

# Package ‘ClustAll’

May 1, 2024

**Type** Package

**Title** ClustAll: Data driven strategy to find groups of patients within complex diseases

**Version** 1.1.0

**Imports** FactoMineR, bigstatsr, clValid, doSNOW, parallel, foreach, dplyr, fpc, mice, modeest, flock, networkD3, methods, ComplexHeatmap, cluster, RColorBrewer, circlize, grDevices, ggplot2, grid, stats, utils, pbapply

**Suggests** RUnit, knitr, BiocGenerics, rmarkdown, BiocStyle, roxygen2

**Description** Data driven strategy to find hidden groups of patients with complex diseases using clinical data. ClustAll facilitates the unsupervised identification of multiple robust stratifications. ClustAll, is able to overcome the most common limitations found when dealing with clinical data (missing values, correlated data, mixed data types).

**Depends** R (>= 4.2.0)

**License** GPL-2

**Encoding** UTF-8

**biocViews** Software, StatisticalMethod, Clustering, DimensionReduction, PrincipalComponent

**RoxygenNote** 7.3.1

**VignetteBuilder** knitr

**git\_url** <https://git.bioconductor.org/packages/ClustAll>

**git\_branch** devel

**git\_last\_commit** 86f35b6

**git\_last\_commit\_date** 2024-04-30

**Repository** Bioconductor 3.20

**Date/Publication** 2024-05-01

**Author** Asier Ortega-Legarreta [aut, cre]  
(<<https://orcid.org/0009-0000-3563-5362>>),  
Sara Palomino-Echeverria [aut]

**Maintainer** Asier Ortega-Legarreta <aortegal@navarra.es>

## Contents

addValidationData . . . . .	2
characterOrNA . . . . .	3
ClustAllObject-class . . . . .	4
cluster2data . . . . .	4
createClustAll . . . . .	5
dataImputed . . . . .	6
dataOriginal . . . . .	7
dataValidation . . . . .	8
initialize,ClustAllObject-method . . . . .	8
JACCARD_DISTANCE_F . . . . .	9
listOrNULL . . . . .	10
logicalOrNA . . . . .	11
matrixOrNULL . . . . .	11
nImputation . . . . .	11
numericOrCharacter . . . . .	12
numericOrNA . . . . .	13
obj_noNA1 . . . . .	13
obj_noNA1simplify . . . . .	14
obj_noNA1Validation . . . . .	14
plotJACCARD . . . . .	15
plotSANKEY . . . . .	16
processed . . . . .	17
resStratification . . . . .	17
runClustAll . . . . .	18
show,ClustAllObject-method . . . . .	19
showData . . . . .	20
summary_clusters . . . . .	20
validateStratification . . . . .	21
wdbc . . . . .	22
wdbcMIDS . . . . .	23
wdbcNA . . . . .	24
<b>Index</b>	<b>25</b>

---

addValidationData	<i>Add the validation data into the ClustAllObject</i>
-------------------	--

