Map Making for Social Scientists

American Sociological Association
Anaheim, 20 August 2001

Waldo Tobler
Professor Emeritus of Geography
University of California at Santa Barbara
http://www.geog.ucsb.edu
Some hot topics in contemporary cartography

Animation of geographical objects

Three dimensional visualization

Map making on the internet

Map generalization
I will emphasize three other subjects

Map projections

Dealing with aggregate data
  Spatial filtering
  Estimating densities
  Converting to other units

Depicting movement
First, very quickly, map projections
The mapping process
Common Surfaces Used in Cartography

[Diagram showing Earth, Ellipsoid, Sphere, Map Plane U,V, and Map Plane X,Y]
The **surface** of the earth is two-dimensional
Sphere or Ellipsoid?

The departure of the earth from a sphere is approximately one part in three hundred

This is $\frac{3}{10}$ of one percent

This can be used as a rule of thumb:
Is your work accurate to better than one percent?
Sphere or Map?

This is equivalent to asking whether you want to work in latitude and longitude or in plane coordinates. Programs exist, for example, to convert from street address to lat/lon. There are also programs to convert from lat/lon to X, Y, and visa versa. Many kinds of analysis are very simple on a sphere. This includes such things as distance, direction, or area computation. A plane is a sufficiently good approximation to a sphere for a small area. What is small? You can glue a postage stamp, without wrinkling it, on a 20 cm globe.
Many analytical problems can be solved directly in geographic coordinates.

This is often easy when the earth is considered spherical.

It is more difficult to work in ellipsoidal coordinates.

Some people like to work in plane, Euclidean, coordinates. Then a map projection is needed.

Of course the projection must be suited to the problem, and there are many choices.
Plane Coordinate Systems Are Based on Map Projections

The two most important ones are

- The Universal Transverse Mercator system
- The State Plane Coordinate system

The equations for both are complicated and based on an ellipsoid

Virtually all countries of the world have similar systems
All map projections result in distorted maps!

Since the time of Ptolemy the objective has been to obtain maps with as little distortion as possible. Most geographic information systems and government mapping agencies take this point of view. But then Mercator changed this by introducing the idea of a systematic distortion to assist in the solution of a problem. Mercator’s famous anamorphose helps solve a navigation problem. His idea caught on.

Anamorphic projections are used to solve problems and are not primarily for display.
One way to use map projections

It is useful to think of a map projection like you are used to thinking of graph paper

Semi logarithmic, logarithmic, probability plots, and so on, are employed to bring out different aspects of data being analyzed

Map projections may be used in the same way

This is not a common use in geographic information systems
Hägerstrand’s Logarithmic Map
A map projection to solve a special problem

The next illustration shows the U.S. population assembled into one degree quadrilaterals.

We would like to partition the U.S. into regions containing the same number of people.

There follows a map projection (anamorphose) that may be useful for this problem.
US Population By One Degree Quadrilaterals
Now Use the Transform-Solve-Invert Paradigm

Transform the graticule, and map, to obtain areas of equal population

Then position a hexagonal tesselation on the map

Then take the inverse transformation

Next topic

Often we deal with data given by areal units
Such as census tracts, counties, states, or other administrative units
It is convenient to think of the data as being binned into these spatial units in a manner similar to the making of histograms
The difference is that the bins are of irregular sizes, shapes, and orientation on the surface of the earth
A choropleth (area filling) map with shading proportional to density
The same data shown as a bivariate geographic histogram
with bin heights proportional to density
I will consider three problems relating to such binnings

1. The filtering of data in the irregular spatial units
   including map generalization

2. Converting to continuous densities

3. Converting between areal units
Spatial filtering typically uses nearby, local, observations

Processing using neighbors is common in image processing. The value in a cell is converted to a weighted average of the values in neighboring cells. Depending on the weights one obtains either smoothing (a.k.a. blurring) or sharpening. Local geographic measures are similar in that they compute a value at each location that depends on nearby values. There are many examples.
Modifying the center cell in the case of pixels

