I. INTRODUCTION AND OVERVIEW

Output devices, such as CRT display terminals and special plotters, which are now available for high-quality graphical displays, provide the analyst new opportunities for presenting and exploring data. In addition to summarizing data and performing statistical evaluations, the computer can be used to present the results through attractive and meaningful graphical displays. These displays are further enhanced by the increasing accessibility of color graphics. As a result, individuals with limited statistical training can obtain simple, but concise, graphical presentations of the data, while a trained analyst can gain an improved understanding of complex data by using graphical tools for exploratory data analysis. Moreover, obtaining such displays on a computer, rather than manually, can significantly reduce the elapsed time to get the desired graphics and allows the analyst to operate in an iterative manner.

The marriage of statistical data analysis and high-quality graphics has become possible through the development of statistical graphics software by software houses, graphics hardware suppliers, universities and others. Two previous papers (Caporal and Hahn [1981a and 1981b]) described software for high-quality graphical displays. These papers dealt principally with offerings for presentation graphics, many of which have only limited capabilities for statistical data analysis. In this paper, we concentrate on software that permits the user to integrate the data analysis and data display functions, i.e., offerings that have non-trivial statistical, as well as data display, capabilities, for a wide variety of computers, ranging from desktops to mainframes. We will, however, deal only with software that provides high-quality graphical output. Thus, we will not, for example, consider programs that only produce plots on standard line printers, although these may often be sufficient in the exploratory stages of a data analysis. Consideration will also be limited to software for general statistical analysis; offerings that perform only a small number of analyses, e.g., regression analysis, analysis of variance, or face plots, will not be considered.


A total of 15 different offerings are reviewed under the following headings:

- Statistical graphics software for desktop computers (Section II).
- Interfaces with, or extensions of, existing statistical packages (Section III).
- Integrated statistical graphics software for minicomputers and mainframes (Section IV).

Tables 1 and 2 give a comprehensive overview of the 15 offerings. Table 1 indicates the primary user of each, indicates whether the offering is a program or a library of subroutines and provides some brief descriptive comments. Table 2 shows the acquisition costs (ranging from $150 to $40,000 or more), the software source and the associated host computers and graphic output devices. Since these tables are self-explanatory, the information contained in them will not be repeated.
The major purpose is to report on the various offerings—not to evaluate them. We have tried to be objective in our comments and descriptions. However, some degree of judgment invariably enters into such a review, especially in deciding which programs have sufficient statistical and graphical content and which are general enough to warrant inclusion. We have aimed at describing all offerings that meet our criteria. Earlier versions of this paper were widely distributed. Suggestions, especially for additions, were solicited and acted upon. Also, in Section V, we indicate some offerings currently under development. Despite this, it is possible that some relevant applicable software, especially of a very recent nature, has escaped our attention. Due to lack of resources, we do not claim hands-on experience with much of the software. However, we have participated in demonstrations of most of the offerings.

II. STATISTICAL GRAPHICS SOFTWARE FOR DESKTOP COMPUTERS

A. Introduction

Boardman (1982) in an article on the future of statistical computing on desktop computers describes various desktop computers and includes a representative list of manufacturers. The use of these computers for data analysis is rapidly increasing. Many suppliers now offer and sell statistical graphics packages. These offerings have the following characteristics:

1. Their use is generally limited only to the desktop computer model for which they were developed (or to systems that emulate it).

2. They are generally directed mainly at data analysts with limited statistical backgrounds, rather than at the (smaller) specialized audience of statisticians. As a result, software packages for desktop computers tend, in general, to be more user-friendly, more flexible, and less statistically sophisticated than most of the offerings for larger computers described in the next two sections. However, the graphical features of the offerings for desktop computers are comparable to those that are discussed subsequently in this paper.

3. The available statistical and graphical features vary greatly.

The offerings with the most extensive graphics and statistical features are:

- The HP Series 70 Statistical Software
- The HP Series 80 Statistical Software
- The HP 9845 Statistics Library
- The HP Series 200 Statistics Library
- The Tektronix Plot 50 Applications Library
- The Trend Spotter '82 Graphics Management Information System.

These offerings are discussed further below. They represent only three manufacturers and only two of the 11 listed by Boardman [1982]. Offerings by other manufacturers have been omitted in this survey because they are of a specialized, rather than general, nature.

B. HP Series 70 Statistical Software

The HP Series 70 Statistical Software consists of libraries for:

- Basic one variable statistics
- Simple Linear Regression
- Analysis of Variance
- Correlation Coefficient
- Probability Distributions

The statistical software contains the capability of creating graphical output.

