

# Package ‘BSDA’

September 5, 2021

**Type** Package

**Title** Basic Statistics and Data Analysis

**Version** 1.2.1

**Date** 2021-09-04

**LazyData** yes

**Maintainer** Alan T. Arnholt <arnholtat@appstate.edu>

**Description** Data sets for book ``Basic Statistics and Data Analysis" by  
Larry J. Kitchens.

**Depends** lattice, R (>= 2.10)

**Imports** e1071

**License** GPL-3

**Suggests** ggplot2 (>= 2.1.0), dplyr, tidyr

**URL** <https://github.com/alanarnholt/BSDA>,  
<https://alanarnholt.github.io/BSDA/>

**BugReports** <https://github.com/alanarnholt/BSDA/issues>

**RoxygenNote** 7.1.1

**NeedsCompilation** no

**Author** Alan T. Arnholt [aut, cre],  
Ben Evans [aut]

**Repository** CRAN

**Date/Publication** 2021-09-05 04:40:01 UTC

## R topics documented:

Abbey . . . . .	8
Abc . . . . .	9
Abilene . . . . .	10
Ability . . . . .	11
Abortion . . . . .	12

Absent . . . . .	13
Achieve . . . . .	14
Adsales . . . . .	14
Aggress . . . . .	15
Aid . . . . .	16
Aids . . . . .	17
Airdisasters . . . . .	18
Airline . . . . .	19
Alcohol . . . . .	20
Allergy . . . . .	20
Anesthet . . . . .	21
Anxiety . . . . .	22
Apolipop . . . . .	22
Append . . . . .	23
Appendec . . . . .	24
Aptitude . . . . .	25
Archaeo . . . . .	25
Arthriti . . . . .	26
Artifici . . . . .	27
Asprin . . . . .	28
Asthmati . . . . .	28
Attorney . . . . .	29
Autogear . . . . .	30
Backtoback . . . . .	31
Bbsalaries . . . . .	31
Bigten . . . . .	32
Biology . . . . .	33
Birth . . . . .	34
Blackedu . . . . .	35
Blood . . . . .	35
Board . . . . .	36
Bones . . . . .	37
Books . . . . .	38
Bookstor . . . . .	38
Brain . . . . .	39
Bumpers . . . . .	40
Bus . . . . .	41
Bypass . . . . .	41
Cabinets . . . . .	42
Cancer . . . . .	43
Carbon . . . . .	44
Cat . . . . .	45
Censored . . . . .	45
Challeng . . . . .	46
Chemist . . . . .	47
Chesapea . . . . .	48
Chevy . . . . .	49
Chicken . . . . .	50

Chipavg . . . . .	50
Chips . . . . .	51
Cigar . . . . .	52
Cigarett . . . . .	53
CSim . . . . .	54
Citrus . . . . .	55
Clean . . . . .	56
Coaxial . . . . .	57
Coffee . . . . .	58
Coins . . . . .	58
Combinations . . . . .	59
Commute . . . . .	60
Concept . . . . .	61
Concrete . . . . .	61
Corn . . . . .	62
Correlat . . . . .	63
Counsel . . . . .	64
Cpi . . . . .	64
Crime . . . . .	65
Darwin . . . . .	66
Dealers . . . . .	67
Defectiv . . . . .	67
Degree . . . . .	68
Delay . . . . .	69
Depend . . . . .	70
Detroit . . . . .	70
Develop . . . . .	71
Devmath . . . . .	72
Dice . . . . .	72
Diesel . . . . .	73
Diplomat . . . . .	74
Disposal . . . . .	75
Dogs . . . . .	76
Domestic . . . . .	77
Dopamine . . . . .	78
Dowjones . . . . .	79
Drink . . . . .	80
Drug . . . . .	80
Dyslexia . . . . .	81
Earthqk . . . . .	82
EDA . . . . .	83
Educat . . . . .	84
Eggs . . . . .	85
Elderly . . . . .	85
Energy . . . . .	86
Engineer . . . . .	87
Entrance . . . . .	88
Epaminicompact . . . . .	89

Epatwoseater . . . . .	90
Executiv . . . . .	91
Exercise . . . . .	91
Fabric . . . . .	92
Faithful . . . . .	93
Family . . . . .	94
Ferraro1 . . . . .	95
Ferraro2 . . . . .	95
Fertility . . . . .	96
Firstchi . . . . .	97
Fish . . . . .	98
Fitness . . . . .	99
Florida2000 . . . . .	100
Fluid . . . . .	101
Food . . . . .	102
Framingh . . . . .	102
Freshman . . . . .	103
Funeral . . . . .	104
Galaxie . . . . .	105
Gallup . . . . .	105
Gasoline . . . . .	106
German . . . . .	107
Golf . . . . .	108
Governor . . . . .	109
Gpa . . . . .	110
Grades . . . . .	111
Graduate . . . . .	112
Greenriv . . . . .	112
Gnriv2 . . . . .	113
Groupabc . . . . .	114
Groups . . . . .	114
Gym . . . . .	115
Habits . . . . .	116
Haptoglo . . . . .	117
Hardware . . . . .	117
Hardwood . . . . .	118
Heat . . . . .	119
Heating . . . . .	120
Hodgkin . . . . .	121
Homes . . . . .	122
Homework . . . . .	123
Honda . . . . .	124
Hostile . . . . .	124
Housing . . . . .	125
Hurrican . . . . .	126
Iceberg . . . . .	127
Income . . . . .	128
Independent . . . . .	129

Indian	130
Indiapol	131
Indy500	131
Inflatio	132
Inletoil	133
Inmate	134
Inspect	135
Insulate	136
Iqgpa	137
Irises	137
Jdpower	138
Jobsat	139
Kidsmoke	140
Kilowatt	141
Kinder	141
Laminect	142
Lead	143
Leader	144
Lethal	144
Life	145
Lifespan	146
Ligntmonth	146
Lodge	147
Longtail	148
Lowabil	149
Magnesi	149
Malpract	150
Manager	151
Marked	151
Math	152
Mathcomp	153
Mathpro	154
Maze	155
Median	155
Mental	156
Mercury	157
Metrent	157
Miller	158
Miller1	159
Moisture	159
Monoxide	160
Movie	161
Music	162
Name	163
Nascar	164
Nervous	164
Newsstand	165
Nfldraf2	166

Nfldraft	166
Nicotine	167
normarea	168
nsize	169
ntester	170
Orange	171
Orioles	171
Oxytocin	172
Parented	173
Patrol	174
Pearson	175
Phone	175
Poison	176
Politic	177
Pollutio	178
Porosity	178
Poverty	179
Precinct	180
Prejudic	181
Presiden	181
Press	182
Prognost	183
Program	184
Psat	184
Psych	185
Puerto	186
Quail	186
Quality	187
Rainks	188
Randd	189
Rat	189
Ratings	190
Reaction	191
Reading	192
Readiq	192
Referend	193
Region	194
Register	195
Rehab	195
Remedial	196
Rentals	197
Repair	198
Retail	198
Ronbrown1	199
Ronbrown2	200
Rural	200
Salary	201
Salinity	202

Sat . . . . .	203
Saving . . . . .	204
Scales . . . . .	204
Schizop2 . . . . .	205
Schizoph . . . . .	206
Seatbelt . . . . .	207
Selfdefe . . . . .	208
Senior . . . . .	208
Sentence . . . . .	209
Shkdrug . . . . .	210
Shock . . . . .	211
Shoplift . . . . .	211
Short . . . . .	212
Shuttle . . . . .	213
SIGN.test . . . . .	214
Simpson . . . . .	216
Situp . . . . .	217
Skewed . . . . .	218
Skin . . . . .	218
Slc . . . . .	219
Smokyph . . . . .	220
Snore . . . . .	221
Snow . . . . .	222
Soccer . . . . .	223
Social . . . . .	223
Sophomor . . . . .	224
South . . . . .	225
Speed . . . . .	225
Spellers . . . . .	226
Spelling . . . . .	227
Sports . . . . .	227
Spouse . . . . .	228
SRS . . . . .	229
Stable . . . . .	230
Stamp . . . . .	230
Statclas . . . . .	231
Statelaw . . . . .	232
Statisti . . . . .	232
Step . . . . .	233
Stress . . . . .	234
Study . . . . .	235
Submarin . . . . .	235
Subway . . . . .	236
Sunspot . . . . .	237
Superbowl . . . . .	238
Supercar . . . . .	238
Tablrock . . . . .	239
Teacher . . . . .	241

Tenness	242
Tensile	243
Test1	243
Thermal	244
Tiaa	245
Ticket	245
Toaster	246
Tonsils	247
Tort	248
Toxic	249
Track	250
Track15	251
Treatments	251
Trees	252
Trucks	253
tsum.test	253
Tv	257
Twin	257
Undergrad	258
Vacation	259
Vaccine	260
Vehicle	261
Verbal	261
Victoria	262
Viscosit	263
Visual	264
Vocab	264
Wastewat	265
Weather94	266
Wheat	267
Windmill	268
Window	268
Wins	269
Wool	270
Yearsunspot	271
z.test	271
zsum.test	274

**Index****278**


---

Abbey	<i>Daily price returns (in pence) of Abbey National shares between 7/31/91 and 10/8/91</i>
-------	--

---

**Description**

Data used in problem 6.39



**Usage**

Abbey

**Format**

A data frame/tibble with 50 observations on one variable

**price** daily price returns (in pence) of Abbey National shares

**Source**

Buckle, D. (1995), Bayesian Inference for Stable Distributions, *Journal of the American Statistical Association*, 90, 605-613.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Abbey$price)
qqline(Abbey$price)
t.test(Abbey$price, mu = 300)
hist(Abbey$price, main = "Exercise 6.39",
      xlab = "daily price returns (in pence)",
      col = "blue")
```

---

Abc

*Three samples to illustrate analysis of variance*

---

**Description**

Data used in Exercise 10.1

**Usage**

Abc

**Format**

A data frame/tibble with 54 observations on two variables

**response** a numeric vector

**group** a character vector A, B, and C

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(response ~ group, col=c("red", "blue", "green"), data = Abc )
anova(lm(response ~ group, data = Abc))
```

---

Abilene

*Crimes reported in Abilene, Texas*

---

## Description

Data used in Exercise 1.23 and 2.79

## Usage

Abilene

## Format

A data frame/tibble with 16 observations on three variables

**crimetype** a character variable with values Aggravated assault, Arson, Burglary, Forcible rape, Larceny theft, Murder, Robbery, and Vehicle theft.

**year** a factor with levels 1992 and 1999

**number** number of reported crimes

## Source

*Uniform Crime Reports*, US Dept. of Justice.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
par(mfrow = c(2, 1))
barplot(Abilene$number[Abilene$year=="1992"],
names.arg = Abilene$crimetype[Abilene$year == "1992"],
main = "1992 Crime Stats", col = "red")
barplot(Abilene$number[Abilene$year=="1999"],
names.arg = Abilene$crimetype[Abilene$year == "1999"],
main = "1999 Crime Stats", col = "blue")
par(mfrow = c(1, 1))

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Abilene, aes(x = crimetype, y = number, fill = year)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))

## End(Not run)
```

---

Ability

*Perceived math ability for 13-year olds by gender*

---

**Description**

Data used in Exercise 8.57

**Usage**

Ability

**Format**

A data frame/tibble with 400 observations on two variables

**gender** a factor with levels girls and boys

**ability** a factor with levels hopeless, belowavg, average, aboveavg, and superior

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
CT <- xtabs(~gender + ability, data = Ability)
CT
chisq.test(CT)
```

---

 Abortion

*Abortion rate by region of country*


---

**Description**

Data used in Exercise 8.51

**Usage**

Abortion

**Format**

A data frame/tibble with 51 observations on the following 10 variables:

**state** a character variable with values alabama, alaska, arizona, arkansas, california, colorado, connecticut, delaware, dist of columbia, florida, georgia, hawaii, idaho, illinois, indiana, iowa, kansas, kentucky, louisiana, maine, maryland, massachusetts, michigan, minnesota, mississippi, missouri, montana, nebraska, nevada, new hampshire, new jersey, new mexico, new york, north carolina, north dakota, ohio, oklahoma, oregon, pennsylvania, rhode island, south carolina, south dakota, tennessee, texas, utah, vermont, virginia, washington, west virginia, wisconsin, and wyoming

**region** a character variable with values midwest northeast south west

**regcode** a numeric vector

**rate1988** a numeric vector

**rate1992** a numeric vector

**rate1996** a numeric vector

**provide1988** a numeric vector

**provide1992** a numeric vector

**lowhigh** a numeric vector

**rate** a factor with levels Low and High

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~region + rate, data = Abortion)
T1
chisq.test(T1)
```

---

Absent	<i>Number of absent days for 20 employees</i>
--------	---

---

**Description**

Data used in Exercise 1.28

**Usage**

Absent

**Format**

A data frame/tibble with 20 observations on one variable

**days** days absent

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
CT <- xtabs(~ days, data = Absent)
CT
barplot(CT, col = "pink", main = "Exercise 1.28")
plot(ecdf(Absent$days), main = "ECDF")
```

---

Achieve *Math achievement test scores by gender for 25 high school students*

---

**Description**

Data used in Example 7.14 and Exercise 10.7

**Usage**

Achieve

**Format**

A data frame/tibble with 25 observations on two variables

**score** mathematics achievement score

**gender** a factor with 2 levels boys and girls

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
anova(lm(score ~ gender, data = Achieve))
t.test(score ~ gender, var.equal = TRUE, data = Achieve)
```

---

Adsales *Number of ads versus number of sales for a retailer of satellite dishes*

---

**Description**

Data used in Exercise 9.15

**Usage**

Adsales

**Format**

A data frame/tibble with six observations on three variables

**month** a character vector listing month

**ads** a numeric vector containing number of ads

**sales** a numeric vector containing number of sales

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(sales ~ ads, data = Adsales, main = "Exercise 9.15")
mod <- lm(sales ~ ads, data = Adsales)
abline(mod, col = "red")
summary(mod)
predict(mod, newdata = data.frame(ads = 6), interval = "conf", level = 0.99)
```

---

Aggress	<i>Aggressive tendency scores for a group of teenage members of a street gang</i>
---------	---

---

## Description

Data used in Exercises 1.66 and 1.81

## Usage

Aggress

## Format

A data frame/tibble with 28 observations on one variable

**aggres** measure of aggressive tendency, ranging from 10-50

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
with(data = Aggress,
      EDA(aggres))
# OR
IQR(Aggress$aggres)
diff(range(Aggress$aggres))
```

---

Aid *Monthly payments per person for families in the AFDC federal program*

---

**Description**

Data used in Exercises 1.91 and 3.68

**Usage**

Aid

**Format**

A data frame/tibble with 51 observations on two variables

**state** a factor with levels Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**payment** average monthly payment per person in a family

**Source**

US Department of Health and Human Services, 1993.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Aid$payment, xlab = "payment", main =  
"Average monthly payment per person in a family",  
col = "lightblue")  
boxplot(Aid$payment, col = "lightblue")  
dotplot(state ~ payment, data = Aid)
```



---

Aids	<i>Incubation times for 295 patients thought to be infected with HIV by a blood transfusion</i>
------	---

---

### Description

Data used in Exercise 6.60

### Usage

Aids

### Format

A data frame/tibble with 295 observations on three variables

**duration** time (in months) from HIV infection to the clinical manifestation of full-blown AIDS

**age** age (in years) of patient

**group** a numeric vector

### Source

Kalbsleich, J. and Lawless, J., (1989), An analysis of the data on transfusion related AIDS, *Journal of the American Statistical Association*, 84, 360-372.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
with(data = Aids,
      EDA(duration)
    )
with(data = Aids,
      t.test(duration, mu = 30, alternative = "greater")
    )
with(data = Aids,
      SIGN.test(duration, md = 24, alternative = "greater")
    )
```

---

Airdisasters

*Aircraft disasters in five different decades*

---

### Description

Data used in Exercise 1.12

### Usage

Airdisasters

### Format

A data frame /tibble with 141 observations on the following seven variables

**year** a numeric vector indicating the year of an aircraft accident

**deaths** a numeric vector indicating the number of deaths of an aircraft accident

**decade** a character vector indicating the decade of an aircraft accident

### Source

2000 *World Almanac and Book of Facts*.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
par(las = 1)
stripchart(deaths ~ decade, data = Airdisasters,
           subset = decade != "1930s" & decade != "1940s",
           method = "stack", pch = 19, cex = 0.5, col = "red",
           main = "Aircraft Disasters 1950 - 1990",
           xlab = "Number of fatalities")
par(las = 0)
```

---

Airline	<i>Percentage of on-time arrivals and number of complaints for 11 airlines</i>
---------	--

---

**Description**

Data for Example 2.9

**Usage**

Airline

**Format**

A data frame/tibble with 11 observations on three variables

**airline** a character variable with values Alaska, Amer West, American, Continental, Delta, Northwest, Pan Am, Southwest, TWA, United, and USAir

**ontime** a numeric vector

**complaints** complaints per 1000 passengers

**Source**

Transportation Department.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
with(data = Airline,  
      barplot(complaints, names.arg = airline, col = "lightblue",  
              las = 2)  
      )  
plot(complaints ~ ontime, data = Airline, pch = 19, col = "red",  
      xlab = "On time", ylab = "Complaints")
```

---

Alcohol	<i>Ages at which 14 female alcoholics began drinking</i>
---------	--

---

**Description**

Data used in Exercise 5.79

**Usage**

Alcohol

**Format**

A data frame/tibble with 14 observations on one variable

**age** age when individual started drinking

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Alcohol$age)
qqline(Alcohol$age)
SIGN.test(Alcohol$age, md = 20, conf.level = 0.99)
```

---

Allergy	<i>Allergy medicines by adverse events</i>
---------	--

---

**Description**

Data used in Exercise 8.22

**Usage**

Allergy

**Format**

A data frame/tibble with 406 observations on two variables

**event** a factor with levels insomnia, headache, and drowsiness

**medication** a factor with levels seldane-d, pseudoephedrine, and placebo

**Source**

Marion Merrel Dow, Inc. Kansas City, Mo. 64114.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~event + medication, data = Allergy)
T1
chisq.test(T1)
```

---

Anesthet

*Recovery times for anesthetized patients*

---

**Description**

Data used in Exercise 5.58

**Usage**

Anesthet

**Format**

A with 10 observations on one variable

**recover** recovery time (in hours)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Anesthet$recover)
qqline(Anesthet$recover)
with(data = Anesthet,
t.test(recover, conf.level = 0.90)$conf
)
```

---

Anxiety

*Math test scores versus anxiety scores before the test*

---

**Description**

Data used in Exercise 2.96

**Usage**

Anxiety

**Format**

A data frame/tibble with 20 observations on two variables

**anxiety** anxiety score before a major math test

**math** math test score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(math ~ anxiety, data = Anxiety, ylab = "score",
      main = "Exercise 2.96")
with(data = Anxiety,
     cor(math, anxiety)
)
linmod <- lm(math ~ anxiety, data = Anxiety)
abline(linmod, col = "purple")
summary(linmod)
```

---

Apolipop

*Level of apolipoprotein B and number of cups of coffee consumed per day for 15 adult males*

---

**Description**

Data used in Examples 9.2 and 9.9

**Usage**

Apolipop

**Format**

A data frame/tibble with 15 observations on two variables

**coffee** number of cups of coffee per day

**apolipB** level of apolipoprotein B

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(apolipB ~ coffee, data = Apolipop)
linmod <- lm(apolipB ~ coffee, data = Apolipop)
summary(linmod)
summary(linmod)$sigma
anova(linmod)
anova(linmod)[2, 3]^0.5
par(mfrow = c(2, 2))
plot(linmod)
par(mfrow = c(1, 1))
```

---

 Append

---

*Median costs of an appendectomy at 20 hospitals in North Carolina*


---

**Description**

Data for Exercise 1.119

**Usage**

Append

**Format**

A data frame/tibble with 20 observations on one variable

**fee** fees for an appendectomy for a random sample of 20 hospitals in North Carolina

**Source**

North Carolina Medical Database Commission, August 1994.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

fee <- Append$fee
ll <- mean(fee) - 2*sd(fee)
ul <- mean(fee) + 2*sd(fee)
limits <-c(ll, ul)
limits
fee[fee < ll | fee > ul]

```

---

Appendec

*Median costs of appendectomies at three different types of North Carolina hospitals*


---

**Description**

Data for Exercise 10.60

**Usage**

Appendec

**Format**

A data frame/tibble with 59 observations on two variables

**cost** median costs of appendectomies at hospitals across the state of North Carolina in 1992

**region** a vector classifying each hospital as rural, regional, or metropolitan

**Source**

*Consumer's Guide to Hospitalization Charges in North Carolina Hospitals* (August 1994), North Carolina Medical Database Commission, Department of Insurance.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

boxplot(cost ~ region, data = Appendec, col = c("red", "blue", "cyan"))
anova(lm(cost ~ region, data = Appendec))

```



---

Aptitude

*Aptitude test scores versus productivity in a factory*

---

**Description**

Data for Exercises 2.1, 2.26, 2.35 and 2.51

**Usage**

Aptitude

**Format**

A data frame/tibble with 8 observations on two variables

**aptitude** aptitude test scores

**product** productivity scores

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(product ~ aptitude, data = Aptitude, main = "Exercise 2.1")
model1 <- lm(product ~ aptitude, data = Aptitude)
model1
abline(model1, col = "red", lwd=3)
resid(model1)
fitted(model1)
cor(Aptitude$product, Aptitude$aptitude)
```

---

Archaeo

*Radiocarbon ages of observations taken from an archaeological site*

---

**Description**

Data for Exercises 5.120, 10.20 and Example 1.16

**Usage**

Archaeo

**Format**

A data frame/tibble with 60 observations on two variables

**age** number of years before 1983 - the year the data were obtained

**phase** Ceramic Phase numbers

**Source**

Cunliffe, B. (1984) and Naylor and Smith (1988).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(age ~ phase, data = Archaeo, col = "yellow",
        main = "Example 1.16", xlab = "Ceramic Phase", ylab = "Age")
anova(lm(age ~ as.factor(phase), data= Archaeo))
```

---

Arthriti

*Time of relief for three treatments of arthritis*

---

**Description**

Data for Exercise 10.58

**Usage**

Arthriti

**Format**

A data frame/tibble with 51 observations on two variables

**time** time (measured in days) until an arthritis sufferer experienced relief

**treatment** a factor with levels A, B, and C

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(time ~ treatment, data = Arthriti,  
col = c("lightblue", "lightgreen", "yellow"),  
ylab = "days")  
anova(lm(time ~ treatment, data = Arthriti))
```

---

Artifici

*Durations of operation for 15 artificial heart transplants*

---

## Description

Data for Exercise 1.107

## Usage

```
Artifici
```

## Format

A data frame/tibble with 15 observations on one variable

**duration** duration (in hours) for transplant

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
stem(Artifici$duration, 2)  
summary(Artifici$duration)  
values <- Artifici$duration[Artifici$duration < 6.5]  
values  
summary(values)
```

---

Asprin

*Dissolving time versus level of impurities in aspirin tablets*

---

### Description

Data for Exercise 10.51

### Usage

Asprin

### Format

A data frame/tibble with 15 observations on two variables

**time** time (in seconds) for aspirin to dissolve

**impurity** impurity of an ingredient with levels 1%, 5%, and 10%

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
boxplot(time ~ impurity, data = Asprin,  
        col = c("red", "blue", "green"))
```

---

Asthmati

*Asthmatic relief index on nine subjects given a drug and a placebo*

---

### Description

Data for Exercise 7.52

### Usage

Asthmati

### Format

A data frame/tibble with nine observations on three variables

**drug** asthmatic relief index for patients given a drug

**placebo** asthmatic relief index for patients given a placebo

**difference** difference between the placebo and drug

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
qqnorm(Asthmati$difference)
qqline(Asthmati$difference)
shapiro.test(Asthmati$difference)
with(data = Asthmati,
      t.test(placebo, drug, paired = TRUE, mu = 0, alternative = "greater")
)
```

---

Attorney

*Number of convictions reported by U.S. attorney's offices*

---

## Description

Data for Example 2.2 and Exercises 2.43 and 2.57

## Usage

Attorney

## Format

A data frame/tibble with 88 observations on three variables

**staff** U.S. attorneys' office staff per 1 million population

**convict** U.S. attorneys' office convictions per 1 million population

**district** a factor with levels Albuquerque, Alexandria, Va, Anchorage, Asheville, NC, Atlanta, Baltimore, Baton Rouge, Billings, Mt, Birmingham, Al, Boise, Id, Boston, Buffalo, Burlington, Vt, Cedar Rapids, Charleston, WVA, Cheyenne, Wy, Chicago, Cincinnati, Cleveland, Columbia, SC, Concord, NH, Denver, Des Moines, Detroit, East St. Louis, Fargo, ND, Fort Smith, Ark, Fort Worth, Grand Rapids, Mi, Greensboro, NC, Honolulu, Houston, Indianapolis, Jackson, Miss, Kansas City, Knoxville, Tn, Las Vegas, Lexington, Ky, Little Rock, Los Angeles, Louisville, Memphis, Miami, Milwaukee, Minneapolis, Mobile, Ala, Montgomery, Ala, Muskogee, Ok, Nashville, New Haven, Conn, New Orleans, New York (Brooklyn), New York (Manhattan), Newark, NJ, Oklahoma City, Omaha, Oxford, Miss, Pensacola, Fl, Philadelphia, Phoenix, Pittsburgh, Portland, Maine, Portland, Ore, Providence, RI, Raleigh, NC, Roanoke, Va, Sacramento, Salt Lake City, San Antonio, San Diego, San Francisco, Savannah, Ga, Scranton, Pa, Seattle, Shreveport, La, Sioux Falls, SD, South Bend, Ind, Spokane, Wash, Springfield, Ill, St. Louis, Syracuse, NY, Tampa, Topeka, Kan, Tulsa, Tyler, Tex, Washington, Wheeling, WVA, and Wilmington, Del

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
par(mfrow=c(1, 2))
plot(convict ~ staff, data = Attorney, main = "With Washington, D.C.")
plot(convict[-86] ~staff[-86], data = Attorney,
main = "Without Washington, D.C.")
par(mfrow=c(1, 1))
```

---

Autogear

*Number of defective auto gears produced by two manufacturers*

---

## Description

Data for Exercise 7.46

## Usage

Autogear

## Format

A data frame/tibble with 20 observations on two variables

**defectives** number of defective gears in the production of 100 gears per day

**manufacturer** a factor with levels A and B

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
t.test(defectives ~ manufacturer, data = Autogear)
wilcox.test(defectives ~ manufacturer, data = Autogear)
t.test(defectives ~ manufacturer, var.equal = TRUE, data = Autogear)
```

---

Backtoback	<i>Illustrates inferences based on pooled t-test versus Wilcoxon rank sum test</i>
------------	--

---

**Description**

Data for Exercise 7.40

**Usage**

Backtoback

**Format**

A data frame/tibble with 24 observations on two variables

**score** a numeric vector

**group** a numeric vector

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
wilcox.test(score ~ group, data = Backtoback)
t.test(score ~ group, data = Backtoback)
```

---

Bbsalaries	<i>Baseball salaries for members of five major league teams</i>
------------	---

---

**Description**

Data for Exercise 1.11

**Usage**

Bbsalaries

**Format**

A data frame/tibble with 142 observations on two variables

**salary** 1999 salary for baseball player

**team** a factor with levels Angels, Indians, Orioles, Redsoxs, and Whitesoxs

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
stripchart(salary ~ team, data = Bbsalaries, method = "stack",
           pch = 19, col = "blue", cex = 0.75)
title(main = "Major League Salaries")
```

---

Bigten	<i>Graduation rates for student athletes and nonathletes in the Big Ten Conf.</i>
--------	---

---

## Description

Data for Exercises 1.124 and 2.94

## Usage

Bigten

## Format

A data frame/tibble with 44 observations on the following four variables

**school** a factor with levels Illinois, Indiana, Iowa, Michigan, Michigan State, Minnesota, Northwestern, Ohio State, Penn State, Purdue, and Wisconsin

**rate** graduation rate

**year** factor with two levels 1984–1985 and 1993–1994

**status** factor with two levels athlete and student

## Source

NCAA Graduation Rates Report, 2000.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.