Neighborhood operators are used frequently in image processing
Neighborhood Operators Can Also Be Used With Resels
First and Second Order Neighbors of Kansas
Choropleth map of university attendance
Adrian Herzog, Zürich
University attendance, adjusted
Adrian Herzog, Zürich
Unemployment, June 2001, by county
USA Today, 20 August 2001, page 4B
US unemployment map, two detail views

Brown: < 3.3%, Tan: 3.3-4.4%, Green: 4.5-6.2%, Red: >6.3%
Now a word about resolution

Average resolution can be calculated as

\[
\left( \frac{\text{area of domain}}{\text{number of observations}} \right)^{1/2}
\]

In three dimensions use the cube root

In effect this measures the average distance influence of each observation

Unequal resolution in different parts of a map has an effect similar to unequal magnification in a microscope
Average resolution ~55 km. Patterns >110 km detectable

In these resels the resolution varies across the US. Patterns within cities cannot be seen.
Social data are often made available in a hierarchy of administrative units

Moving up through the hierarchy changes the resolution and this acts as a low pass spatial filter

The result is a less detailed - more blurred - map

Consequently I recommend using the finest data available
For example
The Dutch administrative hierarchy
Swiss migration at reduced resolution
To emphasize the filtering effect of resolution
Another type of map generalization

 Courtesy of Dr Guido Dorigo, University of Zurich

14.7 km resolution (184 Districts)                           39.2 km resolution (26 Cantons)
Three levels of administrative units and three levels of migration resolution all at once.
An across boundary problem
Courtesy of Dr. Claude Grasland, Paris
In order to “uniformize” the resolution the bins in France are aggregated up the political hierarchy.

They then more nearly match the resolution of the Belgium information.

Had this not been done the resulting density for France would appear to have much more variability than that of Belgium.

But this variability would be an artifact of the difference in resolution.
Population along the French-Belgium border
Courtesy of Dr. Claude Grasland, Paris

VARIATION OF POPULATION 1980-1990 BY CANTONS (FR) AND COMMUNES (BEL)

Sources: EUROSTAT (Sire); MEGRIN (Sabe)
(c) C. Grasland, 2000, UMR Géographie-cités / The Hypercarte Project
Conversion From Areal Units to Densities
A Gaussian kernel function

The data values are assigned to the centroid of the administrative units and then summed using weights taken from a sliding kernel function.

Exponential kernels are also used.
How this works

Position the chosen kernel on the map
Search for all centroids within the kernel
Pick a weight from the kernel depending on the
distance of the centroid from the map location
Multiply the value at a centroid by the kernel weight
Sum all of the weighted values within the kernel
and assign this value to the location of center of the
kernel
Move to the next location and repeat
After all locations have been evaluated you are done
and can contour the results
Density based on a Gaussian kernel with a 5 km span
Using a Gaussian kernel with a 10 km span
Three references for further reading on density estimation techniques


Kernels can also be applied to dot maps

Each dot is assigned a value of one unit
(dots with numerical values can also be used)
The distance of each dot from the center of the kernel is calculated
Then the dot values are modified by the kernel weight
The weighted values within the kernel are summed and assigned to the location of the kernel center
The map is complete when the sum has been calculated for all locations
Thus the dot distribution has been converted to a density map
A uniform kernel is often used but is not recommended because of its effects.

This kernel inverts some peaks and valleys.

There is also a method that avoids the use of kernel functions.

It is sometimes referred to as areal interpolation. From this point of view it is incorrect, in my opinion, to assign areal observations to points (centroids). One criterion to be satisfied is that the resultant maintain the data values within each unit. The method is known as pycnophylactic reallocation.
Pycnophylactic Reallocation
(Mass Preserving)

Allows the production of density or contour maps to be made from areal data.

It is reallocation - and somewhat of a disaggregation operator. My assertion is that it may actually improve the data.

