§sas
THE POWER TO KNOW.

Analytics 101:
New Visual and Wizard-Driven Paradigms for Exploring Data and Developing Analytic Workflows

Insights from a webinar in the SAS Applying Business Analytics series
Originally broadcast in November 2010

Featuring:

Anne Milley, Senior Director, Analytic Strategy,
Worldwide Marketing, SAS

Laura Ryan, Systems Engineer, SAS

§sas

Opportunities and risks abound. Successfully navigating a more complex and dynamic world calls for better, faster decisions. However, many organizations still rely heavily on spreadsheets for guidance, even when high-risk/high-reward decisions are at stake. Costly mistakes have sometimes resulted.

Analytics goes far beyond where spreadsheets leave off. For one, analytic methods can be applied to vast data repositories, yielding deeper insights with less time and effort. Moreover, analytics brings greater sophistication to data investigations, revealing insights that spreadsheets would have missed.

However, analytic expertise is in short supply in most organizations. Quantitative specialists carry a heavy workload and need to focus on the most critical business issues. As a result, other decisions that may be important or influential to the organization – but not deemed critical – are made based on suboptimal information.

What if business analysts throughout the organization could create their own analytical processes within a familiar spreadsheet environment? What if they could interact with the data in a visual and intuitive way to quickly identify important themes? What if quantitative specialists could expedite some analytical processes and offload others – spreading analytic capabilities further across the organization while actually increasing governance?

In short, what if the power and sophistication of analytics could be made simpler, faster, and more intuitive and repeatable than ever? What would that do for the quality and timeliness of your organization's decision making?

That was the topic of a SAS webinar in the Applying Business Analytics series. Anne Milley described new capabilities for visual data exploration, visual programming of structured and unstructured data, and wizard-driven shortcuts for rapid model development. Laura Ryan provided live demonstrations of these capabilities, using SAS® Enterprise Miner™, SAS Rapid Predictive Modeler and SAS Add-In for Microsoft Office.

"By leveraging data and analytics as strategic assets, organizations can make better decisions with greater confidence. No matter what kinds of decisions or challenges you face, analytics can greatly enhance your efficiency and effectiveness, helping you create more value."

Anne Milley, Senior Director, Analytic Strategy, Worldwide Marketing, SAS

Top Five Big Ideas in Analytics

1. Spreadsheets do not offer enough statistical power, scalability or governance for organizations that truly want to use information as a strategic asset. Although many organizations rely heavily on spreadsheets, the current business environment calls for an evolution to analytics-driven decision making.

2. Analytics is about more than building and running models; it is a closed-loop process of data exploration and discovery; model creation and validation; then getting the results to the right people at the right time and learning from the results to further refine the process.

3. Quantitative expertise – the ability to write code and manage this analytical process – is in short supply, relative to the number of decisions that must be made. Organizations need ways to empower more people to participate in this process, even if they aren't data mining or analytical types.

4. Visual discovery, visual programming and wizard-driven analytics enable business analysts to conduct their own data exploration, define their own analytic workflows and create their own models.

5. Packaging sophisticated analytic processes into easy-to-use applications improves upon traditional coding, empowers more people in the organization to generate analytic insights, frees quantitative specialists to focus on the most complex and critical questions, and improves information governance.

Analytics Is More than Models; It's a Process.

To some people, "analytics" means building a model that will reveal new knowledge from data. That's too narrow a definition, said Milley. "Analytics is about solving problems and taking action as part of that. If you built a model to identify which customers are likely to default on a loan, but no one knows who they are and what kind of actions to take as a result, you haven't fully addressed the issue."

SAS conceptualizes analytics as a closed-loop process of continuous learning and improvement, where each phase is instrumental to the cycle:

- The *explore or discovery phase* of analysis helps you understand what questions to ask and what to do when you come across unexpected trends, relationships and anomalies in the data.

- In the *test and learn phase*, models are created, tested, compared – and a champion model chosen to use in scoring data.

