Package 'hsstools'

June 20, 2022

Title Simplify the HSS Data Process

Version 0.0.0.9000

Description A collection of functions to simplify and streamline the process of cleaning and analysing HSS data.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.0

Imports dplyr,
flextable,
forcats,
lubridate,
magrittr,
officer,
purrr,
questionr,
stringr,
tidyr

Suggests knitr,
pkgdown,
rmarkdown,
testthat (>= 3.0.0)

VignetteBuilder knitr

Depends R (>= 2.10)

LazyData true

Config/testthat/edition 3

R topics documented:

`append_tables` *Append all data files stored in the same folder + file name*

Description

Append all data files stored in the same folder + file name

Usage

```
append_tables(
  source_folder,
  destination_folder,
  destination_file,
  file_pattern,
  file_separator
)
```

Arguments

`source_folder`	The folder in which the to-be appended data files are stored
`destination_folder`	The folder in which the appended data file should be stored
`destination_file`	The name of the appended data file
`file_pattern`	Set to ".csv" when appending CSV files
`file_separator`	Set to "," when appending CSV files

Value

A CSV file containing the contents of all source files, incl. a new column containing the file names of the source files. NOTE: If the source files contain column headers, the appended file will contain the header row of each source file

hss_chisq *Runs chi-squared test on selected variables*

Description

This functions runs chi-squared significance tests for one or more variables in a dataframe with a single cross-variable. Variable names can be passed as a single object or as a character string. hss_chisq_formatted returns a formatted character string containing a verbose explanation of the p-value.

Usage

```
hss_chisq(df, var, group, full = FALSE, multi = FALSE)

hss_chisq_formatted(df, var, group)
```

Arguments

df	The dataframe containing the variable(s) of interest
var	The variable(s) of interest. Accepts a single value or character string.
group	The grouping (or disaggregation) variable.
full	should the full results be returned. If set to FALSE, only p.value is returned
multi	Set to FALSE if used for a 'select-one' question. Set to TRUE if used for a 'select-multiple question to look up the corresponding response options.

Value

A vector containing the results of the chi-squared test for the selected variables. If full is set to TRUE this will be a list, otherwise an atomic numeric vector. For hss_chisq_formatted the output is a character vector of length 1.

Examples

```
# Create dummy dictionary
dict_var <- dummy_var
dict_val <- dummy_val
# Calculte p-value for chi-squared test on a 'select-one' question.
hss_chisq(dummydata, "migr_nr", "gender")

# Calculate p-value for chi-squared test on a 'select-multiple' question.
hss_chisq(dummydata, "migr_why_all", "gender", multi = TRUE)

# Chi-squared test with full output
hss_chisq(dummydata, "migr_nr", full = TRUE)

# Formatted output
hss_chisq_formatted(dummydata, "migr_nr", "gender")
```

hss_combine_single *Combine multiple questions with the same response options*

Description

Combine multiple questions with the same response options

Usage

```
hss_combine_single(df, var)

hss_combine_multi(df, var)
```

Arguments

df	The dataframe containing the questions
var	The shared element in all of the question/variable names

Value

This returns a dataframe with the response options and the number of responses for that option.

hss_create_dict *Create a dictionary of HSS variable or value labels*

Description

Create a dictionary from an XLS form. The dictionary is stored as a dataframe. This function needs to be run separately for variable- and value dictionaries.

Usage

```
hss_create_dict(form, type = "var")
```

Arguments

form	Path to the XLS form.
type	Use "var" to create a dictionary of variable names and labels. Use "val" to create a dictionary of value names and labels.

Value

A dataframe containing variable or value names and their associated text labels.

hss_create_question_list

Create a list of questions from XLS form

Description

This function reads the XLS form and provides a list of all questions in the form as well as the type of question.

Usage

```
hss_create_question_list(dict_path)
```

Arguments

dict_path	Path to the XLS form

Value

A named character vector with question type as name and the question/variable name as value.

hss_dummydata

Generates a dataframe for testing purposes.

Description

Generates a dataframe for testing purposes.

Usage

```
hss_dummydata(rows = 500, seed = 1234)
```

Arguments

rows	An integer to set the number of rows to generate. Default is 500.
seed	Set the seed for random number generation. Default is 1234.

Value

A dataframe

Examples

```
df <- testdata(100)
```

hss_export_formatted *Exports a formatted list of tables to .docx*

Description

Exports a formatted list of tables to .docx

Usage

```
hss_export_formatted(list, file = NULL, type = "word")
```

Arguments

list	A list containing flextable objects
file	The file path to save to

Value

Outputs the selected flextable objects to a .docx file. Depending on the number of tables to export, this may take a long time. NOTE: running this function while the chosen output file is opened in another program will cause the R session to abort.

hss_export_tables *Exports a list of data tables to CSV*

Description

Exports a list of HSS tables to CSV. This function uses sink() to write multiple tables to the same file. If an error occurs during execution of the function you may need to close the sink with sink() before proceeding.

Usage

```
hss_export_tables(df_list, path)
```

Arguments

df_list	A named list containing the data tables. The output of hss_write_tables() usually.
path	Where to store the .csv output.

Value

A CSV file stored at the provided location.

See Also

hss_write_tables()

hss_format_single *Format tables using HSS style preferences*

Description

Converts table to a flextable and applies formatting, according to set preferences. There are two separate functions for single-response and multiple-response question tables. Neither one needs any further arguments.

