

Comparing S-PLUS® and R

Insightful Corporation
info@insightful.com
206-283-8802 • 800-569-0123

Edition 1.0

Abstract

The choice of a data analysis platform is a strategic, long-term decision that has a dramatic impact on an organization's competitiveness. The platform must be flexible enough to support an organization's changing needs. Today, this often means choosing between S-PLUS and SAS, or exploring one of the freeware data analysis languages available such as R. This paper compares S-PLUS to its alternatives and describes where each fits in real-world business and research applications.

Table of Contents

1	ONLY S-PLUS® TAKES YOU FROM PROTOTYPING TO PRODUCTION	3
2	S-PLUS ENHANCES PRODUCTIVITY TO REDUCE TIME AND COST OF PROTOTYPING ANALYTIC APPLICATIONS	4
	2.1 S-PLUS's GUI is designed to boost productivity	4
	2.2 Interactive graphics and visualization with support for Microsoft Office®	5
	2.3 S-PLUS provides access to your source data, whatever the format.....	5
	2.4 S-PLUS's persistent environment helps manage large projects.....	5
	2.5 S-PLUS delivers the newest methods in a validated environment for best-in-class competitive advantage.....	5
	2.6 Specialized toolkits for industry-specific problems	6
3	S-PLUS WORKS WITH LARGER DATA SETS THAN R	7
	3.1 Benchmark Methodology.....	7
	3.2 Regression applications.....	8
	3.3 Training Trees	11
	3.4 Variance/Covariance Calculations.....	12
	3.5 Insightful Miner scales S-PLUS functions to very large data sets.....	13
4	PRODUCTION S-PLUS APPLICATIONS CAN BE EASILY DEPLOYED WITHIN YOUR EXISTING INFRASTRUCTURE	14
	4.1 S-PLUS® Server for Web-based or client/server deployment.....	14
	4.2 Insightful Miner deploys applications to non-technical users.....	14
	4.3 S-PLUS has tools for producing reports in HTML, PDF or RTF formats.....	15
	4.4 S-PLUS is the best choice of embedded analytic engine for third-party applications ...	15
	4.5 You must install, compile and maintain R yourself.....	16
5	S-PLUS FULLY COMPLIES WITH INDUSTRY MANDATES. R DOES NOT.	16
	5.1 S-PLUS is validated software to ensure accuracy, reliability and consistency.....	16
	5.2 Insightful services can assist your compliance efforts	16
	5.3 S-PLUS is professionally engineered and supported.....	16
	5.4 Newest statistical methods are made enterprise-ready	17
6	SUMMARY	17
7	ABOUT INSIGHTFUL	18

1 Only S-PLUS® takes you from prototyping to production

The choice of a data analysis platform is a strategic, long-term decision that has a dramatic impact on an organization's competitiveness. The platform must be flexible enough to support an organization's changing needs. Today, this often means choosing between S-PLUS and SAS, or exploring one of the freeware data analysis languages available such as R. This paper compares S-PLUS to its alternatives and describes where each fits in real-world business and research applications.

Although several data analysis programming languages provide the "core" features of statistical algorithms and graphics, businesses need a tightly integrated approach to the design, development, deployment and maintenance of analytic applications to achieve a strong ROI. Companies also seek a validated IT infrastructure and ready expertise from their key technology suppliers, looking to them for top-notch training, support, consulting and "whole-product" capabilities required to achieve the highest efficiency and competitiveness.

Basic Research	Applied Research	Prototype	Validated Production Systems	Packaged Application
R				
S-PLUS and Insightful Miner (Desktop)				
	Insightful Miner and S-PLUS (Servers)			
		Connect/C++/Java		
			SAS	

Only the S-PLUS product family combines modern data analysis programming and visualization techniques for creating new analytics with the ability to deploy these applications into validated production settings.

S-PLUS is a fully-featured commercial data analysis package from Insightful Corp. (NASDAQ:IFUL), a company with a 16-year track record of delivering business-critical software applications to analysts and their senior managers in biopharmaceuticals, finance, manufacturing, telecommunications and government. S-PLUS is unique in that applications built with the desktop version can be efficiently deployed into highly scalable production environments using S-PLUS Server or Insightful Miner. Unlike older languages like SAS, S-PLUS's modern programming paradigm and advanced visualization tools easily support the creation of new applications tailored to individual tasks, and S-PLUS has a rich history of innovation with hundreds of books and papers published describing ground-breaking statistical methods and analysis written using its language.

R is a freeware programming language initially written in the mid 1990's at the University of Auckland's Department of Statistics in New Zealand. It is supported by The R Foundation, a not-for-profit whose general goals are "the exploration of new methodology, teaching and training of statistical computing and the organization of meetings and conferences with a statistical computing orientation." R is popular in academic and research institutions due to its free nature and a strong community that supports the advancement of the statistical sciences and the R programming language.

Both S-PLUS and R are implementations of the "S" programming language. In fact, the languages are similar enough that many applications written in R are easily ported to S-PLUS. S is a very high level language and environment for data analysis and graphics. In 1998, the Association for Computing Machinery (ACM) presented its Software System Award to the designers of S noting they "forever altered the way people analyze, visualize, and manipulate data."

Although statistics and graphics are still vital requirements, organizations today rely on their data analysis platform to provide many additional capabilities such as:

- Ease of use with graphical and visual workflow interfaces tailored to the task at hand
- Speed and scalability to efficiently handle larger data sets
- Simplified access to a wide variety of data sources
- A validated and qualified IT infrastructure to support scalable data management, analysis and reporting; as well as submissions to regulatory organizations.
- Deployment of interactive and batch analytical applications to key decision makers around the organization

Throughout this technical comparison, S-PLUS and R are evaluated based on their ability to provide a solid, integrated infrastructure for accessing, analyzing and reporting on the data that drives business and research decisions.

2 S-PLUS enhances productivity to reduce time and cost of prototyping analytic applications

The S-PLUS product family is the only data analysis platform that supports rapid development of analytic applications that can scale to medium and very large data sets. Unlike SAS, both S-PLUS and R provide a modern and highly flexible programming environment that enables you to build custom functions and applications quickly and affordably. S is the only language specifically designed for data visualization and modeling. It has over 4,200 built-in functions that can be modified and adapted to your specific situation. With S-PLUS, you can operate on your data by rows and by columns. It includes numerous object-oriented data types for vectors, matrices, data arrays, lists, etc., and calling your custom routines from other programs is a made simple with S-PLUS's built-in APIs.

