

Visual Statistical Analysis using **ViSta** The Visual Statistics System

Forrest W. Young & Carla M. Bann
L.L. Thurstone Psychometric Laboratory
University of North Carolina, Chapel Hill, NC USA

Outline

1. **ViSta's audience ranges from novices to experts. It is best suited to teaching multivariate, computational and graphical statistics.**
2. **ViSta runs on MS-Windows, Macs, and under Unix. It is free, open and extensible. The code, documentation and references are available on line at <http://forrest.psych.unc.edu/>.**
3. **ViSta features a Structured Graphical Interface including:**
 - A) **WorkMaps that visually summarize your data analysis session**
 - B) **GuideMaps that visually guide your data analysis**
4. **ViSta features Statistical Visualization and Analysis, including**
 - A) **Univariate - T-tests (etc.), ANOVA, Regression (OLS, robust)**
 - B) **Multivariate - Regression, Principal Components Analysis, Multidimensional Scaling, Correspondence Analysis**
5. **ViSta has a Language Interface including:**
 - A) **ViDAL, ViSta's Data Analysis Language (keyboard or scripts)**
 - B) **XLispStat: object-oriented statistical computing language**

1: ViSta's Audience

ViSta is best for teaching. It is being used in:

- **multivariate data analysis classes.**
- **computational and graphical statistics classes.**
- **introductory statistics (as a supplemental system since ViSta does not yet have a complete selection of basic statistical capabilities).**

ViSta is also useful for research and development in computational statistics and statistical graphics.

ViSta is designed for a wide ranges of users:

- **ViSta provides seamlessly integrated data analysis environments specifically tailored to the user's level of expertise.**
- **Guidance is available for students and novices.**
- **A structured graphical user interface is available for all users.**
- **A menu interface is available for users who don't need guidance.**
- **A command line interface is available for sophisticated users and those who don't like graphical interfaces.**
- **GuideTools are available for teachers and other experts to create guidance for students and novices.**
- **The complete Lisp-Stat (Tierney, 1990) programming environment is available to researchers, graduate students and programmers who wish to extend ViSta's capabilities.**

2: ViSta runs on MS-Windows, Macs & Unix. It is Free, Open & Extensible

ViSta runs under

- **MS-Windows (3.1 & 95)**
- **MacOS (68040, PowerPC)**
- **Unix with X11**

ViSta is free from <http://forrest.psych.unc.edu/>.

- **Code may be freely copied and redistributed, with certain restrictions.**
- **Documentation is also available for free from the above site.**

ViSta is Open:

- **All of the code is available to the programmer. This includes both the Lisp code for ViSta and the C and Lisp code for XLisp-Stat.**
- **This code can be used as the basis for developing new code that extends the system's capabilities.**

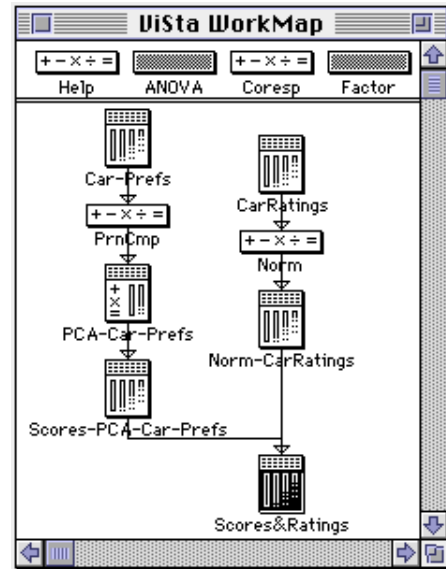
ViSta is Extensible:

- **Programs written in Lisp, FORTRAN or C are accessible from within ViSta.**
- **New ViSta model-objects can be written to extend the range of desired “prepackaged” statistical computations.**
- **New graphical-objects can be written to implement new statistical visualization ideas.**

3: ViSta's Structured Graphical Interface

WorkMaps

- WorkMaps visualize the structure of an on-going data analysis
- WorkMaps help you remember the steps of your analysis
- WorkMaps let you return to earlier steps in the analysis
- WorkMaps help you communicate the analysis to others



GuideMaps

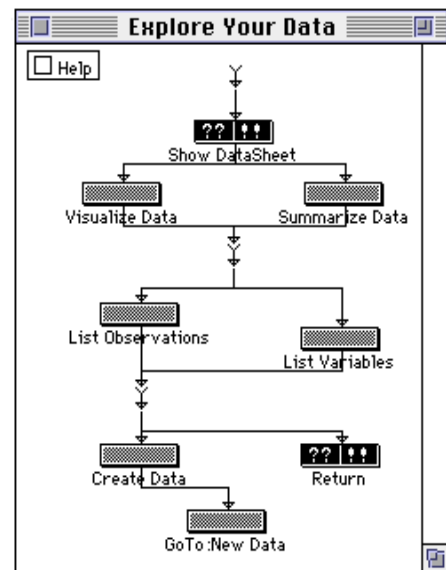
- GuideMaps visually guide users with no knowledge about statistical analysis through the analysis process.
- GuideMaps provide structured context-sensitive help, as well as guidance.

