

Experiments In Migration Mapping By Computer †

Waldo R. Tobler
Geography Department
University of California
Santa Barbara, CA 93106

ABSTRACT. Migration maps represent patterns of geographical movement by arrows or bands between places, using information arriving in “from-to” tables. In the most interesting cases the tables are of large size, suggesting that computer assistance would be useful in the preparation of the maps. A computer program prepared for this purpose shows that graphical representation is feasible for tables as large as fifty by fifty, and possibly larger. The program contains options for alternate forms of movement depiction, and rules are suggested for the parsing of migration tables prior to the cartographic display, without loss of spatial resolution.

KEY WORDS: Computer cartography, geographic movement, thematic maps.

Maps that show patterns of geographical movement function as particularly effective illustrative and research tools. Like most graphical aids their value increases in direct proportion to the complexity of the data. The general style of these maps has not changed much since the time of Harness, Belpaire, or Minard in the last century (Brinton 1914; Robinson 1955, 1967, 1982; Tufte 1983). The areas between which a migration, or other movement, occurs are connected by a “band” whose width represents the quantity moved. What is new today is that computers can aid in their construction.

As a generic case for all movement mapping problems it is assumed here that the data take the form of an N-squared table of geographical interactions. These “from-to” tables indicate all of the N by N possible movements between N geographical areas during some interval of time. When the movements are between only a few of these possible pairs of places, a small subset of the N by N pairs for example, then complete from-to tables do not provide a particularly convenient formulation of the problem. These special cases might include movement of people on a highway network, where only the link volumes are known. Here the detailed path of the movement may be shown but the individual nodes of such a network are directly connected to only a few, adjacent, nodes. Most of the entries in a complete node-to-node table for such a network would be empty. Similarly, travel from many agricultural zones to a few urban centers would result in a from-to table with a very special, sparse, structure. Conceptually it makes sense to think of these cases as large but incomplete N by N from-to tables but from a data storage or data entry point of view this is not very practical, even though the types of displays described here are still relevant.

For the general N by N situation the cartographic challenge is when N is greater than ten, with over one hundred possible flows. Now it begins to pay to use a computer. Tables of smaller size usually have such a coarse spatial resolution as to be geographically uninteresting and the depiction of the few flows not a difficult problem. Thus the U.S. Bureau of the Census often shows migration between only four regions - Northeast, North Central, South, and West, but actually collects data by counties. A county-to-county table of migration in the United States could contain nearly ten million numbers; the actual estimated 1965-1970 county-to-county migration table contains only 540,022 numerical entries (Slater 1984), in part because of the sampled manner in which the data were collected. This is still a large number of movements.

The cartographic problem is then severe, but a solution is also more urgently needed because no one can comprehend this large a table. And yet such tables can easily be stored with current information technology, and several migration tables of this size are available. To reduce the 3141 by 3141 county-to-county table to a four by four region-to-region map, as the Bureau of the Census does, seems a rather excessive parsimony, with a change in mean spatial resolution from 50 kilometers to around 1420 kilometers. It is not only the US Census that makes such reductions. In Holland, for example, van der Erf (1984) has available a 774 by 774 table of movements between municipalities (for several time periods) but displays the results in map form between only four regions, a reduction in mean spatial resolution from around seven kilometers to 100 kilometers.

In the research described here we have found that a fifty by fifty movement table, with 2500 potential entries, can be depicted fairly easily using conventional cartographic techniques in a computer-plotter environment. This size corresponds to a state-to-state migration table. By judicious parsing of such a table (see below) it is usually not necessary to draw more than 25% of the flow arrows. It is possible that larger tables can be accommodated by these techniques, but we have not tried. As the tables get larger the spatial resolution increases and the number of movements grows as a quadratic function of the number of origins and destinations. Ideally a desirable parsing strategy is one which yields only a linear growth in the number of movements needing to be shown. Even then one can imagine taking the limit, "as N approaches infinity", wherein the geographical space is considered to be continuous, not broken into data collection units, and the table no longer finite. This leads one to methods of movement depiction more analogous to those used in fluid dynamics (Figure 1). This vector field approach to migration and general movement mapping has been described elsewhere (Tobler 1978, Tobler 1981, Dorigo and Tobler 1983). These continuous techniques are also useful for movement tables of large but finite size although the present discussion is limited to more conventional mapping methods.

