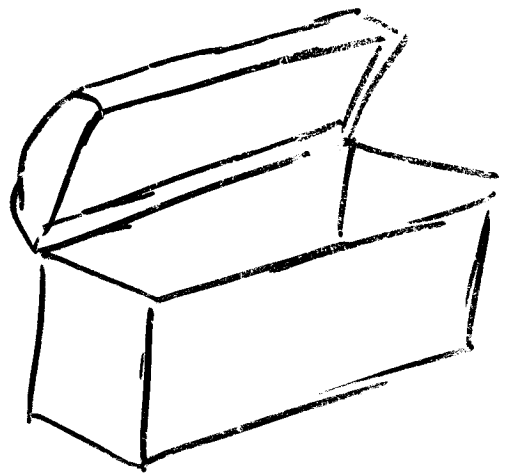


ARTIFACTS
from the

Early History
of

S



Rick Becker

AT&T Labs - Research

Sept 1974



Chambers J M 1214
John



Becker R A 1215
Richard

Joined Statistics Research
at Bell Labs
Murray Hill



Prim R C

Bob

12



Tukey J W

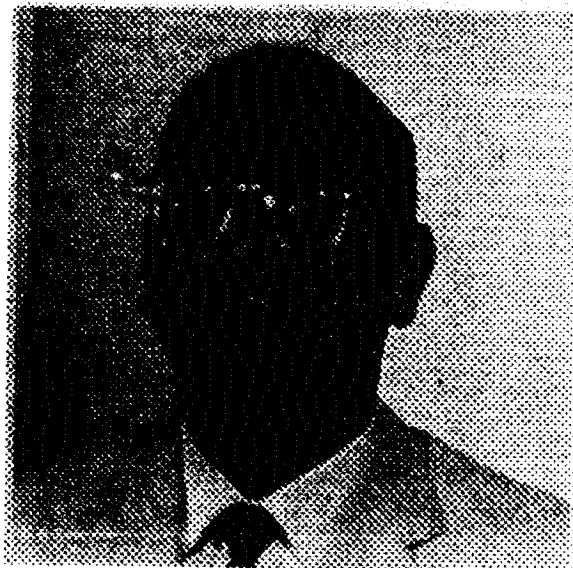
John

12



Pollak H O

121



Gnanadesikan R
Ram

1214



Mallows C L

Colin

1215



Freeny P E 1215
Anne



Landwehr J M 1214
Jim



Denby L 1215
Lorraine



Gross A M 1215
Alan



Kettenring J R **1214**
Jon



Cleveland W S **1215**
William

Computing Environment

Statistical Computing Subroutine library
portable FORTRAN

Put

To analyze data, write a
FORTRAN program to:

- read data
- compute with SCS subroutines
- write results

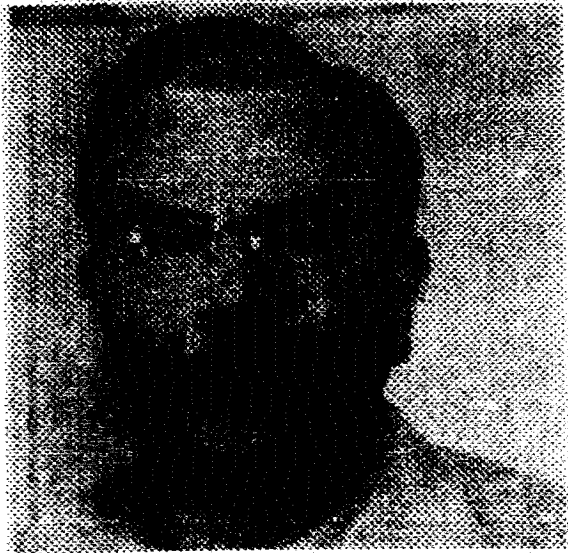
The read/write was often the
hardest part

