Spatial Analysis Visualized

How many words are a graphic worth?

Some examples from my work.

An introduction using eleven examples.

W. Tobler
Geographer
Abstract:

Some quantitative and computational procedures are presented in simple graphical form.

A few examples from my research in geography. Included are spatial and temporal lag effects, comparison of movement patterns, consequences of changing spatial resolution, migration potential and fields, impact of biproportional adjustment on movement tables, coalescent cities, explanation of multidimensional scaling iterations, the transform-solve-invert paradigm, a map comparison method, mass preserving reallocation of spatial data, and potentials from asymmetry.
First Visualization

Spatial and temporal lags in movement tables illustrated and made obvious.
This is an example of a census migration table. There are also (50 by 50) state tables and (3100 by 3100) county by county tables.

Can you comprehend a table of over 9 million numbers?

I know that can’t!
It is well known that spatial autocorrelation is present in migration tables.

See, for example:
Curry, 1972, “A spatial analysis of gravity flows”, Regional Studies, 6: 131-147

Accounting for these effects in a model is difficult.

See, for example:
There is a great deal of spatial coherence in the migration pattern.

In the US case the state boundaries hide the effect.

Therefore they should be omitted.

There is also temporal coherence.

The Population Change Information Can Be Positioned Locationally using centroids.

This type of information is often shown using choropleth maps. I consider this method to be simpler and superior.

Observe the spatial autocorrelation and how this is brought out more clearly by omitting the collection unit boundaries, as on the next map.
The map is even better if the symbol size is made proportional to the magnitude of the change, as on the next map.
Gaining and Losing Migration States
Symbol positioned at the state centroids, and proportional to magnitude of the change.

The map is based on the marginals of a 48 x 48 state to state migration table and shows the accumulation and depletion places. Draw a boundary around the losing states. This demonstrates that states are not the appropriate size for studies of migration and also that there is a great deal of correlation amidst state migration data.
The previous maps make it clear that there is spatial autocorrelation.

It is not necessary to invoke complicated statistics to see this.

The maps make it obvious.

That there is temporal coherence is also easily visualized using movement maps.
The CSISS Mission recognizes the growing significance of space, spatiality, location, and place in social science research. It seeks to develop unrestricted access to tools and perspectives that will advance the spatial analytic capabilities of researchers throughout the social sciences. CSISS is funded by the National Science Foundation under its program of support for infrastructure in the social and behavioral sciences.

<table>
<thead>
<tr>
<th>Core Programs</th>
<th>Learning Resources</th>
<th>Spatial Resources</th>
<th>Spatial Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSISS has six research initiatives and a professional development program for undergraduate instructors.</td>
<td>These introductory materials include CSISS Classics and select video clips from the CSISS summer workshops.</td>
<td>CSISS has compiled e-journals, bibliographies, and other spatial resources for the social sciences.</td>
<td>Spatial Tools Search Engine</td>
</tr>
<tr>
<td>Search Engines</td>
<td>CSISS Events</td>
<td>Community Center</td>
<td>Select Tools</td>
</tr>
<tr>
<td>Try out one of our custom search engines to find spatial analysis resources on the Internet.</td>
<td>Here’s where you’ll find information and registration for workshops, conferences and specialist meetings.</td>
<td>Join the forums, or if your organization relates to our mission and goals, register as a CSISS affiliate.</td>
<td>Links to Portals</td>
</tr>
</tbody>
</table>

GeoDa™

Tobler’s Flow Mapper

CSISS presentations, news, personnel, and sitemap. Our Strategic Plan and Annual Reports are also found here.
Tobler’s Flow Mapper

Background

I claim that geographical movement is of crucial importance. This is because most change in the world is due to movement; the movement of people, ideas, animals, insects, disease, money, energy, or material. One way of depicting and analyzing geographical movement is by way of geographical maps. A convenient and rapid method of displaying movement data on such maps is therefore very useful. A flow mapping program is one approach to this objective. For in depth information see csiss.org/Spatial Tools:

- Flow Mapper program and tutorial; and my web site for several publications on migration.

About Flow Mapper

In 2003 CSISS supported a short effort to produce an interactive flow mapping program. The result is a new Windows-based version of a 1987 program by Waldo Tobler. This original application has been updated by David Jones using Microsoft Visual Basic.Net and Scaleable Vector Graphics for map rendering. It requires as input locational coordinates and information on the interaction between the places. Additional input may include place names and a file of boundary coordinates (for a background map). The user has several menu options for producing a map. The program allows for the production of a total movement maps shown by volume-scaled bands, net movement given by scaled arrows, or simultaneous two-way moves.
Some nice properties of the program

• Simple and quick flow map preparation - GIS Not Needed!
• Extensive color styles available. Black & white too.
• Hovering over a band or arrow gives the magnitude.
• Hovering over a centroid gives its label.
• Two-way, total, or net movement maps.
• Many to many, one to many, or many to one maps.
• Easy threshold choice. Some statistics made available.
• Size dependant only on memory availability.
• Multiple output formats.
• Non-geographic flows within firms, industries, organizations, too.
• Help file included.
• Microsoft Windows compatible.
A sequence of migration maps
Based on US Census Bureau data
The dust bowl time

1935-1940
Net Migration
Only one year of data
In the other cases there are five years of data
Florida shows up
What is going on in the middle of the country?
A sequence of migration maps

1965-1970
Net Migration

1975-1980
Net Migration

1985-1990
Net Migration

1995-2000
Net Migration
Second Visualization

Comparison of spatial movement patterns

Movement of different groups.

