

# Package ‘syllogi’

July 23, 2025

**Type** Package

**Title** Collection of Data Sets for Teaching Purposes

**Version** 1.0.4

**Date** 2025-01-10

**Author** Jared Studyvin [aut, cre]

**Depends** R (>= 3.6.0)

**Maintainer** Jared Studyvin <studyvinstat@gmail.com>

**Description** Collection (syllogi in greek) of real and fictitious data sets for teaching purposes.

The datasets were manually entered by the author from the respective references as listed in the individual dataset documentation.

The fictions datasets are the creation of the author, that he has found useful for teaching statistics.

**License** Apache License (>= 2)

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2025-01-10 17:00:02 UTC

## Contents

alligatorDiet . . . . .	2
alligatorLength . . . . .	3
annualSales . . . . .	4
beer . . . . .	5
bighornSheep . . . . .	5
bladderCancer . . . . .	6
butterflyPlot . . . . .	7
depression . . . . .	8
dogFood . . . . .	9
federalistPapers . . . . .	10
genericData . . . . .	11

golf . . . . .	11
nutritionCancer . . . . .	12
osteosarcoma . . . . .	13
patientSatisfaction . . . . .	14
politicalIdeology . . . . .	14
schoolProgram . . . . .	15
shipDamage . . . . .	16
shipGold . . . . .	17
ski . . . . .	18
weightLoss . . . . .	18
wheat . . . . .	19

<b>Index</b>	<b>21</b>
--------------	-----------

---

alligatorDiet	<i>Study of Diets in Alligators</i>
---------------	-------------------------------------

---

## Description

Data.frame

## Usage

data(alligatorDiet)

## Format

The data frame has 16 rows and 8 variables:

**lake** Lake in Florida of the capture of the alligator.

**gender** Female (F) or Male (M).

**size** small ( $\leq 2.3$ m) or big ( $> 2.3$ m).

**fish** Number of alligators with a primary stomach contents of fish.

**invertabrate** Number of alligators with a primary stomach contents of invertebrate.

**reptile** Number of alligators with a primary stomach contents of reptile.

**bird** Number of alligators with a primary stomach contents of bird.

**other** Number of alligators with a primary stomach contents of other.

## Details

A study done at four lakes in Florida captured 219 alligators. The primary food type found in the alligator's stomach is recorded. Along with the gender, lake of capture, and size of the alligator.

## References

Agresti, A. (2013) Categorical Data Analysis. 3rd Edition, John Wiley & Sons, Hoboken, New Jersey.

**Examples**

```
data("alligatorDiet", package='syllogi')
str(alligatorDiet)
```

---

alligatorLength	<i>Study of Diets in Alligators at Lake George, Florida</i>
-----------------	---

---

**Description**

Data.frame

**Usage**

```
data(alligatorLength)
```

**Format**

The data frame has 63 rows and 3 variables:

**sex** Female (F) or Male (M).

**length** Length of alligator in meters. Subadult alligators have length < 1.83 and adults if > 1.83 meters.

**foodChoice** Primary stomach contents of the alligator.

**Details**

A study in Lake George, Florida caught 63 alligators. Each alligator's stomach contents were classified as fish, invertebrate, or other. The sex and the length of the alligator were also recorded.

**References**

Agresti, A. (2013) Categorical Data Analysis. 3rd Edition, John Wiley & Sons, Hoboken, New Jersey.

**Examples**

```
data("alligatorLength", package='syllogi')
str(alligatorLength)
```

---

`annualSales`*Fictitious Data Set of Annual Sales*

---

**Description**`Data.frame`**Usage**`data(annualSales)`**Format**

The data frame has 12 rows and 3 variables:

**sales** Annual gross sales in \$1000 of dollars.

**advert** Annual cost of advertising in \$1000 of dollars.

**quality** Quality of their store's typical product: 0=very poor quality to 25 = exceptional quality.

**Details**

You are hired as a statistical consultant. Twelve stores in the Fort Collins, CO area have asked you to develop a prediction model for their annual gross sales (sales; measured in \$1000 of dollars). They would like to know if it is possible to predict the amount of their sales by knowing how much they spend annually on advertising (advert; measured in \$1000 of dollars) and the quality of their store's typical product (quality; measure on a scale from 0 = very poor quality to 25 = exceptional quality).

