

Multidimensional Data Visualization Tools

Sanjay Matange, Jim Beamon, Cindy Huffman, SAS Institute Inc., Cary, NC

ABSTRACT

Huge amounts of data are piling up in Data Warehouses all over the corporate world. There are valuable trends about customers, clients, users, patients, products, etc. that are buried in the sheer volume of seemingly overwhelming data. Help is on the way! New interactive Multidimensional Data Visualization tools from SAS Institute are designed to help you extract trends and associations in your data, that are beyond the capabilities of traditional 2D tools.

This presentation will outline some of the features of the Interactive Multidimensional Visualization capabilities that will be included with the Version 7 SAS/GRAPH ® software.

INTRODUCTION

Across the world, companies have amassed a great amount of data over time. Finding valuable information in this data requires new interactive tools. To address this need SAS Institute has added new Multidimensional Data Visualization capabilities to the SAS/GRAPH ® software to help users in this task. The new objects are the ideal tools for extracting trends and associations in the data and presenting them in a visual form.

Four new objects allow users to create a visual representation of their data: Multidimensional Bar Charts, Scatter Plots, Histograms, and Pie Charts. These objects are specifically designed for interactive exploration and presentation of data. The key features of these objects are:

- They provide for Multidimensional and Multivariate Visualizations of data.
- They are interactive, so users can examine the data from various angles, including viewing detailed information about specific items directly.
- They provide many methods to subset and summarize the data so users can slice and dice through their data.

These objects allow the user to view multiple variables in a data set simultaneously against

one, two or three independent variables at a time. The data can be presented in 2D or 3D views. Trends in the data are easier to understand when the data can be viewed from different perspectives. So, these objects are designed to allow users to easily interact with their data. Users can move around and probe the data for detailed information. These objects also provide a rich set of methods to subset the data and allow users to “slice and dice” through their data and to specify exactly how they want to visualize it.

This presentation will introduce many of the features of these objects and how they can be used, individually or together, to visualize several patterns in the same data source. Some of the data used in this presentation is provided by Asiaweek magazine.

MULTIDIMENSIONAL PIE CHART

The Multidimensional Pie Chart Object allows the user to create a visualization of one to three response variables by one to three independent variables at a time. The independent variables are Slice, Row and Column. The dependent variables are Size, Height and Color.

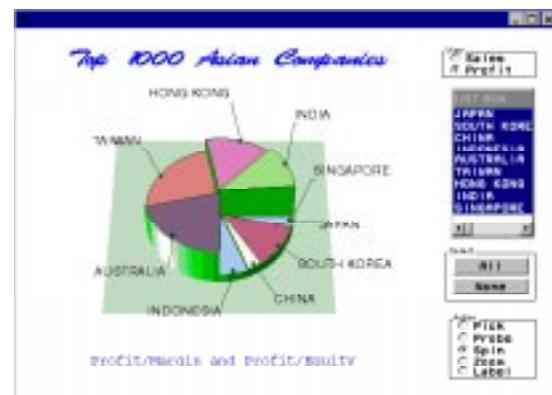


Figure 1 Multidimensional PieChart

Each category of the Slice variable is displayed as an instance of a pie slice. If Row and Column independent variables are used, a grid of pies is generated. The dependent variables Size, Height and Color control the angular size, the height and the color of the pie slice.

Labels can be displayed for each slice of the pie and are controlled by the user. Basic statistics, such as Sum, Mean, and Frequency, are also supported. The Pie Chart can be viewed in 2D, 3D, and as a Pie Grid, which displays several Pie Charts together as one object.

The Multidimensional Pie Chart in Figure 1 is an example of a 3D Pie Chart. The data can be defined through the Attributes screen or through SCL Methods. The user can define up to three dependent variables Size, Height, Color.

In Figure 1, the object is customized to display the Gross Sales and Assets, or the Profit/Margin and Profit/Equity, for the Top 1000 Asian Companies. When “Profit” is selected for the view, the dependent variables assigned are “Profit as % of Sales” to the Size variable and “Profit as % of Equity” for the Height variable. The Mean is taken for both these variables.

When “Sales” is selected for the view, the dependent variables are “Sales in Millions of Dollars” for the Size and “Profit as % of Equity” for the Height variable. The Sum of those variables is displayed.

MULTIDIMENSIONAL CHART

The Bar Chart object allows the user to create a visualization of two to six variables at a time. Any of the variables can be attached to one of three axes and one of three color ranges. Each observation in a data set is an instance of the bar or box for the Box and Bar Charts. The **BY** variable links the observations for the Line and Area Charts. The Scatter type chart allows users to attach marker shapes, size and color to different response variables.

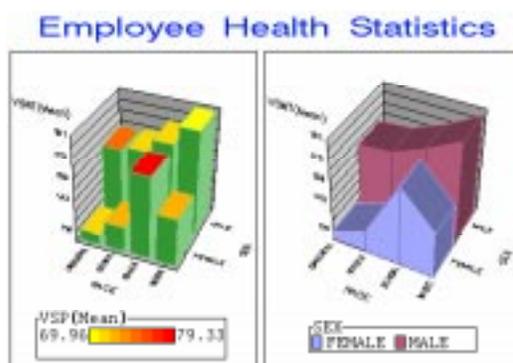


Figure 2 Multidimensional Bar Chart Object – Bar Chart and Area Chart

The Multidimensional Bar Chart object in Figure 2 contains an example of the Bar and Area Charts. In this example, both charts display the mean of the response variable Weight by Race and Sex (independent variables). The Bar Chart also shows the Mean of the Pulse Rate as the color of each bar. A legend of the color range is included at the bottom.

MULTIDIMENSIONAL HISTOGRAM

The Multidimensional Histogram Object allows the user to build a density plot of the data using one to three independent variables to create 2D or 3D visualizations. The variables can be attached to one of three axes. Histograms are ideal for large volumes of data. The data is binned into a grid of bins as defined by the user. The response value can be either the count of the number of observations falling in the bin, or the mean or sum of some other response variable.

