Glossary

abscissa 1) *Math.* the x-coordinate.

address 1) *Comp.* a number that represents a location in memory.

aggregate 1) *Stat.* to combine consecutive observations in a **time series**; for example, a monthly time series of amount sold can be aggregated into an annual series.

AIC 1) *Stat.* "An Information Criterion", a method of combining the **likelihood** and the number of parameters so that non-nested models can be compared, due to H. Akaike. AIC has been criticized for suggesting models with too many parameters. See also **SIC**.

algorithm 1) *Comp.* a recipe for a computation. This can either be specific as in Euclid's algorithm for the greatest common divisor, or be a general way of approaching a problem as in the **EM algorithm**.

alias 1) *Unix.* a name that expands into some (part of a) command; similar to a **macro** in C. 2) *Stat.* in fractional experimental designs, some effects are aliased (or confounded) with others.

alpha version 1) *Comp.* a prototype of some software, numerous and serious **bugs** are to be expected. See **beta version**.

alphanumeric 1) the alphabetic characters plus the ten digits.

ampersand 1) the character "&", a symbol used in S, C and Unix.

argument 1) *Comp.* an input into a function or subroutine. 2) *Math.* the angle of a **complex number** from the positive real axis, calculated in S with `Arg`.

argument matching 1) *Comp.* the process of deciding which arguments in a particular call corresponds to the formal arguments of the routine being called. This is a trivial exercise except in languages like S that allow default values for arguments.

371

array 1) *S.* an object that has a `dim` **attribute**, and represents a (hyper) rectangle of objects. Also a function for creating those objects. 2) *C.* an object that is the equivalent to a **vector** in S. See also **ragged array**.

ASCII 1) *Comp.* a standard for encoding characters. Pronounced "ASS-key", it is an acronym for American Standard Code for Information Interchange. Each character uses 8 bits (1 byte), but many programs (including S) only deal well with the first $128 = 2^7$ characters.

aspect ratio 1) *Comp.* the ratio of the x-axis to the y-axis. This can refer either to the physical size of the graph, or to the units represented by the graph.

assembly code 1) *Comp.* a program in the machine language for a particular chip—very fast, but untidy.

assignment 1) *Comp.* the process of giving a name to some entity. Performed in S with "`<-`".

assignment function 1) *S.* the definition of a function that appears on the left of an assignment. For example:

```
 names(x) <- letters
```

The name of the `names` assignment function is "`names<-`". If `fjj` is a generic assignment function, then "`fjj<-.middle`" would be the assignment function for `fjj` for class `middle`.

association 1) *Math.* an operator $\circ$ is associative if $(x \circ y) \circ z = x \circ (y \circ z)$. 2) *Comp.* an operator $\circ$ associates from left to right if $x \circ y \circ z$ is evaluated the same as $(x \circ y) \circ z$, and it associates from right to left if it evaluates as $x \circ (y \circ z)$.

asymptotic 1) *Stat.* an approximation of a quantity based on the limit as the number of observations goes to infinity. For example, the asymptotic variance of an estimator.

at sign 1) the character "@", a symbol not used in S version 3, but is used to access **attributes** in version 4.

atomic 1) *S.* an object that is one of the five modes `numeric`, `logical`, `character`, `complex`, `null`. All elements of an atomic vector are of the same type.

attach 1) *S.* the process of adding a **database** to the **search list**. This is done with functions `attach`, `library`, `module`.

attributes 1) *S.* a named list that is a subsidiary part of an object. Examples of attributes include `class`, `names`, `dim`. Also the function that retrieves or changes this list.

audit   1) *Comp.* a record of the commands performed, generally with enough information that all states can be reproduced. S keeps such a record in the `.Data/.Audit` file.

axis   1) *Math.* the line in a graph along which one variable is zero. 2) *S.* the information such as tick marks and tick labels that are put along an axis. The axes are numbered 1 for the bottom, 2 for the left, 3 for the top, and 4 for the right.

background   1) *Unix.* a process is in the background if it is running, but not listening to **standard-in**. An example is an S `BATCH` job. A command can be put in the background by placing an **ampersand** at the end of the command.

background color   1) *S.* the color of a blank graph, in S this is denoted color 0.

backquote   1) the symbol '', unused in S, used in Unix to surround commands that are to be interpreted. See also **single quote**.

backslash   1) the symbol " \ " used as an **escape** character by many languages.

backward-compatibility   1) *Comp.* the state such that old code will run properly using a new version of the software.

bang   1) *Comp.* slang for the exclamation mark "!". Also called "screamer".

batch   1) *Comp.* a style of computing in which all instructions are submitted and then results are obtained; compare to **interactive**.

benchmark   1) *Comp.* the process of testing the speed of one computer (or piece of software) relative to another. A task that is harder than it seems since the relative speed of different tasks is often not very similar.

beta test   See **beta version**.

beta version   1) *Comp.* a version of some software that is thought to be close to release quality, but which has only been used by the developers. Allowing a group of people to use the version in return for reporting bugs and giving general comments is the *beta test.*

bias   1) *Stat.* the difference between the mean value of an estimator and the true (unknown) value that is being estimated. Often there is a trade-off between bias and variance when choosing an estimator.

binary file   1) *Comp.* a file that is in a format suitable to some program as opposed to a format for being read by humans. Generally, any file that is not in ASCII format is said to be a binary file. 2) *Comp.* in particular, **object code**.

binary number 1) *Math.* a number in base 2, composed only of the digits 0 and 1.

binary tree   1) *Math.* a relative of a mathematical **graph** in which each **node** has zero, one or two children (other nodes), and "left" children are distinguished from "right" children. It is the distinction of left and right that makes a binary tree different from an ordinary mathematical **tree**.

bit 1) *Comp.* a binary digit. 2) *Comp.* the amount of information contained in a binary digit.

bivariate 1) *Stat.* a situation involving two random variables. The smallest case of **multivariate**.

BLAS 1) *Comp.* acronym for Basic Linear Algebra Subroutine. If your computer has the BLAS, it does not have a case of ennui but rather it has subroutines for doing some linear algebra tasks that are optimized for that machine.

blue book 1) *S.* Becker, Chambers and Wilks (1988) *The New S Language*.

body 1) *S.* the part of a function where the computations are, as opposed to the **arguments**.

bootstrap 1) *Stat.* a way of evaluating the variability of an estimate by resampling the input data. 2) *Comp.* starting a simple process that allows a similar but more complex process to work.

Bourne shell   1) *Unix.* one of the most common Unix **shells**, very popular for programming.

box   1) *S.* the lines around the **plot area** of a **figure**.

box constraint 1) *Math.* constraints on a vector such that there is a minimum and a maximum for each element (independently of the other elements). Compare **linear constraint**.

boxplot   1) *Stat.* a graph indicating the distribution of some data. There are various definitions, but a common one (and the one used in S) is: The ends of the box are the first and third quartiles, the mark within the box is the median. The "whiskers" stretch to the farthest datapoint that is within distance d of the near end of the box, where d is some number times the length of the box. Points farther than d from the box are individually marked.

brace 1) one of the symbols "{" or "}", used in S and in C to bind several commands into one. See also **bracket**, **parenthesis**.

bracket 1) one of the symbols "[" or "]", used in S and in C for **subscripting**. Sometimes called "square bracket" to distinguish from the more general use of the term "bracket" that includes **braces** and **parentheses**.

brain-dead 1) *Comp.* slang for something that is ill-planned and/or ill-executed.

buffer 1) *Comp.* a chunk of memory set aside for a specific purpose. For example printed output in C uses a buffer—characters to be printed are put in the buffer and printing actually takes place when the buffer is full; then the buffer is marked empty so it can accept more characters.

bug 1) *Comp.* a problem with some software or hardware. 2) *Comp.* pejorative for **feature**.

byte 1) *Comp.* the basic unit of memory for a computer, in most cases consisting of 8 **bits**.

