Package 'cohortBuilder'

July 3, 2026

Type Package

Title Data Source Agnostic Filtering Tools

Version 1.0.0

Maintainer Krystian Igras <krystian8207@gmail.com>

Description Common API for filtering data stored in different data models.
Provides multiple filter types and reproducible R code.
Works standalone or with 'shinyCohortBuilder' as the GUI for interactive Shiny apps.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

LazyData true

Imports R6, S7 (>= 0.2.0), jsonlite, purrr, tibble, dplyr (>= 1.0.0),
tidyr, glue, ggplot2, rlang (>= 1.0), formatR, collapse

KeepSource true

Suggests queryBuilder, vdiffr, testthat (>= 3.0.0), knitr, rmarkdown,
ellmer (>= 0.3.0), withr

Config/testthat/edition 3

Collate 'cohortBuilder-package.R' 'cohort_methods.R'
'source_methods.R' 'step.R' 'filter.R' 'ai_tools.R'
'attrition.R' 'bind_keys.R' 'hooks.R' 'list_operators.R'
'repro_code_utils.R' 'source_tblist.R' 'data.R'

VignetteBuilder knitr

URL https://github.com/r-world-devs/cohortBuilder/,
https://r-world-devs.github.io/cohortBuilder/

Depends R (>= 4.1.0)

Config/roxygen2/version 8.0.0

RoxygenNote 7.3.3

NeedsCompilation no

Author Krystian Igras [cre, aut],
Adam Foryś [ctb]

Repository CRAN

Date/Publication 2026-07-03 08:50:26 UTC

Contents

cohortBuilder-package . . . 4
.gen_id . . . 4
.get_item . . . 4
.get_method . . . 5
.if_value . . . 5
.print_filter . . . 6
.propagate_domains . . . 6
add_filter . . . 7
add_source . . . 8
add_step . . . 9
attrition . . . 10
autofilter . . . 10
binding-keys . . . 11
CbFilter . . . 15
CbFilterDateRange . . . 16
CbFilterDatetimeRange . . . 17
CbFilterDiscrete . . . 18
CbFilterDiscreteText . . . 19
CbFilterMultiDiscrete . . . 20
CbFilterQuery . . . 21
CbFilterRange . . . 22
cb_domain_from_data . . . 23
cb_domain_from_stats . . . 23
cb_filter_data . . . 24
cb_filter_to_expr . . . 24
cb_get_filter_data . . . 25
cb_get_filter_defaults . . . 25
cb_get_filter_stats . . . 26
cb_intersect_domain . . . 26
cb_intersect_domain_values . . . 27
cb_plot_filter_data . . . 27
cb_register_tool . . . 28
cb_tool . . . 28
cb_tool_add_filters . . . 29
cb_tool_apply_filters . . . 30
cb_tool_clear_filters . . . 30
cb_tool_describe_state . . . 31
cb_tool_filters_meta . . . 31
cb_tool_get_code . . . 32
cb_tool_get_data_summary . . . 32
cb_tool_remove_filters . . . 33
cb_tool_remove_step . . . 33

cb_tool_run . . . 34
cb_tool_set_filter_values . . . 34
cb_tool_toggle_filters . . . 35
code . . . 35
Cohort . . . 36
cohort-methods . . . 48
create-cohort . . . 48
data_key . . . 49
describe . . . 50
description . . . 50
filter . . . 51
filter_domain . . . 52
filter_effective_value . . . 52
filter_variables . . . 53
get_data . . . 53
get_filter_params . . . 54
get_state . . . 54
hooks . . . 55
librarian . . . 56
managing-cohort . . . 57
managing-source . . . 58
plot_data . . . 58
primary_keys . . . 59
register_filter_type . . . 60
restore . . . 61
rm_filter . . . 61
rm_step . . . 62
run . . . 63
set_source . . . 64
shape . . . 65
Source . . . 66
source-layer . . . 69
stat . . . 72
step . . . 73
sum_up . . . 74
tblist . . . 74
tblist_class . . . 75
update_filter . . . 76
update_source . . . 76
%->% . . . 77

Index **78**

cohortBuilder-package *Create data source cohort*

Description

Create data source cohort

.gen_id *Generate random ID*

Description

Generate random ID

Usage

```
.gen_id()
```

Value

A character type value.

.get_item *Return list of objects matching provided condition.*

Description

Return list of objects matching provided condition.

Usage

```
.get_item(list_obj, attribute, value, operator = `==`)
```

Arguments

list_obj	List of R objects.
attribute	Object attribute name.
value	Object value.
operator	Logical operator - two-argument function taking ‘list_obj‘ attribute value as the first one, and ‘value‘ as the second one.

Value

A subset of list object matching provided condition.

Examples

```
my_list <- list(
  list(id = 1, name = "a"),
  list(id = 2, name = "b")
)
.get_item(my_list, "id", 1)
.get_item(my_list, "name", c("b", "c"), identical)
```

.get_method *Get function definition*

Description

Whenever the function with provided name exists anywhere, the one is returned (or the first one if multiple found). Return NULL otherwise.

Usage

```
.get_method(name)
```

Arguments

name	Name of the function.

Value

Function - when found in any namespace or NULL otherwise.

.if_value *Return default value if values are equal*

Description

Return default value if values are equal

Usage

```
.if_value(x, value, default)
```

Arguments

x	Condition to be compared with value.
value	Value to be compared with x.
default	Default value to be returned when ‘x‘ is identical to ‘value‘.

Value

Evaluated condition or provided default value.

`.print_filter` *Method for printing filter details*

Description

Method for printing filter details

Usage

```
.print_filter(filter, data_objects, to_string = FALSE)

## Default S3 method:
.print_filter(filter, data_objects, to_string = FALSE)
```

Arguments

`filter`	The defined filter object.
`data_objects`	List of data objects for the underlying filtering step.
`to_string`	If 'TRUE', return the output as a character vector instead of printing it. Defaults to 'FALSE'.

`.propagate_domains` *Propagate domains between steps*

Description

Source-layer generic that recomputes the domains of a **target step** from that step's parent (the previous step). Override to implement dynamic domain narrowing.

Usage

```
.propagate_domains(source, data_object, step_id, cohort, ...)

## Default S3 method:
.propagate_domains(source, data_object, step_id, cohort, ...)

## S3 method for class 'tblist'
.propagate_domains(source, data_object, step_id, cohort, mode, ...)
```

Arguments

source	Source object.
data_object	Filtered data of the **parent** step (the input to the target step), used by ‘mode = "data"‘.
step_id	Target step id whose domains should be recomputed.
cohort	Cohort object (for accessing the parent step’s filters and stored statistics).
...	Additional arguments.
mode	Propagation mode: ‘"filter"‘ (from upstream filter values, no data access), ‘"stats"‘ (from the parent’s post-step stored stats), or ‘"data"‘ (scan the parent’s filtered data directly).

Details

The default is a no-op. Only called when ‘propagate_domains != "none"‘ in the Cohort.

Contract: ‘step_id = N‘ recomputes the domains of step ‘N‘’s filters using step ‘N - 1‘ as the source of truth. It is a no-op when ‘N‘ has no parent (i.e. ‘N == "1"‘) or when the target step is absent.

---

add_filter *Add filter definition*

Description

Add filter definition

Usage

```
add_filter(x, filter, step_id, ...)

## S3 method for class 'Cohort'
add_filter(
  x,
  filter,
  step_id,
  run_flow = FALSE,
  hook = list(pre = get_hook("pre_add_filter_hook"), post =
    get_hook("post_add_filter_hook")),
  ...
)

## S3 method for class 'Source'
add_filter(x, filter, step_id, ...)
```

Arguments

x	An object to add filter to.
filter	Filter definition created with filter.
step_id	Id of the step to add the filter to. If missing, filter is added to the last step.
...	Other parameters passed to specific S3 method.
run_flow	If 'TRUE', data flow is run after the filter is added.
hook	List of hooks describing methods to run before/after the filter is added. See hooks for more details.

Value

Method dependent object (i.e. 'Cohort' or 'Source') having filter added in selected step.

See Also

managing-cohort, managing-source

---

add_source *Add source to Cohort object.*

Description

When Cohort object has been created without source, the method allows to attach it.

Usage

```
add_source(x, source)
```

Arguments

x	Cohort object.
source	Source object to be attached.

Value

The 'Cohort' class object with 'Source' attached to it.

See Also

managing-cohort

add_step *Add filtering step definition*

Description

Add filtering step definition

Usage

```
add_step(x, step, ...)

## S3 method for class 'Cohort'
add_step(
  x,
  step,
  run_flow = FALSE,
 hook = list(pre = get_hook("pre_add_step_hook"), post = get_hook("post_add_step_hook")),
  ...
)

## S3 method for class 'Source'
add_step(x, step, ...)
```

Arguments

x	An object to add step to.
step	Step definition created with step.
...	Other parameters passed to specific S3 method.
run_flow	If 'TRUE', data flow is run after the step is added.
hook	List of hooks describing methods to run before/after the step is added. See hooks for more details.

Value

Method dependent object (i.e. 'Cohort' or 'Source') having new step added.

See Also

managing-cohort, managing-source

attrition *Show attrition plot.*

Description

Show attrition plot.

Usage

```
attrition(x, ..., percent = FALSE)
```

Arguments

x	Cohort object.
...	Source specific parameters required to generate attrition.
percent	Should attrition changes be presented with percentage values.

Value

Plot object of class 'ggplot'.

See Also

cohort-methods

autofilter *Generate filters definition based on the Source data*

Description

The method should analyze source data structure, generate proper filters based on the data (e.g. column types) and attach them to source.

Usage

```
autofilter(source, attach_as = c("step", "meta"), ...)

## Default S3 method:
autofilter(source, ...)

## S3 method for class 'tblist'
autofilter(source, attach_as = c("step", "meta"), ...)
```

Arguments

source	Source object.
attach_as	Choose whether the filters should be attached as a new step, or list of available filters (used in filtering panel when ‘new_step = "configure"‘). By default in step.
...	Extra arguments passed to a specific method.

Value

Source object having step configuration attached.

Examples

```
library(cohortBuilder)

iris_source <- set_source(tblist(iris = iris)) |>
  autofilter()
iris_cohort <- cohort(iris_source)
sum_up(iris_cohort)
```

---

binding-keys *Describe data relations with binding keys*

Description

When source consists of multiple datasets, binding keys allow to define what relations occur between them. When binding keys are defined, applying filtering on one dataset may result with updating (filtering) the other ones.