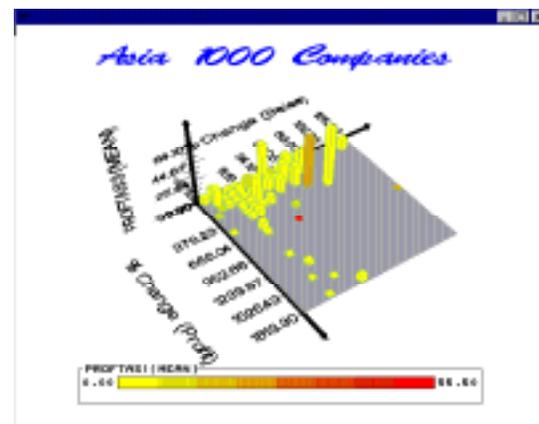


Figure 3 Multidimensional Histogram Object – 2D Data, 3D Graph

The Multidimensional Histogram in Figure 3 is an example of a 3D graph with 2D data. The variables are setup to visualize “Profit as % Equity” (Mean) against “% change in Sales”, and “% change in Profit”. “Profit as % of Sales” (Mean) is displayed by the color of the bars.

So for an aggressive investor, the companies that fall in the tall yellow bin in the right of the graph could be of interest.

MULTIDIMENSIONAL SCATTER PLOT

The Scatter object allows the visualization of individual observations of the data in 2D or 3D. The data can be plotted against one or two independent variables. A marker is produced for each observation. Independent variables can be of categorical or numeric type. The Shape, Size and Color of the marker can be attached to other response variables.



Figure 4 3D Scatter Plot

BUILDING AN APPLICATION

An important feature of the Multidimensional Data Visualization Objects is how they work together. The next part of this presentation will demonstrate how the objects can be used to visualize a large data set, then effectively “drill down” to specific information about their relationships.

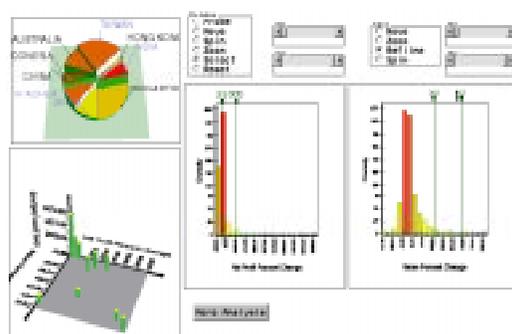


Figure 5 Application with Multidimensional Data Visualization Objects

The application (Figure 5) begins with a Pie Chart, displaying the top 1000 Asian Companies by country. The size, height, and color of the pie

slices are the “Profit per Employee”, “Sales Profit % Change”, and “Equity Profit % Change”, respectively. Two 1D Histograms, one displaying “Net Profit % Change” and the other “Sales % Change”, are driven by the selections from the PieChart. A 3D Histogram displays “Net Profit % Change” and “Sales % Change” and is updated by the reference lines from the 1D Histograms.

An investor could use this application to find specific companies for investment possibilities. Starting with the Pie Chart, the investor would select countries of interest – the countries Taiwan, India, Japan, and South Korea are selected in the pie. The countries, based on their size, height, and color, appear to have a fair amount of companies that have high Profit per Employee, Net Profit % Change, and Sales % Profit.

The 1D Histograms are updated to display the Net Profit % Change and Sales % Change for the countries selected. The investor now wants to see only companies with net profit increase between approximately 30% and 330%. The 3D Histogram (bottom left graph) is updated as the reference lines are dragged on the 1D Histogram.

The Sales % Change is also subset to reflect a Sales % Change from about 30% to 60%. The 3D Histogram is now showing all companies, from the selected countries, whose Net % Profit Change is between 30% - 330% and Sale % Change is from 30% - 60%.

Selecting the “More Analysis” button brings the investor to another “page” (Figure 6) with the 3D Histogram. Along with the 3D Histogram, a Scatter Plot and Table are displayed. The Bar Chart displays specific company information from the data set.

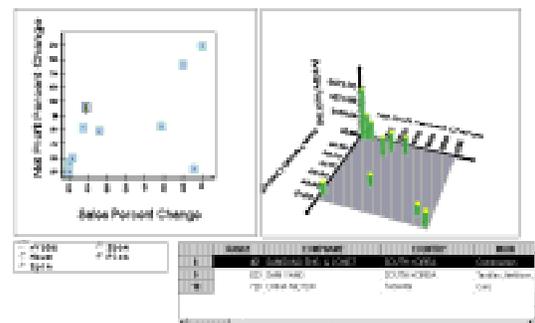


Figure 6 Application with Multidimensional Objects

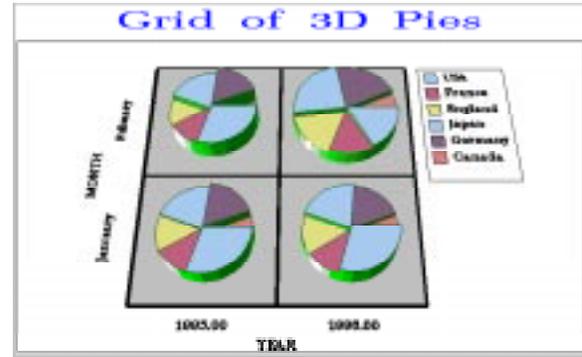
From the 3D Histogram, the investor can see the Net Profit % Change and Sales % Change. In addition, the Z-axis includes Mean Sales in Millions of Dollars. Now the investor can specifically look at companies with large sales, within the profit and sales percent range that they are most interested in investing.