SpreadPlots:

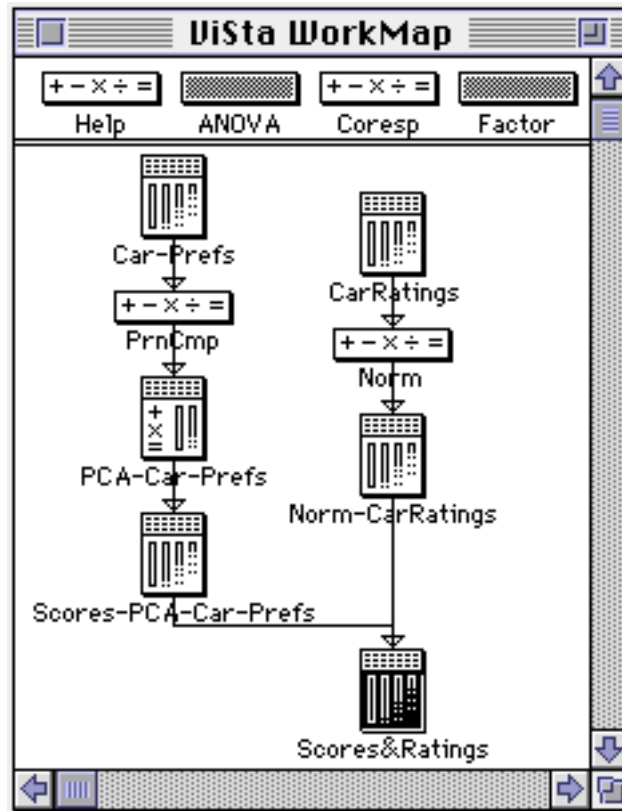
- Dynamic visualizations that reveal data and model structure.

ViSta also has:

- DataSheets to display & edit data.
- Text windows that show statistics, variable and observation names, help, etc.



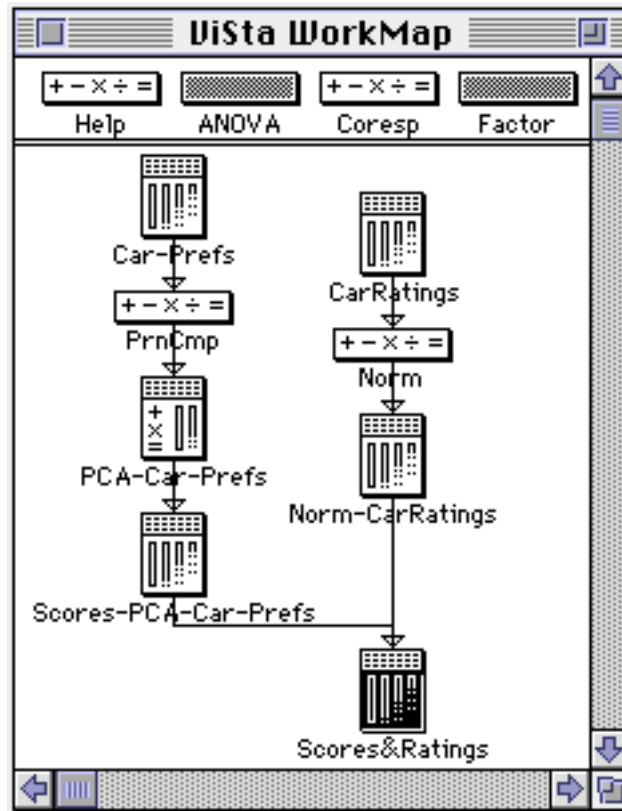
3.1 WorkMaps



ViSta's WorkMaps:

- visualize the structure of an on-going data analysis session.
- are created by ViSta as the data analysis session progresses.
- help you remember the data analysis steps you've taken.
- let you return to earlier steps for new analyses
- help you communicate your analysis to others

