Plot Bubbles map

Plot spatial data on an interactive bubble map

Product: IBM® SPSS® Modeler

Extension type: Output/Visualization

Table of Contents

Description....4
Reauirements....4
Installation....4
R Packages used....4
User interface....5
Result example....13
Important links....16
Learn....16
Get products and technologies....16
Discuss....16

Description

This IBM SPSS Modeler output node allows you to plot data containing latitude and longitude on an interactive bubble map. You can use the same color for all points or use a legend column to specify a color code. This legend may be categorical or continuous. Several color palettes are available (sequential, divergent, qualitative or monochrome) covering all possible use of the node.

More precisely, this node generates an HTML file which can be saved to a specific directory and/or opened in the default browser on execution. This html page is an interactive map, that is to say you can move, zoom in and out, etc.

Note that you can obtain longitude and latitude from an address using a Geocoding Node.

Requirements

- SPSS Modeler v16.0 or later
- SPSS Modeler 'R Essentials' plugin
- R v2.15.2

Installation

Close SPSS Modeler. Save the *.cfe* file in the CDB folder, located by default on Windows in "*C:\ProgramData\IBM\SPSS\Modeler\16\CDB*" or under your Modeler 16 installation directory.

Restart SPSS Modeler: the node will now appear in the output panel.

R Packages used

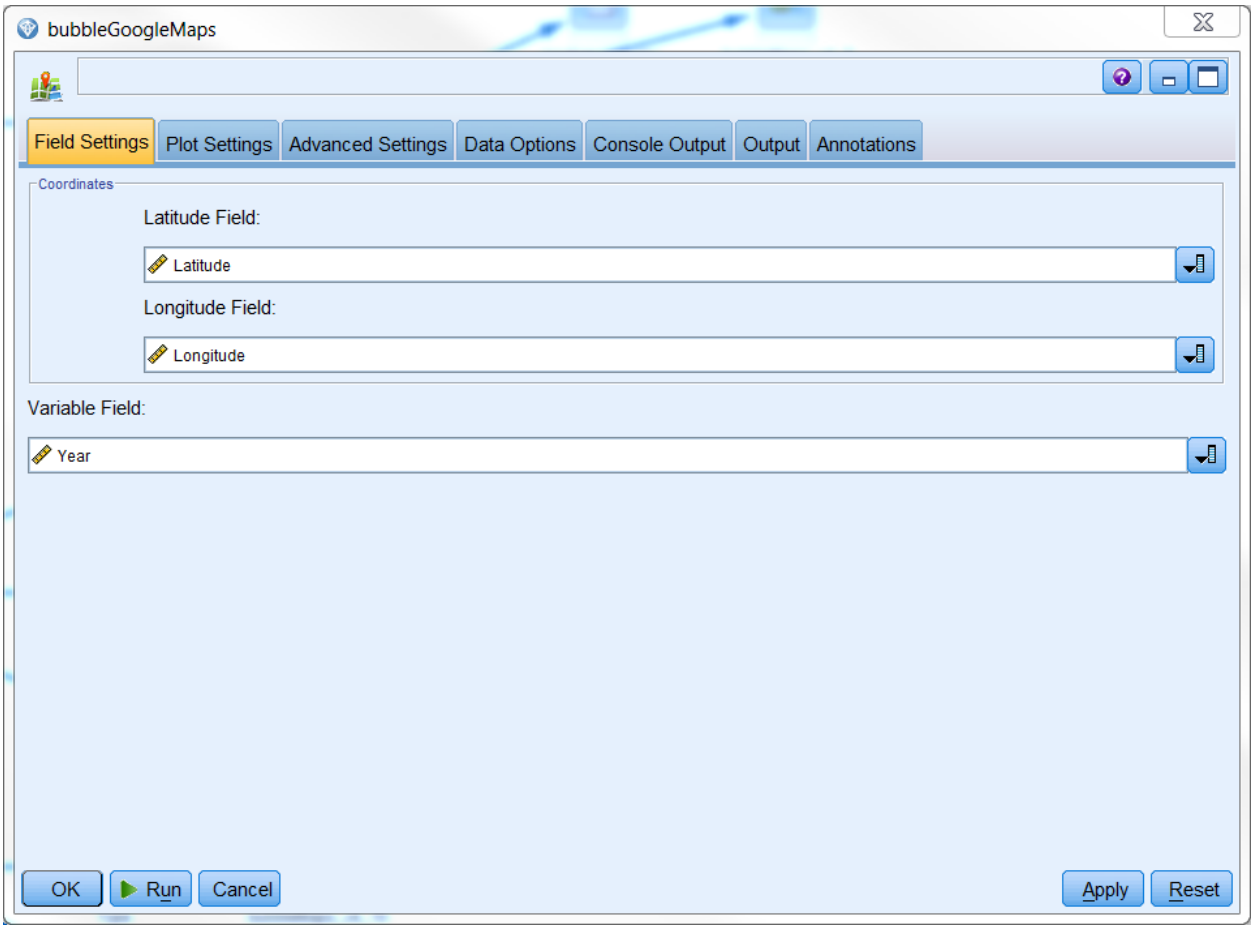
The R packages will be installed the first time the node is used as long as an Internet connection is available.