When the investor makes their selection on the 3D Histogram, the Scatter Plot is updated and displays all the companies that fall within the selected bin. Using the table, the investor can read more information about the type of companies that fall into the selected category and find out more detailed data such as profit, sales, etc. The investor can also select a company from the Scatter Plot, which will highlight the specific company in the Table, for easier look-up.

Again, this is one example of how effective the Multidimensional Data Visualization Objects are with managing a large data set, finding relationships in the data, and breaking it down to specific details to meet a particular need.

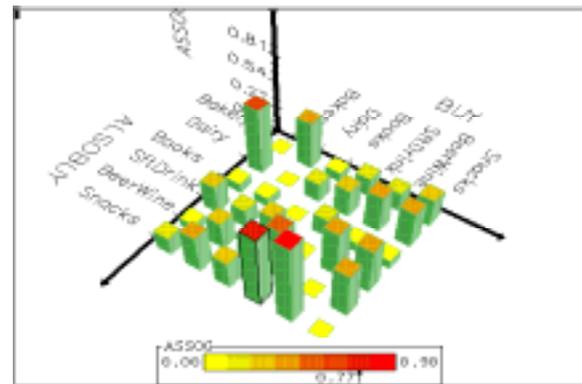
EXAMPLES

There are a variety of chart types and SCL methods for each of the Multidimensional Data Visualization Objects, which are available to meet the needs of the user.



