Package ‘heims’

January 25, 2018

Title Decode and Validate HEIMS Data from Department of Education, Australia
Version 0.4.0
Date 2018-01-25

Description Decode elements of the Australian Higher Education Information Management System (HEIMS) data for clarity and performance. HEIMS is the record system of the Department of Education, Australia to record enrolments and completions in Australia's higher education system, as well as a range of relevant information. For more information, including the source of the data dictionary, see <http://heimshelp.education.gov.au/sites/heimshelp/dictionary/pages/data-element-dictionary>.

Depends R (>= 3.4.0), data.table
Imports hutils, magrittr, fastmatch, bit64, lubridate
License GPL-3
Encoding UTF-8
LazyData true
Suggests testthat, fst
RoxygenNote 6.0.1
NeedsCompilation no
Author Hugh Parsonage [aut, cre]
Maintainer Hugh Parsonage <hugh.parsonage@gmail.com>
Repository CRAN
Date/Publication 2018-01-25 10:05:11 UTC

R topics documented:

  browse_elements ....................................................... 2
  decoders .............................................................. 2
  decode_heims .......................................................... 4
  dummy_enrol ........................................................... 5
  element_decoders ...................................................... 5
browse_elements  

Browse elements for description

Description  

Browse elements for description

Usage  

browse_elements(pattern)

Arguments

pattern  

A case-insensitive perl expression or expressions to match in the long name of heims_data_dict.

Value  

A data.table of all element-long name combinations matching the perl regular expression.

Examples  

browse_elements(c("ProViDer", "Maj"))

decoders  

Decoders

Description  

Decoders
Usage

- E089_decoder
- E095_decoder
- E306_decoder
- E310_decoder
- E312_decoder
- E316_decoder
- E329_decoder
- E327_decoder
- E330_decoder
- E331_decoder
- E337_decoder
- E346_decoder
- E348_decoder
- E355_decoder
- E358_decoder
- E386_decoder
- E392_decoder
- E461_decoder
- E463_decoder
- E464_decoder
- E490_decoder
- U490_decoder
- E551_decoder
- E562_decoder
Format

An object of class `data.table` (inherits from `data.frame`) with 2 rows and 2 columns.

<table>
<thead>
<tr>
<th>decode_heims</th>
<th>Decode HEIMS elements</th>
</tr>
</thead>
</table>

Description

Decode HEIMS elements

Usage

decode_heims(DT, show_progress = FALSE, check_valid = TRUE, selector)

Arguments

- **DT**: A `data.table` with the original HEIMS column names.
- **show_progress**: Display the progress of the function (which is likely to be slow on real data).
- **check_valid**: Check the variable is valid before decoding. Setting to `FALSE` is faster, but should only be done when you know the data has been validated.
- **selector**: Original HEIMS names to restrict the decoding to. Other names will be preserved.

Details

Each variable in DT is validated according `heims_data_dict` before being decoded. Any failure stops the validation.

If DT has a key, the output will have a key, but set on the **decoded** columns and the ordering will most likely change (to reflect the decoded values).

This function will, on the full HEIMS data, take a long time to finish. Typically in the order of 10 minutes for the enrol file.

Value

DT with the values decoded and the names renamed.
### Examples

```r
## Not run:
## (E488 is made up so won't work if validation is attempted.)
decode_heims(dummy_enrol)

## End(Not run)
decode_heims(dummy_enrol, show_progress = TRUE, check_valid = FALSE)
```

---

**dummy_enrol**

*Dummy enrolment file*

---

**Description**

A `data.table` of five fictitious enrolments.

**Usage**

```r
dummy_enrol
```

**Format**

An object of class `data.table` (inherits from `data.frame`) with 5 rows and 56 columns.

---

**element_decoders**

*Make HEIMS element nos human-readable*

---

**Description**

Make HEIMS element nos human-readable

**Usage**

```r
rename_heims(DT)
element2name(v)
```

**Arguments**

- **DT**
  - The data table with original names
- **v**
  - A vector of element names.

**Details**

See `heims_data_dict`. Note that `decode_heims` is generally better, as it decodes the variable if a decoder is present in the dictionary.

`element2name` is the inverse of `browse_elements`: given an element like `E306`, it returns the name (HE/provider_cd).
element_validation

Validate HEIMS elements

Value

DT with the new names or the vector with the names translated.

Description

Return TRUE or FALSE on whether or not each variable in a data.table complies with the HEIMS code limits

Usage

validate_elements(DT, .progress_cat = FALSE)
prop_elements_valid(DT, char = FALSE)
count_elements_invalid(DT, char = FALSE)

Arguments

DT The data.table whose variables are to be validated.
.progress_cat Should the progress of the function be displayed on the console? If TRUE the name of the element about to be validated is shown.
char Return as character vector, in particular marking – any complete or completely absent values.

Details

For early detection of invalid results, the type of the variable (in particular integer vs double) is considered first, vetoing a TRUE result if different.

Value

A named logical vector, whether or not the variable complies with the style requirements. A value of NA indicates the variable was not checked (perhaps because it is absent from heims_data_dict).

Examples

X <- data.frame(E306 = c(0, 1011, 999, 9998))
validate_elements(X) # FALSE
prop_elements_valid(X)
X <- data.frame(E306 = as.integer(c(0, 1011, 999, 9998)))
validate_elements(X) # TRUE
first_levels

Description

See `relevel_heims`.

