Paper SAS1732-2018

Leverage custom geographical polygons in SAS® Visual Analytics

Falko Schulz, SAS Institute Inc., Brisbane, Australia

ABSTRACT

Discover how you can explore geographical maps using your own custom map regions. SAS® Visual Analytics supports several predefined geo codes including various country and sub-division lookups. However often you have your own custom polygons or shape files drawing exact boundaries of the regional overlay you are trying to explore. From generic sales regions, floor plans or even pipe lines - there are many use cases for custom polygons in visual data analysis. Using custom regions is now easier than ever with a user-interface driven support for importing and registering these custom providers. This paper demonstrates not only the different types of custom providers supported but also shows how to leverage custom polygons within SAS® Visual Analytics by showcasing various industry examples.

INTRODUCTION

Location Analytics is becoming more and more important in today's world of data visualization with business data often associated with location information. A location represents an association to customer address, a store location or sales territory. Data can be stored as single point or a series representing waypoints of a route. Furthermore, locations may represent a specific area such as a state or country. Many organizations may also have their own regions defined from public known fire district or police zones to internal sales regions. Being able to visualize and explore such geographical regions in SAS Visual Analytics is central factor for every data exploration and reporting effort.

ABOUT LOCATION ANALYTICS

SAS Visual Analytics provides a robust and powerful platform to achieve location intelligence performed with a combination of GIS mapping technologies like Esri and SAS Analytics. It enables the creation and distribution of dashboards, reports and allows highly interactive exploration of business visualizations. A range of geo-mapping capabilities provides rich geo visualization from plain reference maps, street maps to high resolution satellite imagery. Customers may also choose to deploy their own GIS server for any special mapping requirements.

With more and more organizations understanding the importance of geo-tagged data, being able to combine business data such as events, weather conditions, social-demographics and more is important to make relevant and informed decisions.

MAPPING TECHNOLOGIES

SAS Visual Analytics leverages various mapping technologies to render geographical maps. By default, the environment provides access to maps provided by the OpenStreetMap community as well as public accessible maps from Esri ArcGIS Online. The latter provides a wide range of map styles from light canvas maps, street or terrain maps to high resolution satellite imagery.

An administrator may also choose to configure their own on-premise or cloud deployment of Esri ArcGIS Server which allows the access to any custom or special organizational maps.

Once configured a user can chose an available background map service in the user interface:

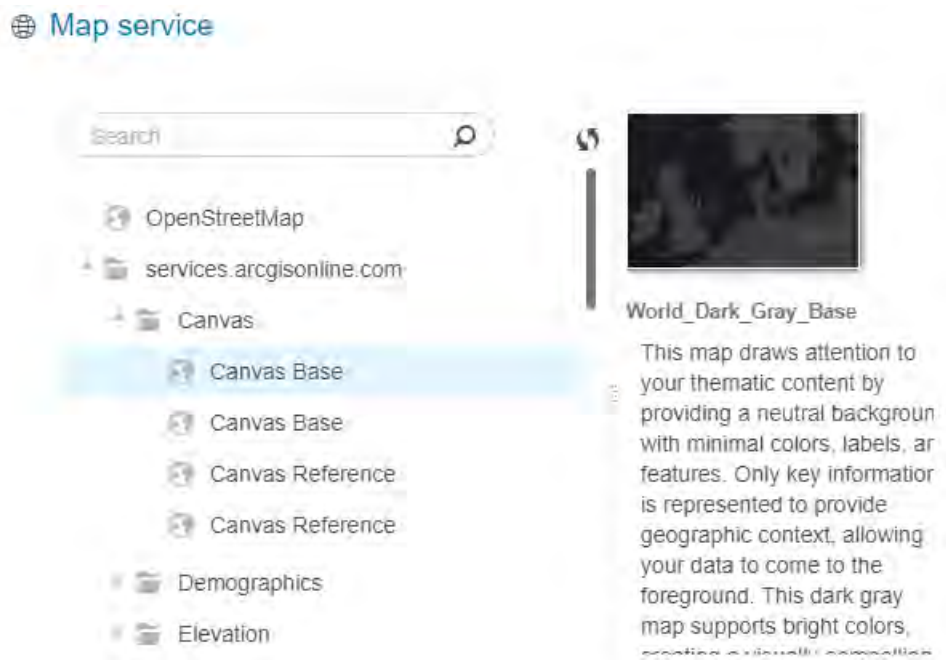


Figure 1 - SAS Visual Analytics map selector

SAS Visual Analytics provides various ways of rendering the data on top of a map. From specific data points or point clustering to regional areas. A user has the choice of the following visualization techniques:

- Bubble plot
- Coordinates plot
- Clustered coordinates plot
- Network plot
- Regional visualization

What type visualization to choose depends on the type of data as well as what location details are available. Network plots are great for things like routes, roads or pipelines but coordinate plots are better suited for customer addresses. Regional or thematic maps are best for geographical areas such as a county, state or sales region.

CUSTOM REGIONAL AREAS

So why do you need the ability to define your own custom regional areas? SAS Visual Analytics provides boundaries for predefined and well known-areas such as countries or state/provinces. Data is provided by GfK GeoMarketing GmbH [Gfk] and regularly updated. The user can select of list of pre-defined types:

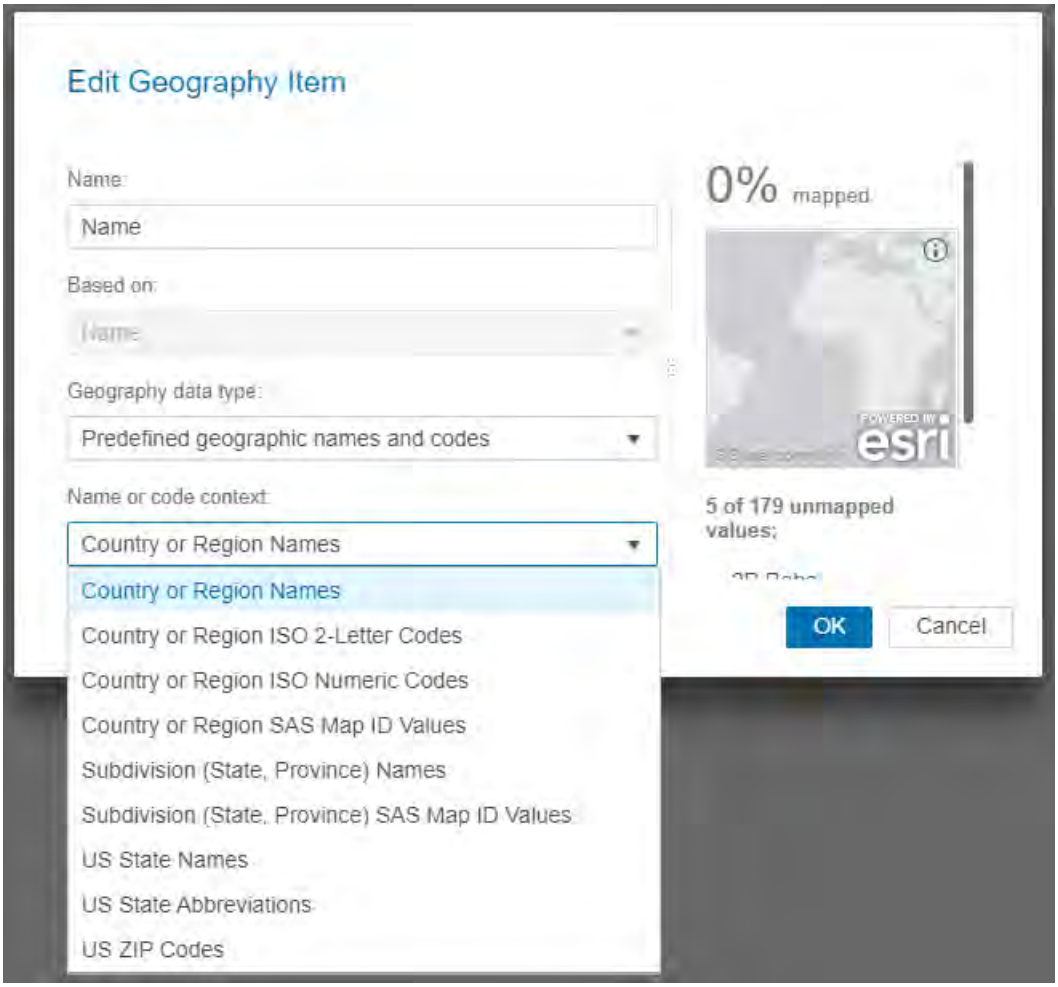


Figure 2 - Geography Item Editor with predefined contexts

For any other regional area or special region, the customer would need to register their own polygon data representing their own custom geographical areas. Examples of custom areas include sales regions, floor plans, property lots, campus maps, districts/zones, etc.