C. HP Series 80 Statistical Software

The HP Series 80 Statistical Library contains sublibraries for:
D. HP 9845 Statistics Library

The HP 9845 Statistics Library contains sublibraries for:
- General Statistics
- Basic Statistics and Data Manipulation
- Regression Analysis

Capabilities for plotting statistical results are contained within the sublibraries:

E. HP Series 200 Statistics Library

The HP Series 200 Statistics Library is similar to the HP 9845 Statistics Library.

F. Tektronix PLOT 50 Applications Library

The Tektronix PLOT 50 Applications Library consists of a statistical library as well as other software. The statistical library consists of sublibraries for:
- Tests and Distributions
- Analysis of Variance
- Multiple Linear Regression
- Nonlinear Estimation

Menus and prompts guide the user. The statistical sublibraries contain capabilities for plotting statistical results. A separate time series analysis and forecasting package is also available.

G. Trend-Spotter '82 Graphics Management Information System

Unlike the other packages for desktop computers described in this section, the Trend-Spotter '82 basic system includes both the actual hardware, i.e., a color CRT, as well as the applicable software. Additional options, such as equipment for hard copy output, can be ordered. As suggested by its name, Trend-Spotter is aimed principally at business people and administrators and is claimed to be 'a graphics system designed by managers, for use by managers.' For example, it permits estimation of trends and forecasts and includes adjustments for cost-of-living, anticipated inflation, and seasonal effects.

III. INTERFACES WITH, OR EXTENSIONS OF, EXISTING STATISTICAL PACKAGES

A. OVERVIEW

Sophisticated computer packages for statistical analysis have become prominent during the last decade and are used extensively by statisticians and other data analysts. Some well-known examples are BMDP (Biomedical Computer Programs P-Series), Minitab, SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences), and P-Stat. These offerings, and many others, are compared by Francis (1981). Most were developed prior to the general availability of high-quality computer graphics. As a result, the following extensions for, or interfaces with, high-quality graphics have been developed:

- Three offerings involving the SAS package:
  - The SAS/Graph extension to SAS
  - An interface between SAS and the Tell-A-Graf graphics program
  - Datagraf, an interface to SAS, SAS/Graph, and the Disspla graphics package.
The interface between SPSS and the SPSS Graphics Option, including an interface option with the Tell-A-Graf graphics program.

The interface between the StatII statistical analysis package and the PlotII program.


Each of these are reviewed in this Section. Plans for the development of graphical capabilities for BMDP, Minitab, and P-Stat are also indicated.

B. The SAS/Graph Extensions to SAS

The statistical analysis capabilities of SAS range from simple descriptive statistics to complex multivariate techniques and include regression, analysis of variance, categorical data analysis, multivariate techniques, discriminant analysis, and clustering.

The SAS/Graph extension to SAS allows users to obtain high-quality graphical displays of the output generated during a SAS run. SAS/Graph procedures are similar to SAS procedures and can be included in SAS runs. SAS/Graph also interfaces with the SAS/ETS (Economic Time Series) Library.

C. Interface Between SAS and Tell-A-Graf

An interface between SAS and the Tell-A-Graf graphics package is achieved through a SAS procedure called PROC TAG. This transforms SAS data files into Tell-A-Graf data sets. PROC TAG is available from ISSCD (see Table 2).

D. Datagraf

Datagraf, developed by M/A-COM Sigma Data, Inc., and available on the Boeing Computer Services national time-sharing service, provides interfaces to SAS, SAS/Graph, and the Disspla graphics subroutine package. The interface is achieved through a conversational language that prompts the user for information and does not require the user to learn either the SAS or the Disspla language. Thus, unlike the other packages described in this section, Datagraf appears as a new offering to the user. It also has various additional features beyond those of SAS and Disspla (e.g., a database consisting of demographic, socioeconomic and other data by county from the Bureau of the Census).

E. The SPSS Graphics Option and Tell-A-Graf Interface

SPSS is a well-known statistical package for analysis of social science data. Its capabilities include descriptive statistics, frequency distributions, cross tabulations, simple and partial correlation, n-way analysis of variance, multiple regression, discriminant analysis, factor analysis, and canonical correlations.

The SPSS Graphics Option permits the use of graphics commands within an SPSS run that are similar in form to the standard SPSS commands. In addition, a graphics permits SPSS users to interface with Tell-A-Graf by allowing the user to write files that can be accessed by Tell-A-Graf.

F. The StatII/PlotII Interface

StatII is a statistical package on the MarkIII time-sharing service of the General Electric Information Services Company (GEISCO). The capabilities of StatII include descriptive statistics, analysis of variance, regression, curve fitting, polynomial fitting, time series analysis, forecasting and data management. It interfaces with PlotII, a graphics package for the same system.