---

### Description

Generic function to add validation data to the [ClustAllObject-class](#) object

### Usage

```
addValidationData(Object, dataValidation)
```

**Arguments**

Object [ClustAllObject-class](#) object  
dataValidation numericOrCharacter

**Details**

addValidationData

**Value**

[ClustAllObject-class](#) object

**See Also**

[ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
label <- as.numeric(as.factor(wdbc$Diagnosis))
wdbc <- wdbc[,-c(1, 2)] # delete patients IDs & label
obj_noNA <- createClustAll(data = wdbc)
obj_noNA <- addValidationData(Object = obj_noNA,
                             dataValidation = label)
```

---

characterOrNA

*characterOrNA Class union of character, null or missing*

---

**Description**

Contains either character, NULL or missing object

**Value**

characterOrNA class object

---

ClustAllObject-class *ClustAllObject*

---

### Description

Stores the original data used, the imputed datasets and the results of the ClustAll pipeline.

### Value

ClustAllObject class object

### Slots

`data` Data Frame of the data used. Maybe modified from the input data.

`dataOriginal` Data Frame of the original data introduced.

`dataImputed` Mids object derived from the mice package that stores the imputed data, in case imputation was applied. Otherwise NULL.

`dataValidation` labelling numericOrNA. Original data labelling.

`nImputation` Number of multiple imputations to be applied.

`processed` Logical if the ClustAll pipeline has been executed previously

`summary_clusters` listOrNULL. List with the resulting stratifications for each combination of clustering methods (distance + clustering algorithm) and depth, in case ClustAll pipeline has been executed previously. Otherwise NULL.

`JACCARD_DISTANCE_F` matrixOrNULL. Matrix containing the Jaccard distances derived from the robust populations stratifications if ClustAll pipeline has been executed previously. Otherwise NULL.

---

cluster2data

*cluster2data*

---

### Description

Returns the original data in a dataframe, including the selected robust stratification(s) as variables. The representative stratification names can be obtained using the method. [resStratification](#)

### Usage

```
cluster2data(Object,
              stratificationName)
```

### Arguments

`Object` [ClustAllObject-class](#) object

`stratificationName`

Character vector with one or more stratification names

**Value**

data.frame

**See Also**[resStratification, plotJACCARD, ClustAllObject-class](#)**Examples**

```

data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc, select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15, 1:8]
obj_noNA <- createClustAll(data = wdbc)

obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = TRUE)
resStratification(Object = obj_noNA1, population = 0.05,
                  stratification_similarity = 0.88, all = FALSE)
df <- cluster2data(Object = obj_noNA1,
                  stratificationName = c("cuts_a_1", "cuts_b_5", "cuts_a_5"))

```

---

createClustAll	<i>Creates ClustAllObject and perform imputations to deal with missing values</i>
----------------	---

---

**Description**

This pipeline creates the ClustAllObject and computes the imputations if the dataset contains missing values. The next step would be [runClustAll](#)

**Usage**

```

createClustAll(data=data,
              nImputation=NULL,
              dataImputed=NULL,
              colValidation=NULL)

```

**Arguments**

data	Data Frame of the using data. It may contain missing (NA) values.
nImputation	Numeric value with the number of imputations to be computed in case the data contains NAs.
dataImputed	mids object created with mice package. The introduced data for the imputation and the data using must be the same.
colValidation	Character value with the original labelling of the input data.

**Value**

An object of class [ClustAllObject-class](#)

**See Also**

[runClustAll](#), [ClustAllObject-class](#)

**Examples**

```
# Scenario 1: data does not contain missing values
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- wdbc[,-c(1,2)]
obj_noNA <- createClustAll(data = wdbc)

# Scenario 2: data contains NAs and imputed data is provided automatically
data("BreastCancerWisconsinMISSING", package = "ClustAll") # load example data
obj_NA <- createClustAll(wdbcNA, nImputation = 5)

# Scenario 3: data contains NAs and imputed data is provided manually
data("BreastCancerWisconsinMISSING", package = "ClustAll") # load the example data
ini <- mice::mice(wdbcNA, maxit = 0, print = FALSE)
pred <- ini$pred # predictor matrix
pred["radius1", c("perimeter1", "area1", "smoothness1")] <- 0 # example of how to remove predictors
imp <- mice::mice(wdbcNA, m=5, pred=pred, maxit=5, seed=1234, print=FALSE)
obj_imp <- createClustAll(data=wdbcNA, dataImputed = imp)
```

---

dataImputed

*Retrieve the imputed data from ClustAllObject*

---

**Description**

Generic function to retrieve the imputed data obtained in [createClustAll](#) from a [ClustAllObject-class](#) object

**Usage**

```
dataImputed(Object)
```

**Arguments**

Object            [ClustAllObject-class](#) object

**Value**

Mids class object with the imputed data or NULL if imputation was not required

**See Also**

[createClustAll](#), [ClustAllObject-class](#), [runClustAll](#)

**Examples**