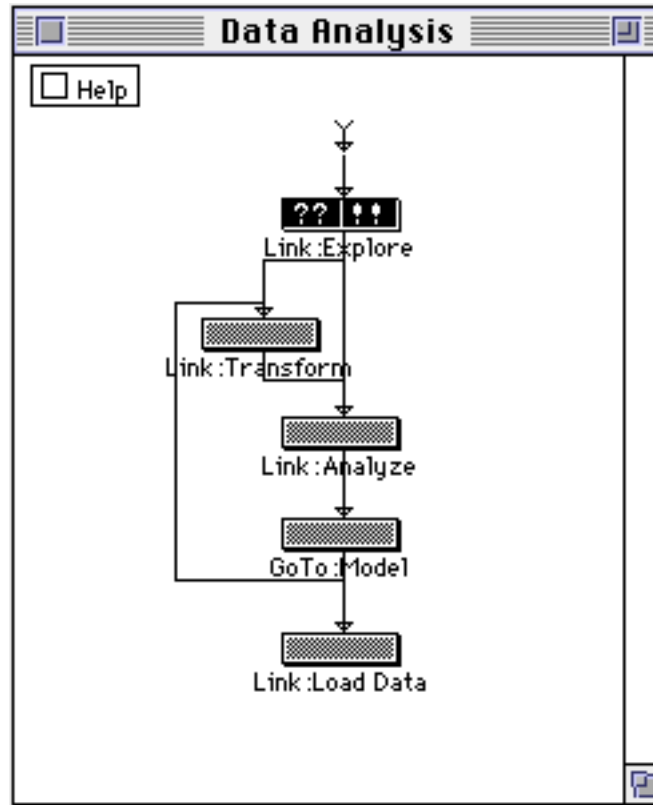
3.1 WorkMaps



In this example the analyst

1. Loaded in data named Car-Prefs
2. Did a principal component analysis (PrnCmp) of these data, producing the model named PCA-Car-Prefs .
3. Created an output dataset of the component scores named Scores-PCA-Car-Prefs
4. Loaded in the CarRatings data.
5. Normalized these data.
6. Merged the normalized data with the component scores, creating the Scores&Ratings data.

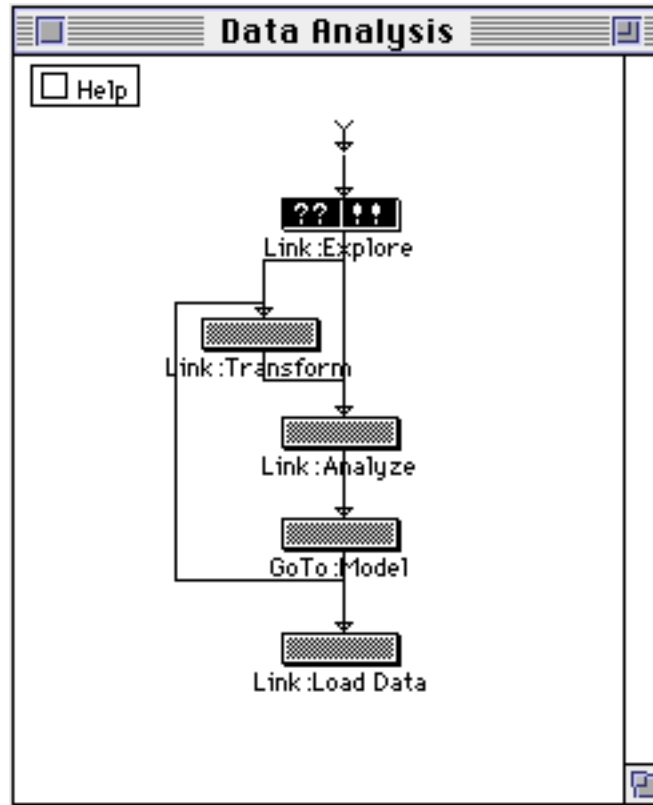
3.2 GuideMaps



ViSta's GuideMaps:

- visually guide you through the steps of your data analysis session.
- buttons show analysis steps: Dark buttons show suggested steps.
- change after a step is taken. The button highlighting changes to show you which actions can be taken next.
- show the step sequence by arrows connecting buttons.
- let you make choices by clicking on highlighted buttons.
- buttons have a !! side which makes a data analysis step happen, or links to another guidemap.
- buttons have a ?? side which gives you help about the step.

