A Video Tour through ViSta 6.4, a Visual Statistical System based on **Lisp-Stat**

J. Gabriel Molina
Universitat de València

Rubén Ledesma
Universidad Nacional de Mar de Plata

Pedro M. Valero
Universitat de València

Forrest W. Young
University of North Carolina at Chapel Hill

Abstract

This paper offers a visual tour throughout **ViSta** 6.4, a freeware statistical program based on Lisp-Stat and focused on techniques for statistical visualization (Young 2004). This travel around **ViSta** is based on screen recordings that illustrate the main features of the program in action. The following aspects of **ViSta** 6.4 are displayed: the program's interface (**ViSta**'s desktop, menubar and pop-up menus, help system); its data management capabilities (data input and editing, data transformations); features associated to data analysis (data description, statistical modeling); and the options for Lisp-Stat development in **ViSta**. The video recordings associated to this tour (.wmv files) can be visualized at http://www.jstatsoft.org/v13/i08/ using the Internet Explorer navigator, or by clicking on the figures in the paper.

Keywords: Lisp-Stat, statistical visualization, statistical package.

1. Introduction

ViSta 6.4 (Young 2004) is a free-distribution, open source program oriented to do statistical data analysis. The development of **ViSta** is based on Lisp-stat and, since its origins in the early $90's$ (Young and Rheingans 1991), it has been growing to currently offer a wide range of statistical procedures and, more specifically, innovative graphical methods associated with the execution of these analyses as well as with the visualization of their results. As an example, Figure/Video 1 shows a spreadplot designed to visualize multivariate data.
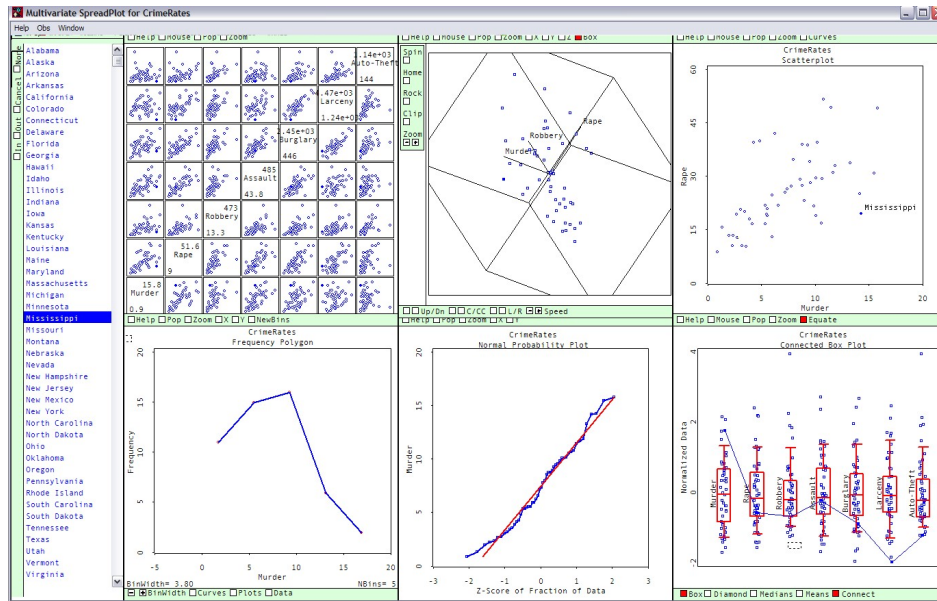
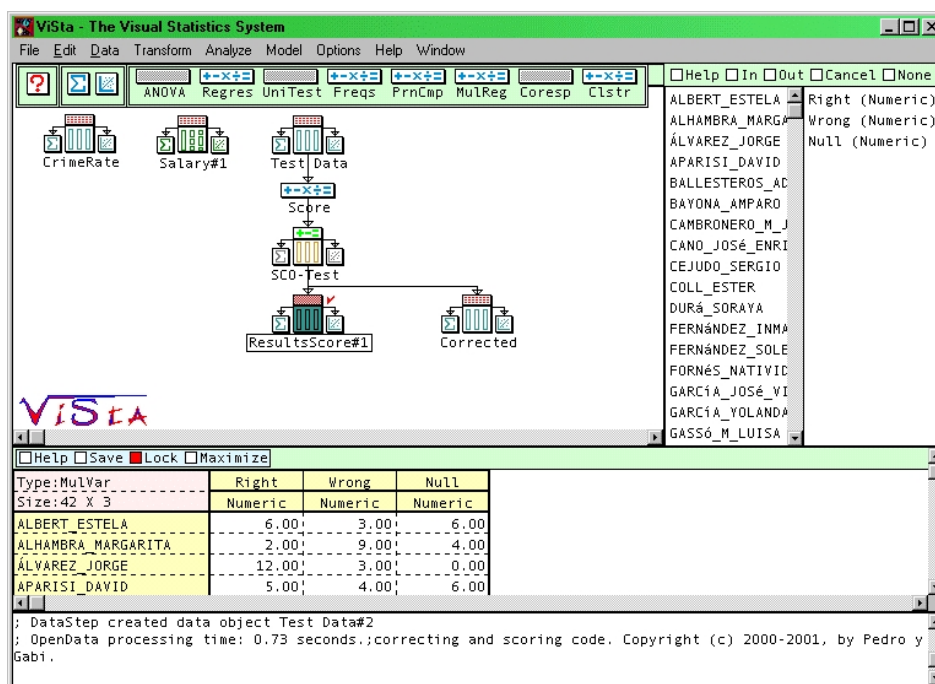Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_1.wmv

