Package 'yamlet'

January 22, 2020

Type Package

Title Versatile Curation of Table Metadata

Version 0.3.3

Author Tim Bergsma

Maintainer Tim Bergsma <bergsmat@gmail.com>

Description The 'yamlet' package implements a file-based mechanism for documenting datasets. It reads and writes YAML-formatted metadata and applies it as data item attributes. Data and metadata are stored independently but can be coordinated by using similar file paths with different extensions. The 'yamlet' dialect is valid 'YAML', but some conventions are chosen to improve readability. Defaults and conventions can be over-ridden by the user. See ?yamlet and ?decorate.data.frame. See ?read_yamlet ?write_yamlet, and ?io_csv.

License GPL-3

Encoding UTF-8

LazyData true

Imports yaml, csv (>= 0.5.4), encode, ggplot2, knitr, dplyr (>= 0.8.1), rlang

RoxygenNote 7.0.2

VignetteBuilder knitr

Suggests testthat (>= 2.1.0), magrittr, table1

NeedsCompilation no

Repository CRAN

Date/Publication 2020-01-22 06:50:03 UTC

R topics documented:

`as_lab.list` *Coerce List to Axis Label*

Description

Coerces list to axis label. Accepts a default label and returns that if nothing better can be done. If the attribute list has one named 'label', it is chosen as a substitute. But if that attribute is itself a list of values, an attempt is made to identify a single relevant value by treating the value names as conditions to evaluate on the supplied data. If a suitable value is found, it is chosen as a substitute. See `singularity` for search logic.

Usage

```
## S3 method for class 'list'
as_lab(
  x,
  default,
  collapse = "\n",
  enclose = getOption("yamlet_enclose", default = c("(", ")")),
  data,
  ...
)
```

Arguments

`x`	list, such as returned by `attributes`.
`default`	a value to return by default
`collapse`	character: separator for collapsing multi-line units

enclose	length-two character for enclosing unit
data	data.frame for resolving competing named values
...	ignored

Value

length-one character

See Also

Other lab: as_lab(), ggplot.decorated(), print.ag(), singularity()

Examples

```
meta <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
x <- decorate(meta)
as_lab(attributes(x$time), 'time', enclose = c('[',']'))
as_lab(attributes(x$time), 'time', enclose = c('[ ',' ]'))
```

---

as_yamlet *Coerce to Yamlet*

Description

Coerces something to yamlet format. If the object or user specifies default keys, these are applied. See as_yamlet.character.

Usage

```
as_yamlet(x, ...)
```

Arguments

x	object
...	passed arguments

Value

a named list

See Also

Other yamlet: as_yamlet.character(), as_yamlet.yam(), print.yamlet()

Examples

```
file <- system.file(package = 'yamlet', 'extdata','quinidine.yaml')
file
identical(as_yamlet(as_yam(file)), as_yamlet(file))

# Read yamlet from storage and apply default keys.
as_yamlet(file)
```

`conditionalize.data.frame`

Conditionalize Attributes of Data Frame

Description

Conditionalizes attributes of data.frame. Creates a conditional `attribute` definition for `column` by mapping `value` to `test`. Only considers records where both `test` and `value` are defined, and gives an error if there is not one-to-one mapping. Can be used with write methods as an alternative to hand-coding conditional metadata.

Usage

```
## S3 method for class 'data.frame'
conditionalize(x, column, attribute, test, value, ...)
```

Arguments

`x`	data.frame
`column`	unquoted name of column to conditionalize
`attribute`	unquoted name of attribute to create for column
`test`	unquoted name of column to test
`value`	unquoted name of column supplying attribute value
`...`	ignored arguments

Details

If the test column is character, individual elements should not contain both single and double quotes. For the conditional expressions, these values will be single-quoted by default, or double-quoted if they contain single quotes.