C shell 1) *Unix.* a popular **shell** for interactive use, less popular for programming.

cache 1) *Comp.* a storage area that is treated specially in some way. Almost always the term is used to mean something that is time efficient—it can be a memory location, a **hash table**, a **lookup table**, etc.

call 1) *S.* the action of invoking a function. 2) *S.* a **mode**, an object of mode `call` contains the information for a specific call to a function. 3) *S.* a **component** or **attribute** of many S objects that is the call that created the object of which it is a part (the mode of this component or attribute is generally `call`).

canonical correlation 1) *Stat.* a way of assessing the realationship between two multivariate sets of data.

carriage return 1) *Comp.* an **ASCII** character rather like **newline**, but not quite the same.

case-sensitive 1) *Comp.* a program is case-sensitive if capital letters are considered to be distinct from lower-case letters. S, Unix and C are case-sensitive; Fortran and DOS are not.

category 1) *Stat.* a variable that takes on a finite number of values and the values are generally non-numeric. Compare to **continuous variable**. The levels can be ordered as in {"low", "medium", "high"}, or not as in {"corgi", "kitty", "human"}. 2) *S.* an object that represents a category; the preferred type of object is one that inherits from class `factor`.

ceiling 1) *Math.* the smallest integer greater than or equal to a number. Compare **floor**.

central processing unit 1) *Comp.* the computer chip that does the actual computation. Abbreviated as "CPU".

chaos 1) *Math.* a process that appears to be random, but is really deterministic. **Pseudorandom numbers** are an example of chaos.

character   1) *S.* an **atomic** mode, each element of a character vector is a character string of arbitrary length.

characteristic   1) *Comp.* the exponent of a number stored in **floating-point**. The other part is the *mantissa*. 2) see **eigen**.

characteristic function   1) *Stat.* the Fourier transform of the density function of a random variable. One use is to compute moments of the distribution.

child   1) *S.* a **memory frame** is the child of the frame that caused it to come into existence. Frame 1 is no one's child. 2) *Unix.* a **process** is the child of the process that spawned it.

class   1) *S.* an **attribute**, the `class` attribute of an object determines which **method** of a **generic function** is dispatched when the object is a particular argument in the function call.

classification tree   1) *Stat.* a statistical model that describes a **categorical** variable by means of a recursive partition of the **explanatory variables** (where each partition is of a single explanatory variable). Fit in S with `tree`. See also **regression tree**.

clustering   1) *Stat.* a technique for grouping datapoints by their similarity; this is called "unsupervised learning" in the machine-learning literature.

coercion   1) *Comp.* the conversion of an object from one concept to another, such as from character to numeric, or from double-precision floating-point to single-precision floating-point.

collision   1) *Comp.* the occurrence of more than one value being encoded to the same location in a **hash table**.

colon   1) the character ":" which is used in S to produce sequences.

column-major order   1) *Comp.* the storage order of the elements of a matrix in which all of the elements of the first column are together, followed by the elements of the second column, etc. S and Fortran store matrices (and higher dimensional arrays) in column-major order. Compare to **row-major order**.

combination   1) *Math.* a selection without replacement of a certain number of elements from a finite set where the order of the selection is ignored. The binomial coefficient gives the number of combinations. Compare with **permutation**.

command line editing   1) *Comp.* the ability to recall past commands, edit them and execute the modified command. In S this can be accomplished with S-mode for emacs, or through the `-e` flag to S-PLUS.

command   1) a non-technical term—something that will make the computer do something.

comment 1) *Comp.* text in software that is ignored in computations, and is used to inform humans reading the code. In S, text following a **pound sign** on a line is a comment.

commitment 1) *S.* an **assignment** of an object to a permanent location is committed when it is actually written to disk, which often is not until S is about to give a prompt rather than at the time the assignment is made. In the interim S acts as if the object were in the specified location even though it isn't physically there.

commute 1) *Math.* an operator $\circ$ commutes if $x \circ y = y \circ x$.

compacting 1) *S.* term used to mean **garbage collection**. Compacting usually occurs only inside of loops.

comparison operator 1) *Comp.* an operator that tests equality or inequality. The comparison operators in S are ==, !=, >, <, >=, <=.

compile 1) *Comp.* the act of converting **source code** into the corresponding instructions for a particular computer chip and operating system. The result is a file of **object code**.

compile-time 1) *Comp.* the time at which a program was **compiled**, as opposed to when it is used which is called **run-time**.

compiled language 1) *Comp.* a language, C or Fortran for example, in which the source code is **compiled** into **object code** which can then be executed. Compare to **interpreted language**.

complex 1) *S.* an **atomic** mode, an object of this mode contains **complex numbers**.

complex number 1) *Math.* a number which is a real number plus a real number times the square root of -1. The square root of -1 is often denoted "i", which is the case in S.

component 1) *S.* an item in a **list** or other recursive object. The length of a list is the number of components it has. Compare with **element**.

condition number 1) *Comp.* a number that states the degree of ill-conditioning in a computation. By convention, small numbers mean good results, and large or infinite values mean poor results. For example, the condition number for inverting a matrix is often defined as the ratio of largest to smallest singular value.

conditional distribution 1) *Stat.* a probability distribution that assumes some condition holds. For example, the distribution of height given that weight is less than some specific amount.

S Poetry ©1998 Patrick J. Burns v1.0

conditioning plot 1) *Comp.* a plot of some sort that is conditional on a value
or set of values of a variable that is not represented in the plot. The
`coplot` function does this in S.

confounded   1) *Stat.* two effects are confounded if the data at hand can
not distinguish if it is only one or the other that has influence on the
response. For example, if you only observe people who either use both
alcohol and tobacco or use neither, then you won't be able to discern
the health effects of one of the drugs. Confounding is often a matter of
degree, not just totally confounded versus not confounded at all.

confusion matrix 1) *Stat.* when a **categorical** variable is being modeled,
a matrix that is a **contingency table** of the true categories versus the
predicted categories from the model of the data.

conjugate 1) *Math.* a number is the (complex) conjugate of another if it has the
same real part but the opposite imaginary part, so $a - bi$ is the conjugate
of $a + bi$. Computed in S with `Conj`. 2) *Stat.* in Bayesian statistics there
are conjugate distributions, meaning that the computations simplify when
the two distributions are used together as **prior** and **likelihood**.

constant 1) *S.* an object—like the number 5.13—that is fixed. 2) *Math.* an
object that is fixed, but not necesarily known.

constraint 1) *Math.* a condition placed on a problem. Two common classes
of constraint are **box constraint** and **linear constraint**.

contingency table   1) *Stat.* a table of counts of the number of occurrences
in the combination of levels of two or more **categorical** variables. When
there are two variables, then the table is represented as a matrix. More
variables means higher dimensional arrays. These are created in S with
`table`. 2) where you eat when your kitchen table has a broken leg.

continuation 1) *S.* an incomplete statement may be given to S and the con-
tinuation of the command given on one or more additional lines. The
default continuation prompt is "`+` ". 2) *Math.* a means of extending a
function over more of the complex plane.

continuous variable 1) *Stat.* a variable that can (conceptually) take on any
value in some range of the real numbers. Compare to **category**.

contour plot 1) *Comp.* a graphic that shows the value of a function of two
variables by drawing lines of equal value of the function.

contrast 1) *Stat.* a coefficient vector that sums to zero. Important because
they are **orthogonal** to the vector of all 1's which corresponds to the
mean. 2) *S.* an **attribute** or **component** of some objects representing a
statistical model that shows the set of contrasts that was used in the fit-
ting process. (These are not necessarily contrasts in the sense of meaning
1.)

control   1) *Comp.* the act of keeping a series of backups of objects such as files of code or documentation. In Unix this is done with SCCS or RCS (and probably other methods).

control operator   1) *S.* one of the operators `&&` or `||` which perform non-vectorized "and" and "or", respectively.

CPU 1) *Comp.* acronym for **central processing unit**.

crash 1) *Comp.* slang for the unintentional or unexpected halt of a program or of hardware.

cross-validation 1) *Stat.* a process of estimating the quality of a model by comparing observations left out of the fitting process with the prediction of that fit for the observations. For example, the available observations are randomly divided into 10 groups, the fitting is performed 10 times with a different group dropped each time. Read carefully when this term comes up—there are some confusing uses of it.

cursor 1) *Comp.* the marker on a computer screen that indicates where input is to go, or the spot where the mouse is pointing.

cut and paste 1) *Comp.* the action when using a window system of marking some text in one spot and making that text appear in another spot.

data frame   1) *S.* an object of class `data.frame` that represents a rectangle of values in which the type of value may be different from column to column. Compare to **matrix**.

database 1) *S.* an item on the **search list**, or something that might be. 2) *Comp.* a program designed to store information—generally in the form of **fields** and **records**, and often using **SQL**.

debug 1) *Comp.* the process of tracking down and fixing a problem. [The entomology of this word is reputed to be that a moth flew into the Mark I computer in 1945, crashing the machine. The problem was fixed by "debugging" the relay that failed. If this story isn't true, it should be.]

decimal number 1) *Math.* a number written in base ten.

declaration 1) *Comp.* the specification of the type of object that a variable is to be. This is done in compiled languages like C and Fortran, but essentially non-existent in S.

decomposition 1) *Math.* writing a matrix as the product of two or more other matrices of special types; for example, **QR decomposition**.

default 1) *S.* the value of an optional argument that is not given in the call.

density   1) *Stat.* a function that defines a probability distribution. If the distribution is discrete, then the density at a point is the probability of the distribution taking that value. If the distribution is continuous, then the density is such that the integral between any two points of the density is the probability of the distribution landing in that interval.

deparse   1) *S.* to change a language object into one or more character strings that represents the object.

dependent   1) *Stat.* two random variables are dependent if knowing information about one helps predict the other. 2) *Math.* a dependent variable is a function of one or more other variables. 3) *Stat.* a **response** in a statistical model. This is in analogy to meaning 2, but confusing because of meaning 1.

deque   1) *Comp.* a data structure similar to a **stack** or a **queue** except that both adding and deleting are allowed at both ends. [I believe the name derives from "double-ended queue."]

dereference   1) *C.* the act of getting the value pointed to by a **pointer**, done with the $*$ operator.

derivative   1) *Math.* the slope of a function at any given point.

deviance   1) *Stat.* a measure of the distance of the data from its fit to a model. The form of the measurement depends on the assumed **error** distribution.

diagnostics   1) *Stat.* statistics and graphics intended to show how well a statistical model fits the data—an important part of the modeling process.

diagonal matrix   1) *Math.* a matrix that is zero except on the diagonal (where the row number is equal to the column number).

dim   1) *S.* the **attribute** that describes the shape of an array. Also the function that retrieves or changes this attribute.

directory   1) *Comp.* a location that contains files and (possibly) sub-directories.

distribution   1) *Stat.* a description of how a particular random variable behaves.

distribution function   1) *Stat.* a function of x that is the probability that a random variable will be less than or equal to x.

distribution-free test   1) *Stat.* a class of hypothesis tests which use ranks of the data. An example is the Wilcoxon signed-rank test. Some times called "nonparametric".

dot plot 1) *Comp.* a plot that displays information as plotting symbols that are each some distance along their respective lines—used as an alternative to barplots and pie charts. These are created in S with `dotchart`.

dot product 1) *Math.* a number which is the sum of the product of corresponding elements of two vectors.