- plotGoogleMaps by Milan Kilibarda
- RColorBrewer by Erich Neuwirth

User interface

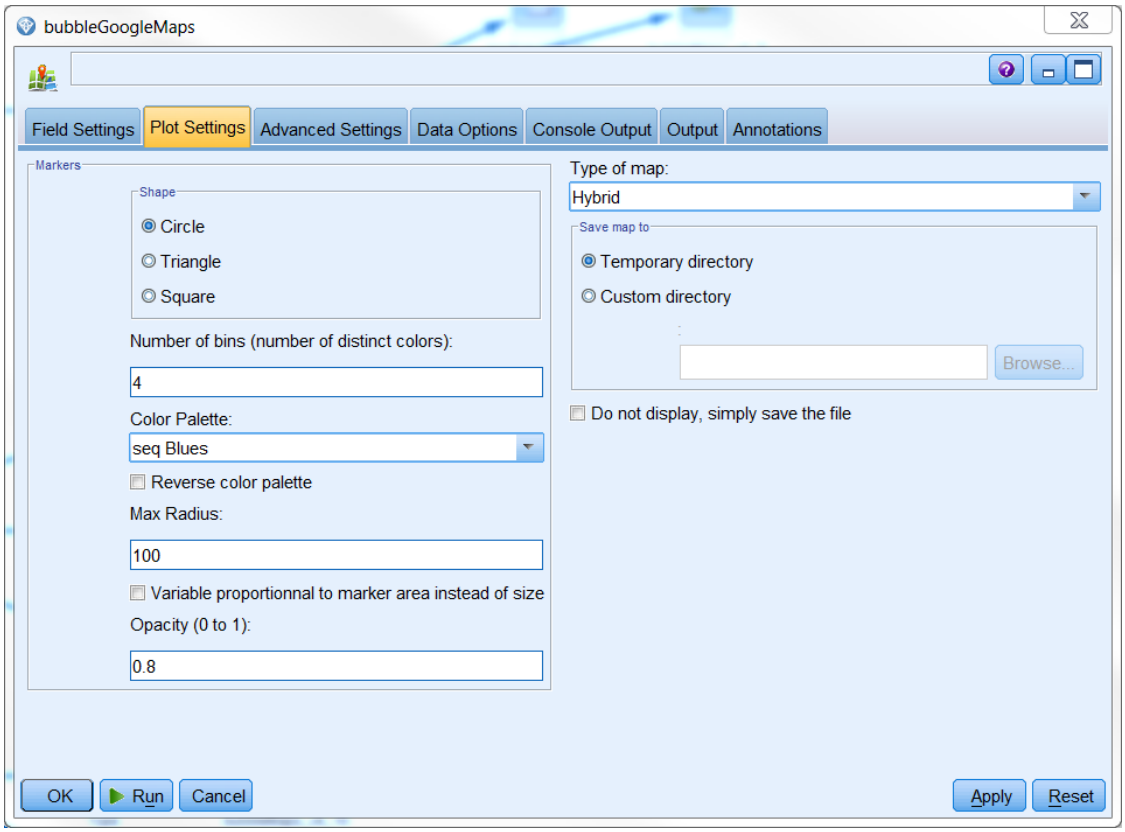
Double click on the node to bring up the user interface.

1. General settings

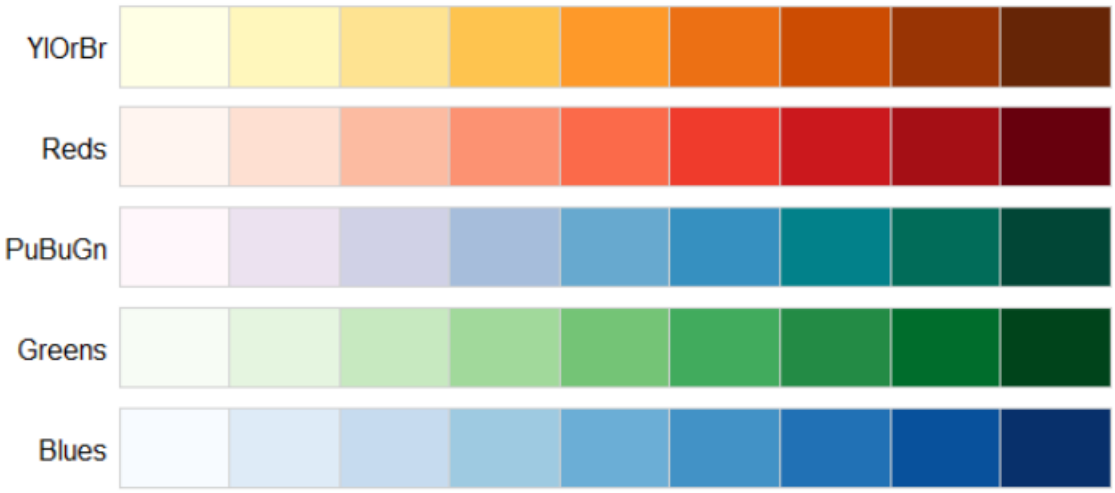


- Latitude and Longitude fields: specify which columns of your data must be used as Latitude and Longitude
- Variable field: the field which will define the color code and the size of the markers. It MUST be numeric.

