SAS Visual Analytics 7.3

An Enterprise Guide project to import Polygon shape files into SAS VA.

Marcel Toledo
me@mrclll.in

Contents

License

Purpose

The purpose of this document is to explain how to import polygon shape files into SAS Visual Analytics v7.3.

An Enterprise Guide project has been created to facilitate the process and it is available for download at https://github.com/mrclll/theSASGitRepository.

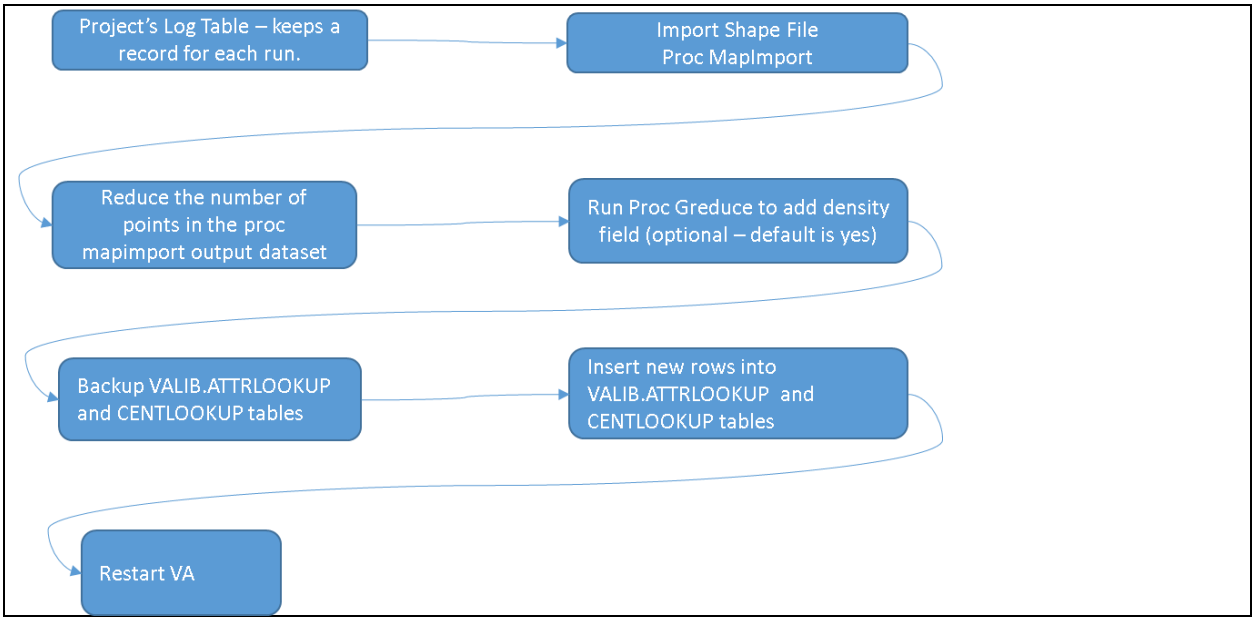
At the end of this document we will have imported two Australians Administrative regions into Visual Analytics.

Go to http://biogeo.ucdavis.edu/data/diva/adm/AUS_adm.zip to download the shape files.

Notice that you must have read the below document and executed its configuration steps before you start the steps shown in this tutorial.

> Document : **SAS Visual Analytics 6.4 Adding Belgium Regions** by David Demeyer. Download it from https://communities.sas.com/t5/SAS-Visual-Analytics/Import-shapefile-in-SAS-Visual-Analytics-v7-1/td-p/224983?attachment-id=8606
>
> Required steps:
>
> - 3.1 Define Custom Maps Library
> - 3.2 Define Geographical Data Library
> - 3.3 Back-up of Lookup Tables

Project Flow overview



Above is a rough view of the E.G. Project. When running the project, you will be asked to enter a sort of value through prompts. These are used to handle our shape file. The parameters are discussed later in this document.

Project prompts

Clean Up <YES|NO>

It is to do with the project's internal log table that keeps track of previous imports that you may have done. By choosing the YES option, it will read the MAPSCSTM._IMPORT_LOG log table and look for a row that contains the same FULL PATH as the file that you are currently importing. If the row is found, it will delete all of the contents related to that specific shape file as follows:

VALIB.ATTRLOOKUP <this Shape's file previous rows>

VALIB.CENTLOOKUP <this Shape's file previous rows>

MAPSCSTM <this Shape's file previous table>

MAPSCSTM._IMPORT_LOG <this Shape's file previous rows>

It is a helpful option when you are importing the same shape file multiple times to sort out problems or tweak density and precision.

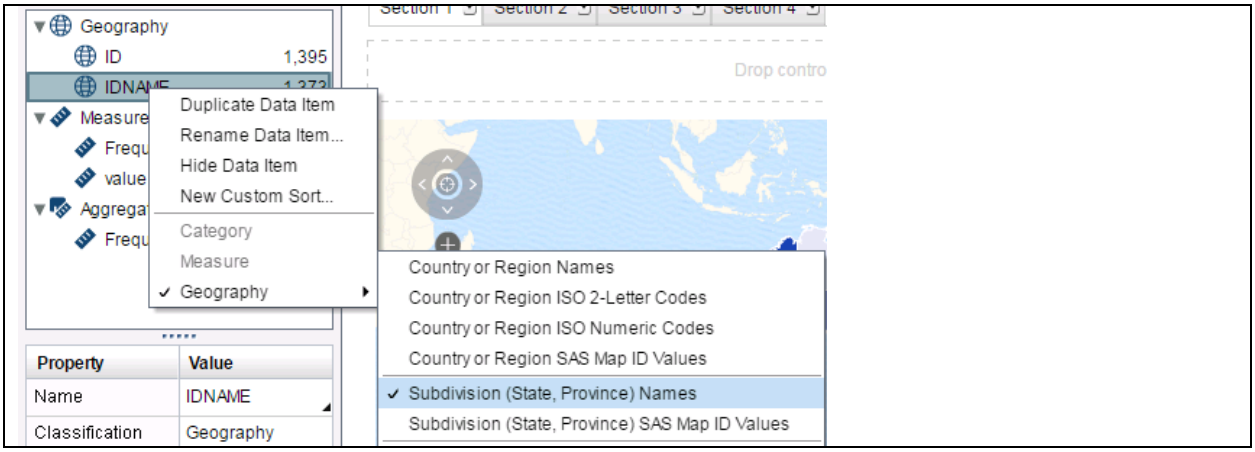
Render maps by density <YES|NO>

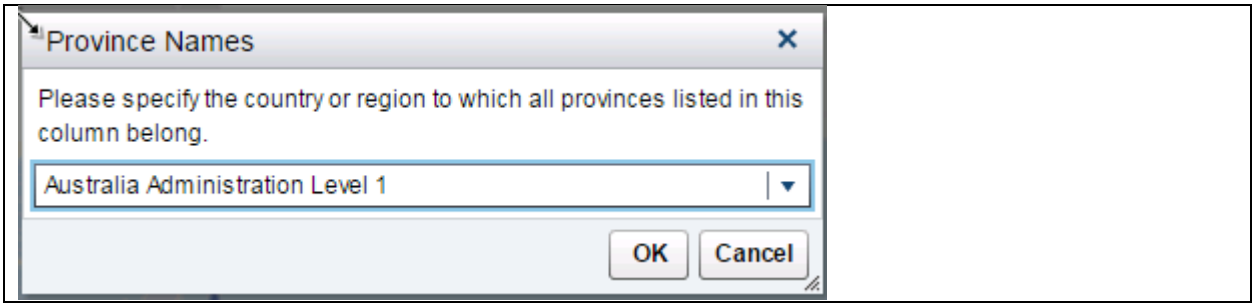
Used only to view how the shape file is rendered by SAS products, it will plot a map - using PROC GMAP – for each density level.

Map Name

This value will be displayed in Visual Analytics when creating a Geography type variable.

