SAS3480-2016

The GEOCODE Procedure and SAS® Visual Analytics

Darrell Massengill, SAS Institute Inc., Cary, NC

ABSTRACT

SAS® Visual Analytics can display maps with your location information. However, you might need to display locations that do not match the categories found in the SAS Visual Analytics interface, such as street address locations or non-US postal code locations. You might also wish to display custom locations that are specific to your business or industry, such as the locations of power grid substations or railway mile markers. Alternatively, you might want to validate your address data that you are using with SAS Visual Analytics.

This paper shows how PROC GEOCODE can be used to simplify geocoding by processing your location information before importing data into SAS Visual Analytics.

INTRODUCTION

SAS Visual Analytics can produce geographic maps. However, it might not produce every map that you need. You might have street addresses of customers, postal codes for the United Kingdom, or worldwide city names that SAS Visual Analytics doesn't handle. You can use PROC GEOCODE to get the latitude and longitude coordinates for these address locations to add to your data set. Or you can use PROC GEOCODE to find the administrative level coordinates and the ID information for a country or state. Then you can move the data set to SAS Visual Analytics and display it in a map.

PROC GEOCODE

In order to produce your map, the first thing that you must do is convert your location information into map coordinates. You use PROC GEOCODE to do this. There are three categories of examples below: US street-level geocoding, administrative level (such as country, state, and county) geocoding, and validating your data with geocoding. See RESOURCES for additional information.

The examples discussed in this paper as well as additional examples can be found with the *Proceedings*. Look up the title of this paper in the *Proceedings*

(<u>http://support.sas.com/resources/papers/proceedings16</u>). Under the title is a link that reads: "Download the data files (ZIP)".

US STREET-LEVEL GEOCODING

In the first example of US street-level geocoding, PROC GEOCODE will convert each street address (of a school) into X and Y coordinates for the map. These coordinates can be shown as a *Bubbles* map or as a *Coordinates* map in SAS Visual Analytics. The METHOD= option value STREET specifies street-level geocoding. DATA= specifies the input data set and OUT= specifies the output data set for the geocoded data. The LOOKUPSTREET= option provides the name of the data set used to look up the street address. The lookup data set, SASHELP.GEOEXM, is a sample data set for Wake County, North Carolina only. The entire country data set, SASHELP.USM, is not installed with SAS. To download it, go to the SAS Maps Online web site. The variable names of Address, City, State, and Zip are the default names used in PROC GEOCODE, so you are not required to specify them.

```
data schools (label='Wake County, NC public schools');
infile datalines dlm=',';
length school $64 address $32 city $24
state $2 zip 5 type $12 color $10;
input school address city state zip type enrollment;
datalines;
Apex Elementary, 700 Tingen Road, Apex, NC, 27502, Elementary, 674
Apex High, 1501 Laura Duncan Road, Apex, NC, 27502, High, 2542
...
```

```
;
run;
proc geocode
    method = street    /* Street method    */
    data = work.schools    /* Address data to geocode    */
    out = work.streetlevel    /* Geocoded output data set    */
    lookupstreet = sashelp.geoexm    /* Street method lookup data    */
    /*lookupstreet = sashelp.usm*/
    ;
run;
guit;
```

This will produce X and Y coordinates along with some other variables:

Obs	Y	X	M_ADDR	M_CITY	M_STATE	M_ZIP	M_OBS	_MATCHED_	_STATUS_	_NOTES_	_SCORE_	school	address	city	state	zip	type	enrollment
1	35.7200	-78.8559	700 Tingen Rd	Apex	NC	27502	18391	Street	Found	AD ZC NM TS	65	Apex Elementary	700 Tingen Road	Apex	NC	27502	Elementary	674
2	35.7439	-78.8367	1501 Laura Duncan Rd	Apex	NC	27502	10520	Street	Found	AD ZC NM TS	65	Apex High	1501 Laura Duncan Road	Apex	NC	27502	High	2542

To see a map in SAS Visual Analytics, choose the **SCHOOL** variable and the **GEOGRAPHY** category. Then choose **CUSTOM**.

ADMINISTRATIVE-LEVEL GEOCODING

The administrative-level geocoding example will use the same data that SAS Visual Analytics uses to find the location. SAS Visual Analytics uses two data sets to match up administrative-level names and locations: ATTRLOOKUP and CENTLOOKUP. These data sets will map to the region or the point data to the centroid of the country or state. To access the location data, assign the libref VAlookup to the location of the ATTRLOOKUP and CENTLOOKUP data sets.