For example having two tables in Source: ‘book(book_id, author_id, title)‘ ‘authors(author_id, name, surname)‘ if we filter ‘authors‘ table, we way want to return only books for the selected authors.

With binding keys you could achieve it by providing ‘binding_keys‘ parameter for Source as below:

```
binding_keys = bind_keys(
  bind_key(
    update = data_key('books', 'author_id'),
    data_key('authors', 'author_id')
  )
)
```

Or if we want to have two-way relation, just define another binding key:

```
binding_keys = bind_keys(
  bind_key(
    update = data_key('books', 'author_id'),
    data_key('authors', 'author_id')
```

```
  ),
  bind_key(
    update = data_key('authors', 'author_id'),
    data_key('books', 'author_id')
  )
)
```

As a result, whenever ‘books‘ or ‘authors‘ is filtered, the other table will be updated as well.

In order to understand binding keys concept we need to describe the following functions:

- data_key - Defines which table column should be used to describe relation.
- `bind_key` - Defines what relation occur between datasets.
- `bind_keys` - If needed, allows to define more than one relation.

- ‘data_key‘ - requires to provide two parameters:

- `dataset` - Name of the dataset existing in Source.
- `key` - Single character string or vector storing column names that are keys, which should be used to describe relation.

For example ‘data_key(’books’, ’author_id’)‘.

- ‘bind_key‘ - requires to provide two obligatory parameters

- `update` - Data key describing which table should be updated.
- `...` - **Triggering data keys**. One or more data keys describing on which dataset(s) the one in ‘update‘ is dependent.

The output of ‘bind_key‘ function is named **binding key**. ‘bind_key‘ offers two extra parameters ‘post‘ and ‘activate‘. See below to learn how these parameters affect the final result.

- ‘bind_keys‘ - takes only binding keys as parameters The function is used to define ‘binding_keys‘ parameter of Source. Whenever you define a single or more binding keys wrap them with ‘bind_keys‘.

It’s worth to mention that binding key describes inner-join like relation. That means the updated table’s key is intersection of its key and keys of remaining tables defined in binding key.

Another important note is that binding keys order matters - binding is performed sequentially, taking into account returned data from the previous bindings.

You may achieve more flexibility with two parameters:

- `activate`
- `post`

Active tables and ‘activate‘ parameter

We name a table ‘active‘ that is attached to at least one active filter (in a step).

When having defined binding key, e.g.

```
bind_key(
  update = data_key('books', 'author_id'),
  data_key('authors', 'author_id')
)
```

the key is taken into account only when at least one triggering table is active. So in the above example binding key will update ‘books‘ only when ‘authors‘ was filtered (more precisely when any filter attached to ‘authors‘ is active).

The ‘activate = TRUE‘ parameter setup, lets us to decide whether ‘update‘ table should be marked as active as well when the binding finish. This allows to build dependency chains between table.

Let’s explain this in the below example. Having defined another table in Source ‘borrowed(book_id, user_id, date)‘ and binding key:

```
bind_keys(
  bind_key(
    update = data_key('books', 'book_id'),
    data_key('borrowed', 'book_id')
  ),
  bind_key(
    update = data_key('authors', 'author_id'),
    data_key('books', 'author_id')
  )
)
```

Let’s consider the case when table ‘borrowed‘ is active, ‘books‘ is not. What happens during the binding process: 1. Based on the first binding key, active ‘borrowed‘ triggers this one. 2. As a result ‘books‘ is modified.

What should happen with the second binding key. We have two options: 1. ‘books‘ could be marked as active as well so it triggers the second key. 2. ‘books‘ could remain inactive so the second key is not triggered. It will be triggered only when ‘books‘ is directly filtered (activated).

You may choose between 1 and 2 with ‘activate = TRUE‘ (the default) and ‘activate = FALSE‘ respectively.

So in the above example (because ‘activate = TRUE‘ by default) the authors table will also be modified by the second binding key.

To turn off this behavior we just need to:

```
bind_keys(
  bind_key(
    update = data_key('books', 'book_id'),
    data_key('borrowed', 'book_id'),
    activate = TRUE
  ),
  bind_key(
    update = data_key('authors', 'author_id'),
    data_key('books', 'author_id')
  )
)
```

Bind filtered on unfiltered data - ‘post‘ parameter

Let’s tart with the below binding key example:

```
bind_keys(
  bind_key(
    update = data_key('authors', 'author_id'),
    data_key('books', 'author_id')
  )
)
```

Let's assume 'authors' table is filtered and we apply filtering for 'books' table. We may want to achieve one of the two results: 1. 'authors' filters should be taken into account while binding. 2. we should take unfiltered 'authors' an apply binding based on 'books' choices.

We can achieve 1 and 2 with defining 'post = TRUE' (the default) and 'post = FALSE' respectively.

So the following setup:

```
bind_keys(
  bind_key(
    update = data_key('authors', 'author_id'),
    data_key('books', 'author_id'),
    post = FALSE
  )
)
```

Whenever 'books' is changed will result with filtering only the authors that written selected books - no extra 'authors' filters will be applied.

There might be the situation when table was already bound but there is another one binding key to be executed on the same table.

In this case 'post = FALSE' case will remain the same - unfiltered table will be taken. More to that filtering and previous binding related to this table will be ignored. In case of 'post = TRUE' the previously bound table will be updated.

Usage

```
bind_keys(...)

bind_key(update, ..., post = TRUE, activate = TRUE)
```

Arguments

`...`	In case of 'bind_keys', binding keys created with 'bind_key'. In case of 'bind_key', data keys describing triggering tables.
`update`	Data key describing table to update.
`post`	Update filtered or unfiltered table.
`activate`	Mark bound table as active.

Value

List of class 'bind_keys' storing 'bind_key' class objects ('bind_keys') or 'bind_key' class list ('bind_key').

`CbFilter` *Base class for all cohortBuilder filters*

Description

Base class for all cohortBuilder filters

Usage

```
CbFilter(
  type = character(0),
  id = character(0),
  name = character(0),
  active = logical(0),
  description = NULL,
  domain = NULL,
  step_id = NULL,
  extra = list(),
  private = list()
)
```

Arguments

`type`	Filter type string.
`id`	Filter identifier.
`name`	Filter display name.
`active`	Whether the filter is active.
`description`	Optional filter description.
`domain`	Optional domain constraining valid filter values. Structure depends on filter type: character vector for discrete, 2-length vector for range, named list for multi_discrete. When set, filter values are intersected with the domain. When value is unset (‘NA‘) and domain is provided, the domain serves as the effective value.
`step_id`	Step identifier (set when filter is attached to a step).
`extra`	Named list of extra parameters.
`private`	Named list of internal parameters, not intended to be set directly by users.

`CbFilterDateRange` *Date range filter class*

Description

Filters data by a date range.

Usage

```
CbFilterDateRange(
  id = NULL,
  name = NULL,
  variable,
  range = NA,
  dataset,
  keep_na = TRUE,
  description = NULL,
  domain = NULL,
  active = getOption("cb_active_filter", default = TRUE),
  ...
)
```

Arguments

`id`	Filter identifier.
`name`	Filter display name (defaults to ‘id‘).
`variable`	Column name to filter on.
`range`	Numeric vector of length 2 (min, max). ‘NA‘ means no filtering.
`dataset`	Dataset name.
`keep_na`	If ‘TRUE‘, NA values are retained.
`description`	Optional description.
`domain`	Optional filter domain (the set of allowed values). ‘NULL‘ means the domain is derived from the data.
`active`	If ‘FALSE‘, filter is skipped.
`...`	Extra parameters.

`CbFilterDatetimeRange` *Datetime range filter class*

Description

Filters data by a datetime (POSIXct) range.

Usage

```
CbFilterDatetimeRange(
  id = NULL,
  name = NULL,
  variable,
  range = NA,
  dataset,
  keep_na = TRUE,
  description = NULL,
  domain = NULL,
  active = getOption("cb_active_filter", default = TRUE),
  ...
)
```

Arguments

`id`	Filter identifier.
`name`	Filter display name (defaults to ‘id‘).
`variable`	Column name to filter on.
`range`	Numeric vector of length 2 (min, max). ‘NA‘ means no filtering.
`dataset`	Dataset name.
`keep_na`	If ‘TRUE‘, NA values are retained.
`description`	Optional description.
`domain`	Optional filter domain (the set of allowed values). ‘NULL‘ means the domain is derived from the data.
`active`	If ‘FALSE‘, filter is skipped.
`...`	Extra parameters.

`CbFilterDiscrete` *Discrete filter class*

Description

Filters data by matching a variable against a set of discrete values.

Usage

```
CbFilterDiscrete(
  id = NULL,
  name = NULL,
  variable,
  value = NA,
  dataset,
  keep_na = TRUE,
  description = NULL,
  domain = NULL,
  active = getOption("cb_active_filter", default = TRUE),
  ...
)
```

Arguments

`id`	Filter identifier.
`name`	Filter display name (defaults to ‘id‘).
`variable`	Column name to filter on.
`value`	Values to keep. ‘NA‘ means no filtering.
`dataset`	Dataset name.
`keep_na`	If ‘TRUE‘, NA values are retained.
`description`	Optional description.
`domain`	Optional filter domain (the set of allowed values). ‘NULL‘ means the domain is derived from the data.
`active`	If ‘FALSE‘, filter is skipped.
`...`	Extra parameters.

`CbFilterDiscreteText` *Discrete text filter class*

Description

Filters data by matching a variable against comma-separated text values.

Usage

```
CbFilterDiscreteText(
  id = NULL,
  name = NULL,
  variable,
  value = NA,
  dataset,
  keep_na = TRUE,
  description = NULL,
  domain = NULL,
  active = getOption("cb_active_filter", default = TRUE),
  ...
)
```

Arguments

`id`	Filter identifier.
`name`	Filter display name (defaults to ‘id‘).
`variable`	Column name to filter on.
`value`	Values to keep. ‘NA‘ means no filtering.
`dataset`	Dataset name.
`keep_na`	If ‘TRUE‘, NA values are retained.
`description`	Optional description.
`domain`	Optional filter domain (the set of allowed values). ‘NULL‘ means the domain is derived from the data.
`active`	If ‘FALSE‘, filter is skipped.
`...`	Extra parameters.

`CbFilterMultiDiscrete` *Multi-discrete filter class*

Description

Filters data by matching multiple variables against sets of discrete values.

