Package 'tangram.pipe'

October 14, 2022

Title Row-by-Row Table Building

Version 1.1.2

Description Builds tables with customizable rows. Users can specify the type
of data to use for each row, as well as how to handle missing
data and the types of comparison tests to run on the table
columns.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.1

Depends dplyr

Suggests knitr, kableExtra, rmarkdown

VignetteBuilder knitr

URL https://github.com/thomasgstewart/tangram.pipe

BugReports https://github.com/thomasgstewart/tangram.pipe/issues

NeedsCompilation no

Author Andrew Guide [aut, cre],
Thomas Stewart [aut]

Maintainer Andrew Guide <andrew.guide@vumc.org>

Repository CRAN

Date/Publication 2022-08-17 17:10:02 UTC

R topics documented:

---

binary_count   *Count summary for a Binary Row*

Description

Summarizes a binary row using counts.

Usage

```
binary_count(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

reference : the name of the row category to use as the reference. Default will use alphabetical first category

ref.label : choice of whether you want the reference label to be in the table. Default is on and includes reference label; off switches it off.

`rowlabel` : the label for the table row name, if different from row_var.

`compact` : if TRUE, data displayed in one row.

`missing` : if TRUE, missing data is considered; FALSE only uses complete cases.

`digits` : significant digits to use.

Value

A dataframe with summary statistics for a binary variable.

See Also

Possible summary functions for binary data:binary_default, binary_pct, binary_jama

---

binary_default	*Default summary for a Binary Row*

Description

Summarizes a binary row using counts and column proportions.

Usage

```
binary_default(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of `tangram.pipe`. Additional arguments should be supplied for this function to work properly.

`reference` : the name of the row category to use as the reference. Default will use alphabetical first category

`ref.label` : choice of whether you want the reference label to be in the table. Default is on and includes reference label; `off` switches it off.

`rowlabel` : the label for the table row name, if different from row_var.

`compact` : if TRUE, data displayed in one row.

`missing` : if TRUE, missing data is considered; FALSE only uses complete cases.

`digits` : significant digits to use.

Value

A dataframe with summary statistics for a binary variable.

See Also

Additional prewritten summary functions for binary data: binary_pct, binary_count, binary_jama

---

binary_diff   *Binary Difference in Proportions*

Description

Default comparison function for binary data.

Usage

```
binary_diff(dt, num_col, reference, digits)
```

Arguments

dt	the name of the dataframe object.
num_col	the number of categorical columns in the data.
reference	the name of the reference row category to use.
digits	significant digits to use.

Value

A dataframe with difference in proportions test results between pairs of columns for binary data, as well as an overall chi-squared test across all groups.

---

binary_jama   *JAMA-style summary for a Binary Row*

Description

Summarizes a binary row using column percentages and the total number in each cell divided by the column total. This is the style used by the Journal of the American Medical Association.

Usage

```
binary_jama(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

reference : the name of the row category to use as the reference. Default will use alphabetical first category

ref.label : choice of whether you want the reference label to be in the table. Default is on and includes reference label; off switches it off.

rowlabel : the label for the table row name, if different from row_var.

compact : if TRUE, data displayed in one row.

missing : if TRUE, missing data is considered; FALSE only uses complete cases.

digits : significant digits to use.

Value

A dataframe with summary statistics for a binary variable.

See Also

Possible summary functions for binary data:binary_default, binary_pct, binary_count

---

binary_or	*Binary Odds Ratio*

Description

Calculates odds ratio across categories for binary data.

Usage

```
binary_or(dt, num_col, reference, digits)
```

Arguments

dt	the name of the dataframe object.
num_col	the number of categorical columns in the data.
reference	the name of the reference row category to use.
digits	significant digits to use.

Value

A dataframe with computed odds ratios between pairs of columns for binary data, as well as an overall chi-squared test across all groups.

---

binary_pct   *Percentage summary for a Binary Row*

Description

Summarizes a binary row using counts and column percentages.

Usage