Value

class 'decorated' 'data.frame'

See Also

Other interface: decorate.character(), decorate.data.frame(), ggplot.decorated(), io_csv.character(), io_csv.data.frame(), io_table.character(), io_table.data.frame(), io_yamlet.character(), io_yamlet.data.frame(), read_yamlet(), resolve.data.frame(), write_yamlet()

Other conditionalize: conditionalize()

Examples

```
library(magrittr)
library(dplyr)
library(csv)
file <- system.file(package = 'yamlet', 'extdata','phenobarb.csv')
x <- as.csv(file)
head(x,3)

# suppose we have an event label stored as a column:

x %<>% mutate(evid = ifelse(
  event == 'dose',
  'dose of drug administered',
  'serum phenobarbital concentration'
 )
)

# We can define a conditional label for 'value'
# by mapping evid to event:

x %<>% conditionalize(value, label, event, evid)

x %>% as_yamlet
x %>% write_yamlet
```

---

decorate.character *Decorate Character*

Description

Treats x as a file path. By default, metadata is sought from a file with the same base but the 'yaml' extension.

Usage

```
## S3 method for class 'character'
decorate(
  x,
  meta = NULL,
  read = getOption("yamlet_import", as.csv),
```

```
  ext = getOption("yamlet_extension", ".yaml"),
  ...
)
```

Arguments

x	file path for table data
meta	file path for corresponding yamlet metadata, or a yamlet object
read	function or function name for reading x
ext	file extension for metadata file, if relevant
...	passed to read (if accepted) and to as_yamlet.character

Value

class 'decorated' 'data.frame'

See Also

Other decorate: decorate.data.frame(), decorate.list(), decorate(), decorations.data.frame(), decorations()

Other interface: conditionalize.data.frame(), decorate.data.frame(), ggplot.decorated(), io_csv.character(), io_csv.data.frame(), io_table.character(), io_table.data.frame(), io_yamlet.character(), io_yamlet.data.frame(), read_yamlet(), resolve.data.frame(), write_yamlet()

Examples

```
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
meta <- system.file(package = 'yamlet', 'extdata','quinidine.yaml')
identical(
  decorate(file),
  decorate(file, meta = meta)
)
identical(
  decorate(file, meta = as_yamlet(meta)),
  decorate(file, meta = meta)
)
a <- decorate(file)
b <- resolve(decorate(file))
c <- resolve(decorate(
  file,
  read = read.table,
  quote = "",
  as.is = FALSE,
  sep = ',',
  header = TRUE,
  na.strings = c('', '\\s', '.','NA'),
  strip.white = TRUE,
  check.names = FALSE
))
```

```
d <- decorate(
  file,
  read = read.table,
  quote = "",
  as.is = FALSE,
  sep = ',',
  header = TRUE,
  na.strings = c('', '\\s', '.','NA'),
  strip.white = TRUE,
  check.names = FALSE
)

# Importantly, b and c are identical with respect to factors
cbind(
  `as.is/!resolve`   = sapply(a, class), # no factors
  `as.is/resolve`    = sapply(b, class), # factors made during decoration
  `!as.is/resolve`   = sapply(c, class), # factors made twice!
  `!as.is/!resolve`  = sapply(d, class)  # factors made during read
)
str(a$Smoke)
str(b$Smoke)
str(c$Smoke)
str(d$Smoke)
levels(c$Creatinine)
levels(d$Creatinine) # level detail retained as 'guide'
```

`decorate.data.frame` *Decorate Data Frame*

Description

Decorates a data.frame. Expects metadata in yamlet format, and loads it onto columns as attributes.

Usage

```
## S3 method for class 'data.frame'
decorate(x, meta = NULL, ...)
```

Arguments

`x`	data.frame
`meta`	file path for corresponding yaml metadata, or a yamlet; an attempt will be made to guess the file path if x has a 'source' attribute
`...`	passed to `decorate.list`

Value

class 'decorated' 'data.frame'

See Also

decorate.list

Other interface: conditionalize.data.frame(), decorate.character(), ggplot.decorated(), io_csv.character(), io_csv.data.frame(), io_table.character(), io_table.data.frame(), io_yamlet.character(), io_yamlet.data.frame(), read_yamlet(), resolve.data.frame(), write_yamlet()