E.g.





Shape File Path

Enter the path to the shape file – extension .SHP - in the operational system folder. In the same folder also must exists the mandatory files .SHX and .DBF.

Unit Area Id <DBF Column>

In the DBF file, it is the column that is unique for each unit area (polygon shape). E.g. the column that identifies a Country, state, county and so on.

Unit Area Description <DBF Column>

In the DBF file, it is the column that contains names or description for a unit area.

Administration Type

It will not be displayed in Visual Analytics however it is necessary to fill out the final shape's file table in the MAPSCSTM library.

Precision

Points (X,Y coordinates) in the Proc MapImport output dataset will be rounded by this number (e.g. 0.001 , 0.01 , 0.1) . It will speed up the Proc GREDUCE execution. It may also impact in the map's render quality. Min (0.001) Max (0.9)

Run Proc GReduce?

Sometimes, you can come across a shape file that the Proc GReduce will not finish up and it will be running for ages. By selecting the NO option it will skip the Proc GReduce step.

Work around : You can increase the Precision value (e.g. from 0.01 to 0.1) to reduce the number of points in the final dataset.

Density

From 1 to 5 it will define the map's render quality. 1 is the lowest quality and 5 is the highest. A high quality will impact the rendering time. Combine the Precision value with Density to get an optimal quality / computing time.

Spatial Reference System

Leave it blank if the current reference system is EPSG:4326, WGS84, or any other form that may refer to the WGS84 reference system. Otherwise, it means that we need to project from the current reference system to the WGS84 which is used by VA. Enter a value as for example:

- +proj=tmerc +lat_0=53.5 +lon_0=-8 +k=1.000035 +x_0=200000 +y_0=250000 +ellps=mod_airy +towgs84=482.5,-130.6,564.6,-1.042,-0.214,-0.631,8.15 +units=m +no_defs
- EPSG:2157

For additional information visit https://epsg.io/ , http://spatialreference.org/ref/epsg/wgs-84/ .

Final table name

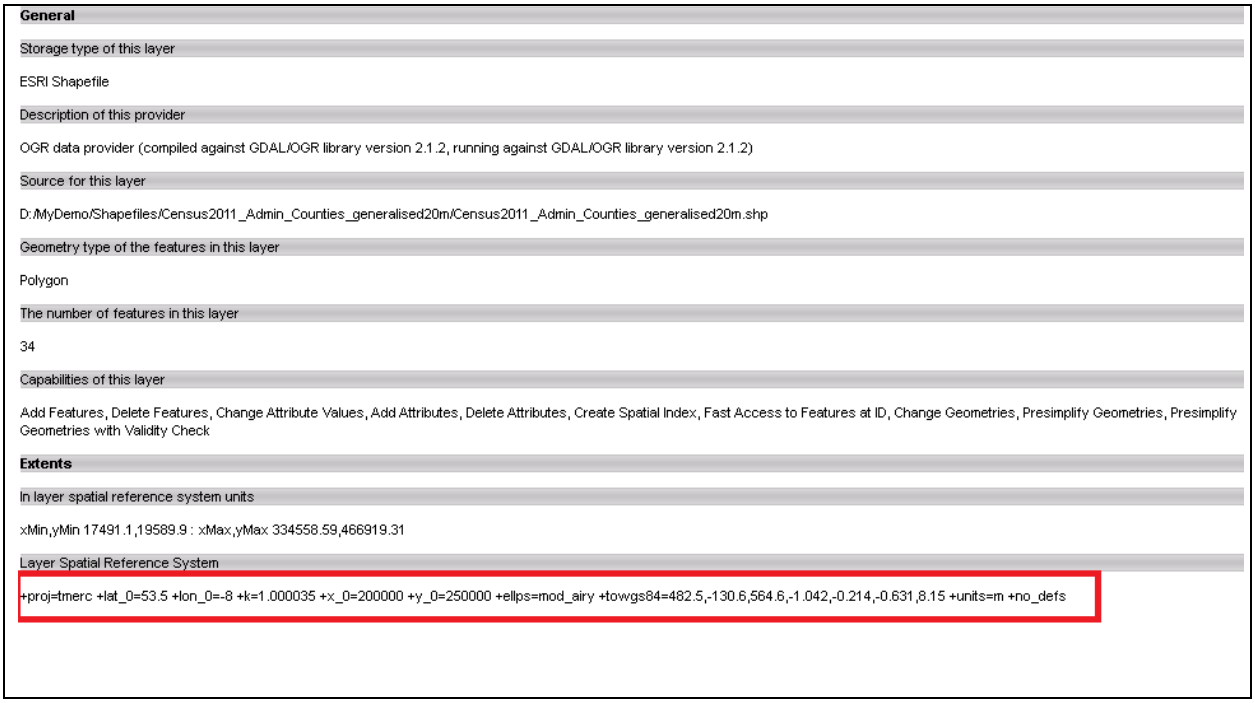
Enter a name for the shape file SAS dataset. It must end with the number 1. E.g. Administration_Level1_1 / Administration_Level2_1.

QGIS

If you want to explore your shape files before you start importing into SAS you can use QGIS to open them. It is a freeware software (even for business) to download it visit http://www.qgis.org/en/site/forusers/download.html.

Spatial Reference System information

In the QGIS Brownser, it is in the Metadata tab.



Importing the AUS_adm1.shp file.

Here are the text values to copy and paste.

Map name : AUSTRALIA ADMINISTRATION LEVEL 1

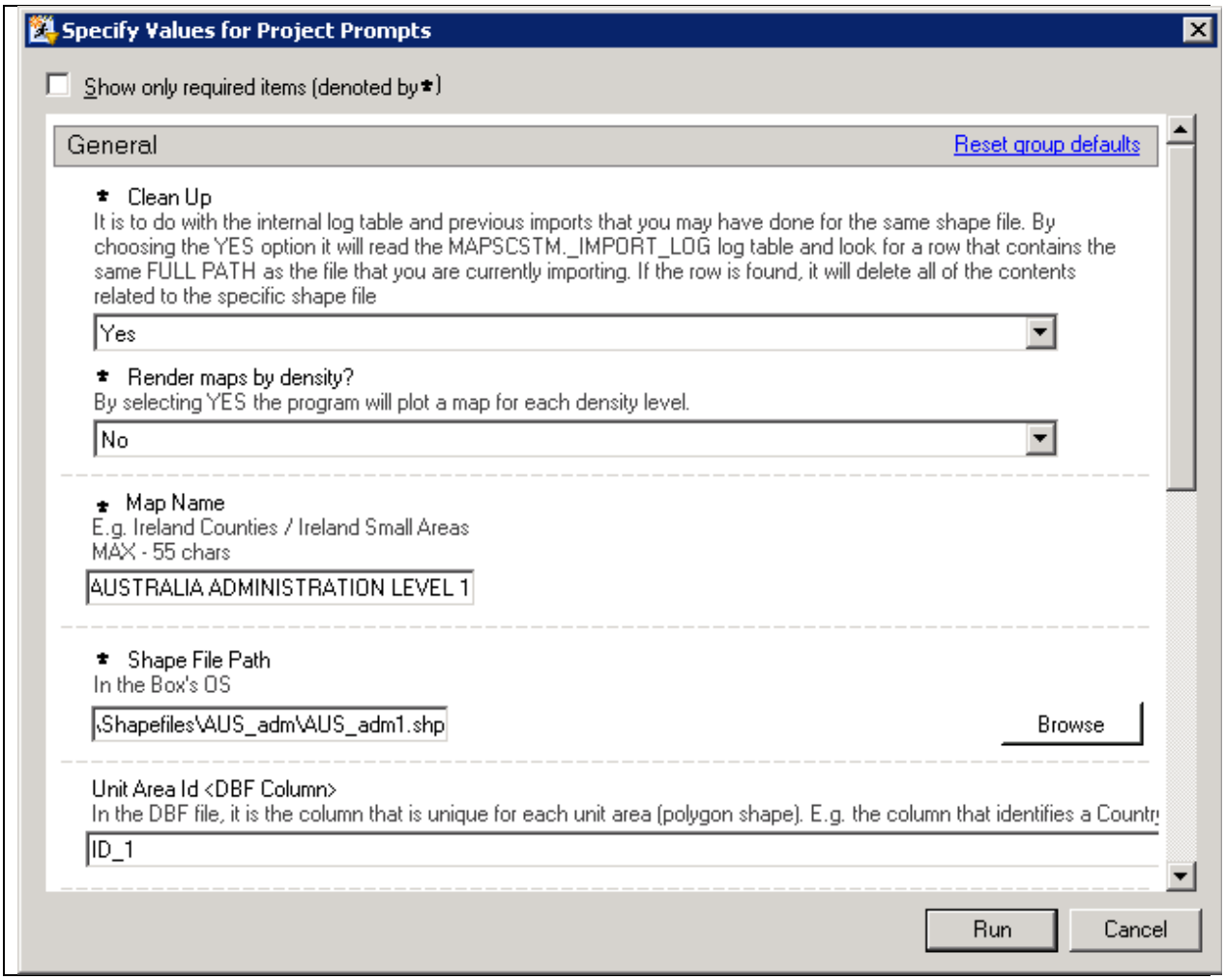
Shape File Path : <PATH TO>\ AUS_adm\AUS_adm1.shp

