Paper 331-2013

10-Minute JMP®

George J. Hurley, The Hershey Company

ABSTRACT

Heard of JMP®, but haven't had time to try it? Don't want to devote 50 minutes to a talk about software that you might not want to use? This is the talk to you. In 10 minutes, you will learn some of the amazing visualization and modeling features in JMP and how to use them. This talk will JMP-start your JMP usage. When it's complete, we suspect you will want to attend some of the longer talks, too.

INTRODUCTION

JMP software is a data analytics and visualization package that was originally conceptualized by John Sall at SAS® Institute. Like SAS®, JMP has scripting capabilities, in the form of JMP Scripting Language, or JSL. JSL allows users of JMP to also submit SAS code if the users have a SAS license. There are many possibilities to build processes using JSL and submissions to SAS; however, what makes JMP powerful to a new user is its point and click access to powerful graphics and analytics. The learning curve for this aspect of JMP is minimal and those attending this talk should get a broad view of its capabilities, and with attendance of a few additional talks, can learn enough to go back to the office and begin using and digging deeper into this product.

This paper will focus on some of the graphics and analytical graphics features available in JMP. While JMP is capable of doing much more than described in this paper, these are some great techniques to get started.

VISUALIZATIONS

JMP GRAPH BUILDER

As with any dataset, it is ideal to visualize the dataset first. The JMP Graph Builder is an excellent tool that offers the flexibility to visualize data in a multitude of ways by simply dragging and dropping variables. To open the JMP Graph Builder, go to the Graph menu and select "Graph Builder".

Once selected, the Graph Builder window will open. The window consists of a region where variables can be selected, a symbol representing their variable type, and an area that will contain graphics that the variables will be dragged to.

The graphics region consists of eleven zones. These zones are called "X", "Y", "Group X", "Group Y", "Wrap", "Overlay", "Color", "Legend", "Shape", "Freq", and then the central data zone, with the caption "Drag and drop variables into graph zones". These regions interact with one another, depending on how variables are dragged into the Graph Builder.

Specifically, the Graph Builder is a "smart" interface in that it will attempt to build the optimal graph based on the variables chosen, their variable types, and which zone they are placed in the graphics region. For those knowledgeable in using Microsoft Excel® Pivot Tables, the functionality of the Graph Builder can be considered similar, but instead of producing tables (or a somewhat dry array of Pivot Charts), the Graph Builder can produce a much richer array of different visualizations. Figure 1 provides visualization based on the Fitness.jmp sample data set. All JMP sample data is found by using the "Help" menu and selecting "Sample Data". The Graph Builder can also produce maps, as seen below in Figure 2.

With limited effort, a large number of graphs representing various visualizations of the data can be created using Graph Builder, generating useful insights. There are multiple examples available, but the easiest way to master this interface is to open up a dataset and drag and drop variables and see the results. Graph Builder is an intuitive tool to learn, with meaningful usage possible within an hour. Graph Builder is covered in more detail in my 2011 paper, "Using JMP® Visualizations to Build a Statistical Model".[1] Also, note that all SAS conference papers referenced in this paper can be found at www.lexjansen.com.
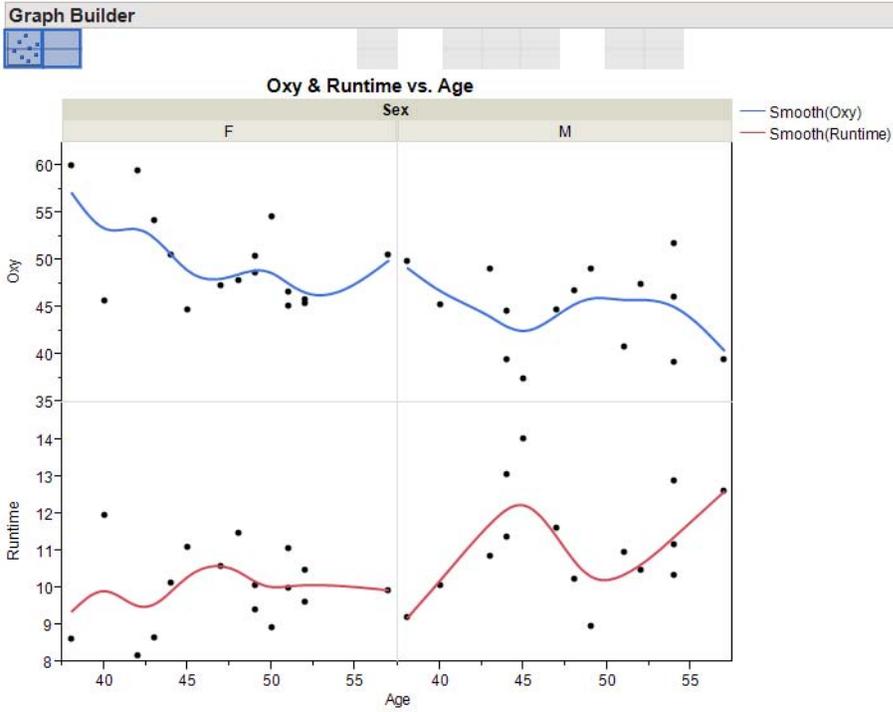
Figure 1. Graph Builder Visualization: Plotting Age, Gender, Oxygen, and Runtime
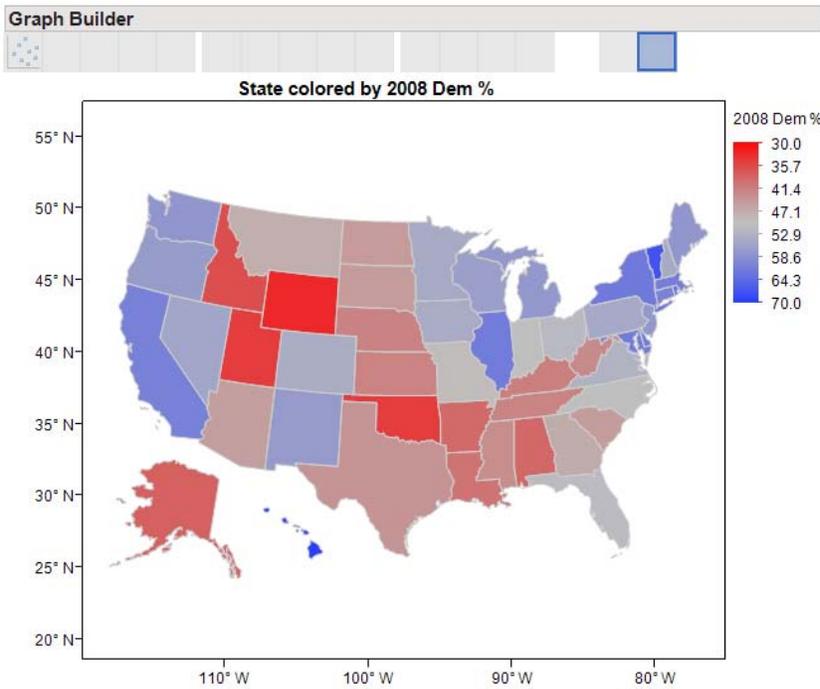