Other decorate: decorate.character(), decorate.list(), decorate(), decorations.data.frame(), decorations()

Examples

```
library(csv)
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
meta <- system.file(package = 'yamlet', 'extdata','quinidine.yaml')
a <- decorate(as.csv(file))
b <- decorate(as.csv(file), meta = as_yamlet(meta))
c <- decorate(as.csv(file), meta = meta)
d <- decorate(as.csv(file), meta = file)
e <- resolve(decorate(as.csv(file)))

# Most import methods are equivalent.
identical(a, b)
identical(a, c)
identical(a, d)
identical(a, e)
```

---

decorations.data.frame

Retrieve Decorations for Data Frame

Description

Retrieve the decorations of a data.frame; i.e., the metadata used to decorate it. Returns a list with same names as the data.frame. By default, `class` attributes are excluded from the result, as this is an attribute you likely don't want to manipulate independently. Consider carefully whether the default handling of factor levels (see `coerce` argument) is appropriate for your application.

Usage

```
## S3 method for class 'data.frame'
decorations(
  x,
  coerce = getOption("yamlet_coerce_decorations", FALSE),
  exclude_attr = getOption("yamlet_exclude_attr", "class"),
  ...
)
```

Arguments

x	data.frame
coerce	logical: whether to coerce factor levels to guide; alternatively, a key for the levels
exclude_attr	attributes to remove from the result
...	ignored

Value

named list

See Also

Other decorate: decorate.character(), decorate.data.frame(), decorate.list(), decorate(), decorations()

Examples

```
library(csv)
library(magrittr)
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
x <- decorate(as.csv(file))[,c('conc','Race')]
y <- decorate(as.csv(file))[,c('conc','Race')] %>% resolve
decorations(x)
decorations(y)
decorations(y, coerce = TRUE)
decorations(y, coerce = 'codelist')
decorations(y, exclude_attr = NULL)
```

---

encode.yamlet *Encode Yamlet*

Description

Encodes yamlet. Each 'guide' element with length > 1 is converted to an encoding, if possible. If data is supplied, conditional guides will be ignored.

Usage

```
## S3 method for class 'yamlet'
encode(x, target = "guide", data = NULL, ...)
```

Arguments

x	yamlet
target	attribute to encode
data	optional data.frame for guide context
...	ignored

Value

yamlet, with guide elements possibly transformed to encodings

See Also

Other encode: `list2encoding()`

Examples

```
meta <- system.file(package = 'yamlet', 'extdata','quinidine.yaml')
meta <- as_yamlet(meta)
meta <- encode(meta)
```

---

explicit_guide *Coerce Guide to Something More Explicit*

Description

Coerces 'guide' to something more explicit. Generic, with methods for data.frame and yamlet. The key 'guide' generally suggests a guide to interpretation of a data item, such as units, formats, codelists, and encodings. The idea here is to replace 'guide' with something explicit in case required downstream.

Usage

```
explicit_guide(x, ...)
```

Arguments

x	object of dispatch
...	passed arguments

Value

see methods

See Also

Other explicit_guide: `explicit_guide.data.frame()`, `explicit_guide.yamlet()`

factorize_codelist *Coerce Codelist to Factor*

Description

Coerces codelist to factor. Generic, with default and data.frame methods.

Usage

```
factorize_codelist(x, ...)
```

Arguments

x	object
...	passed arguments

Value

see methods

See Also

Other factorize_codelist: `factorize_codelist.character()`, `factorize_codelist.data.frame()`, `factorize_codelist.default()`, `factorize_codelist.factor()`

ggplot.decorated *Create a New ggplot for a Decorated Data Frame*

Description

Creates a new ggplot object for a decorated data.frame. This is the ggplot() method for class 'decorated'; it tries to implement automatic labels and units in axes and legends in association with `print.ag`. Use `ggplot(as.data.frame(x))` to get default ggplot() behavior. Use `ggplot(as_decorated(x))` to enforce custom behavior.

