10-2014

A Microsoft VBA Application for Generating Heat Maps

Mark Polczynski
Marquette University, mark.polczynski@marquette.edu

Michael Polczynski
Georgetown University

Transactions in GIS, Vol. 18, No. 5 (October 2014): 783-791. DOI.

A Microsoft VBA Application for Generating Heat Maps

Mark Polczynski
Electrical and Computer Engineering, Marquette University, Milwaukee, WI
Michael Polczynski
Department of History, Georgetown University, Washington, DC

Abstract

A choropleth map is a form of thematic map used to portray the structural characteristics of some particular geographical distribution not apparent in data presented in tabular form. Preparation of a choropleth map starts with the assignment of map features to classes based on the value of a specific feature attribute followed by the association of classes of features with appropriate map colors or symbols. Map features are often geographical regions with naturally or artificially defined boundaries, but choropleth maps can also be prepared by segmenting the area to be mapped into a regular grid of regions. Maps prepared with each grid shaded in an intuitive manner such as blue for grids with the lowest attribute values to red for the highest values can be termed “heat maps”. This technical note describes the HeatMap Microsoft Excel application which converts information contained in a worksheet into a heat map, and then converts the heat map into a file suitable for

display using mapping systems such as Google Earth. An example illustrates how the application can be used to visualize the seventeenth century frontier between the Polish/Lithuanian Commonwealth and the Ottoman Empire.

1. Introduction

A choropleth map is a form of thematic map used to portray the structural characteristics of some particular geographical distribution not apparent in data presented in tabular form (Robinson [5]). Preparation of a choropleth map starts with the assignment of map features to classes based on the value of a specific feature attribute (e.g. population density), followed by the association of classes of features with appropriate map colors or symbols. Map features are often regions with naturally or artificially defined boundaries (e.g. watersheds, countries), but choropleth maps can also be prepared by segmenting the area to be mapped into a regular grid of regions. Maps prepared with each grid shaded in an intuitive manner, such as blue for grids with the lowest feature attribute values to red for grids with the highest values, are termed here “heat maps”, with map “hot spots” being associated with high attribute values.

This technical note describes the HeatMap Microsoft Excel VBA (Visual Basic for Applications) application, which converts information contained in a Microsoft Excel worksheet into a heat map, and then converts the heat map into a KML (Keyhole Markup Language) file suitable for display using mapping systems such as Google Earth. The HeatMap application does not require a user to understand either VBA or KML code, and only requires that the user be able to enter data into an Excel worksheet.

A convenient way to describe the HeatMap application is to examine its use in visualizing the seventeenth century frontier between the Polish/Lithuanian Commonwealth and the Ottoman Empire. Since this frontier was in a state of constant conflict during this period, one approach to visualizing the frontier is to identify networks of defensive sites along the frontier. Here, a defensive site is any area of habitation designed to resist attack. The defensive potential of a site is the capacity of the site to resist attack and to be aided by other sites in resisting attack. The defensive potential of a site depends on several factors, including geographical proximity to other sites, geographical terrain in the vicinity of the site, and site construction elements associated with defensibility such as stone vs. wood construction. A defensive network is a geographical cluster of sites with high defensive potential.

For the example followed here, the Commonwealth/Empire frontier area is segmented into grids and the grid attribute value for each grid is calculated based on the defensive potential of the defensive sites in the region of the grid. Hot spots on the resulting heat map can then be used to identify defensive networks along the frontier. Figure 1 provides a Google Earth map showing the location of 600 defensive sites associated with the frontier, and Figure 2 shows a Google Earth heat map of the frontier. The purpose of this technical note is to describe how the defensive site map of Figure 1 can be converted to the defensive network heat map of Figure 2.

2 Problem

Simple digital maps can be easily produced using applications such as Google Earth's mapmaking tools (earth.google.com/outreach). However, map creation and maintenance using these tools can become a challenge when dozens or hundreds of features need to be mapped (Polczynski and Polczynski [4]), and generation of heat maps such as that shown in Figure 2 is impractical using these tools. Sophisticated systems for creating and maintaining maps such as ArcGIS (www.esri.com/software/arcgis) and GRASS (grass.osgeo.org) provide a more practical alternative, but licenses can be expensive and effective use of such systems requires extensive training and experience (Ballagh et al. [1]). Researchers interested in applying these sophisticated mapping systems may not have the time to invest in acquiring and maintaining map generation expertise, and may not have access to mapping experts or funding to support development of maps on their behalf.

