

Automated equal area cartograms - algorithm-led generation of equal area cartograms for anywhere, any period and any scale

Mitchell, Bruce JW¹, Tzelepis, George²

Office for National Statistics (ONS), Geography Branch

March 22nd 2018

SUMMARY

ONS Geography is developing a robust and dependable methodology for automating the generation of equal area cartograms (EACs). An EAC is a visualisation where every district is represented with an identical geometric object of equal size, shape and area. The intention is to publish the tool via GitHub. ONS Geography is working with the ONS Data Science Campus towards this end.

The ONS tool may be the first to provide UI-fronted and parameter-led method for users to create EACs for any space, for any period and at any scale.

KEYWORDS: equal area cartogram, automation, Python, hexagon, hexmap

1. Introduction

EACs may be created using GIS software, or a vector drawing package such as Adobe Illustrator, but these methods are slow and labour-intensive.

ONS Geography is therefore creating a parameter-driven cartogram creation engine using Python. Input requirements are an overall geographic boundary divided into sub-units, along with the sub-unit centroids, all rendered in a projection appropriate to the location, and in any format supported by OGR's API 'Fiona'³.

Internet resources exist for creating various types of cartogram, but most concentrate on the type where area shapes and sizes are distorted with respect to a given value^{4 5 6}. These may be very striking, but every single one has different spatial relationships, emphasizing the extremes for the variable of interest of the specific cartogram.

EACs appear to be a relatively underexploited area^{7 8 9}. Automating EACs appears to be unexplored in academic literature. Existing methods concentrate on the creation of a mesh over a region.

¹ bruce.mitchell@ons.gov.uk

² george.tzelepis@ons.gov.uk

³ <https://toblerity.org/fiona/README.html>

⁴ <http://www.arcgis.com/home/item.html?id=d348614c97264ae19b0311019a5f2276#overview>

⁵ <http://www.worldmapper.org/about.html>

⁶ http://www.clearlyandsimply.com/clearly_and_simply/2015/06/create-your-own-cartograms-in-tableau.html

⁷ <https://odileeds.org/blog/2017-05-08-mapping-election-with-hexes>

⁸ <http://hindsight.clerestories.com/2014/01/06/chris-mcdowall-hexagonal-maps/>

⁹ <https://github.com/PitchInteractiveInc/tilegrams>

Tableau® does offer the functionality as part of their commercial product, but only for the USA.¹⁰ London has also been ‘hexed’.¹¹ But both of these are relatively simple geographies.

ONS’ Python-based EAC-generation tool puts the user in the designer’s seat – providing a series of parameters that drive the form of the cartogram while ensuring that the result is reminiscent of the original space. The current development version generates hexagon-based EACs (hexmaps). A future version may extend support to square cells.

Development to date has been centred upon three geographies that present different challenges: Germany (at NUTS3 level), Russia (at Oblast) and the EU28 (NUTS2). However, the same methodological challenges apply regardless of period or geography, and from the microscopic to the galactic.

2. The dominance of large areas

Thematic mapping illustrates statistical difference across space. As such, the area of the map is conventionally tiled by administrative areas displaying different geostatistical values. But administrative systems often include both small urban districts and large rural expanses, and the latter tend to dominate maps purely on account of their size and regardless of socio-economic data interest.

At a global scale, and on certain projections, Canada and Russia dominate the whole: many smaller states are completely invisible. Equally, within Canada, densely populated cities along the US border are dwarfed by the vast lands to the north. Within a region, the vast majority of the data of interest may be located at a single location (take Las Vegas in Nevada). At the scale of the city, or the block, or the building, the same applies – there may be a few small areas where most of the interest lies. In all such cases, whatever the scale, a map may be dominated by large and relatively empty areas of little statistical interest.

In many countries, Census geographies (e.g. Output Areas (UK), Census Tracts (USA), Mesh Blocks (Australia)) have been developed to provide a statistically-sound geographical analysis network. But even these are subject (as a direct outworking of their population thresholds) are subject to the visual dominance of large areas. For mapping certain datasets, this may be unwelcome.

3. The cartogram

The visual dominance of such large areas may be set aside completely by scaling scale territories, not according to their actual geographic size, but according to a thematic value of statistical interest. This is known as a cartogram – a hybrid statistical graph / map, a compromise by which the viewer gains more immediate access to the data, but loses some of the geographical context. Cartograms exist in many forms: they may preserve, or ignore, shape, contiguity and orientation^{12 13}.

