to `NULL`, indicating that all values are retained. To specify exclusions for variables identified by `var_stem`, use the corresponding stems or variable names as names in the vector or list. To exclude multiple values from these variables or a grouping variable, supply them as a named list.

`force_pivot` A logical value that enables pivoting to the 'wider' format even when variables have inconsistent value sets. By default, this is set to `FALSE` to prevent reshaping errors when values differ across variables in the returned table. Set to `TRUE` to override this safeguard and pivot to the 'wider' format regardless of value inconsistencies.

Value

A tibble displaying the count and percentage for each category in a multiple response variable, grouped either by a specified variable in the dataset or by matching patterns in variable names.

Author(s)

Ama Nyame-Mensah

Examples

```
select_group_tbl(data = stem_social_psych,
                 var_stem = "belong_belong",
                 group = "\\d",
                 group_type = "pattern",
                 group_name = "wave",
                 na_removal = "pairwise",
                 pivot = "wider",
                 only = "count")

tas_recoded <-
  tas |>
  dplyr::mutate(sex = dplyr::case_when(
    sex == 1 ~ "female",
    sex == 2 ~ "male",
    TRUE ~ NA)) |>
  dplyr::mutate(dplyr::across(
    .cols = dplyr::starts_with("involved_"),
    .fns = ~ dplyr::case_when(
      .x == 1 ~ "selected",
      .x == 0 ~ "unselected",
      TRUE ~ NA)
  ))

select_group_tbl(data = tas_recoded,
                 var_stem = "involved_",
                 group = "sex",
```

```
                  group_type = "variable",
                  na_removal = "pairwise",
                  pivot = "wider")

depressive_recoded <-
  depressive |>
  dplyr::mutate(sex = dplyr::case_when(
    sex == 1 ~ "male",
    sex == 2 ~ "female",
    TRUE ~ NA)) |>
  dplyr::mutate(dplyr::across(
    .cols = dplyr::starts_with("dep_"),
    .fns = ~ dplyr::case_when(
      .x == 1 ~ "often",
      .x == 2 ~ "sometimes",
      .x == 3 ~ "hardly",
      TRUE ~ NA
    )
  ))

select_group_tbl(data = depressive_recoded,
                  var_stem = "dep",
                  group = "sex",
                  group_type = "variable",
                  na_removal = "listwise",
                  pivot = "wider",
                  only = "percent",
                  var_labels =
                    c(dep_1 = "how often child feels sad and blue",
                      dep_2 = "how often child feels nervous, tense, or on edge",
                      dep_3 = "how often child feels happy",
                      dep_4 = "how often child feels bored",
                      dep_5 = "how often child feels lonely",
                      dep_6 = "how often child feels tired or worn out",
                      dep_7 = "how often child feels excited about something",
                      dep_8 = "how often child feels too busy to get everything"))
```

`select_tbl` *Summarize multiple response variables*

Description

`select_tbl()` displays frequency counts and percentages for multiple response variables (e.g., a series of questions where participants answer "Yes" or "No" to each item) as well as ordinal variables (such as Likert or Likert-type items with responses ranging from "Strongly Disagree" to "Strongly Agree", where respondents select one response per statement, question, or item).

Usage

```
select_tbl(
  data,
  var_stem,
  var_input = "stem",
  regex_stem = FALSE,
  ignore_stem_case = FALSE,
  na_removal = "listwise",
  pivot = "longer",
  only = NULL,
  var_labels = NULL,
  ignore = NULL,
  force_pivot = FALSE
)
```

Arguments

data	A data frame.
var_stem	A character vector with one or more elements, where each represents either a variable stem or the complete name of a variable present in data. A variable 'stem' refers to a common naming pattern shared among related variables, typically reflecting repeated measures of the same idea or a group of items assessing a single concept.
var_input	A character string specifying whether the values supplied to var_stem should be treated as variable stems (stem) or as complete variable names (name). By default, this is set to stem, so the function searches for variables that begin with each stem provided. Setting this argument to name directs the function to look for variables that exactly match the provided names.
regex_stem	A logical value indicating whether to use Perl-compatible regular expressions when searching for variable stems. Default is FALSE.
ignore_stem_case	A logical value indicating whether the search for columns matching the supplied var_stem is case-insensitive. Default is FALSE.
na_removal	A character string specifying how missing values are handled. Must be one of listwise or pairwise. Defaults to listwise. • listwise: Removes any row that has at least one missing value across all variables returned or analyzed. (Effectively uses complete cases only.) • pairwise: Handles missing values per variable or per pair of variables, using all available data, even if other variables in the row have missing values.
pivot	A character string that determines the format of the table. By default, longer returns the data in the long format. To receive the data in the wide format, specify wider.
only	A character string or vector of character strings of the types of summary data to return. Default is NULL, which returns both counts and percentages. To return only counts or percentages, use count or percent, respectively.

