

Package: formstools (via r-universe)

October 10, 2026

Type Package

Title Tools for working with ODK XLSForms

Version 0.0.0.9000

Description Set of utility functions for use by EcoHealth Alliance researchers in working with Open Data Kit XLSForms. These functions are aimed at aiding users in the data cleaning and data validation process using information found in the ODK XLSForms.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 2.10)

Imports dplyr, readxl, rlang, stringr, tidyr

Suggests tibble, ruODK

Remotes ropensci/ruODK

LazyData true

RoxygenNote 7.2.3

Config/pak/sysreqs libicu-dev

Repository <https://humanitarian-user-group.r-universe.dev>

Date/Publication 2023-06-02 16:01:30 UTC

RemoteUrl <https://github.com/ecohealthalliance/formstools>

RemoteRef HEAD

RemoteSha 778958d15125bf1e22232a69d045a03df0bee55a

Contents

form_codebook	2
get_choices	2
get_choices_ruodk	3
get_questions	4
match_other_to_choices	5

pizza_data	6
split_multiple_response	7
spread_vector_to_columns	8
Index	9

form_codebook	<i>An example ODK form schema retrieved using ruODK</i>
---------------	---

Description

An example ODK form schema retrieved using ruODK

Usage

form_codebook

Format

A tibble with 7 columns and 10 rows:

| ****Variable**** | ****Description**** | | :— | | ***path*** | XML path to question in ODK | | ***name*** | Variable name used in XForm | | ***type*** | Variable type | | ***binary*** | Are the values for the question binary? | | ***ruodk_name*** | Variable name used by ruODK | | ***label*** | Question used in XForm | | ***choices*** | Named list of choices for select_one and select_multiple questions |

Examples

form_codebook

get_choices	<i>Get choices used for select_one and select_multiple questions in an XLSForm</i>
-------------	--

Description

Get choices used for select_one and select_multiple questions in an XLSForm

Usage

get_choices(xlsform, choice_name = NULL)

Arguments

xlsform	A character value for the path to the XLSForm
choice_name	A character value or vector of values of names given to the set of choices in the supplied/specified XLSForm. Default to NULL in which case all sets of choices are returned.

Value

A tibble of the 'list_name', 'name', and 'label' of the choices for the specified 'select_one' and 'select_multiple' type of questions in the specified XLSForm

Examples

```
## Get all choices
get_choices(
  xlsform = system.file(
    "extdata", "ghana_community_form.xlsx", package = "formstools"
  )
)

## Get choices for gender
get_choices(
  xlsform = system.file(
    "extdata", "ghana_community_form.xlsx", package = "formstools"
  ),
  choice_name = "gender"
)
```

get_choices_ruodk	<i>Get choices for a specified variable in an ODK form</i>
-------------------	--

Description

Get choices for a specified variable in an ODK form

Usage

```
get_choices_ruodk(form_schema, var_name = NULL, choice_name = NULL)
```

Arguments

form_schema	A form schema created by 'ruODK::form_scheme_ext()'
var_name	Specific variable name in ODK form to get choices of
choice_name	Specific variable name in 'form_schema' that holds the information on choices. Forms with multiple languages will have a 'form_schema' with separate columns for choices for each language. Default is NULL which assumes that only one language is used in the form.

Value

A tibble with 'var_name', 'question', 'values', and 'labels'

Examples

```
## Get all choices for select_one and select_multiple type questions
get_choices_ruodk(form_schema = form_codebook)

## Get choices for pizza2 variable
get_choices_ruodk(form_schema = form_codebook, var_name = "pizza2")
```

get_questions	<i>Get questions used for select_one and select_multiple questions in an XLSForm</i>
---------------	--

Description

Get questions used for select_one and select_multiple questions in an XLSForm

Usage

```
get_questions(xlsform, choice_name = NULL)
```

Arguments

xlsform	A character value for the path to the XLSForm
choice_name	A character value or vector of values of names given to the set of choices in the supplied/specified XLSForm. Default to NULL in which case all 'select_one' and 'select_multiple' questions types are returned.

Value

A tibble of the 'type', 'list_name', 'name', and 'label' of the choices for the specified 'select_one' and 'select_multiple' type of questions in the specified XLSForm

Examples

```
## Get all select_one and select_multiple question types
get_questions(
  xlsform = system.file(
    "extdata", "ghana_community_form.xlsx", package = "formstools"
  )
)

## Get all select_one and select_multiple question types for gender list_name
get_questions(
  xlsform = system.file(
```

```

      "extdata", "ghana_community_form.xlsx", package = "formstools"
    ),
    choice_name = "gender"
  )

```

match_other_to_choices

Match other responses to specified choices

Description

Match other responses to specified choices

Usage

```

match_other_to_choices(
  form_schema,
  form_data,
  var_name,
  choice_name = NULL,
  other_var_name,
  id = "id"
)