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i$$

where n is the length of the vectors. This can be done in S via `sum(u*v)`.

double quote 1) the character `"`. See also **backquote** and **single quote**.

double-precision 1) *Comp.* a **floating-point number**, usually consisting of 8 **bytes**. 2) *Comp.* computation performed with double-precision numbers.

dump 1) *S.* an **ASCII** representation of an S object created by `dump` or `data.dump` (these two representations are distinctly different). 2) *S.* the action of representing the state of S when an error occurs. This generally goes to the object `last.dump`. 3) *Comp.* an error action as in a core dump.

dynamic graphics 1) *Comp.* graphics that change spontaneously or from input by the user, usually to display high-dimensional data.

dynamic load 1) *Comp.* the process of adding code to a program that is already running. This is performed in S with `dyn.load` and its relatives. Compare **static load**.

dynamic programming 1) *Comp.* a technique for optimizing over n time points. Given the state at time $n-1$, the best decision to take is computable; so work backwards through time to the present.

echo 1) *Comp.* the act of the computer repeating input so it can be seen. Passwords are not echoed when they are typed in.

edge 1) *Math.* a part of a mathematical **graph** that connects two **nodes**.

effect 1) *Stat.* in 2^k designs, twice the estimated coefficient for a factor or interaction. 2) *S.* a component of an object of class `aov` that is a vector of orthogonal single degree-of-freedom values. This is not the same as definition 1.

efficiency 1) *Comp.* some measure of the amount of computing time, memory or other resources that a program uses. 2) *Stat.* the variance of an estimator (usually) measured relative to the variance of the best possible estimator given the model.

eigenvalue 1) *Math.* When an equation of the form $A\mathbf{x} = \lambda\mathbf{x}$ is satisfied, then λ is an *eigenvalue* of the square matrix A, and $\mathbf{x}$ is an eigenvector of A. If A is real-valued and symmetric, then all of the eigenvalues are real-valued. Eigenvalues are sometimes referred to as "characteristic values".

eigenvector See **eigenvalue**.

element 1) *S.* an item of an **atomic** vector. The **length** of an atomic vector is the number of elements that it contains.

ellipsis 1) the set of characters "...", also known as **three-dots** (which see).

EM algorithm 1) *Stat.* an approach to estimating a statistical model in which there is some form of missing data. The expected value of the missing data is taken (the E step), then the model **likelhood** is maximized given these values (the M step). These two steps are iterated until convergence. Convergence can be slow, but the code can be easier to write than alternatives (if there are any).

environment variable 1) *Unix.* variable that a **process** passes on to its children.

EOF 1) *Comp.* acronym for "End Of File."

error 1) *Stat.* a random disturbance, or the difference between the data and a **model** of the data. In this sense of the word, it does not mean "mistake". Most statistical models specify the distribution of the errors. 2) *Comp.* a condition that causes some program to stop execution—see **error handling** and **warning**. 3) *Comp.* non-technically, a **bug**. See also **numerical error** ,

error handling 1) *Comp.* the actions that happen when a program hits an error condition. See **dump** (meaning 2) for what happens in S.

escape 1) *Comp.* to temporarily leave a program to give an operating system command, done in S with the ! form. 2) *Comp.* a character that allows the following character (or set of characters) to have a meaning other than usual. In S the back-slash is the escape character in character strings.

evaluation 1) *S.* the process of doing the actual computation, after the command has been **parsed**.

evaluation frame 1) *S.* see **memory frame**.

event 1) *Comp.* an action that happens on a computer, such as a mouse click or the cursor entering a certain region. **Version 3** of S knows nothing about events, but **version 4** does.

executable 1) *Comp.* (adjective) the condition when a file can be executed, that is, will perform some action. 2) *Comp.* (noun) the file that executes a program, `Sqpe` is the executable of S.

expected value 1) *Stat.* the mean of a probability distribution.

explanatory variable 1) *Stat.* a variable in a statistical **model** that is used to approximate the **response**. This is also known as a "predictor" or the confusing "independent variable."

exponential notation 1) the same as **scientific notation**.

expression 1) *S.* a **recursive** mode in S that is the **parsed** form of one or more commands. 2) *Comp.* used non-technically to mean a command or part of a command. See also **regular expression**.

extraction See **subscript**.

factor 1) *S.* class of an object that represents a **categorical** variable. 2) *Stat.* a categorical variable in an experimental design (hence meaning 1). 3) *Stat.* a (hypothetical) latent variable in, for example, a factor analysis model. 4) *Math.* relating to a product, as in "an integer can be factored uniquely as the product of prime numbers."

factorial 1) *Math.* n-factorial, written $n!$, is the product of the integers from 1 to n. In S this is found with `gamma(n+1)`. Because the numbers quickly get large, `lgamma(n+1)` is usually more useful.

false convergence 1) *Comp.* state of a derivative-based optimizer in which it knows that it has not found an optimum, but doesn't know in which direction to go. This can be caused by various forms of ill-conditioning, but the surest way to achieve it is to get the sign of the gradient wrong.

family 1) *S.* a class of object used in `glm` and `gam` that describes the **link function** and **error** distribution.

FAQ 1) *Comp.* acronym for Frequently Asked Questions.

feature 1) *Comp.* euphemism for **bug**.

field 1) *Comp.* area of an **ASCII** file that is to be read as a unit, one or more fields comprise a **record**. For example, there could be two fields, species and weight, and each record would correspond to a different individual. 2) *Comp.* a similar concept in any **database**.

FIFO 1) *Comp.* acronym for "First In, First Out", a **queue**.

figure 1) *S.* conceptual area of a **graphics frame** that contains a **plot area** surrounded by a **margin**. There can be one or more figures in a frame, all of which are surrounded by the **outer margin**.

file 1) *Comp.* an object of the operating system that holds information (and often has a name).

FILO 1) *Comp.* acronym for "First In, Last Out", a **stack**.

fixed format 1) *Comp.* a style of **ASCII** file in which the **fields** are each a specified number of characters and in a specific location relative to the left border. This is opposed to a file which divides the fields by a delimitor such as a semicolon.

fixed-width font 1) *Comp.* a font in which each character is the same width.

flag 1) *Unix.* an optional argument to a Unix command, introduced by a dash. For example the "t" in `ls -t`. 2) *Comp.* in programming, a logical variable that signals a certain condition. For example, `has.names` in `c.rationalnum` on page 251 is a flag.

floating-point number 1) *Comp.* a computer representation of a real number composed of a fractional part, called the *mantissa* and an exponent, called the *characteristic*. Because the representation is finite, few numbers are represented exactly, and hence **numerical error** will usually result during computations.

floor 1) *Math.* the largest integer less than or equal to a number. Compare **ceiling**.

font 1) *Comp.* a set of descriptions of how characters are to be represented. This can be either rather general, as in "computer modern font"; or very specific, as in "10 point italic computer modern font."

formula 1) *S.* class of object that results from an expression that uses the **tilde** operator. Used to provide an argument of arbitrary complexity. Operators generally have special meanings within formulas.

fragmented memory 1) *Comp.* the condition of **RAM** in which there are numerous unused portions, making the total use of memory greater.

frame 1) *S.* there are (at least) three types of "frame" in S: **memory frame** (also called "evaluation frame"), **data frame**, and **graphics frame**.

function 1) *S.* **mode** of an object that defines computations. The computations are performed when a call to the function is evaluated. 2) *Math.* a correspondence from one set into another set.

functional language 1) *Comp.* a class of computer languages that avoids **side effects**. S is essentially a functional language, but would not be considered one by a functional language purist.

garbage collection 1) *Comp.* the act of freeing up memory that is no longer being used.