A downside with R is that it has only a rudimentary Graphical User Interface (GUI), available only on Windows systems. By contrast, S-PLUS provides users with both a command line interface (CLI) and a comprehensive GUI on both Windows and UNIX systems for greater productivity. S-PLUS also integrates tightly with Insightful Miner – a highly scalable visual programming environment – to easily create guided analytic workflows that can process data sets of arbitrarily large size.

2.1 S-PLUS's GUI is designed to boost productivity

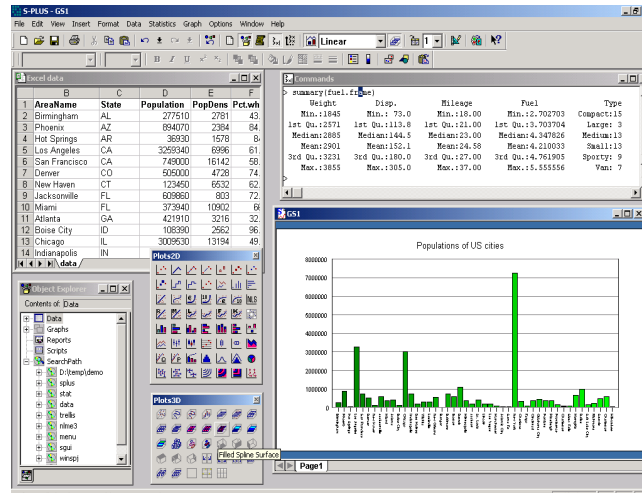
S-PLUS's graphical user interface (GUI) is menu-driven, providing easy access to data manipulation, statistics and graphical functionality. R has only a rudimentary visual interface and requires most interaction to be done via the command line. This means new users can become productive with S-PLUS much more quickly, with access to the command line still available when needed or desired.

S-PLUS's graphical user interface is designed to promote productivity, making commonly-performed tasks convenient and easy to accomplish. Routine tasks like importing data from flat files or databases are better suited to a menu-based interface, where it's easier to select from a wide range of file formats and import options. In S-PLUS, a visual data previewer helps the user validate that the correct data file is being accessed, and that the data is being imported with the correct specifications. Of course, if a file is being imported regularly with the same options, the S-PLUS GUI allows the user to automate this task by converting the interactive use of the dialog into a script in the S language, which can then be run as needed or even linked to a custom toolbar button for instant access.

To apply the full power of statistical analysis to your data, menus and dialogs are restrictive compared to the flexibility offered by the S programming language. S-PLUS fully supports direct access to the programming capabilities of S, with an interactive command line including convenient features like command recall and history logging, plus a script editor which allows you to create, test, and run complete scripts within the graphical environment. R has no facilities for creating or editing scripts that interface closely with the graphical user interface. In short, by combining a menu and dialog graphical user interface with complete access to the command line and script editing environment, S-PLUS gives you the best of both worlds.

2.2 Interactive graphics and visualization with support for Microsoft Office®

Only with S-PLUS can you easily create and modify your graphics with a point-and-click interface, allowing you the fine control needed to create charts to exacting specifications. R requires all modifications be done through the command line. S-PLUS uniquely allows you to automatically create a PowerPoint presentation directly from within your data analysis environment. Furthermore, S-PLUS offers Graphlets™ for interactive, drill-down graphics you can include on web pages; no comparable facility exists in R.



The S-PLUS GUI can be easily customized and lets new users become productive fast.

2.3 S-PLUS provides access to your source data, whatever the format

S-PLUS can directly import more than 30 file types, including ASCII, SAS, SPSS and Excel. You can even open Excel files directly, and apply statistical and graphical features directly to data in your worksheet. S-PLUS also includes direct access to Oracle, Sybase, DB2 and SQL Server databases, in addition to ODBC interfaces to access all major database formats. Packages for file and database access for a few of the data formats supported by S-PLUS are available with R, but there are often difficulties with using them. For example, reading SAS files with R requires an available SAS license, unlike when reading SAS files with S-PLUS.

2.4 S-PLUS's persistent environment helps manage large projects

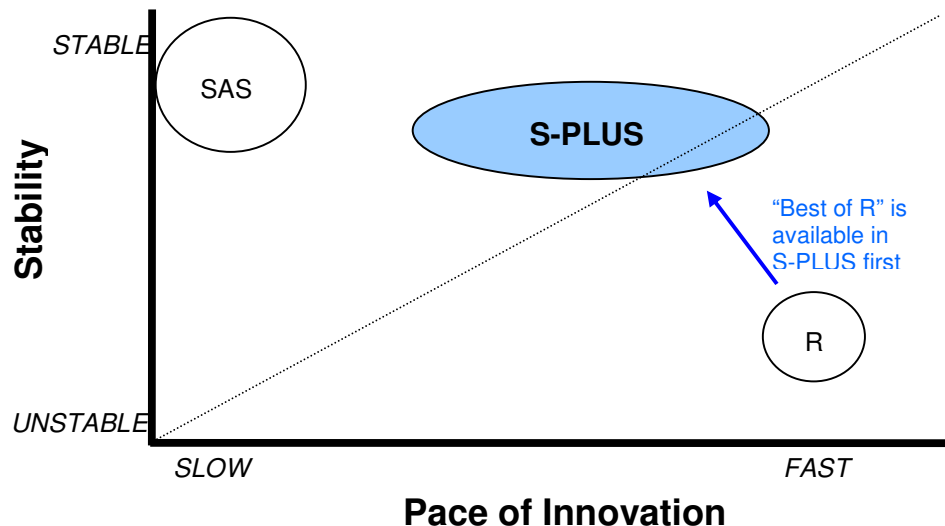
R stores all of its working data and objects in memory, unlike S-PLUS which saves all permanent objects to disk. Not only does this mean that S-PLUS requires less memory when working with large numbers of objects in a project, it also means that you can exit S-PLUS at any time and then return to the same project later with all of the objects remaining intact. This makes it much easier to pick up on a project from where you left off. By contrast, because R keeps all objects in memory and not on disk, if the computer or R crashes, all of your objects are lost.

2.5 S-PLUS delivers the newest methods in a validated environment for best-in-class competitive advantage

R is popular in academic and research circles for those advancing the state-of-the-art in statistical analysis. This highly dynamic environment can spur dramatic improvements in how data is gathered, processed and analyzed across many fields of study. Over the years, some of these new methods will follow the technology adoption curve and move from research to the industry's innovators and then into general acceptance and usage. Like many freeware packages, R can be updated at a rapid-fire pace, so bugs might be fixed faster once identified, but new bugs can be introduced faster as well.

S-PLUS also has a long history of innovation. And because of the close similarity of their base languages, packages written in R can be easily ported to S-PLUS. But S-PLUS is commercial-grade software that goes through extensive quality assurance cycles to validate its results and compatibility with all its available platforms and add-on products. So rather than asking users to stay abreast of frequent or sporadic release cycles, new versions of S-PLUS are distributed on a more even time table. This ensures platform stability and better matches the pace at which business users can effectively deploy updates to a core engine running important applications.