Figure 2. Graph Builder Map Visualization: 2008 Democratic Percent of Presidential Vote by State

TREE MAP

One particularly interesting graphical feature in JMP is tree maps. A tree map is method of plotting outcomes (either categorical or continuous) by the size of their relevance. This, like all JMP features, is activated by an easily mastered user interface.

The 2000 U.S. Presidential Election was one of the more interesting elections in modern U.S. history and illustrated the unique nature of the Electoral College. As the winner is chosen by electoral votes, which are awarded based on winning popular at the state number (with two exceptions), George W. Bush defeated Al Gore, as he had 271 electoral votes compared to 269 electoral votes. However, Mr. Gore had 50,999,897 votes overall, compared to Mr. Bush's 50,456,002. The tree maps below illustrate these differences.

In Figure 3, the red blocks are states that were won by President Bush, sized by their electoral votes. The blue blocks are states won by Mr. Gore. However, in Figure 4, the blocks are sized by popular vote margin of victory, as opposed to electoral votes. This, of course, illustrates the popular vote. While both Figure 3 and Figure 4 indicate a razor-thin margin, it is clear that the allocation of the popular vote margins is generated in a way that is quite different than that of the electoral vote.
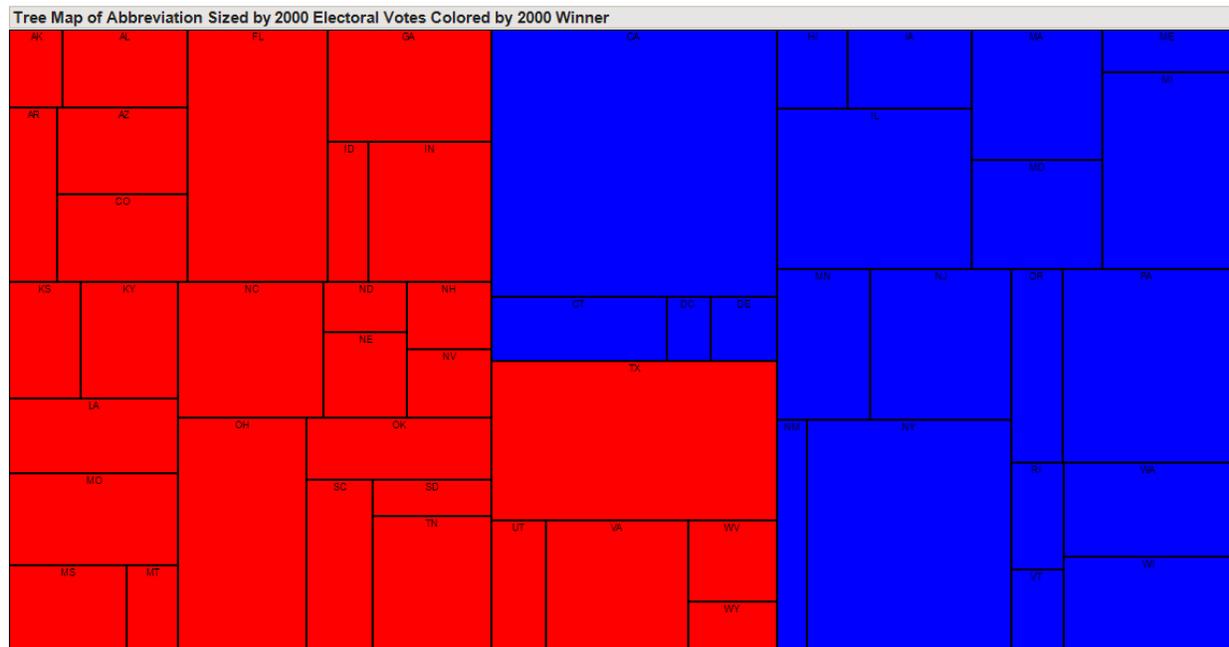

