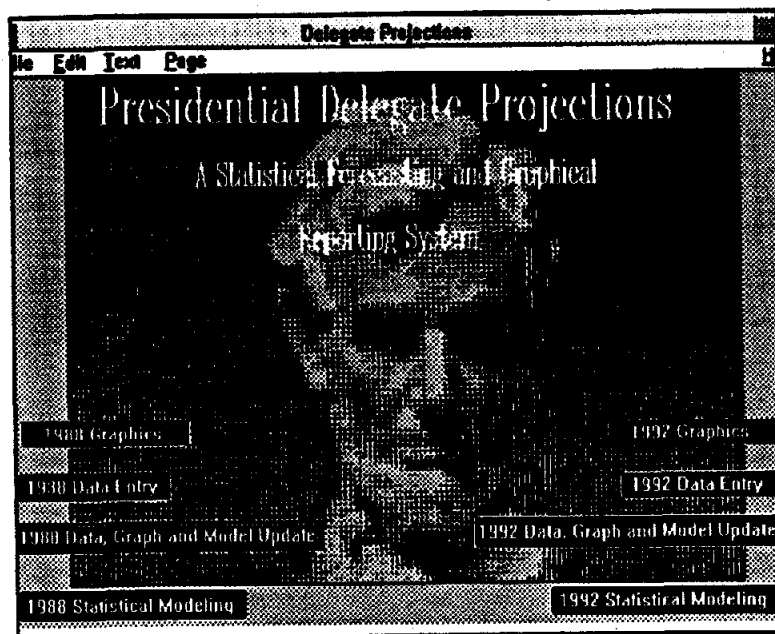


**Designing Data Analysis Systems by
Integrating PC SAS*, Windows 3.0*,
and Toolbook***

**Kathy and Dr. Mack Shelley
Iowa State University**



Introduction

This system has been designed to project candidate delegate counts for presidential primaries and caucuses. It uses statistical modeling to identify which variables are most important in explaining success in the electoral process. Software features include: (1) simplicity of design to facilitate structural modifications, (2) the use of statistical modeling in addition to the more traditional use of graphics for decision making, (3) user-friendliness provided by the ToolBook interface, and (4) built-in versatility which makes the system equally useful in classroom teaching or on the campaign trail. With the assistance of a statistician, the underlying model could easily be changed via PC-SAS to build systems as varied as crop yield testing, surgical procedure evaluation, or visual perception testing. The success of these 1988 forecasting and explanatory models will soon be tested in Iowa and New Hampshire in 1992.

Overall System Design

Two main design goals were to make the entire system as user-friendly as possible and to use statistical modeling and graphics as the basis for decision making. The user interface was designed entirely in Asymetrix ToolBook. The scripts behind the buttons were easily created by using the recording option which generated initial script code. These scripts were then modified as needed. The entire data entry procedure was formed in ToolBook. Note that Windows 3.0 was heavily used in the design of the system. PIF files were created for EMS memory management and the launching of the SAS programs. The Windows paint program was used to import the PCX-scanned image of President Lincoln. During actual operation, multi-tasking is used in the update process, and Windows Notepad is used to display the output from the predictive and explanatory statistical models. Figure 1 displays the overall layout of the programs.

