

Movement Mapping

Waldo Tobler

Geographical movement is of crucial importance. This is because much change in the world is due to movement; the movement of people, ideas, 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. The flow mapping program is one approach to this objective. The French engineer Charles Minard in the mid 1800's was one of the first to use maps to depict social movement phenomena (see Palsky 1996; also the 'CSISS Classic' on Minard though with less detail).

Computerized maps showing geographical movement were introduced in the 1970s & 1980s (Tobler 1987, with some history; also see Marble et al 1995; Granados 2000). In general there are several ways of rendering geographical movement on maps, assuming data tables of flows, or interaction, between places or on networks. A common technique is to use a choropleth map with shading or colors. But these do not actually show movement. Instead they show a change of state, but not the actual moves. A simpler visualization is to just show change by appropriately scaled plus and minus symbols at the proper locations (for examples see Tobler 2001, slides 27-30). Flow maps, *sensu strictu*, are generally of two types, discrete or continuous, depending on how they show the movements. The discrete map type consists of bands or arrows whose width is proportional to the volume moved. The continuous types of map use vector fields or streamlines to show continuous flow patterns (see Tobler 1981, Dorigo & Tobler 1983, Tobler 2001 slides 13-22, 33-39 et passim, for examples). More recently animated maps are being considered in a dynamic computer environment. This is a genre that has not reached its full potential for movement because of the large volume of moves that take place in geographic space. For example the 2000 US Census migration table indicates that over 21,000,000 people migrated in the 1995 - 2000 time period. In all cases the maps can be used for simple display or they can be used as analytical tools for hypothesis generation, and also provide a rapid method of checking information for errors.

The movement data need not be restricted to geographic information. Citations between scientific journals can be used to illustrate this point. Tables of from-journal to to-journal citations are now available in the form of tables, from rather large samples. Based on this information the locations of the journals relative to each other can be estimated via an ordination. Using these locations the table can be used to produce an information flow map. Similar comments apply to intergenerational occupation tables as occasionally used in sociology. Economic information, such as inter-industry linkages or input/output tables are similar, and the data in these can be depicted using a flow map. There are many such examples.

The CSISS flow mapping program

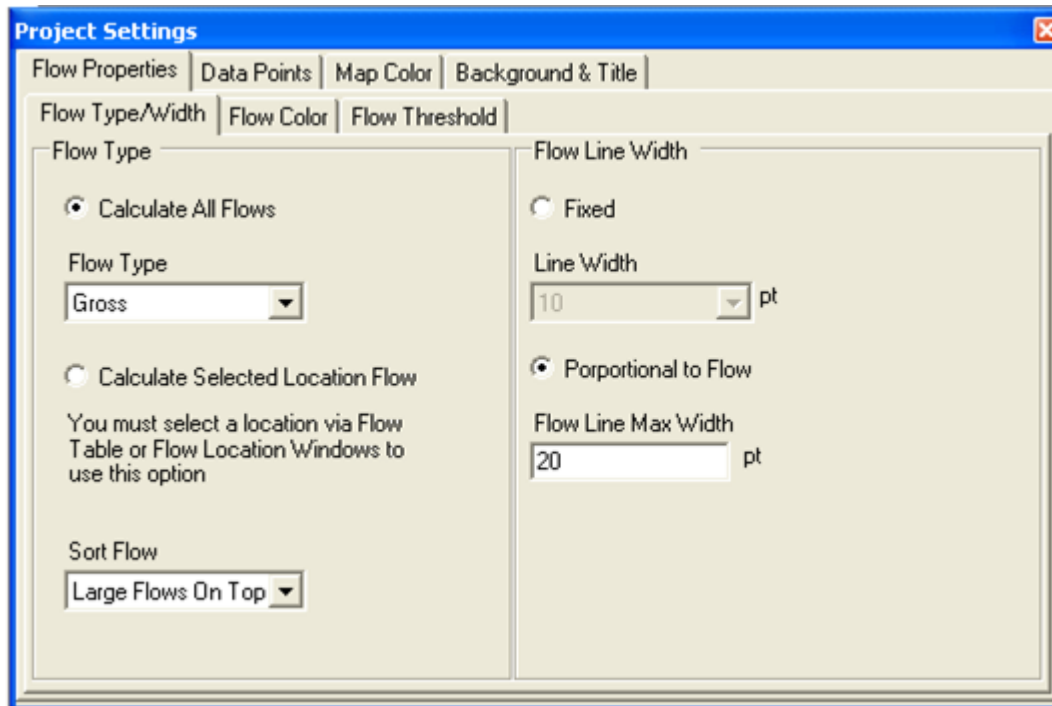
In 2003 the Center for Spatially Integrated Social Science of the Geography Department of the University of California at Santa Barbara supported an effort of somewhat under one-man month to produce an interactive flow mapping program of the discrete display type. This is essentially an updated interactive Microsoft Windows version of the program described in Tobler (1987) with the addition of color. Examples of the result are shown later, and are also available with the program download. The program is written in the Microsoft Visual Basic Net language with SVG for use as a stand-alone program. This program requires as input locational coordinates and a table of interaction between places.

Additional input may include place names for the locations and a file of boundary coordinates. For the making of a map the user has available several options, chosen from menus. Invoking the program then allows for the production of a total movement map shown by volume-scaled bands, or net movement given by scaled arrows, or simultaneous two-way moves. The color of the background map and of the bands or arrows is user selectable. Positioning the cursor on a band or arrow presents the user with the volume of that movement. The point locations can be shown or not shown. If the place names are given then hovering over the place picks up the name. Other options allow the depiction of selected movements, from, or to, one place, or between all places, with a bound on the magnitude of the movement volume to be shown. One can specify a common volume scale for the comparison of one or more tables giving movements - over time or between models. Smaller flows can be drawn on top of larger ones, or larger above smaller (recommended). A distinctive color, or black or white depending on the background, can be used for the edge of the arrows to provide greater clarity. It is expected that most of maps prepared using this program will be made from square arrays with both rows and columns of the interaction table representing the same locations. But rectangular tables, from cities to national parks for example, can also be used. Or from (to) one resolution scale to another; i.e., from all US states (except California) to (from) all California counties. The resulting maps can be saved in several formats for display or for printed copies. The program contains a help file and a power point tutorial is also available. The program may be downloaded at CSISS.org/Tools.

Not in the program

It is anticipated that the user may wish to analyze or modify the input interaction table in various ways, external to the program. For example one would normally wish to compute some statistics from a single array (or several arrays), such as the mean, variance, asymmetry (degree, variance), maximum, minimum, row sum, column sum, trace, or examine a histogram. Or change to a new array, saved in a new file to subsequently be used in the program, by a scalar multiplication, or division, or to delete rows and columns, to aggregate to a smaller array (rows and columns collapsed), convert a rectangular from-to array to a square one; or transpose, reroute, or model the interaction. Or to the classify the flow data into high, medium, and low value classes, or as +/- sigma; or partition the data into above or below the mean, or a specified value, or a percent, all for new flow maps. Given two (or more) tables one could sum or difference them and save this as a new table for mapping, or take a ratio, or model them as a Markov process. There are also many ways that the array(s) can be treated as a network for analysis. It is not anticipated that these types of modifications should be incorporated into the flow mapping program since programs already exist to perform these types of modifications or analyses.

A menu from the flow mapping program:



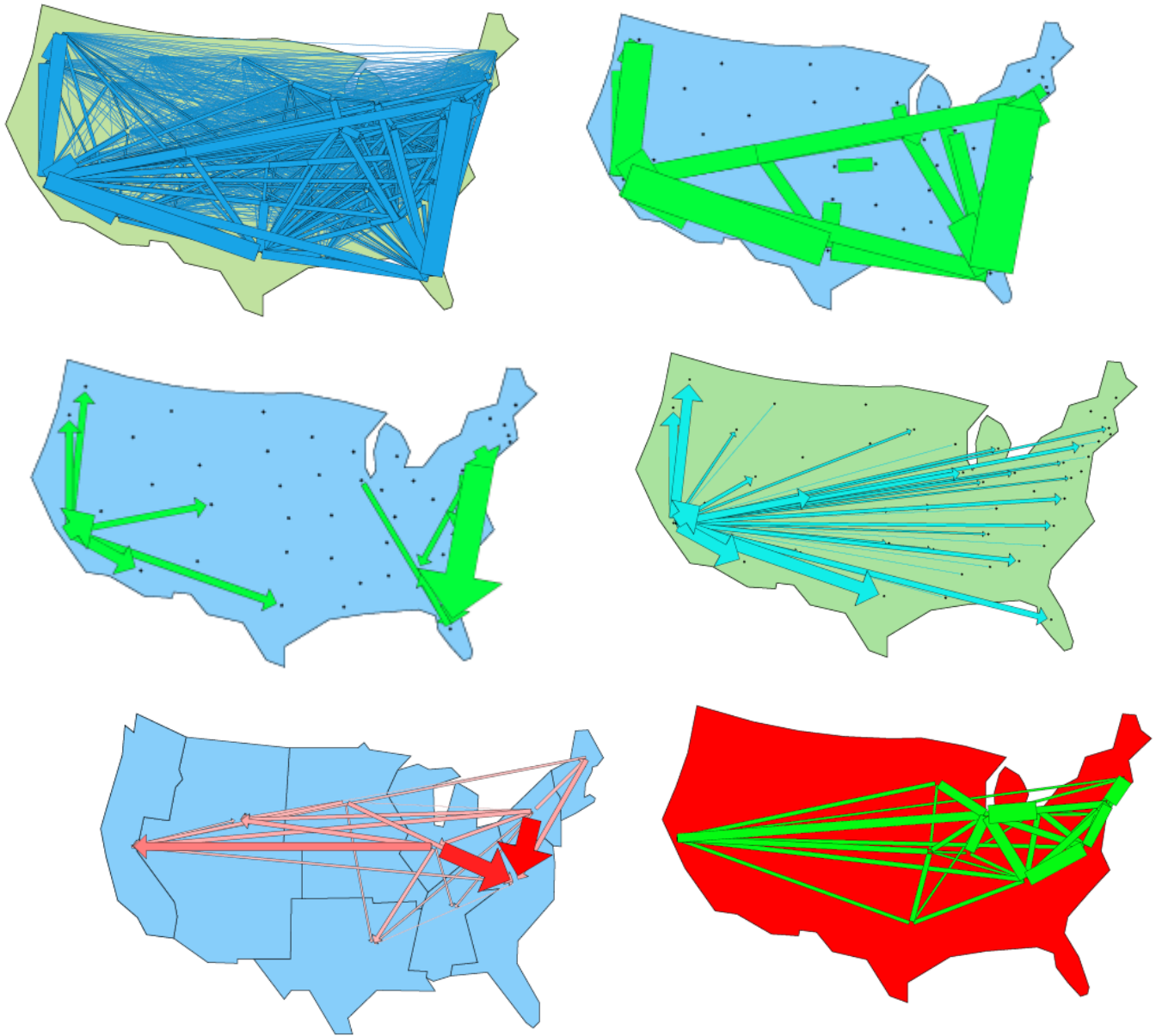
Future Work

Given the modest investment by CSISS in this flow mapping program many options can be considered for the future. Enhancing the program will depend on feedback from users, and financial incentives.

Comments on the attached color maps (it may take a few seconds for the maps to appear on the screen):

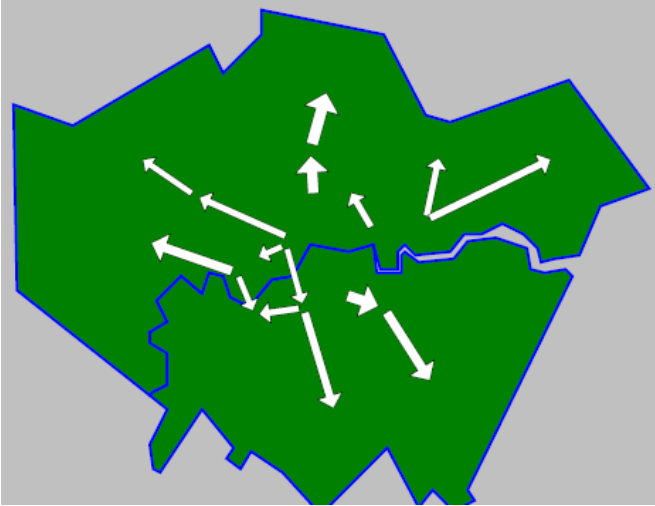
The first four of the colored sample maps use the 2000 US Census Bureau state-to-state tables describing migration for the interval 1995-2000. Alaska, the District of Columbia, and Hawaii are omitted for compatibility with previous studies, as are the diagonal table entries. The 48 by 48 tables yield a possible 2256 arrows. The first map shows the sum of the migrations in both directions between the 48 states for a total of 1128 migration bands (in blue on a light green background), and is very difficult to read because of the many movements shown. The map to the right (with green bands on a blue background with state centroids indicated) uses the same information but shows only the major migrations. The next map, with pointed arrows, shows the net moves during this five-year period – the difference between the incoming and outgoing migrants. The basic movement is from New York to Florida and out of California. Next shown is all of the migration from California to all other states. The next map, including the Census Division boundaries, depicts the net movement of American geographers from division of birth to division of employment (Janelle 1992) and uses a color gradation as well as arrow thickness to indicate the movement volume. The final US map renders visible the movement of US Federal Reserve Notes, 'green backs', between the twelve Federal Reserve Districts in 1976.

The next page includes a map of net moves between 33 London boroughs; everybody seemed to be leaving the town center in the 1965-1966 time period. In Western Washington State the map shows the 1974-1978 movement of cancer patients from their county of residence to the county of treatment. Both a title and islands are included on the map. The final map (bottom) uses a twenty-year journal-to-journal citation table to display journal referencing between 23 scientific fields, after an ordination to produce locational coordinates. The 'Data & Maps' folder included with the program download contains over 50 other examples.

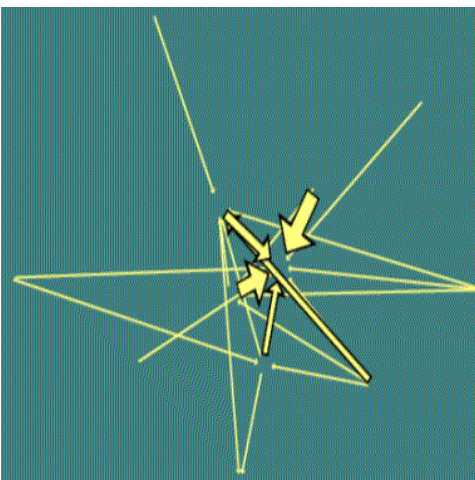
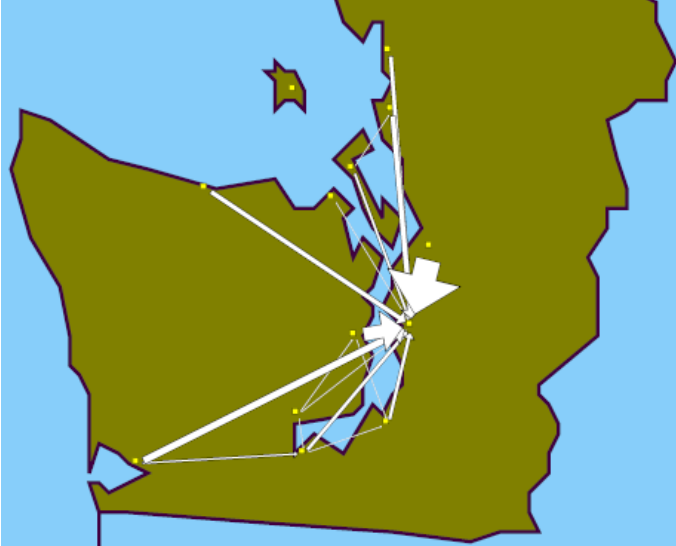


Left to right, top to bottom; 2000 US Census 48 by 48 tables: 1128 total two-way migrations. Same, but parsed to show only major migrations. Major net migrations. From CA migrations. American geographers migrating (Janelle). Dollar bill (green back) moves between 12 FRB regions.

1965-1966 Outmigration



Cancer Patients



Top to bottom: London out-migration; Cancer patients, residence to treatment; Journal-to-journal citations (experimental).

References:

Dorigo, G., W. Tobler, 1983, "Push-Pull Migration Laws", *Annals, Assoc. Am. Geographers*, 73(1): 1-17.

Janelle, D., 1992, "The Peopling of American Geography", pp. 363-390 of R. Abler, M. Marcus, J. Olson, eds, *Geography's Inner Worlds*, Rutgers University Press, New Brunswick.

Granados, M., 2000, "The Use of Analytical Models to Improve the Exploratory Visualization of Interregional Flow Data", Ph.D. Thesis, Geography Department, Ohio State University, Columbus.

Marble, D., Z. Gou, L. Liu, J. Saunders, 1995, "Recent advances in the exploratory analysis of interregional flows in space and time", pp. 75-88 of Z. Kemp, ed., *Innovations in Geographical Information Systems*, Taylor and Francis, London.

Palsky, G., 1996, *Des chiffres et des cartes*, Comité des travaux historiques et scientifiques, Paris

Tobler, W. 1987, "An Experiment in Migration Mapping by Computer", *The American Cartographer*, 14(2): 155-163.

Tobler, W., 2001, "Geographical Movement", power point presentation at <http://www.geog.ucsb.edu/~tobler/presentations>

On Viewing Flow Maps

Waldo Tobler

Abstract:

An essay generated after producing an interactive computer program to depict interaction between geographical areas by means of flow maps. The main emphasis is on the error in the discrete type of such maps, with a few comments on the continuous versions.

