Paper JM-01

JMP® Spatial and Temporal Graphics: The Geocoded Moving Bubble Plot, Proc Geocode, and PowerPoint

George J. Hurley, The Hershey Company, Hershey, PA

Abstract

Some of the primary accolades of JMP are the graphics features, which allow for easily created graphics with minimal coding. This paper will focus on how to create bubble plots in a map environment. These plots will include a temporal dimension, classification dimension, and a magnitude dimension, creating a 5-dimensional plot on a single surface. In addition to creating the plot, the pre-work of using proc geocode to generate map-type data and the post-work of saving the plot as a Flash movie and including it in PowerPoint will also be discussed

Introduction

In today's business environment, there are many questions in which geography plays an important role, including sales, operations, logistics, marketing, and quantitative research. Further, outside of business, geography is an important component in other areas, such as epidemiology, polling/survey research, and government.

JMP offers a wide variety of graphics options that can address geographical questions. This paper will focus on one of the more complicated, yet enlightening graphics options, the geocoded moving bubble plot. Several related topics, including SAS® Proc Geocode and PowerPoint will be considered as well.

What is geocoded data?

According to Wikipedia, "**Geocoding** is the process of finding associated geographic coordinates (often expressed as latitude and longitude) from other geographic data, such as street addresses, or zip codes (postal codes)."[1] Geocoded data often provides useful information tied to geographic points; for example, geocoded data may contain sales data by store, with each store having its unique latitudinal and longitudinal coordinates provided.

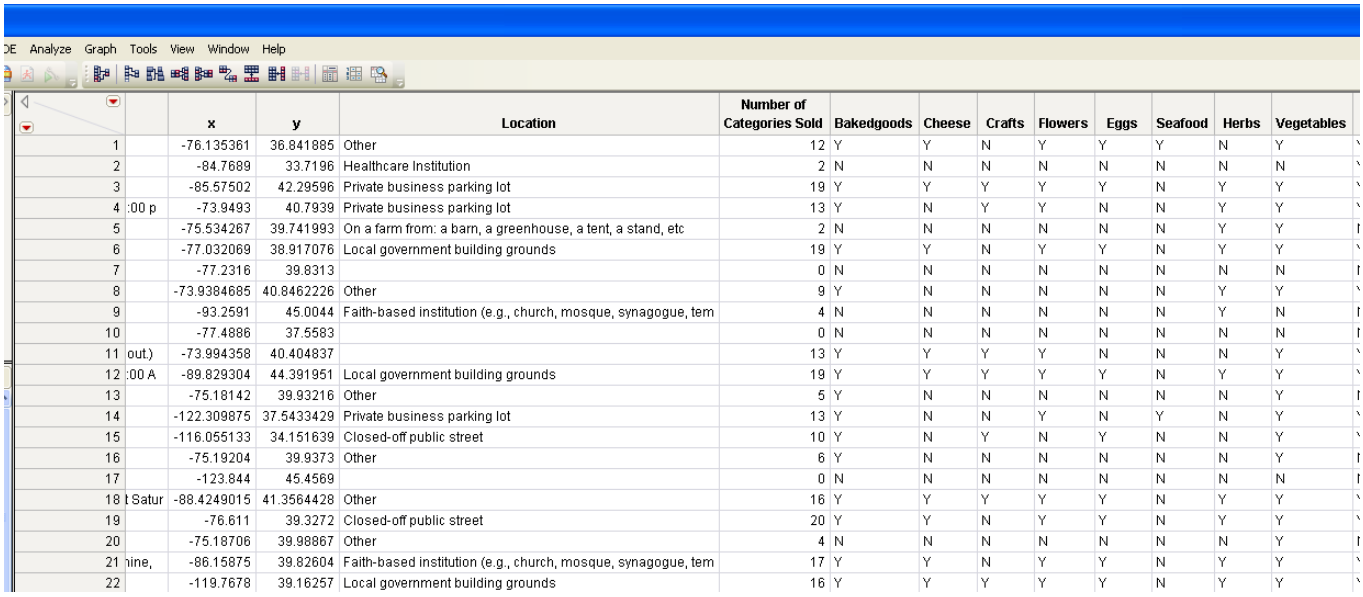
A simple geocoded bubble plot

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 1 illustrates the data in JMP.