3.2 The Data Analysis GuideMap



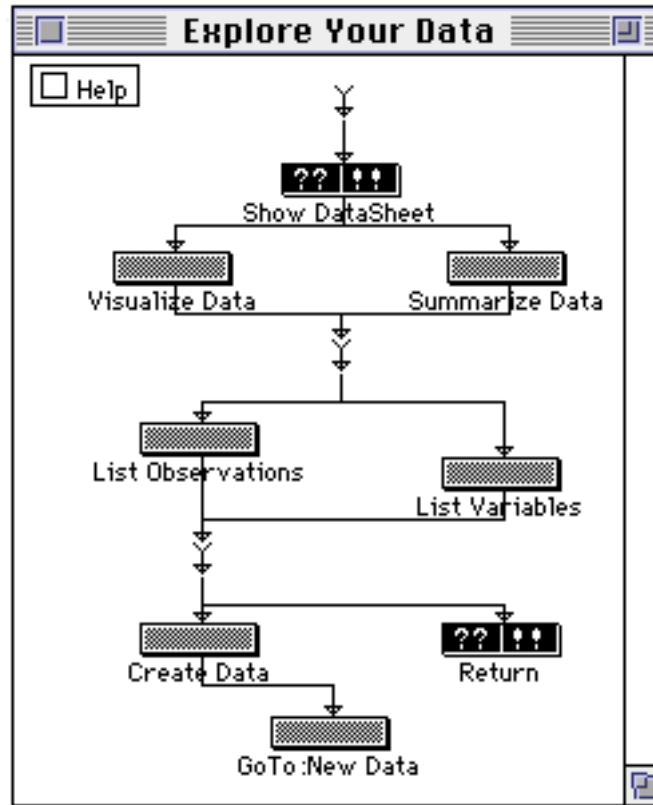
This guidemap, titled “Data Analysis”, is the first guidemap shown to you. It presents the basic data analysis cycle. It guides you to

- 1. Explore the data**
- 2. Transform the data (this step may be skipped)**
- 3. Analyze the data**
- 4. Look at the model created by the analysis.**

These four steps may be cycled many times. Some buttons are links to other guidemaps. Some are data analysis actions. In the “Data Analysis” guidemap all buttons are links to other guidemaps.

- After clicking “Link:Explore” you see a new guidemap:**

3.2 The Data Exploration GuideMap



- After you click on the "Link:Explore" button in the "Data Analysis" guidemap, you see the "Explore Your Data" guidemap.