Value and problems with maps

Geographical flow maps can bring cohesion from, and clarity to, large tables in a very rapid and simple manner. This is one reason for using these maps. But maps of geographical movement, like most maps, suffer from a property that I might call 'crispness'. They are too neat and hide the fact that all data are, to some extent, incorrect. I am mostly concerned here with the type of map that shows discrete flows between areas of geography. This is only one of the class of movement maps¹. These maps are generated from tables that represent the amount of flow occurring between pairs of places. These kinds of patterns represent an instance of phenomena long studied by geographers under the title of 'spatial interaction', but are also of interest in many other fields. My discussion now centers around two themes. The first is concerned with the properties of the defining tables; the other is about the graphical display of the movement in the form of geographical maps.

I do not include here speculation on the reasons for movement, nor interaction models. For this see Howe 1959, Berry 1966, Dorigo and Tobler 1983, Tobler 1988, Tobler 2001, and the references in these publications.

Questions about from-to tables

The migration tables produced by the U.S. Bureau of the Census every decade represent a useful case. These tables are based on a sample subset of the census questionnaires. Along with the tables the census offices provides estimates of the sampling error. But these do not take into consideration the geographical aspects of the errors, and this is my concern here. I am not familiar with many studies of this type of error (but see Bolduc, et al, 1995). Examination of the literature on input-output tables may also be useful, and there is general literature in the geographic information science community on spatial errors (Zhang and Goodchild, 2002).

The Census Bureau publishes the from-to tables at a variety of resolutions, aggregated from the individual questionnaires into patches of geography. These tables are published about three years after the enumeration. They release some of the detailed information on individuals only after about a half of a century or more. They generally do not inquire as to the 'precise' date of a migration or move and only ask for events occurring within the last (usually) five years. Internal Revenue Service information coming from address changes has a finer time distinction but is also somewhat limited. The government in this country also publishes tables of commuting patterns and of commodity flows and trade information. And there is currently a great deal of interest in information movement and patterns of fiscal exchange.

How can the geographic error(s) in these diverse types of tables be measured, or at least recognized? Movement tables are not random but in fact have a great deal of structure. Typically

the largest amount of movement is between the closest places, but this is modulated by the size of the interacting places. And a few large volumes (table entry sizes) account for a dominantly large percentage of the movement, with very many minor items going between the remaining majority of places. There is also usually a great deal of persistence over time in the movement patterns. One need only compare and map the eight decades of state-to-state US Census Bureau migration tables to see this (Tobler 1995).

A method of recognizing the structure in movement tables is to randomly permute the table entries. In one of my early studies (1976) I did this and the structure did in fact disappear, as demonstrated by the incoherent and chaotic map displays. It would seem that Monte Carlo simulations in this vein would be of interest. And there is almost always spatial autocorrelation in geographic variables. I know of few studies of this effect in single, or multiple, from-to tables. Does this affect Markov chains, or variances, when computed from these tables? Do methods related to two-dimensional (spatial) spectral analysis/decomposition offer promise here? Can regional variable theory (crudely, Kriging) be applied to these tables - the geographic locations are, after all, assumed to be known. Can Kalman filters be applied? Are abstract mathematical studies of matrices of use? For example, the from-to tables are generally asymmetric, and an elementary theorem demonstrates that these tables can be separated into symmetrical and skew-symmetrical parts. The degree of asymmetry can be measured and the variance computed separately for the two parts.

What is the effect of alternate partitionings of the geographic domain on the phenomena represented in the ensuing tables? How are the errors propagated during the aggregation process? There are several such studies of statistical inventories, but to my knowledge, not many of movement tables. It is also of interest to compare annual tables with half-decade tables, and how one can move between them. What errors are then introduced? Comparison of aggregate tables and individual itineraries and histories also invites investigation.

Questions about discrete graphical map displays

An obvious deficiency of from-to flow maps relating movement between geographic areas is the use of centroids to depict the location of the ends of the movement patterns. For example a map of all of the migrations to California by state shows the movements going to only one point. This is not very sensible. Not only is there the question of where to put this point, but also it hides the dispersion of the actual movement. A spread out paint-brush-like effect - with strands going to dozens of locations - should be put at the end of the arrow going to California to capture this, but is not very practical. Instead of there just being one line (arrow) going to California the migration could be represented by a bevy of lines, again too complex graphically to be feasible. Increasing or changing the spatial resolution just shifts the problem to another level and does not change this at all. At the lowest level, where every move is shown individually, we would need several millions of lines, to an almost equal number of destinations. The notion of a mean movement field helps to overcome this problem (see Tobler 1979, 1995).

The usual choice for a distinguished point location is either a random place within the bounded area used for the enumeration, or a geometric centroid, or an item-weighted centroid. Apportioning the migration to California to the three largest cities would seem to make sense, but the state-to-state tables do not contain the detail for this. Would a cluster of 'centroid' locations within an area, say one hundred for California, be more reasonable, with the movement

somehow partitioned between these? Or would stochastic, dynamic, variation of 'the' centroid location to generate multiple flow patterns be feasible, and with what sorts of parameters?

In the 1960s it was possible to defocus the CRT display under program control. This seems a near perfect way to reduce the 'crispness' of maps in order to show the fuzziness of the data. This is no longer possible so that computational methods of smoothing and blurring need to be, and have been, explored, but not, as far as I am aware, to movement maps. The method of showing movement by crisp solid bands or arrows in color, or with a pattern, should perhaps be replaced by a graded symbol with a Gaussian density cross section, and fading edges - perhaps varying in width along the course of the arrow.

Animation and simulation clearly offer some advantages in these respects, possibly if done at faster than the 30 frames per second usually employed.

Spatially continuous graphical map depictions

Imagining that the movement tables become infinitely large so that the space becomes continuous, and no longer broken into discrete pieces allows a depiction using continuous vector fields and streamline maps. In my implementation these are produced from the same kinds of interaction from-to tables as discussed earlier. The technique is to lay a fine grid over the area of the movement, and to assign the moves to the nodes of this grid. Then one sets up a large array of linear partial differential equations and solves these by a finite difference method, after adding a boundary condition. A neater solution would be to have the Census Bureau assemble the data directly by a fine raster of latitude/longitude quadrangles and then the spherical equations could be used immediately. The partial differential system gives results similar in appearance to the flow of a non-viscous liquid, with a potential and gradient. Going to this form of continuous representation avoids the centroid-location problem, but some of the same questions about errors can be asked.

In a vector field the structure is immediately obvious, adjacent vectors clearly being correlated in length and direction. Conversely, if this is not the case then it is also obvious. Other properties, such as the curl and rate of change, can also be detected and measured. Domains of divergence or convergence, commuting fields, hinterlands, 'milk sheds', or basins & catchments, backwaters & eddies, and areas of turbulence, are all easily identified. Do Karman vortices occur in the movement of people – the circling about the Ka'ba comes to mind?

Reducing the size of the table (by aggregation of the geographical areas, say) results in a coarser spatial resolution. Now approximating the continuous space to produce the vector or streamline maps for the reduced-resolution table has a clear impact. The result is a two-dimensional spatial frequency filter (Holloway 1958) - a smoothing, or low pass filter – and this is clearly seen on the resultant vector and streamline maps, as shown in Tobler (1990, 2001), with a reduction in spatial detail. Can this interpretation also be applied to the discrete type of movement map? An interesting thought: show the circa 9,060,000 possible county to county migrations for the entire United States on discrete flow maps (with bands for the total, arrows for the net moves, and double arrows for the two-way moves), and compare these with the state to state movement pattern that requires rendering of only about 2500 flow bands/arrows. Does the difference in resolution still appear as an obvious spatial low pass filter? Surely this can be formalized mathematically.

The data errors, in the continuous representations, have not gone away. Perhaps we need fuzzy vector fields. Is there room for random Markov field methods here, perhaps in the complex domain? Possibly these have already been studied in meteorology or by mathematicians

Some references

(Those by me can be found under reprinted articles at <http://www.geog.ucsb.edu/~tobler/publications/>)

Berry, B., 1966, *Essays on Commodity Flows and the Spatial Structure of the Indian Economy*, Department of Geography, University of Chicago Research Paper 111.

Bolduc, D., R. Laferriere, G. Santaross, 1995, "Spatial Autoregressive Error Components in Travel Flow Models...", page 96-106 of L. Anselin & R. Florax, eds., *New Directions in Spatial Econometrics*, Springer, New York.

Dorigo, G. and W. Tobler, 1983, "Push Pull Migration Laws". *Annals*, Assn. Am. Geographers, 73 (1): 1-17.

Holloway, J., 1958, "Smoothing and filtering of time series and space fields", *Advances in Geophysics* 4, 351-389,

Howe, R., 1959, "A field theory of land use and the movement of people", Doctoral Dissertation, Engineering Department, Purdue University, Purdue.

Tobler, W., 1976, "Spatial Interaction patterns", *J. of Environmental Systems*, VI (4): 271-301.

Tobler, W., 1979, "A Geographical Probability Density Function", *Ontario Geography* (Department of Geography, University of Western Ontario, London), 13: 41-46

Tobler, W., 1981, "A Model of Geographical Movement", *Geographical Analysis*, 13 (1): 1-20.

