

# Package ‘cluster.datasets’

October 12, 2022

**Version** 1.0-1

**Date** 2013-10-28

**Author** Frederick Novomestky <fnovomes@poly.edu>

**Maintainer** Frederick Novomestky <fnovomes@poly.edu>

**Depends** R (>= 2.0.1)

**Description** A collection of data sets for teaching cluster analysis.

**Title** Cluster Analysis Data Sets

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2013-10-29 07:55:17

## R topics documented:

acidosis.patients . . . . .	2
airline.distances.1966 . . . . .	3
all.mammals.milk.1956 . . . . .	5
all.us.city.crime.1970 . . . . .	6
amino.acid.sequence.1972 . . . . .	7
animal.cluster.trees . . . . .	9
birth.death.rates.1966 . . . . .	10
british.butterfly.appearance . . . . .	11
cake.ingredients.1961 . . . . .	12
candida.oxidation.fermentation . . . . .	13
cerci.tail.presence . . . . .	15
ct.president.vote.1920.1964 . . . . .	16
european.foods . . . . .	17
hardware.triads . . . . .	18
hartigan.datasets . . . . .	19
indian.caste.measures . . . . .	20
indo.european.languages . . . . .	21
indochina.combat.deaths . . . . .	22

ivy.league.football.1965 . . . . .	23
jigsaw.puzzle.measures . . . . .	24
languages.spoken.europe . . . . .	26
leukemia.youth.mortality.1956.1967 . . . . .	27
life.expectancy.1971 . . . . .	29
life.expectancy.age.sex.1971 . . . . .	30
linguistic.relatedness . . . . .	31
mammal.dentition . . . . .	32
minor.planets.1961 . . . . .	33
mutation.distances.1967 . . . . .	34
nails.screws . . . . .	35
new.haven.school.scores . . . . .	36
nutrients.meat.fish.fowl.1959 . . . . .	37
ohio.croplands.1949 . . . . .	38
olympic.track.1896.1964 . . . . .	39
physical.measure.correlations . . . . .	40
planet.earth.distances.1970 . . . . .	41
planets.moons.1970 . . . . .	42
portable.typewriters . . . . .	43
rda.meat.fish.fowl.1959 . . . . .	44
sample.mammals.milk.1956 . . . . .	45
sample.stock.yields.1959.1969 . . . . .	46
sample.us.city.crime.1970 . . . . .	48
student.questionnaire . . . . .	49
un.votes.1969.1970 . . . . .	50
us.car.repair.1969 . . . . .	51
us.civil.war.battles . . . . .	52
us.congressional.bills . . . . .	53
us.food.cost.nutrients.1959 . . . . .	55
us.links.between.states . . . . .	56
us.per.capita.income.1964 . . . . .	57
us.president.vote.1900.1968 . . . . .	58
us.sector.profitability.1959.1968 . . . . .	59
us.south.demographics.1965 . . . . .	60
vervet.sleeping.groups . . . . .	62
wine.evaluation.1961.1970 . . . . .	63

**Index** **65**

---

acidosis.patients	<i>Hartigan (1975) Acidosis Patients</i>
-------------------	--

---

**Description**

The table contains measures of various compounds in cerebrospinal fluid and blood for acidosis patients. This is Table 14.11 in Chapter 14 of Hartigan (1975) on page 265.

**Usage**

```
data(acidosis.patients)
```

**Format**

A data frame with 40 observations on the following 6 variables.

- ph.cerebrospinal.fluid a numeric vector
- ph.blood a numeric vector
- hco3.cerebrospinal.fluid a numeric vector
- hco3.blood a numeric vector
- co2.cerebrospinal.fluid a numeric vector
- co2.blood a numeric vector

**Details**

Hartigan suggests the use of the direct splitting algorithm with this data set.

**Source**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.  
SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(acidosis.patients)
```

---

airline.distances.1966

*Hartigan (1975) Airline Distance Between Principal Cities of the World*

---

**Description**

The table contains the airline distances in hundreds of miles between the principal cities of the world. This is Table 11.1 in Chapter 11 of Hartigan (1975) on page 192.

**Usage**

```
data(airline.distances.1966)
```

**Format**

A data frame with 30 observations on the following 31 variables.

code a character vector for the cities  
AZ a numeric vector for Azores  
BD a numeric vector for Baghdad  
BN a numeric vector for Berlin  
BY a numeric vector for Bombay  
BS a numeric vector for Buenos Aires  
CO a numeric vector for Cairo  
CN a numeric vector for Capetown  
CH a numeric vector for Chicago  
GM a numeric vector for Guam  
HU a numeric vector for Honolulu  
IL a numeric vector for Istanbul  
JU a numeric vector for Juneau  
LN a numeric vector for London  
MA a numeric vector for Manila  
ME a numeric vector for Melbourne  
MY a numeric vector for Mexico City  
ML a numeric vector for Montreal  
MW a numeric vector for Moscow  
NS a numeric vector for New Orleans  
NY a numeric vector for New York  
PY a numeric vector for Panama City  
PS a numeric vector for Paris  
RO a numeric vector for Rio De Janeiro  
RE a numeric vector for Rome  
SF a numeric vector for San Francisco  
SO a numeric vector for Santiago  
SE a numeric vector for Seattle  
SI a numeric vector for Shanghai  
SY a numeric vector for Sydney  
TO a numeric vector for Tokyo

**Details**

Hartigan uses this data set with the single linkage algorithm.

### Source

*The World Almanac* (1966).

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(airline.distances.1966)
```

---

`all.mammals.milk.1956` *Hartigan (1975) Mammal's Milk*

---

### Description

The table contains a list of animals and the constituents of their milk. A shorter version appears in `jh.table.1.2`. This is Table 16.3 in Chapter 16 of Hartigan (1975) on page 304.

### Usage

```
data(all.mammals.milk.1956)
```

### Format

A data frame with 25 observations on the following 6 variables.

`name` a character vector for the animal name

`water` a numeric vector for the percentage of water

`protein` a numeric vector for the percentage of protein

`fat` a numeric vector for the percentage of fat

`lactose` a numeric vector for the percentage of lactose

`ash` a numeric vector for the percentage of ash.

### Details

Hartigan suggests the use of a joiner-scaler algorithm on this data set.

### Source

Spector, W. S. (1956) *Handbook of Biological Data*, Saunders.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(all.mammals.milk.1956)
```

---

```
all.us.city.crime.1970
```

*Hartigan (1975) City Crime*

---

**Description**

The table records city crime along with population statistics. This is Table 18.6 in Chapter 18 of Hartigan (1975) on page 342.

**Usage**

```
data(all.us.city.crime.1970)
```

**Format**

A data frame with 24 observations on the following 10 variables.

city a character vector for the city name

population a numeric vector for the population in thousands

white.change a numeric vector for the percent change in inner city white population from 1960 to 1970

black.population a numeric vector for the black population in thousands

murder a numeric vector for the murder rate

rape a numeric vector for the rape rate

robbery a numeric vector for the robbery rate

assault a numeric vector for the assault rate

burglary a numeric vector for the burglary rate

car.theft a numeric vector for the car theft rate

**Details**

All rate variables are per 100,000 population. Hartigan suggests using the AID algorithm on this data set.

**Source**

*The Statistical Abstract of the United States* (1971), Bureau of Census, Department of Commerce, Grossett and Dunlop, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(all.us.city.crime.1970)
```

---

```
amino.acid.sequence.1972
```

*Hartigan (1975) Amino Acid Sequence for Vertebrates*

---

**Description**

The table defines the position of amino acids for Cytochrome-c. This is Table 13.4 in Chapter 13 of Hartigan (1975) on page 240.

**Usage**

```
data(amino.acid.sequence.1972)
```

**Format**

A data frame with 17 observations on the following 37 variables.

species a character vector for the species names  
p.1 a factor for position 1 with levels I V  
p.2 a factor for position 2 with levels A E  
p.3 a factor for position 3 with levels I T V  
p.4 a factor for position 4 with levels I T V  
p.5 a factor for position 5 with levels M Q  
p.6 a factor for position 6 with levels A S  
p.7 a factor for position 7 with levels C V  
p.8 a factor for position 8 with levels K N  
p.9 a factor for position 9 with levels T V  
p.10 a factor for position 10 with levels H N S W Y  
p.11 a factor for position 11 with levels F I

- p.12 a factor for position 12 with levels A E P Q V
- p.13 a factor for position 13 with levels F Y
- p.14 a factor for position 14 with levels S T
- p.15 a factor for position 15 with levels A D E
- p.16 a factor for position 16 with levels N S
- p.17 a factor for position 17 with levels I T V
- p.18 a factor for position 18 with levels G K N Q
- p.19 a factor for position 19 with levels E N Q
- p.20 a factor for position 20 with levels D E
- p.21 a factor for position 21 with levels M R
- p.22 a factor for position 22 with levels E I
- p.23 a factor for position 23 with levels I V
- p.24 a factor for position 24 with levels T V
- p.25 a factor for position 25 with levels I L
- p.26 a factor for position 26 with levels K S
- p.27 a factor for position 27 with levels K
- p.28 a factor for position 28 with levels A D E G K S T
- p.29 a factor for position 29 with levels A E Q T V
- p.30 a factor for position 30 with levels D N
- p.31 a factor for position 31 with levels I V
- p.32 a factor for position 32 with levels D E K Q S
- p.33 a factor for position 33 with levels A K T
- p.34 a factor for position 34 with levels A C T
- p.35 a factor for position 35 with levels A K N S
- p.36 a factor for position 36 with levels - A E K S

### Details

The factor levels across the 36 positions common. Hartigan uses the reduced mutation algorithm with this data set.