Usage

```
## S3 method for class 'decorated'
ggplot(data, ...)
```

Arguments

data	data.frame or similar
...	passed to `ggplot`

Value

return value like ggplot

See Also

Other lab: as_lab.list(), as_lab(), print.ag(), singularity()

Other interface: conditionalize.data.frame(), decorate.character(), decorate.data.frame(), io_csv.character(), io_csv.data.frame(), io_table.character(), io_table.data.frame(), io_yamlet.character(), io_yamlet.data.frame(), read_yamlet(), resolve.data.frame(), write_yamlet()

Examples

```
meta <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
x <- decorate(meta)
library(ggplot2)
class(ggplot(data = x) + geom_path(aes(x = time, y = conc)))
class(ggplot(data = x, aes(x = time, y = conc)) + geom_path())
example(print.ag)
```

---

io_csv *Import and Export Documented Tables as CSV*

Description

Imports or exports documented tables as comma-separated variable. Generic, with methods that extend as.csv.

Usage

```
io_csv(x, ...)
```

Arguments

x	object
...	passed arguments

Value

See methods.

See Also

Other io: io_csv.character(), io_csv.data.frame(), io_table.character(), io_table.data.frame(), io_table(), io_yamlet.character(), io_yamlet.data.frame(), io_yamlet.yamlet(), io_yamlet()

Examples

```
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
x <- decorate(file)
out <- file.path(tempdir(), 'out.csv')
foo <- io_csv(x, out)
identical(out, foo)
y <- io_csv(foo)
attr(x, 'source') <- NULL
attr(y, 'source') <- NULL
identical(x, y) # lossless 'round-trip'
```

---

io_table *Import and Export Documented Tables*

Description

Imports or exports documented tables. Generic, with methods that extend `read.table` and `write.table`.

Usage

```
io_table(x, ...)
```

Arguments

x	object
...	passed arguments

Value

See methods.

See Also

Other io: `io_csv.character()`, `io_csv.data.frame()`, `io_csv()`, `io_table.character()`, `io_table.data.frame()`, `io_yamlet.character()`, `io_yamlet.data.frame()`, `io_yamlet.yamlet()`, `io_yamlet()`

Examples

```
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
x <- decorate(file)
out <- file.path(tempdir(), 'out.tab')
foo <- io_table(x, out)
identical(out, foo)
y <- io_table(foo, as.is = TRUE)
attr(x, 'source') <- NULL
rownames(x) <- NULL
rownames(y) <- NULL
identical(x, y) # lossless 'round-trip'
```

io_yamlet *Import and Export Yamlet*

Description

Imports and exports yamlet. Generic, with a read method `read_yaml` for character and a write method `write_yaml` for data.frame.

Usage

```
io_yamlet(x, ...)
```

Arguments

x	object
...	passed arguments

Value

see methods

See Also

Other io: `io_csv.character()`, `io_csv.data.frame()`, `io_csv()`, `io_table.character()`, `io_table.data.frame()`, `io_table()`, `io_yamlet.character()`, `io_yamlet.data.frame()`, `io_yamlet.yamlet()`

Examples

```
file <- system.file(package = 'yamlet', 'extdata','quinidine.yaml')
x <- io_yamlet(file)
tmp <- tempdir()
out <- file.path(tmp, 'tmp.yaml')

# we can losslessly 'round-trip' x using to generic calls
identical(x, io_yamlet(io_yamlet(x, out)))
```

print.ag *Print Automatic Labels and Units for ggplot*

Description

Prints automatic labels and units for ggplot. Reworks the labels as a function of attributes in corresponding data. Default for `labeller` (`as_lab`) will receive existing labels one at a time and corresponding attributes (if any) from data.