`var_labels` An optional named character vector or list used to assign custom labels to variable names. Each element must be named and correspond to a variable included in the returned table. If `var_input` is set to `stem`, and any element is either unnamed or refers to a variable not present in the table, all labels will be ignored and the table will be printed without them.

`ignore` An optional named vector or list indicating values to exclude from variables matching specified stems (or names). Defaults to `NULL`, indicating that all values are retained. To specify exclusions for variables identified by `var_stem`, use the corresponding stems or variable names as names in the vector or list. To exclude multiple values from these variables, supply them as a named list.

`force_pivot` A logical value that enables pivoting to the 'wider' format even when variables have inconsistent value sets. By default, this is set to `FALSE` to prevent reshaping errors when values differ across variables in the returned table. Set to `TRUE` to override this safeguard and pivot to the 'wider' format regardless of value inconsistencies.

Value

A tibble displaying the count and percentage for each category in a multiple response variable.

Author(s)

Ama Nyame-Mensah

Examples

```
select_tbl(data = tas,
           var_stem = "involved_",
           na_removal = "pairwise")

select_tbl(data = depressive,
           var_stem = "dep",
           na_removal = "listwise",
           pivot = "wider",
           only = "percent")

var_label_example <-
  c(dep_1 = "how often child feels sad and blue",
    dep_2 = "how often child feels nervous, tense, or on edge",
    dep_3 = "how often child feels happy",
    dep_4 = "how often child feels bored",
    dep_5 = "how often child feels lonely",
    dep_6 = "how often child feels tired or worn out",
    dep_7 = "how often child feels excited about something",
    dep_8 = "how often child feels too busy to get everything")

select_tbl(data = depressive,
           var_stem = "dep",
           na_removal = "pairwise",
           pivot = "longer",
           var_labels = var_label_example)
```

```
select_tbl(data = depressive,
           var_stem = "dep",
           na_removal = "pairwise",
           pivot = "wider",
           only = "count",
           var_labels = var_label_example)
```

`social_psy_data` *Social Psychological (Simulated) Data*

Description

Simulated data capturing social psychological responses in a real-world college setting. This dataset represents college students' feelings, attitudes, and perceptions related to their experiences in STEM degree programs. It was designed to reflect key psychological factors that influence student engagement, motivation, and persistence in STEM fields.

Usage

```
social_psy_data
```

Format

A tibble with 10,200 rows and 17 columns:

id participant id number)

belong_1 I feel like I belong at this institution (1=Strongly Disagree, 2=Disagree,3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

belong_2 I feel like part of the community (1=Strongly Disagree, 2=Disagree,3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

belong_3 I feel valued by this institution (1=Strongly Disagree, 2=Disagree,3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

identity_1 This institution is a big part of who I am (1=Strongly Disagree,2=Disagree,3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

identity_2 I feel comfortable being myself in this setting (1=Strongly Disagree,2=Disagree,3=Neither agree nor disagree,4=Agree, 5=Strongly Agree)

identity_3 This institution is a big part of who I am (1=Strongly Disagree, 2=Disagree,3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

identity_4 I care about doing well at this institution (1=Strongly Disagree, 2=Disagree,3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

selfEfficacy_1 I am confident about A (1=Strongly Disagree,2=Disagree, 3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

selfEfficacy_2 I am confident about B (1=Strongly Disagree,2=Disagree, 3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

selfEfficacy_3 I am confident about C (1=Strongly Disagree,2=Disagree, 3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

selfEfficacy_4 I am confident about D (1=Strongly Disagree,2=Disagree, 3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

selfEfficacy_5 I am confident about E (1=Strongly Disagree,2=Disagree, 3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

selfEfficacy_6 I am confident about F (1=Strongly Disagree,2=Disagree, 3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

selfEfficacy_7 I am confident about G (1=Strongly Disagree,2=Disagree, 3=Neither agree nor disagree,4=Agree,5=Strongly Agree)

gender Participant's gender identity (1=Woman,2=Man,3=Non-binary, 4=Self-identify,5=Transgender,6=Genderqueer/non-conforming)

citizen Participant's citizenship status (1=U.S. citizen,2=Non-U.S. citizen with permanent residency,3=Non-U.S. citizen with temporary visa,4=Other)

`stem_social_psych` *STEM Social Psychological (Simulated) Data*

Description

Simulated data designed to reflect social psychological responses among college students. These data were generated to model attitudes, perceptions, and experiences of students participating in a Science, Technology, Engineering, and Mathematics (STEM) intervention program. The dataset aims to represent real- world psychological factors relevant to STEM education contexts.