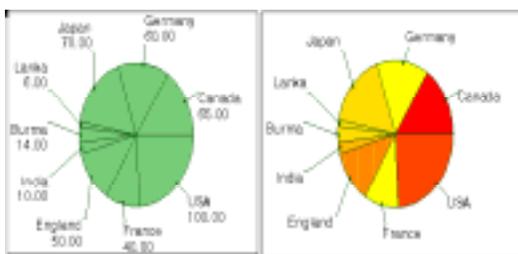
Pie Grid

Bar Charts

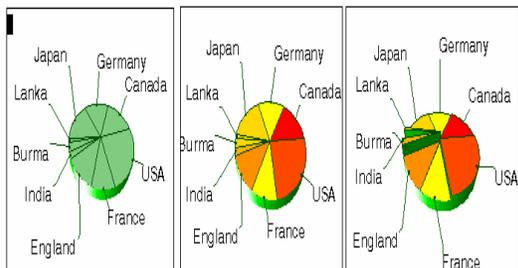


Bar Chart

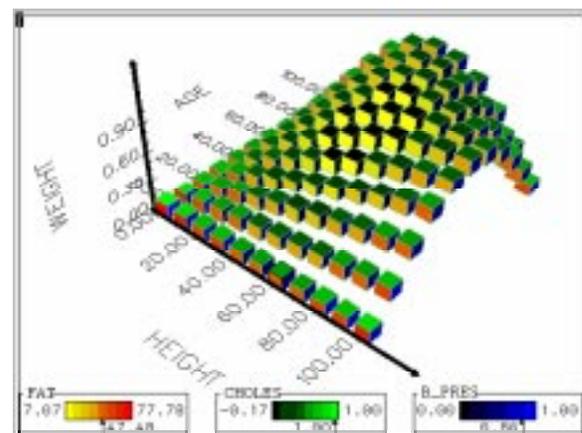
Pie Charts



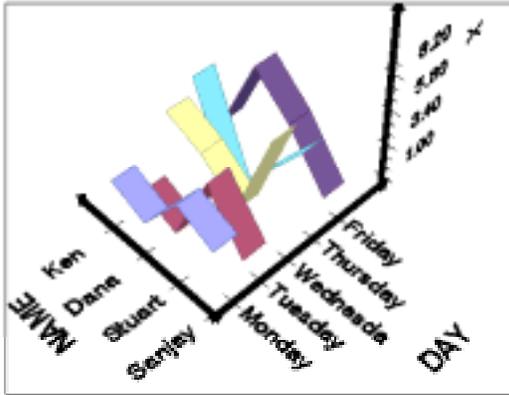
2D Pie Chart



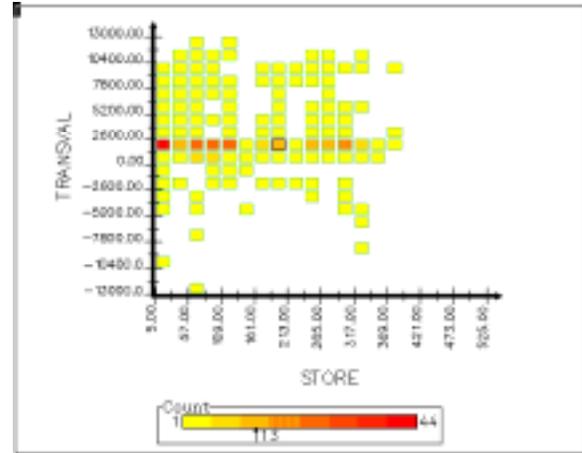
3D Pie Chart



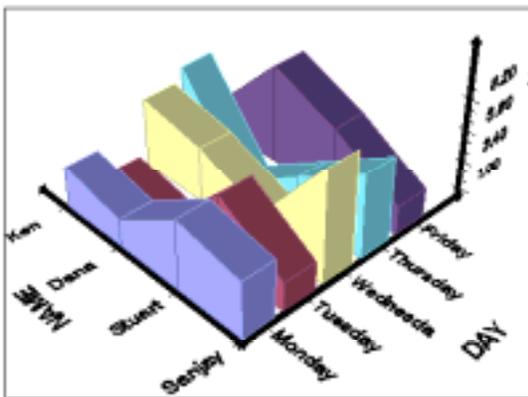
Box Chart



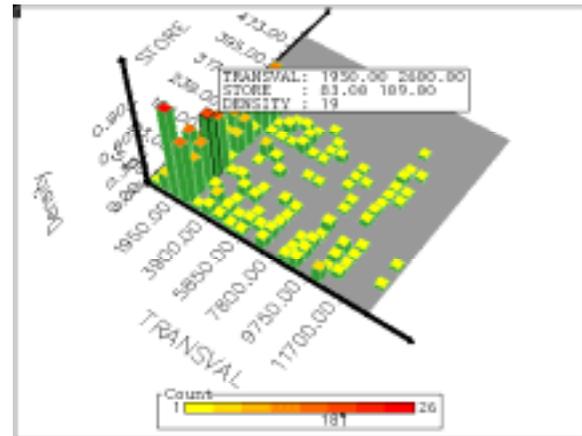
Line Chart



2D Data with 2D Histogram

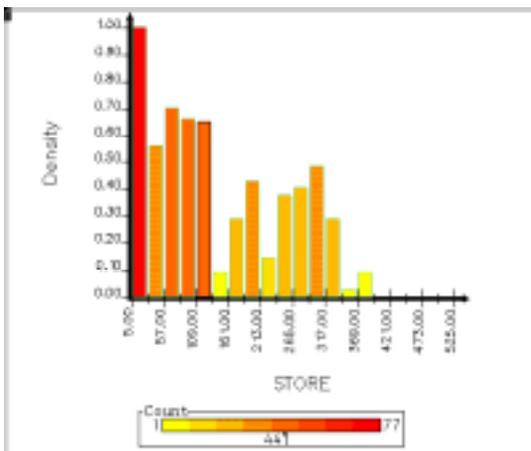


Area Chart

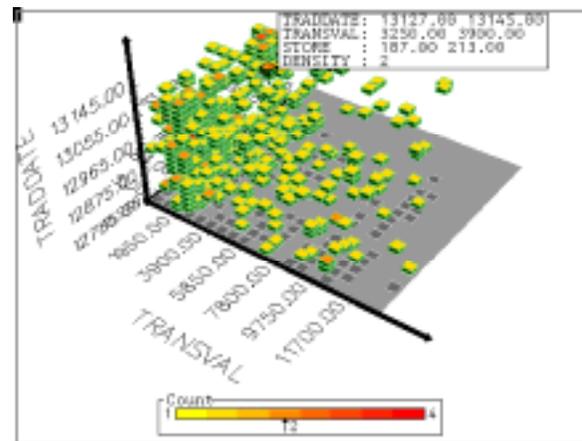


2D Data with 3D Histogram

Histogram Charts

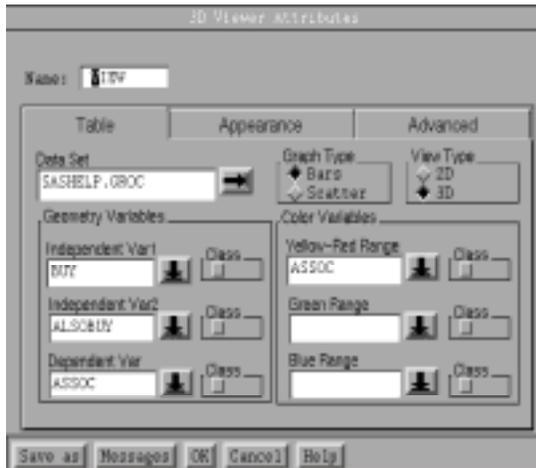


1D Data with 2D Histogram



3D Data with 3D Histogram

Sample V6 Attributes Screen



Common V6 SCL Methods

_SET_DSNAME_	_GET_DSNAME_
_SET_VALUE_	_GET_VALUE_
_SET_INDEP_VAR_	_GET_INDEP_VAR_
_SET_DEP_VAR_	_GET_DEP_VAR_
_SET_GRID_	_SET_COLOR_RANGE_
_HIDE_LEGEND_	_SHOW_LEGEND_
_SAVE_AS_	_SET_LOC_
_SET_TILT_	_GET_TILT_
_SET_ROTATE_	_GET_ROTATE_
_SET_SCALE_	_GET_SCALE_
_SET_GRAPH_TYPE_	_GET_GRAPH_TYPE_
_SET_TITLE_	_GET_TITLE_
_SET_LABEL_	_GET_LABEL_
_SHOW_LABELS_	_HIDE_LABELS_
_SET_COLOR_	_GET_COLOR_
_SET_COLORS_	_GET_COLORS_
_SET_FONT_	_GET_FONT_
_SET_MODE_	_GET_MODE_
_SET_STYLE_	_SET_COLOR_
_SET_ATTRIBUTES_	_GET_ATTRIBUTES_
_SET_STATS_	_GET_STATS_
_SET_SELECTED_VARS_	_GET_SELECTED_VARS_
_SET_SELECT_	_GET_SELECT_
_SELECT_ALL_	_SELECT_
_UNSELECT_ALL_	_UNSELECT_
_WRITE_FILEPATH_	_DO_
_WRITE_DIRPATH_MEMBER_	_WRITE_FILEREF_
PRINT	_WRITE_DIRREF_MEMBER_
HIGHLIGHT	

CONCLUSION

Multidimensional Data Visualization Objects are effective tools that are capable of helping users make sense of their data. These objects provide the user with an efficient, versatile and intelligent means to get a handle on large volumes of multivariate data. The objects provide high levels of interactivity, allowing the user to visualize and explore data in real time. These objects also provide a variety of methods to “slice and dice” through the data to glean the valuable trends hidden in the data.

The Multidimensional Data Visualization Objects are used in SAS Institutes’ Enterprise Miner Product and will be shipped with SAS/EIS and SAS/GRAPH software in Version 7.

ACKNOWLEDGMENTS

Lee Richardson, Director of Asia/Pacific operations, for contributing the data from Asiaweek Magazine.

SAS, SAS/EIS, SAS/GRAPH and SAS/AF are registered trademarks or trademarks of SAS Institute, Inc., Cary, NC, USA.