It is also important for the conversion of data from one set of statistical units to another, as from census tracts to school districts.
1st example

Population density in Kansas
by county

Courtesy of T. Slocum

A piecewise continuous surface
Population density in Kansas after mass preserving reallocation

Each County Still Contains the Same Number of People

A smooth continuous surface, with population pycnophylactically redistributed
Another example

Migration from Illinois shown first as a piecewise continuous bivariate geographical histogram, based on state outlines, with volumes according to Illinois outmigration.

Recall that most migrants in Illinois relocate within the state.

The same data is then shown as pycnophylactically interpolated.

The smoothed surface can be partitioned to yield estimated migration by arbitrary regions - the Great Lakes basin for example.
Bivariate histogram of Illinois outmigration by state
Illinois outmigration pycnophylactically smoothed
Another example

This time using population data by Federal Planning Regions for Germany.

First the data are represented in a perspective view of a bivariate geographical histogram.

This is followed by a similar view of the continuous population density distribution.

How pycnophylactic reallocation works

Philosophically it is based on the notion that people are gregarious, influence each other, are mobile, and tend to congregate. This leads to neighboring and adjacent places being similar.

Mathematically this translates into a smoothness criterion (with small partial derivatives).

It applies to any data exhibiting spatial autocorrelation.
1. Data polygons
2. Rasterized
3. Smoothed
How the smoothing is done.

Imagine that each unit is built up of colored clay, with a different color for each unit. The volume of clay represents the number of people, say, and the height represents the density. In order to obtain smooth densities a spatula is used to smooth the surface, but no clay is allowed to move from one unit into another. Color mixing is not allowed.

This, converted to mathematics, is what the computer program does.
Density from dot maps without using kernels

The pycnophylactic method can also be used to prepare smooth density maps from data given at spot locations.

Step 1. Use the inverse area of Dirichlet (a.k.a. Thiessen) regions as the density for each location.
If weights are attached to the locations divide these by the region area.

Step 2. Smooth the resulting densities by the pycnophylactic reallocation method.
Another important advantage of mass preserving reallocation

A frequent problem is the reassignment of observations from one set of collection units to a different set, when the two sets are not nested nor compatible. For example converting the number of children observed by census tract to a count by school district. Area boundaries also usually change over time, requiring reallocation for compatibility.

The density values obtained using the smooth pycnophylactic method allow an estimate to be made rather simply. A “cookie cutter” can cut the continuous (clay) surface into the new zones with subsequent addition (summation) to get the count.
The last topic is the depiction of geographical movement.

A great deal of change in the world is due to geographical movement.

Movement of information, of people, of money, or of material.

Animation is well suited to depicting this dynamic cartography.
Tables are an important way of recording data on geographic movement

Especially when the rows and columns refer to known geographic locations

The tables are then “square”, having the same number of rows as columns

The entries in the tables record the amount of movement during some period of time

Such tables can be decomposed into two parts, a symmetric part and a skew symmetric part

For the statisticians in the audience the total variance can also be partitioned into these two parts
From B to A is not the same as A to B

(Gary Larson)
An example

In the United States the currency indicates where it was issued.

For bills this is the Federal Reserve District. Coins contain a mint abbreviation.

You can check your wallet to estimate your interaction with the rest of the country.
Dollar Bill
(Federal Reserve Note)
The 12 Federal Reserve Districts
(Alaska and Hawaii Omitted)
### Movement of One Dollar Notes

