

Show Me That! Using SAS® VIYA®, Visual Analytics and Free ESRI Shapefiles to Show Geographic Data

John Schmitz, Luminare Data, Omaha, NE

ABSTRACT

Viya and Visual Analytics include features to connect with free, premium and custom ESRI maps to display geographic information. This paper provides simple examples for generating polygonal maps based on input data and free ESRI shapefiles. The paper reviews key configuration settings that impact ESRI map capabilities, generation and promotion of data for use by the geo-map feature, defining the category field for use with geo-mapping, filtering graph data, and producing choropleth style maps.

INTRODUCTION

Reporting is often enhanced by presenting the geographic patterns which may appear within the data. The power to graph data in maps can often improve the report viewer's ability to assimilate information more quickly. This reporting effort benefits from the ability to leverage relevant known shapes in the graphing process. Depending on the dispersion and desired granularity, users may desire maps at the country, state, county or other levels. Some organizations may require customized area graphs for sales territories, distribution regions, or related needs. Commonly used shapes may be obtained from various sources, while companies that desire custom mapping may construct them using ESRI mapping or similar tools.

In this paper, the focus is on leveraging ESRI style shapefiles to produce choropleth maps for geographic data. Sample geo-maps are created that shows employment by state as well as employment by ZIP code for a user-selected city. The process used is detailed and uses publicly available data sources so that reader may reproduce the effort as a training exercise. The core tools required to produce these maps are included with SAS® Viya® and Visual Analytics (VA), with this effort completed using Viya 3.3 and Visual Analytics 8.2.

This paper is divided into five sections:

- Environment Manager Settings.
- Importing Employment Data.
- State-level Map using Predefined Geographies.
- Obtain and Load a Public Shapefile.
- Generating the Custom Shape Geographic Map.

ENVIRONMENT MANAGER SETTINGS

Figure 1 highlights ESRI map settings within the Map Service section of VIYA Environment Manager. Although this effort does not require modifications to the settings, companies that self-host or subscribe to ESRI shapefiles for their reporting needs will likely require adjustments to these parameters. Organizations that leverage local ESRI map services will also modify these parameters to access their repositories. The VIYA Environment Manager is primarily used by the SAS system administrators, so not all users will have direct access to this environment. More detail on these configurations may be found in the *SAS® Viya® 3.3 Administration* guide (pages 236-237).

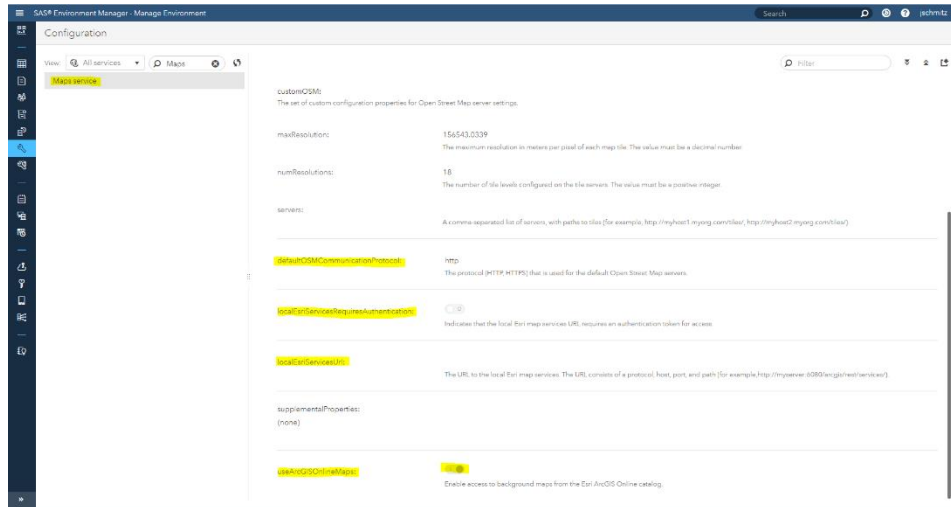


Figure 1. Map Service Settings in VIYA Environment Manager.