Usage

```
CbFilterMultiDiscrete(
  id = NULL,
  name = NULL,
  values,
  variables,
  dataset,
  keep_na = TRUE,
  description = NULL,
  domain = NULL,
  active = getOption("cb_active_filter", default = TRUE),
  ...
)
```

Arguments

`id`	Filter identifier.
`name`	Filter display name (defaults to 'id').
`values`	Named list of values to filter by, keyed by variable name.
`variables`	Vector of column names to filter on.
`dataset`	Dataset name.
`keep_na`	If 'TRUE', NA values are retained.
`description`	Optional description.
`domain`	Optional filter domain (the set of allowed values). 'NULL' means the domain is derived from the data.
`active`	If 'FALSE', filter is skipped.
`...`	Extra parameters.

CbFilterQuery *Query filter class*

Description

Filters data using a queryBuilder query object.

Usage

```
CbFilterQuery(
  id = NULL,
  name = NULL,
  variables,
  value = NA,
  dataset,
  keep_na = TRUE,
  description = NULL,
  domain = NULL,
  active = getOption("cb_active_filter", default = TRUE),
  ...
)
```

Arguments

id	Filter identifier.
name	Filter display name (defaults to 'id').
variables	Vector of column names used in the query.
value	Query object (from queryBuilder package). 'NA' means no filtering.
dataset	Dataset name.
keep_na	If 'TRUE', NA values are retained.
description	Optional description.
domain	Optional filter domain (the set of allowed values). 'NULL' means the domain is derived from the data.
active	If 'FALSE', filter is skipped.
...	Extra parameters.

`CbFilterRange` *Range filter class*

Description

Filters data by a numeric range.

Usage

```
CbFilterRange(
  id = NULL,
  name = NULL,
  variable,
  range = NA,
  dataset,
  keep_na = TRUE,
  description = NULL,
  domain = NULL,
  active = getOption("cb_active_filter", default = TRUE),
  ...
)
```

Arguments

`id`	Filter identifier.
`name`	Filter display name (defaults to ‘id‘).
`variable`	Column name to filter on.
`range`	Numeric vector of length 2 (min, max). ‘NA‘ means no filtering.
`dataset`	Dataset name.
`keep_na`	If ‘TRUE‘, NA values are retained.
`description`	Optional description.
`domain`	Optional filter domain (the set of allowed values). ‘NULL‘ means the domain is derived from the data.
`active`	If ‘FALSE‘, filter is skipped.
`...`	Extra parameters.

cb_domain_from_data *Extract domain from data*

Description

Derives a domain (set of valid values) directly from the data object. Uses dual dispatch on filter type and source type. Custom filter types should implement S7 methods for this generic. The default method returns 'NULL' (no domain).

Usage

```
cb_domain_from_data(filter, source, ...)
```

Arguments

filter	S7 filter object.
source	Source object.
...	Additional arguments. Methods receive a 'data_object' (the data to extract the domain from) here.

Value

Domain value appropriate for the filter type, or 'NULL'.

cb_domain_from_stats *Extract domain from stored filter statistics*

Description

Derives a domain (set of valid values) from previously computed filter statistics. Custom filter types should implement an S7 method for this generic. The default method returns 'NULL' (no domain).

Usage

```
cb_domain_from_stats(filter, ...)
```

Arguments

filter	S7 filter object.
...	Additional arguments. Methods receive 'stats', the list of stored statistics for the filter.

Value

Domain value appropriate for the filter type, or 'NULL'.

cb_filter_data *Apply filter to data object*

Description

Apply filter to data object

Usage

```
cb_filter_data(filter, source, ...)
```

Arguments

filter	S7 filter object.
source	Source object.
...	Additional arguments. Methods receive a ‘data_object‘ (the data to filter) here.

Value

Filtered data object.

cb_filter_to_expr *Generate reproducible code expression for filter*

Description

Generate reproducible code expression for filter

Usage

```
cb_filter_to_expr(filter, source, ...)
```

Arguments

filter	S7 filter object.
source	Source object.
...	Additional arguments.

Value

An R expression representing the filter operation.

cb_get_filter_data *Get filter-related data*

Description

Get filter-related data

Usage

```
cb_get_filter_data(filter, source, ...)
```

Arguments

filter	S7 filter object.
source	Source object.
...	Additional arguments. Methods receive a ‘data_object‘ here.

Value

Filter-related data subset.

cb_get_filter_defaults *Get filter default values*

Description

Get filter default values

Usage

```
cb_get_filter_defaults(filter, source, ...)
```

Arguments

filter	S7 filter object.
source	Source object.
...	Additional arguments. Methods receive a ‘data_object‘ and a ‘cache_object‘ (cached statistics) here.

Value

Named list of default parameter values.

cb_get_filter_stats *Get filter statistics*

Description

Get filter statistics

Usage

```
cb_get_filter_stats(filter, source, ...)
```

Arguments

filter	S7 filter object.
source	Source object.
...	Additional arguments. Methods receive a 'data_object' (the data to compute statistics from) here.

Value

List of statistics.

cb_intersect_domain *Get effective filter value after domain intersection*

Description

Returns the filter's value intersected with its domain. When value is 'NA' and domain is set, returns the domain as the effective value.

Usage

```
cb_intersect_domain(filter, ...)
```

Arguments

filter	S7 filter object.
...	Additional arguments passed to methods.

Details

Custom filter types should implement an S7 method for this generic. The default method returns the raw value with no domain logic.

Value

The effective value for filtering.

cb_intersect_domain_values *Intersect two domain values for a filter type*

Description

Combines two already-computed domain values (not value-vs-domain) into their intersection, dispatching on filter type. Used by ‘"filter"‘-mode domain propagation when the same logical filter appears in more than one upstream step and its effective domains must be combined. Unlike [cb_intersect_domain()], this never emits trimming messages.

Usage

```
cb_intersect_domain_values(filter, ...)
```

Arguments

filter	S7 filter object (used for type dispatch only).
...	Additional arguments. Methods receive two domain values ‘a‘ and ‘b‘ to intersect (either may be ‘NULL‘).

Value

The intersected domain value, or ‘NULL‘.

cb_plot_filter_data *Plot filter data*

Description

Plot filter data

Usage

```
cb_plot_filter_data(filter, source, ...)
```

Arguments

filter	S7 filter object.
source	Source object.
...	Additional arguments passed to plotting functions. Methods receive a ‘data_object‘ (the data to plot) here.

Value

Plot side effect.

cb_register_tool *Register cohortBuilder tools with an ellmer chat*

Description

cb_register_tool registers a single cb_tool with an **ellmer** chat object. cb_register_tools is a convenience wrapper that registers all built-in tools at once.

Usage

```
cb_register_tool(chat, tool)

cb_register_tools(chat, cohort)
```

Arguments

chat	An **ellmer** chat object (e.g. from ellmer::chat_openai()).
tool	A cb_tool object.
cohort	A Cohort object.

Value

The chat object, invisibly (for piping).

Examples

```
## Not run:
source <- set_source(tblist(iris = iris)) |> autofilter(attach_as = "meta")
coh <- cohort(source)
chat <- ellmer::chat_openai()
chat |> cb_register_tools(coh)
chat$chat("Show me the available filters")

## End(Not run)
```

cb_tool *Create a cohortBuilder tool definition*

Description

Constructs a tool object that can be registered with an LLM chat via cb_register_tool.

Usage

```
cb_tool(fun, name, description, arguments = list())

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

Arguments

fun	Function to invoke when the tool is called.
name	Character string identifying the tool.
description	Character string describing what the tool does. The more detail provided, the better the LLM can decide when to use it.
arguments	Named list of argument type definitions created by **ellmer** type_*() functions (e.g. ellmer::type_string()).
x	A cb_tool object.
...	Ignored.

Value

An object of class cb_tool.

---

cb_tool_add_filters *Create a tool for adding filters to a cohort*

Description

Returns a cb_tool that adds selected filters from the source's available_filters to the cohort. The LLM chooses whether to add to a new step or the existing last step via the action argument.

Usage

```
cb_tool_add_filters(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_apply_filters *Create a tool that adds filters and sets their values in one call*

Description

Returns a cb_tool that combines filter addition and value assignment into a single tool call. This avoids issues with LLMs splitting the work across multiple parallel calls.

Usage

```
cb_tool_apply_filters(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_clear_filters *Create a tool for resetting filters to defaults*

Description

Returns a cb_tool that resets filter values to their defaults without removing the filters from the cohort.

Usage

```
cb_tool_clear_filters(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_describe_state

Create a tool returning the current cohort state

Description

Returns a cb_tool whose function takes no arguments and returns a structured text summary of all steps, filters, their active status, and pending state.

Usage

```
cb_tool_describe_state(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_filters_meta

Create a tool returning available filters metadata

Description

Returns a cb_tool whose function takes no arguments and returns a JSON string with filter metadata from shape.

Usage

```
cb_tool_filters_meta(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_get_code *Create a tool returning reproducible filtering code*

Description

Returns a cb_tool that generates reproducible R code for the current cohort filtering pipeline via get_code().

Usage

```
cb_tool_get_code(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_get_data_summary *Create a tool returning row counts per dataset and step*

Description

Returns a cb_tool that reports how many rows remain in each dataset at each step (before and after filtering). This is the primary tool for understanding the impact of applied filters.

Usage

```
cb_tool_get_data_summary(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_remove_filters *Create a tool for removing filters from the cohort*

Description

Returns a cb_tool that removes filters from a step in the cohort. If removing all filters from a step, the entire step is removed.

Usage

```
cb_tool_remove_filters(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_remove_step *Create a tool for removing the last step*

Description

Returns a cb_tool that removes the last step from the cohort, including all its filters.

Usage

```
cb_tool_remove_step(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_run *Create a tool for running the cohort pipeline*

Description

Returns a cb_tool that triggers data calculations for the entire cohort or a specific step. Only functional when the cb_tool_run_cohort option is FALSE; otherwise returns an informative message that the cohort runs automatically.

Usage

```
cb_tool_run(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_set_filter_values *Create a tool for setting filter values*

Description

Returns a cb_tool that updates filter parameter values on the last step of the cohort and triggers the data pipeline.

Usage

```
cb_tool_set_filter_values(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

cb_tool_toggle_filters *Create a tool for activating or deactivating filters*

Description

Returns a cb_tool that toggles the active state of existing filters in the cohort.