```
binary_pct(dt, ...)
```

Arguments

dt   the name of the dataframe object.

...   Additional arguments supplied within the package row functions.

Details

This is an internal function of `tangram.pipe`. Additional arguments should be supplied for this function to work properly.

`reference` : the name of the row category to use as the reference. Default will use alphabetical first category

`ref.label` : choice of whether you want the reference label to be in the table. Default is on and includes reference label; `off` switches it off.

`rowlabel` : the label for the table row name, if different from row_var.

`compact` : if TRUE, data displayed in one row.

`missing` : if TRUE, missing data is considered; FALSE only uses complete cases.

`digits` : significant digits to use.

Value

A dataframe with summary statistics for a binary variable.

See Also

Possible summary functions for binary data:binary_default, binary_count, binary_jama

---

binary_row   *Binary Row*

Description

Adds in a binary row to a `tangram.pipe` table.

Usage

```
binary_row(
  list_obj,
  row_var,
  col_var = NULL,
  newdata = FALSE,
  ref.label = "on",
  rowlabel = NULL,
  summary = NULL,
  reference = NULL,
  compact = TRUE,
  missing = NULL,
  overall = NULL,
  comparison = NULL,
  digits = NULL,
  indent = 5
)
```

Arguments

list_obj	the name of the `tbl_start` object previously initialized.
row_var	the name of the variable to be used in the rows.
col_var	the variable to be used in the table columns. Default is from initialized `tbl_start` object.
newdata	enter new dataset name if different from that initialized in `tbl_start`.
ref.label	toggles the reference label in the table. Default is on, which displays the reference; off switches it off. Only relevant if a compact row is used.
rowlabel	the label for the table row name, if different from `row_var`.
summary	summary function for the data, if different from the one supplied in `tbl_start`.
reference	the name of the row category to use as the reference. Default will use alphabetical first category.
compact	logical: if TRUE, data displayed in one row.
missing	logical: if TRUE, missing data is considered; FALSE only uses complete cases.
overall	logical: if TRUE, an overall column is included.
comparison	the name of the comparison test to use, if different from that initialized in `tbl_start`.

digits	significant digits to use.
indent	number of spaces to indent category names.

Value

A list with the binary row's table information added as a new element to `list_obj`.

See Also

Possible summary functions for binary data:binary_default, binary_pct, binary_count, binary_jama

Other related row-building functions: num_row, cat_row, n_row, empty_row

Starting a `tangram.pipe` table: tbl_start

Examples

```
iris$color <- sample(c("Blue", "Purple"), size=150, replace=TRUE)
x <- tbl_start(iris, "Species", missing=TRUE, overall=TRUE, comparison=TRUE) %>%
  binary_row("color", rowlabel="Color")
```

---

binary_rr	*Binary Risk Ratio*

Description

Calculates risk ratio across categories for binary data.

Usage

```
binary_rr(dt, num_col, reference, digits)
```

Arguments

dt	the name of the dataframe object.
num_col	the number of categorical columns in the data.
reference	the name of the reference row category to use.
digits	significant digits to use.

Value

A dataframe with computed risk ratios between pairs of columns for binary data, as well as an overall chi-squared test across all groups.

cat_comp_default	*Chi-Squared Test for Categorical Variables*

Description

Default comparison function for categorical data.

Usage

```
cat_comp_default(dt, digits)
```

Arguments

dt	the name of the dataframe object.
digits	significant digits to use.

Value

A dataframe calculating relative entropy between column pairs, as well as an overall chi-squared test across all groups.

cat_count	*Count summary for a Categorical Row*

Description

Summarizes a categorical row using counts.

Usage

```
cat_count(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of `tangram.pipe`. Additional arguments should be supplied for this function to work properly.

`rowlabel` : the label for the table row name, if different from row_var.

`missing` : if TRUE, missing data is considered; FALSE only uses complete cases.