Movement at different times.

Maps produced using a mathematical migration model.
Comparison of non-white and white net migration
1935-1940 five year census migration data, by state

Non White migration

White migration
Comparison of net migration patterns at an interval of thirty years.

Migration in the Western United States by State Economic Areas

The effect of resolution on migration data.

Social data are often made available in a hierarchy of administrative units. Moving through the hierarchy changes the resolution and this acts as a spatial filter. This is shown by migration vector fields at several levels of resolution for Switzerland.

41,293 sq. km. Average resolution = sq. root (Area / Units)

3.6 km resolution (3090 Gemeinde)
14.7 km resolution (184 Bezirke)
39.2 km resolution (26 Kantone)

Maps by Guido Dorigo, University of Zürich, based on a program by Waldo Tobler
Migration “Turbulence” in the Alps. 3090 units - 3.6 km resolution
Less of the Fine Detail. 184 units - 14.7 km resolution
The Broad Pattern Only 26 units - 39.2 km resolution
Changing the resolution has the effect of a spatial filter.
Three levels of administrative units and three levels of migration resolution all at once.

Notice that the resolution is not uniform throughout the country.


<table>
<thead>
<tr>
<th></th>
<th>Communities</th>
<th>Districts</th>
<th>Cantons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Units</td>
<td>3090</td>
<td>184</td>
<td>29</td>
</tr>
<tr>
<td>Average Area</td>
<td>3.6 km</td>
<td>14.7 km</td>
<td>39.2 km</td>
</tr>
</tbody>
</table>
County Units
3100+ units
If you got this kind of resolution in photographic film you would reject it, wouldn’t you?

U.S. Counties

Average resolution ~55 km. Patterns >110 km detectable.
Still not sufficient to see movement within cities.
Fourth Visualization

Displaying the results of a migration model.

The model is given as a pair of partial differential equations for a uniform fine mesh covering the region, namely

\[
\frac{\partial^2 R}{\partial u^2} + \frac{\partial^2 R}{\partial v^2} = I(u, v) - 4T(u, v),
\]
\[
\frac{\partial^2 E}{\partial u^2} + \frac{\partial^2 E}{\partial v^2} = O(u, v) - 4T(u, v),
\]

assuming that the \( E \) (‘pulls’ 😊) and \( R \) (‘pushes’ 😞) are differentiable spatial functions and that In-movement, Out-movement, and Turnover are continuous densities given as functions of the Cartesian coordinates \( u \) and \( v \).

This is the model used for the preceding slides.
In this continuous two dimensional migration model we have a coupled system of two simultaneous partial differential equations

These equations can be combined to yield either gross movements or net movements.

For the simultaneous movement in both directions at each pair of places **add** the two equations to get the ‘turnover’ potential, T.

One of the “Laws of Migration” is that in and out movements are nearly the same. Ravenstein 1885.

The result is Helmholtz’s equation.
For the net movement we need only the difference between the ‘in’ and ‘out’ at each node for the ‘attractivity’ potential, as follows:

By subtraction from the previous equations, we have

\[ \frac{\partial^2 A}{\partial u^2} + \frac{\partial^2 A}{\partial v^2} = I(u,v) - O(u,v), \]

where \( A = E - R \) (‘pull’ 😊 minus ‘push’ 😞) can be thought of as the attractivity of each location.

This is the well-known Poisson equation for which numerical solutions are easily obtained.

The right hand side is the amount of change.

Once \( A(u,v) \) - the potential - has been found from this equation, the net movement pattern is given by the vector field,

\[ \mathbf{V} = \text{grad} \ A, \]

or by the difference in potential between each pair of mesh nodes.
Solving the Poisson equation for the potential gives

**The Pressure to Move in the US**

Based on the continuous spatial quadratic model

Movement is from high to low; using state data
An alternate view of the same result

The migration potentials shown as contours
and with gradient vectors connected to give streaklines
Another 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.

For Europe use the Euro coins.
Dollar Bill
(Federal Reserve Note)

Issued by the 8th (St. Louis) Federal Reserve District.
(H is the 8th letter of the alphabet)
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>From:</th>
<th>Boston</th>
<th>New York</th>
<th>Philadelphia</th>
<th>Cleveland</th>
<th>Richmond</th>
<th>Atlanta</th>
<th>Chicago</th>
<th>St. Louis</th>
<th>Minneapolis</th>
<th>Kansas City</th>
<th>Dallas</th>
<th>San Francisco</th>
</tr>
</thead>
<tbody>
<tr>
<td>To:</td>
<td>B</td>
<td>NY</td>
<td>P</td>
<td>Cl</td>
<td>R</td>
<td>A</td>
<td>Ch</td>
<td>SL</td>
<td>M</td>
<td>K</td>
<td>D</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td>2040</td>
<td>289</td>
<td>47</td>
<td>52</td>
<td>137</td>
<td>118</td>
<td>90</td>
<td>16</td>
<td>15</td>
<td>13</td>
<td>13</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>602</td>
<td>1980</td>
<td>231</td>
<td>209</td>
<td>388</td>
<td>307</td>
<td>286</td>
<td>15</td>
<td>48</td>
<td>26</td>
<td>18</td>
<td>261</td>
</tr>
<tr>
<td></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></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></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></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></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></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></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></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></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></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 the continuous version of the gravity model. The result is the system of partial differential equations and 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
Solving the 2000 equations yields the potential
Shown here by contours and gradients.