Usage

```
cb_tool_toggle_filters(cohort)
```

Arguments

cohort	A Cohort object.

Value

A cb_tool object.

code *Return reproducible data filtering code.*

Description

Return reproducible data filtering code.

Usage

```
code(
  x,
  include_source = TRUE,
  include_methods = c(".pre_filtering", ".post_filtering", ".run_binding"),
  include_action = c("pre_filtering", "post_filtering", "run_binding"),
  modifier = .repro_code_tweak,
  mark_step = TRUE,
  ...
)
```

Arguments

x	Cohort object.
include_source	If 'TRUE' source generating code will be included.
include_methods	Which methods definition should be included in the result.
include_action	Which action should be returned in the result. 'pre_filtering'/'.post_filtering' - to include data transformation before/after filtering. s'run_binding' - data binding transformation.
modifier	A function taking data frame (storing reproducible code metadata) as an argument, and returning data frame with 'expr' column which is then combined into a single expression (final result of 'get_code'). See .repro_code_tweak.
mark_step	Include information which filtering step is performed.
...	Other parameters passed to tidy_source.

Value

tidy_source output storing reproducible code for generating final step data.

See Also

cohort-methods

Cohort *R6 class representing Cohort object.*

Description

R6 class representing Cohort object.

R6 class representing Cohort object.

Details

Cohort object is designed to make operations on source data possible.

Public fields

attributes List of Cohort attributes defined while creating a new Cohort object.

Methods

Public methods:

- Cohort$new()
- Cohort$add_source()
- Cohort$update_source()
- Cohort$get_source()
- Cohort$get_propagate_domains_mode()
- Cohort$add_step()
- Cohort$copy_step()
- Cohort$remove_step()
- Cohort$add_filter()
- Cohort$remove_filter()
- Cohort$update_filter()
- Cohort$clear_filter()
- Cohort$clear_step()
- Cohort$sum_up_state()
- Cohort$get_state()
- Cohort$restore()
- Cohort$get_data()
- Cohort$plot_data()
- Cohort$show_attrition()
- Cohort$calc_stats()
- Cohort$show_help()
- Cohort$get_code()
- Cohort$run_flow()
- Cohort$run_step()
- Cohort$bind_data()
- Cohort$describe_state()
- Cohort$get_step()
- Cohort$get_filter()
- Cohort$update_stats()
- Cohort$get_stats()
- Cohort$list_active_filters()
- Cohort$last_step_id()
- Cohort$is_pending()
- Cohort$set_pending()
- Cohort$set_pending_cascade()
- Cohort$set_domain()
- Cohort$propagate_domains_to()
- Cohort$propagate_filter_domains()
- Cohort$modify()
- Cohort$clone()

Method `new()`**:** Create Cohort object.

Usage:

```
Cohort$new(
  source,
  ...,
  run_flow = FALSE,
  compute_stats = TRUE,
  propagate_domains = c("none", "filter", "stats", "data"),
 hook = list(pre = get_hook("pre_cohort_hook"), post = get_hook("post_cohort_hook"))
)
```

Arguments:

`source` Source object created with set_source.

`...` Steps definition (optional). Can be also defined as a sequence of filters - the filters will be added to the first step.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`compute_stats` If 'TRUE' (default), filter statistics are computed and stored after each step. Set to 'FALSE' to skip stats computation (useful for metadata-only operation).

`propagate_domains` Domain propagation mode between steps. One of '"none"' (default, no propagation), '"filter"' (derive from previous step filter values), '"stats"' (derive from stored statistics), or '"data"' (scan filtered data). '"stats"' requires 'compute_stats = TRUE'; use '"data"' for the stats-free equivalent.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Returns: The object of class 'Cohort'.

Method `add_source()`**:** Add Source to Cohort object.

Usage:

```
Cohort$add_source(source)
```

Arguments:

`source` Source object created with set_source.

Method `update_source()`**:** Update Source in the Cohort object.

Usage:

```
Cohort$update_source(
  source,
  keep_steps = !has_steps(source),
  run_flow = FALSE,
  hook = list(pre = get_hook("pre_update_source_hook"), post =
    get_hook("post_update_source_hook"))
)
```

Arguments:

`source` Source object created with set_source.

`keep_steps` If 'TRUE', steps definition remains unchanged when updating source. If 'FALSE' steps configuration is deleted. If vector of type integer, specified steps will remain.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `get_source()`**:** Return Source object attached to Cohort.

Usage:

```
Cohort$get_source()
```

Method `get_propagate_domains_mode()`**:** Return the configured domain propagation mode. One of '"none"', '"filter"', '"stats"' or '"data"'. Read-only; the mode is fixed at construction. UI layers can inspect it (e.g. to validate that a domain-based rendering strategy is compatible with the cohort).

Usage:

```
Cohort$get_propagate_domains_mode()
```

Method `add_step()`**:** Add filtering step definition

Usage:

```
Cohort$add_step(
  step,
  run_flow = FALSE,
  hook = list(pre = get_hook("pre_add_step_hook"), post = get_hook("post_add_step_hook"))
)
```

Arguments:

`step` Step definition created with step.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `copy_step()`**:** Copy selected step.

Usage:

```
Cohort$copy_step(step_id, filters, run_flow = FALSE)
```

Arguments:

`step_id` Id of the step to be copied. If missing the last step is taken. The copied step is added as the last one in the Cohort.

`filters` List of Source-evaluated filters to copy to new step.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

Method `remove_step()`**:** Remove filtering step definition

Usage:

```
Cohort$remove_step(
  step_id,
  run_flow = FALSE,
  hook = list(pre = get_hook("pre_rm_step_hook"), post = get_hook("post_rm_step_hook"))
)
```

Arguments:

`step_id` Id of the step to remove.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `add_filter()`**:** Add filter definition

Usage:

```
Cohort$add_filter(
  filter,
  step_id,
  run_flow = FALSE,
  hook = list(pre = get_hook("pre_add_filter_hook"), post =
    get_hook("post_add_filter_hook"))
)
```

Arguments:

`filter` Filter definition created with filter.

`step_id` Id of the step to add the filter to. If missing, filter is added to the last step.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `remove_filter()`**:** Remove filter definition

Usage:

```
Cohort$remove_filter(
  step_id,
  filter_id,
  run_flow = FALSE,
  hook = list(pre = get_hook("pre_rm_filter_hook"), post =
    get_hook("post_rm_filter_hook"))
)
```

Arguments:

`step_id` Id of the step from which filter should be removed.

`filter_id` Id of the filter to be removed.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `update_filter()`**:** Update filter definition

Usage:

```
Cohort$update_filter(
  step_id,
  filter_id,
  ...,
  active,
  run_flow = FALSE,
```

```
  hook = list(pre = get_hook("pre_update_filter_hook"), post =
    get_hook("post_update_filter_hook")),
  hook_args = list(pre = list(), post = list())
)
```

Arguments:

`step_id` Id of the step where filter is defined.

`filter_id` Id of the filter to be updated.

`...` Filter parameters that should be updated.

`active` Mark filter as active ('TRUE') or inactive ('FALSE').

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

`hook_args` Named list of extra arguments passed to pre/post hooks.

Method `clear_filter()`**:** Reset filter to its default values.

Usage:

```
Cohort$clear_filter(step_id, filter_id, run_flow = FALSE)
```

Arguments:

`step_id` Id of the step where filter is defined.

`filter_id` Id of the filter which should be cleared.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

Method `clear_step()`**:** Reset all filters included in selected step.

Usage:

```
Cohort$clear_step(step_id, run_flow = FALSE)
```

Arguments:

`step_id` Id of the step where filters should be cleared.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

Method `sum_up_state()`**:** Sum up Cohort configuration - Source, steps definition and evaluated data.

Usage:

```
Cohort$sum_up_state()
```

Method `get_state()`**:** Get Cohort configuration state.

Usage:

```
Cohort$get_state(step_id, json = FALSE)
```

Arguments:

`step_id` If provided, the selected step state is returned.

`json` If TRUE, return state in JSON format. Restore Cohort configuration.

Method `restore()`**:**

Usage:

```
Cohort$restore(
  state,
  modifier = function(prev_state, state) state,
  run_flow = FALSE,
 hook = list(pre = get_hook("pre_restore_hook"), post = get_hook("post_restore_hook"))
)
```

Arguments:

`state` List or JSON string containing steps and filters configuration.

`modifier` Function two parameters combining the previous and provided state. The returned state is then restored.

`run_flow` If 'TRUE', data flow is run after the operation is completed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `get_data()`**:** Get step related data

Usage:

```
Cohort$get_data(step_id, state = "post", collect = TRUE)
```

Arguments:

`step_id` Id of the step from which to source data.

`state` Return data before ("pre") or after ("post") step filtering?

`collect` Return raw data source ('FALSE') object or collected (to R memory) data ('TRUE').

Method `plot_data()`**:** Plot filter specific data summary.

Usage:

```
Cohort$plot_data(step_id, filter_id, ..., state = "post")
```

Arguments:

`step_id` Id of the step where filter is defined.

`filter_id` Id of the filter for which the plot should be returned

`...` Another parameters passed to filter specific method.

`state` Generate plot on data before ("pre") or after ("post") step filtering?

Method `show_attrition()`**:** Show attrition plot.

Usage:

```
Cohort$show_attrition(..., percent = FALSE)
```

Arguments:

`...` Source specific parameters required to generate attrition.

`percent` Should attrition changes be presented with percentage values.

Method `calc_stats()`**:** Get Cohort related statistics.

Usage:

```
Cohort$calc_stats(step_id, filter_id, ..., state = "post")
```

Arguments:

`step_id` When ‘filter_id‘ specified, ‘step_id‘ precises from which step the filter comes from. Otherwise data from specified step is used to calculate required statistics.

`filter_id` If not missing, filter related data statistics are returned.

`...` Specific parameters passed to filter related method.

`state` Should the stats be calculated on data before ("pre") or after ("post") filtering in specified step.

Method `show_help()`**:** Show source data or filter description

Usage:

```
Cohort$show_help(
  field,
  step_id,
  filter_id,
  modifier = getOption("cb_help_modifier", default = function(x) x)
)
```

Arguments:

`field` Name of the source description field provided as ‘description‘ argument to set_source. If missing, ‘step_id‘ and ‘filter_id‘ are used to return filter description.

`step_id` Id of the filter step to return description of.

`filter_id` Id of the filter to return description of.