Another approach to map building is to construct digital maps "from scratch" using Keyhole Markup Language (KML), the code language commonly used by mapping systems such as Google Earth (Wernecke [6]). This can be a zero-cost solution for preparing maps, since Google Earth is a free application and KML code can be written using a simple text editor like Microsoft Notepad, but this is a training-intensive approach and is not practical for most researchers employing digital map generation technology. What these researchers need are simple, low-cost, robust, and readily-available tools for automating the process of creating and maintaining digital maps. The purpose of this technical note is to provide such a tool for generating heat maps.

3 Solution

HeatMap is a Microsoft Excel workbook that runs a VBA application which automatically generates a heat map from data in an Excel worksheet. The application converts the worksheet data into a KML file for display by a mapping system such as Google Earth. The application has three major elements:

1. x2003; Generate a grid of geographical regions.
2. x2003; Calculate a grid attribute value for each grid region.
3. x2003; Build a KML file containing grid regions shaded according to their grid attribute values.

A brief description of each element follows:

Generate Geographical Grid Regions

The heat map grid regions used by HeatMap are formed by rectangles arranged in rows and columns as shown in Figure 4. A user sets the grid side length in kilometers. Using this value, HeatMap calculates the number of grid rows and columns required to cover all of the geographical points used to calculate grid attribute values. The application then calculates the four corner coordinates in degrees for all grids such that the grid height and width of each grid is equal to the grid side length.

Calculate Grid Attribute Values

The grid attribute value used to determine the heat map shading of a particular grid (e.g. the defensive potential of the grid) is the sum of the point attribute values of the points in the vicinity of the grid (e.g. the defensive potential of the mapped sites) as calculated per Equation:

$$g_i = \sum_{j=1}^{N} p_j e^{-\left(d_{ij}^2/r^2\right)}$$

(1)

where g_i is the grid attribute value of grid i, p_j is the point attribute value of point j, d_{ij} is the distance between the center of grid i and point j, r is the grid range (described below), and N is the number of points on the map. (See Appendix for a description of how HeatMap calculates d_{ij}).

In general, geographical models are composed of two elements: site characteristics and situation influence. In Equation (1), the point attribute value p accounts for site characteristics (e.g. stone vs. wood construction and terrain in the vicinity of the site for the example being followed here). The exponential weighting factor of Equation (1) accounts for situation influence by applying a weight to the point attribute value p of each point before adding the point attribute value to the grid attribute value g, with the weight depending on the distance d between the point and the center of the grid and on the grid range r.

The effect of the exponential weighting factor can be demonstrated by considering point x in Figure 4. With the grid range r set to three times grid side length, the distance from the point x to the center of the shaded grid in the figure is equal to one half the grid range, yielding:

$$e^{-\left(d_{ij}^2/r^2\right)} = e^{-\left(\frac{1}{2}\right)^2} = 0.779$$

(2)

Thus for grid range = 3 * grid side length, the point attribute value of point x is reduced by 0.779 before being added to grid attribute value. Figure 5 illustrates how the exponential weighting factor varies with $(d/r)^2$.

Build the KML Heat Map File

Figure 6 shows the Point Data worksheet of the HeatMap workbook where a user enters the name, location, and attribute value (here, site defensive potential) of the points being used to generate the heat map. Figure 7 shows the workbook's Control worksheet. Before running the application, the user enters the name of the map as it will appear when displayed and the file path where the generated .kml map file will be saved. The user also enters the grid side length in kilometers and the grid range, where the grid range value entered is the ratio of grid range r to grid side length (grid range = 1 → grid range = grid side length). The user can select either a color or grayscale heat map, and can set the transparency of the grids when they are displayed on the map. Clicking the Create Heat Map button generates the heat map KML file. After the map is generated, the application updates the Grid Data sheet (Figure 9) to show the grid attribute value for each grid. Figure 9 shows the generated KML file displayed by Google Earth. Separate map layers are generated for points (e.g. defensive sites) and the heat map grids (polygons), allowing these map layers to be displayed separately or together. Rolling the cursor over a point on the map reveals a balloon with the point's name and the point attribute value.