### Source

Dickerson, R. E. (1972). The structure and history of an ancient problem, *Scientific American*, 222(4), 58-72.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(amino.acid.sequence.1972)
```



---

animal.cluster.trees    *Hartigan (1975) Cluster of Animals Forming a Tree*

---

### Description

The table is a binary table that identifies which animals are in given cluster. This is Table 8.1 in Chapter 8 of Hartigan (1975) on page 155.

### Usage

```
data(animal.cluster.trees)
```

### Format

A data frame with 13 observations on the following 11 variables.

symbol a character vector for

name a character vector for

c.1 a numeric vector for a binary variable. A value 1 means the animal is in cluster 1 while 0 means that it is not in that cluster

c.2 a numeric vector for a binary variable. A value 1 means the animal is in cluster 2 while 0 means that it is not in that cluster

c.3 a numeric vector for a binary variable. A value 1 means the animal is in cluster 3 while 0 means that it is not in that cluster

c.4 a numeric vector for a binary variable. A value 1 means the animal is in cluster 4 while 0 means that it is not in that cluster

c.5 a numeric vector for a binary variable. A value 1 means the animal is in cluster 5 while 0 means that it is not in that cluster

c.6 a numeric vector for a binary variable. A value 1 means the animal is in cluster 6 while 0 means that it is not in that cluster

c.7 a numeric vector for a binary variable. A value 1 means the animal is in cluster 7 while 0 means that it is not in that cluster

c.8 a numeric vector for a binary variable. A value 1 means the animal is in cluster 8 while 0 means that it is not in that cluster

c.9 a numeric vector for a binary variable. A value 1 means the animal is in cluster 9 while 0 means that it is not in that cluster

### Details

This table is used to construct and present a cluster tree as defined in Hartigan (1975).

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(animal.cluster.trees)
```

---

```
birth.death.rates.1966
```

*Hartigan (1975) Birth and Death Rates Per 1000*

---

**Description**

A table with birth and death rates per 1000 persons for selected countries. This is Table 11.6 in Chapter 11 of Hartigan (1975) on page 197.

**Usage**

```
data(birth.death.rates.1966)
```

**Format**

A data frame with 70 observations on the following 3 variables.

country a character vector for the country name

birth a numeric vector for the birth rates per 1000 persons

death a numeric vector for the death rates per 1000 persons

**Details**

Hartigan recommends that spirical search algorithm be applied to this data set.

**Source**

*Reader's Digest Almanac* (1966)

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(birth.death.rates.1966)
```

---

 british.butterfly.appearance

*Hartigan (1975) Times of Appearance of British Butterflies*


---

### Description

The table defines the metamorphosis sequences of British butterflies. This is Table 7.6 in Chapter 7 of Hartigan (1975) on page 150.

### Usage

```
data(british.butterfly.appearance)
```

### Format

A data frame with 27 observations on the following 13 variables.

name a character vector for the species

jan a factor for January occurrences with levels I L O P

feb a factor for February occurrences with levels I L O P

mar a factor for March occurrences with levels I L O P

apr a factor for April occurrences with levels I L LP O OL P PI

may a factor for May occurrences with levels I L LI LP LPI P PI

jun a factor for June occurrences with levels I IL IOL L LI LP LPI P PI

jul a factor for July occurrences with levels I L LI LP LPI O P PI

aug a factor for August occurrences with levels I L LI LPI O P PI

sep a factor for September occurrences with levels I L LI LP LPI O P PI

oct a factor for October occurrences with levels I L LP LPI O P

nov a factor for November occurrences with levels I L O P

dec a factor for December occurrences with levels I L O P

### Details

Hartigan suggests using this data set to test the ditto algorithm.

### Source

Ford, T. L. E. (1963). *Practical Entomology*, Warne, London, p. 181.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(british.butterfly.appearance)
```

---

```
cake.ingredients.1961 Hartigan (1975) Ingredients in Cakes
```

---

**Description**

The table identifies for each cake which ingredient is used and the quantity. This is Table 12.8 in Chapter 12 of Hartigan (1975) on page 229.

**Usage**

```
data(cake.ingredients.1961)
```

**Format**

A data frame with 18 observations on the following 35 variables.

Cake a character vector for the name of the cake  
AE a numeric vector for the amount of Almond essence in teaspoons  
BM a numeric vector for the amount of Buttermilk in cups  
BP a numeric vector for the amount of Baking powder in teaspoons  
BR a numeric vector for the amount of Butter in cups  
BS a numeric vector for the amount of Bananas in whole bananas  
CA a numeric vector for the amount of Cocoa in tablespoons  
CC a numeric vector for the amount of Cottage Cheese in pounds  
CE a numeric vector for the amount of Chocolate in ounces  
CI a numeric vector for the amount of Crushed Ice in cups  
CS a numeric vector for the amount of Crumbs in cups  
CT a numeric vector for the amount of Cream of tartar in teaspoons  
DC a numeric vector for the amount of Dried currants in tablespoons  
EG a numeric vector for the amount of Eggs in whole eggs  
EY a numeric vector for the amount of Egg white in whole eggs  
EW a numeric vector for the amount of Egg yolk in whole eggs  
FR a numeric vector for the amount of Sifted flour in cups  
GN a numeric vector for the amount of Gelatin in tablespoons  
HC a numeric vector for the amount of Heavy cream in cups  
LJ a numeric vector for the amount of Lemon juice in tablespoons  
LR a numeric vector for the amount of Lemon rind in teaspoons  
MK a numeric vector for the amount of Milk in cups

NG a numeric vector for the amount of Nutmeg in teaspoons  
NS a numeric vector for the amount of Nuts in cups  
RM a numeric vector for the amount of Rum in ounces  
SA a numeric vector for the amount of Soda in teaspoons  
SC a numeric vector for the amount of Sour cream in cups  
SG a numeric vector for the amount of Shortening in tablespoons  
SR a numeric vector for the amount of Granulated sugar in cups  
SS a numeric vector for the amount of Strawberries in quarts  
ST a numeric vector for the amount of Salt in teaspoons  
VE a numeric vector for the amount of Vanilla extract in teaspoons  
WR a numeric vector for the amount of Water in cups  
YT a numeric vector for the amount of Yeast in ounces  
ZH a numeric vector for the amount of Zwiebach in ounces

### Details

For each cake and ingredient, the data frame contains NA if the ingredient is not required or a numeric value.

### Source

Claiborn, C. (1961) *The New York Times Cookbook*, Harper and Row, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(cake.ingredients.1961)
```

---

candida.oxidation.fermentation

*Hartigan (1975) Oxidation-Fermentation Patterns*

---

### Description

The table contains the oxidation-fermentation patterns for a sample of species of *Candida* in terms of acid production. This is Table 15.1 in Chapter 15 Hartigan (1975) on page 279.

### Usage

```
data(candida.oxidation.fermentation)
```

**Format**

A data frame with 8 observations on the following 13 variables.

name a character vector for the species name  
glucose a factor for glucose with levels +  
maltose a factor for maltose with levels - +  
sucrose a factor for sucrose with levels - +  
lactose a factor for lactose with levels - +  
galactose a factor for galactose with levels - +  
melibiose a factor for melibiose with levels - +  
cellobiose a factor for cellobiose with levels - +  
inositol a factor for inositol with levels -  
xylose a factor for xylose with levels - +  
raffinose a factor for raffinose with levels - +  
trehalose a factor for trehalose with levels - +  
dulcitol a factor for dulcitol with levels - +

**Details**

A '+' level means oxidative production of acid where as a '-' level means no acide production. Hartigan suggests using direct joining on this data set.

**Source**

Hall, T. C., Webb, C. D> and Papageorge, C. (1972) Use of oxidation-fermentation medium in the identification of yeasts, *HSMHA Report*, 87, 172 - 176.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(candida.oxidation.fermentation)
```

---

cerci.tail.presence     *Hartigan (1975) Presence of Cerci in Insects*

---

### Description

The table defines the hierarchy of insects classified according to cerci or tail appendages. This is Table 13.1 in Chapter 13 of Hartigan (1975) on page 234.

### Usage

```
data(cerci.tail.presence)
```

### Format

A data frame with 38 observations on the following 4 variables.

index a numeric vector for the insect index

code a character vector for the insect code

name a character vector for the name of the index or family

parent a numeric vector the index of the parent insect

### Details

Hartigan applies the minimum mutation method to this data set.