The earliest computer drawn flow maps are those of the Chicago Area Transportation Study (1959). A special cathode ray tube system, "the cartographatron," was constructed to display several million "desire lines" (Figure 2). On this device the end points of individually desired trips, as determined from interviews coded to the nearest quarter mile on a rectangular grid, were connected by a single light trace, and the cumulative result obtained from a time exposed photograph. These maps were then used to help determine the locations of new expressways. Another mapping program, by Kern and Ruston (1969), showed geographical interactions by single lines drawn on a graphical pen plotter. The early work of Wittick (1976), however, is more in the spirit of the present effort. Related studies in transportation engineering have also occasionally been reported (Beddoe 1978, Noguchi and Schneider 1977, McLaughlin 1977).

In addition to the complete N by N movement table, one is required to have a map showing the boundaries of the data collection units. These boundaries are then digitized and can be drawn as background information on the movement map. Weighted geographic centroids of these regions are calculated and used as rectangular coordinates for the initial and terminal points of the flow lines. A small offset from these origin-destination points is generally desirable in order to avoid excessive overlap where the flow bands come together. The simplest depiction consists of straight connections between origin and destination, but the number of graphical options is still large. Cartographic theory does not provide much guidance so we have done a number of simple experiments. These may be broken down into two general types: (a) experiments with alternate graphical displays, and (b) experiments in reducing the complexity of the movement table. To some extent these categories overlap but they do provide a convenient organization for discussion.

The simplest graphic is the rectangular flow band with width proportional to the flow and stretching from starting centroid to ending centroid, and representing all of the two-way flow from place A to place B and from place B to place A, a combined total. The cartographic problems are minor, except for the large number of flows, $N \times (N-1) / 2$ in this case. It is necessary to choose a scale of flow magnitudes, and this choice clearly impacts the impression obtained from the map. No cartographic rules are known to us that allow an unambiguous choice for this variable. Of course, with a computer one can redo the map several times very quickly until the desired effect is achieved. As a default option, convenient but not necessarily correct, the computer program fixes the width of the largest flow band, making it equal to the distance between the closest centroid pair on the map. All other flow bands are scaled relative to this largest flow. A special scale option is available to maintain compatibility between different maps - the same migration phenomena at different historical times, for example. We have chosen not to represent the individual self moves, from place A to place A, as given by the diagonal entries of the from-to table, nor the in or out totals at the N places (the margin sums of the table) since these latter values do not represent the actual movements and are easily handled by conventional cartographic techniques; e.g., positive and negative symbols, or shading.

Should the width of the flow band be proportional to the magnitude of the movement? One alternative is to make all flow bands the same width, and then to use a variable intensity shading to represent the magnitude, as on choropleth maps, or to use a color variation to indicate the intensity of movements. Alternatively, the shading density times the area shaded (length times width of band) can be made proportional to the movements, this option corresponding to the notion that visual intensity should be proportional to density times area. Or the bands can be chosen to have their area (width times length) proportional to the movement magnitudes, the idea again being that the eye responds to area and not just width. Thus one can have constant width bands with variable shading, or variable width bands with or without constant shading. Psychological testing of the type currently popular in cartography has not been done for these options. Our impression, not supported by real evidence, is that widths proportional to flow magnitudes are interpreted more correctly. Do we now dare mention the possibility of non-linear (e.g., logarithmic) scaling of flow band widths or shading? Once the data are in the computer all of these options are very simple.

In order to be able to represent the movement from place A to place B as something distinct from the movement in the reverse direction we require an asymmetric symbol, the flow arrow being the classic form. Generally there will be $N \times (N-1)$ of these, when the self-moves are omitted. We have found this to be the most difficult graphical problem, and not really solved. The individual arrows are no particular problem for simple or barbed types (Figure 3), either in the form of constant width arrows with variable density shading, or variable width arrows with or without shading. The shading is available in one of four styles in our computer program; lines parallel to the flow, lines perpendicular to the flow, cross-hatching, and, when the movement is directional, as chevrons (a herring bone style). When shading is used our computer program allows one to omit the edge of the shape (arrow or band) for a somewhat fancier style. When shading lines parallel to the flows are used with this option one is nearly able to count lines to get an impression of magnitudes. Other problems and options are similar to those discussed for flow bands; for example, variation in the choice of magnitude scales, as shown in Figure 4. A major difficulty lies in showing the simultaneous two directional movement along the single path connecting places A and B. Half-barbed arrows, each of whose width is proportional to the respective movement, which abut or which are separated by a small gap, do not seem very effective visually, nor does putting the smaller flow arrow on top of the larger one work very well (Figure 5).

