Paper SAS021-2014

More Than a Map: Location Intelligence with SAS® Visual Analytics

Falko Schulz and Anand Chitale, SAS Institute Inc., Cary, NC

ABSTRACT

More organizations are understanding the importance of geo-tagged data and the need for tools that can successfully combine location data with business metrics to provide intelligent outputs that are beyond a simple map. SAS® Visual Analytics provides a robust and powerful platform for achieving location intelligence performed with a combination of SAS® Analytics and GIS mapping technologies such as that offered by Esri. This paper describes the essentials for achieving location intelligence and demonstrates with industry examples how SAS Visual Analytics makes it possible.

INTRODUCTION

Mankind has known the importance and benefits of utilizing location for ages. Cavemen mapped hunting grounds and developed hunting strategies, Romans used maps to identify the best logistics to build super cities.

With maturing business intelligence (BI) technologies and users, visualization of information in the best-fit format is a given requirement. Maps are rapidly becoming the default visualization of business data associated with location information in reports and dashboards. There are many forms of user needs when it comes to map display, and their capabilities range from the simplest format of static map visualization to a highly visual, interactive, and intelligent mapping output. Toward the extreme end of it comes location intelligence, which not only shows a highly visual map, it brings much additional contextual information generated in a combination of business data and location information, including factors such as events, weather conditions, and social demographics. This combination brings a more holistic view of information so the users can make more informed decisions.

THE “WHERE” OF INFORMATION

Businesses are run across locations, and events take place everywhere. Location is a central factor for every business and covers numerous associations with location—a customer’s address, a store’s location, sales territory, tapped and untapped markets, logistics, and so on—and having an awareness of these is critical for a successful business. It helps generate new opportunities and make more informed decisions.

A typical BI system largely handles the queries around What, When, and Who. Such queries include:

- What is our current sales revenue?
- When were the profits higher?
- Who are the top customers by revenue?

When business data covering these queries are combined with geo-spatial information, it allows for new analyses adding the Where information and leading to new discoveries in the enterprise data.

Some queries answered by bringing geo-information include:

- Where do we have more sales?
- Where are the profit-making customers located?
- Where are the new markets popping up?

This especially becomes important in today’s big data world with enormous amount of data coming from many sources and hiding in themselves a lot of location information. It brings the possibility to merge data from any source with a common spatial reference and, when utilized with advanced map-based visualization techniques, this brings vast improvements over standard charts like pie and bar charts, leading to findings in a shorter time and less effort than what typically would require slicing and dicing of information.

In a standard graphical or tabular report format, different reports representing customers, competitors, and demographic information might not tell much separately, but when geocoded and overlaid on a map, one can quickly see where the best customers are, where they are in relation to competitors, and regions that provide the most market potential based on underlying demographics.

SAS® Visual Analytics brings the advantage of exploring large amounts of data using a highly visual and interactive mapping format in combination with other analytical and visualization capabilities, taking BI beyond the boundaries of graphs and tables.

SAS VISUAL ANALYTICS TECHNOLOGY OVERVIEW

SAS Visual Analytics is a new BI solution that uses intelligent ways to help business analysts and nontechnical users visualize data, see patterns and trends, and identify opportunities for further analysis, backed by the power of SAS® Analytics, which is made available to the users in a self-service and approachable manner. It enables the creation and dissemination of dashboards, reports, and the results of investigative exploration either to the web or out to native mobile applications.

Figure 1. Business Visualization Driven by Analytics

SAS Visual Analytics delivers a number of capabilities running on an enterprise class SAS® 9.4 platform that makes SAS Visual Analytics an end-to-end solution for all user needs. It includes the capability to prepare data prior to making it available to end users. It delivers an interface for exploring your data, something often known as data discovery, and it provides an interface for building out highly interactive and visual reports and dashboards. SAS Visual Analytics also includes native applications running on Android devices or the iPad for the delivery of information to augment the web.

All of this is enabled by the revolutionary and unique SAS® In-Memory Analytics engine, SAS® LASR™ Analytic Server, which sits beneath the surface and can scale from handling a few hundred megabytes of data to many terabytes. In other words, this is a solution that can start small and grow with you.

