

UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

XYPLOT: a plotting program for the IBM PC
using Lotus 1-2-3 and a Hewlett-Packard plotter

by

R. Randall Schumann¹

Open-File Report 86-569-A

Although program tests have been made, no guarantee (expressed or implied) is made by the author or the U.S. Geological Survey regarding program correctness, accuracy, or proper execution on all computer systems.

Any use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Geological Survey. This report is preliminary and has not been reviewed for conformity with U.S. Geological Survey editorial standards.

¹Denver, Colorado

INTRODUCTION

Many scientists use Lotus 1-2-3 to enter, manipulate, analyze, and store data. However, Lotus' graphic capabilities are somewhat limited for scientific uses. XYPLOT was developed to augment Lotus' graphics by adding features and capabilities that are commonly used to analyze scientific data. The primary added features are the ability to plot data in semilog and log-log formats (as well as linear, which is supported by Lotus 1-2-3), and the ability to plot multiple pairs of X-Y data on one set of axes. Up to five pairs of X-Y data may be plotted on one graph in black-and-white or color (Fig. 1). A least-squares linear regression may be calculated and plotted for each data set.

HARDWARE REQUIREMENTS

The program was developed on an IBM PC/XT microcomputer using DOS version 3.1 and the IBM Advanced BASIC interpreter (BASICA), then compiled using the Microsoft QuickBASIC compiler to increase program execution speed. The program should run on most "PC-compatible" computers using MS-DOS version 2.0 or later. At least 256K of memory and one disk drive are required; two disk drives are preferred. Any type of display monitor and display adapter may be used. The program was developed using a Hewlett-Packard 7475A 6-pen plotter, but any multiple-pen plotter using the Hewlett-Packard Graphics Language (HPGL) may be used. In addition, a printer is needed if output of the regression equation coefficients is desired.

FILES INCLUDED ON THE DISTRIBUTION DISKETTE

The distribution diskette contains the following files:

XYPLOT.EXE	the XYPLOT program.
XYPLOT.DOC	documentation for the XYPLOT program.
XYDEMO.WK1	a Lotus 1-2-3 worksheet file containing a sample data set.
XYDEMO.DIF	a sample data file, readable by XYPLOT. It was produced by converting the XYDEMO.WK1 worksheet file to a DIF file.

DATA ENTRY AND TRANSFORMATION

Data may be entered in any program that can convert its files to Data Interchange Format (DIF) files in "columnwise" format. The procedure for entering and transforming data using Lotus 1-2-3 is described below.

The data files must be arranged in vertical columns with a non-numeric (label) heading at the top of each column (Table 1). One blank row between the labels and the start of the data is permitted but not necessary. All data must be numeric, but labels (such as "m") or blank cells to denote missing data are permitted. Missing values or labels other than column headings are converted to the value -999, which is used as a flag in the XYPLOT program causing these values to be ignored. In the unlikely event that a data set contains the value -999, that point will not be plotted; however, you may change the data value to -999.00001, or a similar value, to allow the value to be read as valid data without a significant loss of plotting accuracy.