Figure 1: Screenshot of Farmers Market Data


DE Analyze Graph Tools View Window Help

		x	y	Location	Number of Categories Sold	Bakedgoods	Cheese	Crafts	Flowers	Eggs	Seafood	Herbs	Vegetables
1		-76.135361	36.841885	Other	12	Y	Y	N	Y	Y	Y	N	Y
2		-84.7689	33.7196	Healthcare Institution	2	N	N	N	N	N	N	N	N
3		-85.57502	42.29596	Private business parking lot	19	Y	Y	Y	Y	Y	N	Y	Y
4	:00 p	-73.9493	40.7939	Private business parking lot	13	Y	N	Y	Y	N	N	Y	Y
5		-75.534267	39.741993	On a farm from: a barn, a greenhouse, a tent, a stand, etc	2	N	N	N	N	N	N	Y	Y
6		-77.032069	38.917076	Local government building grounds	19	Y	Y	N	Y	Y	N	Y	Y
7		-77.2316	39.8313		0	N	N	N	N	N	N	N	N
8		-73.9384685	40.8462226	Other	9	Y	N	N	N	N	N	Y	Y
9		-93.2591	45.0044	Faith-based institution (e.g., church, mosque, synagogue, tem	4	N	N	N	N	N	N	Y	N
10		-77.4886	37.5583		0	N	N	N	N	N	N	N	N
11	out.)	-73.994358	40.404837		13	Y	Y	Y	Y	N	N	N	Y
12	:00 A	-89.829304	44.391951	Local government building grounds	19	Y	Y	Y	Y	Y	N	Y	Y
13		-75.18142	39.93216	Other	5	Y	N	N	N	N	N	N	Y
14		-122.309875	37.5433429	Private business parking lot	13	Y	N	N	Y	N	Y	N	Y
15		-116.055133	34.151639	Closed-off public street	10	Y	N	Y	N	Y	N	N	Y
16		-75.19204	39.9373	Other	6	Y	N	N	N	N	N	N	Y
17		-123.844	45.4569		0	N	N	N	N	N	N	N	N
18	t Satur	-88.4249015	41.3564428	Other	16	Y	Y	Y	Y	Y	N	Y	Y
19		-76.611	39.3272	Closed-off public street	20	Y	Y	N	Y	Y	N	Y	Y
20		-75.18706	39.98867	Other	4	N	N	N	N	N	N	N	Y
21	hine,	-86.15875	39.82604	Faith-based institution (e.g., church, mosque, synagogue, tem	17	Y	Y	N	Y	Y	N	Y	Y
22		-119.7678	39.16257	Local government building grounds	16	Y	Y	Y	Y	Y	N	Y	Y

This data can be easily plotted in JMP using a simple geocoded bubble plot. To do so, select "Bubble Plot" from the Graph Menu. This opens the Bubble Plot window. To create the plot, first, in the Bubble Plot window, drag y to "y", drag x to "x", and drag Number of Categories Sold to "Sizes". Then, click "OK". Figure 2 illustrates this. However, this brings up the default and non-geocoded Bubble Plot (Figure 3).

Right clicking the plot area and choosing "Background Map" brings up a menu giving various options to choose a background map that corresponds to the data you have. There are a number of default maps that appear. Here, choose "US States", because this is what the data corresponds to. After this selection, when returning to the graph, the circle size can be reduced so that the map can be more easily (Figure 4). It is seen that the denser areas of the farmers markets are located on the East Coast, the Mid-West, and the West Coast.

Note the inclusion of Alaska and Hawaii cause a lot of white space in the map. The simplest way to exclude these states is to double-click each axis and restrain the latitude and longitude range (Figure 5).

