**Title**
Tobler's FlowMapper



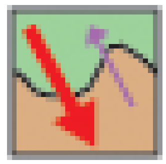
**Author**
Tobler, Waldo

**Publication Date**
2003

Center for Spatially Integrated Social Science

back

# Tobler's Flow Mapper



## Background

Geographical movement is of crucial importance. This is because much change in the world is due to movement; the movement of people, ideas, money, 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:

- Flow Mapper Tutorial, Tobler 2004 4.7 MB **Updated!** (7-19-05)
- Movement Mapping, Tobler 2003 2.5 MB
- Experiments in Migration Mapping by Computer, Tobler 2003 500 kb

**About Flow Mapper**
In 2003 CSISS supported an effort to produce an interactive flow mapping program. The result is an updated Windows-based version of a program originally designed and programmed by Waldo Tobler in 1987. Tobler's original application was updated by David Jones using Microsoft Visual Basic .Net and SVG (Scaleable Vector Graphics) for map rendering. It requires as input locational coordinates and a table of interaction between places. Additional input may include place names for the location coordinates and a file of boundary coordinates (the background map). The user has several menu options for producing a map. The program allows for the production of a total movement map shown by volume-scaled bands, net movement given by scaled arrows, or simultaneous two-way moves.

**Flow Mapper Functionality in ArcGIS** Flow Data Model Tools consist of several ArcGIS 9.x VBA macros. The prototype software integrates the functionality of Flow Mapper into ArcGIS, and allows ArcGIS interaction with the Flow Data Model. Source code and further information.

**Flow Mapper Requirements**

- Microsoft Windows 98SE, ME, 2000 or XP
- Microsoft Dot.Net Framework Installed
- Microsoft Internet Explorer (Required to display maps properly)
- Scalable Vector Graphics Support for Internet Explorer: Adobe SVG Plugin 3.x or higher
- C:\temp folder

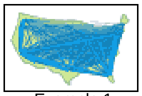
**Installing Flow Mapper**

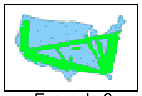
1. Remove any existing version of Flow Mapper (*Start* > *Control Panel* > *Add/Remove Programs*).
2. Verify that your operating system meets the requirements above.
3. Install the .Net Framework if necessary. Many newer Windows XP machines already come pre-installed with it. Go to *Start* > *Control Panel*, the .Net Framework Management Icon will be visible if it is installed (you may need to look in *Administrative Tools*.) If not visible you will need to **Download and install the .Net Framework** (21MB)
4. If you you have not done so previously, install Adobe SVG Plugin. If you are unsure, install it again. **Download and install Adobe SVG Viewer** (2.4MB)
5. Make sure C:\Temp (c:\temp) exists, the application uses this directory to write temporary files.
6. **Download and install Flow Mapper** (22MB), unzip the files o temporary directory and run setup.exe. Flow mapper

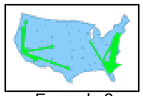
will be installed and a shortcut will be placed in the *Start* Menu under *Flow Mapper* and on your desktop. The documentation, and *Data_Sets* will be automatically installed and links placed in the *Flow Mapper* folder.

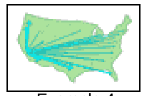
7. After installing Flow Mapper, **Download the Documents and Data update** (8.5MB, 7-19-05) and unzip it to your installation folder (the default folder is C:\Program Files\Tobler\Flow Mapper).

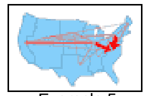
**Examples**


Example 1

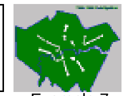

Example 2

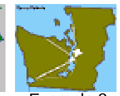

Example 3


Example 4


Example 5


Example 6


Example 7


Example 8


Example 9

**User Contributions**

Boundary Files from GeoDa to FlowMapper

Please send questions and comments to Waldo Tobler

# CSISS.org/Spatial Tools/Tobler's Flow Mapper


Center for Spatially Integrated Social Science
Spatial Resources for the Social Sciences

**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.

**CSISS News**

**Core Programs**
CSISS has six research initiatives and a professional development program for undergraduate instructors.

**Learning Resources**
These introductory materials include **CSISS Classics** and **select video clips** from the CSISS summer workshops.

**Spatial Resources**
CSISS has compiled e-journals, bibliographies, and other spatial resources for the social sciences.

**Spatial Tools**
- Spatial Tools Search Engine
- Select Tools
- Links to Portals
- GeoDa™
- Tobler's Flow Mapper

**Search Engines**
Try out one of our custom search engines to find spatial analysis resources on the Internet.

**CSISS Events**
Here's where you'll find information and registration for workshops, conferences and specialist meetings.

**Community Center**
Join the forums, or if your organization relates to our mission and goals, register as a CSISS affiliate.

CSISS presentations, news, personnel, and sitemap. Our Strategic Plan and Annual Reports are also found here.

Core Programs | Learning Resources | Spatial Resources | Spatial Tools | Search Engines | CSISS Events | Community Center | About CSISS
Site Map | Site Search | Contact CSISS | Plug-ins | Privacy Policy | Site Credits | Home

# Tobler's Flow Mapper

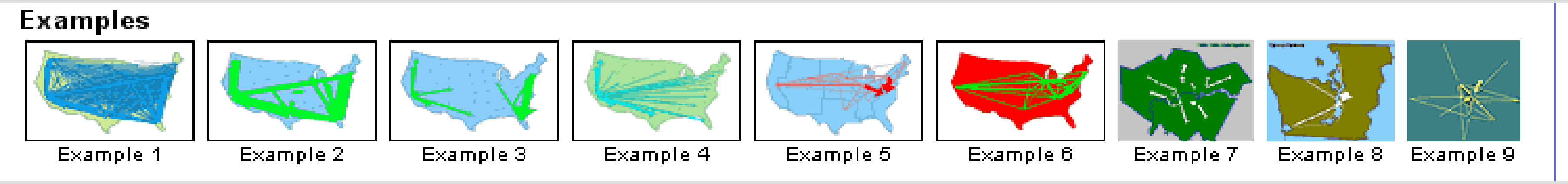
## Background

Geographical movement is of crucial importance. This is because much change in the world is due to movement; the movement of people, ideas, 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 Tutorial, Tobler 2004 4.1 mb - Updated 6-6-05
- Movement Mapping, Tobler 2003 2.5 MB
- Experiments in Migration Mapping by Computer, Tobler 1987, 500 kb

## 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 basnds, net movement given by scaled arrows, or simultaneous two-way moves.



# Flow Mapper Requirements

- Microsoft Windows 98SE, ME, 2000, XP
- Microsoft Dot.Net Framework installed
- Microsoft Internet Explorer
- Scalable Vector Graphics support for Internet Explorer. Adobe SVG plugin 3.x or higher
- C:\temp folder

# Installing Flow Mapper

Remove any existing version of Flow Mapper (*Control Panel > add/remove*).

1. Verify that your operating system is within the requiremts.
2. Install Dot.Net Framework, if necessary. Go to *Start > Control Panel*. The .Net Framework Management icon will be visible if it is installed. (You may need to look in *Administrative Tools*). Many Windows XP and 2000 machines come pre-installed with this.

   Download and install .Net Framework
3. Make sure that you have Microsoft Internet Explorer installed (required to display maps).
4. Install Adobe SVG plugin. If you are not sure that you have it installedt, install it again.

   Download and install Adobe SVG Viewer
5. Make sure that the C:\temp folder exists. It's needed for temporary files.
6. Download and install the Flow Mapper program from the CSISS.org web site. Store it to a directory of your choice ("C:\program files\tobler\flow mapper" is suggested). A shortcut to the Flow Mapper program will appear on your desktop. The program is Flow Mapper.exe Documentation and Data_Sets will also be with the program in this folder. Replace this Data_Sets folder with the newer update from the csiss.org site.

# 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.

# Flow Mapper Tutorial

## Parts I, II, III

To be used in conjunction with the Flow Mapper program developed by Waldo Tobler & David Jones and available for download at CSISS.org/tools

Much of Computer

# Cartography is a Dot-to Dot

Replace the dots by coordinates

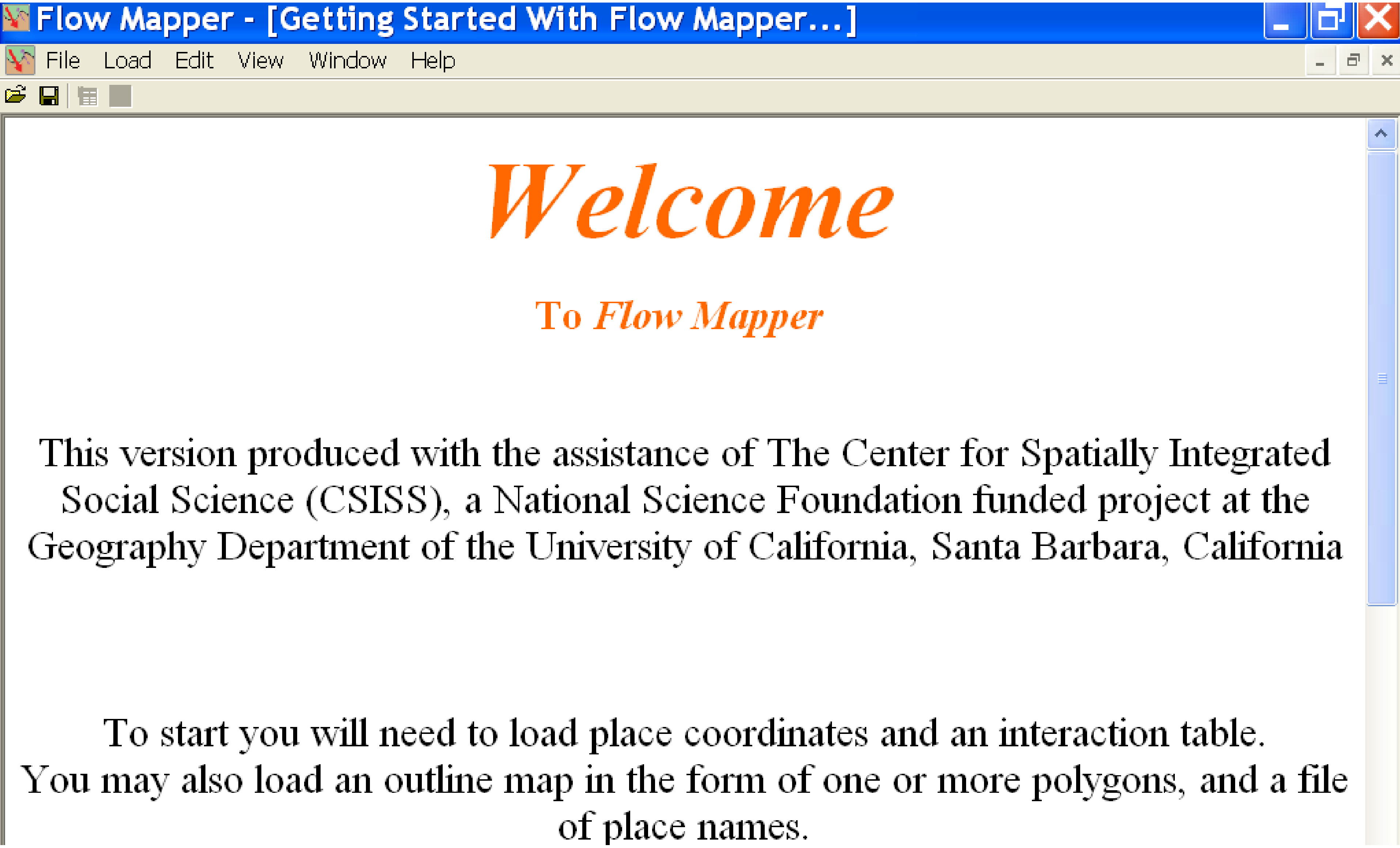

# Tutorial Part I

## General Instructions

## Getting started

The help file also contains instructions

The help file has good instructions & hints. After looking at the help files you can view this Welcome screen. To close it click on the smaller (the lower one) of the two x's in the upper right corner. Then go through this tutorial and start using the program.



# The first steps

You will need to have available coordinates.

And an interaction table, or an origin - destination list.

The order in which you load these is not important.

I usually load a background map first to make certain that I am working with the correct area, as in a subsequent view.

Then I load the place names and locations,

then the interaction table.

# If you have an origin - destination list instead of a complete flow table

Then look under

data_sets\programs\moves\input help programs

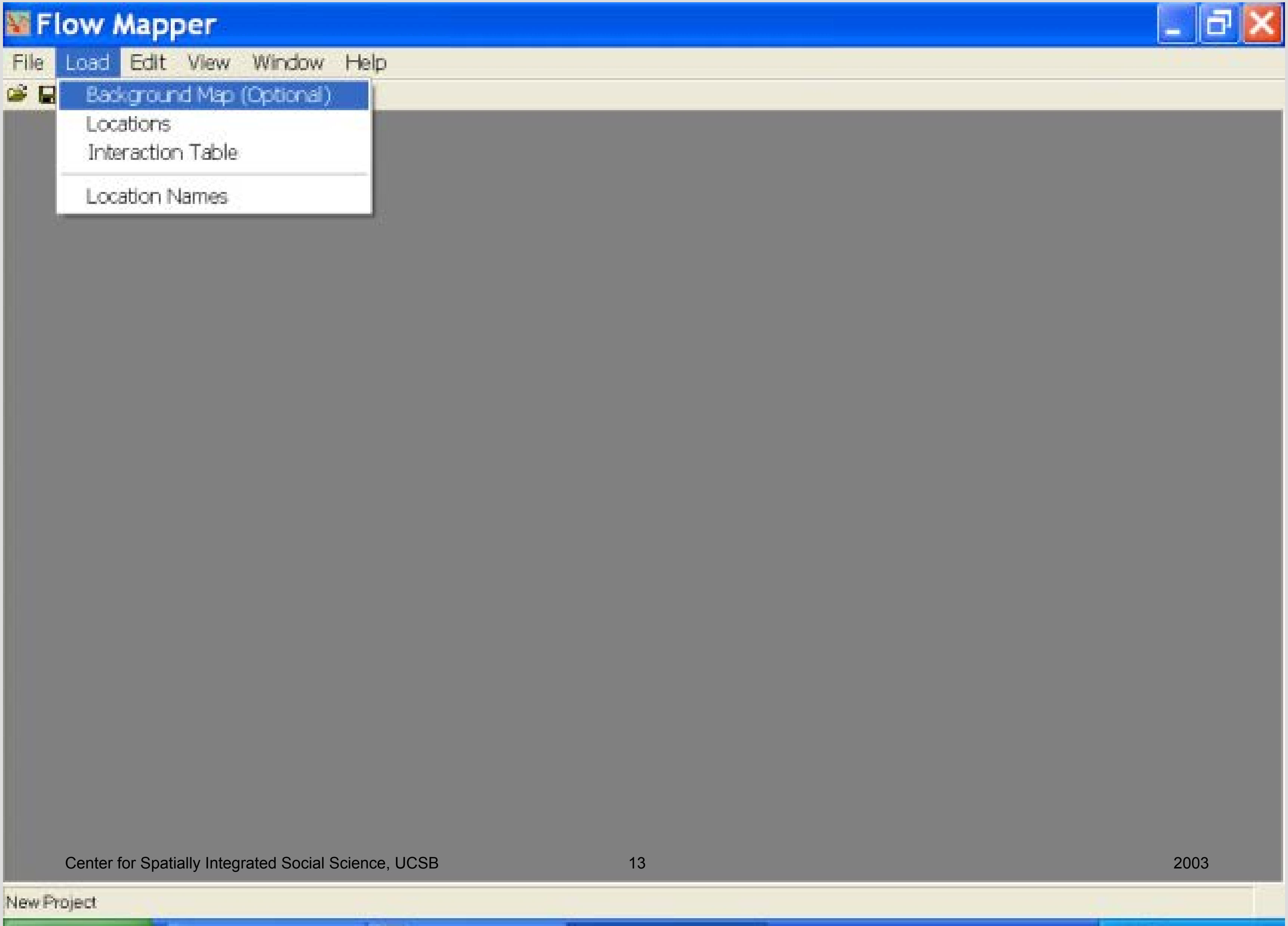
and choose the appropriate program to convert your data.

(do not use tab delimited lists - only comma or space delimited will work)

The program should convert your list to a table in the correct form for use in Flow Mapper.

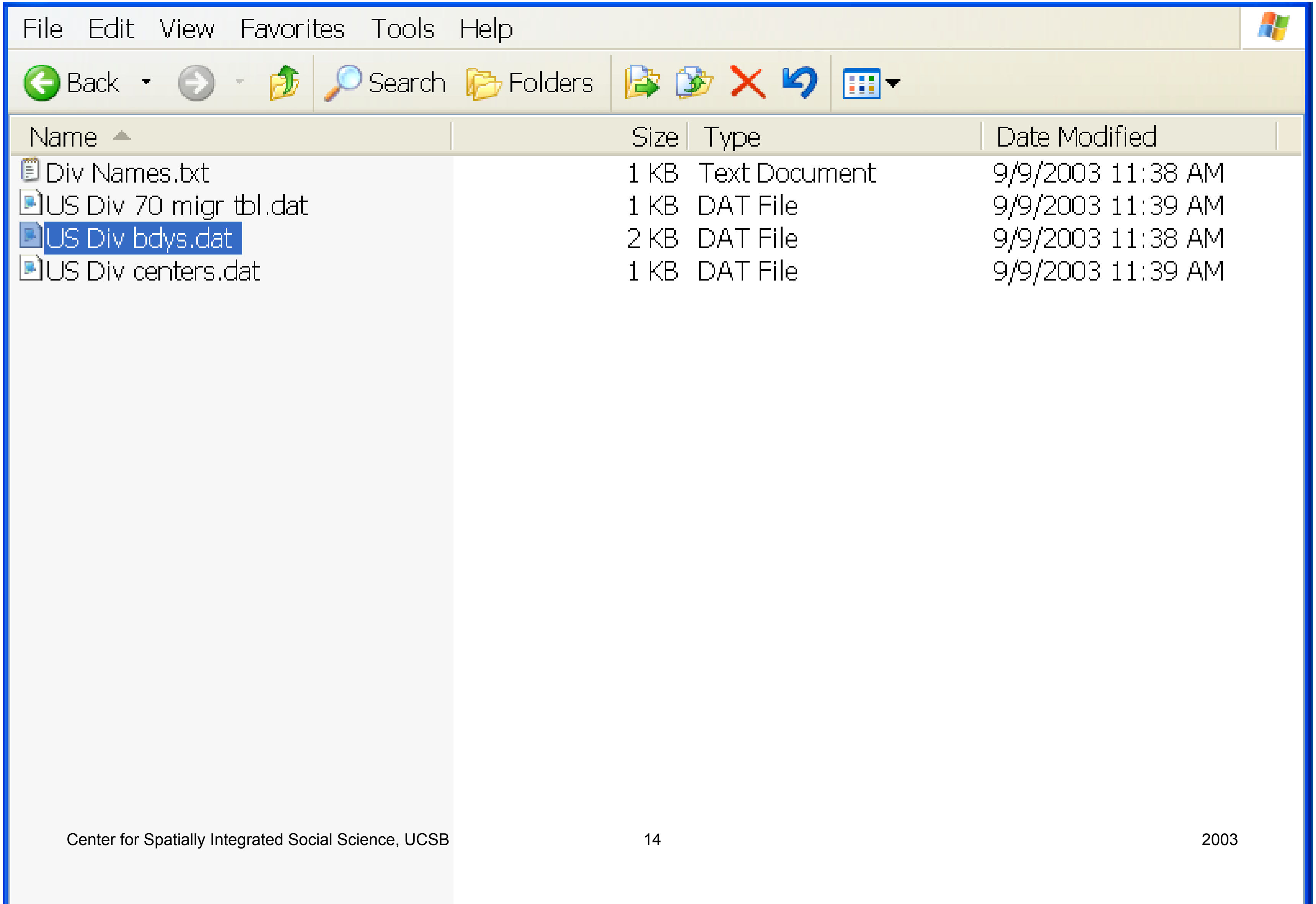
In oder to do this you will need to exit the Flow Mapper program, convert and save the data, and then restart the Flow Mapper program using the movment table that was created..

# Load a background map



# Locate the file containing the background map

Then load it. Or look at it to see the simple format.



# Boundary Coordinates

Number of points, counter-clockwise order, first-last, arbitrary units

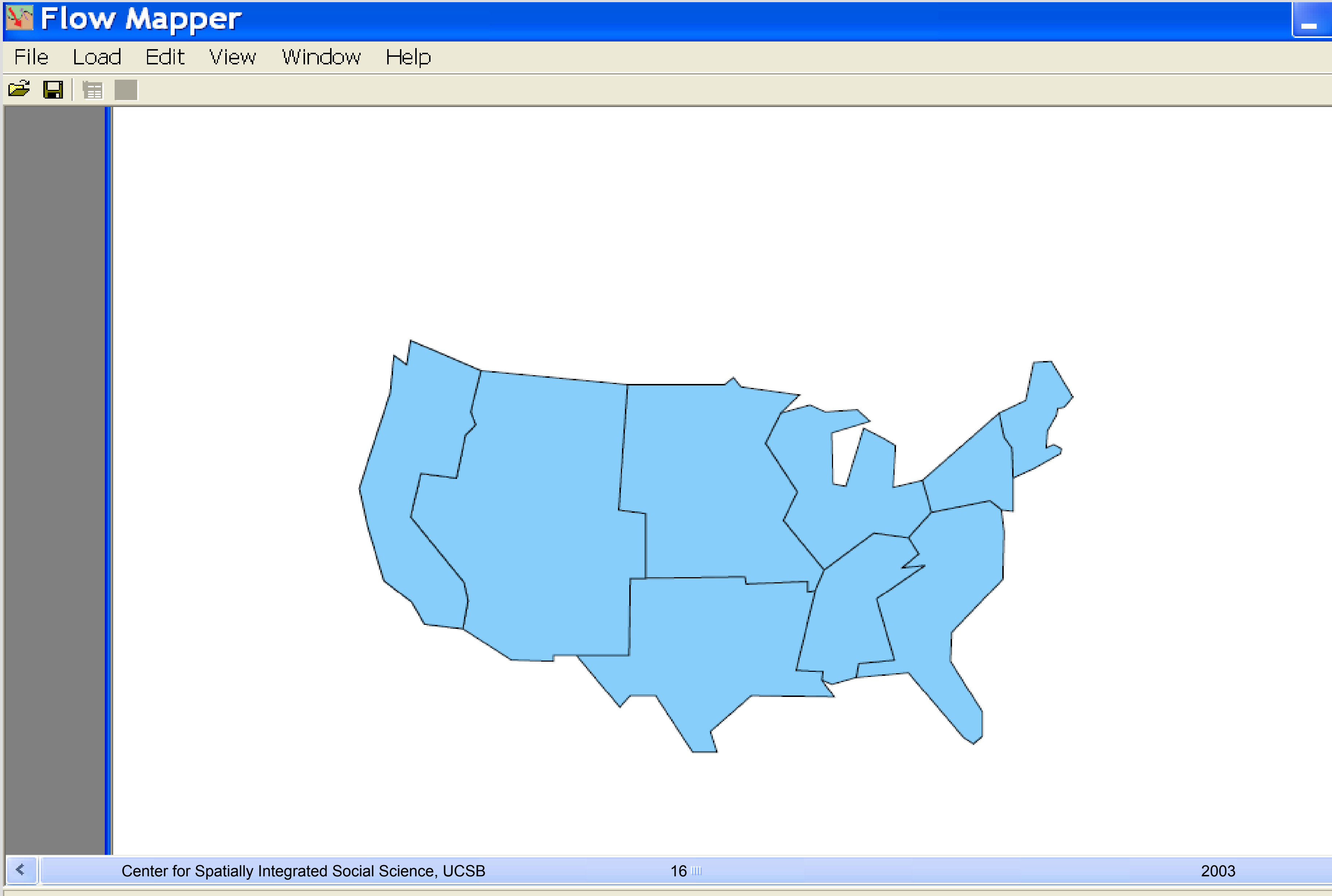
US Div bdys.dat - Notepad

File Edit Format View Help

```
19
65.9 38.5
66.3 36.4
66.9 35.5
67.0 32.9
68.6 33.7
70.7 35.0
70.8 35.4
70.2 35.8
69.6 35.5
69.7 37.0
70.4 38.3
70.5 38.9
71.0 39.0
71.7 39.9
71.2 40.8
70.0 43.0
68.6 42.9
68.0 39.6
65.9 38.5
10
59.9 32.7
60.6 29.9
65.2 30.9
66.1 30.1
67.0 30.0
67.0 32.9
66.9 35.5
```