XYPLOT DEMO

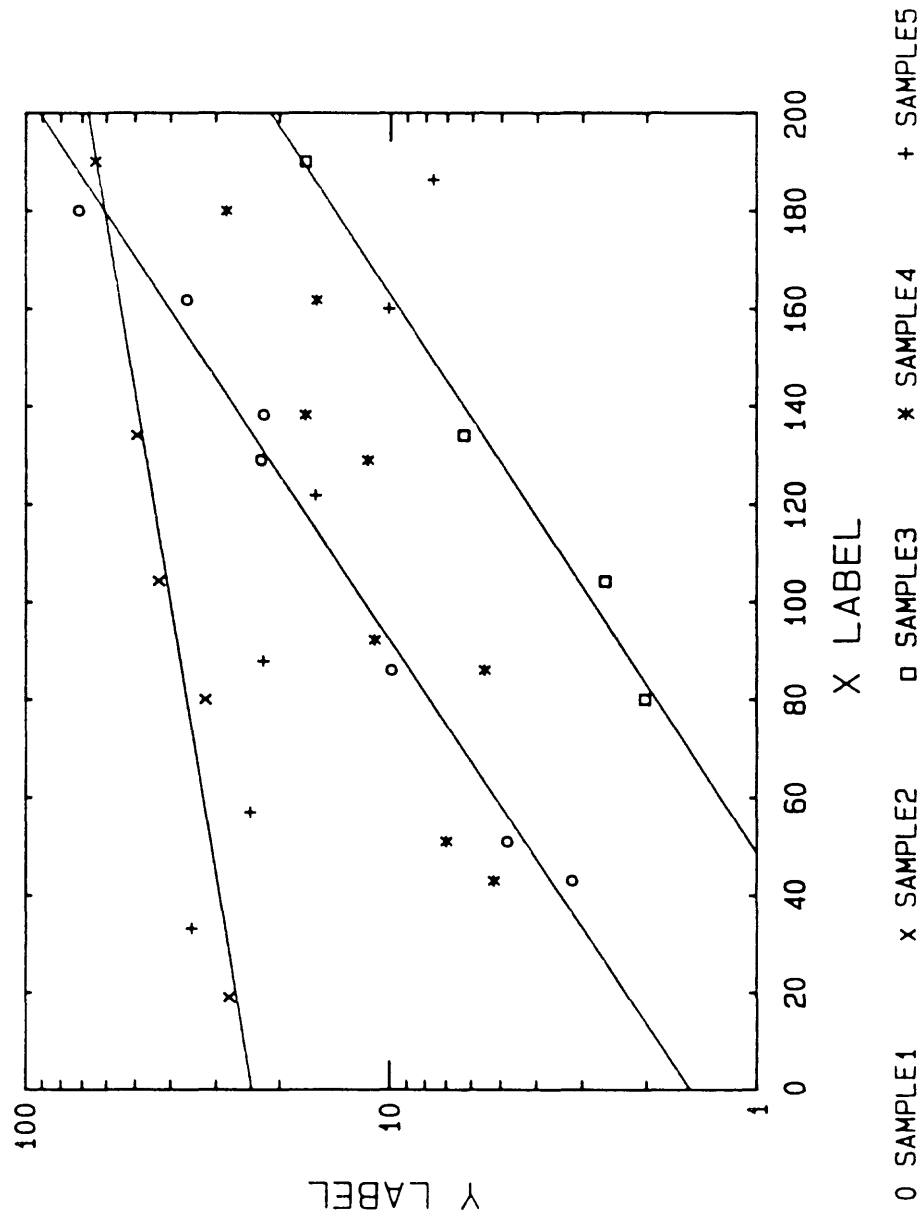


Figure 1. Sample plot generated by the XYPLOT program.

TABLE 1. Sample Lotus 1-2-3 worksheet used for XYPLOT data input.

X1	X2	X3	X4	X5	SAMPLE1	SAMPLE2	SAMPLE3	SAMPLE4	SAMPLE5
43	19	19	43	33.2	3.2	27.5	m	5.2	35
51	80	80	51	57	4.8	32	2.02	7	24
86	104.3	104.3	86	88	9.9	43	2.61	5.52	22.2
92.3	134	134	92.3	122		49.2	6.3	11	16
129	190	190	129	160	22.5	64	17	11.5	10.07
138.2			138.2	186.3	22.12			17	7.6
161.8			161.8		36			15.85	
180			180		71.4			28	

The spreadsheet must begin in cell A1 and all cells must be contiguous, i.e., there can be no totally blank columns. The data may also be taken from a larger spreadsheet by using the Lotus **/File Xtract** command to save the desired part of a larger spreadsheet. If this technique is used, be sure to specify values rather than formulas when the file is saved. The Xtracted file must then be reloaded and resaved (using the same filename) so that the resulting spreadsheet file can be properly converted to a DIF file by the Translate program.

After exiting the spreadsheet, use the Lotus **Translate** program to translate the 1-2-3 worksheet file to a DIF file. Consult the Lotus documentation for details on this program.

USING THE DOS "MODE" COMMAND

The serial port is not automatically initialized by the XYPLOT program, so the DOS "MODE" command must be executed prior to running XYPLOT. The MODE command is used to set communications parameters for the serial port and takes the form

```
MODE COM1: 96,N,8,1,P .
```

In this example, serial port COM1: has been set to 9600 baud, no parity, 8 data bits, and one stop bit. The "P" parameter enables a continuous retry of timeout errors when the serial port (i.e., the plotter) is accessed; it is strongly recommended that this parameter be included in the MODE statement. The DOS reference manual contains a detailed explanation of the MODE command and its parameters; consult your plotter manual for a listing of the recommended communications parameter settings for the plotter. It is recommended that the MODE command be included in an AUTOEXEC.BAT file so that it is executed automatically each time the computer is started up.