Figure 2: Screen Shot of Bubble Plot Window

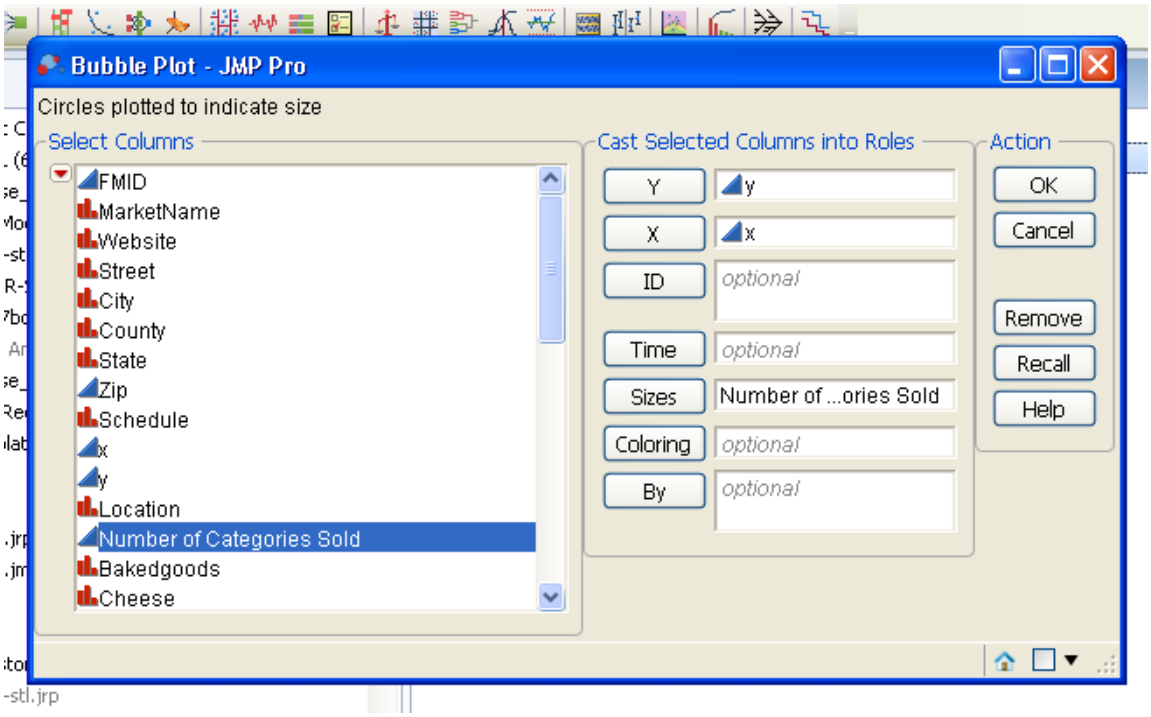


Figure 3: Screen Shot of Bubble Plot Window

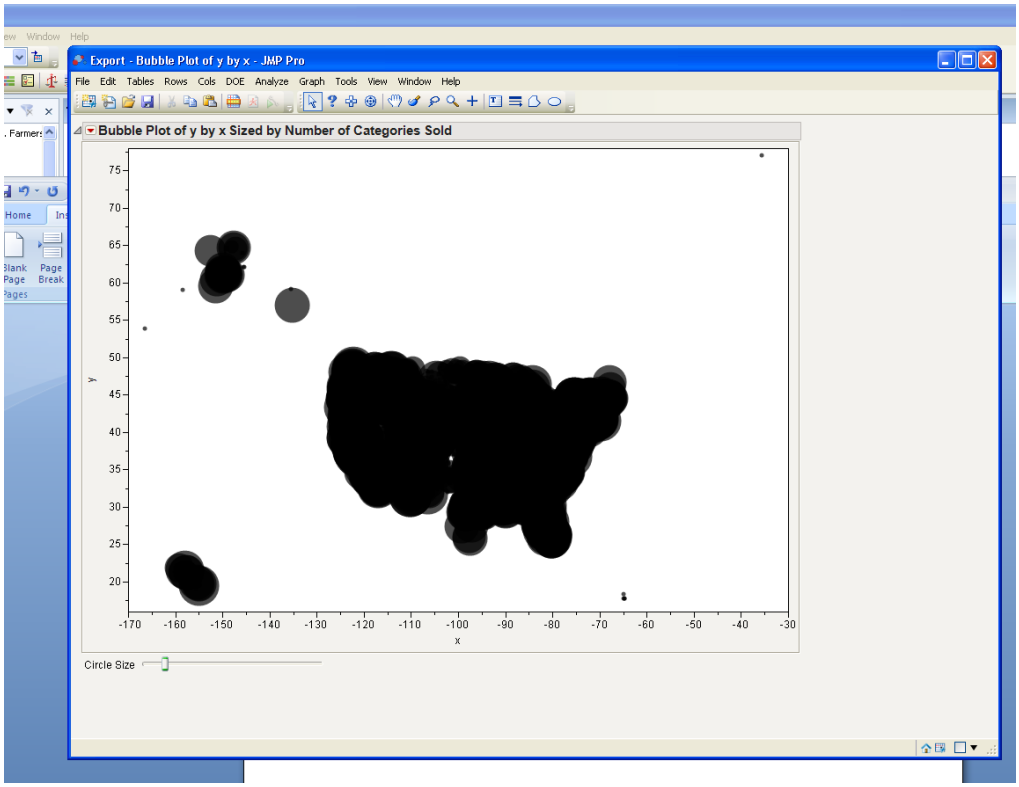


Figure 4: Bubble Plot with US State Map Superimposed

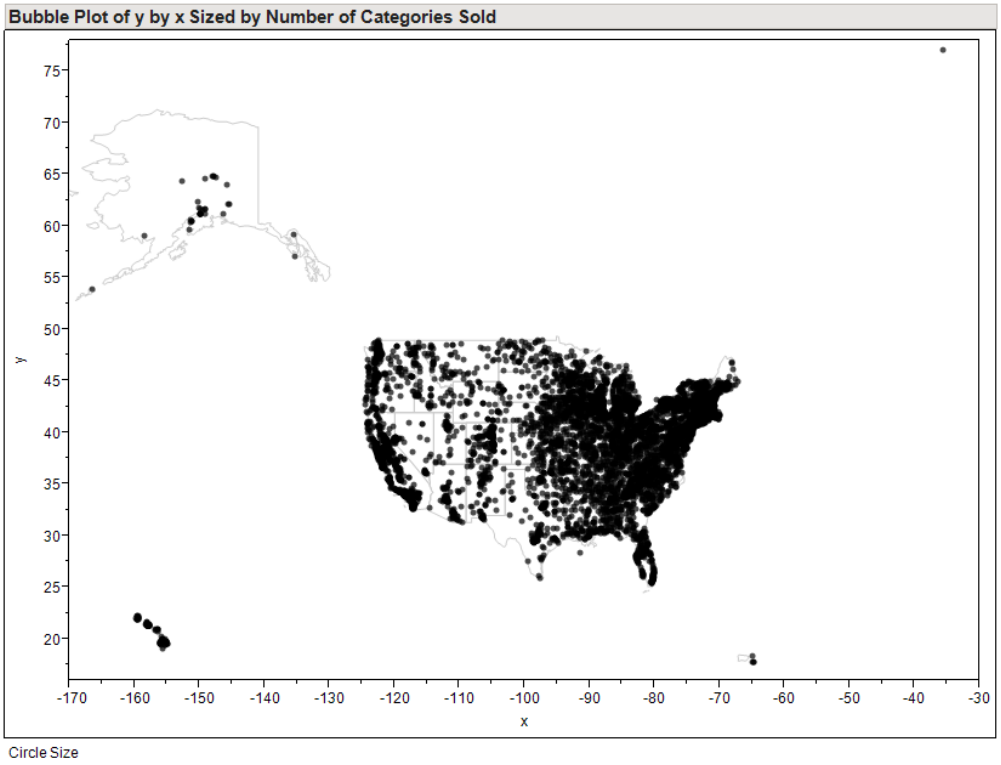
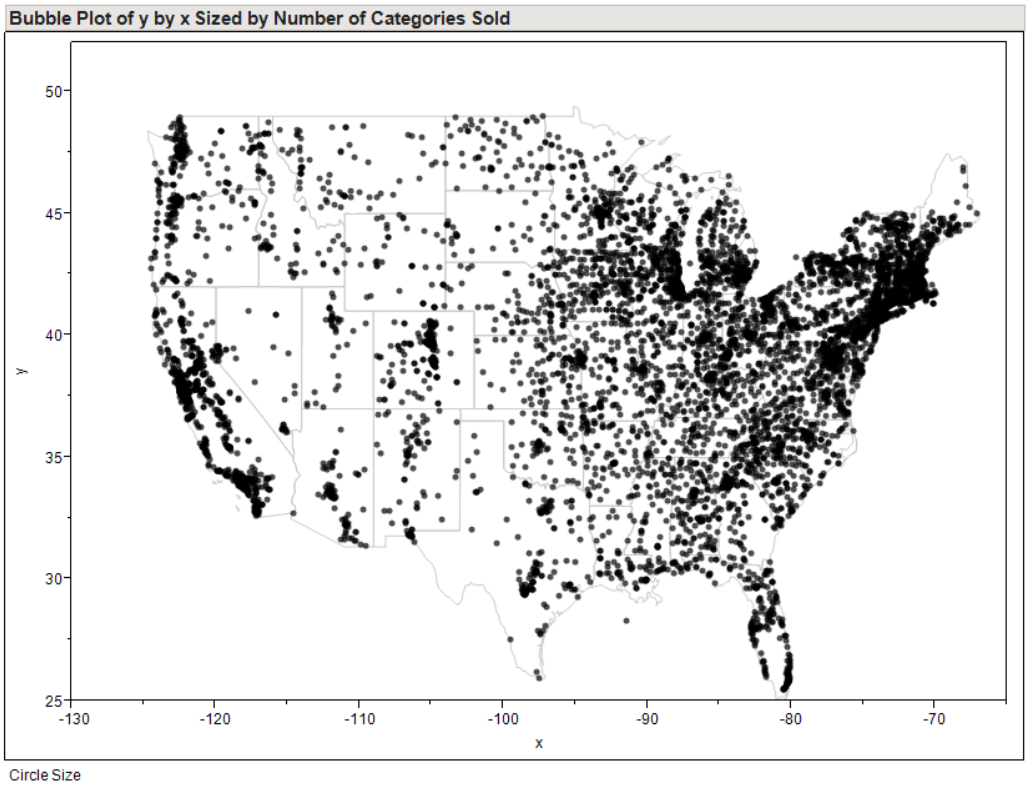


Figure 5: Bubble Plot with US State Map Superimposed (Alaska and Hawaii Excluded)



Of course, the points corresponding to these states could also have been excluded and the plot recreated by the same method described above.

When the US is magnified to exclude Alaska and Hawaii, it becomes apparent that the farmers markets are located very heavily in the Boswash megalopolis, along the Great Lakes, and on the West Coast.

Suppose there is interest in looking at the distribution within Pennsylvania. One could consider reducing this map further, but it may be useful to plot the graph onto a more relevant map, for example Pennsylvania zip codes.