S-PLUS balances rapid innovation in its features with customer's need for a highly stable development and deployment platform



When porting R packages to S-PLUS, Insightful has corrected a number of bugs that would otherwise have gone unnoticed. Phillip Stafford, CEO of BioMining.com said: "The work Insightful did validating and fixing the R team's Bioconductor package [to create the Insightful product S+ArrayAnalyzer] is a good example of the quality that S-PLUS brings to the table versus the free software community."

2.6 Specialized toolkits for industry-specific problems

S-PLUS toolkits provide unique functionality that address analysis problems in areas such as: bioinformatics, econometrics, optimization, clinical trials analysis, environmental statistics and spatial data analysis. Many of these modules provide functionality that is not available in the core R distribution or in contributed libraries:

- **S+ArrayAnalyzer™**: An affordable, accurate, and highly adaptable microarray analysis module.
- **S+FinMetrics™**: Modern and flexible analytics for powerful econometric analysis.
- **S+NuOPT™**: For cutting-edge portfolio optimization, nonlinear and robust statistical modeling and circuit optimization.
- **S+SeqTrial™**: For designing, monitoring, and analyzing clinical trials using group sequential methods.
- **S+EnvironmentalStats**: Powerful yet simple-to-use functions for analyzing environmental data.
- **S+Wavelets™**: A wavelets toolkit for problems involving signals, images and financial time series data.
- **S+SpatialStats™**: A comprehensive statistical package for analyzing spatially correlated data.
- **S-PLUS for ArcView GIS**: Integrates the powerful statistics and graphics of S-PLUS into ArcView GIS 3.2

In addition, Insightful's team of research scientists provide additional libraries than can be downloaded and used with S-PLUS. These research libraries provide additional functionality in evolving areas of

statistical research, including correlated data analysis, Bayesian modeling, resampling techniques and other areas. S-PLUS users can download these libraries from www.insightful.com/downloads/libraries.

3 S-PLUS works with larger data sets than R

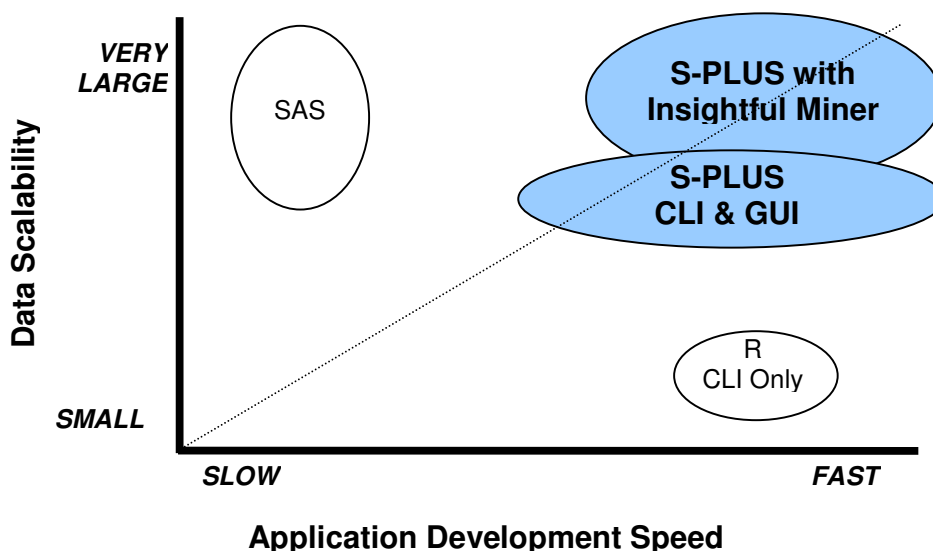
For production usage, it is critical that analytic applications run reliably, with the necessary performance levels, and on data sets of the size typically found in production environments. In this environment, the fact that S-PLUS can handle much larger data sets than R on the same architecture, and completes many common analytic problems on large data sets in a shorter amount of time compared to R, makes it the ideal choice for production deployment of analytics created with the S language.

S-PLUS is faster in typical modeling applications and can process larger data sets than R. Insightful's performance benchmarks (presented below) echo an independent comparison of S-PLUS and R by one of the world's largest three banks. That study found that S-PLUS was faster than R for seven out of nine standard data analysis procedures on medium-sized data sets (10,000 rows x 50 columns). For large data sets (20,000 rows x 100 columns), R crashed with an out-of-memory error for seven out of the nine tests, whereas S-PLUS computed results in a timely manner for all.

Insightful's benchmarking results show that S-PLUS is consistently faster than R in all data set sizes for typical linear and logistic regression problems, with R frequently unable to complete the analysis with larger data sets (over 40,000 rows and over 150 columns). Training regression and classification trees show that R is slightly faster for small data set sizes, but its advantages shrink as data sizes grow, and again it cannot complete analysis on larger data sets. On a typical variance/covariance calculation, we see that S-PLUS and R perform similarly throughout the ranges tested.

Scaling S-PLUS routines and functions to analyze arbitrarily large, gigabyte-class data sets is available through Insightful Miner, Insightful's highly scalable visual data analysis and deployment environment. There is no similar functionality in R.

Only the S-PLUS product family supports rapid development of new applications that are capable of analyzing medium to very large datasets.



3.1 Benchmark Methodology

We compare S-PLUS (version 6.2 for Windows) and R (version 1.8.1 for Windows) for five standard statistical analysis: ordinary least squares regression, logistic regression, training of a classification tree, training of a regression tree, and calculation of a sample variance/covariance matrix.

To create these benchmarks, a random data set of a given size was created and then saved to disk. The random data set consisted of one response column, and p explanatory variable columns, each with n rows. The explanatory variable columns were drawn from independent random Normal variates, and the response column was a suitably transformed linear combination of the explanatory variable columns plus random Normal noise.

All calculations were performed using S-PLUS Version 6.2 for Windows and R version 1.8.1 for Windows, on hardware with the following specifications:

- Pentium III 550MHz
- 768 MB RAM
- 112GB Hard Disk
- Windows XP Professional operating system

The elapsed time to perform the relevant calculation (with no other applications running) is measured in seconds. The machine was restarted between each test to maximize reliability.