It is also sometimes necessary to adjust the Report Data service limits on the Geo Map to increase the geoRegion limit (Figure 2) and other maxRowLookup settings as needed. Reference to them is added here for completeness, but considered beyond the scope of this discussion

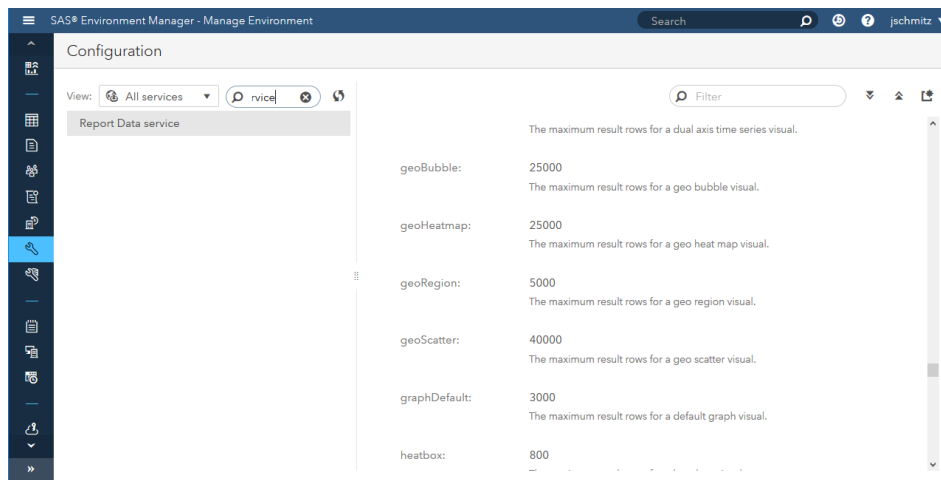


Figure 2. Environment Manager Settings Panel for Report Data Service.

IMPORTING EMPLOYMENT DATA

Our effort will require employment data at the state and ZIP code level. These data are available for download from the US Census Bureau (<https://census.gov>), as part of the County Business Patterns series. The Complete ZIP Code Totals File for 2016 (<https://www2.census.gov/programs-surveys/cbp/datasets/2016/zbp16totals.zip>) is used for this paper. The file was downloaded, unzipped, and stored in the user home directory as “~/SAS_Projects/ESRI_Maps/CBP_Data/zbp16totals.txt” where “~” is the users default home directory path.

Two core steps are required to prepare the data for use in the project, accessing two SAS procedures:

- IMPORT used to import the CSV file as a data set

- CASUTIL to load the data set into memory and promote to PUBLIC

The IMPORT procedure is a straightforward data load:

```
FILENAME CSV "~/SAS_Projects/ESRI_Maps/CBP_Data/zbp16totals.txt"
TERMSTR=LF;

PROC IMPORT DATAFILE=CSV
            OUT=WORK.CBP_totals
            DBMS=CSV
            REPLACE;

RUN;
```

This code produces a SAS data set, stored as WORK.CBP_totals.

The CASUTIL procedure completes 3 distinct steps:

- Delete existing table from PUBLIC (if found),
- Load the CBP_TOTALS table into memory in CASUSER,
- Promote the CASUSER table to PUBLIC

These steps can be accomplished in a single procedure call:

```
proc casutil;
  droptable casdata="CBP_totals" incaslib="public" quiet;
  load data=work.CBP_totals(keep=zip stabbr emp name
    where=(zip NE '99999'))
    outcaslib="casuser" casout="CBP_totals" replace;
  promote casdata="CBP_totals" casout="CBP_totals"
    incaslib="casuser" outcaslib=public keep;
run;
quit;
```

The process here drops any unneeded fields and excludes the record ZIP = '99999' which contains a total for all records. The process complete process includes other code required for setting CAS connections, process cleanups and related items. Complete code used for this step is included at the end of the paper in Appendix 1.

STATE-LEVEL MAP USING PREDEFINED GEOGRAPHIES

Once the Employment Data are loaded into memory, a state-level employment map can be created using a predefined geographic shape in VA. The process begins with a new report in VA environment.

The first step is to select data for the report, which is the CBP_TOTALS data set loaded above. Select the Data icon on the left-hand panel and select CBP_TOTALS on the OPEN DATA SOURCE form. Review the data to insure fields and record counts look good and explore as desired. Press OK to open this data set in the report.

Next, a geography data item is required. Select New Data Item and select Geography item from the dropdown menu (See Figure 3). A New Geography Item form will emerge. Set the NAME to "state_level", select stabbr for Based on, level Geography Data Type as "Predefined geographic names and codes", and select "US State Abbreviations" for the Name or code context. These settings will link the State Abbreviation code (stabbr) contained in CBP_TOTALS to a predefined state-level polygon shapefile. The completed form is shown in Figure 4.

