

## Package 'nohac'

---

Type: Package  
Title: Nominal Hierarchical Agglomerative Clustering  
Version: 1.0.0  
Maintainer: Angga Dwi Mulyanto <angga.dwi.m@mat.uin-malang.ac.id>  
Description: Implements Hierarchical Clustering specifically designed for handling nominal data, using Hamming distance as a measure of dissimilarity.  
License: [GPL-3](#)  
Encoding: UTF-8  
LazyData: true  
RoxygenNote: 7.3.1  
Author: Sri Harini [aut],  
Angga Dwi Mulyanto [aut, cre],  
Hairur Rahman [aut],  
Fachrur Rozi [aut],  
Universitas Islam Negeri Maulana Malik Ibrahim Malang [cph, fnd]  
Built: R 4.4.1; ; 2024-11-10 01:48:13 UTC; windows

## Table of Contents

Compute Hamming Distance Matrix .....	3
Calculate Hamming Distance Between Two Vectors.....	4
Perform Hierarchical Agglomerative Clustering Using Hamming Distance .....	5

# Compute Hamming Distance Matrix

## Description

Calculates a distance matrix for a dataset using the Hamming distance, where each element (i, j) in the matrix represents the distance between row i and row j of the input data.

## Usage

```
compute_hamming_distance_matrix(data)
```

## Arguments

`data` A matrix or data frame of observations (rows) and variables (columns). Data should be suitable for Hamming distance calculation.

## Value

An object of class 'dist' representing the pairwise Hamming distances between all rows in the dataset.

## Examples

[Run examples](#)

```
data <- matrix(sample(0:1, 20, replace = TRUE), nrow = 5)
```

```
distance_matrix <- compute_hamming_distance_matrix(data)
```

# Calculate Hamming Distance Between Two Vectors

## Description

Computes the Hamming distance between two vectors. The Hamming distance is the count of positions at which the corresponding elements are different.

## Usage

```
hamming_distance(vec1, vec2)
```

## Arguments

vec1 First vector for comparison.

vec2 Second vector for comparison, should be of the same length as 'vec1'.

## Value

An integer value representing the Hamming distance between the two vectors.

## Examples

[Run examples](#)

```
vec1 <- c(1, 0, 1, 1, 0)
```

```
vec2 <- c(1, 1, 1, 0, 0)
```

```
hamming_distance(vec1, vec2)
```

# Perform Hierarchical Agglomerative Clustering Using Hamming Distance

## Description

This function performs hierarchical agglomerative clustering on a dataset using Hamming distance as the measure of dissimilarity between observations. It is particularly useful for binary or categorical data. The function computes the Hamming distance matrix, performs hierarchical clustering, and plots the resulting dendrogram.

## Usage

```
nohac(X, method = "ward.D2")
```

## Arguments

- X        A matrix where rows are observations and columns are binary or categorical variables.
- method   The clustering method to be used. Available methods include:
- "ward.D": Minimizes variance within clusters, an older version of "ward.D2".
  - "single": Uses the minimum of the distances between all observations of two clusters.
  - "complete": Uses the maximum of the distances between all observations of two clusters.
  - "average": Uses the average of the distances between all observations of two clusters.
  - "mcquitty": Similar to "average", also known as WPGMA.
  - "median": Uses the median of the distances between all observations of two clusters, known as WPGMC.
  - "centroid": Uses the distance between the centroids of two clusters.
  - "ward.D2": Minimizes the total within-cluster variance, recommended for most cases.

Default is "ward.D2".

## Value

Returns a list with two components: `hamming_distance_matrix`, the computed Hamming distance matrix, and `cluster`, an object of class `hclust` representing the hierarchical clustering model. Note: The dendrogram plot is generated as a side effect and is not directly returned by the function.

## Examples

[Run examples](#)

```
# Generating a sample binary data matrix
```

```
set.seed(123)
```

```
data_sample <- matrix(sample(0:1, 20, replace = TRUE), nrow = 5)
```

```
result <- nohac(data_sample)
```