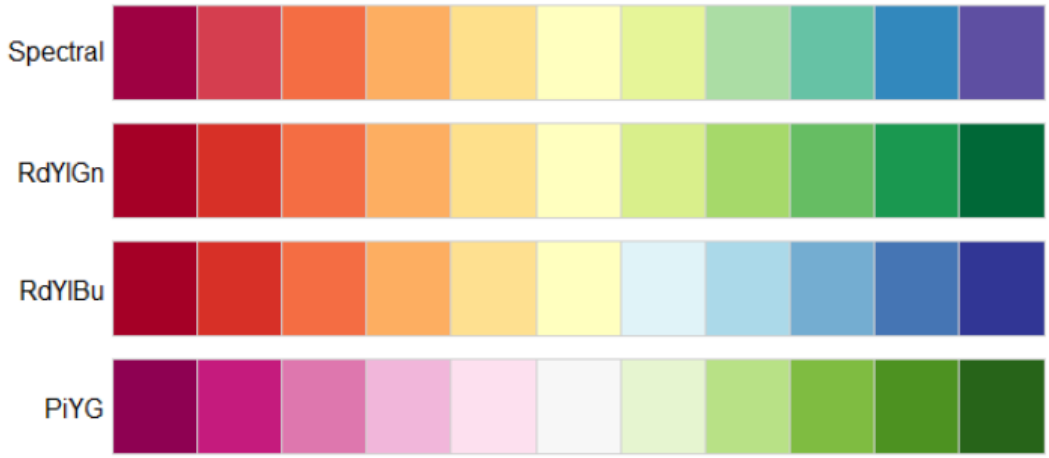
2. Plot settings



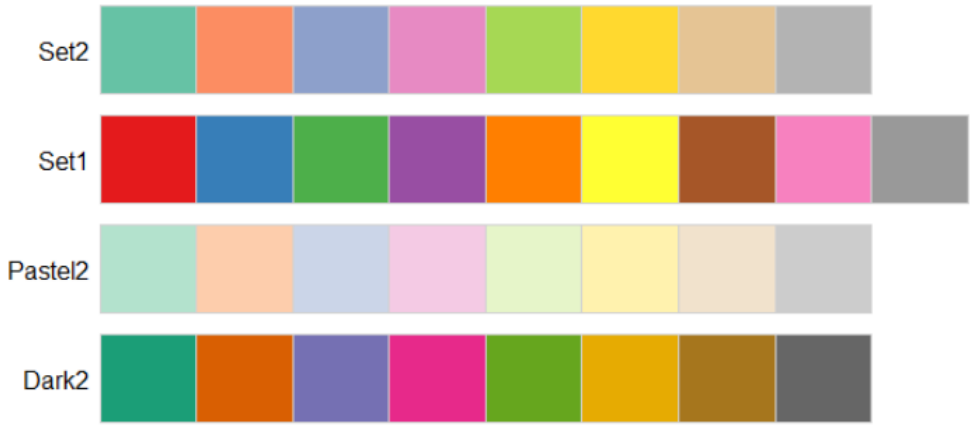
- Type of map:
  - Hybrid (default): satellite view + labels (roads, cities, etc.)
  - Satellite
  - Roadmap
  - Terrain: Roadmap + terrain information
- Shape: choose between circles, triangles or square. The size of the marker will be proportional to the variable field no matter what shape you choose.
- Number of bins: the variable field is binned in N groups, that is to say the values are ranked then divided equally between N groups. Each group will have a distinct size and a distinct color (except if you choose a monochrome palette of course). The number of bins is therefore the number of legend item.
- Color palette: choose a color theme.
  - There are several types of palettes. What follows is from the help of the RColorBrewer package.
    - Sequential palettes are suited to ordered data that progress from low to high. Lightness steps dominate the look of these schemes, with light colors for low data values to dark colors for high data values.



- Diverging palettes put equal emphasis on mid-range critical values and extremes at both ends of the data range. The critical class or break in the middle of the legend is emphasized with light colors and low and high extremes are emphasized with dark colors that have contrasting hues.



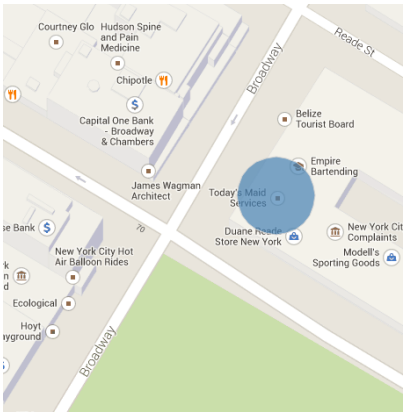
- Qualitative palettes do not imply magnitude differences between legend classes, and hues are used to create the primary visual differences between classes. Qualitative schemes are best suited to representing nominal or categorical data. The 9-colors palettes were hand-picked. If you need a lot of colors (typically more than 9), the color palettes will be programmatically expanded as needed but they will lose in color quality.