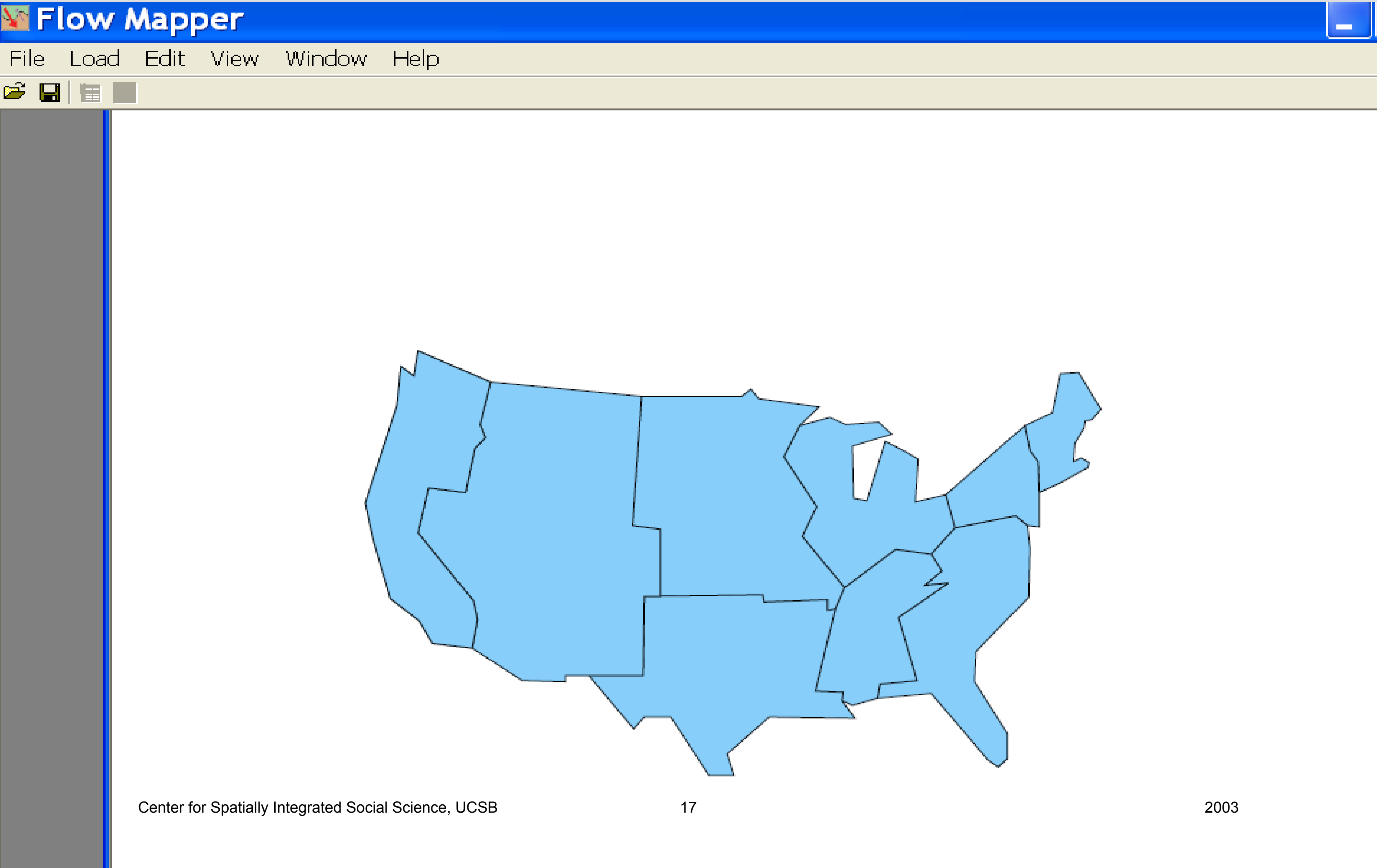
# Background map selected





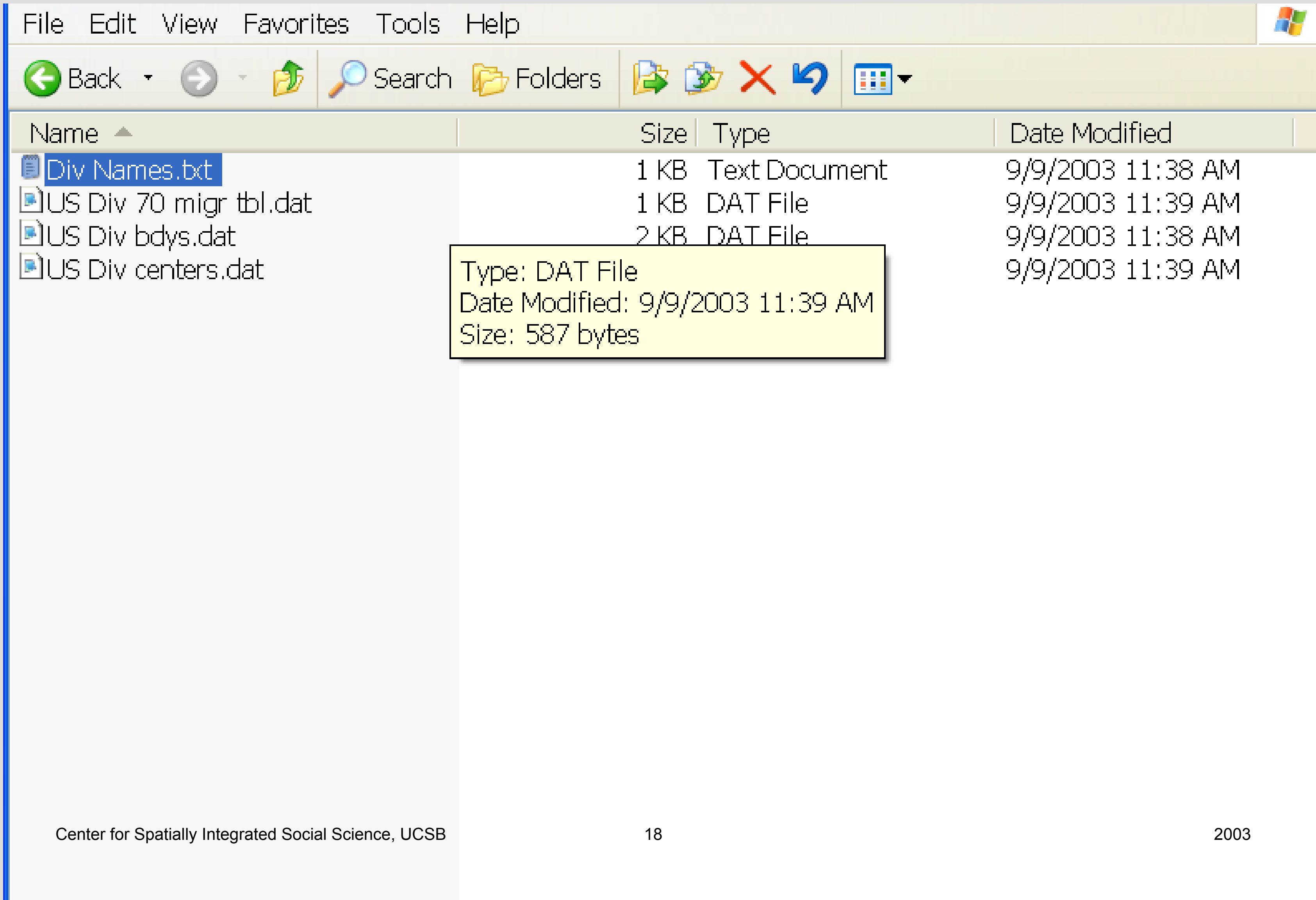
# Then load location names

## If you have them

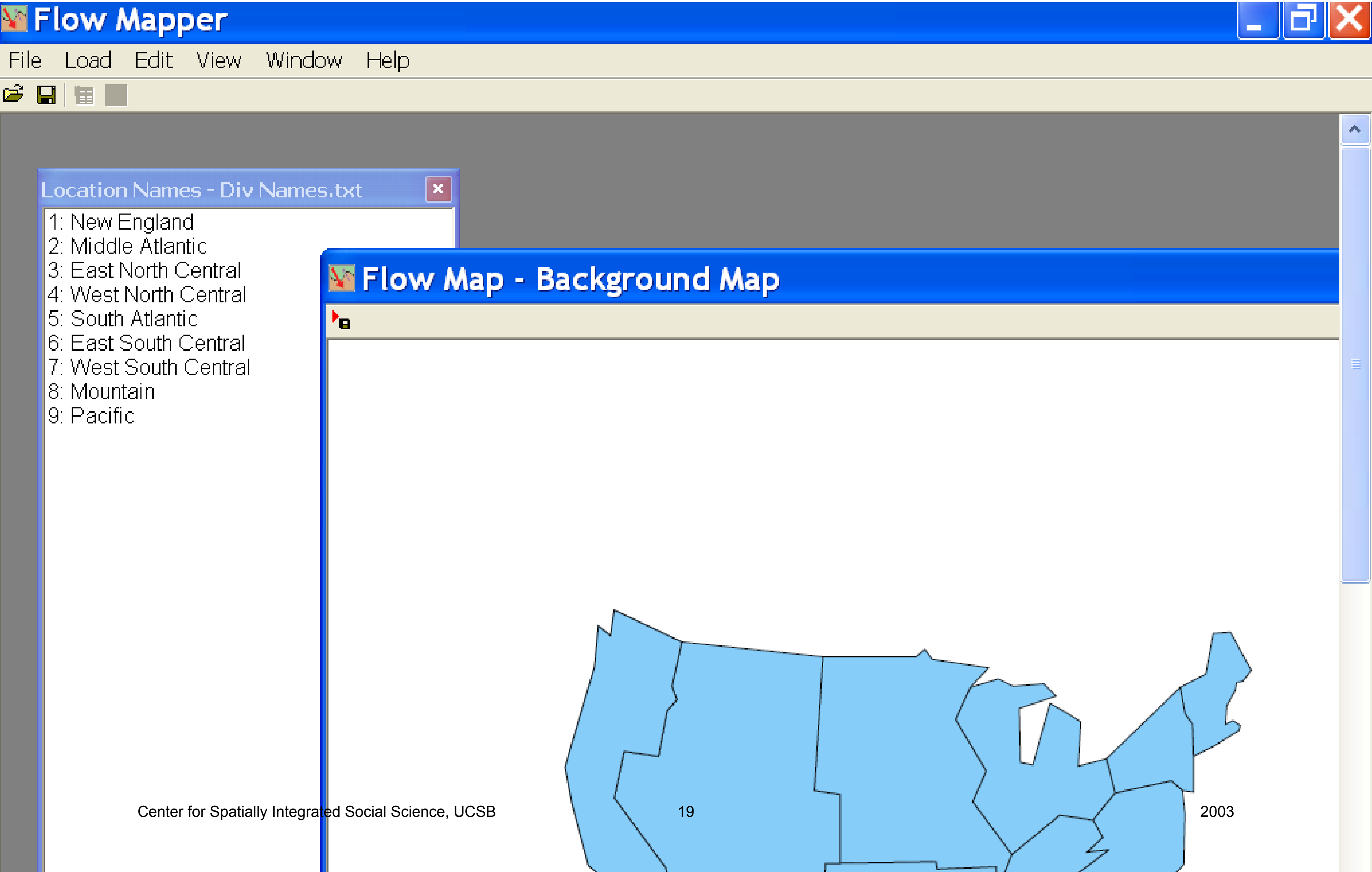


# Select location names file

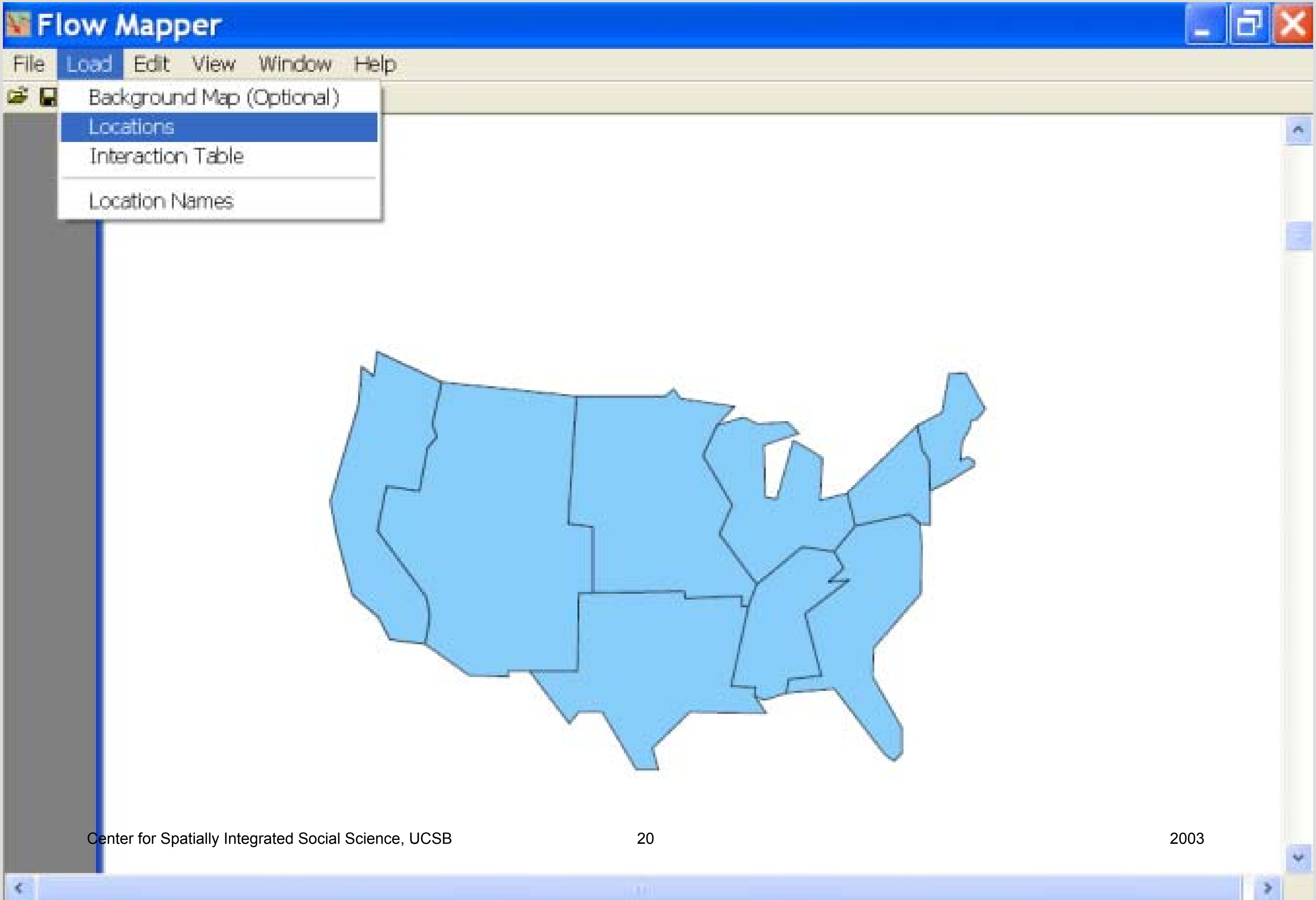
Then load it.



# Location names selected and loaded.

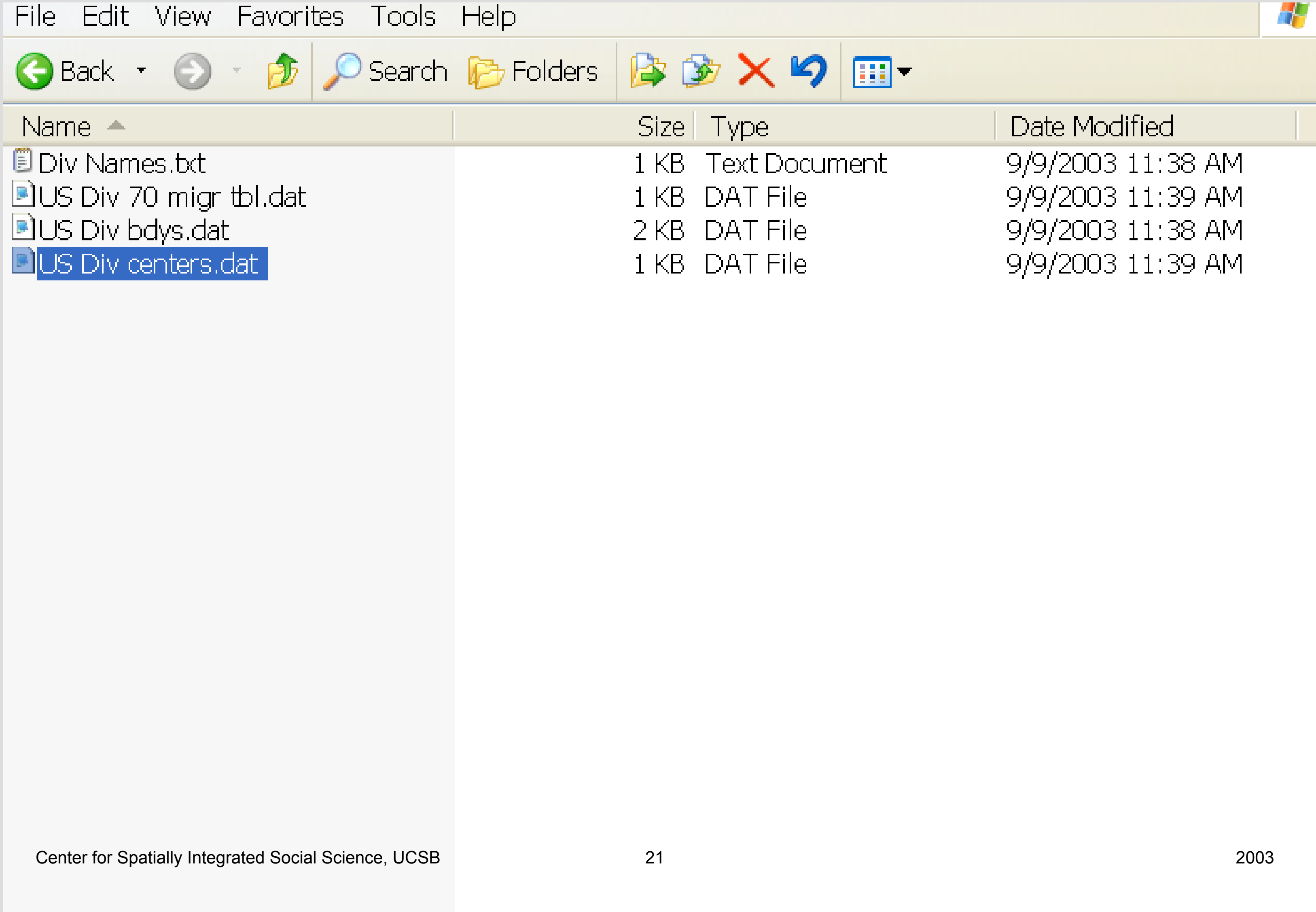


# Next: Load locations



# Select the centroid coordinate file

Then load it,



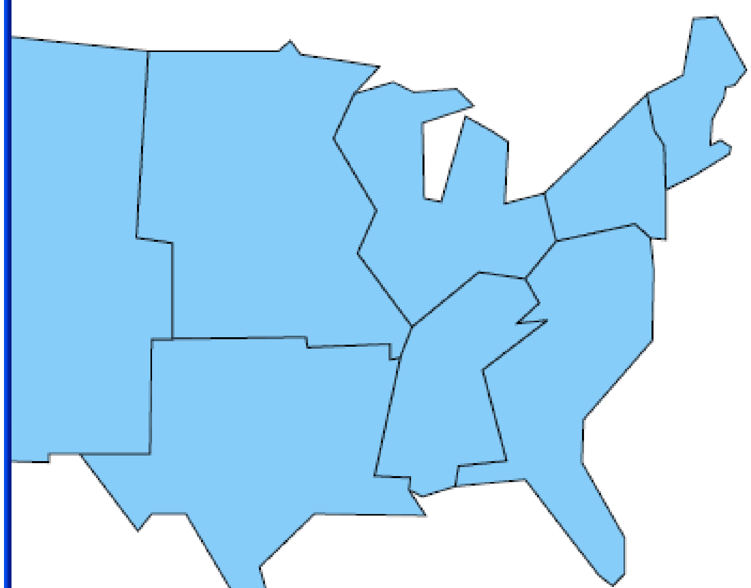
# Locations (centroids) loaded

Flow Mapper

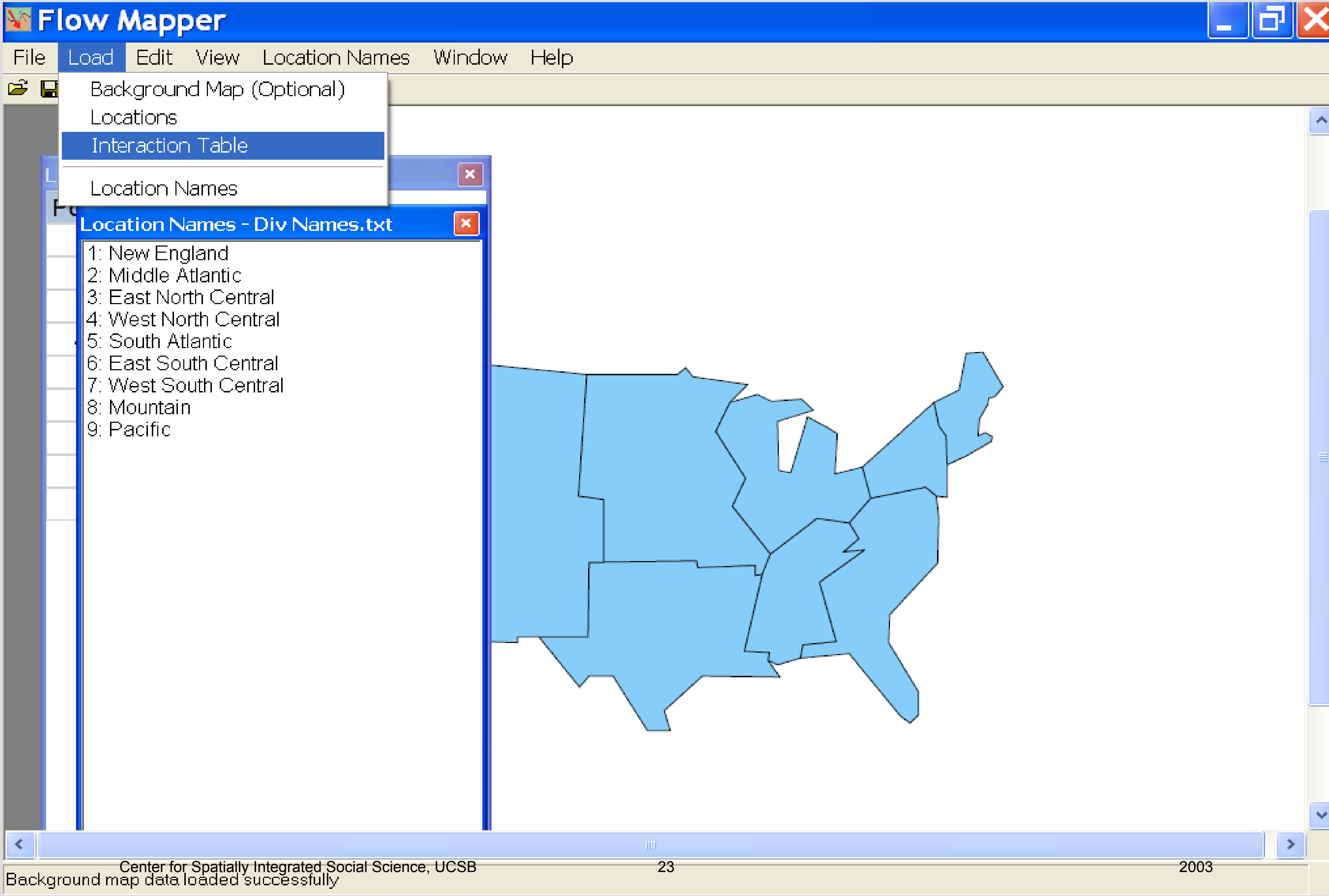
File Load Edit View Locations Window Help

Locations - US Div centers.dat

| Point | X | Y |
|---|---|---|
| 1 | 70 | 40 |
| 2 | 64 | 34 |
| 3 | 53 | 30 |
| 4 | 43 | 35 |
| 5 | 62 | 26 |
| 6 | 55 | 23 |
| 7 | 46 | 19 |
| 8 | 31 | 33 |
| 9 | 19 | 30 |