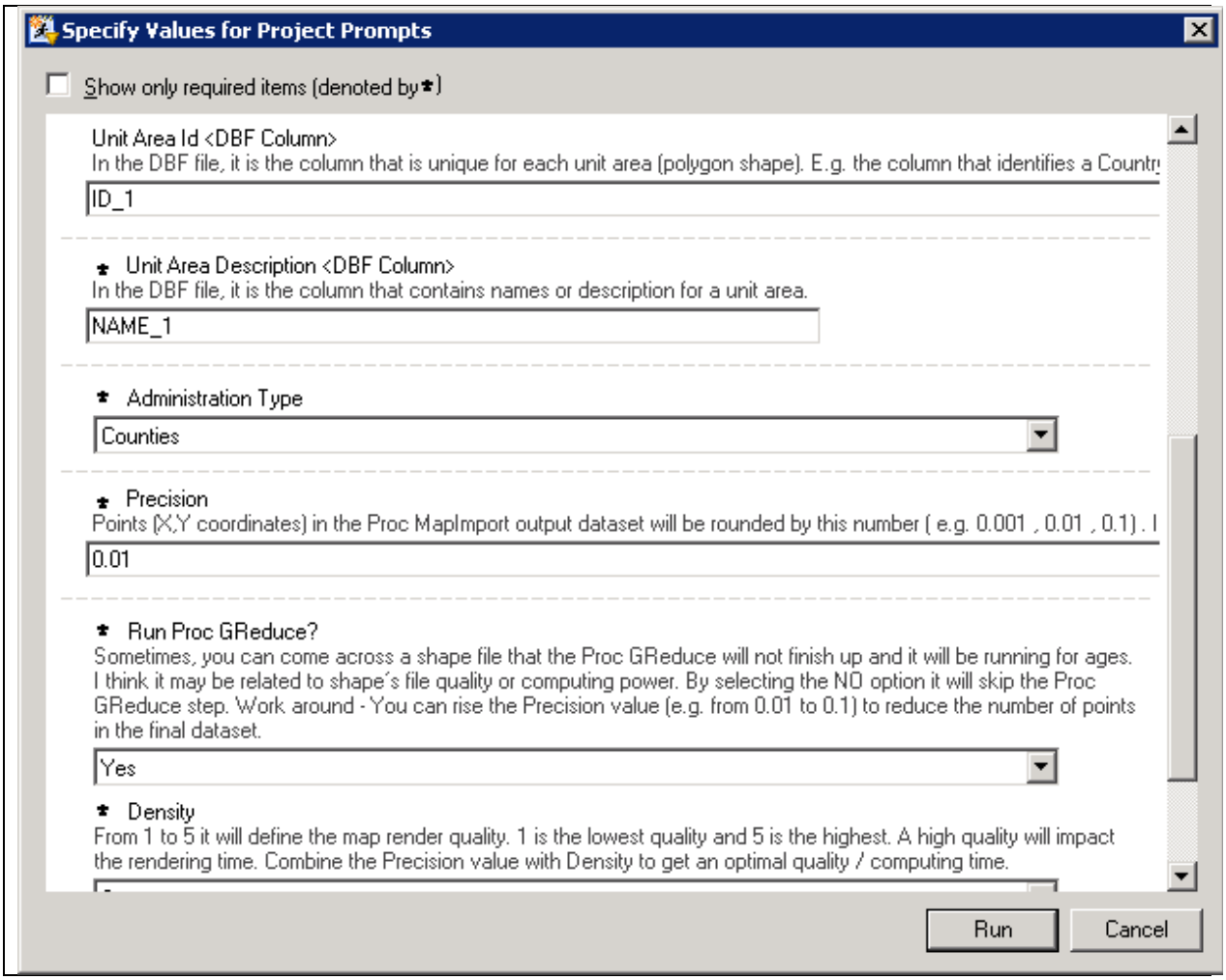
Unit Area Id: ID_1

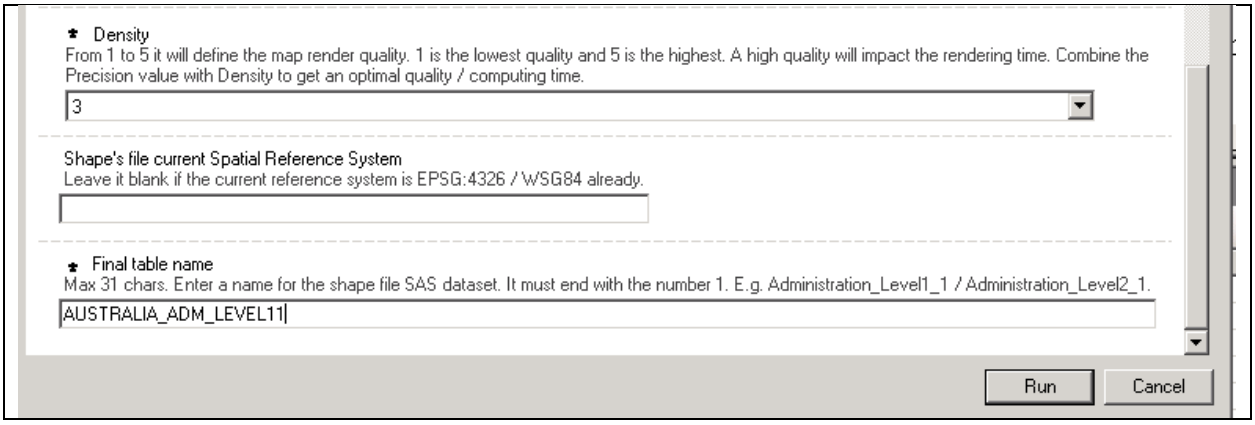
Unit Area Description : NAME_1

Final Table : AUSTRALIA_ADM_LEVEL11

To discover the best column to be used as Unit Area Id or Unit Are Description you may open the DBF file using excel or by opening the shape file using a GIS software like the QGIS aforementioned.







Click on Run, check the SAS Log for errors.

Importing the AUS_adm2.shp file – Skipping Proc GReduce

In my environment, I was not able to run the Proc GReduce for this shape file, every attempt would take more than 2 hours and the job would not finish up. So we are going to run **without running the Proc GReduce**. In addition, we are going to set a low precision to reduce the number of points within each Unit Area we hope that will reduce the rendering time while it keeps a good image quality.

Here are the text values to copy and paste.

