

Package ‘MixedDiversityAnalysis’

October 29, 2026

Title Analysis and Visualization of Mixed-Trait Dissimilarity

Version 0.1.0

Description Provides methods for analysing datasets containing quantitative, qualitative, binary, and ordinal traits. Computes mixed-trait dissimilarities and performs hierarchical clustering using multiple linkage methods. Supports principal coordinates analysis (PCoA), cluster evaluation, trait contribution analysis, and graphical visualization. The package facilitates mixed-trait diversity assessment and characterization of crop germplasm. For more detail see Gower (1971) <[doi:10.2307/2528823](https://doi.org/10.2307/2528823)> and Jukanti et al. (2026) <[doi:10.1007/s44187-025-00715-w](https://doi.org/10.1007/s44187-025-00715-w)>.

License GPL (>= 3)

Encoding UTF-8

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

Imports cluster

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Santosha Rathod [aut],
Prabhat Kumar [aut, cre],
A. K. Jukanti [aut],
Radha Upadhyay [aut],
Prashantkumar S Hanjagi [aut],
Basavaraj PS [aut],
R. N. Singh [aut],
Anil Kumar [aut]

Maintainer Prabhat Kumar <prabhatkkr@gmail.com>

Repository CRAN

Date/Publication 2026-10-29 11:10:07 UTC

Contents

<code>mixed_cluster</code>	2
<code>mixed_distance</code>	4
<code>mixed_diversity</code>	6
<code>mixed_pcoa</code>	8
<code>mixed_visualize</code>	9

Index	13
--------------	--------------------

<code>mixed_cluster</code>	<i>Perform Hierarchical Clustering on Mixed-Trait Data</i>
----------------------------	--

Description

Performs hierarchical clustering using a mixed-distance object created by `mixed_distance()`.

Usage

```
mixed_cluster(  
  distance_object,  
  k = "optimal",  
  method = "average",  
  k_range = NULL,  
  min_cluster_size = 1  
)
```

Arguments

<code>distance_object</code>	An object of class "mixed_distance".
<code>k</code>	Number of clusters, or "optimal" to select the number using average silhouette width.
<code>method</code>	Hierarchical linkage method: "average", "complete", "single", or "mcquitty".
<code>k_range</code>	Integer vector of candidate cluster numbers when k = "optimal". If NULL, a suitable range is generated.
<code>min_cluster_size</code>	Minimum permitted cluster size during automatic selection. Defaults to 1.

Value

An object of class "mixed_cluster" containing the clustering tree, cluster assignments, cluster sizes, selected number of clusters, linkage method, silhouette results, comparison of candidate solutions, and minimum cluster size.

Examples

```

# Prepare the same rice dataset used in mixed_distance()
rice_data <- data.frame(
  PlantHeight = c(
    90, 100, 110, 95, 85, 105, 115, 92, 98, 108,
    88, 102, 112, 97, 83, 107, 118, 93, 101, 99
  ),
  GrainYield = c(
    25, 30, 35, 28, 22, 32, 38, 26, 29, 34,
    24, 31, 36, 27, 21, 33, 40, 25, 30, 28
  ),
  GrainType = factor(c(
    "Long", "Long", "Bold", "Bold", "Medium",
    "Long", "Bold", "Medium", "Long", "Bold",
    "Medium", "Long", "Bold", "Medium", "Bold",
    "Long", "Bold", "Medium", "Long", "Bold"
  )),
  DiseaseResponse = factor(c(
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Resistant", "Susceptible", "Resistant",
    "Susceptible", "Resistant", "Susceptible", "Resistant",
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Susceptible", "Resistant", "Susceptible"
  )),
  Maturity = ordered(
    c(
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Late",
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Medium"
    ),
    levels = c("Early", "Medium", "Late")
  )
)

rownames(rice_data) <- paste0("Rice_", sprintf("%02d", 1:20))

trait_types <- c(
  "quantitative", "quantitative", "qualitative",
  "binary", "ordinal"
)

# Step 1: Calculate dissimilarities
dist_result <- mixed_distance(rice_data, trait_types)

# Step 2: Cluster the rice genotypes
cluster_result <- mixed_cluster(
  dist_result,
  k = 3,
  method = "average"
)

```

```
# Step 3: Inspect cluster assignments
print(cluster_result$clusters)
```

mixed_distance	<i>Calculate Mixed-Trait Dissimilarity</i>
----------------	--

Description

Calculates pairwise dissimilarities for datasets containing quantitative, qualitative, binary, and ordinal traits.

