

Package ‘AutoEDA’

September 12, 2026

Type Package

Title Automatic Exploratory Data Analysis

Version 0.1.1

Description Automatically performs exploratory data analysis for tabular datasets, including data summaries, missing value analysis, descriptive statistics, visualizations, correlation analysis, outlier detection, and automated report generation. The package provides a streamlined workflow for rapid data exploration and produces publication-ready tables and graphics. For methodological details, see Tukey (1977, ISBN:9780201076165) and the cited methodological literature in the package documentation.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

Depends R (>= 4.2)

Imports e1071, rlang, dplyr, ggplot2, tidyr, psych, factoextra, openxlsx, GGally, visdat, igraph

Suggests knitr, mice, rmarkdown, testthat (>= 3.0.0), tibble

Config/testthat/edition 3

URL <https://github.com/vinodhpm/AutoEDA>

BugReports <https://github.com/vinodhpm/AutoEDA/issues>

VignetteBuilder knitr

NeedsCompilation no

Config/roxygen2/version 8.1.0

Author Vinodhkumar Obli Rajendran [aut, cre],
Keerthi Aaradhana [aut]

Maintainer Vinodhkumar Obli Rajendran <vinodhkumar.rajendran@gmail.com>

Repository CRAN

Date/Publication 2026-09-12 14:50:15 UTC

Contents

auto_eda	3
auto_plot	3
bar_plot	4
box_plot	4
categorical_plot	5
categorical_summary	5
cluster_analysis	6
cluster_dendrogram	6
cluster_elbow	7
cluster_plot	7
cluster_silhouette	8
correlation_analysis	8
correlation_network	9
correlation_plot	9
correlation_test	10
create_report	10
density_plot	11
export_excel	11
export_numeric_summary	12
histogram_plot	12
html_report	13
missing_heatmap	13
missing_pattern	14
missing_summary	14
numeric_summary	15
outlier_detection	15
outlier_summary	16
outlier_values	16
pair_plot	17
pca_analysis	17
pca_biplot	18
pca_individual_plot	18
pca_scee_plot	19
pca_variable_plot	19
pdf_report	20
plot_missing	20
plot_outliers	21
print.AutoEDAResult	21
print.CategoricalSummary	22
print.ClusterResult	22
print.CorrelationMatrix	23
print.MissingSummary	23
print.NumericSummary	24
print.OutlierSummary	24
print.PCAResult	25
scatter_plot	25

auto_eda

3

summary_data

26

theme_autoeda

26

word_report

27

Index

28

auto_eda	<i>Automatic Exploratory Data Analysis</i>
----------	--

Description

Performs a complete exploratory data analysis of a dataset.

Usage

auto_eda(data)

Arguments

data A data.frame.

Value

An object of class AutoEDAReport.

auto_plot	<i>Automatic Plot</i>
-----------	-----------------------

Description

Automatically selects an appropriate plot based on the variable type.

Usage

auto_plot(data, variable)

Arguments

data A data.frame.
variable Name of the variable.

Value

A ggplot object.

`bar_plot`*Bar Plot*

Description

Draws a bar plot for a categorical variable.

Usage

```
bar_plot(data, variable)
```

Arguments

<code>data</code>	A data.frame.
<code>variable</code>	Name of the categorical variable.

Value

A ggplot object.

`box_plot`*Box Plot*

Description

Draws a box plot for a numeric variable.

Usage

```
box_plot(data, variable)
```

Arguments

<code>data</code>	A data.frame.
<code>variable</code>	Name of the numeric variable.

Value

A ggplot object.

categorical_plot	<i>Categorical Plot</i>
------------------	-------------------------

Description

Draws a bar plot for a categorical variable.

Usage

```
categorical_plot(data, variable)
```

Arguments

- data A data.frame.
- variable Name of a categorical variable.

Value

A ggplot object.

categorical_summary	<i>Categorical Summary</i>
---------------------	----------------------------

Description

Generates summary statistics for categorical variables.

Usage

```
categorical_summary(data)
```

Arguments

- data A data.frame.

Value

An object of class CategoricalSummary.

cluster_analysis	<i>Cluster Analysis</i>
------------------	-------------------------

Description

Performs K-means clustering on numeric variables.

Usage

```
cluster_analysis(data, centers = 3, scale = TRUE, nstart = 25)
```

Arguments

data	A data.frame.
centers	Number of clusters.
scale	Logical; should variables be scaled?
nstart	Number of random starts.

Value

An object of class `ClusterResult`.

cluster_dendrogram	<i>Hierarchical Clustering Dendrogram</i>
--------------------	---

Description

Performs hierarchical clustering and displays a dendrogram.

Usage

```
cluster_dendrogram(data)
```

Arguments

data	A data.frame.
------	---------------

Value

Invisibly returns the `hclust` object.