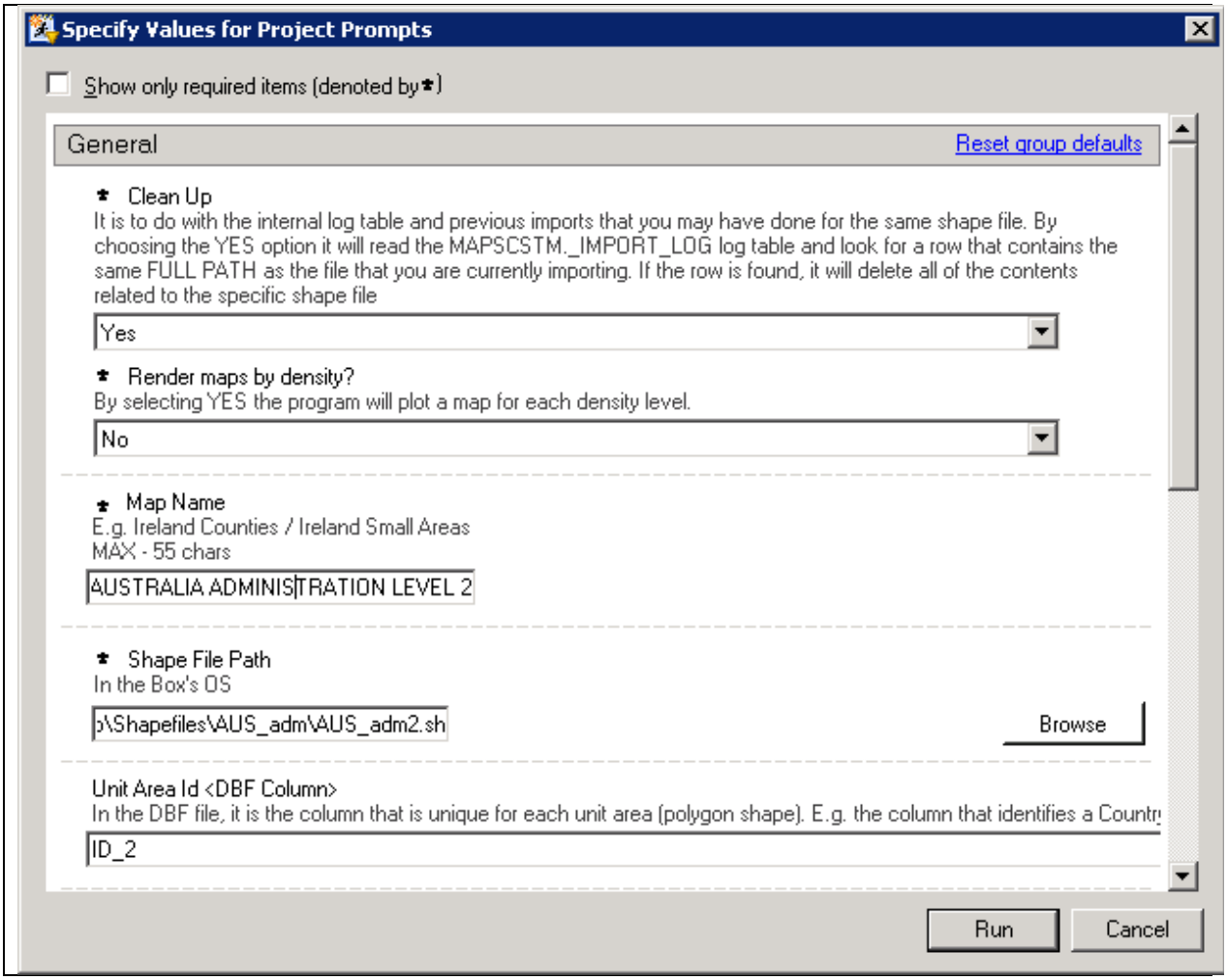
Map name : AUSTRALIA ADMINISTRATION LEVEL 2

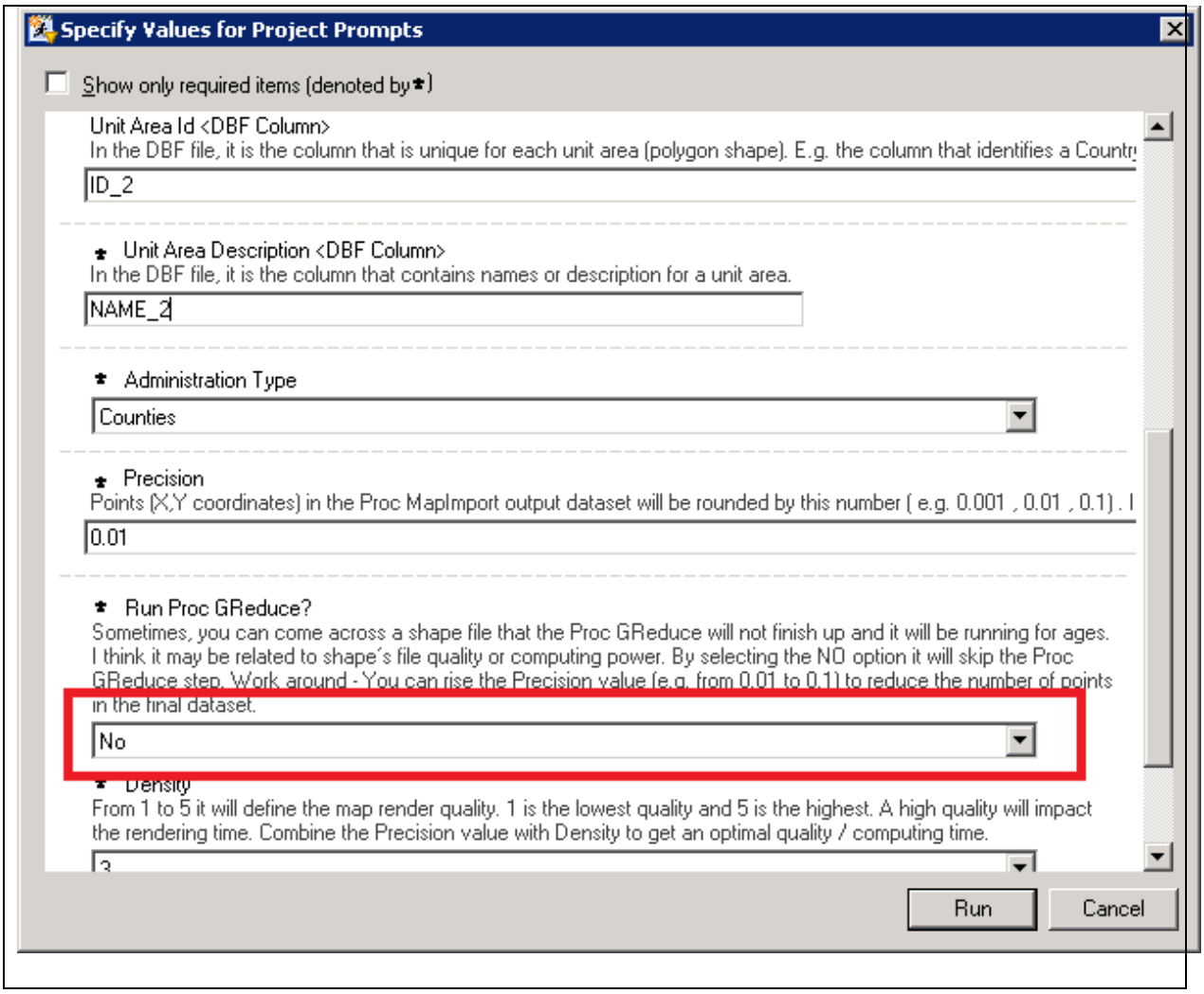
Shape File Path : <PATH TO>\ AUS_adm\AUS_adm2.shp