```
data("BreastCancerWisconsinMISSING", package = "ClustAll")
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=-ID)
obj_NA <- createClustAll(data = wdbcNA, colValidation = "Diagnosis",
                        dataImputed = wdbcMIDS)
dataImputed(obj_NA)
```

---

dataOriginal	<i>Retrieve the initial dataOriginal from ClustAllObject</i>
--------------	--

---

**Description**

Generic function to retrieve the initial data used for [createClustAll](#) from a [ClustAllObject-class](#) object

**Usage**

```
dataOriginal(Object)
```

**Arguments**

Object            [ClustAllObject-class](#) object

**Value**

The Data Frame with the initial data

**See Also**

[createClustAll](#), [ClustAllObject-class](#), [runClustAll](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=-ID)
obj_noNA <- createClustAll(data = wdbc, colValidation = "Diagnosis")
dataOriginal(obj_noNA)
```

---

dataValidation	<i>Retrieve the original data labelling from ClustAllObject</i>
----------------	---

---

**Description**

Generic function to retrieve numeric vector if it has been added with the true labels from a [ClustAllObject-class](#) object

**Usage**

```
dataValidation(Object)
```

**Arguments**

Object            [ClustAllObject-class](#) object

**Value**

numeric vector if true labels have been added. Otherwise NULL

**See Also**

[ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=-ID)
obj_noNA <- createClustAll(data = wdbc, colValidation="Diagnosis")
dataValidation(obj_noNA)
```

---

initialize,ClustAllObject-method	<i>initializeClustAllObject</i>
----------------------------------	---------------------------------

---

**Description**

constructor for [ClustAllObject-class](#)



**Usage**

```
## S4 method for signature 'ClustAllObject'
initialize(
  .Object,
  data,
  dataOriginal,
  dataImputed,
  dataValidation,
  nImputation,
  processed,
  summary_clusters,
  JACCARD_DISTANCE_F
)
```

**Arguments**

<code>.Object</code>	initializing object
<code>data</code>	Data Frame of the data used. Maybe modified from the input data.
<code>dataOriginal</code>	Data Frame of the original data introduced.
<code>dataImputed</code>	Mids object derived from the mice package that stores the imputed data, in case imputation was applied. Otherwise NULL.
<code>dataValidation</code>	labelling numericOrNA. Original data labelling.
<code>nImputation</code>	Number of multiple imputations to be applied.
<code>processed</code>	Logical if the ClustAll pipeline has been executed previously
<code>summary_clusters</code>	listOrNULL. List with the resulting stratifications for each combination of clustering methods (distance + clustering algorithm) and depth, in case ClustAll pipeline has been executed previously. Otherwise NULL.
<code>JACCARD_DISTANCE_F</code>	matrixOrNULL. Matrix containing the Jaccard distances derived from the robust populations stratifications if ClustAll pipeline has been executed previously. Otherwise NULL.

**Value**

An object of class `ClustAllObject-class`

---

<code>JACCARD_DISTANCE_F</code>	<i>Retrieve the matrix with the Jaccard distances derived from the robust populations stratifications in ClustAllObject</i>
---------------------------------	---

---

**Description**

Generic function to retrieve the matrix with the Jaccard distances derived from the robust populations stratifications in `runClustAll` from a `ClustAllObject-class` object

**Usage**

```
JACCARD_DISTANCE_F(Object)
```

**Arguments**

Object            [ClustAllObject-class](#) object

**Value**

Matrix containing the Jaccard distances derived from the robust populations stratifications or NULL if runClustAll method has not been executed yet

**See Also**

[runClustAll](#), [ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15,1:8]
obj_noNA <- createClustAll(data = wdbc)
obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = FALSE)
JACCARD_DISTANCE_F(obj_noNA1)
```