This paper describes the data requirements as well as steps to take for loading such data into SAS Visual Analytics.

Note, the latest version of SAS Visual Analytics 8.2 uses different techniques to load geographical boundaries and as such the steps described for Visual Analytics 7.4 [VA74AdminDoc] are no longer valid.

POLYGON PROVIDER TYPES

SAS Visual Analytics supports two types of polygon providers:

1. SAS Cloud Analytics Services table (CAS Table)
2. Esri ArcGIS Feature Service

If your organization has an existing cloud or on-premise ArcGIS server with feature services containing required polygon data you may choose #2 as your preferred option. An existing feature service doesn't require data import nor any additional storage in the SAS environment and provides quick access to polygon data.

If your polygon data is stored in 3rd party GIS applications or provided in other file formats, you will need to import the data first into SAS Viya.

DATA REQUIREMENTS

Polygon data are usually exported from special GIS applications or CAD systems. As such you may find data in formats such as shape (SHP) or AutoCAD files (DWG/DXF). Regardless of the input format SAS Visual Analytics requires polygon data loaded into CAS and the following structure of the map data set:

Required roles:

- **X/LONG** – represents the x- or longitude coordinate of the data point (Type: *Numeric*)
- **Y/LAT** - represents the y- or latitude coordinate of the data point (Type: *Numeric*)
- **ID** – the unique identifier of the polygon (Type: *Character*/*Numeric*)
- **SEGMNENT** – if a polygon consists of multiple disconnected areas this column identifies each segment (Type: *Numeric*)
- **SEQUENCE** – a column indicating the correct order of the data points. This column is used for sorting and ensures that correct order is maintained after joining (Type: *Numeric*)

Note, that the names used above are just examples and can be defined by the customer.

	X	Y	id	SEGMENT	sequence
1	-117.5525666	34.072260088	1	1	1
2	-117.5525883	34.072262732	1	1	2
3	-117.5525706	34.072364936	1	1	3
4	-117.5526555	34.072375237	1	1	4
5	-117.5526735	34.072377806	1	1	5
6	-117.5526283	34.072636648	1	1	6
7	-117.5525933	34.072643617	1	1	7
8	-117.5525005	34.072631576	1	1	8
9	-117.5525528	34.072333112	1	1	9
10	-117.5525666	34.072260088	1	1	10
11	-117.5501799	34.072434309	2	1	11
12	-117.5501945	34.072424196	2	1	12
13	-117.5502959	34.072437894	2	1	13
14	-117.5502382	34.072730165	2	1	14
15	-117.5501268	34.072714501	2	1	15
16	-117.5501407	34.072638593	2	1	16
17	-117.5501799	34.072434309	2	1	17
18	-117.5531943	34.074004649	3	1	18
19	-117.5529706	34.075175386	3	1	19
20	-117.5528498	34.075159681	3	1	20
21	-117.5528436	34.075194421	3	1	21

Figure 3 - Example of polygon data set

In order to import polygon data you may choose the provided utility macro `%SHPIMPRT` for importing shape files [Viya33ShapeImport] or select your own method of importing dependent on the data source. You may also utilize the SAS/Graph PROC MAPIMPORT [Graph94MapImport] to import shape files.

Note, regardless of the import method selected – all final polygon data sets need to have the columns outlined before especially the **SEQUENCE** column. To create a sequence variable to enable the polygon segments to be read in the correct order you can use a SAS DATA step and reference the _n_ automatic variable. For example, the following DATA step creates a sequence variable for the MYMAP data set:

```
data mymap;
  set mymap;
  sequence = _n_;
run;
```

Optional you may want to consider reducing the polygon data set size by reducing the density of data points required to render each polygon. Subset your polygon data to decrease the level of detail and improve performance. Reducing the level of detail might also enable you to display a greater number of map regions at one time.

If you have a license for SAS/GRAPH, then you can use the GREDUCE procedure to create a DENSITY variable that enables you to reduce the density of your polygon data. Depending on the source of your map data sets, a DENSITY variable might already be present.

You can use the DENSITY variable in a WHERE statement in a DATA step to reduce the detail in your polygon data. For example, the following DATA step reduces the MYMAP data set to exclude segments that are density level 4 or greater:

```
data mymap;
  set mymap;
  where(density<4);
run;
```

Note: By default, SAS Visual Analytics can retrieve up to 250,000 polygon vertices at a time. If you encounter an error message in a geo map object about the number of polygon vertices, then you might need to reduce the density of your polygon data or filter the data query for your geo map object. In some cases, a very wide ID column in your polygon data can further limit the number of polygon vertices that are retrieved. Check the width of your ID column in SAS Data Explorer if you encounter this message.

If you have data in other formats, e.g. AutoCAD you will need to convert the given file into a shape file using GIS software. Applications such as the open-source QGIS or Esri's ArcMap are very popular and powerful tools. The screenshot below shows the import of an AutoCAD .DWG file into ArcMap and related export dialog to save data into a shape file.

