Package 'fds'

October 13, 2022

Type Package

Title Functional Data Sets

Version 1.8

Date 2018-11-1

Depends R (>= 3.4.0), rainbow, RCurl

LazyLoad yes

LazyData yes

LazyDataCompression xz

Author Han Lin Shang and Rob J Hyndman

Maintainer Han Lin Shang <hanlin.shang@anu.edu.au>

Description Functional data sets.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2018-10-31 22:40:03 UTC

R topics documented:

---

fds-package   *Functional data sets*

Description

This package contains a list of functional time series, sliced functional time series, and functional data sets. Functional time series is a special type of functional data observed over time. Sliced functional time series is a special type of functional time series with a time variable observed over time.

Author(s)

Han Lin Shang and Rob J Hyndman

Maintainer: Han Lin Shang <hanlin.shang@anu.edu.au>

References

R. J. Hyndman and H. L. Shang. (2010) "Rainbow plots, bagplots, and boxplots for functional data", *Journal of Computational and Graphical Statistics*, **19**(1), 29-45.

R. J. Hyndman and H. L. Shang (2009) "Forecasting functional time series (with discussion)", *Journal of the Korean Statistical Society*, **38**(3), 199-221.

H. L. Shang and R. J. Hyndman (2009) "Nonparametric time series forecasting with dynamic updating", *Mathematics and Computers in Simulation*, **81**(7), 1310-1324.

Ausmortality   *Australia and Australian state mortality rates*

Description

Age-specific mortality rates for Australia and Australian states.

Format

An object of class `fts`.

Details

The following data sets are included:

ausmale: Australia male log mortality rates (1901-2003).

ausfemale: Australia female log mortality rates (1901-2003).

austotal: Australia total log mortality rates (1901-2003).

nswmale: New South Wales male log mortality rates (1901-2003).

nswfemale: New South Wales female log mortality rates (1901-2003).

nswtotal: New South Wales total log mortality rates (1901-2003).

vicmale: Victoria male log mortality rates (1901-2003).

vicfemale: Victoria female log mortality rates (1901-2003).

victotal: Victoria total log mortality rates (1901-2003).

qldmale: Queensland male log mortality rates (1901-2003).

qldfemale: Queensland female log mortality rates (1901-2003).

qldtotal: Queensland total log mortality rates (1901-2003).

samale: South Australia male log mortality rates (1901-2003).

safemale: South Australia female log mortality rates (1901-2003).

satotal: South Australia total log mortality rates (1901-2003).

wamale: Western Australia male log mortality rates (1901-2003).

wafemale: Western Australia female log mortality rates (1901-2003).

watotal: Western Australia total log mortality rates (1901-2003).

ntmale: Northern Territory male mortality rates (1901-2003).

ntfemale: Northern Territory female mortality rates (1901-2003).

ntotal: Northern Territory total mortality rates (1901-2003).

actmale: Australian Capital Territory male mortality rates (1901-2003).

actfemale: Australian Capital Territory female mortality rates (1901-2003).

actotal: Australian Capital Territory total mortality rates (1901-2003).

tasmale: Tasmania male mortality rates (1901-2003).

tasfemale: Tasmania female mortality rates (1901-2003).

tastotal: Tasmania total mortality rates (1901-2003).

Mortality rates are in logarithm form for Australia, New South Wales, Victoria, Queensland, South Australia, and Western Australia.

Mortality rates without log transformation are: Northern Territory, Australian Captial Territory and Tasmania. These three states have either missing data or zero mortality rates.

All data are from v3.2b of the Australian Demographic Data Bank released 10 February 2005.

Author(s)

Rob J Hyndman

Source

The Australian Demographic Data Bank (courtesy of Len Smith).

References

H. Booth and R. J. Hyndman and L. Tickle and P. De Jong (2006) "Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions", *Demographic Research*, **15**, 289-310.