- The button highlighting in the new guidemap shows you are to use the "Show Datasheet" button (or return to the "Data Analysis" guidemap). When you do this you will see the datasheet.
- After closing the datasheet, the "Show Datasheet" button turns gray and the next two buttons highlight.
- Both of these buttons must be used before the next two.
- In this way you are guided to take data analysis steps in a specific order.

4 SpreadPlots for Statistical Visualization

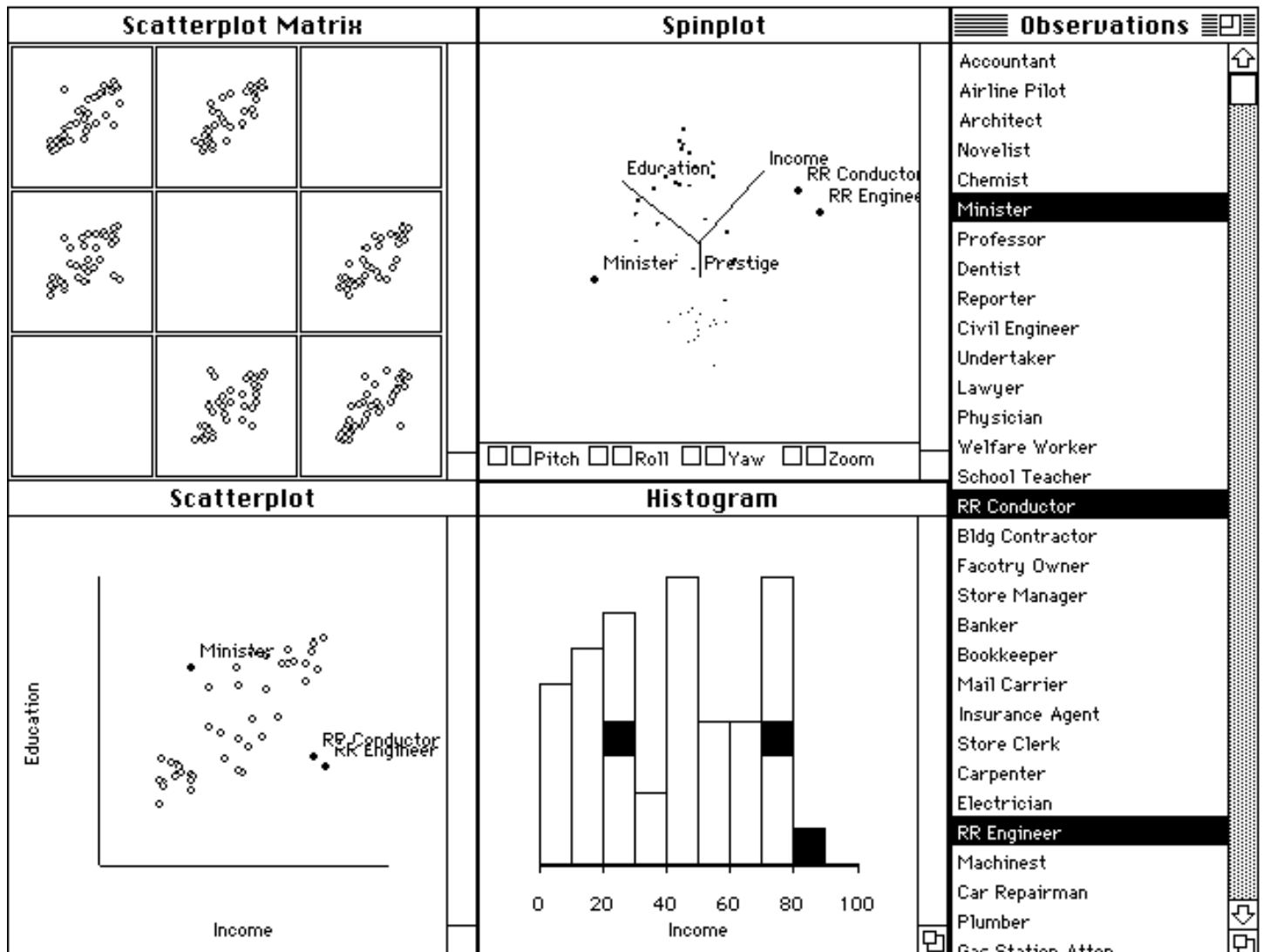
- **ViSta features SpreadPlots, state-of-the art visualization techniques that help you see and explore the structure of data and models:**
- **SpreadPlots are multi-window, linked, dynamic statistical graphics.**
- **Linkage: A SpreadPlot's individual plots can be linked by the data's observations or variables.**
- **Dynamic Graphics: The spinplot spins to communicate 3D structure. All plots can be brushed.**

