

Package ‘tisai’

February 4, 2026

Title AI enabled time series and spacial statistics

Version 0.0.1

Author Xu Liu [aut,cre]

Maintainer Xu Liu <liu.xu@sufe.edu.cn>

Description Datasets used in the book ``AI enabled time series and spacial statistics''

License GPL (>=2)

Encoding UTF-8

Roxxygen list(markdown = TRUE)

RoxxygenNote 7.3.2

Depends R (>= 3.5.0)

Imports Matrix, glmnet

LazyData true

Contents

Gtemp	1
gtemp.month	2
Hare	3
HKflu	4
lap	5
Lynx	6
USeconomic	7
USmacro	8

Index	10
--------------	-----------

Gtemp

Global Annual Temperature Anomalies (1850–2017)

Description

This dataset provides the global annual mean temperature anomalies relative to the 1961–1990 average, sourced from the HadCRUT4 dataset.

Usage

```
data(Gtemp)
```

Format

A data frame with 168 observations on the following 2 variables:

year The year of observation (from 1850 to 2017).

anomaly Global temperature anomaly (degrees Celsius) relative to the 1961–1990 mean.

Details

The dataset reflects the long-term warming trend and potential 60-70 year periodic oscillations in global temperatures.

Source

Climatic Research Unit (CRU), University of East Anglia (HadCRUT4). <https://crudata.uea.ac.uk/cru/data/temperature/>

References

Wang, S., You, J., & Huang, T. (2022). Modelling and applications for non-stationary time series in the presence of trend and period. *Scientia Sinica Mathematica*, 52(2), 177-208.

Examples

```
data(Gtemp)
plot(Gtemp$year, Gtemp$anomaly, type = "l", main = "Global Temperature Anomalies")
```

gtemp.month

Global Temperature Monthly Data

Description

Global average monthly temperature data covering the period from 1975 to 2023. This dataset provides global temperature measurements, including both land and ocean surface temperatures, and is widely used for climate change research and time series analysis.

Usage

```
data(gtemp.month)
```

Format

A data frame with 12 rows (months) and 49 columns (years from 1975 to 2023):

- Rows: Represent months (1 = January, 2 = February, ..., 12 = December)
- Columns: Represent years from 1975 to 2023
- Values: Monthly global average temperatures in degrees Celsius

Details

Global temperature monthly data

This dataset contains global monthly temperature data from 1975 to 2023.

The gtemp.month dataset provides monthly global temperature data from 1975 to 2023. It is commonly used to study climate change trends, seasonal variations, and long-term temperature patterns. The data includes measurements from both land and ocean surfaces, providing a comprehensive view of global temperature changes over time.

Source

Package 'astsa' available at <https://nickpoison.github.io/>

Examples

```
data(gtemp.month)

# Transpose the data for plotting
gtemp_t <- t(gtemp.month)

# Plot the temperature data for January
plot(rownames(gtemp_t), gtemp_t[, 1], type = "l",
      main = "January Global Temperature",
      xlab = "Year", ylab = "Temperature (°C)",
      col = "blue")

# Calculate summary statistics for each month
apply(gtemp.month, 1, summary)

# Plot average temperature across all months for each year
annual_avg <- colMeans(gtemp.month)
plot(names(annual_avg), annual_avg, type = "l",
      main = "Annual Average Global Temperature",
      xlab = "Year", ylab = "Temperature (°C)",
      col = "red")
```

Hare

Snowshoe Hare Population Data

Description

Snowshoe hare population data from the Hudson's Bay Company records, covering the period from 1845 to 1935. This classic dataset is widely used for studying predator-prey population dynamics.

Usage

```
data(Hare)
```

Format

A time series object with the following characteristics:

- Time period: 1845 - 1935
- Frequency: Annual
- Values: Number of snowshoe hare pelts traded (in thousands)

Details

Hare population data

This dataset contains snowshoe hare population data collected by the Hudson's Bay Company.

The Hare dataset provides annual counts of snowshoe hare pelts traded by the Hudson's Bay Company from 1845 to 1935. This dataset, along with the Lynx dataset, forms a classic example of predator-prey population dynamics and is commonly used in time series analysis and ecological modeling.

Source

Package 'astsa' available at <https://nickpoison.github.io/>

Examples

```
data(Hare)

# Plot the hare population data
plot(Hare, main = "Snowshoe Hare Population Data",
     xlab = "Year", ylab = "Pelts Traded (thousands)",
     col = "blue")

# Calculate summary statistics
summary(Hare)
```

Description

This dataset contains weekly records of influenza incidence rates along with various air quality and meteorological indicators in Hong Kong from December 30, 2012, to December 22, 2019.

