# Package 'MOQA'

October 12, 2022

**Type** Package

**Title** Basic Quality Data Assurance for Epidemiological Research

**Version** 2.0.0

**Date** 2017-06-21

**Author**

Martin Bialke <mosaic-projekt@uni-greifswald.de>, Thea Schwaneberg <thea.schwaneberg@uni-greifswald.de>

**Maintainer** Martin Bialke <mosaic-projekt@uni-greifswald.de>

**Description** With the provision of several tools and templates the MOSAIC project (DFG-Grant Number HO 1937/2-1) supports the implementation of a central data management in epidemiological research projects. The 'MOQA' package enables epidemiologists with none or low experience in R to generate basic data quality reports for a wide range of application scenarios. See <https://mosaic-greifswald.de/> for more information. Please read and cite the corresponding open access publication (using the former package-name) in METHODS OF INFORMATION IN MEDICINE by M. Bialke, H. Rau, T. Schwaneberg, R. Walk, T. Bahls and W. Hoffmann (2017) <doi:10.3414/ME16-01-0123>. <https://methods.schattauer.de/en/contents/most-recent-articles/issue/2483/issue/special/manuscript/27573/show.html>.

**License** AGPL-3

**Depends** psych, gplots, grid, readr

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-06-22 13:23:11 UTC

## R topics documented:

---

| codelist | *codelist* |
|----------|-----------|

---

## Description

internal data variable

## Note

internal data variable

## Author(s)

The MOSAIC Project, Martin Bialke

| footnoteString | *footnoteString* |
|---|---|

### Description

internal data variable

### Note

internal data variable

### Author(s)

The MOSAIC Project, Martin Bialke

| labelCounts | *labelCounts* |
|---|---|

### Description

internal label for data variable

### Note

internal label for data variable

### Author(s)

The MOSAIC Project, Martin Bialke

| labelPercentage | *labelPercentage* |
|---|---|

### Description

internal label for data variable

### Note

internal label for data variable

### Author(s)

The MOSAIC Project, Martin Bialke

---

label_boxplot *label_boxplot*

---

### Description

internal label for data variable

### Note

internal label for data variable

### Author(s)

The MOSAIC Project, Martin Bialke

---

label_description *label_description*

---

### Description

internal label for data variable

### Note

internal label for data variable

### Author(s)

The MOSAIC Project, Martin Bialke

---

label_normalverteilung

*label_normalverteilung*

---

### Description

internal label for data variable

### Note

internal label for data variable

### Author(s)

The MOSAIC Project, Martin Bialke

---

| label_qnormplot | *label_qnormplot* |
|---|---|

---

## Description

internal label for data variable

## Note

internal label for data variable

## Author(s)

The MOSAIC Project, Martin Bialke

---

| label_unit | *label_unit* |
|---|---|

---

## Description

internal label for data variable

## Note

internal label for data variable

## Author(s)

The MOSAIC Project, Martin Bialke

---

| MOQA | *Basic Quality Data Assurance for Epidemiological Research* |
|---|---|

---

## Description

With the provision of several tools and templates the MOSAIC project (DFG-Grant Number HO 1937/2-1) supports the implementation of a central data management in epidemiological research projects. The 'MOQA' package enables epidemiologists with none or low experience in R to generate basic data quality reports for a wide range of application scenarios. See <https://mosaic-greifswald.de/> for more information. Please read and cite the corresponding open access publication (using the former package-name) in METHODS OF INFORMATION IN MEDICINE by M. Bialke, H. Rau, T. Schwaneberg, R. Walk, T. Bahls and W. Hoffmann (2017) <doi:10.3414/ME16-01-0123>. <https://methods.schattauer.de/en/contents/most-recent-articles/issue/2483/issue/special/manuscript/27573/show.