Usage

```
## S3 method for class 'ag'
print(x, labeller = getOption("yamlet_labeller", default = as_lab), ...)
```

Arguments

x	class 'ag' from ggplot.decorated
labeller	a function (or its name) like as_lab to generate axis labels
...	passed arguments

Value

used for side effects

See Also

Other lab: as_lab.list(), as_lab(), ggplot.decorated(), singularity()

Examples

```
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
library(ggplot2)
library(dplyr)
library(magrittr)
# par(ask = FALSE)
options(enclose = c('[ ',' ]'))

# Filter() strips 'label' from factors (see legend), but not vectors:

file %>% decorate %>% resolve %>% filter(!is.na(conc)) %>%
ggplot(aes(x = time, y = conc, color = Heart)) + geom_point()

# No factors created here, but print.ag promotes guide to factor if it can:

file %>% decorate %>% filter(!is.na(conc)) %>%
ggplot(aes(x = time, y = conc, color = Heart)) + geom_point()

# Here we try a dataset with conditional labels and units.

file <- system.file(package = 'yamlet', 'extdata','phenobarb.csv')

# Note that there are two elements each for value label and value guide.
#'
file %>% decorate %>% as_yamlet
# Guide might have been mistaken for an attempt to provide codes/decodes
# for a factor.  However, the keys evaluate to logical on the data.frame.
# Seeing that, we test for one of them being all true, and if so we select it.

file %>% decorate %>% ggplot(aes(x = time, y = value, color = event)) + geom_point()

# In the above example, we are plotting doses and concentrations, which have
```

```
# different labels and units, so we can't improve on the y axis label.
# But if we subset to just one of these, then only one of the named conditions
# will be always true (and will therefore be promoted).

file %>% decorate %>%
filter(event == 'conc') %>%
ggplot(aes(x = time, y = value, color = ApgarInd)) + geom_point()

file %>% decorate %>%
filter(event == 'dose') %>%
ggplot(aes(x = time, y = value, color = Wt)) +
geom_point() +
scale_y_log10() +
scale_color_gradientn(colours = rainbow(4))
```

`read_yamlet` *Read Yamlet*

Description

Reads yamlet from file. Similar to `io_yamlet.character` but also reads text fragments.

Usage

```
read_yamlet(
  x,
  default_keys = getOption("yamlet_default_keys", list("label", "guide")),
  ...
)
```

Arguments

`x`	file path for yamlet, or vector of yamlet in storage syntax
`default_keys`	character: default keys for the first n anonymous members of each element
`...`	passed to `as_yamlet`

Value

yamlet: a named list with default keys applied

See Also

`decorate.data.frame`

Other interface: `conditionalize.data.frame()`, `decorate.character()`, `decorate.data.frame()`, `ggplot.decorated()`, `io_csv.character()`, `io_csv.data.frame()`, `io_table.character()`, `io_table.data.frame()`, `io_yamlet.character()`, `io_yamlet.data.frame()`, `resolve.data.frame()`, `write_yamlet()`

Examples

```
library(csv)
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
meta <- system.file(package = 'yamlet', 'extdata','quinidine.yaml')
x <- as.csv(file)
y <- read_yamlet(meta)
x <- decorate(x, meta = y)
identical(x, decorate(file))
```

---

resolve.data.frame *Resolve Guide for Data Frame*

Description

Resolves implicit usage of default key 'guide' to explicit usage for data.frame. Simply calls `explicit_guide` followed by `factorize_codelist`.

Usage

```
## S3 method for class 'data.frame'
resolve(x, ...)
```

Arguments

x	object
...	passed arguments

Value

data.frame

See Also

Other resolve: `resolve()`

Other interface: `conditionalize.data.frame()`, `decorate.character()`, `decorate.data.frame()`, `ggplot.decorated()`, `io_csv.character()`, `io_csv.data.frame()`, `io_table.character()`, `io_table.data.frame()`, `io_yamlet.character()`, `io_yamlet.data.frame()`, `read_yamlet()`, `write_yamlet()`

Examples

```
library(magrittr)
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
x <- decorate(file)
x %>% resolve(default = 'unit') %>% as_yamlet
```

to_yamlet *Coerce to Yamlet Storage Format*

Description

Coerces to yamlet storage format. Generic, with methods for default, null, character and list which together implement the yamlet storage syntax. Always returns length-one character, possibly the empty string.