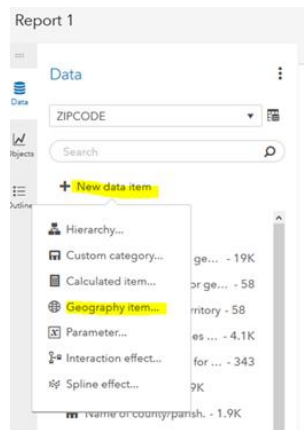


Figure 3. Create a new Geography Item.

Figure 4. Completed New Geography Item Form.

To create the geographic map, find Geo Map within Objects on the left-hand pane and drag it onto the Report. A choropleth map is generated by selecting Options on the right-hand pane, opening Geo Map 1, expanding Map and selecting REGIONS in the dropdown under Type (see Figure 5).

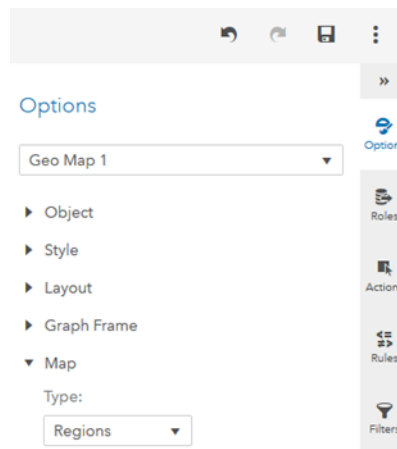


Figure 5. Setting Options for Regions.

Select Roles on the right-hand side. Under Category, add the Geographic item from above (STATE_LEVEL) and in Color, select EMP.

Field names can be changed and formats applied to produce a cleaner map. Select STATE_LEVEL under Geography and expand icon to the right. Rename STATE_LEVEL to STATE. Expand EMP under Measures. Rename it to EMPLOYMENT and change the format to COMMA12. This will update the field names and apply the format throughout the report. These changes also auto generate a more appropriate title 'Employment by State'. If either of the side panels are open, click on the icon that is open to close the panels. Select the SAVE icon on the upper right to save the report if desired.

The final map is shown in Figure 6. Pan and zoom controls in the upper right portion of the map can resize and adjust the map portion as desired. The object also provides a hover feature. Place the mouse over any state and it will show STATE and EMPLOYMENT values for that area.

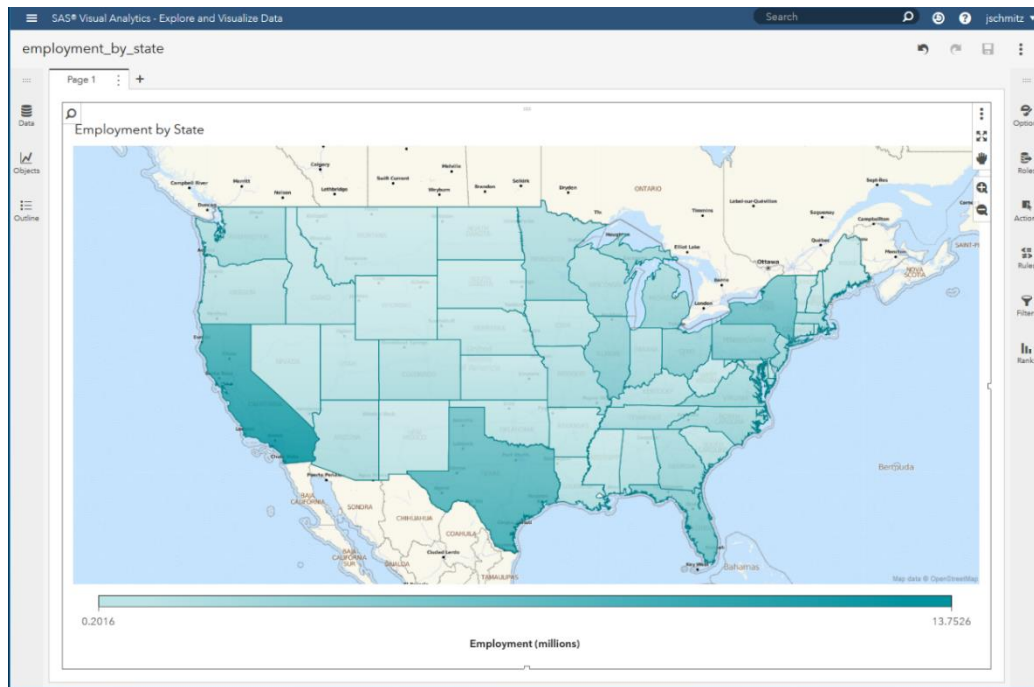


Figure 6. Final Employment by State, Zoomed to Continental US.