3.2 Regression applications

Here we compare how S-PLUS and R perform when computing a standard linear regression and a logistic regression against various data set sizes. The regression and logistic regression analyses were then performed in each system with the respective commands:

```
lm(Y ~ ., data = test.data )
glm( Y ~ ., family=binomial(link=logit), data = test.data )
```

For a given data set of N rows and p columns, the data frame `test.data` was created using the function `dgen.lm` and `dgen.glm` below for linear regression and logistic regression respectively, which in turn use the functions `dgen.X` and `dgen.Y` to create the matrix of explanatory variables and response vector, respectively. This data frame was then saved to disk, and the same data used for analysis in both S-PLUS and R.

```
"dgen.X" <-
## inputs: n = number of rows, p = number of columns
## outputs: returns X = [x_1 x_2 ... x_p], where
## x_i ~ rnorm(n,0,1)
function( n=60, p=2, seed=227 ){
  ## make a matrix of rnorm values
  set.seed( seed )
  return( matrix( rnorm( n * p, 0, 1 ), ncol=p ) )
}

"dgen.Y" <-
## input: takes matrix output from dgen.X
## output: vector Y = 1 + x_2 + 2x_3 + ... + (p-1)x_p
##
function( X, seed=364 ){
  set.seed( seed )
  n <- dim( X )[1]
  p <- dim( X )[2]
  ## multiply columns of X. New cols will be v_i = (i-1)x_i
  ## --also add 1
  X <- ( X %*% 0:(p-1) ) + 1
  ## determine scale factor
  scale.factor = quantile(X, 0, .75)
  return( X/scale.factor )
}

"dgen.lm" <-
## inputs: n = number of records, p = number of regressors
```



```

## outputs: returns data frame X with names(X) = Y, X1, X2, ..., Xp
## this output should be suitable for doing linear regression
## lm( Y ~ ., data=X)
function( n=30, p=2 ){
  ## make the predictors
  X <- dgen.X( n, p )
  ## add the response and add some noise to it too
  X <- cbind( (dgen.Y(X) + rnorm(n,0,1) ), X )
  ## make ourselves a data frame with proper names
  X <- data.frame( X )
  names(X) <- c("Y", paste( "X", 1:p, sep="" ) )
  return( X )
}

"antilogit" <- function( x ){
  expx <- exp(x)
  ifelse(is.inf(expx),1,exp(x)/(1+exp(x)))
}

"dgen.glm" <-
## inputs: n = number of rows, p = number of predictors
## outputs: returns data frame X with names(X) = Y, X1, X2, ..., Xp
## output suitable for logistic regression. That is, Y is a
## binary factor.
function( n=20, p=2 ){
  ## make the predictors
  X <- dgen.X( n, p )
  ## make the response (antilogit() defined above)
  m <- antilogit( dgen.Y(X) )
  Y <- factor( rbinom(n, 1, m) )
  ## convert predictors to data frame and then add response
  X <- data.frame( X )
  X <- cbind( Y, X )
  names(X) <- c("Y", paste( "X", 1:p, sep="" ) )
  return( X )
}

```

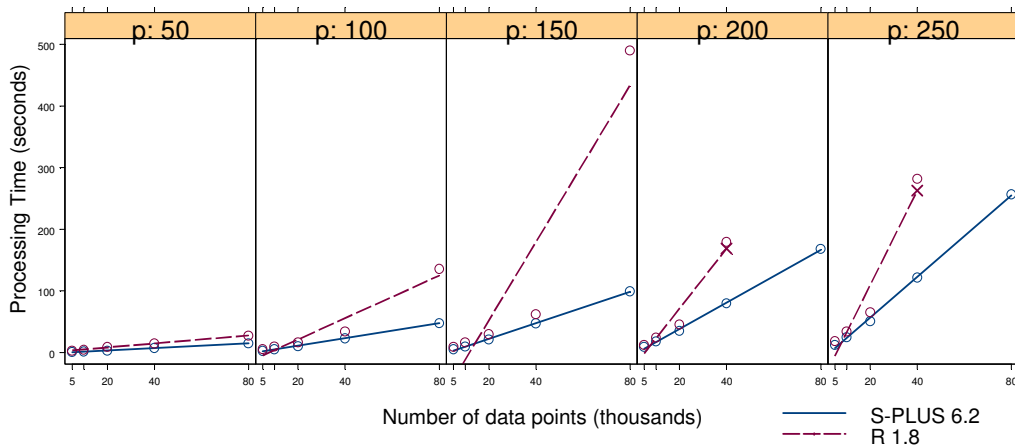
The total elapsed time was measured for each analysis using system tools and given in the tables and charts below. Note that N is reported in thousands of rows, and that the data set size (Mb) is calculated assuming 8 bytes per data point. In cases where the computation did not complete due to an out of memory error ("cannot allocate dynamic memory") we report **DNC**.

In all cases, S-PLUS finished the calculation faster than R, even with small data sets. In nearly all cases, R slows significantly once the number of rows exceed 40,000, while S-PLUS's performance remains linear. We note that R fails to complete the calculation as data sets grow in size.

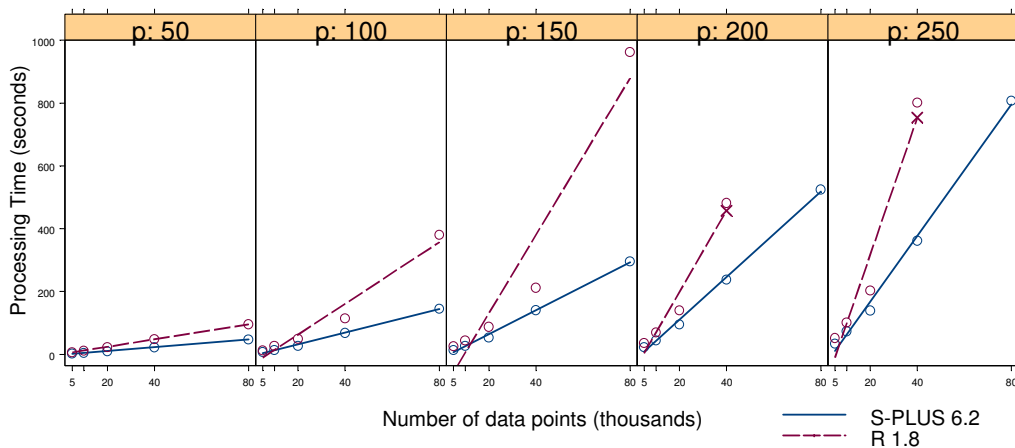
'000s			Linear Regression		Logistic Regression	
	N	p	Elapsed Time (seconds)	Elapsed Time (seconds)	R	S-PLUS
		Mb	R	S-PLUS	R	S-PLUS
5	50	1.91	2.93	0.86	6.86	2.55
10	50	3.81	4.26	1.69	11.82	4.84
20	50	7.63	8.98	3.00	22.95	9.49
40	50	15.26	14.55	7.27	48.09	22.38
80	50	30.52	27.30	14.96	95.63	47.41
5	100	3.81	5.72	2.42	12.70	6.95
10	100	7.63	9.50	5.17	27.10	13.55
20	100	15.26	16.22	10.13	48.93	26.98
40	100	30.52	34.06	22.99	114.44	68.03
80	100	61.04	135.59	47.67	380.36	145.35