The raster is indicated by the tick marks. The arrows are the gradients to the potentials. The earlier streakline map is obtained by connecting the gradient vectors.
Fifth Visualization

Biproportional Adjustment

Start with a square array, such as a migration table. Such a table is asymmetric and the marginal values (row and column sums) are unequal. They can be rendered equal by an iteration (IPFP).

This has the effect of removing size differences.

I have always wondered what the IPFP does to the data. Haven’t you?

My first example is based on migration of employed Frenchmen 1962-1968 (INSEE).

   Rouget, 1972, “Graph Theory and Hierarchisation Models” Regional and Urban Economics, 2,3 :263-296
Paris dominates before the adjustment
What biproportionalizing has done to the data

Adjustment to equal marginal values.
Another Example
US Migration before adjustment.

Above 50%
1995-2000 Migration
When the array is modified by the Iterative Proportional Fitting Procedure the smaller values also appear.

Adjustment to equal marginal values.
Sixth Visualization

Coalescent Cities

A satellite image of adjacent cities, providing visual evidence for urban potential fields, illustrates the dynamics of urban growth.

Appolo VI Hasselblat photograph AS6-2-1462.
Tobler, *Mappemond*, 1/91, pp.46-47
Forth Worth and Dallas
Growing towards each other.
For comparison a schematic diagram with a portion of the expected stream function if Fort Worth and Dallas induce a potential field.

A more recent image of the area.
And a map of the area.
Looking for more examples

Potential examples - some California speculations

San Luis Obispo - Santa Maria
Ventura - Oxnard - Simi Valley
Sacramento - Stockton - Modesto
Los Angeles - Orange County - San Diego
Los Angeles to San Diego
More possibilities?

Seattle-Tacoma
Millwaukee - Chicago
Washington D.C. - Baltimore

Examples in Asia?
South America, Europe, Africa?

Topics for student papers?
Seventh Visualization

Multidimensional scaling
(a.k.a. MDS)

The problem is to find relative locations when given dissimilarities taken to be distances.

That is, given $D_{ij}$ find $X$ and $Y$ when $D_{ij} = [(X_i - X_j)^2 + (Y_i - Y_j)^2]^{1/2}$

In high school one learns to compute distances from coordinates. Why does one not learn the reverse, compute coordinates from distances?

One mathematical procedure is as follows:

1: Guess

Much of the Angst regarding mathematics is because guessing is not taught as a legitimate mathematical procedure.

2: Systematically improve the guess by taking into account the given information.

This can be visualized in the following example.
A Classic Example
Use the road distances to make a map showing the locations.

<table>
<thead>
<tr>
<th></th>
<th>ATL</th>
<th>BOS</th>
<th>DAL</th>
<th>DEN</th>
<th>LAX</th>
<th>MIA</th>
<th>MIN</th>
<th>SFO</th>
<th>SEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>1100</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAL</td>
<td>810</td>
<td>-</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEN</td>
<td>1440</td>
<td>1990</td>
<td>800</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAX</td>
<td>-</td>
<td>-</td>
<td>1430</td>
<td>1170</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>680</td>
<td>-</td>
<td>1330</td>
<td>2090</td>
<td>-</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIN</td>
<td>1130</td>
<td>1400</td>
<td>-</td>
<td>870</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFO</td>
<td>-</td>
<td>-</td>
<td>1770</td>
<td>1270</td>
<td>400</td>
<td>-</td>
<td>2000</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SEA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1370</td>
<td>-</td>
<td>-</td>
<td>1640</td>
<td>820</td>
<td>0</td>
</tr>
</tbody>
</table>

ATL  BOS  DAL  DEN  LAX  MIA  MIN  SFO  SEA
Step 1: Guess at locations.
Step 2: Insert lines to represent distances.
Step 3: Insert estimated distances. Some will be too short, some too long, some missing.
Step 4: Examine and average discrepancies.
Step 5: Apply adjustments.
Step 6: Insert adjusted distances and repeat.
The complete sequence of iterative steps.
For the USA data

The first step is to systematically array the data in a table; only half of the table is needed, since it’s symmetric, and places with only one connection can be ignored:

<table>
<thead>
<tr>
<th></th>
<th>ATL</th>
<th>BOS</th>
<th>DAL</th>
<th>DEN</th>
<th>LAX</th>
<th>MIA</th>
<th>MIN</th>
<th>SFO</th>
<th>SEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>1100</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAL</td>
<td>810</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEN</td>
<td>1440</td>
<td>1990</td>
<td>800</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAX</td>
<td></td>
<td>1430</td>
<td>1170</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>680</td>
<td>1330</td>
<td>2090</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIN</td>
<td>1130</td>
<td>1400</td>
<td></td>
<td>870</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFO</td>
<td></td>
<td>1770</td>
<td>1270</td>
<td>400</td>
<td>2000</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEA</td>
<td></td>
<td></td>
<td>1370</td>
<td></td>
<td>1640</td>
<td>820</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only twenty of the possible thirty six distances are given. As initial coordinates one can use:

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2600</td>
</tr>
<tr>
<td>2</td>
<td>3300</td>
</tr>
<tr>
<td>3</td>
<td>1800</td>
</tr>
<tr>
<td>4</td>
<td>1050</td>
</tr>
<tr>
<td>5</td>
<td>300</td>
</tr>
<tr>
<td>6</td>
<td>3100</td>
</tr>
<tr>
<td>7</td>
<td>1900</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>200</td>
</tr>
</tbody>
</table>

obtained by placing the cities in their approximate location on millimeter paper.
The answer, with estimated coordinates and distances.

After several hundred iterations the root mean square error is 44.4 units (a 99.1% fit to the data) and the solution coordinates are:

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2538</td>
<td>700</td>
</tr>
<tr>
<td>2</td>
<td>3168</td>
<td>1588</td>
</tr>
<tr>
<td>3</td>
<td>1727</td>
<td>473</td>
</tr>
<tr>
<td>4</td>
<td>1211</td>
<td>1073</td>
</tr>
<tr>
<td>5</td>
<td>242</td>
<td>497</td>
</tr>
<tr>
<td>6</td>
<td>3066</td>
<td>233</td>
</tr>
<tr>
<td>7</td>
<td>1794</td>
<td>1637</td>
</tr>
<tr>
<td>8</td>
<td>-19</td>
<td>837</td>
</tr>
<tr>
<td>9</td>
<td>79</td>
<td>1689</td>
</tr>
</tbody>
</table>

Resulting distances are, to the nearest mile,

<table>
<thead>
<tr>
<th></th>
<th>ATL</th>
<th>BOS</th>
<th>DAL</th>
<th>DEN</th>
<th>LAX</th>
<th>MIA</th>
<th>MIN</th>
<th>SFO</th>
<th>SEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>1089</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAL</td>
<td>842</td>
<td>1822</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEN</td>
<td>1378</td>
<td>2023</td>
<td>791</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAX</td>
<td>2304</td>
<td>3122</td>
<td>1485</td>
<td>1127</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>705</td>
<td>1359</td>
<td>1361</td>
<td>2036</td>
<td>2836</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIN</td>
<td>1196</td>
<td>1375</td>
<td>1166</td>
<td>811</td>
<td>1925</td>
<td>1895</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFO</td>
<td>2560</td>
<td>3274</td>
<td>1783</td>
<td>1252</td>
<td>429</td>
<td>3143</td>
<td>1981</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SEA</td>
<td>2650</td>
<td>3090</td>
<td>2048</td>
<td>1289</td>
<td>1202</td>
<td>3323</td>
<td>1715</td>
<td>857</td>
<td>0</td>
</tr>
</tbody>
</table>
The solution
Showing (exaggerated) standard errors.
USA Highway Distance Map
Here is a more elaborate example.

Using a road atlas the student took many values from the table of distances between places. These tables are common in such atlases. Using these distances he then computed the location of the places. The US outline and latitude longitude grid were then interpolated to complete the map.

As for all maps, Tissot’s indicatrix can be used to measure the linear, angular, and areal distortion in this case introduced by the roads.
Road Distance Map of the United States
Student drawing
Another example

Cappadocian Speculation

Given hundreds of cuneiform tablets from 1700 B.C. Use the ones that mention places in Cappadocia (Central Turkey).

To find the location of the places.

Cappadocian Cuneiform Tablet

Some giving trade between places
About 5 by 8 cm
Some tablets include the names of towns and occasionally indications of trade.

The frequency count of town names can be taken to be proportional to their size or population.

When two town names occur on one tablet this implies an interaction between them.

So we have $P_i$ and $P_j$ and $T_{ij}$ in this case relating 62 towns.
Use the gravity model

\[ T_{ij} = k \frac{P_i P_j}{D_{ij}} \]

\( T = \text{trade}, \ P = \text{population}, \ D = \text{distance} \)

An exponent can also be used, with obvious modification. Other model variants could also be tried.

Invert the model to get

\[ D_{ij} = k \frac{P_i P_j}{T_{ij}} \]

Given distances compute the locations.

That is, find the latitudes and longitudes using the (rather sparse) symmetric 62 by 62 table of distances and the very few known locations.
Predicted locations

For standard errors see the published text
Another example

Using an input/output table, consider two industries to be “close” if the quantity of exchanges between them is large. Construct a diagram in which “close” industries are placed near each other, as follows.