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(cerci.tail.presence)
```

---

ct.president.vote.1920.1964

*Hartigan (1975) Connecticut Votes for President*

---

### Description

The table contains presidential votes recorded over 12 elections and for 8 counties in Connecticut. This is Table 14.13 in Chapter 14 of Hartigan (1975) on page 267.

### Usage

```
data(ct.president.vote.1920.1964)
```

### Format

A data frame with 36 observations on the following 10 variables.

year a numeric vector for the election year  
party a character vector for the political party  
fairfield a numeric vector for Fairfield county  
hartford a numeric vector for Hartford county  
litchfield a numeric vector for Litchfield county  
middlesex a numeric vector for Middlesex county  
new.haven a numeric vector for New Haven county  
new.london a numeric vector for New London county  
tolland a numeric vector for Tolland county  
windham a numeric vector for Windham county

### Details

Hartigan recommend the use of the two direct splitting algorithm on this data set.

### Source

Scammon, R. M. (1965) *America at the Polls*, University of Pittsburgh, Pittsburgh.  
SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(ct.president.vote.1920.1964)
```



---

european.foods	<i>Hartigan( 1975) European Food</i>
----------------	--------------------------------------

---

**Description**

The table contains by country the percentage of all households with various foods in house at the time of questionnaire. This is Table 15.9 in Chapter 15 of (Hartigan) on page 289.

**Usage**

```
data(european.foods)
```

**Format**

A data frame with 20 observations on the following 18 variables.

code a character vector for the food code  
name a character vector for the food name  
wg a numeric vector for West Germany  
it a numeric vector for Italy  
fr a numeric vector for France  
ns a numeric vector for Netherlands  
bm a numeric vector for Belgium  
lg a numeric vector for Luxemburg  
gb a numeric vector for Great Britain  
pl a numeric vector for Portugal  
aa a numeric vector for Austria  
sd a numeric vector for Switzerland  
sw a numeric vector for Sweden  
dk a numeric vector for Denmark  
ny a numeric vector for Norway  
fd a numeric vector for Finland  
sp a numeric vector for Spain  
id a numeric vector for Ireland

**Details**

Hartigan suggests applying two way direct joining to this data set.

**Source**

*A Survey of Europe Today*, The Readers' Digest Association Ltd, London.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

## References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

## Examples

```
data(european.foods)
```

---

hardware.triads	<i>Hartigan (1975) Triads Based on Hardware</i>
-----------------	---

---

## Description

The table defines pairs of hardware objects that are most similar along with a dissimilar object. This is Table 10.1 in Chapter 10 of Hartigan (1975) on page 178.

## Usage

```
data(hardware.triads)
```

## Format

A data frame with 20 observations on the following 4 variables.

case a character vector

similar.1 a factor for the first object of similar pair with levels B N P T

similar.2 a factor for the second object of similar pair with levels B F S T

odd a factor for the different object with levels B F N P S T

## Details

Six pieces of hardware were considered. Every possible set of three distinct pieces of hardware was examined, and a judgment was made about which two pieces were most similar. The results were reported by listing the closest pair with parentheses surrounding them, followed by the "odd" item. The hardware objects are identified as follows

1. "N" is a nail
2. "P" is a Phillips head screw
3. "B" is a bolt
4. "T" is a tack
5. "F" is a finishing nail
6. "S" is a screw

These data are used to test the triads algorithm.

**Source**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(hardware.triads)
```

---

hartigan.datasets	<i>Hartigan (1975) Data Sets</i>
-------------------	----------------------------------

---

**Description**

This data frame contains the directory of data sets from Hartigan (1975) that are available in this package.

**Usage**

```
data(hartigan.datasets)
```

**Format**

A data frame with 53 observations on the following 4 variables.

`table.name` a character vector with the table name

`chapter` a numeric vector with the chapter containing the table

`page` a numeric vector with the page on which the table appears

`data.set.name` a character vector the data set name in this package

**Details**

Chapter number 0 is associated with the Introduction of the book.

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(hartigan.datasets)
```

---

indian.caste.measures *Hartigan (1975) Indian Caste Measurements*

---

### Description

The table contains the correlations multiplied by 10000 for 22 caste groups each with 67 to 196 individuals. This is Table 17.6 in Chapter 17 of Hartigan (1975) on page 324.

### Usage

```
data(indian.caste.measures)
```

### Format

A data frame with 9 observations on the following 9 variables.

st a numeric vector for the correlations with stature

sh a numeric vector for the correlations with sitting height

nd a numeric vector for the correlations with basal depth

nh a numeric vector for the correlations with nasal height

hl a numeric vector for the correlations with head length

fb a numeric vector for the correlations with frontal breadth

bb a numeric vector for the correlations with bizygomatic breadth

hb a numeric vector for the correlations with head breadth

nb a numeric vector for the correlations with nasal breadth

### Details

The data frame has as row names the variable names. The actual correlations are recovered by dividing the data frame by 10000. Hartigan suggests performing a factor analysis on the data set as well as performing a joining algorithm.

### Source

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification, *J. Royal Stat. Soc. B*, 10, 159 - 193.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(indian.caste.measures)
```

---

indo.european.languages

*Hartigan (1975) Indo-European Languages*

---

### Description

The table contains foreign language equivalent of the names associated with the column names. This is Table 13.8 in Chapter 13 of Hartigan (1975) on page 243.

### Usage

```
data(indo.european.languages)
```

### Format

A data frame with 13 observations on the following 17 variables.

language a character vector for the foreign language  
all a character vector for the foreign language equivalent  
bad a character vector for the foreign language equivalent  
belly a character vector for the foreign language equivalent  
black a character vector for the foreign language equivalent  
bone a character vector for the foreign language equivalent  
day a character vector for the foreign language equivalent  
die a character vector for the foreign language equivalent  
drink a character vector for the foreign language equivalent  
ear a character vector for the foreign language equivalent  
eat a character vector for the foreign language equivalent  
egg a character vector for the foreign language equivalent  
eye a character vector for the foreign language equivalent  
father a character vector for the foreign language equivalent  
fish a character vector for the foreign language equivalent  
five a character vector for the foreign language equivalent  
foot a character vector for the foreign language equivalent

### Details

Hartigan suggest that the minimum mutation algorithm is applied to this data set.

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(indo.european.languages)
```

---

```
indochina.combat.deaths
```

*Hartigan (1975) Combat Deaths in Indochina*

---

**Description**

Table contains the number of monthly combat deaths for US troops, South Vietnamese troops, third party troops and enemy troops. This is Table 6.4 in Chapter 6 of Hartigan (1975) on page 139.

**Usage**

```
data(indochina.combat.deaths)
```

**Format**

A data frame with 72 observations on the following 5 variables.

month.year a character vector for the year

us a numeric vector for the number of US combat deaths

svn a numeric vector for the number of South Vietnamese combat deaths

third a numeric vector for the number of third party combat deaths

enemy a numeric vector for the number of enemy combat deaths

**Details**

None

**Source**

*Unclassified Statistics on Southeast Asia* (1972), Department of Defense, OASD (Comptroller), Directorate for Information Operations.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(indochina.combat.deaths)
```

---

 ivy.league.football.1965

*Hartigan (1975) Ivy League Football 1965*


---

### Description

The table contains the scores for the first half of the 1965 season of the Ivy League football games. This is Table 12.1 in Chapter 12 of Hartigan (1975) on page 217.

### Usage

```
data(ivy.league.football.1965)
```

### Format

A data frame with 40 observations on the following 4 variables.

home.team a character vector for the home team code

opponent.team a character vector for the opponent team code

home.score a numeric vector for the home team score

opponent.score a numeric vector for the opponent team score

### Details

The following teams are represented in the table

Brown	BN
Bucknell	BL
Colgate	CE
Connecticut	CT
Columbia	CA
Dartmouth	DN
Harvard	HD
New Hampshire	NH
Holy Cross	HO
Lafayette	LE
Pennsylvania	PA
Princeton	PN
Rhode Island	RI
Rutgers	RS
Tufts	TS
Yale	YE

Hartigan applies a joining algorithm to this data set.

**Source**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(ivy.league.football.1965)
```

---

```
jigsaw.puzzle.measures
```

*Hartigan (1975) Jigsaw Puzzle Measurements*

---

**Description**

A table of measurements for each piece in a jigsaw puzzle. This is Table 3.1 in Chapter 3 of Hartigan (1975) on page 76.

**Usage**

```
data(jigsaw.puzzle.measures)
```

**Format**

A data frame with 20 observations on the following 13 variables.

piece a numeric vector for the number of the piece.

L1 a numeric vector for length of the line between the corners.

I1 a numeric vector for the maximum deviation of the line into the piece

O1 a numeric vector for the maximum deviation of the line out of the piece.

L2 a numeric vector for the length of the line between the corners

I2 a numeric vector for the maximum deviation of the line into the piece

O2 a numeric vector for the maximum deviation of the line out of the piece.