`ordering` : Sorts the row variable: options are "ascending" or "descending"

`sortvar` : Column to sort row on. Requires `ordering` to be `ascending` or `descending`. By default, will sort based on overall statistics.

`digits` : significant digits to use.

Value

A dataframe with summary statistics for a categorical variable.

See Also

Additional prewritten summary functions for categorical data: cat_default, cat_pct, cat_jama

---

cat_default *Default summary for a Categorical Row*

Description

Summarizes a categorical row using counts and column proportions.

Usage

```
cat_default(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

rowlabel : the label for the table row name, if different from row_var.

missing : if TRUE, missing data is considered; FALSE only uses complete cases.

ordering : Sorts the row variable: options are "ascending" or "descending"

sortvar : Column to sort row on. Requires ordering to be ascending or descending. By default, will sort based on overall statistics.

digits : significant digits to use.

Value

A dataframe with summary statistics for a categorical variable.

See Also

Additional prewritten summary functions for categorical data: cat_pct, cat_count, cat_jama

---

cat_jama   *JAMA-style summary for a Categorical Row*

Description

Summarizes a categorical row using column percentages and the total number in each cell divided by the column total. This is the style used by the Journal of the American Medical Association.

Usage

```
cat_jama(dt, ...)
```

Arguments

dt   the name of the dataframe object.

...   Additional arguments supplied within the package row functions.

Details

This is an internal function of `tangram.pipe`. Additional arguments should be supplied for this function to work properly.

`rowlabel` : the label for the table row name, if different from row_var.

`missing` : if TRUE, missing data is considered; FALSE only uses complete cases.

`ordering` : Sorts the row variable: options are "ascending" or "descending"

`sortvar` : Column to sort row on. Requires `ordering` to be `ascending` or `descending`. By default, will sort based on overall statistics.

`digits` : significant digits to use.

Value

A dataframe with summary statistics for a categorical variable.

See Also

Additional prewritten summary functions for categorical data: cat_default, cat_pct, cat_count

cat_pct

Percentage summary for a Categorical Row

Description

Summarizes a categorical row using counts and column percentages.

Usage

```
cat_pct(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of `tangram.pipe`. Additional arguments should be supplied for this function to work properly.

`rowlabel` : the label for the table row name, if different from row_var.

`missing` : if TRUE, missing data is considered; FALSE only uses complete cases.

`ordering` : Sorts the row variable: options are "ascending" or "descending"

`sortvar` : Column to sort row on. Requires `ordering` to be `ascending` or `descending`. By default, will sort based on overall statistics.

`digits` : significant digits to use.

Value

A dataframe with summary statistics for a categorical variable.

See Also

Additional prewritten summary functions for categorical data: cat_default, cat_count, cat_jama

cat_row	*Categorical Row*

Description

Adds in a categorical row to a `tangram.pipe` table.

Usage

```
cat_row(
  list_obj,
  row_var,
  col_var = NULL,
  newdata = FALSE,
  rowlabel = NULL,
  summary = NULL,
  missing = NULL,
  overall = NULL,
  comparison = NULL,
  digits = NULL,
  ordering = "none",
  sortcol = NULL,
  indent = 5
)
```

Arguments

`list_obj`	the name of the `tbl_start` object previously initialized.
`row_var`	the name of the variable to be used in the rows.
`col_var`	the variable to be used in the table columns. Default is from initialized `tbl_start` object.
`newdata`	enter new dataset name if different from that initialized in `tbl_start`.
`rowlabel`	the label for the table row name, if different from `row_var`.
`summary`	summary function for the data, if different from the one supplied in `tbl_start`.
`missing`	logical: if TRUE, missing data is considered; FALSE only uses complete cases.
`overall`	logical: if TRUE, an overall column is included.
`comparison`	the name of the comparison test to use, if different from that initialized in `tbl_start`.
`digits`	significant digits to use.
`ordering`	If ascending, will sort by overall ascending order; if descending, will sort by overall descending order. Default is no row sorting.
`sortcol`	Column to sort row on. Requires `ordering` to be ascending or descending. By default, will sort based on overall statistics.
`indent`	number of spaces to indent category names.