This problem is, of course, avoided if only net movements—the larger of A to B minus B to A, or B to A minus A to B—is used. Then there are again only $N \times (N-1) / 2$ movements to be shown and, equally importantly, only one (directed) movement along any single route. Since one of the Laws of Migration is that “...each main current of migration produces a compensating counter current” (Ravenstein, 1885; p. 199), these net movement maps show the extent by which the currents and counter currents differ, whereas maps of the sum of the moves in the two directions (shown by bands) indicate the total volume of the exchanges or turnover. Each of the several types of maps thus emphasizes a different aspect of the movement pattern.

Another difficulty stems from the First Law of Geography, “near places interact more than distant places.” Thus the large movements are often between close places on the map, where there is little room to draw anything. A common cartographic technique used to overcome this problem is to choose a base map, which enlarges areas of high data density. This can be done as easily by computer as by manual methods.

The graphical simplicity of the maps is greatly enhanced if the arrows or bands are shown with overlap deletion. This would hardly bear mention if it were not for the computer programming complexity, in effect the same problem as the “hidden line” problem of computer graphics. For a vector (line drawing) device an explicit routine must be used. On a raster device the problem takes care of itself if the arrows are drawn in sorted order. Our initial guess was that we should have the arrows representing the smaller flows cross over the top of the larger flows. The reverse in fact seems preferred. Why? The map clutter seems reduced for a given size of from-to table, and the more important (larger) movements become more noticeable. Optimal thresholding, described below, may reduce the clutter sufficiently so that placing the remaining smaller arrows on top of the larger ones is effective. This option seems preferred by students who have used the computer program in a cartography class. A graphical nicety is to introduce a small gap where the flow arrows overlap, generally on the southeast side to simulate a three dimensional effect. Such niceties have not been incorporated into our computer program. Other enhancements, with which we have not experimented, would include curved (circular, elliptical, splined) flow bands, or arrows with trajectories “through the air” above a map shown in perspective, or the merging of smaller movements into larger streams going in the same general direction and then splitting to go to different destinations (see Thornthwaite, 1934, for examples of this technique).

The design possibilities are of course legion, and some are easily programmed. Even without these sorts of enhancements, but including those listed below, a calculation based on the options available for input to our program shows that more than 125 distinct and different maps can easily be made from one single from-to table (not including variation in the width magnitude scale, which allows for almost infinite variety). Here is a case in which the computer greatly enlarges the cartographic possibilities. Choosing among so many options is difficult, and the setting of the default cases requires considerable cartographic judgment. Ideally we would also like the map to provide an impression of the lack of precision in the data, which often come from a sample and always contain errors. For this purpose a degree of fuzziness can be obtained on a raster display screen by defocusing under program control; alternately one can use controlled variation in the contents of the refresh buffer to yield a visual instability in the displayed map (by alternating two buffered displays, for example). We have as yet not obtained user opinions from experiments of this type.

The number of migrations, which need to be shown, can be reduced in a number of ways. Instead of showing the entire N -squared possible migrations on one map, all of the $N-1$ movements from one place, or to one place, can be shown, in gross or net form. These possibilities will yield $4N$ distinct maps from one single N by N from-to table; our program

allows us to make all of these maps on one pass through the computer. Two sample maps are shown in Figure 6.

Another, classical, method is to delete all of the movements below some threshold quantity. The problem lies in determining this threshold level. The literature on this topic is sparse, but there seems to be an optimal way of making this choice. The trick is to recognize that geographic movement tables are not random number tables and that they in fact contain a great deal of structure. It is quickly apparent that the entries in these tables approximate a Pareto distribution. That is, there are many small entries and only a relatively few large ones. These few large entries make up a large percentage of the total movement. **The optimal deletion strategy is to remove all movements whose magnitude is less than that of the average table entry.** This delicately balances the deletion of individual migration streams with retention of movement volumes. In our experience we are typically able to remove 75% or more of the flow bands while removing less than 25% of the migrants using this rule (Table 1).

Arbitrary thresholds can now be replaced by the optimal cut-off value, and this simple rule greatly reduces the map clutter while still providing a faithful representation of the geographic situation. Here less than 50 of the possible 1128 arrows need be shown to represent the majority of the net movement (Figure 7). An alternative is to use a theoretical model to produce a table of "expected" movements and then to show only those movements which are significantly different from these expectations.