L3 a numeric vector for the length of the line between the corners.

I3 a numeric vector for the maximum deviation of the line into the piece

O3 a numeric vector for the maximum deviation of the line out of the piece.

L4 a numeric vector for the length of the line between the corners.

I4 a numeric vector for the maximum deviation of the line into the piece

O4 a numeric vector for the maximum deviation of the line out of the piece.



**Details**

A jigsaw puzzle comprises 20 pieces, arranged in a regular array and numbered as follows:

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

Each piece is roughly rectangular. The corners of the piece are called its vertices, and the sides are called its edges. The four edges of each piece are numbered consecutively, starting from the top and moving clockwise.

For each piece, three measurements were made on each of the four edges, estimating the length of the side, and the amount by which the edge cuts into or juts out of the line joining the two vertices on that side. The measurements are in hundredths of an inch.

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(jigsaw.puzzle.measures)
```

---

```
languages.spoken.europe
```

*Hartigan (1975) Languages Spoken in Europe*

---

### Description

The table presents the percentage of the population who claimed to speak a language well enough to be understood. This is Table 15.10 in Chapter 15 of Hartigan (1975) on page 290.

### Usage

```
data(languages.spoken.europe)
```

### Format

A data frame with 16 observations on the following 13 variables.

country a character vector for the country

finnish a numeric vector for speakers of Finnish

swedish a numeric vector for speakers of Swedish

danish a numeric vector for speakers of Danish  
norwegian a numeric vector for speakers of Norwegian  
english a numeric vector for speakers of English  
german a numeric vector for speakers of German  
dutch a numeric vector for speakers of Dutch  
flemish a numeric vector for speakers of Flemish  
french a numeric vector for speakers of French  
italian a numeric vector for speakers of Italian  
spanish a numeric vector for speakers of Spanish  
portuguese a numeric vector for speakers of Portuguese

### Details

Hartigan suggests the use of direct joining for this data set.

### Source

*A Survey of Europe Today*, The Readers' Digest Association Ltd, London.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(languages.spoken.europe)
```

---

```
leukemia.youth.mortality.1956.1967
```

*Hartigan (1975) Mortality Rates from Leukemia AMong Children*

---

### Description

The table contains the mortality rates from Leukemia recorded per million children between the ages of 0 to 14 and between 1956 and 1967. This is Table 18.1 in Chapter 15 of Hartigan (1975) on page 334.

### Usage

```
data(leukemia.youth.mortality.1956.1957)
```

**Format**

A data frame with 18 observations on the following 13 variables.

country a character vector for the country name

y.1956 a numeric vector for the mortality rates in 1956

y.1957 a numeric vector for the mortality rates in 1957

y.1958 a numeric vector for the mortality rates in 1958

y.1959 a numeric vector for the mortality rates in 1959

y.1960 a numeric vector for the mortality rates in 1960

y.1961 a numeric vector for the mortality rates in 1961

y.1962 a numeric vector for the mortality rates in 1962

y.1963 a numeric vector for the mortality rates in 1963

y.1964 a numeric vector for the mortality rates in 1964

y.1965 a numeric vector for the mortality rates in 1965

y.1966 a numeric vector for the mortality rates in 1966

y.1967 a numeric vector for the mortality rates in 1967

**Details**

Hartigan suggests using the adding algorithm on this data set to make a prediction.

**Source**

Spier (1972). Relationship between age of death to calendar year of estimated maximum leukemia mortality rate, *HSMHA Health Report*, 87, 61 - 70.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(leukemia.youth.mortality.1956.1967)
```

---

life.expectancy.1971    *Hartigan (1975) Expectations of Life by Country, Age and Sex*

---

**Description**

A table with remaining life expectancies for males and females of sampled ages. This is Table 4.10 in Chapter 14 of Hartigan (1975) on page 101.

**Usage**

```
data(life.expectancy.1971)
```

**Format**

A data frame with 31 observations on the following 10 variables.

country a character vector for the country

year a numeric vector for the year in which the data were computed

m0 a numeric vector for the remaining life expectancies for a male of age 0

m25 a numeric vector for the remaining life expectancies for a male of age 25

m50 a numeric vector for the remaining life expectancies for a male of age 50

m75 a numeric vector for the remaining life expectancies for a male of age 75

f0 a numeric vector for the remaining life expectancies for a female of age 0

f25 a numeric vector for the remaining life expectancies for a female of age 25

f50 a character vector for the remaining life expectancies for a female of age 50

f75 a numeric vector for the remaining life expectancies for a female of age 75

**Details**

None.

**Source**

Keylitz, N. and Flieger, W. (1971), *Population*, Freeman.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(life.expectancy.1971)
```

---

life.expectancy.age.sex.1971

*Hartigan (1975) Expectation of Life in Various Cities by Age and Sex*

---

### Description

Table defines life expectancy by attained age and sex in various cities in the specified years. This is Table 10.3 in Chapter 10 of Hartigan (1975) on page 182.

### Usage

```
data(life.expectancy.age.sex.1971)
```

### Format

A data frame with 16 observations on the following 10 variables.

city a character vector for the city

year a numeric vector for the year of census

m00 a numeric vector for the male expectancy with attained age 0

m25 a numeric vector for the male expectancy with attained age 25

m50 a numeric vector for the male expectancy with attained age 50

m75 a numeric vector for the male expectancy with attained age 75

f00 a numeric vector for the female expectancy with attained age 0

f25 a numeric vector for the female expectancy with attained age 25

f50 a numeric vector for the female expectancy with attained age 50

f75 a numeric vector for the female expectancy with attained age 75

### Details

This data set can be applied to the triads-leader algorithm.

### Source

Keyfitz, N. and Flieger, W. (1971) *Population*, Freeman, San Francisco.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(life.expectancy.age.sex.1971)
```

---

`linguistic.relatedness`*Hartigan (1975) Relatedness Values of Selected Words*

---

**Description**

Frequencies with which a pair is judged more highly related than other pairs, over many triads and subjects. This is Table 10.4 in Chapter 10 of Hartigan (1975) on page 184.

**Usage**

```
data(linguistic.relatedness)
```

**Format**

A data frame with 6 observations on the following 7 variables.

`word` a character vector for the

`the` a numeric vector for the frequency with which words are related to 'the'

`boy` a numeric vector for the frequency with which words are related to 'boy'

`has` a numeric vector for the frequency with which words are related to 'has'

`lost` a numeric vector for the frequency with which words are related to 'lost'

`a` a numeric vector for the frequency with which words are related to 'a'

`dollar` a numeric vector for the frequency with which words are related to 'dollar'

**Details**

This is an unusual data set to be used with the triads-leader algorithm.

**Source**

Levelt, W. J. M (1967). Psychological representations of syntactic structures, in *The Structure and Psychology of Language*, T. G. Bever and W. Weksel, eds, Holt, Rinehart and Winston, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(linguistic.relatedness)
```

---

mammal.dentition	<i>Hartigan (1975) Dentition of Animals</i>
------------------	---

---

### Description

The table contains for each animal the number of teeth in each major grouping. This is Table 9.1 in Chapter 9 of Hartigan (1975) on page 170.

### Usage

```
data(mammal.dentition)
```

### Format

A data frame with 66 observations on the following 9 variables.

name a character vector for the name of the animal  
top.i a numeric vector for the number of top incisors  
bottom.i a numeric vector for the number of bottom incisors  
top.c a numeric vector for the number of top canines  
bottom.c a numeric vector for the number of bottom canines  
top.pm a numeric vector for the the number of top premolars  
bottom.pm a numeric vector for the number of bottom premolars  
top.m a numeric vector for the number of top molars  
bottom.m a numeric vector for the number of bottom molars

### Details

Hartigan uses this table to illustrate a tree-leader algorithm.

### Source

Palmer, E. I. (1957). *Fieldbook of Mammals*, Dutton, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(mammal.dentition)
```



---

minor.planets.1961      *Hartigan (1975) Minor Planets*

---

### Description

Some minor planets may have been sighted more than once. In the data frame, sightings thought to be of the same planet are listed together. This is Table 1.1 in the Introduction of Hartigan (1975) on page 2.

### Usage

```
data(minor.planets.1961)
```

### Format

A data frame with 19 observations on the following 4 variables.

name a character vector for the year of sighting and astronomer initials

node a numeric vector for the angle in degrees in the earth plane at which the minor planet crosses the earth's orbit

inclination a numeric vector for the angle in degrees between the plane of the earth's orbit and the plane of the planet's orbit

axis a numeric vector for the maximum distance of the minor planet from the sun in astronomical units

### Details

None.

### Source

*Elements of Minor Planets* (1961), University of Cincinnati Observatory

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(minor.planets.1961)
```

---

mutation.distances.1967

*Hartigan (1975) Mutation Distances*

---

### **Description**

The table contains mutation distance between pairs of species. This is Table 11.12 in Chapter of Hartigan (1975) on page 209.