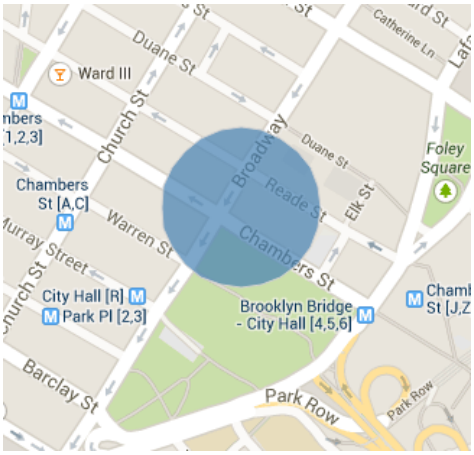
- Monochrome palette: hardly a ‘palette’, they consist of a single color. Available are: blue, red, green, orange and white.

Actual colors: steelblue, indianred, forestgreen, darkorange, floralwhite. These are more esthetic than their saturated counterparts.

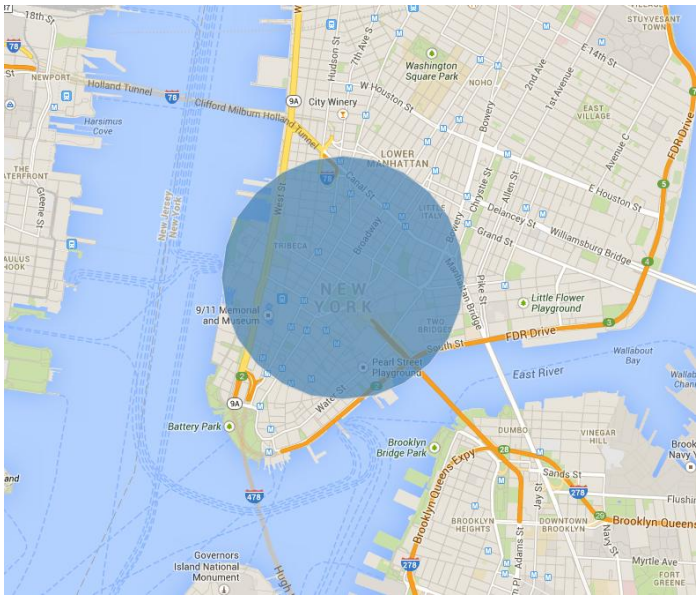
- Reverse color palette: You can reverse the color palette. For example, you may feel that dark colors should represent low values in a sequential color palette, or you may want a divergent color palette to evolve from blue to red (use case: temperature)
- Max radius: The radius of the biggest markers (representing the highest values). This is absolutely key to have a good plot. Since the bubbles do not scale (*ie* keep the same size on the map as you zoom in and out), this must be correctly set.
  - If you do not see your bubbles but there is no error, you bubbles may just be too small to see from the zoom level.
  - Indicative values:
    - 10 is the size of a building