4 Implementation

HeatMap users can find it difficult to select values of grid side length and grid range when setting up to run the application. As shown in Figure 7, the approximate height and width of the map being created is displayed on the Control worksheet. These values are derived from the latitudes and longitudes provided on the Point Data sheet, so can help users in determining an appropriate grid side length when creating the map. The number of rows and columns actually used for the map are also displayed on the Control worksheet.

Selection of appropriate values for grid range can also be challenging. In this regard, it is critical to note that HeatMap is intended to be a visualization tool vs. a quantitative or statistical analysis tool, so the “best” value for grid range is in the eye of the beholder. Grid range can be viewed as a “contrast knob” that the user adjusts along with grid side length to improve the visualization produced by the heat map. The most appropriate way to select values for these parameters can be to experiment until the heat map provides a visualization that reveals the features that the user is interested in identifying (here, defensive site hot-spots along the seventeenth century Polish-Lithuanian Commonwealth/Ottoman Empire frontier).

Users should also be aware of some additional aspects of HeatMap implementation. When calculating the number of rows and columns required to cover all of the map sites, a buffer row is added at the top and bottom of the map and a buffer column is added at the right and left of the map to insure that round-off errors in calculations do not exclude any sites from the map. When selecting color or grayscale values to shade grids, grid attribute values are linearly scaled to values between 0 and 1, although other scaling techniques (e.g. logarithmic) are possible. HeatMap uses five easily-distinguished colors for color maps: black, blue, green, yellow, and red. For grayscale maps, HeatMap converts grid attribute values into a continuous gradation of hexadecimal values from grid attribute value = 0 → hexadecimal value = 000000 (black) to grid attribute value = 1 → hexadecimal value = ffffff (white).

5 VBA Application Pseudocode

High-level pseudo-code for the HeatMap VBA application is as follows. Details of the procedure for converting data into KML code (pseudocode section 3) are provided in Polczynski and Polczynski ([4]).

1. Generate grids
    a. Calculate the number of rows, row corner coordinates, and row center coordinates
    b. Calculate the number of columns, column corner coordinates, and column center coordinates
    c. Find corner coordinates and center coordinates of each grid
2. Calculate grid attribute values
    a. Find the row and column that each point lies in
    b. Find the grid that each point lies in
    c. Calculate the grid attribute value per Equation (1)
    d. Color grids per grid attribute value
3. Build heat map
    a. Define KML code segments
    b. Format map information and start KML file
    c. Build point layer
    d. Build polygon layer
    e. Finish KML file

7 Discussion

The usefulness of the visualization provided by a heat map depends on the values of the grid side length and grid range. A long side length produces a low resolution heat map, but a short side length can yield many grids containing no sites interspersed among grids containing a single point, i.e. a point map vs. a heat map. Allowing points in surrounding grids to contribute to g by using high values of grid range r tends to smooth out the resulting abrupt grid-to-grid changes in heat map shading. The problem with this smoothing effect is that too much smoothing forces all values of g toward a single average value. Ultimately, the best way to enhance the usefulness of visualizations provided by the HeatMap application is simply to experiment with grid side length and grid range. For reference, the map of Figure 2 was generated using a grid range of 5, while the map of Figure 3 used a grid range of 2.

The method used here to calculate g is loosely related to grid-based clustering algorithms commonly used for data mining and machine learning applications (Yue et al. [7]). The technique of calculating g for a particular grid using points in surrounding grids is similar to that used by Ma and Chow ([2]). The technique of using both distance and point attribute values to calculate g also bears resemblance to eigenvector centrality as applied in social network analysis (Prell [3]).

8 Conclusions

The seventeenth century frontier between the Polish/Lithuanian Commonwealth and the Ottoman Empire was one of continuous conflict and turmoil, so for this example using a heat map to visualize this frontier in terms of defensive networks makes sense, since geographical regions with strong defensive networks provide an indication of relatively stable frontier boundaries. But heat maps can be used to reveal other geographically-related features. For example, if level of trade activity is substituted for site defensive potential, then a heat map can be used to reveal strong trading networks. This is can be done using HeatMap by simply replacing site defensive potential values in column D of Figure 6 with site trade level values.

As demonstrated in the example provided here, heat maps can enhance visualization of geographical information. The HeatMap application described in this technical note provides map generation non-experts with a fast and simple way to produce heat maps for display using free and readily-available mapping systems such as Google Earth – users need only be capable of entering data into an Excel spreadsheet. Interested parties are encouraged to experiment with HeatMap by downloading the application at http://www.technologyforge.net/MichaelPolczynski/Publications.