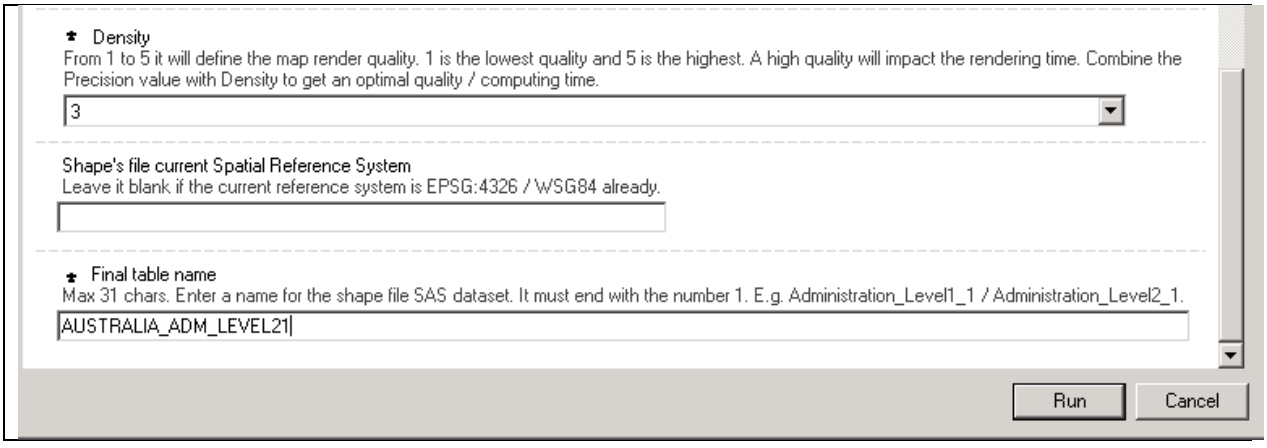
Unit Area Id: ID_2

Unit Area Description : NAME_2

Final Table : AUSTRALIA_ADM_LEVEL21







Click on Run, check the SAS Log for errors.

Restart the VA Server

It must be done in the OS Level. Restart the windows machine or Restart the VA server via bash in the UNIX environment.

Create test data for VA.

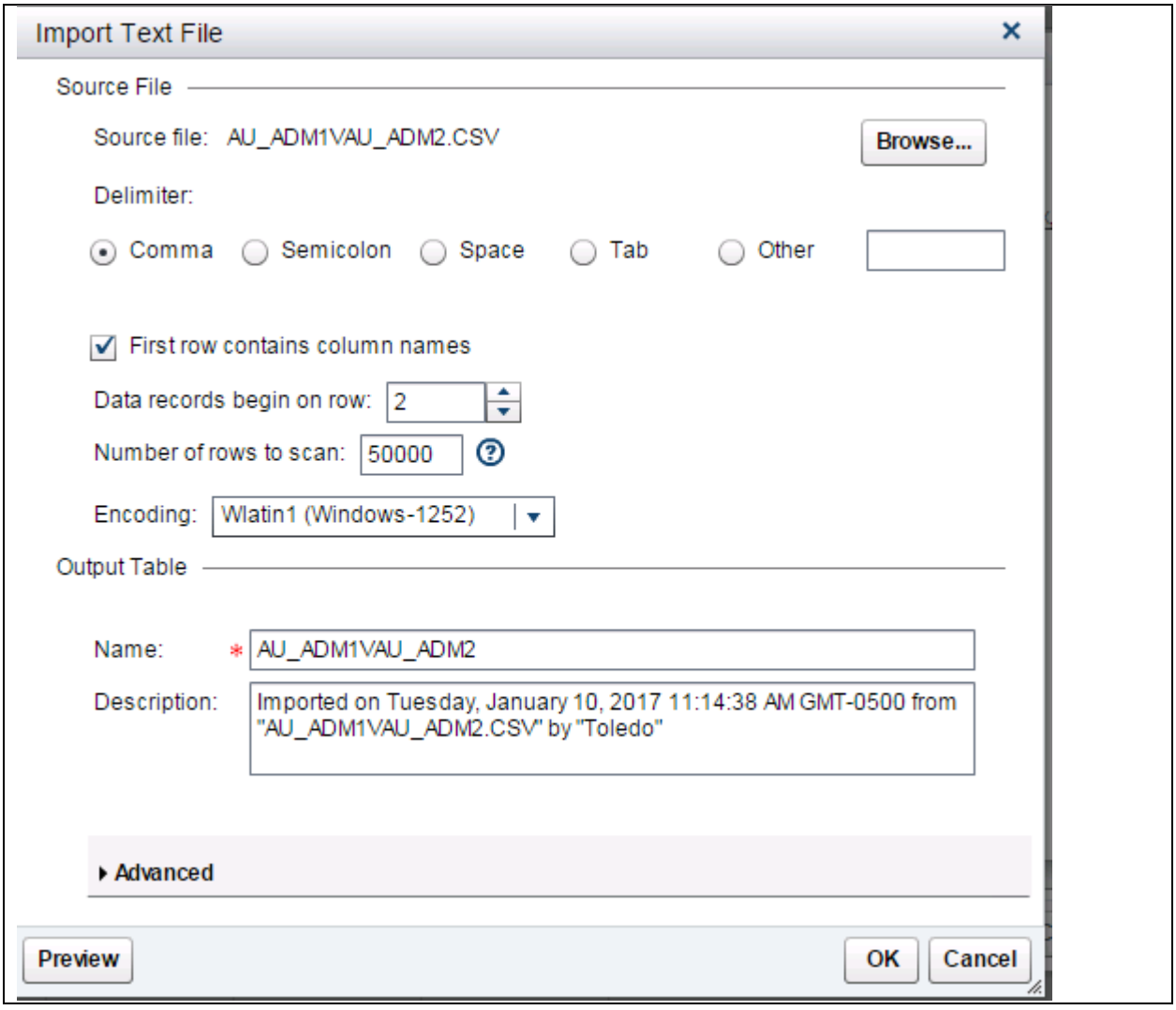
This will create a CSV file that contains the relationship between Level 1 and Level 2 shape files. It will allow us to drill down from the Level 1 areas to Level 2 within Visual Analytics.

```
PROC SQL;
     CREATE TABLE AUSTRALIA_ADM_LEVEL2
     AS
     SELECT DISTINCT ID AS LEVEL2, X, Y FROM
MAPSCSTM.AUSTRALIA_ADM_LEVEL21 GROUP BY 1;
QUIT;

PROC SORT DATA=MAPSCSTM.AUSTRALIA_ADM_LEVEL11
OUT=AUSTRALIA_ADM_LEVEL1(RENAME=(ID=LEVEL1));
     BY ID;
RUN;

PROC GINSIDE DATA=AUSTRALIA_ADM_LEVEL2 MAP=AUSTRALIA_ADM_LEVEL1
OUT=G_LEVEL1VLEVEL2 INSIDEONLY;
     ID LEVEL1;
RUN;

PROC SORT DATA=G_LEVEL1VLEVEL2(WHERE=(LEVEL1 NOT IS MISSING))
OUT=LEVEL1VLEVEL2(KEEP=LEVEL1 LEVEL2) NODUPKEY ;
     BY LEVEL2;
RUN;

PROC SQL;
SELECT COUNT(*) FROM LEVEL1VLEVEL2;
QUIT;
PROC SQL;
     CREATE TABLE CSV_EXPORT
     AS
     SELECT T1.LEVEL1, T1.LEVEL2, T2.IDLABEL AS LEVEL1_LABEL ,
T3.IDLABEL AS LEVEL2_LABEL, INT(RANUNI(0)*123) AS VALUE
     FROM LEVEL1VLEVEL2 AS T1
     INNER JOIN VALIB.ATTRLOOKUP AS T2
     ON T1.LEVEL1 = T2.ID
     INNER JOIN VALIB.ATTRLOOKUP AS T3
     ON T1.LEVEL2 = T3.ID;
QUIT;
PROC EXPORT DATA=CSV_EXPORT FILE="<PATH>\AU_ADM1VAU_ADM2.CSV"
DBMS=CSV REPLACE;
RUN;
```