Gaussian distribution 1) *Stat.* the most common probability distribution in Statistics. Also known as "Normal" and "bell curve".

generalized additive model 1) *Stat.* a statistical model in which the response is approximated by a **link function** of the sum of nonlinear **smooths** of each explanatory variable. The **error** distribution can be one of several. Fit in S with `gam`.

generalized linear model 1) *Stat.* a statistical model in which the response is approximated by a **link function** of a linear combination of the explanatory variables. The **error** distribution can be one of several. Fit in S with `glm`.

generic function 1) *S.* a function whose behavior is determined by the **class** of one or more of its arguments. The class of the relevant argument(s) determines which **method** of the generic function is used.

genetic algorithm 1) *Comp.* a form of **random algorithm** used for optimization that imitates a natural selection process.

gigabyte 1) *Comp.* 10^9 bytes. 2) *Comp.* 2^{30} bytes.

global variable 1) *Comp.* (in S terminology) a variable that is not created locally, and is not passed in as an argument.

gotcha 1) Short for "got you", the "Choppy Water" chapter contains a number of gotchas.

gradient 1) *Math.* vector which is the first derivative of a multivariate function.

graph 1) *Math.* a collection of **nodes** (also called *vertices*) and **edges** which each connect two nodes. A particular type of graph is a **tree**. 2) *Comp.* (or "graphics") a pictorial representation of information. In S you must start a **graphics device** before creating graphics.

graphical parameter 1) *S.* same as **graphics parameter**.

graphical user interface 1) *Comp.* a system of how human and computer communicate, generally consisting of windows, menus, etc. The Motif flavor of X-Windows is an example.

graphics device 1) *S.* a program that renders graphics in some form. For example, some create a window and draw graphics in the window. The `postscript` device translates the graphics into PostScript and then sends it to a printer or a file. 2) *S.* an S function that starts such a program.

graphics frame 1) *S.* if the graphics are printed, this is equivalent to a single page. A frame contains one or more **figures**, and an **outer margin**.

graphics parameter 1) *S.* a number, string or logical value that controls some aspect of how a graph will look, or that describes the state of the graphics device. The `par` function queries and changes graphics parameters.

gray scale 1) *Comp.* the rendering of values by means of different shades of gray. This can be done in S-PLUS with `image`.

greedy algorithm 1) *Comp.* (not strictly an algorithm, but more a description of some algorithms) an optimization algorithm that always does the best it can on each step, but does not look ahead. For some problems this will produce the true optimum, but for lots of problems it will not.

group 1) *S.* a collection of **generic functions** that may have one **method** written for all of them. 2) *Math.* a set for which there is an associative operation, the identity for the operation is in the set, and the inverse of every element in the set is also in the set.

GUI 1) *Comp.* acronym for **Graphical User Interface**.

hack 1) *Comp.* to illegally enter (or attempt to enter) a computer. 2) *Comp.* to discover how some undocumented program works by experimenting with it. 3) *Comp.* a **kludge** (meaning 1).

hacker 1) *Comp.* a person who "hacks" (meaning 1). 2) *Comp.* slang for programmer, as in "S hacker."

Hadamard product 1) *Math.* element by element product of matrices. The operation performed by `*` on matrices in S (the `%*%` operator is used for matrix multiplication).

hang 1) *Comp.* said of a computer or program when nothing happens any more, as in "S hangs when I execute this command." This is probably caused by one of three things: it is not doing anything, there is an infinite loop, the computations take muxh longer than expected.

hard-coded 1) *Comp.* a part of a program that is not easily changed. For example, the functionality performed in a `.Internal` function in S is hard-coded since it can not be changed by users. Another example is a concept in a program that is always used as a constant rather than abstracted into a variable.

hash symbol 1) the character "#", also called "pound sign."

hash table 1) *Comp.* a table of values in which each record has a key on which a mathematical operation is performed to give an address for the record. This provides fast access to any record for which the key is known.

hat 1) slang for the caret character, ˆ. Used in Statistics to mean an estimate, often specifically the **MLE**.

hat matrix 1) *Stat.* the matrix (in linear regression) that puts the "hat" on the vector of observations, that is, the hat matrix times the observation vector equals the fitted observations. $\hat{y} = Hy$. It is mainly the diagonal of the hat matrix that is of interest in diagnostics.

header file 1) *C.* a file usually ending in ".h" that contains definitions of general use—intended to be included in code. The `S.h` file, which can be used for C code to be called by S, is an example.

help file 1) *S.* the on-line documentation on individual objects. The corresponding term in Unix is "man page".

Hermitian matrix 1) *Math.* a matrix that is equal to the **transpose** of its complex **conjugate**. Also called "self-adjoint." For real-valued matrices, this is the same as symmetric.

Hessian 1) *Math.* matrix which is the second derivative of a multivariate function.

hexadecimal 1) *Math.* referring to base 16.

high-level graphics function 1) *S.* a function that makes a new **graphics frame**, and then creates a graph.

high-level graphics parameter 1) *S.* a **graphics parameter** that may not be given to `par`.

histogram 1) *Stat.* a plot showing the distribution of a variable by means of boxes that show the count of observations in a number of adjacent intervals. Created in S by `hist`.

hook 1) *Comp.* a feature of a program that allows some functionality to be easily added later if desired. As in, "`line.integral` has a hook for additional methods of integration."

idempotent 1) *Math.* an object such that the product of itself with itself equals itself. 0 and 1 are idempotent, as are some matrices.

independent 1) *Stat.* two distributions are independent if information on one gives you no information about the other. Mathematically this is expressed by the density of the **joint distribution** being equal to the product of the densities of the **marginal distributions**. 2) *Math.* an independent variable is in the domain of a function. 3) *Stat.* the term "independent variable " is used to mean an **explanatory variable** in analogy with meaning 2, but confusing because of meaning 1.

index 1) *S.* an element or component number. In the S command `x[4]`, 4 is the index into `x`.

indirection 1) *Comp.* a form of abstraction in which a variable is created to avoid **hard-coding** something. For instance, creating a variable for the first part of a path to a file.

information matrix 1) *Stat.* matrix that measures the variability of estimated parameters. In **maximum likelihood estimation**, this can be approximated using the **Hessian** at the optimum.

inheritance   1) *Comp.* the feature of **object-oriented programming** in which a subclass of objects can use some of the methods for the main class. In S inheritance results when a **class** attribute has length greater than one.

input/output   1) *Comp.* also "I/O", functionality related to information entering or leaving a program.

integer   1) *Math.* one of the numbers $0, \pm 1, \pm 2, \ldots$ 2) *Comp.* a value known to be an integer. Compare to **floating-point**. 3) *S.* sometimes a vector with storage mode `integer`; sometimes a floating-point number that is logically integer.

integer programming   1) *Comp.* an optimization problem with (linear) constraints in which the elements of the answer must be integer-valued.

interactive   1) *Comp.* a program that converses with the user. S is an interactive language since it takes input from the user, responds to it, and then asks for more input. Opposed to **batch**.

interpreted language   1) *Comp.* a language that is not compiled, so that the interpretation into machine instructions is performed on the fly. S is an interpreted language. Compare **compiled language**.

interquartile range   1) *Stat.* the distance between the first quartile (25th percentile) and the third quartile (75th percentile) of a probability distribution or empirical sample.

interrupt   1) *Comp.* a **signal** that tells a program to quit doing what it is doing at the moment. Control-c is the usual key binding for the interrupt signal. Sometimes "interrupt" is used in a less technical sense to include other signals as well, such as control-**backslash**.

invisible   1) *S.* a value that is not printed automatically.

IQR   1) *Stat.* acronym for the **interquartile range**.

iteration   1) *Comp.* performing an operation repetitively in a loop while some things change. Compare to **recursion**. 2) *Comp.* one pass through an iterative computation.

joint distribution   1) *Stat.* a multivariate probability distribution of all of the variables in question. Compare with **marginal distribution**.

julian date   1) *Comp.* an integer that represents the number of days from a certain date.

justify 1) *Comp.* to place text in a certain alignment. For example, left justify means to place along the left margin.

K 1) *Comp.* slang for **kilobyte**.

kernel 1) *Unix.* the heart of the Unix system, the part that is the same no matter what **shell** is used. 2) *Stat.* a function that produces the weights for a moving average-type (kernel) **smooth**.

key binding 1) *Comp.* the association of a key or key combination on the keyboard to a specific meaning such as "erase" or "interrupt."

kilobyte 1) *Comp.* $1024 = 2^{10}$ bytes. 2) *Comp.* 1000 bytes.

kludge 1) *Comp.* a piece of programming that is ugly. 2) *Comp.* a piece of programming that is so ugly, it's sort of cute.

knap-sack problem 1) *Math.* a specific class of optimization problem. The prototypical problem is to maximize the value of what goes into your knap-sack while keeping the weight below a certain limit.

Kronecker product 1) *Math.* a product of two matrices in which the result has dimensions that are the products of the dimensions of the matrices being multiplied. The S-PLUS `kronecker` function generalizes the concept beyond multiplication.

labels 1) *S.* a generic function that creates labels for various objects. 2) *S.* an argument to `factor` and `ordered` that controls the value of the `levels` attribute of the output.

language 1) *S.* an object that possesses one of several **modes** that correspond to the language itself rather than to data. See also **functional language**, **interpreted language**, **compiled language**, **natural language**.