Collapsing a migration table to a smaller size by combining adjacent (geographically contiguous) places is, as noted earlier, a common technique. This deletes detail from the movement pattern at the expense of spatial resolution and a geographically blurred image is obtained. This is generally not a desirable procedure and is to be avoided if at all possible. If it must be done it probably should be done in such a manner as to reduce the variance of the resolution. For example, in the United States the units used for data collection are typically smaller in the eastern than in the western part of the country. Aggregation of these areas can be used to reduce this variability in spatial resolution so that the spatial filtering introduced by the aggregation is at least the same in all parts of the map. This type of aggregation depends only on the spatial units and not on the specific data assembled using these units. But there is a large literature on the topic of spatial aggregation of movement tables (e.g., Masser and Brown 1978, Slater 1984). Unfortunately, this literature does not focus on the cartographic problems and is therefore, from the present perspective, somewhat misleading. The effect of an aggregation of areal units on movement data may be compared to other forms of map generalization (e.g., simplifying topographic maps); this point of view suggests investigation of weighted averaging of geographically neighboring movement streams as an alternative to spatial aggregation (which can be considered a weighted averaging with only 0 or 1 as the weights). Two-dimensional spectral analysis of the movement pattern may also be expected to detect spatial scales at which important events are occurring. Partitioning of migration tables by categories of subgroup (age, sex, etc., or by purpose, or distance of move, and so on) is of course also valuable, and decomposition of the complete table into additive or multiplicative components provides an additional technique worth further exploration. Computer maps can also be made of theoretically computed, smoothed, or derived tables (Markovian transition tables, for example) for comparison with actual tables.

Simplifying a from-to table can also be done by allowing movement to take place only between geographically adjacent places. Migration, for example between Maine and California, would need to be deleted under such a criterion, but an adjustment then needs to be made so that the total number of moves from Maine, and to California, do not change. The migrants are "rerouted" to pass through adjacent places. This can be done in a number of ways, but the final

result is always the same; the movement maps never have crossing bands or arrows. They are similar to highway traffic maps that show the actual movement along each link but never show the origin or destination of a particular movement. Compare Figures 7 and 8. Figure 8 shows the migrations rerouted to pass through nearest adjacent states. Movements below average are deleted for map clarity (from Tobler, 1981).

Instead of drawing arrows from centroid to centroid an interesting variation would now be to place the arrows to just cross the borders of the immediately adjacent regions, as did Ravenstein on his "Currents of Migration" map in the famous "Laws of Migration" paper of 1885 (Figure 9).

After rerouting, the from-to table has relatively few entries and it is probably more efficient to enter coordinates of the end points of the from-to links rather than the entire table. Such a program is much easier to write and the cartographer is not faced with as great a complexity as when the complete N by N table is available and needs to be represented graphically. Alternatively, our program could be changed to allow this form of data entry.

In the foregoing discussion we have left open several research problems. For example, from a computational point of view it is immaterial whether the data table is of within-city movements, or between the provinces of a country, or between several countries within a region, but truly international movement is different because the spherical nature of the earth has not been taken into account. Thus one can imagine drawing the flow arrows along great circles on an oblique orthographic view of a hemisphere; the map projection problem becomes even more difficult when a movement pattern over the entire world must be shown.

We have also assumed that the objective is just to display the migration data contained in a single table. For research purposes it is often desirable to look at the difference between two tables, at distinct time periods, or for populations partitioned by age, sex, or other characteristics, or to compare a theoretically computed table with an observed table. Then there may be positive and negative flow differences and the ingenuity of the cartographer is again challenged. Finally, completely different display techniques seem called for in more dynamic situations when one has a movement table which is regularly updated; by decade, by year, or by month. This calls for real cartographic animation. A record of one hundred years of daily migrations for a country, as might be obtained from a continuous population register, contains enough data for a movie of about one half-hour duration, assuming one days' events per frame. Commuting data also lend themselves to this type of depiction. Given that computer programs now exist for the conversion of residential street addresses into geodetic coordinates, and combining this with the recent advances in high volume data storage and processing capabilities, we can expect more cartographic displays of point-to-point movements rather than the area-to-area moves that are now usually required. This again changes the nature of the cartographic problems. Aggressive experiments are needed in these areas in order to obtain further improvements in our stock of tools for cartographic analysis

ACKNOWLEDGMENTS

M. Fowler, R. Milliff, V. Sefcik, and N. Shagour all assisted in these experiments. Partial support was provided by the Geography and Regional Science Program of the National Science Foundation under grant S0C77-00191.