- In the *inform and act phase*, analytic insights are pushed out to the organization to drive decisions and action. Learning is plowed back into the business to improve the process and address similar challenges.
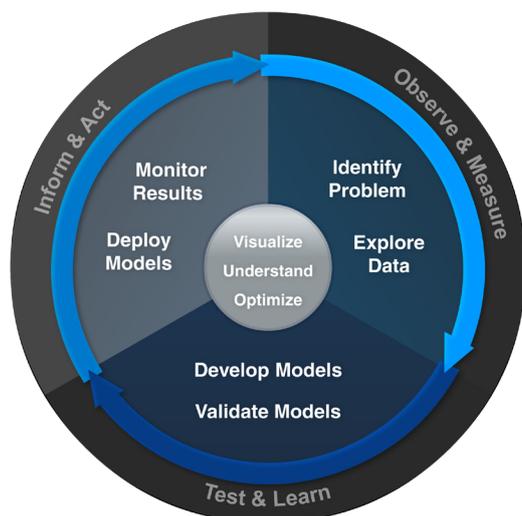
Figure 1. More than just model-building, analytics is a continuously improving, multistage process.

Analytics encompasses many disciplines, such as statistics, data mining, artificial intelligence, behavioral economics, operations research, optimization, mathematics, decision theory, neural computing, econometrics, information theory, measurement theory and others – all applied to problem solving.

"These disciplines borrow from each other and give back new methods based on changing needs," said Milley. "One example of this cross-pollination is discrete choice models, which originated in econometrics or economics and have found use by market researchers for other areas and applications. There are many such examples, but the point is that a diversity of perspectives fuels innovation. Multiple analytic disciplines are required to solve the wide variety of problems an organization faces, while enabling the greatest analytic bandwidth."

So how do you make the most of that analytic bandwidth – the finite capacity of your modeling experts to create, test and run models? In the past, that process revolved around coding – writing models in line-by-line programming logic – which could only be done by specialists. SAS Code Analyzer helps optimize SAS code to make jobs run faster, but until recently, models were still created by SAS programmers to deliver information to be used by others in the organization.

New technologies enable business analysts to be more self-sufficient in exploring data, asking questions, developing workflows and even generating their own analyses. Milley and Ryan presented three of these capabilities, including:

- **Visual discovery**, which makes it point-and-click easy to explore data in custom, graphical views.

- **Visual programming**, which makes it drag-and-drop, fill-in-the-blank easy to define analytic workflows.

- **Wizard-driven programming**, which makes it step-by-step easy for users to run their own analyses.

Visual Discovery – *Graphical and intuitive data exploration*

"The discovery phase of analytics enables you to quickly see what matters, what unexpected patterns and relationships you may find, what additional questions to ask that you hadn't anticipated, what shortcuts may be revealed, and how to best formulate and solve problems," said Milley.

Here is where a picture really is worth a thousand data points – or several million. "A highly visual data exploration environment really complements the way our brains work," said Milley. "We're faster at recognizing patterns, trends and anomalies in dynamic visual displays of information versus in tabular output or even static graphs. What was invisible becomes visible, very quickly."

Even in discovery mode though, some level of governance is necessary if organizations want to treat analytics as a strategic asset. Through automatically created scripts and journals, analyses can be repeated, reproduced and shared with others. This self-documenting capability fosters collaboration and sharing of best practices and results.

A Guided Tour of Visual Discovery with SAS® Enterprise Miner™

Laura Ryan demonstrated SAS visual discovery capabilities using an example of an insurance provider that wants to ensure that its patients are taking their medications as prescribed. The concern is that patients who drop their medications will ultimately incur higher costs for follow-on medical appointments, emergency room visits and more prescriptions. Ryan demonstrated visual data discovery processes that would help determine which patients are likely to discontinue their medications.

The demo shows how point-and-click easy it is to explore the data. Select "Distribution" from the Data Discovery pull-down menu to select variables to display, such as patient's region, age, plan type, gender, generic or brand drug, number of statin prescriptions and total prescriptions over a period of time. Click "Okay" to graph the distribution for each selected variable.

Dynamic linking enables you to interact with the graph, in this case, selecting patients who discontinued their medication, to show the distribution of those patients across the other variables. The sample data set showed that these patients tend to be in regions 2 and 3, under 80 years old, using plan type 2, and with few statin prescriptions or total prescriptions.

Because the JMP® interface in SAS holds data in memory, it is instantaneously responsive, helping you notice what is important.

An easy **graph-builder utility** enables you to display categorical and continuous variables in a variety of ways. Just drag and drop a variable into the X or Y axis and instantly see a graph. In the demo, this graphing capability gives an at-a-glance view of how many people did or did not discontinue their medications.