`modifier` A function taking the description as argument. The function can be used to modify its argument (convert to html, display in browser etc.).

Method `get_code()`**:** Return reproducible data filtering code.

Usage:

```
Cohort$get_code(
  include_source = TRUE,
  include_methods = c(".pre_filtering", ".post_filtering", ".run_binding"),
  include_action = c("pre_filtering", "post_filtering", "run_binding"),
  modifier = .repro_code_tweak,
  mark_step = TRUE,
  ...
)
```

Arguments:

`include_source` If ‘TRUE‘ source generating code will be included.

`include_methods` Which methods definition should be included in the result.

`include_action` Which action should be returned in the result. ‘pre_filtering‘/‘.post_filtering‘ - to include data transformation before/after filtering. s‘run_binding‘ - data binding transformation.

`modifier` A function taking data frame (storing reproducible code metadata) as an argument, and returning data frame with ‘expr‘ column which is then combined into a single expression (final result of ‘get_code‘). See .repro_code_tweak.

`mark_step` Include information which filtering step is performed.

`...` Other parameters passed to tidy_source.

Method `run_flow()`**:** Trigger data calculations sequentially.

Usage:

```
Cohort$run_flow(
  min_step,
 hook = list(pre = get_hook("pre_run_flow_hook"), post = get_hook("post_run_flow_hook"))
)
```

Arguments:

`min_step` Step id starting from the calculation will be started.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `run_step()`**:** Trigger data calculations for selected step.

Usage:

```
Cohort$run_step(
  step_id,
 hook = list(pre = get_hook("pre_run_step_hook"), post = get_hook("post_run_step_hook"))
)
```

Arguments:

`step_id` Id of the step for which to run data calculation.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `bind_data()`**:** Run data binding for selected step. See more at binding-keys.

Usage:

```
Cohort$bind_data(step_id)
```

Arguments:

`step_id` Id of the step for which to bind the data.

Method `describe_state()`**:** Print defined steps configuration.

Usage:

```
Cohort$describe_state(to_string = FALSE)
```

Arguments:

`to_string` If 'TRUE', return the output as a character string instead of printing it. Defaults to 'FALSE'.

Method `get_step()`**:** Get selected step configuration.

Usage:

```
Cohort$get_step(step_id)
```

Arguments:

`step_id` Id of the step to be returned.

Method `get_filter()`**:** Get selected filter configuration.

Usage:

```
Cohort$get_filter(step_id, filter_id, method = function(x) x)
```

Arguments:

`step_id` Id of the step where filter is defined.

`filter_id` If of the filter to be returned.

`method` Custom function taking filters list as argument.

Method `update_stats()`**:** Update filter or step statistics. Computes and stores step and filter attached data statistics such as number of data rows, filter choices or frequencies.

Usage:

```
Cohort$update_stats(step_id, filter_id, state = "post", name = NULL)
```

Arguments:

`step_id` Id of the step for which statistics should be computed. If 'filter_id' is not missing, the parameter describes id of the step where filter should be found.

`filter_id` Id of the filter for which statistics should be computed.

`state` Should statistics be computed on data before ("pre") or after ("post") filtering in specified step.

`name` Optional name(s) of the individual filter statistics to (re)compute. When supplied (filter-level only), only those statistics are computed and merged into the stored entry, leaving any other stored statistics untouched. When missing, the full statistics set is computed.

Method `get_stats()`**:** Return step or filter specific statistics.

Usage:

```
Cohort$get_stats(
  step_id,
  filter_id,
  state = "post",
  .recalc_when_missing = TRUE,
  name = NULL
)
```

Arguments:

`step_id` Id of the step for which stored statistics should be returned If 'filter_id' is not missing, the parameter describes id of the step where filter should be found.

`filter_id` Id of the filter for which stored statistics should be returned.

`state` Should statistics be returned on data before ("pre") or after ("post") filtering in specified step.

`.recalc_when_missing` Should the function compute statistics automatically when not computed yet?

`name` Optional name of a single filter statistic to return (e.g. "choices", "n_data"). When supplied (filter-level only) the method returns just that statistic and, if recomputation is needed, computes only it instead of the whole statistics set. When missing, the full stored entry (a list of all statistics) is returned.

Method `list_active_filters()`**:** List active filters included in selected step.

Usage:

```
Cohort$list_active_filters(step_id)
```

Arguments:

`step_id` Id of the step where filters should be found.

Method `last_step_id()`**:** Return id of the last existing step in Cohort.

Usage:

```
Cohort$last_step_id()
```

Method `is_pending()`**:** Check if step is pending.

Usage:

```
Cohort$is_pending(step_id)
```

Arguments:

`step_id` Id of the step to be checked.

Method `set_pending()`**:** Mark step as pending or resolved.

Usage:

```
Cohort$set_pending(
  step_id,
  pending = TRUE,
  hook = list(pre = get_hook("pre_set_pending_hook"), post =
    get_hook("post_set_pending_hook"))
)
```

Arguments:

`step_id` Id of the step.

`pending` Logical; 'TRUE' to mark pending, 'FALSE' to resolve.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `set_pending_cascade()`**:** Mark a step and all the steps after it as pending.

A change to step 'step_id''s filters invalidates that step and every step downstream (each step's input is the previous step's output). Used by the filter/step mutating methods so the GUI greys all affected steps via the 'post_set_pending_hook'.

Usage:

```
Cohort$set_pending_cascade(step_id)
```

Arguments:

`step_id` Id of the first step to mark pending.

Method `set_domain()`**:** Silently set a filter's domain.

Internal setter used by domain propagation (.propagate_domains). Writes 'filter@domain' directly **without** firing the 'update_filter' hooks and **without** triggering 'run_flow', which prevents the per-hop hook re-entry that would otherwise occur if domains were applied through 'update_filter'. Does not itself trigger further propagation.

Usage:

```
Cohort$set_domain(step_id, filter_id, domain)
```

Arguments:

`step_id` Id of the step where the filter is defined.

`filter_id` Id of the filter whose domain should be set.

`domain` New domain value to assign.

Method `propagate_domains_to()`**:** Recompute a target step's domains from its parent and signal the change.

Thin wrapper around .propagate_domains that runs propagation for 'target_id' (computing its filters' domains from step 'target_id - 1') and then fires 'post_propagate_domains_hook' so UI layers can refresh the affected step's inputs. No-op when propagation is disabled or the target step is absent / has no parent.

Usage:

```
Cohort$propagate_domains_to(
  target_id,
  hook = get_hook("post_propagate_domains_hook")
)
```

Arguments:

`target_id` Id of the step whose domains should be recomputed.

`hook` List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Method `propagate_filter_domains()`**:** Eagerly propagate '"filter"'-mode domains to a target step and all steps after it.

'"filter"' mode needs no computed data, so domains can be (re)established immediately when the step structure changes. Because each step's domain depends on every previous step, this recomputes 'from_target' and cascades downstream ascending. No-op unless the propagation mode is '"filter"'.

Usage:

```
Cohort$propagate_filter_domains(from_target)
```

Arguments:

`from_target` Id of the first step to recompute.

Method `modify()`**:** Helper method enabling to run non-standard operation on Cohort object.

Usage:

```
Cohort$modify(modifier)
```

Arguments:

`modifier` Function of two arguments 'self' and 'private'.

Method `clone()`**:** The objects of this class are cloneable with this method.

Usage:

```
Cohort$clone(deep = FALSE)
```

Arguments:

`deep` Whether to make a deep clone.

cohort-methods *Cohort related methods*

Description

The list of methods designed for getting Cohort-related details.

- plot_data - Plot filter related Cohort data.
- stat - Get Cohort related statistics.
- code - Return reproducible data filtering code.
- get_data - Get step related data.
- sum_up - Sum up Cohort state.
- get_state - Save Cohort state.
- restore - Restore Cohort state.
- attrition - Show attrition plot.
- description - Show Source or filter related description.

Value

Various type outputs dependent on the selected method. See each method documentation for details.

create-cohort *Create new 'Cohort' object*

Description

Cohort object is designed to make operations on source data possible.

Usage

```
cohort(
  source,
  ...,
  run_flow = FALSE,
  compute_stats = TRUE,
  propagate_domains = c("none", "filter", "stats", "data"),
 hook = list(pre = get_hook("pre_cohort_hook"), post = get_hook("post_cohort_hook"))
)
```

Arguments

source	Source object created with set_source.
...	Steps definition (optional). Can be also defined as a sequence of filters - the filters will be added to the first step.
run_flow	If 'TRUE', data flow is run after the operation is completed.
compute_stats	If 'TRUE' (default), filter and step statistics are computed and stored after each step. Set to 'FALSE' for metadata-only operation.
propagate_domains	Domain propagation mode between steps: '"none"' (default), '"filter"' (from previous step filter values), '"stats"' (from stored statistics; requires 'compute_stats = TRUE'), or '"data"' (scan filtered data; the stats-free equivalent).
hook	List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Value

The object of class 'Cohort'.

---

data_key *Define Source dataset key*

Description

Data keys are used to define primary_keys and binding-keys.

Usage

```
data_key(dataset, key)
```

Arguments

dataset	Name of the dataset included in Source.
key	Character or character vector storing column names to be used as table keys.

Value

'data_key' class list of two objects: 'dataset' and 'key' storing name and vector of data key names respectively.

describe *Create a description object*

Description

Helper for building structured description entries used in source definition.

Usage

```
describe(description, label = NULL, ...)
```

Arguments

description	A single string describing the field.
label	Optional short, human-readable label for the field. When the field describes a variable, [autofilter()] reuses the label to fill the generated filter's 'name'. 'NULL' (default) leaves the name untouched.
...	Additional named parameters to include in the description object.

Value

A named list with 'text', an optional 'label', and any extra parameters.

description *Show source data or filter description*

Description

If defined allows to check the provided description related to source data or configured filters.

Usage

```
description(
  x,
  field,
  step_id,
  filter_id,
  modifier = getOption("cb_help_modifier", default = function(x) x)
)
```

Arguments

x	Cohort object.
field	Name of the source description field provided as ‘description‘ argument to set_source. If missing, ‘step_id‘ and ‘filter_id‘ are used to return filter description.
step_id	Id of the filter step to return description of.
filter_id	Id of the filter to return description of.
modifier	A function taking the description as argument. The function can be used to modify its argument (convert to html, display in browser etc.).

Value

Any object (or its subset) attached to Source of filter via description argument.