OBTAIN AND LOAD A PUBLIC SHAPEFILE

There are multiple sources for shapefiles available. One of the common repositories is the TIGER/Line® data maintained by US Census Bureau (<https://www.census.gov/geo/maps-data/>). This site hosts a variety of shapefiles that can be imported into SAS for use with the Geo Map object. The map used here is obtained by navigating on their site to 'Tiger Products', then to 'TIGER/Line Shapefiles'. That location provides the download screen shown in Figure 7.

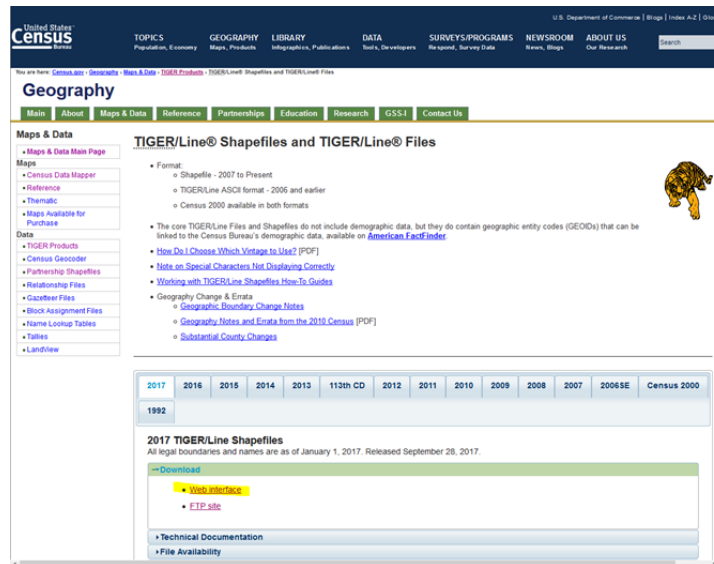


Figure 7. TIGER/Line Shapefiles Download Screen.

The web interface option navigates to a screen where various shapefiles options are available. This exercise uses the 2017 ZIP Code Tabulation Areas. The specific URL to the download page is:

<https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2017&layergroup=ZIP+Code+Tabulation+Areas>.

The download copies a .zip file (tl_2017_us_zcta510.zip) which contains multiple files that define ZIP code geographic boundaries for the United States. Contents of the zipfile are shown in Figure 8. These files need to be unzipped and placed in a location accessible to Viya for the shapefile import process.

Name	Type	Compressed size	Password ...	Size	Ratio	Date modified
tl_2017_us_zcta510.cpg	CPG File	1 KB	No	1 KB	0%	9/18/2017 10:22 AM
tl_2017_us_zcta510.dbf	DBF File	833 KB	No	2,267 KB	64%	9/18/2017 10:23 AM
tl_2017_us_zcta510.prj	PRJ File	1 KB	No	1 KB	20%	9/18/2017 10:22 AM
tl_2017_us_zcta510.shp	SHP File	519,186 KB	No	828,325 KB	38%	9/18/2017 10:23 AM
tl_2017_us_zcta510.shp.ea.iso.xml	XML Document	2 KB	No	16 KB	88%	9/18/2017 10:24 AM
tl_2017_us_zcta510.shp.iso.xml	XML Document	8 KB	No	38 KB	82%	9/18/2017 10:23 AM
tl_2017_us_zcta510.shp.xml	XML Document	5 KB	No	19 KB	74%	9/18/2017 10:23 AM
tl_2017_us_zcta510.shx	SHX File	206 KB	No	260 KB	21%	9/18/2017 10:23 AM

Figure 8. File Contents of Zipped Shapefiles.

The Viya installation provides macros for loading polygon shapefiles into CAS. Documentation is shown in the *SAS® Viya® 3.3 Administration* guide (pages 288-290).

