

Making Business Decisions Using SAS Mapping Technologies

Darrell Massengill, SAS Institute

INTRODUCTION

Location and geographic information have become critically important in making business decisions. SAS has a number of mapping technologies that allow you to integrate geographic information with the analytic capabilities of SAS to aid in making those decisions. This paper gives an overview of the existing the mapping technologies, highlights new 9.2 functionality and enhancements and gives two examples of using the mapping technologies.

OVERVIEW OF SAS MAPPING TECHNOLOGIES

SAS Mapping Technologies include technologies that are built into SAS and third-party technologies that are integrated into SAS. The third-party integration is using several ESRI technologies including their ArcGIS Desktop product, their ArcGIS Server product and their ArcGIS engine product The SAS technologies include SAS/Graph mapping with Proc GMAP and SAS/GIS.

ESRI INTEGRATION TECHNOLOGY

There are currently three areas of integration between ESRI and SAS: SAS Bridge for ESRI, SAS Map Component and Enterprise Guide.

SAS Bridge for ESRI

SAS Bridge for ESRI is integration between ArcGIS Desktop and SAS that allows the ArcGIS user to read and write SAS data and run SAS programs from within the ArcGIS software. Figure 1 illustrates the results of viewing SAS data on an ArcGIS map, selecting some of the data on the map and then running a SAS program on the selected data to generate the barchart. For more information, see: <http://www.sas.com/products/bridgeforesri/>

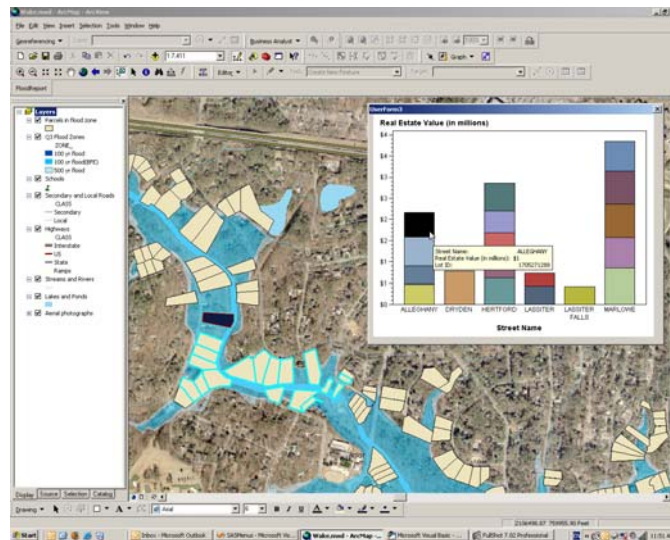


Figure 1 – SAS Bridge for ESRI

ESRI Map Component

The ESRI Map component is integrated with ESRI's ArcGIS Server and SAS. This map component was added to SAS Web OLAP Viewer in SAS 9.1.3SP2 and added to SAS Web Report Studio 3.1 (post-9.1.3SP4). Both products deliver information through a web browser with zero download. Figure 2 and Figure 3 show SAS Web OLAP Viewer and SAS Web Report Studio with OLAP data in the table and the ESRI Map component. For more information, see: http://www.sas.com/technologies/bi/query_reporting/webolapviewer/ and http://www.sas.com/technologies/bi/query_reporting/webreportstudio/

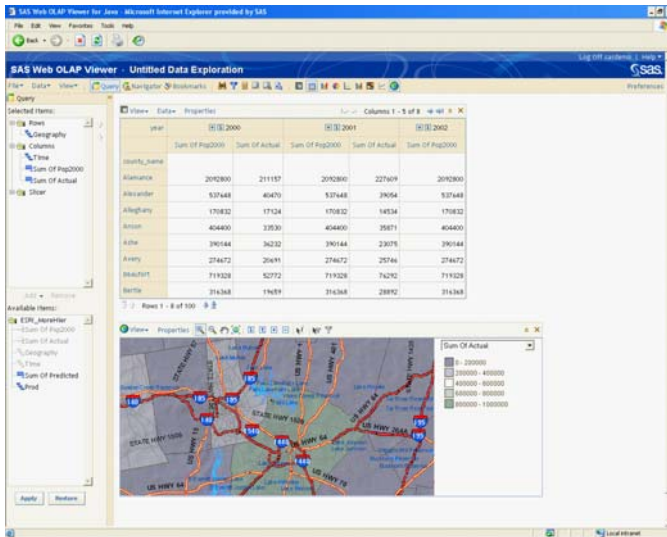


Figure 2 – SAS Web OLAP Viewer

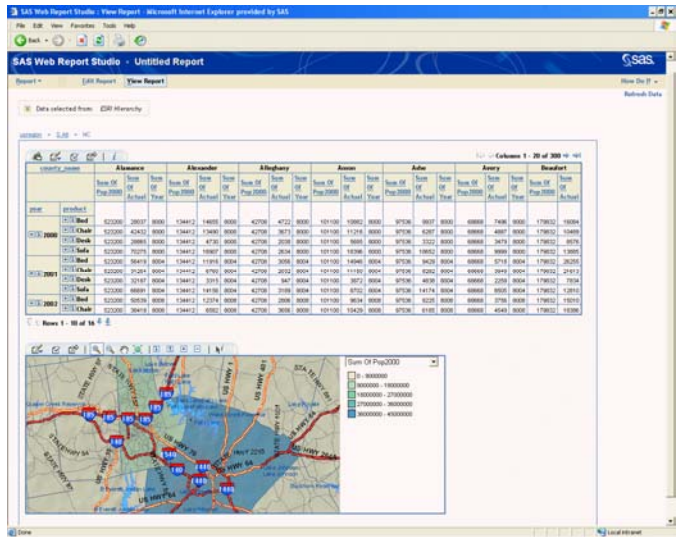


Figure 3 – SAS Web Report Studio

Enterprise Guide

The ESRI map component was added to Enterprise Guide in 4.1 (post 9.1.3SP4). Enterprise Guide interfaces with the ESRI ArcGIS Engine. Figure 4 illustrates a table and an ESRI map. For more information, see: http://www.sas.com/technologies/bi/query_reporting/guide/

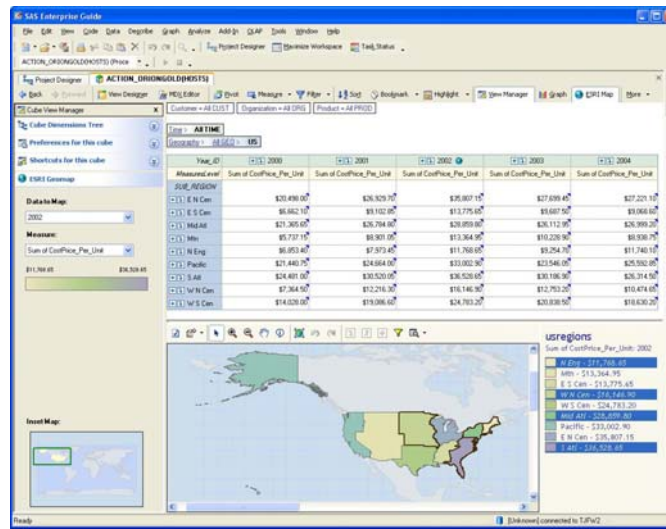


Figure 4 – Enterprise Guide

SAS MAPPING TECHNOLOGY

There are a number of technologies in SAS related to mapping, but the primary way of creating and displaying maps in SAS are SAS/GIS and Proc GMAP in SAS/GRAPH.