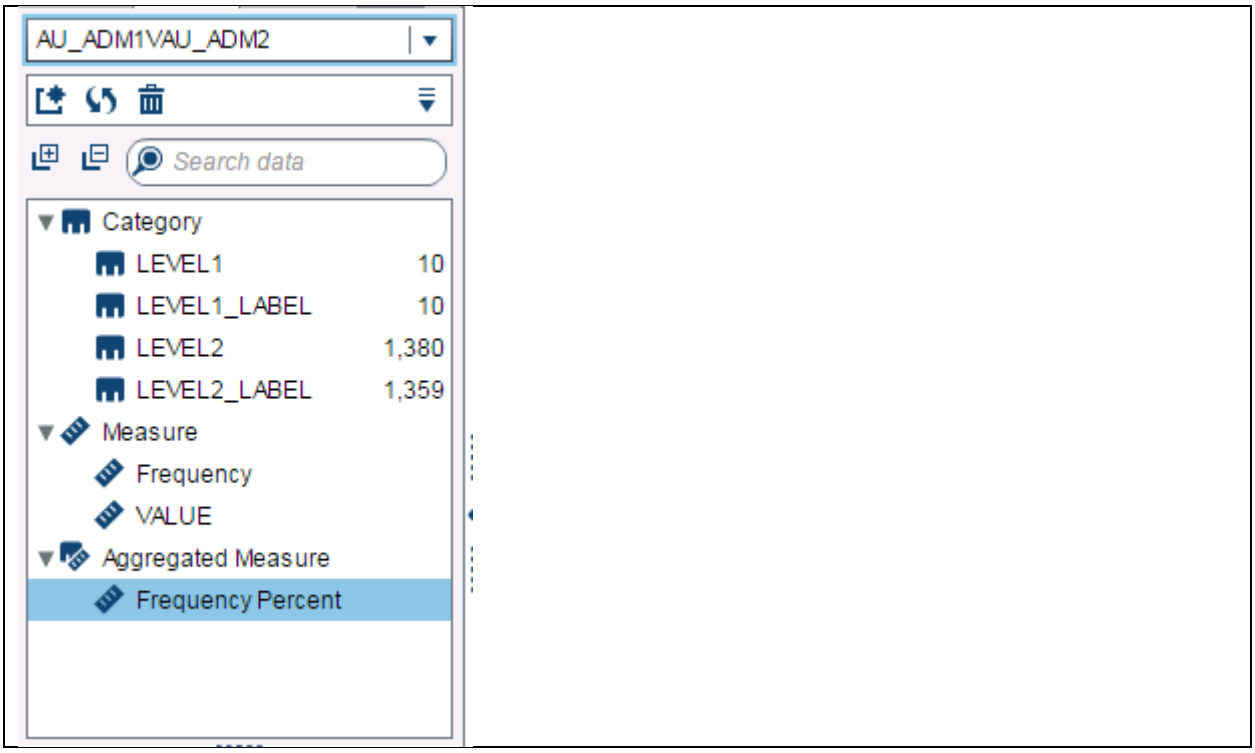
Create report within VA.

Once you have restarted the Visual Analytics box, create a new report within the Report Designer and import the AU_ADM1VAU_ADM2.CSV into a new LASR Table. It should look like below.

Important: Increase the number of row to scan to 50000 – it could be the number of rows in the CSV file too – just make sure that SAS reads the entire file to get the proper length for each column.



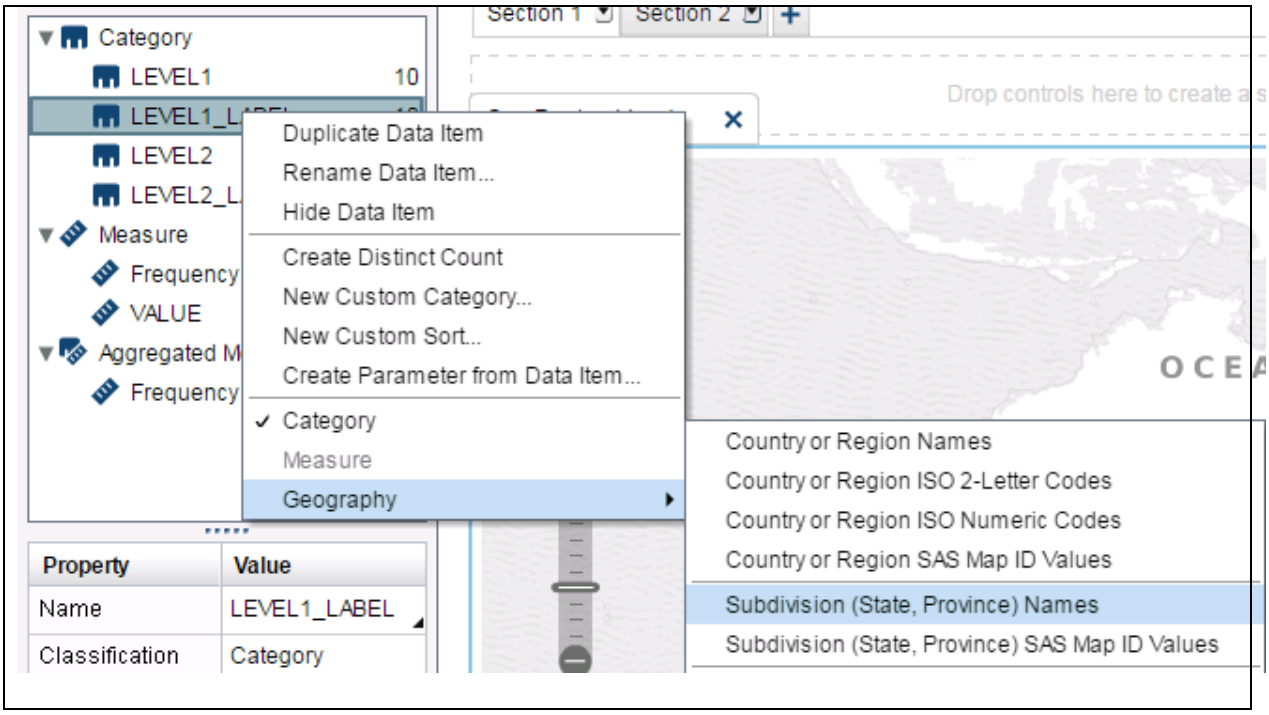
You will get the following columns.



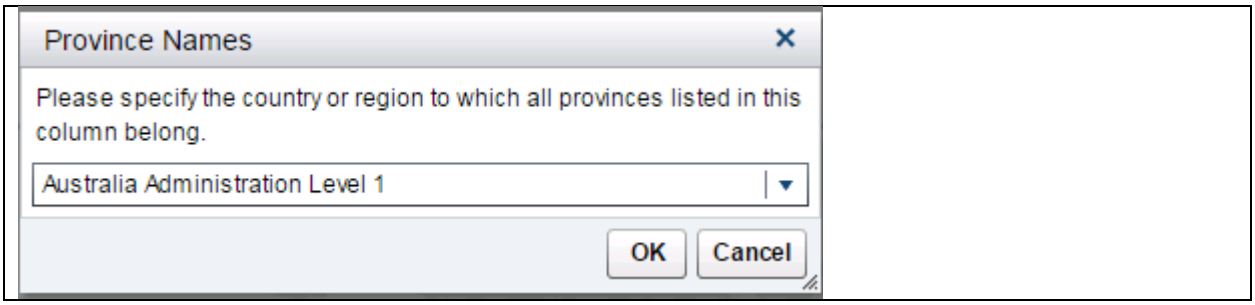
Notice that Level 2 Label was is not a unique identifier as it unique count is 1359 it was supposed to have 1380 as Level 2 (ID).

Add a new region map to your report.