Figure 1: **ViSta**'s spreadplot for the visualization of multivariate datasets.

2. Running ViSta

If you want to install the **ViSta** program in your computer, you can do it by free from the **ViSta** web site (http://www.visualstats.org). **ViSta** 6.4 is available for the Macintosh and MS Windows operating systems. Once installed in your computer, when the program is run the **ViSta** Desktop window is presented (see Figure/Video 2), which is set by a menubar, a toolbar and three window-panes:

1. the WorkMap window, a graphical record of the steps you take during a data analysis session;

2. the Listener window, which displays messages about **ViSta** internal processing and lets you type and execute commands; and

3. the Selector window, which displays lists of the observations and variables in the currently active data file (when there is one). By default, you see the menu-bar and all three window-panes at the same time, however, you can use the various items of the Desktop menu to focus exclusively on one specific pane or to restore your view of all window-panes.

In addition to the **ViSta** Desktop windows, other windows are designed in **ViSta** which are displayed when some tasks are executed by the program user. This is the case of the Datasheet window, the SpreadPlot and Report windows, and the Xlisp-Stat window, some of which will be treated below. Finally, **ViSta** integrates a wide Help system oriented to give support to the program user about an important range of subjects (see Figure/Video 3).

Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_2.wmv

Figure 2: **ViSta**'s Desktop window.
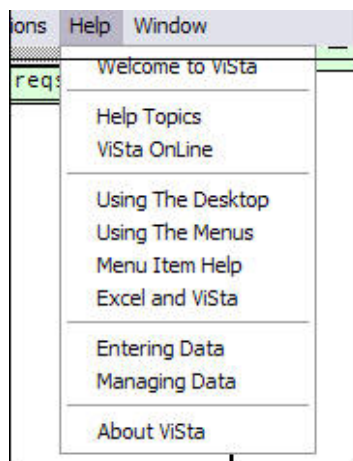
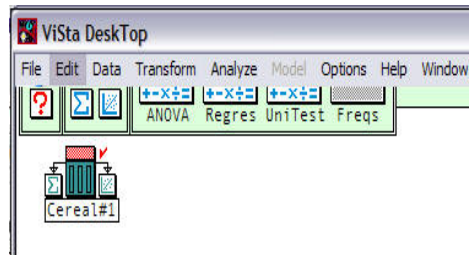Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_3.wmv

Figure 3: Help system in **ViSta**.

3. ViSta's interface main elements

3.1. The Menus and the Toolbar

ViSta desktop has menu-bar and pop-up menus. The menu-bar menus are organized into three groups (see Figure/Video 4):

1. the first two menus (File and Edit) are system menus which provide standard tools for editing and managing your data files;

2. the next four menus (Data, Transform, Analyze, and Model) contain data transformation and analysis tools; finally,

3. the last four menus (Options, Help, DeskTop, and Window) provide tools to support your interactions with **ViSta**.
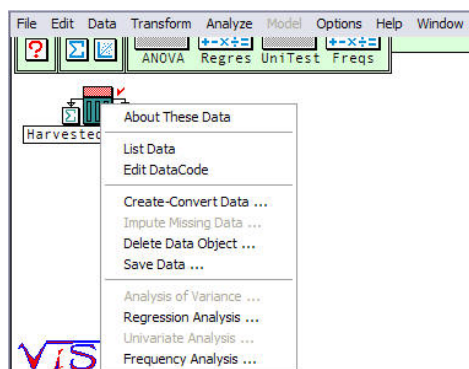

Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_4.wmv

Figure 4: The menu-bar menus in **ViSta**.

