Data Cleaning

With dataMaid

Cheat Sheet

Getting Started

install.packages("dataMaid")
Install the dataMaid package from CRAN.

library(dataMaid)
Load the dataMaid package.

Example Datasets

toyData
Small example data to show the features of dataMaid.

artData
Semi-artificial data about masterpieces of art.

exampleData
Example data with zero-inflated variables.

presidentData
Semi-artificial data about the US presidents.

List All Functions

allClasses( )
Supported variable classes in dataMaid

allCheckFunctions( )
View all available check functions.

allSummaryFunctions( )
View available summary functions

allVisualFunctions( )
Overview of all available visual functions.

Visualize

basicVisual(v, vnam, doEval = TRUE)
Produce distribution plots in the base R (graphics). For all variable classes.

standardVisual(v, vnam, doEval = TRUE)
Produce distribution plots using ggplot from ggplot2. For all variable classes.

Test Battery and Reports

check(v, nMax = 10, checks = setChecks( ), ...)
Perform checks of potential errors in variable or dataframe. For all variable classes.

makeDataReport(...)
Produce a data report. For all variable classes.

summarize(v, reportstyleOutput = FALSE, summaries = setSummaries( ), ...)
Summarize a variable or dataframe. For all variable classes

Check Functions

centralValue(v, ...)
Central values of data. For all variable classes

countMissing(v, ...)
Report number of missing values (i.e. NA). For all variable classes.

identifyCaseIssues(v, nMax = 10)
Identify values appearing multiple times with different case settings. For character, factor, and labeled variables.

identifyLoners(v, nMax = 10)
Identify sparsely represented values (loners). For character, labeled, or factor variables

identifyMissing(v, nMax = 10, ...)
Identify miscoded missing values. For all variable classes.

Check Functions (continued)

identifyNums(v, nVals = 12, ...)
Identify numeric values incorrectly listed as character, factor, or labeled data. For character, factor, or labeled classes.

identifyOutliers(v, nMax = 10, maxDecimals = 2)
Identify outliers. For date, numeric, or integer variable classes.

identifyWhitespace(v, nMax = 10)
Identify white space. For character, labeled, or factor variables.

isEmpty(v)
Identify if a variable contains only a single value. For all variable classes.

isKey(v)
Check if a variable qualifies as a key. For all variable types.

isSupported(v)
Check if a variable has a class supported by dataMaid. For all variable types.

minMax(v, maxDecimals = 2)
Get min and max values. For numeric or integer variable types.

quartiles(v, maxDecimals)
Get 1st and 3rd quartiles. For numeric or integer variable types.

uniqueValues(v)
Count number of unique values. For all variable classes.

Add Your Own Check Functions

checkFunction(f, description = NULL, classes = NULL)
Convert a function, f, into an S3 checkFunction object. This adds f to the overview list returned by an allCheckFunctions() call.

checkResult(ls)
Convert a list resulting from the checks performed in a checkFunction into a checkResult object, thereby supplying it with a print() method.

messageGenerator(problemStatus, message "", nMax = 10)
Helper function for producing output messages for checkFunction type functions.

summaryFunction(f, description, classes = NULL)
Convert a function, f, into an S3 summaryFunction object. This adds f to the overview list returned by an allSummaryFunctions() call.

summaryResult(ls)
Convert a list resulting from the summaries performed in a summaryFunction into a summaryResult object, thereby supplying it with a print() method.

visualFunction(f, description, classes = NULL)
Convert a function, f, into an S3 visualFunction object. This adds f to the overview list returned by an allVisualFunctions() call.