This example has two parts: a macro that calls PROC GEOCODE and does the work, and a program that sets up the data and calls the macro. The macro first sets a default value for the startattr parameter if the data set is not specified. Next the STARTATTR data set is sorted and has an X variable and a Y variable added because PROC GEOCODE requires them. The macro calls PROC GEOCODE twice with a METHOD= value of CUSTOM. The first PROC GEOCODE finds the ID value that matches the ATTRLOOKUP data set. Here is the first part of the macro:

```
proc geocode
    method=custom
    data= &indata
    lookup=attrlookup
    out=work.attr
    lookupvar= &lookupvar
    addressvar= &addrvar
    attribute_var= (ID )
    ;
run;
```

The second time PROC GEOCODE is run to find the X and Y values that match the ID value in the CENTLOOKUP data set:

```
/* Then use the ID value to find the x,y CENTroid and the MAPNAME */
proc geocode
    method=custom
    data=work.attr
    lookup=VAlookup.centlookup
    out=&outdata
    lookupvar= ID
    addressvar=ID
    attribute_var=(mapname)
    ;
run;
%mend;
```

The second part of the program prepares the data that you want to locate on a map, subsets the ATTRLOOKUP data set, and then calls the GEOCODE macro. The DATA step code subsets the ATTRLOOKUP data set; MYATTR will contain data for states (level=1) in the United States.

```
data addr; /* replace this with your data */
length InputAddr $199;
InputAddr="North Carolina"; output;
InputAddr="Virginia"; output;
InputAddr="California"; output;
run;
/*subset out US states only */
data myattr;
  set VAlookup.attrlookup;
  if (isoname="UNITED STATES" and level=1 /*states*/) then output;
run;
/* output data, input data, lookup var, address var, startattr */
```

```
%geocode(usstatename,addr,IDNAME,InputAddr,myattr);
```

Here is a PROC PRINT of the results:

Obs	у	x	mapname	M_OBS	_MATCHED_	ID	InputAddr
1	35.2142	-79.493	MAPSGFK.NAMERICA1	0	ID	US-37	North Carolina
2	38.0034	-78.455	MAPSGFK.NAMERICA1	0	ID	US-51	Virginia
3	37.2719	-119.995	MAPSGFK.NAMERICA1	0	ID	US-06	California

To see a map in SAS Visual Analytics, choose the ID variable and the **GEOGRAPHY** category. Then choose **Subdivision SAS Map ID Values**.

Using the same GEOCODE macro with another program, you can also display countries around the world. The example below creates a data set with selected countries and uses the GEOCODE macro to process them:

```
data addr; /* replace this with your data */
  length InputAddr $199;
  InputAddr="United States"; output;
  InputAddr="Canada"; output;
  InputAddr="Mexico"; output;
  InputAddr="Peru"; output;
  InputAddr="Korea, Republic of"; output;
run;
```

%geocode(countryname,addr,key,InputAddr);

Here are the results with PROC PRINT:

Obs	у	x	mapname	M_OBS	_MATCHED_	ID	InputAddr
1	37.1681	-99.565	MAPSGFK.NAMERICA	0	ID	US	United States
2	56.9545	-110.865	MAPSGFK.NAMERICA	0	ID	CA	Canada
3	23.6244	-102.259	MAPSGFK.NAMERICA	0	ID	MX	Mexico
4	-9.1955	-75.746	MAPSGFK.SAMERICA	0	ID	PE	Peru
5	36.4547	128.000	MAPSGFK.ASIA	0	ID	KR	Korea, Republic of

In SAS Visual Analytics, choose the following ID variable and the **GEOGRAPHY** category. Then choose **Country SAS Map ID Values**.

VALIDATING YOUR DATA WITH GEOCODING

You can also use PROC GEOCODE to validate that your data will work with SAS Visual Analytics. If the administrative level (country or state) does not match the spelling in SAS Visual Analytics, you can be alerted to which variables need checking. The following validation uses the same GEOCODE macro as the administrative-level geocoding in the previous example, but some extra checking is needed. When MATCHED =None, there was not a match between your data and the SAS Visual Analytics data.

```
data addr; /* replace this with your data */
   length Country $200;
   input Country $1-20 population;
datalines;
United States 317466000
Canada 35295770
Mexico 118395054
Peru 30475144
```

```
Korea, Republic of 50219669
Bolivia
                    10027254
;
run;
/* Subset the lookup data to just the country data*/
data myattr;
  set VALOOKUP.attrlookup(where= (level=0)); /* country level */
run;
/* notice the last parm overrides the default lookup data*/
%geocode(CKcountry,addr,IDNAME,Country,myattr);
data CkBadMatch;
  set CKcountry (where= ( MATCHED = "None"));
 if _N_=1 then put "WARNING: Country Name values that do not match:";
 put Country=;
run;
```

This produces the following statement in the SAS Log:

WARNING: Country Name values that do not match: Country=Bolivia

This indicates that "Bolivia" is not the correct name; the correct name is "Bolivia, Plurinational State of".