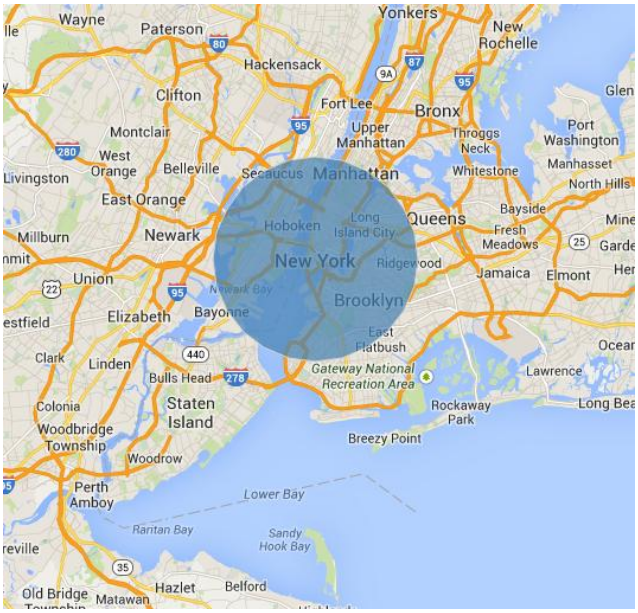
- 100 is the size of a block



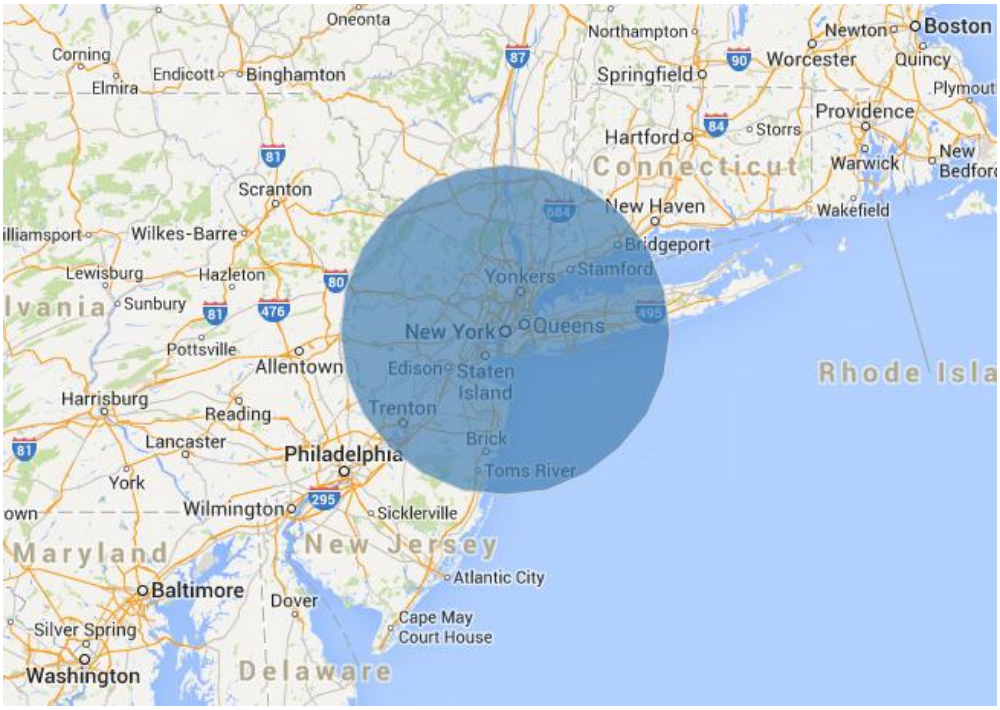
- 1000 is the size of a small town



- 10000 is the size of a big town



- 100000 is the size of a county/small country

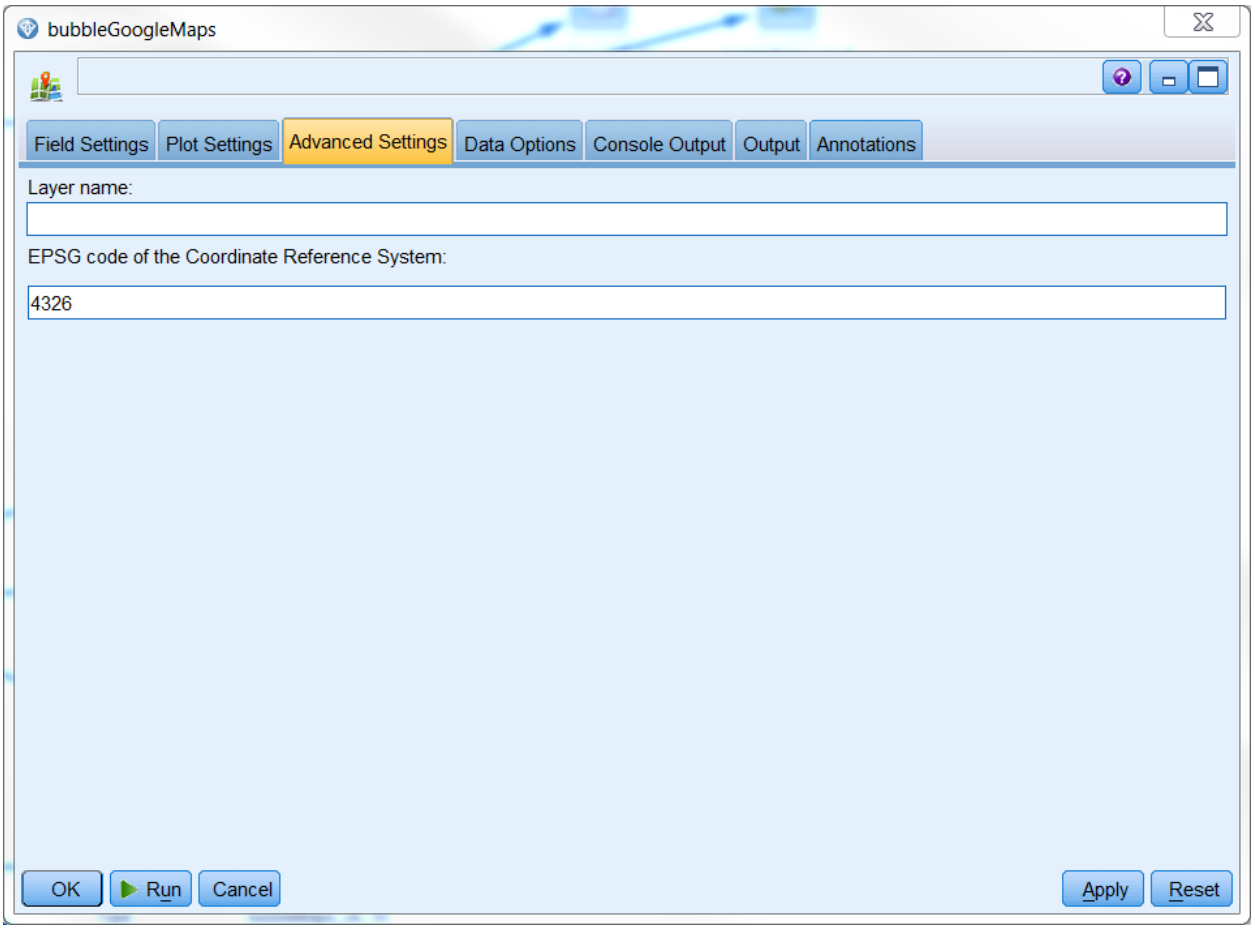


- 1000000 is the size of a country

- 10000000 is probably too big for what you want to do.

