

# R Background



# R Background

---

## **R is an open-source, object-oriented statistical programming language**

- History:

- R is based on the S statistical programming language developed by John Chambers at Bell labs in the 1980's
- The commercial package S-plus is based on the S language
- R is an open-source implementation of the S language
- Developed by Robert Gentleman and Ross Ihaka in New Zealand

- Features:

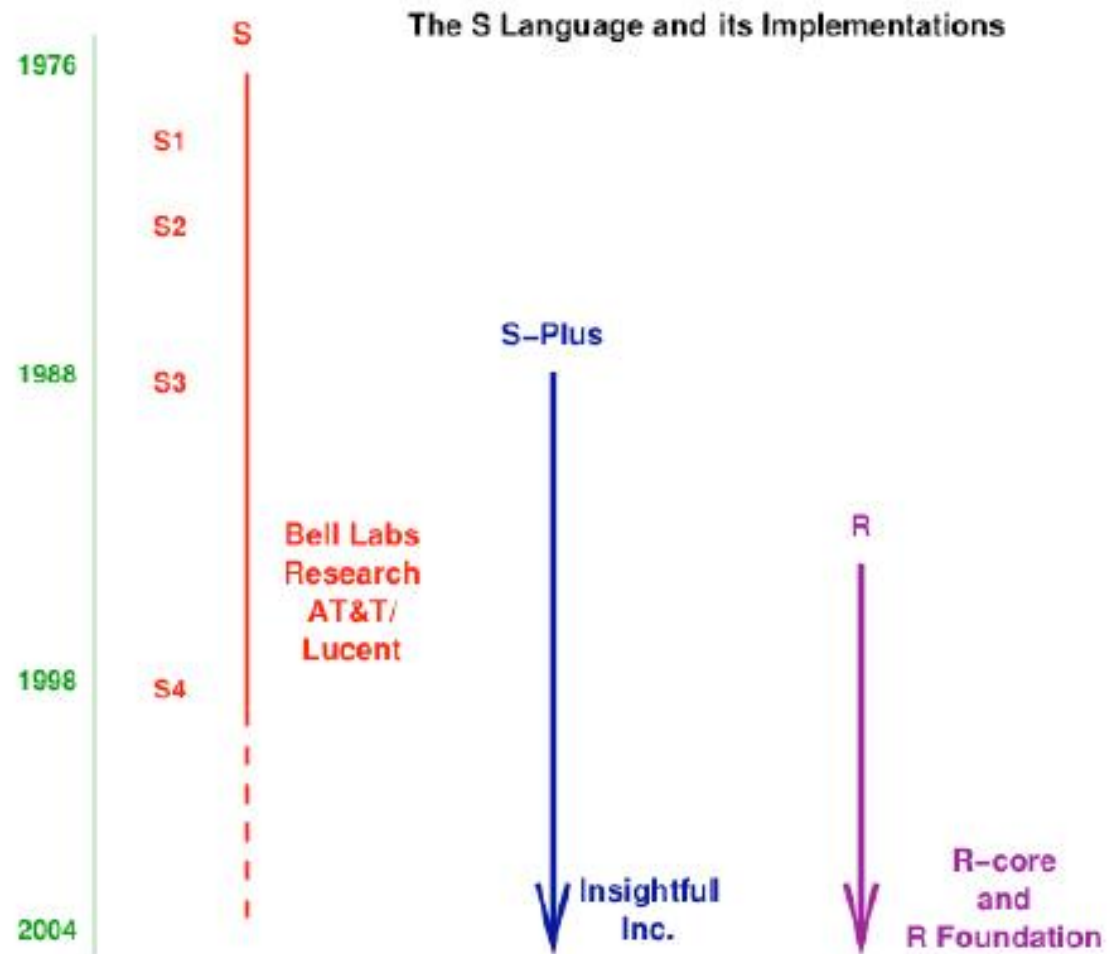
- R is a high-level, object-oriented programming environment
- R has advanced graphical capabilities
- Statisticians around the world contribute add-on packages... therefore:

“The great beauty of R is that you can modify it to do all sorts of things,” said Hal Varian, chief economist at Google. “And you have a lot of prepackaged stuff that’s already available, so you’re standing on the shoulders of giants.”

# R Evolution

---

- S is the original language
- S-plus is a commercial implementation of S
- R is an open-source implementation of S
- R is very similar to, but not identical with, other implementations of S



# Facets of R (why an actuary would care!)

---

1. An *interface* to computational procedures of many kinds
2. *Interactive*, hands-on in real time
3. *Modular*, built from standardized pieces
4. *Collaborative*, a world-wide, open-source effort

# Growing interest in R

---

- August 2006



## Limited Attendance Seminar on Predictive Modeling

August 15-16, 2006  
Chicago, Illinois

**NEW!** The Limited Attendance Seminar on Predictive Modeling is now full. No additional registrations will be taken at this time. Thank you for your interest.