SAS VISUAL ANALYTICS MAPS

Now that you have the data from PROC GEOCODE and SAS, you can use it with SAS Visual Analytics. A way to load the data is shown in screen shots below. First, click on **Select a Data Source**:

👘 🔹 👯 Exploration 1 ×				§sas
File Edit Data Visualization	View Help			Sign Out
	SAS Visual Analytics Explorer		×	
	Start a new exploration			
	Select a Data Source			
	Recent explorations		Browse	
	Name	Location	_	
			- 1	
		One	en Cancel	
		- Opt		

Now under Import Data and Local, select SAS Data Set:

Sas										
File Edit Data	Open Data Source			×	Sign Out					
	Data Sources	P	5	Import Data						
	Name 🔺	Description	Location	Local						
	_STARTPEROD02242_S	Imported on Tuesday, Fe	/SAS Projects/NLS/Autom	Microsoft Excel (*.xls, *.xlsx, *.xlsm						
	02242015PRDSALE_R	Imported on Tuesday, Fe	/SAS Projects/NLS/Autom	Text Files (*.csv, *.txt, *.zip)						
	100 BEST AUTO EXCEL	Imported on Monday, Feb	/Shared Data/SAS Visual	SAS Data Set						
	ABTDATA	Telco_abtdata	/Shared Data/LASR/VAFIT	Server						
	ACME_BANKJUNE2013	Capital Exposure Risk D	/Shared Data/LASR/VAFIT	SAS Data Set						
	ACME_BANKJUNE2013		/SAS Projects/VADemo/M	Aster						
	ACME_PORTFOLIO_STR	Murali Acme Bank portfoli	/Shared Data/LASR/VAFIT	DB2						
	ACMEBANK_STRESSTE	Murali Acme Bank Demo	/Shared Data/LASR/VAFIT	Hadoop						
	ACTUAL_PLAN_2010_2	Murali actual_plan_2010	/Shared Data/LASR/VAFIT	BigInsights						
	ADVERSE	Adverse	/Shared Data/LASR/VAFIT	Cloudera						
	AIRPORT4_NETWORK	Data showing the relatio	/Shared Data/LASR/VAFIT	Cloudera Impala						
	AIRPORT4_NETWORK	Data showing the relatio	/Shared Data/LASR/VAFIT	Other						
	ALCOHOL_DATE_2008	(にほんご) [host:port]	/SAS Projects/NLS/LASR/	Facebook						
	ALERT_INVESTIGATION	Alert investigation extract	/Shared Data/LASR/VAFIT	Google Analytics						
			05 D-t(04.03/61	Twitter						
	587 data sources found									
				Open Cancel						

Then select your data and select **Open**:

ft 🔹 👯 Ex	ploration	n 1 ×									<u>S</u> sas
File Edit	Data	Open Da	ata Source							×	Sign Out
		Data So	urces	ø				5	Import Data		
		Name		 Descriptio 	n	Loca	tion		Local		
		_STARTE	PEROD02242	S Imported o	n Tuesday, Fe	/SAS	Projects/NLS	S/Autom	Microsoft Excel (*.xl	s, *.xlsx, *.xlsm	
								Autom	Text Files (*.csv, *.t	(t, *.zip)	
😂 Select file to u	upload b	y rdcesx1617	1.race.sas.com				X	/isual	SAS Data Set		
Look in:	NA 鷆	Viles		-	G 🦻 🖻 🛙			/VAFIT	Server		
Ca.	Name	2	~		Date modified		Туре	/VAFIT	SAS Data Set		
Recent Places	co	ountryabbrev	.sas7bdat		2/19/2015 1:56	5 PM	SAS Data !	mo/M	Aster	E)	
Hecchi Haces	E, co	ountryname.	sas7bdat		2/19/2015 1:56	PM	SAS Data !	/VAFIT	DB2	•	
	nonuscounties.sas7bdat		2/19/2015 1:56	5 PM	SAS Data !	VAFIT.	Hadoon				
	📑 no	onusstatenan	ne.sas7bdat		2/19/2015 1:56)15 1:56 PM	SAS Data !	A/AFIT	насоор		
Desktop	str	reetlevel.sas7	bdat		2/19/2015 1:59	PM (SAS Data !	/VAFIL	BigInsights	-	
	📑 us	abbrev.sas7t	bdat		2/19/2015 1:56	5 PM	SAS Data !	WAFIT	<u>Cloudera</u>	=.	
	us 🔤	counties.sas	7bdat		2/19/2015 1:56	5 PM	SAS Data !	/VAFIT	Cloudera Impala	•	
Libraries	us us	countyname	e.sas7bdat		2/19/2015 1:56	5 PM	SAS Data !	WAFIT	Other		
	us	statename.s	as7bdat		2/19/2015 1:56	PM	SAS Data !	ASR/	other		
									Facebook		
								/VAFI1	Google Analytics		
D7A558								/:==I 🛄	Twitter		
	4						•				
						_			Ope	en Cancel	
Network	File nam	ne: s	treetlevel.sas7b	dat	_		Open				
	Files of t	type:	SAS Data Set(*.s	as7bdat)	•		Cancel				
	_		_	_		_					