Appendix: Calculating map distances and grid locations

The grid attribute value calculation of Equations (1) and (2) and require measurement of the distance d_{ij} between the center of grid i and point j. In HeatMap, the distance between two points is calculated using the spherical law of cosines:

$$d_{ij} = r^{*}\text{a}\cos\left(\sin(lat_i)^{*}\sin\left(lat_j\right) + \cos(lat_i)^{*}\cos\left(lat_j\right)^{*}\cos\left(lon_i - lon_j\right)\right)$$

(3)

where lat_i and lat_j are the latitudes of points i and j, lon_i and lon_j are the longitudes of the points, and r is the mean radius of the earth = 6,371 km.

The procedure for forming heat map grids requires degrees-to-distance calculations per Equation (3), and also distance-to-degrees latitude and distance-to-degrees longitude calculations, per Equations (4) and (5):

$$lat = 1{,}000^{*}l/111133 - 559.8^{*}\cos(2^{*}lat_c) + 1.2^{*}\cos(4^{*}lat_c)$$

(4)

$$lon = l/(6{,}378.137^{*}(\cos(lat)/\text{sqrt}(1 - 0.00669^{*}\sin(lat^2))$$

(5)

where lat is degrees of latitude, lon is degrees of longitude at latitude lat, lat_c is the latitude of the center of the heat map, and l is length in meters that is to be converted to degrees latitude or longitude, which is the grid side length when forming HeatMap grids.

References

Ballagh L M, Raup B H, Duerr R E, Khalsa S J S, Helm C, Fowler D, and Gupte A 2011 Representing scientific data sets in KML: Methods and challenges. Computers and Geosciences 37 : 57 – 64

Ma E W M and Chow T 2004 A new shifting grid clustering algorithm. Pattern Recognition 37 : 503 – 514

Prell C 2012 Social Network Analysis: History, Theory, and Methodology. London, Sage

Polczynski M and Polczynski M 2013 A Microsoft Excel application for automatically building historical geography GIS maps. Transactions in GIS 17 : 148 – 157

Robinson A 1975 Map design. In Proceedings of the International Symposium on Computer-Assisted Cartography, Reston, Virginia : 9 – 14

Wernecke J 2008 The KML Handbook: Geographic Visualization for the Web. San Francisco, CA, Addison-Wesley

Yue S, Wei M, Wang J, and Wang H 2008 A general grid-clustering approach. Pattern Recognition Letters 29 : 1372 – 1384

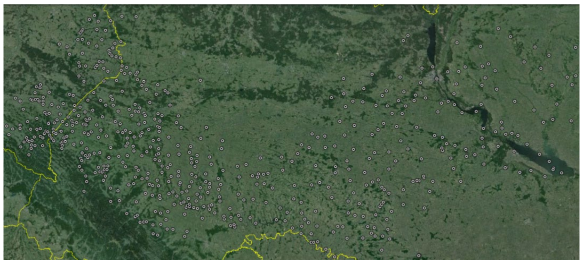

Figure 1. Sites on the seventeenth century Polish/Lithuanian Commonwealth – Ottoman Empire frontier

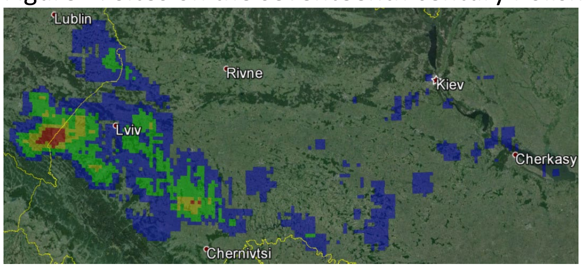

Figure 2. Defensive network heat map of the frontier (grid range = 5)

Figure 3. Defensive network heat map of the frontier (grid range = 2)