### PURPOSE

This two-day limited attendance seminar is designed to impart an understanding of predictive modeling applications in insurance. Examples of personal lines ratemaking and commercial lines underwriting applications will be covered in detail using simulated data. The seminar format will be a mixture of lecture, demonstration, and hands-on exercises. Pre-assignments will be given to the participants to familiarize them with some of the concepts to be covered. Datasets that will be used during the seminar will also be provided.

The seminar will cover both theoretical and practical aspects of insurance predictive modeling. Topics to be covered will include:

- Data scrubbing and manipulation
- Modeling methodologies
- Model validation
- Interpreting and analyzing model results

The statistical software that will be used during the seminar is the widely used, freely available shareware package "R". Prior to the seminar, students will be provided with instructions on how to install and use R. Participants are expected to bring their own laptop to the seminar with the R software installed.

The instructors are James Guszczka, Ph.D., FCAS, and Jun Yan, Ph.D. Both instructors are members of Deloitte Consulting LLP's Advanced Quantitative Services group.

# Growing interest in R

---

- November 2006

<http://www.casact.org/newsletter/index.cfm?fa=viewart&id=5311>

The screenshot shows the header of 'The Actuarial Review' newsletter. At the top right, there is a link: 'CLICK HERE TO DOWNLOAD A .PDF VERSION OF THIS NEWSLETTER.'. The title 'The Actuarial Review' is in a dark blue box with a logo. Below the title is a link 'RETURN TO MAIN PAGE'. The main article title is 'The R Programming Language—My “Go To” Computational Software' by Glenn Meyers, which is circled in red. The article text follows, discussing the author's experience with R and its use in actuarial work.

CLICK HERE TO DOWNLOAD A .PDF VERSION OF THIS NEWSLETTER.

**The Actuarial Review**

[RETURN TO MAIN PAGE](#)

**The R Programming Language—My “Go To” Computational Software**  
Glenn Meyers

My involvement in a number of predictive modeling projects in the past few years has given me the opportunity to work with professional statisticians. These statisticians introduced me to something that I believe will be useful to many actuaries.

The R programming language is a software environment for statistical computing and graphics. R is widely used for statistical software development and data analysis. R’s source code is free and available at the Web Site [www.r-project.org](http://www.r-project.org) where precompiled binary versions are provided for Microsoft Windows, Mac OS X, and other UNIX-like operating systems.

R is the result of a collaborative effort with contributions from all over the world. R was initially written by Robert Gentleman and Ross Ihaka—also known as “R & R” of the Statistics Department of the University of Auckland. Since mid-1997 there has been a core group with access to write the actual source code for R.

R supports a wide variety of statistical and numerical techniques. R is also highly extensible through the use of packages, which are user-submitted libraries for specific functions or areas of study. A core set of packages are included with the installation of R, with over 700 more available at the [Comprehensive R Archive Network](#) (CRAN) as of 2006.

# Growing interest in R

---

- January 2009

[http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?\\_r=1&pagewanted=print](http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=1&pagewanted=print)

**The New York Times**

PRINTER-FRIENDLY FORMAT  
SPONSORED BY

**(500) DAYS OF  
SUMMER**

This copy is for your personal, noncommercial use only. You can order presentation-ready copies for distribution to your colleagues, clients or customers [here](#) or use the "Reprints" tool that appears next to any article. Visit [www.nytreprints.com](http://www.nytreprints.com) for samples and additional information. [Order a reprint of this article now.](#)

January 7, 2009

## Data Analysts Captivated by R's Power

By [ASHLEE VANCE](#)

To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca partly because data mining has entered a golden age, whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as [Google](#), [Pfizer](#), [Merck](#), [Bank of America](#), the InterContinental Hotels Group and Shell use it.

But R has also quickly found a following because statisticians, engineers and scientists without computer programming skills find it easy to use.

"R is really important to the point that it's hard to overvalue it," said Daryl Pregibon, a research scientist at Google, which uses the software widely. "It allows statisticians to do very intricate and complicated analyses without knowing the blood and guts of computing systems."

# Growing interest in R

---

- September-October 2013

Source: Actuarial Review

---

THE LATEST IN ... R BY LEE BOWRON

## Using the Open-Source Statistical Language for Actuarial Work

In March 2013, the Comprehensive R Archive Network (CRAN) released a major new version of R, the free software programming language and environment for statistical computing and graphics. R 3.0.0 takes better

advantage of 64-bit memory and allows data structures that exceed two billion records!

If you've been reluctant to try R because of memory issues, you might want to try the new version of R. If you've

been running an old version, be sure to update your packages under the new version using the following command: **update.packages(checkBuilt=TRUE)**. (Note: Throughout this article, R commands and codes will be in bold type.)

