ABSTRACT
Recent advances in software and hardware technology have provided immediate online access to increasingly large volumes of data, either stand alone or as part of a data warehouse. A typical first step is to use data exploration and visualization techniques to illuminate trends and patterns in the data.

SAS/SPECTRAVIEW software provides a simple interactive framework, which is efficient for both technical and non-technical analysts to visualize important relationships in their data which are often hidden among multiple variables and observations. SAS/SPECTRAVIEW software provides the ability to quickly and efficiently spot key trends and relationships that may not be readily apparent in data.

This paper presents an overview of SAS/SPECTRAVIEW software and data visualization, as well as an example of how to explore a large multi-variable data set and isolate key components resulting in enhanced analysis and decision making.

INTRODUCTION
Advances in information technology are allowing organizations to collect and store larger amounts of detailed data. In order to help the organization make better business decisions this data needs to be mined and explored. As technology has advanced in the areas of collection and storage, the technology for exploring the data has also advanced. Faster and more powerful hardware with increased graphical support has allowed the area of data mining and data visualization to catch the attention of technology users worldwide.

Data mining is a process of sampling, modeling, and exploring large volumes of data in order to better understand the data. One key piece of the data mining process is the area of exploration and visualization. By utilizing advanced graphics methods the technology of data visualization enables you to color code and explore large data sets using sophisticated visualization tools and techniques. The benefit is that you can typically discover patterns, trends and outliers in the data much more quickly when using visualization tools than traditional graphics and reporting methods.

The goal of data mining and data visualization is to extract useful information from large data repositories to assist organizations in making better business decisions. To help meet the data mining and data visualization needs of the SAS user community, SAS Institute has released SAS/SPECTRAVIEW software on several new platforms with additional enhancements and functionality.

OVERVIEW OF SAS/SPECTRAVIEW
SAS/SPECTRAVIEW software is an interactive data visualization tool designed to explore and present data visually. Data is read into the software, color coded and then visualized all with a point and click mouse driven interface. Specialized visualization techniques such as point clouds, cutting planes, and isosurfaces let you see the data in new ways. Two and three-dimensional probes, charts, and graphs let you explore the data to find key relationships and patterns.

Designed to work with both small and large data sets, SAS/SPECTRAVIEW can easily handle a quarter million observations.

It is important here to distinguish the difference between SAS/SPECTRAVIEW and SAS/INSIGHT® which is another of SAS Institute's interactive data analysis tools. While SAS/INSIGHT has some color coding methods for exploring data, it tends to work better with smaller data sets, such as under 10,000 observations. Furthermore, SAS/INSIGHT contains many of the statistical tools to perform number crunching and analysis of the data, while SAS/SPECTRAVIEW focuses on visualization tools such as cutting planes, point clouds, and volume views to explore the data visually. In addition, there are no sophisticated statistical procedures or methods in SAS/SPECTRAVIEW; it is mainly a visual analysis tool. These two tools can work well together. For example, you can use SAS/SPECTRAVIEW to get a first look at the data to spot trends and relationships as well as subset the data. Then you can follow up with SAS/INSIGHT to do the statistical analysis to confirm and expand on the visual analysis done with SAS/SPECTRAVIEW.

Combining both a powerful point and click interface with sophisticated visualization techniques, SAS/SPECTRAVIEW can assist you in getting the most out of your data resources. See figure 1 for a visual of the SAS/SPECTRAVIEW interface. SAS/SPECTRAVIEW's capabilities apply well to a variety of both scientific and business application areas:

- Financial & Business Analysis
- Pharmaceutical Studies
- Oil Exploration
- Chemical Analysis
- Environmental Science
- Medical Science

Figure 1. User Interface.

DATA VISUALIZATION IN BUSINESS
As businesses move forward with technology, they are constantly looking for ways to gain a competitive advantage. Organizations are realizing the importance and potential of
tapping into and exploiting their large volumes of data. The challenge has been the methods and tools for making sense of all of the information. To assist in this process, data visualization tools like SAS/SPECTRAVIEW have been positioned to help business organizations learn more about their data.

SAS/SPECTRAVIEW was originally designed as a scientific data visualization tool. Visualization tools have been used by scientists for years to help them uncover relationships in large volumes of data. The Institute is taking the same technology and making modifications so that business data as well as scientific data can benefit from the power of visualization software. To handle the additional needs of the business community, the Institute has added capabilities to read character data, categorize data, visually probe and subset data as well as many other features. SAS/SPECTRAVIEW has been enhanced to make it easy to read and explore business data. Additionally, unlike other visualization products SAS/SPECTRAVIEW is designed to work across a broad range of hardware platforms including desktop solutions.