G. Interface of Sharp APL Statistical Function Library
and Superplot

The Sharp APL functions for Statistical Analysis is a library of sophisticated routines organized under the following headings:

- General Statistical Functions
- Model Parameter Estimation
- Probability Functions and Distributions
- Analysis of Variance
- Multivariate Analysis
- Time Series Analysis

The statistical library is directed at relatively knowledgeable analysts. It interfaces with Superplot, a Sharp APL graphics package, to provide graphical displays.

H. Plans for BMDP, Minitab and P-STAT

BMDP, Minitab and P-STAT do not currently have interfaces with graphics packages, but there are plans for developing such interfaces. In particular, we received the following responses to our inquiries from the suppliers of these packages:

BMDP:

"BMDP does not currently interface with high-resolution graphics devices. One of our current projects is implementation of BMDP on desktop microcomputer which features a graphics terminal. We expect to be able to drive the graphics terminal with modified versions of our current programs."

Minitab:

"Minitab does not currently have high resolution graphics capabilities. We do plan to add such capabilities, but the first edition of this will not be available in a regular release before January 1983. Several of our users have added graphic capabilities to Minitab. We expect initially to develop graphics for one or two environments, probably Tektronix 4010 and Siggraph Standard, and to begin with some basic capabilities... The attempt will be to: (1) Emphasize data analysis graphics (as opposed to presentation graphics.) (2) Support medium as well as high resolution graphics."

P-STAT: "P-Stat does not yet offer high quality graphics..... We have looked at graphics interfaces, and this seems to be the direction in which we will eventually follow....."

IV. INTEGRATED STATISTICAL GRAPHICS SOFTWARE FOR MINICOMPUTERS AND MAINFRAMES

A. Overview

Ideally, users of statistical graphics would like graphical and statistical capabilities integrated into a single package. This is the way most statistical graphics offerings for desktop computers, such as those described in Section II, are written. In contrast, the software described in Section III marries previously written statistical routines with more recently developed
graphics software. (Datagraf appears to be an in-between case.)

A new generation of software that integrates statistical analysis with high-quality computer graphics into a single package has recently emerged. Prominent among these are:

- Dataplot - developed at the National Bureau of Standards
- Statgraphics - developed at Princeton University
- S - developed at Bell Labs

These three offerings differ appreciably with regard to primary user (see Table 1), cost (see Table 2) and user-friendliness. (Statgraphics is aimed at the most general audience; S is aimed at a more limited audience who have greater sophistication in statistics and computers.)

B. Dataplot

Dataplot uses Tektronix Plot-10 graphics software to create statistical graphics. According to its developers, "Dataplot was developed originally in 1977 in response to data analysis problems encountered at the National Bureau of Standards (NBS). It has subsequently been the most heavily-used interactive graphics and non-linear fitting language at NBS. It is a... tool not only for 'raw' graphics, but also for manuscript preparation, modeling, data analysis, data summarization, and mathematical analysis. Dataplot may be run either in batch mode or interactively, although it was primarily designed for (and is most effectively used in) an interactive environment." Its capabilities are grouped under the following headings:

- General Graphics Diagrammatic
- Diagrammatic Graphics
- Fitting
- Graphical Data Analysis
- Non-Graphical Data Analysis
- Mathematical

Filliben (1980 and 1981) provides more information on Dataplot.

C. Statgraphics

Statgraphics, developed at Princeton University, is a new package with broad statistical and graphical capabilities. It contains an optional menu control scheme for general users. Experienced users can access the APL program functions directly and create new ones. Functions are organized into 24 chapters devoted to analysis of variance, basic plotting functions, cluster analysis, descriptive methods, estimation and testing, distribution functions, simulation and random numbers, forecasting, data input/output, exploratory data analysis, basic draw functions, categorized data analysis, multivariate statistics, nonparametric statistics, numerical analysis, sampling, quality control, regression analysis, smoothing, time series analysis, stochastic modeling, experimental design, special math functions and mathematical programming. Pohlemus (1982a and 1982b) provides more information on Statgraphics.