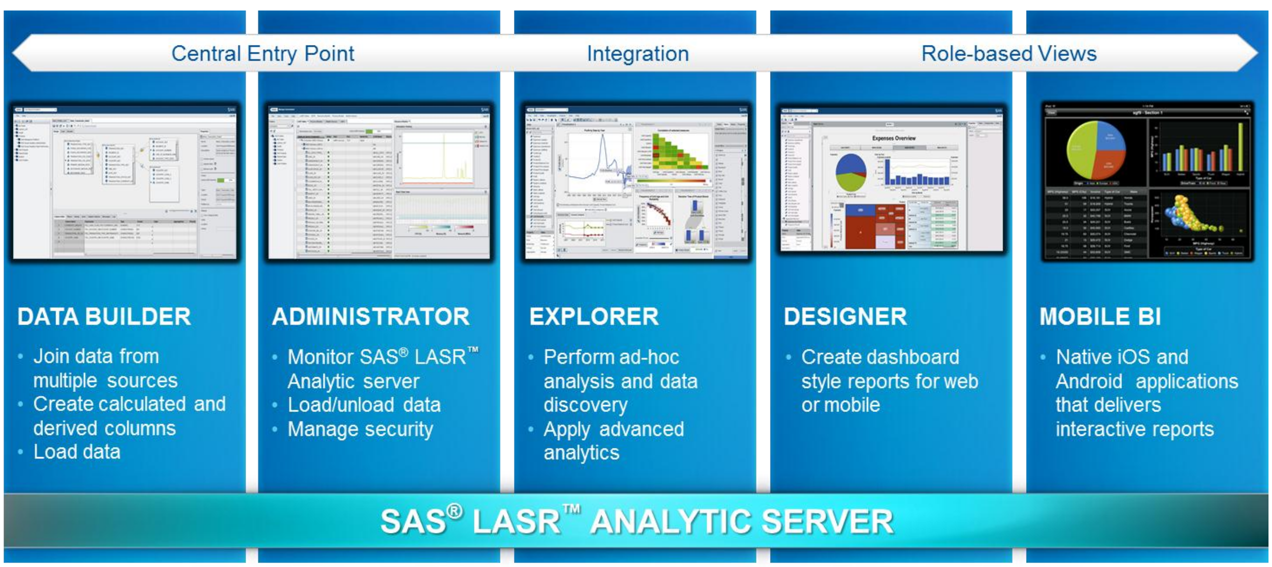


Figure 2. SAS Visual Analytics Capabilities Delivered in a Single Solution for Faster and Smarter Decisions

INTEGRATION WITH MAPPING TECHNOLOGIES

SAS Visual Analytics provides geo-mapping capabilities that are based on integration with two mapping technologies.

OpenStreetMap → OpenStreetMap is based on an open-source project, which is built by a community of mappers that maintain data about locations of all possible mapping entities such as roads, boundaries, trails, shops, and much more from all over the world.

Esri Maps → Esri ArcGIS map services provide an advanced mapping platform for organizations to leverage the highly interactive and informative geographical maps that can be used alongside business data.

Leveraging the above two technologies, SAS Visual Analytics allows its users to view their enterprise data mapped across the various locations on the map.

To be able to leverage the maps from the aforementioned mapping technologies, you need to configure the SAS Visual Analytics environment and point it to the relevant mapping technology. OpenStreetMap server is hosted by SAS and is available as a default configuration with SAS Visual Analytics. Organizations can also choose to host and maintain their own OpenStreetMap server. Organizations can also use an Esri server (ArcGIS for Server, version 10.1 or higher) to get access to maps and display in SAS Visual Analytics.

For details about configuring the map servers, please read the *SAS Visual Analytics 6.3: Administration Guide*.

BASIC DATA REQUIREMENTS

SAS Visual Analytics uses lookup values to match category values in your source data to geographical features. Using these lookup tables will help to map common locations to a geographical map without the need of performing manual geocoding steps. Lookup tables are available online at http://support.sas.com/va63geo.

In some cases you will need to geocode your data before loading into the SAS Visual Analytics. Geocoding is the process of adding geographic coordinates (latitude and longitude values) to an address. This process provides a way to convert address data into map locations. The geographic coordinates typically represent the center of a ZIP code, a city, an address, or any geographic region. After geocoding, the coordinates can be used to display a point on a map or to calculate distances.

As a basic example we have covered here the steps involved in geocoding U.S. cities. The online documentation for SAS/GRAPH® software contains more details and further examples how to geocode other geographical levels.

We assume the following data set containing U.S. cities:

```
data cities;
  infile datalines dlm=',';
  length city  $ 24 state $ 2;
  input city state;
datalines;
Atlanta, GA
Bedminster, NJ
Cary, NC
Chicago, IL
Dayton, OH
Des Moines, IA
Detroit, MI
Hartford, CT
Phoenix, AZ
San Antonio, TX
San Francisco, CA
Seattle, WA
;
run;
```

In order to geocode, we make use of the GEOCODE procedure, available as part of SAS/GRAPH:

```
proc geocode method=city data=cities out=cities_geocoded;
run;
```

The resulting data set thus contains the geographic coordinates that can be used to define a custom geo data item within SAS Visual Analytics:

VIEWTABLE: Work.Cities_geocoded (Geocoded 03Feb2014)

	LAT	LONG	M_OBS	_MATCHED_	city	state
1	33.74889	-84.38806	26936	City	Atlanta	GA
2	40.68056	-74.64583	81289	City	Bedminster	NJ
3	35.79139	-78.78139	93762	City	Cary	NC
4	41.85	-87.65	32440	City	Chicago	IL
5	39.75889	-84.19167	98332	City	Dayton	OH
6	41.60056	-93.60889	40883	City	Des Moines	IA
7	42.331395	-83.04583	65729	City	Detroit	MI
8	41.76361	-72.68556	18459	City	Hartford	CT
9	33.44833	-112.07333	6610	City	Phoenix	AZ
10	29.42389	-98.49333	123723	City	San Antonio	TX
11	37.775	-122.41833	14956	City	San Francisco	CA
12	47.60639	-122.33083	142724	City	Seattle	WA

VARIOUS MAPPING EXAMPLES

In SAS Visual Analytics the user can decide to use either OpenStreetMap or ESRI ArcGIS map services for the underlying map background as stated before. Both of these technologies provide exceptional geographic details with features such as boundaries, roads, rivers, or streets. Report developers often choose light map colors with just the minimum features such as country or state boundaries in order to let the report consumer focus on the geographical overlay, using the display of key metrics as bubble plots or colored mapped segments (also known as choropleth) for each geographic location.