### **Usage**

```
data(mutation.distances.1967)
```

### **Format**

A data frame with 20 observations on the following 22 variables.

code a character vector for specifies identifier  
species a character vector fir the species name  
s.1 a numeric vector for distance to species 1  
s.2 a numeric vector for distance to species 2  
s.3 a numeric vector for distance to species 3  
s.4 a numeric vector for distance to species 4  
s.5 a numeric vector for distance to species 5  
s.6 a numeric vector for distance to species 6  
s.7 a numeric vector for distance to species 7  
s.8 a numeric vector for distance to species 8  
s.9 a numeric vector for distance to species 9  
s.10 a numeric vector for distance to species 10  
s.11 a numeric vector for distance to species 11  
s.12 a numeric vector for distance to species 12  
s.13 a numeric vector for distance to species 13  
s.14 a numeric vector for distance to species 14  
s.15 a numeric vector for distance to species 15  
s.16 a numeric vector for distance to species 16  
s.17 a numeric vector for distance to species 17  
s.18 a numeric vector for distance to species 18  
s.19 a numeric vector for distance to species 19  
s.20 a numeric vector for distance to species 20

**Details**

The distance is defined by the number of positions in the protein molecule cytochrome-c where the two species have different amino acids. Hartigan uses the single-linkage algorithm on this dataset.

**Source**

Fitch and Margoliash (1967) *Science*

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(mutation.distances.1967)
```

---

nails.screws

*Hartigan (1975) Nails and Screws*

---

**Description**

The table contains the attributes for a sample of nails and screws. This is Table 12.7 in Chapter 12 of Hartigan (1975) on page 228.

**Usage**

```
data(nails.screws)
```

**Format**

A data frame with 24 observations on the following 7 variables.

name a character vector for the name of the object

threaded a factor for the presence of threads with levels N Y

head a factor for the type of head with levels F O R U Y

indentation a factor for the head indentation with levels L N T

bottom a factor for the type of bottom with levels F S

length a numeric vector for the length in half inches

brass a factor that determines if the object is made of brass with levels N Y

**Details**

All the attributes, with the exception of length, are factors. The factor values for the threaded variable are as follows.

Y yes  
N no

The factor values for the head variable are as follows.

F flat  
U cut  
O cone  
R round  
Y cylinder

The factor values for the head indentation variable are as follows.

N none  
T star  
L slit

The value values for the brass variable are as follows

Y yes  
N no

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(nails.screws)
```

---

new.haven.school.scores

*Hartigan (1975) Achievement Test Schores, New Haven Schools*

---

### Description

The measurements are in years and months of national averages. There are ten months in the school year. At the beginning of fourth grades, the national average score is 4.0. This is Table 5.1 in Chapter 5 of Hartigan (1975) on page 118.

**Usage**

```
data(new.haven.school.scores)
```

**Format**

A data frame with 25 observations on the following 5 variables.

school a character vector for the name of the school

reading.4 a numeric vector for the reading scores for fourth grade

arithmetic.4 a numeric vector for the arithmetic scores for fourth grade

reading.6 a numeric vector for for the reading scores for sixth grade

arithmetic.6 a numeric vector for the arithmetic scores for sixth grade

**Details**

None.

**Source**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(new.haven.school.scores)
```

---

nutrients.meat.fish.fowl.1959

*Hartigan (1975) Nutrients in Meat, Fish and Fowl*

---

**Description**

A table with the nutrient levels in meat, fish and fowl. Nutrient levels were measured in a 3 ounce portion of various foods. This is Table 4.1 in Chapter 4 of Hartigan (1975) on page 86.

**Usage**

```
data(nutrients.meat.fish.fowl.1959)
```

**Format**

A data frame with 27 observations on the following 6 variables.

name a character vector for the food

energy a numeric vector for the number of calories

protein a numeric vector for the amount of protein in grams

fat a numeric vector for the amount of fat in grams

calcium a numeric vector for the amount of calcium in milligrams

iron a numeric vector for the amount of iron in milligrams

**Details**

None.

**Source**

*The Yearbook of Agriculture* (1959), The United States Department of Agriculture, Washington, DC.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(nutrients.meat.fish.fowl.1959)
```

---

ohio.croplands.1949     *Hartigan (1975) Ohio Croplands*

---

**Description**

The table presents the percentage of cropland devoted to various crops in Ohio counties. This is Table 15.7 in Chapter 15 of Hartigan( 1975) on page 287.

**Usage**

```
data(ohio.croplands.1949)
```

**Format**

A data frame with 15 observations on the following 8 variables.

county a character vector for the county

corn a numeric vector for the percentage of cropland devoted to corn

mixed a numeric vector for the percentage of cropland devoted to mixed crop

wheat a numeric vector for the percentage of cropland devoted to wheat

oats a numeric vector for the percentage of cropland devoted to oats

barley a numeric vector for the percentage of cropland devoted to varley

soy a numeric vector for the percentage of cropland devoted to soy

hay a numeric vector for the percentage of cropland devoted to hay

**Details**

Hartigan suggest the use of direct joining with this data set.

**Source**

*U.S. Census of Agriculture*, 1949.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(ohio.croplands.1949)
```

---

```
olympic.track.1896.1964
```

*Hartigan (1975) Olympic Track 1896 to 1964*

---

**Description**

Olympic track times, in tenths of a second, were recorded orver the years. This is Table 6.1 in Chapter 6 of Hartigan (1975) on page 131.

**Usage**

```
data(olympic.track.1896.1964)
```

**Format**

A data frame with 16 observations on the following 8 variables.

year a character vector for the year

t.100m a numeric vector for the winning time in the 100 m

t.200m a numeric vector for the winning time in the 200 m

t.400m a numeric vector for the winning time in the 400 m

t.800m a numeric vector for the winning time in the 800 m

t.1500m a numeric vector for the winning time in the 1500 m

t.5000m a numeric vector for the winning time in the 5000 m

t.10000m a numeric vector for the winning time in the 10000 m

**Details**

None.

**Source**

*The World Almanac* (1966), New York World-Telegram, New York,

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(olympic.track.1896.1964)
```

---

```
physical.measure.correlations
```

*Hartigan (1975) Correlation Between Physical Measurements*

---

**Description**

The table contains the correlations between various body parts. This is Table 17.1 in Chapter 17 of Hartigan (1975) on page 314.

**Usage**

```
data(physical.measure.correlations)
```



**Format**

A data frame with 7 observations on the following 7 variables.

hl a numeric vector for the correlations with head length  
hb a numeric vector for the correlations with head breadth  
fb a numeric vector for the correlations with face breadth  
ft a numeric vector for the correlations with foot  
fm a numeric vector for the correlations with forearm  
ht a numeric vector for the correlations with height  
fl a numeric vector for the correlations with finger length

**Details**

Hartigan suggests performing factor analysis on this data set to determine the minimum number of principal components. In addition, a joining algorithm can be performed on the data set. Note that the data frame has the variable names as row names. It can be used directly by the `eigen` function.

**Source**

Pearson, K. (1901). On lines and planes of closest fit to points in space. *Philosophical Magazine*, 559 - 572.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(physical.measure.correlations)
```

---

```
planet.earth.distances.1970
```

*Hartigan (1975) Planets and Moons*

---

**Description**

From astronomical knowledge of 1970, a table of planetary moons was compiled. This is the bottom portion of Table 5.5 in Chapter 5 of Hartigan (1975) on page 122.

**Usage**

```
data(planet.earth.distances.1970)
```

**Format**

A data frame with 8 observations on the following 5 variables.

name a character vector for the name of the planet

distance a numeric vector for its distance from the sun in thousands of miles

diameter a numeric vector for its diameter in miles

period a numeric vector for the period of its orbit in hours

mass a numeric vector for the mass, relative to the earth

**Details**

None.

**Source**

Moore, P. (1970). *The Atlas of the Universe*, Rand McNally, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(planet.earth.distances.1970)
```

---

planets.moons.1970      *Hartigan (1975) Planets and Moons*

---

**Description**

From astronomical knowledge of 1970, a table of planetary moons was compiled. This is the top portion of Table 5.5 in Chapter 5 of Hartigan (1975) on page 122.

**Usage**

```
data(planets.moons.1970)
```

**Format**

A data frame with 31 observations on the following 4 variables.

planet.moon a character vector for the planet and the number of the moon

distance a numeric vector for the distance in thousands of miles between the moon and the planet

diameter a numeric vector for the diameter in miles of the moon

period a numeric vector for the period, in days, of the orbit of the moon about the plane

**Details**

None.

