

Package ‘dataqualitycontrol’

December 10, 2019

Type Package

Title Data quality check

Version 0.1.0

Author Alberto Gualtieri

Maintainer Alberto Gualtieri <alberto.gualtieri@reach-initiative.org>

Description

Data quality control according to a predefined list of data quality checks submitted by the user and produces outputs that allow for a straightforward visualisation and mapping of data entry errors. Requires IMPACT ``composer" and ``koboquest".

License GPL-3 + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 6.1.1

Depends R (>= 3.0.0),

Imports dplyr,
reshape2,
questionr,
composr,
koboquest (>= 1.0.1),
kobostandards,
data.table,
surveyweights,
htmltools,
assertthat,
survey,
magrittr,
tidyr,
crayon,
tibble

Remotes github::mabafaba/koboquest,
github::mabafaba/kobostandards,
github::mabafaba/composr,

Suggests testthat,
rmarkdown,
knitr

VignetteBuilder knitr

R topics documented:

batch_issue_checks	2
dataqualitycontrol	2
quality_checks_log_to_long_format	3
run_checks_from_dataframe	3

Index	5
--------------	----------

batch_issue_checks	<i>Batch check of predefined data quality conditions</i>
--------------------	--

Description

The following function checks for the presence of invalid combinations of variables within the supplied data and produced and editable output that can be used for data quality review.

Usage

```
batch_issue_checks(df, conditions, tests, meta_to_keep = c())
```

Arguments

df	the dataframe you wish to check for invalid combinations of variables
conditions	the set of conditions you wish to check
tests	the set of test related to each condition
meta_to_keep	a list of metadata you wish to include in the final output

Value

Standardized issues table for quality check in long format.

Author(s)

Alberto Gualtieri, <alberto.gualtieri@reach-initiative.org>

References

<https://github.com/agualtieri/dataqualitycontrol>

dataqualitycontrol	<i>dataqualitycontrol: A package to produce standardized data quality output</i>
--------------------	--

Description

The data quality control package provides two categories of functions:

- batch quality checks
- quality log to long format

`quality_checks_log_to_long_format`*Function to melt quality check log from long to wide*

Description

The following function turns the quality check output from wide to long format.

Usage

```
quality_checks_log_to_long_format(data, variable_name = "variable",  
  value_name = "value", meta_not_to_transform = c())
```

Arguments

<code>data</code>	the dataset you wish to transform from wide to long
<code>variable_name</code>	the name of the measured variable. The default name is 'variable'
<code>value_name</code>	name of the transformed data value. The default name is 'value'
<code>meta_not_to_transform</code>	a list of metadata you wish to keep as id variables.

Value

A transformed quality check log

Author(s)

Alberto Gualtieri, <alberto.gualtieri@reach-initiative.org>

References

<https://github.com/agualtieri/dataqualitycontrol>

`run_checks_from_dataframe`*Run predefined quality check test on a raw dataset*

Description

The following function checks for the presence of invalid data quality combinations within a dataframe using the supplied condition_list.

Usage

```
run_checks_from_dataframe(df, conditions_df, condition.column,  
  test.name.column, meta_to_keep = c())
```

Arguments

<code>df</code>	the dataset you wish to check for invalid combinations of variables
<code>conditions_df</code>	the dataframe containing the name of the checks and the list of controls to perform
<code>condition.column</code>	the column name that contains the conditions you want to check in the dataset
<code>test.name.column</code>	the column name that contains the description of the checks you want to perform
<code>meta_to_keep</code>	a list of metadata you wish to include in the final output

Details

The `condition_list` must be a dataframe containing at least two variables: `condition` as the name of the condition one wishes to check and `test` as the test applied to each condition.

Value

Standardized issues table for quality check in long format.

Author(s)

Alberto Gualtieri, <alberto.gualtieri@reach-initiative.org>

References

<https://github.com/agualtieri/dataqualitycontrol>

Index

- *Topic **check**,
 - batch_issue_checks, [2](#)
 - quality_checks_log_to_long_format,
[3](#)
 - run_checks_from_dataframe, [3](#)
- *Topic **mcla**,
 - batch_issue_checks, [2](#)
 - quality_checks_log_to_long_format,
[3](#)
 - run_checks_from_dataframe, [3](#)
- *Topic **platypus**
 - batch_issue_checks, [2](#)
 - quality_checks_log_to_long_format,
[3](#)
 - run_checks_from_dataframe, [3](#)
- *Topic **quality**
 - batch_issue_checks, [2](#)
 - quality_checks_log_to_long_format,
[3](#)
 - run_checks_from_dataframe, [3](#)
- *Topic **yemen**,
 - batch_issue_checks, [2](#)
 - quality_checks_log_to_long_format,
[3](#)
 - run_checks_from_dataframe, [3](#)

batch_issue_checks, [2](#)

dataqualitycontrol, [2](#)

dataqualitycontrol-package
(dataqualitycontrol), [2](#)

quality_checks_log_to_long_format, [3](#)

run_checks_from_dataframe, [3](#)