SAS/GIS

SAS/GIS software provides an interactive desktop Geographic Information System (GIS) within the SAS System. SAS/GIS software provides a variety of tools, features, and capabilities that are specific to the business geographic market and can be used to build powerful decision support systems. Figure 5 shows the SAS/GIS desktop interface. For more information, see: <http://www.sas.com/products/gis/>

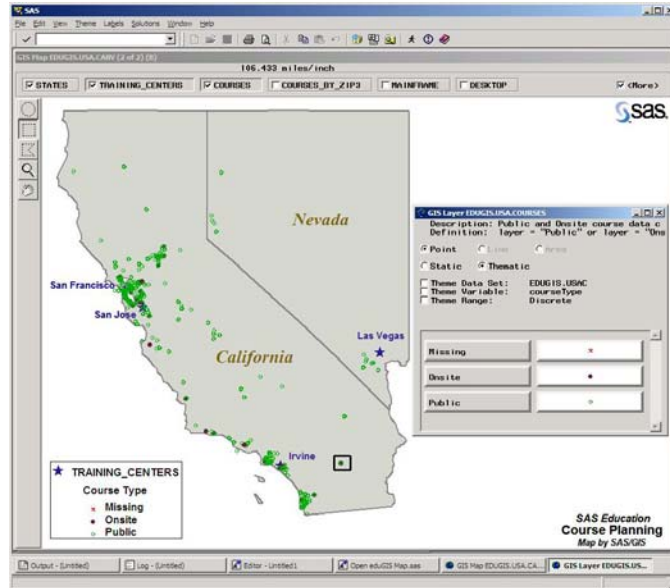


Figure 5 – SAS/GIS

SAS/GRAPH - Proc GMAP

The SAS/Graph mapping functionality allows you to create choropleth, prism, block and surface maps within the SAS System. For product information, see: http://www.sas.com/technologies/bi/query_reporting/graph/index.html. The SUGI presentations in the Resources section below contain more information on SAS/GRAPH mapping. Figure 6 below illustrates an elevation map created with Proc GMAP and Annotate.

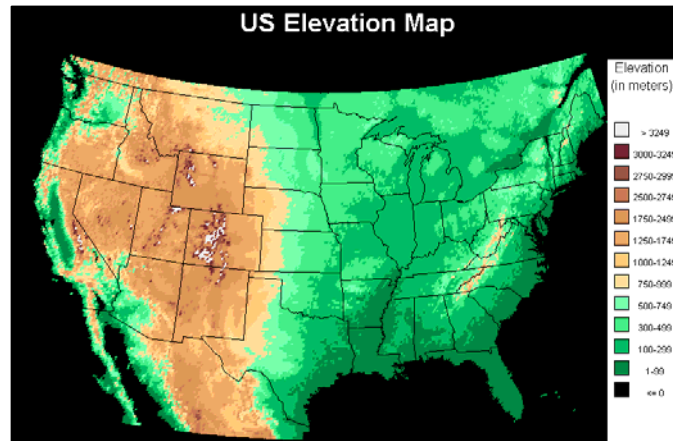


Figure 6 – Proc GMAP

NEW FUNCTIONALITY AND ENHANCEMENTS IN SAS 9.2

SAS 9.2 will contain a number of new procedures and functions for use with your mapping data. In addition, there are enhancements to existing mapping functionality. Below is an overview of key items.

NEW FUNCTIONALITY

Proc GINSIDE

Proc GINSIDE will be new in SAS 9.2. The procedure will compare a dataset with point data to a map data set and identify which polygon contains each point. For example, if you had a list of customers and their location, it could identify which county they were in.

```
PROC GINSIDE MAP=map DATA=points OUT=output ;
  ID idvar idvar2;
```

MAP= Contains the map polygons.
DATA= Contains a list of points.
OUT= Returns the point data with the 'ID' of the polygon.

Proc GEOCODE

Proc GEOCODE converts address information to map locations. This is experimental in 9.2.

```
PROC GEOCODE
  <DATA=address-data-set>
  <ADDRESSCITYVAR=city-var>
  <ADDRESSSTATEVAR=state-var>
  <ADDRESSZIPVAR=zip-var>
  <ATTRIBUTEVAR=attribute-var-list>
  <FORCE>
  <LOOKUP=zip-centroids-data-set>
  <LOOKUPCITYVAR=city-var>
  <LOOKUPSTATEVAR=state-var>
  <LOOKUPXVAR=x-coordinate-var>
  <LOOKUPYVAR=y-coordinate-var>
  <LOOKUPZIPVAR=zip-var>
  <NOCITY>
  <NOSTIMER>
  <OUT=output-data-set>
  <ZIP| ZIP5>;
```

ZIPCITYDISTANCE Function

The ZIPCITYDISTANCE function will be new in SAS 9.2 and allows the user to get a straight-line distance between two ZIP Codes. The ZIP Code distance will be calculated using the coordinates from the ZIP Code centroid file that ships in SASHELP.

```
distance = ZIPCITYDISTANCE(zip1, zip2);
```

GEODIST Function

The GEODIST function will be added in SAS 9.2 and allow the user to get a straight-line distance between any two coordinate pairs.

```
distance = GEODIST(lat1,long1,lat2,long2 [,options])
```

Options:

- D - lat/long are in degrees (default)
- R - lat/long are in radians
- M - distance given is in miles (default)
- K - distance given is in kilometers

ZIPCITY Function

The ZIPCITY function will be added in SAS 9.2 and returns the city for a given zipcode. The returned information is based on the zipcode dataset shipped in SASHELP.

```
city = ZIPCITY(zipcode);
```

ENHANCEMENTS

Proc GMAP

New statements and options will be available for Proc GMAP in 9.2:

- AREA statement
- UNIFORM option
- DENSITY option
- CDEFAULT option

- STRETCH option
- RELZERO option
- Statistics

Proc MAPIMPORT

Proc MAPIMPORT is enhanced in SAS 9.2. The proc imports ESRI Shapefiles into SAS Map data sets. An ID statement was added in 9.2 to force a variable to be the ID for the map polygons.

SAS Bridge for ESRI

In 9.2, SAS Bridge for ESRI will support running BI Stored Processes from within ArcGIS desktop.