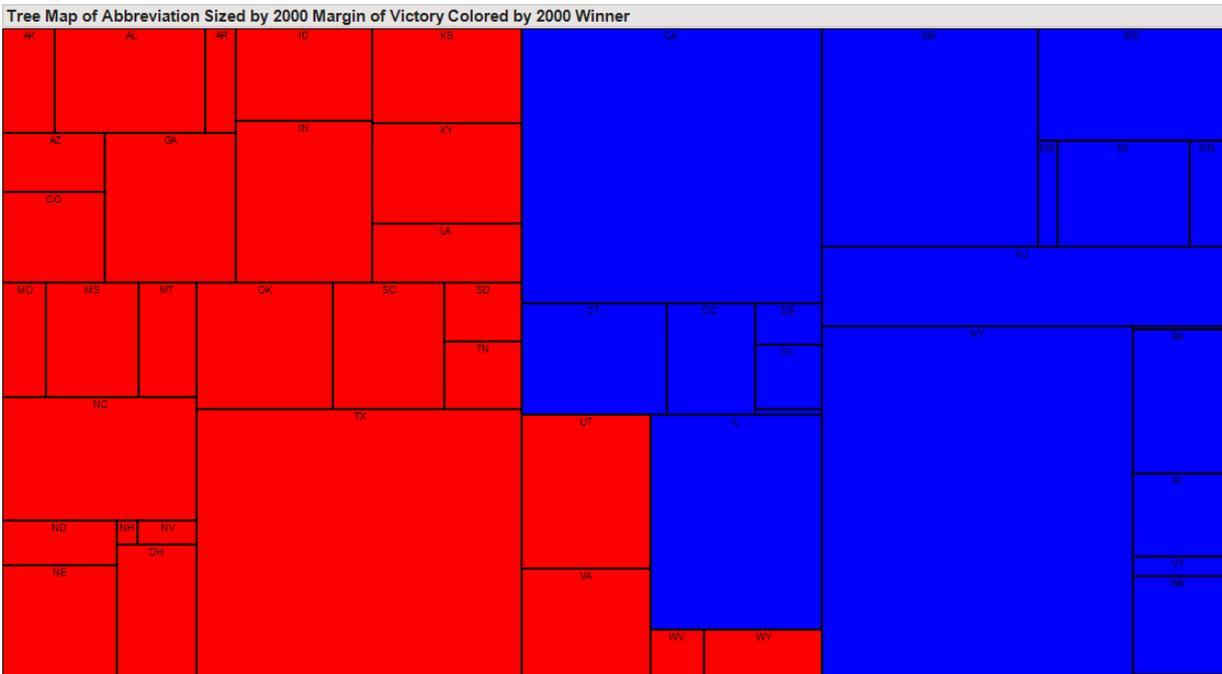**Figure 3. 2000 Election Tree Map Sized on Electoral Votes**

Figure 4. 2000 Election Tree Map Sized on Popular Vote Margin of Victory

GEOCODED BUBBLE PLOTS

A very useful feature in JMP is the ability to make geocoded bubble plots. A geocoded bubble plot depicts data using both metrics (i.e. sales) and spatial information (usually latitude and longitude) on a map (political, physical, structural, etc.). JMP allows for these to differentiate between distinct groups and to run temporally. Ultimately, five dimensions of data can be plotted in one instance.

Consider a USDA dataset that lists over 7,800 U.S. farmers markets' locations and geographic coordinates, based on data from the USDA National Farmers Market Directory[2]. It should be noted that the dataset states that the information is voluntary and self-reported by market managers, representatives from state farmers market agencies and associations, and other key market personnel across the country.

In addition to the geographic coordinates of the markets, this dataset contains binary information answering "Y/N" if a farmers market sells items in a particular category, such as "Cheese", "Eggs", "Maple", "Trees", etc. A column was created in Excel (using the CountIf function) to determine the number of categories that were covered (out of 20 total). It should further be noted that x denotes longitude and y denotes latitude in this dataset. Figure 6 illustrates the data in JMP.