R. J. Hyndman and M. S. Ullah (2007) "Robust forecasting of mortality and fertility rates: A functional data approach", *Computational Statistics and Data Analysis*, **51**(10), 4942-4956.

R. J. Hyndman and H. Booth (2008) "Stochastic population forecasts using functional data models for mortality, fertility and migration", *International Journal of Forecasting*, **24**(3), 323-342.

R. J. Hyndman and H. Shang (2009) "Functional time series forecasting" (with discussion), *Journal of the Korean Statistical Society*, **38**(3), 199-221.

J-M. Chiou and H-G. Muller (2009) "Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting", *Journal of the American Statistical Association*, **104**(486), 572-585.

Examples

```
plot(victotal)
```

Biscuit	*Biscuit dough piece data*

Description

The experiment involved varying the composition of biscuit dough pieces. Two sets of dough pieces were measured, a calibration set and a prediction set. They were created and measured as two distinct sets, on separate occasions, and do not result from a random (or any other) split of a larger set.

Usage

```
data(labp)
data(labc)
data(nirp)
data(nirc)
```

Format

nirp and nirc are objects of class fds.

labp and labc are objects of class matrix.

Details

The data `labc` (c stands for calibration) and `labp` (p stands for prediction) contain the reference data on the composition of the doughs.

The data `nirc` and `nirp` contain 700 point near infrared reflectance (NIR) spectra for the same dough. The spectral range is 1100-2498 nm in steps of 2nm.

The data `labc$y` is 4 rows by 40 columns, the rows being fat, sucrose, flour and water all in percents. The percents do not quite add up to 100, since there are other minor ingredients present, but they add up to nearly 100 percent.

According to Brown et al. (2001), the observation 23 in the calibration set appears as an outlier.

Sample number 21 in the `labp` shows up as a validation set.

Note

We thank Professor Marina Vannucci for the permission to re-distribute this data set.

References

P. J. Brown and T. Fearn and M. Vannucci (2001) "Bayesian wavelet regression on curves with applications to a spectroscopic calibration problem", *Journal of the American Statistical Association*, **96**(454), pp. 398-408.

Examples

```
plot(nirp)
plot(nirc)
```

Cancerrate *Breast Cancer Data*

Description

Age-specific breast cancer rates for Australian females with 9 age groups (45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+) from 1921 to 2001.

Usage

```
data(Cancerrate)
```

Format

An object of class `fts`.

Source

Australian Institute of Health and Welfare (AIHW) website at https://www.aihw.gov.au/reports-statistics/health-conditions-disability-deaths/cancer/data.

References

B. Erbas and R. J. Hyndman and D. Gertig (2007) "Forecasting age-specific breast cancer mortality using functional data models", *Statistics in Medicine*, **26**(2), 458-470.

R. J. Hyndman and M. S. Ullah (2007) "Robust forecasting of mortality and fertility rates: A functional data approach", *Computational Statistics and Data Analysis*, **51**(10), 4942-4956.

Examples

```
plot(Cancerrate)
```

ECBYieldcurve   *Yield curve data spot rate*

Description

Provided by European Central Bank, this data set contains daily yield curve spot rate from 29/12/2006 to 24/07/2009 for government bond, nominal, all triple AAA issued companies,with maturity term at 3, 6 months and 1 to 30 years.

Usage

```
data(ECBYieldcurve)
```

Format

An object of class `fts`.

Note

We thank Mr Sergio S. Guirreri for the permission to re-distribute this data set.

Source

European Central Bank: <http://www.ecb.europa.eu/stats/money/yc/html/index.en.html>.

Examples

```
plot(ECBYieldcurve)
```

Electricityconsumption

Electricity consumption time series

Description

This set of time series focus on the US monthly electricity consumed by the residential and commercial sectors from January 1973 up to February 2001 (336 months). This data set is a part of the original one which can be found at <http://www.economagic.com>.