---

listOrNULL

*Class Union listOrNULL*

---

**Description**

Contains either list, NULL or missing object

**Details**

Class union of list, null or missing

**Value**

listOrNULL class object

---

logicalOrNA	<i>logicalOrNA</i>
-------------	--------------------

---

**Description**

Contains either logical, NULL or missing object

**Details**

Class union of logical, null or missing

**Value**

logicalOrNA class object

---

matrixOrNULL	<i>matrixOrNULL</i>
--------------	---------------------

---

**Description**

Contains either matrix or NULL object

**Details**

Class union of matrix, null or missing

**Value**

matrixOrNULL class object

---

nImputation	<i>Retrieve the number of imputations applied at the imputation step from ClustAllObject</i>
-------------	--

---

**Description**

Generic function to retrieve the number of imputations in `createClustAll` from a `ClustAllObject`-class object

**Usage**

`nImputation(Object)`

**Arguments**

Object [ClustAllObject-class](#) object

**Value**

Numeric vector that contains the number of imputations. 0 in the case of no imputations were required

**See Also**

[createClustAll](#), [ClustAllObject-class](#), [runClustAll](#)

**Examples**

```
data("BreastCancerWisconsinMISSING", package = "ClustAll")
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=-ID)
obj_NA <- createClustAll(data = wdbcNA, colValidation = "Diagnosis",
                        dataImputed = wdbcMIDS)
nImputation(obj_NA)
```

---

numericOrCharacter     *numericOrCharacter*

---

**Description**

Contains either numeric or character object

**Details**

Class union of numeric or character

**Value**

numericOrCharacter class object

---

numericOrNA	<i>Class Union numericOrNA</i>
-------------	--------------------------------

---

**Description**

Contains either numeric, NULL or missing object

**Details**

Class union of numeric, null or missing

**Value**

numericOrNA class object

---

obj_noNA1	<i>obj_noNA1: Processed wdbc dataset for testing purposed</i>
-----------	---

---

**Description**

Processed wdbc as appear in vignette

**Usage**

```
data("testData", package = "ClustAll")
```

**Format**

A processed ClustAllObject

**Value**

ClustAllObject Object

obj\_noNA1simplify      *obj\_noNA1simplify: Processed wdbc dataset for testing purposed*

---

**Description**

Processed wdbc as appear in vignette, with simplify TRUE parameter

**Usage**

```
data("testData", package = "ClustAll")
```

**Format**

A processed ClustAllObject

**Value**

ClustAllObject Object

---

obj\_noNA1Validation      *obj\_noNA1Validation: Processed wdbc dataset for testing purposed*

---

**Description**

Processed wdbc as appear in vignette, with no validation data

**Usage**

```
data("testData", package = "ClustAll")
```

**Format**

A processed ClustAllObject

**Value**

ClustAllObject Object

---

plotJACCARD	<i>Correlation matrix heatmap showing the Jaccard distance between robust stratifications in the ClustAllObject</i>
-------------	---

---

### Description

This function plots the correlation matrix heatmap showing the Jaccard Distance between robust stratifications

### Usage

```
plotJACCARD(Object,  
             paint=TRUE,  
             stratification_similarity=0.7)
```

### Arguments

Object	<a href="#">ClustAllObject-class</a> object
paint	Logical vector with the annotation for the different stratifications
stratification_similarity	The minimum Jaccard Distance value to consider two stratifications similar. Default is 0.7.

### Value

plot

### See Also

[resStratification](#), [cluster2data](#), [ClustAllObject-class](#)

### Examples

```
data("BreastCancerWisconsin", package = "ClustAll")  
wdbc <- subset(wdbc,select=c(-ID, -Diagnosis))  
wdbc <- wdbc[1:15,1:8]  
obj_noNA <- createClustAll(data = wdbc)  
  
obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = TRUE)  
plotJACCARD(obj_noNA1, paint = TRUE, stratification_similarity = 0.9)
```

---

plotSANKEY	<i>Plots Sankey Diagram showing the cluster distribution and shifts between a pair of stratifications derived from ClustAllObject</i>
------------	---

---

### Description

This function plots the Sankey Diagram with the cluster distribution and shifts between a pair of stratifications

### Usage

```
plotSANKEY(Object,
            clusters,
            validationData=FALSE)
```

### Arguments

Object	<a href="#">ClustAllObject-class</a> object
clusters	Character vector with the names of a pair of stratifications. Check <code>resStratification</code> to obtain the stratification names.
validationData	Logical value to use original labelling data to compare with the ClustALL selected stratification.