This data can be easily plotted in JMP using a simple geocoded bubble plot. To learn to create these, reference "JMP® Spatial and Temporal Graphics: The Geocoded Moving Bubble Plot, Proc Geocode, and PowerPoint"[3]. Figure 5 illustrates a basic map of the U.S. that can be created using this method, while Figure 6 illustrates a customized map of Pennsylvania zip codes generated from a shape file. Shape files for many maps are available free online. Finally, Figure 7 is a screenshot of a plot that has been turned into a 5-dimensional moving bubble plot. This plot can be dynamically changed, or "played" through time like a movie.
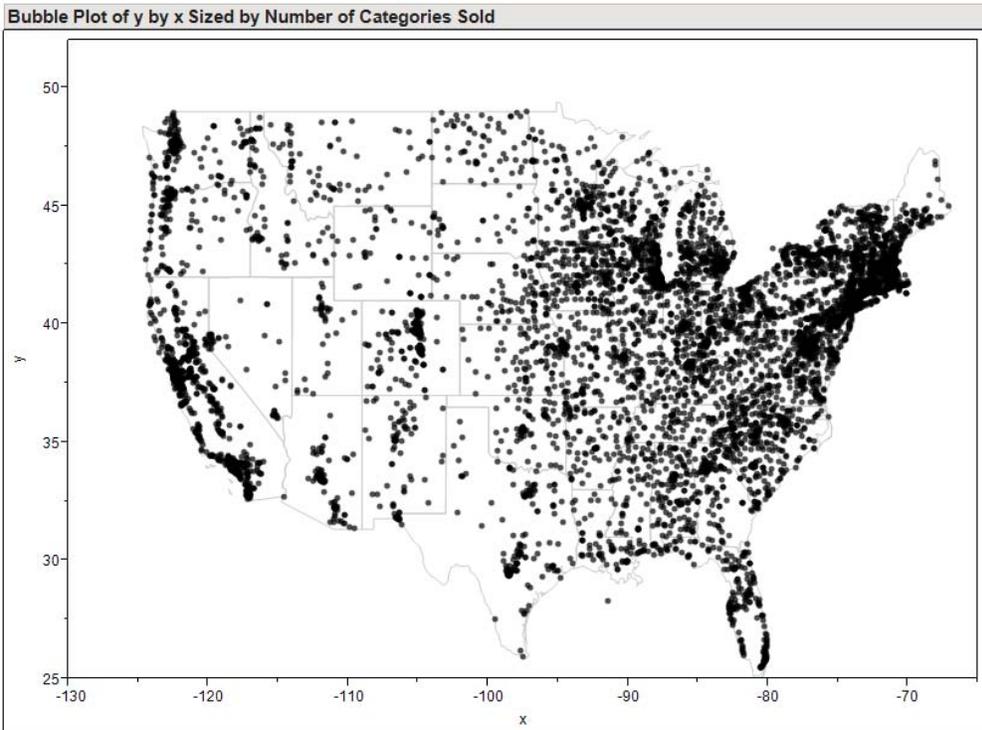
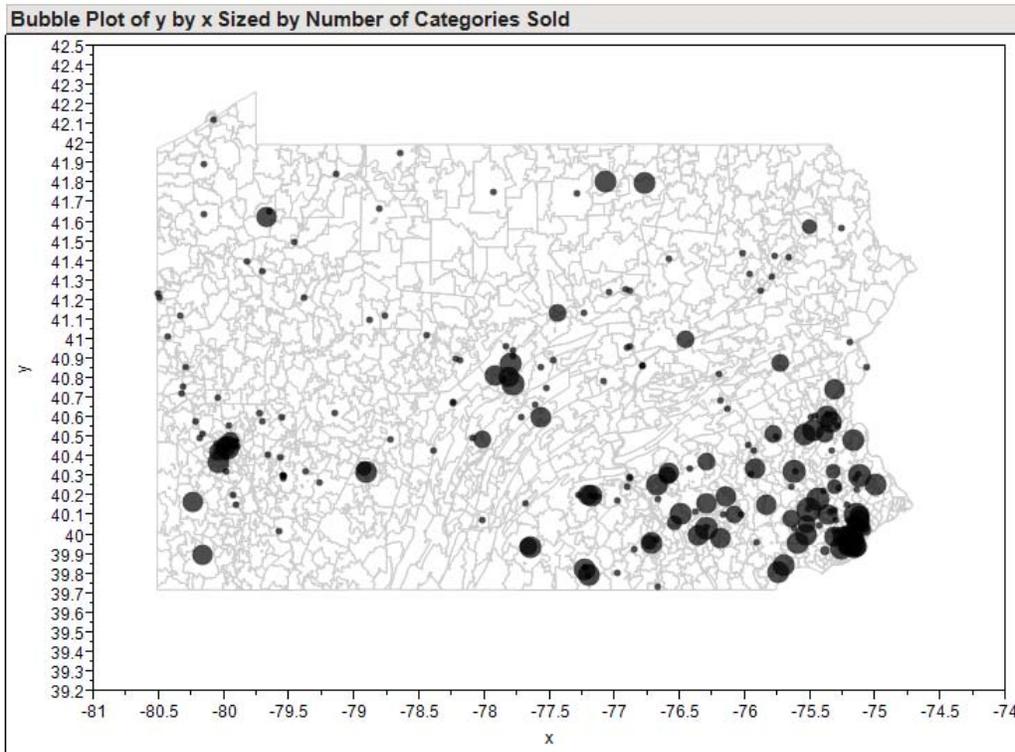**Figure 5. Basic Geocoded Bubble Plot of Farmers Markets**
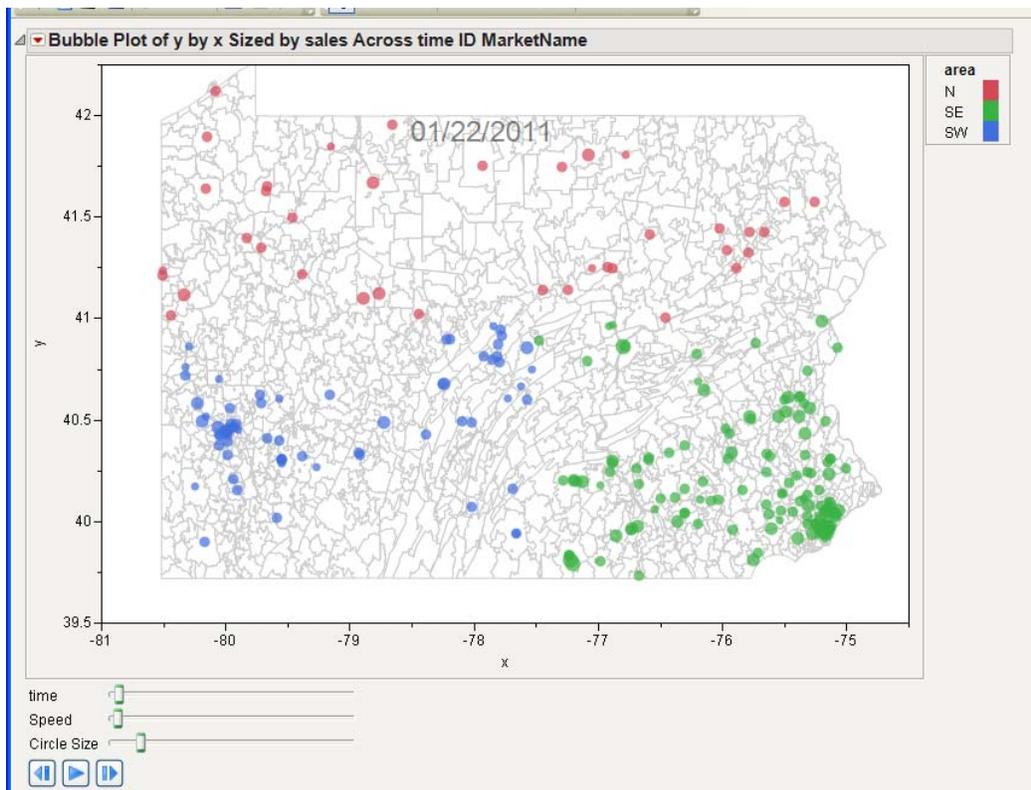