Custom map files

JMP allows for the creation of custom map files. Typically, these come from ESRI® shapefiles and SAS/Graph® map data sets[3]. In this case, the Census provides free shape files for each US state.[4] Pennsylvania is selected and downloaded. The following process, which is taken directly from JMP's website is followed to create the new JMP map file[5]:

> *To convert an ESRI shapefile to a JMP map file:*
>
> *1.Open the .shp file in JMP.*
> *2.Make sure that the Shape column is the first column in the **.shp** file. Add formatting and axis settings for the X and Y columns (optional). Graph Builder uses those settings for the X and Y axes.*
> *3.Save the .shp file as a JMP data table to the Maps folder with a name that ends in -XY.jmp.*
> *4.Open the .dbf file.*
> *5.Add a Shape ID column as the first column in the table. This column should be the row numbers from 1 to n, the number of rows in the data table (Note - You can use Cols >New Column > Initialize Data > Sequence Data).*
> *6.Assign the **Map Role** column property to any column that you use for place names in the Shape role of Graph Builder. To do this, right-click at the top of the column and choose Column Properties > Map Role.*
> *7.Choose **Shape Name Definition** from the drop-down box in the property definition.*
> *8.Save the table as a JMP data table with a name that matches the earlier table and that ends in -Name.jmp.*
> *JMP looks for these files in two locations. One location is shared by all users on a machine. This location is:*
> *•Windows: C:\Program Files\SAS\JMP\<Version Number>\Maps*
> *•Mac: /Library/Application Support/JMP/<Version Number>/Maps*

Once completed, Data Filter (under Rows Menu) can be used to select Pennsylvania (based on the state variable). The bubble plot is made in the same way, except, now, when background map is selected, "PA ZipCode," which is the name I chose for the shape file is now an option (Figure 6). From looking at Pennsylvania, it can be seen that the many of the larger markets are near Philadelphia, stretching west across the southern part of the state through Harrisburg and somewhat west. One hypothesis for this concentration is that the region has terrain more conducive to farming relative to the rest of the state, as well as a larger population of people who practice traditional farming, such as the Amish.

Figure 6: Screen Shot of New Boundaries Options

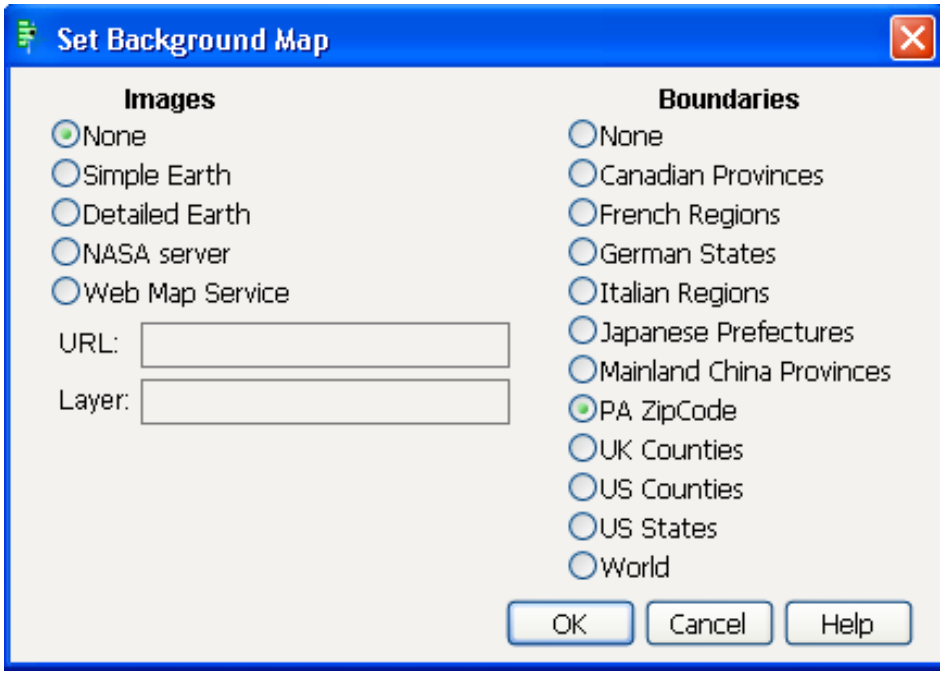
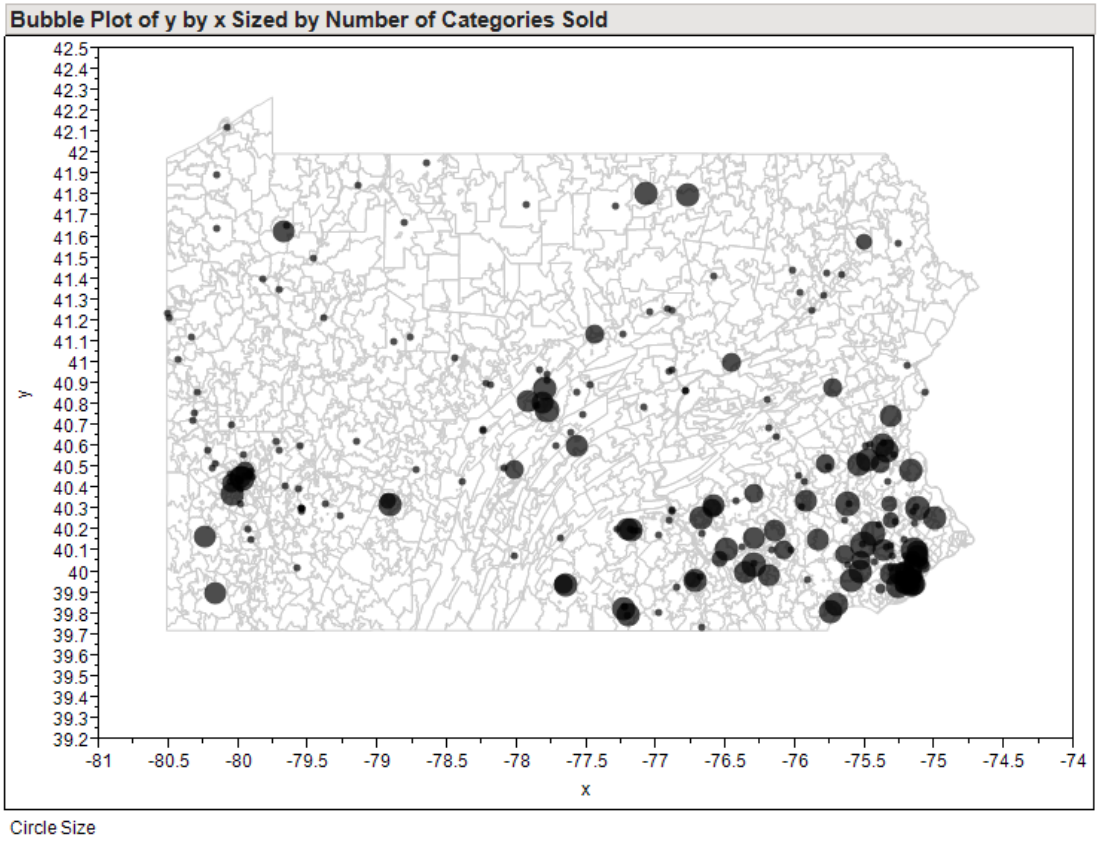