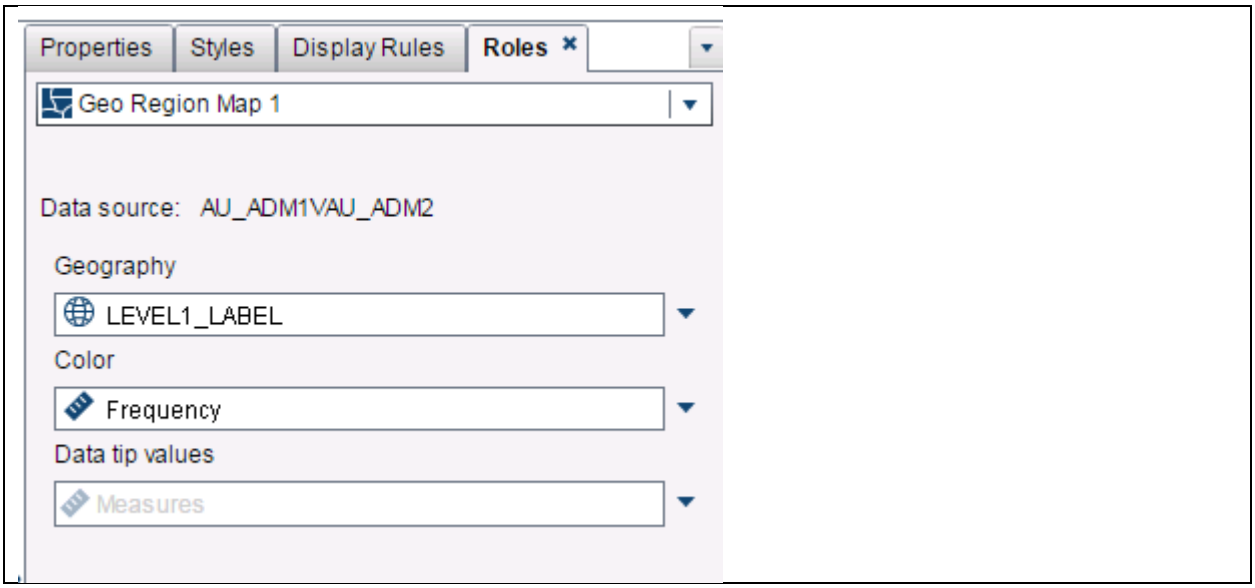
Define the LEVEL1_LABEL data item as Geography Subdivision (State, Province) names.



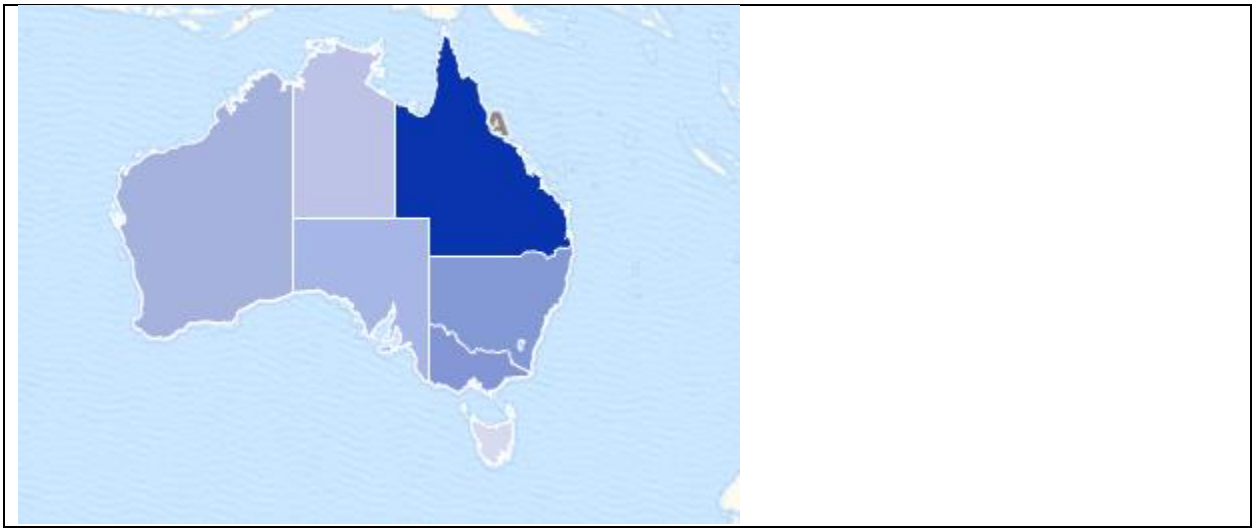
Select Australia Administration Level 1.



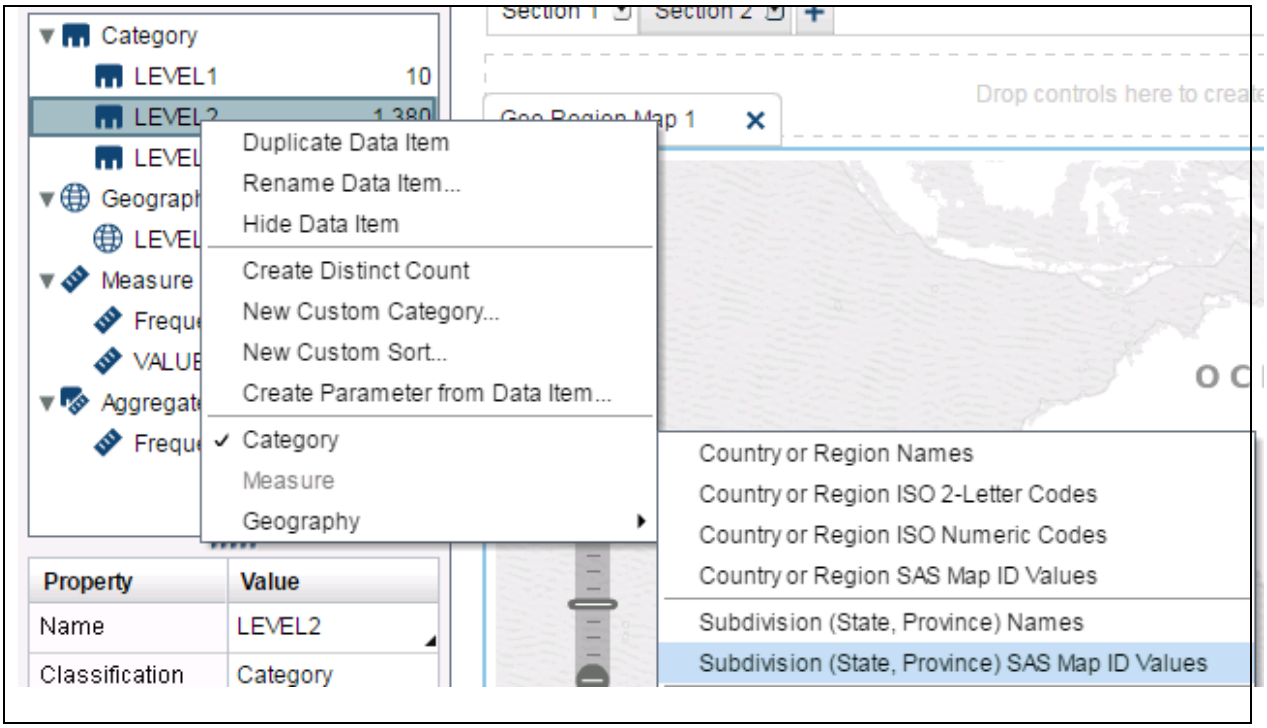
Add this Geography data item to the Geo Region Map.