Figure 6. PA Farmers Market Distribution

Figure 7. Screen Shot of Five-Dimensional Moving Plot

ANALYTICAL GRAPHICS

FITTING A REGRESSION MODEL

The model fitting platform in JMP provides an intuitive platform in which statistical models can be developed.

In this case, a dataset is generated from within JMP. Here, SAS code is ran within JSL, which is a scripting language internal to JMP. Here, the JSL is using "sassubmit" to run regular SAS code. The implication is that SAS code can be ran within JMP. This code is below.

```
sassubmit(
    "data sgf2011c (drop=i);
    do i=1 to 10000;
          x1=rannor(1978)+2;
          x2=rannor(1999)-1;
          x3=rannor(1971)+1;
          x4=rannor(2006);
          x5=rannor(1888);
          x6=rannor(1234);
          y=x1*2+x2*3+x3+x4*6+rannor(2009);
    output;
    end;
    run;"
)
;
sas import data("work", "sgf2011c")
;
```

As can be seen, this dataset is designed to setup a multiple linear regression model. This model was chosen due to its applicability to a broad audience. However, JMP handles a wide arrange of models and fitting.

Since data was created for this example, plot examples as shown above would be redundant. However, for a real dataset, one would want to use the graph builder and distribution plots to get a feel for the data and verify that it is reasonably distributed; there are no outlier issues to consider, etc. In this instance, we will proceed straight to modeling and will consider linear models. To do this, the Analyze Menu is chosen. Under Analyze, Fit Model is chosen. Here, Y is chosen as the Y variable and each of the x variables is chosen as the x variables.

