Package 'panelbuild'

July 3, 2026

Description Provides tools for auditing, validating, and preparing panel datasets before statistical analysis. Functions identify duplicate unit-time observations, missing unit-time cells, panel gaps, and balance issues. The package also provides audit summaries, row-level diagnostic flags, panel-completion utilities, and a concise audit report.

Title Panel Data Auditing, Validation, and Preparation Tools

Version 0.1.0

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Imports dplyr, tidyr, tibble, rlang

Suggests testthat (>= 3.0.0), knitr, rmarkdown

Config/testthat/edition 3

VignetteBuilder knitr

URL https://desirajulavanya.github.io/panelbuild/, https://github.com/desirajulavanya/panelbuild

BugReports https://github.com/desirajulavanya/panelbuild/issues

Depends R (>= 4.1.0)

LazyData true

NeedsCompilation no

Author Lavanya Desiraju [aut, cre]

Maintainer Lavanya Desiraju <desirajulavanya@gmail.com>

Repository CRAN

Date/Publication 2026-07-03 13:00:02 UTC

Contents

`audit_panel` *Audit a panel dataset*

Description

`audit_panel()` checks whether a dataset has the expected structure of a panel dataset. It reports the number of panel units, time periods, observed rows, unique unit-time cells, expected unit-time cells, missing unit-time cells, duplicate unit-time cells, and whether the panel is balanced.

Usage

```
audit_panel(data, id, time)
```

Arguments

`data`	A data frame or tibble.
`id`	Unquoted column name identifying the panel unit, such as a person, firm, district, county, or country.
`time`	Unquoted column name identifying the time period, such as a year, month, quarter, or date.

Details

A panel is treated as balanced when every observed panel unit appears in every observed time period exactly once. Missing cells are unit-time combinations that are implied by the full unit-by-time grid but are not present in the data. Duplicate cells are unit-time combinations that appear more than once.

`audit_panel()` does not modify the input data. It returns an audit object that can be summarized with `audit_summary()` and inspected with accessor functions such as `missing_cells()` and `duplicate_cells()`.

Value

An object of class `panelbuild_panel_audit`. The object is a list containing panel metadata, balance information, counts of missing and duplicate unit-time cells, and data frames containing the missing and duplicate cells.

See Also

audit_summary(), missing_cells(), duplicate_cells(), duplicate_summary(), gap_summary(), complete_panel()

Examples

```
audit_panel(example_panel, id = id, time = year)
```

---

audit_report *Create a panel audit report*

Description

audit_report() prints a concise, human-readable report from an audit object created by audit_panel().

Usage

```
audit_report(x)
```

Arguments

x	An object created by audit_panel().

Details

The report summarizes the panel structure, balance status, missing unit-time cells, duplicate unit-time cells, and recommended next steps. This is a lightweight console report and does not create files or modify the data.

Value

Invisibly returns x, the input audit object.

See Also

audit_panel(), audit_summary(), missing_cells(), duplicate_cells()

Examples

```
audit <- audit_panel(example_panel, id = id, time = year)
audit_report(audit)
```

audit_summary *Summarize a panel audit*

Description

audit_summary() converts an audit object created by audit_panel() into a one-row tibble of panel diagnostics.

Usage

```
audit_summary(x)
```

Arguments

x	An object created by audit_panel().

Details

This function is useful when users want a compact, tabular summary of a panel audit. The resulting tibble can be printed, saved, joined with other metadata, or combined across multiple datasets.

The summary includes the number of units, number of time periods, observed rows, observed unit-time cells, expected unit-time cells, missing cells, duplicate cells, and a logical indicator for whether the panel is balanced.

Value

A one-row tibble with the following columns:

data Name of the audited object.

id Name of the panel unit column.

time Name of the time column.

n_units Number of unique panel units.

n_periods Number of unique time periods.

observed_rows Number of rows in the original data.

observed_id_time_cells Number of unique observed unit-time cells.

expected_cells Number of cells in the full unit-by-time grid.

missing_cells Number of missing unit-time cells.

duplicate_cells Number of duplicate unit-time cells.

balanced Logical indicator for whether the panel is balanced.

See Also

audit_panel(), missing_cells(), duplicate_cells()

Examples

```
audit <- audit_panel(example_panel, id = id, time = year)
audit_summary(audit)
```

`complete_panel` *Complete a panel dataset with an audit trail*

Description

`complete_panel()` expands a panel dataset so that every observed panel unit appears in every observed time period. Newly created unit-time cells are flagged with audit columns, and substantive variables are left missing.