This distortion of geography can make it hard to navigate a cartogram - the reader can easily get lost and misinterpret the data. Where your original geographical shapes are complex, the distortion of geography in a cartogram may be so severe as to present a barrier to assessing value. WorldMapper’s

¹⁰ <https://www.tableau.com/about/blog/2017/1/viz-whiz-hex-tile-maps-64713>

¹¹ <https://twitter.com/benjmoss/status/689843030636023808>

¹² http://www.dannydorling.org/?page_id=92

¹³ <http://www.worldmapper.org/archive/>

“Predicted Distribution for the Estimated World Population in 2050”¹⁴, show Angola and Canada to have almost the same projected populations, but it is impossible to judge that from the cartogram. Which is a pity, as value comparison is precisely its purpose.

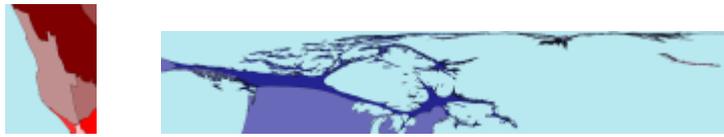


Figure 1 Angola and Canada taken from WorldMapper’s map “Estimated world population in 2050”

The fact that a cartogram is scaled on a variable (areas with high statistical values expand and those with low values shrink), makes it analogous to the conventional geographic map – where the ‘scaling variable’ is surface.

4. The equal area cartogram

This form of cartogram completely dispenses with the original shapes of the geographical components and substitutes geometry for geography. Large areas lose their dominance, and small areas are brought forward to equality. Each area is treated equally, and may then be shaded or coloured in the manner of a choropleth map to represent data values.

At the heart of the EAC lies the *tessellation of the plane* - the covering of a surface with one or more geometric shape (tile) with no overlaps and leaving no gaps. All tessellations (tilings) that require more than one shape to achieve this are eliminated, because all areas need to be compared on a common basis. Complex tile shapes are also eliminated - tiles have to be simple, identical in shape and identical in orientation – nothing should obstruct a pure interface with the data.

5. Python-based EAC Generation

The ONS EAC Generation tool is built on Python, with *GeoPandas*, *ShapeLy*, *NumPy*, *SciPy* and *MatPlot*. All the geospatial elements of *GeoPandas* are based on *OGR*’s API ‘*Fiona*’.

The geometric ‘input centroids’ of the areas to be represented in the EAC are translated into a *GeoPandas* *GeoDataFrame*, retaining their real-world coordinates and an appropriate projection.

The geometric median location of the input centroids is calculated using *SciPy* and a transformation function (comprising ‘gravity’ and ‘friction’ elements) shifts the input centroids towards the median location.

A ‘step distance’ unit is calculated from the number of input locations and the overall extent of the geography being represented (whether it be microns or parsecs). A regular grid of points is cast across the map using this unit, behind the geographic centroids that are to be mapped. With a shorter step distance unit, a dense grid will be generated, resulting in smaller hexagons. With a longer unit, the grid will be sparser and the hexes larger. The step distance must be set to produce at least as many grid points as there are input geographic centroids.

¹⁴ <http://www.worldmapper.org/display.php?selected=11>

The grid points are now used as seeds to generate hexagons¹⁵, which are translated into a *GeoPandas* *GeoDataFrame* and given the coordinate system of the input centroids.

The input centroids are sorted by distance from their median centre location, and sequentially snapped to the nearest available grid point.

3.1 Influences on the shape of the hexmap

3.1.1 Projection

The first step is to determine and select the appropriate projection for the geography to be represented. The significance of this will depend on how large that area is and where it is located. Guidance on the selection of appropriate projections for terrestrial maps is readily obtainable from the Internet¹⁶.

3.1.2 Interactions between a set of fixed and flexible parameters.

With a projection selected, certain general spatial characteristics of the geography are settled and may be considered *fixed parameters*.

3.1.2.a The external polygon boundary

The representation of countries such as Chile, Russia and Egypt all present different challenges. Enclaves, exclaves and ‘doughnuts’ may need special handling.

3.1.2.b The spatial distribution of the input centroids

A perfectly distributed set of input centroids – where all are equally distant from each other – will rarely occur. They more usually cluster around a single area (a unimodal pattern), or around several distinct areas (multimodal). The ONS model currently assumes a unimodal distribution, but future updates will provide a method for handling multimodal data.

3.1.3 Flexible parameters

These may all be set within the tool

3.1.3.a Contiguity

While the grid behind a hexmap will cover the entire geography, not every cell need be occupied. Spaces may be left – and this allows a non-contiguous thematic hexmap to be created that occupies a space more reminiscent of the original geography.