The stepwise platform will be taken advantage of to five stopping rules available for model selection, BIC (Schwarz's Bayesian Information Criterion), AICc (Corrected Akaike Criterion), p-Value threshold, Max Validation R-Square, and Max K-fold R-Square. In this case, since a validation sample was not specified, the two R-Square methodologies are grayed out. Further, other options that one typically associates with stepwise regression are available, such as direction, entering all variables to start, etc. Here, AICc will be chosen with forward selection, and no variables entered. Each step in the model selection can be taken using the "Step" button, or alternatively, the "Go" button allows JMP to quickly determine the model based on the specified criterion. The model chosen by JMP is as expected, with x1, x2, x3, x4 all selected having coefficients of 1.99, 3.01, 1.00, and 5.99, respectively. Further, a very high r-square is seen. However, often, using real data, one may want to look at the effects of non-selected variables in the model. This can be accomplished by manually checking the boxes in the "Current Estimates" section of the window. Further, if there is a variable of interest that is being assessed (as in the case of a hypothesis test), it can be "locked" in, while the model is built around it. Here, we accept the model selecting x1, x2, x3, and x4, by choosing "Run Model" with those four selected. This brings up the "Fit Model" Window.

For comparison, this analysis would be done using proc reg in SAS. For more complicated datasets with a lot of exploratory analysis necessary, it would take multiple iterations of proc reg.

JMP MODELING GRAPHICS

JMP statistical graphics include a number of features. Here, the prediction profiler, interaction profiler, and ANN interface will be highlighted.

Once a statistical model is built, the prediction profiler allows the user to see how changing levels of any variable or combination of variables will affect the response, as well as providing another way to conceptualize interactions. That is, if an interaction term is in the model, as the level of one interacting effect is changed, the slope of the other interacting effect changes, as appropriate. This is another way to think about interaction effects in a visual manner. The prediction profiler is seen in Figure 8.

Likewise, more traditional interaction plots can be created using the interaction profiler (See Figure 9).

Finally, even Artificial Neural Networks (ANN) can be visualized using this technology. Figure 10 provides a screen shot of the ANN interface.

There are many additional graphical features available for statistical modeling within JMP. In general, like all JMP features, the prediction profiler, interaction profiler, and ANN interface, as well as these other features, are easy to get up to speed on. To learn more about these features, once again, see, "Using JMP® Visualizations to Build a Statistical Model".[1]
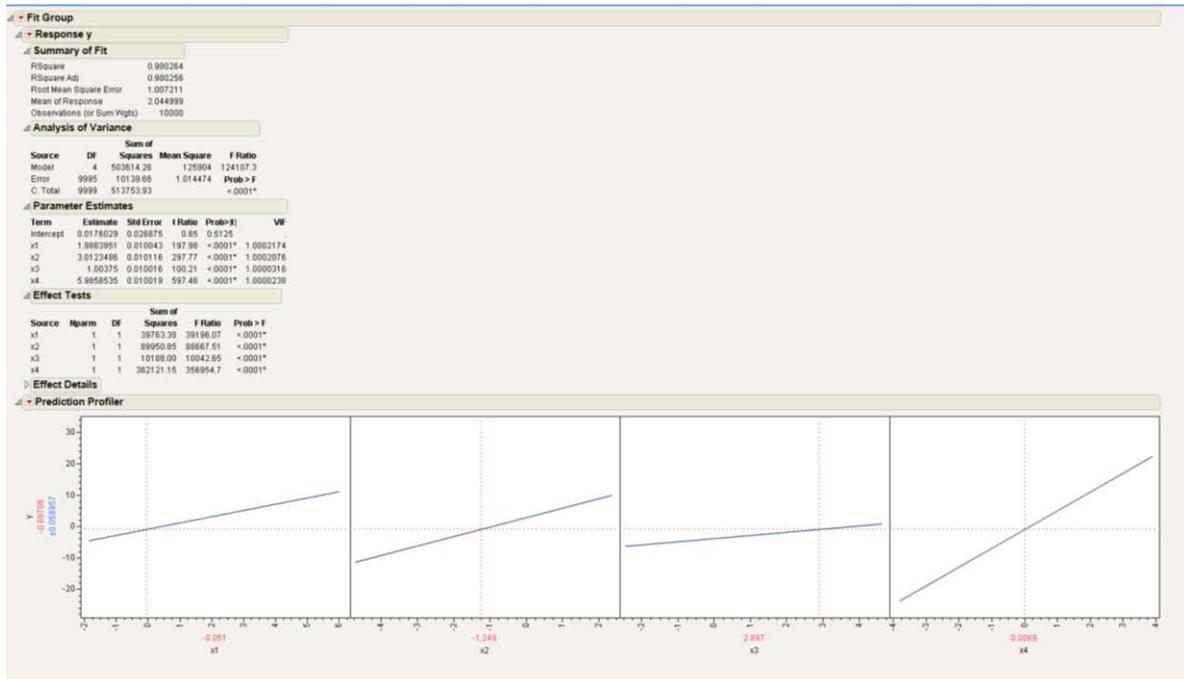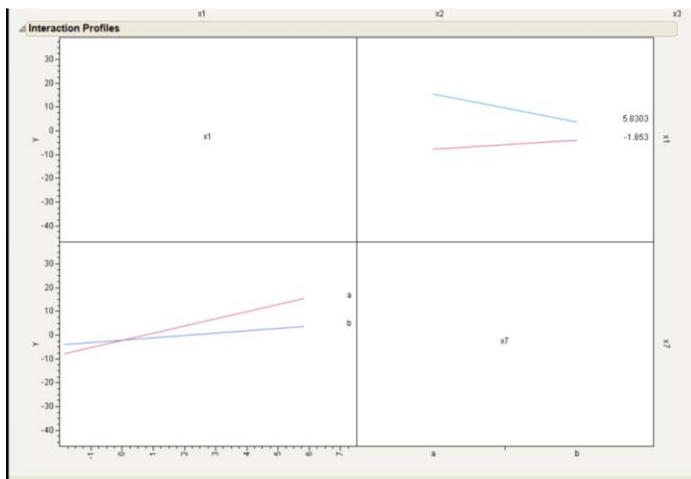
Figure 8. JMP Prediction Profiler

