Welcome to TimesPeople
What's this?
TimesPeople Lets You Share and Discover the Best of NYTimes.com 9:49 AM

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS | Get Home D

The New York Times
Business Computing
Search

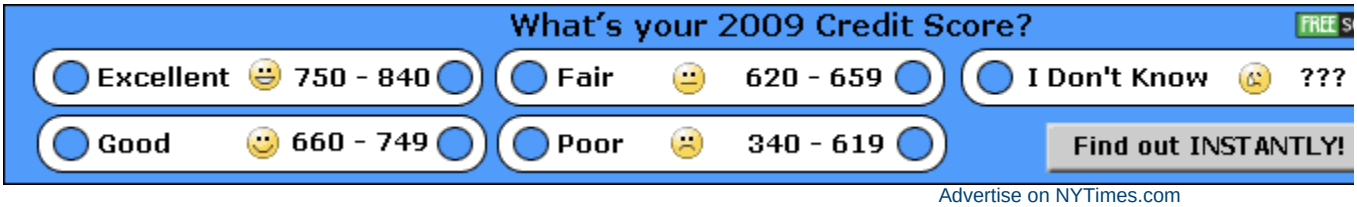
WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | JO
AUTOS

Search Technology
Go

Inside Technology
Internet | Start-Ups | Business Computing | Companies

Bits Blog »

Personal Tech »
Cellphones, Camera



Advertise on NYTimes.com

Data Analysts Captivated by R's Power

More Articles in Technology »

Stuart Isett for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.

By ASHLEE VANCE
Published: January 6, 2009

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.

SIGN IN TO E-MAIL OR SAVE THIS
PRINT
SINGLE PAGE
REPRINTS

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.

SHARE

ARTICLE TOOLS SPONSORED BY
MICKEY ROURKE

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.”

It is also free. R is an open-source program, and its popularity reflects a shift in the type of software used inside corporations. Open-source software is free for anyone to use and modify. I.B.M., Hewlett-Packard and Dell make billions of dollars a year selling servers that run the open-source Linux operating system, which competes with Windows from Microsoft. Most Web sites are displayed using an open-source application called Apache, and companies increasingly rely on the open-source MySQL database to store their critical information. Many people view the end results of all this technology via the Firefox Web browser, also open-source software.

R is similar to other programming languages, like C, Java and Perl, in that it helps people perform a wide variety of computing tasks by giving them access to various commands. For statisticians, however, R is particularly useful because it contains a number of built-in mechanisms for organizing data, running calculations on the information and creating graphical representations of data sets.

Some people familiar with R describe it as a supercharged version of Microsoft’s Excel spreadsheet software that can help illuminate data trends more clearly than is possible by entering information into rows and columns.

What makes R so useful — and helps explain its quick acceptance — is that statisticians, engineers and scientists can improve the software’s code or write variations for specific

Politics E-Mail

Keep up with the 2008 p
daily Politics e-mail new

See Sample | Privacy P

Subscribe to Technology I

Technology News
Internet
Start-Ups
Business Computing
Companies
Bits Blog
Personal Tech
Pogue’s Posts

MOST POPULAR - TECHNOLOGY

E-MAILED | BLOGGED

1. Want to Copy iTunes Mus
2. Data Analysts Captivated
3. The Facebook Virus Sprea Safe
4. Doctors Will Make Web C
5. Blu-ray’s Fuzzy Future
6. Nine Web sites IT pros sh
7. Basics: There’s Lots of Te
8. Google Hopes to Open a T
9. Apple’s Jobs Explains His
10. In Silicon Valley, Venture and Focus on the Short T

Go to Complete List »

The New York Times

tasks. Packages written for R add advanced algorithms, colored and textured graphs and mining techniques to dig deeper into databases.

Close to 1,600 different packages reside on just one of the many Web sites devoted to R, and the number of packages has grown exponentially. One package, called BiodiversityR, offers a graphical interface aimed at making calculations of environmental trends easier.

Another package, called Emu, analyzes speech patterns, while GenABEL is used to study the human genome.

The financial services community has demonstrated a particular affinity for R; dozens of packages exist for derivatives analysis alone.

“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 first appeared in 1996, when the statistics professors Ross Ihaka and Robert Gentleman of the University of Auckland in New Zealand released the code as a free software package.

According to them, the notion of devising something like R sprang up during a hallway conversation. They both wanted technology better suited for their statistics students, who needed to analyze data and produce graphical models of the information. Most comparable software had been designed by computer scientists and proved hard to use.

Lacking deep computer science training, the professors considered their coding efforts more of an academic game than anything else. Nonetheless, starting in about 1991, they worked on R full time. “We were pretty much inseparable for five or six years,” Mr. Gentleman said. “One person would do the typing and one person would do the thinking.”

Some statisticians who took an early look at the software considered it rough around the edges. But despite its shortcomings, R immediately gained a following with people who saw the possibilities in customizing the free software.

John M. Chambers, a former Bell Labs researcher who is now a consulting professor of statistics at Stanford University, was an early champion. At Bell Labs, Mr. Chambers had helped develop S, another statistics software project, which was meant to give researchers

The Critics' picks for th
Also in Movies:
How Brad Pitt aged with Benja
Caped crusading
Continuous coverage of award

ADVERTISEMENTS

Need to know more?
Get 50% off home delivery of The Times.