`cluster_elbow`*Elbow Method*

Description

Displays the Within-Cluster Sum of Squares (WSS) to help determine the optimal number of clusters.

Usage

```
cluster_elbow(data)
```

Arguments

`data` A `data.frame`.

Value

A `ggplot` object.

`cluster_plot`*Cluster Plot*

Description

Plots clustering results.

Usage

```
cluster_plot(cluster)
```

Arguments

`cluster` A `ClusterResult` object.

Value

A `ggplot` object.

cluster_silhouette	<i>Silhouette Plot</i>
--------------------	------------------------

Description

Displays the average silhouette width for different numbers of clusters to help identify the optimal clustering solution.

Usage

```
cluster_silhouette(data)
```

Arguments

data	A data.frame.
------	---------------

Value

A ggplot object.

correlation_analysis	<i>Correlation Analysis</i>
----------------------	-----------------------------

Description

Computes the Pearson correlation matrix for numeric variables.

Usage

```
correlation_analysis(data, method = "pearson")
```

Arguments

data	A data.frame.
method	Correlation method ("pearson", "spearman", or "kendall").

Value

An object of class CorrelationMatrix.

correlation_network	<i>Correlation Network</i>
---------------------	----------------------------

Description

Builds a correlation network from numeric variables.

Usage

```
correlation_network(data, cutoff = 0.7)
```

Arguments

data	A data.frame.
cutoff	Minimum absolute correlation.

Value

An igraph object.

correlation_plot	<i>Correlation Heatmap</i>
------------------	----------------------------

Description

Displays a correlation heatmap.

Usage

```
correlation_plot(data, method = "pearson")
```

Arguments

data	A data.frame.
method	Correlation method.

Value

A ggplot object.

correlation_test	<i>Pairwise Correlation Tests</i>
------------------	-----------------------------------

Description

Computes pairwise Pearson correlation coefficients, p-values, and confidence intervals for all numeric variables in a dataset.

Usage

```
correlation_test(data)
```

Arguments

data	A data.frame.
------	---------------

Value

An object of class CorrelationTest.

create_report	<i>Create AutoEDA Report</i>
---------------	------------------------------

Description

Creates a comprehensive AutoEDA report.

Usage

```
create_report(data)
```

Arguments

data	A data.frame.
------	---------------

Value

An object of class AutoEDAReport.

density_plot	<i>Density Plot</i>
--------------	---------------------

Description

Draws a density plot for a numeric variable.

Usage

```
density_plot(data, variable)
```

Arguments

data	A data.frame.
variable	Name of the numeric variable.

Value

A ggplot object.

export_excel	<i>Export Report to Excel</i>
--------------	-------------------------------

Description

Exports an AutoEDA report to an Excel workbook.

Usage

```
export_excel(report, file = "AutoEDA_Report.xlsx")
```

Arguments

report	An AutoEDAReport object returned by auto_eda() or create_report().
file	Output Excel filename.

Value

Invisibly returns the normalized output filename.

`export_numeric_summary`*Export Numeric Summary*

Description

Exports a numeric summary table to a CSV file.

Usage

```
export_numeric_summary(x, file)
```

Arguments

<code>x</code>	An object of class <code>NumericSummary</code> .
<code>file</code>	Character string specifying the output CSV file path. This argument must be supplied explicitly.

Value

Invisibly returns the normalized output file path.

`histogram_plot`*Histogram Plot*

Description

Draws a histogram for a numeric variable.

Usage

```
histogram_plot(data, variable)
```

Arguments

<code>data</code>	A <code>data.frame</code> .
<code>variable</code>	Name of the numeric variable.

Value

A `ggplot` object.

html_report	<i>Export HTML Report</i>
-------------	---------------------------

Description

Creates an HTML report from an AutoEDAReport object.

Usage

```
html_report(report, file = "AutoEDA_Report.html")
```

Arguments

report	An AutoEDAReport object.
file	Output HTML filename.

Value

Invisibly returns the output filename.

missing_heatmap	<i>Missing Values Heatmap</i>
-----------------	-------------------------------

Description

Displays a heatmap showing the pattern of missing values in a dataset.

Usage

```
missing_heatmap(data)
```

Arguments

data	A data.frame.
------	---------------

Value

A ggplot object.

missing_pattern	<i>Missing Value Pattern</i>
-----------------	------------------------------

Description

Displays the pattern of missing values in a dataset.

Usage

```
missing_pattern(data)
```

Arguments

data	A data.frame.
------	---------------

Details

This function is a wrapper around `mice::md.pattern()` and displays the missing-data pattern.

Value

Invisibly returns the missing-value pattern matrix produced by `mice::md.pattern()`.