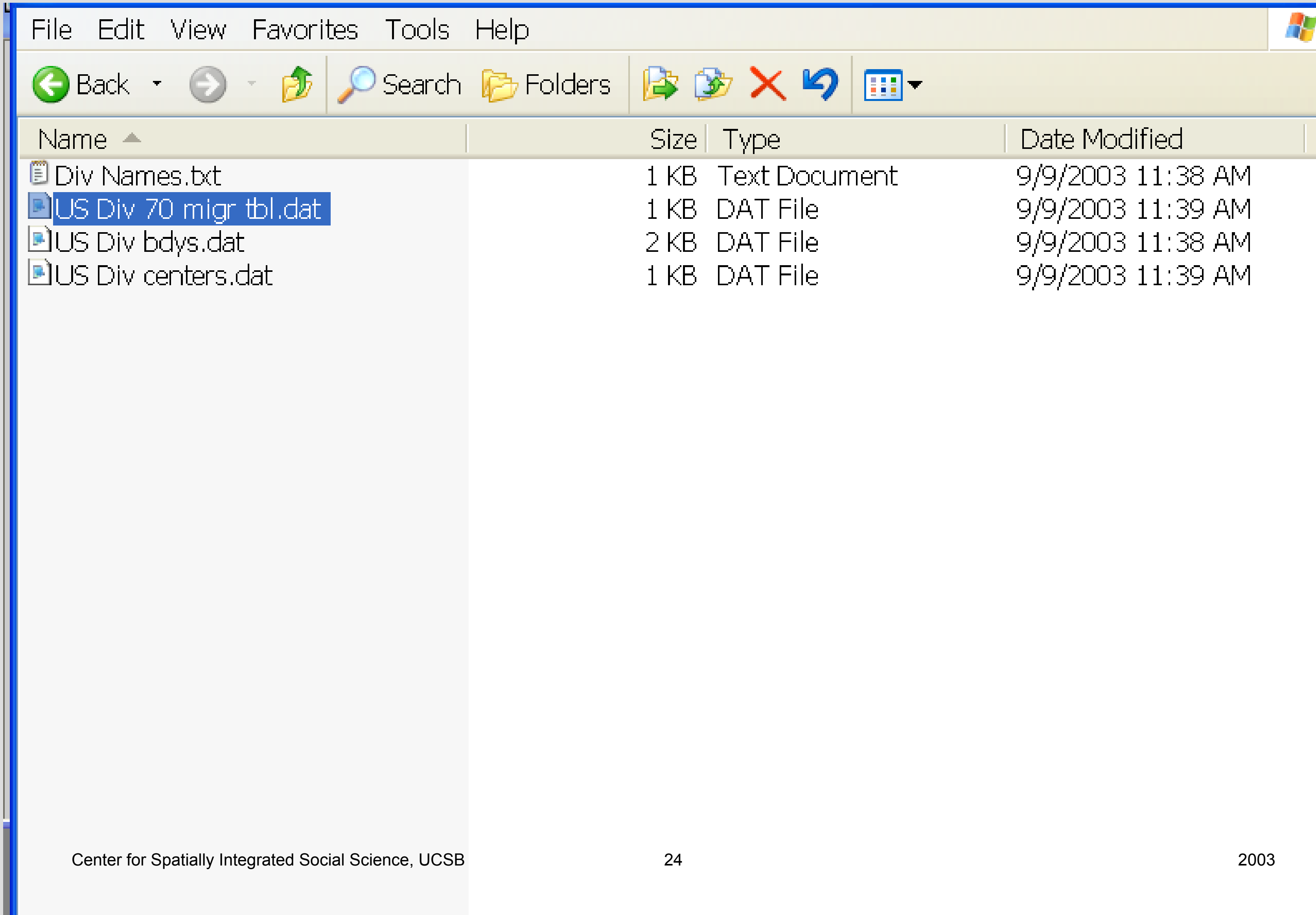


# Load interaction table

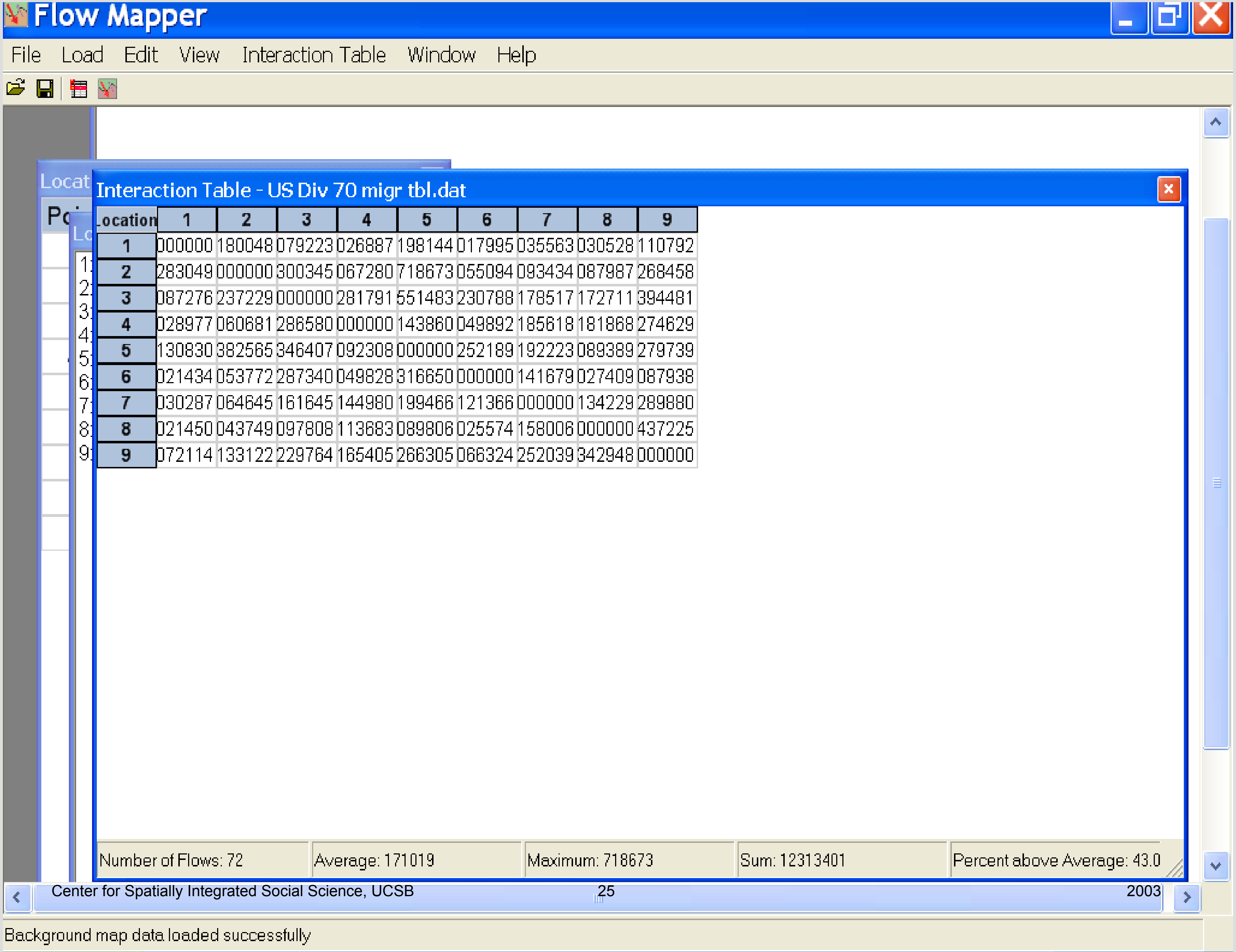


# Select interaction table

Then load it

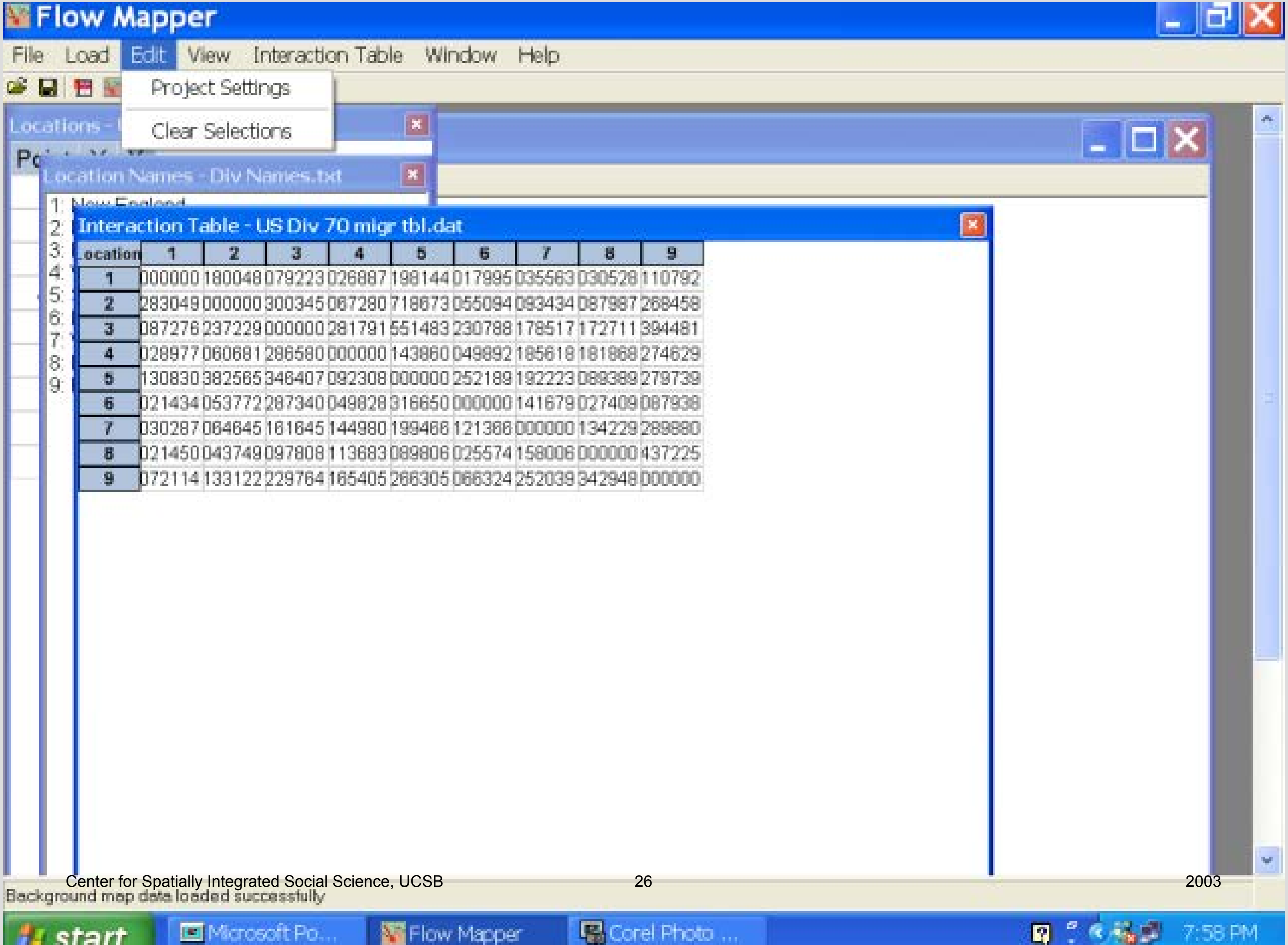


# Interaction table loaded

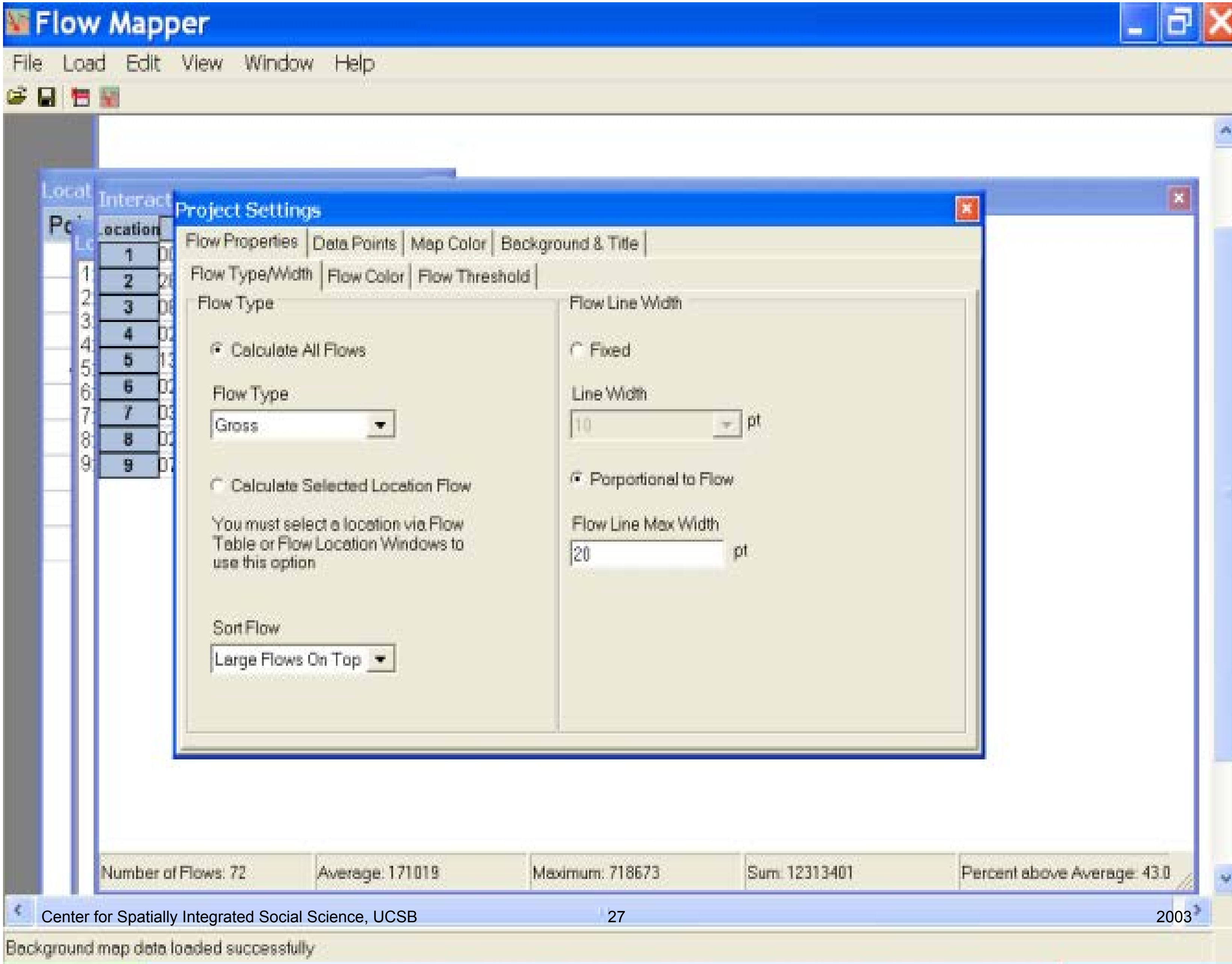



| Location | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 000000 | 180048 | 079223 | 026887 | 198144 | 017995 | 035563 | 030528 | 110792 |
| 2 | 283049 | 000000 | 300345 | 067280 | 718673 | 055094 | 093434 | 087987 | 268458 |
| 3 | 087276 | 237229 | 000000 | 281791 | 551483 | 230788 | 178517 | 172711 | 394481 |
| 4 | 028977 | 060681 | 286580 | 000000 | 143860 | 049892 | 185618 | 181868 | 274629 |
| 5 | 130830 | 382565 | 346407 | 092308 | 000000 | 252189 | 192223 | 089389 | 279739 |
| 6 | 021434 | 053772 | 287340 | 049828 | 316650 | 000000 | 141679 | 027409 | 087938 |
| 7 | 030287 | 064645 | 161645 | 144980 | 199466 | 121366 | 000000 | 134229 | 289880 |
| 8 | 021450 | 043749 | 097808 | 113683 | 089806 | 025574 | 158006 | 000000 | 437225 |
| 9 | 072114 | 133122 | 229764 | 165405 | 266305 | 066324 | 252039 | 342948 | 000000 |

# Select EDIT from the menu



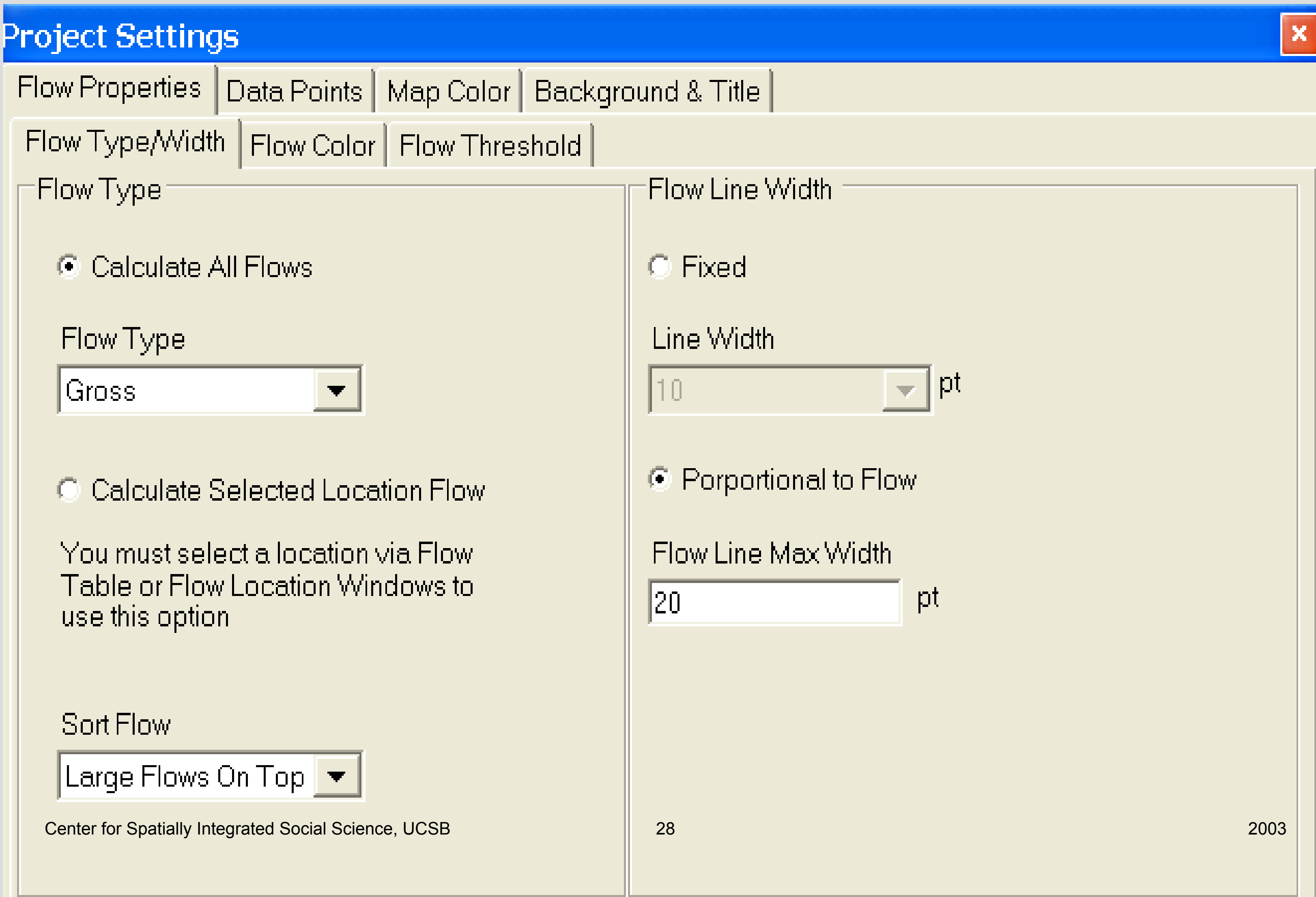
# Project settings menu selected



**Flow types:** Gross, net, two-way; single row or column

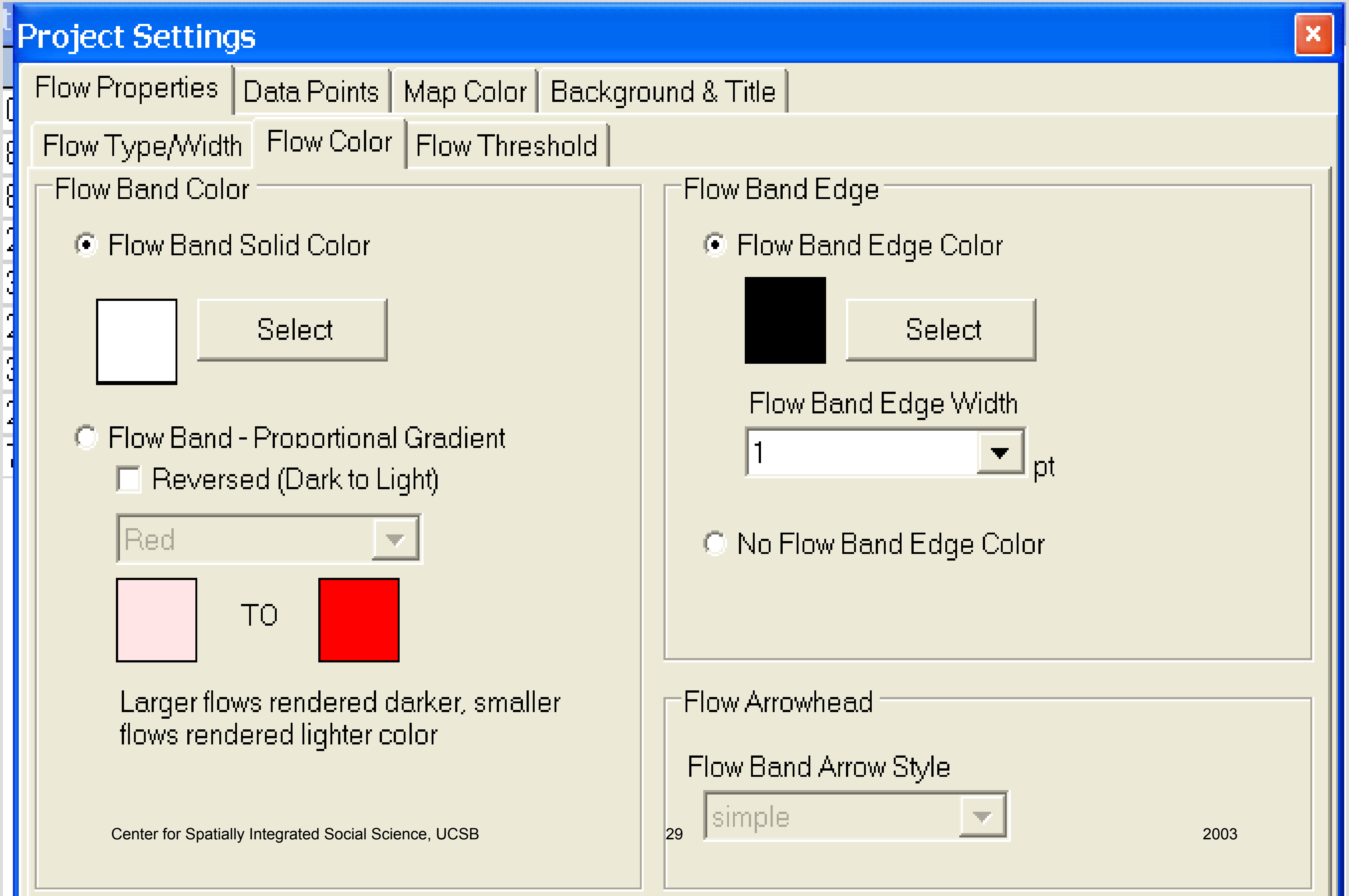
**Sort:** Large/small on top, large recommended