SAS Web OLAP Viewer

- AJAX support
- Display of crossing information
- Improved Bookmarking
- Rectangular selections

SAS Web Report Studio

- AJAX support
- Conditional Highlighting
- Execute Stored Processes (from selections)
- Multiple Maps in a section
- Rectangular selections
- Report Linking

MAKING BUSINESS DECISIONS WITH MAPPING TECHNOLOGY

Mapping Technologies allows you to examine and explore your location-based data visually. You can see patterns, trends and influences that are not otherwise apparent. Below are two simple examples using maps to explore data for making business decisions.

CONFERENCE LOCATION

Company X wants to know if the location of their user conference has an impact on which users will attend. The company tends to alternate between the East and West Coast and they want to know if there is a shift in users who attend based on which coast the conference is on.

A first step in answering this question is to map where the users come from for each conference over a period of years. The logic to do this is fairly straightforward:

- Get the zip code for each conference attendee.
- Get the map location of the zip codes/attendees by using Proc GEOCODE (experimental in 9.2).
- Because multiple people attend from a single zip code, the attendee location data needs to be randomized into a dot-density map. This is done using the fuzz macro. (This macro is available with the 'Cheap Geocoding' paper listed in the Resources section below).
- Convert the attendee location data into an Annotate dataset.
- Display the US map with the customer locations using Proc GMAP.

These steps were run over many years of data; the results from the last two years are shown in Figures 7 and 8.

At first glance, it is difficult to see much difference in the two maps. With a side-by-side comparison, it appears that there may be more attendees from the northern east coast for the East coast conference and more California attendees for the West coast conference. Unfortunately, it requires careful study of a side-by-side comparison to determine these differences.

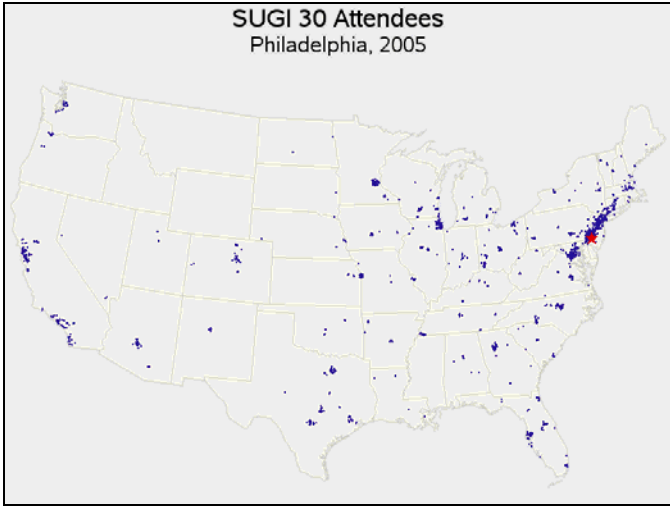


Figure 7 – SUGI 30 Attendee Location (East Coast)

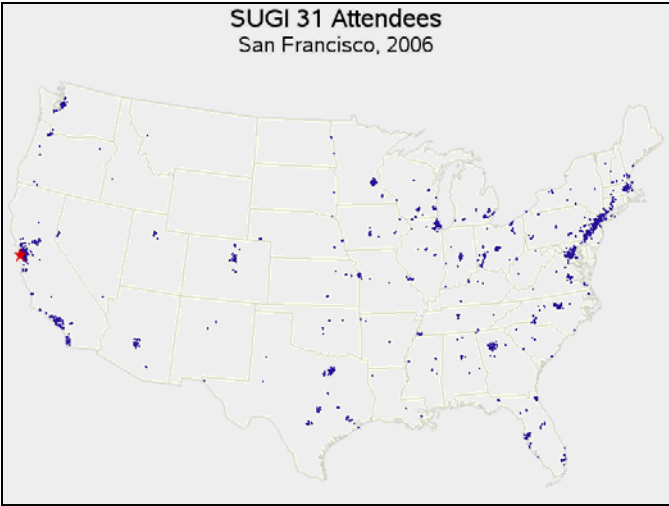


Figure 8 – SUGI 31 Attendee Location (West Coast)

In order to increase readability and to understand the magnitude of the shifts, the program will need changes for second step. The change will mean showing green dots on the map when a zip code has increased in attendees from the previous year and showing red dots when it has decreased. Blue dots mean the attendees remained constant. The changes needed are:

- Sort the data by zip code and year.
- Process the data and compare the number of attendees in a zip code to the number of attendees for the prior year and create a 'difference' value. The difference is negative if there was a reduction in attendees.
- When creating the annotate dataset, examine the 'difference' value. Set the color to RED if negative, GREEN is positive and BLUE if no change.

Figures 9 and 10 show the results for two years. As can be clearly seen, there is a coastal shift along the coastal edges of the country depending on the location of the conference. The changes in the middle of the country are less dramatic.

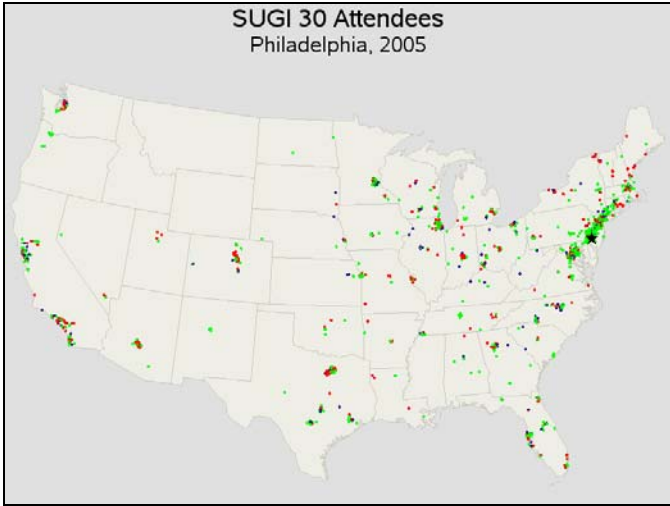


Figure 9 – SUGI 30 Change from last year

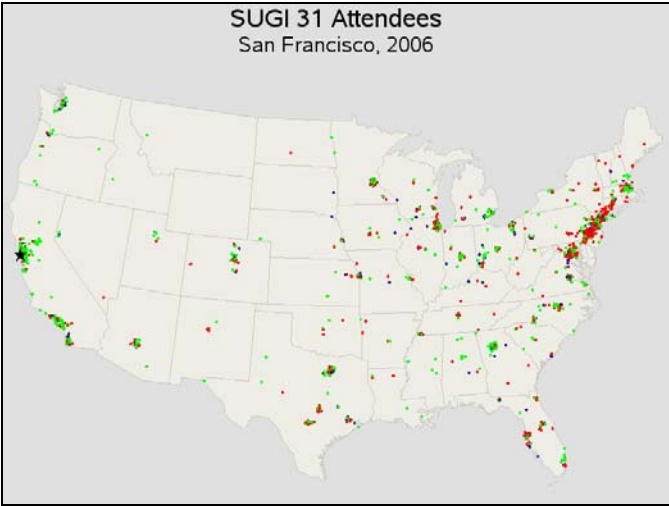


Figure 10 – SUGI 31 Change from last year

SALES ANALYSIS

Company Z sells widgets all over the Continental US. Sales require a face to face visit. The company has nine sales regions and they track sales and sales expenses at the county level. Company Z now wants to examine their sales and improve their profitability. Because of the costly face to face visit, they may need to eliminate sales in some areas. Note that the analysis and solution presented here is extremely limited due to time constraints but gives an idea of what is possible.