The pop-up menus in **ViSta** provide a context-sensitive approach to the same tools, organized according to the specific objects you are interacting with, rather than according to the nature of the tool you are using. You can pop-up a menu by right-clicking anywhere you wish to on the desktop. The menu that pops up depends on what you clicked on (see Figure/Video 5). Finally, **ViSta** toolbar contains buttons which provide instant access to some **ViSta** specific functions. The left three buttons provide access to Help and to a statistical report and visualization for the selected icon in the WorkMap. The right-hand group of buttons provide quick access to various statistical analysis models, which correspond to items of the Analysis menu. Right-clicking the toolbar gives you a menu which lets you redefine the number of buttons and the button actions.

3.2. The WorkMap

The WorkMap is the heart of **ViSta**'s data analysis and visualization environment. It provides an up-to-date picture of your data analysis session, and, with the help of the Toolbar, is the control center where you interact with **ViSta** to understand your data. The WorkMap grows as your analysis progresses. At first, before you start analyzing your data, there is no map. As you step through your data analysis, icons representing the statistical objects created by your data analysis steps are added to the WorkMap. The icons are usually connected to previous icons by lines that show the flow of your data analysis steps. Thus, the WorkMap records

Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_5.wmv

Figure 5: **ViSta**'s toolbar and pop-up menus.

the steps you take during a data analysis session, nevertheless, it is more than just an aid for remembering what you have done: you can interact with the map to re-visit previous steps in your analysis, and, when you wish, to analyze your data in new ways (see Figure/Video 6).
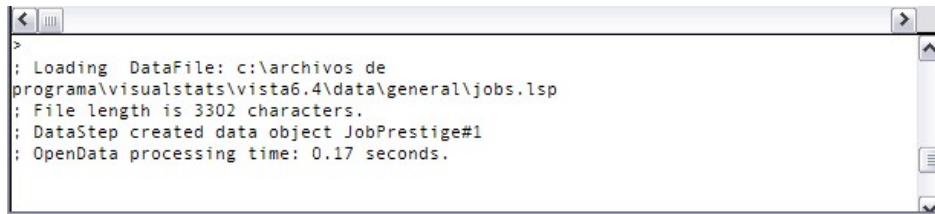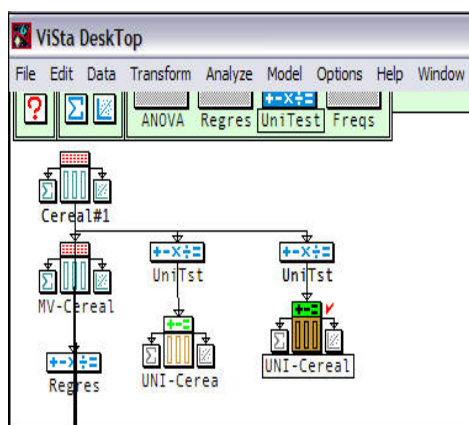

Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_6.wmv

Figure 6: Working with the **ViSta** WorkMap.

3.3. The Listener

The lower pane of **ViSta**'s desktop window is called the Listener. The Listener listens to the state of **ViSta**, displays messages about the system, and lets you type commands in **ViSta**'s underlying XLisp-Stat language. Really, there are two Listener windows: the latter, the desktop Listener (see Figure/Video 7), and the XLisp-Stat editor window, which must be activated through the Window menu's XLisp-Stat Window item. Both listeners work in the same way, however, the XLisp-Stat editor window provides extended functions for editing and evaluating sequences of commands. You can choose to use both approaches simultaneously, since it can be useful to use the desktop Listener to test statements in the program and the XLisp-Stat editor window to construct and save programs. Only one of these is active at a time.
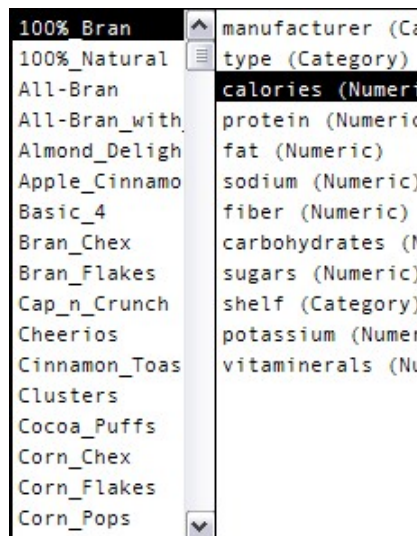
Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_7.wmv

Figure 7: Interacting with **ViSta** through the Listener.