Usage

```
data(HKflu)
```

Format

A data frame with 365 observations on the following 11 variables:

week A numeric vector indicating the sequence of monitored weeks.

influenza Influenza intensity, measured as the number of laboratory-confirmed influenza cases per 1,000 recorded consultations.

FSP Fine Suspended Particulates (PM2.5) daily average concentration.

RSP Respirable Suspended Particulates (PM10) daily average concentration.

NO2 Nitrogen Dioxide (\$NO_2\$) daily average concentration.

CO Carbon Monoxide (CO) daily average concentration.

SO2 Sulphur Dioxide (\$SO_2\$) daily average concentration.

NOX Nitrogen Oxides (\$NO_x\$) daily average concentration.

O3 Ozone (\$O_3\$) daily average concentration.

temperature Weekly average temperature in degrees Celsius.

humidity Weekly average relative humidity percentage.

Source

Hong Kong Environmental Protection Department and relevant health authorities.

References

Li, J., Wang, S., & You, J. (2022). Nonparametric additive models for discrete-time series with periodic features. *Acta Mathematica Sinica, Chinese Series*, 65(1), 177-204.

Examples

```
data(HKflu)
summary(HKflu)
plot(HKflu$temperature, HKflu$influenza)
```

lap*Los Angeles Pollution and Mortality Data*

Description

Pollution and mortality data from Los Angeles, California, covering the period from 1970 to 1979. This dataset contains weekly mortality rates and associated pollutant concentrations, and is used to study the relationship between air pollution and public health, particularly cardiovascular and respiratory diseases.

Usage

```
data(lap)
```

Format

A multivariate time series (mts) object with the following characteristics:

- Time period: 1970 - 1979
- Frequency: Weekly (52 observations per year)
- Number of variables: 11
- Variables:
 - tmort: Total mortality rate (per 100,000 people per week)
 - rmort: Respiratory mortality rate (per 100,000 people per week)
 - cmort: Cardiovascular mortality rate (per 100,000 people per week)
 - tempr: Temperature (degrees Fahrenheit)
 - rh: Relative humidity (percentage)
 - co: Carbon monoxide concentration (ppm)
 - so2: Sulfur dioxide concentration (ppm)
 - no2: Nitrogen dioxide concentration (ppm)
 - hycarb: Hydrocarbon concentration (ppm)
 - o3: Ozone concentration (ppm)
 - part: Particulate matter concentration (ppm)

Details

LAP data

This dataset contains pollution and mortality data from Los Angeles, California, from 1970 to 1979.

The lap dataset provides weekly measurements of various pollution and health indicators from Los Angeles, California, from 1970 to 1979. It is commonly used in environmental health studies to analyze the effects of pollution on mortality, particularly for cardiovascular and respiratory diseases. The dataset includes measurements of multiple pollutants along with mortality rates, allowing researchers to investigate the relationships between specific pollutants and health outcomes.

Source

Package 'astsa' available at <https://nickpoison.github.io/>

Examples

```
data(lap)

# Plot total mortality over time
plot(lap[, "tmort"], main = "Total Mortality in Los Angeles",
     xlab = "Time", ylab = "Mortality Rate (per 100,000 per week)",
     col = "red")

# Calculate correlation between ozone and cardiovascular mortality
cor(lap[, "o3"], lap[, "cmort"], use = "complete.obs")

# Plot mortality rates together
plot(lap[, c("tmort", "rmort", "cmort")],
     main = "Mortality Rates in Los Angeles")

# Plot pollutants together
plot(lap[, c("co", "so2", "no2", "o3")],
     main = "Pollutant Concentrations in Los Angeles")
```

Description

Canadian lynx population data from the Hudson's Bay Company records, covering the period from 1845 to 1935. This classic dataset is widely used for studying predator-prey population dynamics alongside the Hare dataset.

Usage

```
data(Lynx)
```

Format

A time series object with the following characteristics:

- Time period: 1845 - 1935
- Frequency: Annual
- Values: Number of lynx pelts traded (in thousands)

Details

Lynx population data

This dataset contains Canadian lynx population data collected by the Hudson's Bay Company.

The Lynx dataset provides annual counts of Canadian lynx pelts traded by the Hudson's Bay Company from 1845 to 1935. Together with the Hare dataset, this forms a classic example of predator-prey population dynamics, where lynx (predators) and snowshoe hares (prey) exhibit cyclical population fluctuations.