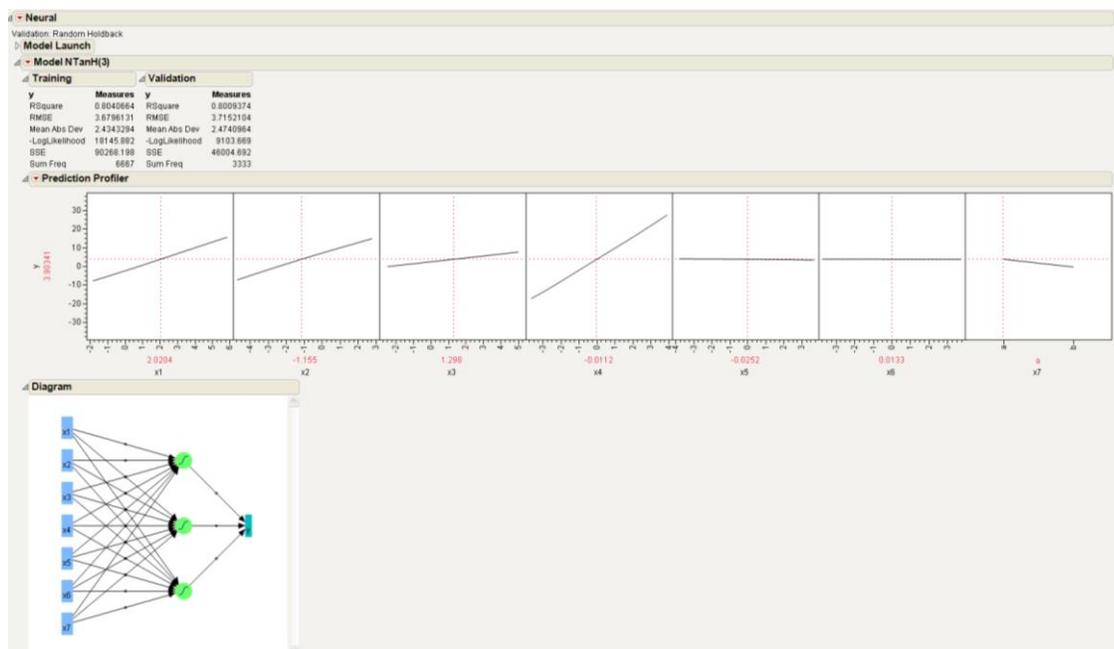**Figure 9. JMP Interaction Profiler**

Figure 10. ANN Interface

CLUSTERING

JMP offers five clustering methods. These are hierarchical clustering, k-means clustering, normal mixtures, robust normal mixtures, and self-organizing maps. These methods are all available in JMP® Pro. As this paper is a survey of JMP, details of these methods will not be included here. However, the paper, "A Brief Survey of Clustering and Segmentation for Customer Intelligence"[4], provides additional information on this topic. However, as is requisite for even brief clustering references, see Figure 11 for a Dendrogram created in JMP.

OTHER JMP FEATURES

JMP offers numerous other analytical features, ranging from experimental design to data mining. All of which are easy to use and interface with. One particularly interesting feature in JMP Pro is the implementation of the "Bootstrap Forest" algorithm, which is a repeated partitioning algorithm, which is useful for certain data mining and modeling applications. More information on this method can be found in "JMP® Pro Bootstrap Forest"[5].