**Line Width:** fixed, proportional, maximum size



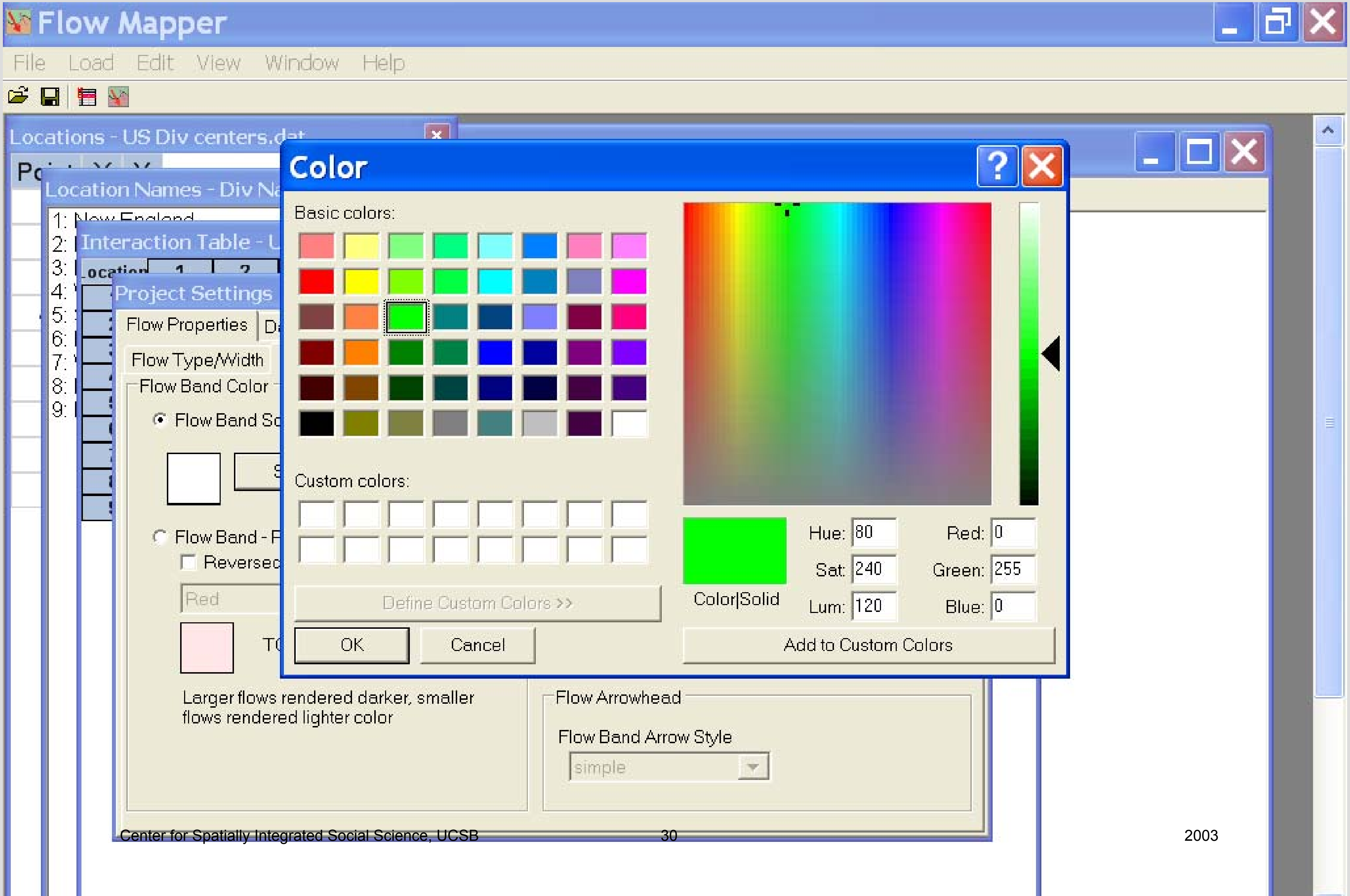
# Flow band properties

Solid color, gradient, arrowhead style, edge color options



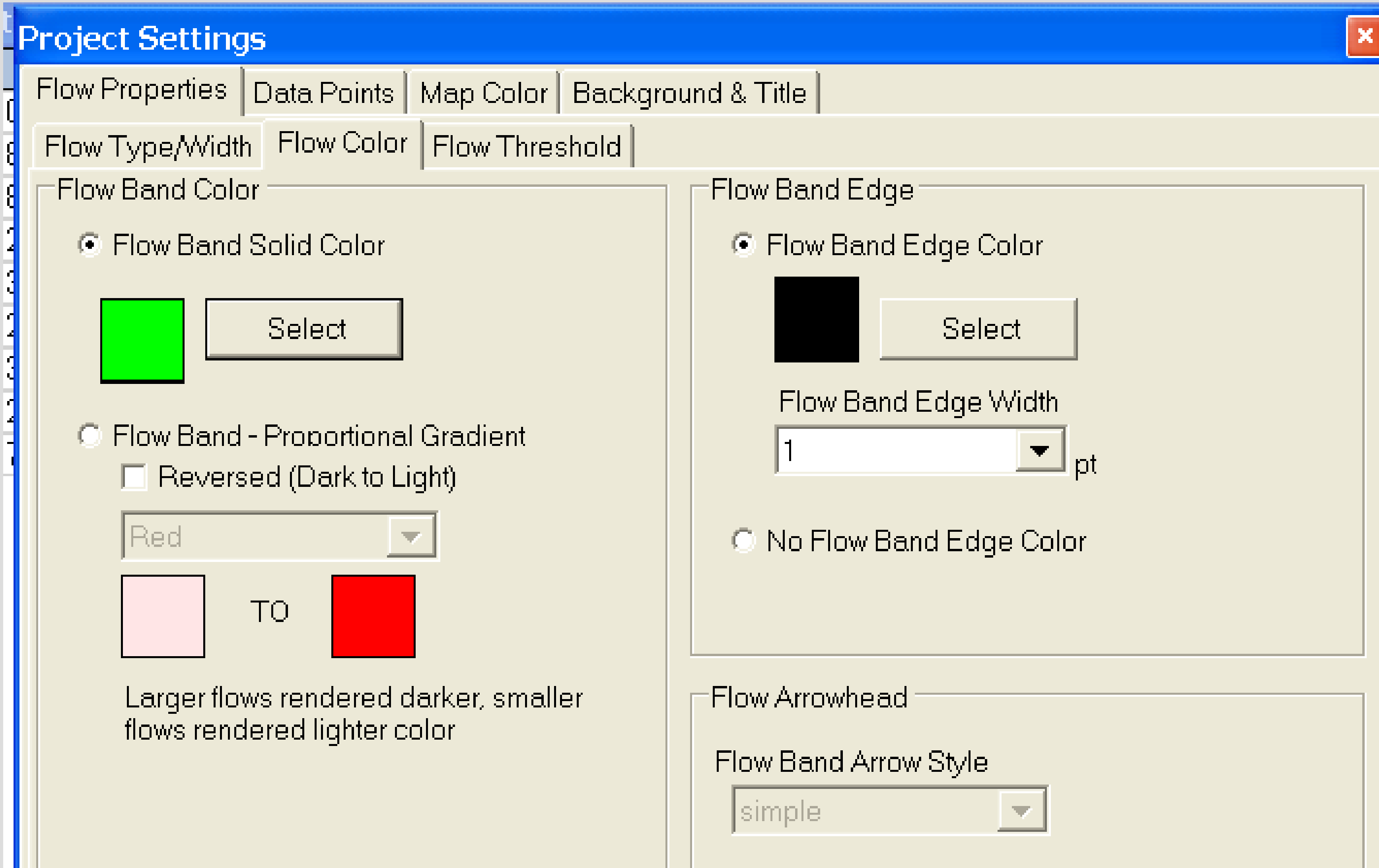
# Color selection menu

Note RGB values. Click OK after choosing.



# Color appears in the flow band box

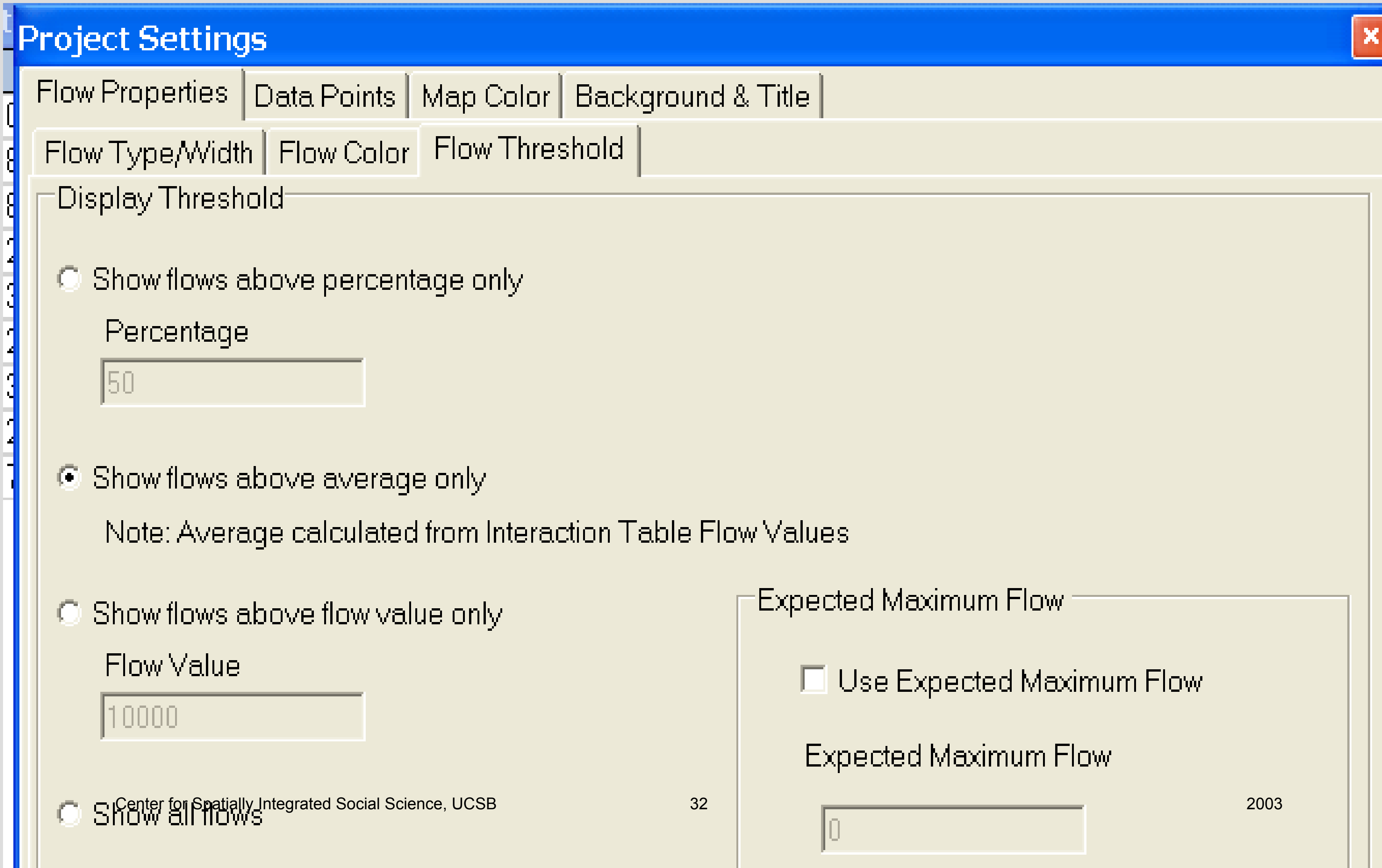
Gradient available in three colors. Edge color helpful when overlaps occur.



# Threshold

## None (all flows), average, percent, specific, maximum expected.

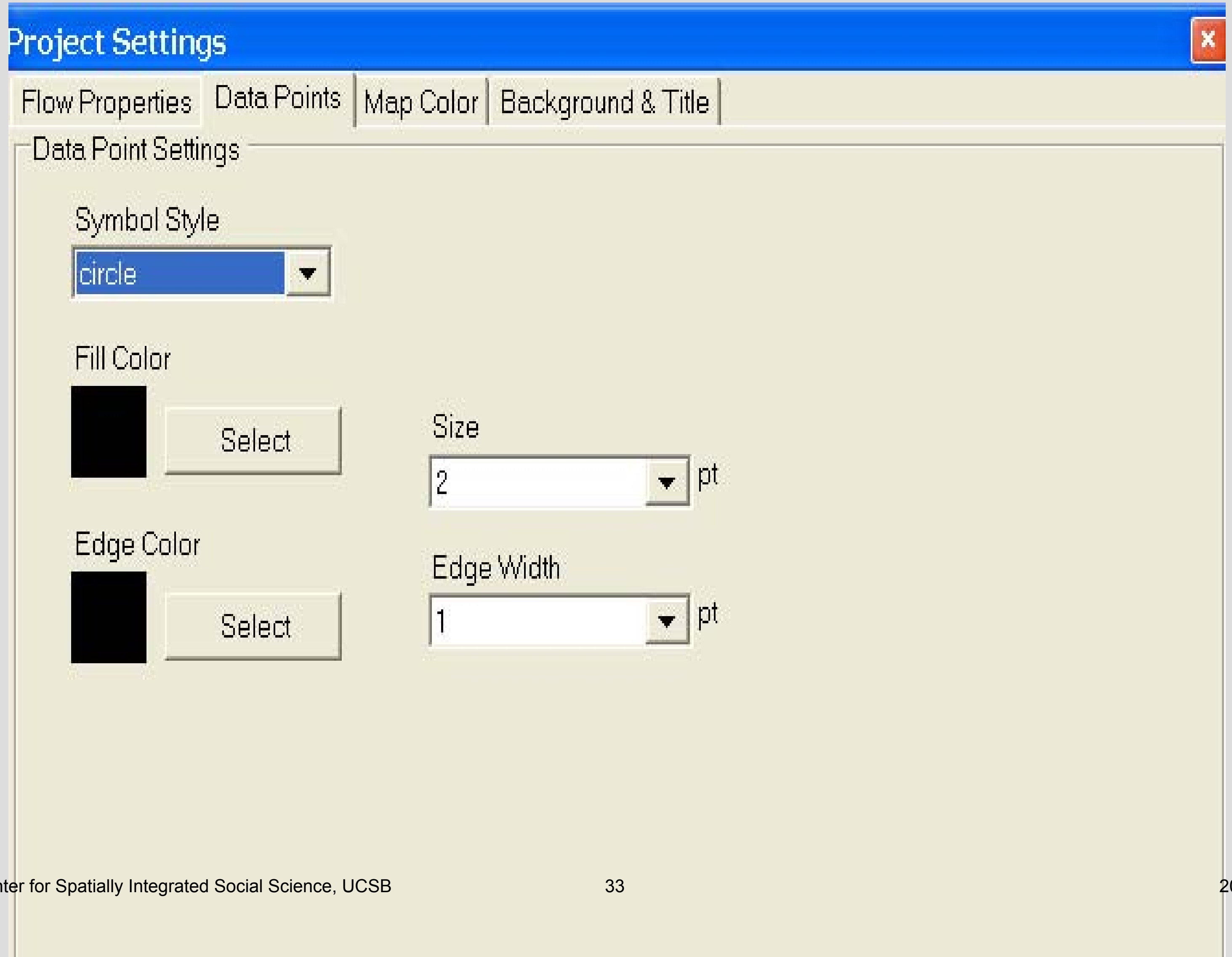
Note that the average calculated from the interaction table is of <u>all array entries</u> and that the gross flows may exceed this and net flows can be much smaller.



# Centroid point display

None, circle, square, triangle; color, size, edge

A centroid must be displayed for it to be identified when the mouse hovers on it

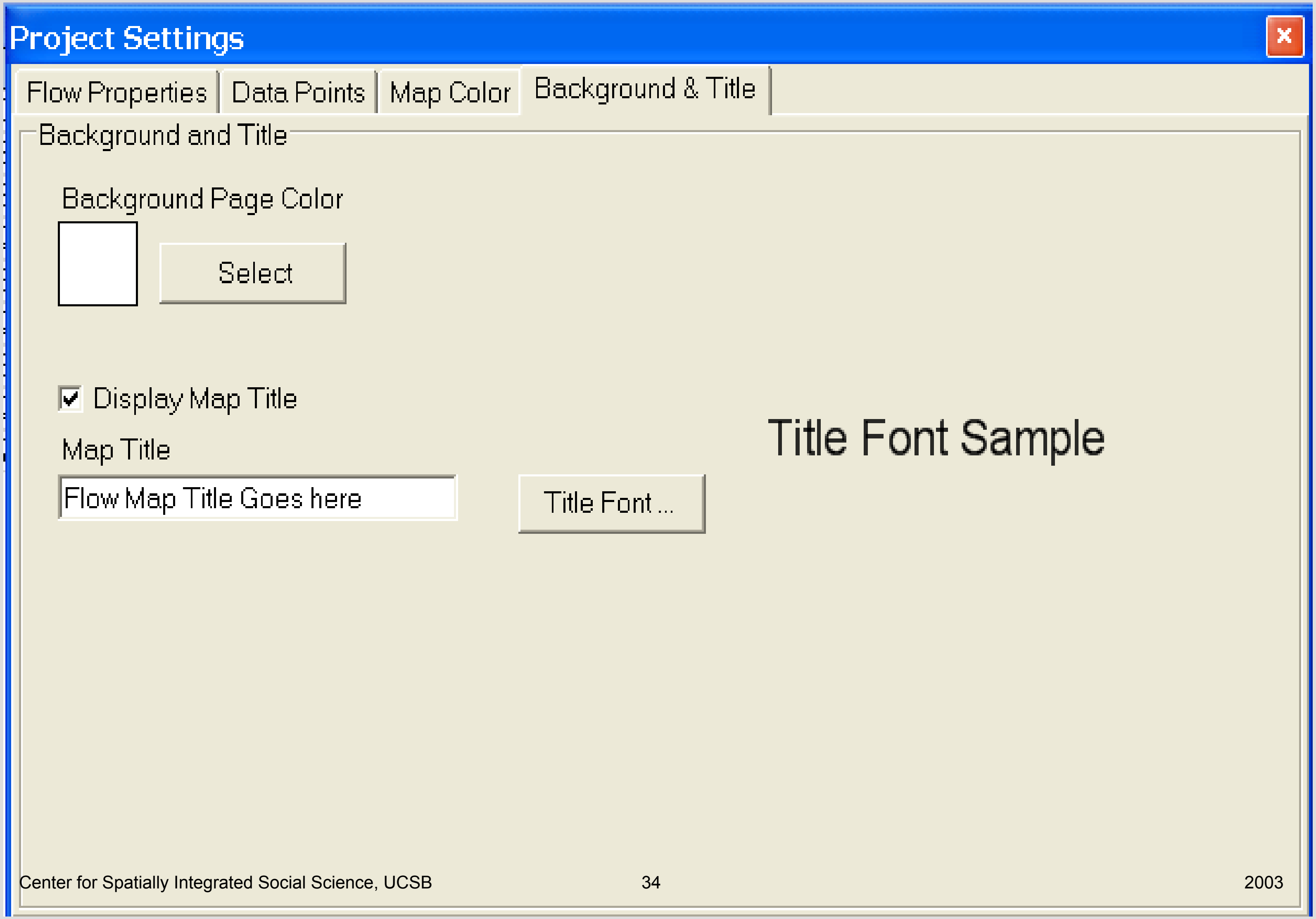


# Background color and title



One or two line title moves with FGVT keys when mouse is the on map.

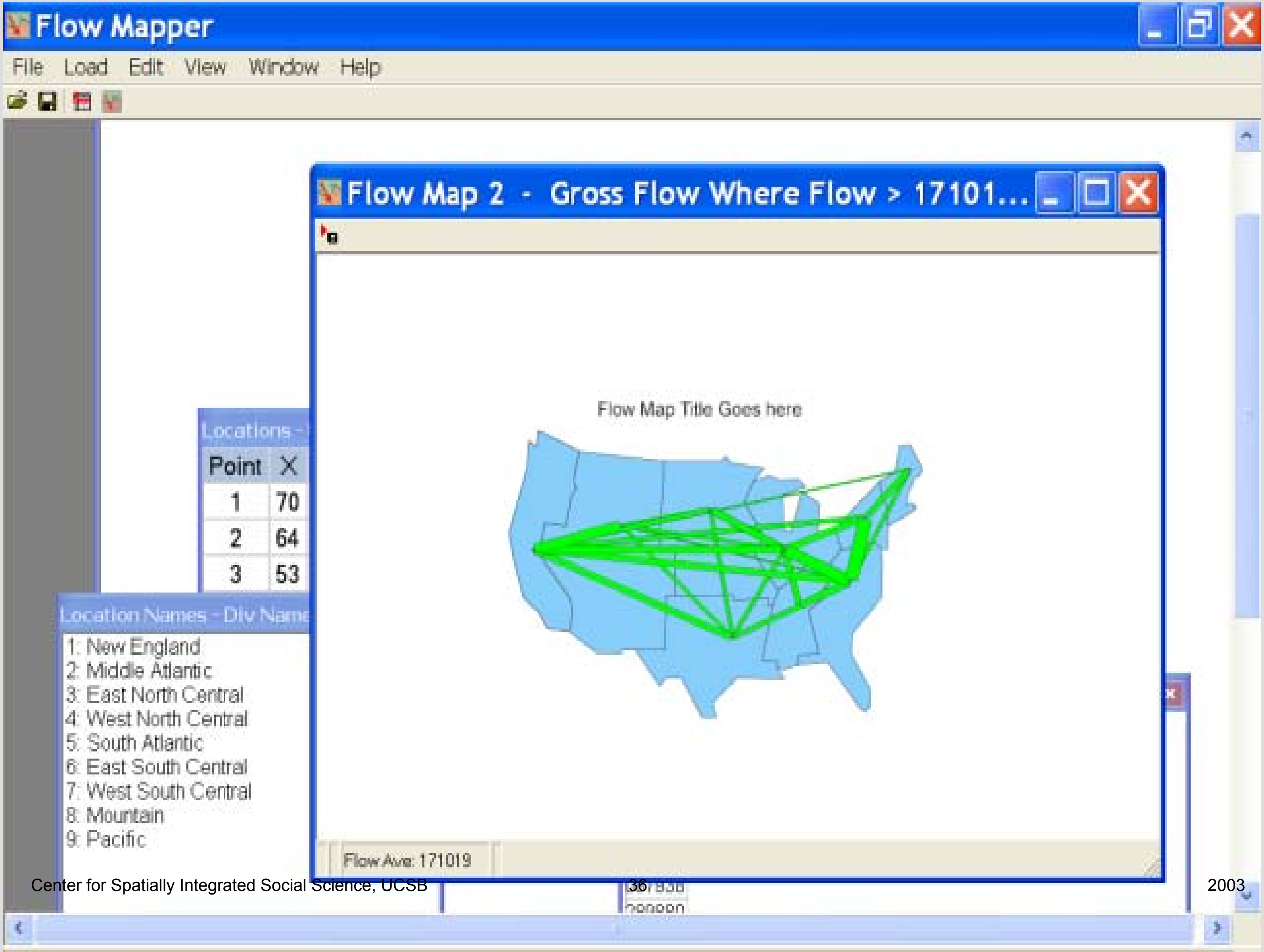
Back-slash separates title lines



# To make a map click on the rightmost icon on the second line in the upper left.

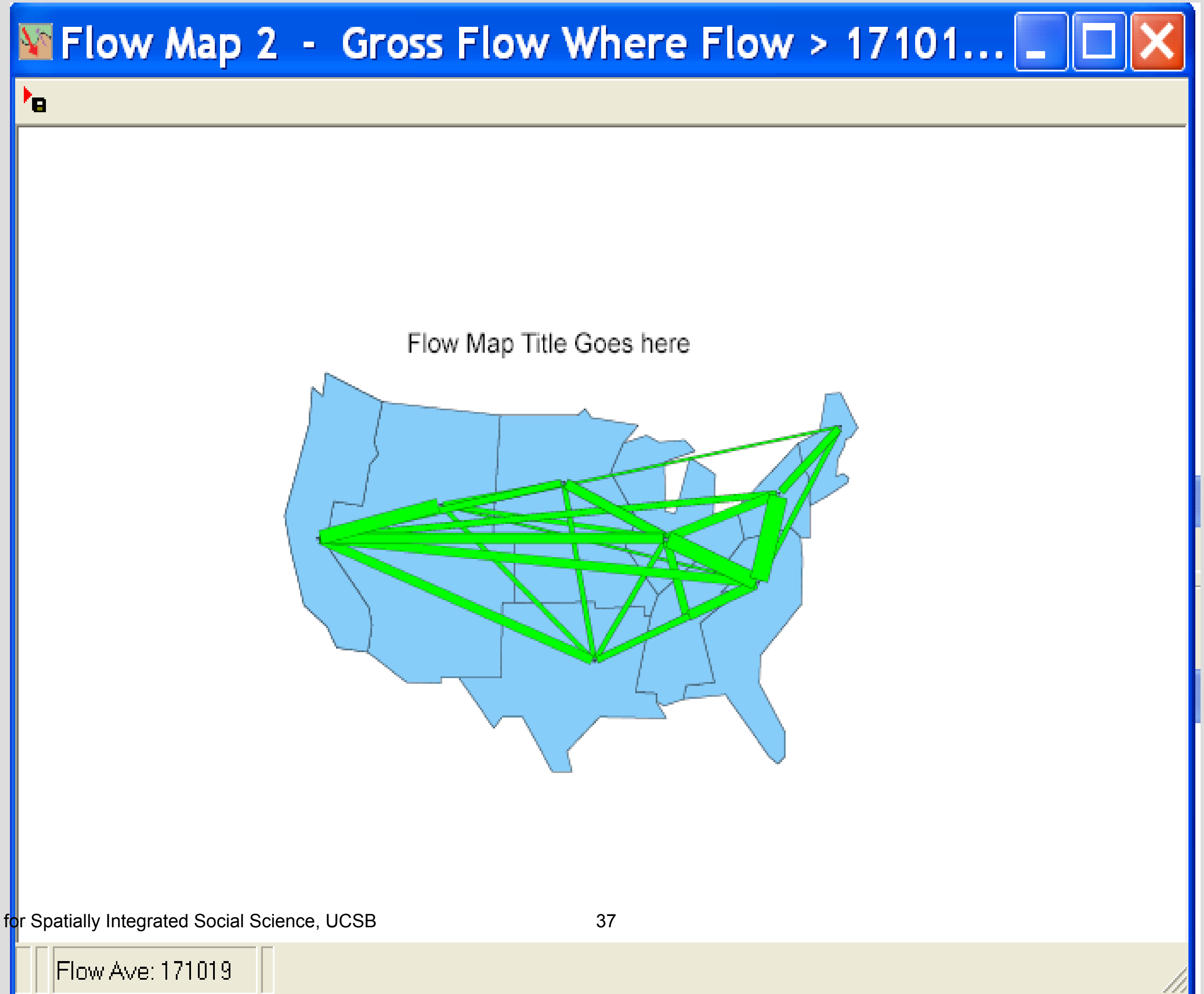
# Here is a map on the screen

Ctrl & Alt keys and right mouse can modify it to make it fit. Use right scroll bar too.