Figure 7: PA Farmers Market Distribution



The five-dimensional moving bubble plot

The previous examples have displayed data in three dimensions: latitude, longitude, and size (in the example given, number of categories). However, often times, data involves various classifications and time. This type of data can also be addressed in JMP.

The Pennsylvania farmers market data had sales simulated by market for it by date for one year. The following SAS code was used to generate these sales numbers. These numbers are purely fictitious and used for illustration purposes only. Furthermore, this code breaks the markets into three regions, "SE" for southeast, "SW" for southwest, and "N" for north.

```
%macro abc;
data dd1.farmers markets;
set %do i=1 %to 52;
dd1.PA_Farmers_Markets (in=a&i)
%end;
;
if a1 then time=18628; /* Jan 1, 2011 */
%do i=2 %to 52;
else if a&i then time=18628+%eval(7*&i);
%end;
format time mmddyy10.;
sales=10000+max(rannorm(1978)*10*_n_, rannorm(2009)*1000, -5000);
if y < 41 and x > -77.5 then area="SE";
else if y<41 and x <=-77.5 then area="SW";
else area="N ";
run;
%mend;

%abc;
```

The data can then be opened in JMP to create the five-dimensional moving bubble plot. To make this plot, again choose bubble plot and assign x and y as before, but assign sales to "Sizes", time to "Time", area to "Coloring", and MarketName to "ID" (Figure 8). It must be noted that having IDs for individual markets is critical for this method to work.

Once "OK" is selected, the new graph window will appear. A background map can be chosen as before. Here, PA ZipCode is chosen again. In the new window, there are now three bars. There is the one for circle size, as before, but two new ones for time and Speed. The time button manually changes the date. Speed changes the speed if the "play" button on the bottom is selected. The "play" button advances the sales data through time. Finally, the farmers markets are colored according to the geographical area they were assigned to. A screen shot of this is seen in Figure 9. The entire movie can be seen in the accompanying PowerPoint presentation to this paper. This is available upon request. A discussion of how to save and insert into PowerPoint can be found in the next section.