**References**

fictitious data set

**Examples**

```
data("annualSales", package='syllogi')
str(annualSales)
```

---

beer

*Beer*

---

### Description

Data.frame

### Usage

```
data(beer)
```

### Format

The data frame has 86 rows and 5 variables:

**brand** Brand name of the beverage

**brewery** Brewery of the beverage

**percentAlcohol** Percent alcohol by volume

**calories** Total calories

**carbohydrates** Total carbohydrates

### Details

Does a beer with more carbohydrates tend to have more alcohol? To answer this question the carbohydrates and percent alcohol from several different beer brands were measured.

### Examples

```
data("beer", package='syllogi')
str(beer)
```

---

bighornSheep

*Bighorn Sheep*

---

### Description

Bighorn Sheep data

### Usage

```
data(bighornSheep)
```

**Format**

The data frame has 8000 rows (a geographic sample unit) and 15 variables:

**sampleUnit** Sample unit ID, 150m circles randomly overlaid across the study area

**count** Count of use by bighorn sheep.

**slope** Average slope (degrees) within the sampling unit

**elev** Average elevation (m) within the sampling unit

**distBurn** Sampling unit center to nearest (m) burned habitat edge calculated after fire event

**distRoad** Sampling unit center to nearest (m) road

**distEscp** Sampling unit center to nearest (m) escape terrain (slope > 27 degrees)

**distWater** Sampling unit center to nearest (m) perennial water source

**aspect** Dominant cardinal direction within each sampling unit

**fire** 1 = after fire, 0 = before fire

**season** Season, summer or winter

**Details**

Twelve female bighorn sheep are radio collared and tracked. Location of use of points is recorded before and after a forest fire.

**References**

Clapp, J.G., Beck, J.L. Short-Term Impacts of Fire-Mediated Habitat Alterations on an Isolated Bighorn Sheep Population. *fire ecol* 12, 80–98 (2016). <https://doi.org/10.4996/fireecology.1203080>

**Examples**

```
data('bighornSheep', package='syllogi')
str(bighornSheep)
```

---

bladderCancer

*Study of Recurrence of Bladder Cancer*

---

**Description**

Data.frame

**Usage**

```
data(bladderCancer)
```

**Format**

The data frame has 31 rows and 3 variables:

**Size** 0 = small primary tumor (< 3 cm) and 1 = large primary tumor (> 3cm).

**Tumors** Number of tumors.

**Time** Follow up time in months.

**Details**

Study on recurrence of bladder cancer tumor patients. Each patient had perviously received surgery to remove a primary tumor. The size of the primary removed tumor was recorded. After different follow up times the number of recurring tumors were recorded.

**References**

Bilder, C.R., & Loughin, T.M. (2014). Analysis of Categorical Data with R (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b17211>

**Examples**

```
data("bladderCancer", package='syllogi')
str(bladderCancer)
```

---

butterflyPlot

*Fictitious Data Set of Butterfly Counts*

---

**Description**

Data.frame

**Usage**

```
data(butterflyPlot)
```

**Format**

The data frame has 40 rows and 2 variables:

**area** Plot area size in hectares.

**numSpecies** Count of number of unique species.

**Details**

Plots ranging in size from 1ha to 1000ha, were left uncut in a larger landscape of logged tropical rainforest. In each plot the number of unique butterfly species was recorded. What is the relationship between plot size and unique species count?

**References**

fictitious data set

**Examples**

```
data("butterflyPlot", package='syllogi')
str(butterflyPlot)
```

---

depression

*Self Reported Depression*

---

**Description**

Self reported level of depression and other associated metrics.

