

Free Mapping Software for Butterfly Atlases: Combining Quantum GIS with Google Fusion Tables

Alan Macnaughton

49 Northforest Trail, Kitchener, Ontario, Canada N2N 2Y7 amacnaug@uwaterloo.ca

Overview

Obtaining the latitude and longitude for the locations of butterfly observations has become increasingly easy in recent years with the advent of portable GPS devices and map websites. Turning these observations into dots on a map is the next step. Thus, the websites of both the Butterflies and Moths of North America (BAMONA) and the North American Butterfly Association (NABA) provide a system for having these point observations appear on online maps.¹ However, butterfly atlas developers generally prefer to aggregate observations over geographic units in order to facilitate the preparation of species lists and to avoid disclosure of the exact locations of rare and endangered species.² Usually these geographic units are squares of 10 km per side,³ although political units are sometimes used.⁴

Mapping these 10 km squares has required the use of costly and knowledge-intensive commercial geographic information system (GIS) software programs such as ArcGIS, even for simple printed or PDF maps which do not involve user interaction ("static maps"). Resource requirements have been still higher for the production of dynamic maps for online access which, for example, display more detailed geographic information on the computer screen as the user zooms in to examine particular locations. As a result, butterfly atlases have generally been produced only by well-funded organizations with access to professional mapping staff.⁵ Butterfly atlases, regrettably, have been much rarer than bird atlases.

New software tools allow the production of dynamic butterfly atlas maps at no cost and with much less technical knowledge. In particular, open-source GIS software such as Quantum GIS ("QGIS") will assign individual observations to geographic units and Google Fusion Tables ("Fusion Tables") will layer those units on top of standard dynamic maps (Google Maps) for web access (without requiring the user to download Google Earth).⁶ The use of Fusion Tables to draw the maps is the key simplification. Microsoft Excel (or other spreadsheet software, such as the free product Calc from the OpenOffice.org suite) provides additional data manipulations and communication between QGIS and Fusion Tables. All of the software can be used on computers running Windows, Mac OS X and Linux.

Figures 1 and 2 are screen captures of maps prepared using this methodology.⁷ Figure 1 shows in red the eighty-nine 10 km squares in southern Ontario, Canada for which

at least one observation of the Silver-spotted Skipper has been recorded in the atlas period (primarily 2002-2009).

Figure 2 shows the additional capabilities made possible by Fusion Tables and its dynamic maps. This figure shows a screen capture where the user has zoomed in to examine the geographic details (roads, parks, etc.) of a particular square in northwest Toronto. The user has also clicked on the square to display its full species list as a text bubble, which partly covers the square. To see the general topography of the area, the user could click on the box on the top right of the map to switch to satellite view. To see photographs of the view from roads (including rural roads), just metres away from the subject, the user could click on the human figure on the top left to switch to street view.

Implementation

Although general documentation for QGIS, Fusion Tables and Excel is available, the challenge is determining the best way for these three applications to work together to create a butterfly atlas. The following set of 10 steps is based on the experience of developing, with collaborators Colin Jones and Ross Layberry, an atlas of Ontario's butterflies for the Toronto Entomologists' Association (TEA). Steps 1 to 5 involve QGIS and Excel, while steps 6 to 10 involve Fusion Tables.

Computer implementation details are set out in detailed endnotes, and the underlying computer programs are available on request. The resulting maps, from which Figures 1 and 2 are drawn, are accessed through the TEA website.⁸

Step 1 is to record all observations in a single Excel spreadsheet. In the case of the TEA, this involved 65,000 observations of 149 species. Approximately another 20 species occur in Ontario occasionally but were not present in the data.

For each butterfly observation, decimal latitude and longitude values ("georeferences") must be assigned.⁹ Typos are hard to avoid, and so some preliminary mapping of data points is needed to identify them. In particular, missing the negative sign before the longitude will put a point in Asia rather than Ontario.

Step 2 is to determine the required definitions of suitable geographic subunits. Sharing information with other researchers and naturalists' organizations is best facilitated if an established grid is used. For the TEA project, Bird

Studies Canada supplied a division of Ontario into 10,747 10 km Universal Transverse Mercator (UTM) squares in the industry-standard “shape file” format. Records were available for 1,334 of these 10 km squares. Few records were available for Ontario’s sparsely-populated northern area.