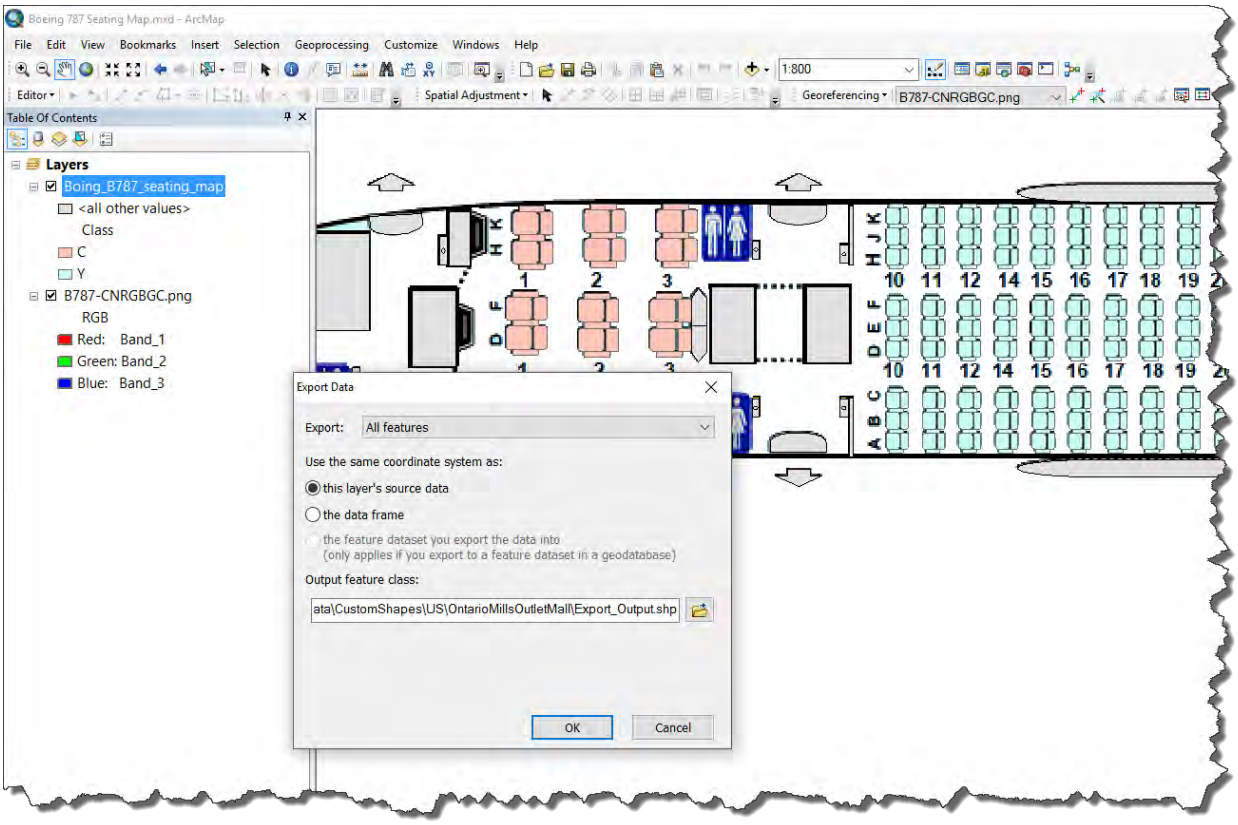


Figure 4 - Esri ArcMap - Export to Shape File

You may also use GIS applications to adjust the projection used in the data. The most common way to reference data points is using unprojected data (latitude/longitude) which also represents the easiest way to register the data later. Users who are not familiar with coordinates spaces and related projections may find it difficult to use your polygon data otherwise.

REGISTRATION

To utilize custom polygonal shapes an administrator will need to register each polygon data source as provider. SAS Visual Analytics 8.2 provides an interface for registration which is accessible from within the geographical item editor. If you select **Custom polygonal shapes** as the geography data type, a list of existing providers is shown.

An administrator can manage and add new polygon providers. An organization may nominate other users for maintaining polygon providers if required. The administrator guide contains details how to setup required permission and group membership.

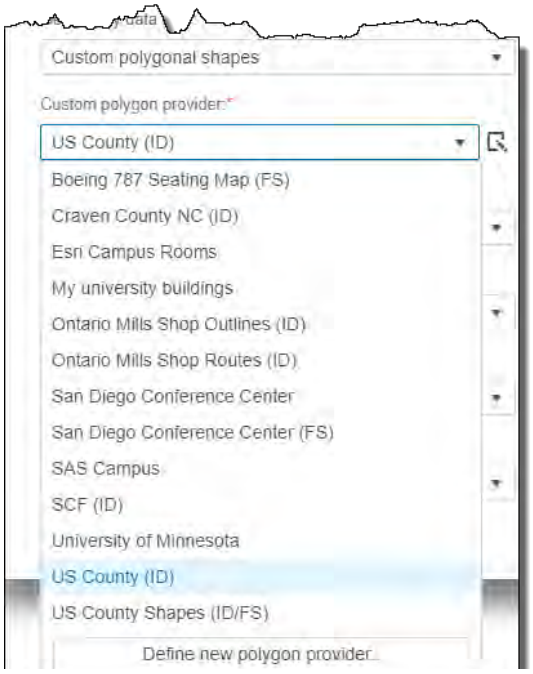
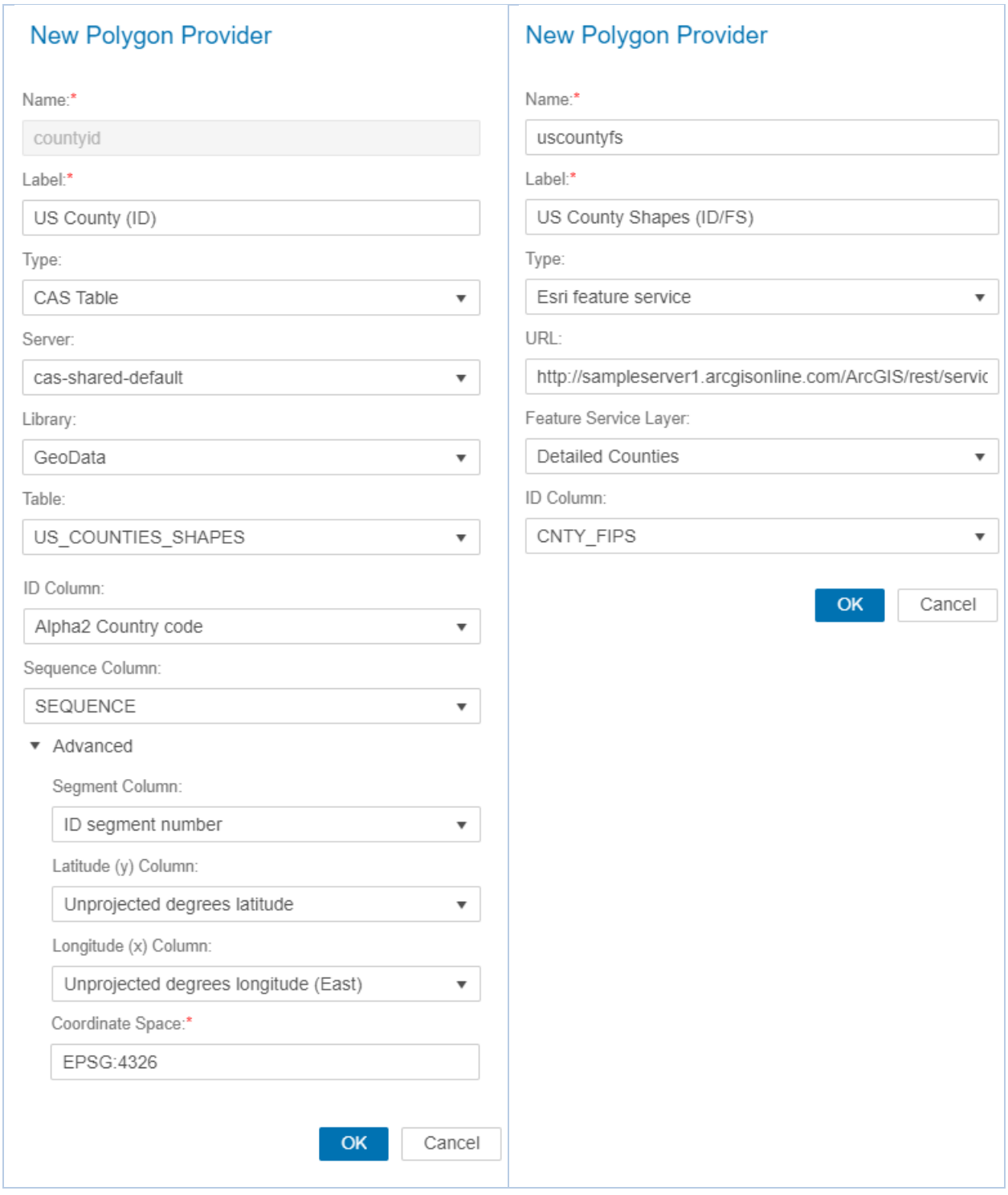


Figure 5 – List of available polygon providers

Given required permissions, registering a new provider is done by nominating a unique name and a user-friendly label. Dependent on the type of the provider (CAS Table vs Esri Feature Service) – you will either need to reference a table or enter the URL of the Esri feature service. The CAS table provider requires additional column mappings and specifying the projection used in the data with the default being WGS84.