Tobler, W., 1985, "Derivation of a Spatially Continuous Transportation Model, *Transportation Research*, 19A (2): 169-172.

Tobler, W., 1988, "The Quadratic Transportation Problem as a Model of Spatial Interaction Patterns", pp. 75-88 of W. Coffey, ed., *Geographical systems and Systems of Geography: Essays in Honor of William Warntz*, University of Western Ontario, London.

Tobler, W., 1990, "Frame Independent Spatial Analysis", pp 115-122 of M. Goodchild & S. Gopal, eds, *Accuracy of Spatial Data Bases*, Taylor and Francis, London, UK.

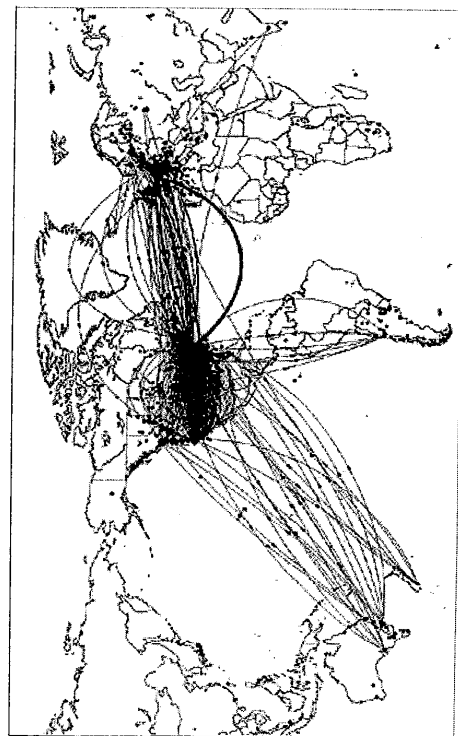
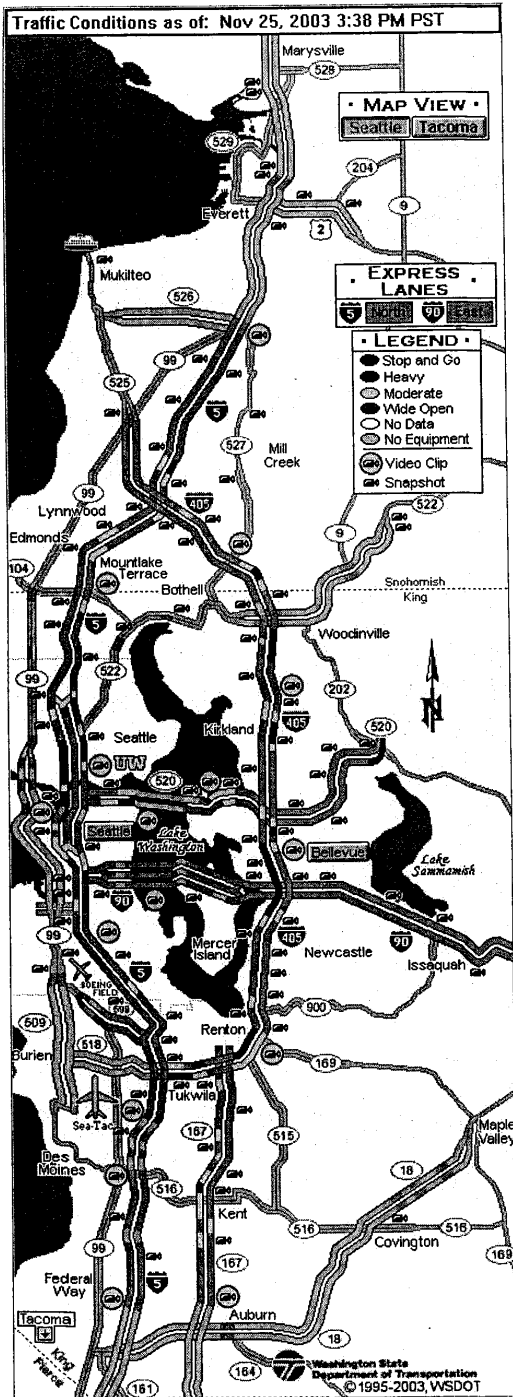
Tobler, W., 1999, "Migration: Ravenstein, Thornthwaite, and Beyond", *Urban Geography*, 14 (4): 327-343.

Tobler, W., 2001, "Movement", power point presentation at <http://www.geog.ucsb.edu/~tobler>.

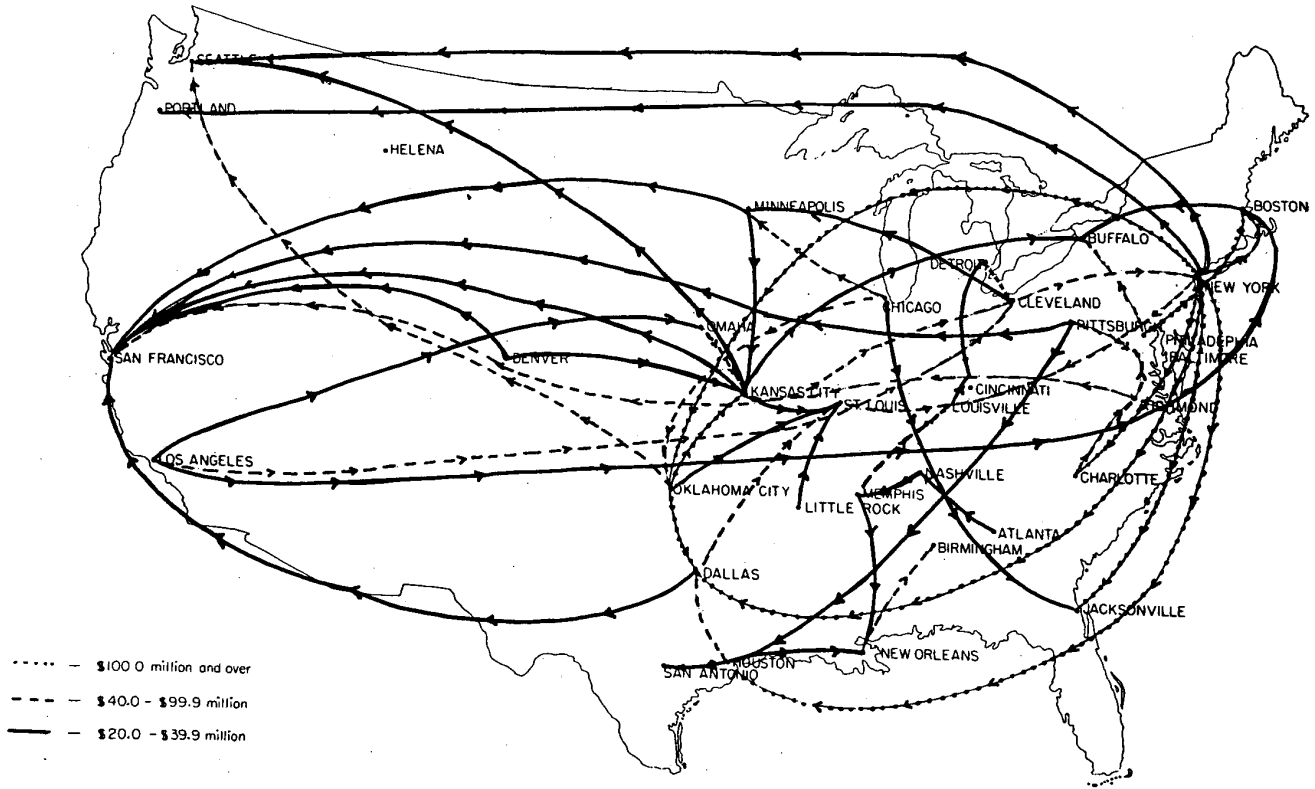
Zhang, J., M. Goodchild, 2002, *Uncertainty in Geographical Information*, Taylor and Francis, New York.