Examples

```
missing_pattern(iris)

dat <- iris
dat$Sepal.Length[1:10] <- NA
missing_pattern(dat)
```

missing_summary	<i>Missing Value Summary</i>
-----------------	------------------------------

Description

Summarizes missing values in a dataset.

Usage

```
missing_summary(data)
```

Arguments

data	A data.frame.
------	---------------

Value

An object of class MissingSummary.

numeric_summary	<i>Numeric Summary</i>
-----------------	------------------------

Description

Generates descriptive statistics for all numeric variables.

Usage

```
numeric_summary(data)
```

Arguments

data	A data.frame.
------	---------------

Value

An object of class NumericSummary.

outlier_detection	<i>Outlier Detection</i>
-------------------	--------------------------

Description

Detects outliers in all numeric variables.

Usage

```
outlier_detection(data, method = "IQR", threshold = 3)
```

Arguments

data	A data.frame.
method	Outlier detection method. One of "IQR", "ZScore", or "MAD".
threshold	Numeric threshold used for ZScore and MAD methods.

Value

An object of class OutlierSummary.

outlier_summary	<i>Outlier Summary</i>
-----------------	------------------------

Description

Detects and summarizes outliers.

Usage

```
outlier_summary(data, method = "IQR", threshold = 3)
```

Arguments

- data A data.frame.
- method Method used for outlier detection ("iqr" or "zscore").
- threshold Threshold for outlier detection.

Value

A data.frame summarizing detected outliers.

outlier_values	<i>Outlier Values</i>
----------------	-----------------------

Description

Returns all observations identified as outliers for a specified numeric variable using the $1.5 \times \text{IQR}$ rule.

Usage

```
outlier_values(data, variable)
```

Arguments

- data A data.frame.
- variable Name of a numeric variable.

Value

A data.frame containing the outlier observations.

`pair_plot`*Pair Plot*

Description

Creates a scatterplot matrix for all numeric variables.

Usage

```
pair_plot(data)
```

Arguments

`data` A data.frame.

Value

A GGally ggmatrix object.

`pca_analysis`*Principal Component Analysis*

Description

Performs Principal Component Analysis (PCA) on numeric variables.

Usage

```
pca_analysis(data, scale = TRUE)
```

Arguments

`data` A data.frame.
`scale` Logical; should variables be scaled?

Value

An object of class `PCAResult`.

`pca_biplot`*PCA Biplot*

Description

Displays a PCA biplot showing both variables and observations.

Usage

```
pca_biplot(pca)
```

Arguments

`pca` A PCAResult object.

Value

A ggplot object.

`pca_individual_plot`*PCA Individual Plot*

Description

Displays the observations in PCA space.

Usage

```
pca_individual_plot(pca)
```

Arguments

`pca` A PCAResult object.

Value

A ggplot object.

pca_scree_plot	<i>Scree Plot</i>
----------------	-------------------

Description

Draws a scree plot showing the percentage of variance explained by each principal component.

Usage

```
pca_scree_plot(pca)
```

Arguments

pca	A PCAResult object.
-----	---------------------

Value

A ggplot object.

pca_variable_plot	<i>PCA Variable Plot</i>
-------------------	--------------------------

Description

Displays variable contributions to the principal components.

Usage

```
pca_variable_plot(pca)
```

Arguments

pca	A PCAResult object.
-----	---------------------

Value

A ggplot object.

pdf_report	<i>Export PDF Report</i>
------------	--------------------------

Description

Creates a PDF report from an AutoEDAReport object.

Usage

```
pdf_report(report, file = "AutoEDA_Report.pdf")
```

Arguments

report	An AutoEDAReport object.
file	Output PDF filename.

Value

Invisibly returns the output filename.

plot_missing	<i>Missing Values Plot</i>
--------------	----------------------------

Description

Visualizes the number of missing values in each variable.

Usage

```
plot_missing(data)
```

Arguments

data	A data.frame.
------	---------------

Value

A ggplot object.

plot_outliers	<i>Plot Outliers</i>
---------------	----------------------

Description

Creates boxplots for all numeric variables to visualize potential outliers.

Usage

```
plot_outliers(data)
```

Arguments

data	A data.frame.
------	---------------

Value

A ggplot object.

print.AutoEDAResult	<i>Print AutoEDA Report</i>
---------------------	-----------------------------

Description

Print AutoEDA Report

Usage

```
## S3 method for class 'AutoEDAResult'  
print(x, ...)
```

Arguments

x	AutoEDAResult object.
...	Additional arguments.

Value

Invisibly returns the AutoEDAResult object.

```
print.CategoricalSummary
```

Print Categorical Summary

Description

Prints the categorical summary generated by AutoEDA.

Usage

```
## S3 method for class 'CategoricalSummary'  
print(x, ...)
```

Arguments

x	A CategoricalSummary object.
...	Additional arguments passed to print.data.frame().