### Value

plot

### See Also

[resStratification](#), [cluster2data](#), [ClustAllObject-class](#)

### Examples

```
data("BreastCancerWisconsin", package = "ClustAll")
label <- as.numeric(as.factor(wdbc$Diagnosis))
wdbc <- subset(wdbc,select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15,1:8]
label <- label[16:30]
obj_noNA <- createClustAll(data = wdbc)

obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = TRUE)
resStratification(Object = obj_noNA1, population = 0.05,
                  stratification_similarity = 0.88, all = FALSE)
plotSANKEY(Object = obj_noNA1, clusters = c("cuts_a_1","cuts_b_5"))

obj_noNA1 <- addValidationData(obj_noNA1, label)
plotSANKEY(Object = obj_noNA1, clusters = "cuts_a_1", validationData=TRUE)
```



---

processed	<i>Retrieve logical if runClustAll has been executed considering ClustAllObject as input</i>
-----------	--

---

**Description**

Generic function to retrieve the logical if `runClustAll` have been runned from a `ClustAllObject-class` object

**Usage**

```
processed(Object)
```

**Arguments**

Object            `ClustAllObject-class` object

**Value**

TRUE if `runClustAll` has been already executed. Otherwise FALSE

**See Also**

[runClustAll](#), [ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15,1:8]
obj_noNA <- createClustAll(data = wdbc)
processed(obj_noNA)
```

---

resStratification	<i>Show the stratification representatives from the ClustAllObject</i>
-------------------	--

---

**Description**

This function returns the stratifications representatives by keeping those clusters with a minimum percentage of the population. Default is 0.05. It returns all the robust stratification (TRUE) or the representative for each group of stratifications (FALSE). Default is FALSE

**Usage**

```
resStratification(Object,
                   population=0.05,
                   all=FALSE,
                   stratification_similarity=0.7)
```

**Arguments**

Object	<a href="#">ClustAllObject-class</a> object
population	Numeric vector with the minimum percentage of the total population that a stratification must have to be considered as representative
all	Logical vector to return all the representative stratifications per group of clusters. If it is FALSE, only the centroid stratification of each group of clusters is returned
stratification_similarity	The minimum Jaccard distance value to consider two groups similar. Default is 0.7

**Value**

list

**See Also**

[plotJACCARD,cluster2data, ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15,1:8]
obj_noNA <- createClustAll(data = wdbc)

obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = TRUE)
resStratification(Object = obj_noNA1, population = 0.05,
                  stratification_similarity = 0.88, all = FALSE)
```

---

runClustAll

*ClustAll: Data driven strategy to find hidden subgroups of patients within complex diseases using clinical data*

---

**Description**

This method runs the ClustAll pipeline

**Usage**

```
runClustAll(Object,
            threads=1,
            simplify=FALSE)
```

**Arguments**

Object            [ClustAllObject-class](#) object  
threads          Numeric vector that indicates the number of cores to use  
simplify         if TRUE computes one out of four depths of the dendrogram

**Value**

An object of class [ClustAllObject-class](#)

**See Also**

[resStratification](#), [plotJACCARD](#), [cluster2data](#), [ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15,1:8]

obj_noNA <- createClustAll(data = wdbc)
obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = TRUE)
```

---

show,ClustAllObject-method

*show method for ClustAllObject*

---

**Description**

Show method for a [ClustAllObject-class](#) object

**Usage**

```
## S4 method for signature 'ClustAllObject'
show(object)
```

**Arguments**

object            [ClustAllObject-class](#) object

**Value**

summarize information about the object

---

showData	<i>Retrieve the initial data from ClustAllObject</i>
----------	--

---

**Description**

Generic function to retrieve the initial data used for `createClustAll` from a `ClustAllObject-class` object

**Usage**