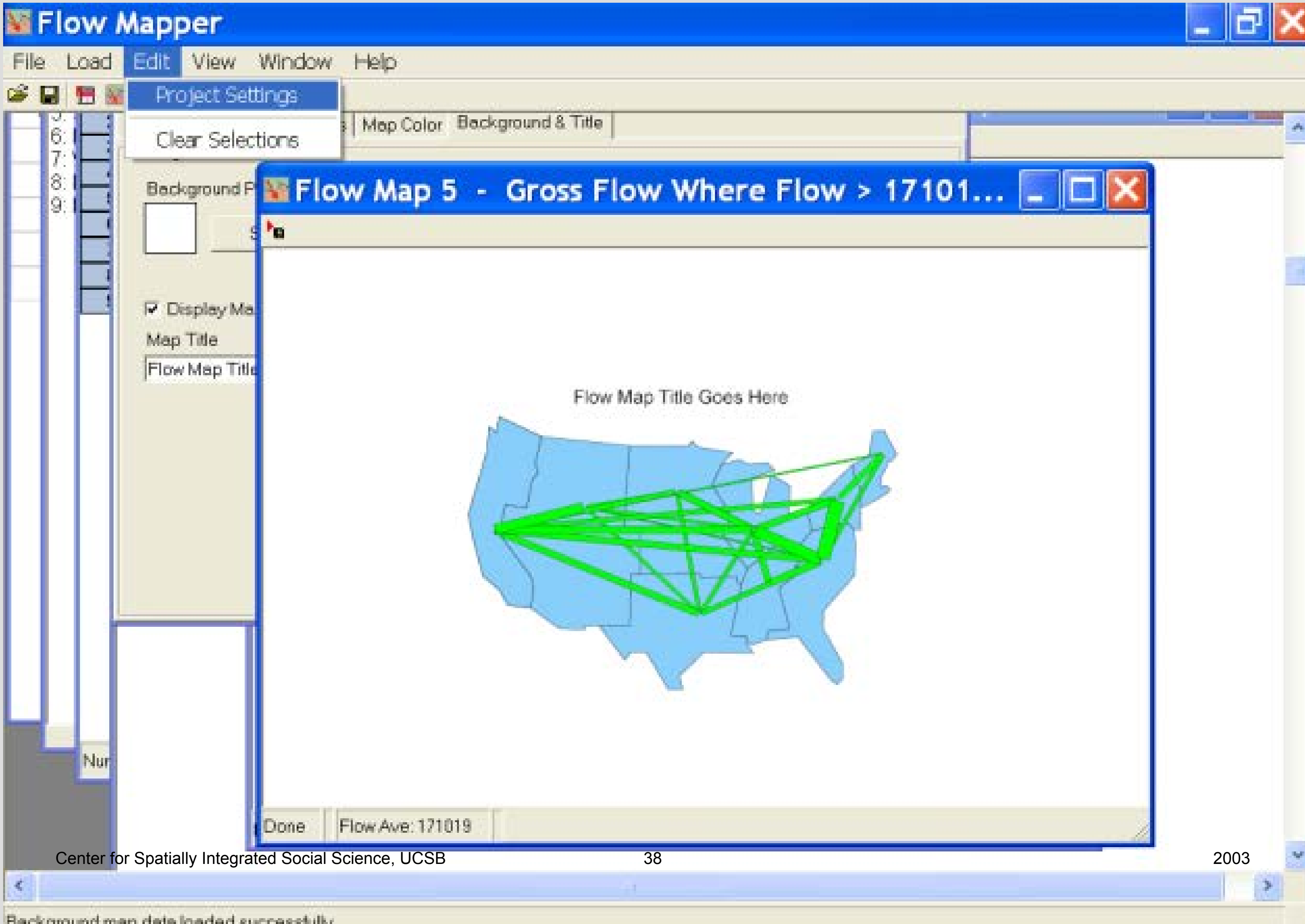
# To save the map

Use the little flagged box at the upper left corner; name it with with an extension. All of the map must be on view on the screen! Later cropping may be desirable.



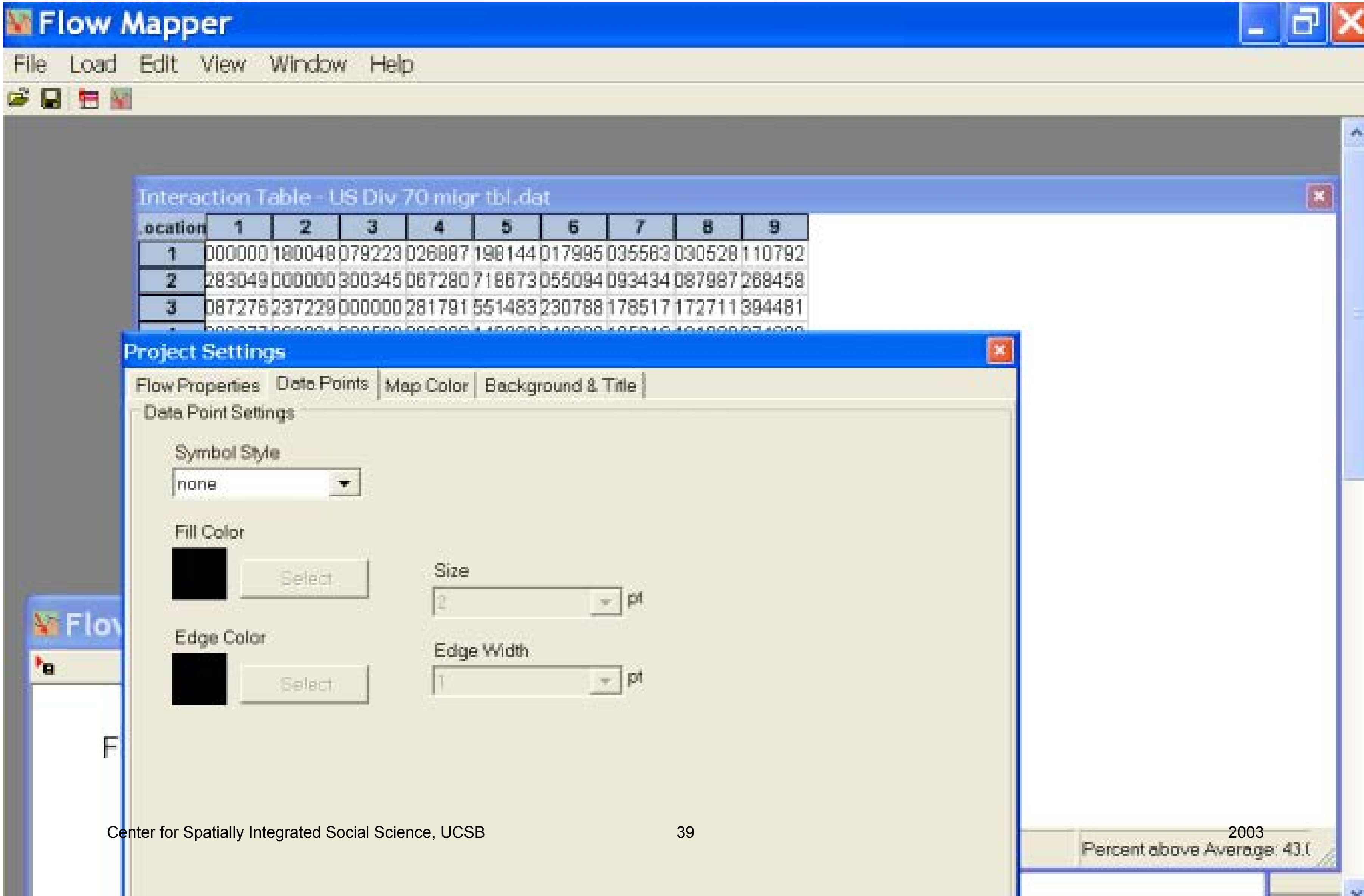
# For a new map click the Edit option again.

This brings up the Project Setting menu. Change the settings as desired for a new map.



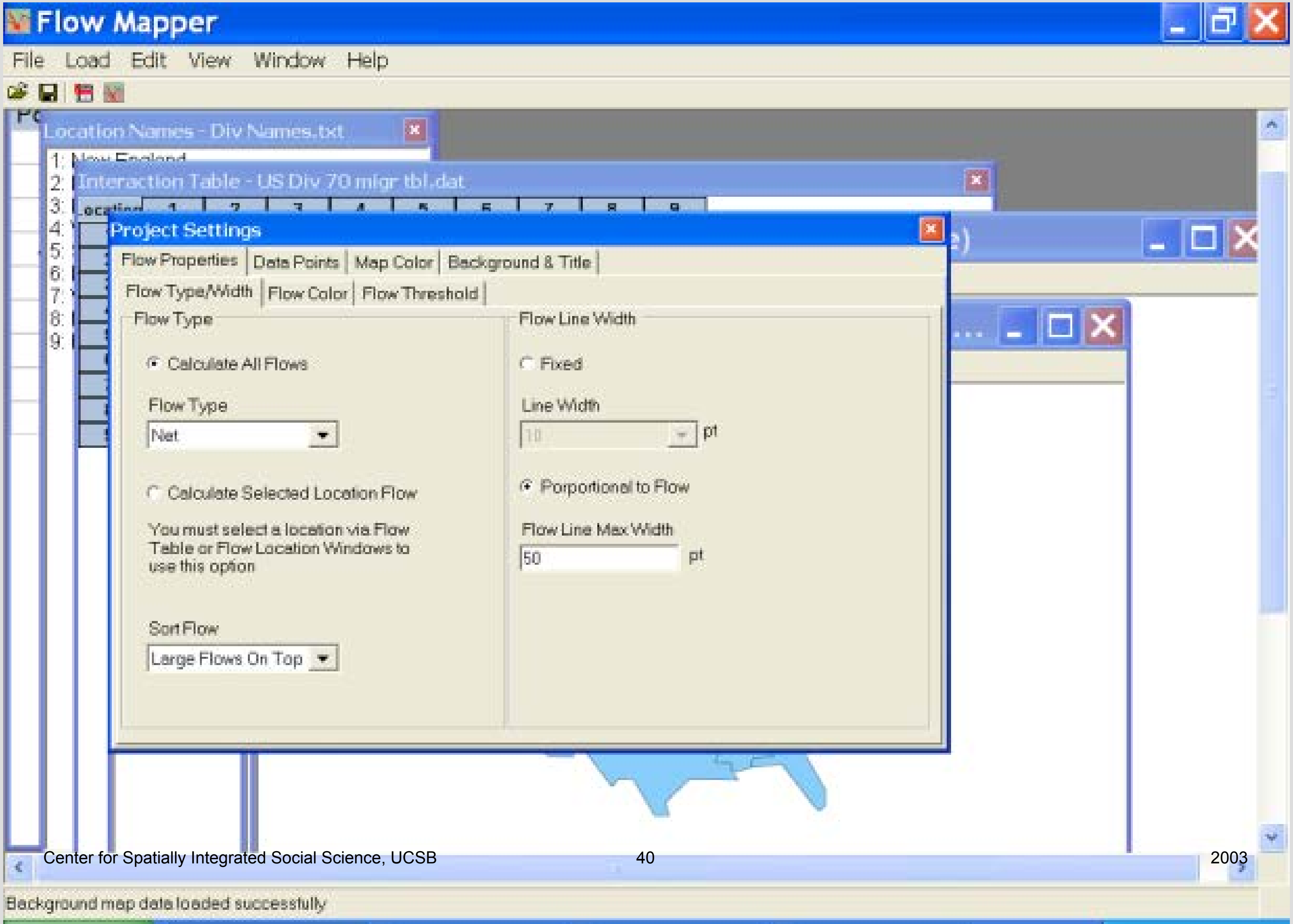
# To move back to first menu click on flow properties

Upper left just below ‘Project Settings’



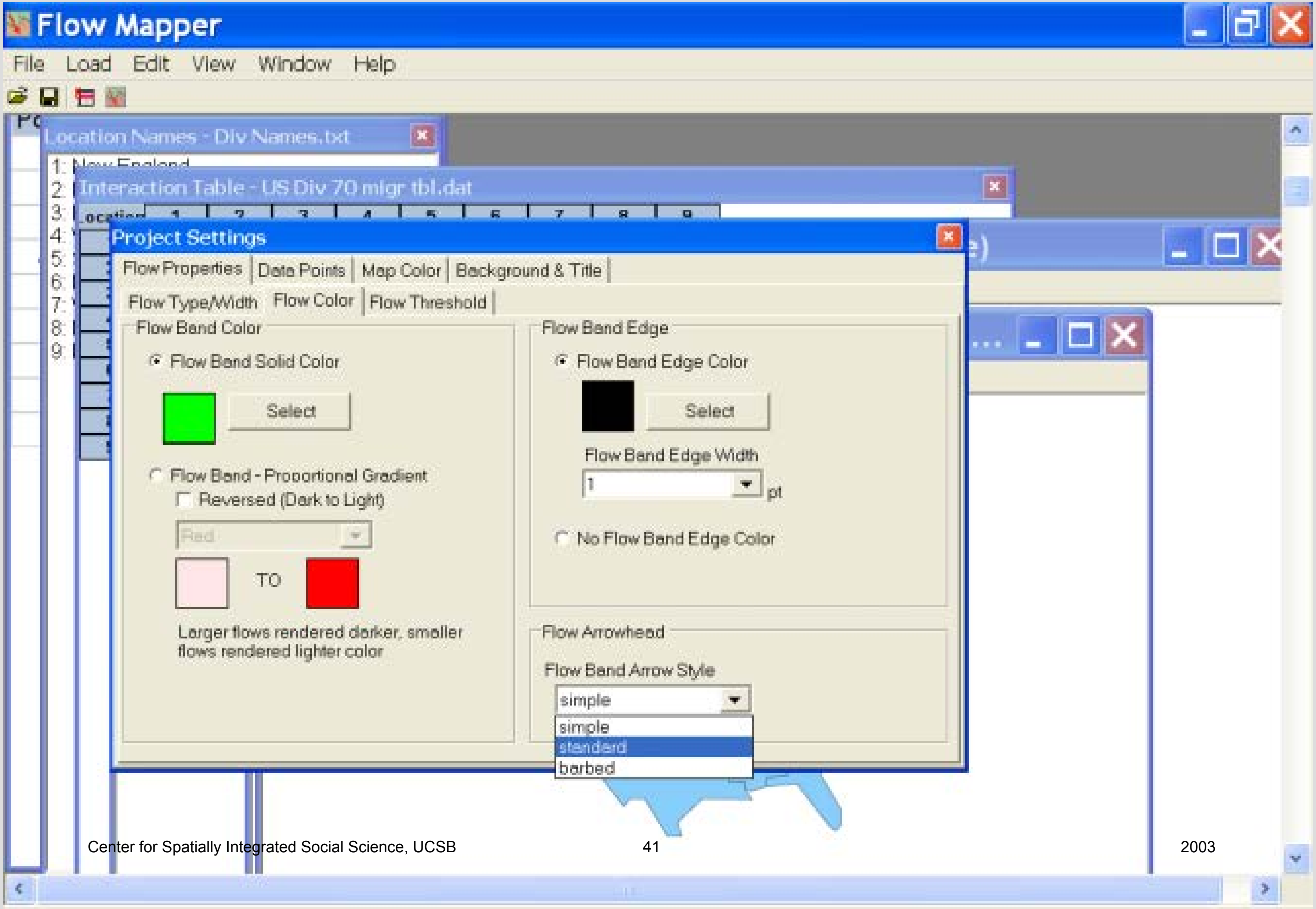
# Settings changed for a net flow map

with changed symbol width

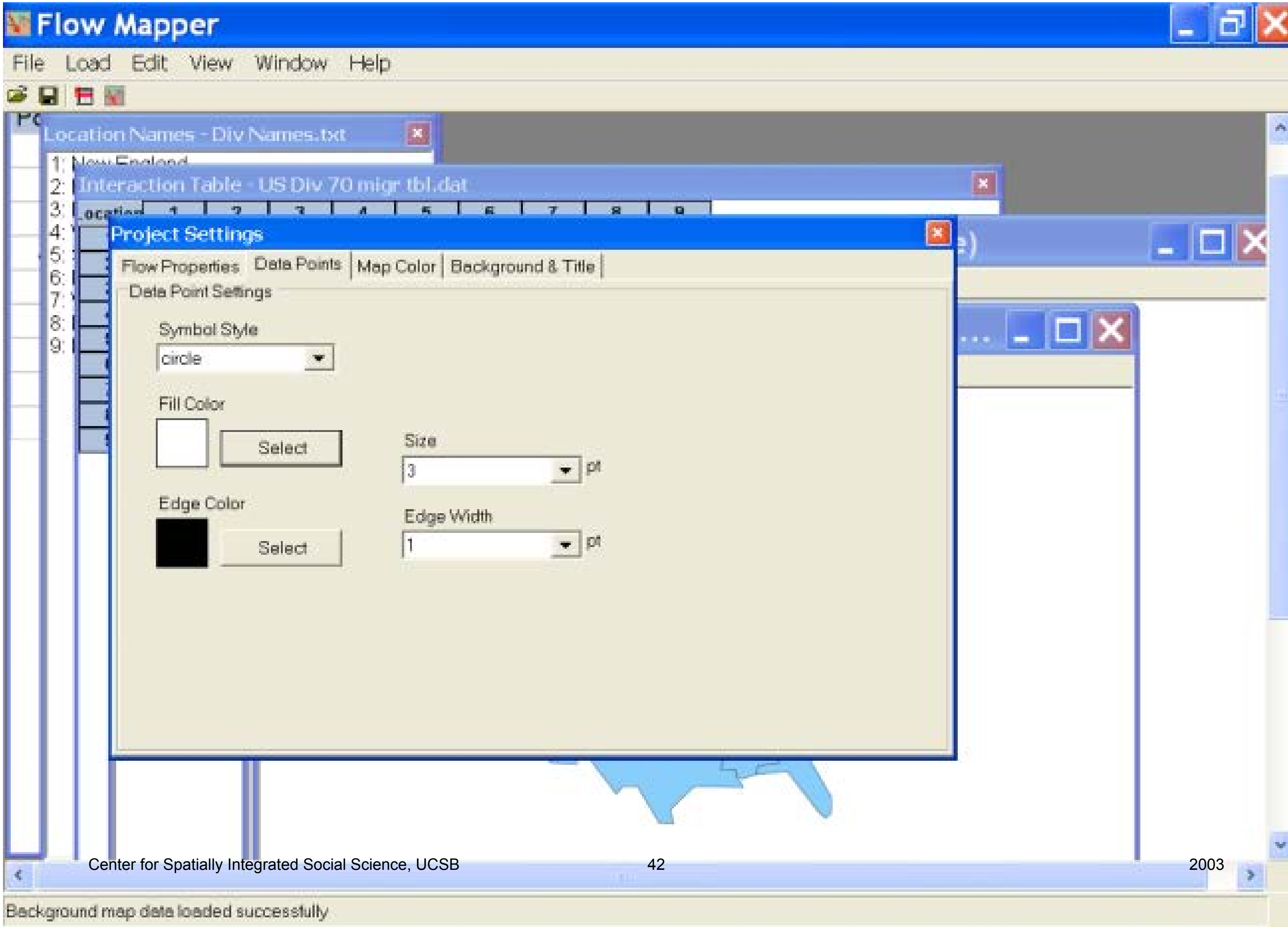


# Arrow style changed

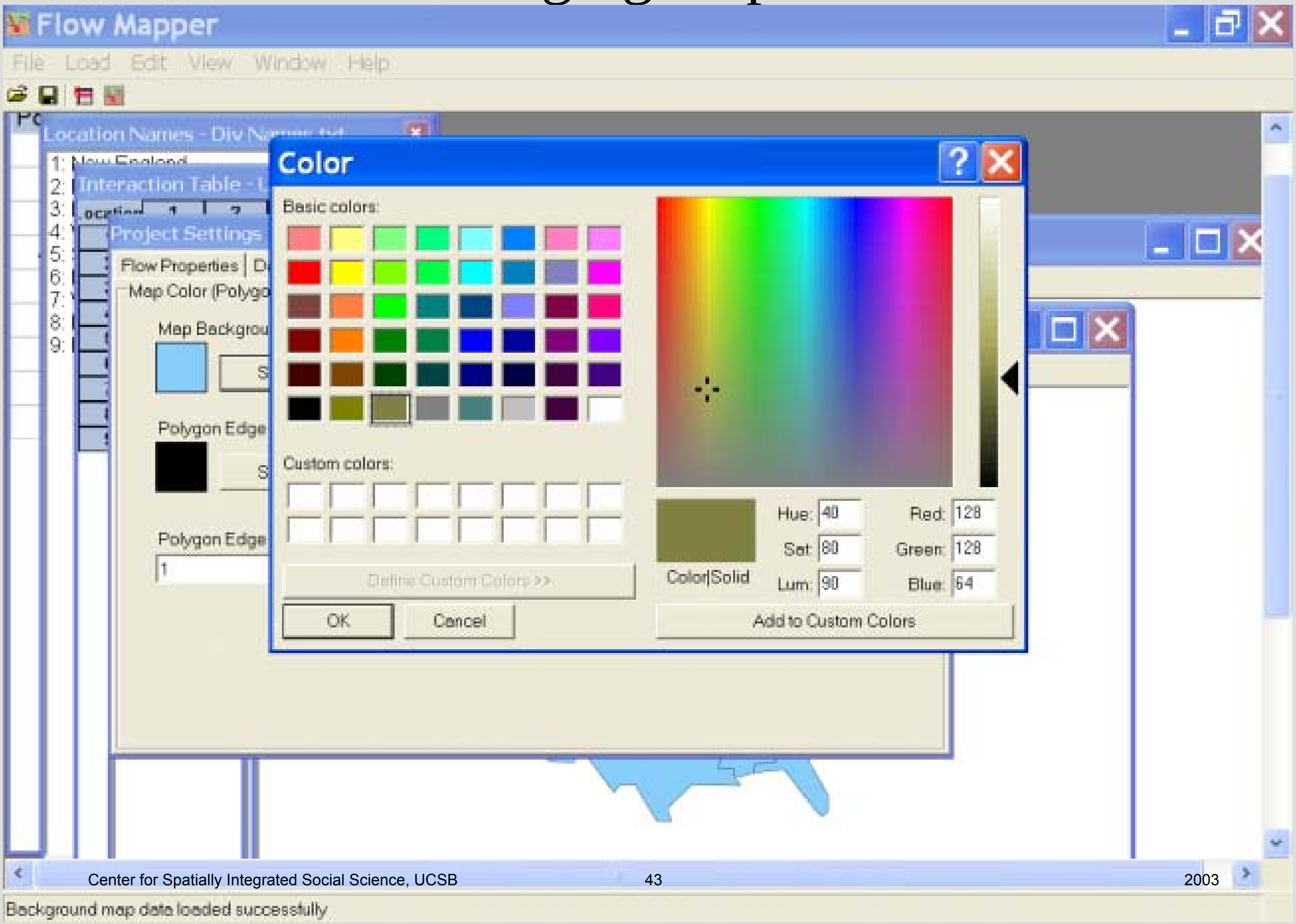
simple, standard, barbed



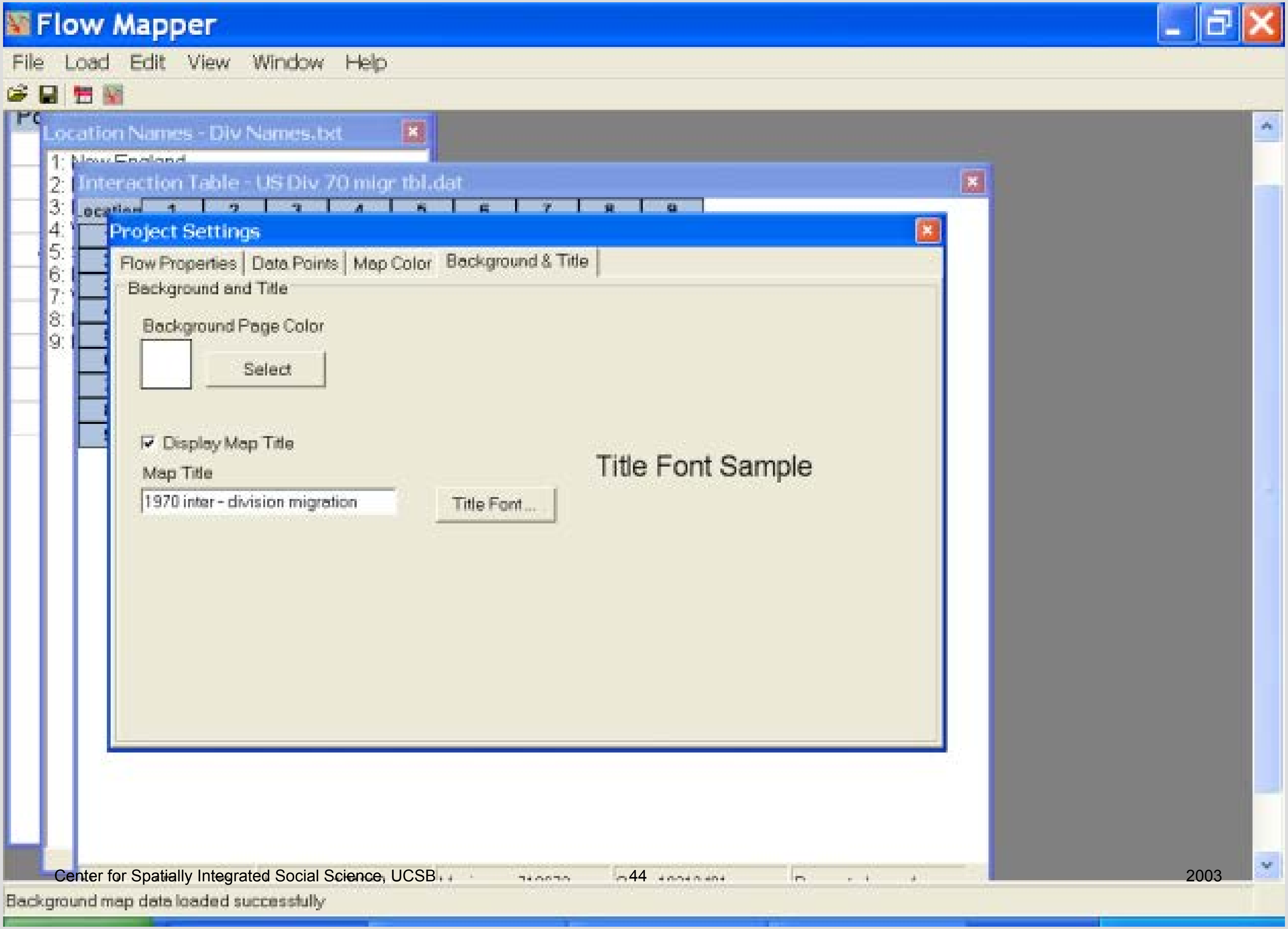
# Displaying locations with a white circle.