3.4. The Selector

The Selector, the upper-right pane of **ViSta**'s DeskTop window, display lists of the observations and variables in the currently active data file (the selected one in the WorkMap). It is possible to select items in these lists to form a subset of active observations and variables. Thus, active variables and observations are those which are highlighted in the Selector lists, or if none are highlighted, those which are listed. The important point is that only the active variables will be used in subsequent transformations and analyses in **ViSta**. For example, you can create a new dataset containing only the desired observations and variables by selecting them first in the Selector window and, then, choosing the Create Data menu item in the Data menu (see Figure/Video 8).

Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_8.wmv

Figure 8: **ViSta**'s Selector windows.

4. Data management with ViSta

4.1. Defining data

ViSta offers different options to help the program users to define their datasets:

1. Entering and editing datasets: **ViSta** provides a datasheet window with the standard functions for entering new data as well as for modifying existing data (see Figure/Video 9). This datasheet does not have all the capabilities of spreadsheets programs like Excel, however, the user can take advantage of Vista's facility to import Excel data files, so this powerful spreadsheet could be used if preferred for creating and editing datasets.

□Help □Save ■Lock □Restore					
Type:MulVar	Murder	Rape	Robbery	Assault	Burglary
Size:50 X 7	Numeric	Numeric	Numeric	Numeric	Numeric
Alabama	14.20	25.20	96.80	278.30	1135.50
Alaska	10.80	51.60	96.80	284.00	1331.70
Arizona	9.50	34.20	138.20	312.30	2346.10
Arkansas	8.80	27.60	83.20	203.40	972.60

Play video (WMV format): `http://www.jstatsoft.org/v13/i08/video_9.wmv`

Figure 9: Entering and editing datasets.

2. Importing data: **ViSta**'s Import Data function (see Figure/Video 10) allows to open a text file with data organized according to some common standards. This enables defining data with programs can save the information in text format as it is the case of text processors, databases, and many others.
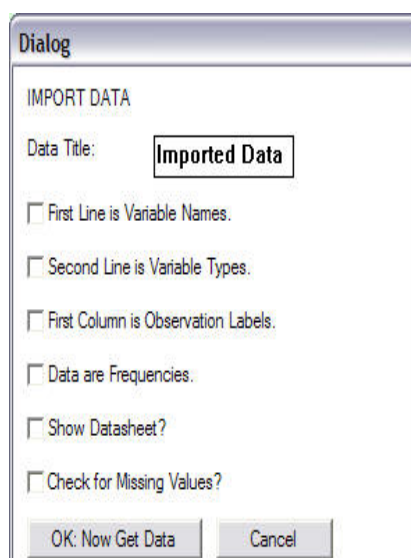
Play video (WMV format): `http://www.jstatsoft.org/v13/i08/video_10.wmv`

Figure 10: Importing a text data file.

3. Simulating data: **ViSta** lets to generate new data by simulating the process of sampling from a population according to a specific theoretical distribution (uniform, normal, F, chi-square, binomial, beta) and sample size defined by the program user (see Figure/Video 11) .
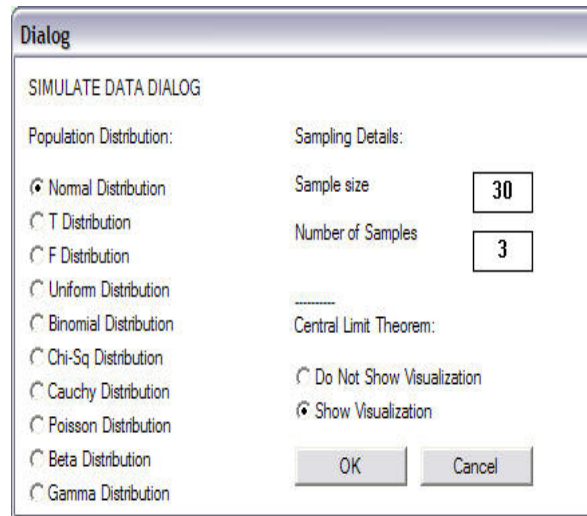
Figure 11: Options for simulating data in **ViSta**.