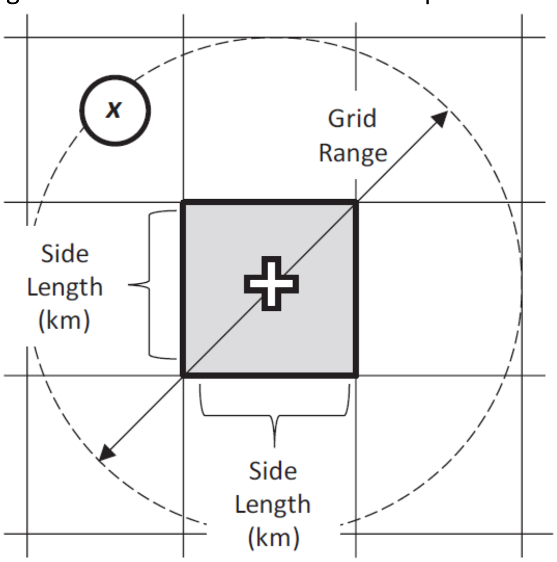

Figure 4. Geographical grid regions

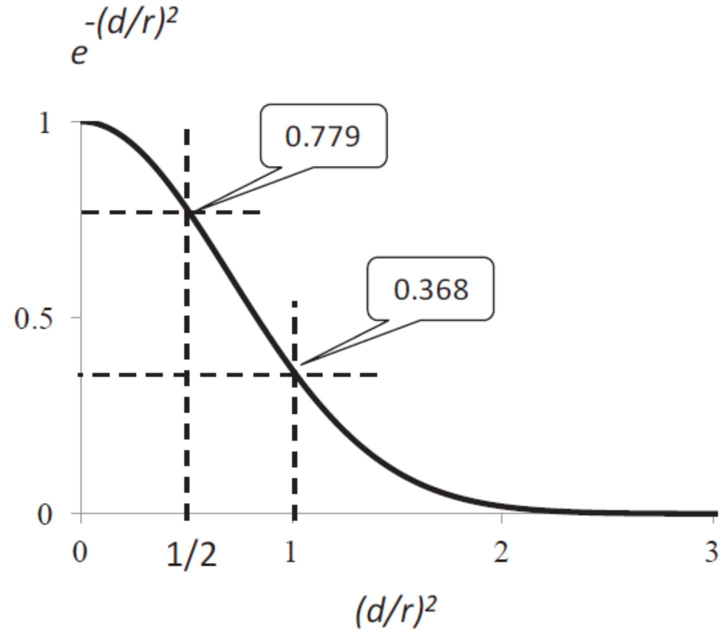

Figure 5. Effect of grid range r on point attribute value weighting factor

	A	B	C	D
1	Point name	Point latitude	Point longitude	Point att. val.
2	Achtyrka	50.30630	34.89743	1
3	Aleksandro'wski	48.52792	29.09889	1
4	Antono'w	49.63061	29.79006	1
5	Bo'brka	49.63347	24.29316	1

Control | **Point Data** | Grid Data | Instructions

Figure 6. HeatMap Point Data worksheet

	A	B
1	**HeatMapV01**	
2	Map name	Frontier Defensive Network
3	File path	
4	Map author	Michael Polczynski © 2013
5	Grid side length (km)	5
6	Grid range	2
7	Color or Grayscale	10
8	Grid transparency (0-100)	20
9	Create Heat Map	
10	Rows	76
11	Columns	198
12	Map width (km)	1218
13	Map height (km)	671

Control | Point Data | Grid Data | Instructions

Figure 7. HeatMap Control worksheet

	A	B	C	D
1	Grid Number	Grid Latitude	Grid Longitude	Grid Att.Val.
2	1	51.27228	21.27065	0
3	2	51.27228	21.34233	0
4	3	51.27228	21.41402	0

Control | Point Data | **Grid Data**

Figure 8. HeatMap Grid Data worksheet

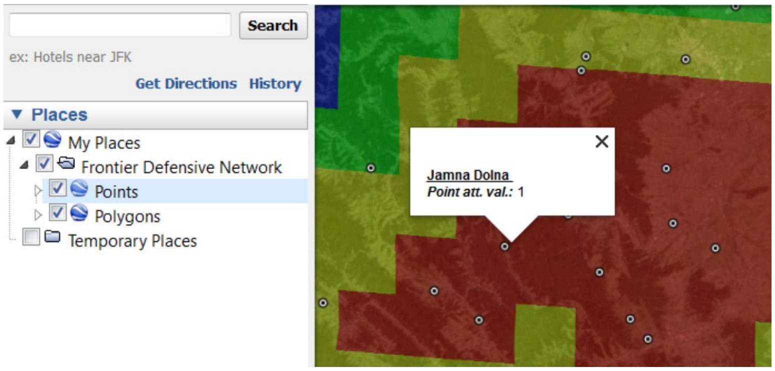

Figure 9. HeatMap KML file displayed by GoogleEarth