- Variable proportional to marker area instead of size
  - If unchecked, the marker diameter will be proportional to the variable field
  - If checked, the markers *area* will be proportional to the variable field, that is to say the marker diameter will be proportional to the square-root of the variable field.
- Opacity: sets the degree of transparence of the marker. 0 is totally transparent and 1 is totally opaque. Default is 0.8.
- Save map: by default, the html file is written to temporary directory. If you want to keep the html file, you can specify a directory here. Note that they are actually 2 files: the .html and .png which contains the legend and which is saved along with the html.
- Do not display: if checked, the map won't be opened in the default browser when the node is run. The map will still be generated and you will still be able to open the html file manually, but this will be difficult if you did not specify a custom directory to save the file.

3. Advanced Settings



- Layer name: you can specify a custom layer name. It will be printed above the legend. This is really useful only for multi-layered map, a feature not yet available.
- EPSG code of the Coordinate Reference System (CRS)
    - Earth is a three-dimensional body, roughly spherical in shape, yet the vast majority of maps are flat (2-dimensional). A Coordinate Reference System (CRS) defines a method of projecting all or part of the Earth onto a 2D surface. (source: openstreetmaps.org). Every time you use spatial data (for this node, this means data with latitude and longitude), a specific CRS was used to generate the coordinates. To plot your data at the correct location, you must know the CRS. There is no way to determine it from the data, the CRS information is pure metadata.
    - All the common CRS are referenced by an EPSG code. If you know what CRS was used, a simple search will give you the EPSG code. If your points are not where they are supposed to be, the EPSG code of the CRS may be wrong. For example, US federal institutions use a specific CRS when sharing data, British institutions use another, most web APIs use EPSG 4326 or EPSG 3857…

- The default is 4326, corresponding to the CRS WGS84 which is very common. This choice is motivated by the fact that the Google Maps service as well the Openstreetmaps service use this CRS. Therefore, this node is compatible by default with the geocoding node and with any data using spatial data generated with the Google Maps API or the Openstreetmaps API.