REFERENCES

- Beddoe, D. 1978. An alternative cartographic method to portray Origin-Destination data, MA thesis, University of Washington, Seattle.
- Brinton, W. 1914. *Graphic Methods for Presenting Facts*, The Engineering Magazine Co., New York.
- Chicago Area Transportation Study, 1959. *Final Report*, Vol. I, Survey Findings, Chicago, pp.96-99.
- Dorigo, G., and Tobler, W. 1983. Push-Pull Migration Laws. *Annals, Assoc. Am. Geographers* 73 (1): 1-17.
- van der Erf, R. 1984. Internal migration in the Netherlands: Measurement and main characteristics. pp. 47-68 of H. ter Heide and F. Willekens, eds., *Demographic Research and Spatial Policy*, Academic Press, London.
- Kern, R., and Ruston, G. 1969. A computer program for the production of flow maps, dot maps, and graduated symbol maps. *The Cartographic Journal*, 131-136.
- Masser, I., and Brown, P. eds., 1978. *Spatial Representation and Spatial Interaction*. Nijhoff, Boston.
- McLaughlin, M., ed., 1977. *Applications of interactive graphics*. Transportation Research Record, #657, National Research Council, Washington, D.C.
- Noguchi, T., and Schneider, J. 1977. Data display techniques for transportation analysis and planning. *Transportation Planning and Technology*, 4:23-36.
- Ravenstein, E. 1885. The laws of migration. *Journal of the Statistical Society* 48: 167—235.
- Robinson, A. H. 1955 The 1837 maps of Henry Drury Harness. *Geographical Journal* 121: 440-450.
- Robinson, A. H. 1967. The thematic maps of Charles Joseph Minard, *Imago Mundi* 21: 95-108.
- Robinson, A. H. 1982. *Early Thematic Mapping in the History of Cartography*, University of Chicago, Chicago.
- Slater, P. 1984. A partial hierarchical regionalization of 3140 US counties on the basis of 1965-1970 inter-county migration. *Environment and Planning, A* 16: 545-550.
- Thornthwaite, W. 1934. *Internal Migration in the United States*, University of Pennsylvania, Philadelphia.
- Tobler, W. 1978. Migration fields. pp. 215-232 of W. Clark and E. Moore, eds., *Population Mobility and Residential Change*, Northwestern University Studies in Geography No. 24, Evanston.
- Tobler, W. 1981, A model of geographic movement. *Geographical Analysis* 13(1): 1-20.
- Tufte, E. 1983. *The Visual Display of Quantitative Information*, Graphics Press, Cheshire Conn.
- Wittick, R. 1976, A computer system for mapping and analyzing transportation networks. *Southeastern Geographer*, XVI, 1 (May): 74-81.

† The American Cartographer, 1987, 14(2): 155-163

Figure 1. Estimated state to state net migration depicted as a vector field with scalar potential shown by contour lines, and estimated trajectories of the 1965-1970 net movements. Computed as described in Tobler, 1981.

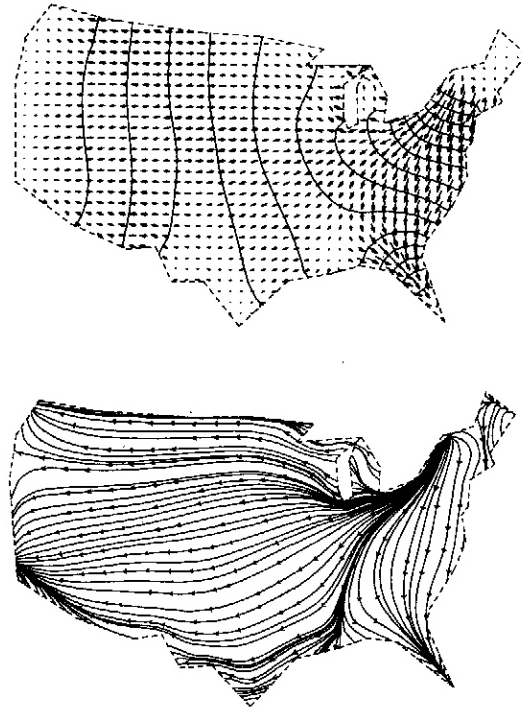


Figure 2. Cartographatron display of 9,931,000 desire line traces of personal trips in Chicago. *Chicago Area Transportation Study*, 1959; fig. 23, p. 46.