4. Borrowing data: **ViSta** comes with a library of more than a hundred sets of data. The Open Data item of the File menu shows a folder named DataLibary containing several folders with names such as anova, corresp, etc. These folders contain sets of data that are especially appropriate for the analysis method associated to the folder name.

4.2. Transforming data

ViSta includes a set of functions directly accessible from the Transform menu, which are oriented to satisfy the potential analyst's needs to transform the collected data before proceeding with its analysis (see Figure/Video 12). Some of these transformations are variable-oriented (rank, normal scores, dummy coding, etc.), whereas other are file-oriented (sort-permute, transpose, distances, etc.).
 ViStat's transformation functions generally work in a direct way, producing a new dataset that includes the new transformed variables, however, some of them also provide visualizations that allow to dynamically choosing the parameters of the transformation function in order to find the best result. This is the case of the Box-Cox transformations, where the visualization includes a slider that controls the exponent of the Box-Cox transformation for the currently selected variable, a histogram of the variable after the transformation, a line plot of the original versus the transformed data, a scatterplot matrix of the variables in the dataset, and a list of the labels of the observations (see Figure/Video 13).

Additionally, Vista integrates a variable language (ViVa) which allows to create or modify variables from the existing ones in the active dataset. ViVa statements are like algebraic equations and should be typed in the DeskTop Listener window.
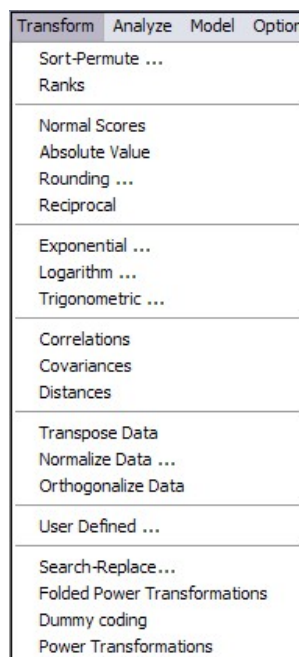
Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_12.wmv

Figure 12: Examples of data transformations in **ViSta**.

5. Data analysis with ViSta

5.1. Describing data

A double perspective is offered in **ViSta** in the description of any dataset, one based on a list of summary statistics for the variables included in the dataset, the another based on SpreadPlots, a **ViSta** graphical concept consisting of linked list and plot windows that shows different aspects of the selected dataset (Young, Valero-Mora, Faldowski, and Bann 2003). This double approximation corresponds to the two small icons associated to any icon representing a dataset in **ViSta**'s WorkMap, as well as to the Data menu items 'Summarize Data' and 'Visualize Data', respectively (see Figure/Video 14).

The list of summary statistics include, for each numeric variable, the variable's mean, standard deviation, variance, skewness, kurtosis, and the number of observations. In addition, the five number summary (the minimum, 2nd quartile, median, 4th quartile and maximum) is presented for each ordinal and numeric variable. It is possible to choose to also get information about ranges, interquartile ranges, correlations and covariances.