Figure 8: Screen Shot of Bubble Plot Window for Five-Dimensional Moving Plot

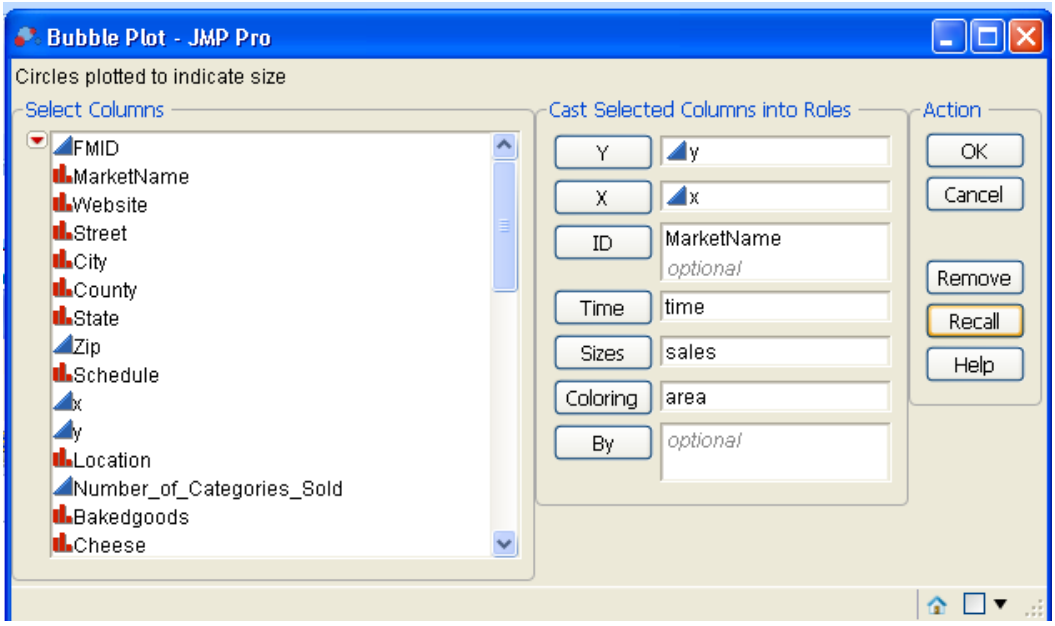
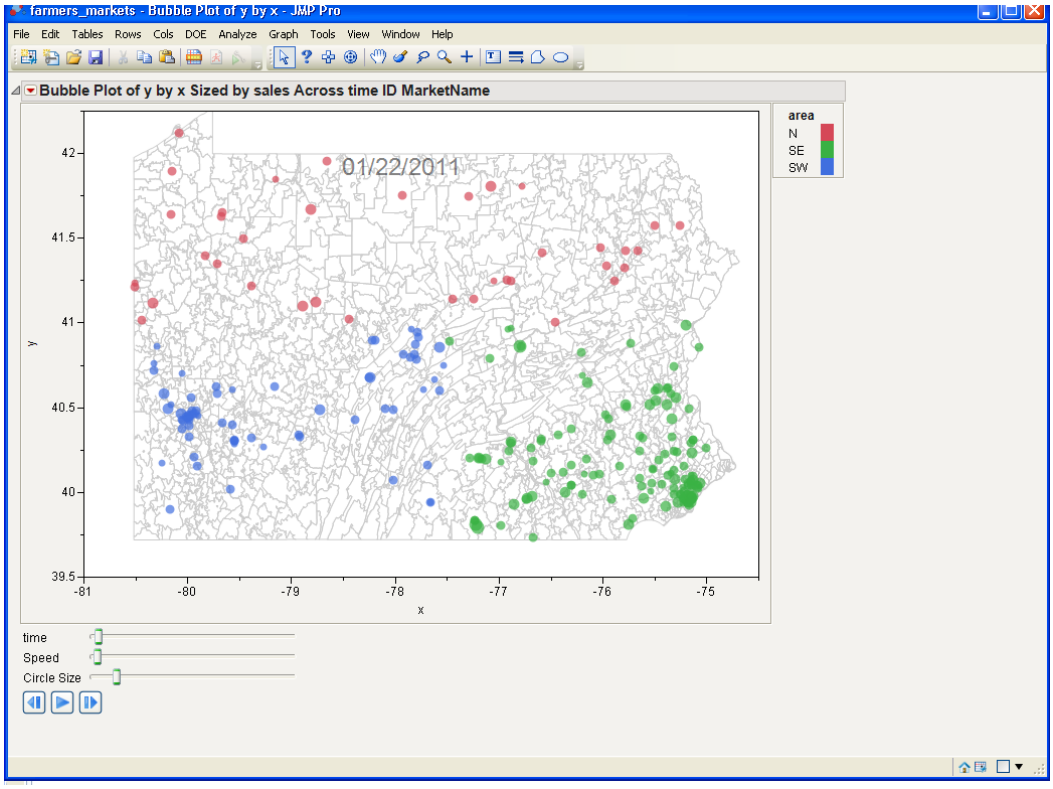


Figure 9: Screen Shot of Five-Dimensional Moving Plot



Presenting the five-dimensional moving bubble plot

While the five-dimensional moving bubble plot is useful in JMP to analyze data temporally, it is often more practically presented in PowerPoint. To present this type of chart in PowerPoint, the chart must be saved as an Adobe Shockwave Flash file. To do so, choose the red triangle in the upper left hand corner of the plot and select "Save for Adobe Flash Platform (.SWF)…". The companion HTML file is not necessary to save in PowerPoint, but offers an alternative vehicle to transfer /show the graph to others who do not have JMP on their machine.

Once the Flash file is created, the following methods can be used to insert it into PowerPoint. For any method, Flash Player should be installed. Either method requires PowerPoint to be used in Slide Show view to work.