Next select OK:

🚹 🔹 👯 Exploratio	n 1 ×		
File Edit Data	Open Data So	urce ×	
	Data Sources	Dimport Data	
	Name	Import SAS Data Set ×	
	_STARTPEROD	Source File	
	02242015PRDS	Source data set: streetlevel.sas7bdat Browse	
	100 BEST AUTO	Output Tabla	
	ABTDATA		
	ACME_BANKJU	Name: * streetlevel	
	ACME_DARKSO	Description: Imported on Wednesday, February 25, 2015 01:04:27 PM	
	ACMEBANK_ST	GMT-0500 from "streetlevel.sas7bdat" by "Massengill, A. Darrell"	
	ACTUAL_PLAN		
	ADVERSE	Store the table in a private location (so others cannot access the data)	
	AIRPORT4_NET	Preview OK Cancel	
	AIRPORT4_NET		
	ALCOHOL_DAT	E_2008 (RtgtAcC) (nostport) //SAS Projects/NLS/LASR/ Facebook	
	587 data sources	s found	
		Open Cancel	

Next click school ...

👘 💌 👯 * Exploration 1 *		<u>s</u> sas
File Edit Data Visu	alization View Help	<u>Sign Out</u>
闘 🖕 🖬 🍽 🖉 📗		: • fa 🖉 🖩 🖉 🛛 🖗
Data	Visualization 1 🐺 🗕 🗆 🗙	🔊 Roles × 🌱 👯 🗄 🇣 🖾 👻
STREETLEVEL •	₹	Automatic Chart
🖽 🕒 🔎 Search data		Data source: STREETLEVEL
🔻 📊 Category (12)		Catagorias
MATCHED_	1	Categories
T _NOTES_	3	Mageurae
		A Hessure
ity address		✓ Measure
Color		:
M_ADDR 1		t I
M_CITY		
M_STATE	Drop a data item here	1
school		
Property Value		
Name		
Classification		
Model type		
Format		
Aggregation		Auto-update Update
Expression		100%

1 1 100 * E	Sas.									
File Edit	File Edit Data Visualization View Help Sign Out									
👯 🖆 🖬 🖷	░ृॄॖॖॖॖॖॖॖॣॱॖॖॖॣॱॖॖॣॱॖॖॾॱॗॾॱ॒ऻॖॖॣॱऻढ़ॕ॒ॱऻऒॿॿॎज़ज़ॎॹड़ॎक़ज़ॎख़क़ॶॿॎॶख़ॱॎऄ									
Data	📕 🗸 Visualiz	ation 1	₹_ ¤ ×	💀 Roles * 🌱 🚼 🗄 🗣 🖾						
STREETLEVEL	▼ ₹			射 Automatic Chart						
🖽 🕒 🔎 Searc	h data			Data source: STREETLEVEL	-					
🔻 🌐 Geography (1)									
 school Measure (7) SCORE enrollme Geocode Geocode M_OBS M_ZIP zin 	Add to Visualization Add as Filter on Visualization Add as Filter on STREETLEVEL Add as Rank on Visualization Create New Custom Category Duplicate Data Item Rename	□ ◊		Categories Category Measures Measure	•					
	Hide	Country or Region Names								
Property	Category Measure	Country or Region ISO Numeric Codes								
Name	Document Collection	Country or Region SAS Map ID Values								
Classification	• Geography >	Subdivision (State, Province) Names								
Model type	Set as Unique Row Identifier	Subdivision (State, Province) SAS Map ID Values								
Format	Colors	Custom								
Aggregation	None	US State Names US State Abbreviations US ZIP Codes		V Auto-update	Update					

And then right-click and select Geography. Then select Custom.