Usage

```
mixed_distance(data, types)
```

Arguments

data	A data frame with observations in rows and traits in columns.
types	A character vector specifying the type of each trait: "quantitative", "qualitative", "binary", or "ordinal".

Details

This function implements a custom mixed-trait dissimilarity method and does not use Gower dissimilarity.

Quantitative differences are divided by the observed range. Qualitative and binary traits receive a dissimilarity of zero for matching values and one for different values. Binary traits are treated symmetrically.

Ordinal traits must be ordered factors. Differences between their integer ranks are divided by the number of rank intervals.

For each pair, the overall dissimilarity is the mean of the available trait-specific dissimilarities. Missing comparisons are omitted. If no trait can be compared for a pair, its dissimilarity is NA. The diagonal is always zero.

Value

An object of class "mixed_distance", containing:

- dissimilarity** A pairwise dissimilarity matrix.
- data** The input data used in the calculation.
- types** The specified type of each trait.
- method** The name of the dissimilarity method.

Examples

```

# Create an expanded dataset with mixed trait types
rice_data <- data.frame(
  PlantHeight = c(
    90, 100, 110, 95, 85,
    105, 115, 92, 98, 108,
    88, 102, 112, 97, 83,
    107, 118, 93, 101, 99
  ),
  GrainYield = c(
    25, 30, 35, 28, 22,
    32, 38, 26, 29, 34,
    24, 31, 36, 27, 21,
    33, 40, 25, 30, 28
  ),
  GrainType = factor(c(
    "Long", "Long", "Bold", "Bold", "Medium",
    "Long", "Bold", "Medium", "Long", "Bold",
    "Medium", "Long", "Bold", "Medium", "Bold",
    "Long", "Bold", "Medium", "Long", "Bold"
  )),
  DiseaseResponse = factor(c(
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Resistant", "Susceptible", "Resistant",
    "Susceptible", "Resistant", "Susceptible", "Resistant",
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Susceptible", "Resistant", "Susceptible"
  )),
  Maturity = ordered(
    c(
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Late",
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Medium"
    ),
    levels = c("Early", "Medium", "Late")
  )
)

# Assign genotype names
rownames(rice_data) <- paste0("Rice_", sprintf("%02d", 1:20))

# Specify the type of each trait in column order
trait_types <- c(
  "quantitative",
  "quantitative",
  "qualitative",
  "binary",
  "ordinal"
)

# Calculate mixed-trait dissimilarities

```

```
result <- mixed_distance(rice_data, trait_types)

# Display the dissimilarity matrix
round(result$dissimilarity, 3)
```

mixed_diversity	<i>Measure Trait Contributions to Cluster Separation</i>
-----------------	--

Description

Quantifies how strongly individual traits distinguish clusters using separation scores appropriate for different trait types.

Usage

```
mixed_diversity(data, types, clusters)
```

Arguments

data	A data frame containing the traits to be evaluated.
types	A character vector specifying the type of each trait. Supported types are "quantitative", "qualitative", "binary", and "ordinal".
clusters	A vector assigning each observation to a cluster. Its length must equal the number of rows in data, and values must not be missing.

Details

For quantitative and ordinal traits, separation is measured as the proportion of total variation attributable to differences between clusters, using the ratio of between-cluster sum of squares to total sum of squares.

For qualitative and binary traits, separation is measured using squared Cramer's V, calculated from the Pearson chi-squared statistic. This score describes the association between trait categories and cluster membership.

Scores range from zero to one, with larger values indicating stronger observed separation among clusters. These scores are descriptive measures of association and do not establish causality or statistical significance.

Value

A data frame with three columns:

trait The name of each trait.

type The specified trait type.

separation A descriptive separation score between zero and one, or NA when a score cannot be calculated for an entirely missing trait.

Traits are ordered by decreasing separation score.