The New York

"Robert Joh
Buy Now

Ads by Google

Business Analytics
Learn More About Business Ana

of all stripes an accessible data analysis tool. It was, however, not an open-source project.

The software failed to generate broad interest and ultimately the rights to S ended up in the hands of Tibco Software. Now R is surpassing what Mr. Chambers had imagined possible with S.

“The diversity and excitement around what all of these people are doing is great,” Mr. Chambers said.

1 | 2 NEXT PAGE »

A version of this article appeared in print on January 7, 2009, on page B6 of the New York edition.

More Articles in Technology »

Click here to enjoy the convenience of home delivery of The Times for less than $1 a day.

Ads by Google what's this?

Congruent Software WA
Project Accounting Software Microsoft Gold Certified Partner
www.CongruentSoft.com

University of Phoenix
Get A Degree In Soft. Engineering Flexible Hours, Online Or On Campus
Phoenix.edu

Data Mining Software
Data Mining Tools & Applications To Improve Decision Making. Free Demo!
www.SPSS.com/Data_Mining

www.SAS.com

Open Source Software
Innovate with Linux on Intel. Gra
www.Intel.com

Data Mining Software
Data Mining Tools & Applications
Free Demo!
www.SPSS.com/Data_Mining

Related Articles

FROM THE NEW YORK TIMES

INSIDE NYTIMES.COM

TELEVISION »

REAL ESTATE »

DINING & WINE »

OPINION »
Editorial: America’s Colonial Capital
Washington deserves full voting rights in Congress.

POLITICS »

Welcome to TimesPeople
What's this?
TimesPeople Lets You Share and Discover the Best of NYTimes.com 9:52 AM

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS | Get Home D

The New York Times
Business Computing
Search

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | JO
AUTOS

Search Technology
Go

Inside Technology
Internet | Start-Ups | Business Computing | Companies

Bits Blog »

Personal Tech »
Cellphones, Camera

Advertise on NYTimes.com

Data Analysts Captivated by R's Power

Published: January 6, 2009

SIGN IN TO E-MAIL
OR SAVE THIS
PRINT
SINGLE PAGE
REPRINTS
SHARE
ARTICLE TOOLS SPONSORED BY
MICKEY ROURKE

More Articles in Technology »

(Page 2 of 2)

While it is difficult to calculate exactly how many people use R, those most familiar with the software estimate that close to 250,000 people work with it regularly. The popularity of R at universities could threaten SAS Institute, the privately held business software company that specializes in data analysis software. SAS, with more than $2 billion in annual revenue, has been the preferred tool of scholars and corporate managers.

"R has really become the second language for people coming out of grad school now, and there's an amazing amount of code being written for it," said Max Kuhn, associate director of nonclinical statistics at Pfizer. "You can look on the SAS message boards and see there is a proportional downturn in traffic."

SAS says it has noticed R's rising popularity at universities, despite educational discounts on its own software, but it dismisses the technology as being of interest to a limited set of people working on very hard tasks.

"I think it addresses a niche market for high-end data analysts that want free, readily

available code," said Anne H. Milley, director of technology product marketing at SAS. She adds, “We have customers who build engines for aircraft. I am happy they are not using freeware when I get on a jet.”

But while SAS plays down R’s corporate appeal, companies like Google and Pfizer say they use the software for just about anything they can. Google, for example, taps R for help understanding trends in ad pricing and for illuminating patterns in the search data it collects. Pfizer has created customized packages for R to let its scientists manipulate their own data during nonclinical drug studies rather than send the information off to a statistician.

The co-creators of R express satisfaction that such companies profit from the fruits of their labor and that of hundreds of volunteers.

Mr. Ihaka continues to teach statistics at the University of Auckland and wants to create more advanced software. Mr. Gentleman is applying R-based software, called Bioconductor, in work he is doing on computational biology at the Fred Hutchinson Cancer Research Center in Seattle.

“R is a real demonstration of the power of collaboration, and I don’t think you could construct something like this any other way,” Mr. Ihaka said. “We could have chosen to be commercial, and we would have sold five copies of the software.”

« PREVIOUS PAGE 1 | 2

A version of this article appeared in print on January 7, 2009, on page B6 of the New York edition.

More Articles in Technology »

Click here to enjoy the convenience of home delivery of The Times for less than $1 a day.

Ads by Google what's this?

Congruent Software WA
Project Accounting Software Microsoft Gold Certified Partner
www.CongruentSoft.com

University of Phoenix
Get A Degree In Soft. Engineering Flexible Hours, Online Or On Campus
Phoenix.edu

Data Mining Software
Data Mining Tools & Applications To Improve Decision Making. Free Demo!
www.SPSS.com/Data_Mining

Breaking News Alerts by
Sign up to be notified wh
Privacy Policy

Subscribe to Technology R
Technology News
Internet Start-Ups
Business Computing Companies
Bits Blog
Personal Tech
Pogue’s Posts

MOST POPULAR - TECHNOLOGY
E-MAILED BLOGGED
1. Want to Copy iTunes Musi
2. Data Analysts Captivated b
3. The Facebook Virus Sprea Safe
4. Doctors Will Make Web Ca
5. Blu-ray’s Fuzzy Future
6. Nine Web sites IT pros sho
7. Basics: There’s Lots of Tec
8. Google Hopes to Open a T
9. Apple’s Jobs Explains His
10. In Silicon Valley, Venture and Focus on the Short Te
Go to Complete List »

The New York Times