## Details

The DESCRIPTION file:

| | |
|---|---|
| Package: | MOQA |
| Type: | Package |
| Title: | Basic Quality Data Assurance for Epidemiological Research |
| Version: | 2.0.0 |
| Date: | 2017-06-21 |
| Author: | Martin Bialke <mosaic-projekt@uni-greifswald.de>, Thea Schwaneberg <thea.schwaneberg@uni-greifs |
| Maintainer: | Martin Bialke <mosaic-projekt@uni-greifswald.de> |
| Description: | With the provision of several tools and templates the MOSAIC project (DFG-Grant Number HO 1937/2 |
| License: | AGPL-3 |
| Depends: | psych, gplots, grid, readr |
| NeedsCompilation: | no |
| Repository: | CRAN |

Index of help topics:

```
MOQA.env                MOQA.env
codelist                codelist
footnoteString          footnoteString
labelCounts             labelCounts
labelPercentage         labelPercentage
label_boxplot           label_boxplot
label_description       label_description
label_normalverteilung
                        label_normalverteilung
label_qnormplot         label_qnormplot
label_unit              label_unit
moqa                    Basic Quality Data Assurance for
                        Epidemiological Research
mosaic.addFootnote      addFootnote
mosaic.beginPlot        beginPlot
mosaic.countValue       countValue
mosaic.createSimplePdfCategorical
                        createSimplePdfCategorical
mosaic.createSimplePdfCategoricalDataframe
                        createSimplePdfCategoricalDataframe
mosaic.createSimplePdfMetric
                        createSimplePdfMetric
mosaic.createSimplePdfMetricDataframe
                        createSimplePdfMetricDataframe
mosaic.finishPlot       finishPlot
mosaic.generateCategoricalPlot
                        generateCategoricalPlot
mosaic.generateMetricPlots
                        generateMetricPlots
mosaic.generateMetricTablePlot
                        generateMetricTablePlot
mosaic.getTimestamp     getTimestamp
```

```
mosaic.importToolboxSpssDataFile
                       importToolboxSpssDataFile
mosaic.info           info
mosaic.loadCsvData    loadCsvData
mosaic.preProcessCategoricalData
                       preProcessCategoricalData
mosaic.preProcessMetricData
                       preProcessMetricData
mosaic.setGlobalCodelist
                       setGlobalCodelist
mosaic.setGlobalDescription
                       setGlobalDescription
mosaic.setGlobalMissingTreshold
                       setGlobalMissingTreshold
mosaic.setGlobalUnit  setGlobalUnit
outputPrefix          outputPrefix
qualifiedMissingsTreshold
                       qualifiedMissingsTreshold
```

The aim of the MOQA R-Package is to provide a basic assessment of data quality and to generate a set of informative graphs. Especially, there should be no demand for the potential researcher to master R. This R-package enables researchers to generate reports for various kinds of metric and categorical data. Additionally, general reports for multivariate input data and, if needed, detailed results for single-variable data can be produced.

CSV-files as well as dataframes can be used as input format to create a report. The results are instantly saved in an automatically generated PDF-file. For each study variable within the data input file a separate PDF-file with standard or, if applicable, customized plots and tables is produced. These standard reports enable the user to monitor and report the data integrity and completeness. However, for more specific reports the knowledge of metadata is necessary, including definition of units, variables, descriptions, code lists and categories of qualified missings.

Version 1.2 ———— ADDED Support for metric and categorical dataframes BUGFIX Aborted report generation in case of non-existent missings in datacolumn

Version 2.0 ———— RENAME Official Renaming of former package-name mosaicQA to MOQA ADDED new function importToolboxSpssDataFile

### Author(s)

Martin Bialke <mosaic-projekt@uni-greifswald.de>, Thea Schwaneberg <thea.schwaneberg@uni-greifswald.de>, Rene Walk <rene.walk@uni-greifswald.de>