			Linear Regression		Logistic Regression	
'000s			Elapsed Time (seconds)		Elapsed Time (seconds)	
N	p	Mb	R	S-PLUS	R	S-PLUS
5	150	5.72	9.25	5.08	25.63	13.82
10	150	11.44	16.59	9.42	44.56	26.62
20	150	22.89	29.66	21.12	87.40	53.51
40	150	45.78	62.07	47.23	212.16	140.77
80	150	91.55	489.94	99.35	961.70	295.67
5	200	7.63	12.40	9.15	36.07	22.73
10	200	15.26	24.33	17.83	69.14	44.48
20	200	30.52	45.43	34.78	139.79	94.96
40	200	61.04	179.47	79.77	482.16	237.90
80	200	122.07	DNC	168.19	DNC	524.65
5	250	9.54	18.57	12.22	51.49	34.16
10	250	19.07	34.11	24.54	101.40	73.31
20	250	38.15	65.20	50.47	203.63	139.06
40	250	76.29	281.87	121.53	801.11	361.39
80	250	152.59	DNC	256.76	DNC	807.93

Linear Regression



Logistic Regression



3.3 Training Trees

We used R and S-PLUS to train trees using both regression and binary classification algorithms. The commands used to train the trees were:

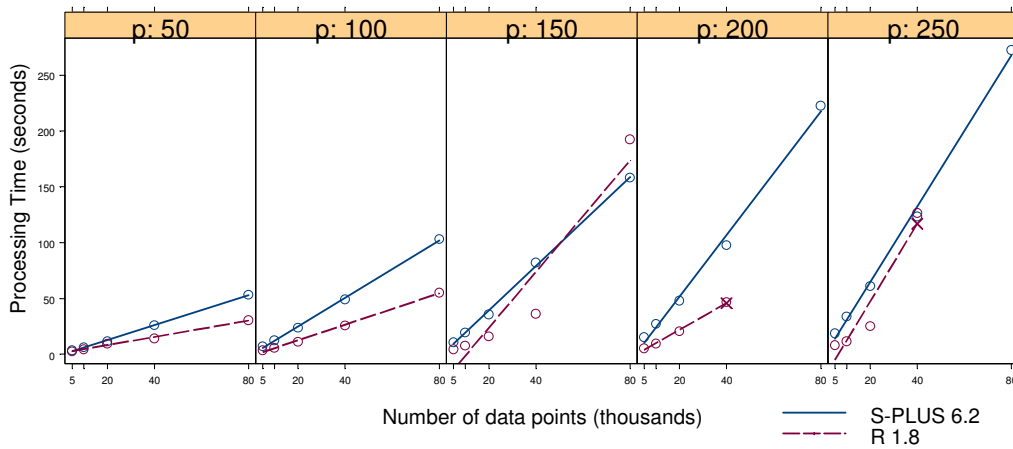
```
tree(Y ~ ., data = test.data ) in S-PLUS
rpart(Y ~ ., data = test.data ) in R
```

The data frame test.data was generated using the same methodology as for linear regression and logistic regression. The times for completion are given in the tables below; again, we report **DNC** when the calculation fails to complete due to out-of-memory errors.

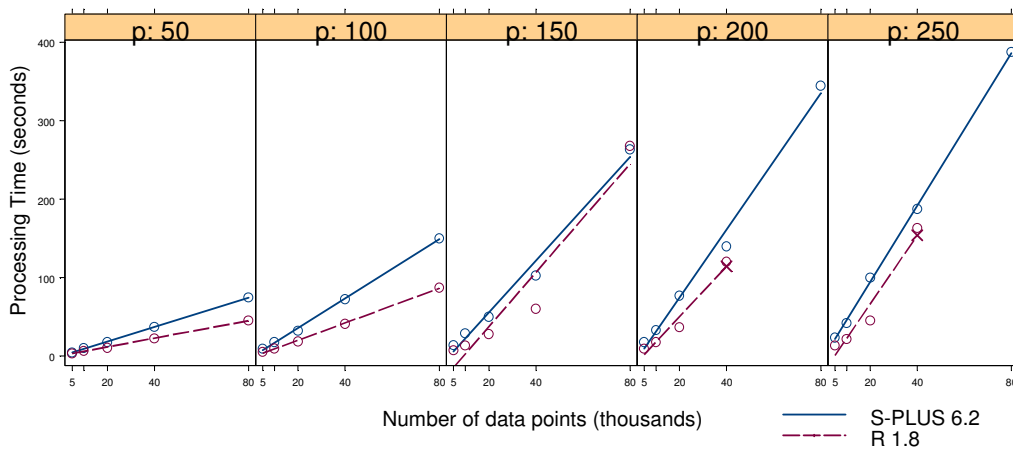
'000s	N	p	Mb	Regression Tree		Classification Tree	
				Elapsed Time (seconds)	Elapsed Time (seconds)	R	S-PLUS
5	50	1.91	2.39	3.27	3.33	4.48	
10	50	3.81	4.34	6.15	6.42	10.11	
20	50	7.63	9.16	11.46	10.46	17.73	
40	50	15.26	14.07	26.04	22.38	37.22	
80	50	30.52	30.45	53.31	45.35	74.54	
5	100	3.81	3.37	7.02	5.26	9.35	
10	100	7.63	5.52	12.35	9.59	17.80	
20	100	15.26	11.00	23.53	18.57	32.12	
40	100	30.52	25.56	48.95	41.13	72.47	
80	100	61.04	55.14	103.04	87.23	149.96	
5	150	5.72	4.18	10.55	7.24	13.85	
10	150	11.44	7.52	19.35	13.62	28.99	
20	150	22.89	15.79	35.50	27.92	49.90	
40	150	45.78	36.08	81.96	60.17	102.63	
80	150	91.55	192.41	158.12	267.64	263.19	
5	200	7.63	5.21	15.12	9.48	17.69	
10	200	15.26	9.46	27.14	17.83	32.85	
20	200	30.52	20.42	47.86	36.61	76.91	
40	200	61.04	46.59	97.64	120.00	139.82	
80	200	122.07	DNC	222.61	DNC	344.48	
5	250	9.54	8.00	18.60	13.46	23.42	
10	250	19.07	11.49	33.77	21.68	42.05	
20	250	38.15	24.98	60.96	45.32	99.83	
40	250	76.29	126.54	123.40	163.07	187.29	
80	250	152.59	DNC	272.44	DNC	387.25	

Overall behavior of both products were similar in each case. For the smallest data sets, R completes the computation slightly faster than S-PLUS, but this disparity decreases as the dimension of the problem (p) gets larger. For the largest data sets R again fails to finish its analysis, with out-of-memory errors causing no result to be computed.