RUNNING THE XYPLOT PROGRAM

At the DOS prompt, type **XYPLOT** and press RETURN to start the XYPLOT program. When running the program for the first time, a setup routine titled "SYSTEM INFORMATION" is executed. You will be prompted for two pieces of information concerning your system's configuration.

The first prompt requests a drive name (and subdirectory, if desired) as a default location for data files. The default is drive B and may be selected by pressing RETURN. Any other drive specifier or directory name may be typed in place of the "B", for example "C:\PLOT\DATA", but the drive letter (C: in this example) must always be included. This tells the program where to find data files if a drive (and/or path) is not specified in response to the prompt for the name of the data file. The default may be overridden during program execution by adding a drive and path when specifying the data file name.

The second prompt is for the number corresponding to the serial port to which the plotter is connected. The choices are COM1 and COM2. To choose COM1 (the default), press RETURN. Enter a "2" and press RETURN to select COM2.

This system information is stored in a file named PLOTTER.INF and is read by the program in all subsequent runs. As long as the program is able to read

this file, the setup routine will not be executed. To change the default values, delete the PLOTTER.INF file and rerun the XYPLOT program. The setup routine will then be executed and a new PLOTTER.INF file created.

NOTE: the PLOTTER.INF file is an ASCII file containing the responses to the two prompts in the setup routine. As an alternative to deleting the PLOTTER.INF file, the parameters in this file may be changed using any text editor (word processor, EDLIN, etc.) capable of reading and writing ASCII files.

SELECTING PLOTTING OPTIONS

After the title screen has appeared and the setup routine has been executed (if necessary), the XYPLOT main menu appears (Figure 2). If your computer has a color display, the prompts will be in green, and the user responses will be colored blue. The current menu option ("active prompt") is indicated by a "double right arrow" symbol (>>), shown pointing to the "DATA FILE:" prompt in Figure 2. The pointer is moved to each menu item by using the down or right cursor keys to go forward (down through the menu), or the up or left cursor keys to go back (up through the menu). Pressing RETURN at any prompt moves the pointer forward to the next prompt, as does pressing the Y or N keys in response to prompts requiring only a yes or no response. Going "back" from the "DATA FILE:" prompt will move the pointer to the "PLOT GRAPH ON PLOTTER" prompt, and going "forward" from the "PLOT GRAPH ON PLOTTER" prompt returns the pointer to the "DATA FILE:" prompt. To end the program at any point, press the Esc key. You cannot move the cursor to any prompt other than "DATA FILE:" until the name of a data file has been entered. The menu prompts and appropriate responses to them are discussed below:

DATA FILE: the first step in setting up a plot is to enter the name of the .DIF file containing the data to be plotted, then press RETURN. If the file is on a disk in the default drive (as specified in the setup routine), the drive does not need to be specified. If the file is located on a drive or subdirectory other than the specified default, the drive and path must be specified with the file name. The ".DIF" file extension does not need to be entered unless you have used a file extension other than .DIF for the data file (this is not recommended; remember that XYPLOT can only read files in the DIF format). For example, if the default drive has been set to B, then XYDEMO, B:XYDEMO, XYDEMO.DIF, and B:XYDEMO.DIF are all acceptable entries if the file XYDEMO.DIF is on the diskette currently in drive B. The data file is read, then the lower part of the menu screen will be temporarily replaced by a listing of the column labels in the data file. Each column label will be preceded by a number; enter these numbers in response to the "ENTER THE NUMBER OF THE X(n) DATA SET:" and "ENTER THE NUMBER OF THE Y(n) DATA SET:" prompts (where n will be 1 for the first data set, up to 5 for five data sets). Enter one number per prompt; you will be prompted separately for each X and Y data column number. When you have selected all of the data sets (X-Y pairs) to be plotted, press RETURN at the next "ENTER THE NUMBER OF THE X(n) DATA SET:" prompt, and the main menu will reappear. Although the data file may contain any number of columns, a maximum of five pairs of data columns may be plotted in each program run (the program will automatically return to the main menu after five pairs of columns have been entered). A column

XYPLOT

»DATA FILE: B:XYDEMO.DIF

PLOT TYPE: LINEAR
REGRESSION: N HARD-COPY OF REGRESSION DATA: N
SYMBOLS AT DATA POINTS: Y CONNECT POINTS WITH LINE: N
PLOT AXES: Y PLOT SAMPLE LABELS: N
PLOT GRAPH IN COLOR: Y PEN SPEED: 10
X-AXIS LABEL:
Y-AXIS LABEL:
PLOT TITLE:
SCALING: AUTOMATIC
X AXIS- MIN: 0 MAX: 200 INC: 20
Y AXIS- MIN: 0 MAX: 80 INC: 10

PLOT GRAPH ON PLOTTER

press Esc to quit...