Figure 3 shows an example of the ESRI ArcGIS map service selector in SAS Visual Analytics, which can be used to toggle the geo map background.

Figure 3. Map Service Selector

EXAMPLE 1: STORE LOCATIONS

A common example for geographical representation of data is to simply plot locations of stores, factories, facilities on a map. In this example we are using fictitious company Insight Toys, which sells toys around the world. The data contains information about factories and sales facilities including their city, region, and country locations.

SAS Visual Analytics provides an auto mapping feature that allows the geographical reference of each location (either by name or by ID). This means that the original data source does not require to carry the geographical coordinates (latitude and longitude). In order to use geo map visualization, the source column needs to be configured as geographical column by assigning one of the geographical roles. In the example shared here, for a column "Facility Country" we select "Country or Region Names" as a geographical role using the pop-up menu as shown in Figure 4.

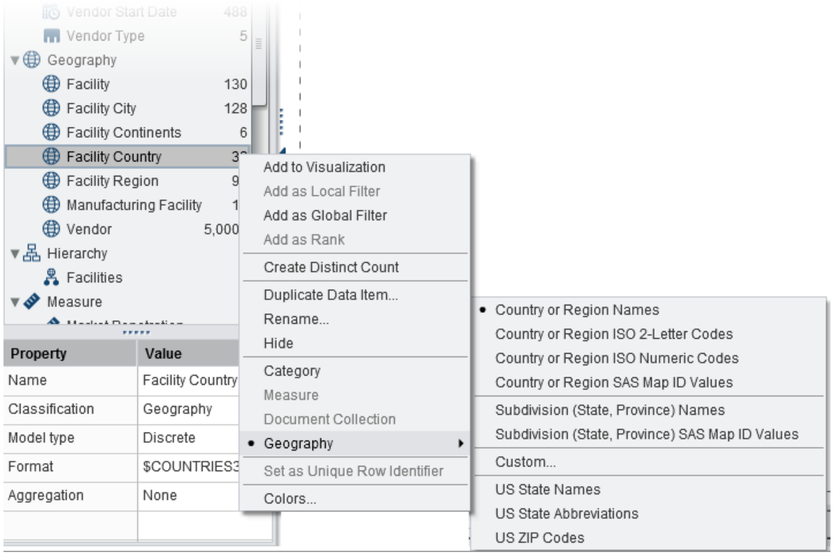


Figure 4. Geographical Roles

Once the column has been configured as geographical item, it can be used to visualize the facility locations on a map (Figure 5).

Figure 5. Number of Facilities by Country

For further analysis you might also add additional metrics in order to answer some of the typical business questions such as:

- Where do we sell well?
- Which facility region is causing high production costs?

Figure 6 shows an example of how to use the bubble size and color to represent two of those metrics.

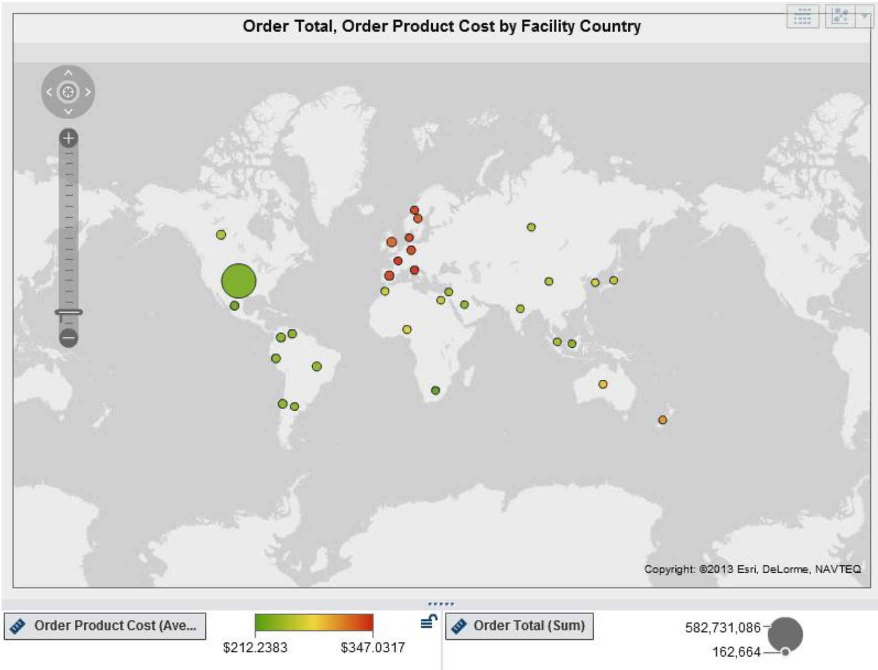


Figure 6. Product Cost and Total Number of Orders by Country

EXAMPLE 2: THEMATIC MAPS

Often geographical data are not bound to a specific location and rather represent a regional area such as state or country. SAS Visual Analytics uses choropleth mapping to show statistical data aggregated over these regions, by coloring or shading such regions.