LAPACK 1) *Comp.* a package of Fortran routines for performing linear algebra, available from **netlib**. Some LAPACK **symbols** are in S-PLUS.

lazy evaluation 1) *Comp.* strategy in which objects are not actually **evaluated** until necessary. S uses lazy evaluation.

least absolute deviations 1) *Stat.* principle of estimation in which the sum of the absolute value of the estimated errors is minimized. This typically overcomes some of the statistical **robustness** problems of least squares, but has other statistical deficiencies and is a little more expensive to compute. Also known as LAD, least sum of absolute errors, LSAE, $\mathcal{L}_1$.

least squares 1) *Stat.* principle of estimation where the sum of squares of the estimated errors is minimized. Useful because of its computational simplicity in many situations, and its theoretical optimality given a Gaussian distribution, but it often fails to be statistically **robust**.

legend 1) *Comp.* a part of a graph that explains the meaning of the symbols used in the main part of the graph.

length 1) *S.* a fundamental property of an S object. In numeric vectors, for example, it states how many numbers the object contains.

levels 1) *S.* **attribute** of a categorical object that is a character vector containing the possible values that the variable can take on. The `labels` argument to `factor` and `ordered` can be used to control the resulting `levels` attribute.

library 1) *S.* a directory containing S objects that is set up so that the `library` function can **attach** it. 2) *Comp.* a set of compiled routines. A user's **object code** is **linked** against a library to get **symbols** that are defined in the library.

likelihood 1) *Stat.* the density of a statistical model thought of as a function of the parameters of the distribution with the data known. The **maximum likelihood estimate** is the value of the parameters that maximizes this function. 2) *Stat.* the value of the likelihood (previous meaning) or its logarithm when it is maximized.

limitation 1) *Comp.* something that a program might be expected to do but doesn't. A **bug** of omission.

linear constraint 1) *Math.* a constraint on vectors $\mathbf{x}$ that can be written as $\mathbf{c}\mathbf{x} < a$ or $\mathbf{c}\mathbf{x} \leq a$ or $\mathbf{c}\mathbf{x} = a$. A particularly simple and common set of linear constraints is a **box constraint**.

linear programming 1) *Comp.* an optimization problem in which the objective function is a linear function of the variables, and there are linear constraints on the variables. Abbreviated as "LP".

link 1) *Comp.* to combine files of **object code** together so that **symbols** from one that are used in another are resolved. 2) *Unix.* a file name that points to a file with another name. This can be used for abstraction, or to save space.

link function 1) *Stat.* in a **generalized linear model**, the function of the linear combination of the **explanatory variables** that approximates the **expectation** of the **response**. A function that performs a similar task in other models, such as **generalized additive models**.

linked list 1) *Comp.* a data structure in which each item in the list contains a pointer to the location of the next item. Specializations include doubly linked lists in which each item has a pointer to the previous item as well as the next item, and circular lists in which the final item has a pointer to the first item.

literal 1) *Comp.* in a language, text that represents some specific thing, as opposed to a variable. For example, "3.1" is a literal for the number 3.1.

load 1) *Comp.* to make compiled routines available to a program by making sure that the **symbol table** of the program contains the routines. This can be done statically or dynamically. See **dynamic load**, **static load**.

logical 1) *Comp.* something that is True or False. In S a third logical value is NA. 2) *S.* an **atomic** mode.

long-tailed distribution 1) *Stat.* a probability distribution that has a higher probability of points far from the mean than a **Gaussian** distribution with the same variance. Such distributions are problematic for **least squares** estimators.

lookup table 1) *Comp.* a table of values used in a computing strategy in which values are stored and retrieved rather than being computed each time.

low-level graphics function 1) *S.* a function that does not switch to a new **graphics frame**, but adds to a graph instead. Examples include `points`, `abline`, `axis`.

LP 1) *Comp.* acronym for **linear programming**.

machine dependent 1) *Comp.* something that changes between types of computers. For example, the order of bits in a **floating-point number** is different on some machines than on others. The order of bits for **ASCII** characters is the same on all computers, so ASCII is not machine dependent (which is the main point of ASCII).

macro 1) *C.* a definition that is substituted into the code in the **preprocessing**.

man page 1) *Unix.* On-line help for a Unix command.

mantissa 1) *Comp.* the fractional part of a number in a **floating-point number**. The exponential part is called the *characteristic*.

margin 1) *S.* the conceptual part of a graph that surrounds the **plot area** but is within a **figure**. Axis labels and titles are in the margin. See also **outer margin**.

marginal distribution 1) *Stat.* the univariate probability distribution of a variable (within a multivariate context). Compare with **joint distribution**.

masking 1) *S.* an object masks another object if they have the same name (and sometimes the same **mode**), and the first is earlier on the **search list** than the second.

mathematical graph 1) *Math.* See **graph**.

matrix 1) *Math.* a rectangular display of numbers that can be treated as a single entity. 2) *S.* an object that has a `dim` of length two. (This is the definition `is.matrix` uses, includes **data frames**.) 3) *S.* an object that has a `dim` attribute of length two and is also an **array**. (This is the definition `as.matrix` uses, excludes data frames.)

maximum likelihood estimation 1) *Stat.* a principle of estimation which uses the most probable value of a set of parameters given the data and a statistical model. In some circumstances, models exist such that the maximum likelihood estimator is equivalent to **least squares**, or to **least absolute deviations**.

mean 1) *Stat.* a measure of the central location of a probability distribution, for empirical data it is the sum of the observations divided by the number of observations. For a single sample it is a **least squares** estimate.

median 1) *Stat.* a measure of the central location of a probability distribution, it is the 50th percentile. For empirical data, at least half the data are bigger than or equal to the median, and at least half are smaller than or equal to it. The median is not uniquely defined for all datasets. It is a **least absolute deviation** estimate for a single sample.

meg 1) *Comp.* slang for megabyte. As in, "My S function gobbled 55 megs."

megabyte 1) *Comp.* 2^{20} bytes. 2) *Comp.* 1,000,000 bytes.

memory   1) *Comp.* storage of objects. This is often **RAM**, but may be **swap-space**, disk-space, etc.

memory frame   1) *S.* the collection of values that are local to a particular function during an evaluation. Almost every function call results in a new frame as it is evaluated. Also called "evaluation frame."

menu 1) *Comp.* a device in interactive computing that provides a number of choices to the user from which the user is to select.

method 1) *S.* a function that performs the computation of a **generic function** for a specific **class**.

micro- 1) a prefix meanining one-millionth (10^{-6}).

missing   1) *S.* **mode** of an object that is an **argument** which was not included in the **call**.

missing value 1) *S.* a value written `NA` that represents a value that is unknown. See also **not-a-number**.

MLE 1) *Stat.* acronym for **Maximum Likelihood Estimate**.

mode 1) *S.* a fundamental property of an S object which states what type of object it is. 2) *Stat.* the most likely value of a probability distribution. One way to estimate this for continuous distributions is to find the maximum of a density estimate of the data. For discrete distributions use `table`.

model 1) *Stat.* a hypothetical picture of how some particular set of data is generated, usually consisting of a functional description plus a probability distribution for the errors from the functional part. G. E. P. Box says, "All models are wrong, some models are useful."

modular programming 1) *Comp.* discipline in programming that allows pieces of the code to work independently of each other.

modulo 1) *Math.* a number which can be thought of as the remainder after a division. The `%%` operator performs this in S.

modulus 1) *Math.* the distance of a complex number from the origin, equivalent to the absolute value for real numbers. In S the `Mod` function returns this.

monadic 1) *Math.* **unary**.

Monte Carlo 1) *Stat.* evaluation by means of a random mechanism. Basically, a synonym for simulation. For example, a multivariate integral can be approximated by sampling uniformly over the integration space and averaging the function values of those points—this is Monte Carlo integration.

multitasking 1) *Comp.* Using the same resources for more than one job by a time-sharing mechanism. Unix arranges for its processes to be multitasked.

multivariate 1) *Stat.* pertaining to more than one variable. Opposed to **univariate**, but includes **bivariate**.

NA 1) *S.* how **missing values** and **not-a-numbers** are printed.

name 1) *S.* **mode** of an object that is a variable within an **expression**. Not to be confused with the **names** attribute.

names 1) *S.* an **attribute** of many objects that labels the elements or components of the object. Also the function that retrieves or sets this attribute.

NaN 1) *Comp.* a common way that **not-a-number** is printed. In S they are printed `NA`, but can be distinguished from ordinary missing values with `is.nan`.

nano- 1) a prefix meaning one-billionth (10^{-9}), as in "π seconds are a nanocentury."

natural language 1) a language that humans have developed to communicate with each other. Examples are English, French and Northern Sahaptin. Opposed to computer language.

netlib 1) *Comp.* on-line repository of a large amount of software and other information. Much of the software is numeric. The world wide web address is `http://www.netlib.org/`.

newline 1) *Comp.* a character that means a new line of text is to be started. Represented in S and in C by backslash-n. It is different than **carriage return**.

node 1) *Math.* a part of a mathematical **graph** that is usually represented as a point. Also called "vertex".

nonparametric 1) *Stat.* a means of estimation that does not depend on optimizing parameters, many **smooths** are nonparametric. 2) *Stat.* a **distribution-free** procedure.

norm 1) *Math.* a number which gives the "size" of something. There are various norms for vectors, and for matrices.

not-a-number 1) *Comp.* one of a class of "numbers" that represents a limit that does not exist. Zero divided by zero, and infinity minus infinity are examples. Often printed as `NaN`, but in S printed as `NA`. S has the function `is.nan` to distinguish where not-a-numbers are.

