How to create a diffusion cartogram using Newman and Gastner's generously provided program

Carl Mason

March 19, 2007

Contents

1 Introduction

Diffusion cartograms are maps which are very cleverly distorted so as to maintain, to the extent possible, the original shapes (or recognizability) of boundaries while adjusting the areas of places to reflect underlying data.

The diffusion cartogram algorithm was devised by Newman and Gastner and reported in the PNAS in 2004. It represents a very elegant solution to a rather difficult problem. See

`http://www.pnas.org/cgi/content/short/101/20/7499` for a complete explanation.

2   HowTo

There are 4 steps to creating a diffusion cartogram:

1. Obtain an "ESRI *shape file*" of the regions you wish to map

2. "ungenerate" the shape file (i.e. convert it to an ASCII format)

3. create a data file with values corresponding to each of the regions that you wish to distort

4. run the cartogram program (written by Gastner and Newman)

None of the steps are particularly challenging, but the procedure is a bit rough around the edges, and of course this document is riddled with errors.

3   Obtaining a shape file

The Gastner and Newman program (which does the real work) requires as input an *ungenerated shape file*. Shape files refer to a format developed by ESRI for Arc/info but are used fairly widely. By "ungenerated" ESRI means that the data are in ASCII format instead of some unreadable binary format.

An easy way to obtain a shape file for a chunk of US Census Geography is by exporting a thematic map from the CensusCD program which we license from Geolytics. (See Section 3.2 for instructions). But that is not the only way – nor is it a particularly useful way to get a shape file for something *other than* US Census geography. For example school districts in California or villages in Rwanda are not well covered by CensusCD[1].

Luckily there are lots and lots of sites on the internet from which shape files are available. The USGS, for example:
`http://nationalatlas.gov/atlasftp.html`.

If you can get a shapefile, you can make a diffusion cartogram. Of course it is helpful if the shapefile you have is of a geographical region that you and your audience are familiar with and for which you have some interesting data.

---

[1]The US Census Bureau does have shape files for CA school districts available for downloading. The data are just not in the CensusCD program

3.1 What is a shapefile?

Shapefiles generally consist of at least 3 and often more files, the 3 that have to be there are .shp, .shx and .dbf. The .shp and .shx contains actual geographic information – instructions on how to draw polygons or lines or points. The .dbf file contains data that relate to the lines, points or polygons in the .shp file.

The .dbf file can be opened and manipulated as a spreadsheet using `scalc` or some other spreadsheet program. To create an interesting cartogram, You'll probably want to do exactly that (unless you have already done the data manipulation in the program from which you acquired the shapefile e.g. CensusCD2000).

3.2 Using CensusCD2000

Just type:

> **@:>** `censuscd2000`

The manual for CencusCD can be found at
`http:www.demog.berkeley.edu/VirtualLab`

A nice short step by step guide might still be findable on the web at
`http://www.dartmouth.edu/~maproom/guides/censuscd.html`.

But if you are too lazy to read anything, you can probably figure it out anyway. The most unintuitive bit is that you need to create a "Request" file to hold the data you are working with. File→New Request let's you do this.

Once you have managed to create a thematic map and have it displayed in the map viewer, select Map→Export→Shape File

You do not get to specify a name, CensusCD will create **three** files in your home directory (regardless of where you launched censuscd from). They will be called `whatever.dbf`, `WHATEVER.shx` and `WHATEVER.shp` where `whatever` is the name you specified for you `.req` file. NOTE the odd cases.

4 Ungenerating the shape file

To convert the three files (.dbf, .shp, and .shx) which comprise the "shape" file, you can use the perl program which I wrote for this purpose[2].

---

[2] my perl program relies very heavily on the Geo::ShapeFile module written by Jason Kohls

1. rename the three files so that they all have the same filesytem. That is make them all upper or all lower case. **(This is probably only necessary if you generated your shapefiles using CensusCD)**

2. run dos2unix on the .dbf file – or else you'll get a message to the effect that there is one extra character in that file. **(This too is probably only necessary if you generated your shape files using CensusCD or some other Windows program)** Here's the command to do that:

   @:> `dos2unix whatever.dbf`

3. run ungenerate-shape-file.

   @:> `ungenerate -s whatever`

   where `whatever` is the common name of your three components of the shape file with no suffix. `ungenerate` will create a subdirectory called `whatever-CART`, and in that subdirectory you *should* find two interesting files:

   (a) `map.gen` is the the *ungenerated* shape file.
   (b) `census.dat` is a rough stab at the data file that you need in order to run `cartogram`. Section 5 explains what you need to do in order to convert this file into something usable.

5   Preparing your cenus.dat file

the file `census.dat` is an ASCII file with a simple format:

- column 1 contains the number corresponding to a polygon. Your shape file is composed of polygons, each of which corresponds to a geographical entity. If your map is of the US states, you probably have 51 polygons (Washington DC is generally included). Each polygon has to have a number generally these start at 1 and march up to the number of polygons in the file (51 in the example above).

  So Column 1 is just a sequence of integers from 1..

4

- column2 contains the data by which your map is to be distorted. For example if your cartogram is to show the number of foreign born persons in each US state, then column 2 would contain these counts of foreign born persons.

- column 3... are ignored. ONLY the first 2 columns matter to the cartogram program. But if you are lucky, `census.dat` has at least one other column that gives you some clue as to which polygon/geographic entity the row refers to. If you are **really** lucky, `census.dat` might have another column containing the actual data that you want to use.

`ungenerate` produces a `census.dat` file with the second column filled with 100's. Thus the file as it arrives would be appropriate for creating a cartogram which converts all of the geographic entities into blobs with the same area. If you want to create a different cartogram, then you will want to change the entries in the second column.

The easiest way to edit `census.dat` is with a spreadsheet program. To use StarOffice, **From the whatever-CART directory** do the following:

@:$>$ `ln -s census.dat census.csv`

@:$>$ `scalc census.csv`

You should be asked to specify the delimiters, check `space` and `Merge` clear all other boxes.

When you are ready to save the file be careful to save it as a "Text CSV" file – spreadsheets always try to trick you into saving in their preferred format.

the census.dat file also contains all of the columns in your .dbf file - but to make life interesting, the column heading have been removed. So before you run `ungenerate` make sure the column you want to distort by is near the left side and maybe write down a number or two just to be sure you get the right one.

6   Running the cartogram program

Once you have constructed the ungenerated shape file which must be called `map.gen` and the data file which must be called `census.dat` you can create your cartogram by simply typing **from inside the whatever-CART directory**

5

@:> `cartogram`

After a few minutes of fascinating output, the result will be three files:

1. map.ps is a postscript file of the original map based on the polygons in the shape file (map.gen).

2. cart.ps is a postscript file of the diffusion cartogram that you just created

3. cartogram.gen is an ungenerated shape file of the cartogram you just created (you probably don't need it but who knows).

you can view these files with `gsview`:

@:> `gsview cart.ps`

and you can print them or include them in a LaTeX file just as you would with any postscript file. You can also convert the cart.ps file to .pdf so you can email it to your mom:

@:> `ps2pdf cart.ps`

7   More information

The program that Gastner and Newman provided comes with a readme document. If you want to see it use acroread to open: `/usr/local/Cartogram-Gastner-Newman/readm`

6