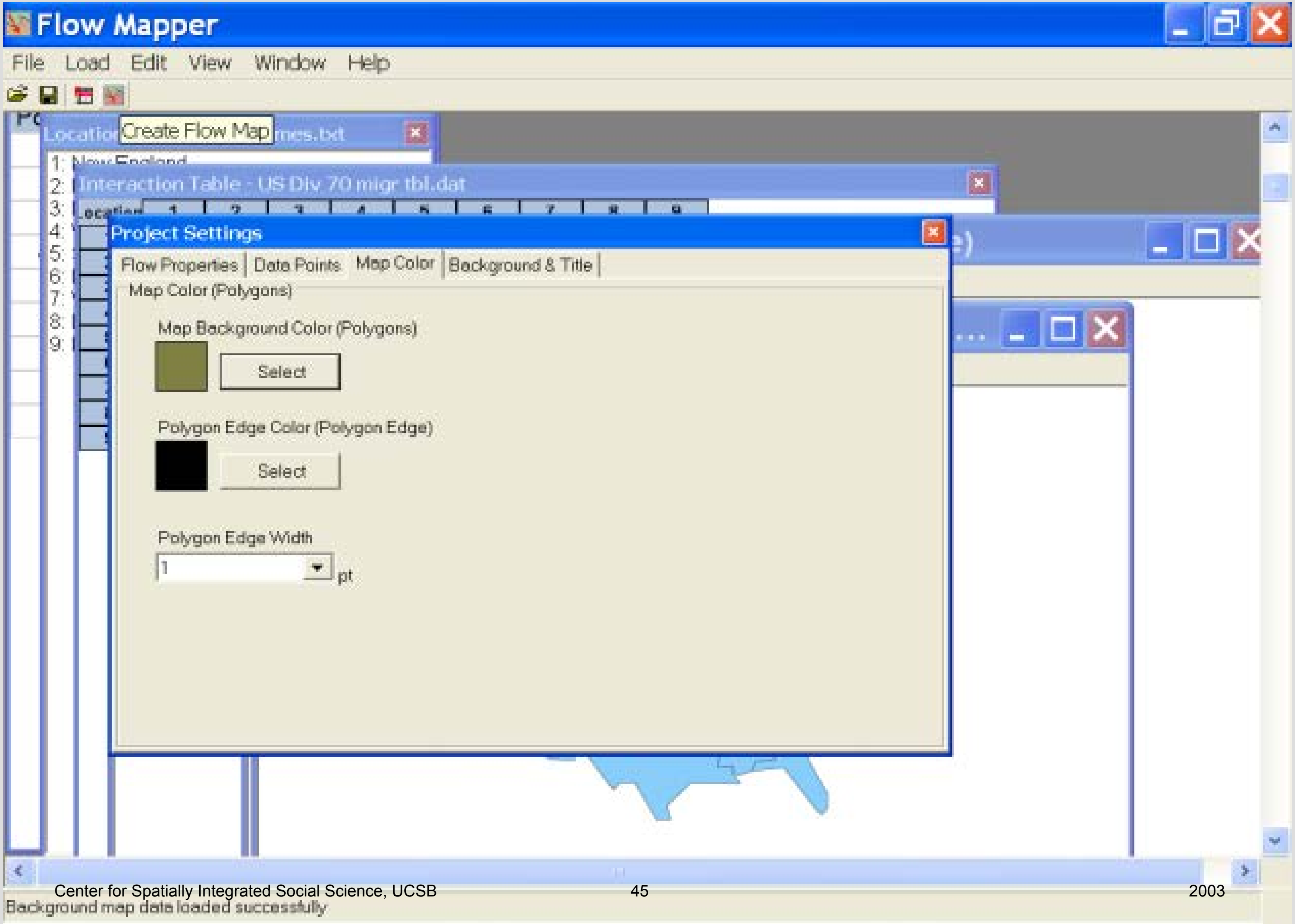
# Changing map color



# Changing title

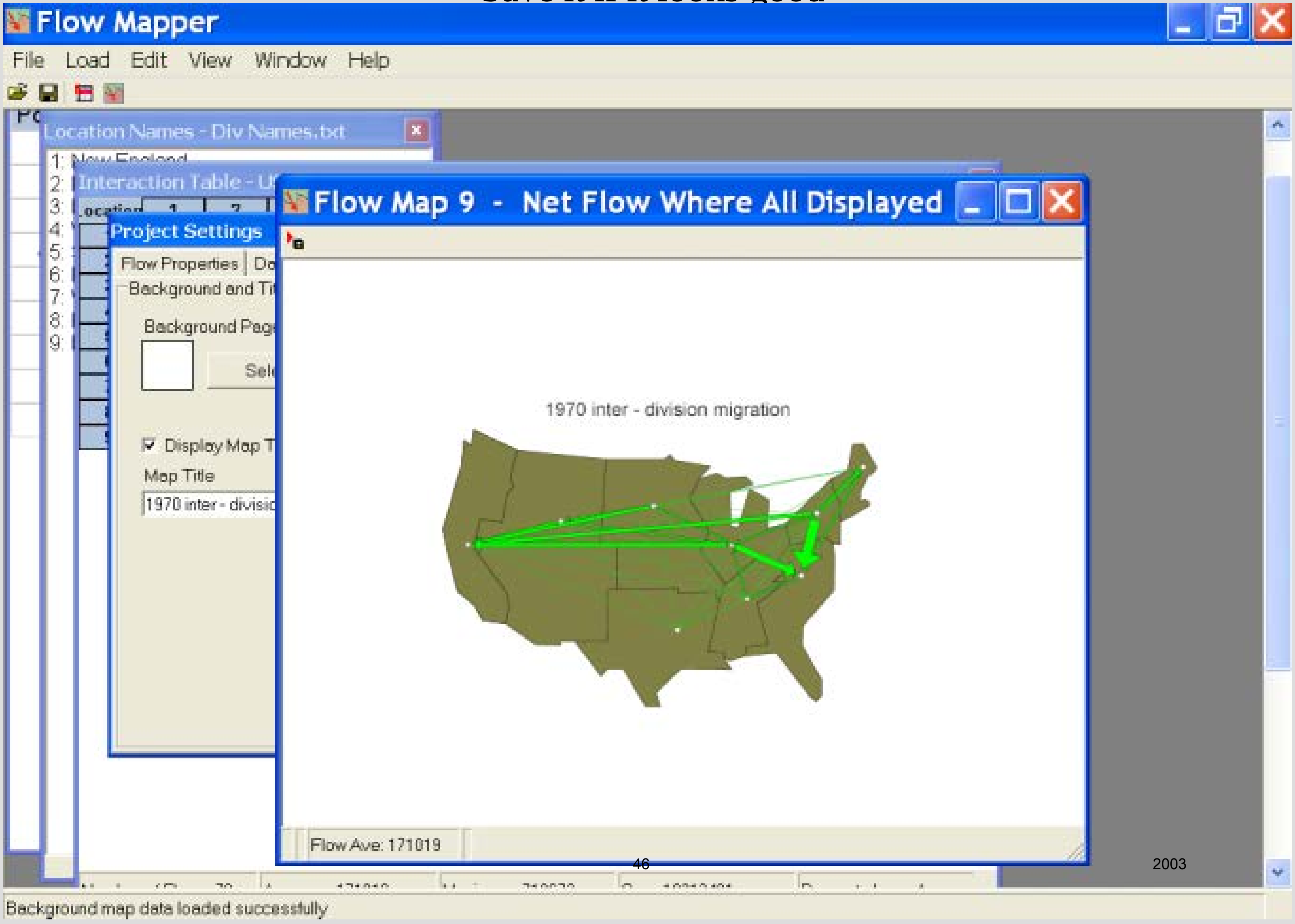


# Creating new map.

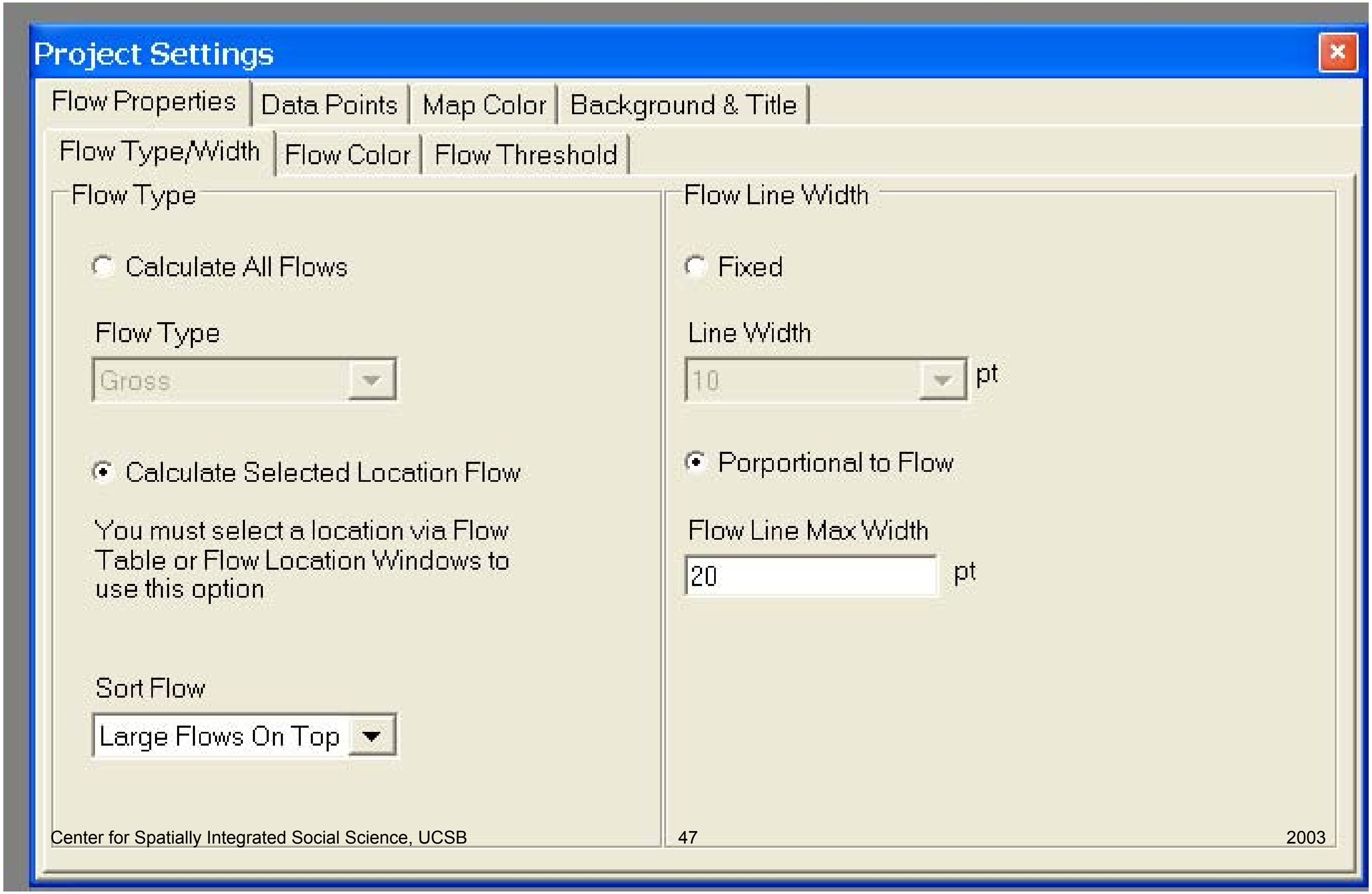


# New map displayed

Save it if it looks good

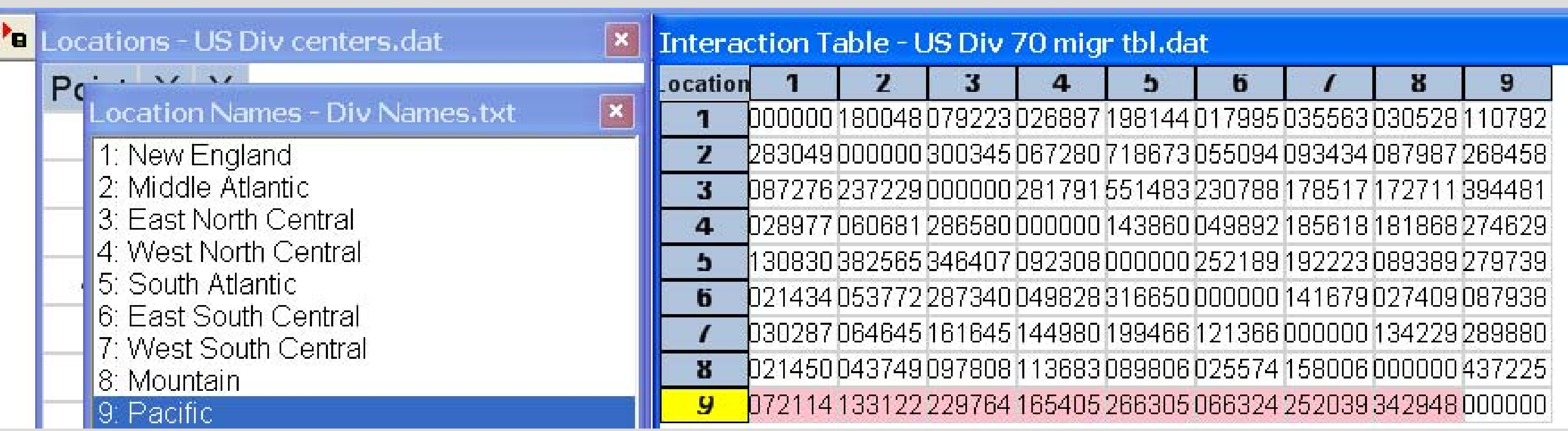


To get moves from (or to) only one place use the ‘**Calculate Selected Location Flow**’ on the ‘**Flow Type**’ menu



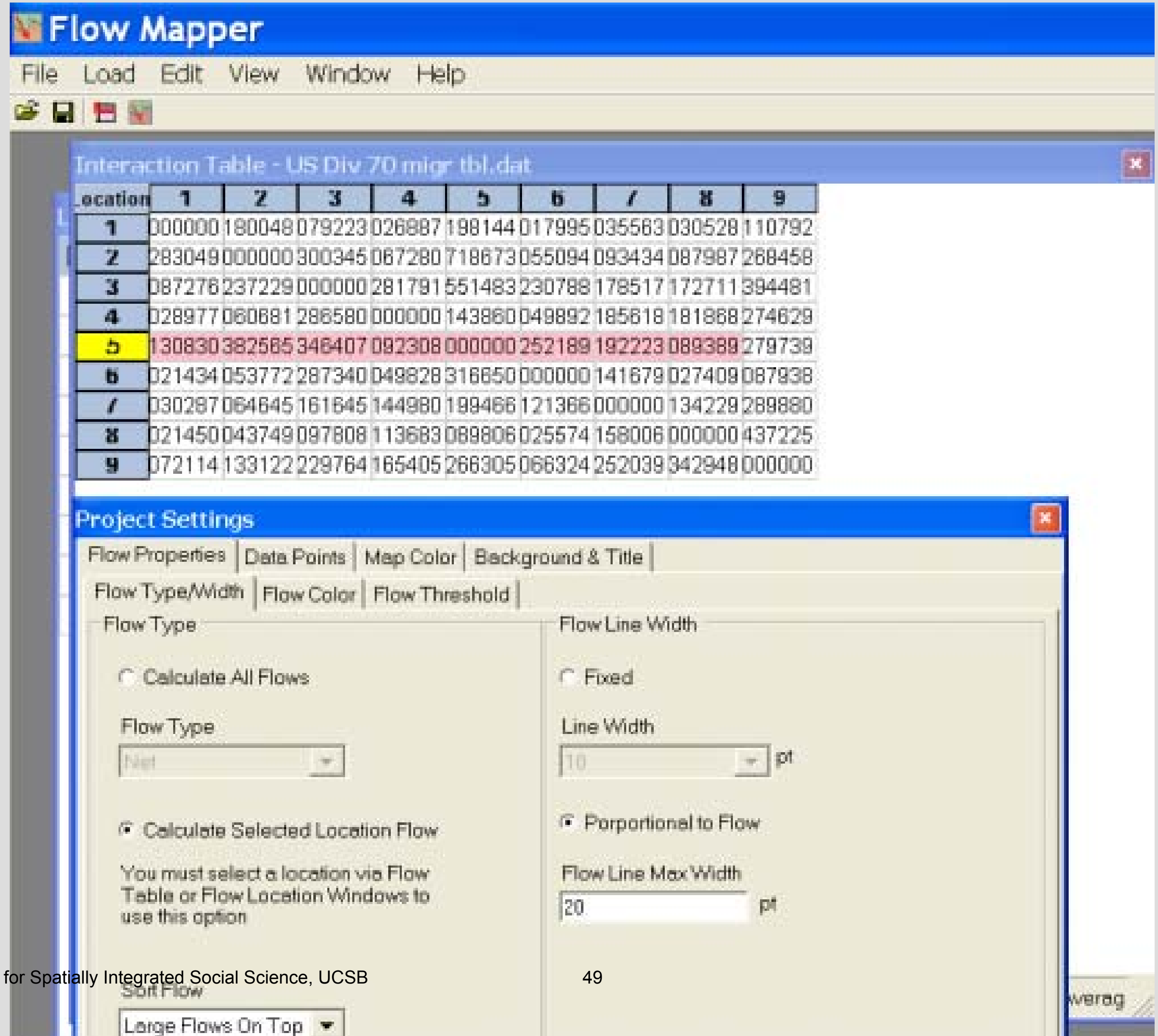
# Next highlight a row (for ‘from’ a place) or a column (for ‘to’ a place) on the interaction table.

Or click on the place in the location table. One click gets you the ‘to’ place, two gets the ‘from’ place. If you cannot see the interaction table use the ‘view’ tab in the top line. The map that you get will be of the net flow, so chose an arrowhead style.