The following dialog box comes up. Use the drop-down menu to select **Geocoded Latitude (DEGREES)** and **Geocoded Longitude (DEGREES)**:

Custom Geography ×								
Latitude	Geocoded Latitude (DEGREES)	-						
Longitude	enrollment							
Coordinate Space	Geocoded Latitude (DEGREES) 🗧							
Coordinate Space	Geocoded Longitude (DEGREES)	≣						
	M_OBS							
	M_ZIP	•						
Drop a data item here								

Click the map icon at the top. Then drag **school** to the map. Once the map appears, you can change the **Map style**.



For the other types of maps, you use **ID** instead of **school**.

ft v 腺 * E	Exploratio	on 1 ×					
File Edit	Data	Visua	alizati	on	Vie		
👯 🖀 🖬 🗏) (°		٦				
Data					<u>M</u> •		
USSTATENAME	₹						
🖽 🖪 🔎 Searc	\supset						
🔻 🖬 Category (4)							
	ED_		1				
🖬 ID			3				
nputAddi 📊	nputAddr 3						
📊 mapnam	е		1				
🔻 🔷 Measure (3)							
💸 Geocode	d Latitud	le		ł.			
🔗 Geocode	d Longit	ude		1			
🔷 M_OBS							
				1			
Property	Value						
Name							
Classification							
Model type							
Format							
Aggregation							
Expression							

For USSTATENAME, you would select **Geography** and then **Subdivision (State, Province) SAS Map ID Values**:



You would have three map styles to choose from: **Coordinates**, **Bubbles**, and **Regions**.





Here are two of the map styles: **Regions** and **Bubbles**.



If you have a country, such as the data set CountryName, then you can select **Country or Region SAS Map ID Values**.

	Exploration 1 •	¢		S sas
File Edit	Data Vis	ualization View Help		<u>Sign Out</u>
🐝 🚘 🖬 🛛 🗳	୨ ୯ 🛛 🔳		▦▦ਘどど≿҂╳ҍёіі;;;;;	L9A & # • •
Data		Visualization 1	₹ _ 🗆 × 🞼 R	oles * 🌱 👯 🗄 🗣 🖾 🔽
COUNTRYNAME	•] 🔻		💹 Automatic Chart
🖽 🖪 🔎 Sean	ch data		Dat	a source: COUNTRYNAME
🔻 🖬 Category (3)			
MATCH	IED_	1	Cat	egones
📊 InputAdo	ir	5		Category 🔻
📊 mapnan	ne	3	Mea	asures
🔻 🌐 Geography	(1)	Add to Visualization		Measure 🔻
DI 🌐		Add as Filter on Visualization		
🔻 🔗 Measure (3)	Add as Filter on COUNTRYNAME		
💸 Geocode	ed Latitude	Add as Rank on Visualization		
Seocode 🔗	ed Longitude	Create		
♦ M_OBS		New Custom Category	Country or Region Names	1
·		Duplicate Data Item	Country or Region ISO 2-Letter Codes	
Property	Value	Rename	Country or Region ISO Numeric Codes	
Name	ID	Hide	Country or Region SAS Map ID Values	_
Classification	Geography	Category	Subdivision (State, Province) Names	
Model type	Discrete	Measure	Subdivision (State, Province) SAS Map ID Values	
Format	\$23	Document Collection	Custom	
Aggregation	None	• Geography •	US State Names	to-update Update
		Set as Unique Row Identifier	US State Abbreviations	
		Colors	US ZIP Codes	100%
				A 40001



Here are two map styles: **Coordinates** and **Regions**.



CONCLUSION

As seen above, PROC GEOCODE can be used to obtain the latitude and longitude coordinates from the location information to display a SAS Visual Analytics geographic map. PROC GEOCODE can also obtain the ID information for the administrative areas used by SAS Visual Analytics. Validation of the data, including matching the variable names of your data with expected variable names in SAS Visual Analytics, can also be done with PROC GEOCODE.

RESOURCES

"PROC GEOCODE: Finding locations outside the US." SAS Presentations at SAS Global Forum 2013, Cary, NC: SAS Institute Inc. Available <u>http://support.sas.com/rnd/papers.</u>

Documentation on PROC GEOCODE:

http://support.sas.com/documentation/cdl/en/graphref/67881/HTML/default/viewer.htm#p087i29802bnfgn 1qn9isdqzde0h.htm

SAS Maps Online web site: http://support.sas.com/mapsonline

CONTACT INFORMATION

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

Darrell Massengill SAS Institute Inc. SAS Campus Drive Cary, NC 27513 Darrell.Massengill@sas.com http://www.sas.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.