¹ Types of movement maps:

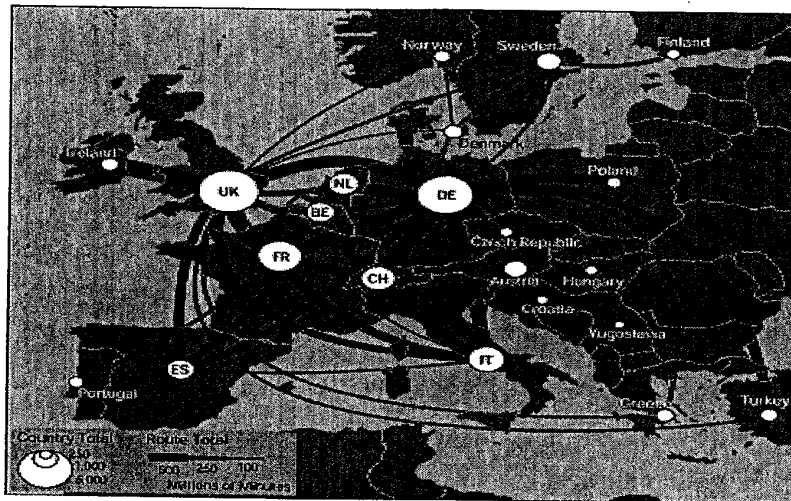
There are lots of them. On Google one finds single line arrows, movement along city road networks (some 'live', and in real time), stream and river flows, and a few of interaction between geographical areas of the type examined here. But the more exciting ones are of climatic elements - e.g. storms, tracking a tornado live - or rainfall and clouds. Or tracking individual automobiles, truck, or trains, and more recently GPS equipped people. These are no longer aggregated moves - the ultimate resolution. A few examples are appended and show some alternatives including some from my power point presentations. Some of these were produced manually, others by computer. The last map is from a recent (July 2003) development found on the Internet at <http://www.fhwa.dot.gov/ctpp/cfs8x.htm>. Also look at 'GeoFreight: The intermodal Freight Display Tool' available from the Bureau of Transportation statistics of the Department of transportation. And for more Journey-to-Work flows look at <http://www.fhwa.dot.gov/ctpp/flow.htm>



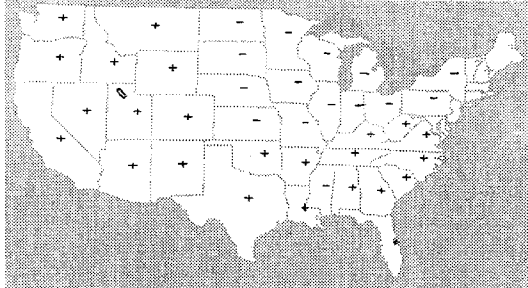
Map of complete aggregate news flow, worldwide, from 13 May 1993. (Courtesy of Brian Reid)



Map 20. Net inflows through Interdistrict Settlement Fund to Federal Reserve Offices except New York, Chicago, Philadelphia, Detroit, Pittsburgh, and Atlanta, June-July 1954. Source: N. N. Bowsher, J. D. Daane, R. Einzig [7], p. 144.

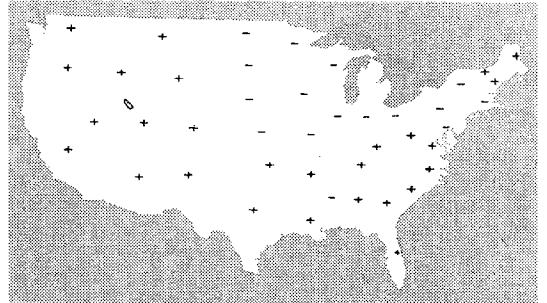


The Population Change Information Can Be Positioned Locationally using centroids



Observe the spatial autocorrelation and how this is brought out more clearly by omitting the collection unit boundaries, as on the next map.

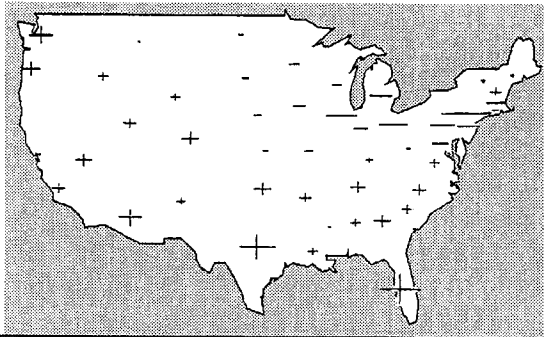
Population Change at State Centroids



The map is even better if the symbol size is made proportional to the magnitude of the change, as on the next map

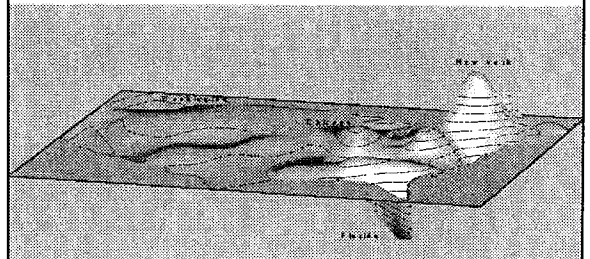
Gaining and Losing States

Based on the marginals of a 48 by 48 migration table



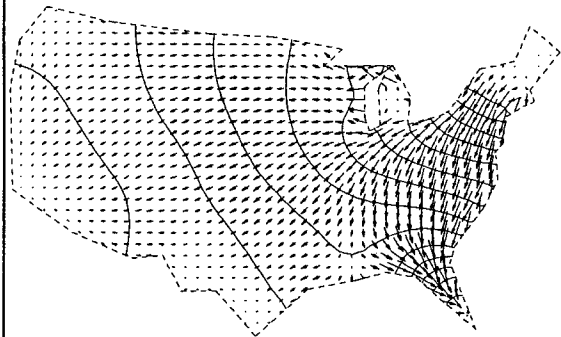
The Pressure to Move in the US

Based on a continuous spatial gravity model

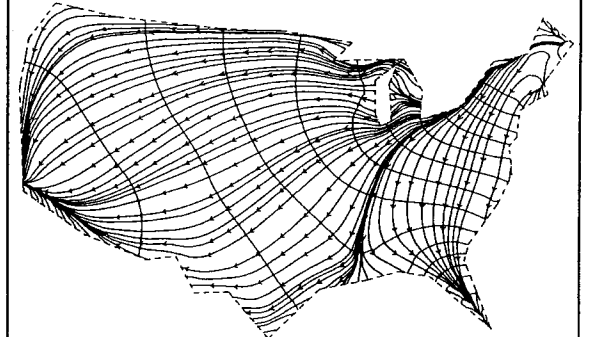


Another view

The Migration Potentials and Gradients
with the potentials shown as contours



Migration Potentials and Streaklines
with the gradient vectors connected to form streaklines



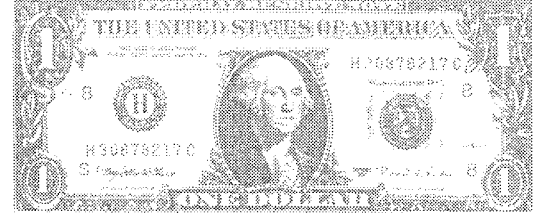
As an example of geographic movement.

In the United States the Currency
Indicates Where It Was Issued

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

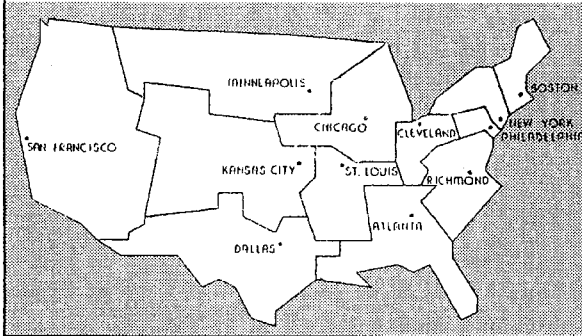
You can check your wallet to estimate your
interaction with the rest of the country.

Dollar Bill (Federal Reserve Note)



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)



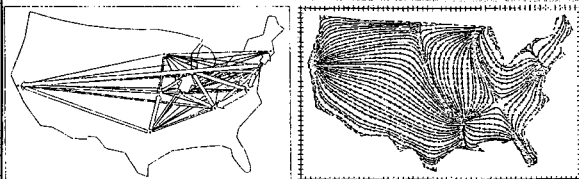
Movement of One Dollar Notes between Federal Reserve Districts, in hundreds, Feb. 1976

	To:	B	NY	P	CI	R	A	Ch	SL	M	K	D	SF
From: Boston	360	239	27	32	27	118	90	16	16	15	17	26	
New York	602	1880	224	209	336	202	286	15	15	126	15	251	
Philadelphia	113	212	569	34	522	130	124	3	25	10	10	80	
Cleveland	68	102	27	126	21	122	403	16	22	21	19	31	
Richmond	150	266	153	228	289	324	295	20	62	24	22	137	
Atlanta	122	129	157	186	219	124	229	30	31	28	102	29	
Chicago	97	153	29	496	122	266	820	22	22	100	46	280	
St. Louis	31	36	14	142	89	201	593	322	36	123	47	209	
Minneapolis	12	26	11	32	29	22	295	10	128	51	14	28	
Kansas City	20	41	3	55	40	22	215	33	120	62	36	227	
Dallas	31	41	14	38	46	165	125	20	37	253	28	215	
San Francisco	32	87	22	32	114	106	251	22	122	123	42	580	

Dollar Bill Movement in the US net flow

Discrete representation

Continuous pattern representation



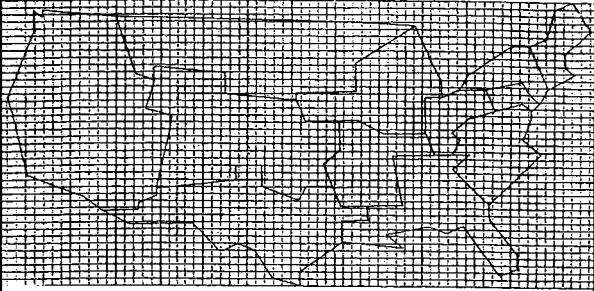
Based on computer programs from the 1980s

The Table of Dollar Bill Movements
 was obtained from MacDonalds outlets throughout the
 United States.