The process uses two predefined macros:

- %SHPCNTNT (SHAPEFILEPATH=)
- %SHPIMPRT (SHAPEFILEPATH=, ID=, OUTTABLE=, CASHOST=, CASPORT=, CASLIB=, REDUCE=)

The first macro displays file contents of the shapefile pointed to in SHAPEFILEPATH. Results can be used to identify observations and variable in the shape file, including the name of the appropriate ID column for use in %SHPIMPRT.

The macro reports the shapefile contains 52,898,242 records and 12 variables. The observations define all the coordinates required to define all the US ZIP code polygons needed to define the polygon graph regions. Figure 9 shows a field list and sample records obtained from %SHPCNTNT. The Field GEOID10 contains ZIP code values that will be used as the primary ID in the import phase.

	X	Y	SEGMENT	ALAND10	AWATER10	CLASSFP10	FUNCSTAT10	GEOID10	INTPFLAT10	INTPFLON10	MTFCC10	ZCTASCE10
1	-83.708733	41.327326	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
2	-83.708547	41.327311	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
3	-83.704204	41.327265	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
4	-83.700799	41.327232	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
5	-83.698948	41.327207	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
6	-83.694229	41.327116	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
7	-83.691732	41.327109	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
8	-83.691004	41.327092	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
9	-83.690707	41.327094	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
10	-83.689463	41.327072	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
11	-83.689373	41.327073	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
12	-83.689389	41.327137	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
13	-83.689406	41.327176	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
14	-83.689006	41.327702	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
15	-83.688776	41.328442	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451
16	-83.686518	41.328939	1	63411475	157689	B5	S	43451	+41.3183010	-083.6174935	06350	43451

Figure 9. Data from %SHPCNTNT showing Field Names and Sample Records.

The %SHPIMPRT macro establishes a separate CAS connection, based on the CASHOST and CASPORT values provided. This session is disconnected when the macro terminates. Internally, it uses the .shp file, as well as related data files contained in that same directory, plus the ID column identified by ID= to convert and load the polygon data. The REDUCE= option can be used to reduce the size of the polygon data by omitting certain values but requires SAS/GRAPH. The final table is loaded to the location set by CASLIB= and OUTTABLE= for libname and tablename, respectively.

Each of the macros create files in SASWORK. The %SHPCNTNT macro leaves a data set named TEMP while the %SHPIMPRT leaves a table matching the name in OUTTABLE. They can be removed with the DATASETS procedure.

Complete code for this process is included in Appendix 2.

GENERATING THE CUSTOM SHAPE GEOGRAPHIC MAP

Generating the Geo Map will follow the same basic process as before; however, this case will require defining a custom polygonal shape based on the shapefile downloaded above. The high number of shapes will require that filtering options be added to the graph so that records can be narrowed to a specific City of interest. As before, process begins by opening a blank report, selecting the CBP_TOTALS data set and adding a Geo_Map object to the report.

USING THE CUSTOM SHAPEFILE DATA

As before, a new Geographic item link is needed to join the input data to the geographic shape data. Select New Data Item and then Geographic item from the dropdown menu. On the form, NAME is set to ZIP5 and select ZIP for BASED ON. However, this case will require that Geography Data Type be set to **Custom Polygonal Shapes**. Switching to the Custom Polygonal Shapes yields additional items to be completed on the form. Select Custom polygon provider set to **Define new polygon provider**. Changing the polygon provider opens a new form (New Polygon Provider) which is used to define the new provider.

This form is completed with:

- NAME: **ZIP5**
- LABEL: **ZIP5**
- TYPE: **CAS Table**
- SERVER: leave as default

- LIBRARY: **CASUSER**
- TABLE: **GEO_ZIP5**
- ID COLUMN: **GEOID10**
- SEQUENCE COLUMN: **_seq_**
- ADVANCED: left as defaults

When completed, select OK to return to the New Geography Item form. On the Geography form requires Region ID which is set to ZIP. This value will be used to complete the link between ZIP from the CBP_TOTALS table with GEOID10 from the Shapefile to map data to polygon regions. The remaining values on the form are not required for this map. The final geography item is shown in Figure 10.

Figure 10. Final Geography Item form.

Once entered, the map on the right should show mapped data and a percentage mapped. This example shows 5 unmapped areas of 6330 and an 84% mapping rate. Clearly, there are some inconsistencies between the two data sources in terms of ZIP codes included.

BUILD THE GEO MAP OBJECT

Now that data files are load and the Geography item has been added, the Geo Map object can be used again to create a choropleth map. As a brief summary, the steps are:

- Drag the Geo Map object onto the Report page.
- Under Options, set the Map Type to REGIONS.
- Under Roles, set the Category to ZIP5 and the Color to EMP.