**Examples**

```
boxplot(rate ~ status, data = subset(Bigten, year = "1993-1994"),
horizontal = TRUE, main = "Graduation Rates 1993-1994")
with(data = Bigten,
      tapply(rate, list(year, status), mean)
)
```

---

Biology

*Test scores on first exam in biology class*

---

**Description**

Data for Exercise 1.49

**Usage**

Biology

**Format**

A data frame/tibble with 30 observations on one variable

**score** test scores on the first test in a beginning biology class

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Biology$score, breaks = "scott", col = "brown", freq = FALSE,
main = "Problem 1.49", xlab = "Test Score")
lines(density(Biology$score), lwd=3)
```

---

 Birth
 

---

*Live birth rates in 1990 and 1998 for all states*


---

**Description**

Data for Example 1.10

**Usage**

Birth

**Format**

A data frame/tibble with 51 observations on three variables

**state** a character with levels Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**rate** live birth rates per 1000 population

**year** a factor with levels 1990 and 1998

**Source**

*National Vital Statistics Report, 48, March 28, 2000, National Center for Health Statistics.*

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
rate1998 <- subset(Birth, year == "1998", select = rate)
stem(x = rate1998$rate, scale = 2)
hist(rate1998$rate, breaks = seq(10.9, 21.9, 1.0), xlab = "1998 Birth Rate",
     main = "Figure 1.14 in BSDA", col = "pink")
hist(rate1998$rate, breaks = seq(10.9, 21.9, 1.0), xlab = "1998 Birth Rate",
     main = "Figure 1.16 in BSDA", col = "pink", freq = FALSE)
lines(density(rate1998$rate), lwd = 3)
rm(rate1998)
```

---

Blackedu

*Education level of blacks by gender*

---

**Description**

Data for Exercise 8.55

**Usage**

Blackedu

**Format**

A data frame/tibble with 3800 observations on two variables

**gender** a factor with levels Female and Male

**education** a factor with levels High school dropout, High school graudate, Some college, Bachelor's degree, and Graduate degree

**Source**

Bureau of Census data.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~gender + education, data = Blackedu)
T1
chisq.test(T1)
```

---

Blood

*Blood pressure of 15 adult males taken by machine and by an expert*

---

**Description**

Data for Exercise 7.84

**Usage**

Blood

**Format**

A data frame/tibble with 15 observations on the following two variables

**machine** blood pressure recorded from an automated blood pressure machine

**expert** blood pressure recorded by an expert using an at-home device

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
DIFF <- Blood$machine - Blood$expert
shapiro.test(DIFF)
qqnorm(DIFF)
qqline(DIFF)
rm(DIFF)
t.test(Blood$machine, Blood$expert, paired = TRUE)
```

---

Board

*Incomes of board members from three different universities*

---

**Description**

Data for Exercise 10.14

**Usage**

Board

**Format**

A data frame/tibble with 7 observations on three variables

**salary** 1999 salary (in \$1000) for board directors

**university** a factor with levels A, B, and C

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(salary ~ university, data = Board, col = c("red", "blue", "green"),
        ylab = "Income")
tapply(Board$salary, Board$university, summary)
anova(lm(salary ~ university, data = Board))
## Not run:
library(dplyr)
dplyr::group_by(Board, university) %>%
  summarize(Average = mean(salary))

## End(Not run)
```

---

Bones	<i>Bone density measurements of 35 physically active and 35 non-active women</i>
-------	--

---

## Description

Data for Example 7.22

## Usage

Bones

## Format

A data frame/tibble with 70 observations on two variables

**density** bone density measurements

**group** a factor with levels active and nonactive

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
t.test(density ~ group, data = Bones, alternative = "greater")
t.test(rank(density) ~ group, data = Bones, alternative = "greater")
wilcox.test(density ~ group, data = Bones, alternative = "greater")
```

---

Books

*Number of books read and final spelling scores for 17 third graders*

---

### Description

Data for Exercise 9.53

### Usage

Books

### Format

A data frame/tibble with 17 observations on two variables

**book** number of books read

**spelling** spelling score

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
plot(spelling ~ book, data = Books)
mod <- lm(spelling ~ book, data = Books)
summary(mod)
abline(mod, col = "blue", lwd = 2)
```

---

Bookstor

*Prices paid for used books at three different bookstores*

---

### Description

Data for Exercise 10.30 and 10.31

### Usage

Bookstor

**Format**

A data frame/tibble with 72 observations on two variables

**dollars** money obtained for selling textbooks

**store** a factor with levels A, B, and C

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(dollars ~ store, data = Bookstor,
        col = c("purple", "lightblue", "cyan"))
kruskal.test(dollars ~ store, data = Bookstor)
```

---

Brain

*Brain weight versus body weight of 28 animals*


---

**Description**

Data for Exercises 2.15, 2.44, 2.58 and Examples 2.3 and 2.20

**Usage**

Brain

**Format**

A data frame/tibble with 28 observations on three variables

**species** a factor with levels African elephant, Asian Elephant, Brachiosaurus, Cat, Chimpanzee, Cow, Diplodocus, Donkey, Giraffe, Goat, Gorilla, Gray wolf, Guinea Pig, Hamster, Horse, Human, Jaguar, Kangaroo, Mole, Mouse, Mt Beaver, Pig, Potar monkey, Rabbit, Rat, Rhesus monkey, Sheep, and Triceratops

**bodyweight** body weight (in kg)

**brainweight** brain weight (in g)

**Source**

P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection* (New York: Wiley, 1987).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(log(brainweight) ~ log(bodyweight), data = Brain,
     pch = 19, col = "blue", main = "Example 2.3")
mod <- lm(log(brainweight) ~ log(bodyweight), data = Brain)
abline(mod, lty = "dashed", col = "blue")
```

---

 Bumpers

*Repair costs of vehicles crashed into a barrier at 5 miles per hour*


---

**Description**

Data for Exercise 1.73

**Usage**

Bumpers

**Format**

A data frame/tibble with 23 observations on two variables

**car** a factor with levels Buick Century, Buick Skylark, Chevrolet Cavalier, Chevrolet Corsica, Chevrolet Lumina, Dodge Dynasty, Dodge Monaco, Ford Taurus, Ford Tempo, Honda Accord, Hyundai Sonata, Mazda 626, Mitsubishi Galant, Nissan Stanza, Oldsmobile Calais, Oldsmobile Ciere, Plymouth Acclaim, Pontiac 6000, Pontiac Grand Am, Pontiac Sunbird, Saturn SL2, Subaru Legacy, and Toyota Camry

**repair** total repair cost (in dollars) after crashing a car into a barrier four times while the car was traveling at 5 miles per hour

**Source**

Insurance Institute of Highway Safety.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Bumpers$repair)
stripchart(Bumpers$repair, method = "stack", pch = 19, col = "blue")
library(lattice)
dotplot(car ~ repair, data = Bumpers)
```



---

Bus

*Attendance of bus drivers versus shift*

---

### Description

Data for Exercise 8.25

### Usage

Bus

### Format

A data frame/tibble with 29363 observations on two variables

**attendance** a factor with levels absent and present

**shift** a factor with levels am, noon, pm, swing, and split

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
T1 <- xtabs(~attendance + shift, data = Bus)
T1
chisq.test(T1)
```

---

Bypass

*Median charges for coronary bypass at 17 hospitals in North Carolina*

---

### Description

Data for Exercises 5.104 and 6.43

### Usage

Bypass

**Format**

A data frame/tibble with 17 observations on two variables

**hospital** a factor with levels Carolinas Med Ct, Duke Med Ct, Durham Regional, Forsyth Memorial, Frye Regional, High Point Regional, Memorial Mission, Mercy, Moore Regional, Moses Cone Memorial, NC Baptist, New Hanover Regional, Pitt Co. Memorial, Presbyterian, Rex, Univ of North Carolina, and Wake County

**charge** median charge for coronary bypass

**Source**

*Consumer's Guide to Hospitalization Charges in North Carolina Hospitals* (August 1994), North Carolina Medical Database Commission, Department of Insurance.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Bypass$charge)
t.test(Bypass$charge, conf.level=.90)$conf
t.test(Bypass$charge, mu = 35000)
```

---

Cabinets

*Estimates of costs of kitchen cabinets by two suppliers on 20 prospective homes*

---

**Description**

Data for Exercise 7.83

**Usage**

Cabinets

**Format**

A data frame/tibble with 20 observations on three variables

**home** a numeric vector

**supplA** estimate for kitchen cabinets from supplier A (in dollars)

**supplB** estimate for kitchen cabinets from supplier A (in dollars)

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
DIF <- Cabinets$supplA - Cabinets$supplB
qqnorm(DIF)
qqline(DIF)
shapiro.test(DIF)
with(data = Cabinets,
      t.test(supplA, supplB, paired = TRUE)
)
with(data = Cabinets,
      wilcox.test(supplA, supplB, paired = TRUE)
)
rm(DIF)
```

---

Cancer

*Survival times of terminal cancer patients treated with vitamin C*

---

## Description

Data for Exercises 6.55 and 6.64

## Usage

Cancer

## Format

A data frame/tibble with 64 observations on two variables

**survival** survival time (in days) of terminal patients treated with vitamin C

**type** a factor indicating type of cancer with levels breast, bronchus, colon, ovary, and stomach

## Source

Cameron, E and Pauling, L. 1978. "Supplemental Ascorbate in the Supportive Treatment of Cancer." *Proceedings of the National Academy of Science*, 75, 4538-4542.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(survival ~ type, Cancer, col = "blue")
stomach <- Cancer$survival[Cancer$type == "stomach"]
bronchus <- Cancer$survival[Cancer$type == "bronchus"]
boxplot(stomach, ylab = "Days")
SIGN.test(stomach, md = 100, alternative = "greater")
SIGN.test(bronchus, md = 100, alternative = "greater")
rm(bronchus, stomach)
```

---

Carbon

*Carbon monoxide level measured at three industrial sites*

---

## Description

Data for Exercise 10.28 and 10.29

## Usage

Carbon

## Format

A data frame/tibble with 24 observations on two variables

**CO** carbon monoxide measured (in parts per million)

**site** a factor with levels SiteA, SiteB, and SiteC

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(CO ~ site, data = Carbon, col = "lightgreen")
kruskal.test(CO ~ site, data = Carbon)
```

---

Cat	<i>Reading scores on the California achievement test for a group of 3rd graders</i>
-----	---

---

**Description**

Data for Exercise 1.116

**Usage**

Cat

**Format**

A data frame/tibble with 17 observations on one variable

**score** reading score on the California Achievement Test

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Cat$score)
fivenum(Cat$score)
boxplot(Cat$score, main = "Problem 1.116", col = "green")
```

---

Censored	<i>Entry age and survival time of patients with small cell lung cancer under two different treatments</i>
----------	---

---

**Description**

Data for Exercises 7.34 and 7.48

**Usage**

Censored

**Format**

A data frame/tibble with 121 observations on three variables

**survival** survival time (in days) of patients with small cell lung cancer

**treatment** a factor with levels armA and armB indicating the treatment a patient received

**age** the age of the patient

**Source**

Ying, Z., Jung, S., Wei, L. 1995. "Survival Analysis with Median Regression Models." *Journal of the American Statistical Association*, 90, 178-184.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(survival ~ treatment, data = Censored, col = "yellow")
wilcox.test(survival ~ treatment, data = Censored, alternative = "greater")
```

---

Challeng

*Temperatures and O-ring failures for the launches of the space shuttle Challenger*

---

**Description**

Data for Examples 1.11, 1.12, 1.13, 2.11 and 5.1

**Usage**

Challeng

**Format**

A data frame/tibble with 25 observations on four variables

**flight** a character variable indicating the flight

**date** date of the flight

**temp** temperature (in fahrenheit)

**failures** number of failures

**Source**

Dalal, S. R., Fowlkes, E. B., Hoadley, B. 1989. "Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure." *Journal of the American Statistical Association*, 84, No. 408, 945-957.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Challeng$temp)
summary(Challeng$temp)
IQR(Challeng$temp)
quantile(Challeng$temp)
fivenum(Challeng$temp)
stem(sort(Challeng$temp)[-1])
summary(sort(Challeng$temp)[-1])
IQR(sort(Challeng$temp)[-1])
quantile(sort(Challeng$temp)[-1])
fivenum(sort(Challeng$temp)[-1])
par(mfrow=c(1, 2))
qqnorm(Challeng$temp)
qqline(Challeng$temp)
qqnorm(sort(Challeng$temp)[-1])
qqline(sort(Challeng$temp)[-1])
par(mfrow=c(1, 1))
```

---

 Chemist

*Starting salaries of 50 chemistry majors*


---

**Description**

Data for Example 5.3

**Usage**

Chemist

**Format**

A data frame/tibble with 50 observations on one variable

**salary** starting salary (in dollars) for chemistry major

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Chemist$salary)
```

---

Chesapea	<i>Surface salinity measurements taken offshore from Annapolis, Maryland in 1927</i>
----------	--

---

**Description**

Data for Exercise 6.41

**Usage**

Chesapea

**Format**

A data frame/tibble with 16 observations on one variable

**salinity** surface salinity measurements (in parts per 1000) for station 11, offshore from Annapolis, Maryland, on July 3-4, 1927.

**Source**

Davis, J. (1986) *Statistics and Data Analysis in Geology, Second Edition*. John Wiley and Sons, New York.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Chesapea$salinity)
qqline(Chesapea$salinity)
shapiro.test(Chesapea$salinity)
t.test(Chesapea$salinity, mu = 7)
```



---

Chevy	<i>Insurance injury ratings of Chevrolet vehicles for 1990 and 1993 models</i>
-------	--

---

**Description**

Data for Exercise 8.35

**Usage**

Chevy

**Format**

A data frame/tibble with 67 observations on two variables

**year** a factor with levels 1988-90 and 1991-93

**frequency** a factor with levels much better than average, above average, average, below average, and much worse than average

**Source**

Insurance Institute for Highway Safety and the Highway Loss Data Institute, 1995.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~year + frequency, data = Chevy)
T1
chisq.test(T1)
rm(T1)
```

---

Chicken

*Weight gain of chickens fed three different rations*

---

### Description

Data for Exercise 10.15

### Usage

Chicken

### Format

A data frame/tibble with 13 observations on three variables

**gain** weight gain over a specified period

**feed** a factor with levels ration1, ration2, and ration3

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
boxplot(gain ~ feed, col = c("red", "blue", "green"), data = Chicken)
anova(lm(gain ~ feed, data = Chicken))
```

---

Chipavg

*Measurements of the thickness of the oxide layer of manufactured integrated circuits*

---

### Description

Data for Exercises 6.49 and 7.47

### Usage

Chipavg

**Format**

A data frame/tibble with 30 observations on three variables

**wafer1** thickness of the oxide layer for wafer1

**wafer2** thickness of the oxide layer for wafer2

**thickness** average thickness of the oxide layer of the eight measurements obtained from each set of two wafers

**Source**

Yashchin, E. 1995. "Likelihood Ratio Methods for Monitoring Parameters of a Nested Random Effect Model." *Journal of the American Statistical Association*, 90, 729-738.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Chipavg$thickness)
t.test(Chipavg$thickness, mu = 1000)
boxplot(Chipavg$wafer1, Chipavg$wafer2, name = c("Wafer 1", "Wafer 2"))
shapiro.test(Chipavg$wafer1)
shapiro.test(Chipavg$wafer2)
t.test(Chipavg$wafer1, Chipavg$wafer2, var.equal = TRUE)
```

---

Chips

*Four measurements on a first wafer and four measurements on a second wafer selected from 30 lots*

---

**Description**

Data for Exercise 10.9

**Usage**

Chips

**Format**

A data frame/tibble with 30 observations on eight variables

**wafer11** first measurement of thickness of the oxide layer for wafer1

**wafer12** second measurement of thickness of the oxide layer for wafer1

**wafer13** third measurement of thickness of the oxide layer for wafer1

**wafer14** fourth measurement of thickness of the oxide layer for wafer1

**wafer21** first measurement of thickness of the oxide layer for wafer2

**wafer22** second measurement of thickness of the oxide layer for wafer2

**wafer23** third measurement of thickness of the oxide layer for wafer2

**wafer24** fourth measurement of thickness of the oxide layer for wafer2

**Source**

Yashchin, E. 1995. "Likelihood Ratio Methods for Monitoring Parameters of a Nested Random Effect Model." *Journal of the American Statistical Association*, 90, 729-738.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
with(data = Chips,
      boxplot(wafer11, wafer12, wafer13, wafer14, wafer21,
              wafer22, wafer23, wafer24, col = "pink")
      )
```

---

Cigar

*Milligrams of tar in 25 cigarettes selected randomly from 4 different brands*

---

**Description**

Data for Example 10.4

**Usage**

Cigar

**Format**

A data frame/tibble with 100 observations on two variables

**tar** amount of tar (measured in milligrams)

**brand** a factor indicating cigarette brand with levels brandA, brandB, brandC, and brandD

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(tar ~ brand, data = Cigar, col = "cyan", ylab = "mg tar")
anova(lm(tar ~ brand, data = Cigar))
```

---

Cigarette

*Effect of mother's smoking on birth weight of newborn*

---

**Description**

Data for Exercise 2.27

**Usage**

Cigarette

**Format**

A data frame/tibble with 16 observations on two variables

**cigarettes** mothers' estimated average number of cigarettes smoked per day

**weight** children's birth weights (in pounds)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

plot(weight ~ cigarettes, data = Cigarette)
model <- lm(weight ~ cigarettes, data = Cigarette)
abline(model, col = "red")
with(data = Cigarette,
      cor(weight, cigarettes)
)
rm(model)

```

---

CIsim

*Confidence Interval Simulation Program*


---

**Description**

This program simulates random samples from which it constructs confidence intervals for one of the parameters mean ( $\mu$ ), variance ( $\sigma$ ), or proportion of successes ( $\pi$ ).

**Usage**

```

CIsim(
  samples = 100,
  n = 30,
  mu = 0,
  sigma = 1,
  conf.level = 0.95,
  type = "Mean"
)

```

**Arguments**

<code>samples</code>	the number of samples desired.
<code>n</code>	the size of each sample.
<code>mu</code>	if constructing confidence intervals for the population mean or the population variance, <code>mu</code> is the population mean (i.e., <code>type</code> is one of either "Mean", or "Var"). If constructing confidence intervals for the population proportion of successes, the value entered for <code>mu</code> represents the population proportion of successes ( $\pi$ ), and as such, must be a number between 0 and 1.
<code>sigma</code>	the population standard deviation. <code>sigma</code> is not required if confidence intervals are of type "Pi".
<code>conf.level</code>	confidence level for the graphed confidence intervals, restricted to lie between zero and one.
<code>type</code>	character string, one of "Mean", "Var" or "Pi", or just the initial letter of each, indicating the type of confidence interval simulation to perform.

**Details**

Default is to construct confidence intervals for the population mean. Simulated confidence intervals for the population variance or population proportion of successes are possible by selecting the appropriate value in the type argument.

**Value**

Graph depicts simulated confidence intervals. The number of confidence intervals that do not contain the parameter of interest are counted and reported in the commands window.

**Author(s)**

Alan T. Arnholt

**Examples**

```
CIsim(100, 30, 100, 10)
# Simulates 100 samples of size 30 from
# a normal distribution with mean 100
# and standard deviation 10. From the
# 100 simulated samples, 95% confidence
# intervals for the Mean are constructed
# and depicted in the graph.

CIsim(100, 30, 100, 10, type="Var")
# Simulates 100 samples of size 30 from
# a normal distribution with mean 100
# and standard deviation 10. From the
# 100 simulated samples, 95% confidence
# intervals for the variance are constructed
# and depicted in the graph.

CIsim(100, 50, .5, type="Pi", conf.level=.90)
# Simulates 100 samples of size 50 from
# a binomial distribution where the population
# proportion of successes is 0.5. From the
# 100 simulated samples, 90% confidence
# intervals for Pi are constructed
# and depicted in the graph.
```

**Description**

Data for Exercise 9.7

**Usage**

Citrus

**Format**

A data frame/tibble with nine observations on two variables

**age** age of children

**percent** percent peak bone density

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
model <- lm(percent ~ age, data = Citrus)
summary(model)
anova(model)
rm(model)
```

---

Clean

*Residual contaminant following the use of three different cleansing agents*

---

**Description**

Data for Exercise 10.16

**Usage**

Clean

**Format**

A data frame/tibble with 45 observations on two variables

**clean** residual contaminants

**agent** a factor with levels A, B, and C

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.



**Examples**

```
boxplot(clean ~ agent, col = c("red", "blue", "green"), data = Clean)
anova(lm(clean ~ agent, data = Clean))
```

---

Coaxial

*Signal loss from three types of coaxial cable*

---

**Description**

Data for Exercise 10.24 and 10.25

**Usage**

Coaxial

**Format**

A data frame/tibble with 45 observations on two variables

**signal** signal loss per 1000 feet

**cable** factor with three levels of coaxial cable typeA, typeB, and typeC

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(signal ~ cable, data = Coaxial, col = c("red", "green", "yellow"))
kruskal.test(signal ~ cable, data = Coaxial)
```

---

Coffee	<i>Productivity of workers with and without a coffee break</i>
--------	--

---

**Description**

Data for Exercise 7.55

**Usage**

Coffee

**Format**

A data frame/tibble with nine observations on three variables

**without** workers' productivity scores without a coffee break

**with** workers' productivity scores with a coffee break

**differences** with minus without

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Coffee$differences)
qqline(Coffee$differences)
shapiro.test(Coffee$differences)
t.test(Coffee$with, Coffee$without, paired = TRUE, alternative = "greater")
wilcox.test(Coffee$with, Coffee$without, paired = TRUE,
alterantive = "greater")
```

---

Coins	<i>Yearly returns on 12 investments</i>
-------	---

---

**Description**

Data for Exercise 5.68

**Usage**

Coins

**Format**

A data frame/tibble with 12 observations on one variable

**return** yearly returns on each of 12 possible investments

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Coins$return)
qqline(Coins$return)
```

---

Combinations

*Combinations*

---

**Description**

Computes all possible combinations of n objects taken k at a time.

**Usage**

```
Combinations(n, k)
```

**Arguments**

n                    a number.  
k                    a number less than or equal to n.

**Value**

Returns a matrix containing the possible combinations of n objects taken k at a time.

**See Also**

[SRS](#)

**Examples**

```
Combinations(5,2)
# The columns in the matrix list the values of the 10 possible
# combinations of 5 things taken 2 at a time.
```

---

 Commute

*Commuting times for selected cities in 1980 and 1990*


---

**Description**

Data for Exercises 1.13, and 7.85

**Usage**

Commute

**Format**

A data frame/tibble with 39 observations on three variables

**city** a factor with levels Atlanta, Baltimore, Boston, Buffalo, Charlotte, Chicago, Cincinnati, Cleveland, Columbus, Dallas, Denver, Detroit, Hartford, Houston, Indianapolis, Kansas City, Los Angeles, Miami, Milwaukee, Minneapolis, New Orleans, New York, Norfolk, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland, Providence, Rochester, Sacramento, Salt Lake City, San Antonio, San Diego, San Francisco, Seattle, St. Louis, Tampa, and Washington

**year** year

**time** commute times

**Source**

Federal Highway Administration.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stripplot(year ~ time, data = Commute, jitter = TRUE)
dotplot(year ~ time, data = Commute)
bwplot(year ~ time, data = Commute)
stripchart(time ~ year, data = Commute, method = "stack", pch = 1,
           cex = 2, col = c("red", "blue"),
           group.names = c("1980", "1990"),
           main = "", xlab = "minutes")
title(main = "Commute Time")
boxplot(time ~ year, data = Commute, names=c("1980", "1990"),
        horizontal = TRUE, las = 1)
```

---

Concept	<i>Tennessee self concept scale scores for a group of teenage boys</i>
---------	--

---

**Description**

Data for Exercise 1.68 and 1.82

**Usage**

Concept

**Format**

A data frame/tibble with 28 observations on one variable

**self** Tennessee self concept scores

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
summary(Concept$self)
sd(Concept$self)
diff(range(Concept$self))
IQR(Concept$self)
summary(Concept$self/10)
IQR(Concept$self/10)
sd(Concept$self/10)
diff(range(Concept$self/10))
```

---

Concrete	<i>Compressive strength of concrete blocks made by two different methods</i>
----------	--

---

**Description**

Data for Example 7.17

**Usage**

Concrete

**Format**

A data frame/tibble with 20 observations on two variables

**strength** comprehensive strength (in pounds per square inch)

**method** factor with levels new and old indicating the method used to construct a concrete block

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
wilcox.test(strength ~ method, data = Concrete, alternative = "greater")
```

---

Corn

*Comparison of the yields of a new variety and a standard variety of corn planted on 12 plots of land*

---

**Description**

Data for Exercise 7.77

**Usage**

Corn

**Format**

A data frame/tibble with 12 observations on three variables

**new** corn yield with new meathod

**standard** corn yield with standard method

**differences** new minus standard

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(Corn$differences)
qqnorm(Corn$differences)
qqline(Corn$differences)
shapiro.test(Corn$differences)
t.test(Corn$new, Corn$standard, paired = TRUE, alternative = "greater")
```

---

Correlat

*Exercise to illustrate correlation*

---

**Description**

Data for Exercise 2.23

**Usage**

Correlat

**Format**

A data frame/tibble with 13 observations on two variables

**x** a numeric vector

**y** a numeric vector

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(y ~ x, data = Correlat)
model <- lm(y ~ x, data = Correlat)
abline(model)
rm(model)
```

---

Counsel

*Scores of 18 volunteers who participated in a counseling process*

---

**Description**

Data for Exercise 6.96

**Usage**

Counsel

**Format**

A data frame/tibble with 18 observations on one variable

**score** standardized psychology scores after a counseling process

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Counsel$score)
t.test(Counsel$score, mu = 70)
```

---

Cpi

*Consumer price index from 1979 to 1998*

---

**Description**

Data for Exercise 1.34

**Usage**

Cpi

**Format**

A data frame/tibble with 20 observations on two variables

**year** year

**cpi** consumer price index



**Source**

Bureau of Labor Statistics.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(cpi ~ year, data = Cpi, type = "l", lty = 2, lwd = 2, col = "red")
barplot(Cpi$cpi, col = "pink", las = 2, main = "Problem 1.34")
```

---

Crime

*Violent crime rates for the states in 1983 and 1993*

---

**Description**

Data for Exercises 1.90, 2.32, 3.64, and 5.113

**Usage**

Crime

**Format**

A data frame/tibble with 102 observations on three variables

**state** a factor with levels Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, DC, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**year** a factor with levels 1983 and 1993

**rate** crime rate per 100,000 inhabitants

**Source**

U.S. Department of Justice, Bureau of Justice Statistics, *Sourcebook of Criminal Justice Statistics*, 1993.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(rate ~ year, data = Crime, col = "red")
```

---

Darwin

*Charles Darwin's study of cross-fertilized and self-fertilized plants*

---

## Description

Data for Exercise 7.62

## Usage

Darwin

## Format

A data frame/tibble with 15 observations on three variables

**pot** number of pot

**cross** height of plant (in inches) after a fixed period of time when cross-fertilized

**self** height of plant (in inches) after a fixed period of time when self-fertilized

## Source

Darwin, C. (1876) *The Effect of Cross- and Self-Fertilization in the Vegetable Kingdom*, 2nd edition, London.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
differ <- Darwin$cross - Darwin$self
qqnorm(differ)
qqline(differ)
shapiro.test(differ)
wilcox.test(Darwin$cross, Darwin$self, paired = TRUE)
rm(differ)
```

---

Dealers	<i>Automobile dealers classified according to type dealership and service rendered to customers</i>
---------	---

---

**Description**

Data for Example 2.22

**Usage**

Dealers

**Format**

A data frame/tibble with 122 observations on two variables

**type** a factor with levels Honda, Toyota, Mazda, Ford, Dodge, and Saturn

**service** a factor with levels Replaces unnecessarily and Follows manufacturer guidelines

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
xtabs(~type + service, data = Dealers)
T1 <- xtabs(~type + service, data = Dealers)
T1
addmargins(T1)
pt <- prop.table(T1, margin = 1)
pt
barplot(t(pt), col = c("red", "skyblue"), legend = colnames(T1))
rm(T1, pt)
```

---

Defectiv	<i>Number of defective items produced by 20 employees</i>
----------	---

---

**Description**

Data for Exercise 1.27

**Usage**

Defectiv

**Format**

A data frame/tibble with 20 observations on one variable

**number** number of defective items produced by the employees in a small business firm

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~ number, data = Defectiv)
T1
barplot(T1, col = "pink", ylab = "Frequency",
xlab = "Defective Items Produced by Employees", main = "Problem 1.27")
rm(T1)
```

---

Degree

*Percent of bachelor's degrees awarded women in 1970 versus 1990*

---

**Description**

Data for Exercise 2.75

**Usage**

Degree

**Format**

A data frame/tibble with 1064 observations on two variables

**field** a factor with levels Health, Education, Foreign Language, Psychology, Fine Arts, Life Sciences, Business, Social Science, Physical Sciences, Engineering, and All Fields

**awarded** a factor with levels 1970 and 1990

**Source**

U.S. Department of Health and Human Services, National Center for Education Statistics.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~field + awarded, data = Degree)
T1
barplot(t(T1), beside = TRUE, col = c("red", "skyblue"), legend = colnames(T1))
rm(T1)
```

---

Delay

*Delay times on 20 flights from four major air carriers*

---

**Description**

Data for Exercise 10.55

**Usage**

Delay

**Format**

A data frame/tibble with 80 observations on two variables

**delay** the delay time (in minutes) for 80 randomly selected flights

**carrier** a factor with levels A, B, C, and D

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(delay ~ carrier, data = Delay,
        main = "Exercise 10.55", ylab = "minutes",
        col = "pink")
kruskal.test(delay ~ carrier, data = Delay)
```

---

Depend                      *Number of dependent children for 50 families*

---

**Description**

Data for Exercise 1.26

**Usage**

Depend

**Format**

A data frame/tibble with 50 observations on one variable

**number** number of dependent children in a family

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~ number, data = Depend)
T1
barplot(T1, col = "lightblue", main = "Problem 1.26",
xlab = "Number of Dependent Children", ylab = "Frequency")
rm(T1)
```

---

Detroit                      *Educational levels of a sample of 40 auto workers in Detroit*

---

**Description**

Data for Exercise 5.21

**Usage**

Detroit

**Format**

A data frame/tibble with 40 observations on one variable

**educ** the educational level (in years) of a sample of 40 auto workers in a plant in Detroit

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Detroit$educ)
```

---

Develop	<i>Demographic characteristics of developmental students at 2-year colleges and 4-year colleges</i>
---------	---

---

**Description**

Data used for Exercise 8.50

**Usage**

```
Develop
```

**Format**

A data frame/tibble with 5656 observations on two variables

**race** a factor with levels African American, American Indian, Asian, Latino, and White

**college** a factor with levels Two-year and Four-year

**Source**

*Research in Development Education* (1994), V. 11, 2.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~race + college, data = Develop)
T1
chisq.test(T1)
rm(T1)
```

---

Devmath	<i>Test scores for students who failed developmental mathematics in the fall semester 1995</i>
---------	--

---

**Description**

Data for Exercise 6.47

**Usage**

Devmath

**Format**

A data frame/tibble with 40 observations on one variable

**score** first exam score

**Source**

Data provided by Dr. Anita Kitchens.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Devmath$score)
t.test(Devmath$score, mu = 80, alternative = "less")
```

---

Dice	<i>Outcomes and probabilities of the roll of a pair of fair dice</i>
------	--

---

**Description**

Data for Exercise 3.109

**Usage**

Dice



**Format**

A data frame/tibble with 11 observations on two variables

**x** possible outcomes for the sum of two dice

**px** probability for outcome x

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
roll1 <- sample(1:6, 20000, replace = TRUE)
roll2 <- sample(1:6, 20000, replace = TRUE)
outcome <- roll1 + roll2
T1 <- table(outcome)/length(outcome)
remove(roll1, roll2, outcome)
T1
round(t(Dice), 5)
rm(roll1, roll2, T1)
```