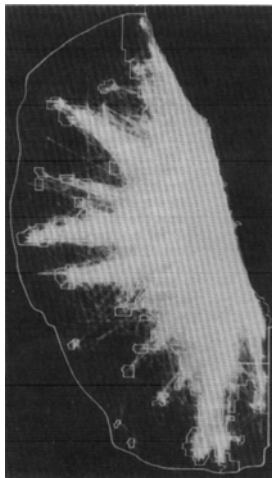


Figure 3. Arrow and shading types.

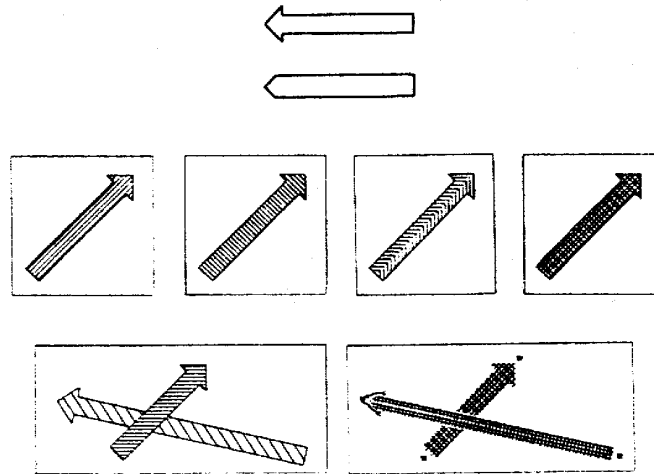


Figure 4. Net migration between states 1965-1970, showing the effect of varying arrow widths. Minor flows deleted.

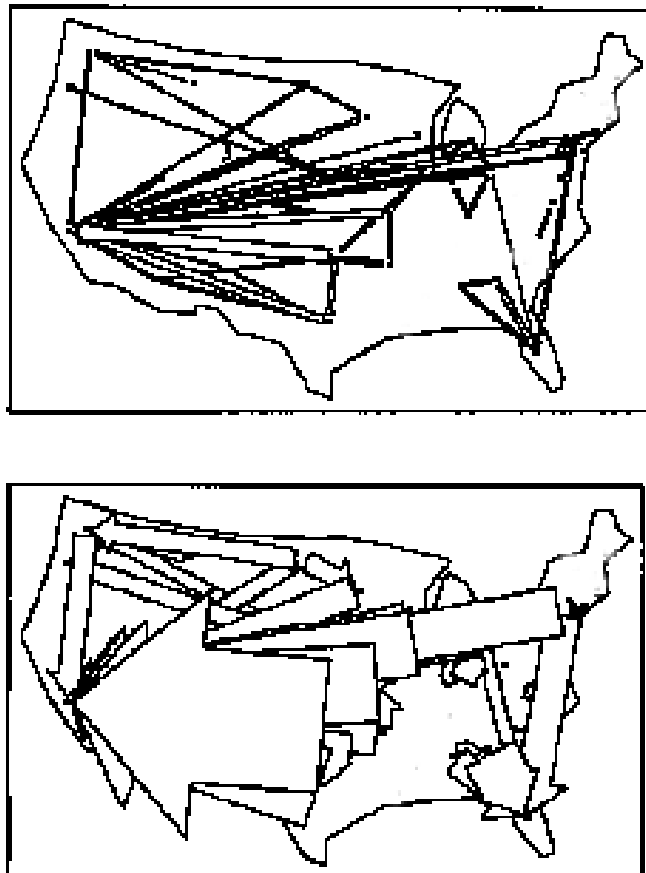


Figure 5. Bi-directional arrows.

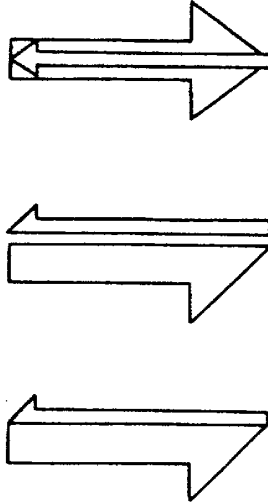


Figure 6. Gross migration to and from California, 1965-1970, with constant density shading.