Step 3 is to assign each observation to its proper square using the mapping program QGIS.¹⁰ The output of this step is a file detailing, for each observation, the species name or number, latitude, longitude, and square ID (the 6-digit UTM-standard name).

Step 4 is to summarize the observations regarding the occurrence of a species in a particular square. Formally, the task is to create a table (a cross-tabulation) for which the rows are the squares, the columns are the species, and each cell entry is the number of records for that square-species combination.¹¹ In the TEA’s case, this created a table of 1,334 rows by 149 columns. A typical row (one square), such as the one appearing in the info window of Figure 2, includes the information that UTM square 17PJ14 has 1 Silver-spotted Skipper record, 1 Juvenal’s Duskywing record, and so on, for all 20 species recorded in that square.

Step 5 is to create a file which provides the computer description (polygon shape) for each square in which one or more species of butterfly has been recorded.¹² Fusion Tables requires this information to draw the map.

Step 6 is to create the first map in Fusion Tables. This is done by setting up a Google account, signing in, uploading the two files created in steps 4 and 5 (the species-squares file and the file of polygon shapes), and then merging the former to the latter by using the square ID as the common variable. The merged file is given a new name. With this merged file open as a table of data, click on “Visualize” and then “Map.” This first map shows all squares where any butterfly species has been recorded.¹³

Step 7 is to arrange for the display of a species list for a particular square when the user clicks on that square. Clicking on a square causes Fusion Tables to display an “info window” which provides, for that square, the names of certain variables (column labels of the table) and the values of those variables. The variables displayed are those chosen using the command “Configure Info Window”. An appropriate choice of variables can be used to display a species list for each square.¹⁴

Step 8 is to give permission in Fusion Tables for the public to view the map. This is done by executing the “share” command in Fusion Tables for each of the 3 files referred to in step 6. This also makes the underlying data publicly viewable, but this data is only the number of records of each species for each square—the actual observations are not uploaded to Fusion Tables.

Step 9 is to produce a map for each particular species by filtering out all observations but those for that species.¹⁵ Finally, step 10 is to create maps for all species and to put them on the web.¹⁶

Refinements

Executing the above 10 steps produces a basic butterfly atlas. Some extensions used in the maps displayed on the TEA website are:

- For each square, the info window lists the major contributors of observations by their initials. This serves as both a thank-you for the observations submitted to date and an incentive to submit more observations.
- Some maps display not only the species recorded for that square but also species which are likely to be observed there, such as those other species recorded in one or more of the 8 adjacent squares.
- The observations for a particular species are disaggregated using multiple square colors (e.g., red if the species has been observed in that square in the past 10 years, and yellow otherwise)¹⁷ and multiple maps (e.g., sight vs. photo vs. specimen, or for specific observers).¹⁸
- Maps are produced for counties as well as 10 km squares.