---

Diesel

*Diesel fuel prices in 1999-2000 in nine regions of the country*

---

**Description**

Data for Exercise 2.8

**Usage**

Diesel

**Format**

A data frame/tibble with 650 observations on three variables

**date** date when price was recorded

**pricepergallon** price per gallon (in dollars)

**location** a factor with levels California, CentralAtlantic, Coast, EastCoast, Gulf, LowerAtlantic, NatAvg, NorthEast, Rocky, and WesternMountain

**Source**

Energy Information Administration, National Energy Information Center: 1000 Independence Ave., SW, Washington, D.C., 20585.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
par(las = 2)
boxplot(pricepergallon ~ location, data = Diesel)
boxplot(pricepergallon ~ location,
        data = droplevels(Diesel[Diesel$location == "EastCoast" |
        Diesel$location == "Gulf" | Diesel$location == "NatAvg" |
        Diesel$location == "Rocky" | Diesel$location == "California", ]),
        col = "pink", main = "Exercise 2.8")
par(las = 1)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Diesel, aes(x = date, y = pricepergallon,
        color = location)) +
  geom_point() +
  geom_smooth(se = FALSE) +
  theme_bw() +
  labs(y = "Price per Gallon (in dollars)")

## End(Not run)
```

---

Diplomat

*Parking tickets issued to diplomats*

---

## Description

Data for Exercises 1.14 and 1.37

## Usage

Diplomat

## Format

A data frame/tibble with 10 observations on three variables

**country** a factor with levels Brazil, Bulgaria, Egypt, Indonesia, Israel, Nigeria, Russia, S. Korea, Ukraine, and Venezuela

**number** total number of tickets

**rate** number of tickets per vehicle per month

## Source

*Time*, November 8, 1993. Figures are from January to June 1993.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
par(las = 2, mfrow = c(2, 2))
stripchart(number ~ country, data = Diplomat, pch = 19,
           col= "red", vertical = TRUE)
stripchart(rate ~ country, data = Diplomat, pch = 19,
           col= "blue", vertical = TRUE)
with(data = Diplomat,
     barplot(number, names.arg = country, col = "red"))
with(data = Diplomat,
     barplot(rate, names.arg = country, col = "blue"))
par(las = 1, mfrow = c(1, 1))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Diplomat, aes(x = reorder(country, number),
                                     y = number)) +
  geom_bar(stat = "identity", fill = "pink", color = "black") +
  theme_bw() + labs(x = "", y = "Total Number of Tickets")
ggplot2::ggplot(data = Diplomat, aes(x = reorder(country, rate),
                                     y = rate)) +
  geom_bar(stat = "identity", fill = "pink", color = "black") +
  theme_bw() + labs(x = "", y = "Tickets per vehicle per month")

## End(Not run)
```

---

Disposal

*Toxic intensity for manufacturing plants producing herbicidal preparations*

---

## Description

Data for Exercise 1.127

## Usage

Disposal

## Format

A data frame/tibble with 29 observations on one variable

**pounds** pounds of toxic waste per \$1000 of shipments of its products

**Source**

Bureau of the Census, *Reducing Toxins*, Statistical Brief SB/95-3, February 1995.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Disposal$pounds)
fivenum(Disposal$pounds)
EDA(Disposal$pounds)
```

---

Dogs

*Rankings of the favorite breeds of dogs*

---

**Description**

Data for Exercise 2.88

**Usage**

Dogs

**Format**

A data frame/tibble with 20 observations on three variables

**breed** a factor with levels Beagle, Boxer, Chihuahua, Chow, Dachshund, Dalmatian, Doberman, Huskie, Labrador, Pomeranian, Poodle, Retriever, Rotweiler, Schnauzer, Shepherd, Shetland, ShihTzu, Spaniel, Springer, and Yorkshire

**ranking** numeric ranking

**year** a factor with levels 1992, 1993, 1997, and 1998

**Source**

*The World Almanac and Book of Facts*, 2000.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

cor(Dogs$ranking[Dogs$year == "1992"], Dogs$ranking[Dogs$year == "1993"])
cor(Dogs$ranking[Dogs$year == "1997"], Dogs$ranking[Dogs$year == "1998"])
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Dogs, aes(x = reorder(breed, ranking), y = ranking)) +
  geom_bar(stat = "identity") +
  facet_grid(year ~. ) +
  theme(axis.text.x = element_text(angle = 85, vjust = 0.5))

## End(Not run)

```

---

Domestic

*Rates of domestic violence per 1,000 women by age groups*


---

**Description**

Data for Exercise 1.20

**Usage**

Domestic

**Format**

A data frame/tibble with five observations on two variables

**age** a factor with levels 12–19, 20–24, 25–34, 35–49, and 50–64

**rate** rate of domestic violence per 1000 women

**Source**

U.S. Department of Justice.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

barplot(Domestic$rate, names.arg = Domestic$age)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Domestic, aes(x = age, y = rate)) +
  geom_bar(stat = "identity", fill = "purple", color = "black") +
  labs(x = "", y = "Domestic violence per 1000 women") +

```

```
theme_bw()

## End(Not run)
```

---

Dopamine	<i>Dopamine b-hydroxylase activity of schizophrenic patients treated with an antipsychotic drug</i>
----------	---

---

### Description

Data for Exercises 5.14 and 7.49

### Usage

Dopamine

### Format

A data frame/tibble with 25 observations on two variables

**dbh** dopamine b-hydroxylase activity (units are nmol/(ml)(h)/(mg) of protein)

**group** a factor with levels nonpsychotic and psychotic

### Source

D.E. Sternberg, D.P. Van Kammen, and W.E. Bunney, "Schizophrenia: Dopamine b-Hydroxylase Activity and Treatment Response," *Science*, 216 (1982), 1423 - 1425.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
boxplot(dbh ~ group, data = Dopamine, col = "orange")
t.test(dbh ~ group, data = Dopamine, var.equal = TRUE)
```

---

Dowjones	<i>Closing yearend Dow Jones Industrial averages from 1896 through 2000</i>
----------	---

---

### Description

Data for Exercise 1.35

### Usage

Dowjones

### Format

A data frame/tibble with 105 observations on three variables

**year** date

**close** Dow Jones closing price

**change** percent change from previous year

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
plot(close ~ year, data = Dowjones, type = "l", main = "Exercise 1.35")
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Dowjones, aes(x = year, y = close)) +
  geom_point(size = 0.5) +
  geom_line(color = "red") +
  theme_bw() +
  labs(y = "Dow Jones Closing Price")

## End(Not run)
```

---

Drink	<i>Opinion on referendum by view on moral issue of selling alcoholic beverages</i>
-------	--

---

**Description**

Data for Exercise 8.53

**Usage**

Drink

**Format**

A data frame/tibble with 472 observations on two variables

**drinking** a factor with levels ok, tolerated, and immoral

**referendum** a factor with levels for, against, and undecided

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~drinking + referendum, data = Drink)
T1
chisq.test(T1)
rm(T1)
```

---

Drug	<i>Number of trials to master a task for a group of 28 subjects assigned to a control and an experimental group</i>
------	---

---

**Description**

Data for Example 7.15

**Usage**

Drug



**Format**

A data frame/tibble with 28 observations on two variables

**trials** number of trials to master a task

**group** a factor with levels control and experimental

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(trials ~ group, data = Drug,
        main = "Example 7.15", col = c("yellow", "red"))
wilcox.test(trials ~ group, data = Drug)
t.test(rank(trials) ~ group, data = Drug, var.equal = TRUE)
```

---

Dyslexia

*Data on a group of college students diagnosed with dyslexia*

---

**Description**

Data for Exercise 2.90

**Usage**

Dyslexia

**Format**

A data frame/tibble with eight observations on seven variables

**words** number of words read per minute

**age** age of participant

**gender** a factor with levels female and male

**handed** a factor with levels left and right

**weight** weight of participant (in pounds)

**height** height of participant (in inches)

**children** number of children in family

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(height ~ weight, data = Dyslexia)
plot(words ~ factor(handed), data = Dyslexia,
      xlab = "hand", col = "lightblue")
```

---

Earthqk

*One hundred year record of worldwide seismic activity(1770-1869)*

---

**Description**

Data for Exercise 6.97

**Usage**

Earthqk

**Format**

A data frame/tibble with 100 observations on two variables

**year** year seismic activity recorded

**severity** annual incidence of severe earthquakes

**Source**

Quenoille, M.H. (1952), *Associated Measurements*, Butterworth, London. p 279.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Earthqk$severity)
t.test(Earthqk$severity, mu = 100, alternative = "greater")
```

**Description**

Function that produces a histogram, density plot, boxplot, and Q-Q plot.

**Usage**

```
EDA(x, trim = 0.05)
```

**Arguments**

x	numeric vector. NAs and Infs are allowed but will be removed.
trim	fraction (between 0 and 0.5, inclusive) of values to be trimmed from each end of the ordered data. If trim = 0.5, the result is the median.

**Details**

Will not return command window information on data sets containing more than 5000 observations. It will however still produce graphical output for data sets containing more than 5000 observations.

**Value**

Function returns various measures of center and location. The values returned for the Quartiles are based on the definitions provided in *BSDA*. The boxplot is based on the Quartiles returned in the commands window.

**Note**

Requires package **e1071**.

**Author(s)**

Alan T. Arnholt

**Examples**

```
EDA(rnorm(100))  
# Produces four graphs for the 100 randomly  
# generated standard normal variates.
```

---

Educat	<i>Crime rates versus the percent of the population without a high school degree</i>
--------	--

---

### Description

Data for Exercise 2.41

### Usage

Educat

### Format

A data frame/tibble with 51 observations on three variables

**state** a factor with levels Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, DC, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**nodegree** percent of the population without a high school degree

**crime** violent crimes per 100,000 population

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
plot(crime ~ nodegree, data = Educat,  
     xlab = "Percent of population without high school degree",  
     ylab = "Violent Crime Rate per 100,000")
```

---

Eggs	<i>Number of eggs versus amounts of feed supplement</i>
------	---

---

**Description**

Data for Exercise 9.22

**Usage**

Eggs

**Format**

A data frame/tibble with 12 observations on two variables

**feed** amount of feed supplement

**eggs** number of eggs per day for 100 chickens

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(eggs ~ feed, data = Eggs)
model <- lm(eggs ~ feed, data = Eggs)
abline(model, col = "red")
summary(model)
rm(model)
```

---

Elderly	<i>Percent of the population over the age of 65</i>
---------	---

---

**Description**

Data for Exercise 1.92 and 2.61

**Usage**

Elderly

**Format**

A data frame/tibble with 51 observations on three variables

**state** a factor with levels Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**percent1985** percent of the population over the age of 65 in 1985

**percent1998** percent of the population over the age of 65 in 1998

**Source**

U.S. Census Bureau Internet site, February 2000.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
with(data = Elderly,
stripchart(x = list(percent1998, percent1985), method = "stack", pch = 19,
           col = c("red", "blue"), group.names = c("1998", "1985"))
)
with(data = Elderly, cor(percent1998, percent1985))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Elderly, aes(x = percent1985, y = percent1998)) +
  geom_point() +
  theme_bw()

## End(Not run)
```

---

Energy

*Amount of energy consumed by homes versus their sizes*

---

**Description**

Data for Exercises 2.5, 2.24, and 2.55

**Usage**

Energy

**Format**

A data frame/tibble with 12 observations on two variables

**size** size of home (in square feet)

**kilowatt** kilowatt-hours per month

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(kilowatt ~ size, data = Energy)
with(data = Energy, cor(size, kilowatt))
model <- lm(kilowatt ~ size, data = Energy)
plot(Energy$size, resid(model), xlab = "size")
```

---

Engineer

*Salaries after 10 years for graduates of three different universities*

---

**Description**

Data for Example 10.7

**Usage**

Engineer

**Format**

A data frame/tibble with 51 observations on two variables

**salary** salary (in \$1000) 10 years after graduation

**university** a factor with levels A, B, and C

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(salary ~ university, data = Engineer,  
        main = "Example 10.7", col = "yellow")  
kruskal.test(salary ~ university, data = Engineer)  
anova(lm(salary ~ university, data = Engineer))  
anova(lm(rank(salary) ~ university, data = Engineer))
```

---

Entrance

*College entrance exam scores for 24 high school seniors*

---

**Description**

Data for Example 1.8

**Usage**

Entrance

**Format**

A data frame/tibble with 24 observations on one variable

**score** college entrance exam score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Entrance$score)  
stem(Entrance$score, scale = 2)
```



---

Epaminicompact

*Fuel efficiency ratings for compact vehicles in 2001*

---

### Description

Data for Exercise 1.65

### Usage

Epaminicompact

### Format

A data frame/tibble with 22 observations on ten variables

**class** a character variable with value MINICOMPACT CARS

**manufacturer** a character variable with values AUDI, BMW, JAGUAR, MERCEDES-BENZ, MITSUBISHI, and PORSCHE

**carline** a character variable with values 325CI CONVERTIBLE, 330CI CONVERTIBLE, 911 CARRERA 2/4, 911 TURBO, CLK320 (CABRIOLET), CLK430 (CABRIOLET), ECLIPSE SPYDER, JAGUAR XK8 CONVERTIBLE, JAGUAR XKR CONVERTIBLE, M3 CONVERTIBLE, TT COUPE, and TT COUPE QUATTRO

**displ** engine displacement (in liters)

**cyl** number of cylinders

**trans** a factor with levels Auto(L5), Auto(S4), Auto(S5), Manual(M5), and Manual(M6)

**drv** a factor with levels 4(four wheel drive), F(front wheel drive), and R(rear wheel drive)

**cty** city mpg

**hwy** highway mpg

**cmb** combined city and highway mpg

### Source

EPA data.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
summary(Epaminicompact$cty)
plot(hwy ~ cty, data = Epaminicompact)
```

---

 Epatwoseater

*Fuel efficiency ratings for two-seater vehicles in 2001*


---

**Description**

Data for Exercise 5.8

**Usage**

Epatwoseater

**Format**

A data frame/tibble with 36 observations on ten variables

**class** a character variable with value TWO SEATERS

**manufacturer** a character variable with values ACURA, AUDI, BMW, CHEVROLET, DODGE, FERRARI, HONDA, LAMBORGHINI, MAZDA, MERCEDES-BENZ, PLYMOUTH, PORSCHE, and TOYOTA

**carline** a character variable with values BOXSTER, BOXSTER S, CORVETTE, DB132/144 DIABLO, FERRARI 360 MODENA/SPIDER, FERRARI 550 MARANELLO/BARCHETTA, INSIGHT, MR2 ,MX-5 MIATA, NSX, PROWLER, S2000, SL500, SL600, SLK230 KOMPRESSOR, SLK320, TT ROADSTER, TT ROADSTER QUATTRO, VIPER CONVERTIBLE, VIPER COUPE, Z3 COUPE, Z3 ROADSTER, and Z8

**displ** engine displacement (in liters)

**cyl** number of cylinders

**trans** a factor with levels Auto(L4), Auto(L5), Auto(S4), Auto(S5), Auto(S6), Manual(M5), and Manual(M6)

**drv** a factor with levels 4(four wheel drive) F(front wheel drive) R(rear wheel drive)

**cty** city mpg

**hwy** highway mpg

**cmb** combined city and highway mpg

@source Environmental Protection Agency.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
summary(Epatwoseater$cty)
plot(hwy ~ cty, data = Epatwoseater)
boxplot(cty ~ drv, data = Epatwoseater, col = "lightgreen")
```

---

Executiv	<i>Ages of 25 executives</i>
----------	------------------------------

---

**Description**

Data for Exercise 1.104

**Usage**

Executiv

**Format**

A data frame/tibble with 25 observations on one variable

**age** a numeric vector

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Executiv$age, xlab = "Age of banking executives",  
     breaks = 5, main = "", col = "gray")
```

---

Exercise	<i>Weight loss for 30 members of an exercise program</i>
----------	--

---

**Description**

Data for Exercise 1.44

**Usage**

Exercise

**Format**

A data frame/tibble with 30 observations on one variable

**loss** a numeric vector

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
stem(Exercise$loss)
```

---

Fabric	<i>Measures of softness of ten different clothing garments washed with and without a softener</i>
--------	---

---

## Description

Data for Example 7.21

## Usage

Fabric

## Format

A data frame/tibble with 20 observations on three variables

**garment** a numeric vector

**softner** a character variable with values with and without

**softness** a numeric vector

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
wilcox.test(softness ~ softner, data = Fabric,
            paired = TRUE, alternative = "greater")

## Not run:
library(tidyr)
T7 <- tidyr::spread(Fabric, softner, softness) %>%
mutate(di = with - without, adi = abs(di), rk = rank(adi),
       srk = sign(di)*rk)
T7
t.test(T7$srk, alternative = "greater")

## End(Not run)
```

---

Faithful

*Waiting times between successive eruptions of the Old Faithful geyser*

---

### Description

Data for Exercise 5.12 and 5.111

### Usage

Faithful

### Format

A data frame/tibble with 299 observations on two variables

**time** a numeric vector

**eruption** a factor with levels 1 and 2

### Source

A. Azzalini and A. Bowman, "A Look at Some Data on the Old Faithful Geyser," *Journal of the Royal Statistical Society, Series C*, 39 (1990), 357-366.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
t.test(time ~ eruption, data = Faithful)
hist(Faithful$time, xlab = "wait time", main = "", freq = FALSE)
lines(density(Faithful$time))

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Faithful, aes(x = time, y = ..density..)) +
  geom_histogram(binwidth = 5, fill = "pink", col = "black") +
  geom_density() +
  theme_bw() +
  labs(x = "wait time")

## End(Not run)
```

---

Family

*Size of family versus cost per person per week for groceries*

---

## Description

Data for Exercise 2.89

## Usage

Family

## Format

A data frame/tibble with 20 observations on two variables

**number** number in family

**cost** cost per person (in dollars)

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(cost ~ number, data = Family)
abline(lm(cost ~ number, data = Family), col = "red")
cor(Family$cost, Family$number)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Family, aes(x = number, y = cost)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()

## End(Not run)
```

---

Ferraro1

*Choice of presidential ticket in 1984 by gender*

---

### Description

Data for Exercise 8.23

### Usage

Ferraro1

### Format

A data frame/tibble with 1000 observations on two variables

**gender** a factor with levels Men and Women

**candidate** a character vector of 1984 president and vice-president candidates

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
T1 <- xtabs(~gender + candidate, data = Ferraro1)
T1
chisq.test(T1)
rm(T1)
```

---

Ferraro2

*Choice of vice presidential candidate in 1984 by gender*

---

### Description

Data for Exercise 8.23

### Usage

Ferraro2

**Format**

A data frame/tibble with 1000 observations on two variables

**gender** a factor with levels Men and Women

**candidate** a character vector of 1984 president and vice-president candidates

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~gender + candidate, data = Ferraro2)
T1
chisq.test(T1)
rm(T1)
```

---

Fertility

*Fertility rates of all 50 states and DC*

---

**Description**

Data for Exercise 1.125

**Usage**

Fertility

**Format**

A data frame/tibble with 51 observations on two variables

**state** a character variable with values Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**rate** fertility rate (expected number of births during childbearing years)

**Source**

Population Reference Bureau.



**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Fertility$rate)
fivenum(Fertility$rate)
EDA(Fertility$rate)
```

---

Firstchi

*Ages of women at the birth of their first child*

---

**Description**

Data for Exercise 5.11

**Usage**

Firstchi

**Format**

A data frame/tibble with 87 observations on one variable

**age** age of woman at birth of her first child

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Firstchi$age)
```

---

Fish

*Length and number of fish caught with small and large mesh codend*

---

### Description

Data for Exercises 5.83, 5.119, and 7.29

### Usage

Fish

### Format

A data frame/tibble with 1534 observations on two variables

**codend** a character variable with values `smallmesh` and `largemesh`

**length** length of the fish measured in centimeters

### Source

R. Millar, "Estimating the Size - Selectivity of Fishing Gear by Conditioning on the Total Catch," *Journal of the American Statistical Association*, 87 (1992), 962 - 968.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
tapply(Fish$length, Fish$codend, median, na.rm = TRUE)
SIGN.test(Fish$length[Fish$codend == "smallmesh"], conf.level = 0.99)
## Not run:
dplyr::group_by(Fish, codend) %>%
  summarize(MEDIAN = median(length, na.rm = TRUE))

## End(Not run)
```

---

Fitness	<i>Number of sit-ups before and after a physical fitness course</i>
---------	---

---

### Description

Data for Exercise 7.71

### Usage

Fitness

### Format

A data frame/tibble with 18 observations on the three variables

**subject** a character variable indicating subject number

**test** a character variable with values After and Before

**number** a numeric vector recording the number of sit-ups performed in one minute

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
t.test(number ~ test, data = Fitness, alternative = "greater", paired = TRUE)
## Not run:
Wide <- tidyr::spread(Fitness, test, number) %>%
mutate(diff = After - Before)
Wide
qqnorm(Wide$diff)
qqline(Wide$diff)
t.test(Wide$diff, alternative = "greater")

## End(Not run)
```

---

Florida2000

*Florida voter results in the 2000 presidential election*

---

### Description

Data for Statistical Insight Chapter 2

### Usage

Florida2000

### Format

A data frame/tibble with 67 observations on 12 variables

**county** a character variable with values ALACHUA, BAKER, BAY, BRADFORD, BREVARD, BROWARD, CALHOUN, CHARLOTTE, CITRUS, CLAY, COLLIER, COLUMBIA, DADE, DE SOTO, DIXIE, DUVAL, ESCAMBIA, FLAGLER, FRANKLIN, GADSDEN, GILCHRIST, GLADES, GULF, HAMILTON, HARDEE, HENDRY, HERNANDO, HIGHLANDS, HILLSBOROUGH, HOLMES, INDIAN RIVER, JACKSON, JEFFERSON, LAFAYETTE, LAKE, LEE, LEON, LEVY, LIBERTY, MADISON, MANATEE, MARION, MARTIN, MONROE, NASSAU, OKALOOSA, OKEECHOBEE, ORANGE, OSCEOLA, PALM BEACH, PASCO, PINELLAS, POLK, PUTNAM, SANTA ROSA, SARASOTA, SEMINOLE, ST. JOHNS, ST. LUCIE, SUMTER, SUWANNEE, TAYLOR, UNION, VOLUSIA, WAKULLA, WALTON, and WASHINGTON

**gore** number of votes

**bush** number of votes

**buchanan** number of votes

**nader** number of votes

**browne** number of votes

**hagelin** number of votes

**harris** number of votes

**mcreynolds** number of votes

**moorehead** number of votes

**phillips** number of votes

**total** number of votes

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
plot(buchanan ~ total, data = Florida2000,  
     xlab = "Total votes cast (in thousands)",  
     ylab = "Votes for Buchanan")
```

---

Fluid	<i>Breakdown times of an insulating fluid under various levels of voltage stress</i>
-------	--

---

**Description**

Data for Exercise 5.76

**Usage**

Fluid

**Format**

A data frame/tibble with 76 observations on two variables

**kilovolts** a character variable showing kilowats

**time** breakdown time (in minutes)

**Source**

E. Soofi, N. Ebrahimi, and M. Habibullah, 1995.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
DF1 <- Fluid[Fluid$kilovolts == "34kV", ]
DF1
# OR
DF2 <- subset(Fluid, subset = kilovolts == "34kV")
DF2
stem(DF2$time)
SIGN.test(DF2$time)
## Not run:
library(dplyr)
DF3 <- dplyr::filter(Fluid, kilovolts == "34kV")
DF3

## End(Not run)
```

Food

*Annual food expenditures for 40 single households in Ohio*

---

**Description**

Data for Exercise 5.106

**Usage**

Food

**Format**

A data frame/tibble with 40 observations on one variable

**expenditure** a numeric vector recording annual food expenditure (in dollars) in the state of Ohio.**Source**

Bureau of Labor Statistics.

**References**Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.**Examples**EDA(Food\$expenditure)

---

Framingh

*Cholesterol values of 62 subjects in the Framingham Heart Study*

---

**Description**

Data for Exercises 1.56, 1.75, 3.69, and 5.60

**Usage**

Framingh

**Format**

A data frame/tibble with 62 observations on one variable

**cholest** a numeric vector with cholesterol values

**Source**

R. D'Agostino, et al., (1990) "A Suggestion for Using Powerful and Informative Tests for Normality," *The American Statistician*, 44 316-321.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Framingh$cholest)
boxplot(Framingh$cholest, horizontal = TRUE)
hist(Framingh$cholest, freq = FALSE)
lines(density(Framingh$cholest))
mean(Framingh$cholest > 200 & Framingh$cholest < 240)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Framingh, aes(x = factor(1), y = cholest)) +
  geom_boxplot() +           # boxplot
  labs(x = "") +           # no x label
  theme_bw() +             # black and white theme
  geom_jitter(width = 0.2) + # jitter points
  coord_flip()             # Create horizontal plot
ggplot2::ggplot(data = Framingh, aes(x = cholest, y = ..density..)) +
  geom_histogram(fill = "pink", binwidth = 15, color = "black") +
  geom_density() +
  theme_bw()

## End(Not run)
```

---

Freshman

*Ages of a random sample of 30 college freshmen*

---

**Description**

Data for Exercise 6.53

**Usage**

Freshman

**Format**

A data frame/tibble with 30 observations on one variable

**age** a numeric vector of ages

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
SIGN.test(Freshman$age, md = 19)
```

---

Funeral

*Cost of funeral by region of country*

---

**Description**

Data for Exercise 8.54

**Usage**

Funeral

**Format**

A data frame/tibble with 400 observations on two variables

**region** a factor with levels Central, East, South, and West

**cost** a factor with levels less than expected, about what expected, and more than expected

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~region + cost, data = Funeral)
T1
chisq.test(T1)
rm(T1)
```



---

Galaxie

*Velocities of 82 galaxies in the Corona Borealis region*

---

**Description**

Data for Example 5.2

**Usage**

Galaxie

**Format**

A data frame/tibble with 82 observations on one variable

**velocity** velocity measured in kilometers per second

**Source**

K. Roeder, "Density Estimation with Confidence Sets Explained by Superclusters and Voids in the Galaxies," *Journal of the American Statistical Association*, 85 (1990), 617-624.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Galaxie$velocity)
```

---

Gallup

*Results of a Gallup poll on possession of marijuana as a criminal offense conducted in 1980*

---

**Description**

Data for Exercise 2.76

**Usage**

Gallup

**Format**

A data frame/tibble with 1,200 observations on two variables

**demographics** a factor with levels National, Gender: Male Gender: Female, Education: College, Education: High School, Education: Grade School, Age: 18-24, Age: 25-29, Age: 30-49, Age: 50-older, Religion: Protestant, and Religion: Catholic

**opinion** a factor with levels Criminal, Not Criminal, and No Opinion

**Source**

George H. Gallup *The Gallup Opinion Index Report No. 179* (Princeton, NJ: The Gallup Poll, July 1980), p. 15.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~demographics + opinion, data = Gallup)
T1
t(T1[c(2, 3), ])
barplot(t(T1[c(2, 3), ]))
barplot(t(T1[c(2, 3), ]), beside = TRUE)

## Not run:
library(dplyr)
library(ggplot2)
dplyr::filter(Gallup, demographics == "Gender: Male" | demographics == "Gender: Female") %>%
ggplot2::ggplot(aes(x = demographics, fill = opinion)) +
  geom_bar() +
  theme_bw() +
  labs(y = "Fraction")

## End(Not run)
```

---

Gasoline

*Price of regular unleaded gasoline obtained from 25 service stations*

---

**Description**

Data for Exercise 1.45

**Usage**

Gasoline

**Format**

A data frame/tibble with 25 observations on one variable

**price** price for one gallon of gasoline

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Gasoline$price)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Gasoline, aes(x = factor(1), y = price)) +
  geom_violin() +
  geom_jitter() +
  theme_bw()

## End(Not run)
```

---

German

*Number of errors in copying a German passage before and after an experimental course in German*

---

**Description**

Data for Exercise 7.60

**Usage**

German

**Format**

A data frame/tibble with ten observations on three variables

**student** a character variable indicating student number

**when** a character variable with values Before and After to indicate when the student received experimental instruction in German

**errors** the number of errors in copying a German passage

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
t.test(errors ~ when, data = German, paired = TRUE)
wilcox.test(errors ~ when, data = German)

## Not run:
T8 <- tidyr::spread(German, when, errors) %>%
mutate(di = After - Before, adi = abs(di), rk = rank(adi), srk = sign(di)*rk)
T8
qqnorm(T8$di)
qqline(T8$di)
t.test(T8$srk)

## End(Not run)
```

---

Golf

*Distances a golf ball can be driven by 20 professional golfers*

---

## Description

Data for Exercise 5.24

## Usage

Golf

## Format

A data frame/tibble with 20 observations on one variable

**yards** distance a golf ball is driven in yards

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
stem(Golf$yards)
qqnorm(Golf$yards)
qqline(Golf$yards)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Golf, aes(sample = yards)) +
  geom_qq() +
  theme_bw()

## End(Not run)
```

---

Governor

*Annual salaries for state governors in 1994 and 1999*

---

## Description

Data for Exercise 5.112

## Usage

Governor

## Format

A data frame/tibble with 50 observations on three variables

**state** a character variable with values Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**year** a factor indicating year

**salary** a numeric vector with the governor's salary (in dollars)

## Source

*The 2000 World Almanac and Book of Facts.*

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

boxplot(salary ~ year, data = Governor)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Governor, aes(x = salary)) +
  geom_density(fill = "pink") +
  facet_grid(year ~ .) +
  theme_bw()

## End(Not run)

```

---

Gpa

*High school GPA versus college GPA*


---

**Description**

Data for Example 2.13

**Usage**

Gpa

**Format**

A data frame/tibble with 10 observations on two variables

**hsgpa** high school gpa

**collgpa** college gpa

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

plot(collgpa ~ hsgpa, data = Gpa)
mod <- lm(collgpa ~ hsgpa, data = Gpa)
abline(mod)           # add line
yhat <- predict(mod)  # fitted values
e <- resid(mod)       # residuals
cbind(Gpa, yhat, e)    # Table 2.1
cor(Gpa$hsgpa, Gpa$collgpa)

## Not run:

```

```
library(ggplot2)
ggplot2::ggplot(data = Gpa, aes(x = hsgpa, y = collgpa)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()

## End(Not run)
```

---

Grades

*Test grades in a beginning statistics class*

---

### Description

Data for Exercise 1.120

### Usage

Grades

### Format

A data frame with 29 observations on one variable

**grades** a numeric vector containing test grades

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
hist(Grades$grades, main = "", xlab = "Test grades", right = FALSE)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Grades, aes(x = grades, y = ..density..)) +
  geom_histogram(fill = "pink", binwidth = 5, color = "black") +
  geom_density(lwd = 2, color = "red") +
  theme_bw()

## End(Not run)
```

---

Graduate

*Graduation rates for student athletes in the Southeastern Conf.*

---

**Description**

Data for Exercise 1.118

**Usage**

Graduate

**Format**

A data frame/tibble with 12 observations on three variables

**school** a character variable with values Alabama, Arkansas, Auburn, Florida, Georgia, Kentucky, Louisiana St, Mississippi, Mississippi St, South Carolina, Tennessee, and Vanderbilt

**code** a character variable with values Al, Ar, Au Fl, Ge, Ke, LSt, Mi, MSt, SC, Te, and Va

**percent** graduation rate

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
barplot(Graduate$percent, names.arg = Graduate$school,  
       las = 2, cex.names = 0.7, col = "tomato")
```

---

Greenriv

*Varve thickness from a sequence through an Eocene lake deposit in the Rocky Mountains*

---

**Description**

Data for Exercise 6.57

**Usage**

Greenriv



**Format**

A data frame/tibble with 37 observations on one variable

**thick** varve thickness in millimeters

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Greenriv$thick)
SIGN.test(Greenriv$thick, md = 7.3, alternative = "greater")
```

---

Grnriv2	<i>Thickness of a varved section of the Green river oil shale deposit near a major lake in the Rocky Mountains</i>
---------	--

---

**Description**

Data for Exercises 6.45 and 6.98

**Usage**

Grnriv2

**Format**

A data frame/tibble with 101 observations on one variable

**thick** varve thickness (in millimeters)

**Source**

J. Davis, *Statistics and Data Analysis in Geology*, 2nd Ed., Jon Wiley and Sons, New York.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Grnriv2$thick)
t.test(Grnriv2$thick, mu = 8, alternative = "less")
```

---

Groupabc

*Group data to illustrate analysis of variance*

---

### Description

Data for Exercise 10.42

### Usage

Groupabc

### Format

A data frame/tibble with 45 observations on two variables

**group** a factor with levels A, B, and C

**response** a numeric vector

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
boxplot(response ~ group, data = Groupabc,  
        col = c("red", "blue", "green"))  
anova(lm(response ~ group, data = Groupabc))
```

---

Groups

*An illustration of analysis of variance*

---

### Description

Data for Exercise 10.4

### Usage

Groups

### Format

A data frame/tibble with 78 observations on two variables

**group** a factor with levels A, B, and C

**response** a numeric vector

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(response ~ group, data = Groups, col = c("red", "blue", "green"))
anova(lm(response ~ group, data = Groups))
```

---

Gym

*Children's age versus number of completed gymnastic activities*

---

**Description**

Data for Exercises 2.21 and 9.14

**Usage**

Gym

**Format**

A data frame/tibble with eight observations on three variables

**age** age of child

**number** number of gymnastic activities successfully completed

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(number ~ age, data = Gym)
model <- lm(number ~ age, data = Gym)
abline(model, col = "red")
summary(model)
```

---

Habits

*Study habits of students in two matched school districts*

---

### Description

Data for Exercise 7.57

### Usage

Habits

### Format

A data frame/tibble with 11 observations on four variables

**A** study habit score

**B** study habit score

**differ** B minus A

**signrks** the signed-ranked-differences

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
shapiro.test(Habits$differ)
qqnorm(Habits$differ)
qqline(Habits$differ)
wilcox.test(Habits$B, Habits$A, paired = TRUE, alternative = "less")
t.test(Habits$signrks, alternative = "less")

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Habits, aes(x = differ)) +
  geom_dotplot(fill = "blue") +
  theme_bw()

## End(Not run)
```

---

Haptoglo

*Haptoglobin concentration in blood serum of 8 healthy adults*

---

**Description**

Data for Example 6.9

**Usage**

Haptoglo

**Format**

A data frame/tibble with eight observations on one variable

**concent** haptoglobin concentration (in grams per liter)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
shapiro.test(Haptoglo$concent)
t.test(Haptoglo$concent, mu = 2, alternative = "less")
```

---

Hardware

*Daily receipts for a small hardware store for 31 working days*

---

**Description**

Daily receipts for a small hardware store for 31 working days

**Usage**

Hardware

**Format**

A data frame with 31 observations on one variable

**receipt** a numeric vector of daily receipts (in dollars)

**Source**

J.C. Miller and J.N. Miller, (1988), *Statistics for Analytical Chemistry*, 2nd Ed. (New York: Halsted Press).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Hardware$receipt)
```

---

Hardwood

*Tensile strength of Kraft paper for different percentages of hardwood in the batches of pulp*

---

**Description**

Data for Example 2.18 and Exercise 9.34

**Usage**

Hardwood

**Format**

A data frame/tibble with 19 observations on two variables

**tensile** tensile strength of kraft paper (in pounds per square inch)

**hardwood** percent of hardwood in the batch of pulp that was used to produce the paper

**Source**

G. Joglekar, et al., "Lack-of-Fit Testing When Replicates Are Not Available," *The American Statistician*, 43(3), (1989), 135-143.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(tensile ~ hardwood, data = Hardwood)
model <- lm(tensile ~ hardwood, data = Hardwood)
abline(model, col = "red")
plot(model, which = 1)
```

---

Heat	<i>Primary heating sources of homes on indian reservations versus all households</i>
------	--

---

## Description

Data for Exercise 1.29

## Usage

Heat

## Format

A data frame/tibble with 301 observations on two variables

**fuel** a factor with levels Utility gas, LP bottled gas, Electricity, Fuel oil, Wood, and Other

**location** a factor with levels American Indians on reservation, All U.S. households, and American Indians not on reservations

## Source

Bureau of the Census, *Housing of the American Indians on Reservations*, Statistical Brief 95-11, April 1995.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
T1 <- xtabs(~ fuel + location, data = Heat)
T1
barplot(t(T1), beside = TRUE, legend = TRUE)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Heat, aes(x = fuel, fill = location)) +
```

```
geom_bar(position = "dodge") +  
labs(y = "percent") +  
theme_bw() +  
theme(axis.text.x = element_text(angle = 30, hjust = 1))  
  
## End(Not run)
```

---

Heating

*Fuel efficiency ratings for three types of oil heaters*

---

### Description

Data for Exercise 10.32

### Usage

Heating

### Format

A data frame/tibble with 90 observations on the two variables

**type** a factor with levels A, B, and C denoting the type of oil heater

**efficiency** heater efficiency rating

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
boxplot(efficiency ~ type, data = Heating,  
        col = c("red", "blue", "green"))  
kruskal.test(efficiency ~ type, data = Heating)
```



---

Hodgkin

*Results of treatments for Hodgkin's disease*

---

### Description

Data for Exercise 2.77

### Usage

Hodgkin

### Format

A data frame/tibble with 538 observations on two variables

**type** a factor with levels LD, LP, MC, and NS

**response** a factor with levels Positive, Partial, and None

### Source

I. Dunsmore, F. Daly, *Statistical Methods, Unit 9, Categorical Data*, Milton Keynes, The Open University, 18.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
T1 <- xtabs(~type + response, data = Hodgkin)
T1
barplot(t(T1), legend = TRUE, beside = TRUE)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Hodgkin, aes(x = type, fill = response)) +
  geom_bar(position = "dodge") +
  theme_bw()

## End(Not run)
```

---

Homes	<i>Median prices of single-family homes in 65 metropolitan statistical areas</i>
-------	--

---

### Description

Data for Statistical Insight Chapter 5

### Usage

Homes

### Format

A data frame/tibble with 65 observations on the four variables

**city** a character variable with values Akron OH, Albuquerque NM, Anaheim CA, Atlanta GA, Baltimore MD, Baton Rouge LA, Birmingham AL, Boston MA, Bradenton FL, Buffalo NY, Charleston SC, Chicago IL, Cincinnati OH, Cleveland OH, Columbia SC, Columbus OH, Corpus Christi TX, Dallas TX, Daytona Beach FL, Denver CO, Des Moines IA, Detroit MI, El Paso TX, Grand Rapids MI, Hartford CT, Honolulu HI, Houston TX, Indianapolis IN, Jacksonville FL, Kansas City MO, Knoxville TN, Las Vegas NV, Los Angeles CA, Louisville KY, Madison WI, Memphis TN, Miami FL, Milwaukee WI, Minneapolis MN, Mobile AL, Nashville TN, New Haven CT, New Orleans LA, New York NY, Oklahoma City OK, Omaha NE, Orlando FL, Philadelphia PA, Phoenix AZ, Pittsburgh PA, Portland OR, Providence RI, Sacramento CA, Salt Lake City UT, San Antonio TX, San Diego CA, San Francisco CA, Seattle WA, Spokane WA, St Louis MO, Syracuse NY, Tampa FL, Toledo OH, Tulsa OK, and Washington DC

**region** a character variable with values Midwest, Northeast, South, and West

**year** a factor with levels 1994 and 2000

**price** median house price (in dollars)

### Source

National Association of Realtors.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
tapply(Homes$price, Homes$year, mean)
tapply(Homes$price, Homes$region, mean)
p2000 <- subset(Homes, year == "2000")
p1994 <- subset(Homes, year == "1994")
## Not run:
```

```
library(dplyr)
library(ggplot2)
dplyr::group_by(Homes, year, region) %>%
  summarize(AvgPrice = mean(price))
ggplot2::ggplot(data = Homes, aes(x = region, y = price)) +
  geom_boxplot() +
  theme_bw() +
  facet_grid(year ~ .)

## End(Not run)
```

---

Homework	<i>Number of hours per week spent on homework for private and public high school students</i>
----------	---

---

### Description

Data for Exercise 7.78

### Usage

Homework

### Format

A data frame with 30 observations on two variables

**school** type of school either private or public

**time** number of hours per week spent on homework

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
boxplot(time ~ school, data = Homework,
        ylab = "Hours per week spent on homework")
#
t.test(time ~ school, data = Homework)
```

---

Honda	<i>Miles per gallon for a Honda Civic on 35 different occasions</i>
-------	---

---

**Description**

Data for Statistical Insight Chapter 6

**Usage**

Honda

**Format**

A data frame/tibble with 35 observations on one variable

**mileage** miles per gallon for a Honda Civic

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
t.test(Honda$mileage, mu = 40, alternative = "less")
```

---

Hostile	<i>Hostility levels of high school students from rural, suburban, and urban areas</i>
---------	---

---

**Description**

Data for Example 10.6

**Usage**

Hostile

**Format**

A data frame/tibble with 135 observations on two variables

**location** a factor with the location of the high school student (Rural, Suburban, or Urban)

**hostility** the score from the Hostility Level Test

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(hostility ~ location, data = Hostile,  
        col = c("red", "blue", "green"))  
kruskal.test(hostility ~ location, data = Hostile)
```

---

Housing

*Median home prices for 1984 and 1993 in 37 markets across the U.S.*

---

**Description**

Data for Exercise 5.82

**Usage**

Housing

**Format**

A data frame/tibble with 74 observations on three variables

**city** a character variable with values Albany, Anaheim, Atlanta, Baltimore, Birmingham, Boston, Chicago, Cincinnati, Cleveland, Columbus, Dallas, Denver, Detroit, Ft Lauderdale, Houston, Indianapolis, Kansas City, Los Angeles, Louisville, Memphis, Miami, Milwaukee, Minneapolis, Nashville, New York, Oklahoma City, Philadelphia, Providence, Rochester, Salt Lake City, San Antonio, San Diego, San Francisco, San Jose, St Louis, Tampa, and Washington

**year** a factor with levels 1984 and 1993

**price** median house price (in dollars)

**Source**

National Association of Realtors.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
stripchart(price ~ year, data = Housing, method = "stack",
           pch = 1, col = c("red", "blue"))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Housing, aes(x = price, fill = year)) +
  geom_dotplot() +
  facet_grid(year ~ .) +
  theme_bw()

## End(Not run)
```

---

Hurrican	<i>Number of storms, hurricanes and El Nino effects from 1950 through 1995</i>
----------	--

---

## Description

Data for Exercises 1.38, 10.19, and Example 1.6

## Usage

Hurrican

## Format

A data frame/tibble with 46 observations on four variables

**year** a numeric vector indicating year

**storms** a numeric vector recording number of storms

**hurricane** a numeric vector recording number of hurricanes

**elnino** a factor with levels cold, neutral, and warm

## Source

National Hurricane Center.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~hurrican, data = Hurrican)
T1
barplot(T1, col = "blue", main = "Problem 1.38",
        xlab = "Number of hurricanes",
        ylab = "Number of seasons")
boxplot(storms ~ elnino, data = Hurrican,
        col = c("blue", "yellow", "red"))
anova(lm(storms ~ elnino, data = Hurrican))
rm(T1)
```

---

Iceberg	<i>Number of icebergs sighted each month south of Newfoundland and south of the Grand Banks in 1920</i>
---------	---

---

**Description**

Data for Exercise 2.46 and 2.60

**Usage**

Iceberg

**Format**

A data frame with 12 observations on three variables

**month** a character variable with abbreviated months of the year

**Newfoundland** number of icebergs sighted south of Newfoundland

**Grand Banks** number of icebergs sighted south of Grand Banks

**Source**

N. Shaw, *Manual of Meteorology*, Vol. 2 (London: Cambridge University Press 1942), 7; and F. Mosteller and J. Tukey, *Data Analysis and Regression* (Reading, MA: Addison - Wesley, 1977).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(Newfoundland ~ `Grand Banks`, data = Iceberg)
abline(lm(Newfoundland ~ `Grand Banks`, data = Iceberg), col = "blue")
```

Income

*Percent change in personal income from 1st to 2nd quarter in 2000***Description**

Data for Exercise 1.33

**Usage**

Income

**Format**

A data frame/tibble with 51 observations on two variables

**state** a character variable with values Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, and Wyoming

**percent\_change** percent change in income from first quarter to the second quarter of 2000

**Source**

US Department of Commerce.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
Income$class <- cut(Income$percent_change,
                    breaks = c(-Inf, 0.5, 1.0, 1.5, 2.0, Inf))
T1 <- xtabs(~class, data = Income)
T1
barplot(T1, col = "pink")
## Not run:
library(ggplot2)
DF <- as.data.frame(T1)
DF
ggplot2::ggplot(data = DF, aes(x = class, y = Freq)) +
  geom_bar(stat = "identity", fill = "purple") +
  theme_bw()
```



```
## End(Not run)
```

---

Independent

*Illustrates a comparison problem for long-tailed distributions*

---

### Description

Data for Exercise 7.41

### Usage

Independent

### Format

A data frame/tibble with 46 observations on two variables

**score** a numeric vector

**group** a factor with levels A and B

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
qqnorm(Independent$score[Independent$group=="A"])
qqline(Independent$score[Independent$group=="A"])
qqnorm(Independent$score[Independent$group=="B"])
qqline(Independent$score[Independent$group=="B"])
boxplot(score ~ group, data = Independent, col = "blue")
wilcox.test(score ~ group, data = Independent)
```

---

Indian	<i>Educational attainment versus per capita income and poverty rate for American Indians living on reservations</i>
--------	---

---

## Description

Data for Exercise 2.95

## Usage

Indian

## Format

A data frame/tibble with ten observations on four variables

**reservation** a character variable with values Blackfeet, Fort Apache, Gila River, Hopi, Navajo, Papago, Pine Ridge, Rosebud, San Carlos, and Zuni Pueblo

**percent high school** percent who have graduated from high school

**per capita income** per capita income (in dollars)

**poverty rate** percent poverty

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
par(mfrow = c(1, 2))
plot(`per capita income` ~ `percent high school`, data = Indian,
     xlab = "Percent high school graduates", ylab = "Per capita income")
plot(`poverty rate` ~ `percent high school`, data = Indian,
     xlab = "Percent high school graduates", ylab = "Percent poverty")
par(mfrow = c(1, 1))
```

---

Indiapol	<i>Average miles per hour for the winners of the Indianapolis 500 race</i>
----------	--

---

**Description**

Data for Exercise 1.128

**Usage**

Indiapol

**Format**

A data frame/tibble with 39 observations on two variables

**year** the year of the race

**speed** the winners average speed (in mph)

**Source**

The World Almanac and Book of Facts, 2000, p. 1004.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(speed ~ year, data = Indiapol, type = "b")
```

---

Indy500	<i>Qualifying miles per hour and number of previous starts for drivers in 79th Indianapolis 500 race</i>
---------	--

---

**Description**

Data for Exercises 7.11 and 7.36

**Usage**

Indy500

**Format**

A data frame/tibble with 33 observations on four variables

**driver** a character variable with values andretti, bachelart, boesel, brayton, c.guerrero, cheever, fabi, fernandez, ferran, fittipaldi, fox, goodyear, gordon, gugelmin, herta, james, johansson, jones, lazier, luyendyk, matsuda, matsushita, pruet, r.guerrero, rahal, ribeiro, salazar, sharp, sullivan, tracy, vasser, villeneuve, and zampedri

**qualif** qualifying speed (in mph)

**starts** number of Indianapolis 500 starts

**group** a numeric vector where 1 indicates the driver has 4 or fewer Indianapolis 500 starts and a 2 for drivers with 5 or more Indianapolis 500 starts

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stripchart(qualif ~ group, data = Indy500, method = "stack",
           pch = 19, col = c("red", "blue"))
boxplot(qualif ~ group, data = Indy500)
t.test(qualif ~ group, data = Indy500)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Indy500, aes(sample = qualif)) +
  geom_qq() +
  facet_grid(group ~ .) +
  theme_bw()

## End(Not run)
```

---

 Inflatio

---

*Private pay increase of salaried employees versus inflation rate*


---

**Description**

Data for Exercises 2.12 and 2.29

**Usage**

Inflatio

**Format**

A data frame/tibble with 24 observations on four variables

**year** a numeric vector of years

**pay** average hourly wage for salaried employees (in dollars)

**increase** percent increase in hourly wage over previous year

**inflation** percent inflation rate

**Source**

Bureau of Labor Statistics.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(increase ~ inflation, data = Inflatio)
cor(Inflatio$increase, Inflatio$inflation, use = "complete.obs")
```

---

Inletoil

*Inlet oil temperature through a valve*

---

**Description**

Data for Exercises 5.91 and 6.48

**Usage**

Inletoil

**Format**

A data frame/tibble with 12 observations on one variable

**temp** inlet oil temperature (Fahrenheit)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Inletoil$temp, breaks = 3)
qqnorm(Inletoil$temp)
qqline(Inletoil$temp)
t.test(Inletoil$temp)
t.test(Inletoil$temp, mu = 98, alternative = "less")
```

---

Inmate

*Type of drug offense by race*

---

**Description**

Data for Statistical Insight Chapter 8

**Usage**

Inmate

**Format**

A data frame/tibble with 28,047 observations on two variables

**race** a factor with levels white, black, and hispanic

**drug** a factor with levels heroin, crack, cocaine, and marijuana

**Source**

C. Wolf Harlow (1994), *Comparing Federal and State Prison Inmates*, NCJ-145864, U.S. Department of Justice, Bureau of Justice Statistics.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~race + drug, data = Inmate)
T1
chisq.test(T1)
rm(T1)
```

---

Inspect	<i>Percent of vehicles passing inspection by type inspection station</i>
---------	--

---

**Description**

Data for Exercise 8.59

**Usage**

Inspect

**Format**

A data frame/tibble with 174 observations on two variables

**station** a factor with levels auto inspection, auto repair, car care center, gas station, new car dealer, and tire store

**passed** a factor with levels less than 70%, between 70% and 84%, and more than 85%

**Source**

*The Charlotte Observer*, December 13, 1992.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~ station + passed, data = Inspect)
T1
barplot(T1, beside = TRUE, legend = TRUE)
chisq.test(T1)
rm(T1)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Inspect, aes(x = passed, fill = station)) +
  geom_bar(position = "dodge") +
  theme_bw()

## End(Not run)
```

---

Insulate

*Heat loss through a new insulating medium*

---

### Description

Data for Exercise 9.50

### Usage

Insulate

### Format

A data frame/tibble with ten observations on two variables

**temp** outside temperature (in degrees Celcius)

**loss** heat loss (in BTUs)

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
plot(loss ~ temp, data = Insulate)
model <- lm(loss ~ temp, data = Insulate)
abline(model, col = "blue")
summary(model)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Insulate, aes(x = temp, y = loss)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw()

## End(Not run)
```



---

Iqgpa

*GPA versus IQ for 12 individuals*

---

**Description**

Data for Exercises 9.51 and 9.52

**Usage**

Iqgpa

**Format**

A data frame/tibble with 12 observations on two variables

**iq** IQ scores

**gpa** Grade point average

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(gpa ~ iq, data = Iqgpa, col = "blue", pch = 19)
model <- lm(gpa ~ iq, data = Iqgpa)
summary(model)
rm(model)
```

---

Irises

*R.A. Fishers famous data on Irises*

---

**Description**

Data for Examples 1.15 and 5.19

**Usage**

Irises

**Format**

A data frame/tibble with 150 observations on five variables

**sepal\_length** sepal length (in cm)

**sepal\_width** sepal width (in cm)

**petal\_length** petal length (in cm)

**petal\_width** petal width (in cm)

**species** a factor with levels setosa, versicolor, and virginica

**Source**

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II, 179-188.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
tapply(Iris$sepal_length, Iris$species, mean)
t.test(Iris$sepal_length[Iris$species == "setosa"], conf.level = 0.99)
hist(Iris$sepal_length[Iris$species == "setosa"],
     main = "Sepal length for\n Iris Setosa",
     xlab = "Length (in cm)")
boxplot(sepal_length ~ species, data = Iris)
```

---

Jdpower

*Number of problems reported per 100 cars in 1994 versus 1995s*

---

**Description**

Data for Exercise 2.14, 2.17, 2.31, 2.33, and 2.40

**Usage**

Jdpower

**Format**

A data frame/tibble with 29 observations on three variables

**car** a factor with levels Acura, BMW, Buick, Cadillac, Chevrolet, Dodge Eagle, Ford, Geo, Honda, Hyundai, Infiniti, Jaguar, Lexus, Lincoln, Mazda, Mercedes-Benz, Mercury, Mitsubishi, Nissan, Oldsmobile, Plymouth, Pontiac, Saab, Saturn, and Subaru, Toyota Volkswagen, Volvo

**1994** number of problems per 100 cars in 1994

**1995** number of problems per 100 cars in 1995

**Source**

*USA Today*, May 25, 1995.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
model <- lm(`1995` ~ `1994`, data = Jdpower)
summary(model)
plot(`1995` ~ `1994`, data = Jdpower)
abline(model, col = "red")
rm(model)
```

---

Jobsat

*Job satisfaction and stress level for 9 school teachers*

---

**Description**

Data for Exercise 9.60

**Usage**

Jobsat

**Format**

A data frame/tibble with nine observations on two variables

**wspt** Wilson Stress Profile score for teachers

**satisfaction** job satisfaction score

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(satisfaction ~ wspt, data = Jobsat)
model <- lm(satisfaction ~ wspt, data = Jobsat)
abline(model, col = "blue")
summary(model)
rm(model)
```

---

Kidsmoke

*Smoking habits of boys and girls ages 12 to 18*

---

## Description

Data for Exercise 4.85

## Usage

Kidsmoke

## Format

A data frame/tibble with 1000 observations on two variables

**gender** character vector with values female and male

**smoke** a character vector with values no and yes

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
T1 <- xtabs(~smoke + gender, data = Kidsmoke)
T1
prop.table(T1)
prop.table(T1, 1)
prop.table(T1, 2)
```

---

 Kilowatt

*Rates per kilowatt-hour for each of the 50 states and DC*


---

**Description**

Data for Example 5.9

**Usage**

Kilowatt

**Format**

A data frame/tibble with 51 observations on two variables

**state** a factor with levels Alabama Alaska, Arizona, Arkansas California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa Kansas Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia Washington, West Virginia, Wisconsin, and Wyoming

**rate** a numeric vector indicating rates for kilowatt per hour

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Kilowatt$rate)
```

---

 Kinder

*Reading scores for first grade children who attended kindergarten versus those who did not*


---

**Description**

Data for Exercise 7.68

**Usage**

Kinder

**Format**

A data frame/tibble with eight observations on three variables

**pair** a numeric indicator of pair

**kinder** reading score of kids who went to kindergarten

**nokinder** reading score of kids who did not go to kindergarten

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(Kinder$kinder, Kinder$nokinder)
diff <- Kinder$kinder - Kinder$nokinder
qqnorm(diff)
qqline(diff)
shapiro.test(diff)
t.test(Kinder$kinder, Kinder$nokinder, paired = TRUE)
# Or
t.test(diff)
rm(diff)
```

---

Laminect

*Median costs of laminectomies at hospitals across North Carolina in 1992*

---

**Description**

Data for Exercise 10.18

**Usage**

Laminect

**Format**

A data frame/tibble with 138 observations on two variables

**area** a character vector indicating the area of the hospital with Rural, Regional, and Metropol

**cost** a numeric vector indicating cost of a laminectomy

**Source**

*Consumer's Guide to Hospitalization Charges in North Carolina Hospitals* (August 1994), North Carolina Medical Database Commission, Department of Insurance.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(cost ~ area, data = Laminect, col = topo.colors(3))
anova(lm(cost ~ area, data = Laminect))
```

---

Lead	<i>Lead levels in children's blood whose parents worked in a battery factory</i>
------	--

---

**Description**

Data for Example 1.17

**Usage**

Lead

**Format**

A data frame/tibble with 66 observations on the two variables

**group** a character vector with values exposed and control

**lead** a numeric vector indicating the level of lead in children's blood (in micrograms/dl)

**Source**

Morton, D. et al. (1982), "Lead Absorption in Children of Employees in a Lead-Related Industry," *American Journal of Epidemiology*, 115, 549-555.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(lead ~ group, data = Lead, col = topo.colors(2))
```

---

Leader	<i>Leadership exam scores by age for employees on an industrial plant</i>
--------	---

---

**Description**

Data for Exercise 7.31

**Usage**

Leader

**Format**

A data frame/tibble with 34 observations on two variables

**age** a character vector indicating age with values under35 and over35

**score** score on a leadership exam

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(score ~ age, data = Leader, col = c("gray", "green"))
t.test(score ~ age, data = Leader)
```

---

Lethal	<i>Survival time of mice injected with an experimental lethal drug</i>
--------	--

---

**Description**

Data for Example 6.12

**Usage**

Lethal

**Format**

A data frame/tibble with 30 observations on one variable

**survival** a numeric vector indicating time survived after injection (in seconds)



**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
SIGN.test(Lethal$survival, md = 45, alternative = "less")
```

---

Life

*Life expectancy of men and women in U.S.*

---

**Description**

Data for Exercise 1.31

**Usage**

Life

**Format**

A data frame/tibble with eight observations on three variables

**year** a numeric vector indicating year

**men** life expectancy for men (in years)

**women** life expectancy for women (in years)

**Source**

National Center for Health Statistics.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(men ~ year, type = "l", ylim = c(min(men, women), max(men, women)),
     col = "blue", main = "Life Expectancy vs Year", ylab = "Age",
     xlab = "Year", data = Life)
lines(women ~ year, col = "red", data = Life)
text(1955, 65, "Men", col = "blue")
text(1955, 70, "Women", col = "red")
```

Lifespan

*Life span of electronic components used in a spacecraft versus heat*

---

**Description**

Data for Exercise 2.4, 2.37, and 2.49

**Usage**

Lifespan

**Format**

A data frame/tibble with six observations two variables

**heat** temperature (in Celcius)

**life** lifespan of component (in hours)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(life ~ heat, data = Lifespan)
model <- lm(life ~ heat, data = Lifespan)
abline(model, col = "red")
resid(model)
sum((resid(model))^2)
anova(model)
rm(model)
```

---

Ligntmonth*Relationship between damage reports and deaths caused by lightning*

---

**Description**

Data for Exercise 2.6

**Usage**

Ligntmonth

**Format**

A data frame/tibble with 12 observations on four variables

**month** a factor with levels 1/01/2000, 10/01/2000, 11/01/2000, 12/01/2000, 2/01/2000, 3/01/2000, 4/01/2000, 5/01/2000, 6/01/2000, 7/01/2000, 8/01/2000, and 9/01/2000

**deaths** number of deaths due to lightning strikes

**injuries** number of injuries due to lightning strikes

**damage** damage due to lightning strikes (in dollars)

**Source**

*Lighting Fatalities, Injuries and Damage Reports in the United States, 1959-1994*, NOAA Technical Memorandum NWS SR-193, Dept. of Commerce.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(deaths ~ damage, data = Lightmonth)
model = lm(deaths ~ damage, data = Lightmonth)
abline(model, col = "red")
rm(model)
```

---

Lodge

*Measured traffic at three prospective locations for a motor lodge*

---

**Description**

Data for Exercise 10.33

**Usage**

Lodge

**Format**

A data frame/tibble with 45 observations on six variables

**traffic** a numeric vector indicating the amount of vehicles that passed a site in 1 hour

**site** a numeric vector with values 1, 2, and 3

**ranks** ranks for variable traffic

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(traffic ~ site, data = Lodge, col = cm.colors(3))
anova(lm(traffic ~ factor(site), data = Lodge))
```

---

Longtail

*Long-tailed distributions to illustrate Kruskal Wallis test*

---

## Description

Data for Exercise 10.45

## Usage

Longtail

## Format

A data frame/tibble with 60 observations on three variables

**score** a numeric vector

**group** a numeric vector with values 1, 2, and 3

**ranks** ranks for variable score

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(score ~ group, data = Longtail, col = heat.colors(3))
kruskal.test(score ~ factor(group), data = Longtail)
anova(lm(score ~ factor(group), data = Longtail))
```

---

Lowabil	<i>Reading skills of 24 matched low ability students</i>
---------	--

---

**Description**

Data for Example 7.18

**Usage**

Lowabil

**Format**

A data frame/tibble with 12 observations on three variables

**pair** a numeric indicator of pair

**experiment** score of the child with the experimental method

**control** score of the child with the standard method

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
diff = Lowabil$experiment - Lowabil$control
qqnorm(diff)
qqline(diff)
shapiro.test(diff)
t.test(Lowabil$experiment, Lowabil$control, paired = TRUE)
# OR
t.test(diff)
rm(diff)
```

---

Magnesi	<i>Magnesium concentration and distances between samples</i>
---------	--

---

**Description**

Data for Exercise 9.9

**Usage**

Magnesi

**Format**

A data frame/tibble with 20 observations on two variables

**distance** distance between samples

**magnesium** concentration of magnesium

**Source**

Davis, J. (1986), *Statistics and Data Analysis in Geology*, 2d. Ed., John Wiley and Sons, New York, p. 146.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(magnesium ~ distance, data = Magnesium)
model = lm(magnesium ~ distance, data = Magnesium)
abline(model, col = "red")
summary(model)
rm(model)
```

---

Malpract

*Amounts awarded in 17 malpractice cases*

---

**Description**

Data for Exercise 5.73

**Usage**

Malpract

**Format**

A data frame/tibble with 17 observations on one variable

**award** malpractice reward (in \$1000)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
SIGN.test(Malpract$award, conf.level = 0.90)
```

---

Manager	<i>Advertised salaries offered general managers of major corporations in 1995</i>
---------	---

---

**Description**

Data for Exercise 5.81

**Usage**

Manager

**Format**

A data frame/tibble with 26 observations on one variable

**salary** random sample of advertised annual salaries of top executives (in dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Manager$salary)
SIGN.test(Manager$salary)
```

---

Marked	<i>Percent of marked cars in 65 police departments in Florida</i>
--------	---

---

**Description**

Data for Exercise 6.100

**Usage**

Marked

**Format**

A data frame/tibble with 65 observations on one variable

**percent** percentage of marked cars in 65 Florida police departments

**Source**

*Law Enforcement Management and Administrative Statistics, 1993*, Bureau of Justice Statistics, NCJ-148825, September 1995, p. 147-148.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Marked$percent)
SIGN.test(Marked$percent, md = 60, alternative = "greater")
t.test(Marked$percent, mu = 60, alternative = "greater")
```

---

Math

*Standardized math test scores for 30 students*

---

**Description**

Data for Exercise 1.69

**Usage**

Math

**Format**

A data frame/tibble with 30 observations on one variable

**score** scores on a standardized test for 30 tenth graders

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.