Source: S. Pignatello, 1977, *Mathematical Modeling for
 Management of the Quality of Circulating Currency*, Federal
 Reserve Bank, Philadelphia

From the table we can compute a movement map.

First the Federal Reserve Districts Are "Rasterized"



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

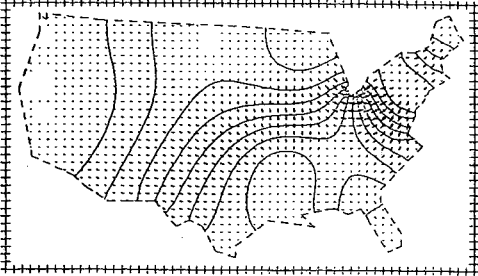
The Map is Computed Using a Continuous
 Version of the Gravity Model

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.

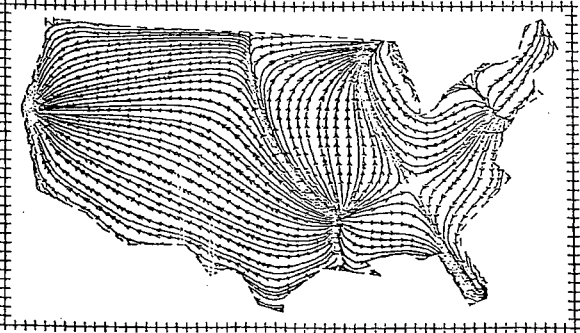
W. Tobler, 1981, "A Model of Geographic Movement", *Geogr. Analysis*, 13 (1): 1- 20
 G. Dorigo, & Tobler, W., 1983, "Push Pull Migration Laws", *Annals, AAG*, 73(1):1-17.

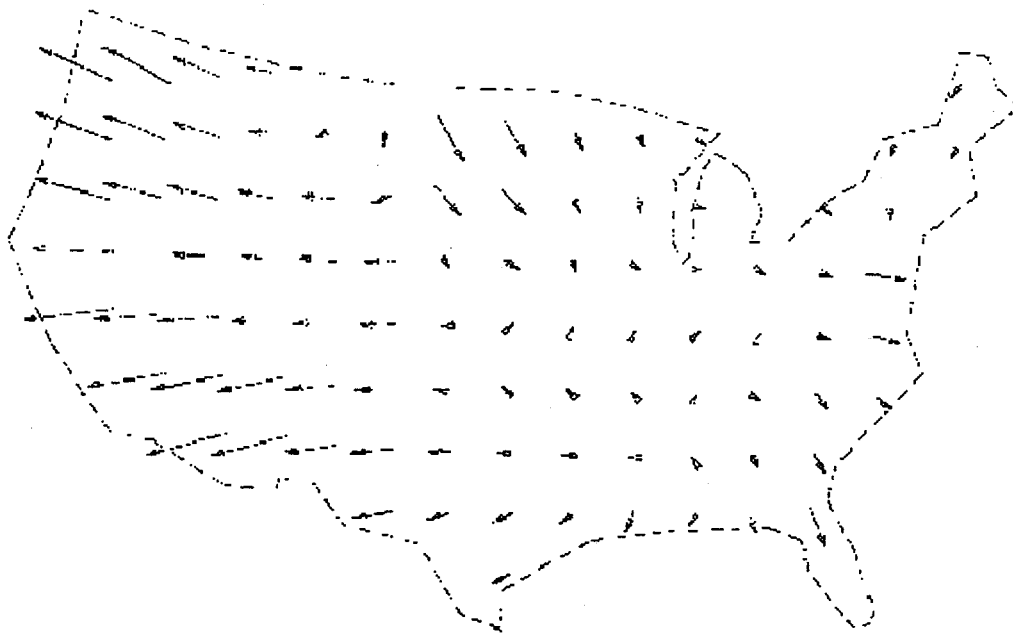
Solving the Equations Yields the Potential
 Shown here by contours



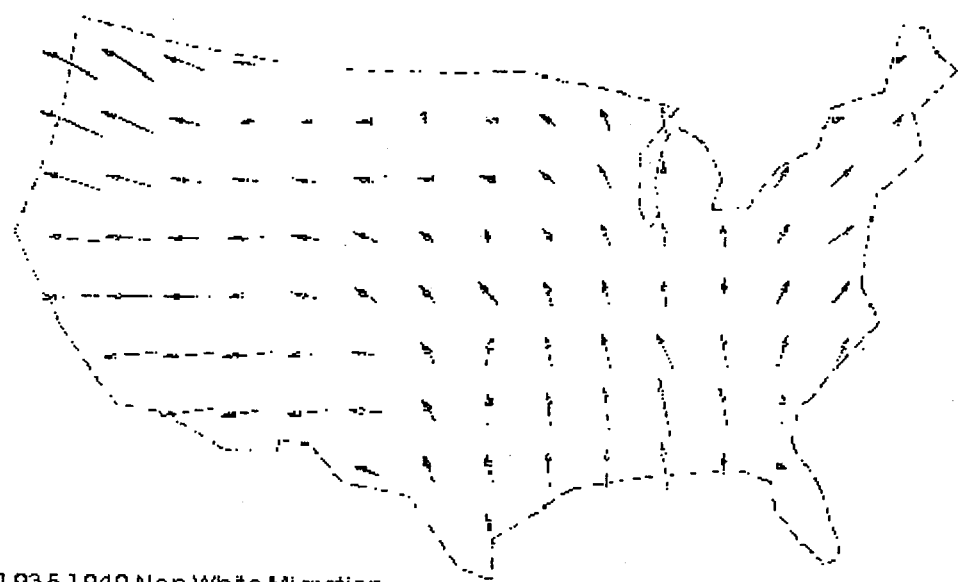
The raster is indicated by the tick marks. The arrows are the gradients to the
 potentials. The streakline map is obtained by connecting the gradient vectors.

Dollar Bill Movement in the U.S.



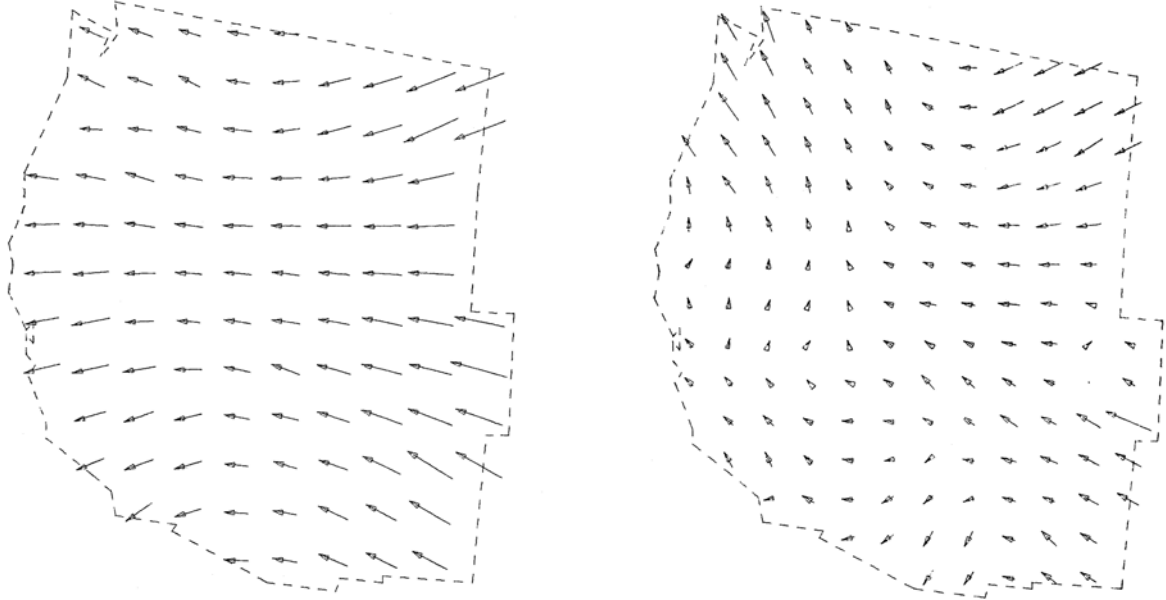


1935-1940 White Migration



1935-1940 Non-White Migration

Migration in the Western United States by State Economic Area, computed using W. Tobler's vector field program. Left 1935-1940, right 1965-1970.