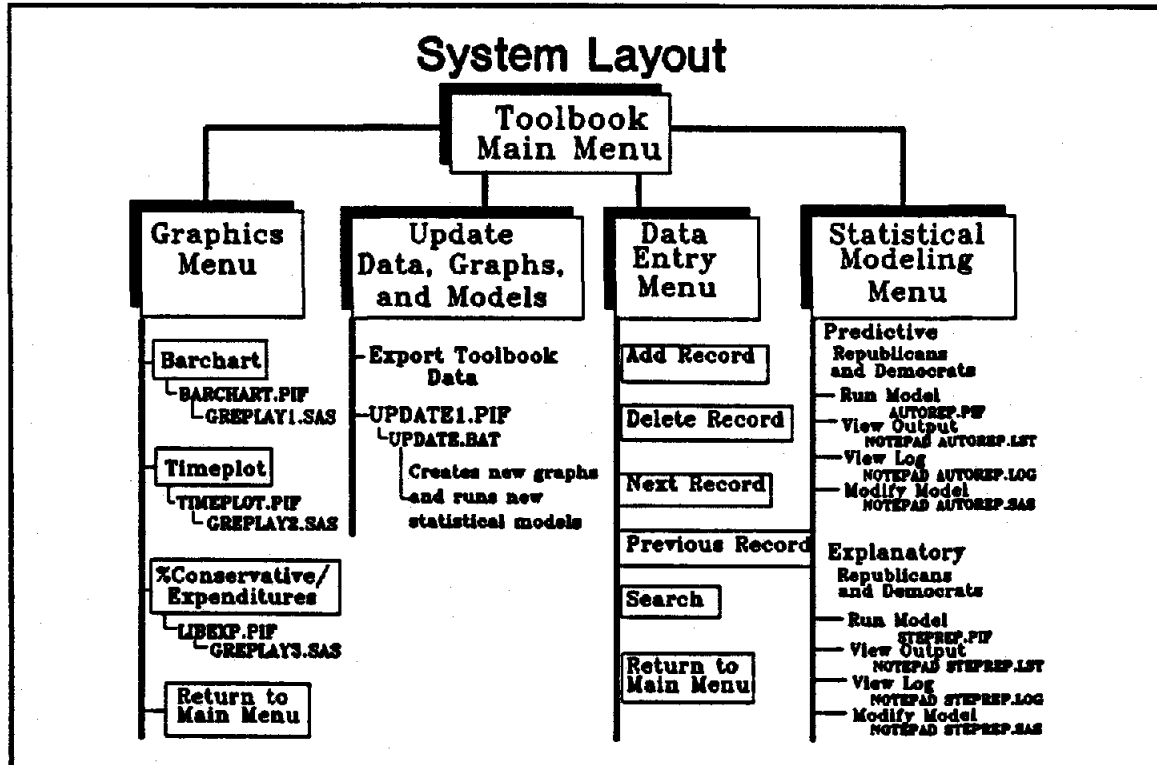


Figure 1

The Power of Graphics in Delegate Estimation

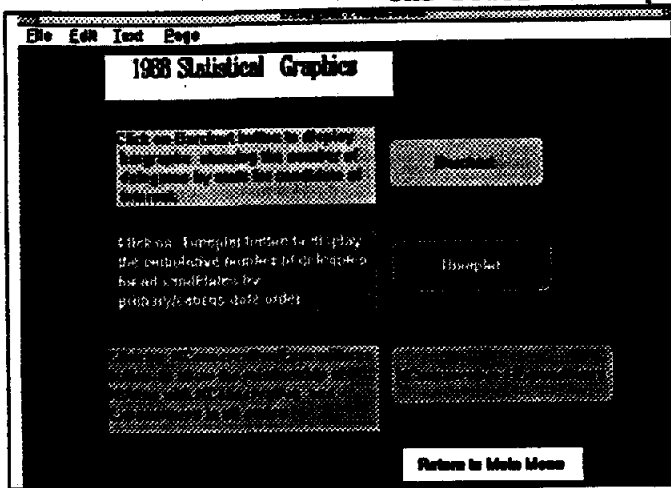


Figure 2

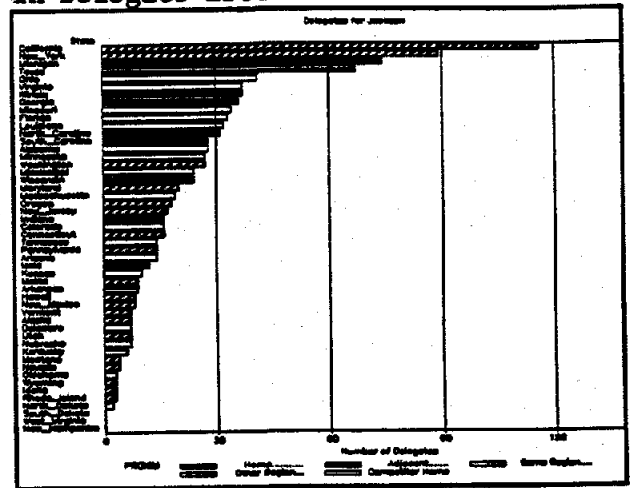


Figure 3

When chosen, the first button on the main menu displays the ToolBook screen shown in Figure 2. Each of the three types of graphs gives different information about the candidates' electoral success. These visual displays provided the motivation for viewing the data problem as both an explanatory regression model and also as a predictive autoregressive time series model. The bar graphs and three-dimensional plots are displayed for each candidate, and the timeplot of total cumulative delegates is viewed for each political party. PC-SAS was used to create three catalogues which contain these graphs. When the user presses the Barchart, Timeplot, or %Conservative/Expenditures button, a menu pops up displaying the list of available graphs. Upon selection, the graph is immediately displayed on the screen. This methodology optimizes SAS's usually slower graph display time by using graph catalogues.