4.1 Statistical Visualization and Analysis: Exploratory & Descriptive

Exploratory and Descriptive Data Analysis & Visualization

- **Dynamic Exploratory SpreadPlots include linked Histograms, Boxplots, Diamond Plots, Dotplots, Scatterplots, Biplots, Spinplots, Scatterplot Matrices. These plots support brushing and labeling, and are dynamically linked.**
- **In the example on the next slide, labels in the Observations window have been selected. These observations also appear in the spinplot, scatterplot and histogram. Points in these three plots may be brushed, labeled, colored, and given distinct symbols. These plots are linked with the other plots.**
- **Variable Linkage: Cells of the scatterplot-matrix can be clicked on to select variables for display in the other plots.**
- **Descriptive Statistics including Means, Standard Deviations, Variances, Ranges, Quartiles, Medians, Correlations, Covariances, Distances.**

A SpreadPlot for Exploring Multivariate Data



4.2 Statistical Visualization and Analysis: Univariate

Univariate Analyses and Visualizations

- **Univariate Tests including T- and Z-tests (and confidence intervals) for single sample, paired samples and two independent samples data, with Wilcoxon Signed-Rank and Mann-Whitney tests in appropriate situations.**

Univariate SpreadPlot -

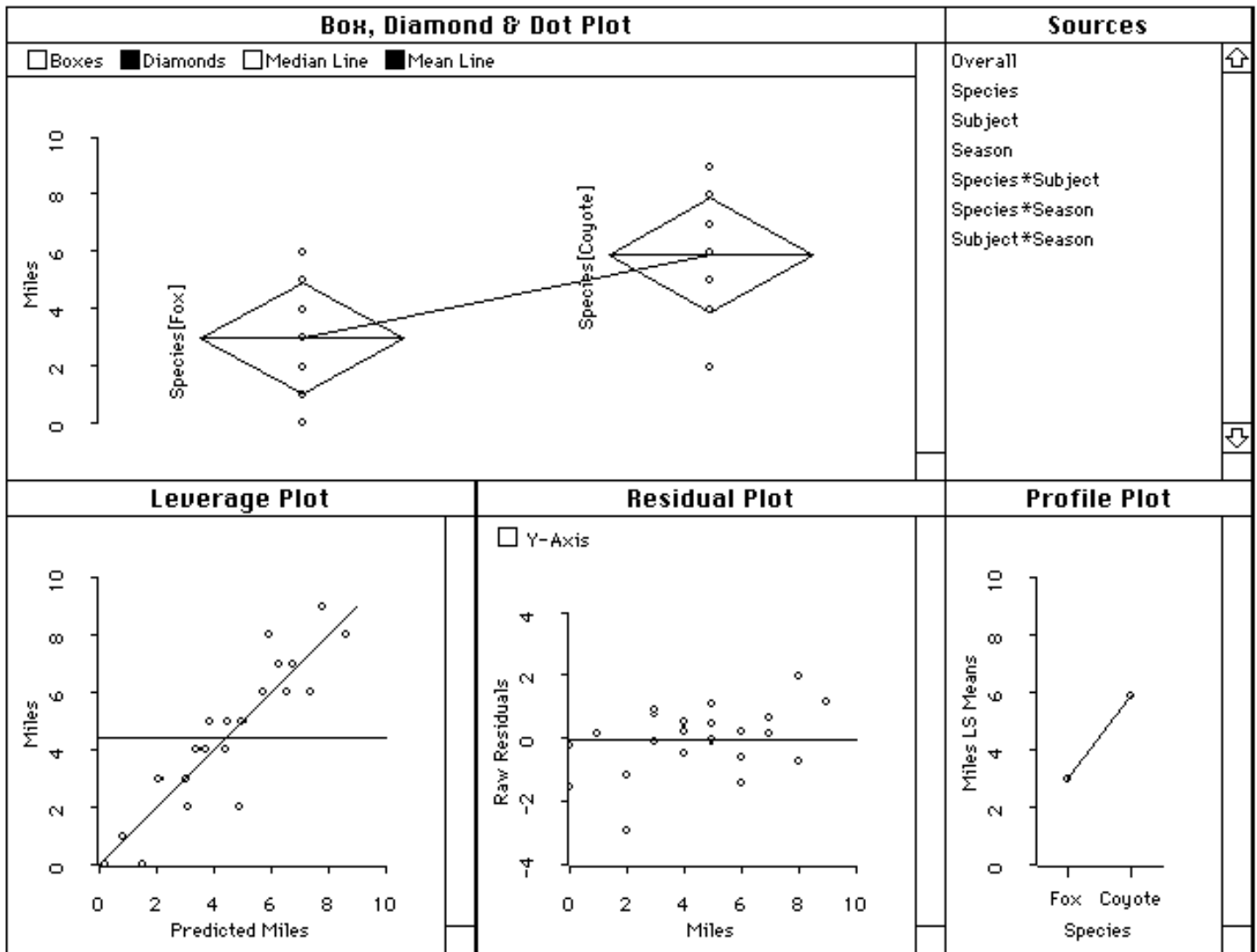
- **ANOVA - Univariate Analysis of Variance for orthogonal one- or multi-way data. Model may or may not include two-way (but not higher-way) interactions.**

ANOVA SpreadPlot - The ANOVA visualization is a spreadplot composed of a boxplot, diamond plot, quantile plot, quantile-quantile plot and effects plot.