See Also

cohort-methods

---

filter *Define Cohort filter*

Description

Creates an S7 filter object of the specified type.

Usage

```
filter(type, ...)
```

Arguments

type	Type of filter to use (e.g., "discrete", "range", "date_range").
...	Filter type-specific parameters.

Value

An S7 filter object inheriting from ‘CbFilter‘.

`filter_domain` *Get a filter's domain*

Description

Returns the declared domain (universe of valid values) attached to a filter, or 'NULL' when no domain is set. This is a thin accessor over the 'domain' property, intended for downstream consumers (e.g. GUIs) that render filter inputs from the domain without reaching into S7 internals.

Usage

```
filter_domain(filter)
```

Arguments

`filter`	S7 filter object.

Value

The filter's domain, or 'NULL' when unset. Structure depends on the filter type (e.g. a character vector for discrete filters, a two-element vector for range filters).

See Also

[filter_effective_value()], [cb_intersect_domain()]

`filter_effective_value` *Get a filter's effective value*

Description

Returns the value that should be used to pre-select the filter input, accounting for the domain. This is the filter's value intersected with its domain; when the value is unset ('NA') and a domain is present, the domain is returned as the effective value. Thin wrapper over [cb_intersect_domain()] for use by GUIs.

Usage

```
filter_effective_value(filter)
```

Arguments

`filter`	S7 filter object.

Value

The effective value for the filter, suitable for pre-selecting an input. Structure depends on the filter type.

See Also

[filter_domain()], [cb_intersect_domain()]

---

filter_variables *Get the variables a filter operates on*

Description

Returns the column name(s) a filter is bound to, regardless of whether the filter stores them in a single-variable property (‘variable‘) or a multi-variable property (‘variables‘).

Usage

```
filter_variables(filter)
```

Arguments

filter	S7 filter object.

Value

Character vector of variable (column) names. Empty when the filter declares no variables.

---

get_data *Get step related data*

Description

Get step related data

Usage

```
get_data(x, step_id, state = "post", collect = FALSE)
```

Arguments

x	Cohort object.
step_id	Id of the step from which to source data.
state	Return data before ("pre") or after ("post") step filtering?
collect	Return raw data source (‘FALSE‘) object or collected (to R memory) data (‘TRUE‘).

Value

Subset of Source-specific data connection object or its evaluated version.

See Also

cohort-methods

`get_filter_params` *Get filter parameters as a list*

Description

Extracts all user-facing properties from an S7 filter object.

Usage

```
get_filter_params(filter, name)
```

Arguments

`filter`	S7 filter object.
`name`	Optional parameter name to retrieve a single value.

Value

Named list of filter parameters, or a single value if ‘name‘ is given. Properties stored in ‘filter@private‘ are always excluded.

`get_state` *Get Cohort configuration state.*

Description

Get Cohort configuration state.

Usage

```
get_state(x, step_id, json = FALSE)
```

Arguments

`x`	Cohort object.
`step_id`	If provided, the selected step state is returned.
`json`	If TRUE, return state in JSON format.

Value

List object of character string being the list conversion to JSON format.

See Also

cohort-methods

`hooks` *Cohort hooks.*

Description

In order to make integration of ‘cohortBuilder‘ package with other layers/packages easier, hooks system was introduced.

Usage

```
add_hook(name, method)

get_hook(name)
```

Arguments

`name`	Name of the hook. See Details section.
`method`	Function to be assigned as hook.

Details

Many Cohort methods allow to define ‘hook‘ parameter. For such method, ‘hook‘ is a list containing two values: ‘pre‘ and ‘post‘, storing functions (hooks) executed before and after the method is run respectively.

Each ‘hook‘ is a function of two obligatory parameters:

- `public` - Cohort object.
- `private` - Private environment of Cohort object.

When Cohort method, for which hook is defined, allow to pass custom parameters, the ones should be also available in hook definition (with some exclusions, see below).

For example ‘Cohort$remove_step‘ has three parameters:

- `step_id`
- `run_flow`
- `hook`

By the implementation, the parameters that we should skip are ‘run_flow‘ and ‘hook‘, so the hook should have three parameters ‘public‘, ‘private‘ and ‘step_id‘.

There are two ways of defining hooks for the specific method. The first one is to define the method ‘hook‘ directly as its parameter (while calling the method).

The second option can be achieved with usage of ‘add_hook‘ (and ‘get_hook‘) function. The default ‘hook‘ parameter for each method is constructed as below:

```
remove_step = function(step_id, run_flow = FALSE,
  hook = list(
    pre = get_hook("pre_rm_step_hook"),
    post = get_hook("post_rm_step_hook")
  )
)
```

’Pre’ hooks are defined with ’pre_<method_name>_hook’ and ’Post’ ones as ’post_<method_name>_hook’. As a result calling:

```
add_hook(
  "pre_remove_step_hook",
  function(public, private, step_id) {...}
)
```

will result with specifying a new pre-hook for ‘remove_step‘ method.

You may add as many hooks as you want. The order of hooks execution is followed by the order or registering process. If you want to check currently registered hooks for the specific method, just use:

```
get_hook("pre_<method_name>_hook")
```

Value

No returned value (‘add_hook‘) or the list of functions (‘get_hook‘).

`librarian` *Sample of library database*

Description

A list containing four data frames reflecting library management database.

Usage

```
librarian
```

Format

A list of four data frames:

books - books on store

`isbn` book ISBN number
`title` book title
`genre` comma separated book genre
`publisher` name of book publisher
`author` name of book author
`copies` total number of book copies on store

borrowers - registered library members

`id` member unique id
`registered` date the member joined library
`address` member address
`name` full member name
`phone_number` member phone number
`program` membership program type (standard, premium or vip)

issues - borrowed books events

`id` unique event id
`borrower_id` id of the member that borrowed the book
`isbn` is of the borrowed book
`date` date of borrow event

returns - returned books events

`id` event id equal to borrow issue id
`date` date of return event

`managing-cohort` *Managing the Cohort object*

Description

The list of methods designed for managing the Cohort configuration and state.

- add_source - Add source to Cohort object.
- update_source - Update Cohort object source.
- add_step - Add step to Cohort object.
- rm_step - Remove step from Cohort object.
- add_filter - Add filter to Cohort step.
- rm_filter - Remove filter from Cohort step.
- update_filter - Update filter configuration.
- run - Run data filtering.

Value

The object of class ‘Cohort‘ having the modified configuration dependent on the used method.

`managing-source` *Managing the Source object*

Description

The list of methods designed for managing the Source configuration and state.

- add_step - Add step to Source object.
- rm_step - Remove step from Source object.
- add_filter - Add filter to Source step.
- rm_filter - Remove filter from Source step.
- update_filter - Update filter configuration.

Value

The object of class ‘Source‘ having the modified configuration dependent on the used method.

See Also

managing-cohort

`plot_data` *Plot filter related Cohort data.*

Description

For specified filter the method calls filter-related plot method to present data.

Usage

```
plot_data(x, step_id, filter_id, ..., state = "post")
```

Arguments

`x`	Cohort object.
`step_id`	Id of step in which the filter was defined..
`filter_id`	Filter id.
`...`	Another parameters passed to filter plotting method.
`state`	Generate plot based on data before ("pre") or after ("post") filtering.

Value

Filter-specific plot.

See Also

cohort-methods

`primary_keys` *Define Source datasets primary keys*

Description

Primary keys can be defined as ‘primary_keys‘ parameter of set_source method. Currently, primary keys are used only to show keys information in attrition plot (See attrition).

Usage

```
primary_keys(...)
```

Arguments

...	Data keys describing tables primary keys.

Value

List of class ‘primary_keys‘ storing data_keys objects.

Examples

```
primary_keys(
  data_key('books', 'book_id'),
  data_key('borrowed', c('user_id', 'books_id', 'date'))
)
```

`register_filter_type` *Register a custom filter type*

Description

Registers an S7 filter constructor so it can be used with `filter("type", ...)`. The constructor must return an object inheriting from CbFilter.

Usage

```
register_filter_type(type, constructor)
```

Arguments

`type`	Character string identifying the filter type.
`constructor`	S7 class constructor (e.g. created with [S7::new_class()]).

Examples

```
## Not run:
MyCbFilter <- S7::new_class("MyCbFilter",
  parent = CbFilter,
  package = "mypackage",
  properties = list(dataset = S7::class_character, variable = S7::class_character),
  constructor = function(id = NULL, name = NULL, variable, dataset,
                         description = NULL, domain = NULL, ...) {
    id <- id %||% .default_filter_id(dataset, variable)
    name <- name %||% id
    S7::new_object(S7::S7_object(),
      type = "my_filter", id = id, name = name,
      dataset = dataset, variable = variable,
      active = TRUE, description = description, domain = domain,
      extra = list(...), private = list(input_param = "value")
    )
  }
)
register_filter_type("my_filter", MyCbFilter)
# Now filter("my_filter", ...) works

## End(Not run)
```

restore *Restore Cohort object.*

Description

The method allows to restore Cohort object with provided configuration state.

Usage

```
restore(
  x,
  state,
  modifier = function(prev_state, state) state,
  run_flow = FALSE
)
```

Arguments

x	Cohort object.
state	List or JSON string containing steps and filters configuration. See get_state.
modifier	Function two parameters combining the previous and provided state. The returned state is then restored.
run_flow	If TRUE, filtering flow is applied when the operation is finished.

Value

The 'Cohort' class object having the state restored based on provided config.

See Also

cohort-methods

rm_filter *Remove filter definition*

Description

Remove filter definition

Usage

```
rm_filter(x, step_id, filter_id, ...)

## S3 method for class 'Cohort'
rm_filter(
  x,
  step_id,
  filter_id,
  run_flow = FALSE,
  hook = list(pre = get_hook("pre_rm_filter_hook"), post =
    get_hook("post_rm_filter_hook")),
  ...
)

## S3 method for class 'Source'
rm_filter(x, step_id, filter_id, ...)
```

Arguments

x	An object from which filter should be removed.
step_id	Id of the step from which filter should be removed.
filter_id	Id of the filter to be removed.
...	Other parameters passed to specific S3 method.
run_flow	If 'TRUE', data flow is run after the filter is removed.
hook	List of hooks describing methods to run before/after the filter is removed. See hooks for more details.

Value

Method dependent object (i.e. 'Cohort' or 'Source') having selected filter removed.