Discussion

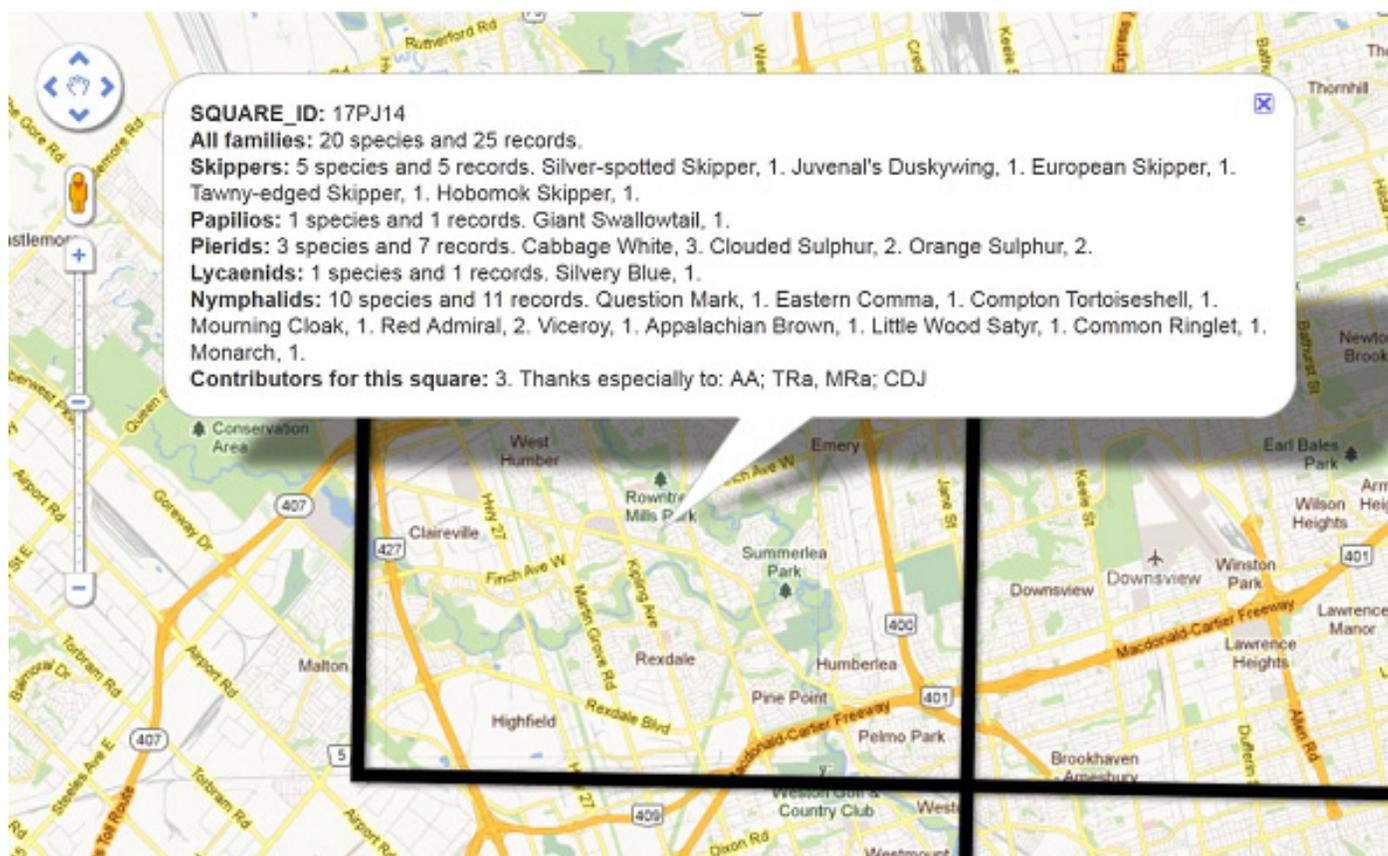
The above approach using free software is, of course, not as good as commercial mapping software. However, both QGIS and Fusion Tables have steadily been improved since their release in 2009.¹⁹ At this time, the issues with this software are as follows:

- QGIS is not as well-documented as commercial software in terms of help and instructions, although the basics are explained in many user-generated YouTube videos and more difficult questions can be posted on a users’ forum.²⁰ Fusion Tables presents fewer problems in this regard because the software is not as complex.²¹
- Sometimes QGIS will fail to process a file of data points if it contains formatting irregularities, and finding those irregularities becomes a time-consuming trial-and-error process. However, simply cutting and pasting the data into a new spreadsheet will often solve the problem.
- Fusion Tables has occasional problems in online display.²² Sometimes the filter will not take effect, that is, the map will still show the squares for which any species has been recorded, not the species in

Figure 1: 10 Km Squares in Ontario with Records of the Silver-spotted Skipper



Figure 2: Zoomed-in View of a Square, Showing the Species List



question. Clicking “refresh” seems to solve that problem, but this has to be repeated each time the problem appears. Also, the view of the map may be blocked by a message saying that the data is still loading; zooming in and out resolves this problem. Finally, for the display of the info window with the Firefox browser, the text will sometimes overflow the space and become unreadable.

- The use of three separate software products (QGIS, Fusion Tables and Excel) implies that updating the maps for new data takes perhaps three hours of the developer's time (and thus cannot be done weekly, as some bird atlases do).
- Fusion Tables limits any cell to a maximum of 1 million characters. By displaying the QGIS map side-by-side with the Fusion Tables map to check accuracy, it has been found that this limit means that only about a thousand 10 km squares can be viewed by a single mapping request; beyond that, squares may disappear without warning. This number of squares would only cover an area of about 320 km by 320 km (since 32 x 32 is about 1,000) if the species in question is recorded in every square. A partial solution is that with a Fusion Tables API, up to 5 separate mapping requests (“layers”) may be used to construct a single map. All 3,251 US counties have been displayed on a single Fusion Tables map in this way. Ontario bird atlases address this problem by showing the province as a whole in 169 blocks of 100 km per side; 10 km squares are used only for smaller areas.