Figure 2. The XYPLOT main menu screen.

specifier may be used more than once, i.e., if column 1 is labelled "Time" and you wish to plot the concentrations of Sulfur (col. 2), Oxygen (col. 3), and Hydrogen (col. 4) over time, enter 1,2; 1,3; and 1,4 at the prompts. However, it is also possible to use different columns for the X axis data sets. For example, if you wish to plot the concentration of oxygen over time during two runs of an experiment, using different plot symbols for each run, the data may be entered in four separate columns labelled Time1 (column 1), Oxygen1 (col. 2), Time2 (col. 3), and Oxygen2 (col. 4) in the data file, then enter 1,2; 3,4 at the prompts. This would plot column 2 versus column 1 and column 4 versus column 3 on the same set of axes, but with different plotting symbols (o and x, in this case) for each set.

PLOT TYPE: four axis types are available in XYPlot. They are Linear, Log X, Log Y, and Log-Log. When the main menu first appears, "LINEAR" will be displayed as the current plot type. To change this setting, press L when the cursor is pointing to the "PLOT TYPE:" prompt. Each press of the L key will toggle another plot type. Continue to press the L key until the type of plot you wish to make is displayed, then press RETURN or one of the cursor movement (arrow) keys to continue.

REGRESSION: press Y for yes or N for no to select whether or not a linear regression will be calculated and plotted for the selected data set(s). You may simply press RETURN or one of the cursor movement keys if the response displayed on the screen does not need to be changed (this is true for all of the prompts on the main menu). Any data transformations to be performed will follow the type of plot selected, i.e., a log transformation will be performed on the data for each log axis selected (example: if a LOG X is selected for the plot type, the X data set(s) will be log transformed, but not the Y data set(s); if LOG-LOG is selected, both X and Y data sets will be log transformed for the regression calculations).

HARD-COPY OUTPUT OF REGRESSION DATA: if a regression is to be performed, press Y or N to select whether the regression equation(s) and r-squared value(s) for each data set will be sent to the printer. If N is selected, the regression lines will be plotted on the plotter, but the regression equations and r-squared values will not be printed. The printer is the only output device that regression data is sent to; it will not appear on the screen at any time.

SYMBOLS AT DATA POINTS: press Y or N to select whether symbols will be plotted at each data point. The symbols are o, x, □, *, and + (on the HP 7475A plotter; a different symbol may be plotted rather than the square on some models), and will be plotted in that order for each of the up to five data sets to be plotted. The only case in which a no (N) response is appropriate is when the data points are to be connected by lines (you may also have the data points connected by lines with plot symbols at each point).

CONNECT POINTS WITH LINE: enter Y or N to select whether data points will be connected by lines. If you wish lines only to be plotted, respond no to the "SYMBOLS AT DATA POINTS:" prompt and yes to this prompt. If you respond no to this prompt, the points will not be connected by lines

(remember to respond yes to the "SYMBOLS AT DATA POINTS:" prompt in this case!).

PLOT AXES: enter Y or N to select whether the axis and labels will be plotted before data is plotted. This option allows you to plot more than five sets of data on one pair of axes. To do this, select a new group of data sets (pairs of columns) from the current data file or a new file by moving the cursor to the "DATA FILE:" prompt and pressing RETURN (same data file) or entering a new file name. If you are plotting in color, you may wish to switch the pens to different positions or insert different colored pens in the pen carousel so that the plot symbols will be different colors for the next group of data sets to be plotted (this way, a red "o" may be distinguished from a blue "o"; note that if you are plotting in one color only and using plot symbols, it is somewhat impractical to plot more than five sets of data on one pair of axes because the first set of plot symbols will not be distinguishable from the second set). If distinguishing one data set from another on a plot is not important, there is virtually no limit to the number of data sets that may be plotted on one set of axes by selecting data sets five at a time, then entering N at the "PLOT AXES:" prompt after the first plot has been made.

PLOT SAMPLE LABELS: enter Y or N to select whether sample labels will be plotted, with their corresponding plot symbols, below the X axis at the bottom of the plot (see Fig. 1 for an example). The sample labels are the column labels for the "Y" column of each data set. If only one set of data is to be plotted, a label will not be plotted even if Y is entered in response to this prompt. No more than 5 sample labels may be plotted when plotting more than five sets of data on one pair of axes, because the sample labels from the first run will be overwritten by those from subsequent plot runs.