The following table shows examples for both provider types and selected values. Each provider may be different so make sure to reference values correctly here or geographical maps may not render correctly.

CAS Table	Esri Feature Service



Pay special attention to the **ID** column selected which is required for both types. The value in this column is used as mapping column. Each unique value should match corresponding values in your report data source. The geography item editor will show mapping statistics in the right panel to give you an indication how well a mapping would work:

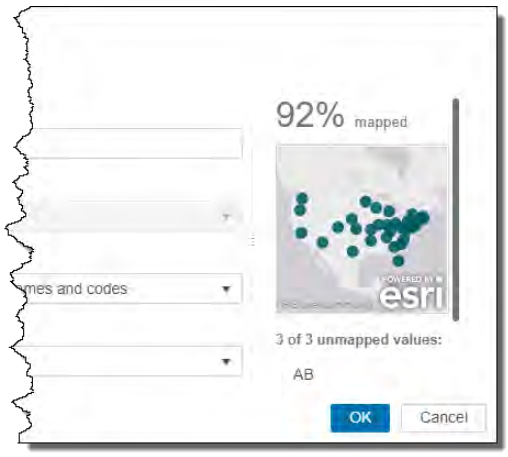


Figure 6 - Geography Item Editor - Mapping Statistics

Note, if you get less than 100% mapped, some data points may be missing during visualization rendering. A related warning sign in the bottom right corner will indicate the missing data.

For more information about registering custom polygon shapes see the VA administration guide [VA82AdminGuide]

USAGE

For a business user, there isn't any additional configuration required to use existing polygon providers. As with any other data item which represents a geographical location the user will need to change the data item classification to be **Geography**. The geography item editor will provide various options to select the correct mapping. As per latest release of Visual Analytics you can select from the following geography data types:

1. Predefined geographic names and codes
2. Custom polygonal shapes
3. Custom coordinates

This paper focuses on type #2 which represent all custom polygon providers which may have been setup before. Select a provider which matches your data source and select the corresponding region ID. Make sure the column selected contains ID values matching the one in the configured provider. A related mapping statistic will give you an indication how well the selected column matches the provider.

VARIOUS MAPPING EXAMPLES

1. US COUNTY BOUNDARIES

This example shows how to register a polygon provider representing the shapes of the United States counties. For the purpose of the example an existing Esri feature service is used provided by the public accessible Esri ArcGIS online examples. The feature service http://sampleserver1.arcgisonline.com/ArcGIS/rest/services/Demographics/ESRI_Census_USA/MapServer/4 provides such boundary details with the field **FIPS** (5-digit) being the unique identifier for each polygon. Values in this field represent the combined FIPS code for state (2-digit) and county (3-digit). This means data loaded into Visual Analytics will also need to include such ID for correct lookup.

The example uses a data set containing all postal codes of the United States. You can retrieve a copy of the data set if required [ZIPCODE].

The data set also comes with fields for the state FIPS code and county FIPS code. In order for Visual Analytics to lookup county polygons we need to provide a column containing the exact ID as per provider data source. This means it's required to combine both fields into one 5-digit long data item. The following formula is used for a new calculated field named CNTY_FIPS.

The following table shows a preview of the data available (first two columns) and the data item generated (3rd column):

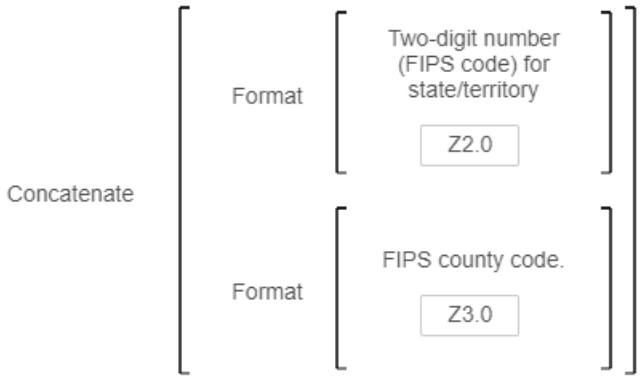


Figure 7 - Calculated field formula to combine state and county FIPS

Two-digit number (FIPS code) for state...	FIPS county code.	CNTY_FIPS
36	103	36103
36	103	36103
72	1	72001
72	3	72003
72	5	72005
72	5	72005
72	5	72005
72	93	72093
72	11	72011
72	141	72141
72	13	72013

Figure 8 - Preview of calculated field CNTY_FIPS

With this new column in place we can change its classification to Geography and create a new custom polygon provider. The following options were selected for this Esri feature service based provider:

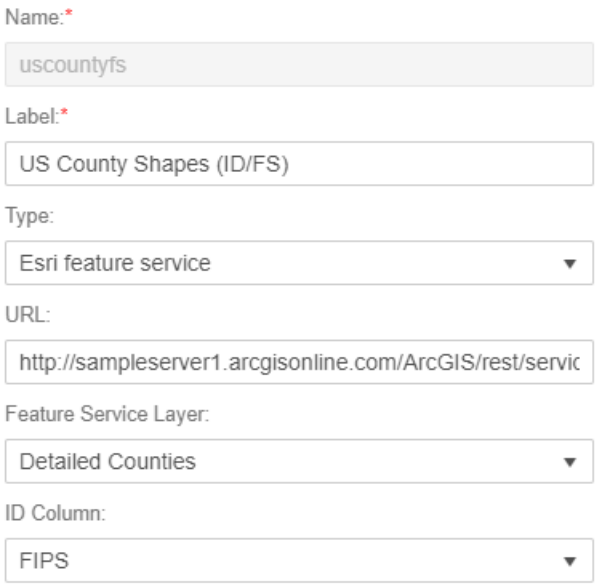


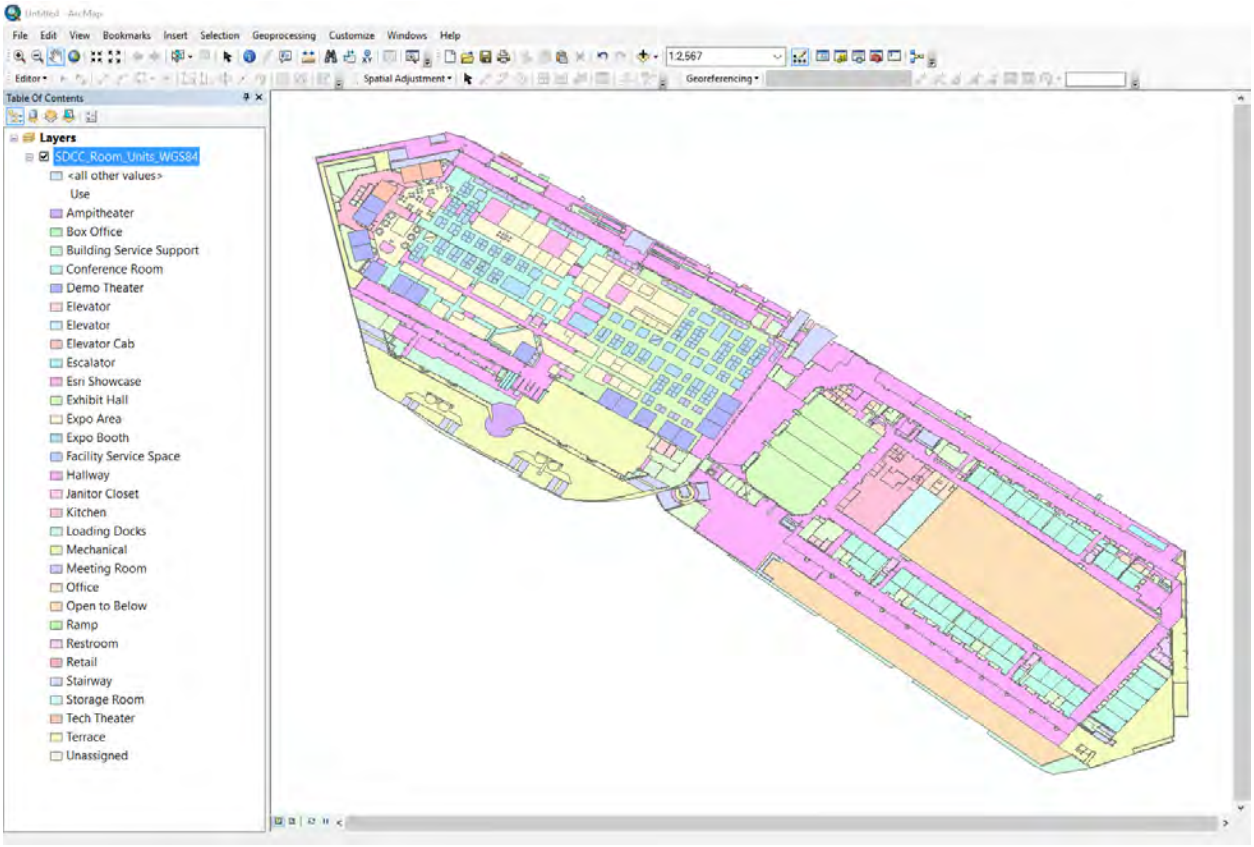
Figure 9 - Esri feature service provider for US counties