Figure 7 shows U.S. states with higher order product costs in dark red on a choropleth map. Using this method assumes a relatively even distribution of the aggregated values within each region.

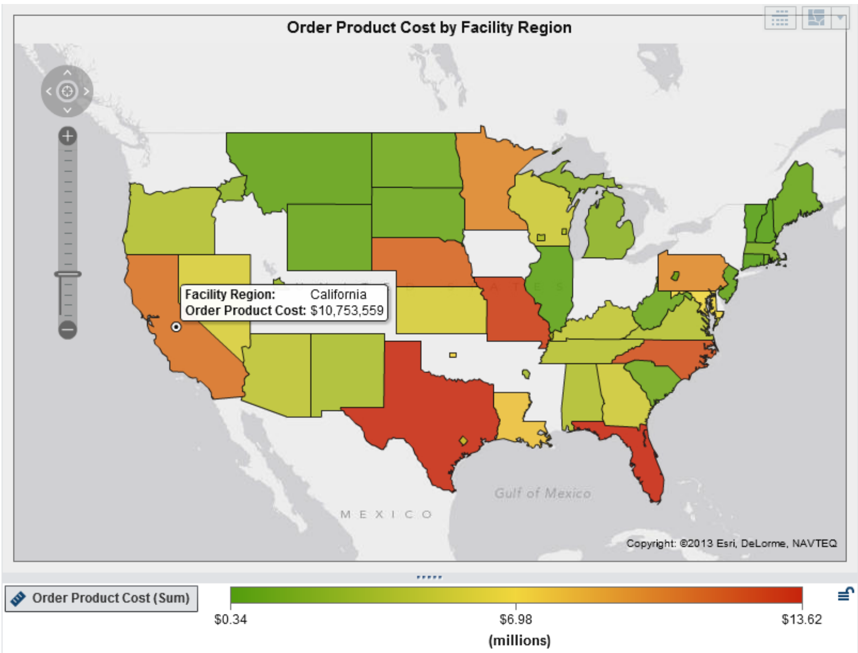


Figure 7. Order Product Cost by Region

EXAMPLE 3: NETWORKS

From social networks to supply chains to text analytics, network analysis is becoming a critical requirement—and network visualization is one of the best ways to make sense of the results. The SAS Visual Analytics visualization option **Network diagram** shows links between related nodes as well as additional attributes such as color, size, or labels. For specific industries such as transportation, nodes can be mapped to spatial coordinates. This is how SAS Visual Analytics creates network diagrams overlaid over geographical maps.

The example in Figure 8 shows a network of the U.S. Interstate Highway System. This network of freeways consists of 230 nodes (closest city at the start or end) as well as 392 edges (representing the interstate sections). The nodes are colored based on the number of interstates connected to (degree) and edges are colored by interstate number.

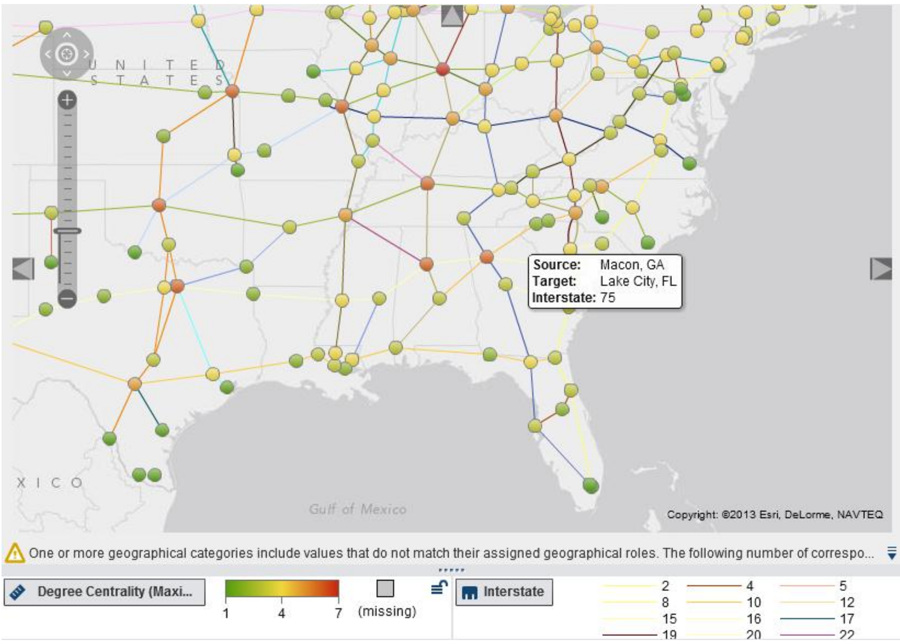


Figure 8. U.S. Interstate Highway System

The data used in the example was geocoded and enriched with network metrics such as degree in order to visualize well-connected interstate sections. Besides the geocode techniques described in the data requirement section, we also used the OPTGRAPH procedure to calculate some basic network metrics.