**Source**

Moore, P. (1970). *The Atlas of the Universe*, Rand McNally, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(planets.moons.1970)
```

---

portable.typewriters    *Hartigan (1975) Portable Typewriters*

---

**Description**

The table contains the features in a collection of portable typewriters. This is Table 10.5 in Chapter 10 of Hartigan (1975) on page 186.

**Usage**

```
data(portable.typewriters)
```

**Format**

A data frame with 20 observations on the following 21 variables.

model a character vector for the typewriter model

HT a numeric vector for the height in inches

WH a numeric vector for the width in inches

DH a numeric vector for the depth in inches

WT a numeric vector for the weight in pounds

PL a numeric vector for the platen length

KS a numeric vector for the number of keys

PE a factor for the pica or elite type with levels 1

TA a factor for the availability of tabulator with levels 0 1

TP a factor for the availability of touch pressure control with levels 0 1

PR a factor for the availability of platen release with levels 0 1

HH a factor for the availability of horizontal half spacing with levels 0 1  
 VH a factor for the availability of vertical half spacing with levels 0 1  
 PI a factor for the availability of page end indicator with levels 0 1  
 PG a factor for the availability of paper guide with levels 0 1  
 PB a factor for the availability of paper bail with levels 0 1  
 PS a factor for the availability of paper support with levels 0 1  
 EP a factor for the availability of erasure plate with levels 0 1  
 TC a factor for the availability of two carriage releases with levels 0 1  
 MR a factor for the availability of margin release with levels 0 1  
 CL a factor for the availability of carriage lock with levels 0 1

### Details

Hartigan suggests that the triads algorithm be used with this data set. The factor variables are binary variables. If the value is 1, then the associated feature is available. If the value is 0, then the associated feature is not available.

### Source

*Consumers' Reports Buying Guide* (1967), Consumers' Union, Mount Vernon, NY.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(portable.typewriters)
```

---

```
rda.meat.fish.fowl.1959
```

*Hartigan (1975) Nutrients in Meat, Fish and Fowl Percent RDA*

---

### Description

A table with the nutrient levels in meat, fish and fowl. Nutrient levels were measured in a 3 ounce portion of various foods. Values are percentages of recommended daily allowances. This is Table 4.2 in Chapter 4 of Hartigan (1975) on page 87.

### Usage

```
data(rda.meat.fish.fowl.1959)
```

### Format

A data frame with 27 observations on the following 6 variables.

name a character vector for the food

energy a numeric vector for the number of calorie

protein a numeric vector for the amount of protein

fat a numeric vector for the amount of fat

calcium a numeric vector for the amount of calcium

iron a numeric vector for the amount of iron

### Details

None.

### Source

*The Yearbook of Agriculture* (1959), The United States Department of Agriculture, Washington, DC.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(rda.meat.fish.fowl.1959)
```

---

sample.mammals.milk.1956

*Hartigan (1975) Mammals Milk*

---

### Description

Selected animals have been clustered by similarity of percentage constituents in milk. This is Table 1.2 in the Introduction of Hartigan (1975) on page 6.

### Usage

```
data(sample.mammals.milk.1956)
```

**Format**

A data frame with 16 observations on the following 5 variables.

name a character vector for the name of the animals

water a numeric vector for the water content in the milk sample

protein a numeric vector for the amount of protein in the milk sample

fat a numeric vector for the fat content in the milk sample

lactose a numeric vector for the amount of lactose in the milk sample

**Details**

None

**Source**

Spector, W. S. (1956). *Handbook of Biological Data*, Saunders, London

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(sample.mammals.milk.1956)
```

---

```
sample.stock.yields.1959.1969
```

*Hartigan (1975) Yield of Stocks*

---

**Description**

The table contains the dividend by average price for each year and for a sample of stocks. This is Table 11.13 in Chapter 11 of Hartigan (1975) on page 210.

**Usage**

```
data(sample.stock.yields.1959.1969)
```

### Format

A data frame with 34 observations on the following 12 variables.

stock a character vector for the company name  
y.1959 a numeric vector for the dividend yield in 1959  
y.1960 a numeric vector for the dividend yield in 1960  
y.1961 a numeric vector for the dividend yield in 1961  
y.1962 a numeric vector for the dividend yield in 1962  
y.1963 a numeric vector for the dividend yield in 1963  
y.1964 a numeric vector for the dividend yield in 1964  
y.1965 a numeric vector for the dividend yield in 1965  
y.1966 a numeric vector for the dividend yield in 1966  
y.1967 a numeric vector for the dividend yield in 1967  
y.1968 a numeric vector for the dividend yield in 1968  
y.1969 a numeric vector for the dividend yield in 1969

### Details

Hartigan proposes applying the single linkage algorithm to this data set.

### Source

Moody's *Handbook of Common Stocks*

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(sample.stock.yields.1959.1969)
```

---

sample.us.city.crime.1970

*Hartigan (1975) City Crime*

---

### Description

A list of cities and the number of crimes per 100,000 population, as of 1970. This is Table 1.1 in Chapter 1 of Hartigan (1975) on page 28.

### Usage

```
data(sample.us.city.crime.1970)
```

### Format

A data frame with 16 observations on the following 8 variables.

city a character vector for the names of the cities

murder a numeric vector for the murder rates

rape a numeric vector for the rape rates

robbery a numeric vector for the robbery rates

assault a numeric vector for the assault rates

burglary a numeric vector for the burglary rates

larceny a numeric vector for the larceny rates

auto a numeric vector for the auto crime rates

### Details

None.

### Source

*United States Statistical Abstracts* (1970).

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(sample.us.city.crime.1970)
```



---

student.questionnaire *Hartigan (1975) Student Questionnaire*

---

### Description

The table contains student responses to a questionnaire about a data analysis course. This is Table 12.4 in Chapter 12 of Hartigan (1975) on page 224.

### Usage

```
data(student.questionnaire)
```

### Format

A data frame with 31 observations on the following 10 variables.

question a numeric vector for the question number

text a character vector for the question text

s.1 a numeric vector for the response from student 1

s.2 a numeric vector for the response from student 2

s.3 a numeric vector for the response from student 3

s.4 a numeric vector for the response from student 4

s.5 a numeric vector for the response from student 5

s.6 a numeric vector for the response from student 6

s.7 a numeric vector for the response from student 7

s.8 a numeric vector for the response from student 8

### Details

Student responses to the questionnaires are evaluated using the following scores.

- |   |                   |
|---|-------------------|
| 1 | strongly disagree |
| 2 | disagree          |
| 3 | neutral           |
| 4 | agree             |
| 5 | strongly agree    |

Hartigan applies the adding algorithm to this data set.

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(student.questionnaire)
```

---

un.votes.1969.1970      *Hartigan (1975) Selected Votes in the United Nations*

---

**Description**

The table contains the votes for selected propositions by country in the United Nations between 1969 and 1970. This is Table 16.5 in Chapter 16 of Hartigan (1975) on page 306.

**Usage**

```
data(un.votes.1969.1970)
```

**Format**

A data frame with 23 observations on the following 11 variables.

country a character vector for the country name  
 p.1 a factor for proposition 1 with levels A N Y  
 p.2 a factor for proposition 2 with levels A N Y  
 p.3 a factor for proposition 3 with levels A N Y  
 p.4 a factor for proposition 4 with levels A N Y  
 p.5 a factor for proposition 5 with levels A N Y  
 p.6 a factor for proposition 6 with levels A N Y  
 p.7 a factor for proposition 7 with levels A N Y  
 p.8 a factor for proposition 8 with levels A N Y  
 p.9 a factor for proposition 9 with levels A N Y  
 p.10 a factor for proposition 10 with levels A N Y

**Details**

The propositions that were voted on were as follows.

p.1	to adopt USSR proposal to delete item on Korean unification
p.2	to call upon the UK to use force against Rhodesia
p.3	to declare the China admission question an important question
p.4	to recognize mainland China and expel Formosa
p.5	to make a study commission on China admission important
p.6	to forma a study comssion on Portuguese colonialism

- p.7 convention on no statutory limit on ware crimes
- p.8 condemn Portuguese colonialism
- p.9 to defer consideration of South Africa expulsion
- p.10 South Africa expulsion is important question

The factor levels are the outcomes for the proposition. Y implies yes, N is no and A is abstain..

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(un.votes.1969.1970)
```

---

us.car.repair.1969      *Hartigan (1975) Frequency of Car Repairs*

---

### Description

The table contains the frequency of car repairs in 1969. Plus means above average. Minus means below average. This is Chapter 9 Table 9.4 in Chapter 9 of Hartigan (1975) on page 174.

### Usage

```
data(us.car.repair.1969)
```

### Format

A data frame with 33 observations on the following 14 variables.

model a character vector for the model of the vehicle

BR a factor for break system with levels - +

FU a factor for fuel system with levels - +

EL a factor for electrical with levels - +

EX a factor for exhaust with levels - +

ST a factor for steering with levels - +

EM a factor for engine, mechanical with levels - +

RS a factor for rattles and squeeks with levels - +

RA a factor for real axle with levels - +  
 RU a factor for rust with levels - +  
 SA a factor for shock absorbers with levels - +  
 TC a factor for transmission, clutch with levels - +  
 WA a factor for wheel alignment with levels - +  
 OT a factor for other with levels - +

### Details

This table is used to illustrate the tree-leader algorithm.

