

Package ‘retentionflow’

July 16, 2026

Title Retention Flow Tables and Sankey Diagrams

Version 0.1.25

Author Alexander Nanni [aut, cre]

Maintainer Alexander Nanni <alexander.nanni@gmail.com>

Description Creates transition tables, summaries, and interactive Sankey diagrams from longitudinal person-term-state data. Sankey diagrams visualize flows between states with link widths proportional to flow counts; see Kennedy and Sankey (1898) ``The Thermal Efficiency of Steam Engines" <doi:10.1680/imotp.1898.19100> and Schmidt (2008) ``The Sankey Diagram in Energy and Material Flow Management: Part I: History" <doi:10.1111/j.1530-9290.2008.00004.x>. The minimum input schema is one row per person per term with an identifier, term, and categorical state.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.1.0)

Suggests htmlwidgets, plotly, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Repository CRAN

Date/Publication 2026-07-16 13:40:02 UTC

Contents

make_transitions	2
plot_sankey	2
retention_example	5
stem_retention	5
summarize_transitions	6
validate_flow_data	6

Index	8
--------------	----------

make_transitions	Create term-to-term transition records
------------------	--

Description

Converts longitudinal person-term-state data into one row per person per consecutive term pair.

Usage

```
make_transitions(
  data,
  id = "ID",
  term = "Term",
  state = "State",
  term_order = NULL,
  missing_label = "Not enrolled"
)
```

Arguments

data	A data frame with one row per person per term.
id	Name of the identifier column.
term	Name of the term/time-period column.
state	Name of the categorical state column.
term_order	Optional vector giving the intended order of terms.
missing_label	Label used when a person is absent from the next term.

Value

A data frame with ID, from_term, to_term, from_state, to_state, and status.

Examples

```
x <- retention_example(); make_transitions(x)
```

plot_sankey	Plot an interactive Sankey retention diagram
-------------	--

Description

Creates a Plotly Sankey diagram from longitudinal person-term-state data. The default formatting follows the project Sankey style: term labels across the top, wide HTML-oriented output, fixed node colors for target state, graduation states, not enrolled, and other states, and detailed hover text.

Usage

```
plot_sankey(
  data,
  id = "ID",
  term = "Term",
  state = "State",
  term_order = NULL,
  start_term = NULL,
  target_state = NULL,
  missing_label = "Not Enrolled",
  unknown_label = "Unknown",
  graduated_other_label = "Graduated (Other)",
  graduated_in_target_label = NULL,
  title = NULL,
  width = NULL,
  height = 700,
  font_size = 11,
  title_font_size = 22,
  title_y = 0.975,
  top_margin = 65,
  term_label_y = 1.02,
  sankey_domain_top = 0.96,
  node_pad = 12,
  node_thickness = 18,
  save_html = FALSE,
  html_file = NULL,
  save_png = FALSE,
  png_file = NULL
)
```

Arguments

<code>data</code>	A data frame with one row per person per term.
<code>id</code>	Name of the identifier column. Default is "ID".
<code>term</code>	Name of the term/time-period column. Default is "Term".
<code>state</code>	Name of the categorical state column. Default is "State".
<code>term_order</code>	Optional vector giving the intended order of terms. If omitted, terms are ordered by first appearance in data.
<code>start_term</code>	Optional term defining the starting cohort. If omitted, the first value in <code>term_order</code> is used.
<code>target_state</code>	Optional state to highlight. When supplied, the starting cohort is restricted to people whose state equals <code>target_state</code> in <code>start_term</code> , and links leaving that state are highlighted in blue.
<code>missing_label</code>	Label assigned when a person is not observed in a later term. Default is "Not Enrolled".
<code>unknown_label</code>	Label assigned to missing/blank states. Default is "Unknown".