Examples

```

# Prepare a dataset of 20 rice genotypes
rice_data <- data.frame(
  PlantHeight = c(
    90, 100, 110, 95, 85, 105, 115, 92, 98, 108,
    88, 102, 112, 97, 83, 107, 118, 93, 101, 99
  ),
  GrainYield = c(
    25, 30, 35, 28, 22, 32, 38, 26, 29, 34,
    24, 31, 36, 27, 21, 33, 40, 25, 30, 28
  ),
  GrainType = factor(c(
    "Long", "Long", "Bold", "Bold", "Medium",
    "Long", "Bold", "Medium", "Long", "Bold",
    "Medium", "Long", "Bold", "Medium", "Bold",
    "Long", "Bold", "Medium", "Long", "Bold"
  )),
  DiseaseResponse = factor(c(
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Resistant", "Susceptible", "Resistant",
    "Susceptible", "Resistant", "Susceptible", "Resistant",
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Susceptible", "Resistant", "Susceptible"
  )),
  Maturity = ordered(
    c(
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Late",
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Medium"
    ),
    levels = c("Early", "Medium", "Late")
  )
)

rownames(rice_data) <- paste0("Rice_", sprintf("%02d", 1:20))

# Define the type of each trait
trait_types <- c(
  "quantitative", "quantitative", "qualitative",
  "binary", "ordinal"
)

# Calculate mixed-trait dissimilarities
dist_result <- mixed_distance(rice_data, trait_types)

# Identify three clusters
cluster_result <- mixed_cluster(
  dist_result,
  k = 3,
  method = "average"
)

```

```
# Measure the separation contributed by each trait
trait_result <- mixed_diversity(
  data = rice_data,
  types = trait_types,
  clusters = cluster_result$clusters
)

# Display trait separation scores
print(trait_result)
```

mixed_pcoa

Perform Principal Coordinates Analysis

Description

Performs Principal Coordinates Analysis (PCoA) using a mixed-trait dissimilarity object.

Usage

```
mixed_pcoa(distance_object, dimensions = 2)
```

Arguments

distance_object	An object created by mixed_distance().
dimensions	Number of coordinate dimensions to return.

Value

An object of class "mixed_pcoa" containing the coordinates, eigenvalues, and percentage of variation explained.

Examples

```
# Prepare the same rice dataset used in mixed_distance()
rice_data <- data.frame(
  PlantHeight = c(
    90, 100, 110, 95, 85, 105, 115, 92, 98, 108,
    88, 102, 112, 97, 83, 107, 118, 93, 101, 99
  ),
  GrainYield = c(
    25, 30, 35, 28, 22, 32, 38, 26, 29, 34,
    24, 31, 36, 27, 21, 33, 40, 25, 30, 28
  ),
  GrainType = factor(c(
    "Long", "Long", "Bold", "Bold", "Medium",
    "Long", "Bold", "Medium", "Long", "Bold",
    "Medium", "Long", "Bold", "Medium", "Bold",

```

```

      "Long", "Bold", "Medium", "Long", "Bold"
    )),
  DiseaseResponse = factor(c(
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Resistant", "Susceptible", "Resistant",
    "Susceptible", "Resistant", "Susceptible", "Resistant",
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Susceptible", "Resistant", "Susceptible"
  )),
  Maturity = ordered(
    c(
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Late",
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Medium"
    ),
    levels = c("Early", "Medium", "Late")
  )
)

# Assign genotype names
rownames(rice_data) <- paste0("Rice_", sprintf("%02d", 1:20))

# Specify trait types in column order
trait_types <- c(
  "quantitative", "quantitative", "qualitative",
  "binary", "ordinal"
)

# Step 1: Calculate mixed-trait dissimilarities
dist_result <- mixed_distance(rice_data, trait_types)

# Step 2: Perform Principal Coordinates Analysis
pcoa_result <- mixed_pcoa(
  dist_result,
  dimensions = 2
)

# Step 3: Inspect the PCoA results
print(pcoa_result)

# View the genotype coordinates
head(pcoa_result$coordinates)

# View the percentage of variation explained
pcoa_result$percent_explained

```

Description

Creates graphical summaries of hierarchical clustering, silhouette widths, principal coordinates analysis (PCoA), and trait contributions to cluster separation.