Selecting this provider and using the previous created CNTY_FIPS column as Region ID, provides our new geography data item for US counties. Dropping this data item on the report canvas produces the expected boundaries for the United States counties:

Figure 10 - Example visualization for US county boundaries

2. BUILDING FLOOR PLANS

The following examples shows the boundaries of the San Diego Convention Center [SDCC] including all rooms and levels. The original project was loaded into Esri ArcMap and related boundaries were exported as shape files.

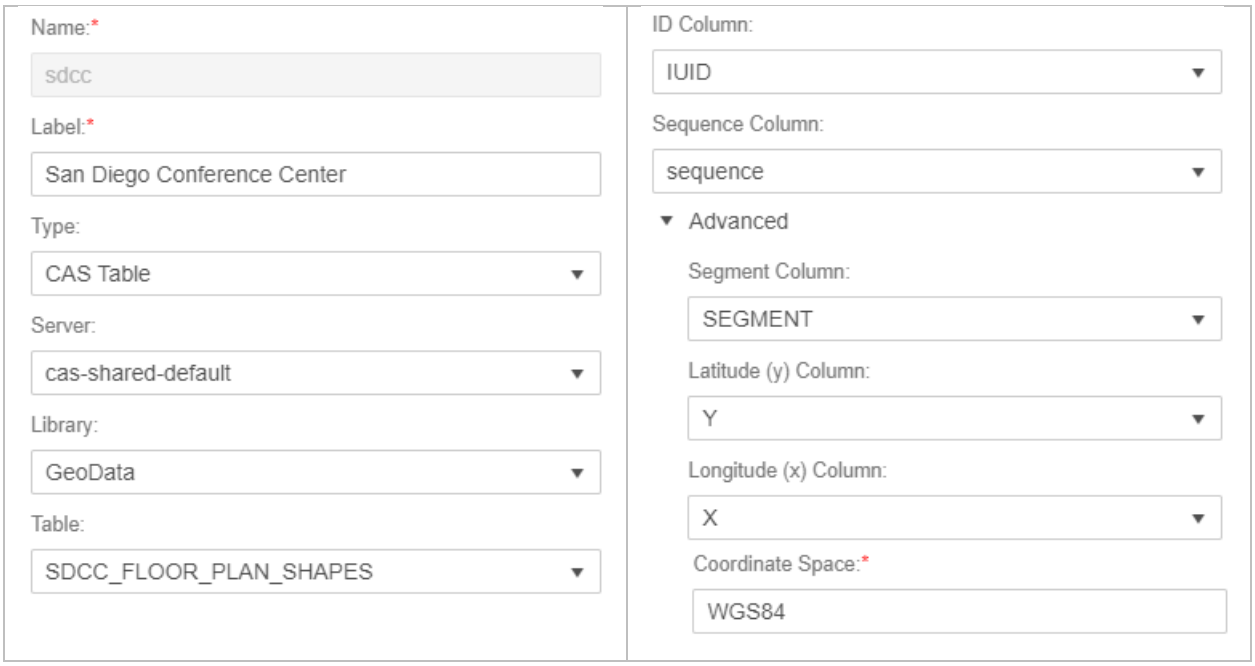


The polygonal data now being available as Esri shape file – we can use the following command to import the data into SAS:

```
proc mapimport datafile="C:\ConferenceFloors\sdcc_floor_plan_wgs84.shp"
      out=work.sdcc_floor_plan;
      id iuid;
run;

proc sql;
      create table mapscstm.sdcc_floor_plan_shapes as
      select x, y, iuid, segment, monotonic() as sequence
      from work.sdcc_floor_plan;
quit;run;
```

Once the data is loaded into CAS, the administrator will be able to configure a new polygon provider referencing this table:



To showcase the usage of boundary data the author created a simple SAS Visual Analytics report allowing the selection of the desired floor, a room search facility as well as usage selection:

Figure 11 - SAS Visual Analytics report showing floor plans for San Diego Convention Center

4. CAMPUS AND AREA MAPS

This example uses data for the buildings of the University of Minnesota [UOM] campus. Similar to the previous example the data has been loaded into Esri's ArcMap first and relevant polygon data have been exported into shape files.