**Usage**

```
data(depression)
```

**Format**

An object of class `data.frame` with 50 rows and 13 columns.

**Details**

This is a fictitious dataset useful for teaching how to use and interpret linear statistical models. The variables are:

**educate** Level of Education: (1) professional degree (non-college), (2) 2 years of college, (3) 2+ years of college, but not a BS degree, (4) BS degree, (5) MS degree

**income** Annual Income: 1 = \$10,000 to \$19,999; 2 = \$20,000 to \$29,999; ... 9 = \$90,000 to \$99,999; 10 = \$100,000 or more

**trauma** Experience of Trauma; Percent of Life Events Viewed as Traumatic: 0 = 0%, 1 = 10%, 2 = 20%, ..., 9 = 90%, 10 = 100%

**satisfac** Satisfied with your Life: 0 = No, 1 = Yes

**control** Feeling of Control; How much do you feel in control: 0 = Not at all, 1 = A Little, 2 = Some, 3 = A Lot, 4 = Completely

**history** Family History of Depression: 0 = No, 1 = Yes

**exercise** Weekly Amount of Exercise: 0 = None, 1 = 1 Hour, 2 = 2 Hours, 3 = 3 Hours, 4 = 4 Hours, 5 = 5 or more Hours

**mhpg** 3-methoxy-4-hydroxyphenylethyleneglycol, Depression Related Chemical Secreted in Urine; milligrams secreted per 24 hour period, labeled as mg/24h: 0 = 0 mg/24h, 1 = 100 mg/24h, ..., 9 = 900 mg/24h, 10 = 1000+ mg/24h

**sleep** Amount of Sleep Problems: 0 = None, 1 = 10% of the time, ... , 9 = 90% of the time, 10 = 100% of the time



**depress** Perceived Level of Depression: 0 = None, 1 = 10% of the time, ... , 9 = 90% of the time, 10 = 100% of the time

**depressYes** Do I consider myself depressed: 0 = No, 1 = Yes

**welbeing** Feeling of Well Being; how often do you feel good about yourself: 0 = None, 1 = 10% of the time, ... , 9 = 90% of the time, 10 = 100% of the time

**gender** Your Sex: 0 = Male, 1 = Female

## References

fictitious data set

---

dogFood

*Fictitious Data Set Comparing Dog Food Brands*

---

## Description

Data.frame

## Usage

```
data(dogFood)
```

## Format

The data frame has 25 rows and 2 variables:

**type** The type of dog food: our dog food or one of the four top sellers.

**gain** The percent weight gain.

## Details

You are hired as a statistical consultant for a dog food manufacturing company. The engineers who designed the company's dog food would like to know how it compares to the current top selling dog food brands on the market? To answer this question, 25 puppies of the same breed and age (within a week of each other) were chosen for this study. Five puppies were assigned to each dog food type. After 4 weeks the percent of weight gained for each puppy was determined.

## References

fictitious data set

## Examples

```
data("dogFood", package='syllogi')
str(dogFood)
```

---

federalistPapers      *Federalist Papers*

---

### Description

List of the Federalist Papers

### Usage

```
data(federalistPapers)
```

### Format

The list has 86 elements, each element is a list with 2 elements. The paper element is the text of the paper. The meta element is a data frame:

**number** Paper number.

**author** Author of the paper.

**title** Title of the paper.

**journal** Newspaper that published the paper.

**date** Date of publication.

### Details

The Project Gutenberg version of the Federalist Papers attributes paper No. 58 to Madison, but Mosteller and Wallace consider this paper to have disputed authorship. Thus, this version considers No. 58 authorship to be disputed.

The Project Gutenberg has two slightly different versions of No. 70, both included.