With regard to the SpreadPlots, there are different data visualizations depending on the selection of the variable types active at the time the visualization is chosen. When all variables are numeric, there are four possible visualizations: a multivariate visualization for data with 3 or more numeric variables (see Figure/Video 1); a bivariate visualization for data with two numeric variables; a univariate visualization for data with one numeric variable; and a Guided Tour visualization for data which have 6 or more numeric variables. Additionally, there is a classification visualization for data which have a numeric variable and one or more category
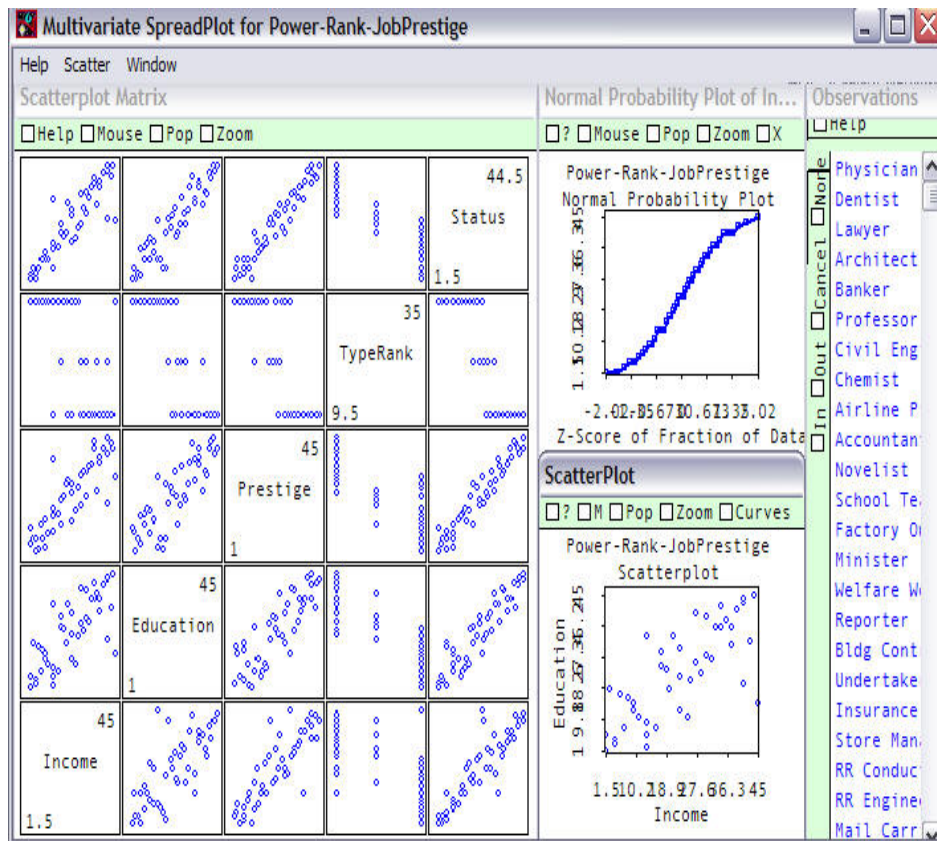
Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_13.wmv
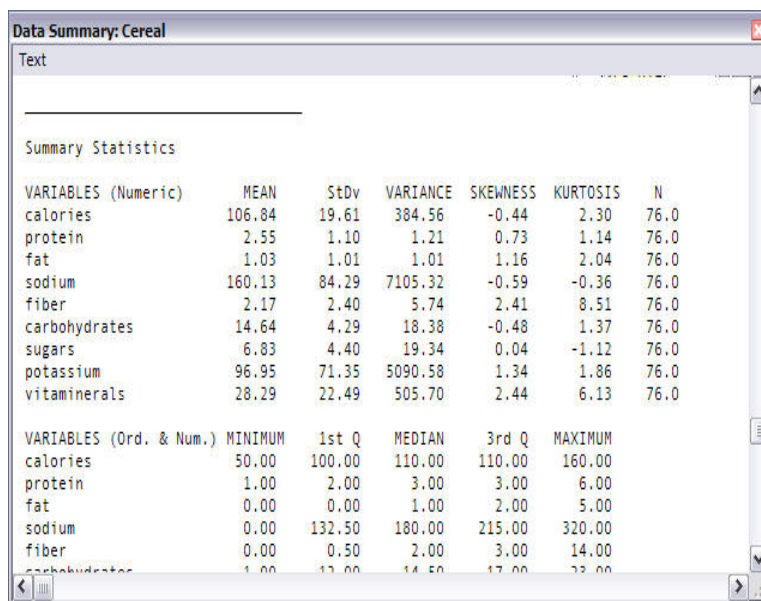
Figure 13: Box-Cox transformation function.

variables; a frequency visualization for frequency data (data which have numeric variables that specify frequency values); and a category visualization for category data (data which have one or more category variables and no numeric variables).

5.2. Modeling data

ViSta performs data analysis for a standard suite of statistical models, including: Univariate Tests (T, Z, non-parametric tests); ANOVA (N-way, balanced/unbalanced); Univariate and Multivariate Regression; Frequency Table Analysis; Principal Components Analysis; Correspondence Analysis; and Multidimensional Scaling. Additionally, after version 6.4, Vista features a plugin-based architecture so that the program user can write its own code implementing some new data analysis method and fairly integrate it in **ViSta** as plugin. Homogeneity Analysis; Psychometric Item Analysis, and Log-linear Analysis are some of the statistical models that have been already added to **ViSta** taking advantage of its plugin-based expanding capability.