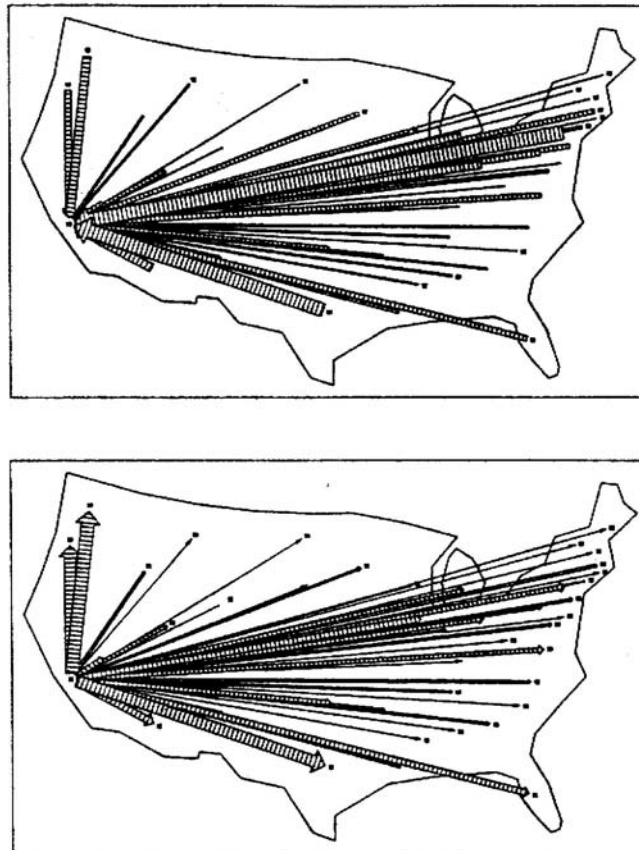


Figure 7. Net migration 1965-1970, with all volumes below the average deleted. Less than 60 of the possible 1128 ($48 \times 47 / 2$) arrows need be shown to represent the majority of the migration.