**Examples**

```
stem(Math$score)
hist(Math$score, main = "Math Scores", xlab = "score", freq = FALSE)
lines(density(Math$score), col = "red")
CharlieZ <- (62 - mean(Math$score))/sd(Math$score)
CharlieZ
scale(Math$score)[which(Math$score == 62)]
```

---

Mathcomp

*Standardized math competency for a group of entering freshmen at a small community college*

---

**Description**

Data for Exercise 5.26

**Usage**

Mathcomp

**Format**

A data frame/tibble with 31 observations one variable

**score** scores of 31 entering freshmen at a community college on a national standardized test

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Mathcomp$score)
EDA(Mathcomp$score)
```

---

Mathpro

*Math proficiency and SAT scores by states*

---

### Description

Data for Exercise 9.24, Example 9.1, and Example 9.6

### Usage

Mathpro

### Format

A data frame/tibble with 51 observations on four variables

**state** a factor with levels Conn, D.C., Del, Ga, Hawaii, Ind, Maine, Mass, Md, N.C., N.H., N.J., N.Y., Ore, Pa, R.I., S.C., Va, and Vt

**sat\_math** SAT math scores for high school seniors

**profic** math proficiency scores for eighth graders

**group** a numeric vector

### Source

National Assessment of Educational Progress and The College Board.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
model <- lm(sat_math ~ profic, data = Mathpro)
plot(sat_math ~ profic, data = Mathpro, ylab = "SAT", xlab = "proficiency")
abline(model, col = "red")
summary(model)
rm(model)
```

---

Maze	<i>Error scores for four groups of experimental animals running a maze</i>
------	--

---

**Description**

Data for Exercise 10.13

**Usage**

Maze

**Format**

A data frame/tibble with 32 observations on two variables

**score** error scores for animals running through a maze under different conditions

**condition** a factor with levels CondA, CondB, CondC, and CondD

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(score ~ condition, data = Maze, col = rainbow(4))
anova(lm(score ~ condition, data = Maze))
```

---

Median	<i>Illustrates test of equality of medians with the Kruskal Wallis test</i>
--------	---

---

**Description**

Data for Exercise 10.52

**Usage**

Median

**Format**

A data frame/tibble with 45 observations on two variables

**sample** a vector with values Sample1, Sample 2, and Sample 3

**value** a numeric vector

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(value ~ sample, data = Median, col = rainbow(3))
anova(lm(value ~ sample, data = Median))
kruskal.test(value ~ factor(sample), data = Median)
```

---

Mental

*Median mental ages of 16 girls*

---

**Description**

Data for Exercise 6.52

**Usage**

Mental

**Format**

A data frame/tibble with 16 observations on one variable

**age** mental age of 16 girls

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
SIGN.test(Mental$age, md = 100)
```

---

Mercury	<i>Concentration of mercury in 25 lake trout</i>
---------	--

---

**Description**

Data for Example 1.9

**Usage**

Mercury

**Format**

A data frame/tibble with 25 observations on one variable

**mercury** a numeric vector measuring mercury (in parts per million)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Mercury$mercury)
```

---

Metrent	<i>Monthly rental costs in metro areas with 1 million or more persons</i>
---------	---

---

**Description**

Data for Exercise 5.117

**Usage**

Metrent

**Format**

A data frame/tibble with 46 observations on one variable

**rent** monthly rent in dollars

**Source**

U.S. Bureau of the Census, *Housing in the Metropolitan Areas, Statistical Brief SB/94/19*, September 1994.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(Metrent$rent, col = "magenta")
t.test(Metrent$rent, conf.level = 0.99)$conf
```

---

Miller	<i>Miller personality test scores for a group of college students applying for graduate school</i>
--------	--

---

**Description**

Data for Example 5.7

**Usage**

Miller

**Format**

A data frame/tibble with 25 observations on one variable

**miller** scores on the Miller Personality test

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Miller$miller)
fivenum(Miller$miller)
boxplot(Miller$miller)
qqnorm(Miller$miller, col = "blue")
qqline(Miller$miller, col = "red")
```

---

Miller1	<i>Twenty scores on the Miller personality test</i>
---------	---

---

**Description**

Data for Exercise 1.41

**Usage**

Miller1

**Format**

A data frame/tibble with 20 observations on one variable

**miller** scores on the Miller personality test

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Miller1$miller)
stem(Miller1$miller, scale = 2)
```

---

Moisture	<i>Moisture content and depth of core sample for marine muds in eastern Louisiana</i>
----------	---

---

**Description**

Data for Exercise 9.32

**Usage**

Moisture

**Format**

A data frame/tibble with 16 observations on four variables

**depth** a numeric vector

**moisture** g of water per 100 g of dried sediment

**Inmoist** a numeric vector

**depthsq** a numeric vector

**Source**

Davis, J. C. (1986), *Statistics and Data Analysis in Geology*, 2d. ed., John Wiley and Sons, New York, pp. 177, 185.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(moisture ~ depth, data = Moisture)
model <- lm(moisture ~ depth, data = Moisture)
abline(model, col = "red")
plot(resid(model) ~ depth, data = Moisture)
rm(model)
```

---

Monoxide

*Carbon monoxide emitted by smoke stacks of a manufacturer and a competitor*

---

**Description**

Data for Exercise 7.45

**Usage**

Monoxide

**Format**

A data frame/tibble with ten observations on two variables

**company** a vector with values manufacturer and competitor

**emission** carbon monoxide emitted

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.



**Examples**

```

boxplot(emission ~ company, data = Monoxide, col = topo.colors(2))
t.test(emission ~ company, data = Monoxide)
wilcox.test(emission ~ company, data = Monoxide)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Monoxide, aes(x = company, y = emission)) +
  geom_boxplot() +
  theme_bw()

## End(Not run)

```

---

 Movie

---

*Moral attitude scale on 15 subjects before and after viewing a movie*


---

**Description**

Data for Exercise 7.53

**Usage**

Movie

**Format**

A data frame/tibble with 12 observations on three variables

**before** moral aptitude before viewing the movie

**after** moral aptitude after viewing the movie

**differ** a numeric vector

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

qqnorm(Movie$differ)
qqline(Movie$differ)
shapiro.test(Movie$differ)
t.test(Movie$after, Movie$before, paired = TRUE, conf.level = 0.99)
wilcox.test(Movie$after, Movie$before, paired = TRUE)

```

---

Music	<i>Improvement scores for identical twins taught music recognition by two techniques</i>
-------	--

---

### Description

Data for Exercise 7.59

### Usage

Music

### Format

A data frame/tibble with 12 observations on three variables

**method1** a numeric vector measuring the improvement scores on a music recognition test

**method2** a numeric vector measuring the improvement scores on a music recognition test

**differ** method1 - method2

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
qqnorm(Music$differ)
qqline(Music$differ)
shapiro.test(Music$differ)
t.test(Music$method1, Music$method2, paired = TRUE)
# Or
t.test(Music$differ)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Music, aes(x = differ)) +
  geom_dotplot() +
  theme_bw()

## End(Not run)
```

---

Name	<i>Estimated value of a brand name product and the company's revenue</i>
------	--

---

**Description**

Data for Exercises 2.28, 9.19, and Example 2.8

**Usage**

Name

**Format**

A data frame/tibble with 42 observations on three variables

**brand** a factor with levels Band-Aid, Barbie, Birds Eye, Budweiser, Camel, Campbell, Carlsberg, Coca-Cola, Colgate, Del Monte, Fisher-Price, Gordon's, Green Giant, Guinness, Haagen-Dazs, Heineken, Heinz, Hennessy, Hermes, Hershey, Ivory, Jell-o, Johnnie Walker, Kellogg, Kleenex, Kraft, Louis Vuitton, Marlboro, Nescafe, Nestle, Nivea, Oil of Olay, Pampers, Pepsi-Cola, Planters, Quaker, Sara Lee, Schweppes, Smirnoff, Tampax, Winston, and Wrigley's

**value** value in billions of dollars

**revenue** revenue in billions of dollars

**Source**

Financial World.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(value ~ revenue, data = Name)
model <- lm(value ~ revenue, data = Name)
abline(model, col = "red")
cor(Name$value, Name$revenue)
summary(model)
rm(model)
```

---

Nascar

*Efficiency of pit crews for three major NASCAR teams*

---

**Description**

Data for Exercise 10.53

**Usage**

Nascar

**Format**

A data frame/tibble with 36 observations on six variables

**time** duration of pit stop (in seconds)

**team** a numeric vector representing team 1, 2, or 3

**ranks** a numeric vector ranking each pit stop in order of speed

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(time ~ team, data = Nascar, col = rainbow(3))
model <- lm(time ~ factor(team), data = Nascar)
summary(model)
anova(model)
rm(model)
```

---

Nervous

*Reaction effects of 4 drugs on 25 subjects with a nervous disorder*

---

**Description**

Data for Example 10.3

**Usage**

Nervous

**Format**

A data frame/tibble with 25 observations on two variables

**react** a numeric vector representing reaction time

**drug** a numeric vector indicating each of the 4 drugs

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(react ~ drug, data = Nervous, col = rainbow(4))
model <- aov(react ~ factor(drug), data = Nervous)
summary(model)
TukeyHSD(model)
plot(TukeyHSD(model), las = 1)
```

---

Newsstand

*Daily profits for 20 newsstands*

---

**Description**

Data for Exercise 1.43

**Usage**

Newsstand

**Format**

A data frame/tibble with 20 observations on one variable

**profit** profit of each newsstand (in dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Newsstand$profit)
stem(Newsstand$profit, scale = 3)
```

---

Nfldraf2	<i>Rating, time in 40-yard dash, and weight of top defensive linemen in the 1994 NFL draft</i>
----------	--

---

**Description**

Data for Exercise 9.63

**Usage**

Nfldraf2

**Format**

A data frame/tibble with 47 observations on three variables

**rating** rating of each player on a scale out of 10

**forty** forty yard dash time (in seconds)

**weight** weight of each player (in pounds)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(rating ~ forty, data = Nfldraf2)
summary(lm(rating ~ forty, data = Nfldraf2))
```

---

Nfldraft	<i>Rating, time in 40-yard dash, and weight of top offensive linemen in the 1994 NFL draft</i>
----------	--

---

**Description**

Data for Exercises 9.10 and 9.16

**Usage**

Nfldraft

**Format**

A data frame/tibble with 29 observations on three variables

**rating** rating of each player on a scale out of 10

**forty** forty yard dash time (in seconds)

**weight** weight of each player (in pounds)

**Source**

*USA Today*, April 20, 1994.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(rating ~ forty, data = Nfldraft)
cor(Nfldraft$rating, Nfldraft$forty)
summary(lm(rating ~ forty, data = Nfldraft))
```

---

Nicotine

*Nicotine content versus sales for eight major brands of cigarettes*

---

**Description**

Data for Exercise 9.21

**Usage**

Nicotine

**Format**

A data frame/tibble with eight observations on two variables

**nicotine** nicotine content (in milligrams)

**sales** sales figures (in \$100,000)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
model <- lm(sales ~ nicotine, data = Nicotine)
plot(sales ~ nicotine, data = Nicotine)
abline(model, col = "red")
summary(model)
predict(model, newdata = data.frame(nicotine = 1),
        interval = "confidence", level = 0.99)
```

---

normarea

*Normal Area*

---

## Description

Function that computes and draws the area between two user specified values in a user specified normal distribution with a given mean and standard deviation

## Usage

```
normarea(lower = -Inf, upper = Inf, m, sig)
```

## Arguments

lower	the lower value
upper	the upper value
m	the mean for the population
sig	the standard deviation of the population

## Author(s)

Alan T. Arnholt

## Examples

```
normarea(70, 130, 100, 15)
# Finds and P(70 < X < 130) given X is N(100,15).
```



---

nsize	<i>Required Sample Size</i>
-------	-----------------------------

---

**Description**

Function to determine required sample size to be within a given margin of error.

**Usage**

```
nsize(b, sigma = NULL, p = 0.5, conf.level = 0.95, type = "mu")
```

**Arguments**

b	the desired bound.
sigma	population standard deviation. Not required if using type "pi".
p	estimate for the population proportion of successes. Not required if using type "mu".
conf.level	confidence level for the problem, restricted to lie between zero and one.
type	character string, one of "mu" or "pi", or just the initial letter of each, indicating the appropriate parameter. Default value is "mu".

**Details**

Answer is based on a normal approximation when using type "pi".

**Value**

Returns required sample size.

**Author(s)**

Alan T. Arnholt

**Examples**

```
nsize(b=.03, p=708/1200, conf.level=.90, type="pi")
# Returns the required sample size (n) to estimate the population
# proportion of successes with a 0.9 confidence interval
# so that the margin of error is no more than 0.03 when the
# estimate of the population propotion of successes is 708/1200.
# This is problem 5.38 on page 257 of Kitchen's BSDA.
```

```
nsize(b=.15, sigma=.31, conf.level=.90, type="mu")
# Returns the required sample size (n) to estimate the population
# mean with a 0.9 confidence interval so that the margin
# of error is no more than 0.15. This is Example 5.17 on page
# 261 of Kitchen's BSDA.
```

---

`ntester`*Normality Tester*

---

**Description**

Q-Q plots of randomly generated normal data of the same size as the tested data are generated and plotted on the perimeter of the graph while a Q-Q plot of the actual data is depicted in the center of the graph.

**Usage**

```
ntester(actual.data)
```

**Arguments**

`actual.data` a numeric vector. Missing and infinite values are allowed, but are ignored in the calculation. The length of `actual.data` must be less than 5000 after dropping nonfinite values.

**Details**

Q-Q plots of randomly generated normal data of the same size as the tested data are generated and plotted on the perimeter of the graph sheet while a Q-Q plot of the actual data is depicted in the center of the graph. The p-values are calculated from the Shapiro-Wilk W-statistic. Function will only work on numeric vectors containing less than or equal to 5000 observations.

**Author(s)**

Alan T. Arnholt

**References**

Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52** : 591-611.

**Examples**

```
ntester(rexp(50,1))
# Q-Q plot of random exponential data in center plot
# surrounded by 8 Q-Q plots of randomly generated
# standard normal data of size 50.
```

---

Orange	<i>Price of oranges versus size of the harvest</i>
--------	--

---

**Description**

Data for Exercise 9.61

**Usage**

Orange

**Format**

A data frame/tibble with six observations on two variables

**harvest** harvest in millions of boxes

**price** average price charged by California growers for a 75-pound box of navel oranges

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(price ~ harvest, data = Orange)
model <- lm(price ~ harvest, data = Orange)
abline(model, col = "red")
summary(model)
rm(model)
```

---

Orioles	<i>Salaries of members of the Baltimore Orioles baseball team</i>
---------	---

---

**Description**

Data for Example 1.3

**Usage**

Orioles

**Format**

A data frame/tibble with 27 observations on three variables

**first name** a factor with levels Albert, Arthur, B. J., Brady, Cal, Charles, dl-Delino, dl-Scott, Doug, Harold, Heathcliff, Jeff, Jesse, Juan, Lenny, Mike, Rich, Ricky, Scott, Sidney, Will, and Willis

**last name** a factor with levels Amaral, Anderson, Baines, Belle, Bones, Bordick, Clark, Conine, Deshields, Erickson, Feters, Garcia, Guzman, Johns, Johnson, Kamieniecki, Mussina, Orosco, Otanez, Ponson, Reboulet, Rhodes, Ripken Jr., Slocumb, Surhoff, Timlin, and Webster

**1999salary** a numeric vector containing each player's salary (in dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stripchart(Orioles$`1999salary`, method = "stack", pch = 19)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Orioles, aes(x = `1999salary`)) +
  geom_dotplot(dotsize = 0.5) +
  labs(x = "1999 Salary") +
  theme_bw()

## End(Not run)
```

---

Oxytocin	<i>Arterial blood pressure of 11 subjects before and after receiving oxytocin</i>
----------	---

---

**Description**

Data for Exercise 7.86

**Usage**

Oxytocin

**Format**

A data frame/tibble with 11 observations on three variables

**subject** a numeric vector indicating each subject

**before** mean arterial blood pressure of subject before receiving oxytocin

**after** mean arterial blood pressure of subject after receiving oxytocin

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
diff = Oxytocin$after - Oxytocin$before
qqnorm(diff)
qqline(diff)
shapiro.test(diff)
t.test(Oxytocin$after, Oxytocin$before, paired = TRUE)
rm(diff)
```

---

 Parented

*Education backgrounds of parents of entering freshmen at a state university*

---

**Description**

Data for Exercise 1.32

**Usage**

Parented

**Format**

A data frame/tibble with 200 observations on two variables

**education** a factor with levels 4yr college degree, Doctoral degree, Grad degree, H.S grad or less, Some college, and Some grad school

**parent** a factor with levels mother and father

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~education + parent, data = Parented)
T1
barplot(t(T1), beside = TRUE, legend = TRUE, col = c("blue", "red"))
rm(T1)
## Not run:
library(ggplot2)
```

```

ggplot2::ggplot(data = Parented, aes(x = education, fill = parent)) +
  geom_bar(position = "dodge") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 85, vjust = 0.5)) +
  scale_fill_manual(values = c("pink", "blue")) +
  labs(x = "", y = "")

## End(Not run)

```

---

Patrol	<i>Years of experience and number of tickets given by patrolpersons in New York City</i>
--------	--

---

### Description

Data for Example 9.3

### Usage

Patrol

### Format

A data frame/tibble with ten observations on three variables

**tickets** number of tickets written per week

**years** patrolperson's experience (in years)

**log\_tickets** natural log of tickets

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```

model <- lm(tickets ~ years, data = Patrol)
summary(model)
confint(model, level = 0.98)

```

---

Pearson

*Karl Pearson's data on heights of brothers and sisters*

---

**Description**

Data for Exercise 2.20

**Usage**

Pearson

**Format**

A data frame/tibble with 11 observations on three variables

**family** number indicating family of brother and sister pair

**brother** height of brother (in inches)

**sister** height of sister (in inches)

**Source**

Pearson, K. and Lee, A. (1902-3), On the Laws of Inheritance in Man, *Biometrika*, 2, 357.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(brother ~ sister, data = Pearson, col = "lightblue")
cor(Pearson$brother, Pearson$sister)
```

---

Phone

*Length of long-distance phone calls for a small business firm*

---

**Description**

Data for Exercise 6.95

**Usage**

Phone

**Format**

A data frame/tibble with 20 observations on one variable

**time** duration of long distance phone call (in minutes)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Phone$time)
qqline(Phone$time)
shapiro.test(Phone$time)
SIGN.test(Phone$time, md = 5, alternative = "greater")
```

---

Poison

*Number of poisonings reported to 16 poison control centers*

---

**Description**

Data for Exercise 1.113

**Usage**

Poison

**Format**

A data frame/tibble with 226,361 observations on one variable

**type** a factor with levels Alcohol, Cleaning agent, Cosmetics, Drugs, Insecticides, and Plants

**Source**

Centers for Disease Control, Atlanta, Georgia.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.



**Examples**

```

T1 <- xtabs(~type, data = Poison)
T1
par(mar = c(5.1 + 2, 4.1, 4.1, 2.1))
barplot(sort(T1, decreasing = TRUE), las = 2, col = rainbow(6))
par(mar = c(5.1, 4.1, 4.1, 2.1))
rm(T1)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Poison, aes(x = type, fill = type)) +
  geom_bar() +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 85, vjust = 0.5)) +
  guides(fill = FALSE)

## End(Not run)

```

---

 Politic

*Political party and gender in a voting district*


---

**Description**

Data for Example 8.3

**Usage**

Politic

**Format**

A data frame/tibble with 250 observations on two variables

**party** a factor with levels republican, democrat, and other

**gender** a factor with levels female and male

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

T1 <- xtabs(~party + gender, data = Politic)
T1
chisq.test(T1)
rm(T1)

```

---

Pollutio	<i>Air pollution index for 15 randomly selected days for a major western city</i>
----------	---

---

**Description**

Data for Exercise 5.59

**Usage**

Pollutio

**Format**

A data frame/tibble with 15 observations on one variable

**inde** air pollution index

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Pollutio$inde)
t.test(Pollutio$inde, conf.level = 0.98)$conf
```

---

Porosity	<i>Porosity measurements on 20 samples of Tensleep Sandstone, Pennsylvanian from Bighorn Basin in Wyoming</i>
----------	---

---

**Description**

Data for Exercise 5.86

**Usage**

Porosity

**Format**

A data frame/tibble with 20 observations on one variable

**porosity** porosity measurement (percent)

**Source**

Davis, J. C. (1986), *Statistics and Data Analysis in Geology*, 2nd edition, pages 63-65.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Porosity$porosity)
fivenum(Porosity$porosity)
boxplot(Porosity$porosity, col = "lightgreen")
```

---

Poverty

*Percent poverty and crime rate for selected cities*

---

**Description**

Data for Exercise 9.11 and 9.17

**Usage**

Poverty

**Format**

A data frame/tibble with 20 observations on four variables

**city** a factor with levels Atlanta, Buffalo, Cincinnati, Cleveland, Dayton, O, Detroit, Flint, Mich, Fresno, C, Gary, Ind, Hartford, C, Laredo, Macon, Ga, Miami, Milwaukee, New Orleans, Newark, NJ, Rochester, NY, Shreveport, St. Louis, and Waco, Tx

**poverty** percent of children living in poverty

**crime** crime rate (per 1000 people)

**population** population of city

**Source**

Children's Defense Fund and the Bureau of Justice Statistics.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(poverty ~ crime, data = Poverty)
model <- lm(poverty ~ crime, data = Poverty)
abline(model, col = "red")
summary(model)
rm(model)
```

---

Precinct

*Robbery rates versus percent low income in eight precincts*

---

## Description

Data for Exercise 2.2 and 2.38

## Usage

Precinct

## Format

A data frame/tibble with eight observations on two variables

**rate** robbery rate (per 1000 people)

**income** percent with low income

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(rate ~ income, data = Precinct)
model <- (lm(rate ~ income, data = Precinct))
abline(model, col = "red")
rm(model)
```

---

Prejudic

*Racial prejudice measured on a sample of 25 high school students*

---

**Description**

Data for Exercise 5.10 and 5.22

**Usage**

Prejudic

**Format**

A data frame with 25 observations on one variable

**prejud** racial prejudice score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Prejudic$prejud)
EDA(Prejudic$prejud)
```

---

Presiden

*Ages at inauguration and death of U.S. presidents*

---

**Description**

Data for Exercise 1.126

**Usage**

Presiden

**Format**

A data frame/tibble with 43 observations on five variables

**first\_initial** a factor with levels A., B., C., D., F., G., G. W., H., J., L., M., R., T., U., W., and Z.

**last\_name** a factor with levels Adams, Arthur, Buchanan, Bush, Carter, Cleveland, Clinton, Coolidge, Eisenhower, Fillmore, Ford, Garfield, Grant, Harding, Harrison, Hayes, Hoover, Jackson, Jefferson, Johnson, Kennedy, Lincoln, Madison, McKinley, Monroe, Nixon, Pierce, Polk, Reagan, Roosevelt, Taft, Taylor, Truman, Tyler, VanBuren, Washington, and Wilson

**birth\_state** a factor with levels ARK, CAL, CONN, GA, IA, ILL, KY, MASS, MO, NC, NEB, NH, NJ, NY, OH, PA, SC, TEX, VA, and VT

**inaugural\_age** President's age at inauguration

**death\_age** President's age at death

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
pie(xtabs(~birth_state, data = Presiden))
stem(Presiden$inaugural_age)
stem(Presiden$death_age)
par(mar = c(5.1, 4.1 + 3, 4.1, 2.1))
stripchart(x=list(Presiden$inaugural_age, Presiden$death_age),
           method = "stack", col = c("green", "brown"), pch = 19, las = 1)
par(mar = c(5.1, 4.1, 4.1, 2.1))
```

---

Press

*Degree of confidence in the press versus education level for 20 randomly selected persons*

---

**Description**

Data for Exercise 9.55

**Usage**

Press

**Format**

A data frame/tibble with 20 observations on two variables

**education\_yrs** years of education

**confidence** degree of confidence in the press (the higher the score, the more confidence)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(confidence ~ education_yrs, data = Press)
model <- lm(confidence ~ education_yrs, data = Press)
abline(model, col = "purple")
summary(model)
rm(model)
```

---

Prognost	<i>Klopper's prognostic rating scale for subjects receiving behavior modification therapy</i>
----------	---

---

**Description**

Data for Exercise 6.61

**Usage**

Prognost

**Format**

A data frame/tibble with 15 observations on one variable

**kprs\_score** Klopper's Prognostic Rating Scale score

**Source**

Newmark, C., et al. (1973), Predictive Validity of the Rorschach Prognostic Rating Scale with Behavior Modification Techniques, *Journal of Clinical Psychology*, 29, 246-248.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Prognost$kprs_score)
t.test(Prognost$kprs_score, mu = 9)
```

---

Program	<i>Effects of four different methods of programmed learning for statistics students</i>
---------	---

---

### Description

Data for Exercise 10.17

### Usage

Program

### Format

A data frame/tibble with 44 observations on two variables

**method** a character variable with values method1, method2, method3, and method4

**score** standardized test score

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
boxplot(score ~ method, col = c("red", "blue", "green", "yellow"), data = Program)
anova(lm(score ~ method, data = Program))
TukeyHSD(aov(score ~ method, data = Program))
par(mar = c(5.1, 4.1 + 4, 4.1, 2.1))
plot(TukeyHSD(aov(score ~ method, data = Program)), las = 1)
par(mar = c(5.1, 4.1, 4.1, 2.1))
```

---

Psat	<i>PSAT scores versus SAT scores</i>
------	--------------------------------------

---

### Description

Data for Exercise 2.50

### Usage

Psat



**Format**

A data frame/tibble with seven observations on the two variables

**psat** PSAT score

**sat** SAT score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
model <- lm(sat ~ psat, data = Psat)
par(mfrow = c(1, 2))
plot(Psat$psat, resid(model))
plot(model, which = 1)
rm(model)
par(mfrow = c(1, 1))
```

---

Psych

*Correct responses for 24 students in a psychology experiment*

---

**Description**

Data for Exercise 1.42

**Usage**

Psych

**Format**

A data frame/tibble with 23 observations on one variable

**score** number of correct responses in a psychology experiment

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Psych$score)
EDA(Psych$score)
```

---

Puerto	<i>Weekly incomes of a random sample of 50 Puerto Rican families in Miami</i>
--------	---

---

**Description**

Data for Exercise 5.22 and 5.65

**Usage**

Puerto

**Format**

A data frame/tibble with 50 observations on one variable

**income** weekly family income (in dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Puerto$income)
boxplot(Puerto$income, col = "purple")
t.test(Puerto$income, conf.level = .90)$conf
```

---

Quail	<i>Plasma LDL levels in two groups of quail</i>
-------	---

---

**Description**

Data for Exercise 1.53, 1.77, 1.88, 5.66, and 7.50

**Usage**

Quail

**Format**

A data frame/tibble with 40 observations on two variables

**group** a character variable with values placebo and treatment

**level** low-density lipoprotein (LDL) cholesterol level

**Source**

J. McKean, and T. Vidmar (1994), "A Comparison of Two Rank-Based Methods for the Analysis of Linear Models," *The American Statistician*, 48, 220-229.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(level ~ group, data = Quail, horizontal = TRUE, xlab = "LDL Level",
        col = c("yellow", "lightblue"))
```