Device - Independent Graphics GR-Z

On Honeywell 645 Computer
running ~~G~~ECOS

- printer
- storage scope (Tektronix)
- pen plotter
- microfilm

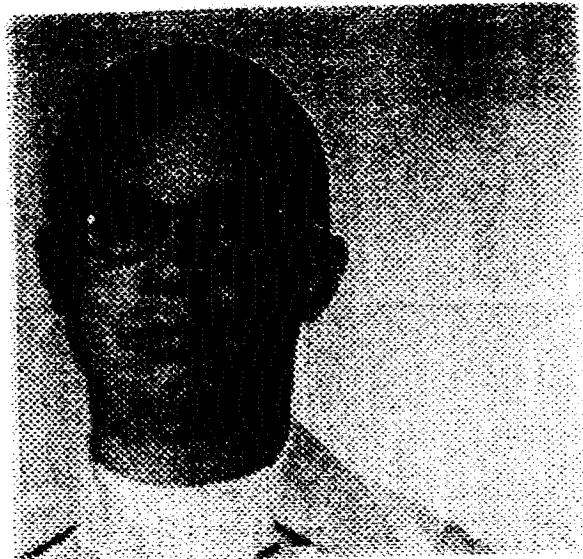
Vector of control parameters



Mc Gill R 1214
 Bob

SCS
Library

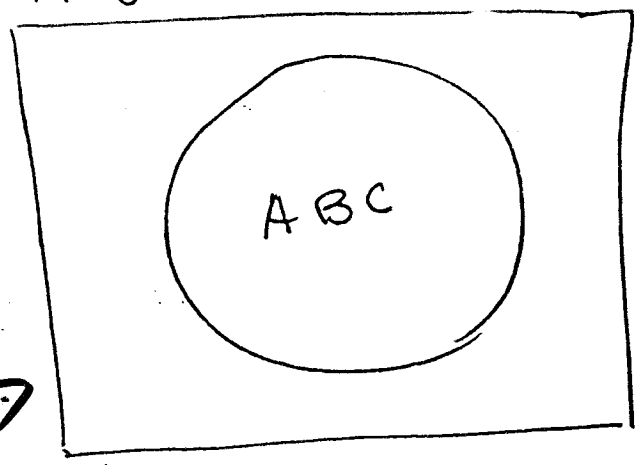
Scatter
Storage



Warner J L 1214
 Jack

Implementation

Big Idea

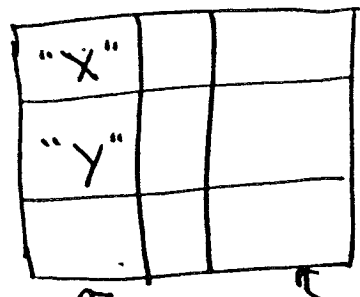


ABC: general
(FORTRAN)
algorithm

XABC: FORTRAN
subroutine to
provide interface
between ABC &
Language and/or
utility programs

XABC (INSTR, OUTSTR)

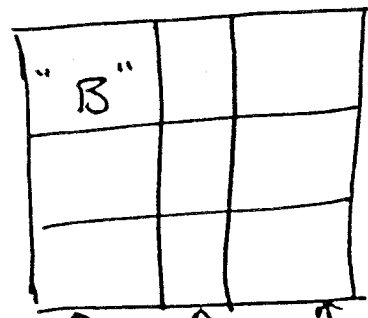
Input INSTR →
Note input &
output look
the same