As businesses continue with the trend of building data warehouses and setting up sophisticated client/server networks to gain access to large data repositories, data visualization tools such as SAS/SPECTRAVIEW are well positioned to assist in turning that raw data into meaningful information.

THE DATA VISUALIZATION PROCESS
There are three main activities used in the process of data visualization with SAS/SPECTRAVIEW software: reading the data, color coding the data, and exploring and visualizing the data using specialized visualization tools and techniques.

Reading the Data
The first step in the visualization process is selecting and reading your data into the SAS/SPECTRAVIEW environment. A visual interface guides you through selecting numeric as well as character variables to load. You then have the option to categorize or perform summary statistics on the data as well as perform WHERE clause processing to subset the data. These features make it easier to load large data sets. The goal is to allow you to read in virtually any data set quickly and easily.

When loading data you may select up to five variables to investigate. Three variables represent the x (horizontal), y (vertical) and z (depth) axis of the 3D viewing area. A fourth variable represents the response variable and is color coded based on its value. In addition you can also read in an optional BY variable which will allow you to animate the data, for example, over time or geographic region.

Color Coding the Data
The next step in the process is the color coding of the data. Using a color palette, the response variable is assigned a default color scheme, which you customize based on the data. For example if the response variable is sales data, you could color the low values in red, the medium values in yellow and the high values in green. This color coding allows you to work with the data in a visual manner to spot trends and relationships not evident in standard reports and graphs.

Visualizing and Exploring the Data
Once you have read and color coded the data its time to explore and visualize the data. SAS/SPECTRAVIEW provides several interactive visualization tools to search through the data. Cutting planes are used to slice through the data, point clouds to display concentrations of data, and a 3D volume gives an overall view of the data. You can combine these tools with interactive 2D and 3D bar, contour, and surface charts as well as legends, titles and other annotation to gain additional information.

SAS/SPECTRAVIEW also provides data probes to pinpoint actual data values within the various visual displays. Furthermore, you have the ability to animate and render the data. All of these methods provide an interactive visual interface for searching and exploring data. Once some interesting results are detected you can visually subset and write out a new SAS data set and explore it further with any of the other products in the SAS System. In addition to outputting a SAS data set, you can render high quality graphic images in TIFF or PostScript format for presentation purposes.

A DATA VISUALIZATION EXAMPLE
Now that we have introduced SAS/SPECTRAVIEW and many of its features and functions, its time to give a detailed example to demonstrate how this technology can help identify important relationships in a large data set.

Financial Example:
The following data examines a financial data set which contains information on over 45,000 loans. The data set contains information detailing the type of loan, the interest rate of the loan, the loan's status, its geographic origin, as well as other relevant information.

A few simple mouse selections allow you to select the library containing the data, the data set and then the four desired variables. For the first example, we'll select the loan type (FHA, VA, Conventional and Low Income), the District (geographic area) and interest rate (from 5.5% to 13.5%) as the axis variables in SAS/SPECTRAVIEW. These are the independent variables. The response variable (dependent variable) is the original loan balance.

We'll choose to read the loan type (LOANTYPE) and district (DIST) as they exist in the data set. For interest rate (RATE), however, we will categorize this variable. We are less concerned with the exact interest rate, only that they fall in some range. This ability to categorize variables allows SAS/SPECTRAVIEW to group values for easier visual analysis. We'll use the default of ten groups (users can specify any number from 2 to 100 groups). Once categorized, multiple observations may fall in the same category, so we'll select to use the mean value as the resultant response value. You can select from eight separate statistic types depending on how you want to look at the data. After completing these steps and reading the data SAS/SPECTRAVIEW produces a data grid containing 5,600 data values from the original 45,681 observations.

The next step is to setup the color coding to most effectively match your data. A simple mouse click takes you to the palette menu. Here a data ramp shows the range of response values and allows you to define up to 32 colors which map to these values. In this example, we will select green for values below $50,000. Lower values in the range from 0-$50,000 being bright green and values closer to $50,000 being dark green. Original
loan amounts from $50,000 to $100,000 map from yellow to orange, respectively. Finally, values above $100,000 are color coded from dark to bright red. Now that we have established a logical color coding scheme to the data it's time to explore and visualize.