Maintainer: Martin Bialke <mosaic-projekt@uni-greifswald.de>

### See Also

mosaic-greifswald.de

### Examples

```
## Example 1: Generate pdf with graphs for a single metric data column, e.g. data of body height

# load MOQA package
library('MOQA')

# specify the csv import file with metric data, use one column per variable
metric_datafile='c:/mosaic/metric_single_var.csv'

#specify output folder
outputFolder='c:/mosaic/outputs/'

#set missing threshold, optional, default is 99900
mosaic.setGlobalMissingTreshold(99900)

#set variable unit, optional
mosaic.setGlobalUnit('(cm)')

#set variable description, optional, if not uses the name of the variable is displayed in
#table heading
mosaic.setGlobalDescription('Height')

#create PDF-report,
#uncomment to start report-generation
#mosaic.createSimplePdfmetric(metric_datafile, outputFolder)



## Example 2: Generate pdf with graphs for a single categorical data column

# load MOQA package
library('MOQA')

# specify the import file with Categorical data
# first row has to contain variable names without special characters
Categorical_datafile='c:/mosaic/cat_single_var_en.csv'

#specify output folder
outputFolder='c:/mosaic/outputs/'

#set treshold to detect missings, default is 99900 (adjust this line to change this global value,
#but be careful)
mosaic.setGlobalMissingTreshold(99900)

#set description of var
mosaic.setGlobalCodelist(c('1=yes','2=no','99996=not specified','99997=not acquired'))

# create simple pdf file foreach variable column in Categorical data file,
# uncomment to start report-generation
# mosaic.createSimplePdfCategorical(Categorical_datafile,outputFolder)
```

```
## Example 3: Generate pdf with graphs for a multiple metric data columns, generates one pdf for
# each column using the variable name for table headings

# load MOQA package
library('MOQA')

# specify the import file with metric data
# use one column per variable, first row should contain variable name, following rows should
# contain data, csv Files with multiple rows are supported, decimal values should be formated
# for example : 25.4
metric_datafile='c:/mosaic/metric_multi_var.csv'

#specify output folder
outputFolder="c:/mosaic/outputs/"

# set treshold to detect missings, default is 99900 (adjust this line to change this global value
# but be careful)
mosaic.setGlobalMissingTreshold(99900)

# create PDF-Files for vars,
# uncomment to start report-generation
#mosaic.createSimplePdfmetric(metric_datafile, outputFolder)




## Example 4: Generate pdf with graphs for a multiple metric dataframe, generates one pdf for
# each column using the variable name for table headings

# load MOQA package
library('MOQA')

# specify the metric dataframe with 1-n columns, here sample data is generated
metric_data=data.frame(matrix(rnorm(20), nrow=10))

#specify output folder
outputFolder="c:/mosaic/outputs/"

# set treshold to detect missings, default is 99900 (adjust this line to change this global value
# but be careful)
mosaic.setGlobalMissingTreshold(99900)

# create PDF-Files for vars,
# uncomment to start report-generation
#mosaic.createSimplePdfMetricDataframe(metric_data, outputFolder)




## Example 5: Import data from SPSS Export file generated by Toolbox for Research
# and generate report for specific variable

# load MOQA package
library('MOQA')
```

```
# specify import dat-file
importfile="c:/mosaic/import/all_in_one.dat"

# specify output folder
outputFolder="c:/mosaic/outputs/"

# import data
#importdata=mosaic.importToolboxSpssDataFile(importfile)

# generate report for a specifc variable e.e. patient.age
# pass data as dataframe to use already given column name for a more descriptive output
#mosaic.createSimplePdfMetricDataframe(as.data.frame(importdata$ve_temperature_ear),outputFolder)
```

---

MOQA.env                     *MOQA.env*

---

### Description

local environment to handle MOQA-internal variables

### Note

local environment

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.addFootnote          *addFootnote*

---

### Description

Add a Footnote to plot using footnotestring and current timestamp.

### Usage

```
mosaic.addFootnote()
```

### Note

Function call type: internal

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.beginPlot *beginPlot*

---

### Description

begin plotting the configured graphs for loaded data and generate the output PDF-File.

### Usage

```
mosaic.beginPlot(varname,outputfolder)
```

### Arguments

varname  name of the studyitem or csv column loaded to plot graphs for.