The following input table was used containing start and end city of each interstate section:

VIEWTABLE: Data.Interstates_network

	origin	destination	OriginCity	OriginState	DestCity	DestState	Interstate
1	Abilene, TX	Fort Worth, TX	Abilene	TX	Fort Worth	TX	20
2	Akron, OH	Cambridge, OH	Akron	OH	Cambridge	OH	77
3	Akron, OH	North Jackson, OH	Akron	OH	North Jackson	OH	76E
4	Albany, NY	Newburgh, NY	Albany	NY	Newburgh	NY	87
5	Albany, NY	Springfield, MA	Albany	NY	Springfield	MA	90
6	Albert Lea, MN	Des Moines, IA	Albert Lea	MN	Des Moines	IA	35
7	Albert Lea, MN	La Crosse, WI	Albert Lea	MN	La Crosse	WI	90
8	Albuquerque, NM	Amarillo, TX	Albuquerque	NM	Amarillo	TX	40
9	Albuquerque, NM	Las Cruces, NM	Albuquerque	NM	Las Cruces	NM	25
10	Amarillo, TX	Lubbock, TX	Amarillo	TX	Lubbock	TX	27
11	Amarillo, TX	Oklahoma City, OK	Amarillo	TX	Oklahoma City	OK	40
12	Angola, IN	Indianapolis, IN	Angola	IN	Indianapolis	IN	69
13	Angola, IN	Toledo, OH	Angola	IN	Toledo	OH	80
14	Angola, IN	Toledo, OH	Angola	IN	Toledo	OH	90
15	Ardmore, OK	Dallas, TX	Ardmore	OK	Dallas	TX	35
16	Ardmore, OK	Fort Worth, TX	Ardmore	OK	Fort Worth	TX	35
17	Arvada, CO	Denver, CO	Arvada	CO	Denver	CO	70
18	Arvada, CO	Welby, CO	Arvada	CO	Welby	CO	76W
19	Asheville, NC	Charlotte, NC	Asheville	NC	Charlotte	NC	26
20	Asheville, NC	Statesville, NC	Asheville	NC	Statesville	NC	40
21	Atlanta, GA	Augusta, GA	Atlanta	GA	Augusta	GA	20
22	Atlanta, GA	Macon, GA	Atlanta	GA	Macon	GA	75
23	Atlanta, GA	Montgomery, AL	Atlanta	GA	Montgomery	AL	85

The CENTRALITY option in PROC OPTGRAPH will calculate related measures such as degree as seen in the following code example. We used the DATA_LINKS_VAR option to specify what column defines the source and target nodes.

```
proc optgraph
        loglevel = moderate
        data_links = interstates
        out_links = interstates_edges
        out_nodes = interstates_nodes;
        data_links_var
                from       = OriginCity
                to         = DestCity;
        performance nthreads = 2;
        centrality
                degree = out;
run;
```

As a result, we receive two data sets (interstates_nodes, interestates_edges) holding node and edge attributes. As degree is a node attribute, you will need to join the new network metric to your source table using standard SQL join. Once the table is reloaded into the SAS Visual Analytics environment, users will be able to utilize this measure as part of their data exploration.

EXAMPLE 4: NETWORK PATH ANALYSIS

You might also utilize optimization routines as part of SAS/OR® software within SAS Visual Analytics. The OPTNET procedure includes a number of graph theory, combinatorial optimization, and network analysis algorithms.

Network optimization routines can be used to analyze relationships between entities. As shown in the previous example, even common things such as roads and rivers can be expressed and visualized via networks. As such we can also apply network routines such as short path analysis to calculate the shortest path between two locations (nodes). Similar routines are used in common in-car navigation systems.

For this example we are using the same input data set as shown in the previous example although we applied geocoding as shown in the data preparation section so that each origin and destination city has latitude and longitude coordinates assigned.

