```
help(co2)
```

co2 {datasets}	R Documentation

Mauna Loa Atmospheric CO2 Concentration

Description

Atmospheric concentrations of CO_2 are expressed in parts per million (ppm) and reported in the preliminary 1997 SIO manometric mole fraction scale.

Usage

```
co2
```

Format

A time series of 468 observations; monthly from 1959 to 1997.

Details

The values for February, March and April of 1964 were missing and have been obtained by interpolating linearly between the values for January and May of 1964.

Source

Keeling, C. D. and Whorf, T. P., Scripps Institution of Oceanography (SIO), University of California, La Jolla, California USA 92093-0220.

https://scrippsco2.ucsd.edu/data/atmospheric_co2/.

Note that the data are subject to revision (based on recalibration of standard gases) by the Scripps institute, and hence may not agree exactly with the data provided by R.

References

Cleveland, W. S. (1993) *Visualizing Data*. New Jersey: Summit Press.

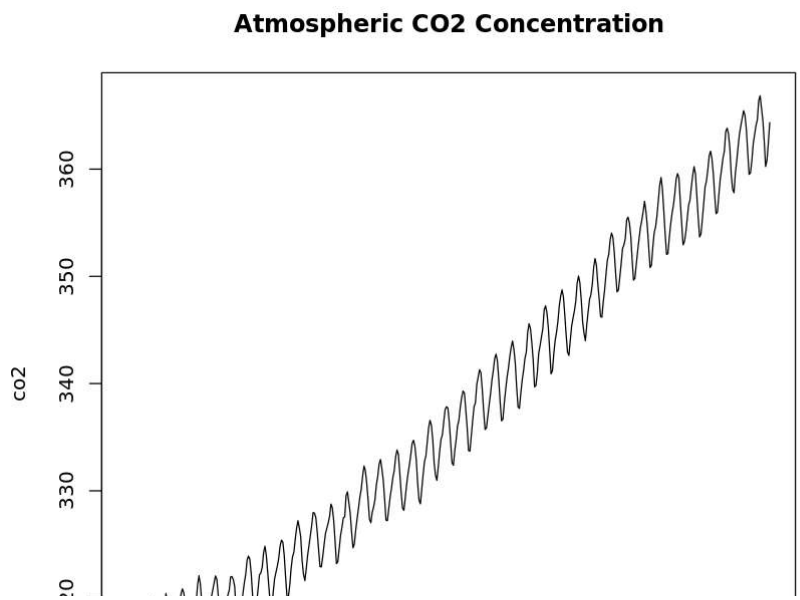
Examples

```
require(graphics)
plot(co2, ylab = expression("Atmospheric concentration of CO"[2]),
     las = 1)
title(main = "co2 data set")
```

[Package *datasets* version 4.2.3]

```
plot(co2, main='Atmospheric CO2 Concentration')
```

Download




```
# Develop your linear model

(co2.linear.model = lm(co2 ~ time(co2)))
```

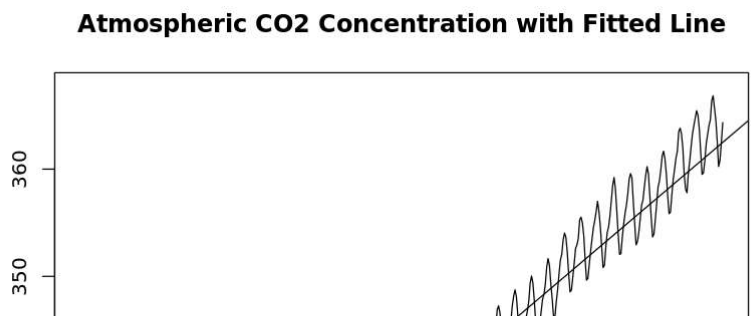
```
Call:
lm(formula = co2 ~ time(co2))

Coefficients:
(Intercept)    time(co2)
  -2249.774        1.307
```

```
# Plot your line with your data:

plot(co2, main='Atmospheric CO2 Concentration with Fitted Line')
abline(co2.linear.model)
```