Multidimensional Data Visualization Tools

Sanjay Matange, Jim Beamon, Cindy Huffman, SAS Institute Inc., Cary, NC

ABSTRACT

Huge amounts of data are piling up in Data Warehouses all over the corporate world. There are valuable trends about customers, clients, users, patients, products, etc. that are buried in the sheer volume of seemingly overwhelming data. Help is on the way! New interactive Multidimensional Data Visualization tools from SAS Institute are designed to help you extract trends and associations in your data, that are beyond the capabilities of traditional 2D tools.

This presentation will outline some of the features of the Interactive Multidimensional Visualization capabilities that will be included with the Version 7 SAS/GRAPH® software.

INTRODUCTION

Across the world, companies have amassed a great amount of data over time. Finding valuable information in this data requires new interactive tools. To address this need SAS Institute has added new Multidimensional Data Visualization capabilities to the SAS/GRAPH® software to help users in this task. The new objects are the ideal tools for extracting trends and associations in the data and presenting them in a visual form.

Four new objects allow users to create a visual representation of their data: Multidimensional Bar Charts, Scatter Plots, Histograms, and Pie Charts. These objects are specifically designed for interactive exploration and presentation of data. The key features of these objects are:

- They provide for Multidimensional and Multivariate Visualizations of data.
- They are interactive, so users can examine the data from various angles, including viewing detailed information about specific items directly.
- They provide many methods to subset and summarize the data so users can slice and dice through their data.

These objects allow the user to view multiple variables in a data set simultaneously against

one, two or three independent variables at a time. The data can be presented in 2D or 3D views. Trends in the data are easier to understand when the data can be viewed from different perspectives. So, these objects are designed to allow users to easily interact with their data. Users can move around and probe the data for detailed information. These objects also provide a rich set of methods to subset the data and allow users to “slice and dice” through their data and to specify exactly how they want to visualize it.

This presentation will introduce many of the features of these objects and how they can be used, individually or together, to visualize several patterns in the same data source. Some of the data used in this presentation is provided by Asiaweek magazine.

MULTIDIMENSIONAL PIE CHART

The Multidimensional Pie Chart Object allows the user to create a visualization of one to three response variables by one to three independent variables at a time. The independent variables are Slice, Row and Column. The dependent variables are Size, Height and Color.

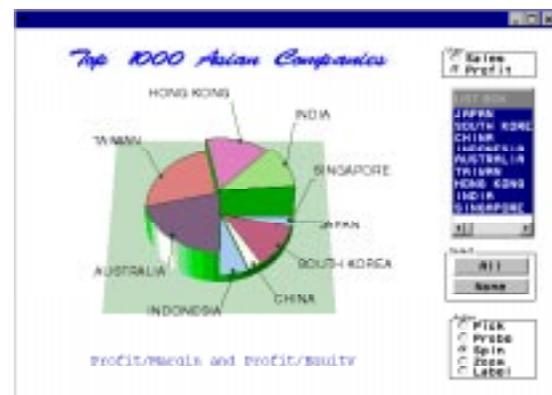


Figure 1 Multidimensional PieChart

Each category of the Slice variable is displayed as an instance of a pie slice. If Row and Column independent variables are used, a grid of pies is generated. The dependent variables Size, Height and Color control the angular size, the height and the color of the pie slice.

Labels can be displayed for each slice of the pie and are controlled by the user. Basic statistics, such as Sum, Mean, and Frequency, are also supported. The Pie Chart can be viewed in 2D, 3D, and as a Pie Grid, which displays several Pie Charts together as one object.

The Multidimensional Pie Chart in Figure 1 is an example of a 3D Pie Chart. The data can be defined through the Attributes screen or through SCL Methods. The user can define up to three dependent variables Size, Height, Color.

In Figure 1, the object is customized to display the Gross Sales and Assets, or the Profit/Margin and Profit/Equity, for the Top 1000 Asian Companies. When “Profit” is selected for the view, the dependent variables assigned are “Profit as % of Sales” to the Size variable and “Profit as % of Equity” for the Height variable. The Mean is taken for both these variables.

When “Sales” is selected for the view, the dependent variables are “Sales in Millions of Dollars” for the Size and “Profit as % of Equity” for the Height variable. The Sum of those variables is displayed.

MULTIDIMENSIONAL CHART

The Bar Chart object allows the user to create a visualization of two to six variables at a time. Any of the variables can be attached to one of three axes and one of three color ranges. Each observation in a data set is an instance of the bar or box for the Box and Bar Charts. The **BY** variable links the observations for the Line and Area Charts. The Scatter type chart allows users to attach marker shapes, size and color to different response variables.

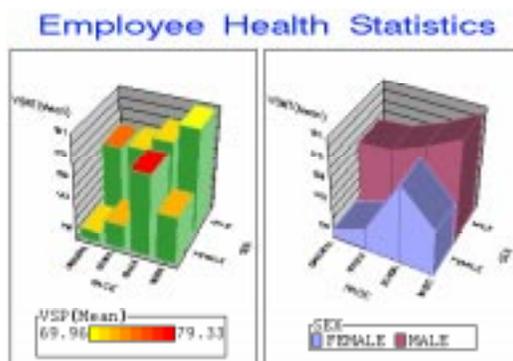


Figure 2 Multidimensional Bar Chart Object – Bar Chart and Area Chart

The Multidimensional Bar Chart object in Figure 2 contains an example of the Bar and Area Charts. In this example, both charts display the mean of the response variable Weight by Race and Sex (independent variables). The Bar Chart also shows the Mean of the Pulse Rate as the color of each bar. A legend of the color range is included at the bottom.

MULTIDIMENSIONAL HISTOGRAM

The Multidimensional Histogram Object allows the user to build a density plot of the data using one to three independent variables to create 2D or 3D visualizations. The variables can be attached to one of three axes. Histograms are ideal for large volumes of data. The data is binned into a grid of bins as defined by the user. The response value can be either the count of the number of observations falling in the bin, or the mean or sum of some other response variable.