Value

A list with the categorical row's table information added as a new element to list_obj.

See Also

Possible summary functions for categorical data:cat_default, cat_pct, cat_count, cat_jama

Other related row-building functions: num_row, binary_row, n_row, empty_row

Starting a tangram.pipe table: tbl_start

Examples

```
iris$Stem.Size <- sample(c("Small", "Medium", "Medium", "Large"), size=150, replace=TRUE)
x <- tbl_start(iris, "Species", missing=TRUE, overall=TRUE, comparison=TRUE) %>%
  cat_row("Stem.Size", rowlabel="Stem Size")
```

---

empty_row   *Empty Row*

Description

Produces a empty dividing row in a tangram.pipe table. May have a row header.

Usage

```
empty_row(list_obj, header = NULL)
```

Arguments

list_obj   the name of the tbl_start object previously initialized.

header   a header to include for the empty row.

Value

If a header is included, a list object is returned with a one-element dataframe containing the header as the most recent entry to list_obj. Otherwise, a list is returned containing a blank character as the last element of list_obj.

See Also

Other related row-building functions: num_row, cat_row, binary_row, n_row

Starting a tangram.pipe table: tbl_start

---

num_date *Date summary for a Numeric Row*

Description

Summarizes a numeric row using the five-number summary for a date object.

Usage

```
num_date(dt, ...)
```

Arguments

dt   the name of the dataframe object.

...   Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

rowlabel : the label for the table row name, if different from row_var.

missing : if TRUE, missing data is considered; FALSE only uses complete cases.

Value

A dataframe with summary statistics for a numeric variable.

See Also

Additional prewritten summary functions for numeric data: num_default, num_mean_sd, num_medianiqr, num_minmax

---

num_default *Default summary for a Numeric Row*

Description

Summarizes a numeric row using the five-number summary, mean, and standard deviation.

Usage

```
num_default(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

rowlabel : the label for the table row name, if different from row_var.

missing : if TRUE, missing data is considered; FALSE only uses complete cases.

digits : significant digits to use.

Value

A dataframe with summary statistics for a numeric variable.

See Also

Additional prewritten summary functions for numeric data: num_mean_sd, num_medianiqr, num_minmax, num_date

---

num_diff	*Numeric Difference in Means*

Description

Default comparison function for numeric data.

Usage

```
num_diff(dt, num_col, row_var, digits)
```

Arguments

dt	the name of the dataframe object.
num_col	the number of categorical columns in the data.
row_var	the name of the row variable in the data.
digits	significant digits to use.

Value

A dataframe calculating the difference in means between column pairs, as well as an overall one-way ANOVA across all groups.

num_mean_sd *Mean/SD summary for a Numeric Row*

Description

Summarizes a numeric row using the mean and standard deviation.

Usage

```
num_mean_sd(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

rowlabel : the label for the table row name, if different from row_var.

missing : if TRUE, missing data is considered; FALSE only uses complete cases.

digits : significant digits to use.

Value

A dataframe with summary statistics for a numeric variable.

See Also

Additional prewritten summary functions for numeric data: num_default, num_medianiqr, num_minmax, num_date

num_medianiqr *Median/IQR summary for a Numeric Row*

Description

Summarizes a numeric row using the median and interquartile range.