Place the 1\textsuperscript{st} industry at random, then the next “closest” on a circle of chosen radius, then the 3\textsuperscript{rd} where circles about the first two intersect. The 4\textsuperscript{th} industry goes where more circles intersect. Obviously this gets very tedious and can’t go on. So use the iterative procedure to place “close” industries near each other.
How can one comprehend such a large table without long and tedious study?
Here is one method: the spatial result of a non-spatial example. A MDS map of a 50 by 50 Inter-Industry Input - Output Table. 2450 possible interactions.

New England I/O table
Isard 1960 Table 3

Would applying this to the Leontiev inverse be of interest?

For more examples see www.geog.ucsb.edu/~tobler/publications/cartography/survey-adjustment
From the Inter-Industry Input - Output table the MDS map shows that **15 of the 50 industries dominate** among 2450 possible interactions.
Iterative procedures, such as MDS, lend themselves particularly well to animation.

In the present instance two modes of animation are possible.

First, the rate of convergence of the discrepancy between the original guess and the current state can be shown. This should tend to a minimum function.

Secondly, and more interestingly, the simultaneous movement of the several objects between locations as they move towards equilibrium can make an amusing dynamic display.
Eighth Visualization

The **transform - solve - invert paradigm**

This is a classic way of solving problems.

Change to a more appropriate coordinate system where the problem becomes simpler.

Solve the problem then revert to the original coordinates.

A map projection can be used in this manner.

Think of map projections as different ways of producing different types of graph paper for spheres.
The conventional satellite tracking chart

The meridians and parallels are straight.
The satellite tracks are curves.
Bend the meridians and parallels so that the satellite tracks are straight. It is then easier to track the satellite.
Mercator’s projection is a famous anamorphose. It is designed to solve a navigation problem. The seaman plots the course as a straight line on the distorted Mercator map and follows the indicated directions. This same course is a curve on the earth but it appears as a straight line on the map, and this is much easier to draw. Other anamorphoses (map projections) can be used to solve other problems.
Another map projection example to solve another problem in the transform-solve-invert paradigm.

The next schematic illustrates the U.S. population density in perspective.

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

There follows a map projection (cartogram) that may be useful for this purpose.
Map Projections and Cartograms

The equal area condition for a map projection in spherical and plane rectangular coordinates is

$$\frac{\partial x}{\partial \lambda} \frac{\partial y}{\partial \phi} - \frac{\partial x}{\partial \phi} \frac{\partial y}{\partial \lambda} = R^2 \cos(\phi)$$

The condition equation for an areal cartogram is

$$\frac{\partial x}{\partial \lambda} \frac{\partial y}{\partial \phi} - \frac{\partial x}{\partial \phi} \frac{\partial y}{\partial \lambda} = R^2 D(\phi, \lambda) \cos(\phi)$$

Where $D(\phi, \lambda)$ is the density distribution on the earth, considered spherical.

Clearly, when the density distribution is constant, then the cartogram becomes an equal area map projection.

In both cases the one condition does not suffice to yield the two equations $x = f(\phi, \lambda), y = g(\phi, \lambda)$ needed to completely define a map projection. The obvious second condition is to require that the angular distortion be minimized.
U.S. Population Density
by one degree latitude/longitude quadrangles
The Lat/Lon Grid in the Two Spaces
Left, the usual grid. Right, transformed according to population.
U.S. Map in the Two Spaces
Left, the usual map. Right, the transform.
Map on the right adjusted to equalize the population.
The Inversion
On the right: Uniform hexagons in the transformed space.
On the left: The solution.
The inverse transform partitions the U.S. into cells of approximately equal population.

Ninth Visualization

Map Matching

This example is from the field known as ‘Mental Mapping’

A list of the sixty largest US cities, in alphabetical order, is given to students
# Cities and Locations

Coordinates not given to students.

<table>
<thead>
<tr>
<th></th>
<th>City</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AKRON</td>
<td>41.066</td>
<td>-81.516</td>
</tr>
<tr>
<td>2</td>
<td>ALBUQUERQUE</td>
<td>35.083</td>
<td>-106.633</td>
</tr>
<tr>
<td>3</td>
<td>ATLANTA</td>
<td>35.749</td>
<td>-84.383</td>
</tr>
<tr>
<td>4</td>
<td>AUSTIN</td>
<td>30.299</td>
<td>-97.783</td>
</tr>
<tr>
<td>5</td>
<td>BALTIMORE</td>
<td>39.299</td>
<td>-76.633</td>
</tr>
<tr>
<td>6</td>
<td>BIRMINGHAM</td>
<td>33.499</td>
<td>-86.916</td>
</tr>
<tr>
<td>7</td>
<td>BOSTON</td>
<td>42.333</td>
<td>-71.083</td>
</tr>
<tr>
<td>8</td>
<td>BUFFALO</td>
<td>42.866</td>
<td>-78.916</td>
</tr>
<tr>
<td>9</td>
<td>CHARLOTTE</td>
<td>35.049</td>
<td>-80.833</td>
</tr>
<tr>
<td>10</td>
<td>CHICAGO</td>
<td>41.833</td>
<td>-87.749</td>
</tr>
<tr>
<td>11</td>
<td>CINCINNATI</td>
<td>39.166</td>
<td>-84.500</td>
</tr>
<tr>
<td>12</td>
<td>CLEVELAND</td>
<td>41.499</td>
<td>-81.683</td>
</tr>
</tbody>
</table>
Instructions to the Students

Work without any reference materials.