CONCLUSION

JMP has an extensive array of analytical and visualization capabilities. This paper has illustrated a survey of these capabilities. While JMP has tremendous depth and power, the learning curve to use JMP is not steep. In most cases, a user can run many analyses within a day's time. For any user that would like to increase their analytical and visualization capabilities, JMP is a great addition!
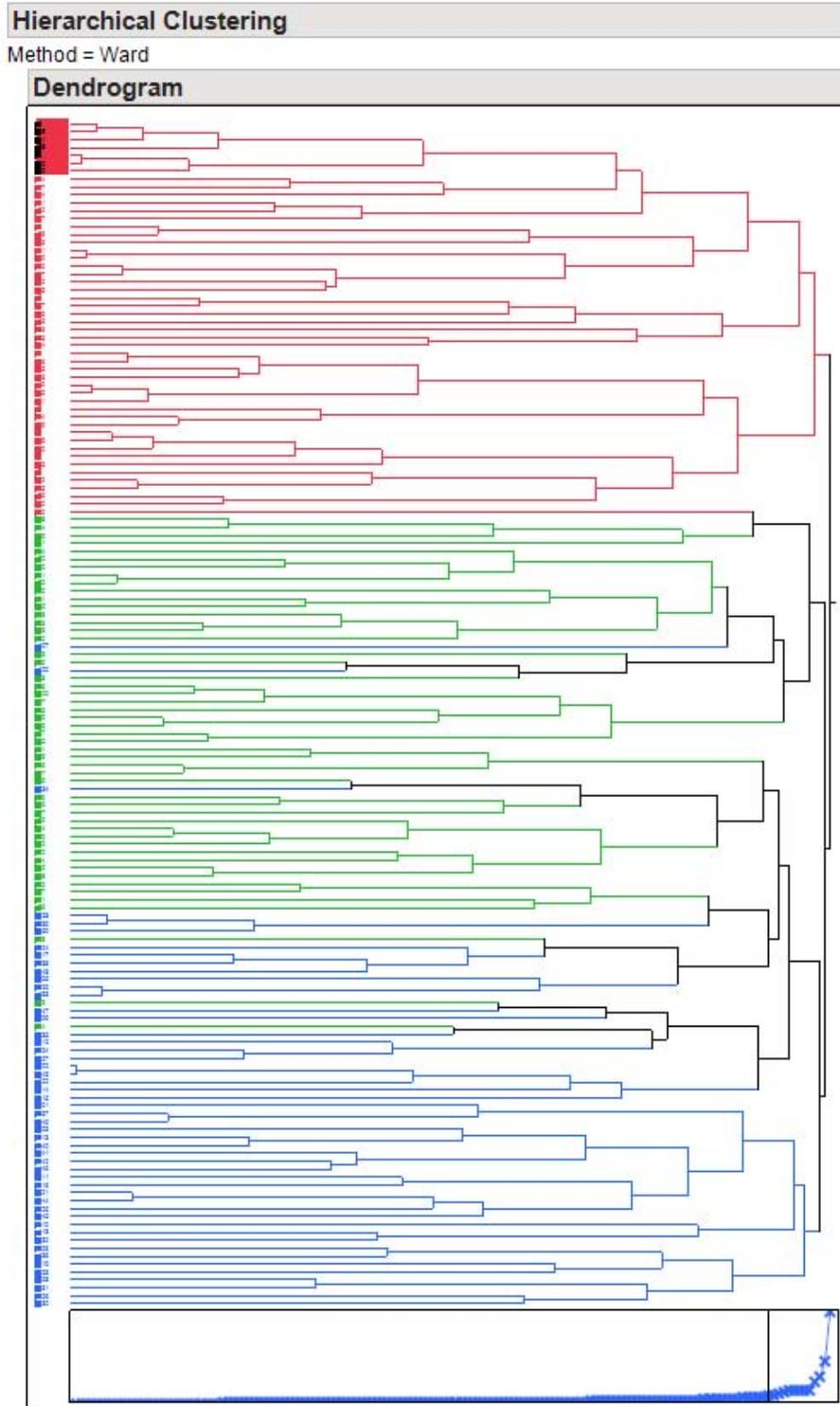
Figure 11. A Dendrogram Created in JMP

REFERENCES <HEADING 1>

[1] Hurley, G.J., "Using JMP® Visualizations to Build a Statistical Model", SAS Global Forum 2011, Paper 298-2011

[2] https://explore.data.gov/Agriculture/Farmers-Markets-Geographic-Data/wfna-38ey

[3] Hurley, G.J., "Using JMP® Visualizations to Build a Statistical Model", MWSUG 2012, Paper JM-01

[4] Hurley, G.J., "A Brief Survey of Clustering and Segmentation for Customer Intelligence", MWSUG 2012, Paper CI-08

5 Hurley, G.J., "A JMP® Pro Bootstrap Forest", MWSUG 2012, Paper JM-04

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

George J. Hurley
The Hershey Company
19 E Chocolate Ave.
Hershey, PA 17033
717.534.5337
717.534.6991
ghurley@hersheys.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.