### References

<https://www.gutenberg.org/ebooks/18>

Mosteller, F. and D. L. Wallace. Inference and Disputed Authorship: The Federalist. Reading, MA., 1964

### Examples

```
data("federalistPapers", package='syllogi')  
str(federalistPapers)
```

---

genericData	<i>Generic Data Set</i>
-------------	-------------------------

---

**Description**

Generic data set with four ratio predictors (X1,X2,X3,X4), two categorical predictors (A,B) and one ratio response variable (Y).

**Usage**

```
data(genericData)
```

**Format**

An object of class `data.frame` with 60 rows and 7 columns.

**Details**

This is a fictitious dataset useful for teaching how to use and interpret linear statistical models.

**References**

fictitious data set

**Examples**

```
data("genericData", package='syllogi')
str(genericData)
```

---

golf	<i>Golfing</i>
------	----------------

---

**Description**

Data.frame

**Usage**

```
data(golf)
```

**Format**

The data frame has 18 rows and 3 variables:

**clubs** clubs used for that round of golf

**course** course for the round of golf

**score** score or strokes for 18 holes

**Details**

I purchased new golf clubs last summer, which I believe will significantly improve my game. I recorded my score after three rounds of golf with my new clubs and my old clubs. I also played at three different courses.

**References**

fictitious data set

**Examples**

```
data("golf", package='syllogi')
str(golf)
```

---

nutritionCancer

*Nutrition Cancer Study*

---

**Description**

Data.frame

**Usage**

```
data(nutritionCancer)
```

**Format**

The data frame has 50 rows and 6 variables:

**id** ID number of each patient.

**age** The age of the patient in years.

**length** The duration or time in months the patient has had breast cancer.

**servings** The number of servings the patient eats of fruits and vegetables in a typical day.

**familyHistory** Does or did any blood relatives (i.e. mother, grandmother, aunt, etc.) have or had breast cancer?

**stage** The stage of the cancer: 0-non-invasive to IV-very invasive or "advanced" cancer.

**Details**

Fictitious data set for teaching purposes. The fictitious scenario:

The purpose of a medical study is to examine the relationship between eating fruits and vegetables and breast cancer. To study the relationship, 1500 caucasian women with breast cancer were randomly selected from the list of cancer patients in the U.S. The first 50 patients have been measured.

**References**

Fictitious data set

**Examples**

```
data("nutritionCancer", package='syllogi')
str(nutritionCancer)
```

---

 osteosarcoma

*Study of Nonmetastatic Osteosarcoma*


---

**Description**

Data.frame

**Usage**

```
data(osteosarcoma)
```

**Format**

The data frame has 8 rows and 5 variables:

**lymphocyticInfiltration** Patient has high or low lymphocytic infiltration.

**gender** Female (F) or Male (M).

**osteoblasticPathology** Patient has osteoblastic pathology yes or no.

**diseaseFreeYes** Number of patients that are disease free after three years.

**diseaseFreeNo** Number of patients that are not disease free after three years.

**Details**

A study of nonmetastatic osteosarcoma was done. They recorded if the patient was disease free after three years. They recorded the gender, level of lymphocytic infiltration, and if there is osteoblastic pathology or not. Can the probability of being disease free after 3 years be predicted?

**References**

A M Goorin, A Perez-Atayde, M Gebhardt, J W Andersen, R H Wilkinson, M J Delorey, H Watts, M Link, N Jaffe, and E Frei 3rd *Journal of Clinical Oncology* 1987 5:8, 1178-1184

Agresti, A. (2002) *Categorical Data Analysis*. 2nd Edition, John Wiley & Sons, Inc., New York, 320-332. <http://dx.doi.org/10.1002/0471249688>

**Examples**

```
data("osteosarcoma", package='syllogi')
str(osteosarcoma)
```

---

patientSatisfaction    *Patient Satisfaction*

---

**Description**

Data.frame

**Usage**

```
data(patientSatisfaction)
```

**Format**

The data frame has 46 rows and 4 variables:

**satisfaction** Patient's level of satisfaction, higher value means more satisfied.

**age** Patient's age in years.

**severityIllness** Patient's severity of illness, higher value means more sever.

**anxietyLevel** Patient's anxiety level, higher value means more sever.

**Details**

A hospital administrator wants to predict patient's satisfaction using their age, severity of illness, and anxiety level. Forty six patients were selected for the study.