Argument Names or
Blank

Pointers/Values

OUTSTR →



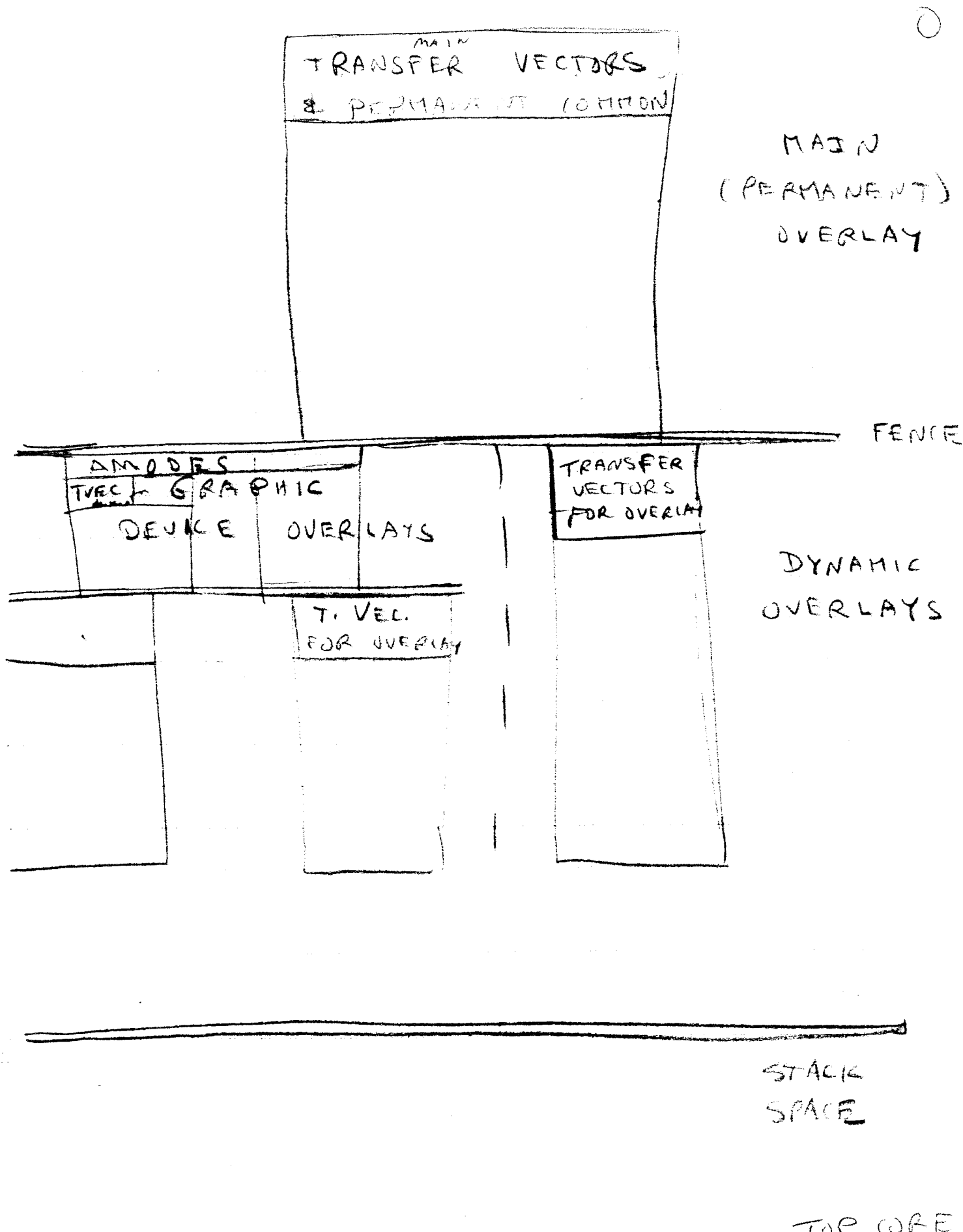
Result Names

Types (Modes)

Pointers/Values

Note: Names are
meaningful to Algorithm,
not necessarily to
Language

How to do it on GCOS



Quality!

Number: 1

"S" Control Sheet

Type of Problem

- Documentation
- Error
- Suggestion

Name: RAB

Date: 10/28/76

Summary of problem:
(attach printouts, plots, etc.)

Prompt should be different when parser is expecting a continuation of previous line.
(eg instead of '=' use '>')

Action:	Chapter	Routine	Fixed
Modify RBVF - the read routine for the lexical analyzer Add QCHACT to MTV.D	Lang	RBVF (lexer)	-
By: <u>RAB</u>			

Chapter SYSTEM _____

Command YYPARS _____

Test Date 11/9/76 _____

***** Installation: By JMS *****

Date 11/9/76 _____

Confirmed RAB _____

AN INTERACTIVE LANGUAGE FOR DATA ANALYSIS

A new interactive, general, extensible computing language is being developed for data analysis. It is designed to make quick, exploratory analysis of data possible without the details needed in conventional languages (FORTRAN). At the same time it is sufficiently powerful and flexible to serve as a vehicle for research in data analysis and for the development of specialized sub-languages.

