

Package ‘DataSum’

July 14, 2026

Type Package

Title Modern Data Summaries and Diagnostic Reports for Statistical Analysis

Version 1.0.0

Description Provides robust, NA-aware data summaries, variable diagnostics, normality decisions, missingness and outlier checks, and reproducible diagnostic report scaffolding for statistical analysis. DataSum is designed for researchers, professors, scientists, and analysts who need trustworthy first-pass insight into tabular data before modeling or publication.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.3

URL <https://github.com/Uzairkhan11w/DataSum>,
<https://uzairkhan11w.github.io/DataSum/>

BugReports <https://github.com/Uzairkhan11w/DataSum/issues>

Depends R (>= 4.1.0)

Imports nortest

Suggests knitr, quarto, rmarkdown, shiny, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Uzair Javid Khan [aut, cre],
Immad Ahmad Shah [aut],
Sukhdev Mishra [aut]

Maintainer Uzair Javid Khan <uzairkhan11w@gmail.com>

Repository CRAN

Date/Publication 2026-07-13 22:30:08 UTC

Contents

DataSum-package	2
datasum_report	2
profile_data	3
run_datasum_app	4
summarize_data	4
summarize_vector	5
Index	6

DataSum-package	<i>DataSum: Modern Data Summaries and Diagnostic Reports</i>
-----------------	--

Description

DataSum provides robust, NA-aware summaries, data profiles, normality diagnostics, outlier checks, reproducible report scaffolding, and an optional Shiny app for statistical data inspection.

See Also

`summarize_vector()`, `summarize_data()`, `profile_data()`, `datasum_report()`, `run_datasum_app()`

datasum_report	<i>Create a DataSum Diagnostic Report</i>
----------------	---

Description

Creates a self-contained Quarto source file containing dataset diagnostics, variable summaries, warnings, and formula definitions. Set `render = TRUE` to render the report through the optional quarto package.

Usage

```
datasum_report(  
  data,  
  path = NULL,  
  format = c("qmd", "html", "pdf", "docx"),  
  title = "DataSum Diagnostic Report",  
  by = NULL,  
  alpha = 0.05,  
  digits = 3,  
  render = FALSE  
)
```

Arguments

data	A data frame or tibble.
path	Output path. When render = FALSE, this should usually end in .qmd. If omitted, a temporary .qmd file is created.
format	One of "qmd", "html", "pdf", or "docx".
title	Report title.
by	Optional grouping columns passed to profile_data().
alpha	Significance level for normality decisions.
digits	Optional number of digits used to round numeric output.
render	Logical; if TRUE, render the Quarto document. Rendering requires the optional quarto package and a working Quarto installation.

Value

The generated file path, invisibly.

Examples

```
report <- datasum_report(iris, render = FALSE)
file.exists(report)
```

profile_data	<i>Profile a Data Frame</i>
--------------	-----------------------------

Description

Builds a dataset-level profile containing variable summaries, dataset shape, missingness, duplicate-row counts, type counts, and warnings that deserve analyst attention.

Usage

```
profile_data(data, by = NULL, alpha = 0.05, digits = NULL)
```

Arguments

data	A data frame or tibble.
by	Optional grouping columns passed to summarize_data().
alpha	Significance level for normality decisions.
digits	Optional number of digits used to round numeric output.

Value

A datasum_profile list with dataset, summary, and warnings.

Examples

```
profile <- profile_data(iris)
profile$dataset
```

run_datasum_app	<i>Launch the DataSum Shiny App</i>
-----------------	-------------------------------------

Description

Opens an interactive Shiny interface for uploading a CSV file, inspecting dataset diagnostics, visualizing a selected variable, and downloading a reproducible DataSum report source file.

Usage

```
run_datasum_app(data = NULL)
```

Arguments

data	Optional data frame used as the starting dataset. If omitted, the app starts with <code>datasets::iris</code> until a CSV is uploaded.
------	--

Value

A Shiny application object.

Examples

```
if (requireNamespace("shiny", quietly = TRUE)) {
  app <- run_datasum_app(iris)
}
```

summarize_data	<i>Summarize a Data Frame</i>
----------------	-------------------------------

Description

Applies `summarize_vector()` to every column in a data frame. Optional grouped summaries are supported by passing one or more grouping column names to `by`.

Usage

```
summarize_data(data, by = NULL, alpha = 0.05, digits = NULL)
```

Arguments

data	A data frame or tibble.
by	Optional character vector of grouping columns.
alpha	Significance level for normality decisions.
digits	Optional number of digits used to round numeric output.

Value

A data.frame, one row per summarized variable and group.

Examples

```
summarize_data(iris)
summarize_data(iris, by = "Species")
```

summarize_vector	<i>Summarize a Single Vector</i>
------------------	----------------------------------

Description

Computes a one-row, NA-aware diagnostic summary for one vector. Numeric vectors receive robust and classical statistics, outlier counts, and a normality diagnostic. Non-numeric vectors receive safe type, missingness, uniqueness, and mode summaries.

Usage

```
summarize_vector(x, name = NA_character_, alpha = 0.05, digits = NULL)
```

Arguments

x	A vector.
name	Optional variable name to store in the variable column.
alpha	Significance level for the normality decision. Defaults to 0.05.
digits	Optional number of digits used to round numeric output. By default, numeric values are not rounded.

Value

A one-row data.frame with summary statistics and diagnostics.

Examples

```
summarize_vector(c(1, 2, 2, NA, 5), name = "score")
summarize_vector(factor(c("control", "treatment", "control")))
```

Index

* **package**

DataSum-package, [2](#)

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

DataSum-package, [2](#)

datasum_report, [2](#)

profile_data, [3](#)

run_datasum_app, [4](#)

summarize_data, [4](#)

summarize_vector, [5](#)