The company's sales data is geographically hierarchical in nature. They have regions that are made up of states and states that are made up of counties. This hierarchy makes the data a natural fit for an OLAP cube. They are using SAS Web OLAP Viewer (SWOV) as their mapping technology. Their OLAP cube is linked to an ESRI map. The levels in the data are mapped to layers in the map: region sales data to region map layer, state data to state map layer and county data to county map layer. Figure 11 shows the OLAP cube in SWOV. Figure 12 shows the map associated with the region view of Sale data and Figure 13 shows the region view of expense data. Note that the map and the cube can also be drilled and expanded. The dropdown box on the upper right of the map allows you to switch between the sales and expense views of the map.

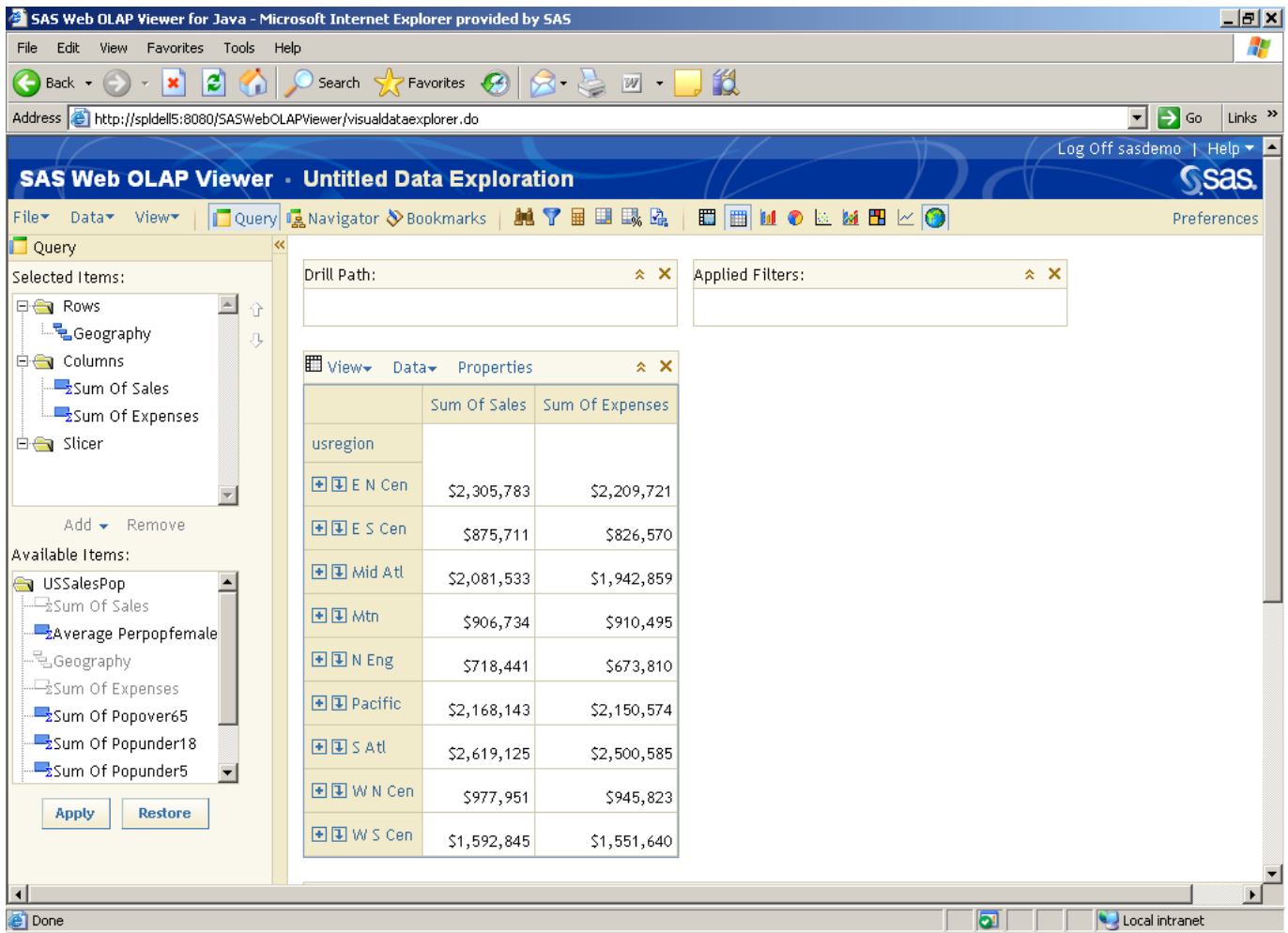


Figure 11 – SAS Web OLAP Viewer – Company Sales and Expenses