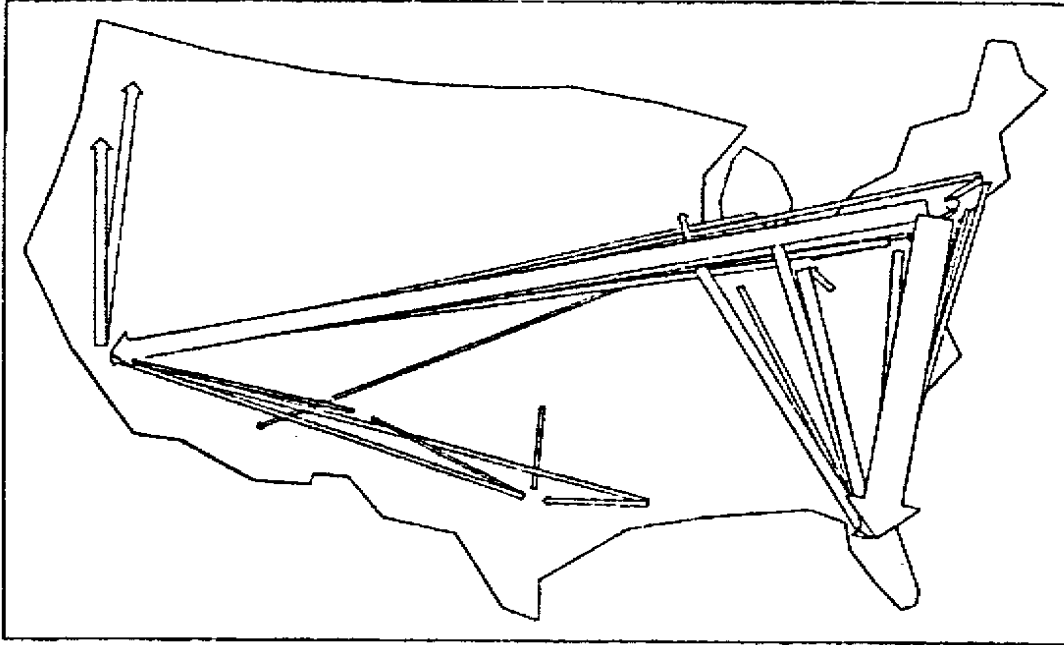


Figure 8. State to state migration, 1965-1970

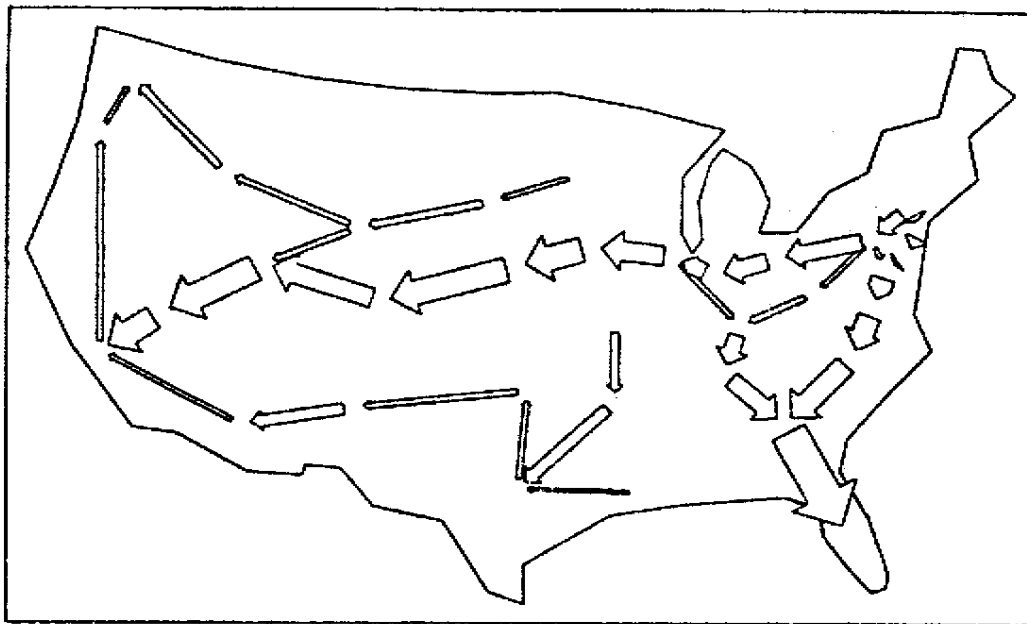


Figure 9. Ravenstein (1885), Map 5, page 183.

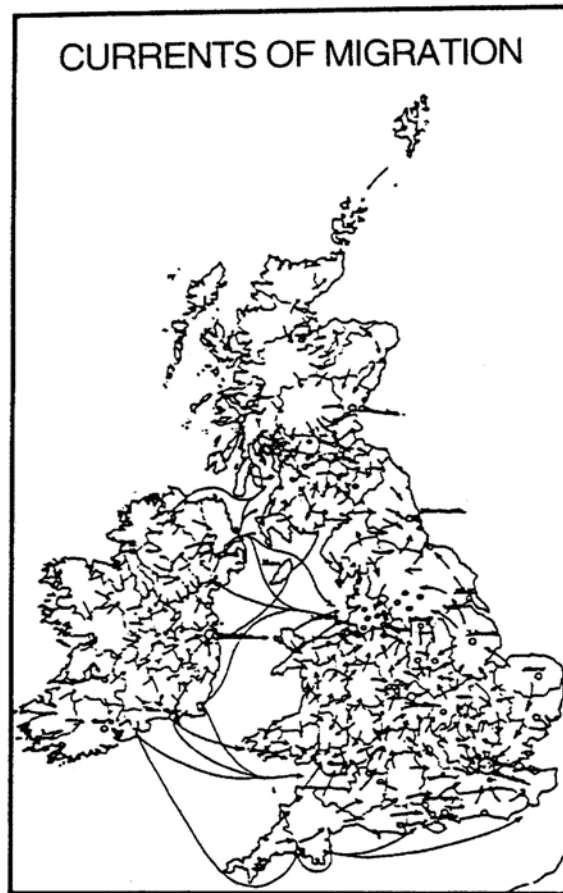


Table 1.

Thresholding of a Migration Table. For the 48 by 48 table of 1975 to 1980 migrations between states of the coterminous United States the following results are obtained:

	Total Moves	Above Mean	% Above Mean	% of migrants above the mean
Bidirectional $M_{I,J}$	2250	535	23.7%	78%
Gross $M_{I,J} + M_{J,I}$	1128	280	24.8%	78.6%
Net $M_{I,J} - M_{J,I}$	1127*	228	20.2%	81.8%

*One would expect 1128 here but the Arizona to New Hampshire movement exactly equals that in the opposite direction in the published table, yielding a net movement of zero.