### Source

*Consumer Reports Buying Guide* (1969)

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(us.car.repair.1969)
```

---

us.civil.war.battles    *Hartigan (1975) Civil War Battles in Chronological Order*

---

### Description

This table contains the Union and Confederate forces and numbers shot This is Table 5.4 in Chapter 5 Hartigan (1975) on page 121.

### Usage

```
data(us.civil.war.battles)
```

### Format

A data frame with 46 observations on the following 5 variables.

battle a character vector for the battle names

union.forces a numeric vector for the Union forces deployed

union.shot a numeric vector for the Union soldiers shot

confederate.forces a numeric vector for the Confederate forces deployed

confederate.shot a numeric vector for the Confederate soldiers shot

**Details**

The data are in chronological order.

**Source**

Livermore, T L. (1957). *Numbers and Losses in the Civial War*, Indiana University Press, Bloomington.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(us.civil.war.battles)
```

---

```
us.congressional.bills
```

*Hartigan (1975) Congressman b y Bills*

---

**Description**

The table contains the behavior of various bill sponsors in the 90th Congress. This is Table 13.7 in Chapter 13 of Hartigan (1975) on page 242.

**Usage**

```
data(us.congressional.bills)
```

**Format**

A data frame with 17 observations on the following 16 variables.

sponsor a character vector for the congressman sponsor

b.1 a factor for the congressman behavior for bill 1 with levels 1 5 7 8

b.2 a factor for the congressman behavior for bill 2 with levels 1 5 6 7

b.3 a factor for the congressman behavior for bill 3 with levels 1 5 6 7

b.4 a factor for the congressman behavior for bill4 with levels 1 7

b.5 a factor for the congressman behavior for bill 5 with levels 1 6 7

b.6 a factor for the congressman behavior for bill 6 with levels 1 6 7

b.7 a factor for the congressman behavior for bill 7 with levels 1 6 7

b.8 a factor for the congressman behavior for bill 8 with levels 1 6 7

b.9 a factor for the congressman behavior for bill 9 with levels 1 6 9

- b.10 a factor for the congressman behavior for bill 10 with levels 1 6 9
- b.11 a factor for the congressman behavior for bill 11 with levels 1 6 9
- b.12 a factor for the congressman behavior for bill 12 with levels 1 6 9
- b.13 a factor for the congressman behavior for bill 13 with levels 1 6 9
- b.14 a factor for the congressman behavior for bill 14 with levels 1 6 9
- b.15 a factor for the congressman behavior for bill 15 with levels 1 6 9

### Details

The bills, sponsoring congressmen and bill titles are as follows.

b.1	Aspinall	Authorize Biscayne National Monument in Florida
b.2	Perkins	Promote health and safety in building trades
b.3	Patman	Sr extend 2 years auth. reg. interest and dividend rates
b.4	Dingell	Rel Dev fish protein concentrate
b.5	Perkins	Establish commission on Negro history and culture
b.6	Aspinall	Designate parts of Morris City, NJ, as wilderness
b.7	Udall	Provide overtime and standby pay for transportation department
b.8	Edwards	Amend bill for relief of sundry claimants
b.9	Gross	Amend omnibus claims bill
b.10	Gross	Strike title 8 of omnibus claims bill
b.11	Hall	Strike title 9 of omnibus claims bill
b.12	Gross	Strike title 10 of omnibus claims bill
b.13	Hall	Strike title 11 of omnibus claims bill
b.14	Talcott	Strike title 14 of omnibus claims bill
b.15	Poage	Take FD and AG ACT AMD SPKRS TBLE AGREE S CONF

The behavior is represented by a factor with the following values

- 1 yes
- 2 pair yes
- 3 announced yes
- 4 announced no
- 5 pair no
- 6 no
- 7 general pair
- 8 abstain
- 9 absent
- 0 sponsor absent

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(us.congressional.bills)
```

---

```
us.food.cost.nutrients.1959
```

*Hartigan (1975) Cost and Nutrient Contribution for Selected Foods*

---

**Description**

The table contains the cost and nutrient content, in percent daily allowance, of various foods reported in 1959. This is Table 8.5 in Chapter 8 of Hartigan (1975) on page 160.

**Usage**

```
data(us.food.cost.nutrients.1959)
```

**Format**

A data frame with 10 observations on the following 8 variables.

food a character vector for the food name

cost a numeric vector for the cost of serving in U.S. cents

size a character vector for for the portion size

protein a numeric vector for % recommended daily allowance of protein

iron a numeric vector for for % recommended daily allowance of iron

thiamine a numeric vector for for % recommended daily allowance of thiamine

riboflavin a numeric vector for for % recommended daily allowance of riboflavin

niacin a numeric vector for for % recommended daily allowance of niacin

**Details**

The table is used to construct trees and distances as described in Hartigan (1975).

**Source**

*Yearbook of Agriculture* (1959).

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(us.food.cost.nutrients.1959)
```

---

```
us.links.between.states
```

*Hartigan (1975) Links Between States*

---

**Description**

The table defines the neighbors for each state. This is Table 11.10 in Chapter 11 of Hartigan (1975) on page 207.

**Usage**

```
data(us.links.between.states)
```

**Format**

A data frame with 50 observations on the following 11 variables.

code a character vector for the state code

name a character vector for the state name

neighbors a numeric vector for the number of neighboring states

n.1 a character vector for the first neighbor

n.2 a character vector for the second neighbor

n.3 a character vector for the third neighbor

n.4 a character vector for the fourth neighbor

n.5 a character vector for the fifth neighbor

n.6 a character vector for the sixth neighbor

n.7 a character vector for the seventh neighbor

n.8 a character vector for the eighth neighbor

**Details**

Hartigan combines this data set with the per capita data set in Table 11.9 and applies the single linkage algorithm.

**Source**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>



## References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

## Examples

```
data(us.links.between.states)
```

---

```
us.per.capita.income.1964
```

*Hartigan (1975) U.S. Per Capita Income in Dollars 1964*

---

## Description

The table contains the per capita income in the United States in 1964. This is Table 11.9 in Chapter 11 of Hartigan (1975) on page 206

## Usage

```
data(us.per.capita.income.1964)
```

## Format

A data frame with 50 observations on the following 3 variables.

code a character vector for the state codes

name a character vector for the state names

income a numeric vector for the income per capita

## Details

Hartigan applies density contour trees and single linkage clustering to this data set.

## Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

## References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

## Examples

```
data(us.per.capita.income.1964)
```

---

us.president.vote.1900.1968

*Hartigan (1975) Republican Vote for President*

---

### Description

The table contains the Republican percentage of the Presidential vote over 18 elections and for southern states. This is Table 14.1 in Chapter 14 of Hartigan (1975) on page 252.

### Usage

```
data(us.president.vote.1900.1968)
```

### Format

A data frame with 16 observations on the following 20 variables.

code a character vector for the state code

state a character vector for the state name

y.1900 a numeric vector for the Republican percentage in 1900

y.1904 a numeric vector for the Republican percentage in 1904

y.1908 a numeric vector for the Republican percentage in 1908

y.1912 a numeric vector for the Republican percentage in 1912

y.1916 a numeric vector for the Republican percentage in 1916

y.1920 a numeric vector for the Republican percentage in 1920

y.1924 a numeric vector for the Republican percentage in 1924

y.1928 a numeric vector for the Republican percentage in 1928

y.1932 a numeric vector for the Republican percentage in 1932

y.1936 a numeric vector for the Republican percentage in 1936

y.1940 a numeric vector for the Republican percentage in 1940

y.1944 a numeric vector for the Republican percentage in 1944

y.1948 a numeric vector for the Republican percentage in 1948

y.1952 a numeric vector for the Republican percentage in 1952

y.1956 a numeric vector for the Republican percentage in 1956

y.1960 a numeric vector for the Republican percentage in 1960

y.1964 a numeric vector for the Republican percentage in 1964

y.1968 a numeric vector for the Republican percentage in 1968

### Details

Hartigan suggests that the direct splitting algorithm is applied to this data set.

**Source**

Peterson, S. (1969). *A Statistical History of the American Presidential Elections*, Ungar, New York  
SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(us.president.vote.1900.1968)
```

---

```
us.sector.profitability.1959.1968
```

*Hartigan (1975) Profitability of U.S. Economic Sectors*

---

**Description**

The table contains the profit as a percentage of stockholder's equity for various economic sectors for the years 1959 through 1968. This is Table 14.12 in Chapter 14 of Hartigan (1975) on page 266.

**Usage**

```
data(us.sector.profitability.1959.1968)
```

**Format**

A data frame with 24 observations on the following 12 variables.

code a character vector for the sector code

sector a character vector for the sector name

y.1959 a numeric vector for the profits in year 1959

y.1960 a numeric vector for the profits in year 1960

y.1961 a numeric vector for the profits in year 1961

y.1962 a numeric vector for the profits in year 1962

y.1963 a numeric vector for the profits in year 1963

y.1964 a numeric vector for the profits in year 1964

y.1965 a numeric vector for the profits in year 1965

y.1966 a numeric vector for the profits in year 1966

y.1967 a numeric vector for the profits in year 1967

y.1968 a numeric vector for the profits in year 1968

**Details**

Hartigan suggests that the direct splitting algorithm be applied to this data set.

**Source**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(us.sector.profitability.1959.1968)
```

---

```
us.south.demographics.1965
```

*Hartigan (1975) Demographic Data for the South*

---

**Description**

A table of demographic information for southern states for the period 1960 to 1965. This is Table 2.2 in Chapter 2 of Hartigan (1975) on page 59.