Acknowledgements

Many people deserve thanks. Colin Jones conceived of the atlas project, established the database, and encouraged data contributions for almost ten years. Ross Layberry further developed the database and contributed a massive amount of personal records. Over 800 people contributed to the TEA seasonal summaries; more data is still welcome. Steve Xu and other staff of the Map Library of the University of Waterloo suggested the use of QGIS and Google Fusion Tables and helped with problems. Mandi Gingerich provided text-editing services.

For Further Research

- Newman Library, Baruch College, City University of New York, *GIS Practicum Manual: Introduction to GIS Using Open Source Software* (www.baruch.cuny.edu/geoportal/practicum/)
- Toronto Entomologists' Association, *Atlas Online* (www.ontarioinsects.org/atlas_online.htm)
- Wisconsin State Cartographer's Office, *Quick and Easy Web Maps with Google Fusion Tables*. SCO

Technical Paper, August 2011, 10 pages (www.sco.wisc.edu/images/stories/publications/SCO_quick_and_easy_web_maps_v1.0.pdf)

NOTES:

- 1 The sites are www.butterfliesandmoths.org and www.nababis.org. The BAMONA website provides for authentication of observations through the review of photographs by regional coordinators. A Canadian website is in development (www.macroecology.ca/butterflies/).
- 2 Mapping points rather than squares reduces the loading time of the map. However, mapping points requires that observations on rare and threatened species be shown as occurring at the centre point of the relevant geographic unit (BAMONA's policy) or be suppressed entirely. Either alternative raises thorny issues of whether one should disclose to users which observations have been geographically censored (although researchers, who are given access to the true data, are not affected).
- 3 A few “seam squares” near UTM zone boundaries are less than 10 km by 10 km in size and appear more like triangles.
- 4 The advantage of political units such as counties is that they are better-known. Thus, the Wisconsin Butterflies website (www.wisconsinbutterflies.org) is able to provide much distribution information using only a thumbnail map of the state showing the outlines of each county. However, counties may be much larger (e.g., for an average US county, 30 times the area of a 10 km square), may be inconsistent in size (e.g., from 122 sq km to 51,960 sq km in California), and may not be defined for some areas (e.g., northern Ontario). Shape files of counties for the US and Canada are available at <http://support.google.com/fusiontables/?hl=en>.
- 5 Two printed atlases are J. Asher, M. Warren, R. Fox, P. Harding, G. Jeffcoate and S. Jeffcoate (eds), *Millennium Atlas of Butterflies in Britain and Ireland* (Oxford University Press, 2001) and C. W. Leahy, B. Cassie, and R. K. Walton (eds), *Massachusetts Butterfly Atlas 1986-1990* (Massachusetts Audubon Society, 2006). Dynamic maps for the latter atlas are available online (www.massaudubon.org/butterflyatlas).
- 6 The sites for QGS and Fusion Tables are www.qgis.org and www.google.com/fusiontables/Home/ respectively. Fusion Tables covers the entire world. No copyright fees are charged if Google's crediting requirements are followed.
- 7 See, e.g., Snipping Tool (included in Windows Vista and Windows 7).
- 8 http://www.ontarioinsects.org/atlas_online.htm.
- 9 Some conversions of geographic-location data may be necessary. Latitude-longitude pairs in degrees, minutes and seconds must be converted to decimal values (www.cpearson.com/excel/LatLong.aspx). UTM easting and northing values also must be converted to decimal latitude and longitude (<http://www.uwgb.edu/dutchs/usefuldata/utmformulas.htm>). Decimal latitude-longitude pairs computed using the NAD27 datum need to be converted to the NAD83 or WGS84 datum (e.g., using the free Windows-based NADCON software: www.ngs.noaa.gov/TOOLS/Nadcon/Nadcon.shtml) to avoid errors in determining the proper square in step 3.
- 10 For the TEA application, this involved several steps. First, an ESRI shape file of the province of Ontario was uploaded to QGIS through the “Add vector layer” menu choice. (The province layer was made transparent, so that the 10 km squares could be seen on top of it, as follows: right-click on the province layer; choose Properties; click on the Change button showing the image of a wrench; and, under Fill Style, choose No Brush.) Second,