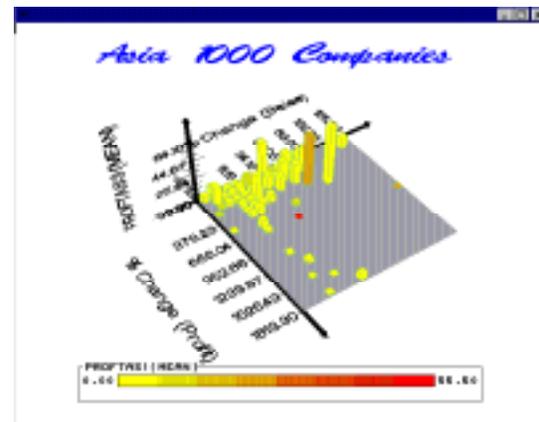


Figure 3 Multidimensional Histogram Object – 2D Data, 3D Graph

The Multidimensional Histogram in Figure 3 is an example of a 3D graph with 2D data. The variables are setup to visualize “Profit as % Equity” (Mean) against “% change in Sales”, and “% change in Profit”. “Profit as % of Sales” (Mean) is displayed by the color of the bars.

So for an aggressive investor, the companies that fall in the tall yellow bin in the right of the graph could be of interest.

MULTIDIMENSIONAL SCATTER PLOT

The Scatter object allows the visualization of individual observations of the data in 2D or 3D. The data can be plotted against one or two independent variables. A marker is produced for each observation. Independent variables can be of categorical or numeric type. The Shape, Size and Color of the marker can be attached to other response variables.



Figure 4 3D Scatter Plot

BUILDING AN APPLICATION

An important feature of the Multidimensional Data Visualization Objects is how they work together. The next part of this presentation will demonstrate how the objects can be used to visualize a large data set, then effectively “drill down” to specific information about their relationships.

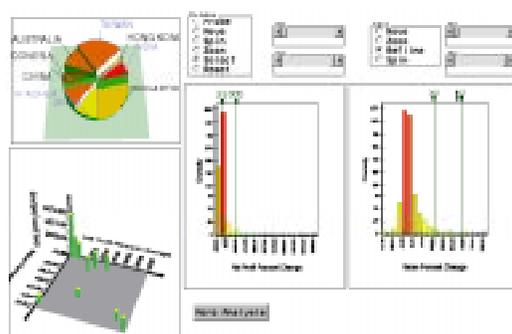


Figure 5 Application with Multidimensional Data Visualization Objects

The application (Figure 5) begins with a Pie Chart, displaying the top 1000 Asian Companies by country. The size, height, and color of the pie

slices are the “Profit per Employee”, “Sales Profit % Change”, and “Equity Profit % Change”, respectively. Two 1D Histograms, one displaying “Net Profit % Change” and the other “Sales % Change”, are driven by the selections from the PieChart. A 3D Histogram displays “Net Profit % Change” and “Sales % Change” and is updated by the reference lines from the 1D Histograms.

An investor could use this application to find specific companies for investment possibilities. Starting with the Pie Chart, the investor would select countries of interest – the countries Taiwan, India, Japan, and South Korea are selected in the pie. The countries, based on their size, height, and color, appear to have a fair amount of companies that have high Profit per Employee, Net Profit % Change, and Sales % Profit.

The 1D Histograms are updated to display the Net Profit % Change and Sales % Change for the countries selected. The investor now wants to see only companies with net profit increase between approximately 30% and 330%. The 3D Histogram (bottom left graph) is updated as the reference lines are dragged on the 1D Histogram.

The Sales % Change is also subset to reflect a Sales % Change from about 30% to 60%. The 3D Histogram is now showing all companies, from the selected countries, whose Net % Profit Change is between 30% - 330% and Sale % Change is from 30% - 60%.

Selecting the “More Analysis” button brings the investor to another “page” (Figure 6) with the 3D Histogram. Along with the 3D Histogram, a Scatter Plot and Table are displayed. The Bar Chart displays specific company information from the data set.

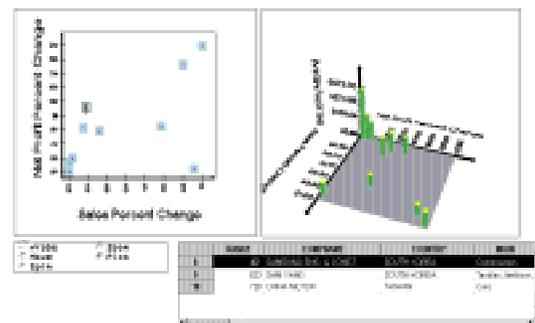


Figure 6 Application with Multidimensional Objects

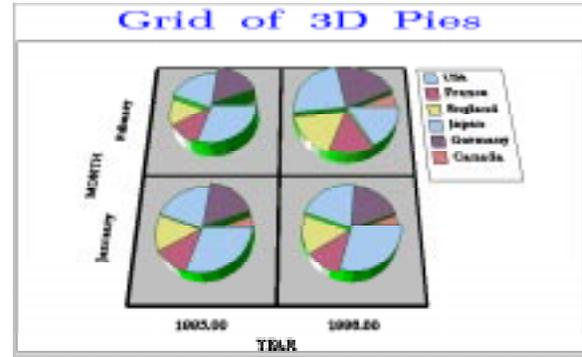
From the 3D Histogram, the investor can see the Net Profit % Change and Sales % Change. In addition, the Z-axis includes Mean Sales in Millions of Dollars. Now the investor can specifically look at companies with large sales, within the profit and sales percent range that they are most interested in investing.