D. The S Package

S is an interactive language and system developed by the Bell Laboratories statistics research departments to reflect modern concepts in data analysis, data display techniques, and computing. S is an expression language; the expressions used include algebraic and functional expressions, which are interpreted and evaluated by S. The emphasis is not on all-inclusiveness of currently available capabilities, but on making it easy for a knowledgeable user to add desired capabilities. External support for S is not provided by Bell Labs. In fact, the description of S states that "It is essential that any group hoping to make the best use of
V. SOFTWARE UNDER DEVELOPMENT

Computer-generated statistical graphics is a rapidly growing field. Software is currently under development that utilizes 3-dimensional capabilities and motion. Specific offerings that are currently being developed and have been reported on at recent conferences are:

- **Grafstat**, under development at the IBM Watson Research Center, is "an APL workspace, designed for interactive scientific and engineering plotting and interactive data analysis. The workspace runs with a full-screen editing interface on the IBM 3277GA dual screen terminal." [see Heidleberger, Welch and Wu (1982)].

- **ALDS** [Analysis of Large Data Sets], under development at the Pacific Northwest Laboratory, is a "flexible data management/statistical analysis software system...that utilizes color, glyphs and motion to glimpse complex structure of multivariate large data collections." [see Nicholson (1982) and Nicholson and Littlefield (1982)].

- **The Orion I workstation**, is "an experimental computer graphics system built at the Stanford Linear Accelerator Center...to study applications of recent developments in computer graphics technology to statistics." Features include "real time motion graphics to display three dimensional scatter plots." [see Friedman, McDonald and Stuetzle (1982) and, also, Kolata (1982)].

- **Andrews (1982)** is conducting research that "illustrates the potential of microcomputers for simulating motions and thus for the display of 3-dimensional characteristics of data,..." using clouds of points.

VI. CONCLUDING REMARKS

This survey has provided an overview of advanced statistical graphics software, in terms of its capabilities, for integrating general statistical analysis with high-quality graphics. Considered have been:

1. Offerings for both presentation graphics and data analysis/exploration graphics.
2. Software for desktop computers, minicomputers and mainframes.
3. Packages directed at managers and business people, engineers and scientists, data analysts and statisticians, and various combinations thereof.
4. Offerings where the required level of statistical and/or computing knowledge of the user varies from little to fairly extensive.
5. Offerings with acquisition costs ranging from $150 to $40,000 or more (various time-sharing and licensing arrangements are also available for many of the offerings).

Most of the offerings provide color outputs, even though they may not currently be used in that manner, due to equipment limitations. These limitations will gradually disappear. New offerings will emphasize motion and 3-dimensional capabilities.

Only one of the 15 offerings (Statgraphics) allows the user to input data through graphics input devices such as joysticks, thumbwheels, and graphic tablets. We expect that new offerings will have increasing capabilities of this type.

We hope these comments will provide a useful starting point for those who would like to select statistical graphics software that best meets current and future needs.

Anyone desiring more information...
can contact P.M. Caporal, General Electric Company, 1000 Western Avenue, Mail Drop 37416, Lynn, Ma 01910 or G.J. Hahn, General Electric Company, Building K-1, Room 3A27, Schenectady, N.Y. 12345.