Usage

```
mixed_visualize(
  object,
  type = c("dendrogram", "silhouette", "pcoa", "trait"),
  clusters = NULL,
  labels = NULL,
  main = NULL,
  ellipse = FALSE,
  ...
)
```

Arguments

object	An object returned by <code>mixed_cluster()</code> , <code>mixed_pcoa()</code> , or <code>mixed_diversity()</code> , depending on type.
type	Character string specifying the plot type: "dendrogram", "silhouette", "pcoa", or "trait".
clusters	Optional vector assigning observations to clusters. Used to colour PCoA points and optionally draw group ellipses.
labels	Optional vector of observation labels for the PCoA plot. Must contain one non-missing value per observation.
main	Optional character string specifying the plot title.
ellipse	Logical; if TRUE, attempts to draw 95% group ellipses on the PCoA plot when cluster assignments are supplied. Ellipses are omitted for groups with insufficient or degenerate coordinate data.
...	Additional graphical parameters passed to the relevant base plotting function.

Details

The "dendrogram" option displays the hierarchical clustering tree and highlights the resulting clusters. The "silhouette" option displays observation-level silhouette widths and the mean silhouette width. The "pcoa" option plots the first two principal coordinate dimensions and can colour observations by cluster. Optional ellipses summarize the spread of groups in the plotted coordinate space. The "trait" option displays trait separation scores calculated by `mixed_diversity()`.

Plot appearance uses base R graphics. The function restores the previous graphical parameters on exit.

Value

Invisibly returns the supplied object. The function is called primarily for its graphical side effect.

Examples

```

# Prepare a dataset of 20 rice genotypes
rice_data <- data.frame(
  PlantHeight = c(
    90, 100, 110, 95, 85, 105, 115, 92, 98, 108,
    88, 102, 112, 97, 83, 107, 118, 93, 101, 99
  ),
  GrainYield = c(
    25, 30, 35, 28, 22, 32, 38, 26, 29, 34,
    24, 31, 36, 27, 21, 33, 40, 25, 30, 28
  ),
  GrainType = factor(c(
    "Long", "Long", "Bold", "Bold", "Medium",
    "Long", "Bold", "Medium", "Long", "Bold",
    "Medium", "Long", "Bold", "Medium", "Bold",
    "Long", "Bold", "Medium", "Long", "Bold"
  )),
  DiseaseResponse = factor(c(
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Resistant", "Susceptible", "Resistant",
    "Susceptible", "Resistant", "Susceptible", "Resistant",
    "Resistant", "Susceptible", "Resistant", "Susceptible",
    "Resistant", "Susceptible", "Resistant", "Susceptible"
  )),
  Maturity = ordered(
    c(
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Late",
      "Early", "Medium", "Late", "Medium", "Early",
      "Late", "Late", "Early", "Medium", "Medium"
    ),
    levels = c("Early", "Medium", "Late")
  )
)

rownames(rice_data) <- paste0("Rice_", sprintf("%02d", 1:20))

# Specify trait types
trait_types <- c(
  "quantitative", "quantitative", "qualitative",
  "binary", "ordinal"
)

# Calculate mixed-trait dissimilarities
dist_result <- mixed_distance(rice_data, trait_types)

# Perform hierarchical clustering
cluster_result <- mixed_cluster(
  dist_result,
  k = 3,
  method = "average"
)

```

```
# Plot the dendrogram
mixed_visualize(
  object = cluster_result,
  type = "dendrogram",
  main = "Rice Genotype Clustering"
)

# Plot the PCoA ordination
pcoa_result <- mixed_pcoa(dist_result, dimensions = 2)
mixed_visualize(
  object = pcoa_result,
  type = "pcoa",
  clusters = cluster_result$clusters,
  labels = rownames(rice_data),
  main = "PCoA of Rice Genotypes"
)

# Plot trait separation scores
trait_result <- mixed_diversity(
  data = rice_data,
  types = trait_types,
  clusters = cluster_result$clusters
)
mixed_visualize(
  object = trait_result,
  type = "trait",
  main = "Trait Contributions to Cluster Separation"
)
```

Index

`mixed_cluster`, [2](#)
`mixed_distance`, [4](#)
`mixed_diversity`, [6](#)
`mixed_pcoa`, [8](#)
`mixed_visualize`, [9](#)