Regression Tree



Classification Tree



3.4 Variance/Covariance Calculations

The final test compares S-PLUS and R in the calculation of a variance-covariance matrix from large data sets. The variance-covariance matrix was calculated using the function call

```
var( test.data )
```

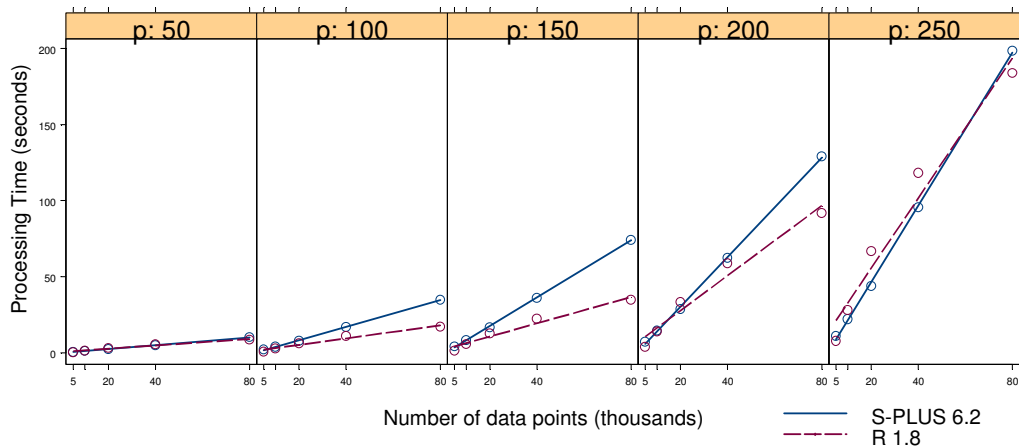
in both S-PLUS and R, with the `test.data` data frame calculated using the `dgen.X` function of previous tests. The elapsed computation times are shown in the table below.

			Covariance	
000s			Elapsed Time (seconds)	
N	p	Mb	R	S-PLUS
5	50	1.91	0.34	0.55
10	50	3.81	1.37	1.22
20	50	7.63	3.07	2.40
40	50	15.26	5.43	4.95
80	50	30.52	8.63	10.17
5	100	3.81	0.70	2.08

Covariance					
000s			Elapsed Time (seconds)		
N	p	Mb	R	S-PLUS	
10	100	7.63	2.82	4.00	
20	100	15.26	6.22	7.88	
40	100	30.52	10.99	16.96	
80	100	61.04	17.07	34.79	
5	150	5.72	1.42	4.18	
10	150	11.44	5.63	8.33	
20	150	22.89	12.68	16.66	
40	150	45.78	22.39	36.02	
80	150	91.55	34.75	74.22	
5	200	7.63	3.86	7.21	
10	200	15.26	13.99	14.48	
20	200	30.52	33.32	28.67	
40	200	61.04	58.83	62.34	
80	200	122.07	91.88	129.01	
5	250	9.54	7.76	11.06	
10	250	19.07	28.06	22.06	
20	250	38.15	66.79	43.82	
40	250	76.29	118.17	95.73	
80	250	152.59	183.88	198.38	

Although R computes the result faster than S-PLUS for many of these tests, we note that the disparity diminishes as the dimension of the problem gets larger. This difference in performance is attributable to the fact that S-PLUS's `var` function, by using a hierarchical multi-pass algorithm, is optimized for accuracy rather than speed.

Covariance



3.5 Insightful Miner scales S-PLUS functions to very large data sets

Insightful Miner is a highly-scalable data analysis workbench from Insightful Corporation. Although Insightful Miner and S-PLUS are separate products, Insightful Miner is designed to work with an existing S-PLUS installation. The main benefit for S-PLUS users wishing to analyze large amount of data is that Insightful Miner can process extremely large data sets.

Insightful Miner's pipeline engine allows it to process extremely large amounts of data without requiring large amounts of RAM. You can therefore use Insightful Miner's capabilities to scale S-PLUS analyses to datasets hundreds of megabytes to gigabytes in size. You can learn more about how Insightful Miner

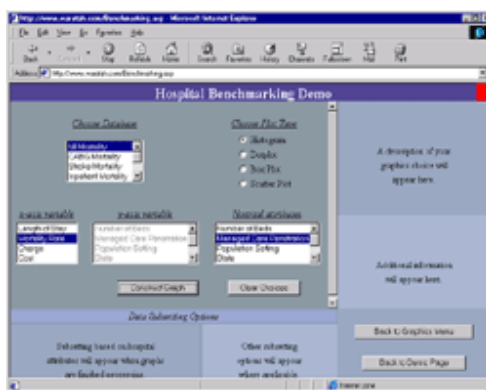
scales S-PLUS in the archived webcast *Scaling S-PLUS in a Production Environment* at www.insightful.com/news_events/webcasts/gen04/scaling.asp and in the white paper Using Insightful Miner to Apply S-PLUS to Large Data Sets available at www.insightful.com/support/whitpapers/Scaling_S-PLUS.pdf.

4 Production S-PLUS applications can be easily deployed within your existing infrastructure

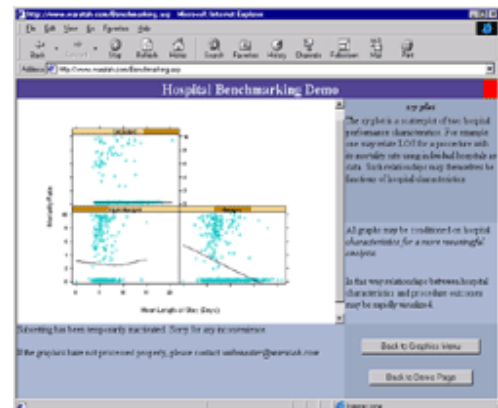
Instead of tasking statisticians to manually produce each analysis in response to individual requests, the S-PLUS product family can deploy analytical tools to decision makers who can then run shared analytical models, view custom reports, and generate graphics from a familiar environment, such as a Web browser, Microsoft Excel or a dedicated client application. By enabling users to run their own analyses productivity and consistency improves and best practices are shared.