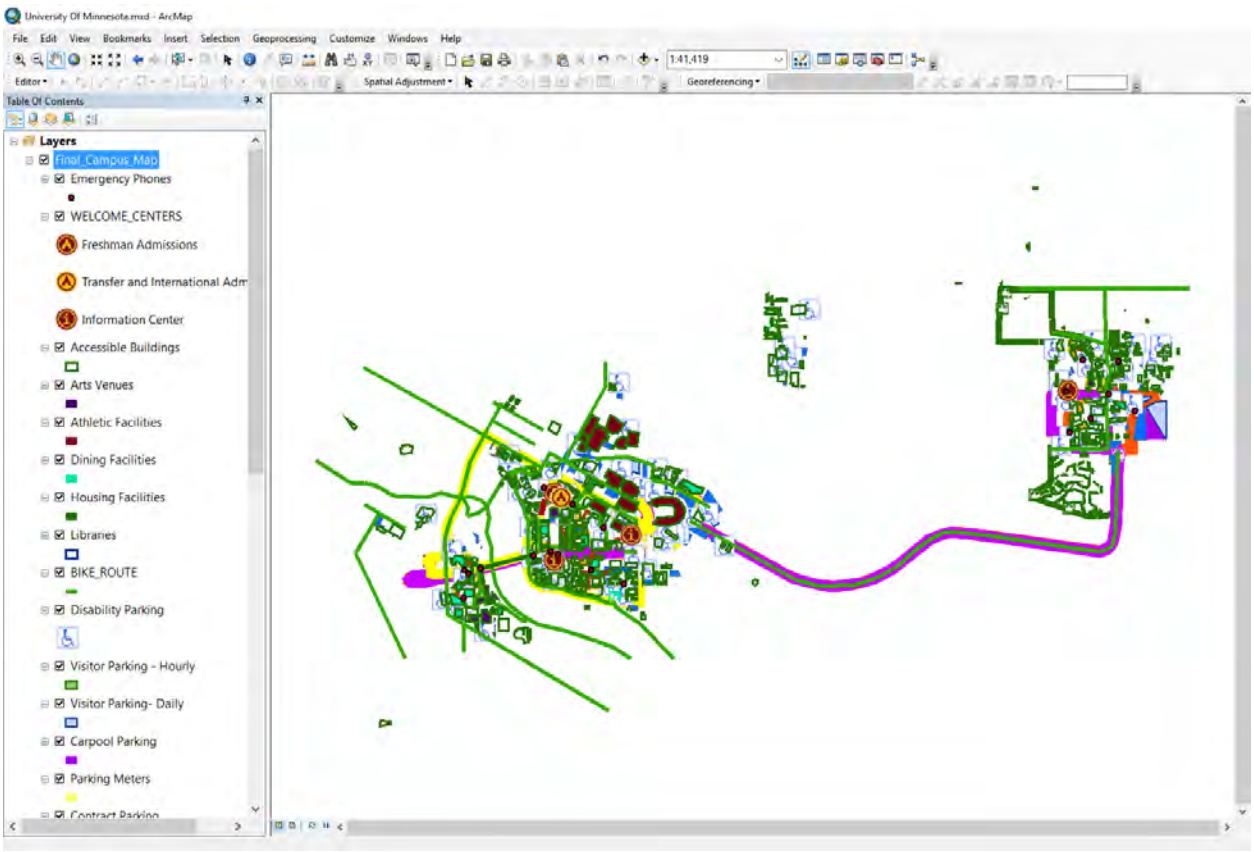


Figure 12 - Esri ArcMap project with polygonal data for campus buildings

A final SAS Visual Analytics report has been created to explore the campus given details available.

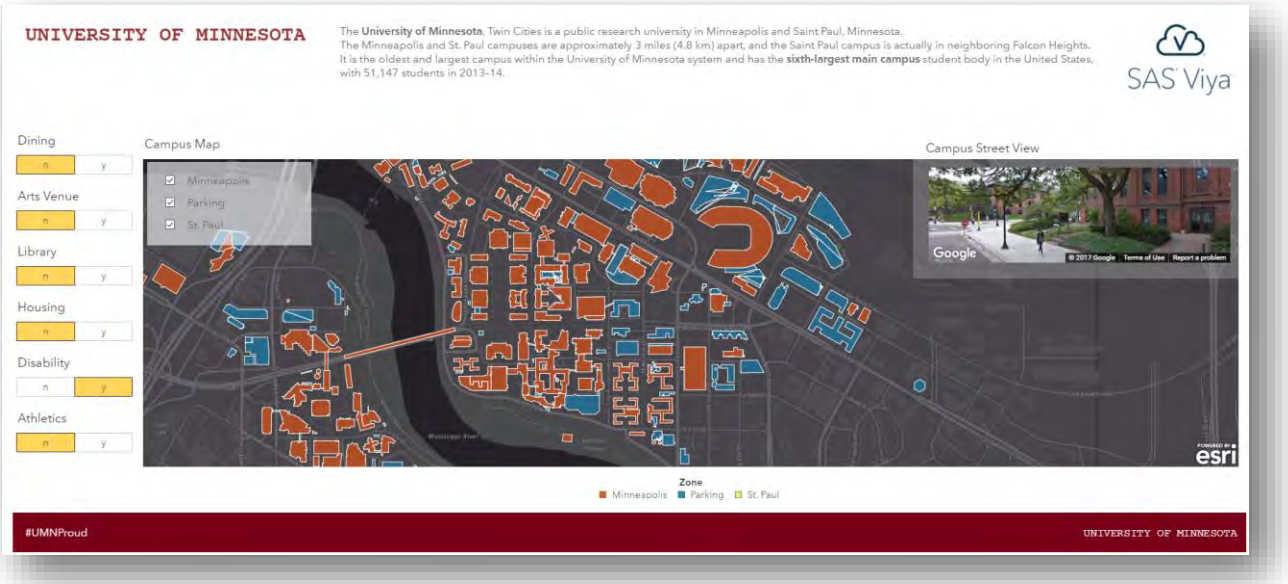


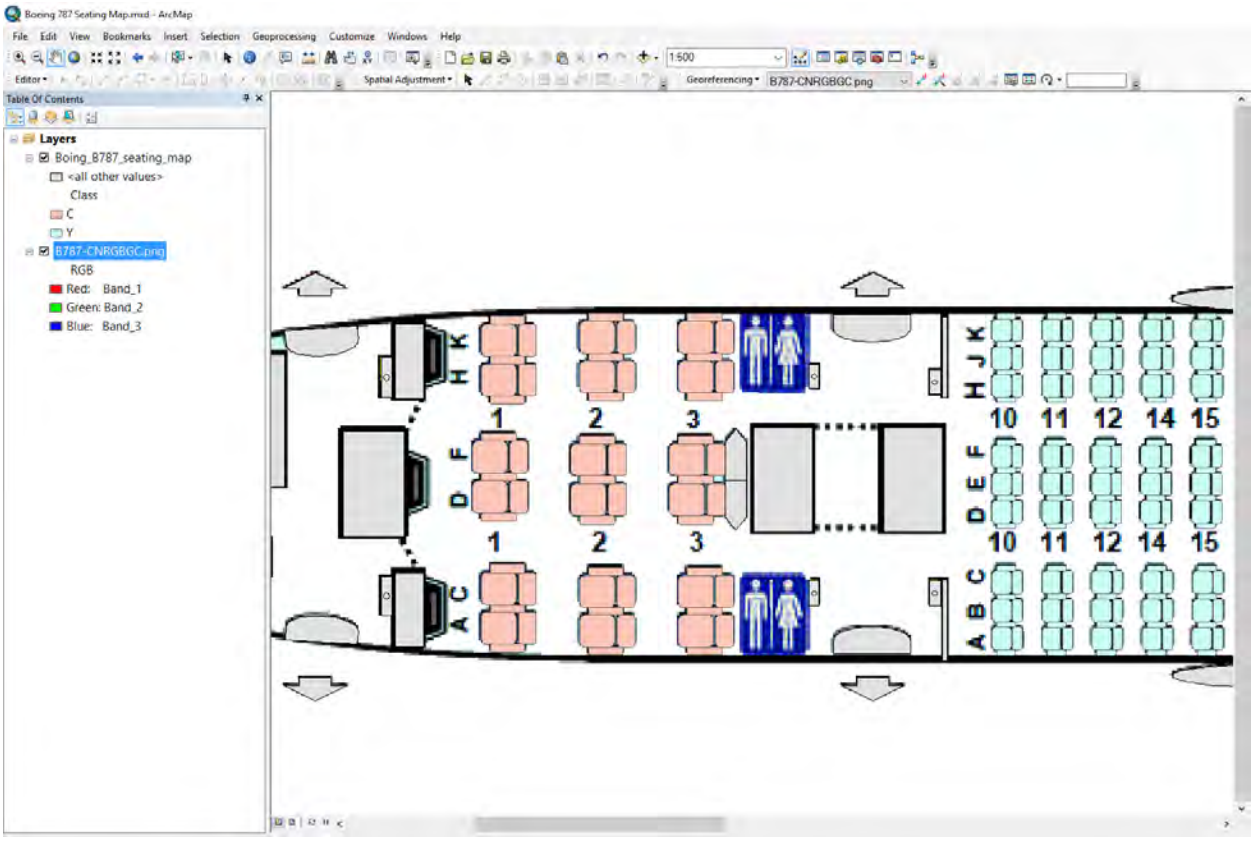
Figure 13 - SAS Visual Analytics report showing the University of Minnesota campus

3. SPECIAL FLOOR PLANS

This is an example using non-standard floor plans. Sometimes you have non-geo referenced floorplans which aren't aligned to any location on earth. As such you can't use a standard background map as you would with any other geo referenced data source.

As example, consider an aircraft seating map with the actual seat being the polygonal shape we are trying to analyze. It may be used by airlines or travel agencies to determine popular seats or simply to lookup seats given a criterion (e.g. exit or seat with disability access).