**Usage**

```
data(us.south.demographics.1965)
```

**Format**

A data frame with 16 observations on the following 24 variables.

state a character vector for an abbreviation for the states

mean.altitude a numeric vector for the mean altitude above sea level, in tens of feet

mean.temperature a numeric vector for the mean annual temperature, in degrees Fahrenheit

mean.precipitation a numeric vector for the mean annual precipitation, in inches

population.density a numeric vector for the number of persons per square mile.

african.americans a numeric vector for the percentage of African-Americans

median.age a numeric vector for the median age in years

urban.population a numeric vector for the percentage urban population

births a numeric vector for the number of births per 1000 population

rural.population a numeric vector for the percentage rural farm population

manufacturing.employment a numeric vector for the percentage of employment in manufacturing

automobiles a numeric vector for the number of automobiles per 100 population

telephones a numeric vector for the number of telephones per 100 population

income a numeric vector for the average income in hundreds of dollars

federal.revenue a numeric vector for the federal revenue per 100 dollars of state and local revenue

lawyers a numeric vector for the number of lawyers per 100,000 population

doctors a character vector for the number of doctors per 100,000 population

white.infant.mortality a numeric vector for the white infant mortality per 1000 births

school.years a numeric vector for the school years completed, in tenths of a year

education.expense a numeric vector for the education expenditure per pupil in tens of dollars

sound.plumbing a numeric vector for the percentage of houses with sound plumbing.

gop.1960.president a numeric vector for the percentage Republican vote in the 1960 presidential election

gop.1964.president a numeric vector for the percentage Republican vote in the 1964 presidential election

gop.1962.1964.governor a numeric vector for the percentage Republican vote in the 1962/1964 governor elections

### Details

None.

### Source

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

### References

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

### Examples

```
data(us.south.demographics.1965)
```

---

 vervet.sleeping.groups

*Hartigan (1975) Vervet Sleeping Groups*


---

### Description

The table defines vervet sleeping groups measured over a set of dates. This is Table 7.5 in Chapter 7 of Hartigan (1975) on page 149.

### Usage

```
data(vervet.sleeping.groups)
```

### Format

A data frame with 22 observations on the following 18 variables.

date a character vector for the date in yy/mm/dd format

I a factor for adult males with levels A B C D E

II a factor for older adult males with levels A B C D

III a factor for adult males with levels A B C D

IV a factor for adult females with levels A B C D E F

V a factor for juvenile males with levels A B C D F

VI a factor for adult females with levels A B C D E

VII a factor for young juvenile females with levels A B C D E

VIII a factor for young juvenile females with levels A B C D E

IX a factor for young juvenile females with levels A B C D E

X a factor for juvenile females with levels A B C D E F G

XI a factor for subadult females with levels A B C D E

XII a factor for adult females with levels A B C D E

XIII a factor with levels A B C D E F

XIV a factor for invant male, son of IV with levels A B C D E F

XV a factor for infant male, son of XII with levels A B C D E F

XVI a factor for infant female from IV with levels A B C D E

XVII a factor with levels A B C D E

### Details

Hartigan suggests using this data set to test the ditto algorithm.

**Source**

Struhsaker, T. T. (1967). Behavior of servet monkeys and other cercopithecines, *Science* 156, 1197 - 1203.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(vervet.sleeping.groups)
```

---

```
wine.evaluation.1961.1970
```

*Hartigan (1975) Evaluation of Wines*

---

**Description**

The table contains the evaluations of various wines from 1961 to 1970. This is Table 7.1 in Chapter 7 of Hartigan (1975) on page 144.

**Usage**

```
data(wine.evaluation.1961.1970)
```

**Format**

A data frame with 15 observations on the following 12 variables.

code a character vector

name a character vector

r . 61 a factor with levels A E G

r . 62 a factor with levels A G P

r . 63 a factor with levels A D P

r . 64 a factor with levels D E G P

r . 65 a factor with levels A D G P

r . 66 a factor with levels A G

r . 67 a factor with levels A G

r . 68 a factor with levels A D G P

r . 69 a factor with levels A G

r . 70 a factor with levels G

**Details**

Hartigan uses this data set to illustrate the ditto algorithm.

**Source**

*Gourmet Magazine* (August 1971) pp 30-33.

SPAETH2 Cluster Analysis Datasets <http://people.sc.fsu.edu/~jburkardt/datasets/spaeth2/spaeth2.html>

**References**

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley, New York.

**Examples**

```
data(wine.evaluation.1961.1970)
```



# Index

## \* datasets

- acidosis.patients, 2
- airline.distances.1966, 3
- all.mammals.milk.1956, 5
- all.us.city.crime.1970, 6
- amino.acid.sequence.1972, 7
- animal.cluster.trees, 9
- birth.death.rates.1966, 10
- british.butterfly.appearance, 11
- cake.ingredients.1961, 12
- candida.oxidation.fermentation, 13
- cerci.tail.presence, 15
- ct.president.vote.1920.1964, 16
- european.foods, 17
- hardware.triads, 18
- hartigan.datasets, 19
- indian.caste.measures, 20
- indo.european.languages, 21
- indochina.combat.deaths, 22
- ivy.league.football.1965, 23
- jigsaw.puzzle.measures, 24
- languages.spoken.europe, 26
- leukemia.youth.mortality.1956.1967, 27
- life.expectancy.1971, 29
- life.expectancy.age.sex.1971, 30
- linguistic.relatedness, 31
- mammal.dentition, 32
- minor.planets.1961, 33
- mutation.distances.1967, 34
- nails.screws, 35
- new.haven.school.scores, 36
- nutrients.meat.fish.fowl.1959, 37
- ohio.croplands.1949, 38
- olympic.track.1896.1964, 39
- physical.measure.correlations, 40
- planet.earth.distances.1970, 41
- planets.moons.1970, 42
- portable.typewriters, 43
- rda.meat.fish.fowl.1959, 44
- sample.mammals.milk.1956, 45
- sample.stock.yields.1959.1969, 46
- sample.us.city.crime.1970, 48
- student.questionnaire, 49
- un.votes.1969.1970, 50
- us.car.repair.1969, 51
- us.civil.war.battles, 52
- us.congressional.bills, 53
- us.food.cost.nutrients.1959, 55
- us.links.between.states, 56
- us.per.capita.income.1964, 57
- us.president.vote.1900.1968, 58
- us.sector.profitability.1959.1968, 59
- us.south.demographics.1965, 60
- vervet.sleeping.groups, 62
- wine.evaluation.1961.1970, 63
- acidosis.patients, 2
- airline.distances.1966, 3
- all.mammals.milk.1956, 5
- all.us.city.crime.1970, 6
- amino.acid.sequence.1972, 7
- animal.cluster.trees, 9
- birth.death.rates.1966, 10
- british.butterfly.appearance, 11
- cake.ingredients.1961, 12
- candida.oxidation.fermentation, 13
- cerci.tail.presence, 15
- ct.president.vote.1920.1964, 16
- european.foods, 17
- hardware.triads, 18
- hartigan.datasets, 19
- indian.caste.measures, 20
- indo.european.languages, 21

indochina.combat.deaths, 22  
ivy.league.football.1965, 23

jigsaw.puzzle.measures, 24

languages.spoken.europe, 26  
leukemia.youth.mortality.1956.1967, 27  
life.expectancy.1971, 29  
life.expectancy.age.sex.1971, 30  
linguistic.relatedness, 31

mammal.dentition, 32  
minor.planets.1961, 33  
mutation.distances.1967, 34

nails.screws, 35  
new.haven.school.scores, 36  
nutrients.meat.fish.fowl.1959, 37

ohio.croplands.1949, 38  
olympic.track.1896.1964, 39

physical.measure.correlations, 40  
planet.earth.distances.1970, 41  
planets.moons.1970, 42  
portable.typewriters, 43

rda.meat.fish.fowl.1959, 44

sample.mammals.milk.1956, 45  
sample.stock.yields.1959.1969, 46  
sample.us.city.crime.1970, 48  
student.questionnaire, 49

un.votes.1969.1970, 50  
us.car.repair.1969, 51  
us.civil.war.battles, 52  
us.congressional.bills, 53  
us.food.cost.nutrients.1959, 55  
us.links.between.states, 56  
us.per.capita.income.1964, 57  
us.president.vote.1900.1968, 58  
us.sector.profitability.1959.1968, 59  
us.south.demographics.1965, 60

vervet.sleeping.groups, 62

wine.evaluation.1961.1970, 63