Usage

```
num_medianiqr(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

rowlabel : the label for the table row name, if different from row_var.

missing : if TRUE, missing data is considered; FALSE only uses complete cases.

digits : significant digits to use.

Value

A dataframe with summary statistics for a numeric variable.

See Also

Additional prewritten summary functions for numeric data: num_default, num_mean_sd, num_minmax, num_date

---

num_minmax   *Min-Max summary for a Numeric Row*

Description

Summarizes a numeric row using the minimum and maximum values.

Usage

```
num_minmax(dt, ...)
```

Arguments

dt	the name of the dataframe object.
...	Additional arguments supplied within the package row functions.

Details

This is an internal function of tangram.pipe. Additional arguments should be supplied for this function to work properly.

rowlabel : the label for the table row name, if different from row_var.

missing : if TRUE, missing data is considered; FALSE only uses complete cases.

digits : significant digits to use.

Value

A dataframe with summary statistics for a numeric variable.

See Also

Additional prewritten summary functions for numeric data: num_default, num_mean_sd, num_medianiqr, num_date

---

num_row   *Numeric Row*

Description

Adds in a numeric row to a `tangram.pipe` table.

Usage

```
num_row(
  list_obj,
  row_var,
  col_var = NULL,
  newdata = FALSE,
  rowlabel = NULL,
  summary = NULL,
  missing = NULL,
  overall = NULL,
  comparison = NULL,
  digits = NULL
)
```

Arguments

list_obj	the name of the `tbl_start` object previously initialized.
row_var	the name of the variable to be used in the rows.
col_var	the variable to be used in the table columns. Default is from initialized `tbl_start` object.
newdata	enter new dataset name if different from that initialized in `tbl_start`.
rowlabel	the label for the table row name, if different from `row_var`.
summary	summary function for the data, if different from the one supplied in `tbl_start`.
missing	logical: if TRUE, missing data is considered; FALSE only uses complete cases.
overall	logical: if TRUE, an overall column is included.
comparison	the name of the comparison test to use, if different from that initialized in `tbl_start`.
digits	significant digits to use.

Value

A list with the numeric row's table information added as a new element to `list_obj`.

See Also

Possible summary functions for numeric data: num_default, num_mean_sd, num_medianiqr, num_minmax, num_date

Other related row-building functions: cat_row, binary_row, n_row, empty_row

Starting a `tangram.pipe` table: tbl_start

Examples

```
x <- tbl_start(iris, "Species", missing=TRUE, overall=TRUE, comparison=TRUE) %>%
  num_row("Sepal.Length", rowlabel="Sepal Length")
```

---

n_row   *Row counter*

Description

Counts the instances of each column variable of the dataframe to be used in a `tangram.pipe` table (if applicable), and gives an overall row count.

Usage

```
n_row(
  list_obj,
  col_var = NULL,
  newdata = FALSE,
  missing = NULL,
  overall = NULL
)
```

Arguments

list_obj	the name of the tbl_start object previously initialized.
col_var	the variable to be used in the table columns. Default is from initialized tbl_start object.
newdata	enter new dataset name if different from that initialized in tbl_start.
missing	logical: if TRUE, missing data in the column variable is considered; FALSE only uses complete cases.
overall	logical: if TRUE, an overall column is included.

Value

A list with the row counts added as a new element to `list_obj`.

See Also

Other related row-building functions: num_row, cat_row, binary_row, empty_row

Starting a tangram.pipe table: tbl_start

Examples

```
x <- tbl_start(iris, "Species", missing=TRUE, overall=TRUE, comparison=TRUE) %>%
  n_row()
```

---

print.tangram.pipe   *Printing a Table*

Description

Prints a finished table created from tangram.pipe.

Usage

```
## S3 method for class 'tangram.pipe'
print(x, ...)
```

Arguments

x	the name of the tbl_start object previously initialized.
...	further arguments passed to or from other methods.

Value

A dataframe object containing the information from the last element of a tangram.pipe class object created using tbl_out(). This is the finalized table object.

Examples