NP-complete 1) *Comp.* the definition is a convoluted statement involving polynomial-time algorithms. The practical effect is that if you have a problem that is NP-complete, then the best you are likely to do is get a reasonable approximation to the solution. To be sure that you have the real solution is too hard in all but the smallest of problems.

null 1) *S.* **mode** of the object of zero length that is printed `NULL`.

numeric 1) *S.* **atomic** mode that represents real numbers. This contains the storage modes `double`, `single` and `integer`. Complex numbers are excluded.

numerical differentiation 1) *Comp.* approximating the derivative of a function at specific locations by performing arithmetic on function values. Not a trivial task since it involves dividing by approximately zero.

numerical error 1) *Comp.* the difference between an actual result as computed compared to what the answer would be if computations were performed with numbers of infinite precision. Note that "error" here is not used in the sense of a **bug**.

numerical integration 1) *Comp.* the approximation of an integral by some means. Also called "quadrature."

object 1) *S.* almost everything in S is an object—if it has a **mode**, it is an object.

object code 1) *Comp.* the **compiled** version of some code in a **compiled language**. Typically the file name ends with `.o` in Unix.

object-oriented programming 1) *Comp.* a style of software in which the action of some "message" (or function) depends on the type of object that receives the message. In S this is accomplished with **generic functions**, **methods** for generic functions and the `class` attribute.

octal 1) *Math.* pertaining to numbers in base 8.

OOP 1) *Comp.* acronym for **Object-Oriented Programming**.

operating system 1) *Comp.* software that controls the general operation of a computer. Unix and MacOS are operating systems.

operator 1) *S.* a function that does not need to use the usual format when called. Examples are `+`, `$` and `!` (when it means "not").

option 1) *S.* one of several values that control some aspect of the S session. These are queried and set by the `options` function.

optional argument 1) *Comp.* an argument to a function that need not be given. When it isn't given, then usually some default value is used for it.

orthogonal matrix 1) *Math.* a square matrix A is orthogonal (also called "unitary") if

$$A'A = I$$

that is, if the transpose of A is equal to the inverse of A. Some call this "orthonormal" and allow an orthogonal matrix B to be such that $B'B$ is some diagonal matrix.

orthogonal vectors 1) *Math.* two vectors are orthogonal if their **dot product** is zero.

outer margin 1) *S.* the conceptual area that is at the very edges of a **graphics frame**. This is usually only used (that is, has non-zero area) when there is more than one **figure** in the frame. See also **margin**.

outlier 1) *Stat.* a datapoint that is far enough from its predicted value in a statistical model that it strains the credibility of the model for that point.

overflow 1) *Comp.* the condition when an operation causes the answer to go out of bounds of the range of the number representation. This most commonly happens with **integers** since the range is relatively small, but may also happen with **floating-point numbers**. In S at least, overflow with floating-point numbers is little problem since the value gets set to infinity; but with integers the value is capricious.

overload 1) *Comp.* the action of making the same thing do more than one job. **Generic functions** in S are overloaded.

p-value 1) *Stat.* in hypothesis testing, the probability that something more extreme than the test statsitic would be observed if the null hypothesis were true. A small p-value is evidence against the null hypothesis being true (if the conditions assumed for the calculation of the p-value are close enough to being true).

packed 1) *Comp.* one of numerous ways of writing a **sparse** or repetitious matrix in a small amount of space.

paging 1) *Comp.* see **swapping**.

parent 1) *S.* the **memory frame** that caused the memory frame in question to come into existence. 2) *Unix.* the **process** that caused the process in question to come into existence.

parenthesis 1) either of the characters "(" or ")", which are respectively an "opening parenthesis" and a "closing parenthesis". Can also be used in the broader sense that includes **braces** and **brackets**.

parse 1) *Comp.* the process of breaking a group of characters into parts with respect to the **syntax** of the language.

paste 1) *Comp.* see **cut and paste**. 2) Both S and Unix have `paste` routines to combine text.

permission 1) *Unix.* the ability of the owner of a file, members of the owner's group and users at large to read, write and execute the file.

permutation 1) *Math.* a particular ordering of objects. There are n-**factorial** permutations of n distinguishable objects. See also **combination**.

permutation test 1) *Stat.* a test of "no effect" that is performed by permuting observations and comparing the real statistic with the distribution of statistics from the permuted data. It has the advantages that it is valid under very broad assumptions and it is easy to compute, but is limited in applicability. It is similar to the **bootstrap**.

pico- 1) prefix meaning 10^{-12}, as in picosecond.

pipe 1) *Unix.* a command that uses the | operator to take the **standard-out** from the process on the left as the **standard-in** for the process on the right.

pivot 1) *Comp.* there are various meanings, but commonly the reordering of computations in order to ensure the computation is as numerically **stable** as possible.

platform 1) *Comp.* non-technical term meaning the type of hardware, and possibly **operating system**. For example, Pentium machine running Linux.

plot area 1) *S.* the portion of a graph where the actual data go, surrounded by the **margin**. See also **figure**, **graphics frame**.

pointer 1) *C.* an object that holds the address of the real object of interest.

polar coordinates 1) *Math.* the designation of points by angle and distance from the origin. Compare to **rectangular coordinates**.

pop 1) *Comp.* the process of removing an item from a **stack**.

portable 1) *Comp.* code that can easily be adapted to numerous computers.

positive definite 1) *Math.* a symmetric matrix A is positive definite if $\mathbf{x}'A\mathbf{x} > 0$ for all $\mathbf{x}$. This is equivalent to all of the eigenvalues of A being positive.

positive semidefinite 1) *Math.* a symmetric matrix A is positive semidefinite if $\mathbf{x}'A\mathbf{x} \geq 0$ for all $\mathbf{x}$. This is equivalent to all of the eigenvalues of A being non-negative.

posterior distribution 1) *Stat.* in Bayesian statistics, the distribution of the model parameters after taking the data into account. The posterior is a combination of the **prior distribution** and the **likelihood**.

PostScript 1) *Comp.* a language that describes how text and graphics are to be rendered.

pound sign 1) The character "#", also called "hash symbol", used for comments in many languages including S and Unix, but not C.

precedence 1) *Comp.* the order in which calculations take place in an otherwise ambiguous context. For example * takes precedence over + in S and in C.

predictor 1) *Stat.* synonym for **explanatory variable**.

preprocessor 1) *C.* a process that changes the text of a source file before it is **compiled**. This is used, for example, to handle machine dependencies. The **pound sign** is used to indicate preprocessor commands.

principal components 1) *Stat.* a multivariate technique that creates a set of new variables (the principal components) that are uncorrelated with each other and are ordered such that the variance of each principal component is no larger than the preceeding one. The first principal component has as large of variance as possible.

prior distribution 1) *Stat.* in Bayesian statistics, the hypothesized distribution of the model parameters before the data are consulted. See **posterior distribution**.

process   1) *Unix.* the basic "being" in Unix, executing a command starts one or more processes.

projection   1) *Math.* a specific way of mapping a space into a smaller subspace. For example, the least squares **hat matrix** is a projection matrix.

projection pursuit   1) *Stat.* a multivariate technique that searches for "interesting" linear combinations of the variables. There are a few such techniques in use, for example projection pursuit regression.

prompt   1) *Comp.* characters used to indicate that an interactive program is ready for input from the user. The default prompt in S is "`>` ".

pseudorandom number   1) *Comp.* one of a series of numbers generated by a mathematical function. The series is hoped to possess the most important empirical features of a truly random series of numbers.

push   1) *Comp.* the process of adding an item to a **stack**.

QP   1) *Comp.* acronym for **quadratic programming**.

QQplot   1) *Stat.* a plot to compare an empirical distribution to a theoretical probability distribution (or another empirical distribution). If the points are along a straight line, then the two distributions match (except possibly different means and variances). Points in the tails are much more variable than those in the middle. Functions in S for this include `qqnorm`, `qqplot` and `ppoints`.

QR decomposition   1) *Comp.*   a class of algorithms for converting a rectangular matrix into an **orthogonal matrix** times an upper **triangular matrix**. This is convenient and numerically **stable** for least squares regression calculations.

quadratic programming   1) *Math.*   the solution to a problem such as the minimum over all $\mathbf{x}$ of $\mathbf{g}'\mathbf{x} - .5\mathbf{x}'H\mathbf{x}$ subject to a set of linear constraints on $\mathbf{x}$.

quadrature   1) *Comp.*   univariate **numerical integration**.

quantile   1) *Stat.* a value such that a random variable is less than the value a given percent of the time. The 95 percent quantile of a standard Gaussian distribution is approximately 1.64. For empirical data there are a number of ways of calculating the quantiles which are slightly different from each other.

quartile   1) *Stat.* a **quantile** that is at the 25th, 50th or 75th percentile. See also **interquartile range**.

queue   1) *Comp.* a data structure in which items are added at one end and removed from the other. Also called "FIFO" for "first in, first out". Compare to **stack**.

quick call 1) *S.* an S function that does not create a separate **memory frame**.

