

# Package ‘TDAstats’

November 5, 2018

**Type** Package

**Title** Pipeline for Topological Data Analysis

**Version** 0.4.0

**Description** A comprehensive toolset for any useR conducting topological data analysis, specifically via the calculation of persistent homology in a Vietoris-Rips complex. The tools this package currently provides can be conveniently split into three main sections: (1) calculating persistent homology; (2) conducting statistical inference on persistent homology calculations; (3) visualizing persistent homology and statistical inference. The published form of TDAstats can be found in Wadhwa et al. (2018) <doi:10.21105/joss.00860>. For a general background on computing persistent homology for topological data analysis, see Otter et al. (2017) <doi:10.1140/epjds/s13688-017-0109-5>. To learn more about how the permutation test is used for nonparametric statistical inference in topological data analysis, read Robinson & Turner (2017) <doi:10.1007/s41468-017-0008-7>. To learn more about how TDAstats calculates persistent homology, you can visit the GitHub repository for Ripser, the software that works behind the scenes at <https://github.com/Ripser/riper>. This package has been published as Wadhwa et al. (2018) <doi:10.21105/joss.00860>.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 3.3)

**Imports** ggplot2 (>= 2.2.1), Rcpp (>= 0.12.15)

**URL** <https://github.com/rrrlw/TDAstats>

**BugReports** <https://github.com/rrrlw/TDAstats/issues>

**LinkingTo** Rcpp

**RoxygenNote** 6.1.0

**SystemRequirements** C++11

**Suggests** testthat (>= 2.0.0), knitr, rmarkdown, covr

**VignetteBuilder** knitr

**NeedsCompilation** yes

**Author** Raoul Wadhwa [aut, cre],  
Andrew Dhawan [aut],  
Drew Williamson [aut],  
Jacob Scott [aut]

**Maintainer** Raoul Wadhwa <raoulwadhwa@gmail.com>

**Repository** CRAN

**Date/Publication** 2018-11-05 19:20:11 UTC

## R topics documented:

calculate_homology . . . . .	2
circle2d . . . . .	3
permutation_test . . . . .	4
phom.dist . . . . .	5
plot_barcode . . . . .	5
plot_persist . . . . .	6
sphere3d . . . . .	7
TDAstats . . . . .	8
unif2d . . . . .	8
unif3d . . . . .	8
<b>Index</b>	<b>10</b>

---

calculate\_homology      *Calculate Persistent Homology of a Point Cloud*

---

### Description

Calculates the persistent homology of a point cloud, as represented by a Vietoris-Rips complex. This function is an R wrapper for Ulrich Bauer's Ripser C++ library for calculating persistent homology. For more information on the C++ library, see <<https://github.com/Ripser/ripser>>.

### Usage

```
calculate_homology(mat, dim = 1, threshold = -1, format = "cloud",
  standardize = FALSE)
```

**Arguments**

mat	numeric matrix containing point cloud or distance matrix
dim	maximum dimension of features to calculate
threshold	maximum diameter for computation of Vietoris-Rips complexes
format	format of 'mat', either "cloud" for point cloud or "distmat" for distance matrix
standardize	boolean determining whether point cloud size should be standardized

**Details**

The 'mat' parameter should be a numeric matrix with each row corresponding to a single point, and each column corresponding to a single dimension. Thus, if 'mat' has 50 rows and 5 columns, it represents a point cloud with 50 points in 5 dimensions. The 'dim' parameter should be a positive integer. Alternatively, the 'mat' parameter could be a distance matrix (upper triangular half is ignored); note: 'format' should be specified as "ldm".

**Value**

3-column matrix, with each row representing a TDA feature

**Examples**

```
# create a 2-d point cloud of a circle (100 points)
num.pts <- 100
rand.angle <- runif(num.pts, 0, 2*pi)
pt.cloud <- cbind(cos(rand.angle), sin(rand.angle))

# calculate persistent homology (num.pts by 3 numeric matrix)
pers.hom <- calculate_homology(pt.cloud)
```

---

circle2d

*2-dimensional point cloud of a unit circle*


---

**Description**

A dataset containing the Cartesian coordinates of 100 points uniformly distributed on the circumference of a unit circle.

**Usage**

```
circle2d
```

**Format**

A matrix with 100 rows and 2 columns: the x- and y-coordinates

**Source**

<https://github.com/rrrlw/TDAstats/blob/master/data-raw/circle2d.R>

---

permutation\_test      *Statistical Inference for Topological Data Analysis*

---

### Description

Conducts a permutation test for nonparametric statistical inference of persistent homology in topological data analysis.

### Usage

```
permutation_test(data1, data2, iterations, exponent = 1, update = 0,
  ...)
```

### Arguments

data1	first dataset
data2	second dataset
iterations	number of iterations for distribution in permutation test
exponent	parameter 'p' that returns Wasserstein-p metric
update	if greater than zero, will print a message every 'update' iterations
...	arguments for 'calculate_homology' used for each permutation; this includes the 'format', 'dim', and 'threshold' parameters

### Details

The persistent homology of two point clouds are compared with the Wasserstein metric (where Wasserstein-1 is also known as the Earth Mover's Distance). However, the magnitude of the metric for a single pair of point clouds is meaningless without a reference distribution. This function uses a permutation test (permuting the points between the two clouds) as a nonparametric hypothesis test for statistical inference.

For more details on permutation tests for statistical inference in topological data analysis, see Robinson A, Turner K. Hypothesis testing for topological data analysis. *J Appl Comput Topology*. 2017; 1(2): 241-261.<doi:10.1007/s41468-017-0008-7>

### Value

list containing results of permutation test

---

phom.dist	<i>Calculate Distance between Homology Matrices</i>
-----------	---

---

**Description**

Calculates the distance between two matrices containing persistent homology features, usually as returned by the ‘calculate\_homology’ function.