REFERENCES


### Table 1
**SUMMARY OF SOFTWARE FOR STATISTICAL GRAPHICS**

<table>
<thead>
<tr>
<th>SOFTWARE</th>
<th>PRIMARY USERS</th>
<th>PROGRAM (P) OR SUBROUTINE LIBRARY (S)</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Software for Desktop Computers:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP Series 70 Statistical Software</td>
<td>General</td>
<td>P</td>
<td>Basic statistical capabilities for HP 75</td>
</tr>
<tr>
<td>HP Series 80 Statistical Software</td>
<td>General</td>
<td>P</td>
<td>Basic statistical capabilities for HP 85, 86, and 87</td>
</tr>
<tr>
<td>HP Series 200 Statistics Library</td>
<td>General</td>
<td>P</td>
<td>Extensive statistical capabilities for HP 9816, 9826, and 9836</td>
</tr>
<tr>
<td>HP 9845 Statistics Library</td>
<td>General</td>
<td>P</td>
<td>Extensive statistical capabilities for HP 9845</td>
</tr>
<tr>
<td>Tektronix Plot 50 Applications Library</td>
<td>General</td>
<td>P</td>
<td>Extensive statistical capabilities for Tektronix 4050 series</td>
</tr>
<tr>
<td>Trend-Spotter '82</td>
<td>Business</td>
<td>P</td>
<td>Software for Trend-Spotter '82 Graphics Management Information System</td>
</tr>
<tr>
<td><strong>Extensions to Statistical Packages:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAS and SAS/Graph Interface</td>
<td>General and statisticians</td>
<td>P</td>
<td>Graphics interface for SAS statistical package</td>
</tr>
<tr>
<td>Datagraf</td>
<td>General and statisticians</td>
<td>P</td>
<td>Conversational program interfacing SAS, SAS/GRAPH, and DISSPLA.</td>
</tr>
<tr>
<td>SPSS and SPSS Graphics Option</td>
<td>General and statisticians</td>
<td>P</td>
<td>Graphics interface for SPSS statistical package with Tell-A-Graf interface option</td>
</tr>
<tr>
<td>STATII/PlotII Interface</td>
<td>General and statisticians</td>
<td>P</td>
<td>Graphics interface for STATII program on GE Mark III Time-sharing service</td>
</tr>
<tr>
<td>Sharp APL Statistics Library and Superplot Interface</td>
<td>Mainly statisticians</td>
<td>S</td>
<td>Graphics interface for Sharp APL Statistics Library</td>
</tr>
<tr>
<td><strong>Integrated Software:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataplot</td>
<td>Mainly engineers, scientists, and statisticians</td>
<td>P</td>
<td>Developed at National Bureau of Standards; mathematical, as well as extensive statistical capabilities</td>
</tr>
<tr>
<td>Statgraphics</td>
<td>General and statisticians</td>
<td>S, P</td>
<td>Developed at Princeton University; extensive statistical capabilities</td>
</tr>
<tr>
<td>S Package</td>
<td>Mainly statisticians</td>
<td>S</td>
<td>Expression language developed at Bell Labs; add-on capabilities emphasized; not supported</td>
</tr>
<tr>
<td>SOFTWARE</td>
<td>SOURCE</td>
<td>ACQUISITION COST*</td>
<td>HOST COMPUTER</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>-------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>HP Series 70 Statistical Software</td>
<td>Hewlett-Packard</td>
<td>$35</td>
<td>Desktop Computer</td>
</tr>
<tr>
<td>HP Series 80 Statistical Software</td>
<td>Hewlett-Packard</td>
<td>$295</td>
<td>Desktop Computer</td>
</tr>
<tr>
<td>HP Series 200 Statistical Library</td>
<td>Hewlett-Packard</td>
<td>$1500</td>
<td>Desktop Computer</td>
</tr>
<tr>
<td>HP 9845 Statistics Library</td>
<td>Hewlett-Packard</td>
<td>$350</td>
<td>Desktop Computer</td>
</tr>
<tr>
<td>Tektronix Plot-80 Applications Library</td>
<td>Tektronix, Inc.</td>
<td>$1250</td>
<td>Desktop Computer</td>
</tr>
<tr>
<td>Tandem PLOT ’82</td>
<td>Computer Graphics Corporation</td>
<td></td>
<td>Desktop Computer</td>
</tr>
</tbody>
</table>

**Extensions to Statistical Packages**

<table>
<thead>
<tr>
<th>SOFTWARE</th>
<th>SOURCE</th>
<th>ACQUISITION COST*</th>
<th>HOST COMPUTER</th>
<th>GRAPHICS OUTPUT DEVICES</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS and SAS/GRAF Interface</td>
<td>SAS Institute, Inc.</td>
<td>$5000</td>
<td>IBM 360/370 mainframe</td>
<td>Many output devices</td>
</tr>
<tr>
<td>SAS and Tell-E-Graph Interface</td>
<td>IBM</td>
<td>$5000</td>
<td>IBM 360/370 mainframe</td>
<td>Many output devices</td>
</tr>
<tr>
<td>CATstat</td>
<td>Marketing Dept.</td>
<td>$5000</td>
<td>IBM 360/370 mainframe</td>
<td>Many output devices</td>
</tr>
<tr>
<td>SPSS and SPSS Graphics Option Interface</td>
<td>Marketing Dept.</td>
<td>$5000</td>
<td>IBM 360/370 mainframe</td>
<td>Many output devices</td>
</tr>
<tr>
<td>Stat/It Update Interface</td>
<td>DEC</td>
<td>$500</td>
<td>DEC 10000 series</td>
<td>Many output devices</td>
</tr>
<tr>
<td>Integrated Software</td>
<td>National Technical Information Service</td>
<td>$990</td>
<td>IBM, CDC, DEC, Honeywell</td>
<td>Many output devices</td>
</tr>
<tr>
<td>Datagraf</td>
<td>Marketing Dept.</td>
<td>$14,000</td>
<td>DEC 10000 series</td>
<td>Many output devices</td>
</tr>
<tr>
<td>5 Package</td>
<td></td>
<td>$130</td>
<td>DEC PDP-11 and VAX</td>
<td>Many output devices</td>
</tr>
</tbody>
</table>

*Approximate cost as of Nov. 1981. Some offerings are available under other arrangement (e.g., rental or time-sharing service).