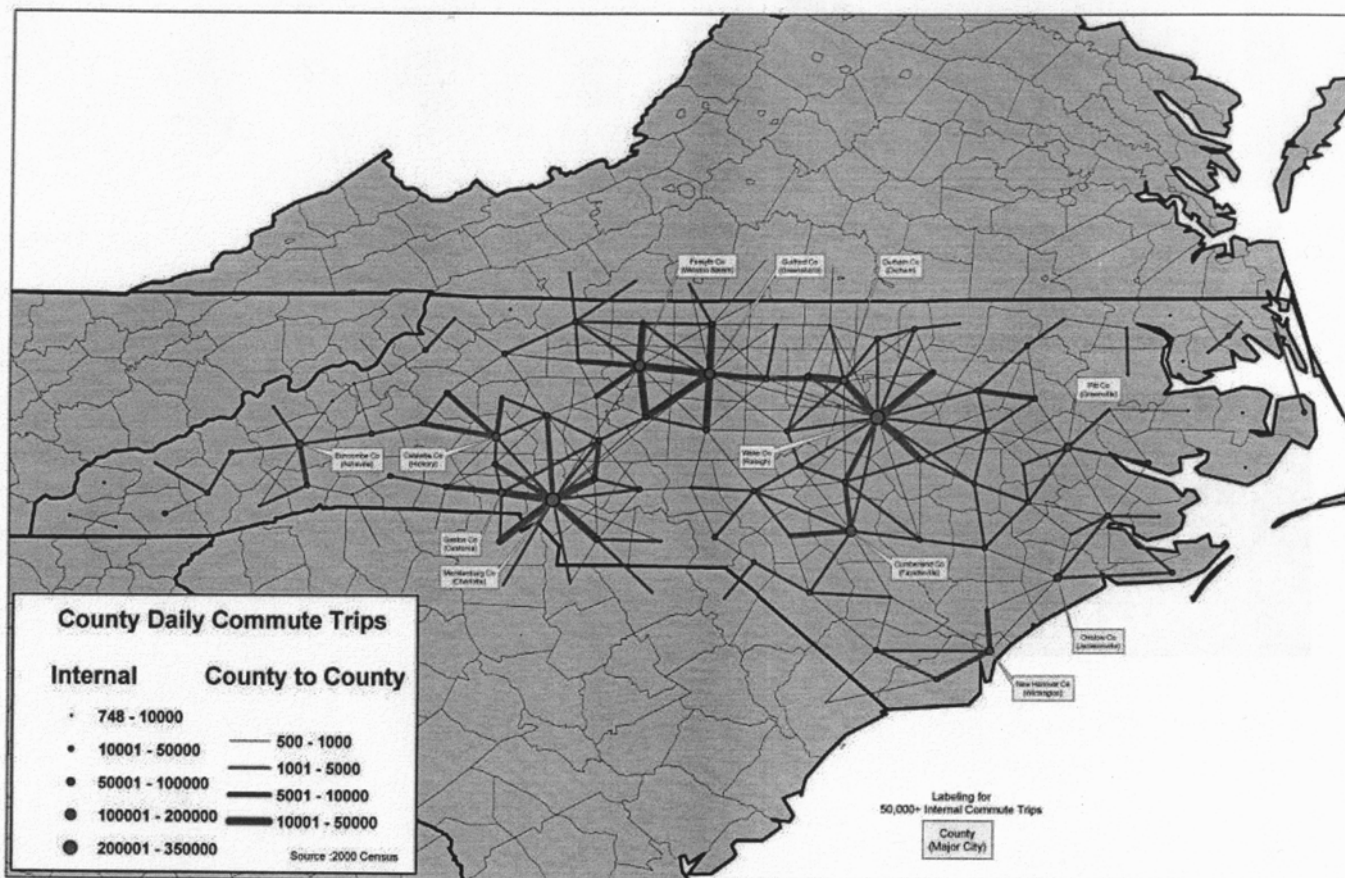


EXHIBIT A – North Carolina County-to-County Commuting

Optimal Parsing of Large Arrays

Waldo Tobler
Geography Department
University of California
Santa Barbara, CA 93106-4060
<http://www.geog.ucsb.edu/~tobler>

Current technology permits the collection and retention of large arrays of data. It is convenient use graphical methods for an overview of this information. In this note I present one method of thresholding for best visual representation for one class of arrays. The information in this case represents geographical movement between spatial locations, i.e., movement tables as are used to represent migration. In this case the arrays are square but not symmetrical. A similar method may be used in other cases. For geographical interaction the appropriate graphic is a flow map, for movement between areas or on a network. The particular motivation stems from the recent completion of an interactive program, 'Flow Mapper' (available for download at [CSISS.ORG/Spatial Tools/](http://CSISS.ORG/SpatialTools/)).

By thresholding is meant the level below which information is to be suppressed, discarded, or not used. So the question becomes: what is the best threshold value to use? In the case of migration, and most kinds of spatial interaction, we know that the distribution is such that there are many moves involving a small number of people and only a few moves with a large number of people. Ignoring the numerous small moves does not detract substantially from the overall situation.

The algorithm is as follows. First compute the total number of moves in your table. Then find the smallest and largest table entry. Do this for all moves for each contemplated type of map. That is, do it for all entries (M_{ij}), or for the sum of the two-way moves ($M_{ij} + M_{ji}$), or for the net moves $|M_{ij} - M_{ji}|$. There will be at most $(N*N-N)$ moves when looking at all M_{ij} , at most $(N*N-N)/2$ when looking at the other types of move, ignoring the diagonal entries (usually large but not describing a move).

Now, starting with the next to smallest movement volume (the move involving the almost the smallest number of people), compute the number of moves below that value, and the corresponding volume of movement. Do this for each potential threshold (i.e., each movement volume) - it is convenient to convert all the values to percentages of the maximum for this. The computation must then be performed, as many times as there are unique entries in the table. Sorting the values by magnitude is convenient for this computation. Take the difference between the percentage of volume (number of people) and the percent of moves. Make a graph of this difference as a function of the volume of movement. The inflection point on this curve allows one to pick out the optimal

thresholding level, below which moves are to be ignored on the map.

As an example the 48 by 48 table (omitting Alaska, Hawaii, and the District of Columbia) of migration between the years 1995 and 2000 in the contiguous United States, as estimated and reported by the Census Bureau, may be used. The table contains 2,256 individual moves. The smallest volume, zero, is between North Dakota and Rhode Island; the largest is 308,230 between New York and Florida. The number of combined back and forth moves is 1,128, of which 13 is the smallest (again between North Dakota and Rhode Island) and the largest (378,448) is between New York and Florida. There are also 1,128 net moves, the difference between the in-movement and the out-movement at each of the states. The smallest (South Carolina and West Virginia) and largest in absolute value are zero and 238,012.

The accompanying figures give the graphs of the difference score for these three cases. The computation puts the threshold values at 9,456, 18,898, and 3,745, respectively. In each case these values are extremely close to the average flow magnitude. This is also the case for other situations examined.

To see the effect on the flow maps I present additional figures. In each case there is a map showing - that is, attempting to show - all of the moves, and a parsed map using the average as the threshold value.

An alternative to thresholding is to reduce the diversity of information by aggregating the areas in the system of moves. There is some literature on this, and some attempt at optimal aggregation, although most studies of aggregation do not treat information concerning pairs of places (but see Masser and Brown 1975, 1978; Broadbent 1969; Masser 1976; Masser and Scheuwater 1980; Openshaw 1977; Slater 1981). The difficulty here is that aggregation operates as a low-pass spatial filter and thus obscures fine detail that is often of interest. I do not recommend aggregation methods.

References:

Broadbent, T., 1969, "Zone size and singly constrained interaction models" CES-WN-132, Centre for Environmental Studies, London, UK.

Masser, I., P. Brown, eds, 1975, "Hierarchical aggregation procedures for interaction data", *Environment and Planning*, A: 7: 509-523.

Masser, I., P. Brown, eds, 1978, *Spatial Representation and Spatial Interaction*, Kroese, Leiden.