VIEWTABLE: Data.Interstates_network

	DestCityLat	DestCityLng	OriginCityLat	OriginCityLng	origin	destination	Interstate
1	32.72528	-97.32056	32.44861	-99.73278	Abilene, TX	Fort Worth, TX	20
2	40.03111	-81.58861	41.08139	-81.51917	Akron, OH	Cambridge, OH	77
3	41.1	-80.8575	41.08139	-81.51917	Akron, OH	North Jackson, OH	76E
4	41.50333	-74.01083	42.6525	-73.75667	Albany, NY	Newburgh, NY	87
5	42.10139	-72.59028	42.6525	-73.75667	Albany, NY	Springfield, MA	90
6	41.60056	-93.60889	43.64806	-93.36806	Albert Lea, MN	Des Moines, IA	35
7	43.80139	-91.23944	43.64806	-93.36806	Albert Lea, MN	La Crosse, WI	90
8	35.22194	-101.83083	35.08444	-106.65056	Albuquerque, NM	Amarillo, TX	40
9	32.31222	-106.77778	35.08444	-106.65056	Albuquerque, NM	Las Cruces, NM	25
10	33.577784	-101.85472	35.22194	-101.83083	Amarillo, TX	Lubbock, TX	27
11	35.4675	-97.51611	35.22194	-101.83083	Amarillo, TX	Oklahoma City, OK	40
12	39.768329	-86.158052	41.63472	-84.99944	Angola, IN	Indianapolis, IN	69
13	41.66389	-83.55528	41.63472	-84.99944	Angola, IN	Toledo, OH	80
14	41.66389	-83.55528	41.63472	-84.99944	Angola, IN	Toledo, OH	90
15	32.78333	-96.8	34.17417	-97.14333	Ardmore, OK	Dallas, TX	35
16	32.72528	-97.32056	34.17417	-97.14333	Ardmore, OK	Fort Worth, TX	35
17	39.73917	-104.98417	39.80278	-105.08694	Arvada, CO	Denver, CO	70
18	39.83667	-104.95861	39.80278	-105.08694	Arvada, CO	Welby, CO	76W
19	35.22694	-80.84333	35.60083	-82.55417	Asheville, NC	Charlotte, NC	26
20	35.782497	-80.887506	35.60083	-82.55417	Asheville, NC	Statesville, NC	40
21	33.47083	-81.975	33.74889	-84.38806	Atlanta, GA	Augusta, GA	20
22	32.84056	-83.6325	33.74889	-84.38806	Atlanta, GA	Macon, GA	75
23	32.36667	-86.3	33.74889	-84.38806	Atlanta, GA	Montgomery, AL	85
24	34.00056	-81.035	33.47083	-81.975	Augusta, GA	Columbia, SC	20

As a road network example, the links are the roads and the nodes are intersections between roads. In order to calculate the shortest path we also assigned the actual distance (or time to travel) between two intersections. Calculating the distance between two locations on the planet can be achieved by using the GEODIST function in SAS.

The following example shows how to calculate the distance (in miles) as well as an average time to travel that distance (using an average car speed of 60 miles per hour based on normal traffic patterns).

```
data interstates_path;
        set interstates_network;
        Miles = geodist(OriginCityLat, OriginCityLng, DestCityLat, DestCityLng, 'M');
        miles_per_hour = 60;
        time_to_travel = miles/miles_per_hour * 60 * 60;
        format time_to_travel time9.;
        label miles_per_hour="MPH" time_to_travel="Time To Travel";
        if destination ne "" and miles ne .;
run;
```

The following use of PROC OPTNET calculates the shortest path between the given locations:

```
proc optnet
   data_links    = interstates_path;
   data_links_var
      from       = Origin
      to         = Destination
      weight     = time_to_travel;
   shortpath
          weight2    = edge
      out_paths = ShortPath
      source     = "Las Vegas, NV"
      sink       = "Washington, DC";
run;
```

The result is a list of route stops indicating the shortest path between Las Vegas and Washington:

VIEWTABLE: Work.Shortpath

	source	sink	order	origin	destination	Time To
1	Las Vegas, NV	Washington, DC	1	Las Vegas, NV	Barstow, CA	2:17:59
2	Las Vegas, NV	Washington, DC	2	Barstow, CA	Flagstaff, AZ	5:05:14
3	Las Vegas, NV	Washington, DC	3	Flagstaff, AZ	Gallup, NM	2:45:51
4	Las Vegas, NV	Washington, DC	4	Gallup, NM	Albuquerque, NM	2:02:06
5	Las Vegas, NV	Washington, DC	5	Albuquerque, NM	Amarillo, TX	4:33:05
6	Las Vegas, NV	Washington, DC	6	Amarillo, TX	Oklahoma City, OK	4:04:20
7	Las Vegas, NV	Washington, DC	7	Oklahoma City, OK	Joplin, MO	3:21:21
8	Las Vegas, NV	Washington, DC	8	Joplin, MO	Springfield, MO	1:07:41
9	Las Vegas, NV	Washington, DC	9	Springfield, MO	Saint Louis, MO	3:15:24
10	Las Vegas, NV	Washington, DC	10	Saint Louis, MO	Mount Vernon, IL	1:13:24
11	Las Vegas, NV	Washington, DC	11	Mount Vernon, IL	Louisville, KY	2:50:59
12	Las Vegas, NV	Washington, DC	12	Louisville, KY	Lexington, KY	1:12:13
13	Las Vegas, NV	Washington, DC	13	Lexington, KY	Charleston, WV	2:36:56
14	Las Vegas, NV	Washington, DC	14	Charleston, WV	Lexington, VA	2:05:38
15	Las Vegas, NV	Washington, DC	15	Lexington, VA	Staunton, VA	0:32:21
16	Las Vegas, NV	Washington, DC	16	Staunton, VA	Middletown, VA	1:14:12
17	Las Vegas, NV	Washington, DC	17	Middletown, VA	Washington, DC	1:07:38

In order to visualize the new shortest route using SAS Visual Analytics, we merged this data set to our main network data set and used a link attribute to highlight the route as shown in Figure 9.