Between Federal Reserve Districts, in hundreds, Feb. 1976

<table>
<thead>
<tr>
<th>To:</th>
<th>B</th>
<th>NY</th>
<th>P</th>
<th>Cl</th>
<th>R</th>
<th>A</th>
<th>Ch</th>
<th>SL</th>
<th>M</th>
<th>K</th>
<th>D</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston</td>
<td>2040</td>
<td>289</td>
<td>47</td>
<td>52</td>
<td>137</td>
<td>118</td>
<td>90</td>
<td>10</td>
<td>16</td>
<td>15</td>
<td>13</td>
<td>138</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>143</td>
<td>414</td>
<td>860</td>
<td>84</td>
<td>342</td>
<td>130</td>
<td>134</td>
<td>8</td>
<td>25</td>
<td>10</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>Cleveland</td>
<td>68</td>
<td>192</td>
<td>47</td>
<td>1296</td>
<td>171</td>
<td>177</td>
<td>618</td>
<td>16</td>
<td>44</td>
<td>43</td>
<td>19</td>
<td>131</td>
</tr>
<tr>
<td>Richmond</td>
<td>150</td>
<td>266</td>
<td>158</td>
<td>226</td>
<td>3899</td>
<td>578</td>
<td>295</td>
<td>20</td>
<td>62</td>
<td>54</td>
<td>22</td>
<td>152</td>
</tr>
<tr>
<td>Atlanta</td>
<td>122</td>
<td>159</td>
<td>57</td>
<td>186</td>
<td>319</td>
<td>3741</td>
<td>439</td>
<td>30</td>
<td>51</td>
<td>78</td>
<td>102</td>
<td>189</td>
</tr>
<tr>
<td>Chicago</td>
<td>97</td>
<td>155</td>
<td>39</td>
<td>496</td>
<td>143</td>
<td>266</td>
<td>5630</td>
<td>74</td>
<td>278</td>
<td>100</td>
<td>40</td>
<td>290</td>
</tr>
<tr>
<td>St. Louis</td>
<td>31</td>
<td>56</td>
<td>14</td>
<td>142</td>
<td>80</td>
<td>201</td>
<td>573</td>
<td>342</td>
<td>46</td>
<td>128</td>
<td>47</td>
<td>109</td>
</tr>
<tr>
<td>Minneapolis</td>
<td>14</td>
<td>26</td>
<td>11</td>
<td>32</td>
<td>29</td>
<td>41</td>
<td>295</td>
<td>10</td>
<td>1438</td>
<td>51</td>
<td>14</td>
<td>138</td>
</tr>
<tr>
<td>Kansas city</td>
<td>20</td>
<td>41</td>
<td>8</td>
<td>55</td>
<td>40</td>
<td>71</td>
<td>215</td>
<td>33</td>
<td>120</td>
<td>811</td>
<td>86</td>
<td>247</td>
</tr>
<tr>
<td>Dallas</td>
<td>31</td>
<td>41</td>
<td>8</td>
<td>38</td>
<td>46</td>
<td>165</td>
<td>125</td>
<td>20</td>
<td>37</td>
<td>253</td>
<td>788</td>
<td>203</td>
</tr>
<tr>
<td>San Francisco</td>
<td>82</td>
<td>81</td>
<td>23</td>
<td>84</td>
<td>114</td>
<td>106</td>
<td>251</td>
<td>22</td>
<td>127</td>
<td>128</td>
<td>43</td>
<td>5380</td>
</tr>
</tbody>
</table>
The table of dollar bill movements was obtained from MacDonalds outlets throughout the United States.


From the table we can compute a movement map.
Dollar Bill Movement in the U.S.
The map is computed using a continuous version of the gravity model.

The result is a system of partial differential equations solved by a finite difference iteration to obtain the potential field.

This can be contoured and its gradient computed and drawn on a map.

First the Federal Reserve Districts are “rasterized”

There will be one finite difference equation for each node on this raster (2088 simultaneous equations)
Solving the equations yields the potential

Shown here by contours

The raster is indicated by the tick marks. The arrows are the gradients to the potentials. The streakline map is obtained by connecting the gradient vectors.
The same technique can be applied to other types of movement