```

Arguments

form_schema	A form schema created by 'ruODK::form_scheme_ext()'
form_data	A tibble of the data retrieved from ODK Central using ruODK
var_name	Specific variable name in ODK form to get choices of
choice_name	Specific variable name in 'form_schema' that holds the information on choices. Forms with multiple languages will have a 'form_schema' with separate columns for choices for each language. Default is NULL which assumes that only one language is used in the form.
other_var_name	Variable name in 'form_data' for other response in 'var_name'
id	A character vector of variable names in 'form_data' to use as identifying data for the output

Value

A tibble with number of columns the same as the number of possible choices for 'var_name' in 'form_data' plus the number of values in 'id' used to identify each row of the output. The number of rows is equal to the number of rows of data in 'form_data'. Names of resulting tibble is the concatenation of 'recode_' and the choice values for 'var_name'

Examples

```
## Match other toppings responses (pizza3) to choices for
## pizza toppings (pizza2)
match_other_to_choices(
  form_schema = form_codebook, form_data = pizza_data,
  var_name = "pizza2", other_var_name = "pizza3"
)
```

pizza_data	<i>An example data collected using ODK and aggregated using ODK Central</i>
------------	---

Description

An example data collected using ODK and aggregated using ODK Central

Usage

```
pizza_data
```

Format

A tibble with 21 columns and 17 rows:

```
| **Variable** | **Description** | :— | :— | | *id* | Identifier | | *start* | Start date and time |
| *end* | End date and time | | *today* | Date of data entry | | *pizza1* | Do you like pizza? | |
| *pizza2* | Pizza toppings you like (select multiple) | | *pizza3* | Any other toppings? | | *closing* | Closing statement | |
| *meta_audit* | Meta audit | | *meta_instance_id* | Meta instance ID | |
| *system_submission_date* | System submission date | | *system_updated_at* | System updated at | |
| *system_submitter_id* | System submitter ID | | *system_submitter_name* | System submitter name | |
| *system_attachments_present* | System attachments present | | *system_attachments_expected* | System attachments expected | |
| *system_status* | System status | | *system_review_state* | System review state | |
| *system_device_id* | System device ID | | *system_edits* | System edits | |
| *odata_context* | OData context |
```

Examples

```
pizza_data
```

split_multiple_response*Split a vector of values from an ODK select multiple type of response*

Description

Split a vector of values from an ODK select multiple type of response

Usage

```
split_multiple_response(x, sep = " ", fill, na_rm = FALSE, prefix)
```

```
split_multiple_responses(x, sep = " ", fill, na_rm = FALSE, prefix)
```

Arguments

x	A select_multiple response or vector of responses with multiple responses
sep	Separator used to separate multiple responses. Default to " ". Regular expressions can be used to detect more than one possible separator.
fill	A vector of all the possible responses to the select multiple question
na_rm	Logical. Should an NA response be reported in its own column? Default to FALSE.
prefix	A character value for prefix to append to names of resulting data.frame.

Value

A tibble with number of columns the same as the number of possible choices for the select multiple question (same as length of 'fill') and number of rows equal to the length of 'x'. Names of resulting tibble is the concatenation of the 'prefix' and the values for 'fill'

Examples

```
## Split the multiple responses of pizza toppings
split_multiple_responses(
  x = pizza_data$pizza2,
  sep = ", ",
  fill = c("cheese", "tomatoes", "pepperoni", "mushrooms",
           "artichoke", "olives", "pineapple", "other"),
  prefix = "toppings"
)
```

`spread_vector_to_columns`*Convert character vector of categorical responses into unique variables*

Description

Function transforms a vector of categorical responses into ‘n’ number of new columns/variables equal to the number of unique categorical values.

Usage

```
spread_vector_to_columns(x, fill = NULL, na_rm = FALSE, prefix)
```

Arguments

<code>x</code>	Vector of categorical values
<code>fill</code>	A vector of all the possible responses to the select multiple question
<code>na_rm</code>	Logical. Should an NA response be reported in its own column? Default to FALSE.
<code>prefix</code>	A character string to prepend to the names of the new columns to be created

Value

A tibble with number of columns the same as the number of unique values ‘x’ and number of rows equal to the length of ‘x’. Names of resulting tibble is the concatenation of the ‘prefix’ and the values of ‘x’

Examples

```
spread_vector_to_columns(  
  x = c("cat", "cat", "dog", "dog", "dog", NA_character_),  
  fill = c("cat", "dog", "monkey"),  
  na_rm = TRUE,  
  prefix = "pets"  
)
```

Index

* datasets

form_codebook, [2](#)

pizza_data, [6](#)

form_codebook, [2](#)

get_choices, [2](#)

get_choices_ruodk, [3](#)

get_questions, [4](#)

match_other_to_choices, [5](#)

pizza_data, [6](#)

split_multiple_response, [7](#)

split_multiple_responses

(split_multiple_response), [7](#)

spread_vector_to_columns, [8](#)