---

Quality

*Quality control test scores on two manufacturing processes*

---

**Description**

Data for Exercise 7.81

**Usage**

Quality

**Format**

A data frame/tibble with 15 observations on two variables

**process** a character variable with values Process1 and Process2

**score** results of a quality control test

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(score ~ process, data = Quality, col = "lightgreen")
t.test(score ~ process, data = Quality)
```

---

Rains	<i>Rainfall in an area of west central Kansas and four surrounding counties</i>
-------	---

---

**Description**

Data for Exercise 9.8

**Usage**

Rains

**Format**

A data frame/tibble with 35 observations on five variables

**rain** rainfall (in inches)

**x1** rainfall (in inches)

**x2** rainfall (in inches)

**x3** rainfall (in inches)

**x4** rainfall (in inches)

**Source**

R. Picard, K. Berk (1990), Data Splitting, *The American Statistician*, 44, (2), 140-147.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
cor(Rains)
model <- lm(rain ~ x2, data = Rains)
summary(model)
```

---

Randd	<i>Research and development expenditures and sales of a large company</i>
-------	---

---

**Description**

Data for Exercise 9.36 and Example 9.8

**Usage**

Randd

**Format**

A data frame/tibble with 12 observations on two variables

**rd** research and development expenditures (in million dollars)

**sales** sales (in million dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(sales ~ rd, data = Randd)
model <- lm(sales ~ rd, data = Randd)
abline(model, col = "purple")
summary(model)
plot(model, which = 1)
rm(model)
```

---

Rat	<i>Survival times of 20 rats exposed to high levels of radiation</i>
-----	--

---

**Description**

Data for Exercise 1.52, 1.76, 5.62, and 6.44

**Usage**

Rat

**Format**

A data frame/tibble with 20 observations on one variable

**survival\_time** survival time in weeks for rats exposed to a high level of radiation

**Source**

J. Lawless, *Statistical Models and Methods for Lifetime Data* (New York: Wiley, 1982).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Rat$survival_time)
qqnorm(Rat$survival_time)
qqline(Rat$survival_time)
summary(Rat$survival_time)
t.test(Rat$survival_time)
t.test(Rat$survival_time, mu = 100, alternative = "greater")
```

---

Ratings

*Grade point averages versus teacher's ratings*


---

**Description**

Data for Example 2.6

**Usage**

Ratings

**Format**

A data frame/tibble with 250 observations on two variables

**rating** character variable with students' ratings of instructor (A-F)

**gpa** students' grade point average

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(gpa ~ rating, data = Ratings, xlab = "Student rating of instructor",
        ylab = "Student GPA")
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Ratings, aes(x = rating, y = gpa, fill = rating)) +
  geom_boxplot() +
  theme_bw() +
  theme(legend.position = "none") +
  labs(x = "Student rating of instructor", y = "Student GPA")

## End(Not run)
```

---

Reaction

*Threshold reaction time for persons subjected to emotional stress*

---

**Description**

Data for Example 6.11

**Usage**

Reaction

**Format**

A data frame/tibble with 12 observations on one variable

**time** threshold reaction time (in seconds) for persons subjected to emotional stress

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Reaction$time)
SIGN.test(Reaction$time, md = 15, alternative = "less")
```

---

Reading

*Standardized reading scores for 30 fifth graders*

---

**Description**

Data for Exercise 1.72 and 2.10

**Usage**

Reading

**Format**

A data frame/tibble with 30 observations on four variables

**score** standardized reading test score

**sorted** sorted values of score

**trimmed** trimmed values of sorted

**winsoriz** winsorized values of score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Reading$score, main = "Exercise 1.72",  
     col = "lightgreen", xlab = "Standardized reading score")  
summary(Reading$score)  
sd(Reading$score)
```

---

Readiq

*Reading scores versus IQ scores*

---

**Description**

Data for Exercises 2.10 and 2.53

**Usage**

Readiq



**Format**

A data frame/tibble with 14 observations on two variables

**reading** reading achievement score

**iq** IQ score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(reading ~ iq, data = Readiq)
model <- lm(reading ~ iq, data = Readiq)
abline(model, col = "purple")
predict(model, newdata = data.frame(iq = c(100, 120)))
residuals(model)[c(6, 7)]
rm(model)
```

---

Referend

*Opinion on referendum by view on freedom of the press*

---

**Description**

Data for Exercise 8.20

**Usage**

Referend

**Format**

A data frame with 237 observations on two variables

**choice** a factor with levels A, B, and C

**response** a factor with levels for, against, and undecided

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~choice + response, data = Referend)
T1
chisq.test(T1)
chisq.test(T1)$expected
```

---

Region	<i>Pollution index taken in three regions of the country</i>
--------	--

---

**Description**

Data for Exercise 10.26

**Usage**

Region

**Format**

A data frame/tibble with 48 observations on three variables

**pollution** pollution index

**region** region of a county (west, central, and east)

**ranks** ranked values of pollution

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(pollution ~ region, data = Region, col = "gray")
anova(lm(pollution ~ region, data = Region))
```

---

Register

*Maintenance cost versus age of cash registers in a department store*

---

**Description**

Data for Exercise 2.3, 2.39, and 2.54

**Usage**

Register

**Format**

A data frame/tibble with nine observations on two variables

**age** age of cash register (in years)

**cost** maintenance cost of cash register (in dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(cost ~ age, data = Register)
model <- lm(cost ~ age, data = Register)
abline(model, col = "red")
predict(model, newdata = data.frame(age = c(5, 10)))
plot(model, which = 1)
rm(model)
```

---

Rehab

*Rehabilitative potential of 20 prison inmates as judged by two psychiatrists*

---

**Description**

Data for Exercise 7.61

**Usage**

Rehab

**Format**

A data frame/tibble with 20 observations on four variables

**inmate** inmate identification number

**psych1** rating from first psychiatrist on the inmates rehabilitative potential

**psych2** rating from second psychiatrist on the inmates rehabilitative potential

**differ** psych1 - psych2

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(Rehab$differ)
qqnorm(Rehab$differ)
qqline(Rehab$differ)
t.test(Rehab$differ)
# Or
t.test(Rehab$psych1, Rehab$psych2, paired = TRUE)
```

---

Remedial

*Math placement test score for 35 freshmen females and 42 freshmen males*

---

**Description**

Data for Exercise 7.43

**Usage**

Remedial

**Format**

A data frame/tibble with 84 observations on two variables

**gender** a character variable with values female and male

**score** math placement score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(score ~ gender, data = Remedial,
col = c("purple", "blue"))
t.test(score ~ gender, data = Remedial, conf.level = 0.98)
t.test(score ~ gender, data = Remedial, conf.level = 0.98)$conf
wilcox.test(score ~ gender, data = Remedial,
conf.int = TRUE, conf.level = 0.98)
```

---

Rentals

*Weekly rentals for 45 apartments*

---

**Description**

Data for Exercise 1.122

**Usage**

Rentals

**Format**

A data frame/tibble with 45 observations on one variable

**rent** weekly apartment rental price (in dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Rentals$rent)
sum(Rentals$rent < mean(Rentals$rent) - 3*sd(Rentals$rent) |
Rentals$rent > mean(Rentals$rent) + 3*sd(Rentals$rent))
```

---

Repair	<i>Recorded times for repairing 22 automobiles involved in wrecks</i>
--------	---

---

**Description**

Data for Exercise 5.77

**Usage**

Repair

**Format**

A data frame/tibble with 22 observations on one variable

**time** time to repair a wrecked in car (in hours)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Repair$time)
SIGN.test(Repair$time, conf.level = 0.98)
```

---

Retail	<i>Length of employment versus gross sales for 10 employees of a large retail store</i>
--------	---

---

**Description**

Data for Exercise 9.59

**Usage**

Retail

**Format**

A data frame/tibble with 10 observations on two variables

**months** length of employment (in months)

**sales** employee gross sales (in dollars)

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(sales ~ months, data = Retail)
model <- lm(sales ~ months, data = Retail)
abline(model, col = "blue")
summary(model)
```

---

Ronbrown1

*Oceanography data obtained at site 1 by scientist aboard the ship Ron Brown*

---

## Description

Data for Exercise 2.9

## Usage

Ronbrown1

## Format

A data frame/tibble with 75 observations on two variables

**depth** ocean depth (in meters)

**temperature** ocean temperature (in Celsius)

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(temperature ~ depth, data = Ronbrown1, ylab = "Temperature")
```

---

Ronbrown2

*Oceanography data obtained at site 2 by scientist aboard the ship Ron Brown*

---

### Description

Data for Exercise 2.56 and Example 2.4

### Usage

Ronbrown2

### Format

A data frame/tibble with 150 observations on three variables

**depth** ocean depth (in meters)

**temperature** ocean temperature (in Celcius)

**salinity** ocean salinity level

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
plot(salinity ~ depth, data = Ronbrown2)
model <- lm(salinity ~ depth, data = Ronbrown2)
summary(model)
plot(model, which = 1)
rm(model)
```

---

Rural

*Social adjustment scores for a rural group and a city group of children*

---

### Description

Data for Example 7.16

### Usage

Rural



**Format**

A data frame/tibble with 33 observations on two variables

**score** child's social adjustment score

**area** character variable with values ci ty and rural

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(score ~ area, data = Rural)
wilcox.test(score ~ area, data = Rural)
## Not run:
library(dplyr)
Rural <- dplyr::mutate(Rural, r = rank(score))
Rural
t.test(r ~ area, data = Rural)

## End(Not run)
```

---

Salary

*Starting salaries for 25 new PhD psychologist*

---

**Description**

Data for Exercise 3.66

**Usage**

Salary

**Format**

A data frame/tibble with 25 observations on one variable

**salary** starting salary for Ph.D. psychologists (in dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Salary$salary, pch = 19, col = "purple")
qqline(Salary$salary, col = "blue")
```

---

Salinity

*Surface-water salinity measurements from Whitewater Bay, Florida*

---

**Description**

Data for Exercise 5.27 and 5.64

**Usage**

Salinity

**Format**

A data frame/tibble with 48 observations on one variable

**salinity** surface-water salinity value

**Source**

J. Davis, *Statistics and Data Analysis in Geology*, 2nd ed. (New York: John Wiley, 1986).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Salinity$salinity)
qqnorm(Salinity$salinity, pch = 19, col = "purple")
qqline(Salinity$salinity, col = "blue")
t.test(Salinity$salinity, conf.level = 0.99)
t.test(Salinity$salinity, conf.level = 0.99)$conf
```

---

Sat	<i>SAT scores, percent taking exam and state funding per student by state for 1994, 1995 and 1999</i>
-----	---

---

### Description

Data for Statistical Insight Chapter 9

### Usage

Sat

### Format

A data frame/tibble with 102 observations on seven variables

**state** U.S. state

**verbal** verbal SAT score

**math** math SAT score

**total** combined verbal and math SAT score

**percent** percent of high school seniors taking the SAT

**expend** state expenditure per student (in dollars)

**year** year

### Source

*The 2000 World Almanac and Book of Facts*, Funk and Wagnalls Corporation, New Jersey.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
Sat94 <- Sat[Sat$year == 1994, ]
Sat94
Sat99 <- subset(Sat, year == 1999)
Sat99
stem(Sat99$total)
plot(total ~ percent, data = Sat99)
model <- lm(total ~ percent, data = Sat99)
abline(model, col = "blue")
summary(model)
rm(model)
```

---

Saving	<i>Problem asset ration for savings and loan companies in California, New York, and Texas</i>
--------	---

---

**Description**

Data for Exercise 10.34 and 10.49

**Usage**

Saving

**Format**

A data frame/tibble with 65 observations on two variables

**par** problem-asset-ratio for Savings & Loans that were listed as being financially troubled in 1992

**state** U.S. state

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(par ~ state, data = Saving, col = "red")
boxplot(par ~ state, data = Saving, log = "y", col = "red")
model <- aov(par ~ state, data = Saving)
summary(model)
plot(TukeyHSD(model))
kruskal.test(par ~ factor(state), data = Saving)
```

---

Scales	<i>Readings obtained from a 100 pound weight placed on four brands of bathroom scales</i>
--------	---

---

**Description**

Data for Exercise 1.89

**Usage**

Scales

**Format**

A data frame/tibble with 20 observations on two variables

**brand** variable indicating brand of bathroom scale (A, B, C, or D)

**reading** recorded value (in pounds) of a 100 pound weight

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

boxplot(reading ~ brand, data = Scales, col = rainbow(4),
ylab = "Weight (lbs)")
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Scales, aes(x = brand, y = reading, fill = brand)) +
  geom_boxplot() +
  labs(y = "weight (lbs)") +
  theme_bw() +
  theme(legend.position = "none")

## End(Not run)

```

---

Schizop2

*Exam scores for 17 patients to assess the learning ability of schizophrenics after taking a specified dose of a tranquilizer*

---

**Description**

Data for Exercise 6.99

**Usage**

Schizop2

**Format**

A data frame/tibble with 17 observations on one variable

**score** schizophrenics score on a second standardized exam

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Schizop2$score, xlab = "score on standardized test after a tranquilizer",
     main = "Exercise 6.99", breaks = 10, col = "orange")
EDA(Schizop2$score)
SIGN.test(Schizop2$score, md = 22, alternative = "greater")
```

---

 Schizoph

*Standardized exam scores for 13 patients to investigate the learning ability of schizophrenics after a specified dose of a tranquilizer*

---

**Description**

Data for Example 6.10

**Usage**

Schizoph

**Format**

A data frame/tibble with 13 observations on one variable

**score** schizophrenics score on a standardized exam one hour after receiving a specified dose of a tranquilizer.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Schizoph$score, xlab = "score on standardized test",
     main = "Example 6.10", breaks = 10, col = "orange")
EDA(Schizoph$score)
t.test(Schizoph$score, mu = 20)
```

---

Seatbelt                      *Injury level versus seatbelt usage*

---

**Description**

Data for Exercise 8.24

**Usage**

Seatbelt

**Format**

A data frame/tibble with 86,759 observations on two variables

**seatbelt** a factor with levels No and Yes

**injuries** a factor with levels None, Minimal, Minor, or Major indicating the extent of the drivers injuries

**Source**

Jobson, J. (1982), *Applied Multivariate Data Analysis*, Springer-Verlag, New York, p. 18.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~seatbelt + injuries, data = Seatbelt)
T1
chisq.test(T1)
rm(T1)
```

---

Selfdefe	<i>Self-confidence scores for 9 women before and after instructions on self-defense</i>
----------	---

---

**Description**

Data for Example 7.19

**Usage**

Selfdefe

**Format**

A data frame/tibble with nine observations on three variables

**woman** number identifying the woman

**before** before the course self-confidence score

**after** after the course self-confidence score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
Selfdefe$differ <- Selfdefe$after - Selfdefe$before
Selfdefe
t.test(Selfdefe$differ, alternative = "greater")
t.test(Selfdefe$after, Selfdefe$before,
       paired = TRUE, alternative = "greater")
```

---

Senior	<i>Reaction times of 30 senior citizens applying for drivers license renewals</i>
--------	---

---

**Description**

Data for Exercise 1.83 and 3.67

**Usage**

Senior



**Format**

A data frame/tibble with 31 observations on one variable

**reaction** reaction time for senior citizens applying for a driver's license renewal

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Senior$reaction)
fivenum(Senior$reaction)
boxplot(Senior$reaction, main = "Problem 1.83, part d",
        horizontal = TRUE, col = "purple")
```

---

Sentence

*Sentences of 41 prisoners convicted of a homicide offense*

---

**Description**

Data for Exercise 1.123

**Usage**

Sentence

**Format**

A data frame/tibble with 41 observations on one variable

**months** sentence length (in months) for prisoners convicted of homicide

**Source**

U.S. Department of Justice, Bureau of Justice Statistics, *Prison Sentences and Time Served for Violence*, NCJ-153858, April 1995.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
stem(Sentence$months)
ll <- mean(Sentence$months)-2*sd(Sentence$months)
ul <- mean(Sentence$months)+2*sd(Sentence$months)
limits <- c(ll, ul)
limits
rm(ul, ll, limits)
```

---

Shkdrug

*Effects of a drug and electroshock therapy on the ability to solve simple tasks*

---

## Description

Data for Exercises 10.11 and 10.12

## Usage

Shkdrug

## Format

A data frame/tibble with 64 observations on two variables

**treatment** type of treatment Drug/NoS, Drug/Shk, NoDg/NoS, or NoDrug/S

**response** number of tasks completed in a 10-minute period

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(response ~ treatment, data = Shkdrug, col = "gray")
model <- lm(response ~ treatment, data = Shkdrug)
anova(model)
rm(model)
```

---

Shock

*Effect of experimental shock on time to complete difficult task*

---

**Description**

Data for Exercise 10.50

**Usage**

Shock

**Format**

A data frame/tibble with 27 observations on two variables

**group** grouping variable with values of Group1 (no shock), Group2 (medium shock), and Group3 (severe shock)

**attempts** number of attempts to complete a task

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(attempts ~ group, data = Shock, col = "violet")
model <- lm(attempts ~ group, data = Shock)
anova(model)
rm(model)
```

---

Shoplift

*Sales receipts versus shoplifting losses for a department store*

---

**Description**

Data for Exercise 9.58

**Usage**

Shoplift

**Format**

A data frame/tibble with eight observations on two variables

**sales** sales (in 1000 dollars)

**loss** loss (in 100 dollars)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(loss ~ sales, data = Shoplift)
model <- lm(loss ~ sales, data = Shoplift)
summary(model)
rm(model)
```

---

Short

*James Short's measurements of the parallax of the sun*

---

**Description**

Data for Exercise 6.65

**Usage**

Short

**Format**

A data frame/tibble with 158 observations on two variables

**sample** sample number

**parallax** parallax measurements (seconds of a degree)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Short$parallax, main = "Problem 6.65",  
     xlab = "", col = "orange")  
SIGN.test(Short$parallax, md = 8.798)  
t.test(Short$parallax, mu = 8.798)
```

---

Shuttle	<i>Number of people riding shuttle versus number of automobiles in the downtown area</i>
---------	--

---

**Description**

Data for Exercise 9.20

**Usage**

Shuttle

**Format**

A data frame/tibble with 15 observations on two variables

**users** number of shuttle riders

**autos** number of automobiles in the downtown area

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(autos ~ users, data = Shuttle)  
model <- lm(autos ~ users, data = Shuttle)  
summary(model)  
rm(model)
```

SIGN.test

*Sign Test***Description**

This function will test a hypothesis based on the sign test and reports linearly interpolated confidence intervals for one sample problems.

**Usage**

```
SIGN.test(
  x,
  y = NULL,
  md = 0,
  alternative = "two.sided",
  conf.level = 0.95,
  ...
)
```

**Arguments**

x	numeric vector; NAs and Infs are allowed but will be removed.
y	optional numeric vector; NAs and Infs are allowed but will be removed.
md	a single number representing the value of the population median specified by the null hypothesis
alternative	is a character string, one of "greater", "less", or "two.sided", or the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, alternative refers to the true median of the parent population in relation to the hypothesized value of the median.
conf.level	confidence level for the returned confidence interval, restricted to lie between zero and one
...	further arguments to be passed to or from methods

**Details**

Computes a "Dependent-samples Sign-Test" if both x and y are provided. If only x is provided, computes the "Sign-Test".

**Value**

A list of class `htest_S`, containing the following components:

statistic	the S-statistic (the number of positive differences between the data and the hypothesized median), with names attribute "S".
p.value	the p-value for the test

conf.int	is a confidence interval (vector of length 2) for the true median based on linear interpolation. The confidence level is recorded in the attribute <code>conf.level</code> . When the alternative is not "two.sided", the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values $k$ for which one would not reject the null hypothesis that the true mean or difference in means is $k$ . Here infinity will be represented by <code>Inf</code> .
estimate	is a vector of length 1, giving the sample median; this estimates the corresponding population parameter. Component <code>estimate</code> has a <code>names</code> attribute describing its elements.
null.value	is the value of the median specified by the null hypothesis. This equals the input argument <code>md</code> . Component <code>null.value</code> has a <code>names</code> attribute describing its elements.
alternative	records the value of the input argument <code>alternative</code> : "greater", "less", or "two.sided"
data.name	a character string (vector of length 1) containing the actual name of the input vector <code>x</code>
Confidence.Intervals	a 3 by 3 matrix containing the lower achieved confidence interval, the interpolated confidence interval, and the upper achieved confidence interval

### Null Hypothesis

For the one-sample sign-test, the null hypothesis is that the median of the population from which  $x$  is drawn is  $md$ . For the two-sample dependent case, the null hypothesis is that the median for the differences of the populations from which  $x$  and  $y$  are drawn is  $md$ . The alternative hypothesis indicates the direction of divergence of the population median for  $x$  from  $md$  (i.e., "greater", "less", "two.sided".)

### Note

The reported confidence interval is based on linear interpolation. The lower and upper confidence levels are exact.

### Author(s)

Alan T. Arnholt

### References

- Gibbons, J.D. and Chakraborti, S. (1992). *Nonparametric Statistical Inference*. Marcel Dekker Inc., New York.
- Kitchens, L.J.(2003). *Basic Statistics and Data Analysis*. Duxbury.
- Conover, W. J. (1980). *Practical Nonparametric Statistics, 2nd ed.* Wiley, New York.
- Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden and Day, San Francisco.

**See Also**

[z.test](#), [zsum.test](#), [tsum.test](#)

**Examples**

```
x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7., 6.4, 7.1, 6.7, 7.6, 6.8)
SIGN.test(x, md = 6.5)
# Computes two-sided sign-test for the null hypothesis
# that the population median for 'x' is 6.5. The alternative
# hypothesis is that the median is not 6.5. An interpolated 95%
# confidence interval for the population median will be computed.

reaction <- c(14.3, 13.7, 15.4, 14.7, 12.4, 13.1, 9.2, 14.2,
             14.4, 15.8, 11.3, 15.0)
SIGN.test(reaction, md = 15, alternative = "less")
# Data from Example 6.11 page 330 of Kitchens BSDA.
# Computes one-sided sign-test for the null hypothesis
# that the population median is 15. The alternative
# hypothesis is that the median is less than 15.
# An interpolated upper 95% upper bound for the population
# median will be computed.
```

---

Simpson

*Grade point averages of men and women participating in various sports-an illustration of Simpson's paradox*

---

**Description**

Data for Example 1.18

**Usage**

Simpson

**Format**

A data frame/tibble with 100 observations on three variables

**gpa** grade point average

**sport** sport played (basketball, soccer, or track)

**gender** athlete sex (male, female)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.



**Examples**

```

boxplot(gpa ~ gender, data = Simpson, col = "violet")
boxplot(gpa ~ sport, data = Simpson, col = "lightgreen")
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Simpson, aes(x = gender, y = gpa, fill = gender)) +
  geom_boxplot() +
  facet_grid(.~sport) +
  theme_bw()

## End(Not run)

```

---

Situp

*Maximum number of situps by participants in an exercise class*


---

**Description**

Data for Exercise 1.47

**Usage**

Situp

**Format**

A data frame/tibble with 20 observations on one variable

**number** maximum number of situps completed in an exercise class after 1 month in the program

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

stem(Situp$number)
hist(Situp$number, breaks = seq(0, 70, 10), right = FALSE)
hist(Situp$number, breaks = seq(0, 70, 10), right = FALSE,
     freq = FALSE, col = "pink", main = "Problem 1.47",
     xlab = "Maximum number of situps")
lines(density(Situp$number), col = "red")

```

---

Skewed

*Illustrates the Wilcoxon Rank Sum test*

---

**Description**

Data for Exercise 7.65

**Usage**

Skewed

**Format**

A data frame/tibble with 21 observations on two variables

**C1** values from a sample of size 16 from a particular population

**C2** values from a sample of size 14 from a particular population

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(Skewed$C1, Skewed$C2, col = c("pink", "lightblue"))
wilcox.test(Skewed$C1, Skewed$C2)
```

---

Skin

*Survival times of closely and poorly matched skin grafts on burn patients*

---

**Description**

Data for Exercise 5.20

**Usage**

Skin

**Format**

A data frame/tibble with 11 observations on four variables

**patient** patient identification number

**close** graft survival time in days for a closely matched skin graft on the same burn patient

**poor** graft survival time in days for a poorly matched skin graft on the same burn patient

**differ** difference between close and poor (in days)

**Source**

R. F. Woolon and P. A. Lachenbruch, "Rank Tests for Censored Matched Pairs," *Biometrika*, 67(1980), 597-606.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Skin$differ)
boxplot(Skin$differ, col = "pink")
summary(Skin$differ)
```

---

 Slc

*Sodium-lithium countertransport activity on 190 individuals from six large English kindred*

---

**Description**

Data for Exercise 5.116

**Usage**

Slc

**Format**

A data frame/tibble with 190 observations on one variable

**slc** Red blood cell sodium-lithium countertransport

**Source**

Roeder, K., (1994), "A Graphical Technique for Determining the Number of Components in a Mixture of Normals," *Journal of the American Statistical Association*, 89, 497-495.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
EDA(Slc$slc)
hist(Slc$slc, freq = FALSE, xlab = "sodium lithium countertransport",
     main = "", col = "lightblue")
lines(density(Slc$slc), col = "purple")
```

---

Smokyph

*Water pH levels of 75 water samples taken in the Great Smoky Mountains*

---

## Description

Data for Exercises 6.40, 6.59, 7.10, and 7.35

## Usage

Smokyph

## Format

A data frame/tibble with 75 observations on three variables

**waterph** water sample pH level

**code** character variable with values low (elevation below 0.6 miles), and high (elevation above 0.6 miles)

**elev** elevation in miles

## Source

Schmoyer, R. L. (1994), Permutation Tests for Correlation in Regression Errors, *Journal of the American Statistical Association*, 89, 1507-1516.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
summary(Smokyph$waterph)
tapply(Smokyph$waterph, Smokyph$code, mean)
stripchart(waterph ~ code, data = Smokyph, method = "stack",
           pch = 19, col = c("red", "blue"))
t.test(Smokyph$waterph, mu = 7)
SIGN.test(Smokyph$waterph, md = 7)
t.test(waterph ~ code, data = Smokyph, alternative = "less")
t.test(waterph ~ code, data = Smokyph, conf.level = 0.90)

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Smokyph, aes(x = waterph, fill = code)) +
  geom_dotplot() +
  facet_grid(code ~ .) +
  guides(fill = FALSE)

## End(Not run)
```

---

Snore

*Snoring versus heart disease*

---

## Description

Data for Exercise 8.21

## Usage

Snore

## Format

A data frame/tibble with 2,484 observations on two variables

**snore** factor with levels nonsnorer, ocassional snorer, nearly every night, and snores every night

**heartdisease** factor indicating whether the indiviudal has heart disease (no or yes)

## Source

Norton, P. and Dunn, E. (1985), Snoring as a Risk Factor for Disease, *British Medical Journal*, 291, 630-632.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~ heartdisease + snore, data = Snore)
T1
chisq.test(T1)
rm(T1)
```

---

Snow

*Concentration of microparticles in snowfields of Greenland and Antarctica*

---

**Description**

Data for Exercise 7.87

**Usage**

Snow

**Format**

A data frame/tibble with 34 observations on two variables

**concent** concentration of microparticles from melted snow (in parts per billion)

**site** location of snow sample (Antarctica or Greenland)

**Source**

Davis, J., *Statistics and Data Analysis in Geology*, John Wiley, New York.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(concent ~ site, data = Snow, col = c("lightblue", "lightgreen"))
```

---

Soccer                      *Weights of 25 soccer players*

---

**Description**

Data for Exercise 1.46

**Usage**

Soccer

**Format**

A data frame/tibble with 25 observations on one variable

**weight** soccer players weight (in pounds)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Soccer$weight, scale = 2)
hist(Soccer$weight, breaks = seq(110, 210, 10), col = "orange",
     main = "Problem 1.46 \n Weights of Soccer Players",
     xlab = "weight (lbs)", right = FALSE)
```

---

Social                      *Median income level for 25 social workers from North Carolina*

---

**Description**

Data for Exercise 6.63

**Usage**

Social

**Format**

A data frame/tibble with 25 observations on one variable

**income** annual income (in dollars) of North Carolina social workers with less than five years experience.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
SIGN.test(Social$income, md = 27500, alternative = "less")
```

---

Sophomor	<i>Grade point averages, SAT scores and final grade in college algebra for 20 sophomores</i>
----------	--

---

**Description**

Data for Exercise 2.42

**Usage**

Sophomor

**Format**

A data frame/tibble with 20 observations on four variables

**student** identification number

**gpa** grade point average

**sat** SAT math score

**exam** final exam grade in college algebra

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
cor(Sophomor)
plot(exam ~ gpa, data = Sophomor)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Sophomor, aes(x = gpa, y = exam)) +
  geom_point()
ggplot2::ggplot(data = Sophomor, aes(x = sat, y = exam)) +
  geom_point()

## End(Not run)
```



---

South

*Murder rates for 30 cities in the South*

---

**Description**

Data for Exercise 1.84

**Usage**

South

**Format**

A data frame/tibble with 31 observations on one variable

**rate** murder rate per 100,000 people

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(South$rate, col = "gray", ylab = "Murder rate per 100,000 people")
```

---

Speed

*Speed reading scores before and after a course on speed reading*

---

**Description**

Data for Exercise 7.58

**Usage**

Speed

**Format**

A data frame/tibble with 15 observations on four variables

**before** reading comprehension score before taking a speed-reading course

**after** reading comprehension score after taking a speed-reading course

**differ** after - before (comprehension reading scores)

**signranks** signed ranked differences

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
t.test(Speed$differ, alternative = "greater")
t.test(Speed$signranks, alternative = "greater")
wilcox.test(Speed$after, Speed$before, paired = TRUE, alternative = "greater")
```

---

Spellers

*Standardized spelling test scores for two fourth grade classes*

---

## Description

Data for Exercise 7.82

## Usage

Spellers

## Format

A data frame/tibble with ten observations on two variables

**teacher** character variable with values Fourth and Colleague

**score** score on a standardized spelling test

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(score ~ teacher, data = Spellers, col = "pink")
t.test(score ~ teacher, data = Spellers)
```

---

Spelling	<i>Spelling scores for 9 eighth graders before and after a 2-week course of instruction</i>
----------	---