Source

Package 'astsa' available at <https://nickpoison.github.io/>

Examples

```
data(Lynx)

# Plot the lynx population data
plot(Lynx, main = "Canadian Lynx Population Data",
     xlab = "Year", ylab = "Pelts Traded (thousands)",
     col = "red")

# Calculate summary statistics
summary(Lynx)

# Plot both hare and lynx populations together
plot(Hare, main = "Hare and Lynx Population Dynamics",
      xlab = "Year", ylab = "Pelts Traded (thousands)",
      col = "blue", ylim = c(0, max(Hare, Lynx)))
lines(Lynx, col = "red")
legend("topleft", legend = c("Hare", "Lynx"),
      col = c("blue", "red"), lty = 1)
```

Description

This dataset provides a structural (quarterly) snapshot of the U.S. economy from 2000.1.1 to 2025.7.1, consisting of 103 observations.

It integrates the major components of Gross Domestic Product (GDP): - Aggregate Output: Nominal and Real GDP. - Domestic Demand: Personal consumption (PCE) and Private investment. - External Sector: Real exports and imports of goods and services.

The dataset allows students to explore the nonlinear dynamics of investment shocks and the global trade balance using advanced AI structural models.

Usage

```
data(USeconomic)
```

Format

A data frame with aligned quarterly observations on the following 7 variables:

```
date Start date of the quarter (YYYY-MM-DD).
gdp Gross Domestic Product (Billions of Dollars).
real_gdp Real Gross Domestic Product (Billions of Chained 2017 Dollars).
pce Personal Consumption Expenditures (Billions of Dollars).
invest Real Gross Private Domestic Investment (Billions of Chained 2017 Dollars).
export Real Exports of Goods and Services (Billions of Chained 2017 Dollars).
import Real Imports of Goods and Services (Billions of Chained 2017 Dollars).
```

Source

U.S. Bureau of Economic Analysis (BEA) via Federal Reserve Bank of St. Louis (FRED). Data IDs: GDP, GDPC1, PCE, GPDIC1, EXPGSC1, and IMPGSC1. Visit: <https://fred.stlouisfed.org/>

References

U.S. Bureau of Economic Analysis, "National Income and Product Accounts". Retrieved from FRED: <https://fred.stlouisfed.org/categories/106>

Examples

```
data(USeconomic)
# Calculate Investment-to-GDP ratio
inv_ratio <- USeconomic$invest / USeconomic$real_gdp
plot(USeconomic$date, inv_ratio, type="l", ylab="Investment/GDP Ratio")
```

Description

This dataset provides a high-frequency (monthly) snapshot of the U.S. economy from 2000.1.1 to 2025.10.1, consisting of 310 observations.

The variables cover four critical dimensions: 1. Labor Market: Unemployment rate and Nonfarm payrolls. 2. Price Level: Consumer Price Index (CPI) for inflation tracking. 3. Monetary Policy: Federal Funds Rate and 10-Year Treasury yields. 4. Real Activity: Industrial production and Housing starts.

This data is ideal for teaching lead-lag relationships, regime switching, and high-frequency forecasting using LSTM or Transformer architectures.

Usage

```
data(USmacro)
```

Format

A data frame with aligned monthly observations on the following 8 variables:

`date` First day of the month (YYYY-MM-DD).
`unrate` Civilian Unemployment Rate (Percent).
`cpi` Consumer Price Index for All Urban Consumers (Index 1982-1984=100).
`interest` Federal Funds Effective Rate (Percent).
`indpro` Industrial Production Index (Index 2017=100).
`employ` All Employees, Total Nonfarm (Thousands of Persons).
`housing` New Privately Owned Housing Units Started (Thousands of Units).
`bond10y` 10-Year Treasury Constant Maturity Rate (Percent).

Source

Federal Reserve Bank of St. Louis (FRED). Data accessible via the following IDs: UNRATE, CPI-AUCSL, FEDFUNDS, INDPRO, PAYEMS, HOUST, and GS10. Visit: <https://fred.stlouisfed.org/>

References

Federal Reserve Bank of St. Louis, Federal Reserve Economic Data (FRED). <https://fred.stlouisfed.org>

Examples

```
data(USmacro)
# Visualize the relationship between interest rates and housing starts
plot(USmacro$interest, USmacro$housing,
     main="Interest Rates vs Housing Starts")
```

Index

* datasets

Gtemp, 1
gtemp.month, 2
Hare, 3
HKflu, 4
lap, 5
Lynx, 6
USeconomic, 7
USmacro, 8

Gtemp, 1
gtemp.month, 2

Hare, 3
HKflu, 4

lap, 5
Lynx, 6

USeconomic, 7
USmacro, 8