Compared to existing statistical languages, its novel features are: general hierarchical, self-defining data structures, with common structures provided automatically (e.g., time-series); syntactic power and flexibility; modularity which allows development of user profiles and sub-languages; emphasis on portability in the design; and an interface language which allows inclusion of new commands in the system starting from FORTRAN algorithms (not written for the language) in a simple way, without detailed knowledge of the language structure.

R. A. Becker	MH 2C-320	x5512
J. M. Chambers	MH 2C-326	x2681
D. M. Dunn	MH 2C-479	x3320
J. E. McRae	MH 2C-323	x6180
J. M. Schilling	MH 2C-322	x3587

Getting the
word
out



Dunn D M 1214
Doug



Mc Rae J E 1214
Jean



Schilling J M 1214
Judy

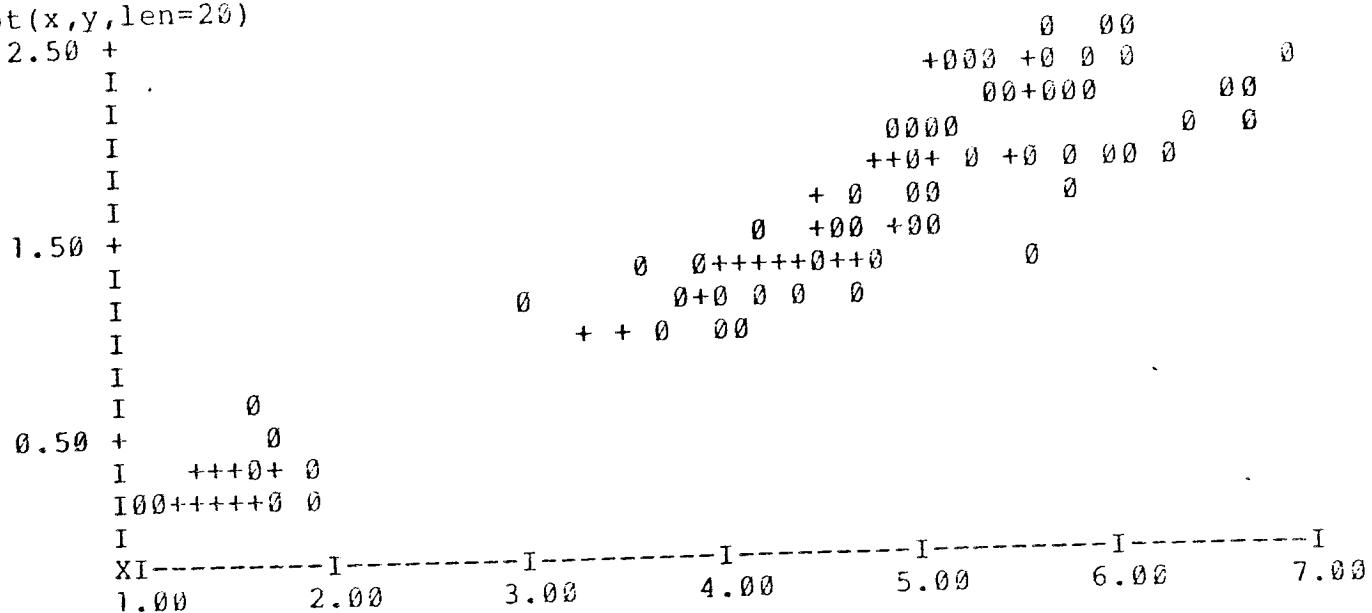
SYSTEM ?s/run -e

Run (in express mode)

```
'S' VERSION 1.0 - 12/1/76
> x <- read("jmc/iris", "(8x,f4.1)")
NUMBER OF ITEMS: 150
> y <- read("jmc/iris", "(12x,f4.1)")
NUMBER OF ITEMS: 150
> listw
166 Y
166 X
WORDS AVAIL: 8564, 96.3 %
> prplot(x,y,len=20)
```

} read from file
(optional format; default is free format)