Application of any of the statistical models considered in Vista works as follows: having a data object already selected in the WorkMap, selection of an Analysis menu item (or the corresponding ToolBar button) produces a new WorkMap icon representing the statistical model applied. Sometimes, the user will have to deal with a dialog that requests further

Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_14.wmv

Figure 14: Reports and visualizations (SpreadPlots) describing different kinds of data.

arguments needed for carrying out the analysis. On other hand, **ViSta** can decide to a limited extent what analyses are appropriate for the kind of data you are using. This decision is reflected in the changing appearance of the ToolBar buttons and the corresponding items in the Analysis menu (see Figure/Video 15).

As it is standard in **ViSta**, two ways of visualizing the results associated to the application of a statistical model are available to the user:

- a listing or report of numerical information about the analysis, that is, the way many classical programs offer the results;

- a specific SpreadPlot for the model applied to analyze the data, that is, a dynamic graphical visualization specifically designed to maximize understanding of the analysis results.

Each data analysis model has its own report and visualization (SpreadPlot) as it is the case in Figure/Video 15 for the Principal Component Analysis.

6. Enhancing ViSta

ViSta is an open software system, that is, its code is open to those who wish to enhance the system. Actually, **ViSta** design consists of a core engine plus plugins: the core is stable, while the plugins provide the path to growth. This program architecture lets **ViSta** be both stable and expandable, and provides for an obvious organization of **ViSta** developers into Application developers and System developers. Applications developers can write plugins that introduce new data analysis capabilities. The interested developer just has to write his/her Lisp-Stat code following a few specific rules and put this file in the **ViSta** plugins folder, in order to get
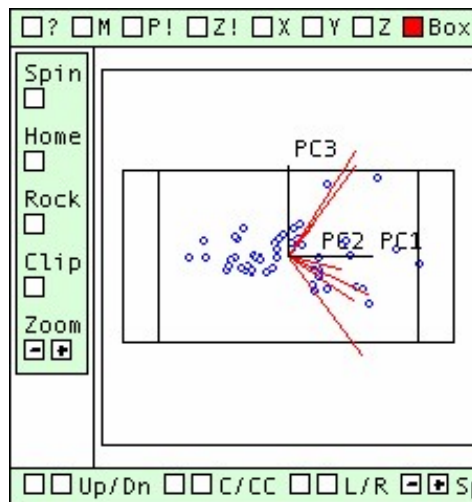
Play video (WMV format): http://www.jstatsoft.org/v13/i08/video_15.wmv

Figure 15: Modelling data: A complete Principal Components Analysis example.

his/her developments available in the **ViSta** environment. The way plugins must be written in order for **ViSta** to be able to read and execute these code files is described in the **ViSta** program documentation (Young 2004). An important feature of this programming capability is that the potential programmer can take advantage of the functions already available in this statistical program, avoiding unnecessary repetition.

Developers interested to enhance **ViSta**'s core engine can get access to portions of the source code. Anyway, because of the critical nature of systems development, and the importance of the core engine to the entire system, the system development effort is coordinated through the use of CVS, a version control system. CVS permits individuals who are part of a widely distributed development effort to work independently and simultaneously on a common set of code. The code is on your machine, where your CVS client coordinates your development with that of other developers, all the while permitting you to work independently from other developers. When you have completed your changes, the central CVS server will review all code changes to detect changes which conflict with those made by other system developers.

Acknowledgements

References

Young F (2004). "**ViSta**, the Visual Statistic System." URL http://www.visualstats.org.

Young F, Rheingans P (1991). "Visualizing High-Dimensional Data." *The IBM Journal of Research and Development*, **35(1/2)**(3), 513–518.

Young F, Valero-Mora P, Faldowski R, Bann C (2003). "Gossip: The Architecture of Spread-plots." *Journal of Computational and Graphical Statistics*, **12**, 80–100.

Affiliation:

J. Gabriel Molina
Facultad de Psicología
Universitat de València
Avda. Blasco Ibanez, 21
46010 Valencia, Spain
E-mail: gabriel.molina@uv.es
URL: http://www.visualstats.org