One of the first steps in analyzing the data is to examine all data values. The point cloud is ideal for this task. This interactive visualization technique allows you to visualize individual data points. The data points display the response variable's values as colored markers. A simple mouse click turns on a point cloud. You can use the response histogram to subset the point cloud based upon the response value. You simply grab either the lower or upper end of the histogram and drag it where you would like that limit to reside and the point cloud is subsetted between the selected range. For example, let's look at the higher loan amounts. We simply slide the lower limit up to $100,000. Now only loans greater than this value are shown in the point cloud. The result is shown in figure 2.

It is immediately evident that no such loans exist for the higher interest rates (above 9.5%). Perhaps our clients simply don't take out such high dollar loans unless the interest rates are attractive enough. Now let's do the same for the lower loans (< $50,000). The point cloud for these loans is shown in figure 3. Many more loans exist than did the higher amount loans. The loans extend over all interest rates. It is clear, however, that the large majority of loans over 9.5% are conventional loans. Only a handful of loans over this interest rate are non-conventional and no low income loans exist over this rate. Since FHA, VA and low income loans are often utilized by clients with lower incomes, perhaps they simply cannot afford the higher rate loans. SAS/SPECTRAVIEW assists you in quickly uncovering these types of relationships in your data in a very visual and intuitive manner.

Let's look at the same data set, but in a different way. We'll select loan type and interest rate again as axis variables, but loan status (Active, Delinquent, etc.) as the third axis variable. We'll use original loan balance as the response variable. Again we'll categorize the interest rates and take the mean of values that fall in the same groups. Doing this produces a data grid of 240 values from the 45,681 observations. Original loan amounts range up to just under $100,000. We'll use a similar color scheme, under $50,000 (green) and over $50,000 (red). If we use the point cloud feature we see that the majority of foreclosed and delinquent loans are red (i.e. over $50,000).

We can use another visualization technique, the cutting plane, to examine the loan status in greater detail. Again, using a point and click interface we select the variable "STATUS" and turn on a cutting plane. This holds loan status constant and shows how the loan amounts vary based on the other two independent variables, loan type and interest rate. Cutting planes are good tools for isolating the effects of independent variables. You can see that as the cutting plane moves through the data, the chart or graph updates. This technique provides additional detailed views of the data.

Let's now move the cutting plane until the status is "Foreclosed". The result is figure 4. Ten bars are shown, eight of these are red indicating original balances over $50,000. We could use the interactive data probe to examine the exact value for each bar if we desired such information. Let's also examine delinquent loans. Using the mouse we slide the cutting plane until the 3D bar chart heading reads "Delinquent". Now the chart shows 25 bars (figure 5), 17 of which are red. For the most part loans seem much more likely to run into problems if...
they exceed $50,000. This is not always true, however, as seen by the conventional loan column. This column has eight bars, six of which are green. Finally, the cutting plane shows that most of the loans are at the lower end of the interest rates.

![Delinquent Loans 3D-Bar Chart](image)

Figure 5. Delinquent Loans 3D-Bar Chart.

In a very short SAS/SPECTRAVIEW session we have discovered several key relationships in this financial data set which can assist in more effective lending practices. Here are a few of the relationships discovered in the data:

- No loans over $100,000 were made at interest rates over 9.5%
- Majority of loans less than $50,000 were made at interest rates below 9.5%
- Majority of loans over 9.5% were conventional loans
- Majority of high risk loans (defaulted, delinquent) were over $50,000

Let's now generate a report on delinquent loans based on the above information. Using the interactive annotation features we'll add a title and a legend to the 3D bar chart of delinquent loans. After saving our graph as an output file we then save the corresponding data values into a SAS data set for further investigation within the SAS System.

CONCLUSIONS

SAS/SPECTRAVIEW software provides a powerful, interactive environment for organizations to explore and visualize data. Having the ability to work with large data files in an interactive and visual manner assists both scientific and business organizations in uncovering key relationships in their data. The power of this data visualization is that it reduces the turnaround time for answering questions about an organization's data.

As technology advances in the areas of data warehousing, data mining, and data visualization, software tools like SAS/SPECTRAVIEW can provide a vehicle for organizations to turn their huge data repositories into useful information that can be used to make better business decisions.

SAS, SAS/INSIGHT and SAS/SPECTRAVIEW are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.