For example the migratory movement of people.
Nine Region Migration Table
US Census 1965-1970
(Note asymmetry. There are places of depletion and accumulation.)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New England</td>
<td>—</td>
<td>180,048</td>
<td>79,223</td>
<td>26,887</td>
<td>198,144</td>
<td>17,995</td>
<td>35,563</td>
<td>30,528</td>
</tr>
<tr>
<td>2</td>
<td>Mid-Atlantic</td>
<td>283,049</td>
<td>—</td>
<td>300,345</td>
<td>67,280</td>
<td>718,673</td>
<td>55,094</td>
<td>93,434</td>
<td>87,987</td>
</tr>
<tr>
<td>3</td>
<td>East North Central</td>
<td>87,267</td>
<td>237,229</td>
<td>—</td>
<td>281,791</td>
<td>551,483</td>
<td>230,788</td>
<td>178,517</td>
<td>172,711</td>
</tr>
<tr>
<td>4</td>
<td>West North Central</td>
<td>28,977</td>
<td>60,681</td>
<td>286,580</td>
<td>—</td>
<td>143,860</td>
<td>49,892</td>
<td>185,618</td>
<td>181,868</td>
</tr>
<tr>
<td>5</td>
<td>South Atlantic</td>
<td>130,830</td>
<td>382,565</td>
<td>346,407</td>
<td>92,308</td>
<td>—</td>
<td>252,189</td>
<td>192,223</td>
<td>89,389</td>
</tr>
<tr>
<td>6</td>
<td>East South Central</td>
<td>21,434</td>
<td>53,772</td>
<td>287,340</td>
<td>49,828</td>
<td>316,650</td>
<td>—</td>
<td>141,679</td>
<td>27,409</td>
</tr>
<tr>
<td>7</td>
<td>West South Central</td>
<td>30,287</td>
<td>64,645</td>
<td>161,645</td>
<td>144,980</td>
<td>199,466</td>
<td>121,366</td>
<td>—</td>
<td>134,229</td>
</tr>
<tr>
<td>8</td>
<td>Mountain</td>
<td>21,450</td>
<td>43,749</td>
<td>97,808</td>
<td>136,683</td>
<td>89,806</td>
<td>25,574</td>
<td>158,006</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>Pacific</td>
<td>72,114</td>
<td>133,122</td>
<td>229,764</td>
<td>165,405</td>
<td>266,305</td>
<td>66,324</td>
<td>252,039</td>
<td>342,948</td>
</tr>
</tbody>
</table>

This is an example of a census migration table. There are also (50 by 50) state tables and county by county tables.
There is a great deal of spatial coherence in the migration pattern.

Choropleth maps do not show this clearly. In the U.S. case the state boundaries hide the effect. Therefore a clearer picture emerges if they are omitted.

There is also temporal coherence.

Gaining and Losing States

Based on the marginals of a 48 by 48 migration table
State centroids used with symbol magnitude proportional to the amount of change
The conventional net movement map
Based on movement between state centroids
(Computer sketch. Optimum deletion: values below mean ignored)
This information can be converted to a potential field and its gradient.

For this a model is required.

The model is, in essence, a continuous version of the familiar gravity model.

The gradients can also be connected to give a streakline map.

The next maps are based on the same observations as the previous map.
The pressure to move in the US
A continuous spatial gravity model from a movement table
Recall that several million people migrate during the 5 year census period.

The next map shows an ensemble average, not the path of any individual.

But observe, not unrealistically, that the people to the East of Detroit tend to go to the Southeast, and Minnesotans to the Northwest, and the remainder to the Southwest.
Migration potentials and streaklines
The streaklines are drawn by connecting the gradients to the potentials
By the insertion of arbitrary areal boundaries, and by measuring the amount of flux across these boundaries, one can obtain information not contained in the original data, i.e., make a prediction.

It’s like using a cookie cutter pressed into the continuous flow model to look at an arbitrary piece and computing the flow across its borders.

The next map is an example, using state boundaries. The US Census Bureau does not provide this information.

The model is used to make the prediction.
Major Flux Across State Boundaries
Predicted from the model and table marginals
If we used the 3,141 counties of the United States the migration table could contain 9,862,740 numbers.

This is not a lot for a computer, but for humans?

We need models and visualization techniques!

Cartography provides excellent visualization and always requires a model.
To conclude

I have emphasized three topics

Map projections

Dealing with aggregate data
  Spatial filtering
  Estimating densities
  Converting to other units

Maps of movement
Thank you for your attention

For more detailed information go to
http://www.geog.ucsb.edu/~tobler/presentations
and
http://www.geog.ucsb.edu/~tobler/publications/reprinted articles

Waldo Tobler