the shape file of 10 km squares from Bird Studies Canada was similarly uploaded and made transparent. Third, the Excel file of butterfly observations was saved in CSV format, and then this CSV file was uploaded to QGIS using the "Add delimited text layer" icon on the Plugins toolbar. Finally, the "Join attributes by location" command in the Vector>Data Management Tools section was used to assign each butterfly observation to a 10 km square. For this command, the "Target vector layer" was the observations file and the "Join vector layer" was the squares file. The dbf file produced in this process was then opened in Excel to become the spreadsheet used in Step 4.

11 Excel does this through the pivot-table command. In Excel 2010, for example, select the square ID and species name-number columns with the mouse, then do Insert > Pivot Tables > OK. Enter the square ID in Row Labels, enter the species name or number in Column Labels, and enter the square ID in Summation of Values.

12 This is a multi-step process. First, create a CSV-format file in which each row is the latitude and longitude of any point in a particular square (including in the file only squares in which one or more species of butterfly has been observed). Centre points of the squares are a good choice for this purpose, as they are likely to be provided as part of the shape file of the squares. Second, upload this file to QGIS and use the "Join attributes by location" command to assign each point to the polygon shape of its associated square. This use of the command is different from that in step 3 in that the "Target vector layer" is the squares layer and the "Join vector layer" is the points file. Third, delete the latitude and longitude of the centre points from the new layer (using Table Manager – get it using Plugins > Fetch Python Plugins > Repositories > Add 3rd Party Repositories > OK); otherwise, on subsequent upload to Fusion Tables, the centre points will appear on the map instead of the squares. Finally, the resulting layer should be saved in the KML format required for Fusion Tables.

13 By default, each square is a solid color which is 50% opaque. A better approach is to choose 1% opaque with a border width of 8 px (pixels) -- when viewing the map, use Configure Styles > Polygons. Then from the zoomed-out view (Figure 1) each square will be almost a solid color and thus still easy to see, but from a zoomed-in view (Figure 2) the borders of each square will narrow to reveal the geographic details inside the square.

14 The method used for the TEA application is to create a variable for each butterfly family name which contains text listing the species in that family found in that square. This involves adding extra columns to the square-species table created in Step 4. Text functions in Excel were used to create these columns.

15 For an illustration using the TEA application, consider filtering for the Silver-spotted Skipper, i.e., displaying the set of squares for which the number of records of this species is greater than zero. To do this, first click on Options > Filter. The left-most drop-down menu contains the column labels. Click on the arrow beside it and choose the Silver-spotted Skipper. In the middle drop-down menu, click on the arrow to toggle from "=" (equals) to ">" (greater than). Next, in the right-most box, type in the number "0" (zero). Finally, click "Apply", and the map appears (Figure 1).

16 There are 3 choices here. The ideal approach is to create a Fusion Tables API on the insect organization's website that retrieves each species map. Alternatively, the developer could, for each species, filter to produce the map, click the "get embeddable link" command to display the required HTML code, and then paste that link into the HTML code of the web page. A third approach is to use the "get link" command in Fusion Tables to create a single link which directs users to a page on the Fusion Tables site. Users must then do their own filtering (step 9). The TEA application presently uses this approach but the plan is to

develop a Fusion Tables API.

17 Chose Configure Styles>Polygons Fill Color>Buckets.

18 An additional variable must be added to the table for each potential map. However, tables exceeding about 3,000 columns become slow to load. Thus, multiple tables may be required.

19 Other free software may be available. The BAMONA website uses OpenLayers to superimpose butterfly data points on a standard Google Map, and ArcGIS Online advertises some free functionality for non-profit applications.

20 <http://forum.qgis.org>

21 Fusion Tables also has a users' group which, despite its name, covers all types of issues. See Fusion Tables API Users Group (<https://groups.google.com/forum/#!forum/fusion-tables-users-group>).

22 It is still described by Google as "beta" (unfinished) software more than two years after its release. However, this is a frequent Google labelling practice.