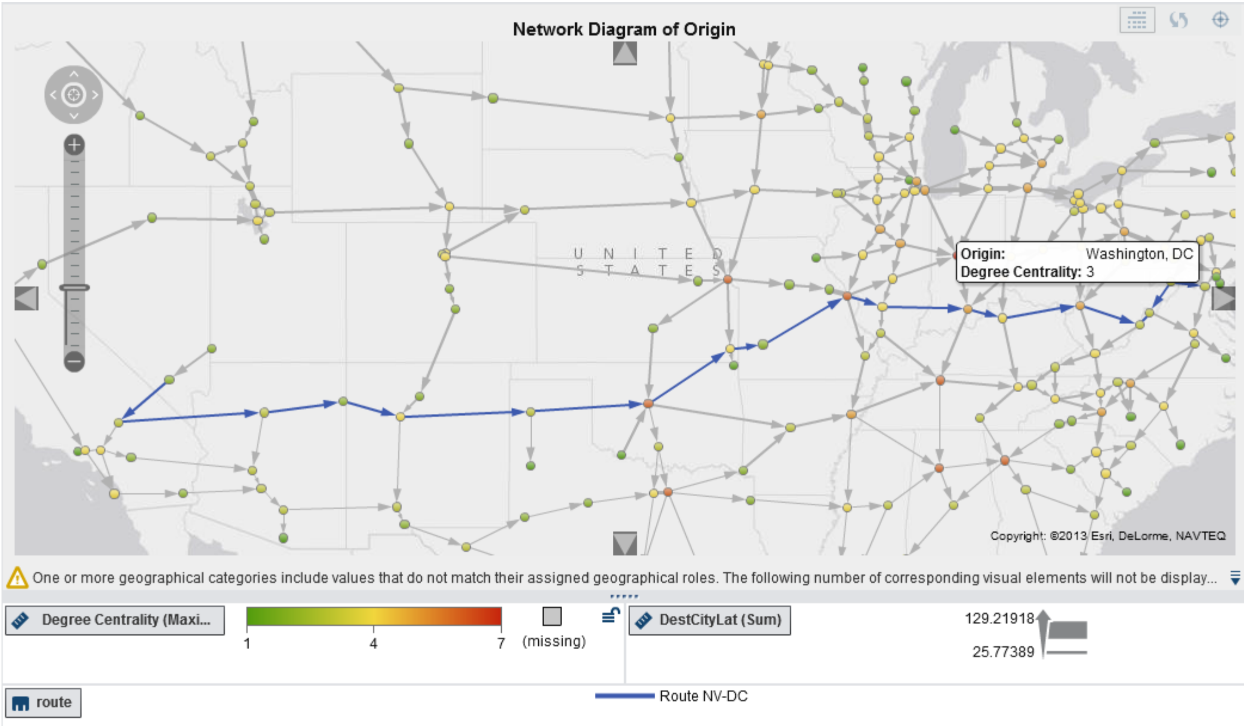
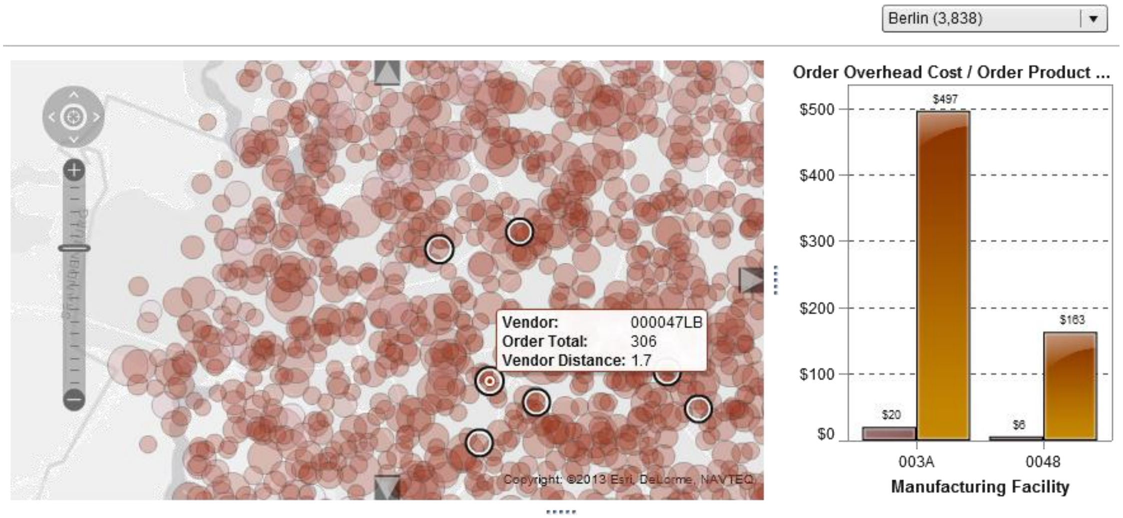


Figure 9. Network Path Analysis

INTERACTING WITH GEO VISUALIZATION

SAS Visual Analytics allows creating logical interaction links across the various visual objects displayed in a report. It is a very common user interaction seen where they would like to click on a location on the geographical map and get more detailed data associated with that location either in detailed tabular format or filtering the information on other reports objects.