```
showData(Object)
```

**Arguments**

Object            `ClustAllObject-class` object

**Value**

The Data Frame with the initial data

**See Also**

`createClustAll`, `ClustAllObject-class`, `runClustAll`

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc, select=-ID)
obj_noNA <- createClustAll(data = wdbc, colValidation = "Diagnosis")
showData(obj_noNA)
```

---

summary_clusters	<i>Retrieve the resulting stratifications for each combination of clusterings method (distance + clustering algorithm) and depth from ClustAllObject</i>
------------------	--

---

**Description**

Generic function to retrieve the resulting stratifications for each combination of clusterings method (distance + clustering algorithm) and depth of `runClustAll` from a `ClustAllObject-class` object

**Usage**

```
summary_clusters(Object)
```

**Arguments**

Object [ClustAllObject-class](#) object

**Value**

List with the resulting stratifications for each combination of clusterings method (distance + clustering algorithm) and depth methods or NULL if runClustAll method has not been executed yet.

**See Also**

[runClustAll](#), [ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
wdbc <- subset(wdbc,select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15,1:8]
obj_noNA <- createClustAll(data = wdbc)
obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = FALSE)
summary_clusters(obj_noNA1)
```

---

validateStratification  
*validateStratification*

---

**Description**

Returns the sensitivity and specificity of the selected stratification the original labelling. The representative stratification names can be obtained using the method [resStratification](#)

**Usage**

```
validateStratification(Object,  
                       stratificationName)
```

**Arguments**

Object [ClustAllObject-class](#) object  
stratificationName  
Character vector with the name a stratification. Check [resStratification](#) to obtain stratification names.

**Value**

numeric

**See Also**

[resStratification](#), [plotJACCARD](#), [ClustAllObject-class](#)

**Examples**

```
data("BreastCancerWisconsin", package = "ClustAll")
label <- as.numeric(as.factor(wdbc$Diagnosis))
wdbc <- subset(wdbc, select=c(-ID, -Diagnosis))
wdbc <- wdbc[1:15, 1:8]
label <- label[16:30]
obj_noNA <- createClustAll(data = wdbc)

obj_noNA1 <- runClustAll(Object = obj_noNA, threads = 1, simplify = TRUE)
resStratification(Object = obj_noNA1, population = 0.05,
                  stratification_similarity = 0.88, all = FALSE)
obj_noNA1 <- addValidationData(Object = obj_noNA1,
                              dataValidation = label)
validateStratification(obj_noNA1, "cuts_a_1")
```

---

wdbc

*wdbc: Diagnostic Wisconsin Breast Cancer Database.*

---

**Description**

A dataset containing Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

**Usage**

```
data("BreastCancerWisconsin", package = "ClustAll")
```

**Format**

A data frame with 660 rows and 31 variables

**Details**

The dataset comprises two types of features —categorical and numerical— derived from a digitized image of a fine needle aspirate (FNA) of a breast mass from 659 patients. Each patient is characterized by 31 features (10x3) and belongs to one of two target classes: ‘malignant’ or ‘benign’.

**Value**

wdbc dataset

**Source**

<<https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>>

- Diagnosis Label says tumor is malignant or benignant
- radius. Mean of distances from the center to points on the perimeter
- perimeter
- area
- smoothness. Local variation in radius lengths
- compactness.  $(\text{Perimeter}^2 / \text{Area}) - 1.0$
- concavity. Severity of concave portions of the contour
- concave points. Number of concave portions of the contour
- symmetry.
- fractal dimension. "Coastline approximation" - 1.

---

wdbcMIDS

*wdbcMIDS: Diagnostic Wisconsin Breast Cancer Database with imputed values*

---

**Description**

We introduced imputed random values to the wdbcNA dataset. Using Mice. It is a mids object.  
[wdbc](#)

**Usage**