---

**Description**

Data for Exercise 7.56

**Usage**

Spelling

**Format**

A data frame/tibble with nine observations on three variables

**before** spelling score before a 2-week course of instruction

**after** spelling score after a 2-week course of instruction

**differ** after - before (spelling score)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Spelling$differ)
qqline(Spelling$differ)
shapiro.test(Spelling$differ)
t.test(Spelling$before, Spelling$after, paired = TRUE)
t.test(Spelling$differ)
```

---

Sports	<i>Favorite sport by gender</i>
--------	---------------------------------

---

**Description**

Data for Exercise 8.32

**Usage**

Sports

**Format**

A data frame/tibble with 200 observations on two variables

**gender** a factor with levels male and female

**sport** a factor with levels football, basketball, baseball, and tennis

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~gender + sport, data = Sports)
T1
chisq.test(T1)
rm(T1)
```

---

Spouse

*Convictions in spouse murder cases by gender*

---

**Description**

Data for Exercise 8.33

**Usage**

Spouse

**Format**

A data frame/tibble with 540 observations on two variables

**result** a factor with levels not prosecuted, pleaded guilty, convicted, and acquitted

**spouse** a factor with levels husband and wife

**Source**

Bureau of Justice Statistics (September 1995), *Spouse Murder Defendants in Large Urban Counties*, Executive Summary, NCJ-156831.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~result + spouse, data = Spouse)
T1
chisq.test(T1)
rm(T1)
```

---

SRS

*Simple Random Sampling*

---

**Description**

Computes all possible samples from a given population using simple random sampling.

**Usage**

```
SRS(POPvalues, n)
```

**Arguments**

POPvalues      vector containing the population values.  
n                the sample size.

**Value**

Returns a matrix containing the possible simple random samples of size n taken from a population POPvalues.

**Author(s)**

Alan T. Arnholt

**See Also**

[Combinations](#)

**Examples**

```
SRS(c(5,8,3),2)
# The rows in the matrix list the values for the 3 possible
# simple random samples of size 2 from the population of 5,8, and 3.
```

---

Stable	<i>Times of a 2-year old stallion on a one mile run</i>
--------	---

---

**Description**

Data for Exercise 6.93

**Usage**

Stable

**Format**

A data frame/tibble with nine observations on one variable

**time** time (in seconds) for horse to run 1 mile

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
SIGN.test(Stable$time, md = 98.5, alternative = "greater")
```

---

Stamp	<i>Thicknesses of 1872 Hidalgo stamps issued in Mexico</i>
-------	--

---

**Description**

Data for Statistical Insight Chapter 1 and Exercise 5.110

**Usage**

Stamp

**Format**

A data frame/tibble with 485 observations on one variable

**thickness** stamp thickness (in mm)

**Source**

Izenman, A., Sommer, C. (1988), Philatelic Mixtures and Multimodal Densities, *Journal of the American Statistical Association*, 83, 941-953.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Stamp$thickness, freq = FALSE, col = "lightblue",
     main = "", xlab = "stamp thickness (mm)")
lines(density(Stamp$thickness), col = "blue")
t.test(Stamp$thickness, conf.level = 0.99)
```

---

Statclas

*Grades for two introductory statistics classes*

---

**Description**

Data for Exercise 7.30

**Usage**

```
Statclas
```

**Format**

A data frame/tibble with 72 observations on two variables

**class** class meeting time (9am or 2pm)

**score** grade for an introductory statistics class

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
str(Statclas)
boxplot(score ~ class, data = Statclas, col = "red")
t.test(score ~ class, data = Statclas)
```

---

Statelaw	<i>Operating expenditures per resident for each of the state law enforcement agencies</i>
----------	---

---

**Description**

Data for Exercise 6.62

**Usage**

Statelaw

**Format**

A data frame/tibble with 50 observations on two variables

**state** U.S. state

**cost** dollars spent per resident on law enforcement

**Source**

Bureau of Justice Statistics, *Law Enforcement Management and Administrative Statistics, 1993*, NCJ-148825, September 1995, page 84.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Statelaw$cost)
SIGN.test(Statelaw$cost, md = 8, alternative = "less")
```

---

Statisti	<i>Test scores for two beginning statistics classes</i>
----------	---

---

**Description**

Data for Exercises 1.70 and 1.87

**Usage**

Statisti



**Format**

A data frame/tibble with 62 observations on two variables

**class** character variable with values Class1 and Class2

**score** test score for an introductory statistics test

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

boxplot(score ~ class, data = Statisti, col = "violet")
tapply(Statisti$score, Statisti$class, summary, na.rm = TRUE)
## Not run:
library(dplyr)
dplyr::group_by(Statisti, class) %>%
  summarize(Mean = mean(score, na.rm = TRUE),
            Median = median(score, na.rm = TRUE),
            SD = sd(score, na.rm = TRUE),
            RS = IQR(score, na.rm = TRUE))

## End(Not run)

```

---

Step

*STEP science test scores for a class of ability-grouped students*

---

**Description**

Data for Exercise 6.79

**Usage**

Step

**Format**

A data frame/tibble with 12 observations on one variable

**score** State test of educational progress (STEP) science test score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Step$score)
t.test(Step$score, mu = 80, alternative = "less")
wilcox.test(Step$score, mu = 80, alternative = "less")
```

---

Stress	<i>Short-term memory test scores on 12 subjects before and after a stressful situation</i>
--------	--

---

**Description**

Data for Example 7.20

**Usage**

Stress

**Format**

A data frame/tibble with 12 observations on two variables

**prestress** short term memory score before being exposed to a stressful situation

**poststress** short term memory score after being exposed to a stressful situation

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
diff <- Stress$prestress - Stress$poststress
qqnorm(diff)
qqline(diff)
t.test(diff)
t.test(Stress$prestress, Stress$poststress, paired = TRUE)
## Not run:
wilcox.test(Stress$prestress, Stress$poststress, paired = TRUE)

## End(Not run)
```

---

Study	<i>Number of hours studied per week by a sample of 50 freshmen</i>
-------	--

---

**Description**

Data for Exercise 5.25

**Usage**

Study

**Format**

A data frame/tibble with 50 observations on one variable

**hours** number of hours a week freshmen reported studying for their courses

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Study$hours)
hist(Study$hours, col = "violet")
summary(Study$hours)
```

---

Submarin	<i>Number of German submarines sunk by U.S. Navy in World War II</i>
----------	--

---

**Description**

Data for Exercises 2.16, 2.45, and 2.59

**Usage**

Submarin

**Format**

A data frame/tibble with 16 observations on three variables

**month** month

**reported** number of submarines reported sunk by U.S. Navy

**actual** number of submarines actually sunk by U.S. Navy

**Source**

F. Mosteller, S. Fienberg, and R. Rourke, *Beginning Statistics with Data Analysis* (Reading, MA: Addison-Wesley, 1983).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
model <- lm(actual ~ reported, data = Submarin)
summary(model)
plot(actual ~ reported, data = Submarin)
abline(model, col = "red")
rm(model)
```

---

Subway

*Time it takes a subway to travel from the airport to downtown*

---

**Description**

Data for Exercise 5.19

**Usage**

Subway

**Format**

A data frame/tibble with 30 observations on one variable

**time** time (in minutes) it takes a subway to travel from the airport to downtown

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Subway$time, main = "Exercise 5.19",
xlab = "Time (in minutes)", col = "purple")
summary(Subway$time)
```

---

Sunspot

*Wolfer sunspot numbers from 1700 through 2000*

---

## Description

Data for Example 1.7

## Usage

Sunspot

## Format

A data frame/tibble with 301 observations on two variables

**year** year

**sunspots** average number of sunspots for the year

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(sunspots ~ year, data = Sunspot, type = "l")
## Not run:
library(ggplot2)
lattice::xyplot(sunspots ~ year, data = Sunspot,
               main = "Yearly sunspots", type = "l")
lattice::xyplot(sunspots ~ year, data = Sunspot, type = "l",
               main = "Yearly sunspots", aspect = "xy")
ggplot2::ggplot(data = Sunspot, aes(x = year, y = sunspots)) +
  geom_line() +
  theme_bw()

## End(Not run)
```

---

Superbowl

*Margin of victory in Superbowls I to XXXV*

---

**Description**

Data for Exercise 1.54

**Usage**

Superbowl

**Format**

A data frame/tibble with 35 observations on five variables

**winning\_team** name of Superbowl winning team

**winner\_score** winning score for the Superbowl

**losing\_team** name of Superbowl losing team

**loser\_score** score of losing team a numeric vector

**victory\_margin** winner\_score - loser\_score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Superbowl$victory_margin)
```

---

Supercar

*Top speeds attained by five makes of supercars*

---

**Description**

Data for Statistical Insight Chapter 10

**Usage**

Supercar

**Format**

A data frame/tibble with 30 observations on two variables

**speed** top speed (in miles per hour) of car without redlining

**car** name of sports car

**Source**

*Car and Driver* (July 1995).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(speed ~ car, data = Supercar, col = rainbow(6),
        ylab = "Speed (mph)")
summary(aov(speed ~ car, data = Supercar))
anova(lm(speed ~ car, data = Supercar))
```

---

 Tablrock

---

*Ozone concentrations at Mt. Mitchell, North Carolina*


---

**Description**

Data for Exercise 5.63

**Usage**

Tablrock

**Format**

A data frame/tibble with 719 observations on the following 17 variables.

**day** date

**hour** time of day

**ozone** ozone concentration

**tmp** temperature (in Celcius)

**vdc** a numeric vector

**wd** a numeric vector

**ws** a numeric vector

**amb** a numeric vector

**dew** a numeric vector

**so2** a numeric vector

**no** a numeric vector

**no2** a numeric vector

**nox** a numeric vector

**co** a numeric vector

**co2** a numeric vector

**gas** a numeric vector

**air** a numeric vector

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
summary(Tablrock$ozone)
boxplot(Tablrock$ozone)
qqnorm(Tablrock$ozone)
qqline(Tablrock$ozone)
par(mar = c(5.1 - 1, 4.1 + 2, 4.1 - 2, 2.1))
boxplot(ozone ~ day, data = Tablrock,
        horizontal = TRUE, las = 1, cex.axis = 0.7)
par(mar = c(5.1, 4.1, 4.1, 2.1))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Tablrock, aes(sample = ozone)) +
  geom_qq() +
  theme_bw()
ggplot2::ggplot(data = Tablrock, aes(x = as.factor(day), y = ozone)) +
  geom_boxplot(fill = "pink") +
  coord_flip() +
  labs(x = "") +
  theme_bw()

## End(Not run)
```



---

Teacher

*Average teacher's salaries across the states in the 70s 80s and 90s*

---

### Description

Data for Exercise 5.114

### Usage

Teacher

### Format

A data frame/tibble with 51 observations on three variables

**state** U.S. state

**year** academic year

**salary** average salary (in dollars)

### Source

National Education Association.

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
par(mfrow = c(3, 1))
hist(Teacher$salary[Teacher$year == "1973-74"],
     main = "Teacher salary 1973-74", xlab = "salary",
     xlim = range(Teacher$salary, na.rm = TRUE))
hist(Teacher$salary[Teacher$year == "1983-84"],
     main = "Teacher salary 1983-84", xlab = "salary",
     xlim = range(Teacher$salary, na.rm = TRUE))
hist(Teacher$salary[Teacher$year == "1993-94"],
     main = "Teacher salary 1993-94", xlab = "salary",
     xlim = range(Teacher$salary, na.rm = TRUE))
par(mfrow = c(1, 1))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Teacher, aes(x = salary)) +
  geom_histogram(fill = "purple", color = "black") +
  facet_grid(year ~ .) +
  theme_bw()
```

```
## End(Not run)
```

---

Tenness

*Tennessee self concept scores for 20 gifted high school students*

---

### Description

Data for Exercise 6.56

### Usage

Tenness

### Format

A data frame/tibble with 20 observations on one variable

**score** Tennessee Self-Concept Scale score

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
hist(Tenness$score, freq= FALSE, main = "", col = "green",
xlab = "Tennessee Self-Concept Scale score")
lines(density(Tenness$score))
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Tenness, aes(x = score, y = ..density..)) +
  geom_histogram(binwidth = 2, fill = "purple", color = "black") +
  geom_density(color = "red", fill = "pink", alpha = 0.3) +
  theme_bw()

## End(Not run)
```

---

Tensile

*Tensile strength of plastic bags from two production runs*

---

**Description**

Data for Example 7.11

**Usage**

Tensile

**Format**

A data frame/tibble with 72 observations on two variables

**tensile** plastic bag tensile strength (pounds per square inch)

**run** factor with run number (1 or 2)

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(tensile ~ run, data = Tensile,  
        col = c("purple", "cyan"))  
t.test(tensile ~ run, data = Tensile)
```

---

Test1

*Grades on the first test in a statistics class*

---

**Description**

Data for Exercise 5.80

**Usage**

Test1

**Format**

A data frame/tibble with 25 observations on one variable

**score** score on first statistics exam

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
stem(Test1$score)
boxplot(Test1$score, col = "purple")
```

---

Thermal

*Heat loss of thermal pane windows versus outside temperature*

---

## Description

Data for Example 9.5

## Usage

Thermal

## Format

A data frame/tibble with 12 observations on the two variables

**temp** temperature (degrees Celcius)

**loss** heat loss (BTUs)

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
model <- lm(loss ~ temp, data = Thermal)
summary(model)
plot(loss ~ temp, data = Thermal)
abline(model, col = "red")
rm(model)
```

---

Tiaa	<i>1999-2000 closing prices for TIAA-CREF stocks</i>
------	--

---

**Description**

Data for your enjoyment

**Usage**

Tiaa

**Format**

A data frame/tibble with 365 observations on four variables

**crefstk** closing price (in dollars)

**crefgwt** closing price (in dollars)

**tiaa** closing price (in dollars)

**date** day of the year

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
data(Tiaa)
```

---

Ticket	<i>Time to complete an airline ticket reservation</i>
--------	---

---

**Description**

Data for Exercise 5.18

**Usage**

Ticket

**Format**

A data frame/tibble with 20 observations on one variable

**time** time (in seconds) to check out a reservation

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
EDA(Ticket$time)
```

---

Toaster

*Consumer Reports (Oct 94) rating of toaster ovens versus the cost*

---

## Description

Data for Exercise 9.36

## Usage

Toaster

## Format

A data frame/tibble with 17 observations on three variables

**toaster** name of toaster

**score** Consumer Reports score

**cost** price of toaster (in dollars)

## Source

*Consumer Reports* (October 1994).

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(cost ~ score, data = Toaster)
model <- lm(cost ~ score, data = Toaster)
summary(model)
names(summary(model))
summary(model)$r.squared
plot(model, which = 1)
```

---

Tonsils

*Size of tonsils collected from 1,398 children*

---

## Description

Data for Exercise 2.78

## Usage

Tonsils

## Format

A data frame/tibble with 1,398 observations on two variables

**size** a factor with levels Normal, Large, and Very Large

**status** a factor with levels Carrier and Non-carrier

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
T1 <- xtabs(~size + status, data = Tonsils)
T1
prop.table(T1, 1)
prop.table(T1, 1)[2, 1]
barplot(t(T1), legend = TRUE, beside = TRUE, col = c("red", "green"))
## Not run:
library(dplyr)
library(ggplot2)
NDF <- dplyr::count(Tonsils, size, status)
ggplot2::ggplot(data = NDF, aes(x = size, y = n, fill = status)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("red", "green")) +
  theme_bw()

## End(Not run)
```

---

Tort	<i>The number of torts, average number of months to process a tort, and county population from the court files of the nation's largest counties</i>
------	---

---

**Description**

Data for Exercise 5.13

**Usage**

Tort

**Format**

A data frame/tibble with 45 observations on five variables

**county** U.S. county

**months** average number of months to process a tort

**population** population of the county

**torts** number of torts

**rate** rate per 10,000 residents

**Source**

U.S. Department of Justice, *Tort Cases in Large Counties*, Bureau of Justice Statistics Special Report, April 1995.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
EDA(Tort$months)
```



---

Toxic	<i>Hazardous waste sites near minority communities</i>
-------	--

---

**Description**

Data for Exercises 1.55, 5.08, 5.109, 8.58, and 10.35

**Usage**

Toxic

**Format**

A data frame/tibble with 51 observations on five variables

**state** U.S. state

**region** U.S. region

**sites** number of commercial hazardous waste sites

**minority** percent of minorities living in communities with commercial hazardous waste sites

**percent** a numeric vector

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
hist(Toxic$sites, col = "red")
hist(Toxic$minority, col = "blue")
qqnorm(Toxic$minority)
qqline(Toxic$minority)
boxplot(sites ~ region, data = Toxic, col = "lightgreen")
tapply(Toxic$sites, Toxic$region, median)
kruskal.test(sites ~ factor(region), data = Toxic)
```

---

Track

*National Olympic records for women in several races*

---

**Description**

Data for Exercises 2.97, 5.115, and 9.62

**Usage**

Track

**Format**

A data frame with 55 observations on eight variables

**country** athlete's country

**100m** time in seconds for 100 m

**200m** time in seconds for 200 m

**400m** time in seconds for 400 m

**800m** time in minutes for 800 m

**1500m** time in minutes for 1500 m

**3000m** time in minutes for 3000 m

**marathon** time in minutes for marathon

**Source**

Dawkins, B. (1989), "Multivariate Analysis of National Track Records," *The American Statistician*, 43(2), 110-115.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(`200m` ~ `100m`, data = Track)
plot(`400m` ~ `100m`, data = Track)
plot(`400m` ~ `200m`, data = Track)
cor(Track[, 2:8])
```

---

Track15

*Olympic winning times for the men's 1500-meter run*

---

**Description**

Data for Exercise 1.36

**Usage**

Track15

**Format**

A data frame/tibble with 26 observations on two variables

**year** Olympic year

**time** Olympic winning time (in seconds) for the 1500-meter run

**Source**

*The World Almanac and Book of Facts, 2000.*

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(time~ year, data = Track15, type = "b", pch = 19,  
      ylab = "1500m time in seconds", col = "green")
```

---

Treatments

*Illustrates analysis of variance for three treatment groups*

---

**Description**

Data for Exercise 10.44

**Usage**

Treatments

**Format**

A data frame/tibble with 24 observations on two variables

**score** score from an experiment

**group** factor with levels 1, 2, and 3

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(score ~ group, data = Treatments, col = "violet")
summary(aov(score ~ group, data = Treatments))
summary(lm(score ~ group, data = Treatments))
anova(lm(score ~ group, data = Treatments))
```

---

Trees

*Number of trees in 20 grids*

---

**Description**

Data for Exercise 1.50

**Usage**

Trees

**Format**

A data frame/tibble with 20 observations on one variable

**number** number of trees in a grid

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Trees$number)
hist(Trees$number, main = "Exercise 1.50", xlab = "number",
     col = "brown")
```

---

Trucks	<i>Miles per gallon for standard 4-wheel drive trucks manufactured by Chevrolet, Dodge and Ford</i>
--------	---

---

**Description**

Data for Example 10.2

**Usage**

Trucks

**Format**

A data frame/tibble with 15 observations on two variables

**mpg** miles per gallon

**truck** a factor with levels chevy, dodge, and ford

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(mpg ~ truck, data = Trucks, horizontal = TRUE, las = 1)
summary(aov(mpg ~ truck, data = Trucks))
```

---

tsum.test	<i>Summarized t-test</i>
-----------	--------------------------

---

**Description**

Performs a one-sample, two-sample, or a Welch modified two-sample t-test based on user supplied summary information. Output is identical to that produced with `t.test`.

**Usage**

```
tsum.test(
  mean.x,
  s.x = NULL,
  n.x = NULL,
  mean.y = NULL,
  s.y = NULL,
  n.y = NULL,
  alternative = "two.sided",
  mu = 0,
  var.equal = FALSE,
  conf.level = 0.95
)
```

**Arguments**

mean.x	a single number representing the sample mean of x
s.x	a single number representing the sample standard deviation for x
n.x	a single number representing the sample size for x
mean.y	a single number representing the sample mean of y
s.y	a single number representing the sample standard deviation for y
n.y	a single number representing the sample size for y
alternative	is a character string, one of "greater", "less" or "two.sided", or just the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard two-sample tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu. For the one-sample and paired t-tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard and Welch modified two-sample t-tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu. For the one-sample t-tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard and Welch modified two-sample t-tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu.
mu	is a single number representing the value of the mean or difference in means specified by the null hypothesis.
var.equal	logical flag: if TRUE, the variances of the parent populations of x and y are assumed equal. Argument var.equal should be supplied only for the two-sample tests.
conf.level	is the confidence level for the returned confidence interval; it must lie between zero and one.

**Details**

If `y` is `NULL`, a one-sample t-test is carried out with `x`. If `y` is not `NULL`, either a standard or Welch modified two-sample t-test is performed, depending on whether `var.equal` is `TRUE` or `FALSE`.

**Value**

A list of class `htest`, containing the following components:

<code>statistic</code>	the t-statistic, with names attribute <code>"t"</code>
<code>parameters</code>	is the degrees of freedom of the t-distribution associated with <code>statistic</code> . Component <code>parameters</code> has names attribute <code>"df"</code> .
<code>p.value</code>	the p-value for the test.
<code>conf.int</code>	is a confidence interval (vector of length 2) for the true mean or difference in means. The confidence level is recorded in the attribute <code>conf.level</code> . When <code>alternative</code> is not <code>"two.sided"</code> , the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values <code>k</code> for which one would not reject the null hypothesis that the true mean or difference in means is <code>k</code> . Here infinity will be represented by <code>Inf</code> .
<code>estimate</code>	vector of length 1 or 2, giving the sample mean(s) or mean of differences; these estimate the corresponding population parameters. Component <code>estimate</code> has a names attribute describing its elements.
<code>null.value</code>	the value of the mean or difference in means specified by the null hypothesis. This equals the input argument <code>mu</code> . Component <code>null.value</code> has a names attribute describing its elements.
<code>alternative</code>	records the value of the input argument <code>alternative</code> : <code>"greater"</code> , <code>"less"</code> or <code>"two.sided"</code> .
<code>data.name</code>	a character string (vector of length 1) containing the names <code>x</code> and <code>y</code> for the two summarized samples.

**Null Hypothesis**

For the one-sample t-test, the null hypothesis is that the mean of the population from which `x` is drawn is `mu`. For the standard and Welch modified two-sample t-tests, the null hypothesis is that the population mean for `x` less that for `y` is `mu`.

The alternative hypothesis in each case indicates the direction of divergence of the population mean for `x` (or difference of means for `x` and `y`) from `mu` (i.e., `"greater"`, `"less"`, or `"two.sided"`).

**Author(s)**

Alan T. Arnholt

**References**

- Kitchens, L.J. (2003). *Basic Statistics and Data Analysis*. Duxbury.
- Hogg, R. V. and Craig, A. T. (1970). *Introduction to Mathematical Statistics, 3rd ed.* Toronto, Canada: Macmillan.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics, 3rd ed.* New York: McGraw-Hill.

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods, 7th ed.* Ames, Iowa: Iowa State University Press.

### See Also

[z.test](#), [zsum.test](#)

### Examples

```
tsum.test(mean.x=5.6, s.x=2.1, n.x=16, mu=4.9, alternative="greater")
# Problem 6.31 on page 324 of BSDA states: The chamber of commerce
# of a particular city claims that the mean carbon dioxide
# level of air pollution is no greater than 4.9 ppm. A random
# sample of 16 readings resulted in a sample mean of 5.6 ppm,
# and s=2.1 ppm. One-sided one-sample t-test. The null
# hypothesis is that the population mean for 'x' is 4.9.
# The alternative hypothesis states that it is greater than 4.9.

x <- rnorm(12)
tsum.test(mean(x), sd(x), n.x=12)
# Two-sided one-sample t-test. The null hypothesis is that
# the population mean for 'x' is zero. The alternative
# hypothesis states that it is either greater or less
# than zero. A confidence interval for the population mean
# will be computed. Note: above returns same answer as:
t.test(x)

x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7.0, 6.4, 7.1, 6.7, 7.6, 6.8)
y <- c(4.5, 5.4, 6.1, 6.1, 5.4, 5.0, 4.1, 5.5)
tsum.test(mean(x), s.x=sd(x), n.x=11, mean(y), s.y=sd(y), n.y=8, mu=2)
# Two-sided standard two-sample t-test. The null hypothesis
# is that the population mean for 'x' less that for 'y' is 2.
# The alternative hypothesis is that this difference is not 2.
# A confidence interval for the true difference will be computed.
# Note: above returns same answer as:
t.test(x, y)

tsum.test(mean(x), s.x=sd(x), n.x=11, mean(y), s.y=sd(y), n.y=8, conf.level=0.90)
# Two-sided standard two-sample t-test. The null hypothesis
# is that the population mean for 'x' less that for 'y' is zero.
# The alternative hypothesis is that this difference is not
# zero. A 90% confidence interval for the true difference will
# be computed. Note: above returns same answer as:
t.test(x, y, conf.level=0.90)
```



---

Tv	<i>Percent of students that watch more than 6 hours of TV per day versus national math test scores</i>
----	--

---

**Description**

Data for Examples 2.1 and 2.7

**Usage**

Tv

**Format**

A data frame/tibble with 53 observations on three variables

**state** U.S. state

**percent** percent of students who watch more than six hours of TV a day

**test** state average on national math test

**Source**

Educational Testing Services.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(test ~ percent, data = Tv, col = "blue")
cor(Tv$test, Tv$percent)
```

---

Twin	<i>Intelligence test scores for identical twins in which one twin is given a drug</i>
------	---

---

**Description**

Data for Exercise 7.54

**Usage**

Twin

**Format**

A data frame/tibble with nine observations on three variables

**twinA** score on intelligence test without drug

**twinB** score on intelligence test after taking drug

**differ** twinA - twinB

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
qqnorm(Twin$differ)
qqline(Twin$differ)
shapiro.test(Twin$differ)
t.test(Twin$twinA, Twin$twinB, paired = TRUE)
```

---

Undergrad

*Data set describing a sample of undergraduate students*

---

**Description**

Data for Exercise 1.15

**Usage**

Undergrad

**Format**

A data frame/tibble with 100 observations on six variables

**gender** character variable with values Female and Male

**major** college major

**class** college year group classification

**gpa** grade point average

**sat** Scholastic Assessment Test score

**drops** number of courses dropped

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```

stripchart(gpa ~ class, data = Undergrad, method = "stack",
col = c("blue","red","green","lightblue"),
pch = 19, main = "GPA versus Class")
stripchart(gpa ~ gender, data = Undergrad, method = "stack",
col = c("red", "blue"), pch = 19,
main = "GPA versus Gender")
stripchart(sat ~ drops, data = Undergrad, method = "stack",
col = c("blue", "red", "green", "lightblue"),
pch = 19, main = "SAT versus Drops")
stripchart(drops ~ gender, data = Undergrad, method = "stack",
col = c("red", "blue"), pch = 19, main = "Drops versus Gender")

## Not run:
library(ggplot2)
ggplot2::ggplot(data = Undergrad, aes(x = sat, y = drops, fill = factor(drops))) +
  facet_grid(drops ~.) +
  geom_dotplot() +
  guides(fill = FALSE)

## End(Not run)

```

---

Vacation	<i>Number of days of paid holidays and vacation leave for sample of 35 textile workers</i>
----------	--

---

**Description**

Data for Exercise 6.46 and 6.98

**Usage**

Vacation

**Format**

A data frame/tibble with 35 observations on one variable

**number** number of days of paid holidays and vacation leave taken

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
boxplot(Vacation$number, col = "violet")
hist(Vacation$number, main = "Exercise 6.46", col = "blue",
     xlab = "number of days of paid holidays and vacation leave taken")
t.test(Vacation$number, mu = 24)
```

---

Vaccine

*Reported serious reactions due to vaccines in 11 southern states*

---

**Description**

Data for Exercise 1.111

**Usage**

Vaccine

**Format**

A data frame/tibble with 11 observations on two variables

**state** U.S. state

**number** number of reported serious reactions per million doses of a vaccine

**Source**

Center for Disease Control, Atlanta, Georgia.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Vaccine$number, scale = 2)
fn <- fivenum(Vaccine$number)
fn
iqr <- IQR(Vaccine$number)
iqr
```

---

Vehicle	<i>Fatality ratings for foreign and domestic vehicles</i>
---------	---

---

**Description**

Data for Exercise 8.34

**Usage**

Vehicle

**Format**

A data frame/tibble with 151 observations on two variables

**make** a factor with levels domestic and foreign

**rating** a factor with levels Much better than average, Above average, Average, Below average, and Much worse than average

**Source**

Insurance Institute for Highway Safety and the Highway Loss Data Institute, 1995.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
T1 <- xtabs(~make + rating, data = Vehicle)
T1
chisq.test(T1)
```

---

Verbal	<i>Verbal test scores and number of library books checked out for 15 eighth graders</i>
--------	---

---

**Description**

Data for Exercise 9.30

**Usage**

Verbal

**Format**

A data frame/tibble with 15 observations on two variables

**number** number of library books checked out

**verbal** verbal test score

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(verbal ~ number, data = Verbal)
abline(lm(verbal ~ number, data = Verbal), col = "red")
summary(lm(verbal ~ number, data = Verbal))
```

---

Victoria	<i>Number of sunspots versus mean annual level of Lake Victoria Nyanza from 1902 to 1921</i>
----------	--

---

**Description**

Data for Exercise 2.98

**Usage**

Victoria

**Format**

A data frame/tibble with 20 observations on three variables

**year** year

**level** mean annual level of Lake Victoria Nyanza

**sunspot** number of sunspots

**Source**

N. Shaw, *Manual of Meteorology*, Vol. 1 (London: Cambridge University Press, 1942), p. 284; and F. Mosteller and J. W. Tukey, *Data Analysis and Regression* (Reading, MA: Addison-Wesley, 1977).