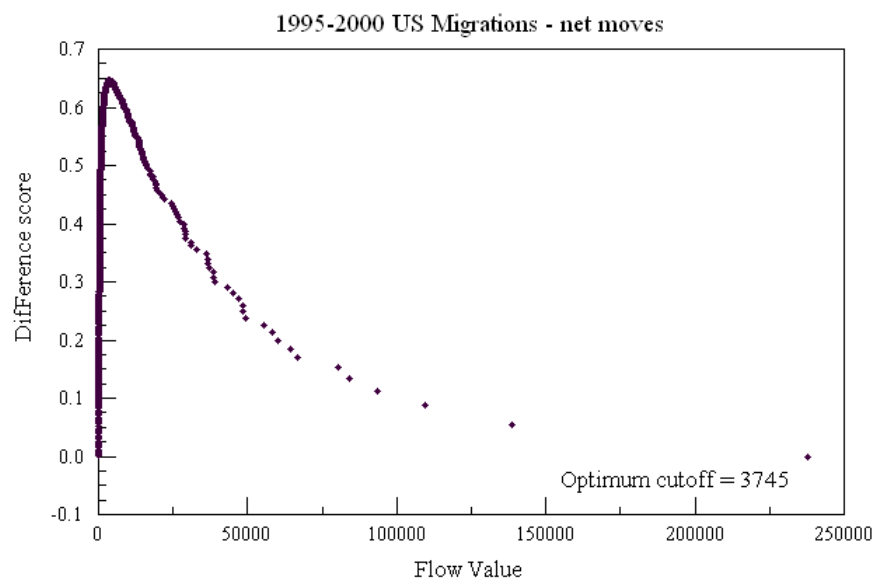
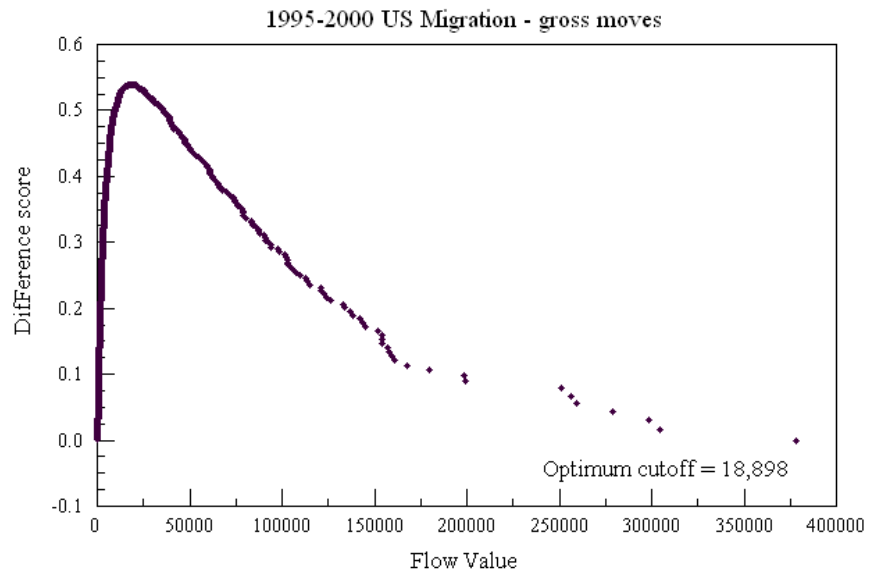
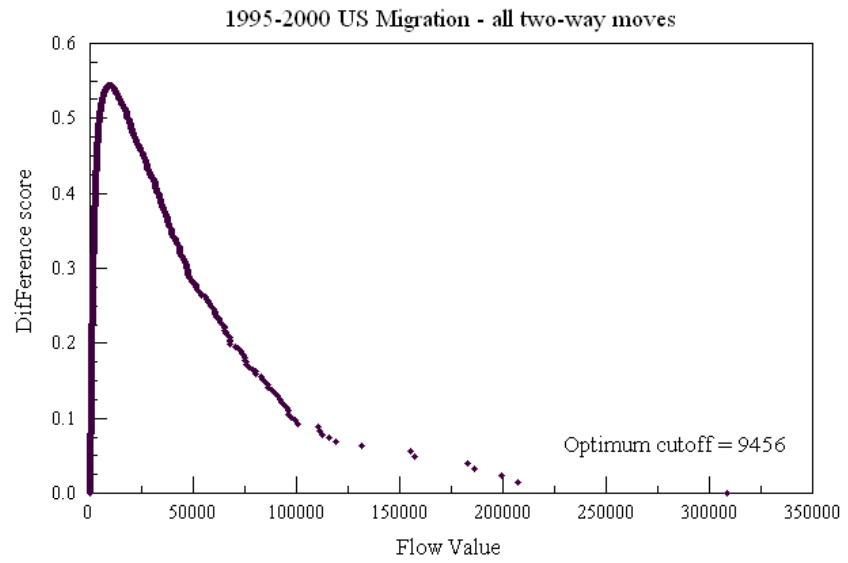
Masser, I., 1976, "The design of spatial systems for internal migration systems", *Regional Studies*, 10, 39-52.

Masser, I., and S. Scheurwater, 1980, "Functional Regionalization of Spatial interaction data...", *Environment and Planning, A*, 12: 1357-1382.

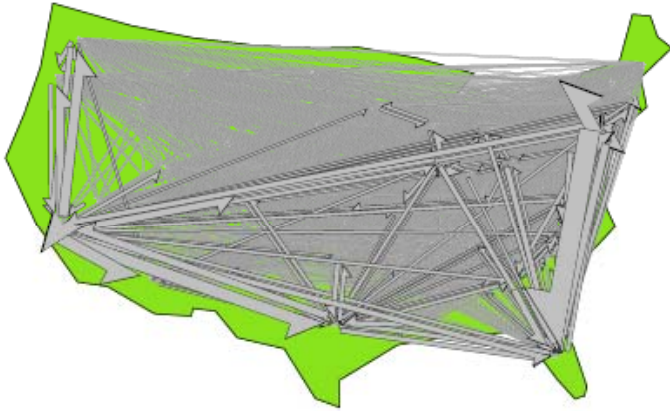
Openshaw, S., 1977, "Optimal zoning systems for spatial interaction models", *Environment and Planning, A*, 9: 169-184.

Slater, P., 1981, "Comparisons of aggregation procedures for interaction data", *Socio-Economic Planning Sciences*, 15: 1-8.

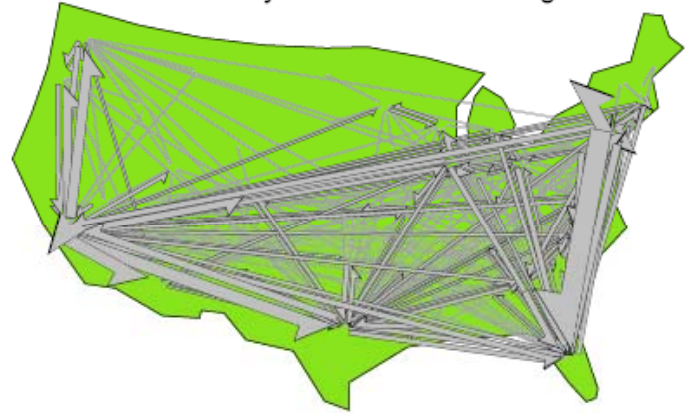
(The computer program used to perform the threshold calculation can be obtained from the author)



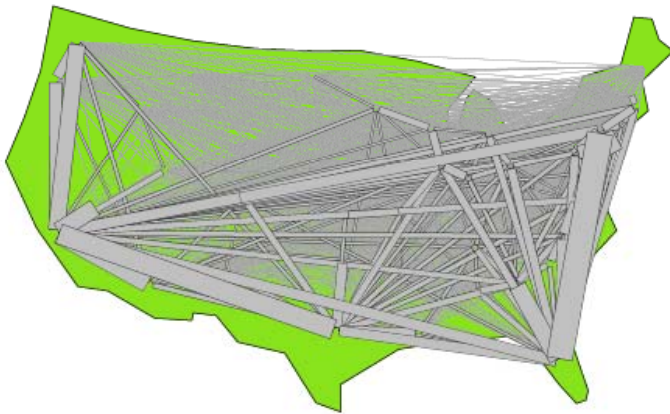
1995-2000 US Migration
All two-way moves



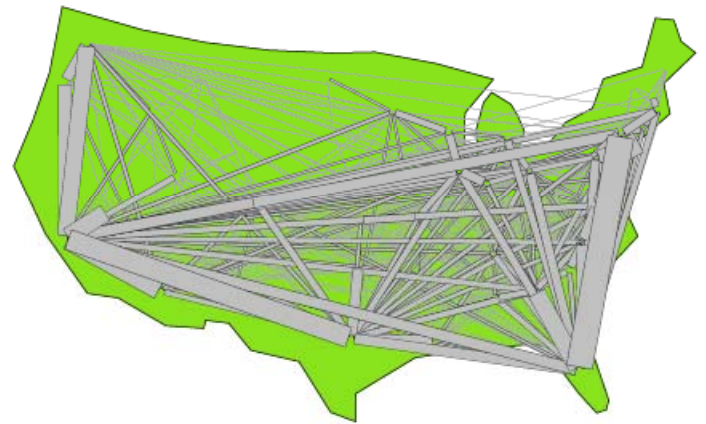
1995-2000 US Migration
Two-way moves above average



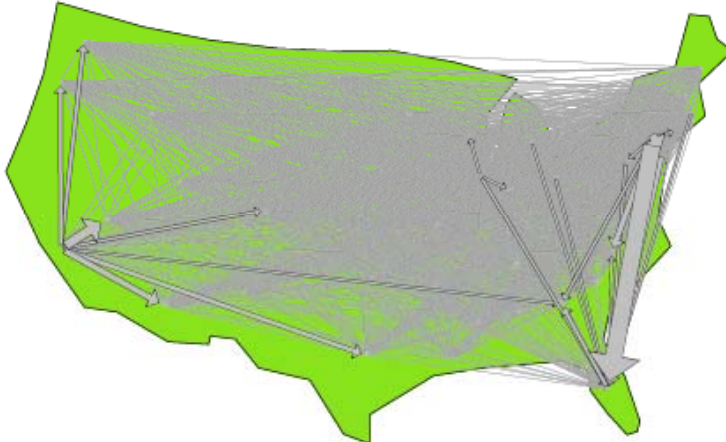
1995-2000 US Migration
All gross moves



1995-2000 US Migration
Above average gross moves



1995-2000 US Migration
All net moves



1995-2000 US Migration
Above average net moves