Download



What We've Learned

- plot time series data
- fit a linear model to a set of ordered pairs

Objective - Perform a simple linear regression with R

- Assess normality of residuals

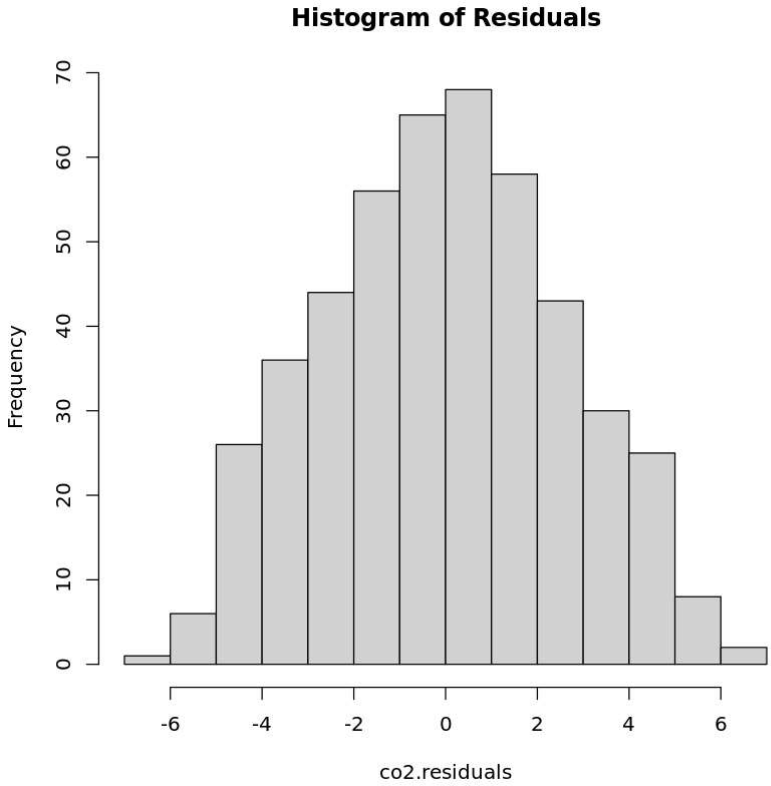
```
plot(co2, main='Atmospheric CO2 Concentration')
co2.lm = lm(co2 ~ time(co2))
abline(co2.lm)
```

Download


```
# Residuals

co2.residuals = resid(co2.lm)
hist(co2.residuals,main='Histogram of Residuals')
```

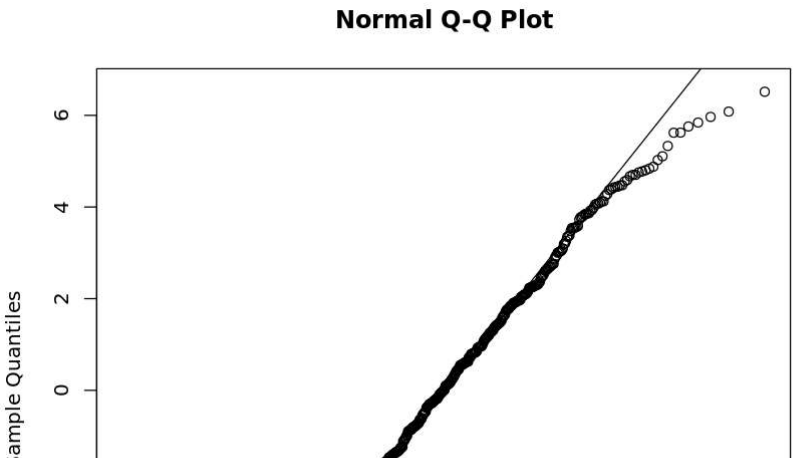
Download




```
# Normal Q-Q Plot

qqnorm(co2.residuals)
qqline(co2.residuals)
```