See Also

managing-cohort, managing-source

---

rm_step *Remove filtering step definition*

Description

Remove filtering step definition

Usage

```
rm_step(x, step_id, ...)

## S3 method for class 'Cohort'
rm_step(
  x,
  step_id,
  run_flow = FALSE,
 hook = list(pre = get_hook("pre_rm_step_hook"), post = get_hook("post_rm_step_hook")),
  ...
)

## S3 method for class 'Source'
rm_step(x, step_id, ...)
```

Arguments

x	An object from which step should be removed.
step_id	Id of the step to remove.
...	Other parameters passed to specific S3 method.
run_flow	If 'TRUE', data flow is run after the step is removed.
hook	List of hooks describing methods before/after the Cohort is created. See hooks for more details.

Value

Method dependent object (i.e. 'Cohort' or 'Source') having selected step removed.

See Also

managing-cohort, managing-source

---

run *Trigger data calculations.*

Description

Trigger data calculations.

Usage

```
run(x, min_step_id, step_id)
```

Arguments

x	Cohort object.
min_step_id	Step id starting from the calculation will be started. Used only when 'step_id' is missing.
step_id	Id of the step for which to run data calculation.

Value

The object of class 'Cohort' having up-to-date data based on the Cohort state.

See Also

managing-cohort

---

set_source *Create Cohort source*

Description

Source is an object storing information about data source such as source type, primary keys and relations between stored data.

Usage

```
set_source(
  dtconn,
  ...,
  primary_keys = NULL,
  binding_keys = NULL,
  source_code = NULL,
  description = NULL,
  available_filters = NULL,
  compute_meta_stats = getOption("cb.source_filters_meta_stats", TRUE)
)

## S3 method for class 'tblist'
set_source(
  dtconn,
  primary_keys = NULL,
  binding_keys = NULL,
  source_code = NULL,
  description = NULL,
  available_filters = NULL,
  compute_meta_stats = getOption("cb.source_filters_meta_stats", TRUE),
  ...
)
```

Arguments

dtconn	An object defining source data connection.
...	Source type specific parameters. Available in 'attributes' list of resulting object.
primary_keys	Definition of primary keys describing source data (if valid). When provided, affects the output of attrition data plot. See primary_keys.
binding_keys	Definition of binding keys describing relations in source data (if valid). When provided, affects post filtering data. See binding-keys.
source_code	Expression presenting low-level code for creating source. When provided, used as a part of reproducible code output.
description	A named list storing the source objects description. Can be accessed with description Cohort method.
available_filters	List of filter definitions available for the source. See autofilter for generating them automatically.
compute_meta_stats	Whether to pre-compute metadata statistics for the source 'available_filters'. When 'FALSE', the computation is skipped and filter domains fall back to live computation. Defaults to the 'cb.source_filters_meta_stats' option ('TRUE').

Value

R6 object of class inherited from 'dtconn'.

Examples

```
mtcars_source <- set_source(
  tblist(mtcars = mtcars),
  source_code = quote({
    source <- list(dtconn = list(datasets = mtcars))
  })
)
mtcars_source$attributes
```

---

shape *Describe the structure of a source*

Description

'shape()' is an S3 generic that summarizes a source for programmatic or LLM-based inspection. Called with only a 'source', it returns a structured list 'list(datasets, filters)' where 'datasets' maps each dataset name to its description text and 'filters' is keyed by filter id (each entry describing the filter's 'dataset', 'type', 'description', 'variables', and 'domain').

Usage

```
shape(source, ...)

## Default S3 method:
shape(source, field, subfield, ...)

## S3 method for class 'tblist'
shape(source, field, subfield, domains = TRUE, ...)
```

Arguments

source	A ‘Source‘ object.
...	Extra arguments passed to methods.
field	Optional dataset (or description) name to look up.
subfield	Optional variable name within ‘field‘ to look up.
domains	When ‘TRUE‘ (default), each filter entry includes a ‘domain‘ field (its set of valid values). Set to ‘FALSE‘ to omit domains, e.g. when only descriptive metadata is needed and computing domains would be wasteful.

Details

Called with a ‘field‘ (and optional ‘subfield‘), it instead performs a description-text lookup, returning the description stored for that dataset/variable. This form is used internally by the Cohort ‘show_help()‘ method.

Value

Either a ‘list(datasets, filters)‘ metadata structure or, when ‘field‘ is supplied, the description text for the requested entry.

See Also

[describe()], [autofilter()]

Source *R6 class representing a data source*

Description

R6 class representing a data source

R6 class representing a data source

Details

Source is an object storing information about data source such as source type, primary keys and relations between stored data.

Public fields

`dtconn` Data connection object the Source if based on.

`dtvalue` Evaluated data connection value used for computing stats.

`meta_stats` Computed metadata statistics for available filters.

`compute_meta_stats` Whether metadata statistics for available filters are pre-computed.

`description` Source object description list.

`attributes` Extra source parameters passed when source is defined.

`options` Extra configuration options.

`binding_keys` Source data relations expressed as binding-keys.

`primary_keys` Source data primary keys expressed as primary_keys.

`source_code` An expression which allows to recreate basic source structure.

Active bindings

`available_filters` List of filter definitions available for the source.

Methods

Public methods:

- `Source$new()`
- `Source$get()`
- `Source$get_steps()`
- `Source$add_step()`
- `Source$rm_step()`
- `Source$add_filter()`
- `Source$rm_filter()`
- `Source$update_filter()`
- `Source$calc_meta_stats()`
- `Source$clone()`

Method `new()`**:** Create a new 'Source' object.

Usage:

```
Source$new(
  dtconn,
  ...,
  primary_keys = NULL,
  binding_keys = NULL,
  source_code = NULL,
  description = NULL,
  available_filters = NULL,
  compute_meta_stats = getOption("cb.source_filters_meta_stats", TRUE),
  options = list(display_binding = TRUE)
)
```

Arguments:

`dtconn` An object defining source data connection.

`...` Extra Source parameters. Stored within 'attributes' field.

`primary_keys` Definition of data 'primary_keys', if appropriate. See primary_keys.

`binding_keys` Definition of relations between data, if appropriate. See binding-keys.

`source_code` A quote object that allows to recreate basic source structure. Used as a part of reproducible code output, see code.

`description` A named list storing the source objects description. Can be accessed with description Cohort method.

`available_filters` List of filter definitions available for the source.

`compute_meta_stats` Whether to pre-compute metadata statistics for 'available_filters'. When 'FALSE', 'meta_stats' are skipped (and filter domains fall back to live computation). Defaults to the 'cb.source_filters_meta_stats' option ('TRUE').

`options` List of options affecting methods output. Currently supported only 'display_binding' specifying whether reproducible code should include bindings definition.

Returns: A new 'Source' object of class 'Source' (and 'dtconn' object class appended).

Method `get()`**:** Get selected 'Source' object 'attribute'.

Usage:

```
Source$get(param)
```

Arguments:

`param` Name of the attribute.

Method `get_steps()`**:** Returns filtering steps definition, if defined for 'Source'.

Usage:

```
Source$get_steps()
```

Method `add_step()`**:** Add filtering step definition.

Usage:

```
Source$add_step(step)
```

Arguments:

`step` Step definition created with step.

Method `rm_step()`**:** Remove filtering step definition.

Usage:

```
Source$rm_step(step_id)
```

Arguments:

`step_id` Id of the step to be removed.

Method `add_filter()`**:** Add filter definition to selected step.

Usage:

```
Source$add_filter(filter, step_id)
```

Arguments:

`filter` Filter definition created with filter.

`step_id` Id of the step to include the filter to. If skipped the last step is used.

Method `rm_filter()`**:** Remove filter definition from selected step.

Usage:

```
Source$rm_filter(step_id, filter_id)
```

Arguments:

`step_id` Id of the step where filter is defined.

`filter_id` Id of the filter to be removed.

Method `update_filter()`**:** Update filter definition.

Usage:

```
Source$update_filter(step_id, filter_id, ...)
```

Arguments:

`step_id` Id of the step where filter is defined.

`filter_id` Id of the filter to be updated.

`...` Parameters with its new values.

Method `calc_meta_stats()`**:** Calculate metadata statistics for available filters.

Usage:

```
Source$calc_meta_stats()
```

Method `clone()`**:** The objects of this class are cloneable with this method.

Usage:

```
Source$clone(deep = FALSE)
```

Arguments:

`deep` Whether to make a deep clone.

`source-layer` *Source compatibility methods.*

Description

List of methods that allow compatibility of different source types. Most of the methods should be defined in order to make new source layer functioning. See 'Details' section for more information.

Usage

```
.init_step(source, ...)

## Default S3 method:
.init_step(source, ...)

.collect_data(source, data_object)

## Default S3 method:
.collect_data(source, data_object)

.get_stats(source, data_object)

## Default S3 method:
.get_stats(source, data_object)

.pre_filtering(source, data_object, step_id)

.post_filtering(source, data_object, step_id)

.post_binding(source, data_object, step_id)

.repro_code_tweak(source, code_data)

## Default S3 method:
.pre_filtering(source, data_object, step_id)

## Default S3 method:
.post_filtering(source, data_object, step_id)

## Default S3 method:
.post_binding(source, data_object, step_id)

.get_attrition_label(source, step_id, step_filters, ...)

## Default S3 method:
.get_attrition_label(source, step_id, step_filters, ...)

.get_attrition_count(source, data_stats, ...)

## Default S3 method:
.get_attrition_count(source, data_stats, ...)

.run_binding(source, ...)

## Default S3 method:
.run_binding(source, binding_key, data_object_pre, data_object_post, ...)
```

```
## S3 method for class 'tblist'
.init_step(source, ...)

## S3 method for class 'tblist'
.collect_data(source, data_object)

## S3 method for class 'tblist'
.get_stats(source, data_object)
```

Arguments

source	Source object.
...	Other parameters passed to specific method.
data_object	Object that allows source data access. ‘data_object‘ is the result of ‘.init_step‘ method (or object of the same structure).
step_id	Name of the step visible in resulting plot.
code_data	Data frame storing ‘type‘, ‘expr‘ and filter or step related columns.
step_filters	List of step filters.
data_stats	Data frame presenting statistics for each filtering step.
binding_key	Binding key describing currently processed relation.
data_object_pre	Object storing unfiltered data in the current step (previous step result).
data_object_post	Object storing current data (including active filtering and previously done bindings).