■ You can see the step-by-step demonstration of SAS visual discovery, visual programming and wizard-driven modeling solutions in the Analytics 101 webinar. The webinar is available to watch on demand at **www.sas.com/reg/web/corp/950738**.

You can add another variable, such as number of statin prescriptions, to see how it relates to discontinuation. Because the data is stored in memory, the display responds instantly to each query.

At any point, one can ask the software to save a script – the behind-the-scenes code that generates the plot – to a data table. Any graph created earlier and saved to the data table can be re-displayed with just one click.

Data can be color-coded, shape-coded and mapped to geographies, such as a visualization that shows distribution by state of patients who discontinue their medications. This information helps clarify where the insurance company should focus its intervention efforts.

A **dynamic bubble plot** can show many dimensions on a two-dimensional display. In addition to X and Y axes, you have an identification variable a time variable to move the scatter plot through time, a size variable, and a color variable. In the demo image, the bubbles represent different regions and the relative size of each bubble shows average patient age. Because the time variable is service data, you can click "Go" to trigger a moving animation – a data movie – that shows how all these variables change over time.

Decision trees provide another approach to visual data exploration. Select "Option-Partition" from the "Analyze" pull-down menu, select predictor variables, click "Okay," and you get a graph of all the data points – 3,000 observations in this case – and how they relate to the response variable: discontinuation.

The decision tree systematically narrows the scope of the analysis by splitting the patient population on the most influential variables at each level, with greater detail at each step. This display reveals which variables to focus on when targeting patients for interventions.

Now the prediction formula can be saved and used to score the data. Three new columns were added to the data table, including probability that the patient will discontinue medication, and a score. The new columns will populate automatically.

The **data visualizations**, even the data movie, can be saved and e-mailed as PDF, SMS or HTML files. The recipient does not need to have SAS to view the visualization, only a Web browser.

"Visual exploration helps the data come alive to tell the story," said Milley. "If relationships change over time, you can see that, whereas in tabular output, you might miss that. These dynamic data visualizations especially are very memorable. When you tell the story in a compelling way, people remember it. They know what they need to know and the actions they need to take."

■ Visual Discovery

- Dynamic, interactive environment optimized for data discovery.
- Visual query and exploration of data.
- In-memory processing for instantaneous response.
- Animated graphics and data movies.
- Scalable for very large data sets.
- Support for both visual and coding paradigms.
- Integrated with SAS and Microsoft Excel.
- Self-documenting scripts and journals.
- Flash output to e-mail and share by Web browser.

Visual Programming – *Creating analytical workflows in a graphical environment*

There will always be a need for people to write code, but the paradigm is changing with the introduction of new and more user-friendly options, said Milley. "The power to program is increasingly surfaced in graphical user interfaces in the form of visual programming.

"Visual programming provides more powerful functionality than just writing code, because it enables you to logically lay out chunks of code and subroutines without the need to repeat coding logic again and again. This paradigm is about letting metadata work harder so you don't have to. You also avoid the typical mistakes of the pure coding approach, such as failure to run predecessor routines when the data has changed."

Visual programming is catching on fast. "Even longtime coders who thought they would never use any GUI have come to be huge fans of the benefits such a flexible environment provides," said Milley. Visual programming combines the power to code (custom code can still be embedded in code nodes) with the convenience of metadata flow-through and visual logic – a clear, birds-eye view of the whole process.

"One of the biggest advantages of visual programming is the time savings from having the score code automatically generated, including all of the pre-modeling steps, such as variable creation and all the transformation logic," said Milley. "This automation greatly reduces the potential for errors and helps the models get put into production much faster."

A Guided Tour of Visual Programming with SAS® Enterprise Miner™

Suppose a telecommunications company would like to reduce churn by identifying customers who are likely to change carriers. If the company could identify desirable customers who were at risk of leaving, and understand them, it could devise the right incentives to retain them.

Ryan showed a workflow created to mine the data to discover exactly that: which customers are likely to churn. The visual programming template is based on the five-step SEMMA process:

- *Sample* – Create training and test sample data sets with high predictive value.

- *Explore* – Interactively explore relationships and anomalies in the data.

- *Modify* – Create, transform and select the most appropriate variables for analysis.