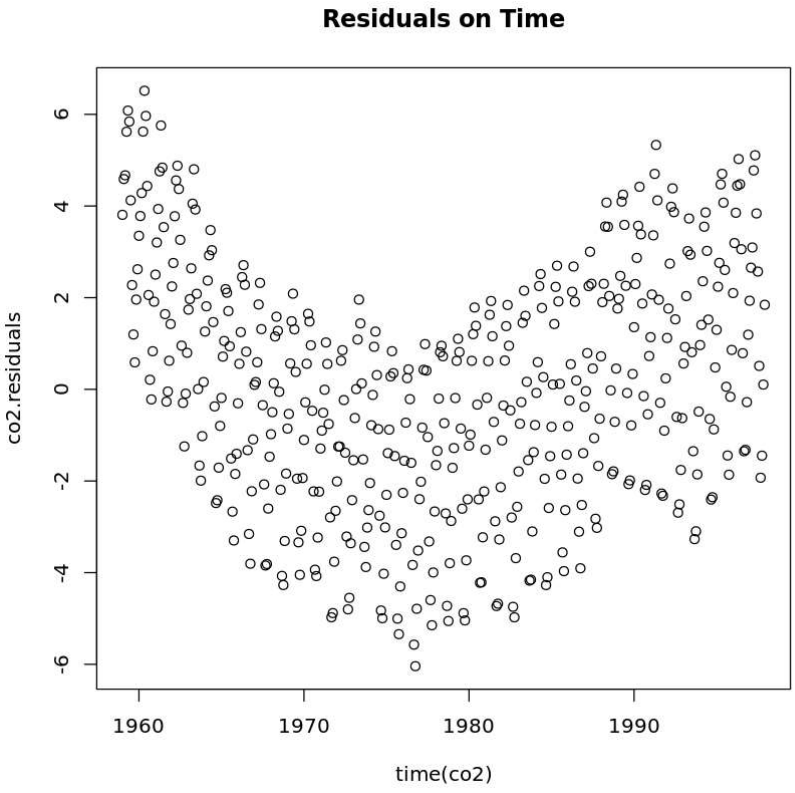
Download



Normality of Data is important for Regression.

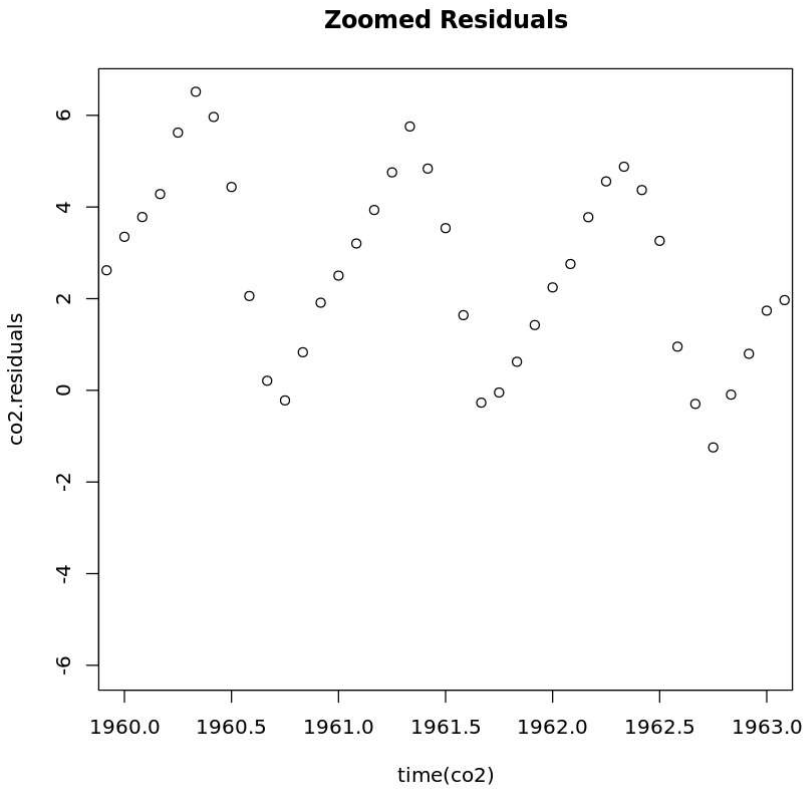
```
plot(co2.residuals ~ time(co2), main='Residuals on Time')
```

Download




```
plot(co2.residuals ~ time(co2), xlim=c(1960, 1963), main="Zoomed Residua
```

Download



Objectives - Review Basic Inferential Statistics

- Develop Graphical Intuition
- Perform a Hypothesis Test Concerning Means

```
help(sleep)
```

sleep {datasets}	R Documentation

Student's Sleep Data

Description

Data which show the effect of two soporific drugs (increase in hours of sleep compared to control) on 10 patients.

