

What is R?

R is a software package for scientific calculations, graphics, and statistical analyses.

R is free and open source.

Versions of R are available for Windows, Mac OS, and Linux/Unix operating systems. (download R here: <https://www.r-project.org/>)

R is in the form of a command-driven programming language with vast statistical, mathematical, and graphical resources built in as pre-programmed functions:

- R has vector and matrix calculations built in (like Matlab).

- R features fantastic graphical displays of data.

- R has calculations (random numbers, percentiles, probabilities, etc.) built in for many different statistical distributions.

- R has special mathematical functions of all sorts built in (complex numbers, gamma & log-gamma function, numerical integration, optimization, etc.).

R has grown an enormous, international group of users (critical mass).

Lots of online forums for getting help.

Many tutorial websites (often posted by university instructors for courses).

Many instructional videos (Youtube especially).

Many books.

Online courses. Even some MOOCs.

Vast numbers of contributed R routines, posted by users as functions, for producing the latest specialized statistical analyses (genetics, finance/investment, multivariate/social sciences, data mining, etc.)

History of R

Bell Laboratories (now Nokia Bell Laboratories) invented/discovered:

radio astronomy, transistor, information theory, charge-coupled device, laser, cosmic microwave background

Unix, C, C++

The S statistical programming language (John Chambers & others).

S-PLUS

MathSoft distributes commercial version of S, called S-PLUS (around 1995).

S-PLUS spreads rapidly in the statistics world; features dazzling (at the time) graphics such as rotating 3D plots and painting of scatter plots, along with statistical programming for numerically intensive research.

R created by Ross Ihaka and Robert Gentleman (later 1990s).

Uses the syntax of S-PLUS & S (syntax can't be copyrighted).

Free, open source, part of the GNU project.

Currently developed by the *R Development Core Team*
(John Chambers is a member).

The R Journal is started in 2001. Many R users by 2005.

"Data analytics" becomes big in the late 2000s; NYT features an article about R in 2009.

Early reputation was that R was slow (in comparison to Matlab), but recent versions of R are substantially speeded up. (Microsoft has issued a free distribution of R that is reportedly quite fast)

R is an *interpreted* programming language rather than a *compiled* language.

R competitors:

S-PLUS (expensive).

Matlab (expensive).

Gauss (expensive).

Octave (free, copies Matlab syntax, graphics not as extensive).

Python (free interpreted programming language, with features that programmers like more than R, but beginner level documentation is poor, and the library of user-contributed statistical routines is much smaller).

SAS (expensive but has established itself in the 70s, 80s, and 90s as the most extensive and powerful statistical analysis and data management software).