**Usage**

```
phom.dist(phom1, phom2, limit.num = 0)
```

**Arguments**

phom1	3-by-n numeric matrix containing persistent homology for first dataset
phom2	3-by-n numeric matrix containing persistent homology for second dataset
limit.num	limit comparison to only top ‘limit.num’ features in each dimension

**Details**

Note that the absolute value of this measure of distance is not meaningful without a null distribution or at least another value for relative comparison (e.g. finding most similar pair within a triplet).

**Value**

distance vector (1 element per dimension) between ‘phom1’ and ‘phom2’

---

plot_barcode	<i>Plot Persistent Homology as Topological Barcode</i>
--------------	--

---

**Description**

Plots a feature matrix as a topological barcode. See ‘plot\_persist’ for an alternate visualization method of persistent homology.

**Usage**

```
plot_barcode(feature.matrix)
```

**Arguments**

feature.matrix	nx3 matrix representing persistent homology features
----------------	--

## Details

The 'feature.matrix' parameter should be a numeric matrix with each row corresponding to a single feature. It should have 3 columns corresponding to feature dimension (col 1), feature birth (col 2), and feature death (col 3). The first column should be filled with integers, and the next two columns should be filled with numeric values. The output from the 'calculate\_homology' function in this package will be a valid value for the 'feature.matrix' parameter.

This function uses the ggplot2 framework to generate persistence diagrams. For details, see: Wickham H (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag: New York, NY.

## Value

ggplot instance representing topological barcode

## Examples

```
# create a 2-d point cloud of a circle (100 points)
num.pts <- 100
rand.angle <- runif(num.pts, 0, 2*pi)
pt.cloud <- cbind(cos(rand.angle), sin(rand.angle))

# calculate persistent homology (num.pts by 3 numeric matrix)
pers.hom <- calculate_homology(pt.cloud)

# plot calculated homology features as persistence diagram
plot_barcode(pers.hom)
```

---

plot\_persist

*Plot Persistent Homology as Persistence Diagram*

---

## Description

Plots a feature matrix as a persistence diagram. See 'plot\_barcode' for an alternate visualization method of persistent homology.

## Usage

```
plot_persist(feature.matrix)
```

## Arguments

feature.matrix nx3 matrix representing persistent homology features

## Details

The 'feature.matrix' parameter should be a numeric matrix with each row corresponding to a single feature. It should have 3 columns corresponding to feature dimension (col 1), feature birth (col 2), and feature death (col 3). The first column should be filled with integers, and the next two columns should be filled with numeric values. The output from the 'calculate\_homology' function in this package will be a valid value for the 'feature.matrix' parameter.

This function uses the ggplot2 framework to generate persistence diagrams. For details, see: Wickham H (2009, ISBN:9780387981413). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag: New York, NY.

## Value

ggplot instance representing persistence diagram

## Examples

```
# create a 2-d point cloud of a circle (100 points)
num.pts <- 100
rand.angle <- runif(num.pts, 0, 2*pi)
pt.cloud <- cbind(cos(rand.angle), sin(rand.angle))

# calculate persistent homology (num.pts by 3 numeric matrix)
pers.hom <- calculate_homology(pt.cloud)

# plot calculated homology features as persistence diagram
plot_persist(pers.hom)
```

---

sphere3d

*3-dimensional point cloud of a unit sphere*

---

## Description

A dataset containing the Cartesian coordinates of 100 points uniformly distributed on the surface of a unit sphere.

## Usage

```
sphere3d
```

## Format

A matrix with 100 rows and 3 columns: the x-, y-, and z-coordinates

## Source

<https://github.com/rrrlw/TDAstats/blob/master/data-raw/sphere3d.R>

---

TDAstats	<i>Statistical Inference for Persistent Homology in Topological Data Analysis</i>
----------	---

---

**Description**

This package aims to be a comprehensive toolset for any user conducting topological data analysis, specifically via the calculation of persistent homology in a Vietoris-Rips complex. The tools this package currently provides can be conveniently split into three main sections: (1) calculating persistent homology; (2) conducting statistical inference on persistent homology calculations; (3) visualizing persistent homology and statistical inference.

---

unif2d	<i>2-dimensional point cloud of a unit square</i>
--------	---

---

**Description**

A dataset containing the Cartesian coordinates of 100 points uniformly distributed within a unit square.

**Usage**

```
unif2d
```

**Format**

A matrix with 100 rows and 2 columns: the x- and y-coordinates

**Source**

<https://github.com/rrrlw/TDAstats/blob/master/data-raw/unif2d.R>

---

unif3d	<i>3-dimensional point cloud of a unit cube</i>
--------	---

---

**Description**

A dataset containing the Cartesian coordinates of 100 points uniformly distributed within a unit cube.

**Usage**

```
unif3d
```



**Format**

A matrix with 100 rows and 3 columns: the x-, y-, and z-coordinates

**Source**

<https://github.com/rrrlw/TDAstats/blob/master/data-raw/unif3d.R>

# Index

## \*Topic **datasets**

circle2d, [3](#)

sphere3d, [7](#)

unif2d, [8](#)

unif3d, [8](#)

calculate\_homology, [2](#)

circle2d, [3](#)

permutation\_test, [4](#)

phom.dist, [5](#)

plot\_barcode, [5](#)

plot\_persist, [6](#)

sphere3d, [7](#)

TDAstats, [8](#)

TDAstats-package (TDAstats), [8](#)

unif2d, [8](#)

unif3d, [8](#)