```
iris$color <- sample(c("Blue", "Purple"), size=150, replace=TRUE)
iris$Stem.Size <- sample(c("Small", "Medium", "Medium", "Large"), size=150, replace=TRUE)
iris$Leaf.Color <- "Green"
x <- tbl_start(iris, "Species", missing=TRUE, overall=TRUE, comparison=TRUE) %>%
  num_row("Sepal.Length", rowlabel="Sepal Length") %>%
  empty_row() %>%
  num_row("Sepal.Width", rowlabel="Sepal Width") %>%
  empty_row() %>%
  num_row("Petal.Length", rowlabel="Petal Length") %>%
  empty_row() %>%
  num_row("Petal.Width", rowlabel="Petal Width") %>%
  empty_row() %>%
  cat_row("Stem.Size", rowlabel="Stem Size") %>%
  empty_row() %>%
  binary_row("color", rowlabel="Color") %>%
  tbl_out() %>%
  print()
```

---

tangram_styling   *Tangram Styling*

Description

Used to preprocess a `tangram.pipe` table for HTML formatting.

Usage

```
tangram_styling(df)
```

Arguments

df   The output data frame object to be printed in HTML form.

Value

A dataframe containing HTML formatting code where applicable.

---

tbl_out   *Output Table*

Description

Produces a finalized `tangram.pipe` table.

Usage

```
tbl_out(list_obj)
```

Arguments

list_obj   the name of the tbl_start object previously initialized.

Value

A tangram.pipe class object with the finalized table as a dataframe added as the most recent element
of `list_obj`.

Examples

```
iris$color <- sample(c("Blue", "Purple"), size=150, replace=TRUE)
iris$Stem.Size <- sample(c("Small", "Medium", "Medium", "Large"), size=150, replace=TRUE)
iris$Leaf.Color <- "Green"
x <- tbl_start(iris, "Species", missing=TRUE, overall=TRUE, comparison=TRUE) %>%
  num_row("Sepal.Length", rowlabel="Sepal Length") %>%
  empty_row() %>%
  num_row("Sepal.Width", rowlabel="Sepal Width") %>%
  empty_row() %>%
  num_row("Petal.Length", rowlabel="Petal Length") %>%
  empty_row() %>%
  num_row("Petal.Width", rowlabel="Petal Width") %>%
  empty_row() %>%
  cat_row("Stem.Size", rowlabel="Stem Size") %>%
  empty_row() %>%
  binary_row("color", rowlabel="Color") %>%
  tbl_out()
```

---

tbl_start   *Table Initialization*

Description

Initializes a `tangram.pipe` table by specifying the desired elements and data components.

Usage

```
tbl_start(
  data,
  col_var,
  missing = FALSE,
  overall = TRUE,
  comparison = FALSE,
  digits = 2,
  default_num_summary = num_default,
  default_cat_summary = cat_default,
  default_binary_summary = binary_default
)
```

Arguments

data	The dataset to be used in the table.
col_var	The variable to be used in the table columns. NULL if single summary column desired.
missing	logical: if TRUE, missing data is considered; FALSE only uses complete cases.
overall	logical: if TRUE, an overall column is included.
comparison	logical: if TRUE, a comparison test is conducted between columns.

digits   The default number of digits to use in the table. By default, the package will use
                       2 significant digits.

default_num_summary

                       The default summary function to use for numerical rows. By default, the pack-
                       age will use `num_default()`, but the user can also choose `num_minmax`, `num_medianiqr`,
                       `num_mean_sd`, or write a custom function to use for the rows.

default_cat_summary

                       The default summary function to use for categorical rows. By default, the pack-
                       age will use `cat_default()`, but the user can also choose `cat_pct` or write a
                       custom function to use for the rows.

default_binary_summary

                       The default summary function to use for binary rows. By default, the package
                       will use `binary_default()`, but the user can also choose `binary_pct` or write
                       a custom function to use for the rows.

Value

A list containing separate entries holding information provided in the function's arguments, as well
as a calculated number of column categories to include for the initialized table.

Examples

```
x <- tbl_start(iris, "Species", missing=TRUE, overall=TRUE, comparison=TRUE)
```

Index