ADDING A FILTERS FOR CITY

The settings above complete the required elements for a choropleth graph, but the region limits exceed predefined values for the Geo Map object. The number of regions and detail required to produce a US based ZIP code map is extensive.

One alternative is to create a Text Entry filters to allow City-level selections for the map. A Text Entry control is a convenient method to quickly add this filtering capabilities. To begin, simply drag the Text Entry Control from Objects to the top of the Geo Map. To set up the map filter, under Options, set Title to Custom and enter a title of 'Enter City', under Roles, select NAME for the Category, and under Actions, check Automatic actions and select one-way filter. Also uncheck Display filter breadcrumbs under Actions. Once set up, users can type the initial characters of the desired city name. The text control will display names that match the entry until the user selects the desired name from the list. Once a City name is selected, the map will update with the selected region.

A FINAL MAP

A final map can now be produced. Similar to the previous map, change the name of the measure EMP to EMPLOYMENT and the format to COMMA12. In this case, the Geography item can remain named ZIP5. Finally, type **INDI** in the Text filter box. It will provide a list of ZIP region names that begin with INDI. Select **INDIANAPOLIS, IN** from the list to view a graph of the Indianapolis area. Figure 11 shows the final graph with employment levels by ZIP code for the Indianapolis area.

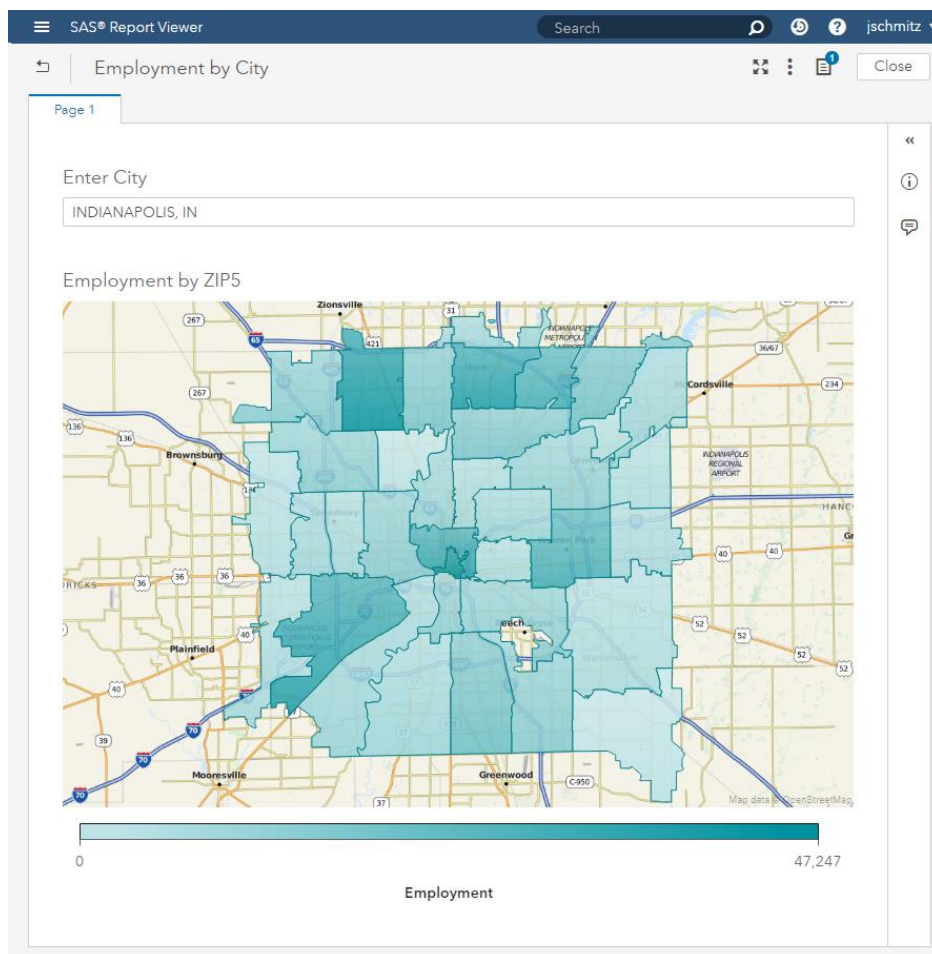


Figure 11. Graph of Employment by ZIP code for the Indianapolis Area.