The data for this example has been prepared using Esri's ArcMap using the built-in editor capabilities. A similar shaped aircraft has been used as reference and as a result seat boundaries may not reflect a real seating layout.



The main difference compared to the previous examples - is that we can't use a standard geographical map background here given the polygons aren't geo referenced. Given that SAS Visual Analytics also supports custom Esri map services the author geo-referenced a custom drawn aircraft image (PNG) and published a new tile-map service.

After polygon provider registration, a new geo map can now reference a custom map service using the service selector:

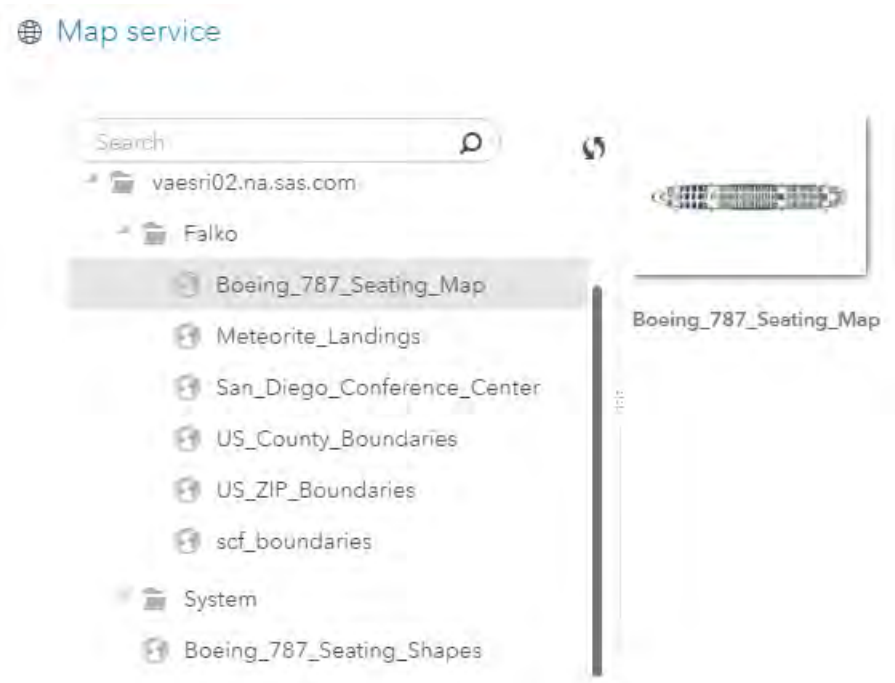


Figure 14 - Visual Analytics map service selector

Using the new polygon provider and new map service in a SAS Visual Analytics report allows analysis of the seats and built-in interactivity allows quick selection using filter options. Note the data for this report have been created using random data for demonstration purposes.

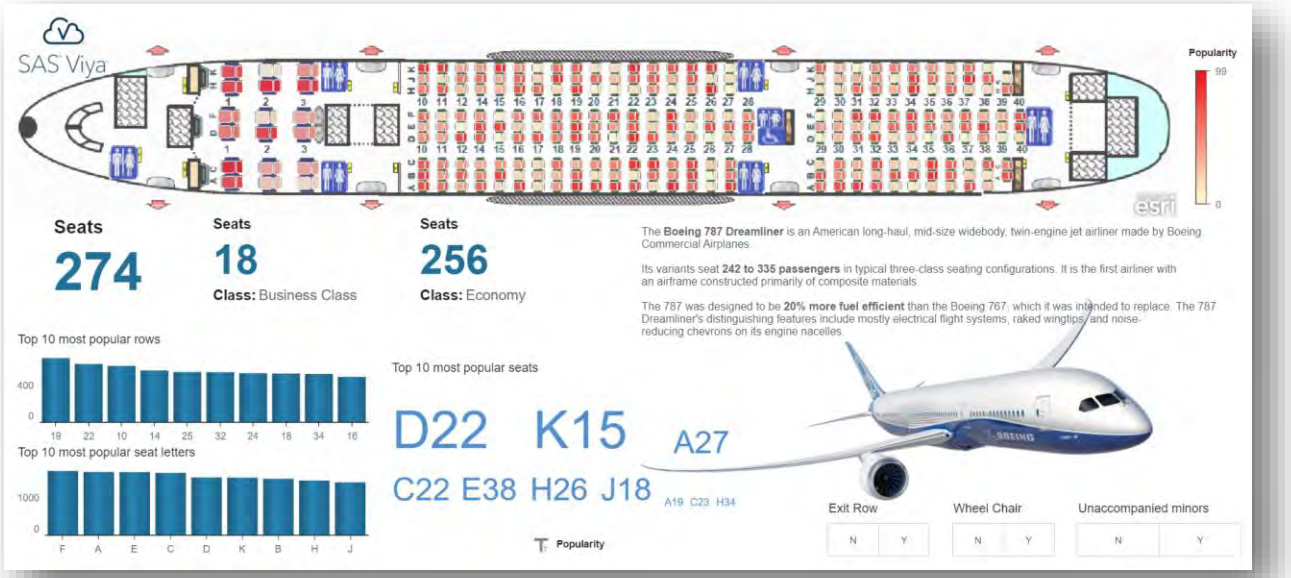


Figure 15 - Visual Analytics report showing aircraft seating map

4. STORE AND TRAFFIC ANALYSIS

The last example shows the building layout of an outlet mall [Ontario] in combination with a special polygon layer showing customer route density. Such layers can be generated using Esri ArcMap and may add additional detail to the geographical area selected for analysis. In the example of the Ontario Mills outlet mall data has been generated to showcase this.

Similar to the example before data has been prepared in ArcMap but this time two of the polygonal layers have been exported to shape files. One representing the building shapes and one the customer route hotspots:

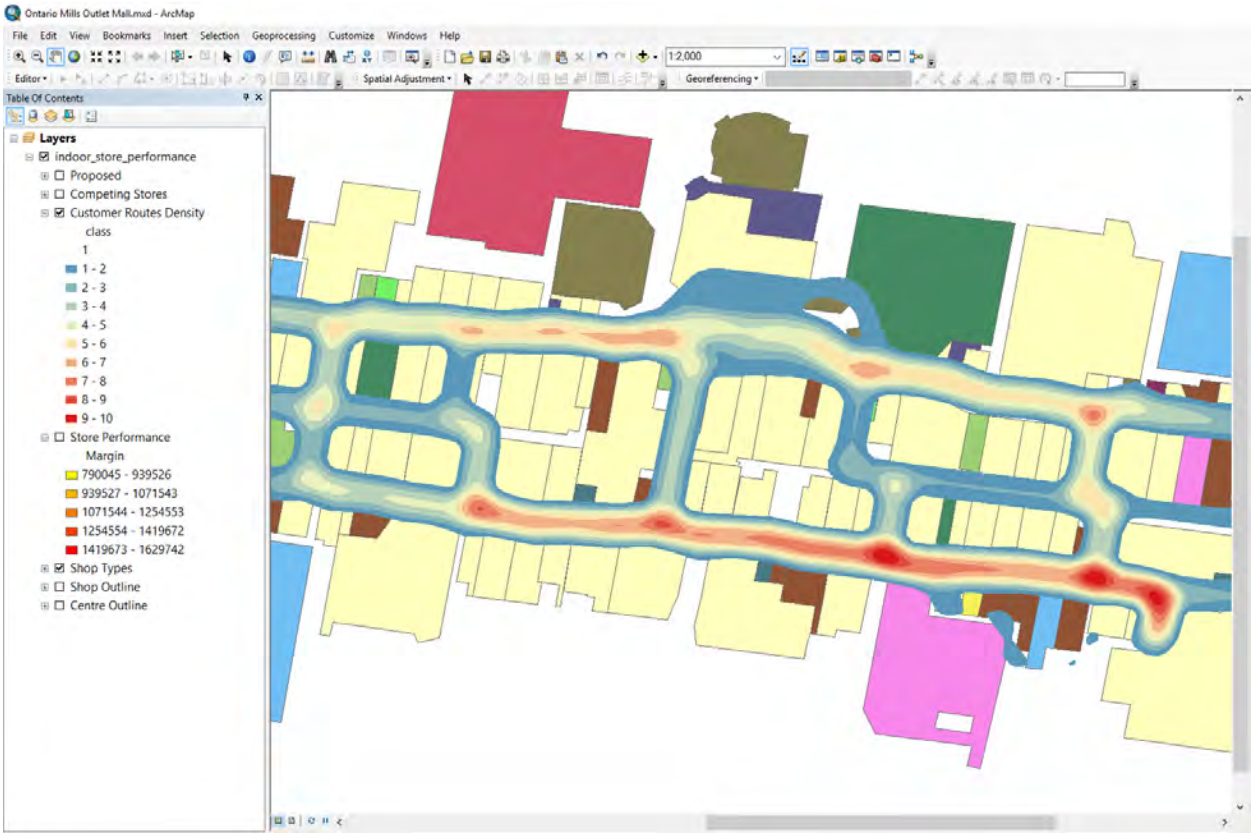


Figure 16 - Esri ArcMap project for outlet map building shapes and customer routes

With custom providers registered and two geo maps created using these providers – the following SAS Visual Analytics sample reports shows how retail store performance analysis could look like in such scenario:

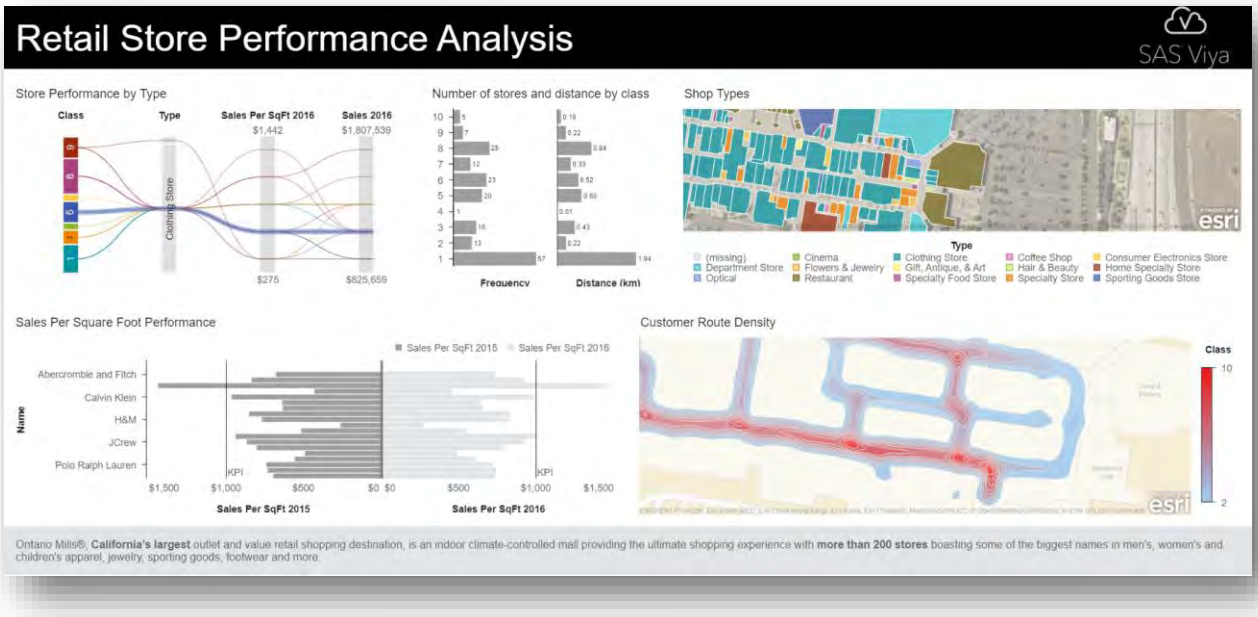


Figure 17 - SAS Visual Analytics report showing retail store performance analysis

CONCLUSION

The ability to import any form of shapes or boundary data into SAS Visual Analytics is a powerful feature. Not only can customers now analyze and visualize non-standard geographical areas but they can also fully understand complex business scenarios and make better decisions based on accurate information. With the support for any type of shape data, from county or province boundaries to complex floor plans, SAS Visual Analytics 8.2 can now be used for advanced location analytics.

With the combination of built-in support for location analysis tools such as clustering, drive-time/live-traffic selection and more – SAS visual Analytics adds valuable insights into your geo-tagged data and creates real value for organizations.

REFERENCES

Schulz, Falko. “More Than a Map: Location Intelligence with SAS® Visual Analytics”. August 19, 2014. http://support.sas.com/resources/papers/proceedings14/SAS021-2014.pdf .

[GfK] The map data sets in library MAPSGFK are based on the digital, vector-based maps from GfK GeoMarketing GmbH and are covered by their copyright. http://support.sas.com/mapsonline/gfklicense .

[VA74AdminDoc] SAS Visual Analytics 7.4 Administration Guide, Adding Custom Polygon Data for Geographic Maps, November 2nd 2017, https://support.sas.com/documentation/cdl/en/vaag/69958/HTML/default/viewer.htm#n0g5kjtnbvsrwbn1lxl2wahyc5du.htm

[Graph94MapImport] SAS 9.4 Programming, SAS/GRAPH: Mapping Reference, MAPIMPORT Procedure

http://go.documentation.sas.com/?cdcId=pgmmvacdc&cdcVersion=9.4&docsetId=grmapref&docsetTarget=p031mmm914jrkwn1e4r5ec19xcvm.htm&locale=en

[Viya33ShapeImport] SAS® Viya™ 3.3 Administration, Loading Geographic Polygon Data as a CAS Table

http://go.documentation.sas.com/?docsetId=caldatamgmtcas&docsetTarget=p1dwawsidsczlpn121j0glleicxp.htm&docsetVersion=3.3&locale=en

[VA82AdminGuide] SAS VA Administration, Working with Geography Data Items, Create a Geography Data Item By Using Custom Polygonal Shapes

http://go.documentation.sas.com/?cdcId=vacdc&cdcVersion=8.2&docsetId=vareportdata&docsetTarget=p031vp9uc5y5iun0zipy3c1trkqn.htm&locale=en#p1a7fw0vj2w6aln18sl77suu45o3

[ZIPCODE] SAS Technical Support, Latest United States zipcode dataset, http://support.sas.com/rnd/datavisualization/mapsonline/html/misc.html

[SDCC] Esri ArcGIS Map Gallery, SDCC_RoomUnits, http://www.arcgis.com/home/item.html?id=7890a4576813473f8f0384f26a474c2c, data obtained November 2017

[UOM] Esri ArcGIS Map Gallery, TC_Campus_Map_Layers_ALL, http://www.arcgis.com/home/item.html?id=84e9c3e04c394e5c972627082d0ef6a8, data obtained November 2017

[Ontario] Esri ArcGIS Map Gallery, OntarioMillsShop, http://www.arcgis.com/home/item.html?id=bced8dd994f843cea0c7595a13f56a05, data obtained November 2017

ACKNOWLEDGMENTS

I would like to thank Jeff Phillips and Eric Short for their assistance about details on geo analysis and their critical role in improving the Geographical visualization component and associated map algorithms.

RECOMMENDED READING

- *SAS® Viya™ 3.3 Administration*

CONTACT INFORMATION

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

Falko Schulz
SAS Institute Australia
1 Eagle St
Brisbane, QLD, 4001
Phone: +61 7-3233 320
Falko.Schulz@sas.com
http://au.linkedin.com/in/falkoschulz