Usage

```
data(Electricityconsumption)
```

Format

An object of class sfts.

Details

We eliminated the heteroscedasticity and the linear trend by differencing the log transformed data.

Note

We thank Professor Frederic Ferraty for the permission to re-distribute this data set.

Source

NonParametric Functional Data Analysis website at http://www.math.univ-toulouse.fr/staph/npfda/index.html.

References

F. Ferraty and A. Rabhi and P. Vieu (2005) "Conditional quantiles for dependent functional data with application to the climate El Nino phenomenon", *Sankhya: The Indian Journal of Statistics*, **67**, 378-398.

F. Ferraty and P. Vieu (2007) *Nonparametric functional data analysis*, New York: Springer.

Examples

```
plot(Electricityconsumption)
```

Electricitydemand   *Electricity demand in Adelaide*

Description

These data sets consist of half-hourly electricity demands from Sunday to Saturday in Adelaide between 6/7/1997 and 31/3/2007.

Usage

```
data(mondaydemand)
data(tuesdaydemand)
data(wednesdaydemand)
data(thursdaydemand)
data(fridaydemand)
data(saturdaydemand)
data(sundaydemand)
data(SAelectdemand)
```

Format

An object of class `sfts`.

Details

In Adelaide, the electricity demands in summer are very volatile and highly dependent on their associated temperatures. Analyses were performed to test whether or not, under different temperature scenarios, there will be enough capacity to satisfy the electricity demands.

References

L. Magnano and J. Boland and R. J. Hyndman (2008) "Generation of symthetic sequences of half-hourly temperature", *Environmetrics*, **19**(8), 818-835.

Examples

```
plot(mondaydemand)
plot(tuesdaydemand)
plot(wednesdaydemand)
plot(thursdaydemand)
plot(fridaydemand)
plot(saturdaydemand)
plot(sundaydemand)
plot(SAelectdemand)
```

Fat   *Fat content spectrometric data*

Description

This data set is a part of the original one which can be found at http://lib.stat.cmu.edu/datasets/tecator.

Usage

```
data(Fatspectrum)
data(Fatvalues)
```

Format

Fatspectrum is an object of class fds.

Fatvalues is a numeric object.

Details

For each unit, we observe one spectrometric curve which corresponds to the absorbance measured at 100 wavelengths (from 852 to 1050 in step of 2nm). For each measurement, we have at hand its fat content obtained by an analytic chemical processing.

Note

We thank Professor Frederic Ferraty for the permission to re-distribute this data set.

Source

NonParametric Functional Data Analysis website at http://www.lsp.ups-tlse.fr/staph/npfda/.

References

C. Goutis (1998) "Second-derivative functional regression with applications to near infra-red spectroscopy", *Journal of the Royal Statistical Society: Series B*, **60**(1), 103-114.

F. Ferraty and P. Vieu (2002) "The functional nonparametric model and application to spectrometric data", *Computational Statistics*, **17**(4), 545-564.

F. Ferraty and P. Vieu (2003) "Curve discrimination: A nonparametric functional approach", *Computational Statistics and Data Analysis*, **44**(1-2), 161-173.

F. Ferraty and P. Vieu (2003) "Functional nonparametric statistics: A double infinite dimensional framework", *Recent advances and trends in nonparametric statistics*, Ed M. G. Akritas and D. N. Politis, Amsterdam, The Netherlands, 61-76.

F. Rossi and N. Delannay and B. Conan-Guez and M. Verleysen (2005) "Representation of functional data in neural networks", *Neurocomputing*, **64**, 183-210.

F. Ferraty and P. Vieu (2007) *Nonparametric functional data analysis*, New York: Springer.

H. Matsui and Y. Araki and S. Konishi (2008) "Multivariate regression modeling for functional data", *Journal of Data Science*, **6**, 313-331.