The first comes directly from JMP's website.[6]

> *Go to the slide where you want to embed the .SWF file.*
>
> *Click the Developer tab at the top of the PowerPoint window.*
> *If the Developer tab is not showing, do one of the following:*
>
> *In PowerPoint 2010, click the File tab and select Options. In the Options window, select Customize Ribbon. In the Customize the Ribbon list, select Developer, and then click OK.*
>
> *In PowerPoint 2007, click the round Microsoft Office button at the top left corner of the window and select PowerPoint Options. In the Popular section, select Show Developer tab in the Ribbon, and then click OK.*
>
> *On the Developer tab, click the More Controls button in the Controls group.*
>
> *In the More Controls window, select Shockwave Flash Object.*
>
> *Click OK.*
> *The cursor changes to crosshairs.*
>
> *Click and drag the cursor to draw a rectangle where you want the Flash object to appear. The Flash object needs to be large enough to show the entire .SWF picture area.*
>
> *Step 4: Embed the .SWF file in the Flash Object*
>
> *A Flash object now appears on the slide. In this step, you embed the .SWF file in the PowerPoint 2010 slide (or link the file in PowerPoint 2007) and assign properties.*
>
> *Right-click on the Flash object and select Properties.*
> *The Properties window appears.*
>
> *On the Alphabetic tab, enter the full pathname of the .SWF file next to Movie (for example, C:\Users\Smith\Desktop\Flash\Plane.swf).*
>
> *(PowerPoint 2010) Set the EmbedMovie property to True.*
> *This property embeds the .SWF file in the slide, so the .SWF file is now part of the slide.*
>
> *Set the Playing property to True.*
>
> *Press F5 to view the .SWF file in Slide Show view.*
> *The Profiler, Bubble Plot, or Histogram .SWF file now appears on the PowerPoint slide.*

JMP's website also includes some useful illustrations to walk through this process and if using this method, it would be useful to view the URL (see reference 6).

The second method is to use the iSpringPro® add-in for Excel[7]. This is a useful add-in that allows inserting this type of file into PowerPoint. The add-in is not free, but certainly saves time and simplifies this process. The add-in also has many other useful features. This is the method preferred by this author.

What if there is no latitude and longitude coordinates

Through this point, all examples have assumed latitude and longitude will be provided. It is often the case in practice that the user has addresses, but not latitude and longitude coordinate. SAS added proc geocode in version 9.2 to address this issue. Proc geocode provides latitude and longitude coordinates at a variety of levels, including street level (which requires maintenance release 3 of SAS 9.2 or higher).

Specific files must be installed to use this feature, however, for street level geocoding, SAS provides a set of these files free[8].

Once these files are installed, the basic syntax of proc geocode is as follows:

```
proc geocode                          /* Invoke geocoding procedure       */
   method=STREET                     /* Specify geocoding method         */
   data=dd1.store_info              /* Input data set of addresses      */
   out=WORK.GEOCODED2               /* Output data set with X/Y values  */
   lookupstreet=dd3.usm            /* Primary street lookup data set   */
   attributevar=( Street City State Store_Name zip amt_sold n_sold) /*
variables you want to maintain   */
ADDRESSVAR=   Street
ADDRESSZIPVAR= zip
ADDRESSCITYVAR= City
ADDRESSSTATEVAR= State;
run;
```

This will output a file with latitude and longitude appended to the attribute variables specified that can then be used in JMP. There are many variants of proc geocode, and many options. From this author's experience, proc geocode is a "smart" procedure and has a tendency to match at the best levels it can. For example, if there is a missing value for addressvar, it will often match at a higher level, such as zip code.

This is a complicated procedure which as much more power than the syntax provided above. However, the goal of this paper is to make the reader aware of its existence and usefulness and not provide extensive detail into its functionality.

For greater detail on proc geocode, geocoding and files available for proc geocode , the 2010 SAS Global Forum paper, "PROC GEOCODE: Now with Street-Level Geocoding" by Darrell Massengill and Ed Odom[9] is an excellent source.

Summary

JMP offers powerful graphics options for geocoded data, including the five-dimensional moving bubble plot. In this paper, generations of that plot and the simpler three-dimensional plot have been discussed. In addition, methods for inserting these plots into PowerPoint for presentation have been detailed. Finally, proc geocode was briefly described as a method to geocode data.

References

[1] http://en.wikipedia.org/wiki/Geocoding

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

[3] http://www.jmp.com/support/help/Creating_Maps.shtml#384676

[4] http://www.census.gov/geo/www/cob/z52000.html

[5] http://www.jmp.com/support/help/Creating_Maps.shtml#360725

[6] http://www.jmp.com/support/swfhelp/en/powerpoint.shtml

[7] http://www.ispringsolutions.com/ispring-pro

[8] http://support.sas.com/rnd/datavisualization/mapsonline/html/geocode.html

[9] http://support.sas.com/resources/papers/proceedings10/332-2010.pdf

Contact Info

Please contact the author with any questions.

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

Trademark Citation