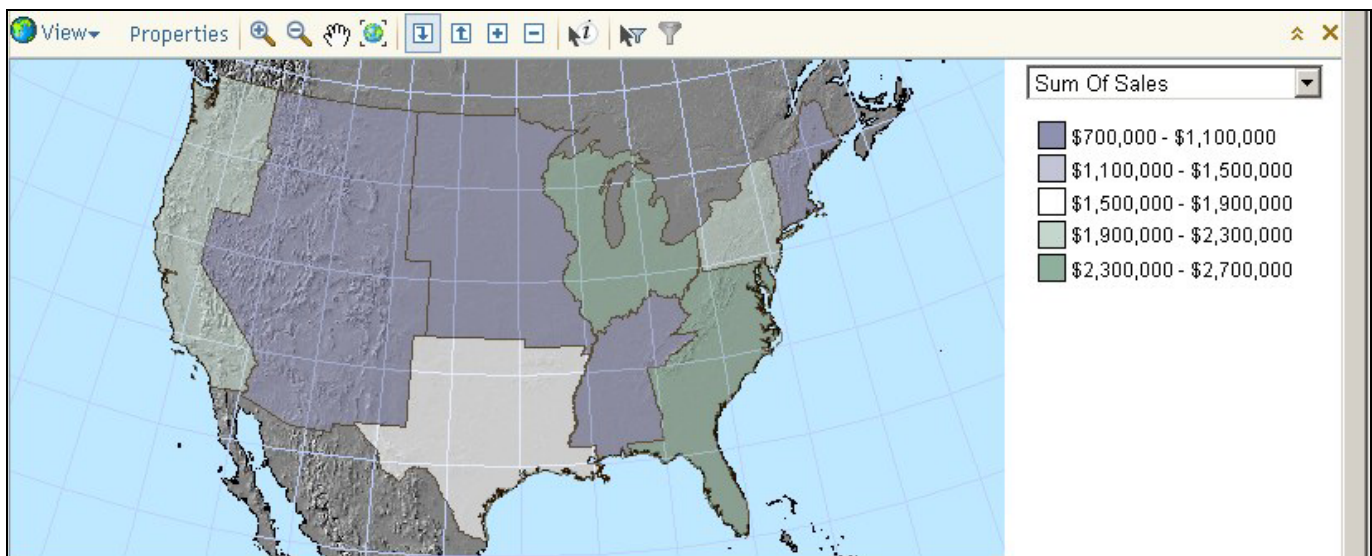


Figure 12 – Sales by Region

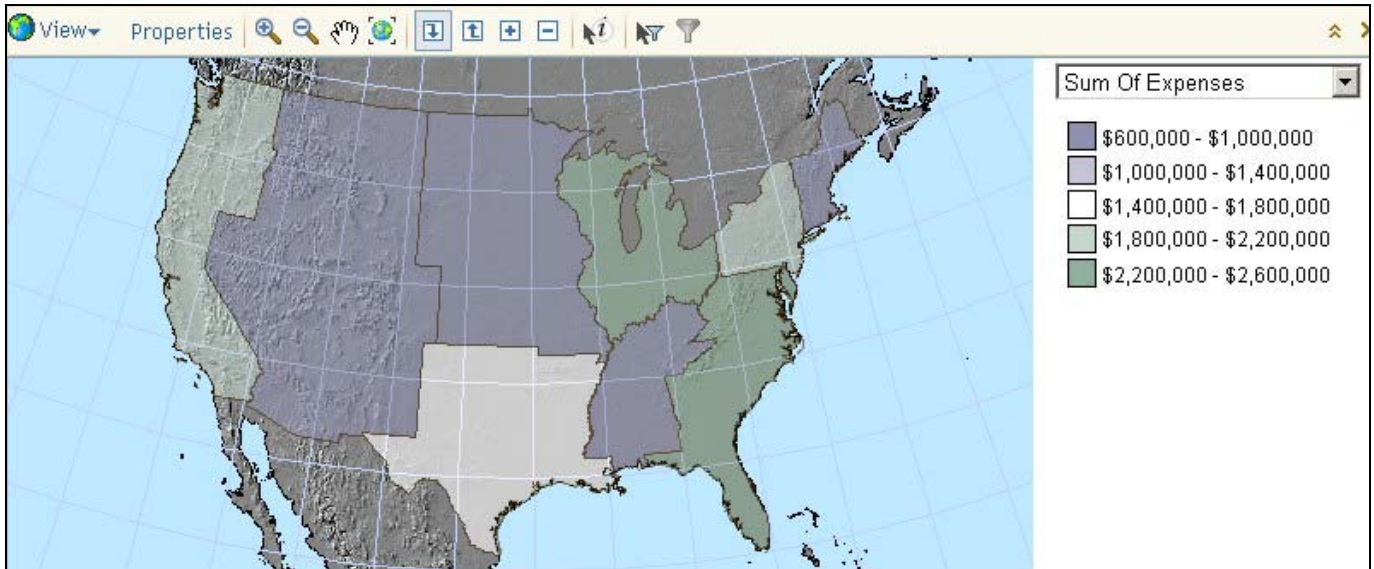


Figure 13 – Expenses by Region

Not surprisingly, the two maps look identical with the exception of the values in the legend. Expense increases as sales increases. This data doesn't help us improve our profitability. Our data doesn't include a column for profit. This data does include some other information that is currently not used and one of those values is the population of each county. Fortunately, SWOV allows the creation of calculated measures. Adding a few calculated measures will help us to analyze our data further. One important measure is profit. Several other measures, Profit per Person and Expenses per Person are calculated based on the population at the county level, but are not used in the example due to time. Figure 14 shows the Measure window. Figure 15 shows the new values calculated in the data based on the new measures.