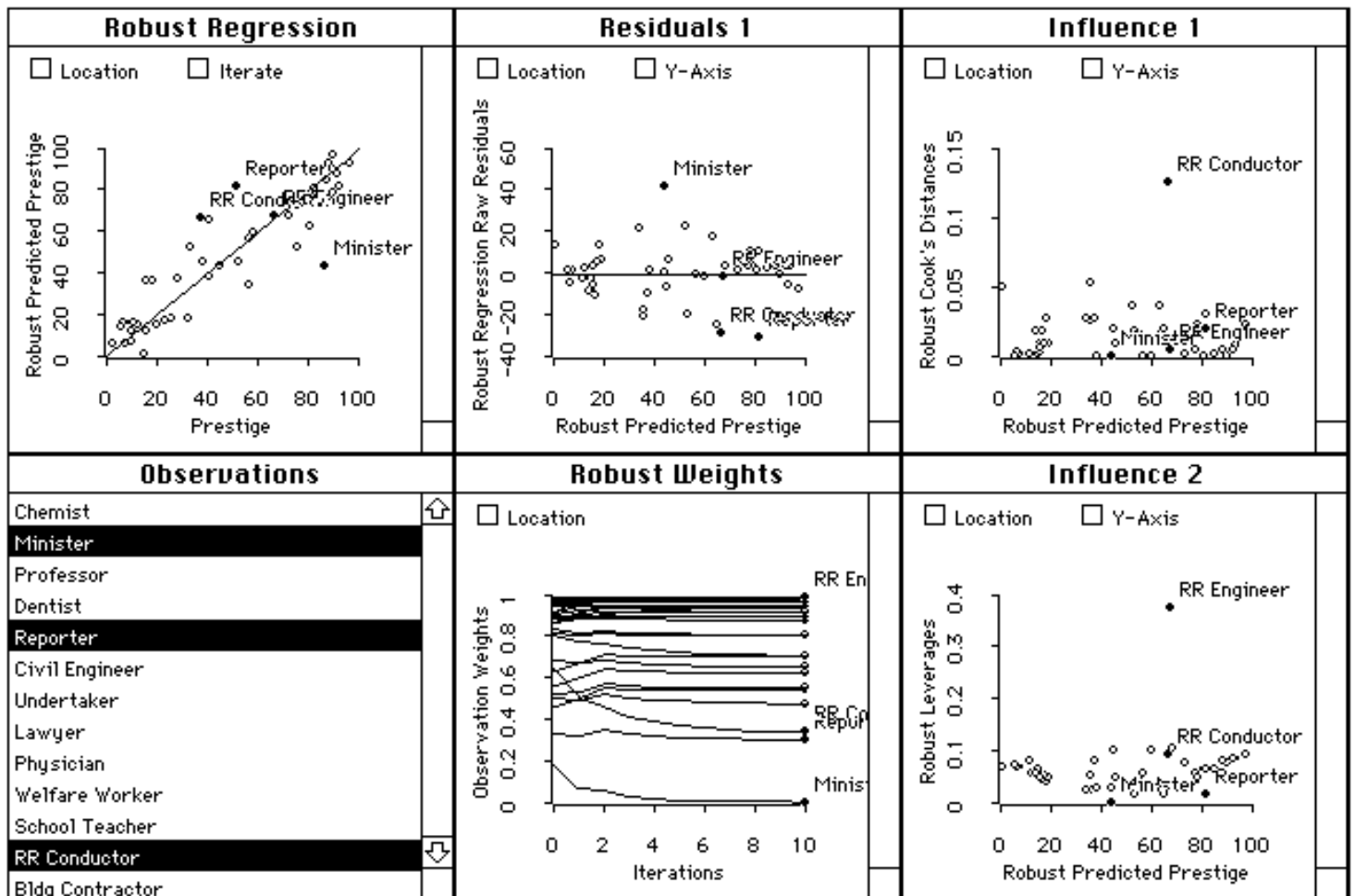
- **Multiple Regression (univariate) - Univariate regression includes simple, multiple, robust, and monotonic regression.**

Regression SpreadPlot - The regression spreadplot is comprised of a regression plot, influence plots, and residuals plots. Weight plots are also included for robust and monotonic regression.