R 1) *Comp.* a language that is very similar to S—much S code can run properly in R. It is free software available via **statlib**, and was originally developed by Robert Gentleman and Ross Ihaka.

radian 1) *Math.* a unit of measure for angles, there are 2π radians in a circle. The trigonometric functions in S use radians.

radix 1) *Math.* the base of the representation of a number. Examples include decimal, binary, hexadecimal.

ragged array 1) *Comp.* a data structure in which there is some number of dimensions (as in an **array**), but not all of the "cells" contain the same number of items. These are discussed further on page 79.

RAM 1) *Comp.* acronym for **random access memory**.

random access memory 1) *Comp.* also called "main memory", the location where data needs to be in order to be accessed by a **central processing unit**. In S all of the objects involved in a computation must be held in RAM. 2) *Comp.* more generally (and accurately) a memory bank that can be accessed by address rather than sequentially.

random algorithm 1) *Comp.* an algorithm that uses (psuedo) random numbers, and hence often arrives at different solutions on different runs. Examples are **genetic algorithms** and **simulated annealing**. 2) *Comp.* used loosely to mean an algorithm that uses randomness, but all of the randomness is **independent**, unlike in a genetic algorithm or simulated annealing.

random seed 1) *Comp.* an object that holds the state of a pseudorandom number generator. In S the random seed is kept in the object `.Random.seed`.

range 1) *Stat.* the minimum and maximum of a dataset, or the distance between the two. 2) *Math.* set of values taken on by a function.

rank 1) *Math.* the maximum number of independent linear combinations that can be made with rows or with columns of a matrix. In mathematics there is a definite rank to a matrix. Computationally the rank is a much fuzzier concept due to **numerical error**—there is not a universally applicable algorithm to get the rank of a matrix. 2) *Stat.* the spot in the ordered sample that an observation falls; ties create fractional values. These are often used for **distribution-free** tests and statistically **robust** estimates. Computed in S with `rank`.

real-time 1) *Comp.* calculations that pertain to the situation at the moment that the calculations are available. For example, a monitor that shows current temperature is real-time, while one that shows the history

of yesterday's temperature is not. The time lag longer than which the computation is not "real-time" is application dependent.

record 1) *Comp.* in a **database**, a collection of **fields** constituting one observation.

rectangular coordinates 1) *Math.* a coordinate system in which the axes are linear and perpendicular to each other. Compare to **polar coordinates**.

recursion 1) *Comp.* a computation that is self-referential—a recursive function calls itself. Some problems are easily solved with recursion, but recursive algorithms tend to be inefficient in terms of computing resources. A recursion can always be recast as an iterative process.

recursive object 1) *S.* an object of one of several **modes** such as `list` and `expression` that can contain an object of the same mode. A list, for example, can have a **component** that is a list.

redirect 1) *Unix.* the action of changing how **standard-in**, **standard-out** and/or **standard-erro**r behave. For example standard-out can be redirected to a file named `my.out`.

register 1) *Comp.* a special memory location where access by the **central processing unit** is extremely fast.

regression test 1) *Comp.* a test of the quality of software that checks if a **bug** occurs that was previously found in the software. 2) It is certainly feasible that this phrase could appear in a statistical context.

regression tree 1) *Stat.* the same as a **classification tree** except that the **response** is **continuous** rather than **categorical**, fit in S with `tree`.

regular expression 1) *Unix.* the pattern matching facility used by `grep` and several other commands. This is not the same as the **wildcarding** that Unix uses with, for example, the `ls` command.

replacement 1) *S.* the use of **subscripts** on the left of an **assignment** to change the values in part of an object. This is using subscripting as an **assignment function**.

required argument 1) *Comp.* an argument to a function that must be given in its calls. When it isn't given, then an **error** results.

reserved word 1) *Comp.* a group of characters that can not be used as a variable name in a language because they mean something particular in the language. Examples of reserved words in S are `return`, `break`, `T`, `NULL`.

response 1) *Stat.* a variable that is approximated in a statistical **model** by a function of one or more **explanatory variables**. This is also confusingly called "dependent variable".

return value 1) *Comp.* the value that is left in place of a function call after it is evaluated. Compare **side effect**.

robust 1) *Stat.* an estimate or hypothesis test that is relatively reliable even when (some) assumptions are violated. For example, the median is robust to outliers while the mean is not. 2) *Comp.* code that works even given extreme conditions. Often this refers to code that is smart enough to retain as much precision as possible. The opposite of **unstable**. 3) *Comp.* an algorithm that works even under adverse conditions. For example, **genetic algorithms** are said to be robust because they work even with multiple local optima and if the objective is not differentiable.

root 1) *Math.* a value that makes a function zero. 2) *Math.* a **node** in a **tree** that is treated specially as the primary node.

rounding 1) *Math.* the process of dropping the least significant digits at some particular location relative to the decimal point. If a 5 is the single digit to be dropped, then it is not clear what to do and there is more than one convention—round to even, and round up are the two most common. Compare **truncate**.

rounding error 1) *Comp.* same as **numerical error**.

row-major order 1) *Comp.* the order of elements in a matrix in which the first row is filled, then the second row, etc. The IMSL C routines use row-major ordering. Compare **column-major ordering**.

RTFM 1) *Comp.* a suggestion to Read The Manual.

rug 1) *Comp.* marks along an **axis** of a plot indicating the **marginal distribution** of the data along that axis.

run-time 1) *Comp.* the time at which a program is used, as opposed to when it is compiled, which is called **compile-time**.

S-news 1) a mailing list for discussion of and questions about S and S-PLUS. At press time, the mechanism to subscribe is to send email to:
`S-news-request@wubios.wustl.edu`
with the body of the message equal to:
`subscribe s-news`
You should get back confirmation in an explanatory message.

sample 1) *Stat.* a selection of observations from some population. Usually used in the sense of a selection from a known population with a specific random mechanism.

scalar 1) *Math.* a single number, as opposed to a **vector** or **matrix**.

scientific notation 1) *Math.* a number written as a number within a certain range times 10 to some power. The number 123.4 is written in scientific notation as 1.234×10^2 or as $.1234 \times 10^3$, depending on the convention used. S (and many other programs) would write it as `1.234e2`.

script 1) *Unix.* a program written in a **shell** language.

search list 1) *S.* the collection of **databases** that an S **session** will search, in order, for objects. The analogue in Unix is the `path` variable.

seed 1) *Comp.* see **random seed**.

selection 1) *Comp.* see **subscript**.

semantics 1) *Comp.* the meaning attached to expressions in a language. The semantics affects the **evaluation** of an expression. Compare **syntax**.

semicolon 1) the character ";" used in S and in C to separate commands.

session 1) *S.* a process that starts when S is started (interactively or in batch) and ends when S is exited.

shell 1) *Unix.* the program that controls the interaction with the user. Examples are the **Bourne shell** and the **C shell**. It is a "shell" because it covers the **kernel** which is unchangeable.

SIC 1) *Stat.* the "Schwarz Information Criterion", sometimes denoted as BIC (but there is another slightly different BIC also). This, like **AIC**, is a way of combining the **likelihood** and the number of parameters so that non-nested models may be compared. In practice, this often suggests smaller models than does AIC.

side effect 1) *Comp.* any action of a program (function) except returning a value. Compare **return value**.

signal 1) *Comp.* a special message given to a program to change what it should do. An **interrupt** signal is an example.

significant digits 1) *Stat.* the number of digits in a number that are not just random noise. 2) *Comp.* used as in "most significant digits" or "least significant digits" (or bits) to mean the digits representing the left-most or right-most places.

simplex algorithm 1) *Comp.* a specific method of solving a **linear programming** problem. 2) *Comp.* an algorithm introduced by Nelder and Mead (1965) to solve general optimization problems without the assumption of differentiability. The algorithm starts with one more solution vector than there are parameters; each step replaces one of the solutions with a new one until a local minimum is found.

simulated annealing 1) *Comp.* a class of optimization algorithm that randomly explores around the current best value, and the size of the random jumps tends to be smaller and smaller as the optimization progresses. The analogy is to a material finding its minimum energy state as it cools. A competitor to **genetic algorithms**.

single quote 1) the character ', almost always on the same key as the **double quote**, not to be confused with the **backquote**.

single-precision 1) *Comp.* a **floating-point number**, generally using 4 bytes. 2) *Comp.* computations done with single-precision numbers.

singular matrix 1) *Math.* a square matrix that is not invertible (has determinant zero). 2) *Comp.* a square matrix that is numerically not invertible. The distinction between singular and non-singular is fuzzy with finite-precision numbers.

singular value decomposition 1) *Math.* the decomposition of a rectangular matrix into the product of an **orthogonal matrix** times a **diagonal matrix** times another orthogonal matrix.

slash 1) the character "/". Used in S and in C to mean division, and in Unix within path names. 2) *Stat.* a probability distribution that is equivalent to a standard Gaussian divided by a uniform(0,1).

smooth 1) *Stat.* a non-linear model of data, often created with a **non-parametric** method. This is most commonly **univariate**, but need not be.

source 1) *S.* the process of making S objects from an **ASCII** representation of the objects. This is done with the `source` function, but more loosely may refer to use of `restore`, `data.restore` and **redirection** of **standard-in**.

source code 1) *Comp.* the software as written in a language like C or Fortran that created a program. This usually refers to code for a **compiled language**, but may also refer to code from an **interpreted language** like S. See also **compile**, **object code**.

sparse matrix 1) *Math.* a matrix that contains a large number of zeros.

special value 1) *Comp.* a number-like entity: an infinity or **not-a-number**.

spline 1) *Math.* a function that is a piece-wise polynomial of a certain degree, often with the constraint that the function and a certain number of its derivatives are continuous.