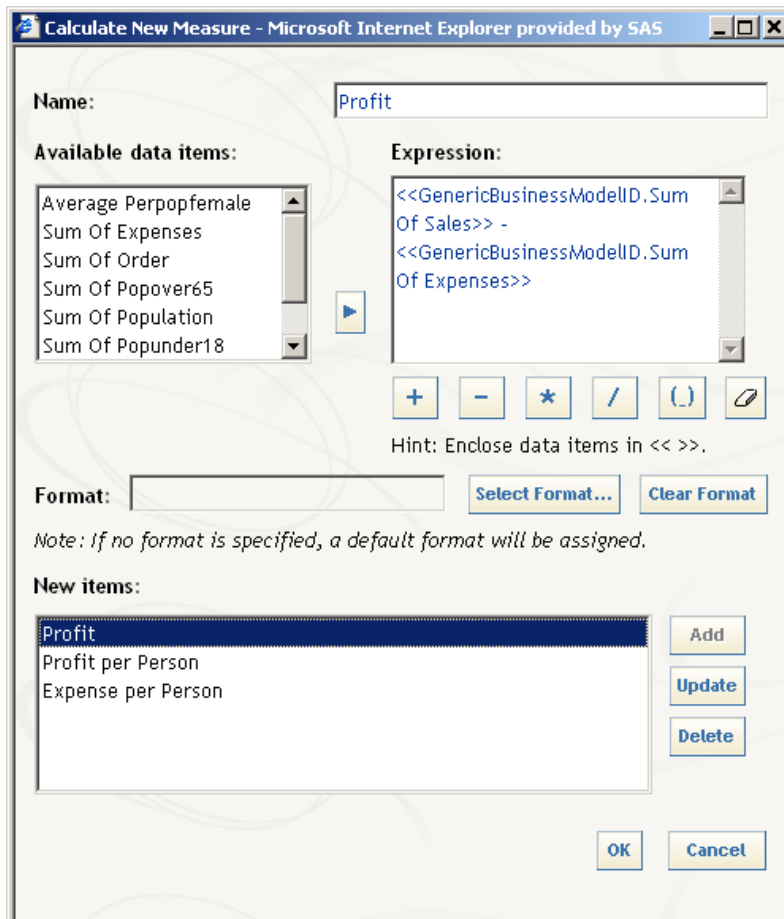


Figure 14 – Calculate New Measures

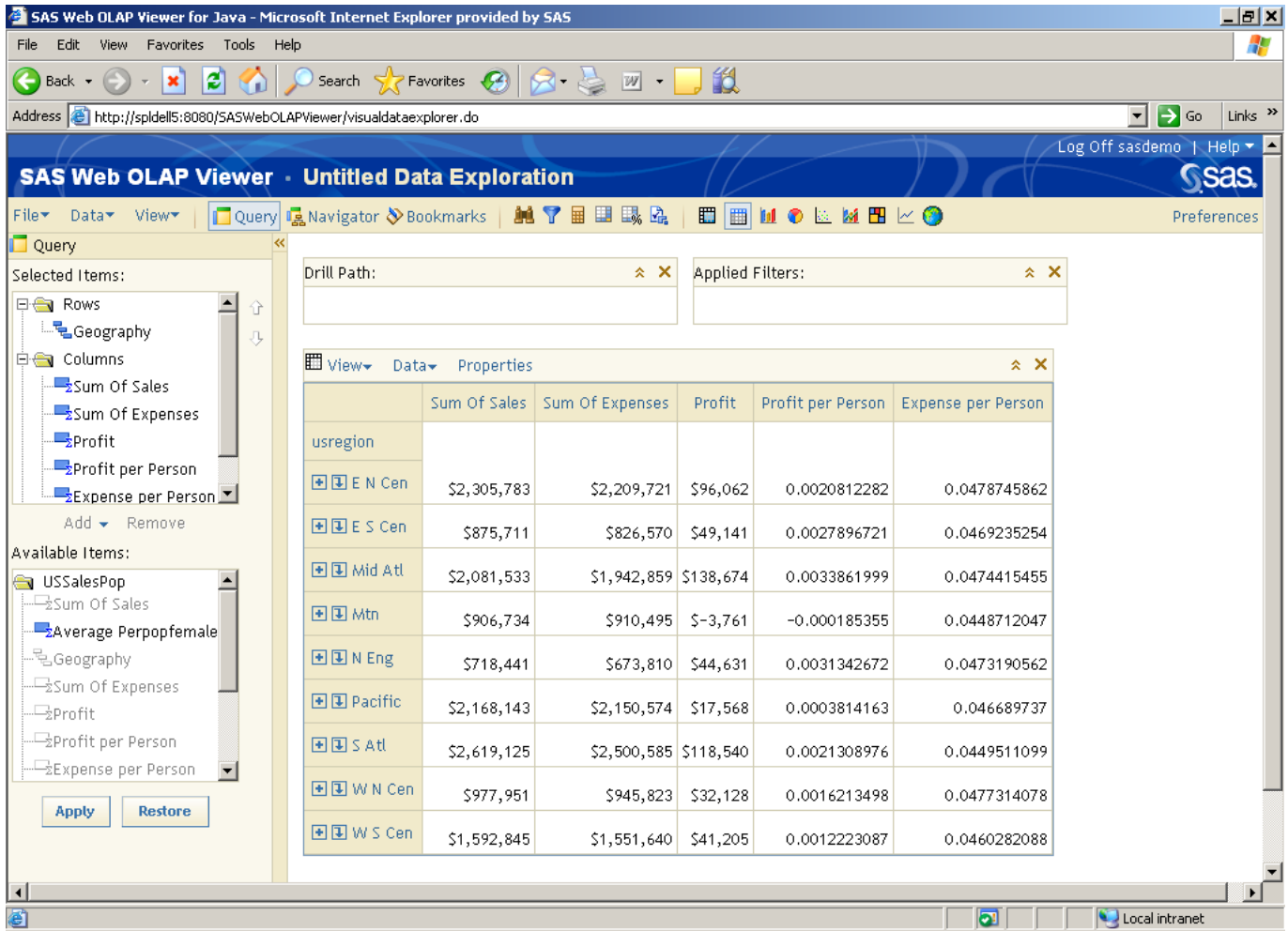


Figure 15 – Table with New Measures

We can now use the new profit measure to look at the profitability of each region. Figure 16 shows that the profit is good on the east coast and bad in the western part of the country. Looking at the table in Figure 15, you can see that the “Mtn” region needs further examination. If you drill on that region by clicking on the map, you will see the states that make up that region in Figure 17.

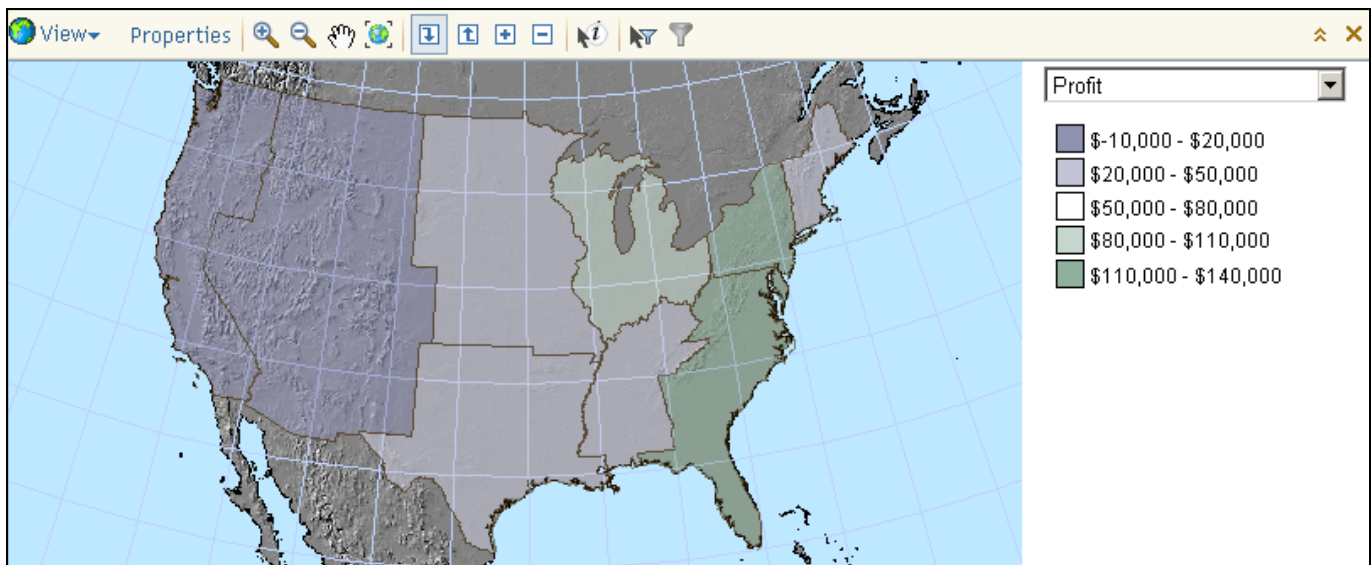


Figure 16 – Profit of Regions

If you just examine sales for this region, Colorado looks like it is doing well. If you examine the Profit for each state, you will see that the company is losing money in that state (Figure 18). Colorado needs to be examined to better understand which parts of the state are profitable and which are not.