4.1 S-PLUS[®] Server for Web-based or client/server deployment

The S-PLUS servers use Web and client/server technologies to deploy analytical expertise throughout an organization. All analysis steps associated with data processing, analytical models and visualization can be prototyped using the S-PLUS desktop and then stored in a central server for access by a wide variety of users who can apply these data analysis techniques from a familiar environment.



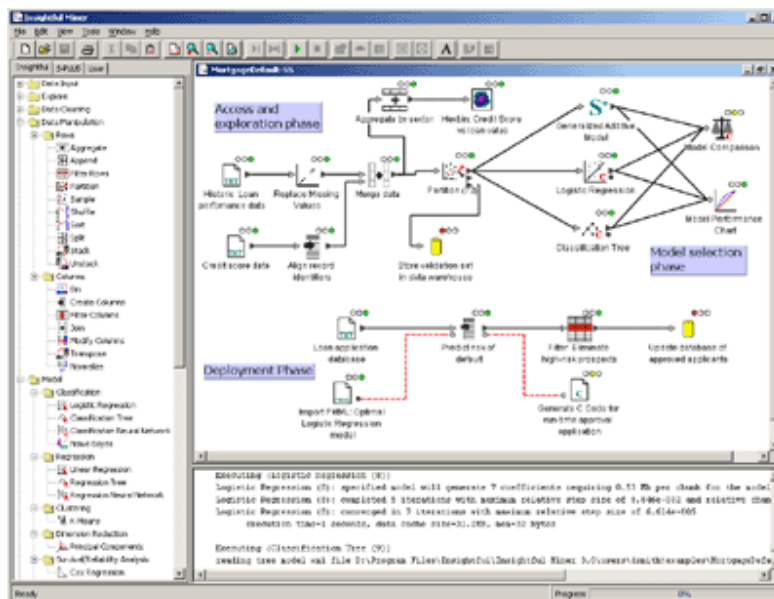
Dynamic HTML-based interfaces can be created to allow users to easily choose the data sets, parameters, and types of analyses to run from a browser.



The specified analysis is run and results displayed immediately. Output data can be viewed as HTML or downloaded in a variety of formats, including Excel.

4.2 Insightful Miner deploys applications to non-technical users

Insightful Miner is a highly scalable data analysis workbench that gives analysts the ability to deploy predictive intelligence throughout the enterprise using an intuitive, visual workmap interface. Insightful Miner can operate in single-user desktop and multi-user Windows or Solaris server environments. Insightful Miner worksheets are platform-independent, so you can easily develop on Windows and deploy on Solaris.



Insightful Miner’s scalable architecture uses an intuitive visual workmap interface to speed and simplify the process of building and deploying sophisticated data analysis applications

Insightful Miner's unique architecture handles the details of splitting data sets into manageable chunks that fit into the computer’s memory, and passes them sequentially to analysis components to form an analytical “pipeline.” This means that predictive models created in S-PLUS can be applied to gigabyte-class data sets. The only limitation is the amount of space available on disk to store the results.

Insightful Miner also helps you manage your S-PLUS applications. By embedding S-PLUS scripts into Insightful Miner's self-documenting visual workflows, you can see exactly how you built your application. This precise visual history makes it easy to replicate all or parts of your analysis and share this work with less-technical decision makers throughout the organization. By deploying best practices, more time is now spent analyzing results, not obtaining them, improving your team's productivity.

You can learn more about Insightful Miner at the multimedia demonstration found at www.insightful.com/products/iminer/iMiner_registration.asp.

4.3 S-PLUS has tools for producing reports in HTML, PDF or RTF formats

Insightful’s proprietary S-PLUS libraries make it simple to construct reports containing tables and rich graphics. Using Microsoft Word, users can edit narrative text easily in surrounding tables and graphs while adhering to a standard, user-defined reporting layout. Custom and automated reports can be created using the XML Style Language (XSL) and converted to standardized, validated reporting templates that support PDF, RTF, HTML, PS, Microsoft Word, PowerPoint and other formats.

For an example of how S-PLUS is used in clinical reporting applications within the pharmaceutical industry, see the archived webcast *Enhanced Clinical Data Analysis & Reporting Using S-PLUS 6.2* available at www.insightful.com/news_events/webcasts/pharm03/moc.asp.

4.4 S-PLUS is the best choice of embedded analytic engine for third-party applications

Because R is licensed under the open-source GPL (GNU Public License), any application which embeds R must in turn be licensed under the GPL. If you plan to embed R in a packaged application, you risk exposing your intellectual property to the GPL, which requires you to make the source code for

your application freely available. Getting your IT department involved in this discussion can help avoid the risk of exposing your IP to free software requirements.

Similarly, due to the restrictions of the GPL, you can not use R to embed analytics in a commercial application without also licensing the application under the GPL and making the source code freely available. This makes R a problematic choice as an analytic engine for a commercial application. Likewise, SAS has a history of licensing policies that can make it prohibitively expensive as an analytic engine for third-party applications. In contrast, Insightful offers flexible licensing for every situation. Our licensing policies deliver affordable, simple solutions for global, distributed, networked and OEM situations.

4.5 You must install, compile and maintain R yourself

S-PLUS works out-of-the-box for all its supported platforms with a modern, easy-to-use installer program. Unlike R, there's no need to compile from source code yourself, or rely on inconsistently-updated binary files downloaded from the web.

5 S-PLUS fully complies with industry mandates. R does not.

Regulations like those that govern the biopharmaceutical industry have a significant impact on the IT infrastructure of companies competing in these markets. The S-PLUS product family has been proven to fully comply with these regulatory frameworks including those provided by 21 CFR Part 11 and the FDA's Guidance for Industry documents, among others. R has no such proof of compliance, and it is questionable if this compliance is possible. Only S-PLUS 6.2 provides these capabilities while offering an enhanced operating environment and specific validation features for managing clinical data, programs, logs, documents and reports.

5.1 S-PLUS is validated software to ensure accuracy, reliability and consistency

Insightful has been audited by outside agencies and proven compliant with the need for validated systems that "ensure accuracy, reliability and consistent intended performance." As part of the audit processes, Insightful provides full standard operating procedures (SOP's) for software development and Quality Assurance, complete training records including CVs for our development and QA staff, documentation of product plans and specifications, documentation of code reviews during development, QA test plans and procedures, proof of physical security for build machines and developer machines, and proof of 21 CFR Part 11 training for QA and development employees.

While these efforts are typically not in the charters of freeware communities, Insightful takes its role in serving regulated industries seriously. Rules that govern technologies continue to evolve and new features specifically addressing these requirements will be a part of releases of the S-PLUS product family. For example, the extensive batch processing and verbose logging capabilities in S-PLUS provide an ideal platform for validated statistical data analysis in production and regulated environments.