A SpreadPlot for Analysis of Variance



A SpreadPlot for Univariate Regression

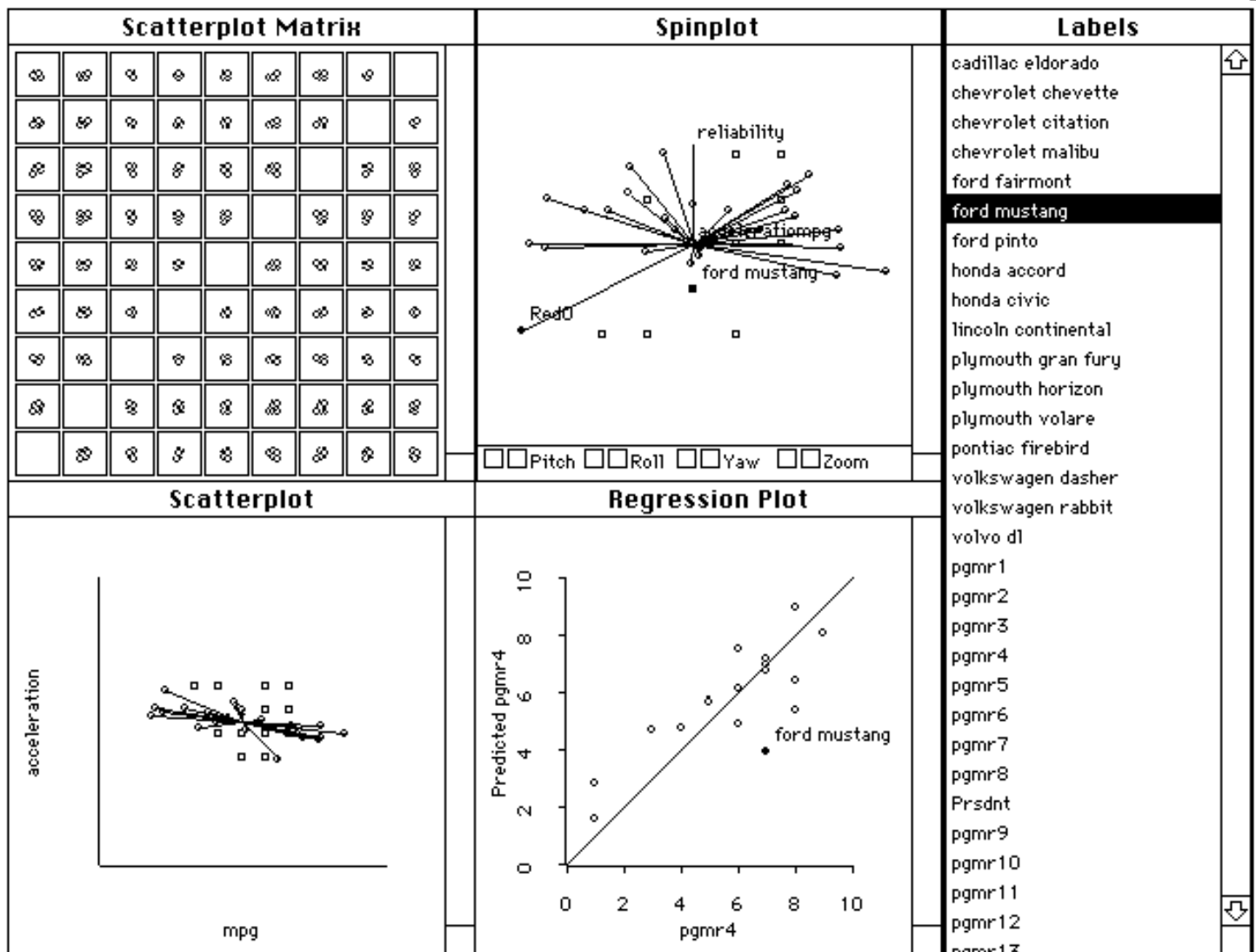


4.3 Statistical Visualization and Analysis: Multivariate

Multivariate Analyses and Visualizations

- **Multiple Regression (multivariate) - Multivariate Multiple Regression Analysis.** The spreadplot consists of a biplot, spinplot, histogram and scatterplot-matrix.
- **Principal Component Analysis of correlations or covariances.** The model visualization is a spreadplot composed of a biplot, spin-plot, scree-plot and scatterplot-matrix.
- **Multidimensional Scaling of one or more symmetric or asymmetric matrices.** The model visualization is a spreadplot composed of a scatterplot, spin-plot, scree-plot and scatterplot-matrix. The spreadplot supports graphical re-estimation of model parameters.
- **Correspondence Analysis (simple) of two-way contingency tables.** The model visualization is a spreadplot composed of a biplot, spinplot, residuals plot and scree-plot. The spreadplot supports graphical re-estimation of model parameters.

A SpreadPlot for Multivariate Regression



5: ViSta's Languages

ViDAL: ViSta's Data Analysis Language

- **ViDAL consists of functions which correspond directly to menu items. Every menu item is paralleled by an identically-named function (except spaces are replaced by dashes). The “Show Datasheet” menu item is paralleled by the (show-datasheet) function.**
- **ViDAL lets you type at the keyboard rather than use the point-and-click graphical interface.**
- **ViDAL also lets you define script files so that you can run analyses automatically without requiring user interaction.**

XLispStat: The Lisp language underlying ViSta

- **XLisp-Stat is an object-oriented environment for statistical computing and dynamic graphics that is open, extensible and freely available.**
- **XLisp-Stat is meant for high-level statistical programming and data analysis by sophisticated statistical data analysts.**
- **XLisp-Stat is based on XLisp, a freely available version of Common Lisp, a functional object-oriented computing language.**
- **ViSta is written in XLisp, using the XLisp-Stat system. ViSta is also an open and extensible system which provides all of the power of XLisp and XLisp-Stat for the programmer who wishes to extend ViSta's capabilities.**

6: Conclusions

- **ViSta is designed for data analysts ranging from novice to expert, presenting a structured graphical interface consisting of GuideMaps, WorkMaps and SpreadPlots, but including traditional Menus, Command Lines and Scripts.**

The audience for complex data analysis software continues to become wider and more naive as the price and availability of software and hardware improves.

Thus, ViSta addresses an important data analysis problem.

- **Our guiding principal and our main hypothesis is that data analyses performed in an environment that visually guides and structures analyses will be**
 - **more accurate**
 - **more accessible**
 - **more satisfying**
- **than analyses performed in a unstructured environment, especially for inexperienced data analysts.**
- **But even more importantly, it is our hypothesis that the most powerful data analysis environment will combine**
 - **the visually guided and structured environment with**
 - **traditional data analysis techniques proven useful over the years**
- **We are planning on performing usability studies to test this hypothesis.**