Usage

```
sleep
```

Format

A data frame with 20 observations on 3 variables.

[, 1]	extra	numeric	increase in hours of sleep
[, 2]	group	factor	drug given
[, 3]	ID	factor	patient ID

Details

The group variable name may be misleading about the data: They represent measurements on 10 persons, not in groups.

Source

Cushny, A. R. and Peebles, A. R. (1905) The action of optical isomers: II hyoscines. *The Journal of Physiology* **32**, 501–510.

Student (1908) The probable error of the mean. *Biometrika*, **6**, 20.

References

Scheffé, Henry (1959) *The Analysis of Variance*. New York, NY: Wiley.

Examples

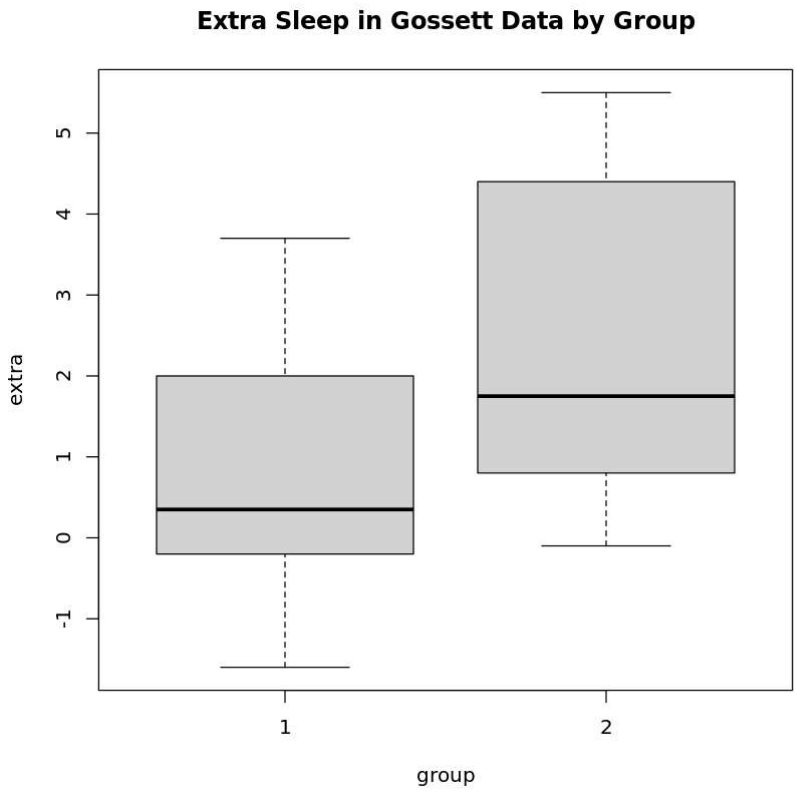
```
require(stats)
## Student's paired t-test
with(sleep,
     t.test(extra[group == 1],
            extra[group == 2], paired = TRUE))

## The sleep *prolongations*
sleep1 <- with(sleep, extra[group == 2] - extra[group == 1])
summary(sleep1)
stripchart(sleep1, method = "stack", xlab = "hours",
           main = "Sleep prolongation (n = 10)")
boxplot(sleep1, horizontal = TRUE, add = TRUE,
        at = .6, pars = list(boxwex = 0.5, staplewex = 0.25))
```

[Package *datasets* version 4.2.3]

```
plot(extra ~ group, data=sleep, main="Extra Sleep in Gossett Data by Gro
attach(sleep)
extra.1=extra[group==1]
extra.2=extra[group==2]
```