Usage

```
hss_format_single(table)

hss_format_multi(table)

hss_format_group()
```

Value

A flextable object with formatting applied.

hss_label *Apply text labels to tables.*

Description

Apply text labels to the specified table. Labels are taken from the dictionaries created from the XLS forms. This converts the table to a flextable object. Works for EN or AR text labels. If used with AR text labels some optional formatting is applied to correctly display text.

Usage

```
hss_label(table, var, grouping, lang = "en")
```

Arguments

table	the table for which labels should be applied
var	the variable name. Used to determine the question label and to look up the appropriate response labels.
grouping	the grouping variable. Will be used to apply appropriate column headers
language	To determine the language of labels to be applied.

Value

A flextable object with the original table values and appropriate question & response labels.

hss_mergetranslated *Merges HSS dataframe with dataframe containing translated columns.*

Description

Merge the translated columns with the original dataframe. All translated columns are appended after their associated original column. Currently this only works for Arabic -> English translation.

Usage

```
hss_mergetranslated(df, df_translated)
```

Arguments

df	The original HSS dataframe
df_translated	The dataframe containing "_ar" and translated "_en" columns

Value

A merged dataframe containing HSS data and all translated columns.

hss_overview_multi *Create overview table for 'select-multiple' questions*

Description

This creates overview tables for multiple questions that contain the same response options. For example when the same question is repeated for different events. This implementation is specifically for 'select-multiple' questions. For 'select-one' questions see hss_overview_single

Usage

```
hss_overview_multi(df, vars, percent = TRUE)
```

Arguments

df	The dataframe containing the questions.
vars	A commom character string that is shared between all relevant variables.
percent	Set to TRUE to display percentages, set to FALSE to display counts. Default is TRUE

Value

A dataframe with responses for the selected variables.

hss_overview_single *Create overview table for a group of similar 'select-one' questions*

Description

This creates overview tables for multiple questions that contain the same response options. For example when the same question is repeated for different events. This implementation is specifically for 'select-one' questions. For 'select-multiple' questions see hss_overview_multi

Usage

```
hss_overview_single(df, vars, percent = TRUE)
```

Arguments

df	The dataframe containing the questions
vars	A commom character string that is shared between all relevant variables.
percent	Set to TRUE to display percentages, set to FALSE to display counts. Default is TRUE

Value

A dataframe with responses for the selected variables.

hss_surveyduration *Calculate survey duration*

Description

Loads raw HSS datafile and performs some basic cleaning: proper encoding of Arabic, dates formatted as date-time objects, calculate survey duration.

Usage

```
hss_surveyduration(path, skip = 0)
```

Arguments

path	Path to the HSS data file. Expects a .csv file
skip	Number of rows to skip. Default is 0. Use this if you know how many rows contain test answers.

Value

A dataframe with an added SurveyDuration column

hss_table_multi *Generate contingency table for multiresponse questions*

Description

Generate contingency table for multiresponse questions

Usage

```
hss_table_multi(df, var, group, percent = TRUE, digits = 1)
```

Arguments

df	The dataframe containing the multiresponse questions
group	A grouping (or disaggregation) variable.
percent	Set to TRUE to show percentages. Set to FALSE to show counts
digits	The number of (significant) digits to display. Trailing zeroes are always removed. Note that 'digits' does not mean 'decimals', so digits = 3 will display as 'mm.d' not 'mm.ddd'
resp	A character string of all response variables to include

Value

A contingency table containing the multiresponse answers and a grouping variable

hss_table_single *HSS Data Table Generation*

Description

HSS Data Table Generation

Usage

```
hss_table_single(df, var, group, percent = TRUE, digits = 1)
```

Arguments

df	A dataframe containing the variable of interest and grouping variable.
var	A character string with the variable name of interest.
group	A character string with the grouping (or disaggregation) variable.
percent	Set to TRUE to show percentages. Set to FALSE to show counts.
digits	The number of (significant) digits to display. Trailing zeroes are always removed. Note that 'digits' does not mean 'decimals', so digits = 3 will display as 'mm.d' not 'mm.ddd'

Value

A contingency table with the variable of interest and grouping variable.

hss_translate *Translate HSS content from Arabic to English*

Description

Translate HSS content from Arabic to English

Usage

```
hss_translate(df, apikey)
```

Arguments

df	the dataframe containing Arabic text columns. These columns are expected to end in "_ar"

Value

A dataframe containing Arabic text columns and their English translations.

hss_write_formatted *Write formatted tables to a list for selected variables*

Description

Write formatted tables to a list for selected variables

Usage

```
hss_write_formatted(
  df,
  questions,
  group,
  percent = TRUE,
  digits = "1",
  lang = "en"
)
```

Arguments

df	The dataframe containing relevant variables
questions	A named character vector containing variable names and the table type required.
group	The desired grouping/disaggregation variable.
percent	Logical vector if the table should be created with percentage values. If set to FALSE, counts are shown.

Value

A named list of flextable objects

`hss_write_tables` *Write tables to a list for selected variables*

Description

Write tables to a list for selected variables

Usage

```
hss_write_tables(df, questions, group, percent = TRUE)
```

Arguments

`df`	The dataframe containing relevant variables
`questions`	A named character vector containing variable names and the table type required.
`group`	The desired grouping/disaggregation variable.
`percent`	Logical vector if the table should be created with percentage values. If set to FALSE, counts are shown.

Value

A named list of dataframes.

Index