PLOT GRAPH IN COLOR: enter Y to plot in color or N to plot in black-and-white (one color). When plotting in color, pen number 1 is used to plot the axes and labels (a black pen is recommended), then pens 2 through 6 are used to plot each data set, i.e., data set 1 will be plotted with pen 2, data set 2 with pen 3, etc. A notice appears at the bottom of the screen indicating how many pens to insert into the pen holders.

PEN SPEED: this controls the speed of the pen as it plots; the pen speed factor ranges from 1 (slowest) to 10 (fastest). Enter a pen speed value or press RETURN or one of the cursor movement keys to retain the displayed value. A pen speed factor of 1 corresponds to approximately 4 cm/sec, and each increment of the pen speed factor increases pen speed by about 4 cm/sec up to the maximum pen speed of 38.1 cm/sec. The default value for the pen speed factor is 10, which is suitable for drafts of plots. A slower pen speed generally produces darker, neater lines. A pen speed factor of 1 or 2 is recommended when using rapidograph-type drafting pens and mylar drafting film.

X-AXIS LABEL, Y-AXIS LABEL, PLOT TITLE: for each of these prompts, enter the desired label, or press RETURN or use the cursor movement keys to retain the displayed label. To change a label, enter the new label and press RETURN. Press Ctrl-backspace to clear a label from the input line.

SCALING: press A for automatic scaling of the axes or M for manual scaling.

Automatic plot scaling is the default option. The scaling algorithm used is modified from that of Bean and Harvey (1984). Press M, then RETURN, to change some or all of the automatically scaled values. The cursor movement keys may be used to skip over values that do not need to be changed. For each axis, the minimum, maximum, and increment are displayed, and the minimum and maximum values for the selected data sets are displayed below the menu to guide in the selection of axis values. No value for increment is displayed, nor can one be entered, for log axes. In addition, only whole log cycles may be plotted. The following examples show how these values correspond to the scaling and labelling of plot axes:

- 1) linear axis: entering 0, 20, 5 will label the plot 0, 5, 10, 15, 20.
- 2) log axis: entering .01, 10 will label the plot .01, .1, 1, 10.
- 3) linear axis: entering 0, 18, 5 will label the plot 0, 5, 10, 15 (the scale is automatically adjusted down to the nearest multiple of the increment).
- 4) log axis: entering 1, 50 will label the plot 1, 10, 100 (the scale is automatically adjusted up to the next whole log cycle).

PLOT GRAPH ON PLOTTER: Check your responses to each prompt to insure that the plot is set up as desired, and check to make sure that the plotter is turned on and loaded with paper and pens, then press P to plot the graph on the plotter, or use the cursor movement keys to move to another prompt on the menu and make appropriate changes. If the regression option has been selected and you have chosen to send the regression data to the printer, the regression equations and r-squared values for each data set will be printed as each data set is plotted (check to be sure the printer is turned on and ready at this point). The prompt will brighten (on color displays, the prompt will turn from green to bright yellow) while the plotter is working; when the plot has been completed, the message "*** PLOT COMPLETED ***" will be displayed and the prompt will return to its normal color and intensity.

SOME NOTES ON PROGRAM EXECUTION

A least-squares linear regression is performed on each data set when that option is selected. The regression equation is of the form

$$Y = a + bX$$

and is calculated using these equations:

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

and

$$a = \bar{Y} - b\bar{X}$$

The coefficient of determination (r^2) is calculated using the following formula:

$$r^2 = \frac{b \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}$$

If a log scale is used for the plot, a log transformation is automatically performed on the data prior to calculating the regression. Some care must be taken in the interpretation of results presented in this situation. Although only nonzero, positive values may be plotted on log axes, it is possible to select log axes for data sets which include zero or negative values, and these values will be excluded from the plot. They will also be excluded from the regression calculations when log transformations are performed, so that the reported regression equation may not be truly representative of the data set.

When manually scaling plot axes, it is possible to specify origins and maxima that are greater than or less than the minimum and maximum values in the data sets, respectively. In this way, outliers which would otherwise drastically change the scale of a plot may be excluded from the plot if desired (this is sometimes referred to as data "clipping"). Those points will also be excluded from the regressions if the regression option is selected. After plotting has been completed, the total number of points excluded from the plot is reported on the screen.

REFERENCE CITED

Bean, S.R., and Harvey, K.C., 1984, An improved technique for scaling Cartesian graphics: American Laboratory, v. 16, no. 12, p. 68.