Usage

```
stem_social_psych
```

Format

A tibble with 786 rows and 37 columns:

id student id number)

belong_belongStem_w1 I feel like I belong in STEM (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

belong_outsiderStem_w1 I feel like an outsider in STEM (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

identity_identityStem_w1 STEM is a big part of who I am. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

belong_welcomedStem_w1 I feel welcomed in STEM workplaces (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

identity_noCommonStem_w1 I do not have much in common with the other students in my STEM classes.(1=Strongly disagree,2=Somewhat disagree,3=Neither disagree nor agree, 4=Somewhat agree, 5=Strongly agree)

selfEfficacy_passStemCourses_w1 pass my STEM courses.(1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_learnConcepts_w1 learn the foundations and concepts of scientific thinking. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree, 4=Somewhat agree, 5=Strongly agree)

selfEfficacy_stemField_w1 do well in a stem-related field. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

selfEfficacy_learnScience_w1 quickly learn new science areas, systems, techniques or concepts on my own. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree, 4=Somewhat agree, 5=Strongly agree)

selfEfficacy_contributeProject_w1 contribute to a science project. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_commScience_w1 clearly communicate scientific problems and findings to varied audiences (1=Strongly disagree,2=Somewhat disagree, 3=Neither disagree nor agree, 4=Somewhat agree,5=Strongly agree)

selfEfficacy_scientist_w1 become a scientist. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

selfEfficacy_completeUG_w1 complete an undergraduate STEM degree. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_admitGrad_w1 get admitted to a graduate STEM program. (1=Strongly disagree,2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_successGrad_w1 be successful in a graduate STEM program. (1=Strongly disagree,2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

belong_belongStem_w2 I feel like I belong in STEM (1=Strongly disagree, 2=Somewhat disagree, 3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

belong_outsiderStem_w2 I feel like an outsider in STEM. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

identity_identityStem_w2 STEM is a big part of who I am. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

belong_welcomedStem_w2 I feel welcomed in STEM workplaces. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

identity_noCommonStem_w2 I do not have much in common with the other students in my STEM classes.(1=Strongly disagree,2=Somewhat disagree,3=Neither disagree nor agree, 4=Somewhat agree, 5=Strongly agree)

selfEfficacy_passStemCourses_w2 pass my STEM courses. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_learnConcepts_w2 learn the foundations and concepts of scientific thinking. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree, 4=Somewhat agree, 5=Strongly agree)

selfEfficacy_stemField_w2 do well in a stem-related field. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

selfEfficacy_learnScience_w2 quickly learn new science areas, systems, techniques or concepts on my own. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree, 4=Somewhat agree, 5=Strongly agree)

selfEfficacy_contributeProject_w2 contribute to a science project. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_commScience_w2 clearly communicate scientific problems and findings to varied audiences (1=Strongly disagree,2=Somewhat disagree, 3=Neither disagree nor agree, 4=Somewhat agree,5=Strongly agree)

selfEfficacy_scientist_w2 become a scientist. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree,5=Strongly agree)

selfEfficacy_completeUG_w2 complete an undergraduate STEM degree. (1=Strongly disagree, 2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_admitGrad_w2 get admitted to a graduate STEM program. (1=Strongly disagree,2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

selfEfficacy_successGrad_w2 be successful in a graduate STEM program. (1=Strongly disagree,2=Somewhat disagree,3=Neither disagree nor agree,4=Somewhat agree, 5=Strongly agree)

is_male Participant's current sex (0=Not Male,1=Male)

has_disability Whether participant has a disability (0=No, 1=Yes)

firstGen Whether participant is a first generation college student (0=No, 1=Yes)

stemMajor Whether participant is a STEM Major (0=No, 1=Yes)

expLearning Whether student has participated in an experiential learning program, such as an internship, research, or leadership opportunity. (0=No, 1=Yes)

urm Whether participant is Asian, Middle Eastern/Arab or White (0) vs. Black, Indigenous, Hispanic/Latino, or Mixed Race (1)

`tas` *Panel Study of Income Dynamics (PSID) Transition into Adulthood Supplement (TAS) Data*

Description

Subset of data from the Panel Study of Income Dynamics (PSID) Transition into Adulthood Supplement. This dataset includes information from young adults about how they spend their free time, including participation in organized activities such as clubs, sports or athletic teams, social-action groups, and other structured extracurricular engagements. For more information about the Panel Study of Income Dynamics, visit: https://psidonline.isr.umich.edu/GettingStarted.aspx.

Usage

```
tas
```

Format

A tibble with 2,526 rows and 8 columns:

pid personal identification number)

sex sex of individual (1 = female, 2 = male)

involved_arts whether the individual participated in any organized activities related to art, music, or the theater in the last 12 months (1 = yes, 0 = no)

involved_sports whether the individual was a member of any athletic or sports teams in the last 12 months (1 = yes, 0 = no)

involved_schoolClubs whether the individual was involved with any high school or college clubs or student government in the last 12 months (1 = yes, 0 = no)

involved_election whether the individual voted in the national election in November 2016 that was held to elect the President (1 = yes, 0 = no)

involved_socialActionGrps whether the individual was involved in any political groups, solidarity or ethnic-support groups or social-action groups in the last 12 months (1 = yes, 0 = no)

involved_volunteer whether the individual was involved in any unpaid volunteer or community service work in the last 12 months (1 = yes, 0 = no)

Index