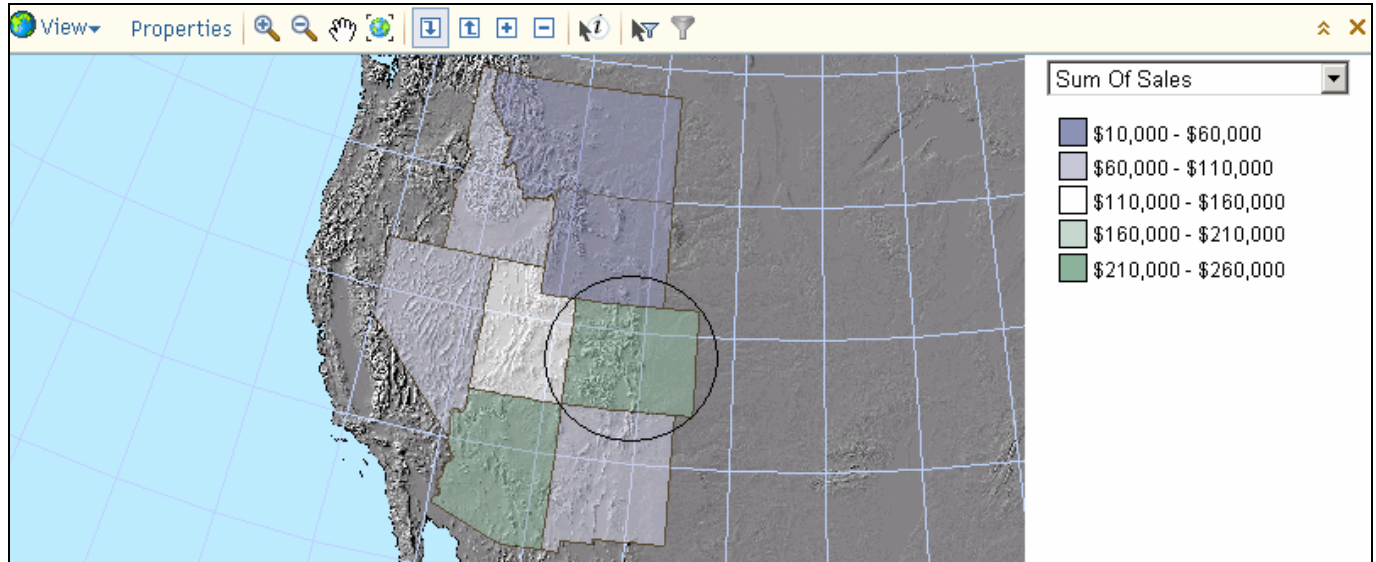


Figure 17 – Sales of Mountain Region by State

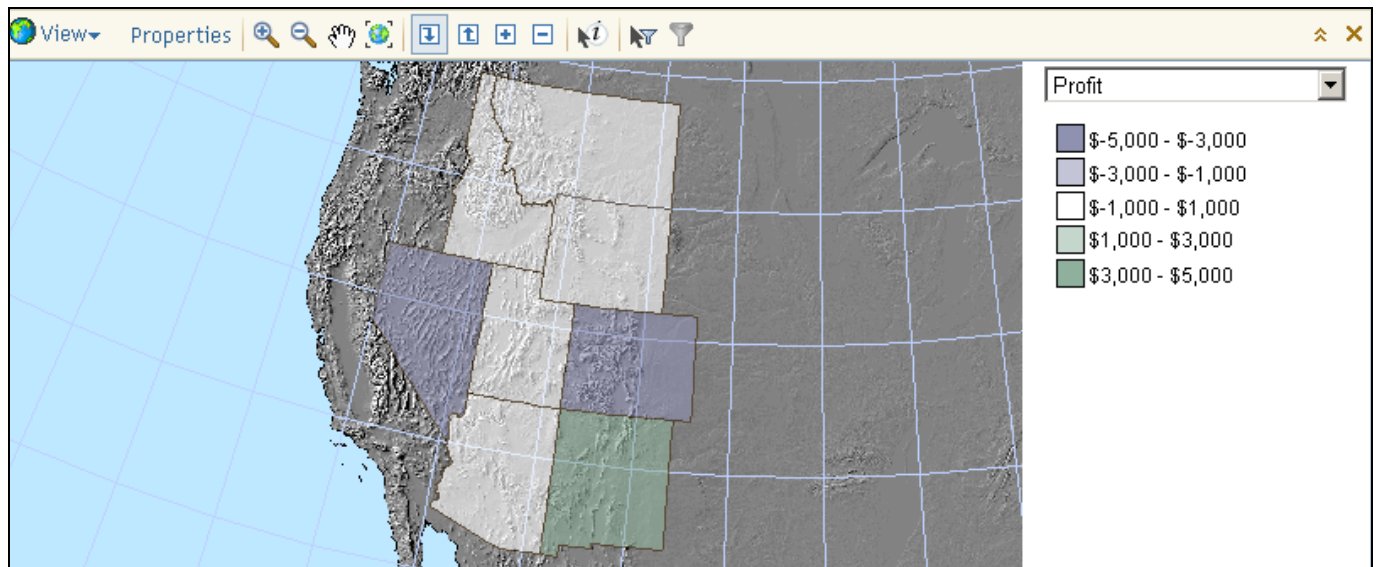


Figure 18 – Profit of Mountain Region by State

Drilling into Colorado will give you the county view of profit by county. Figure 19 shows that most counties are not profitable. The four most profitable counties are highlighted with the red line. Company Z would like to know what factors impact the profitability. Examining other data for these counties will help to uncover possible factors. The goal would be to increase sales and decrease expenses. When compared to the sales (Figure 20) you will note that two of the profitable counties are also the highest sales. However, there are two other counties with high sales and low profitability. High expense is a factor for these two counties and Company Z needs to determine why expenses are so high there.

One item that might impact sales is the population of the county. Higher sales are more likely where the population is high and the sales person can reach more people with less travel. When comparing the population in Figure 21 with the sales in Figure 20, you will see that higher sales are in higher population areas. An item that may increase expenses would be related to travel and travel time. Note that many of the unprofitable counties are to the western part of the state where the mountains are a factor. This could mean that transportation for face-to-face sales is just too expensive in these areas. Note that one county on the western edge of the state is profitable. This might be explained by Figure 22 which shows that an Interstate highway runs through that area, making it easier for the sales person to reach this location.