Usage

```
to_yamlet(x, ...)
```

Arguments

x	object
...	ignored

Value

length-one character

See Also

Other to_yamlet: to_yamlet.NULL(), to_yamlet.character(), to_yamlet.default(), to_yamlet.list()

write_yamlet *Write Yamlet*

Description

Writes yamlet to file. Similar to io_yamlet.data.frame but returns invisible storage format instead of invisible storage location.

Usage

```
write_yamlet(
  x,
  con = stdout(),
  eol = "\n",
  useBytes = FALSE,
  default_keys = getOption("yamlet_default_keys", list("label", "guide")),
  fileEncoding = getOption("encoding"),
  ...
)
```

Arguments

x	something that can be coerced to class 'yamlet', like a yamlet object or a decorated data.frame
con	passed to writeLines
eol	end-of-line; passed to writeLines as sep
useBytes	passed to writeLines
default_keys	character: default keys for the first n anonymous members of each element
fileEncoding	if con is character, passed to file as encoding
...	passed to as_yamlet

Value

invisible character representation of yamlet (storage syntax)

See Also

decorate.list

Other interface: conditionalize.data.frame(), decorate.character(), decorate.data.frame(), ggplot.decorated(), io_csv.character(), io_csv.data.frame(), io_table.character(), io_table.data.frame(), io_yamlet.character(), io_yamlet.data.frame(), read_yamlet(), resolve.data.frame()

Examples

```
library(csv)
file <- system.file(package = 'yamlet', 'extdata','quinidine.csv')
meta <- system.file(package = 'yamlet', 'extdata','quinidine.yaml')
x <- as.csv(file)
y <- read_yamlet(meta)
x <- decorate(x, meta = y)
identical(x, decorate(file))
tmp <- tempfile()
write_yamlet(x, tmp)
identical(read_yamlet(meta), read_yamlet(tmp))
```

---

yamlet *yamlet: Versatile Curation of Table Metadata*

Description

The **yamlet** package supports storage and retrieval of table metadata in yaml format. The most important function is decorate.character: it lets you 'decorate' your data by attaching attributes retrieved from a file in yaml format. Typically your data will be of class 'data.frame', but it could be anything that is essentially a named list.

Storage Format

Storage format for 'yamlet' is a text file containing well-formed yaml. Technically, it is a map of sequences. Though well formed, it need not be complete, and therefore has utility over a longer life cycle of data development.

In the simplest case, the data specification consists of a list of column (item) names, followed by semicolons. Perhaps you only have one column:

```
mpg:
```

or maybe several:

```
mpg:
cyl:
disp:
```

If you know descriptive labels for your columns, provide them (skip a space after the colon).

```
mpg: fuel economy
cyl: number of cylinders
disp: displacement
```

If you know units, create a sequence with square brackets.

```
mpg: [ fuel economy, miles/gallon ]
cyl: number of cylinders
disp: [ displacement , in^3 ]
```

If you are going to give units, you probably should give a key first, since the first anonymous element is 'label' by default, and the second is 'guide'. (A guide can be units for numeric variables, factor levels/labels for categorical variables, or a format string for dates, times, and datetimes.) You could give just the units but you would have to be specific:

```
mpg: [unit: miles/gallon]
```

You can over-ride default keys by providing them in your data:

```
mpg: [unit: miles/gallon]
_keys: [label, unit]
```

Notice that stored yamlet can be informationally defective while syntactically correct. If you don't know an item key at the time of data authoring, you can omit it:

```
race: [race,[white: 0,black: 1,2,asian: 3 ]]
```

Or perhaps you know the key but not the value:

```
race: [race,[white: 0,black: 1,asian: 2,? other ]]
```

Notice that `race` is factor-like; the factor sequence is nested within the attribute sequence. Equivalently:

```
race: [label: race,guide: [white: 0,black: 1,asian: 2,? other ]]
```

To get started using yamlet, see `?as_yamlet.character` and examples there. See also `?decorate` which adds yamlet values to corresponding items in your data. See also `?print.ag` which uses labels and guides to autogenerate axis labels.

Note: the quinidine and phenobarb datasets in the examples are borrowed from **nlme** (`?Quinidine`, `?Phenobarb`), with some reorganization.

Index