```
data("BreastCancerWisconsinMISSING", package = "ClustAll")
```

**Format**

A data frame with 660 rows and 31 variables

**Value**

wdbcMIDS dataset

---

wdbcNA	<i>wdbcNA: Diagnostic Wisconsin Breast Cancer Database with missing values</i>
--------	--

---

**Description**

We introduced random missing values to the wdbc dataset. [wdbc](#)

**Usage**

```
data("BreastCancerWisconsinMISSING", package = "ClustAll")
```

**Format**

A data frame with 660 rows and 31 variables

**Value**

wdbcNA dataset



# Index

## \* datasets

- obj\_noNA1, [13](#)
- obj\_noNA1simplify, [14](#)
- obj\_noNAno1Validation, [14](#)
- wdbc, [22](#)
- wdbcMIDS, [23](#)
- wdbcNA, [24](#)

addValidationData, [2](#)

addValidationData, ClustAllObject, numericOrCharacter-method  
(addValidationData), [2](#)

characterOrNA, [3](#)

characterOrNA-class (characterOrNA), [3](#)

ClustAllObject-class, [4](#)

cluster2data, [4](#), [15](#), [16](#), [18](#), [19](#)

cluster2data, ClustAllObject, character-method  
(cluster2data), [4](#)

createClustAll, [5](#), [6](#), [7](#), [11](#), [12](#), [20](#)

createClustAll, data.frame, numericOrNA, ANY, characterOrNA-method  
(createClustAll), [5](#)

dataImputed, [6](#)

dataImputed, ClustAllObject-method  
(dataImputed), [6](#)

dataOriginal, [7](#)

dataOriginal, ClustAllObject-method  
(dataOriginal), [7](#)

dataValidation, [8](#)

dataValidation, ClustAllObject-method  
(dataValidation), [8](#)

initialize, ClustAllObject-method, [8](#)

JACCARD\_DISTANCE\_F, [9](#)

JACCARD\_DISTANCE\_F, ClustAllObject-method  
(JACCARD\_DISTANCE\_F), [9](#)

listOrNULL, [10](#)

listOrNULL-class (listOrNULL), [10](#)

logicalOrNA, [11](#)

logicalOrNA-class (logicalOrNA), [11](#)

matrixOrNULL, [11](#)

matrixOrNULL-class (matrixOrNULL), [11](#)

nImputation, [11](#)

nImputation, ClustAllObject-method  
(nImputation), [11](#)

numericOrCharacter, [12](#)

numericOrCharacter-class  
(numericOrCharacter), [12](#)

numericOrNA, [13](#)

numericOrNA-class (numericOrNA), [13](#)

obj\_noNA1, [13](#)

obj\_noNA1simplify, [14](#)

obj\_noNAno1Validation, [14](#)

plotJACCARD, [5](#), [15](#), [18](#), [19](#), [22](#)

plotJACCARD, ClustAllObject, logicalOrNA, numericOrNA-method  
(plotJACCARD), [15](#)

plotSANKEY, [16](#)

plotSANKEY, ClustAllObject, character, logicalOrNA-method  
(plotSANKEY), [16](#)

processed, [17](#)

processed, ClustAllObject-method  
(processed), [17](#)

resStratification, [4](#), [5](#), [15](#), [16](#), [17](#), [19](#), [21](#),  
[22](#)

resStratification, ClustAllObject, numericOrNA, logicalOrNA, n  
(resStratification), [17](#)

runClustAll, [5-7](#), [9](#), [10](#), [12](#), [17](#), [18](#), [20](#), [21](#)

runClustAll, ClustAllObject, numericOrNA, logicalOrNA-method  
(runClustAll), [18](#)

show, ClustAllObject-method, [19](#)

showData, [20](#)

showData, ClustAllObject-method  
(showData), [20](#)

summary\_clusters, [20](#)

summary\_clusters, ClustAllObject-method  
(summary\_clusters), [20](#)

validateStratification, [21](#)

validateStratification, ClustAllObject, characterOrNA-method  
(validateStratification), [21](#)

wdbc, [22](#), [23](#), [24](#)

wdbcMIDS, [23](#)

wdbcNA, [24](#)