A fully contiguous hex-map contains no gaps, though this makes it much harder to reflect the shape of the geography being represented, so navigation becomes more difficult. On the other hand, this approach permits larger hexes and larger labels, enhancing readability.

3.1.3.b Orientation

Hexagons may be oriented either horizontally (pointing side-to-side) or vertically (up-down) and either alignment will produce a different arrangement of the hexes.

¹⁵ Using a ‘*Spatial Voronoi*’ technique within SciPy, and then ‘*Polygonize*’

¹⁶ For example: <http://projectionwizard.org/>

3.1.3.c. Hex size

Smaller hexes may be placed more accurately in geographical terms. The spatial relationships of larger hexes are increasingly generalised and they are more likely to have to shift location.

3.1.3.d/e Gravity and Friction functions

The gravitational attraction exerted by the geometric median location on the individual centroids and the friction over distance exerted on the individual centroids. The combination of the two determines the resting places of the area centroids.

As the fixed parameters and the settings for the flexible parameters combine, different results will be generated. It will be for the user to assess how successfully this is achieved.

The substitution of geometry for geography means that real-world spatial relationships cannot be properly represented – a hexagon has a maximum of six neighbours, while a square has a maximum of eight. Where an area has more than this maximum number of neighbours, the symbols have to shift to compensate.

3.2 Examples

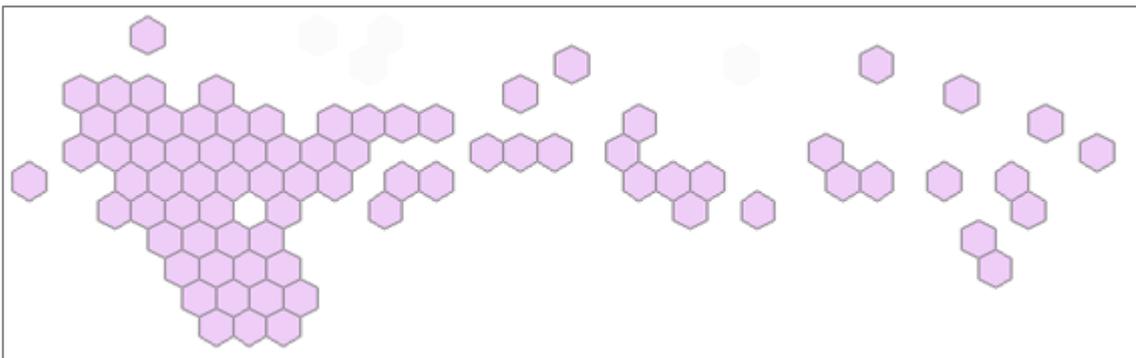


Figure 2: Equal area cartogram of Russia at district level

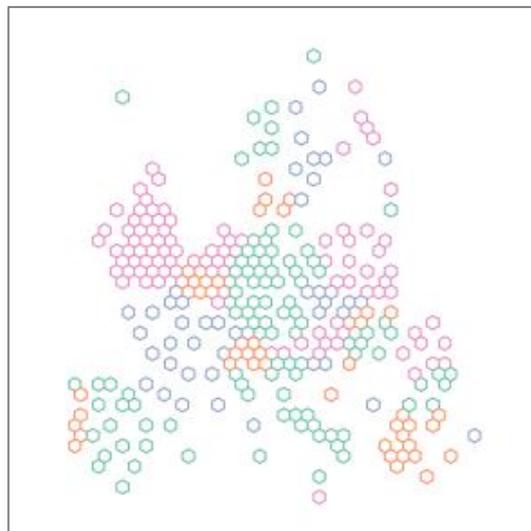


Figure 3: Equal Area Cartogram of the European Union and EFTA (NUTS2)

4. Summary

The few existing automated hex mesh creation methods are proprietary and are limited in scope to simple geographies. Manual methods are slow and laborious. The ONS prototype method is open source, very fast, and has been successfully used for complex geographies.

ONS is developing a partnership with the ONS Data Science Campus and expects to publish a robust UI-fronted version during the coming year. Future versions will extend its functionality to include square-based EACs.

5. Biographies

Bruce Mitchell heads the ONS GIS and Mapping Unit (GISMU). He has experience in small area estimation, spatial analysis, data visualisation and agent-based modelling of Dutch elm disease control on the Isle of Man.

George Tzelepis is a geoscientist and a GI researcher. He manages projects for a variety of clients, from small companies to HM Government, and his research interests focus on data science, machine learning and music.