CONCLUSIONS

The combination of Viya, Visual Analytics and ESRI mapping provide a power combination for presenting geographic-based data. The examples here server only to introduce users to the capabilities at a rudimentary level. The tools provide rich graphing capabilities that can be tailored to meet any number of business demands.

REFERENCES

SAS Institute Inc, 2017 *SAS® Viya® 3.3 Administration*, Cary, NC: SAS Institute, Inc.

CONTACT INFORMATION

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

John Schmitz
Luminare Data
john.schmitz@luminaredata.com
www.luminaredata.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.

APPENDIX 1 – CODE TO LOAD CBP DATA

```
*****;  
** LOAD CBP Data FILE **;  
*****;  
** DATA SOURCE **;  
** https://census.gov/data/datasets/2016/econ/cbp/2016-cbp.html **;  
** DOWNLOAD URL **;  
** https://www2.census.gov/programs-surveys/cbp/datasets/2016/  
** zbp16totals.zip **;  
** FILE PATH **;  
** ~/SAS_Projects/ESRI_Maps/CBP_Data/zbp16totals.txt **;  
*****;  
  
*****;  
** CONNECT TO CAS **;  
*****;  
options cashost=("cas","casbk") casport=5570;  
cas casauto;  
caslib _all_ assign;  
  
*****;  
** FOR CSV Files uploaded from Unix/MacOS **;  
*****;  
FILENAME CSV "~/SAS_Projects/ESRI_Maps/CBP_Data/zbp16totals.txt"  
TERMSTR=LF;  
  
** Import the CSV file. **;  
PROC IMPORT DATAFILE=CSV  
                OUT=WORK.CBP_totals  
                DBMS=CSV  
                REPLACE;  
  
RUN;  
  
** Unassign the file reference. **;  
FILENAME CSV;  
  
*****;  
** LOAD NEW TABLE **;  
*****;  
** LOAD DROPS TOTALS RECORD AND UNNEEDED FIELDS **;  
** TABLE IS PROMOTED TO PUBLIC **;  
*****;  
proc casutil;  
    droptable casdata="CBP_totals" incaslib="public" quiet;  
    load data=work.CBP_totals(keep=zip stabbr emp name  
        where=(zip NE '99999'))  
        outcaslib="casuser" casout="CBP_totals" replace;  
    promote casdata="CBP_totals" casout="CBP_totals"  
        incaslib="casuser" outcaslib=public keep;  
run;  
quit;  
  
*****;  
** TERMINATE CAS SESSION **;  
*****;  
cas casauto terminate;
```

APPENDIX 2 – CODE TO LOAD TIGER/LINE SHAPEFILE

```
*****;
** SHAPEFILE SOURCE **;
*****;
** https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2017& **;
** layergroup=ZIP+Code+Tabulation+Areas **;
** FILE UNZIPPED LOCATION: ~/SAS_Projects/ESRI_Maps/tl_2017_us_zcta510/ **;
*****;

*****;
** CONNECT TO CAS **;
*****;

options cashost=("cas","casbk") casport=5570;
cas casauto;
caslib _all_ assign;

*****;
** VIEW CONTENTS OF SHAPEFILE **;
*****;

%shpcntnt (SHAPEFILEPATH=
~/SAS_Projects/ESRI_Maps/tl_2017_us_zcta510/tl_2017_us_zcta510.shp);
** OBS= 52898242 VARS= 12 ID=GEOID10 **;

*****;
** RUN SHAPEFILE IMPORT **;
*****;

* DROP PREVIOUS TABLE IF FOUND *;
proc casutil;
droptable casdata="GEO_ZIP5" incaslib="casuser" quiet;
run;

* RUN IMPORT MACRO *;
%shpimprt (SHAPEFILEPATH=
~/SAS_Projects/ESRI_Maps/tl_2017_us_zcta510/tl_2017_us_zcta510.shp,
id=GEOID10, OUTTABLE=GEO_ZIP5, cashost=(cas casbk), casport=5570,
caslib=casuser);

*****;
** DROP TABLE FROM SASWORK **;
*****;
proc datasets lib=work nolist nowarn;
delete temp;
run;
delete GEO_ZIP5;
run;
quit;

*****;
** TERMINATE CAS SESSION **;
*****;
cas casauto terminate;
```