Insightful carefully balances regulatory needs with usability and performance requirements so that our solutions thoroughly satisfy our diverse user communities. For more information on Insightful's validation procedures and features, see our validation data sheets and white papers on our web site at www.insightful.com.

5.2 Insightful services can assist your compliance efforts

Insightful's S-PLUS Validation Solution will ensure the structural integrity of your software installation and provide detailed test scripts, summary reports and documentation required to comply with submission regulations and to meet audit demands. The process includes documentation and testing for installation, operation, security, functionality and performance of S-PLUS within your environment.

5.3 S-PLUS is professionally engineered and supported

S-PLUS includes professional installation programs and access to Insightful's software maintenance and technical support services. Our engineers are available by email or phone where 80% of questions

are answered within 24 hours. When compared to using freeware packages, significant cost savings can result from eliminating the time spent on managing your software and its operating environment. Remember, R is really only free only to those whose time is free.

5.4 Newest statistical methods are made enterprise-ready

Owing to its own research and its close association with many of the world's thought leaders in data analysis, S-PLUS offers a wealth of important cutting-edge statistical methods unavailable anywhere else, including Robust Analysis, Missing Data Analysis, Wavelets, Financial Data Analysis, Environmental Statistics, Multiple Comparisons, Quality Control Charts, and Scaleable Optimization.

R is a hotbed for advanced research in the statistical sciences and R libraries are being written to tackle a variety of emerging data analysis problems. Insightful tracks these new methods and incorporates them into S-PLUS once they become proven. This makes R a valuable testing ground for the latest methods and analysis techniques, and because R and S-PLUS are syntactically very close, porting new methods from R to S-PLUS is straightforward.

For example, in 2003 Insightful ported the R-based *Bioconductor* library and incorporated these capabilities into its S+ArrayAnalyzer desktop and server products for microarray analysis in the life sciences. Analysts benefit from using the most modern and accurate techniques combined with the ease of use, tailored GUI professional installation, support and training offered by Insightful's dedicated staff of engineers and services staff.

6 Summary

The choice of a data analysis platform has a dramatic impact on an organization's competitiveness. While numerous packages promise basic algorithms and graphics, a strong ROI is only possible when the tasks of designing, developing, deploying and maintaining analytic applications are integrated. Besides offering a well thought-out roadmap to users of its products, the technology supplier must also supply a validated IT infrastructure and application-specific training, support and consulting to deliver the highest efficiency and competitiveness to its users.

Only the S-PLUS product family combines modern data analysis programming and visualization techniques for readily creating new analytics with the ability to deploy these applications into large-scale, validated production settings. The fact that S-PLUS and R are rooted in a common object-oriented language designed specifically for data analysis delivers great advantages. Both lend themselves to the creation of new analytic functions and applications tailored to an individual task. The research-focused R community is a fount of innovation, with the best of these advancements often being included in the stable and validated S-PLUS platform years before appearing in competitive products such as SAS. This balance of innovation and stability helps give organizations the competitive edge needed to succeed in today's marketplace.

S-PLUS is unique in that applications built with the desktop version can be efficiently deployed into highly scalable production environments using S+ Servers or Insightful Miner. We also saw that the sophisticated memory management features in S-PLUS allowed it to outperform R across many common analysis tasks, and S-PLUS is the only choice for analyzing large to very large data sets.

R must be assembled, installed and maintained by the user, and its GPL licensing means you should not embed R into your internal applications without thoroughly understanding the risks to your company's intellectual property. R is not validated for use in regulatory environments, while S-PLUS can prove its compliance with numerous regulatory standards and has services expertise to assist you prepare for and pass your industry audits.

S-PLUS has exclusive add-ons for solving industry-specific problems in a wide range of fields from bioinformatics and econometrics to reliability testing and environmental statistics. Its dedicated libraries for structured reporting and managing batch applications make it the first-choice for production settings where automation and efficiency are prized.

7 About Insightful

Insightful Corporation (NASDAQ: IFUL) provides enterprises with scalable data analysis solutions that drive better decisions faster by revealing patterns, trends and relationships. The company is a leading supplier of software and services for statistical data mining, business analytics and information retrieval, enabling clients to achieve information superiority.

Insightful products include the award-winning S-PLUS[®], S-PLUS[®] Server, Insightful Miner, and InFact[™]. These products add sophisticated data analysis into data warehousing, business intelligence and content management initiatives. Insightful consulting services provide proven expertise and processes for the design, development and deployment of customized analytical solutions.

Insightful has been at the technical forefront of analytics for more than sixteen years. The company's clients include over 6,000 companies in financial services, pharmaceuticals, biotechnology, telecommunications, manufacturing, plus government and research institutions, who need cutting-edge systems that can quickly and accurately identify risks and reveal opportunities.

The reality is, effective data analysis makes your entire IT investment pay off. If you use lesser tools, you'll get less information. Making a decision with less information, or bad information can have huge consequences. Getting locked into someone else's "black box" analytical application is equally problematic if you can't tune it to keep pace with new requirements. We show you things no one else can show you. We'll tell you things no one else can tell you. We develop solutions no one else can build for you.

Insightful Corporation gives you the information superiority that leads to growth and profitability. By solving data analysis problems, we help you identify opportunities. We help improve your judgment. We increase your confidence. We accelerate positive action, sparked by better analysis. You end up with a smarter, more competitive organization.

For more information contact Insightful at an office near you, visit www.insightful.com or email info@insightful.com.

Global Headquarters

1700 Westlake Avenue North
Suite 500
Seattle WA 98109
USA
Tel: 206.283.8802 • 800.569.0123
Fax: 206.283.6310
Email: info@insightful.com

Insightful UK

5th Floor, Network House
Basing View
Basingstoke, Hampshire RG21 4HG
United Kingdom
Tel: +44 (0) 1256 339800
Fax: +44 (0) 1256 339839
Email: info.uk@insightful.com

Insightful Switzerland

Christoph Merian Ring 11
Ch - 4153 Reinach
Switzerland
Tel: +41 61 717 9340
Fax: +41 61 717 9341
Email: info.ch@insightful.com

Insightful France

7, rue Auber
31000 Toulouse
France
Tel: +33 (0) 1 55 69 37 33
Fax: +33 (0) 1 56 69 37 34
Email: info.fr@insightful.com

S-PLUS is a registered trademark of Insightful Corporation. Insightful, the Insightful logo, "intelligence from data" and InFact are trademarks of Insightful Corporation. Other product names mentioned herein are trademarks or registered trademarks of their respective companies. Printed in the United States of America. This data sheet is for informational purposes only. Insightful makes no warranties, expressed or implied, in this summary. © 2004 Insightful Corporation. All rights reserved.