SQL 1) *Comp.* acronym for "Structured Querry Language", the lingua franca of **database** management programs.

stable 1) *Comp.* a procedure that is **robust** (meaning 2).

stable distribution   1) *Stat.*   a family of distributions that is very interesting from a theoretical view, but has found little practical use except for the **Gaussian** and Cauchy distributions which are special cases.

stable sort   1) *Comp.*   any sorting method that ensures that, in the case of ties in the variables used to sort, observations within a tie retain their original order.

stack 1) *Comp.* a data structure in which items are added to and extracted from the same place, also called FILO for "first in, last out". The image is of a spring-loaded container for plates—new items are pushed onto the stack, and items are popped off of the stack. Compare **queue**. See also **push**, **pop**.

standard deviation   1) *Stat.* the variability of a probability distribution, defined as the square root of the variance.

standard error 1) *Stat.* the variability of an estimate, defined as the square root of the variance.

standard-error 1) *Unix.* the place where error messages go. Usually this is the screen.

standard-in 1) *Unix.* the place where information comes from. Usually this is the keyboard.

standard-out 1) *Unix.* the place where information goes. Usually this is the screen.

star 1) *Comp.*   slang for the asterisk, the character "*". Often used to mean multiplication, but it has other meanings as well.

static load   1) *Comp.* the addition of code (**symbols**) to the **executable** of a program when it is not running. That is, a new, larger executable is created. In S this is done with the `LOAD` utility. Compare **dynamic load**.

statlib 1) a repository of software and other information pertaining to Statistics. It includes a large section of contributed software for S. Accessible on the world wide web with `http://lib.stat.cmu.edu/`.

status 1) *Unix.* the **return value** of a Unix program that indicates the **error** state—zero, by convention, means no error.

storage mode 1) *S.* sub-types of mode `numeric`. The possibilities are `integer`, `double` and `single`.

stream 1) *Comp.* a conceptually unending series of pieces of information. An example is input from a keyboard.

string 1) *Comp.* a sequence of characters.

subscript 1) *Comp.* the extraction (or replacement) of part of an object. In S and C this is done with **brackets**. 2) text that is written lower (and generally in a smaller **font**) than the regular text. In this sense, opposed to superscript.

swap-space 1) *Comp.* portion of disk-space reserved for overflow from **RAM**.

swapping 1) *Comp.* the process of switching information between **RAM** and **swap-space** so that all information required for the pending computation is in RAM. Also known as "paging".

symbol 1) *Comp.* a character string that identifies a **compiled** routine—often with **underscores** added to the routine's name. 2) *S.* the mark put at the location of a datapoint in a plot.

symbol table 1) *Comp.* a table kept by a program that relates **symbols** to **addresses**.

symbolic computation 1) *Comp.* mathematical computation, such as simplification and factoring, where some of the "numbers" are variable names.

symmetric matrix 1) *Math.* a square matrix that is equal to its own **transpose**. Compare **Hermitian matrix**.

syntax 1) *Comp.* the rules governing how expressions are made in a language, without regard to what the expressions mean. The syntax affects the **parsing** of an expression. Compare **semantics**.

system terminating 1) *S.* the action that S takes when it thinks something bad (usually inside calls to `.C` or `.Fortran`) has happened. S kills itself to make sure that no permanent data become corrupted.

tab 1) character written as backslash-t in S and C that represents some number of blank spaces. Counts as **white space**.

table See **contingency table**, **symbol table**, **lookup table**.

tera- 1) prefix meaning 10^{12} (or 2^{40}), as in terabytes of data.

three-dots 1) *S.* the construct "..." used to accept an arbitrary number of **arguments** in a function.

tick 1) *Comp.* mark along the **axis** of a graph indicating a value for that location.

tilde 1) the character "~" used in S to denote a **formula**.

time 1) *Comp.* an ambiguous concept on time-sharing computers since there is a difference between the total elapsed time from start to finish of a task ("wall time") and the time actually expended on the task ("computer time").

time series 1) *Stat.* a dataset where observations are at a sequence of time points, generally evenly spaced time points.

title 1) *S.* one of two **high-level graphics parameters** that label a graph. The `main` title is written at the top of the **figure**, the `sub` title is at the bottom. The titles appear in the **margin**.

token 1) *Comp.* an element of a **parsed** expression.

trace 1) *Math.* the sum of the elements on the diagonal of a matrix. 2) *S.* the process of modifying functions so that they produce some information when called, used for debugging.

transpose 1) *Math.* operation in which the rows of a **matrix** are interchanged with its columns.

tree 1) *Math.* a (connected) mathematical **graph** that contains no loops. See also **binary tree**, **classification tree**, **regression tree**.

trellis 1) *S.* system of **graphics functions** that focus on displaying **multivariate** data in a simple and comprehensible manner.

triangular matrix 1) *Math.* a matrix in which there are all zeros above the diagonal (lower triangular), or below the diagonal (upper triangular).

tridiagonal matrix 1) *Math.* a matrix in which the only non-zero elements are on the diagonal and the diagonals immediately above and below the diagonal. Compare to **diagonal matrix**.

truncate 1) *Math.* a relative of **rounding**, but the least **significant** digits or bits are just dropped with no change to what is left. Comare **rounding**.

type 1) *S.* a **high-level graphics parameter** that specifies how the data are to be represented in the graph. Choices include points, lines, **high-density** or none.

unary 1) *Math.* an operation on a single entity, for example "unary minus" to change the sign of a number. Also called "monadic".

underflow 1) *Comp.* the event in which a non-zero number becomes indistinguishable from zero due to finite-precision arithmetic. Generally, this is of little consequence.

underscore 1) the character "_", used in S to mean **assignment**, used in C within variable names.

univariate 1) *Stat.* a situation involving a single distribution. Compare to **multivariate**.

unstable 1) *Comp.* a computation that can be inaccurate due to **numerical error** or some other phenomenon. For example, dividing by a number close to zero is an unstable operation. The opposite of **robust** (meaning 2).

utility 1) *S.* an operating system command started by S. Examples are BATCH and LOAD.

value 1) *Comp.* often used in the sense of what is returned by a function (**return value**), as opposed to its **side effects**.

vector 1) *S.* an **atomic** object (common usage). 2) *S.* an object with no **attributes** (the definition that as.vector and is.vector use). 3) *S.* an object of one of many **modes** that allows an arbitrary **length**. 4) *Math.* an ordered collection of numbers.

vectorization 1) *S.* a function is vectorized when the output can contain an arbitrary number of answers. 2) *Comp.* computations are vectorized by some (super) computers. 3) *Math.* the columns of a **matrix** stacked on top of each other to form a **vector**. This is indicated by the "vec" operator, and is equivalent to an S matrix being treated as a simple vector.

version 3 1) *S.* the version of S that has existed since the early 1990's.

version 4 1) *S.* the version of S first appearing in 1998 that allows the equal sign to mean assignment and numerous other changes. Confusingly, S-PLUS Version 4.0 is based on Version 3 of S.

vertex (plural *vertices* or *vertexes.*) 1) *Math.* a part of a mathematical **graph** that is usually represented by a point—also called *node*.

virtual memory 1) *Comp.* **swap-space**.

wall time 1) see **time**.

warning 1) *Comp.* an **error**-like condition that does not interrupt computation, but merely issues a message.

white book 1) *S.* Chambers and Hastie (1992) *Statistical Models in S.*

white space 1) *Comp.* any of the characters "space", "tab", and sometimes "newline".

wildcard 1) *Unix.* The style of specification of multiple file names used by ls and some other commands. Compare to **regular expression**.

word 1) *Comp.* a **byte**.

workaround 1) *Comp.* a temporary fix for a **bug**. 2) *Comp.* a method of avoiding a bug by going a different route.

working database   1) *S.* the **database** that is first on the **search list**.

WYSIWYG   1) *Comp.* (pronounced WIZZ-ee-Wig) an acronym for "What you see is what you get."