outputfolder  name of the output folder

### Note

Function call type: internal

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.countValue *countValue*

---

### Description

Count occurrence of search value in data column

### Usage

```
mosaic.countValue(searchvalue, data_column)
```

### Arguments

searchvalue  value to search for

data_column  name of study item or data column to search in

### Details

useful to find qualified missings in data column

### Value

count of occurences of specified value in specified data column

**Note**

Function call type: internal

**Author(s)**

The MOSAIC Project, Martin Bialke

---

mosaic.createSimplePdfCategorical

*createSimplePdfCategorical*

---

**Description**

Create simple PDF-file for categorical data

**Usage**

```
mosaic.createSimplePdfCategorical(inputfile, outputfolder)
```

**Arguments**

inputfile          path to input csv-file

outputfolder      path to output folder

**Note**

Function call type: user

**Author(s)**

The MOSAIC Project, Martin Bialke

**Examples**

```
# load MOQA package
library('MOQA')

# specify the import file with categorial data
# first row has to contain variable names without special characters
categorial_datafile='c:/mosaic/cat_single_var_en.csv'

# specify output folder
outputFolder='c:/mosaic/outputs/'

# set treshold to detect missings, default is 99900 (adjust this line to change this global value,
# but be careful)
mosaic.setGlobalMissingTreshold(99900)
```

```
# set description of var
mosaic.setGlobalCodelist(c('1=yes','2=no','99996=not specified','99997=not acquired'))

# create simple pdf file foreach variable column in categorial data file, uncomment to start
# report-generation
# mosaic.createSimplePdfCategorical(categorial_datafile,outputFolder)
```

---

mosaic.createSimplePdfCategoricalDataframe

*createSimplePdfCategoricalDataframe*

---

### Description

Create simple PDF-file for categorical data

### Usage

```
mosaic.createSimplePdfCategoricalDataframe(df, outputfolder)
```

### Arguments

df            dataframe

outputfolder  path to output folder

### Note

Function call type: user

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.createSimplePdfMetric

*createSimplePdfMetric*

---

### Description

Create simple PDF-file for metric data

### Usage

```
mosaic.createSimplePdfMetric(inputfile, outputfolder)
```

## Arguments

inputfile        path to input csv file

outputfolder     path to output folder

## Note

Function call type: user

## Author(s)

The MOSAIC Project, Martin Bialke

## Examples

```
# load MOQA package
library('MOQA')

# specify the csv import file with metric data, use one column per variable
metric_datafile='c:/mosaic/metric_single_var.csv'

#specify output folder
outputFolder='c:/mosaic/output/'

#set missing threshold, optional, default is 99900
mosaic.setGlobalMissingTreshold(99900)

#set variable unit, optional
mosaic.setGlobalUnit('(cm)')

#set variable description, optional
mosaic.setGlobalDescription('Height')

#create PDF-report, uncomment to start report-generation
#mosaic.createSimplePdfMetric(metric_datafile, outputFolder)
```

---

mosaic.createSimplePdfMetricDataframe

*createSimplePdfMetricDataframe*

---

## Description

Create simple PDF-file for metric data

## Usage

```
mosaic.createSimplePdfMetricDataframe(df, outputfolder)
```

## Arguments

df            path to input csv file

outputfolder     path to output folder

## Note

Function call type: user

## Author(s)

The MOSAIC Project, Martin Bialke

## Examples

```
# load MOQA package
library('MOQA')

# specify the metric dataframe with 1-n columns, here sample data is generated
metric_data=data.frame(matrix(rnorm(20), nrow=10))

#specify output folder
outputFolder="c:/mosaic/outputs/"

# set treshold to detect missings, default is 99900 (adjust this line to change this global value
# but be careful)
mosaic.setGlobalMissingTreshold(99900)

# create PDF-Files for vars,
# uncomment to start report-generation
#mosaic.createSimplePdfMetricDataframe(metric_data, outputFolder)
```

---

mosaic.finishPlot        *finishPlot*

---

## Description

Finish plotting, close PDF-file

## Usage

```
mosaic.finishPlot()
```

## Note

Function call type: internal

## Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.generateCategoricalPlot

*generateCategoricalPlot*

---

### Description

Generate Statistics and Create plots for categorical data

### Usage

```
mosaic.generateCategoricalPlot(dataframe, varname)
```

### Arguments

| | |
|---|---|
| dataframe | data table with one or more columns (first row should contain column names/study item names/variable names) |
| varname | selected column/study item/variable to plot graph for |

### Note

Function call type: internal

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.generateMetricPlots

*generateMetricPlots*

---

### Description

calculate statistics and generate graphs for metric data

### Usage

```
mosaic.generateMetricPlots(data_snippet, var_name)
```

### Arguments

| | |
|---|---|
| data_snippet | data table with one or more columns (first row should contain column names/study item names/variable names) |
| var_name | selected column/study item/variable to plot graph for |

## Note

Function call type: internal

## Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.generateMetricTablePlot

*generateMetricTablePlot*

---

## Description

Generate missing-ratio table for metric data (data, num of columns, column index, varname)