Use Graph Paper, wide Margin at top.

Plot Cities with ID Number on the Graph Paper.

USA Outline may be drawn, but is not required.
An Anonymous Student’s Map
A very good student
Analysis of Student Data

Displacement vectors

Interpolated vectors

Now Tissot’s indicatrix can be found and, by integration, a potential calculated.

Displaced grid
The Student Map
Shows displacement vectors

These vectors could also show change of address coordinates, due to a move, or commuting.

Or they could be home to shopping moves, etc.

Or they could be material moved or information and/or messages sent between places.

Thus there are many possible interpretations of this kind of vector displacement.

And there are methods of estimating the amount of strain and warping implied and calculation of an implicit potential field.
With student maps in hand

How to score?
Compute correlation, \( R^2 \), between actual and student estimates? How to do this?
Do you know how to compute correlation between fields of vectors?
Correlation between scores of different students? Factor analyze?
Compute vector field variance, etc., to determine degree of fuzziness?
Average vectors over all students?
I’m not going to answer these questions here, but
Is There a Method of Computing the Correlation Between Vector Fields?

The question comes up not only in meteorology and oceanography but also for the comparison of the student’s maps, for comparison of old maps, and in many other situations. There are in fact such correlation methods, and associated with these are regression-like predictors. One of these is my bidimensional regression. Statistical significance tests are also available. For other approaches see, for example B. Hanson, et al, 1992, “Vector Correlation”, *Annals*, AAG, 82(1):103-116.
To illustrate the scoring concept for students I have built

The Map Machine
The Map Machine
Detail View 1

Showing the one to one correspondence between the images
The Map Machine

Detail View 2

The front panel is transparent, back panel is white, strings are black.
One set of locations is ‘correct’, the other is the estimate.
The Map Machine
Detail View 3
Releasing the back panel and pulling the strings together
The Map Machine

The Final View

Corresponds to the computer image of displacements
Connecting the estimate to the ‘correct’ locations
My Bidimensional Regression Computer Program

Now available for download online

Differs from ordinary regression in that both the independent and the dependent variable each have two components. One obtains the same effect as the map machine by estimating the difference between original and image components. The original can be a modern map and the image an old map. Or depiction of many kinds movements. The difference between pairs of locations is used to compute displacements. An interpolation can then performed to stretch one space to fit the other.

This is also known as ‘rubber sheeting’

The stretch can then examined mathematically, and the distortion calculated.
Vectors Appear in Map Matching.
Here is an example
Map and Image
The difference between the map and the image shown as discrete vectors.
Difference Vectors by themselves, without the grid. These can be interpolated to yield a vector field.
Interpolated Vector Field
Great Lakes Displaced

The grid has been ‘pushed’ by the interpolated vector field
Benincasa Portolan Chart 1482
Comparing positions on an old map with actual ones yields vectors.
Mediterranean Chart Displacements
After Loomer
Interpolated Vector Field
Based on Mediterranean displacements from Loomer’s thesis
Computed and drawn at half scale using the Bidimensional Regression computer program
Warped Grid of the Portolan Chart
As ‘pushed’ by the interpolated vector field
Computed and drawn using the Bidimensional Regression computer program
Distortion of chart contoured.

Using the sum of squares of the partial derivatives of the displacement vectors from the image.
Computed and drawn by the Bidimensional Regression program.
Tissot’s measure of distortion for the chart.
Small circles in the original become ellipses in the transform.

Obtained from the displacement vectors and illustrating distance, angle, and area distortion at each location.
Calculated and drawn using the Bidimensional Regression program.
Tenth Visualization
Mass Preserving Reallocation using Areal Data.

Given observations for bounded areas, for example within census tracts or school districts, that is, not at point locations or centroids, produce a quasi-continuous field on a raster for contouring, or for data conversion between these areal units.

Both of these uses are considered important for regional studies. For example, area boundaries change frequently, or zones do not coincide, requiring interpolation.

Begin by covering the polygons by a raster. Then smooth the data transition between the polygonal units but keep the individual volumes intact.

Mathematically this requires the constrained minimum of an integral.

In other words, the objective requires that the total content within each region must remain its value, or

$$\int\int f(x, y) \, dx \, dy = V_i$$

for each region $i$.

The smoothness requirement is given by the LaPlace equation

$$\nabla^2 x + \nabla^2 y = 0.$$

This says that the neighboring locations have similar values - or, in a raster, that the central value is the average of those surrounding it, or that the partial derivatives change slowly.

Recognize that this is a hypothesis about the phenomena under study. But it immediately yields a computational algorithm.
What the Mathematics Means

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 modeling spatula is used, but no clay is allowed to move from one unit into another.

Color mixing is not allowed.
The smoothing is done using an iterative process.

The first step is to “rasterize” the region.