Download




```
# Test your Hypothesis!

t.test(extra.1, extra.2, paired=TRUE, alternative="two.sided")
```

```
        Paired t-test

data:  extra.1 and extra.2
t = -4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -2.4598858 -0.7001142
sample estimates:
mean difference
          -1.58
```

General Framework for Hypothesis Tests

- State clearly what your variables are (define your terms)
- State the null and alternative hypotheses
- Decide upon a level of significance, alpha.
- Compute a test statistic (z, t, X^2, and F are popular stats)
- Find the p-value corresponding to your test statistic (for left/right/or two tailed test).
- Form a conclusion: if p < alpha (improbable data) reject H-not otherwise do not reject. We typically do not accept, just like the courts never say that someone is innocent.

Measuring Linear Association with the Correlation Function

Objectives

- Plot data pairwise to visually explore the associations between variables
- calculate and interpret covariance and correlation

```
help(trees)
```

trees {datasets} R Documentation

Diameter, Height and Volume for Black Cherry Trees

Description

This data set provides measurements of the diameter, height and volume of timber in 31 felled black cherry trees. Note that the diameter (in inches) is erroneously labelled Girth in the data. It is measured at 4 ft 6 in above the ground.

Usage

```
trees
```

Format

A data frame with 31 observations on 3 variables.

[,1]	Girth	numeric	Tree diameter (rather than girth, actually) in inches
[,2]	Height	numeric	Height in ft
[,3]	Volume	numeric	Volume of timber in cubic ft

Source

Ryan, T. A., Joiner, B. L. and Ryan, B. F. (1976) *The Minitab Student Handbook*. Duxbury Press.

References

Atkinson, A. C. (1985) *Plots, Transformations and Regression*. Oxford University Press.

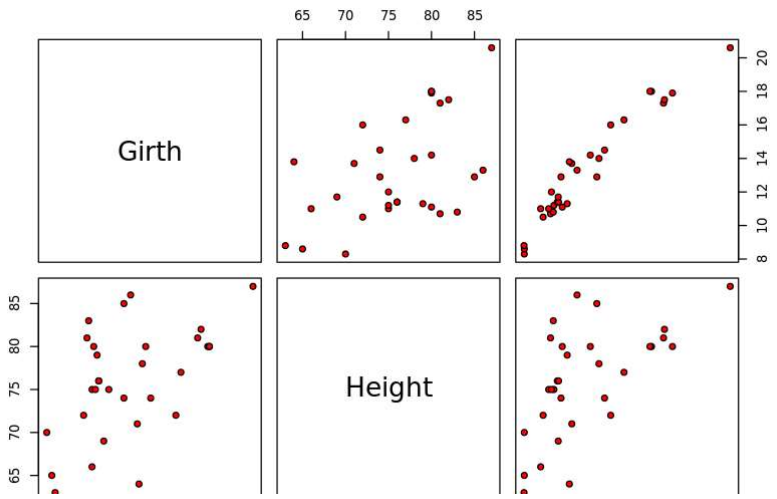
Examples

```
require(stats); require(graphics)
pairs(trees, panel = panel.smooth, main = "trees data")
plot(Volume ~ Girth, data = trees, log = "xy")
coplot(log(Volume) ~ log(Girth) | Height, data = trees,
       panel = panel.smooth)
summary(fm1 <- lm(log(Volume) ~ log(Girth), data = trees))
summary(fm2 <- update(fm1, ~ . + log(Height), data = trees))
step(fm2)
## i.e., Volume ~= c * Height * Girth^2  seems reasonable
```

[Package *datasets* version 4.2.3]

```
pairs(trees, pch = 21, bg = c("red"))
```

Download



Girth is very strongly associated with volume.

```
# covariance
cov(trees)
```

A matrix: 3 × 3 of type dbl

	Girth	Height	Volume
Girth	9.847914	10.38333	49.88812
Height	10.383333	40.60000	62.66000
Volume	49.888118	62.66000	270.20280

What We've Learned

- plot data pairwise to visually explore the associations between variables
- calculate and interpret covariance and correlation