- *Model* – Apply a range of modeling techniques to identify patterns in the data.

- *Assess* – Validate the usefulness and reliability of findings from the data mining process.

With visual programming, the process is intuitive. Simply drag and drop options or nodes into the workflow and connect them to the data as appropriate. In the visual Gantt-type display, it is easy to add steps to the workflow and track the steps that have been taken.

The first step in the sample workflow in the demo is to partition the data into a training set and a validation set to assess the validity of the model(s) under consideration. Click on the node at any time to display and change its properties. Some best practices and smart defaults are built in, but the analyst can modify the metadata as necessary.

Other nodes/steps in the sample workflow include general data exploration, some variable transformation, missing value imputation, variable selection and variable rejection, and then model creation.

At this point, the path splits, because this sample workflow is set up to create more than one model. "This is what makes the process so powerful," said Ryan. "You can create as many models as you like and feed them all into the model comparison node. You can even bring in models from other software packages and compare them to models created in this environment to see which is the champion model."

Just right-click the model comparison node and select "Results" to get both numeric and graphical displays that compare model performance. You can assess models on many measures. A graphical display gives an at-a-glance view of how the champion model (in this case, a forward regression model), along with other contenders, compares to a baseline model.

Once you have a champion model, you can score new observations and surface the results wherever needed, so appropriate business decisions can be made.

Right-click the score node and select "Results" to see the optimized SAS code. "You might not ever work directly with this SAS code, but it's here if you need it," said Ryan. "It is documented, optimized and can be converted into C, Java or HTML, depending on how you need to deploy your model. It's very flexible."

This visual programming process is scalable to very large data sets, Ryan noted. "The quantity of data is not going to be a problem. In fact, SAS Enterprise Miner expects that you will have very large data sets. This software is built for heavy lifting. It can handle large amounts of data, and ensemble models, which can be very CPU-intensive."

Furthermore, process flow diagrams are all logically laid out and automatically documented. "Self-documentation provides an audit trail of what has been done, so you're not as reliant on the person who created the workflow to document it," said Ryan. "This capability allows for easier collaboration, sharing of work and improved governance, which is necessary if you're really going to leverage analytics as a strategic asset."

■ Visual Programming

- Smart automation, to write less code.

- Automatic run of predecessor routines.

- Code generation, customization, optimization.

- Time-stamped, self-documented process flows.

- Support for embedded custom code nodes.

- Model comparisons, reporting and management.

- Automatic score-code generation.

- Faster Dev-Test-Qual-Prod cycles.

- No risk of translation error.

- Grid-enabled for distributable enterprise deployments.

- Support for both structured and unstructured data.

Visual programming can also be used with unstructured data (text, audio, video, images, etc.), such as from customer satisfaction surveys, social media, contact center records, insurance claims, etc. "Unstructured data is great for brand tracking, understanding sentiment and identifying important themes," said Milley. "Visual programming provides an effective way for more people to tap into the value of this textual data and, if needed, to combine textual data with structured data."

Wizard-Driven Analytics – *The user-friendly way to share or get started*

As analytics become more prevalent across the organization, there is a growing need for business analysts and subject matter experts to have greater self-sufficiency in predictive analysis. Analysts shouldn't have to rely on statisticians and modelers every time they need to develop new insights from the data, because the turnaround often needs to be fast, and lots of models need to be developed.

Visual programming can be a good start for some, especially those who don't write code, but many who want to get started with analytics will prize even greater simplicity and ease-of-use. Wizard-driven approaches enable more people to benefit from analytical best practices instead of spreadsheets.

With SAS Rapid Predictive Modeler, business analysts can build models without in-depth statistical knowledge. The platform automatically steps through a workflow of analytical tasks and draws on SAS Enterprise Miner behind the scenes to come up with the right model to yield the best results. Business analysts can find on-demand insights and act on them quickly and effectively in a self-sufficient manner. Quantitative specialists find this capability useful to quickly generate baseline models.

You can choose from several levels of automation, from a fast-and-simple mode to a more compute-intensive, advanced mode. When used with the SAS Add-In for Microsoft Office, this analysis can take place within a familiar Excel environment to benefit from smart automation with repeatable, reproducible and auditable results.