The bar graphs in Figure 3 provide the candidates with information regarding their delegate count in descending order for each state. This graph gives immediate visual feedback to determine in which states they were strongest. The patterning in the bars shows state delegate counts by a proximity variable which we constructed to measure competition from the other candidates. The values for proximity are: 1-home state, 2-adjacent state, 3-same region, 4-other region, and 5-competitor's home state.

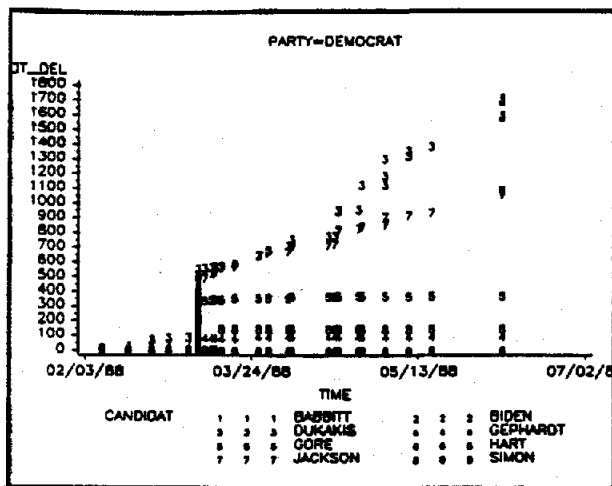


Figure 4

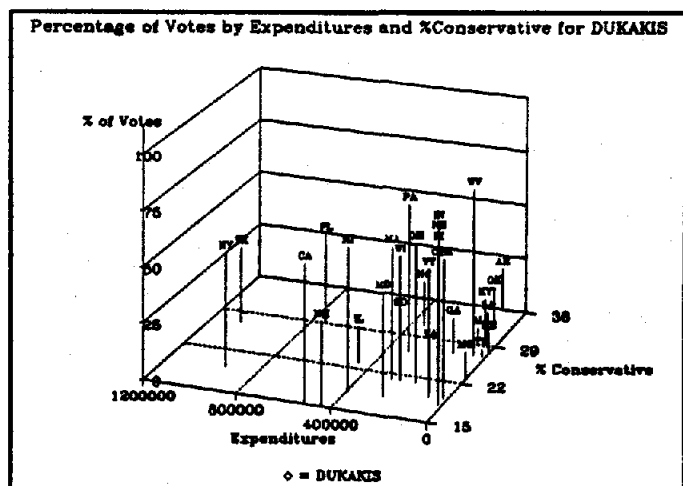


Figure 5

The timeplot graph in Figure 4 tracks each candidate's cumulative delegate count versus cumulative delegates for each of the other candidates. These data are presented by primary or caucus date. By plotting cumulative data over time, one can immediately see that "Super Tuesday" represented a significant differentiating date for the 1988 nomination contest. In the case of the Democrats, it also shows that Jackson provided the only serious competition for Dukakis. Figure 5 actually represents data in four dimensions. Each candidate's percentage of votes is graphed against expenditures and voters' self-defined conservatism by state. By looking at the low and high values, the viewer can decide if a candidate's support is regional, is based on conservative ideology, depends on the amount of expenditures, or is a combination of these.

The next two buttons on the main menu are for entering data and for updating the data, graphs, and statistical models. The data are entered and maintained within the ToolBook Data Entry Screen. Scripts were written for the add, delete, and search buttons. When the user selects the "Data, Graph and Model Update" button from the main menu, a script exports the ToolBook dataset to a text data file. Then the data are read by a SAS program which updates the master text data file, re-computes the graph catalogues, and re-runs the statistical models. This updating is the most time-consuming component of the software and is made much more tolerable to the user by

multi-tasking the SAS programs in the background. Figure 6 displays the Data Entry screen.