The example in Figure 10 shows interactive filtering and brushing between selected vendors in a tabular result and geographical map.



Vendor	Vendor Credit Rating	Vendor Distance	Vendor Rating	Vendor Satisfaction ▼
000047LW	19	0.9	30%	93%
000047LB	21	1.7	57%	93%
000047L6	7	2.4	93%	92%
000047LJ	21	3.2	82%	91%
000047NE	10	2.5	68%	91%
000047OQ	6	4.0	29%	90%
000047PO	10	3.4	12%	90%
	Sum: 51886	Sum: 23,583.2	Average: 50%	Average: 48%

Figure 10. Interaction with Geo Maps

Interactivity between multiple report objects can be configured in both directions. The dialog box shown in Figure 11 is used to configure filter or brushing direction. It shows that vendor selections in the table will highlight each vendor in the linked geographical map which also filters related manufacturers in the resulting bar chart.

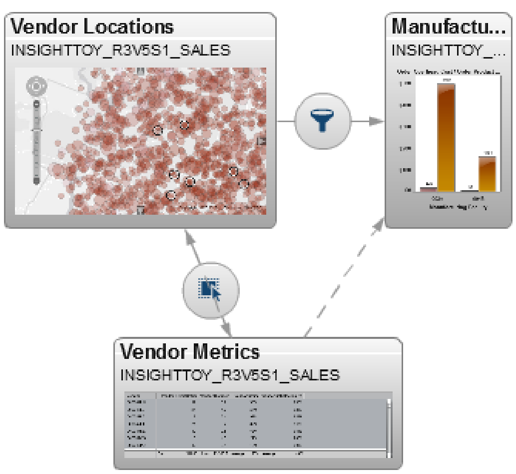


Figure 11. Interactions View

CONCLUSION

SAS Visual Analytics is a unique BI offering from SAS that provides insights on large amount of data at a very high speed, enabling a fact-based and analytical-driven BI solution for all organizations. Integrating with GIS technologies further enhances your data, providing location intelligence and allowing business users to visualize and understand the importance of location information associated with their enterprise data.

It is to be noted that even when a BI technology provides great mapping visuals, some aspect of data preparation is required to make sure the enterprise data holds the required location information to support the various data needs, from a simple display of a geographical map to advanced location intelligence.

RECOMMENDED READING

Chitale, Anand, and Christopher Redpath. 2013. “Whirlwind Tour Around SAS® Visual Analytics.” *Proceedings of the SAS Global Forum 2013 Conference*. Cary, NC: SAS Institute, Inc. Available at http://support.sas.com/resources/papers/proceedings13/057-2013.pdf.

SAS Institute Inc. 2014. Geocoding Data. Available at http://support.sas.com/rnd/datavisualization/mapsonline/html/geocode.html.

SAS Institute Inc. 2013. "GEOCODE Procedure." In *SAS/GRAPH 9.4: Reference, Second Edition*. Cary, NC: SAS Institute Inc. Available at http://support.sas.com/documentation/cdl/en/graphref/66521/HTML/default/viewer.htm#p087i29802bnfgn1qn9isdqzde0h.htm.

SAS Institute Inc. 2013. *SAS Visual Analytics 6.3: Administration Guide*. Cary, NC: SAS Institute Inc. Available at http://support.sas.com/documentation/onlinedoc/va/index.html.

SAS Institute Inc. 2013. *SAS Visual Analytics 6.3: User Guide*. Cary, NC: SAS Institute Inc. Available at http://support.sas.com/documentation/cdl/en/vaug/66720/HTML/default/viewer.htm#titlepage.htm.

SAS Institute Inc. 2013. SAS Social Network Analysis, product documentation. Available at https://support.sas.com/documentation/onlinedoc/socialnetworkinfo/index.html.

SAS Institute Inc. 2013. "The OPTNET Procedure." In *SAS/OR 12.1 User's Guide: Network Optimization Algorithms*. Cary, NC: SAS Institute Inc. Available at http://support.sas.com/documentation/cdl/en/ornoaug/65289/HTML/default/viewer.htm#ornoaug_optnet_gettingstarted01.htm.

CONTACT INFORMATION

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

Anand Chitale
SAS Institute Inc.
SAS Campus Drive
Cary, NC, 27513
anand.chitale@sas.com

Falko Schulz
SAS Institute Inc.
SAS Campus Drive
Cary, NC, 27513
falko.schulz@sas.com