**References**

Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005). Applied linear statistical models (5th ed.). McGraw-Hill Irwin.

**Examples**

```
data("patientSatisfaction", package='syllogi')
str(patientSatisfaction)
```

---

politicalIdeology    *Political Ideology*

---

**Description**

Data.frame

**Usage**

```
data(politicalIdeology)
```

**Format**

The data frame has 20 rows and 4 variables:

**gender** Female (F) or Male (M).

**party** Democrat (D) or Republican (R)

**ideol** Very liberal (VL), Slightly Liberal (SL), Moderate (M), Slightly conservative (SC), or Very conservative (VC).

**count** Count of people.

**Details**

A 1991 U.S. General Social survey that cross classifies people according to gender, political party, and political ideology.

**References**

Bilder, C.R., & Loughin, T.M. (2014). Analysis of Categorical Data with R (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b17211>

**Examples**

```
data("politicalIdeology", package='syllogi')
str(politicalIdeology)
```

---

schoolProgram	<i>High School and Beyond Survey</i>
---------------	--------------------------------------

---

**Description**

A survey conducted on high school seniors by the National Center of Education Statistics.

**Usage**

```
data(schoolProgram)
```

**Format**

The data frame has 200 rows (a student) and 11 variables:

**id** Student ID.

**gender** Student's gender.

**race** Student's race.

**ses** Socio economic status of the student's family, with levels low, middle, and high.

**schtype** Type of school: public or private.

**prog** Type of program the student wants to attend after high school.

**read** Student's standardized reading score.  
**write** Student's standardized writing score.  
**math** Student's standardized math score.  
**science** Student's standardized science score  
**scost** Student's standardized social studies score

### Details

Two hundred students were randomly selected from the whole cohort in the survey.

### References

<https://www.openintro.org/data/index.php?data=hsb2>  
 UCLA Institute for Digital Research & Education - Statistical Consulting.

### Examples

```
data("schoolProgram", package='syllogi')
str(schoolProgram)
```

---

shipDamage	<i>Wave Damage of Ships</i>
------------	-----------------------------

---

### Description

Data.frame

### Usage

```
data(shipDamage)
```

### Format

The data frame has 20 rows and 5 variables:

**shipType** Type of ship  
**constYear** Year of construction  
**operation** Period of operation  
**months** Aggregate months of service  
**incidents** Number of damage incidents

### Details

Cargo carrying vessel's damage to the forward section due to waves. These data are only for the period of operation 1975 to 1979.



**References**

McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models. 2nd Edition, Chapman and Hall, London.

**Examples**

```
data("shipDamage", package='syllogi')
str(shipDamage)
```

---

shipGold	<i>Ships and Gold</i>
----------	-----------------------

---

**Description**

Data.frame

**Usage**

```
data(shipGold)
```

**Format**

The data frame has 20 rows (a ship) and 2 variables:

**shipSize** Size of the ship measured in inches on the horizon.

**gold** Amount of gold pieces on the ship.

**Details**

Fictitious data set for teaching purposes. The fictitious scenario:

Captain Buck Tooth has taken you prisoner aboard his pirate ship, the Lucky Lemon. He sees from your college transcripts you have taken a couple of statistics courses. Captain Buck Tooth wants you to predict the amount of gold a ship is carrying based on the size of the ship. Specifically, he thinks bigger ships carry more gold. For the last several ships he has looted he measured the height in inches when the ship was still way off on the horizon. The captain also has a good memory and remembers how much gold was taken from each ship in number of pieces.