Usage

```
complete_panel(data, id, time)
```

Arguments

`data`	A data frame or tibble.
`id`	Unquoted column name identifying the panel unit, such as a person, firm, district, county, or country.
`time`	Unquoted column name identifying the time period, such as a year, month, quarter, or date.

Details

The function first audits the panel using `audit_panel()`. If duplicate unit-time cells are present, the function stops with an error. This is intentional: completing a panel with duplicate unit-time observations can produce ambiguous results.

`complete_panel()` does not impute outcomes, covariates, treatment variables, or any other substantive variables. It only creates the missing unit-time rows implied by the full unit-by-time grid. Newly created rows are flagged using audit columns.

Value

A tibble containing the completed panel grid. The returned data include the original columns plus the following audit columns:

`panelbuild_original_row` Logical indicator for rows present in the original data.

`panelbuild_completed_cell` Logical indicator for rows created by `complete_panel()`.

`panelbuild_audit_action` Character label describing whether a row was original or added during panel completion.

The returned tibble also includes attributes documenting the panel identifier, time variable, number of completed cells, and audit note.

See Also

audit_panel(), missing_cells(), gap_summary(), duplicate_summary()

Examples

```
panel_unique <- example_panel |>
  dplyr::distinct(id, year, .keep_all = TRUE)

complete_panel(panel_unique, id = id, time = year)
```

---

duplicate_cells *Extract duplicate unit-time cells from a panel audit*

Description

duplicate_cells() extracts duplicate unit-time combinations stored in an audit object created by audit_panel().

Usage

```
duplicate_cells(x)
```

Arguments

x	An object created by audit_panel().

Details

Duplicate cells are unit-time combinations that appear more than once in the original data. The returned table includes a count column n showing how many rows are present for each duplicated unit-time cell.

This function does not re-audit the original dataset. It simply extracts the duplicate-cell table already stored in the audit object.

Value

A tibble containing duplicate unit-time combinations and a count column n.

See Also

audit_panel(), panel_duplicates(), duplicate_summary(), flag_panel_issues()

Examples

```
audit <- audit_panel(example_panel, id = id, time = year)
duplicate_cells(audit)
```

duplicate_summary *Summarize duplicate unit-time cells by panel unit*

Description

`duplicate_summary()` reports how many duplicate unit-time cells each panel unit has.

Usage

```
duplicate_summary(data, id, time)
```

Arguments

`data`	A data frame or tibble.
`id`	Unquoted column name identifying the panel unit.
`time`	Unquoted column name identifying the time period.

Details

This function summarizes duplicate cells at the panel-unit level. It is useful when users want to identify which units contribute most to duplicate unit-time observations.

The output reports both the number of duplicated cells and the number of extra rows implied by those duplicates. For example, if one unit-time cell appears three times, it counts as one duplicate cell and two extra rows.

The function does not modify, drop, aggregate, or impute the data.

Value

A tibble with one row per panel unit that has duplicate cells. The output includes:

`panelbuild_duplicate_cells` Number of duplicated unit-time cells for the unit.

`panelbuild_duplicate_extra_rows` Number of extra rows caused by duplicates.

If no duplicates are present, the function returns all units with zero duplicate cells and zero extra rows.

See Also

`audit_panel()`, `panel_duplicates()`, `duplicate_cells()`, `flag_panel_issues()`

Examples

```
duplicate_summary(example_panel, id = id, time = year)
```

example_panel *Example Panel Dataset*

Description

A small example panel dataset for demonstrating panel-data auditing.

Usage

```
example_panel
```

Format

A data frame with 9 rows and 4 variables:

id Panel unit identifier.

year Time period.

outcome Example outcome variable.

treatment Example treatment indicator.

Details

The dataset intentionally includes one duplicate unit-time observation and missing unit-time cells so that users can test `panelbuild` diagnostics.

Examples

```
data(example_panel)
audit_panel(example_panel, id = id, time = year)
```

flag_panel_issues *Flag row-level panel data issues*

Description

`flag_panel_issues()` adds row-level audit flags to a panel dataset. It identifies duplicate unit-time observations while preserving the original data structure.