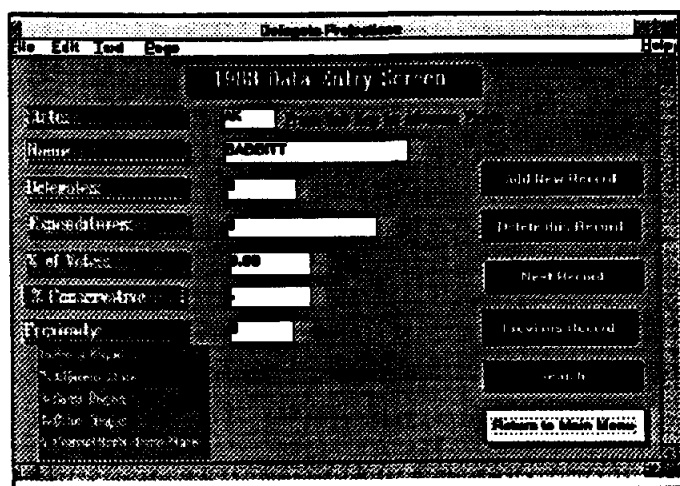


Figure 6

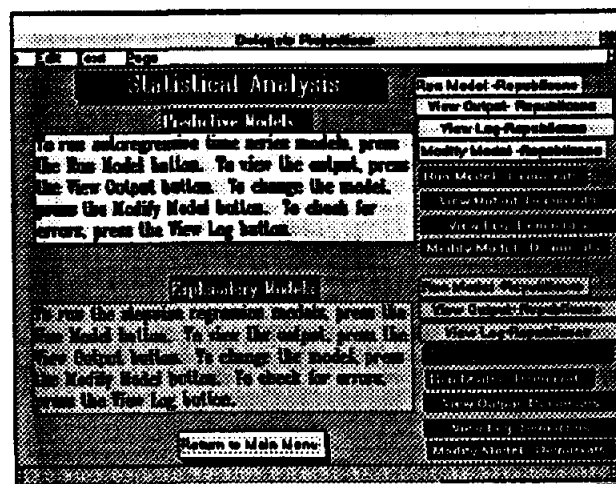


Figure 7

Statistical Modeling Menu Options

Even though PC-SAS is used for all statistical computations, Windows Notepad is used to modify the statistical models and is also used to view the log files and the output. This design feature enables non-SAS users to be able to change and re-estimate the models without needing to know the SAS display manager. By selecting the "Modify Model" button shown in Figure 7, a user only needs to change one or two statements or variable names and then select the "Run Model" button to estimate a new model. As previously mentioned, the models are automatically estimated any time the "Data, Graph and Model Update" button is chosen.

Results of Statistical Modeling

Two different linear models applications were pursued, in order to explain and predict the outcomes of state presidential nomination contests in 1988. These models also provide a framework with which to analyze presidential nominations in 1992 and beyond. We present results separately for the Democratic and Republican nominations in 1988. The predictive statistical model employed is a multiple

regression with autoregressive error structure, using PROC AUTOREG in PC-SAS. The explanatory statistical procedure, employed only for states holding primaries (as opposed to caucuses) is stepwise multiple regression, executed by PROC STEPWISE in PC-SAS. These models are discussed in Mack C. Shelley's paper entitled, "Presidential Elections and Campaign Money: Statistical and Graphical Evidence", which was presented at the 1991 annual American Statistical Association meeting held in Atlanta, Georgia.

The 1992 Campaign

With the advent of powerful laptop computers, complex applications are now possible while on the road. This was aptly demonstrated by the widespread use of laptops in Operation Desert Storm. In a similar manner, presidential candidate campaign staff can install ToolBook, Windows, and SAS software on a 386-based notebook PC. A 20-MHz, 60-megabyte hard drive notebook computer would allow campaign staff immediately to update election results as they came in. Remote communication software could be used to transfer the updated data to campaign headquarters via an internal modem. Early in the campaign, the 1988 models could be used to project 1992 delegates. As the campaign progresses and more 1992 data values are entered, the models could then be based upon 1992 data.

* PC SAS is a product of SAS Institute, Windows 3.0 is a product of Microsoft, Inc. and Toolbook is a product of Asymmetric Corporation.

The authors can be reached as follows:

Kathy Shelley Address: 117 Snedecor Hall, ISU, Ames, IA 50011 Ph:515-294-2260
E-Mail : SL\$KDS@ISUMVS
Mack Shelley Address: 210 Snedecor Hall, ISU, Ames, IA 50011 Ph:515-294-8346
E-Mail : S3\$MCS@ISUMVS