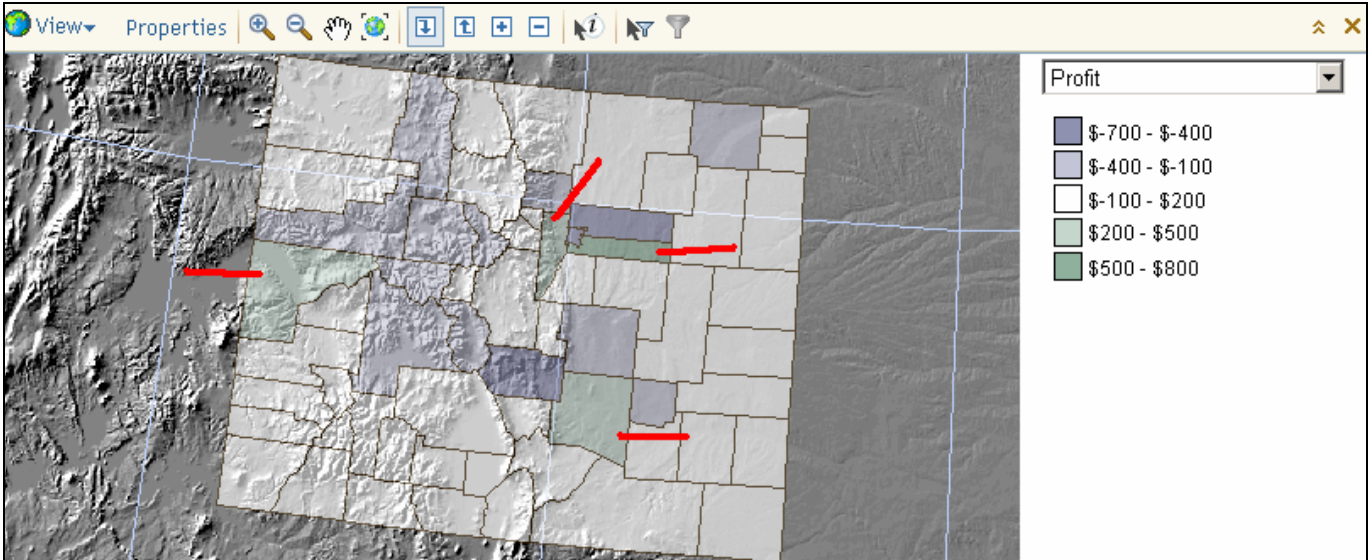


Figure 19 – Colorado Profit by County

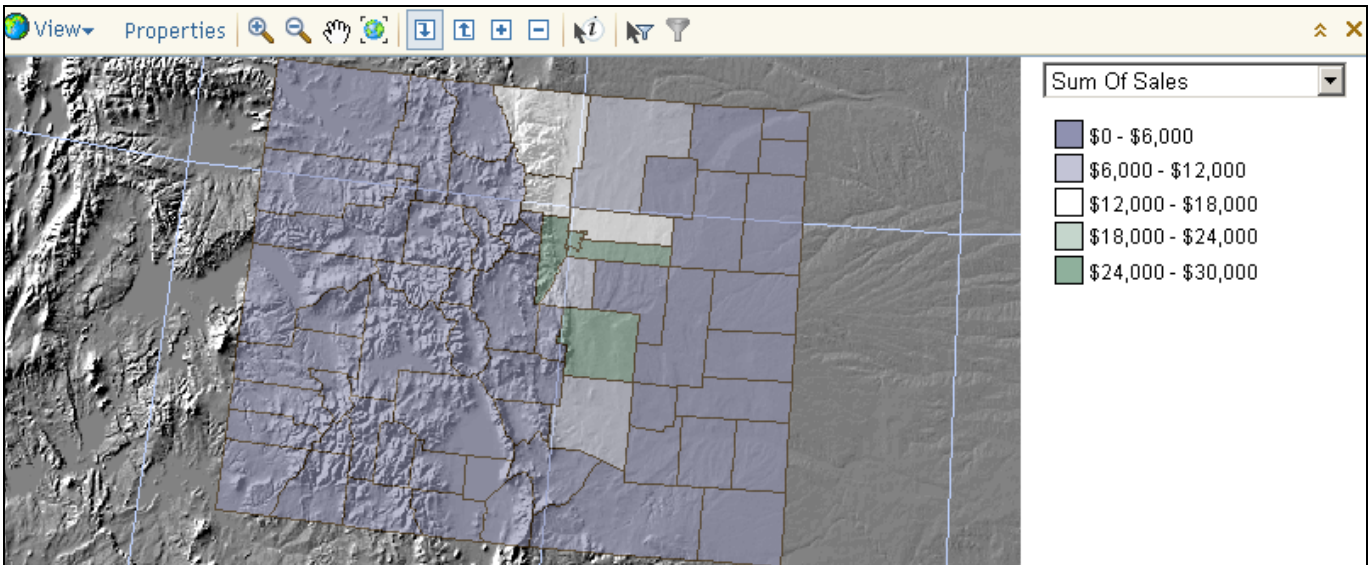


Figure 20 – Colorado Sales by County

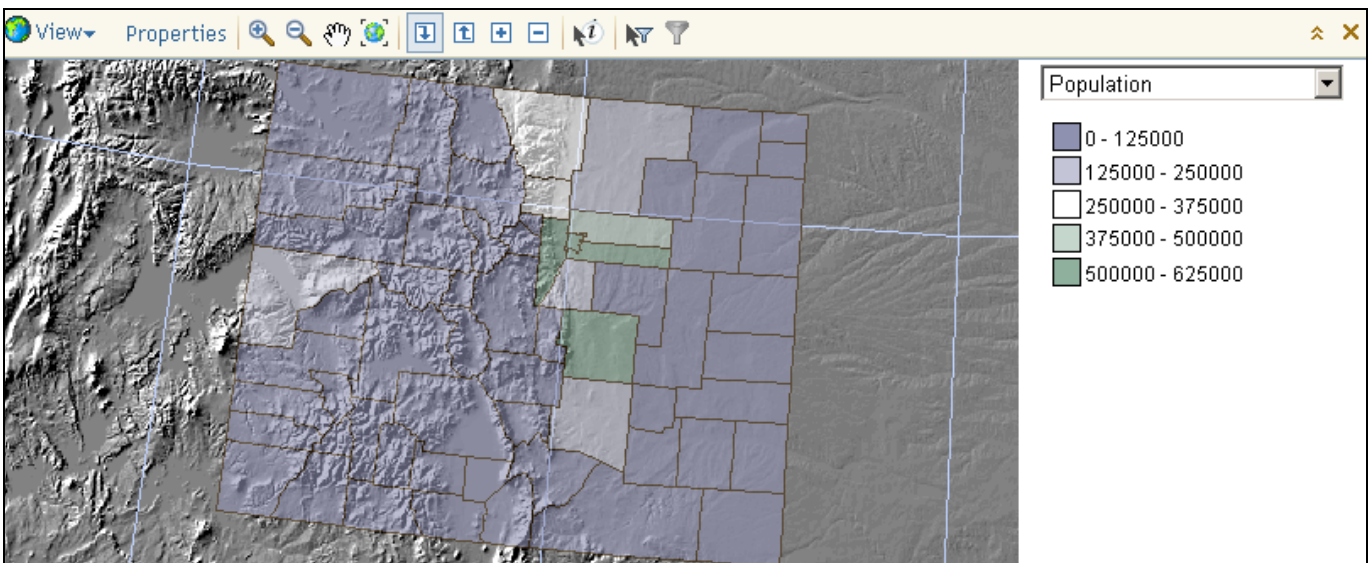


Figure 21 – Colorado Population by County

Code Samples and Technical Tips (<http://support.sas.com/sassamples/index.html>).

Maps OnLine. <http://support.sas.com/maponline>

SAS Bridge for ESRI documentation: <http://support.sas.com/rnd/datavisualization/BridgeForESRI/V2/>

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.