Usage

```
flag_panel_issues(data, id, time)
```

Arguments

data	A data frame or tibble.
id	Unquoted column name identifying the panel unit, such as a person, firm, district, county, or country.
time	Unquoted column name identifying the time period, such as a year, month, quarter, or date.

Details

This function is useful when users want to inspect problematic rows directly rather than only receiving a summary table. It adds diagnostic columns that indicate how many times each unit-time cell appears and whether the row is part of a duplicate cell.

`flag_panel_issues()` does not add rows, remove rows, complete the panel, or impute missing values.

Value

A tibble containing the original data plus row-level audit columns:

`panelbuild_row_id` Integer row identifier based on the original row order.

`panelbuild_id_time_n` Number of rows with the same unit-time combination.

`panelbuild_duplicate_cell` Logical indicator for rows that belong to a duplicate unit-time cell.

The returned tibble also includes attributes documenting the panel identifier, time variable, and audit note.

See Also

`audit_panel()`, `duplicate_summary()`, `duplicate_cells()`, `complete_panel()`

Examples

```
flag_panel_issues(example_panel, id = id, time = year)
```

---

gap_summary *Summarize missing panel periods by unit*

Description

`gap_summary()` reports how many time periods are missing for each panel unit.

Usage

```
gap_summary(data, id, time)
```

Arguments

data	A data frame or tibble.
id	Unquoted column name identifying the panel unit.
time	Unquoted column name identifying the time period.

Details

This function summarizes the missing unit-time cells returned by panel_gaps() at the panel-unit level. It is useful for identifying which units contribute most to panel imbalance.

The function does not modify, complete, or impute the input data.

Value

A tibble with one row per panel unit and a column panelbuild_missing_periods giving the number of missing time periods for that unit. If no gaps are present, all units are returned with zero missing periods.

See Also

audit_panel(), panel_gaps(), missing_cells(), complete_panel()

Examples

```
gap_summary(example_panel, id = id, time = year)
```

---

missing_cells *Extract missing unit-time cells from a panel audit*

Description

missing_cells() extracts the missing unit-time combinations stored in an audit object created by audit_panel().

Usage

```
missing_cells(x)
```

Arguments

x	An object created by audit_panel().

Details

Missing cells are unit-time combinations that are implied by the full unit-by-time grid but are not present in the original data.

This function does not re-audit the original dataset. It simply extracts the missing-cell table already stored in the audit object.

Value

A tibble containing missing unit-time combinations.

See Also

audit_panel(), panel_gaps(), gap_summary(), complete_panel()

Examples

```
audit <- audit_panel(example_panel, id = id, time = year)
missing_cells(audit)
```

---

panel_duplicates *Identify duplicate unit-time cells*

Description

panel_duplicates() returns unit-time combinations that appear more than once in a panel dataset.

Usage

```
panel_duplicates(data, id, time)
```

Arguments

data	A data frame or tibble.
id	Unquoted column name identifying the panel unit.
time	Unquoted column name identifying the time period.

Details

Duplicate unit-time cells occur when the same panel unit appears more than once in the same time period. These duplicates can create problems for panel completion, fixed effects models, difference-in-differences designs, and other longitudinal-data workflows.

The function does not modify, drop, aggregate, or impute the data.

Value

A tibble containing duplicate unit-time combinations and a count column n.

See Also

audit_panel(), duplicate_summary(), duplicate_cells(), flag_panel_issues()

Examples

```
panel_duplicates(example_panel, id = id, time = year)
```

panel_gaps *Identify missing unit-time cells*

Description

`panel_gaps()` returns the missing unit-time combinations implied by the full panel grid.

Usage

```
panel_gaps(data, id, time)
```

Arguments

`data`	A data frame or tibble.
`id`	Unquoted column name identifying the panel unit.
`time`	Unquoted column name identifying the time period.

Details

A missing unit-time cell is a combination of an observed panel unit and an observed time period that does not appear in the data. For example, if unit `A` appears in 2020 and 2022, and 2021 is an observed time period elsewhere in the dataset, then `A`-2021 is treated as a missing unit-time cell.

This function is a data-frame interface to the missing-cell information produced by `audit_panel()`. It does not modify, complete, or impute the input data.

Value

A tibble containing missing unit-time combinations.

See Also

`audit_panel()`, `missing_cells()`, `gap_summary()`, `complete_panel()`

Examples

```
panel_gaps(example_panel, id = id, time = year)
```

Index