Examples

```
plot(Fatspectrum)
```

FedYieldcurve   *Federal Reserve interest rate*

Description

This data set contains monthly interest rate of the Federal Reserve from January 1982 to June 2009.

Usage

```
data(FedYieldcurve)
```

Format

An object of class `fts`.

Note

We thank Mr Sergio S. Guirreri for the permission to re-distribute this data set.

Source

Board of Governors of the Federal Reserve System at http://www.federalreserve.gov/.

Data set is also available in Excel format at http://www.guirreri.host22.com/web_documents/fedrates.xls.

Examples

```
plot(FedYieldcurve)
```

`hmdcountry`	*Function to read a bundle of data sets from the Human Mortality Database*

Description

This function returns a list of relevant demographic data currently available in the HMD, related to a specified country.

Usage

```
hmdcountry(Country, sex, username, password)
```

Arguments

`Country`	A specified country.
`sex`	Possible options are "Male", "Female", "Total".
`username`	Authenticate username.
`password`	Authenticate password.

Details

In order to read the data sets, users are required to create their account via the HMD website (http://www.mortality.org/), and obtain a valid username and password.

Value

List of objects of class `fts`.

Author(s)

Han Lin Shang and Rob J Hyndman

See Also

read.hmd, hmdstatistic

---

hmdstatistic   *Function to read a bundle of data sets from the Human Mortality Database*

Description

This function returns a list of all the countries currently available in the HMD, related to a specified data type.

Usage

```
hmdstatistic(sex, type = c("birth count", "death count", "population", "exposure",
 "mortality rate", "life expectancy"), username, password)
```

Arguments

sex	Possible options are "Male", "Female", "Total".
type	Type of data.
username	Authenticate username.
password	Authenticate password.

Details

In order to read the data sets, users are required to create their account via the HMD website (http://www.mortality.org/), and obtain a valid username and password.

Value

List of objects of class fts.

Author(s)

Han Lin Shang and Rob J Hyndman

See Also

read.hmd, hmdcountry

Moisture	*Moisture content spectrometric data*

Description

This data set consists of near-infrared reflectance spectra of 100 wheat samples, measured in 2 nm intervals from 1100 to 2500nm, and an associated response variables, the samples' moisture content.

Usage

```
data(Moisturespectrum)
data(Moisturevalues)
```

Format

Moisturespectrum is an object of class `fds`.

Moisturevalues is a numeric object.

Note

We thank Professor John Kalivas for the permission to re-distribute this data set.

References

J. H. Kalivas (1997) "Two data sets of near infrared spectra", *Chemometrics and Intelligent Laboratory Systems*, **37**(2), 255-259.

P. Reiss and T. Odgen (2007) "Functional principal component regression and functional partial least squares", *Journal of the American Statistical Association*, **102**(479), 984-996.

P. Reiss and T. Odgen (2008) "Smoothing parameter selection for a class of semiparametric linear models", *Journal of Royal Statistical Society: Series B*, **71**(2), 505-523.

Examples

```
plot(Moisturespectrum)
```

Octane   *Octane content spectrometric data*

Description

This data set comprises spectra from 60 gasoline samples, measured in 2 nm intervals from 900 to 1700 nm. The response variable is the octane numbers of the samples.

Usage

```
data(Octanespectrum)
data(Octanevalues)
```

Format

Octanespectrum is an object of class `fds`.

Octanevalues is a numeric object.

Note

We thank Professor John Kalivas for the permission to re-distribute this data set.

References

J. H. Kalivas (1997) "Two data sets of near infrared spectra", *Chemometrics and Intelligent Laboratory Systems*, **37**(2), 255-259.

P. Reiss and T. Odgen (2008) "Smoothing parameter selection for a class of semiparametric linear models", *Journal of Royal Statistical Society: Series B*, **71**(2), 505-523.