When the investor makes their selection on the 3D Histogram, the Scatter Plot is updated and displays all the companies that fall within the selected bin. Using the table, the investor can read more information about the type of companies that fall into the selected category and find out more detailed data such as profit, sales, etc. The investor can also select a company from the Scatter Plot, which will highlight the specific company in the Table, for easier look-up.

Again, this is one example of how effective the Multidimensional Data Visualization Objects are with managing a large data set, finding relationships in the data, and breaking it down to specific details to meet a particular need.

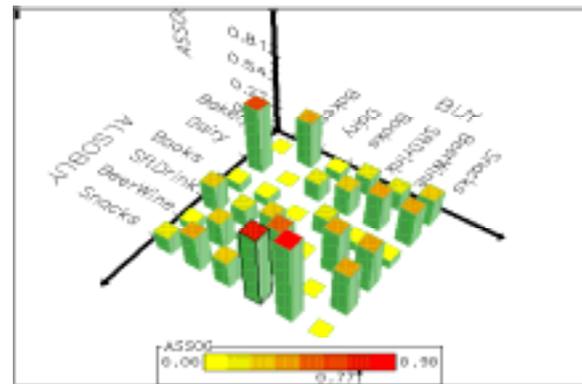
EXAMPLES

There are a variety of chart types and SCL methods for each of the Multidimensional Data Visualization Objects, which are available to meet the needs of the user.