Then the smoothing is done on this raster, all the while maintaining the “population”.

The number of iteration steps depends on the size of the largest region, in raster units.

That is because the smoothing must cross from edge to edge of the largest region. The finer the raster, the higher the resolution, and the longer the iteration time.
Zero Iterations
Rasterized polygons of different content
Five Iterations
Ten Iterations
Fifteen Iterations
Twenty Iterations
From zero to 60 iterations

The final result is quite smooth
Colored Clay Before Smoothing
Colored Clay After Smoothing

Convert to an alternate set of zones by imposing new boundaries (as if using a cookie cutter), then add up the values in the new polygons.
A piecewise continuous surface
Population Density in Kansas by County

Each county still contains the same number of people

A smooth continuous surface, with population pycnophylactically redistributed
Two types of isopleth interpolation compared.
Using data from Kansas counties.

Pycnophylactic reallocation and punctual Kriging from centroids.

The pycnophylactic method does not use centroids, and preserves the county totals.

Figure 8.13 (p. 150) of T. Slocum, “Thematic Cartography and Visualization”, Prentice Hall, 1999.
Another Example
Ann Arbor population by census tract.
Left: Choropleth map.
Right: Bivariate histogram.
Population reallocated pycnophylactically
Left: zero iterations
Right: 200 iterations
Pycnopylactic reallocation yields a smooth field from which one can make a

Left: Contour map
Right: Gradient field
Contours with gradient field and Orthogonals to the contours. Lines of expected growth (Borchert 1961)?
An Image Processing Example
A 20 by 14 Image
Quadrupled from 20 by 14 to 80 by 56
but with the same total “mass”
within each 4 by 4 region
Many types of data come in the form of square tables.

Often these tables are not symmetric.

Movement and interaction tables are geographic examples.

It is useful to take advantage of the asymmetry.
An example of a non-symmetric table.

**Citations among psychology journals**

*Coombs et al 1970*

*Data from 1964*

<table>
<thead>
<tr>
<th></th>
<th><em>AJP</em></th>
<th><em>JASP</em></th>
<th><em>JAP</em></th>
<th><em>JCPP</em></th>
<th><em>JCP</em></th>
<th><em>JEdP</em></th>
<th><em>JExP</em></th>
<th><em>Pka</em></th>
<th><em>Total</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>American Journal of Psychology</strong></td>
<td>119</td>
<td>8</td>
<td>4</td>
<td>21</td>
<td>0</td>
<td>1</td>
<td>85</td>
<td>2</td>
<td>240</td>
</tr>
<tr>
<td><strong>Journal of Abnormal and Social Psychology</strong></td>
<td>32</td>
<td>510</td>
<td>16</td>
<td>11</td>
<td>73</td>
<td>9</td>
<td>119</td>
<td>4</td>
<td>774</td>
</tr>
<tr>
<td><strong>Journal of Applied Psychology</strong></td>
<td>2</td>
<td>8</td>
<td>84</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>16</td>
<td>10</td>
<td>136</td>
</tr>
<tr>
<td><strong>Journal of Comparative and Physiological Psychology</strong></td>
<td>35</td>
<td>8</td>
<td>0</td>
<td>533</td>
<td>0</td>
<td>1</td>
<td>126</td>
<td>1</td>
<td>704</td>
</tr>
<tr>
<td><strong>Journal of Consulting Psychology</strong></td>
<td>6</td>
<td>116</td>
<td>11</td>
<td>1</td>
<td>225</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>385</td>
</tr>
<tr>
<td><strong>Journal of Educational Psychology</strong></td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>52</td>
<td>27</td>
<td>5</td>
<td>107</td>
</tr>
<tr>
<td><strong>Journal of Experimental Psychology</strong></td>
<td>125</td>
<td>19</td>
<td>6</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>586</td>
<td>15</td>
<td>821</td>
</tr>
<tr>
<td><strong>Psychometrika</strong></td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>13</td>
<td>58</td>
<td>98</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>325</td>
<td>683</td>
<td>133</td>
<td>637</td>
<td>321</td>
<td>80</td>
<td>984</td>
<td>102</td>
<td>3,265</td>
</tr>
</tbody>
</table>
Let $M_{ij}$ represent a ‘movement’ table with $i$ rows and $j$ columns. It can be separated into two parts, as follows.

\[ M_{ij} = M^+ + M^- \]

where

\[ M^+ = (M_{ij} + M_{ji})/2 \quad \text{symmetric} \]
\[ M^- = (M_{ij} - M_{ji})/2 \quad \text{skew symmetric} \]

The variance can be computed for each component,
How the two parts are used

I consider the symmetric component as a type of background. The real interest is in the asymmetric part.

In the journal case the position of the places is not known.

Since locations are not given the symmetric part may be used to make an estimate of these positions.

This estimate is made using an ordination, trilateration, or multidimensional scaling algorithm.
The attempt is now made to apply these ideas in a social space.

This can be considered a development of Lewin’s *Topological Psychology* or his *Field Theory in the Social Sciences*.