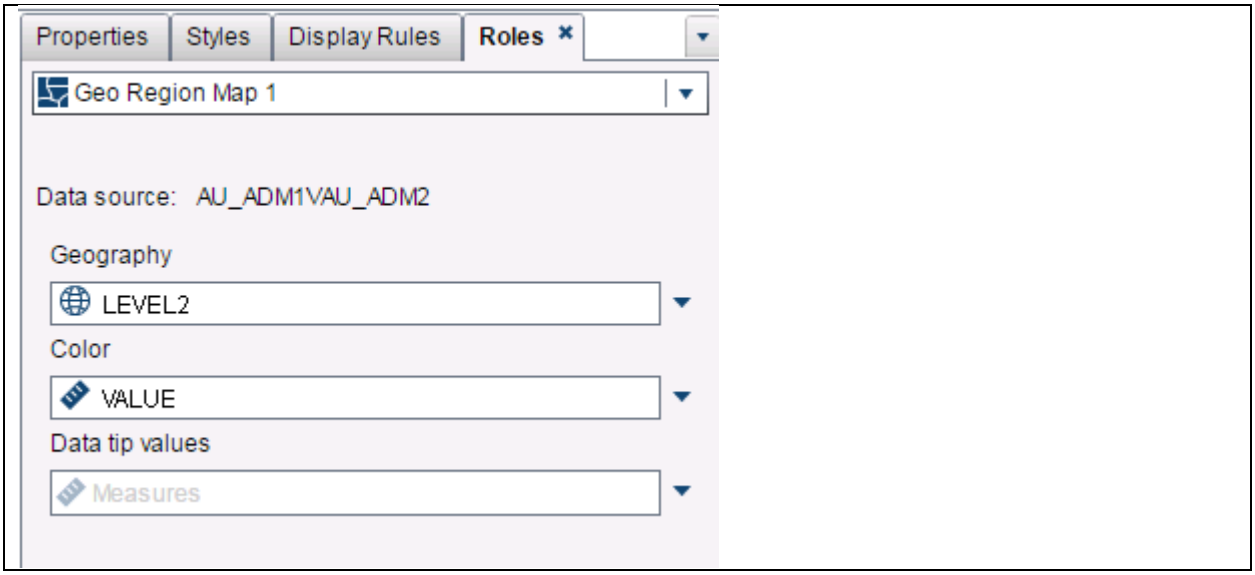
Your map will be updated and render the regions as below.



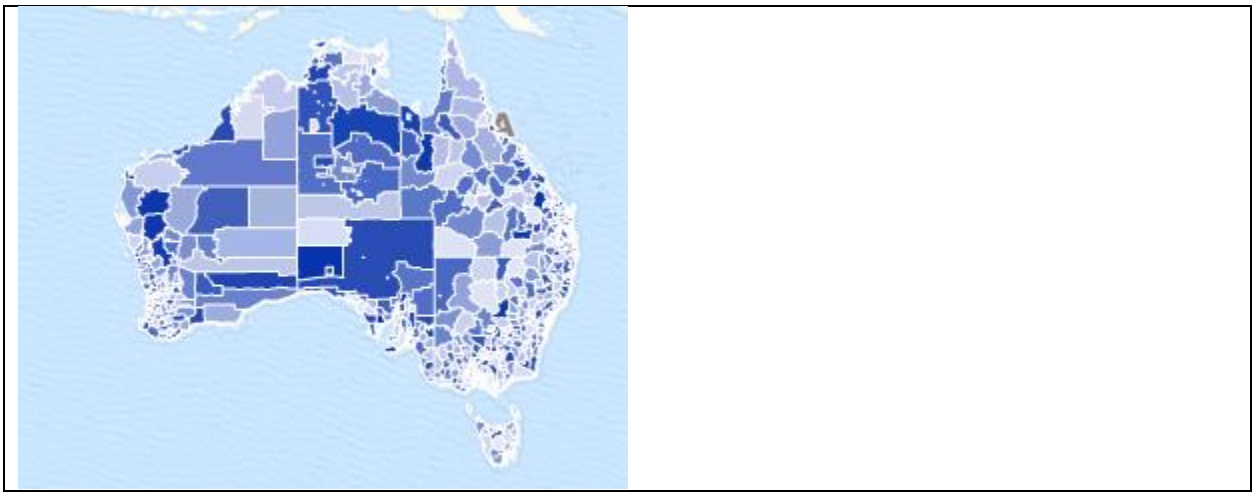
Transform the Level2 category into a geography data item, select Subdivision (State, Province) SAS Map Id Values instead.



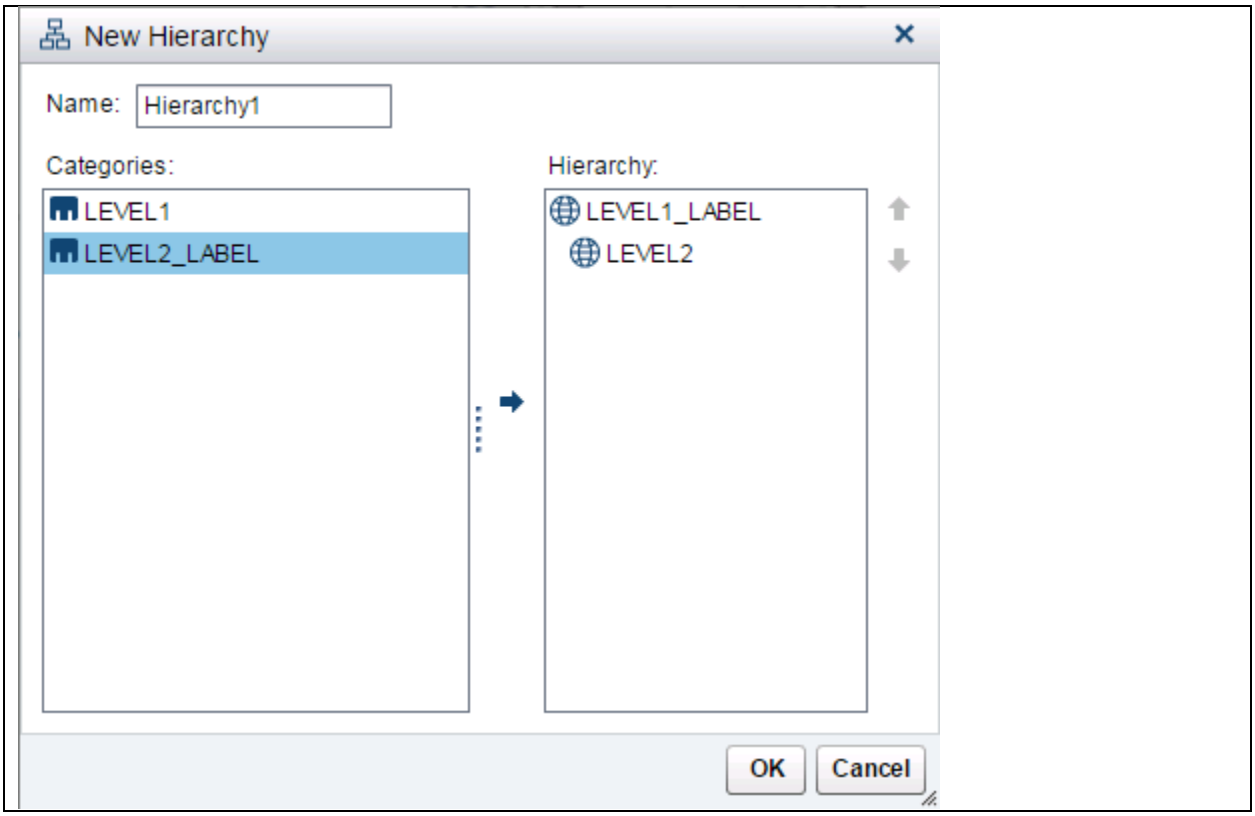
Now, we will replace the variable used to render the map. So drag the LEVEL2 geography data item to the Geo Region Map panel and drop it. Also replace the frequency measure by the value measure in the color role.



You should get the below map.



Next step is to create the hierarchy to drill down from level 1 to level 2. The hierarchy for geography data items started in the VA 7.3.



In the Geo Region Map role tab, replace the current data item by the new hierarchy.

Now, you can drill down into the regions.