Examples

```
plot(Octanespectrum)
```

Phoneme   *Phoneme data*

Description

This data set was formed by selecting five phonemes for classification based on digitized speech. There are $n = 2000$ pairs $(x_i, y_i)_{i=1,\ldots,n}$, where x_i corresponds to the discretized log-periodograms whereas the y_i gives the class membership (five phonemes: aa, ao, dcl, iy, sh).

Usage

```
data(aa)
data(ao)
data(dcl)
data(iy)
data(sh)
```

Format

An object of class `fds`.

Details

The phonemes are transcribed as follows: "sh" as in "she", "dcl" as in "dark", "iy" as the vowel in "she", "aa" as the vowel in "dark", and "ao" as the first vowel in "water".

Note

We thank Professor Frederic Ferraty for the permission to re-distribute this data set.

Source

This data set is a part of the original one from the elements of statistical learning website at `http://www-stat.stanford.edu/ElemStatLearn`.

This data set can also be found at the NonParametric Functional Data Analysis website (`http://www.lsp.ups-tlse.fr/staph/npfda/`).

References

F. Ferraty and P. Vieu (2003) "Curve discrimination: a nonparametric functional approach", *Computational Statistics and Data Analysis*, **44**(1-2), 161-173.

F. Ferraty and P. Vieu (2006) *Nonparametric functional data analysis*, New York: Springer.

T. Hastie and R. Tibshirani and J. Friedman (2009) *The elements of statistical learning: Data mining, inference and prediction*, 2nd edn, New York: Springer.

Examples

```
plot(aa)
plot(ao)
plot(dcl)
plot(iy)
plot(sh)
```

Pigweight   *Pig weight data*

Description

The pig weight data set has 9 repeated weight measures on 48 pigs.

Usage

```
data(Pigweight)
```

Format

An object of class `fds`.

Details

`Pigweight$x`: Number of weeks since measurements commenced.

`Pigweight$y`: Bodyweight(kg) of pig after weeks.

Note

We thank Professor Matt Wand for the permission to re-distribute this data set.

Source

P. J. Diggle and P. Heagerty and K. Liang and S. Zeger (2002) *Analysis of Longitudinal Data*, 2nd edn, Oxford: Oxford University Press.

References

D. Ruppert and M. Wand and R. Carroll. (2003) *Semiparametric Regression*, New York: Cambridge University Press.

Examples

```
plot(Pigweight)
```

---

read.hmd   *Function to read data sets from the Human Mortality Database*

Description

This function allows users to read any data set from the Human Mortality Database (HMD).

Usage

```
read.hmd(country, sex, file = "Mx_1x1.txt", username, password, yname)
```

Arguments

country	Directory abbreviation from the HMD. For instance, Australia = "AUS".
sex	Possible options are "Male", "Female", "Total".
file	Directory abbreviation from the HMD. For instance, mortality rate = "Mx_1x1.txt".
username	Authenticate username.
password	Authenticate password.
yname	Type of data.

Details

In order to read the data sets, users are required to create their account via the HMD website (http://www.mortality.org/), and obtain a valid username and password.

Value

An object of class `fts`.

Author(s)

Han Lin Shang and Rob J Hyndman

See Also

[hmdstatistic](), [hmdcountry]()

Satellite	*Satellite topex/poseidon*

Description

The data were registered by the satellite topex/poseidon around an area of 25 kilometers upon the Amazon River. Each row of the data matrix is represented by its wave (i.e. curve) on the range (0, 70), and the satellite is registering 10 curves each second.

Usage

```
data(Satellite)
```

Format

An object of class `fds`.

Details

Note that each wave is linked with the kind of ground treated by the satellite, and the idea for the Amazonian basin is to use these waveforms for altimetric and hydrological purposes.

Note

We thank Professor Frederic Ferraty for the permission to re-distribute this data set.

Source

F. Frappart (2003). Catalogue des formes d'onde de l'altimetre topex/poseidon sur le bassin amazonien. Technical Report, CNES, Toulouse, France.