Details

The package is designed to make the functionality work with multiple data sources. Data source can be based for example on list of tables, connection to database schema or API service that allows to access and operate on data. In order to make new source type layer functioning, the following list of methods should be defined:

- `.init_source` - Defines how to extract data object from source. Each filtering step assumes to be operating on resulting data object (further named data_object) and returns object of the same type and structure.
- `.collect_data` - Defines how to collect data (into R memory) from ‘data_object‘.
- `.get_stats` - Defines what ‘data_object‘ statistics should be calculated and how. When provided the stats can be extracted using stat.
- `.pre_filtering` - (optional) Defines what operation on ‘data_object‘ should be performed before applying filtering in the step.
- `.post_filtering` - (optional) Defines what operation on ‘data_object‘ should be performed after applying filtering in the step (before running binding).
- `.post_binding` - (optional) Defines what operation on ‘data_object‘ should be performed after applying binding in the step.

- `.run_binding` - (optional) Defines how to handle post filtering data binding. See more about binding keys at binding-keys.
- `.get_attrition_count` and `.get_attrition_label` - Methods defining how to get statistics and labels for attrition plot.
- `.repro_code_tweak` - (optional) Default method passed as a ‘modifier‘ argument of code function. Aims to modify reproducible code into the final format.

Except from the above methods, you may extend the existing or new source with providing custom filtering methods. See ‘vignette("custom-filters")‘. In order to see more details about how to implement custom source check ‘vignette("custom-extensions")‘.

Value

Depends on specific method. See ‘vignette("custom-extensions")‘ for more details.

`stat` *Get Cohort related statistics.*

Description

Display data statistics related to specified step or filter.

Usage

```
stat(x, step_id, filter_id, ..., state = "post")
```

Arguments

`x`	Cohort object.
`step_id`	When ‘filter_id‘ specified, ‘step_id‘ precises from which step the filter comes from. Otherwise data from specified step is used to calculate required statistics.
`filter_id`	If not missing, filter related data statistics are returned.
`...`	Specific parameters passed to filter related method.
`state`	Should the stats be calculated on data before ("pre") or after ("post") filtering in specified step.

Value

List of filter-specific values summing up underlying filter data.

See Also

cohort-methods

step *Create filtering step*

Description

Steps all to perform multiple stages of Source data filtering.

Usage

```
step(...)
```

Arguments

...	Filters. See filter.

Value

List of class 'cb_step' storing filters configuration.

Examples

```
iris_step_1 <- step(
  filter('discrete', dataset = 'iris', variable = 'Species', value = 'setosa'),
  filter('discrete', dataset = 'iris', variable = 'Petal.Length', range = c(1.5, 2))
)
iris_step_2 <- step(
  filter('discrete', dataset = 'iris', variable = 'Sepal.Length', range = c(5, 10))
)

# Add step directly to Cohort
iris_source <- set_source(tblist(iris = iris))
coh <- iris_source |>
  cohort(
    iris_step_1,
    iris_step_2
  ) |>
  run()

nrow(get_data(coh, step_id = 1)$iris)
nrow(get_data(coh, step_id = 2)$iris)

# Add step to Cohort using add_step method
coh <- iris_source |>
  cohort()
coh <- coh |>
  add_step(iris_step_1) |>
  add_step(iris_step_2) |>
  run()
```

sum_up *Sum up Cohort state.*

Description

Sum up Cohort state.

Usage

```
sum_up(x, to_string = FALSE)
```

Arguments

x	Cohort object.
to_string	If 'TRUE', return the output as a character string instead of printing it. Defaults to 'FALSE'.

Value

When 'to_string = FALSE' (default), 'invisible(NULL)' (prints to console). When 'to_string = TRUE', a single character string.

See Also

cohort-methods

tblist *Create in memory tables connection*

Description

Create data connection as a list of loaded data frames. The object should be used as 'dtconn' argument of set_source.

Usage

```
tblist(..., names, .class = NULL)

as.tblist(x, ..., .class = NULL)
```

Arguments

...	additional arguments to be passed to or from methods.
names	A character vector describing provided tables names. If missing names are constructed based on provided tables objects.
.class	The extra (highest priority) class added to the resulting object. Having the extra class defined, enables to implement custom S3 methods for the object having higher priority over the existing methods. Especially useful if you want to change the built-in method behavior.
x	an R object.

Value

Object of class 'tblist' being a named list of data frames.

Examples

```
str(tblist(mtcars))
str(tblist(mtcars, iris))
str(tblist(MT = mtcars, IR = iris))
str(tblist(mtcars, iris, names = c("MT", "IR")))
```

---

tblist_class *S7 class wrapper for the 'tblist' source*

Description

S7 representation of the built-in 'tblist' source class, used as the source side of S7 dual dispatch when registering filter methods (e.g. 'S7::method(cb_filter_data, list(CbFilterDiscrete, tblist_class))'). See 'vignette("custom-filters")'.

Usage

```
tblist_class
```

Format

An S7 S3-class wrapper created with [S7::new_S3_class()].

update_filter *Update filter definition*

Description

Update filter definition

Usage

```
update_filter(x, step_id, filter_id, ...)

## S3 method for class 'Cohort'
update_filter(x, step_id, filter_id, ..., run_flow = FALSE)

## S3 method for class 'Source'
update_filter(x, step_id, filter_id, ...)
```

Arguments

x	An object in which the filter should be updated.
step_id	Id of the step where filter is defined.
filter_id	Id of the filter to be updated.
...	Filter parameters that should be updated.
run_flow	If 'TRUE', data flow is run after the filter is updated.

Value

Method dependent object (i.e. 'Cohort' or 'Source') having selected filter updated.

See Also

managing-cohort, managing-source

update_source *Update source in Cohort object.*

Description

Update source in Cohort object.

Usage

```
update_source(x, source, keep_steps = !has_steps(source), run_flow = FALSE)
```

Arguments

x	Cohort object.
source	Source object to be updated in Cohort.
keep_steps	If 'TRUE', steps definition remain unchanged when updating source. If 'FALSE' steps configuration is deleted. If vector of type integer, specified steps will remain.
run_flow	If 'TRUE', data flow is run after the source is updated.

Value

The 'Cohort' class object with updated 'Source' definition.

See Also

managing-cohort

%->% *Operator simplifying adding steps or filters to Cohort and Source objects*

Description

When called with filter or step object, runs add_filter and add_step respectively.

Usage

```
x %->% object
```

Arguments

x	Source or Cohort object. Otherwise works as a standard pipe operator.
object	Filter or step to be added to 'x'.

Value

An object ('Source' or 'Cohort') having new filter or step added.

Index

∗ **datasets**
librarian, 56
tblist_class, 75
.collect_data *(*source-layer*)*, 69
.gen_id, 4
.get_attrition_count *(*source-layer*)*, 69
.get_attrition_label *(*source-layer*)*, 69
.get_item, 4
.get_method, 5
.get_stats *(*source-layer*)*, 69
.if_value, 5
.init_step *(*source-layer*)*, 69
.post_binding *(*source-layer*)*, 69
.post_filtering *(*source-layer*)*, 69
.pre_filtering *(*source-layer*)*, 69
.print_filter, 6
.propagate_domains, 6, *46, 47*
.repro_code_tweak, *36*, *43*
.repro_code_tweak *(*source-layer*)*, 69
.run_binding *(*source-layer*)*, 69
%->%, 77

add_filter, 7, *57, 58*
add_hook *(*hooks*)*, 55
add_source, 8, *57*
add_step, 9, *57, 58*
as.tblist *(*tblist*)*, 74
attrition, 10, *48*, *59*
autofilter, 10, *65*

bind_key *(*binding-keys*)*, 11
bind_keys *(*binding-keys*)*, 11
binding-keys, 11, *44*, *49*, *65*, *67, 68*, *72*

cb_domain_from_data, 23
cb_domain_from_stats, 23
cb_filter_data, 24
cb_filter_to_expr, 24
cb_get_filter_data, 25
cb_get_filter_defaults, 25
cb_get_filter_stats, 26
cb_intersect_domain, 26
cb_intersect_domain_values, 27
cb_plot_filter_data, 27
cb_register_tool, *28*, 28
cb_register_tools *(*cb_register_tool*)*, 28
cb_tool, *28*, 28, *29–35*
cb_tool_add_filters, 29
cb_tool_apply_filters, 30
cb_tool_clear_filters, 30
cb_tool_describe_state, 31
cb_tool_filters_meta, 31
cb_tool_get_code, 32
cb_tool_get_data_summary, 32
cb_tool_remove_filters, 33
cb_tool_remove_step, 33
cb_tool_run, 34
cb_tool_set_filter_values, 34
cb_tool_toggle_filters, 35
CbFilter, 15, *60*
CbFilterDateRange, 16
CbFilterDatetimeRange, 17
CbFilterDiscrete, 18
CbFilterDiscreteText, 19
CbFilterMultiDiscrete, 20
CbFilterQuery, 21
CbFilterRange, 22
code, 35, *48*, *68*, *72*
Cohort, *28–35*, 36, *55*
cohort *(*create-cohort*)*, 48
cohort-methods, *10*, *36*, 48, *51*, *54, 55*, *59*, *61*, *72*, *74*
cohortBuilder-package, 4
create-cohort, 48

data_key, *12*, 49, *59*
describe, 50
description, *48*, 50, *65*, *68*

filter, *8*, *40*, 51, *60*, *69*, *73*

filter_domain, 52
filter_effective_value, 52
filter_variables, 53

get_data, *48*, 53
get_filter_params, 54
get_hook *(*hooks*)*, 55
get_state, *48*, 54, *61*

hooks, *8, 9*, *38–42*, *44*, *46, 47*, *49*, 55, *62, 63*

librarian, 56

managing-cohort, *8, 9*, 57, *62–64*, *76, 77*
managing-source, *8, 9*, 58, *62, 63*, *76*

plot_data, *48*, 58
primary_keys, *49*, 59, *65*, *67, 68*
print.cb_tool *(*cb_tool*)*, 28

register_filter_type, 60
restore, *48*, 61
rm_filter, *57, 58*, 61
rm_step, *57, 58*, 62
run, *57*, 63

set_source, *38*, *43*, *49*, *51*, *59*, 64, *74*
shape, *31*, 65
Source, 66
source-layer, 69
stat, *48*, *71*, 72
step, *9*, *39*, *68*, 73
sum_up, *48*, 74

tblist, 74
tblist_class, 75
tidy_source, *36*, *43*

update_filter, *57, 58*, 76
update_source, *57*, 76