} list contents of working storage



```
> x2 <- x[ x<2.5 -> t]; y2 <- y[t]
> (1:len(x))[t]
1: 1 2 3 4 5 6 7 8 9 10
11: 11 12 13 14 15 16 17 18 19 20
21: 21 22 23 24 25 26 27 28 29 30
31: 31 32 33 34 35 36 37 38 39 40
41: 41 42 43 44 45 46 47 48 49 50
> ratio <- y/x; circ <- 6.28*sqrt(.5*( x^2 + y^2 ))
+ )
> stem ratio
```

x and y for x < 2.5

} corresponding subscripts

← continuation

N = 150 MEDIAN = 3.030E-01 HINGES = 2.143E-01, 3.571E-01

DECIMAL POINT IS 1 PLACE(S) TO THE LEFT OF THE COLON

```
5 5 0 : 77779
27 22 1 : 122222233333334444444
34 7 1 : 5555778
44 10 2 : 00111113344
65 21 2 : 555677777888999999999
44 3 : 000000000000111111111222222222233333333344
41 28 3 : 566666677777777778888889999
13 11 4 : 00112333344
2 2 4 : 57
```

} stem and leaf plot

← write file called "circ"

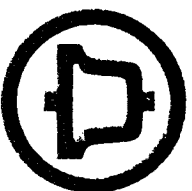
```
> write(circ)
> q
SYSTEM ?list circ
```

```
0.628000000E 01 0.628000000E 01 0.58407376E 01 0.67198934E 01 0.628000000E
0.628000000E 01 0.66757315E 01 0.67198934E 01 0.71603016E
```

SYSTEM ?

Sample 'S' Session 1/13/77

The
UNIX
Era



Bell Laboratories

ect: Use of the Interdata for
Statistical Computing Research

date: December 11, 1978
from: Samuel P. Morgan

2/12
Mr. H. O. Pollak:

In response to your memorandum of November 27, I see no reason why Rick Becker and John Chambers should not use the Interdata 8/32 to put up a UNIX version of the "S" system for interactive data analysis. I understand that the Interdata is now mostly used by a small number of numerical analysts, the original UNIX portability experiment by Ritchie and Johnson having been completed some months ago. A second 67-megabyte disk is being installed to better accommodate the current users. Becker and Chambers are welcome to use the system too, at least up to the level estimated in your memorandum (15% of processor load, 3 megabytes of disk storage).