This data set can also be found at the NonParametric Functional Data Analysis website (http://www.lsp.ups-tlse.fr/staph/npfda/).

References

F. Ferraty and P. Vieu (2006) *Nonparametric functional data analysis*, New York: Springer.

S. Dabo-Niang and F. Ferraty and P. Vieu (2007) "On the using of modal curves for radar waveforms classification", *Computational Statistics and Data Analysis*, **51**(10), 4878-4890.

Examples

```
plot(Satellite)
```

SAtemp   *Temperatures in South Australia*

Description

These data sets consist of half-hourly temperatures measured at Kent Town and Adelaide airport from Sunday to Saturday in Adelaide between 6/7/1997 and 31/3/2007.

Usage

```
data(mondaytempkent)
data(mondaytempairport)
data(tuesdaytempkent)
data(tuesdaytempairport)
data(wednesdaytempkent)
data(wednesdaytempairport)
data(thursdaytempkent)
data(thursdaytempairport)
data(fridaytempkent)
data(fridaytempairport)
data(saturdaytempkent)
data(saturdaytempairport)
data(sundaytempkent)
data(sundaytempairport)
data(tempkent)
data(tempairport)
```

Format

An object of class `sfts`.

Details

In Adelaide, the electricity demands in summer are very volatile and highly dependent on their associated temperatures. Analyses were performed to test whether, under different temperature scenarios, there will be enough capacity to satisfy the demands.

Source

L. Magnano and J. Boland and R. Hyndman (2008) "Generation of symthetic sequences of half-hourly temperature", *Environmetrics*, **19**(8), 818-835.

Examples

```
plot(mondaytempkent)
plot(tuesdaytempkent)
plot(wednesdaytempkent)
plot(thursdaytempkent)
```

```
plot(fridaytempkent)
plot(saturdaytempkent)
plot(sundaytempkent)
plot(tempkent)
```

SOI   *Annual Southern Oscillation Index (SOI) for the period 1900-2004.*

Description

Annual measures on Southern Oscillation Index (SOI): observed annual cycles in period 1900-2004.

Usage

```
data(SOI)
```

Format

An object of class `fts`.

Source

The data are available at the Australian Meteorological Office (http://www.environment.gov.au).

Examples

```
plot(SOI)
```

Spanishmigration   *Spanish migration from 1999 to 2003*

Description

This data set consists of migration number (in thousands) in Spain from 1999 to 2003. This data set contains the migration rates of 9 age groups, namely 0-9, 10-15, 16-19, 20-29, 30-39, 40-49, 50-59, 60-65, and 65+ for both females and males.

Usage

```
data(femalemigration)
data(malemigration)
```

Format

An object of class `fts`.

Details

This data set was calculated at the time in accordance with the European studies of population (EAPS <https://www.eaps.nl>) methodology of 2002, which is hetergeneous with the results calculated using the methodology EAPS in 2005.

Source

Instituto Nacional de Estadistica website at http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t20/p311&file=inebase.

References

D. Reher and M. Requena (2009) "The national immigration survey of Spain. A new data source for migration studies in Europe", *Demographic Research*, **20**, 253-278.

Examples

```
plot(femalemigration)
plot(malemigration)
```

Yieldcurve	*US: Treasury bond*

Description

This data set contains monthly US Treasury bonds from January 1970 through December 2002. Based on the bid-ask midpoint average, the data consist of end of the month price quotes.

Usage

```
data(Yieldcurve)
```

Format

An object of class `fts`.

Details

This data set is filtered to eliminate bonds with special option futures, such as callable and flower bonds. Illiquid securities, such as treasury bills with less than one month on maturity and treasury notes and bonds with less than one year to maturity, are excluded from the samples.

Source

CRSP US Treasury Database (<http://www.crsp.com/products/research-products/crsp-us-treasury-database>).

Examples

```
plot(Yieldcurve)
```

Index