"Now, we are not advocating a black-box approach to solve problems where the upside or downside is large," Milley noted. "In certain cases though, wizard-driven approaches can enable people to make better decisions without having to rely on experts. The experts are often delighted to offload some of the more routine work, to free up time to solve the more challenging problems.

"Collaboration can still take place, with the expert providing the template, or with the more novice user starting the analysis and then having an expert review or extend it if needed. The trick is to automate wisely, and this requires upfront thinking."

■ Modelers and statisticians will always play a key role in analytic data preparation and model development, but there is a need to make their work more readily available in a self-service type of application.

A Guided Tour of SAS® Rapid Predictive Modeler

In a live product demo, Ryan revisited the telecom provider's concerns over customer churn. Suppose sales and marketing analysts want to quickly identify the variables associated with a high propensity to switch carriers. Armed with that knowledge, they would be able to develop promotions, incentives or other interventions to keep them.

Predictive models can be developed from within a familiar interface – in this case, Excel – using a prebuilt task under the SAS tab. Just indicate the input data and launch the task.

In the demo, the target variable is churn, of course. All other variables have been identified as potential predictor variables, except account ID. Choose model options – either a basic, intermediate or advanced data mining flow – for determining the champion model. Ryan selected "Intermediate."

"Once I have indicated the type of model I want the software to build, I can add options, such as lift plot and variable rankings," said Ryan. "All the tasks around this process have been packaged."

From the results, you learn more about the champion model. You can quickly see which variables are the most important for separating churners from non-churners, the relative importance of each variable, and how the champion model performs compared to the baseline model.

The champion model is registered and then used to score future observations. Results can be easily shared wherever they are needed, so the appropriate business decisions can be made.

"It is a nice combination of empowering others in a more governed way," said Milley. "It is so easy to get those results to develop the right strategy, even if you don't have a comfort level with data mining, or the time and resources for a more involved analytic process."

Wizard-driven capabilities are also available for developing forecasts, Milley said. "One survey showed that more than 40 percent of forecasts are still done in spreadsheets, and I'm guessing these are not time series forecasts or statistical forecasts. For forecasters looking to evolve beyond spreadsheets, a wizard-driven approach makes a lot of sense."

■ Organizations that arm business analysts with wizard-driven analytics empower a whole new group of employees to determine the best actions for reducing risk and improving organizational performance.

■ **Wizard-Driven Analytics**

- Quick generation of predictive models at a choice of automation levels.

- Guided model-building with automated best practices and smart defaults.

- Self-sufficiency for business analysts, with a user-friendly visual interface.

- Delivery of analytic results in a simple and consumable way.

- Support for improvement and customization of models using SAS Enterprise Miner.

Closing Thoughts

"There is an overreliance on spreadsheets for many decisions that could be better informed and governed using analytic approaches," said Milley. "Many organizations have admitted significant operational risk exposure and known mistakes using spreadsheets for business applications." The smart ones have taken steps to remedy that, but many find that analytic expertise is in short supply.

"Does your organization have the best mapping of skills and abilities to coding paradigms?" Milley asked. "Are you doing things the same way because that's how you've always done them? Perhaps you're still just coding, and coding is wonderful, but there are ways to improve upon traditional coding, augment your organization's analytic bandwidth, empower more people to generate analytic insights, and improve governance in the process."

For more information: www.sas.com/technologies/analytics

To view the Analytics 101 webinar: www.sas.com/reg/web/corp/950738

About the Presenters

Anne Milley, Senior Director of Analytic Strategy, Worldwide Product Marketing at SAS, works closely with Product Marketing, Product Management and R&D to drive SAS' analytic marketing strategy and direction. She began working with SAS while finishing her thesis on bank failure prediction at the Federal Home Loan Bank of Dallas. She continued her use of SAS at 7-Eleven Inc. as a senior business consultant.

Milley writes and speaks on analytics. She has co-chaired and served on committees for analytical conferences. In 2008, she completed a five-month assignment at a major bank in the UK. She has a Master of Arts in economics from Florida Atlantic University, did postgraduate work at Rheinisch-Westfälische Technische Hochschule Aachen and is proficient in German.

Laura Ryan, Systems Engineer at SAS, was first introduced to JMP as a high school math teacher, and to SAS as a graduate student in applied statistics. After working as a software trainer for almost five years in the education department at SAS, she now provides technical assistance for software pre-sales activities.