The data represent citations between a small set of psychological journals. Larger citation tables are now also available.
In Journal Space
The observed two-way flow between the journals

Journal Flow Map
Coombs et al 1970
<table>
<thead>
<tr>
<th>From</th>
<th>Journal to Journal Citations</th>
<th>Net</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>To</td>
<td></td>
</tr>
<tr>
<td>AJP</td>
<td>119 8 4 21 0 1 85 2</td>
<td>125 910</td>
</tr>
<tr>
<td>JASP</td>
<td>32 510 16 11 73 9 19 4</td>
<td>-1382 -644</td>
</tr>
<tr>
<td>JAP</td>
<td>2 8 84 1 7 8 16 10</td>
<td>-261 -237</td>
</tr>
<tr>
<td>JCPP</td>
<td>35 8 0 533 0 1 126 1</td>
<td>1302 366</td>
</tr>
<tr>
<td>JCP</td>
<td>6 116 11 1225 7 12 7</td>
<td>-924 -2</td>
</tr>
<tr>
<td>JEdP</td>
<td>4 9 7 0 3 52 27 5</td>
<td>-180 324</td>
</tr>
<tr>
<td>JexP</td>
<td>125 19 6 70 0 0 586 15</td>
<td>904 -924</td>
</tr>
<tr>
<td>Pka</td>
<td>2 5 5 0 13 2 13 58</td>
<td>416 207</td>
</tr>
</tbody>
</table>

AJP  Am J of Psychology
JASP J of Abnormal & Social Psychology
JAP  J of Applied Psychology
JCPP J of Comparative & Physiological Psychology
JCP J of Consulting Psychology
JEdP J of Educational Psychology
JexP J of Experimental Psychology
Pka  Psychometrika

The table gives the being-cited journal across the columns. But the information can be considered to move from that journal to the citing journal.

Therefore the transpose is used to produce the source to sink map.
Adding across the table, the column marginals give the outsums (a.k.a. outdegree). Summing down the rows gives the insums (a.k.a indegree).

The ‘sending’ places (rows) are known as ‘sources’, and are shown by negative signs. The ‘receiving’ places (columns) are the ‘sinks’, and are shown by plus signs. The size of the symbol represents the magnitude of the movement volume.
The **net** flow between the journals

Journal Flow Map
Coombs et al data
We now have an assignment problem. How to get 163 citations from JExP, 85 from AJP, & 4 from Pka to the 5 receiving journals, using only the marginals. There are obviously many possibilities

One solution is to use the “Transportation Problem” (Koopmans, Kantorovich, ~1949): Minimize $M_{ij}d_{ij}$, subject to $M_{ij} = O_i$, $M_{ij} = I_j$, $M_{ij} \geq 0$, given the distances computed from the coordinates and using the simplex method for the solution.

A more realistic solution is given by the quadratic transportation problem: Minimize $M^2_{ij}d_{ij}$, subject to the same constraints.

Both of these solutions result in discrete answers, and ‘shadow prices’. I am looking for a spatially continuous solution that allows vectors and streamlines, in order to determine spatial flow fields and a continuous potential.
The next step is to compute the displacements between the cited journals.

This is based on the asymmetry of the citations table.

The fundamental idea being that there exists a ‘wind’ making movement easier in some directions.

The mathematical details are given in a published paper.

The movement from source to sink can be computed to show the direction and magnitude of the movement.

The computation is based on the asymmetry of the movement table.

Small directed vectors represent this movement on the next map.
Displacement between Journal Citations

[Diagram showing various journals such as AJP, JEdP, Pka, JCP, JAP, JASP, JExp, JCPP, and the relationships between them through arrows.]
An interpolation is then performed to obtain a vector field from the isolated individual vectors.

This is done to simplify the mathematical integration needed to obtain the forcing function.
The computed potential should have the vector field as its gradient.

This is a hypothesis that can be tested.

The base level of the potential is determined only up to a constant of integration.

The vector field, to be a gradient field, must be curl free. This can also be tested.
Then the potential is computed by integration.

This potential should be such that its gradient coincides with the displacement vectors.

It may be necessary to use an iteration to obtain this result.
Journal Potential Function
Flow and Potential between Psychological Journals
Some questions

Suppose a new psychological journal were started. Where should it be inserted into in this space? Does it make sense to treat journal citations as being located in a continuous two-dimensional social space?

Can other social data be treated in a similar fashion, for example social mobility tables?

And more general network data?
Conclusion

Spatial studies are readily adaptable to graphic visualization.

This is because they involve geometry.

Graphics often clarify concepts that are otherwise difficult to understand and dramatically suggest topics for further study.

This is well known but often under-utilized by analytical scientists.

For more examples go to http://www.geog.ucsb.edu/~tobler
Thank you for your attention

Eleven examples were demonstrated.

1: Spatial and Temporal Autocorrelation.
2: Comparison of movement patterns.
3: Changing Resolution.
4: Migration potential.
5: Biproportional Adjustment.
6: Coalescent cities.
7: Multidimensional Scaling Iterations.
8: Transform-solve-invert
9: Vector Map Matching.
10: Pycnophylactic Reallocation.
11: Potential from Asymmetry

Questions?