## Usage

```
mosaic.generateMetricTablePlot(data, num_of_columns, index, varname)
```

## Arguments

| | |
|---|---|
| data | preprocessed data frame including 'valid value markers' |
| num_of_columns | absolute number of to be processed data columns |
| index | current column to be processed |
| varname | current name of variable to be used in table heading |

## Note

Function call type: internal

## Author(s)

The MOSAIC Project, Martin Bialke

mosaic.getTimestamp          *getTimestamp*

### Description

get a current timestamp formatted as %Y_%m_%d_%H%M%S

### Usage

```
mosaic.getTimestamp()
```

### Value

timestamp, e.g. '2016_09_09_143458'

### Note

Function call type: internal

### Author(s)

The MOSAIC Project, Martin Bialke

mosaic.importToolboxSpssDataFile
                      *importToolboxSpssDataFile*

### Description

load dat-file from 'toolbox for resarch' spss export with tab-separator with n columns to dataframe

### Usage

```
mosaic.importToolboxSpssDataFile(filename)
```

### Arguments

filename          filename or a complete path to a dat-file

### Note

Function call type: user

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.info *info*

---

### Description

MOSAIC Information

### Usage

```
mosaic.info()
```

### Note

Function call type: user

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.loadCsvData *loadCsvData*

---

### Description

Load data from csv-file is one or more columns. first row should contain the name of the study item, e.g. 'height'

### Usage

```
mosaic.loadCsvData(filename)
```

### Arguments

filename        filename or a complete path to a file

### Note

Function call type: user

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.preProcessCategoricalData
                            *preProcessCategoricalData*

---

### Description

Identify unique values in data column, get absolute, percentage and cumulative statistics

### Usage

    mosaic.preProcessCategoricalData(data)

### Arguments

data            data frame to be processed containing categorical data

### Note

Function call type: internal

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.preProcessMetricData
                            *preProcessMetricData*

---

### Description

Pre-process metric data to allow missing-ratio table

### Usage

    mosaic.preProcessMetricData(data)

### Arguments

data            data frame to be preprocessed containing metric data

### Note

Function call type: internal

### Author(s)

The MOSAIC Project, Martin Bialke

---

mosaic.setGlobalCodelist

*setGlobalCodelist*

---

### Description

set and parse a global code list for categorical data to be used in categorical plot descriptions

### Usage

```
mosaic.setGlobalCodelist(coding)
```

### Arguments

coding          list of code and value pairs, see example for details

### Note

Function call type: user

### Author(s)

The MOSAIC Project, Martin Bialke

### Examples

```
mosaic.setGlobalCodelist(c('1=yes','2=no', '99996=no information'))
```

---

mosaic.setGlobalDescription

*setGlobalDescription*

---

### Description

Set Global Description for variable User (description) data. especially useful when plotting graphs for a selected data column

### Usage

```
mosaic.setGlobalDescription(value)
```

### Arguments

value           string value to be used as study item description, e.g. 'waist circumference'

## Note

Function call type: user

## Author(s)

The MOSAIC Project, Martin Bialke

## Examples

```
mosaic.setGlobalDescription('waist circumference')
```

---

mosaic.setGlobalMissingTreshold
*setGlobalMissingTreshold*

---

## Description

Set Global Threshold for Missings , e.g. 99000

## Usage

```
mosaic.setGlobalMissingTreshold(value)
```

## Arguments

value          threshold to separate missings from valid values

## Note

Function call type: user

## Author(s)

The MOSAIC Project, Martin Bialke

## Examples

```
mosaic.setGlobalMissingTreshold(99000)
```

---

mosaic.setGlobalUnit      *setGlobalUnit*

---

### Description

Set Global Unit Label to be used User in graphs, e.g. '(cm)'

### Usage

```
mosaic.setGlobalUnit(value)
```

### Arguments

value            unit string to be used in graphs

### Note

Function call type: user

### Author(s)

The MOSAIC Project, Martin Bialke

### Examples

```
mosaic.setGlobalUnit('(cm)')
```

---

outputPrefix            *outputPrefix*

---

### Description

internal data variable

### Note

internal data variable

### Author(s)

The MOSAIC Project, Martin Bialke

---

qualifiedMissingsTreshold

*qualifiedMissingsTreshold*

---

## Description

internal data variable

## Note

internal data variable

## Author(s)

The MOSAIC Project, Martin Bialke

# Index