**References**

Fictitious data set

**Examples**

```
data("shipGold", package='syllogi')
str(shipGold)
```

---

ski	<i>Ski Resort</i>
-----	-------------------

---

**Description**

Data.frame

**Usage**

```
data(ski)
```

**Format**

The data frame has 9 rows and 4 variables:

**miles** miles of skiable terrain

**capacity** number of vistors that could be taken per hour from the base of the mountain to the top via the resorts various lifts

**vistors** number of tickets sold per week

**resort** sort ID number

**Details**

Information from local ski resorts in the region. The research question is can weekly visitors be predictors from miles of skiable terrian and/or capacity of lifts?

**References**

ficitious data set

**Examples**

```
data("ski", package='syllogi')  
str(ski)
```

---

weightLoss	<i>Weight Loss Study</i>
------------	--------------------------

---

**Description**

Data.frame

**Usage**

```
data(weightLoss)
```

**Format**

The data frame has 60 rows and 2 variables:

**drug** Which weight loss drug the participant took for 6 weeks.

**loss** Percent of weight loss after the 6 weeks.

**Details**

Fictitious data set for teaching purposes. The fictitious scenario:

You are a statistical consultant. A client comes to you asking for help with their analysis. The client is from a drug company. Their new drug is supposed to help people lose weight. They conducted an experiment with their drug (drug A) and the two best selling weight loss drugs (B and C). Male participants from age 50-60 were used in the study. Each participant took one of the drugs for 6 week and the percent of weight loss was recorded.

**References**

Fictitious data set

**Examples**

```
data("weightLoss", package='syllogi')
str(weightLoss)
```

---

wheat

*Wheat Kernels*


---

**Description**

Data.frame

**Usage**

```
data(wheat)
```

**Format**

The data frame has 275 rows and 7 variables:

**class** hrw = hard red winter wheat and srw = soft red winter wheat.

**density** Density of a kernel.

**hardness** Hardness of a kernel.

**size** Size of a kernel.

**weight** Weight of a kernel.

**moisture** Moisture content of a kernel.

**type** Kernel's condition: Healthy, Sprout (sprouted prematurely), or Scab (infected with a fungus).

**Details**

A study on kernels of wheat was done. There are two classes of wheat: hard and soft red winter wheat. Each kernel measured for density, hardness, size, weight, and moisture content. Each kernel was classified by visual inspection if healthy, sprouted, or scab. A row in the data frame represents a kernel of wheat.

**References**

Bilder, C.R., & Loughin, T.M. (2014). Analysis of Categorical Data with R (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b17211>

Martin, C., Herrman, T.J., Loughin, T. and Oentong, S. (1998), Micropycnometer Measurement of Single-Kernel Density of Healthy, Sprouted, and Scab-Damaged Wheats†. Cereal Chemistry, 75: 177-180. <https://doi-org.libproxy.uwyo.edu/10.1094/CCHEM.1998.75.2.177>

**Examples**

```
data("wheat", package='syllogi')  
str(wheat)
```

# Index

## \* data

alligatorDiet, 2  
alligatorLength, 3  
annualSales, 4  
beer, 5  
bighornSheep, 5  
bladderCancer, 6  
butterflyPlot, 7  
depression, 8  
dogFood, 9  
federalistPapers, 10  
genericData, 11  
golf, 11  
nutritionCancer, 12  
osteosarcoma, 13  
patientSatisfaction, 14  
politicalIdeology, 14  
schoolProgram, 15  
shipDamage, 16  
shipGold, 17  
ski, 18  
weightLoss, 18  
wheat, 19

alligatorDiet, 2  
alligatorLength, 3  
annualSales, 4

beer, 5  
bighornSheep, 5  
bladderCancer, 6  
butterflyPlot, 7

depression, 8  
dogFood, 9

federalistPapers, 10

genericData, 11  
golf, 11

nutritionCancer, 12  
osteosarcoma, 13  
patientSatisfaction, 14  
politicalIdeology, 14  
schoolProgram, 15  
shipDamage, 16  
shipGold, 17  
ski, 18  
weightLoss, 18  
wheat, 19