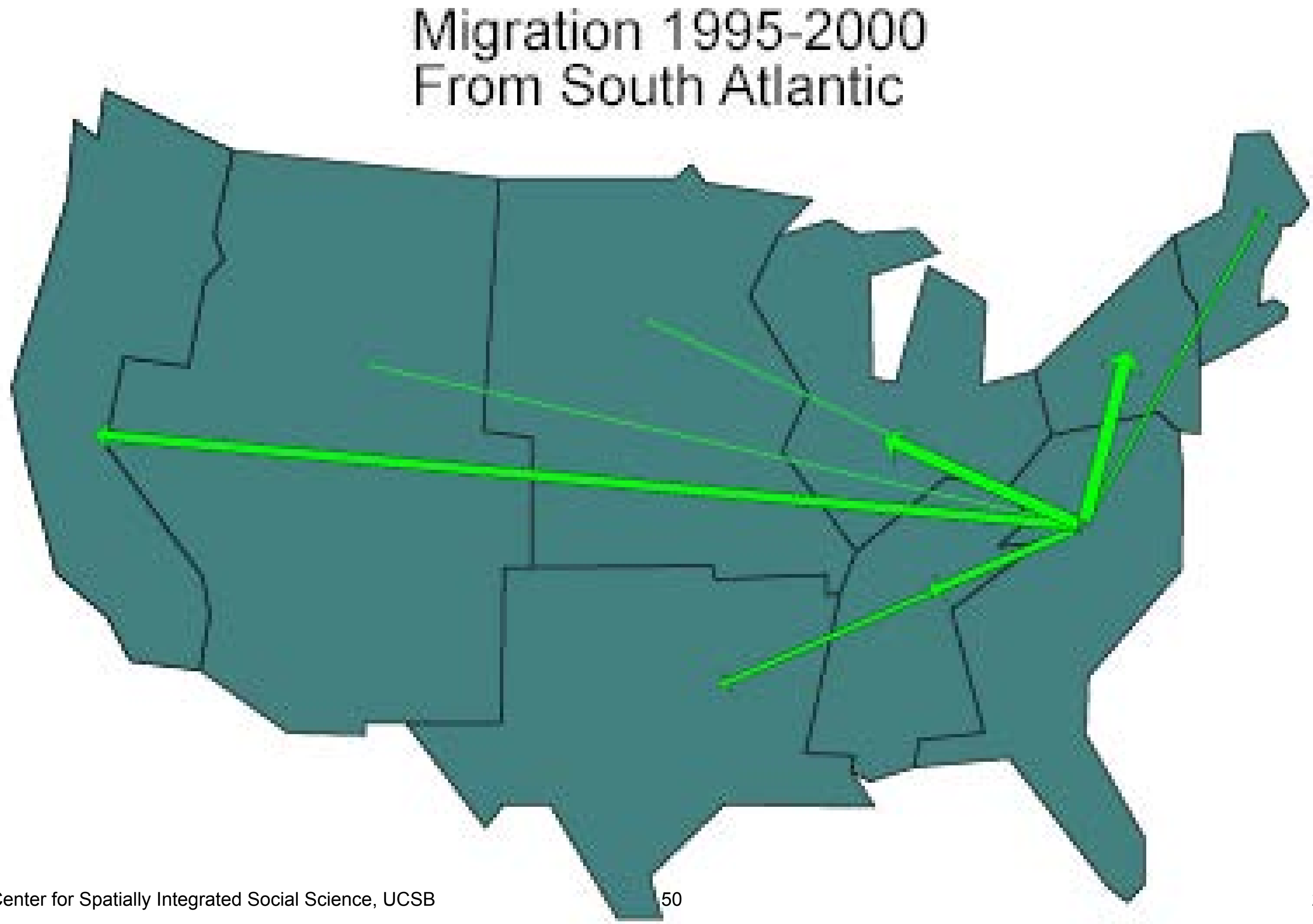



| Location | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 000000 | 180048 | 079223 | 026887 | 198144 | 017995 | 035563 | 030528 | 110792 |
| 2 | 283049 | 000000 | 300345 | 067280 | 718673 | 055094 | 093434 | 087987 | 268458 |
| 3 | 087276 | 237229 | 000000 | 281791 | 551483 | 230788 | 178517 | 172711 | 394481 |
| 4 | 028977 | 060681 | 286580 | 000000 | 143860 | 049892 | 185618 | 181868 | 274629 |
| 5 | 130830 | 382565 | 346407 | 092308 | 000000 | 252189 | 192223 | 089389 | 279739 |
| 6 | 021434 | 053772 | 287340 | 049828 | 316650 | 000000 | 141679 | 027409 | 087938 |
| 7 | 030287 | 064645 | 161645 | 144980 | 199466 | 121366 | 000000 | 134229 | 289880 |
| 8 | 021450 | 043749 | 097808 | 113683 | 089806 | 025574 | 158006 | 000000 | 437225 |
| 9 | 072114 | 133122 | 229764 | 165405 | 266305 | 066324 | 252039 | 342948 | 000000 |

# Or view the interaction table and click on a row

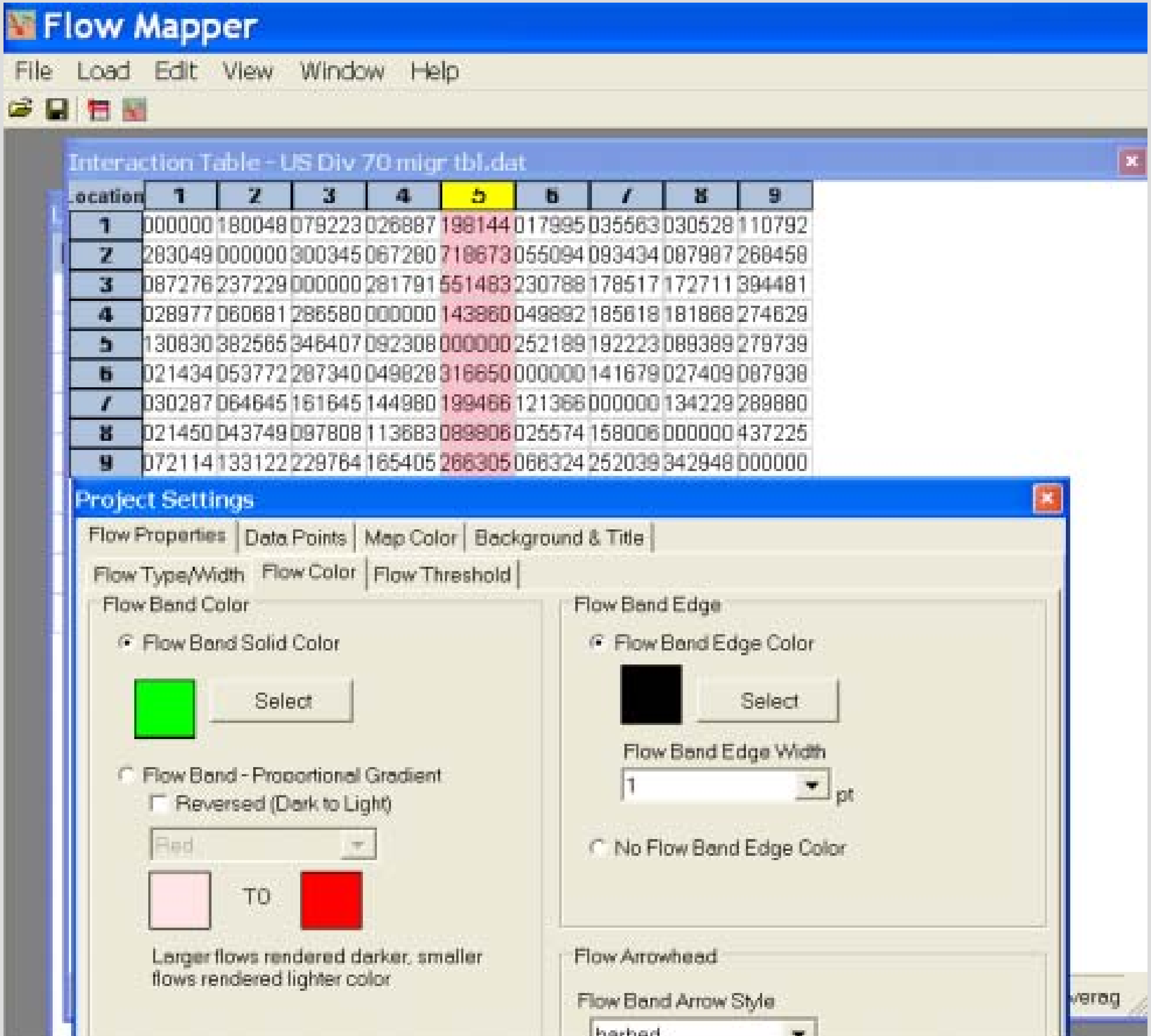


# The moves from the South Atlantic Division



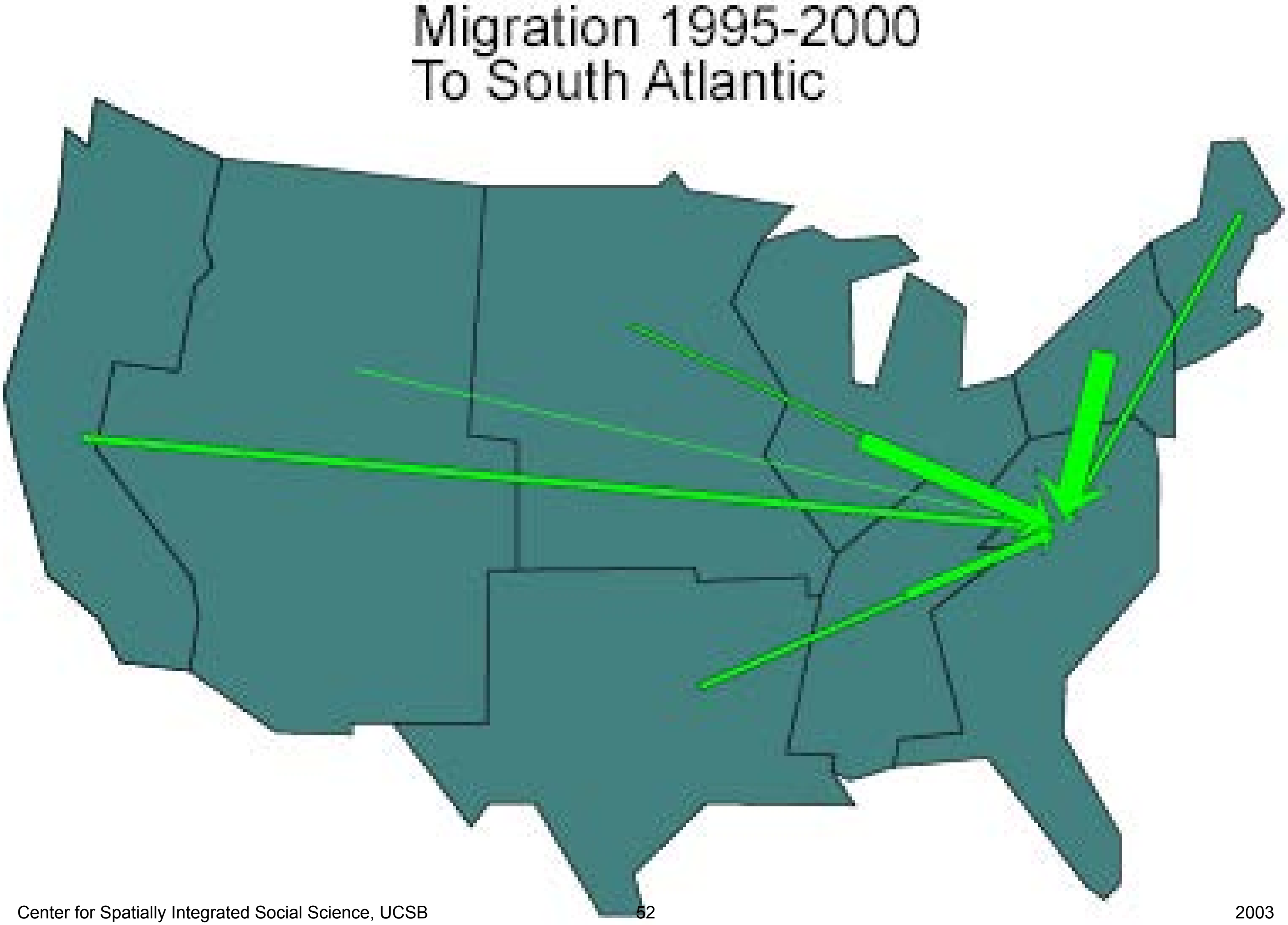
# Or moves to the South Atlantic Division

Notice choice of arrowhead type




2003

# The moves to the South Atlantic Division



The Flow Mapper program can be downloaded from CSISS.org/Spatial tools.
Included are examples and references.
Comments and questions can be directed to
W. Tobler.

http://www.geog.ucsb.edu/~tobler

Program conceived by Waldo Tobler, design & programming by David Jones

# End of part 1 of the tutorial

Now experiment with your own data or try some of the files that came with the program in the Data_Sets folder, or continue with part 2 of the tutorial.

# Tutorial Part II

## An example of using Flow Mapper

by Waldo Tobler

# The life history of a flow mapping project

Locate an interaction table.

Locate a map.

Digitize the map.

Enter the table and coordinates.

Use the flow map program.

Use a model to estimate the movement.

Compare the observed with the estimate.

# Study area in Pennsylvania

## Ten counties containing five parks

324 JOURNAL OF REGIONAL SCIENCE, VOL. 19, NO. 3, 1979

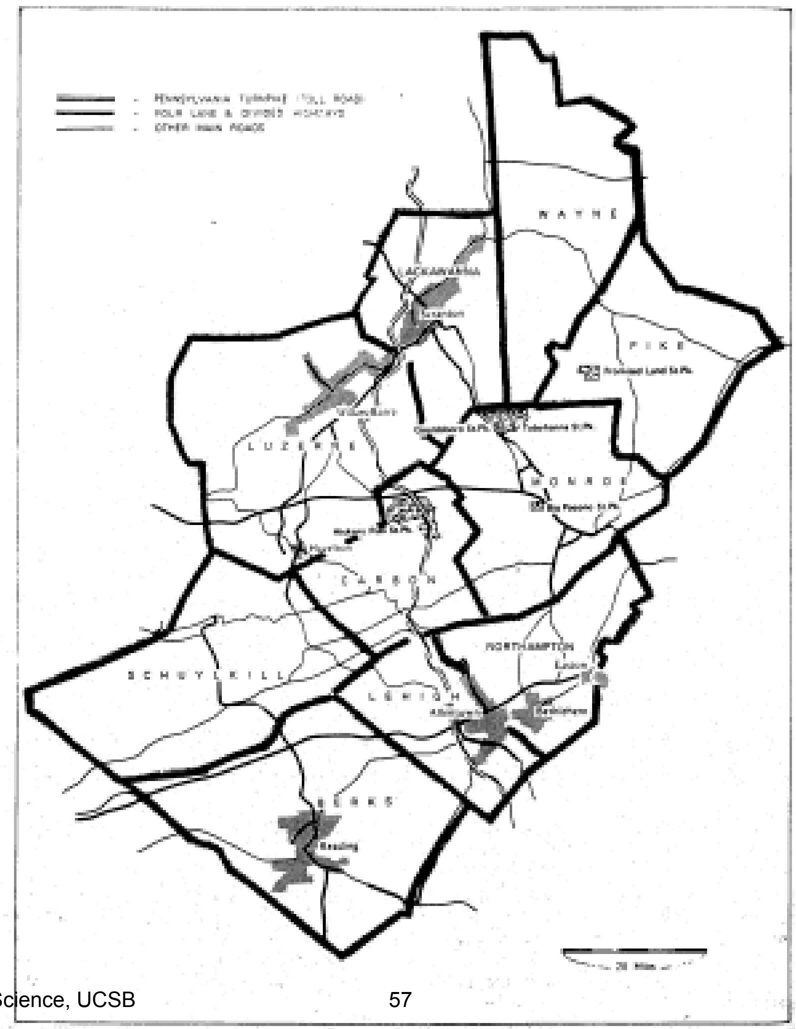
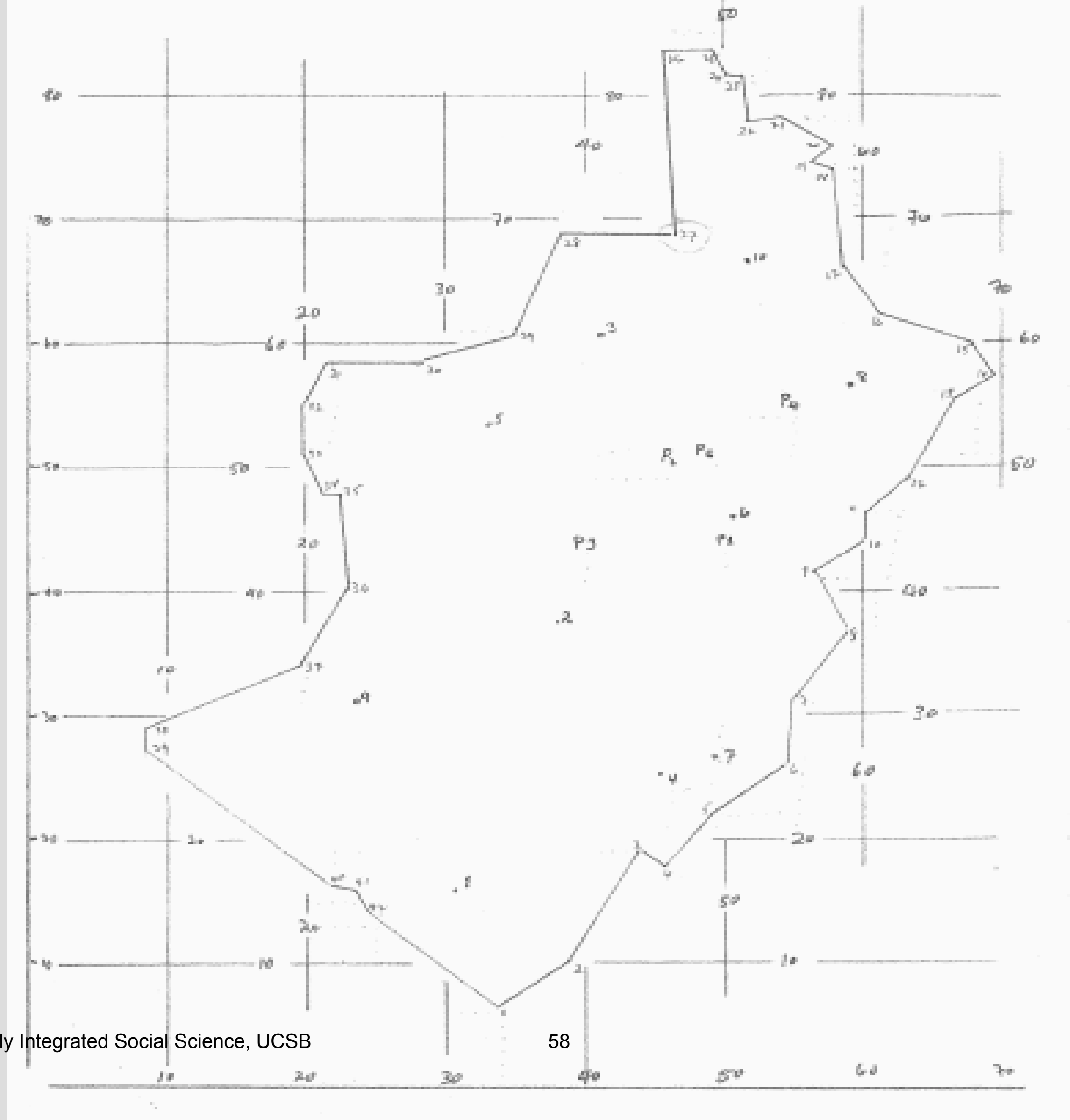



FIGURE 1: Map of the Study Area.

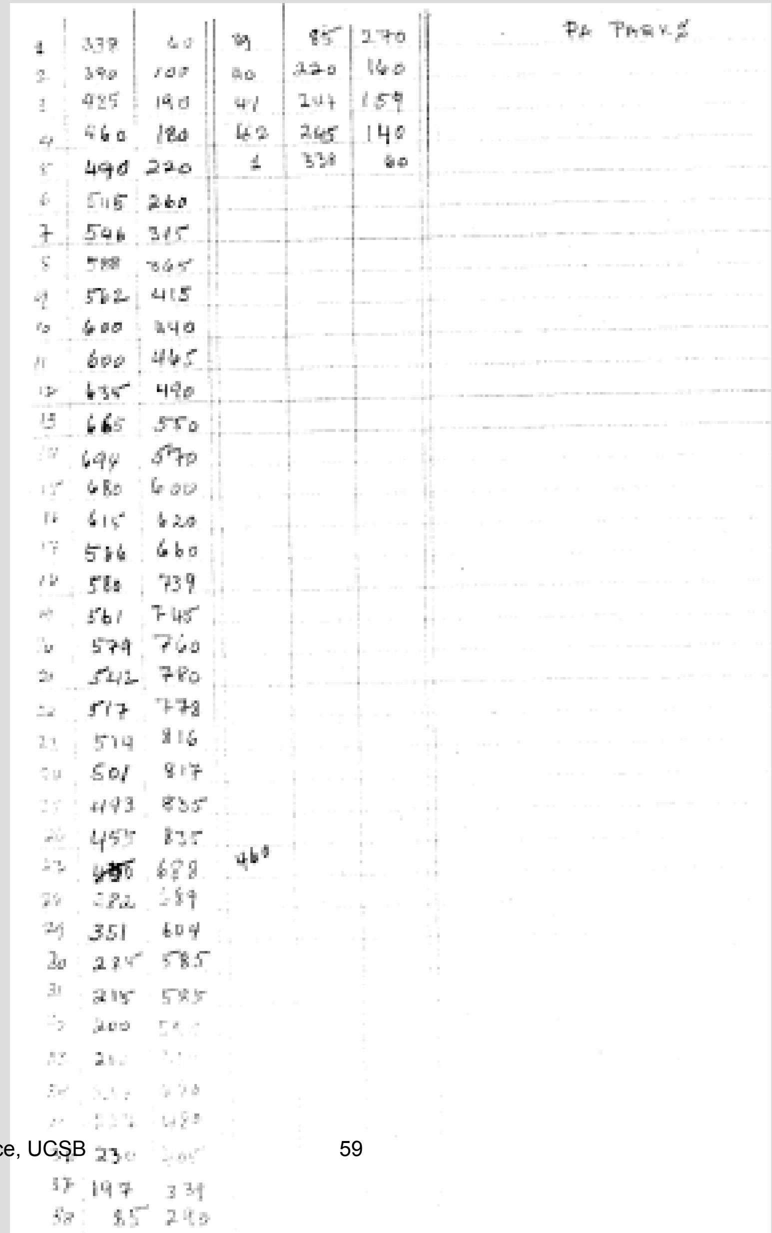
# Getting coordinates

Area outline and centroids, using graph paper.The results go into an ASCII file.

Or use a digitizer but only if you have lots of experience with it.



# My recording of coordinates


PA TRACKS

| | | | | | |
|---|---|---|---|---|---|
| 1 | 338 | 60 | 39 | 85 | 270 |
| 2 | 390 | 100 | 40 | 220 | 160 |
| 3 | 425 | 190 | 41 | 244 | 159 |
| 4 | 460 | 180 | 42 | 265 | 140 |
| 5 | 490 | 220 | 1 | 338 | 60 |
| 6 | 515 | 260 | | | |
| 7 | 546 | 315 | | | |
| 8 | 588 | 365 | | | |
| 9 | 562 | 415 | | | |
| 10 | 600 | 440 | | | |
| 11 | 600 | 465 | | | |
| 12 | 635 | 490 | | | |
| 13 | 665 | 550 | | | |
| 14 | 694 | 570 | | | |
| 15 | 680 | 600 | | | |
| 16 | 615 | 620 | | | |
| 17 | 586 | 660 | | | |
| 18 | 580 | 739 | | | |
| 19 | 561 | 745 | | | |
| 20 | 579 | 760 | | | |
| 21 | 542 | 780 | | | |
| 22 | 517 | 778 | | | |
| 23 | 514 | 816 | | | |
| 24 | 501 | 817 | | | |
| 25 | 493 | 835 | | | |
| 26 | 455 | 835 | | | |
| 27 | 450 | 688 | 460 | | |
| 28 | 282 | 589 | | | |
| 29 | 351 | 604 | | | |
| 30 | 224 | 585 | | | |
| 31 | 215 | 585 | | | |
| 32 | 200 | [illegible] | | | |
| 33 | [illegible] | [illegible] | | | |
| 34 | [illegible] | [illegible] | | | |
| 35 | [illegible] | [illegible] | | | |
| 36 | 230 | [illegible] | | | |
| 37 | 197 | 339 | | | |
| 38 | 85 | 290 | | | |

# Boundary outline coordinates

X and Y stored in an ASCII file

| | |
|---|---|
| 338 | 60 |
| 390 | 100 |
| 435 | 190 |
| 460 | 180 |
| 490 | 220 |
| 545 | 260 |
| 546 | 315 |
| 588 | 365 |
| 562 | 415 |
| 600 | 440 |
| 600 | 465 |
| 635 | 490 |
| 665 | 550 |
| 694 | 570 |
| 680 | 600 |
| 615 | 620 |
| 586 | 660 |
| 580 | 739 |
| 561 | 745 |
| 579 | 760 |
| 542 | 780 |
| 517 | 778 |
| 514 | 816 |
| 501 | 817 |
| 493 | 835 |
| 455 | 835 |
| 460 | 688 |
| 382 | 689 |
| 351 | 604 |
| 285 | 585 |
| 215 | 585 |
| 200 | 550 |
| 200 | 510 |
| 213 | 480 |
| 227 | 480 |
| 230 | 405 |
| 197 | 339 |
| 85 | 290 |
| 85 | 270 |
| 220 | 160 |
| 249 | 159 |
| 245 | 140 |
| 338 | 60 |

# Fifteen centroid coordinates

## Ten counties and five parks

### X and Y also stored in an ASCII file

| | |
|---|---|
| 310 | 158 |
| 380 | 370 |
| 410 | 610 |
| 455 | 245 |
| 330 | 535 |
| 510 | 460 |
| 265 | 490 |
| 590 | 565 |
| 240 | 310 |
| 520 | 660 |
| 500 | 440 |
| 465 | 510 |
| 400 | 440 |
| 540 | 550 |
| 480 | 515 |

# County, then park, names

Berks
Carbon
Lackawanna
Leheigh
Lucerne
Monroe
Northhampton
Pike
Schuykill
Wayne
Big Pocono Park
Gouldsboro Park
Hickory Run Park
Promised Land Park
Tobyhanna Park

# Movement table

## From 10 counties to 5 parks

Table 1: Observed Movements

| To Park / From County | Big Pocono | Gouldsboro | Hickory Run | Promised Land | Tobyhanna | OutSum $O_i$ |
|---|---|---|---|---|---|---|
| Berks | 46 | 35 | 333 | 84 | 69 | 567 |
| Carbon | 50 | 33 | 1670 | 71 | 91 | 1915 |
| Lackawanna | 230 | 6970 | 141 | 977 | 1917 | 10235 |
| Lehigh | 307 | 520 | 1458 | 315 | 387 | 2987 |
| Lucerne | 255 | 3366 | 4586 | 303 | 595 | 9105 |
| Monroe | 376 | 313 | 253 | 150 | 848 | 1940 |
| Northhampton | 385 | 1121 | 1263 | 499 | 981 | 4249 |
| Pike | 17 | 7 | 26 | 87 | 6 | 143 |
| Schuylkill | 63 | 101 | 1886 | 48 | 40 | 2138 |
| Wayne | 8 | 20 | 12 | 124 | 18 | 183 |
| InSum $I_j$ | 1737 | 12486 | 11628 | 2658 | 4952 | 33461 |

Source: Cesario (1973), Table 5, p. 245.

This representa a rather different situation. Given is a ten by five table of interaction relating the residents of ten counties and their attendence at five parks. The table is thus rectangular. But the flow mapping program expects square arrays. A simple short computer program is used to convert the small table into a fifteen by fifteen table, wth the original 10 by 5 in the upper right corner ('Original.tbl') and zeros in the rest of the table.