Ritchie D M 1273
Dennis
C



Thompson K 1271
Ken

Unix



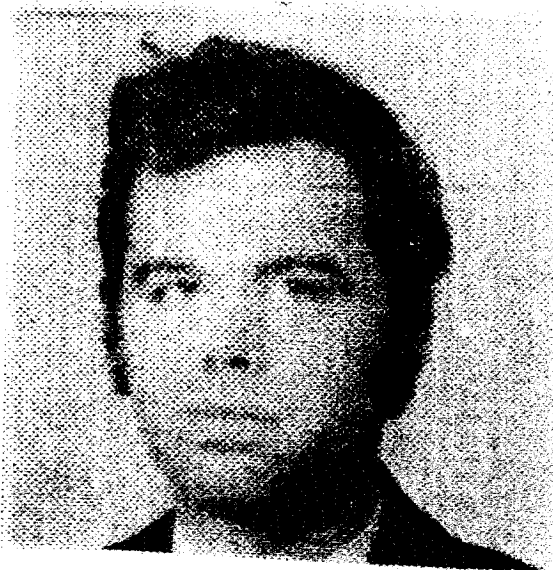
Kernighan B W 1273
Brian

M4
Rat for
C
AVIK



Lesk M E 1274
Michael

Yacc, lex



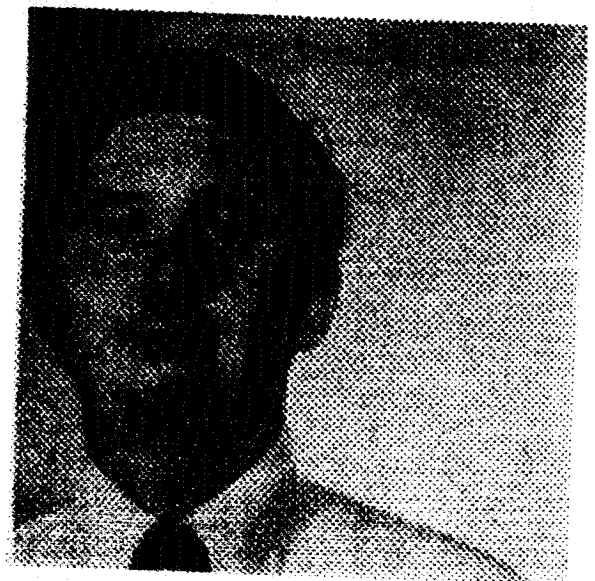
Johnson S C 1273
 Steve
 Yacc, Lex



WEINBERGER, P.J. 1127
 Peter
 F77
 AWK



Baker B S 1271
 Brenda
 Struet



Feldman S I 1274
 F77

BELL LABS NEWS

FOR THE PEOPLE OF BELL LABORATORIES



Vol. 20 No. 38

September 29, 1980

In
this
issue:

United Way campaigns conducted
at all company
locations. **page 2**



Summer Science Program was a
unique experience
for minority
participants. **page 3**

Statistical Software Sleuth Tackles Variety of Cases

Factors contributing to the characteristics of computer programs that affect the amount of necessary debugging and the feasibility of testing by telephone for hearing defects are among the various studies conducted at Bell Labs with the aid of a new software system—simply called *S*.

S was developed at Bell Labs specifically for interactive statistical data analysis, graphics and related scientific computing. Because *S* is interactive, users can see results immediately in order to determine what to do next. They can steer the analysis themselves, without extensive programming experience.

S is being used in a number of Bell System applications. Fran O'Neil, of the Digital Systems Applications Department at Merrimack Valley, used it to develop a model for evaluating the economics of alternative transmission facility and network plans. Tom Hamilton, of AT&T Long Lines' Accounts and Finance Organization at Bedminster, linked *S* with a four-color plotter to enhance the graphical representation of budgetary planning analyses. Last year Brian Godfrey, of the Network Performance Char-



AT&T CASES

Good Ideas

- interactive
- graphics
- general language
- objects (incl. functions, ...) persist
- no built-in limits
- let OS do the work
- port the OS not the language

Aim high

Big Idea

New Blood!



WILKS, A.

11214

Allan

Want more?

See A Brief History of S

at <http://www.research.att.com/~rab>

```
SYSTEM ?S/FUN
9/22/77 NEW:
RDPEN, IDENTIFY, CONTOUR, GS, DISK, BACKSOLVE, DENSITY
'S' VERSION 1.1 - 9/3/77
```

```
> listw
133 BSTART
59 MKTBOOK
53 MKTSP
59 UTILEARN
WORDS USED: 304
```

} AVAILABLE DATA

```
> tek46n
> plot(utilearn,mktbook)
GO?
```

} IDENTIFY GRAPHICS TERMINAL;
SCATTER PLOT

```
> lsr <- reg(utilearn,mktbook)
> abline(lsr$coef)
> text(rdpen(!), label="least;squares")
>
```

} REGRESSION (LEAST SQUARES);
PLOT LINE, IDENTIFY IT.

```
> rr <- rbiwt(utilearn,mktbook)
> abline(rr$coef)
> text(rdpen(!), label="robut@st")
>
```

} ROBUST REGRESSION

```
> stem(lsr$resid)
```

```
N = 45 MEDIAN =-8.101E-03 QINGES =-8.859E-01, 8.030E-01
```

```
DECIMAL POINT IS AT COLON
```

```
1 1 -3 : 5
2 1 -2 : 2
9 7 -1 : 9866211
14 -0 : 99988875554210
22 12 0 : 061122247888
10 6 1 : 035789
4 3 2 : 019
1 1 3 : 3
```

STEM-AND-LEAF DISPLAY
OF LEAST-SQUARES
RESIDUALS

```
> g
SYSTEM ?
```

EXAMPLES OF THE 'S'
SYSTEM FOR DATA ANALYSIS