graduated_other_label	Label used for graduation outside the target state. Default is "Graduated (Other)".
graduated_in_target_label	Optional label used for graduation in the target state. If omitted and target_state is supplied, the label is constructed as "Graduated in {target_state}".
title	Plot title. If omitted, a title is constructed from target_state when available.
width	Plot width. Use NULL for responsive sizing, or a pixel value such as 1650 for wide HTML export. Default is NULL.
height	Plot height in pixels. Default is 700 for RStudio Viewer/browser usability.
font_size	Base font size. Default is 11.
title_font_size	Title font size. Default is 22.
title_y	Vertical position of the plot title in paper coordinates. Default is 0.975. Increase this to move the title up; decrease it to move the title down.
top_margin	Top margin in pixels. Default is 65. Increase this to add more white space above the title and plot area.
term_label_y	Vertical position of the term labels in paper coordinates. Default is 1.02. Increase this to move the term labels farther above the Sankey.
sankey_domain_top	Top of the Sankey drawing region in paper coordinates. Default is 0.96. Decrease this to create more space between the term labels and the colored Sankey regions.
node_pad	Node padding. Default is 12.
node_thickness	Node thickness. Default is 18.
save_html	Logical. If TRUE, the interactive Sankey is saved as an HTML file. Default is FALSE to avoid writing files unless explicitly requested.
html_file	File path for the HTML export. Default is NULL. Required when save_html = TRUE.
save_png	Logical. If TRUE, the function issues a warning that PNG export is not handled directly by retentionflow. Default is FALSE.
png_file	File path for PNG export. Default is NULL.

Value

A plotly htmlwidget.

Examples

```
if (requireNamespace("plotly", quietly = TRUE)) {
  x <- retention_example()
  plot_sankey(x, target_state = "Biology", save_html = FALSE)
}
```

retention_example	<i>Example retention flow data</i>
-------------------	------------------------------------

Description

A small synthetic dataset with one row per student per term.

Usage

```
retention_example()
```

Value

A data frame with ID, Term, and State.

Examples

```
retention_example()
```

stem_retention	<i>Sample STEM retention data</i>
----------------	-----------------------------------

Description

A synthetic longitudinal dataset with 100 students across eight semesters in a four-year STEM program and seven STEM majors: Biology, Chemistry, Computer Science, Engineering, Mathematics, Physics, and Statistics. Some students graduate in fewer than eight semesters, some graduate in the eighth semester, some stop out as Not Enrolled, and some remain enrolled at the end of the fourth year.

Usage

```
data(stem_retention)
```

Format

A data frame with 800 rows and 3 variables:

ID Synthetic student identifier.

Term Academic term.

State Student major or enrollment/completion state.

Examples

```
data(stem_retention)
head(stem_retention)
validate_flow_data(stem_retention)
```

summarize_transitions	<i>Summarize transition records</i>
-----------------------	-------------------------------------

Description

Counts students by transition pair and status.

Usage

```
summarize_transitions(transitions)
```

Arguments

transitions A transition data frame created by make_transitions().

Value

A data frame with transition counts and percentages within each from-term/from-state group.

Examples

```
x <- retention_example(); tr <- make_transitions(x); summarize_transitions(tr)
```

validate_flow_data	<i>Validate retention flow data</i>
--------------------	-------------------------------------

Description

Checks that a data frame has identifier, term, and state columns and no duplicate identifier-term rows.

Usage

```
validate_flow_data(  
  data,  
  id = "ID",  
  term = "Term",  
  state = "State",  
  require_complete = TRUE  
)
```

Arguments

<code>data</code>	A data frame with one row per person per term.
<code>id</code>	Name of the identifier column.
<code>term</code>	Name of the term/time-period column.
<code>state</code>	Name of the categorical state column.
<code>require_complete</code>	If TRUE, missing values in required columns are errors.

Value

Invisibly returns TRUE when validation passes.

Examples

```
x <- retention_example(); validate_flow_data(x)
```

Index

- * **datasets**
 - stem_retention, [5](#)
- make_transitions, [2](#)
- plot_sankey, [2](#)
- retention_example, [5](#)
- stem_retention, [5](#)
- summarize_transitions, [6](#)
- validate_flow_data, [6](#)