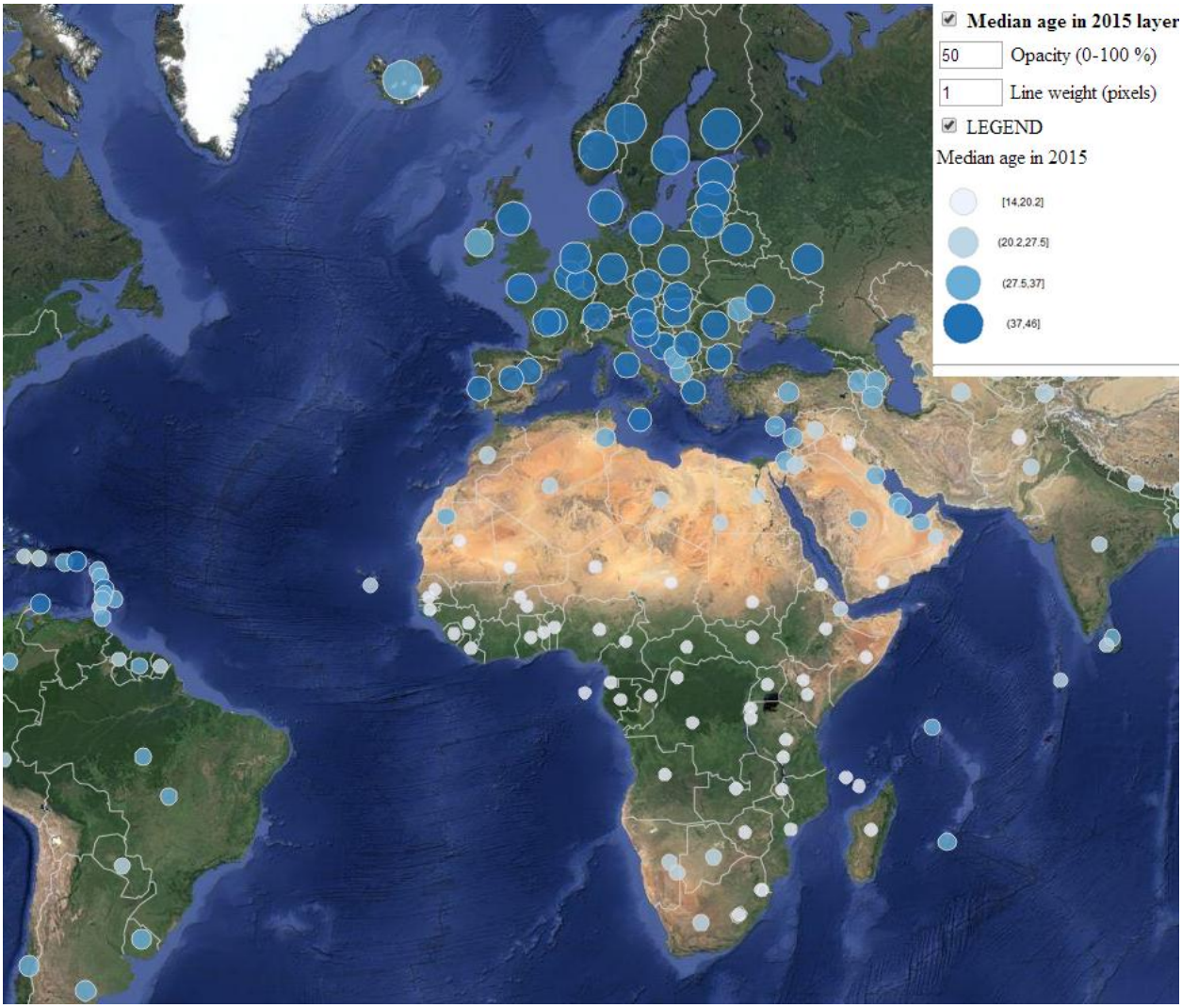
Result example

These are some BubbleMap output that you can easily reproduce. We are using demographic data from UNdata, in that case the SPSS table contains the median age of the population by country or area. Using the geocoding node, the latitude and longitude columns were added. We now have this:

	Country or Area	Latitude	Longitude	Median age in 2015
1	Afghanistan	33.939	67.710	17
2	Africa	-8.783	34.509	19
3	Albania	41.153	20.168	33
4	Algeria	28.034	1.660	27
5	Angola	-11.203	17.874	16
6	Antigua and Barbuda	17.061	-61.796	31
7	Argentina	-38.416	-63.617	31
8	Armenia	40.069	45.038	33
9	Aruba	12.521	-69.968	40
10	Asia	34.048	100.620	30
11	Australia	-25.274	133.775	37
12	Australia/New Zeala...	-26.194	152.670	37
13	Austria	47.516	14.550	43
14	Azerbaijan	40.143	47.577	30
15	Bahamas	25.034	-77.396	32

If you try to use the default settings of bubble map with this data, you won't see any bubble:

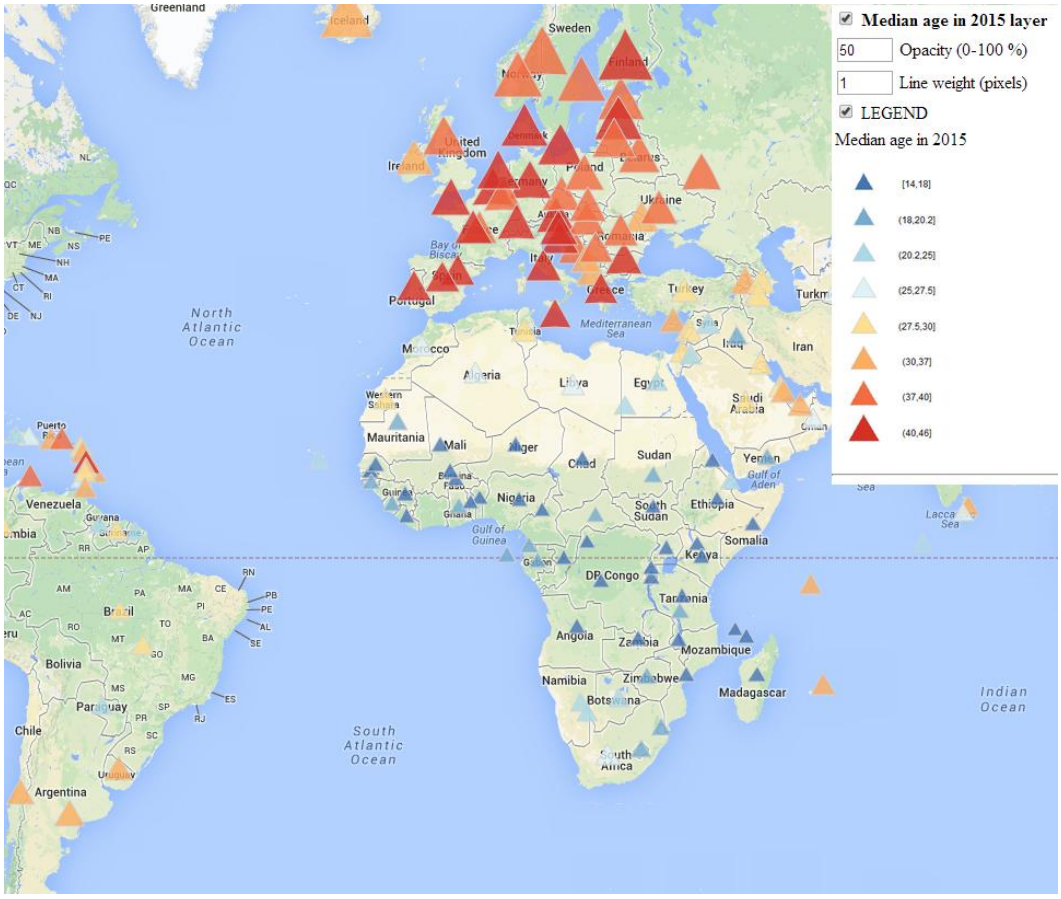
This is because bubbles don't scale as you zoom in or out. They have a fixed size, and the default size is 100, adapted if you're plotting data within a city for example. Choosing a better size, you can get to this:



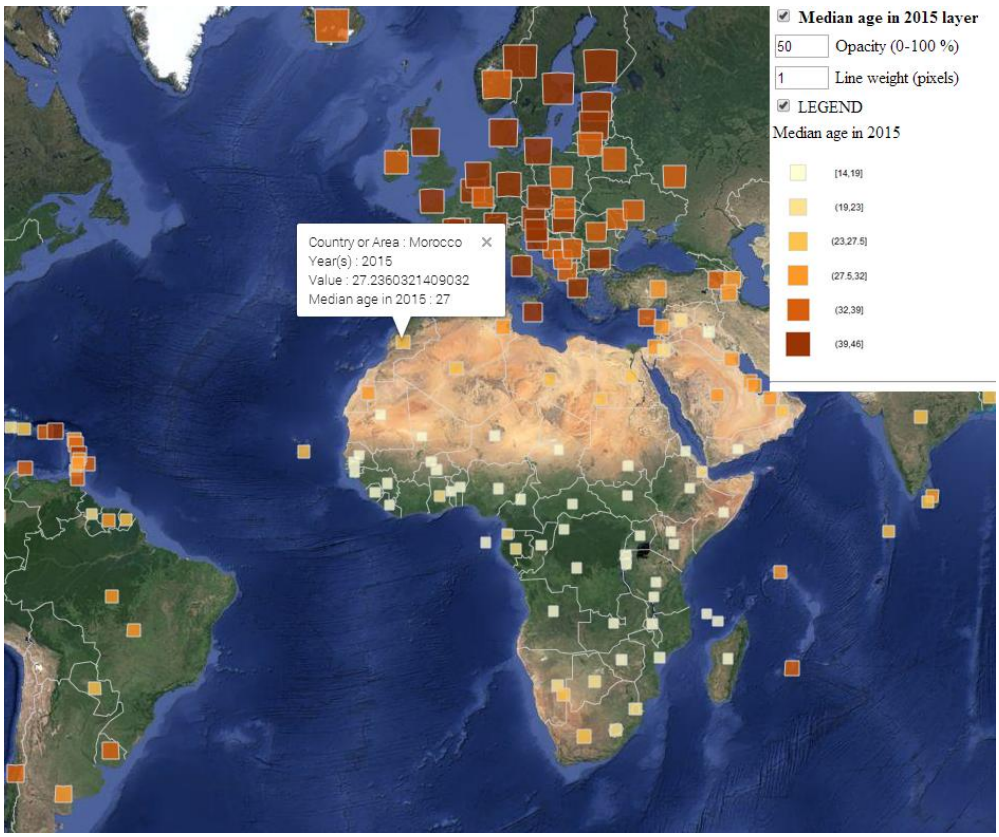
As you can see, the most critical option to set in order to have a good looking plot is the size of the markers. You can do this by going into 'plot settings' and choosing an adapted value for 'max radius'. Here the max radius was set to 150000.

Indicative values for the max radius: with 10 is the size of a house, 1000 is the size of a small town, 100000 is the size of a county/very small country. With 1500000, we have bubbles the size of a country.

Multiple shapes and color palettes are available. You can choose the number of bins (=number of legend items) you want. You can also change the type of map you want to have in the background.



You can move and zoom (but the markers will not scale, which is why the max radius is so important). Click on a bubble to display the data of the corresponding row.



Important links

Learn

- Learn more about SPSS software.
- Visit developerWorks Business analytics for more technical analytics resources for developers.
- The Comprehensive R Archive Network is the main site for the R project and each R package. The help pages and manuals that are associated with `optimx`, `nlmrt`, and `Rcgmin` are detailed. Numerous references are provided.
- Read "Do I need to learn R?" (Catherine Dalzell, developerWorks, September 2013) to learn why R is a valuable tool for data analytics that was expressly designed to reflect the way that statisticians think and work.
- "Calling R from SPSS" describes how to use R code inside IBM SPSS Modeler 16.
- Read "Using Google maps API" to discover how to use Google Maps API with R.
- Read "Create new nodes for IBM SPSS Modeler 16 using R" to learn how to create new extensions easily.

Get products and technologies

- Download the R plug-in for SPSS plugin.
- Download the R 2.15.2 for Windows package.

Discuss

- Visit the IBM SPSS DevCentral developerWorks community to share tips and experiences with other IBM SPSS developers.
- Follow developerWorks on Twitter to be among the first to hear about new resources.