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(level ~ sunspot, data = Victoria)
model <- lm(level ~ sunspot, data = Victoria)
summary(model)
rm(model)
```

---

Viscosit

*Viscosity measurements of a substance on two different days*

---

## Description

Data for Exercise 7.44

## Usage

```
Viscosit
```

## Format

A data frame/tibble with 11 observations on two variables

**first** viscosity measurement for a certain substance on day one

**second** viscosity measurement for a certain substance on day two

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(Viscosit$first, Viscosit$second, col = "blue")
t.test(Viscosit$first, Viscosit$second, var.equal = TRUE)
```

---

Visual	<i>Visual acuity of a group of subjects tested under a specified dose of a drug</i>
--------	---

---

**Description**

Data for Exercise 5.6

**Usage**

Visual

**Format**

A data frame/tibble with 18 observations on one variable

**visual** visual acuity measurement

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
stem(Visual$visual)
boxplot(Visual$visual, col = "purple")
```

---

Vocab	<i>Reading scores before and after vocabulary training for 14 employees who did not complete high school</i>
-------	--

---

**Description**

Data for Exercise 7.80

**Usage**

Vocab

**Format**

A data frame/tibble with 14 observations on two variables

**first** reading test score before formal vocabulary training

**second** reading test score after formal vocabulary training



## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
t.test(Vocab$first, Vocab$second, paired = TRUE)
```

---

Wastewat	<i>Volume of injected waste water from Rocky Mountain Arsenal and number of earthquakes near Denver</i>
----------	---

---

## Description

Data for Exercise 9.18

## Usage

Wastewat

## Format

A data frame/tibble with 44 observations on two variables

**gallons** injected water (in million gallons)

**number** number of earthquakes detected in Denver

## Source

Davis, J. C. (1986), *Statistics and Data Analysis in Geology*, 2 ed., John Wiley and Sons, New York, p. 228, and Bardwell, G. E. (1970), Some Statistical Features of the Relationship between Rocky Mountain Arsenal Waste Disposal and Frequency of Earthquakes, *Geological Society of America, Engineering Geology Case Histories*, 8, 33-337.

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(number ~ gallons, data = Wastewat)
model <- lm(number ~ gallons, data = Wastewat)
summary(model)
anova(model)
plot(model, which = 2)
```

---

Weather94

*Weather casualties in 1994*

---

### Description

Data for Exercise 1.30

### Usage

Weather94

### Format

A data frame/tibble with 388 observations on one variable

**type** factor with levels Extreme Temp, Flash Flood, Fog, High Wind, Hurricane, Lighting, Other, River Flood, Thunderstorm, Tornado, and Winter Weather

### References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

### Examples

```
T1 <- xtabs(~type, data = Weather94)
T1
par(mar = c(5.1 + 2, 4.1 - 1, 4.1 - 2, 2.1))
barplot(sort(T1, decreasing = TRUE), las = 2, col = rainbow(11))
par(mar = c(5.1, 4.1, 4.1, 2.1))
## Not run:
library(ggplot2)
T2 <- as.data.frame(T1)
T2
ggplot2::ggplot(data = T2, aes(x = reorder(type, Freq), y = Freq)) +
  geom_bar(stat = "identity", fill = "purple") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 55, vjust = 0.5)) +
  labs(x = "", y = "count")

## End(Not run)
```

---

Wheat	<i>Price of a bushel of wheat versus the national weekly earnings of production workers</i>
-------	---

---

**Description**

Data for Exercise 2.11

**Usage**

Wheat

**Format**

A data frame/tibble with 19 observations on three variables

**year** year

**earnings** national weekly earnings (in dollars) for production workers

**price** price for a bushel of wheat (in dollars)

**Source**

*The World Almanac and Book of Facts*, 2000.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
par(mfrow = c(1, 2))
plot(earnings ~ year, data = Wheat)
plot(price ~ year, data = Wheat)
par(mfrow = c(1, 1))
```

---

Windmill

*Direct current produced by different wind velocities*

---

**Description**

Data for Exercise 9.34

**Usage**

Windmill

**Format**

A data frame/tibble with 25 observations on two variables

**velocity** wind velocity (miles per hour)

**output** power generated (DC volts)

**Source**

Joglekar, et al. (1989), Lack of Fit Testing when Replicates Are Not Available, *The American Statistician*, 43,(3), 135-143.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
summary(lm(output ~ velocity, data = Windmill))
anova(lm(output ~ velocity, data = Windmill))
```

---

Window

*Wind leakage for storm windows exposed to a 50 mph wind*

---

**Description**

Data for Exercise 6.54

**Usage**

Window

**Format**

A data frame/tibble with nine observations on two variables

**window** window number

**leakage** percent leakage from a 50 mph wind

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
SIGN.test(Window$leakage, md = 0.125, alternative = "greater")
```

---

Wins	<i>Baseball team wins versus seven independent variables for National league teams in 1990</i>
------	--

---

**Description**

Data for Exercise 9.23

**Usage**

Wins

**Format**

A data frame with 12 observations on nine variables

**team** name of team

**wins** number of wins

**batavg** batting average

**rbi** runs batted in

**stole** bases stole

**strkout** number of strikeouts

**caught** number of times caught stealing

**errors** number of errors

**era** earned run average

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
plot(wins ~ era, data = Wins)
## Not run:
library(ggplot2)
ggplot2::ggplot(data = Wins, aes(x = era, y = wins)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw()

## End(Not run)
```

---

Wool

*Strength tests of two types of wool fabric*

---

## Description

Data for Exercise 7.42

## Usage

Wool

## Format

A data frame/tibble with 20 observations on two variables

**type** type of wool (Type I, Type 2)

**strength** strength of wool

## References

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

## Examples

```
boxplot(strength ~ type, data = Wool, col = c("blue", "purple"))
t.test(strength ~ type, data = Wool, var.equal = TRUE)
```

---

Yearsunspot

*Monthly sunspot activity from 1974 to 2000*

---

**Description**

Data for Exercise 2.7

**Usage**

Yearsunspot

**Format**

A data frame/tibble with 252 observations on two variables

**number** average number of sunspots

**year** date

**Source**

NASA/Marshall Space Flight Center, Huntsville, AL 35812.

**References**

Kitchens, L. J. (2003) *Basic Statistics and Data Analysis*. Pacific Grove, CA: Brooks/Cole, a division of Thomson Learning.

**Examples**

```
plot(number ~ year, data = Yearsunspot)
```

---

z.test

*Z-test*

---

**Description**

This function is based on the standard normal distribution and creates confidence intervals and tests hypotheses for both one and two sample problems.

**Usage**

```
z.test(
  x,
  y = NULL,
  alternative = "two.sided",
  mu = 0,
  sigma.x = NULL,
  sigma.y = NULL,
  conf.level = 0.95
)
```

**Arguments**

x	numeric vector; NAs and Infs are allowed but will be removed.
y	numeric vector; NAs and Infs are allowed but will be removed.
alternative	character string, one of "greater", "less" or "two.sided", or the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, alternative refers to the true mean of the parent population in relation to the hypothesized value mu. For the standard two-sample tests, alternative refers to the difference between the true population mean for x and that for y, in relation to mu.
mu	a single number representing the value of the mean or difference in means specified by the null hypothesis
sigma.x	a single number representing the population standard deviation for x
sigma.y	a single number representing the population standard deviation for y
conf.level	confidence level for the returned confidence interval, restricted to lie between zero and one

**Details**

If y is NULL, a one-sample z-test is carried out with x. If y is not NULL, a standard two-sample z-test is performed.

**Value**

A list of class `htest`, containing the following components:

statistic	the z-statistic, with names attribute "z"
p.value	the p-value for the test
conf.int	is a confidence interval (vector of length 2) for the true mean or difference in means. The confidence level is recorded in the attribute <code>conf.level</code> . When alternative is not "two.sided", the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values k for which one would not reject the null hypothesis that the true mean or difference in means is k. Here infinity will be represented by <code>Inf</code> .



estimate	vector of length 1 or 2, giving the sample mean(s) or mean of differences; these estimate the corresponding population parameters. Component estimate has a names attribute describing its elements.
null.value	is the value of the mean or difference in means specified by the null hypothesis. This equals the input argument mu. Component null.value has a names attribute describing its elements.
alternative	records the value of the input argument alternative: "greater", "less" or "two.sided".
data.name	a character string (vector of length 1) containing the actual names of the input vectors x and y

### Null Hypothesis

For the one-sample z-test, the null hypothesis is that the mean of the population from which x is drawn is  $\mu$ . For the standard two-sample z-tests, the null hypothesis is that the population mean for x less that for y is  $\mu$ .

The alternative hypothesis in each case indicates the direction of divergence of the population mean for x (or difference of means for x and y) from  $\mu$  (i.e., "greater", "less", "two.sided").

### Author(s)

Alan T. Arnholt

### References

- Kitchens, L.J. (2003). *Basic Statistics and Data Analysis*. Duxbury.
- Hogg, R. V. and Craig, A. T. (1970). *Introduction to Mathematical Statistics, 3rd ed.* Toronto, Canada: Macmillan.
- Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics, 3rd ed.* New York: McGraw-Hill.
- Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods, 7th ed.* Ames, Iowa: Iowa State University Press.

### See Also

[zsum.test](#), [tsum.test](#)

### Examples

```
x <- rnorm(12)
z.test(x, sigma.x=1)
# Two-sided one-sample z-test where the assumed value for
# sigma.x is one. The null hypothesis is that the population
# mean for 'x' is zero. The alternative hypothesis states
# that it is either greater or less than zero. A confidence
# interval for the population mean will be computed.
```

```

x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7., 6.4, 7.1, 6.7, 7.6, 6.8)
y <- c(4.5, 5.4, 6.1, 6.1, 5.4, 5., 4.1, 5.5)
z.test(x, sigma.x=0.5, y, sigma.y=0.5, mu=2)
  # Two-sided standard two-sample z-test where both sigma.x
  # and sigma.y are both assumed to equal 0.5. The null hypothesis
  # is that the population mean for 'x' less that for 'y' is 2.
  # The alternative hypothesis is that this difference is not 2.
  # A confidence interval for the true difference will be computed.

z.test(x, sigma.x=0.5, y, sigma.y=0.5, conf.level=0.90)
  # Two-sided standard two-sample z-test where both sigma.x and
  # sigma.y are both assumed to equal 0.5. The null hypothesis
  # is that the population mean for 'x' less that for 'y' is zero.
  # The alternative hypothesis is that this difference is not
  # zero. A 90% confidence interval for the true difference will
  # be computed.
rm(x, y)

```

---

zsum.test

*Summarized z-test*


---

### Description

This function is based on the standard normal distribution and creates confidence intervals and tests hypotheses for both one and two sample problems based on summarized information the user passes to the function. Output is identical to that produced with `z.test`.

### Usage

```

zsum.test(
  mean.x,
  sigma.x = NULL,
  n.x = NULL,
  mean.y = NULL,
  sigma.y = NULL,
  n.y = NULL,
  alternative = "two.sided",
  mu = 0,
  conf.level = 0.95
)

```

### Arguments

<code>mean.x</code>	a single number representing the sample mean of x
<code>sigma.x</code>	a single number representing the population standard deviation for x
<code>n.x</code>	a single number representing the sample size for x
<code>mean.y</code>	a single number representing the sample mean of y

<code>sigma.y</code>	a single number representing the population standard deviation for <code>y</code>
<code>n.y</code>	a single number representing the sample size for <code>y</code>
<code>alternative</code>	is a character string, one of "greater", "less" or "two.sided", or the initial letter of each, indicating the specification of the alternative hypothesis. For one-sample tests, <code>alternative</code> refers to the true mean of the parent population in relation to the hypothesized value $\mu$ . For the standard two-sample tests, <code>alternative</code> refers to the difference between the true population mean for <code>x</code> and that for <code>y</code> , in relation to $\mu$ .
<code>mu</code>	a single number representing the value of the mean or difference in means specified by the null hypothesis
<code>conf.level</code>	confidence level for the returned confidence interval, restricted to lie between zero and one

### Details

If `y` is `NULL`, a one-sample z-test is carried out with `x`. If `y` is not `NULL`, a standard two-sample z-test is performed.

### Value

A list of class `htest`, containing the following components:

<code>statistic</code>	the z-statistic, with names attribute <code>z</code> .
<code>p.value</code>	the p-value for the test
<code>conf.int</code>	is a confidence interval (vector of length 2) for the true mean or difference in means. The confidence level is recorded in the attribute <code>conf.level</code> . When <code>alternative</code> is not "two.sided", the confidence interval will be half-infinite, to reflect the interpretation of a confidence interval as the set of all values $k$ for which one would not reject the null hypothesis that the true mean or difference in means is $k$ . Here, infinity will be represented by <code>Inf</code> .
<code>estimate</code>	vector of length 1 or 2, giving the sample mean(s) or mean of differences; these estimate the corresponding population parameters. Component <code>estimate</code> has a names attribute describing its elements.
<code>null.value</code>	the value of the mean or difference in means specified by the null hypothesis. This equals the input argument <code>mu</code> . Component <code>null.value</code> has a names attribute describing its elements.
<code>alternative</code>	records the value of the input argument <code>alternative</code> : "greater", "less" or "two.sided".
<code>data.name</code>	a character string (vector of length 1) containing the names <code>x</code> and <code>y</code> for the two summarized samples

### Null Hypothesis

For the one-sample z-test, the null hypothesis is that the mean of the population from which `x` is drawn is  $\mu$ . For the standard two-sample z-tests, the null hypothesis is that the population mean for `x` less that for `y` is  $\mu$ .

The alternative hypothesis in each case indicates the direction of divergence of the population mean for `x` (or difference of means of `x` and `y`) from  $\mu$  (i.e., "greater", "less", "two.sided").

**Author(s)**

Alan T. Arnholt

**References**

- Kitchens, L. J. (2003). *Basic Statistics and Data Analysis*. Duxbury.
- Hogg, R. V. and Craig, A. T. (1970). *Introduction to Mathematical Statistics, 3rd ed.* Toronto, Canada: Macmillan.
- Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics, 3rd ed.* New York: McGraw-Hill.
- Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods, 7th ed.* Ames, Iowa: Iowa State University Press.

**See Also**

[z.test](#), [tsum.test](#)

**Examples**

```
zsum.test(mean.x=56/30,sigma.x=2, n.x=30, alternative="greater", mu=1.8)
# Example 9.7 part a. from PASWR.
x <- rnorm(12)
zsum.test(mean(x),sigma.x=1,n.x=12)
# Two-sided one-sample z-test where the assumed value for
# sigma.x is one. The null hypothesis is that the population
# mean for 'x' is zero. The alternative hypothesis states
# that it is either greater or less than zero. A confidence
# interval for the population mean will be computed.
# Note: returns same answer as:
z.test(x,sigma.x=1)
#
x <- c(7.8, 6.6, 6.5, 7.4, 7.3, 7.0, 6.4, 7.1, 6.7, 7.6, 6.8)
y <- c(4.5, 5.4, 6.1, 6.1, 5.4, 5.0, 4.1, 5.5)
zsum.test(mean(x), sigma.x=0.5, n.x=11 ,mean(y), sigma.y=0.5, n.y=8, mu=2)
# Two-sided standard two-sample z-test where both sigma.x
# and sigma.y are both assumed to equal 0.5. The null hypothesis
# is that the population mean for 'x' less that for 'y' is 2.
# The alternative hypothesis is that this difference is not 2.
# A confidence interval for the true difference will be computed.
# Note: returns same answer as:
z.test(x, sigma.x=0.5, y, sigma.y=0.5)
#
zsum.test(mean(x), sigma.x=0.5, n.x=11, mean(y), sigma.y=0.5, n.y=8,
conf.level=0.90)
# Two-sided standard two-sample z-test where both sigma.x and
# sigma.y are both assumed to equal 0.5. The null hypothesis
# is that the population mean for 'x' less that for 'y' is zero.
# The alternative hypothesis is that this difference is not
# zero. A 90% confidence interval for the true difference will
```

```
      # be computed. Note: returns same answer as:  
z.test(x, sigma.x=0.5, y, sigma.y=0.5, conf.level=0.90)  
rm(x, y)
```

# Index

## \* datasets

Abbey, 8  
Abc, 9  
Abilene, 10  
Ability, 11  
Abortion, 12  
Absent, 13  
Achieve, 14  
Adsales, 14  
Aggress, 15  
Aid, 16  
Aids, 17  
Airdisasters, 18  
Airline, 19  
Alcohol, 20  
Allergy, 20  
Anesthet, 21  
Anxiety, 22  
Apolipop, 22  
Append, 23  
Appendec, 24  
Aptitude, 25  
Archaeo, 25  
Arthriti, 26  
Artifici, 27  
Asprin, 28  
Asthmati, 28  
Attorney, 29  
Autogear, 30  
Backtoback, 31  
Bbsalaries, 31  
Bigten, 32  
Biology, 33  
Birth, 34  
Blackedu, 35  
Blood, 35  
Board, 36  
Bones, 37  
Books, 38  
Bookstor, 38  
Brain, 39  
Bumpers, 40  
Bus, 41  
Bypass, 41  
Cabinets, 42  
Cancer, 43  
Carbon, 44  
Cat, 45  
Censored, 45  
Challeng, 46  
Chemist, 47  
Chesapea, 48  
Chevy, 49  
Chicken, 50  
Chipavg, 50  
Chips, 51  
Cigar, 52  
Cigaret, 53  
Citrus, 55  
Clean, 56  
Coaxial, 57  
Coffee, 58  
Coins, 58  
Commute, 60  
Concept, 61  
Concrete, 61  
Corn, 62  
Correlat, 63  
Counsel, 64  
Cpi, 64  
Crime, 65  
Darwin, 66  
Dealers, 67  
Defectiv, 67  
Degree, 68  
Delay, 69  
Depend, 70  
Detroit, 70

Develop, 71  
Devmath, 72  
Dice, 72  
Diesel, 73  
Diplomat, 74  
Disposal, 75  
Dogs, 76  
Domestic, 77  
Dopamine, 78  
Dowjones, 79  
Drink, 80  
Drug, 80  
Dyslexia, 81  
Earthqk, 82  
Educat, 84  
Eggs, 85  
Elderly, 85  
Energy, 86  
Engineer, 87  
Entrance, 88  
Epaminicompat, 89  
Epatwoseater, 90  
Executiv, 91  
Exercise, 91  
Fabric, 92  
Faithful, 93  
Family, 94  
Ferraro1, 95  
Ferraro2, 95  
Fertility, 96  
Firstchi, 97  
Fish, 98  
Fitness, 99  
Florida2000, 100  
Fluid, 101  
Food, 102  
Framingh, 102  
Freshman, 103  
Funeral, 104  
Galaxie, 105  
Gallup, 105  
Gasoline, 106  
German, 107  
Golf, 108  
Governor, 109  
Gpa, 110  
Grades, 111  
Graduate, 112  
Greenriv, 112  
Grnriv2, 113  
Groupabc, 114  
Groups, 114  
Gym, 115  
Habits, 116  
Haptoglo, 117  
Hardware, 117  
Hardwood, 118  
Heat, 119  
Heating, 120  
Hodgkin, 121  
Homes, 122  
Homework, 123  
Honda, 124  
Hostile, 124  
Housing, 125  
Hurrican, 126  
Iceberg, 127  
Income, 128  
Independent, 129  
Indian, 130  
Indiapol, 131  
Indy500, 131  
Inflatio, 132  
Inletoil, 133  
Inmate, 134  
Inspect, 135  
Insulate, 136  
Iqgpa, 137  
Iris, 137  
Jdpower, 138  
Jobsat, 139  
Kidsmoke, 140  
Kilowatt, 141  
Kinder, 141  
Laminect, 142  
Lead, 143  
Leader, 144  
Lethal, 144  
Life, 145  
Lifespan, 146  
Lightmonth, 146  
Lodge, 147  
Longtail, 148  
Lowabil, 149  
Magnesiu, 149  
Malpract, 150

- Manager, 151
- Marked, 151
- Math, 152
- Mathcomp, 153
- Mathpro, 154
- Maze, 155
- Median, 155
- Mental, 156
- Mercury, 157
- Metrent, 157
- Miller, 158
- Miller1, 159
- Moisture, 159
- Monoxide, 160
- Movie, 161
- Music, 162
- Name, 163
- Nascar, 164
- Nervous, 164
- Newsstand, 165
- Nfldraf2, 166
- Nfldraft, 166
- Nicotine, 167
- Orange, 171
- Orioles, 171
- Oxytocin, 172
- Parented, 173
- Patrol, 174
- Pearson, 175
- Phone, 175
- Poison, 176
- Politic, 177
- Pollutio, 178
- Porosity, 178
- Poverty, 179
- Precinct, 180
- Prejudic, 181
- Presiden, 181
- Press, 182
- Prognost, 183
- Program, 184
- Psat, 184
- Psych, 185
- Puerto, 186
- Quail, 186
- Quality, 187
- Rainks, 188
- Randd, 189
- Rat, 189
- Ratings, 190
- Reaction, 191
- Reading, 192
- Readiq, 192
- Referend, 193
- Region, 194
- Register, 195
- Rehab, 195
- Remedial, 196
- Rentals, 197
- Repair, 198
- Retail, 198
- Ronbrown1, 199
- Ronbrown2, 200
- Rural, 200
- Salary, 201
- Salinity, 202
- Sat, 203
- Saving, 204
- Scales, 204
- Schizop2, 205
- Schizoph, 206
- Seatbelt, 207
- Selfdefe, 208
- Senior, 208
- Sentence, 209
- Shkdrug, 210
- Shock, 211
- Shoplift, 211
- Short, 212
- Shuttle, 213
- Simpson, 216
- Situp, 217
- Skewed, 218
- Skin, 218
- Slc, 219
- Smokyph, 220
- Snore, 221
- Snow, 222
- Soccer, 223
- Social, 223
- Sophomor, 224
- South, 225
- Speed, 225
- Spellers, 226
- Spelling, 227
- Sports, 227



- Spouse, 228
- Stable, 230
- Stamp, 230
- Statclas, 231
- Statelaw, 232
- Statisti, 232
- Step, 233
- Stress, 234
- Study, 235
- Submarin, 235
- Subway, 236
- Sunspot, 237
- Superbowl, 238
- Supercar, 238
- Tablrock, 239
- Teacher, 241
- Tenness, 242
- Tensile, 243
- Test1, 243
- Thermal, 244
- Tiaa, 245
- Ticket, 245
- Toaster, 246
- Tonsils, 247
- Tort, 248
- Toxic, 249
- Track, 250
- Track15, 251
- Treatments, 251
- Trees, 252
- Trucks, 253
- Tv, 257
- Twin, 257
- Undergrad, 258
- Vacation, 259
- Vaccine, 260
- Vehicle, 261
- Verbal, 261
- Victoria, 262
- Viscosit, 263
- Visual, 264
- Vocab, 264
- Wastewat, 265
- Weather94, 266
- Wheat, 267
- Windmill, 268
- Window, 268
- Wins, 269
- Wool, 270
- Yearsunspot, 271
- \* **distribution**
  - CIsim, 54
  - Combinations, 59
  - normarea, 168
  - ntester, 170
  - SRS, 229
- \* **htest**
  - tsum.test, 253
  - z.test, 271
  - zsum.test, 274
- \* **univar**
  - EDA, 83
  - nsize, 169
- Abbey, 8
- Abc, 9
- Abilene, 10
- Ability, 11
- Abortion, 12
- Absent, 13
- Achieve, 14
- Adsales, 14
- Aggress, 15
- Aid, 16
- Aids, 17
- Airdisasters, 18
- Airline, 19
- Alcohol, 20
- Allergy, 20
- Anesthet, 21
- Anxiety, 22
- Apolipop, 22
- Append, 23
- Appendec, 24
- Aptitude, 25
- Archaeo, 25
- Arthriti, 26
- Artifici, 27
- Asprin, 28
- Asthmati, 28
- Attorney, 29
- Autogear, 30
- Backtoback, 31
- Bbsalaries, 31
- Bigten, 32
- Biology, 33

- Birth, 34  
Blackedu, 35  
Blood, 35  
Board, 36  
Bones, 37  
Books, 38  
Bookstor, 38  
Brain, 39  
Bumpers, 40  
Bus, 41  
Bypass, 41
- Cabinets, 42  
Cancer, 43  
Carbon, 44  
Cat, 45  
Censored, 45  
Challeng, 46  
Chemist, 47  
Chesapea, 48  
Chevy, 49  
Chicken, 50  
Chipavg, 50  
Chips, 51  
Cigar, 52  
Cigarette, 53  
CIsim, 54  
Citrus, 55  
Clean, 56  
Coaxial, 57  
Coffee, 58  
Coins, 58  
Combinations, 59, 229  
Commute, 60  
Concept, 61  
Concrete, 61  
Corn, 62  
Correlat, 63  
Counsel, 64  
Cpi, 64  
Crime, 65
- Darwin, 66  
Dealers, 67  
Defectiv, 67  
Degree, 68  
Delay, 69  
Depend, 70  
Detroit, 70
- Develop, 71  
Devmath, 72  
Dice, 72  
Diesel, 73  
Diplomat, 74  
Disposal, 75  
Dogs, 76  
Domestic, 77  
Dopamine, 78  
Dowjones, 79  
Drink, 80  
Drug, 80  
Dyslexia, 81
- Earthqk, 82  
EDA, 83  
Educat, 84  
Eggs, 85  
Elderly, 85  
Energy, 86  
Engineer, 87  
Entrance, 88  
Epaminicompact, 89  
Epatwoseater, 90  
Executiv, 91  
Exercise, 91
- Fabric, 92  
Faithful, 93  
Family, 94  
Ferraro1, 95  
Ferraro2, 95  
Fertility, 96  
Firstchi, 97  
Fish, 98  
Fitness, 99  
Florida2000, 100  
Fluid, 101  
Food, 102  
Framingh, 102  
Freshman, 103  
Funeral, 104
- Galaxie, 105  
Gallup, 105  
Gasoline, 106  
German, 107  
Golf, 108  
Governor, 109

- Gpa, 110
- Grades, 111
- Graduate, 112
- Greenriv, 112
- Gnriv2, 113
- Groupabc, 114
- Groups, 114
- Gym, 115
  
- Habits, 116
- Haptoglo, 117
- Hardware, 117
- Hardwood, 118
- Heat, 119
- Heating, 120
- Hodgkin, 121
- Homes, 122
- Homework, 123
- Honda, 124
- Hostile, 124
- Housing, 125
- Hurrican, 126
  
- Iceberg, 127
- Income, 128
- Independent, 129
- Indian, 130
- Indiapol, 131
- Indy500, 131
- Inflatio, 132
- Inletoil, 133
- Inmate, 134
- Inspect, 135
- Insulate, 136
- Iqgpa, 137
- Irises, 137
  
- Jdpower, 138
- Jobsat, 139
  
- Kidsmoke, 140
- Kilowatt, 141
- Kinder, 141
  
- Laminect, 142
- Lead, 143
- Leader, 144
- Lethal, 144
- Life, 145
  
- Lifespan, 146
- Ligntmonth, 146
- Lodge, 147
- Longtail, 148
- Lowabil, 149
  
- Magnesi, 149
- Malpract, 150
- Manager, 151
- Marked, 151
- Math, 152
- Mathcomp, 153
- Mathpro, 154
- Maze, 155
- Median, 155
- Mental, 156
- Mercury, 157
- Metrent, 157
- Miller, 158
- Miller1, 159
- Moisture, 159
- Monoxide, 160
- Movie, 161
- Music, 162
  
- Name, 163
- Nascar, 164
- Nervous, 164
- Newsstand, 165
- Nfldraf2, 166
- Nfldraft, 166
- Nicotine, 167
- normarea, 168
- nsize, 169
- ntester, 170
  
- Orange, 171
- Orioles, 171
- Oxytocin, 172
  
- Parented, 173
- Patrol, 174
- Pearson, 175
- Phone, 175
- Poison, 176
- Politic, 177
- Pollutio, 178
- Porosity, 178
- Poverty, 179

- Precinct, 180  
Prejudic, 181  
Presiden, 181  
Press, 182  
Prognost, 183  
Program, 184  
Pstat, 184  
Psych, 185  
Puerto, 186
- Quail, 186  
Quality, 187
- Rainks, 188  
Randd, 189  
Rat, 189  
Ratings, 190  
Reaction, 191  
Reading, 192  
Readiq, 192  
Referend, 193  
Region, 194  
Register, 195  
Rehab, 195  
Remedial, 196  
Rentals, 197  
Repair, 198  
Retail, 198  
Ronbrown1, 199  
Ronbrown2, 200  
Rural, 200
- Salary, 201  
Salinity, 202  
Sat, 203  
Saving, 204  
Scales, 204  
Schizop2, 205  
Schizoph, 206  
Seatbelt, 207  
Selfdefe, 208  
Senior, 208  
Sentence, 209  
Shkdrug, 210  
Shock, 211  
Shoplift, 211  
Short, 212  
Shuttle, 213  
SIGN. test, 214
- Simpson, 216  
Situp, 217  
Skewed, 218  
Skin, 218  
Slc, 219  
Smokyph, 220  
Snore, 221  
Snow, 222  
Soccer, 223  
Social, 223  
Sophomor, 224  
South, 225  
Speed, 225  
Spellers, 226  
Spelling, 227  
Sports, 227  
Spouse, 228  
SRS, 59, 229  
Stable, 230  
Stamp, 230  
Statclas, 231  
Statelaw, 232  
Statisti, 232  
Step, 233  
Stress, 234  
Study, 235  
Submarin, 235  
Subway, 236  
Sunspot, 237  
Superbowl, 238  
Supercar, 238
- Tablrock, 239  
Teacher, 241  
Tennes, 242  
Tensile, 243  
Test1, 243  
Thermal, 244  
Tiaa, 245  
Ticket, 245  
Toaster, 246  
Tonsils, 247  
Tort, 248  
Toxic, 249  
Track, 250  
Track15, 251  
Treatments, 251  
Trees, 252  
Trucks, 253

tsum. test, [216](#), [253](#), [273](#), [276](#)

Tv, [257](#)

Twin, [257](#)

Undergrad, [258](#)

Vacation, [259](#)

Vaccine, [260](#)

Vehicle, [261](#)

Verbal, [261](#)

Victoria, [262](#)

Viscosit, [263](#)

Visual, [264](#)

Vocab, [264](#)

Wastewat, [265](#)

Weather94, [266](#)

Wheat, [267](#)

Windmill, [268](#)

Window, [268](#)

Wins, [269](#)

Wool, [270](#)

Yearsunspot, [271](#)

z. test, [216](#), [256](#), [271](#), [276](#)

zsum. test, [216](#), [256](#), [273](#), [274](#)