Value

Invisibly returns the CategoricalSummary object.

```
print.ClusterResult
```

Print Cluster Analysis Result

Description

Prints a summary of K-means clustering results.

Usage

```
## S3 method for class 'ClusterResult'  
print(x, ...)
```

Arguments

x	A ClusterResult object.
...	Additional arguments (unused).

Value

Invisibly returns the ClusterResult object.

```
print.CorrelationMatrix
```

Print Correlation Matrix

Description

Prints a correlation matrix.

Usage

```
## S3 method for class 'CorrelationMatrix'  
print(x, ...)
```

Arguments

x	A CorrelationMatrix object.
...	Additional arguments (unused).

Value

Invisibly returns the CorrelationMatrix object.

```
print.MissingSummary
```

Print Missing Summary

Description

Prints the missing value summary generated by AutoEDA.

Usage

```
## S3 method for class 'MissingSummary'  
print(x, ...)
```

Arguments

x	A MissingSummary object.
...	Additional arguments (unused).

Value

Invisibly returns the MissingSummary object.

print.NumericSummary *Print Numeric Summary*

Description

Prints the numeric summary generated by AutoEDA.

Usage

```
## S3 method for class 'NumericSummary'  
print(x, ...)
```

Arguments

x	A NumericSummary object.
...	Additional arguments (unused).

Value

Invisibly returns the NumericSummary object.

print.OutlierSummary *Print Outlier Summary*

Description

Prints the outlier summary generated by AutoEDA.

Usage

```
## S3 method for class 'OutlierSummary'  
print(x, ...)
```

Arguments

x	An OutlierSummary object.
...	Additional arguments (unused).

Value

Invisibly returns the OutlierSummary object.

print.PCAResult	<i>Print PCA Result</i>
-----------------	-------------------------

Description

Prints a summary of a principal component analysis.

Usage

```
## S3 method for class 'PCAResult'  
print(x, ...)
```

Arguments

x	A PCAResult object.
...	Additional arguments (unused).

Value

Invisibly returns the PCAResult object.

scatter_plot	<i>Scatter Plot</i>
--------------	---------------------

Description

Draws a scatter plot for two numeric variables.

Usage

```
scatter_plot(data, x, y)
```

Arguments

data	A data.frame.
x	Name of the X variable.
y	Name of the Y variable.

Value

A ggplot object.

summary_data	<i>Dataset Summary</i>
--------------	------------------------

Description

Provides a quick overview of a dataset.

Usage

```
summary_data(data)
```

Arguments

data	A data.frame.
------	---------------

Value

A list containing dataset information.

theme_autoeda	<i>AutoEDA Theme</i>
---------------	----------------------

Description

A consistent ggplot2 theme used throughout AutoEDA.

Usage

```
theme_autoeda()
```

Value

A ggplot2 theme object.

word_report

*Export Word Report***Description**

Creates a Word report from an AutoEDAReport object.

Usage

```
word_report(report, file = "AutoEDA_Report.docx")
```

Arguments

report	An AutoEDAReport object.
file	Output Word filename.

Value

Invisibly returns the output filename.

Index

auto_eda, [3](#)
auto_plot, [3](#)

bar_plot, [4](#)
box_plot, [4](#)

categorical_plot, [5](#)
categorical_summary, [5](#)
cluster_analysis, [6](#)
cluster_dendrogram, [6](#)
cluster_elbow, [7](#)
cluster_plot, [7](#)
cluster_silhouette, [8](#)
correlation_analysis, [8](#)
correlation_network, [9](#)
correlation_plot, [9](#)
correlation_test, [10](#)
create_report, [10](#)

density_plot, [11](#)

export_excel, [11](#)
export_numeric_summary, [12](#)

histogram_plot, [12](#)
html_report, [13](#)

missing_heatmap, [13](#)
missing_pattern, [14](#)
missing_summary, [14](#)

numeric_summary, [15](#)

outlier_detection, [15](#)
outlier_summary, [16](#)
outlier_values, [16](#)

pair_plot, [17](#)
pca_analysis, [17](#)
pca_biplot, [18](#)
pca_individual_plot, [18](#)

pca_scree_plot, [19](#)
pca_variable_plot, [19](#)
pdf_report, [20](#)
plot_missing, [20](#)
plot_outliers, [21](#)
print.AutoEDAResult, [21](#)
print.CategoricalSummary, [22](#)
print.ClusterResult, [22](#)
print.CorrelationMatrix, [23](#)
print.MissingSummary, [23](#)
print.NumericSummary, [24](#)
print.OutlierSummary, [24](#)
print.PCAResult, [25](#)

scatter_plot, [25](#)
summary_data, [26](#)

theme_autoeda, [26](#)

word_report, [27](#)