Usage

first_levels

Format

An object of class `data.table` (inherits from `data.frame`) with 8 rows and 2 columns.

fread_heims

Description

Read raw HEIMS file

Usage

fread_heims(filename)

Arguments

filename A text-delimited file, passed to `fread` from `data.table`.

Details

The strings "" ""NA" ""?" "" ""." ""*" "**" are treated as missing, as well as ZZZZZZZZZZ (so students without a CHESSN will be marked with the integer64 missing value).

Value

A data.table with column names in ascending (lexicographical) order and any columns starting with e will be uppercase.
HEIMS data dictionary

Description

HEIMS data dictionary

Usage

heims_data_dict

Format

A named list each containing 5 elements:

long_name  a human-readable version of the variable; orig_name the element number;
mark_missing a vectorized-function returning TRUE on values of the variable which should be coded as NA;
ad_hoc_prepare  a function to apply before validation;
validate a single-value function returning TRUE or FALSE on vectors which comply with the variable’s coding rules.
ad_hoc_validation_note  If the data dictionary did not cover elements in the file, how the validate function was altered to suffer them.
valid a vectorized function returning TRUE or FALSE on vectors which do not comply with the variable’s coding rules.
decoder  A function of the data.table decoding the variable decoded.
post_fst  A function of the data.table returned by fst to be used (for example to reset attributes).

Details

Abbreviations in long_name:

amt  Amount
cd  Code
det  Detail(s)
F0E  Field of education
Maj  Major

Source

**read_heims_fst**

Read HEIMS data from decoded fst files

**Description**

Read HEIMS data from decoded fst files

**Usage**

read_heims_fst(filename)

**Arguments**

filename File path to .fst file of a decoded HEIMS file (decode_heims) produced by fst::write.fst.

**Value**

A data.table with appropriate attributes.

**relevel_heims**

Relevel categorical variables

**Description**

Changes categorical variables in a data.table to levels with a sensible reference level

**Usage**

relevel_heims(DT)

**Arguments**

DT A data.table post decode_heims.

**Value**

The same data.table with character vectors changed to factors whose first level is the level intended.
| utilities | Utility functions |

**Description**

Only included here because of the unusual nature of `heims_data_dict`.

**Usage**

- `AND()`
- `OR()`
- `never(v)`
- `every(v)`
- `always(v)`
- `is.Date(v)`
- `is.YearMonth(v)`
- `nth_digit_of(x, n)`
- `between(...)`
- `or(...)`
- `and(...)`
- `if_else(...)`
- `coalesce(...)`
- `as.integer64(v)`
- `is.integer64(v)`
- `force_integer(v)`
- `ymd(...)`
Arguments

- v: A vector.
- x, n: vectors
- ...: Passed to other functions
- a: Element suspected to be in tbl
- tbl: A lookup table.

Details

nth_digit_of returns the nth digit of the number starting from the units and going up in magnitude.

Examples

nth_digit_of(503, 1) == 1
Index

*Topic datasets
  decoders, 2
dummy_enrol, 5
first_levels, 7
heims_data_dict, 8
%fin% (utilities), 10
always (utilities), 10
AND (utilities), 10
and (utilities), 10
as.integer64 (utilities), 10
between (utilities), 10
browse_elements, 2, 5
coalesce (utilities), 10
count_elements_invalid
  (element_validation), 6
decode_heims, 4, 5, 9
decoders, 2
dummy_enrol, 5
E089_decoder (decoders), 2
E095_decoder (decoders), 2
E306_decoder (decoders), 2
E310_decoder (decoders), 2
E312_decoder (decoders), 2
E316_decoder (decoders), 2
E327_decoder (decoders), 2
E329_decoder (decoders), 2
E330_decoder (decoders), 2
E331_decoder (decoders), 2
E337_decoder (decoders), 2
E346_decoder (decoders), 2
E348_decoder (decoders), 2
E355_decoder (decoders), 2
E358_decoder (decoders), 2
E386_decoder (decoders), 2
E392_decoder (decoders), 2
E461_decoder (decoders), 2
E463_decoder (decoders), 2
E464_decoder (decoders), 2
E490_decoder (decoders), 2
E551_decoder (decoders), 2
E562_decoder (decoders), 2
E919_decoder (decoders), 2
E920_decoder (decoders), 2
E922_decoder (decoders), 2
element2name (element_decoders), 5
element_decoders, 5
element_validation, 6
every (utilities), 10
first_levels, 7
FOE_uniter (decoders), 2
force_integer (utilities), 10
fread_heims, 7
HE_Provider_decoder (decoders), 2
heims_data_dict, 2, 4, 5, 8, 10
if_else (utilities), 10
is.Date (utilities), 10
is.integer64 (utilities), 10
is.YearMonth (utilities), 10
never (utilities), 10
nth_digit_of (utilities), 10
OR (utilities), 10
or (utilities), 10
prop_elements_valid
  (element_validation), 6
read_heims_fst, 9
relevel_heims, 7, 9
rename_heims (element_decoders), 5
rm_leading_0s (utilities), 10
U490_decoder (decoders), 2
INDEX

utilities, 10

validate_elements(element_validation), 6

ymd(utilities), 10