An algorithm for a spatial mapping using a hexagon tilegram, with application to Australia disease maps

Stephanie

6 December 2018

Abstract

KEYWORDS:

Introduction

Representing the globe in 2D space has presented issues for many years. Methods have been established to visualise geographical space. We present a solution for alternate visualisations when the intention is to represent the spatial distribution.

Motivation

- Spatial distributions are not well presented when areas differ in size.
- We want to understand the distribution among population
- We can benefit from making the focus the distribution, not the geography

- Presenting spatial distributions
- Extremely small areas are lost not easily compared at a high level
- Choropleths; unequal space on the display, difficult to represent Australia
- Cartograms; strong distortions misrepresent Australia

- Unusual population settlement and dispersion over the geographical space: Australia Dense south east, sparse central and west

- Applying alternative maps that allow for dense population areas when cities that are population dense are close together

Background

- Choropleth methods are accurate geographically.
- Cartogram alternative map has been used historically.
- Spatial relationships are preserved.

Proposed Solution

- Like Dorling, same shape will be used. Like tilegram, tesselation using equally sized shapes.

- Allocate each geographic area to a map space of consistent tesselated shape

which is different from existing algorithms: - contiguous, keeps exact neightbours, no recognisable shape - non contiguous, keeps shape, can lose neighbours

- Clear to see spatial distribution, _ Should keep neighbours relationships intact as geographic neighbours will have similar values

Easier to read distribution at a distance, see a much larger area, general relationship at a glance

Works well for America, states are almost homogeneous Works well for England, London is the biggest issue, pushing out rural areas does not disturb population dense areas.

Our solution considers multiple population dense areas.

Allows dispersion to be increased when needed, and distance from origin to be considered

The Dorling maps in [@TVSSS] give examples of displacing central geographically small, population dense areas. This inspired the concept of this algorithm.

Algorithm

The algorithm operates on a set of polygons. There are parameters considered that may be provided, or will be automatically derived. All necessary functions are exported, with a main function `create_hexmap` used to step through automatically.

When utilising individual functions, we recommend the following approach:

Polygon Information

The collection of polygons should be considered first. There are several ways of providing the set of polygons. The following methods are available via the `read_shape` function.

To utilise a shape file from a directory on your device, provide the file path for .shp file. The `st_read` function is applied internally to create the sf object. If the shape file is an .Rda file, it is loaded into the current environment. Empty geometries are removed from the sf object, and the remaining polygons are projected to the Australian standard.

The geometry column now contains projected polygons, the projected centroids are derived for all polygons remaining in the set.

Hexgon grid

Create a grid of hexagon centroids using the `create_grid` function.

The following parameter values are necessary for this step, and will be automatically derived in the `create_hexmap` function if not provided:

- buffer distance: *the distance, in degrees, to extend the hexagon grid beyond the polygon centroids provided*
- size of the hexagons: *a float value in degrees for the width of the hexagons to be placed on the map space*
- filter distance: *the amount of hexagons around each centroid to consider for allocation, by default this is 10 hexagons*

The grid creation takes several steps:

Step 1: Expand the grid

Two sequences are used, one for longitude and another for latitude. The sequences begin at the minimum longitude or latitude, minus the buffer distance. Equally spaced intervals, the size of the hexagons, are created up to the maximum longitude or latitude, plus the buffer distance.

An individual point is created from all combinations of longitude and latitude sequences.

This grid is square, which will not facilitate tesselated hexagons. Every second latitude (row) of points will be shifted right by half of the hexagon size.

There are more points than necessary on this grid. To filter the grid for appropriate hexagons for allocation, the `create_buffer` function is used.

Step 2: Rolling windows

The longitude columns and latitude rows are given integer id values. The rows and columns are then divided into 20 groups, recording the amount of rows in each latitude group and the amount of columns in each longitude group.

The amount of rows and columns are then used to make manual rolling windows.

The first rolling window function finds the minimum and maximum centroid values for the longitude columns and latitude rows.

The second rolling window function finds the average of the rolling minimum and rolling maximum centroid values, for the longitude columns and latitude rows.

Step 3: Filtering the grid

Only the grid points between the rolling average of the minimum and maximum centroid valuesare kept, for each row and column of the grid.

Centroid to focal point distance

For each polygon centroid in the set. Find the distance to each of the focal points provided. The closest focal point name, the distance to the polygon centroid, and the angle from focal point to polygon centroid will be returned and added to the polygon data set in the polygon centroid row.

The distance between the polygon centroid and it's closest focal point data set is used to arrange the data set for allocation. The points are arranged in ascending order, from the centroid closest to any of the focal points, to the furthest.

Allocation of centroids

Allocation of all centroids can now take place using the set of polygon centroids and the hexagon map grid. For each polygon centroid, only the hexagon grid points that have not yet been used can be considered. The filter distance parameter is used to subset possible grid points to only those surrounding the polygon centroid within the filter distance. Using the angle between each polygon centroid, and it's the closest focal point, the subset of points is filtered again. Of these possible points, only those within a specific amount of degrees plus and minus a specific angle range are kept. This angle begins at 30 degrees by default, and may increase if necessary.

If no available hexagon grid point is found within the original filter distance and angle, the distance is expanded, only when a maximum distance is reached will the angle expand to accomodate more possible grid points.

The allocation is returned and combined with the data relating to each polygon.

Applications of algorithm

Examples of Melbourne, Victoria, Australia At different levels, distance from original centroids, change in area

LGA of all australia sa2 for queensland

Evaluation

Measure changes to areas: Each hexagon area in the set of *npolys* is standardised to 1 unit. Where 1 unit is the area of a hexagon of the designated size.

$area_{h_i} = 1$

$area_h = 1 * npolys$

y_i denotes the statistic of interest for polygon i, where i = 1,…,npolygs x_i denotes the spatial area for polygon i, where i = 1,…,npolys

Areal Statistic Distortion is the change in the areal distribution of the set of polygons. Difference between space taken by polygons and space taken by hexagons, with respect to the desired statistic. Setting the desired statistic to area givesthe change in map space used to represent the polygons.

The average when considering area used to display:

$A = \sum_{i=1}^{npolys} y_i x_i / \sum_{i=1}^{npolys} x_i$

$H = \sum_{i=1}^{npolys} y_i / npolys$

Using $w_i = x_i / \sum_{i=1}^{npolys} x_i$

$\sqrt{1/npolys \sum_{i=1}^{npolys} (y_i - \bar{y})^2 w_i}$

Considering only a local area distortion: Where the subset used contains only neighbours within a certain distance of the polygon centroid

Where $\bar{y}_d$ is the average of the polygon statistics considering only the polygons where the distance between polygon centroid x_i and polygon centroid x_p is less than d:

$d(x_i x_p) < d$

$\sqrt{1/npolys \sum_{i=1}^{npolys} (y_i - \bar{y}_d)^2 w_i}$

Summary

Discussion

Reference List