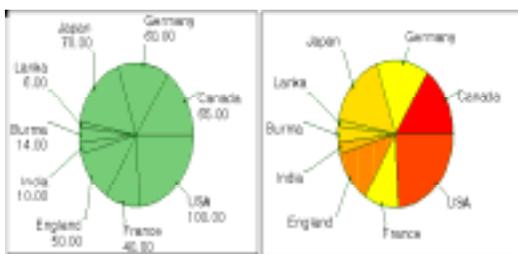
Pie Grid

Bar Charts

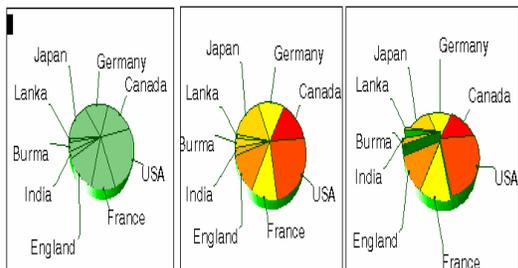


Bar Chart

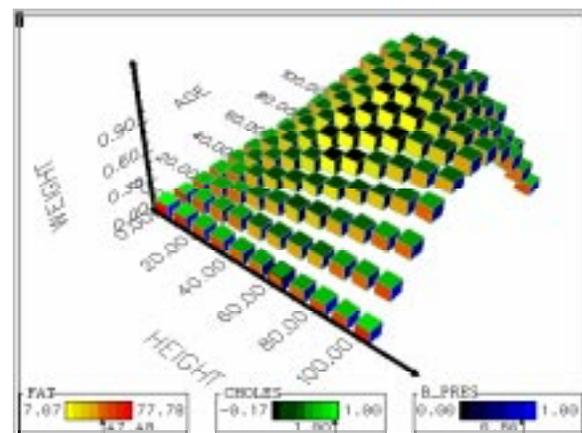
Pie Charts



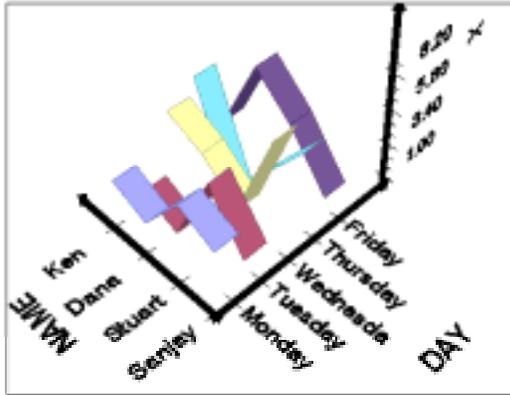
2D Pie Chart



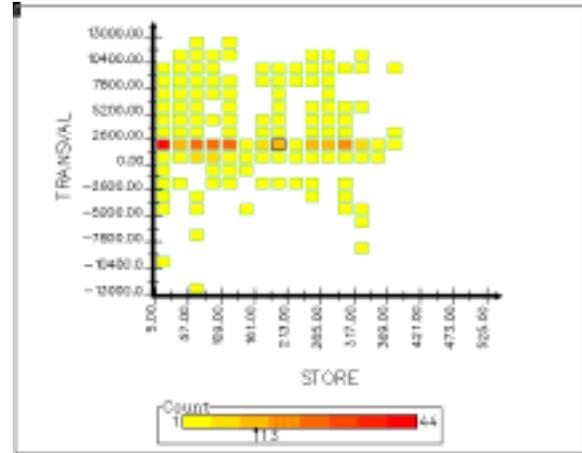
3D Pie Chart



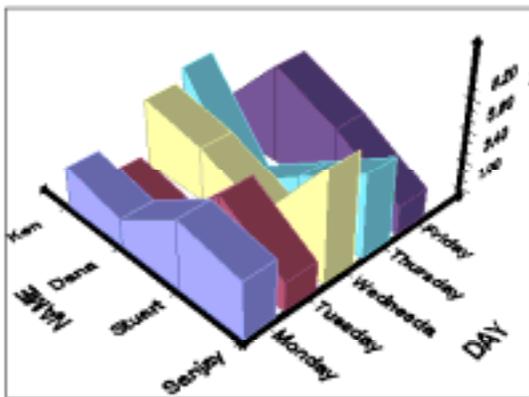
Box Chart



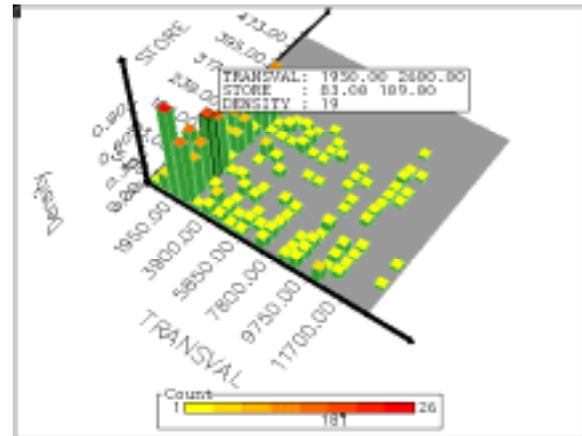
Line Chart



2D Data with 2D Histogram

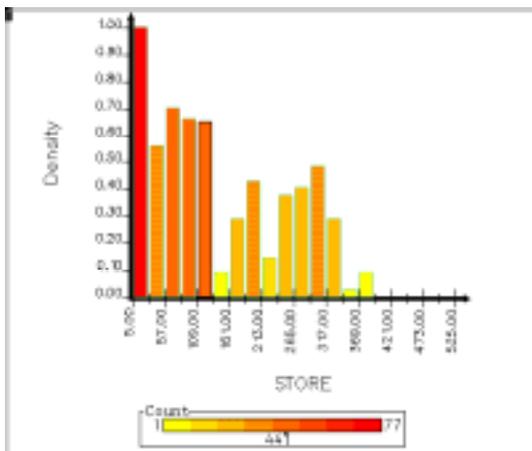


Area Chart

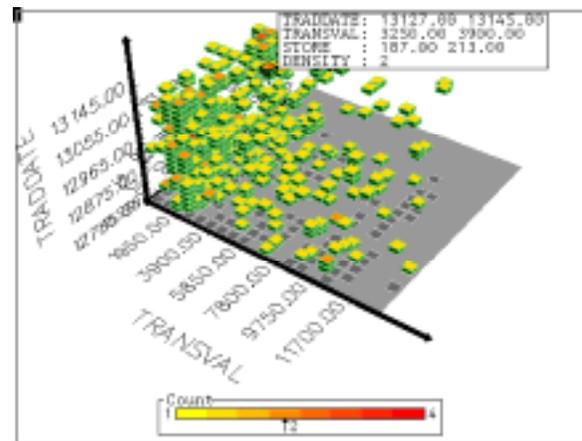


2D Data with 3D Histogram

Histogram Charts

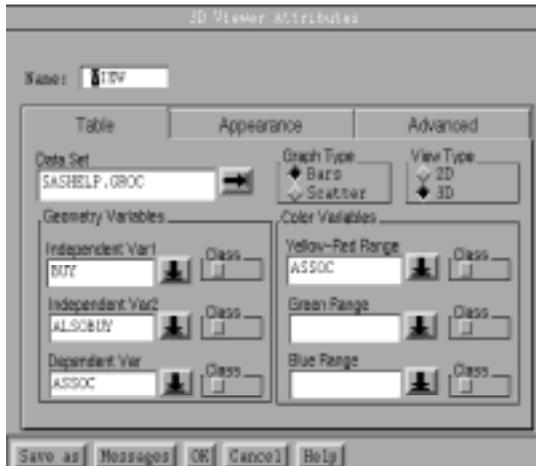


1D Data with 2D Histogram



3D Data with 3D Histogram

Sample V6 Attributes Screen



Common V6 SCL Methods

_SET_DSNAME_	_GET_DSNAME_
_SET_VALUE_	_GET_VALUE_
_SET_INDEP_VAR_	_GET_INDEP_VAR_
_SET_DEP_VAR_	_GET_DEP_VAR_
_SET_GRID_	_SET_COLOR_RANGE_
_HIDE_LEGEND_	_SHOW_LEGEND_
_SAVE_AS_	_SET_LOC_
_SET_TILT_	_GET_TILT_
_SET_ROTATE_	_GET_ROTATE_
_SET_SCALE_	_GET_SCALE_
_SET_GRAPH_TYPE_	_GET_GRAPH_TYPE_
_SET_TITLE_	_GET_TITLE_
_SET_LABEL_	_GET_LABEL_
_SHOW_LABELS_	_HIDE_LABELS_
_SET_COLOR_	_GET_COLOR_
_SET_COLORS_	_GET_COLORS_
_SET_FONT_	_GET_FONT_
_SET_MODE_	_GET_MODE_
_SET_STYLE_	_SET_COLOR_
_SET_ATTRIBUTES_	_GET_ATTRIBUTES_
_SET_STATS_	_GET_STATS_
_SET_SELECTED_VARS_	_GET_SELECTED_VARS_
_SET_SELECT_	_GET_SELECT_
_SELECT_ALL_	_SELECT_
_UNSELECT_ALL_	_UNSELECT_
_WRITE_FILEPATH_	_DO_
_WRITE_DIRPATH_MEMBER_	_WRITE_FILEREF_
PRINT	_WRITE_DIRREF_MEMBER_
HIGHLIGHT	

CONCLUSION

Multidimensional Data Visualization Objects are effective tools that are capable of helping users make sense of their data. These objects provide the user with an efficient, versatile and intelligent means to get a handle on large volumes of multivariate data. The objects provide high levels of interactivity, allowing the user to visualize and explore data in real time. These objects also provide a variety of methods to “slice and dice” through the data to glean the valuable trends hidden in the data.

The Multidimensional Data Visualization Objects are used in SAS Institutes’ Enterprise Miner Product and will be shipped with SAS/EIS and SAS/GRAPH software in Version 7.

ACKNOWLEDGMENTS

Lee Richardson, Director of Asia/Pacific operations, for contributing the data from Asiaweek Magazine.

SAS, SAS/EIS, SAS/GRAPH and SAS/AF are registered trademarks or trademarks of SAS Institute, Inc., Cary, NC, USA.