The location list contains fifteen entries, the first ten being the coordinates for the counties and the last five the coordinates of the parks ('XYs.dat'). The name list is also in this order ('Names.txt'). The parks are all located in Pennsylvania in the area between Wilkes Barre-Scranton-Lackawanna and Berks-Allentown-Bethleham. A map of the study area showing details is given in M. Baxter, G. Ewing, 1979, "Calibration of Production Constrained Trip Distribution Models and the Effect of Intervening Opportunities', Journal of Regional Science, 19(3): 319-330.

A map outline of 43 points is given in a text file ('Boundry.dat'). The units are arbitrary from a piece of graph paper.

Since the information represents the movements from the counties to the parks it is appropriate to use the net flow map ('Observed.tif').

The map of observed movements is then compared to a map ('Estimated.tif') made from the estimated movements computed from the table marginals using a model ('Estimated.tbl'). The model in this instance is the so-called 'Quadratic Transportation Problem'. The absolute value of the difference between these two sets of data ('Difference.tbl') is also represented on a map ('Difference.tif'). The correlation (r-squared) between the model estimate and the observed movements is 0.85

The model, and the data, are completely described in W. Tobler, 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".

# Having found an interaction matrix, the next step is to get it into the computer

If the table is small you can enter it by typing it into notepad.

Larger tables can be entered using a spreadsheet.

Excel tables can be used by converting them to space or comma delimited ASCII files (do not use tab delimited).

# The 15 by 15 observed movement table.

The 10 by 5 table has been forced into a square format.

The movement from the 10 counties to the 5 parks is one directional only.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 46 | 35 | 333 | 84 | 69 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 | 33 | 1670 | 71 | 91 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 230 | 6970 | 141 | 977 | 1917 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 307 | 520 | 1458 | 315 | 387 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 255 | 3366 | 4586 | 303 | 595 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 376 | 313 | 253 | 150 | 848 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 385 | 1121 | 1263 | 499 | 981 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 7 | 26 | 87 | 6 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 63 | 101 | 1886 | 48 | 40 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 20 | 12 | 124 | 18 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The next step is to produce the map, as in the previous tutorial

# There are visitors from 10 counties to 5 Parks

The moves indicated are from the counties to the parks.

This yields a rectangular table.

The flow program expects a square table.

The rectangular table needs to be converted to a square table.

This is done by constructing a 15 by 15 table, of mostly zeros.

An ‘input help program’ does this conversion.

Conversion from origin-destination lists is also available.

The rectangular 10 by 5 table shows up in the upper right hand corner.

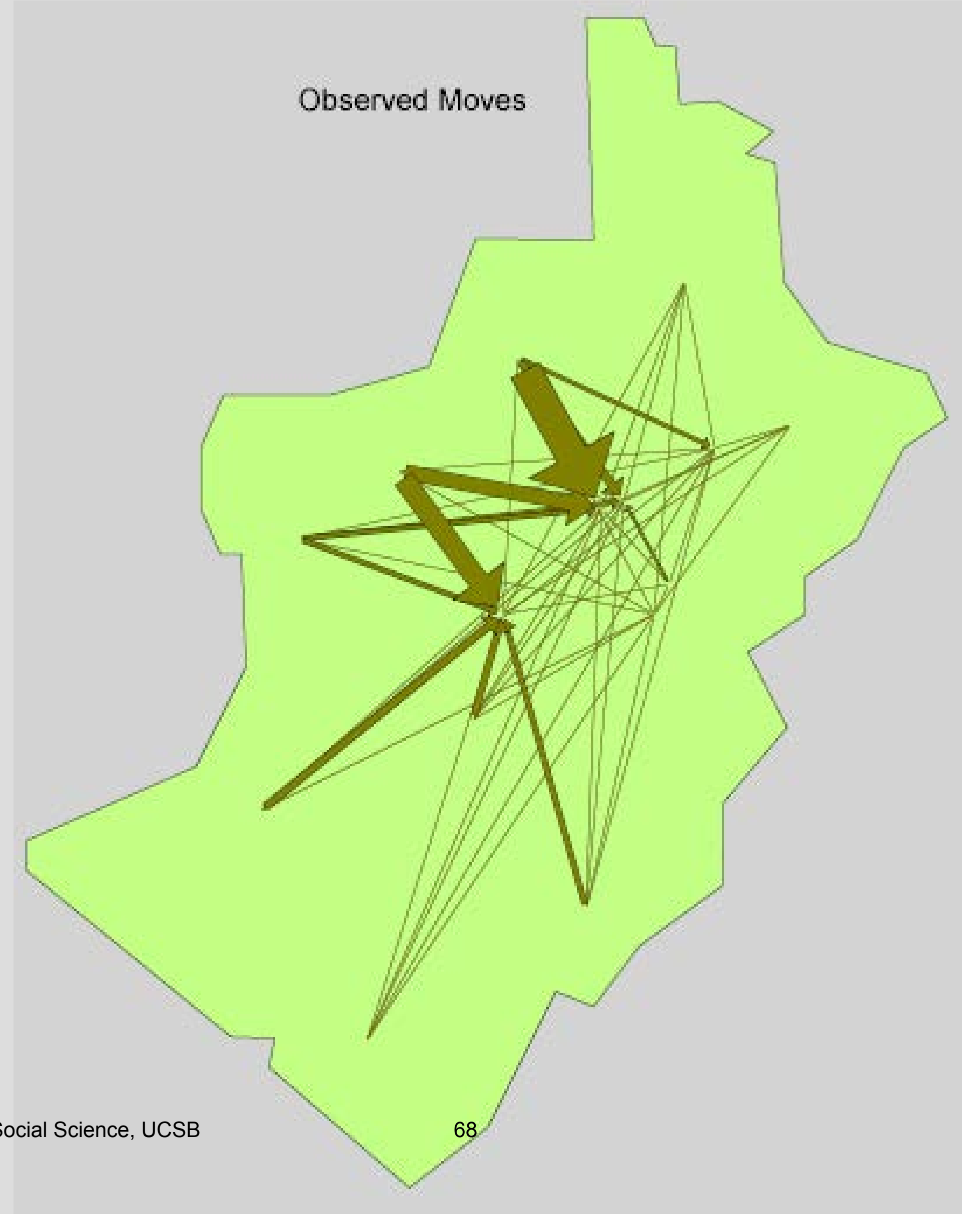
The full table could show moves between counties & between parks.

But these moves are not recorded.

Return moves are implicit but not depicted.

The lower left corner could be used for these, as the transposed table, a 5 by 10 table.

# Visits by county residents to parks



# Distance from parks to counties

Needed for model estimates. The model also uses the table marginals. These values must also be in a computer file.

Table 2: Distance from counties to parks ($C_{ij}$) in miles

| To Parks / From County | Big Pocono | Gouldsboro | Hickory Run | Promised Land | Tobyhanna |
|---|---|---|---|---|---|
| Berks | 95 | 101 | 89 | 115 | 96 |
| Carbon | 40 | 52 | 30 | 71 | 44 |
| Lackawanna | 45 | 21 | 46 | 35 | 29 |
| Lehigh | 47 | 62 | 57 | 70 | 62 |
| Lucerne | 55 | 45 | 25 | 65 | 49 |
| Monroe | 17 | 26 | 45 | 26 | 24 |
| Northhampton | 41 | 56 | 64 | 60 | 52 |
| Pike | 49 | 53 | 80 | 35 | 47 |
| Schuylkill | 70 | 77 | 57 | 85 | 71 |
| Wayne | 53 | 37 | 72 | 22 | 37 |

Source: Cesario (1973), Table 5, p. 245

# Movement table

## From 10 counties to 5 parks

## with marginals: Insums and Outsums noted

Table 1: Observed Movements

| To Park / From County | Big Pocono | Gouldsboro | Hickory Run | Promised Land | Tobyhanna | OutSum $O_i$ |
|---|---|---|---|---|---|---|
| Berks | 46 | 35 | 333 | 84 | 69 | 567 |
| Carbon | 50 | 33 | 1670 | 71 | 91 | 1915 |
| Lackawanna | 230 | 6970 | 141 | 977 | 1917 | 10235 |
| Lehigh | 307 | 520 | 1458 | 315 | 387 | 2987 |
| Lucerne | 255 | 3366 | 4586 | 303 | 595 | 9105 |
| Monroe | 376 | 313 | 253 | 150 | 848 | 1940 |
| Northhampton | 385 | 1121 | 1263 | 499 | 981 | 4249 |
| Pike | 17 | 7 | 26 | 87 | 6 | 143 |
| Schuylkill | 63 | 101 | 1886 | 48 | 40 | 2138 |
| Wayne | 8 | 20 | 12 | 124 | 18 | 183 |
| InSum $I_j$ | 1737 | 12486 | 11628 | 2658 | 4952 | 33461 |

Source: Cesario (1973), Table 5, p. 245.

# Estimated table using the QTP model

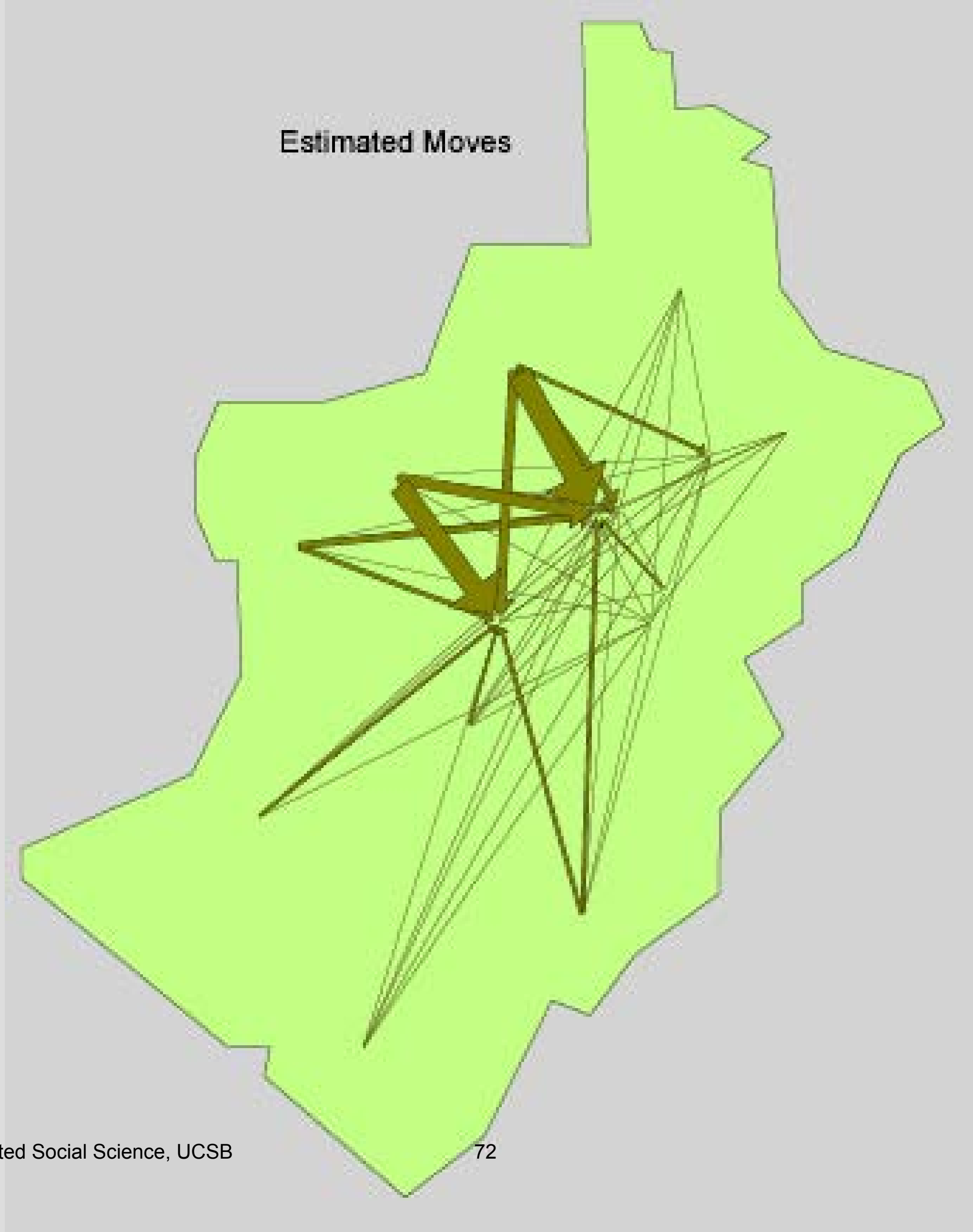
See QTP.doc under reprints on my web site for a description of the model.

Table 3: Results from model

| To Park / From County | Big Pocono | Gouldsboro | Hickory Run | Promised Land | Tobyhanna | **Pushes** |
|---|---|---|---|---|---|---|
| Berks | 31 | 197 | 211 | 42 | 86 | 4.04E-3 |
| Carbon | 110 | 531 | 889 | 109 | 275 | 7.30E-4 |
| Lackawanna | 394 | 5019 | 2242 | 929 | 1650 | 6.78E-5 |
| Lehigh | 208 | 1033 | 1071 | 237 | 438 | 1.85E-3 |
| Lucerne | 347 | 2614 | 4559 | 523 | 1062 | 4.86E-4 |
| Monroe | 191 | 737 | 417 | 23 | 364 | 0 |
| Northhampton | 325 | 1546 | 1292 | 377 | 709 | 1.68E-3 |
| Pike | 9 | 53 | 34 | 21 | 26 | 1.59E-3 |
| Schuylkill | 113 | 682 | 876 | 156 | 311 | 2.35E-3 |
| Wayne | 8 | 72 | 36 | 34 | 32 | 7.75E-4 |
| **Pulls** | 1.93E-3 | 1.58E-3 | 1.67E-3 | 2.32E-3 | 1.82E-3 | |

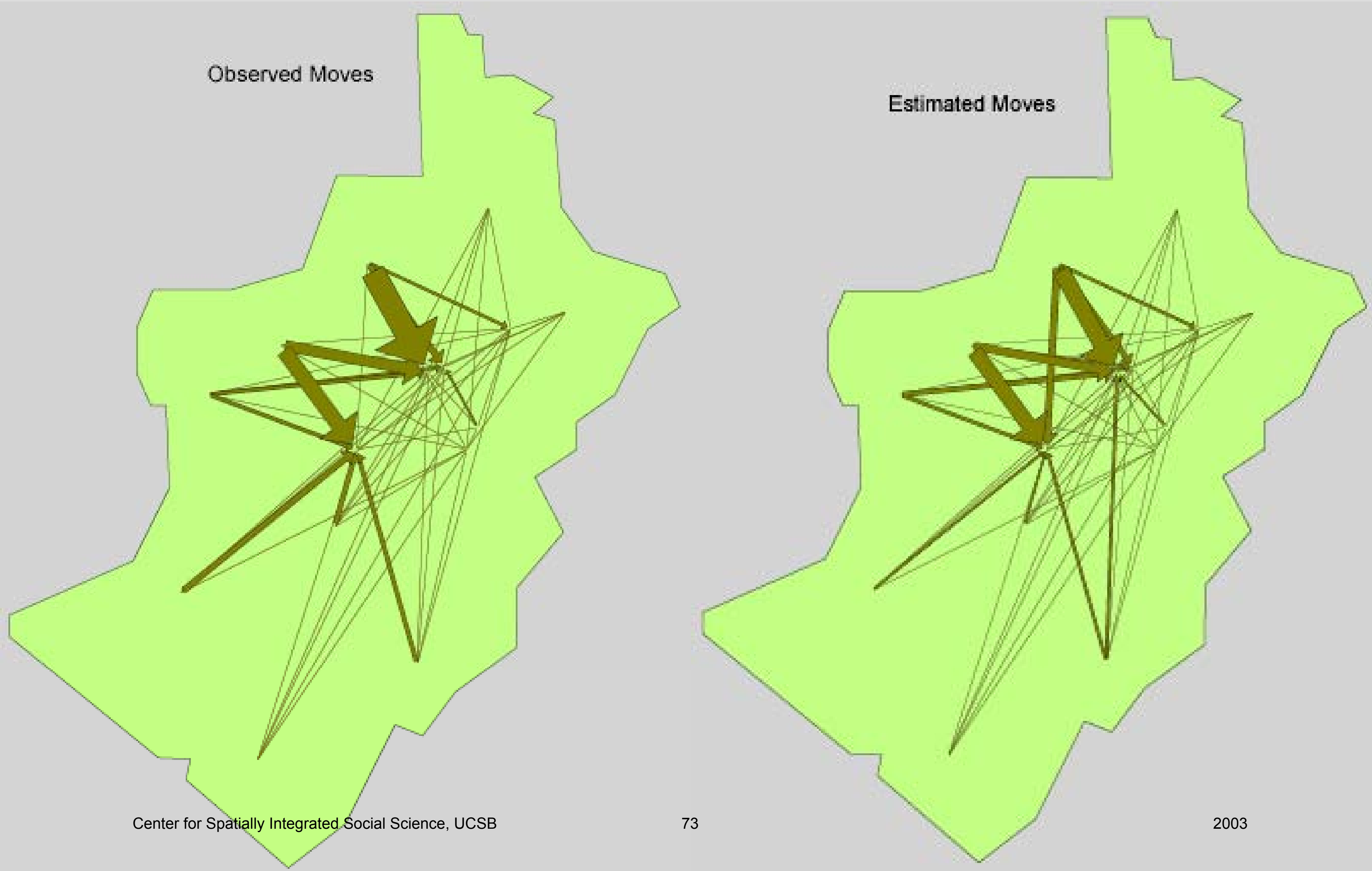
Correlation between observed and estimated: $r^2 = 0.85$, n=50

# Estimated Moves



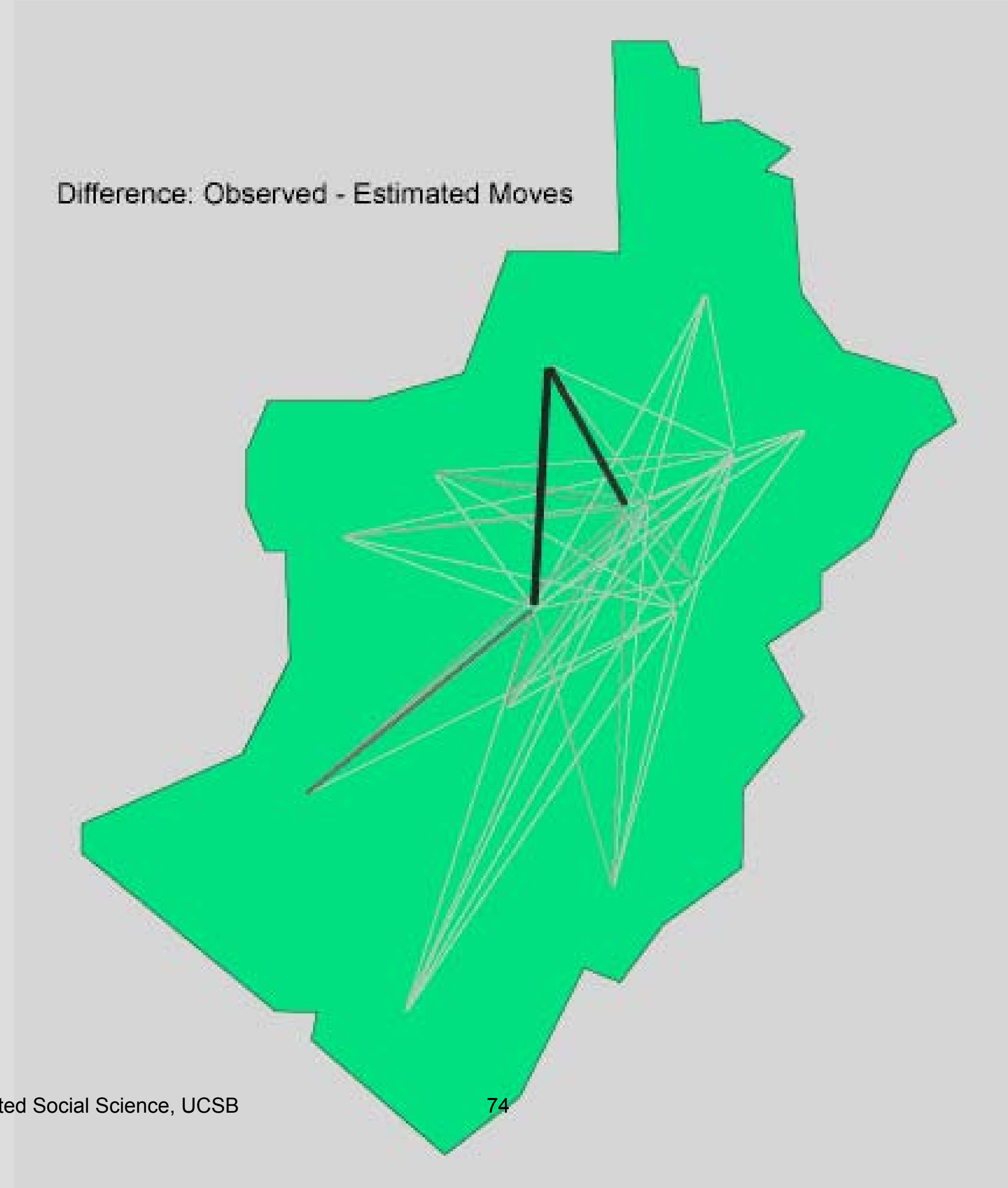
# Observed Moves versus QTP Estimated Moves

## Park attendance



# Difference: Observed minus Estimated

## Quadratic transportation model



Thank you for your attention

Questions can be addressed to:

Waldo Tobler
Professor Emeritus
Geography Department
University of California
Santa Barbara, CA 93016-4060
http://www.geog.ucsb.edu/~tobler

# End of part 2 of the tutorial

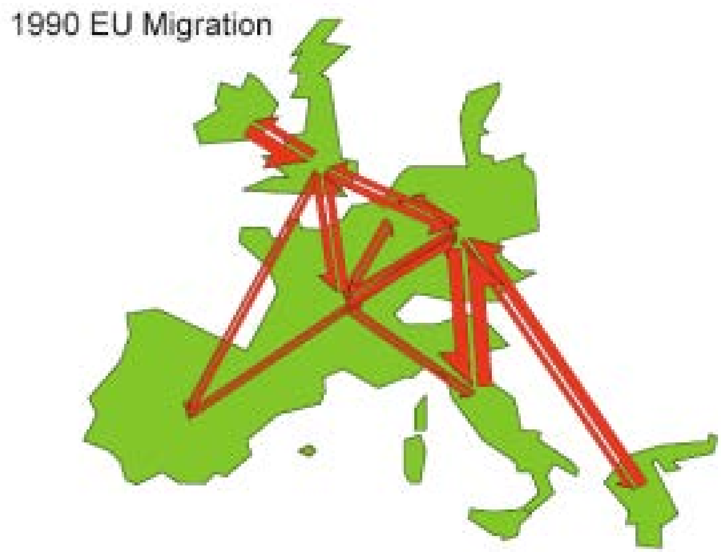
Now experiment with your own data or try some of the files that came with the program in the Data Sets folder, or continue with part 3 of the tutorial

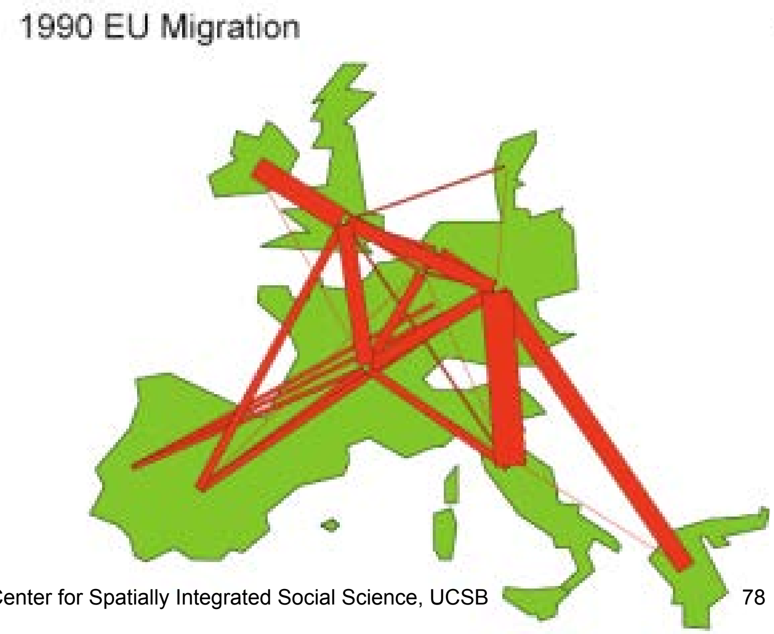
# Tutorial Part III

## Examples produced using the Flow Mapper program.

by Waldo Tobler

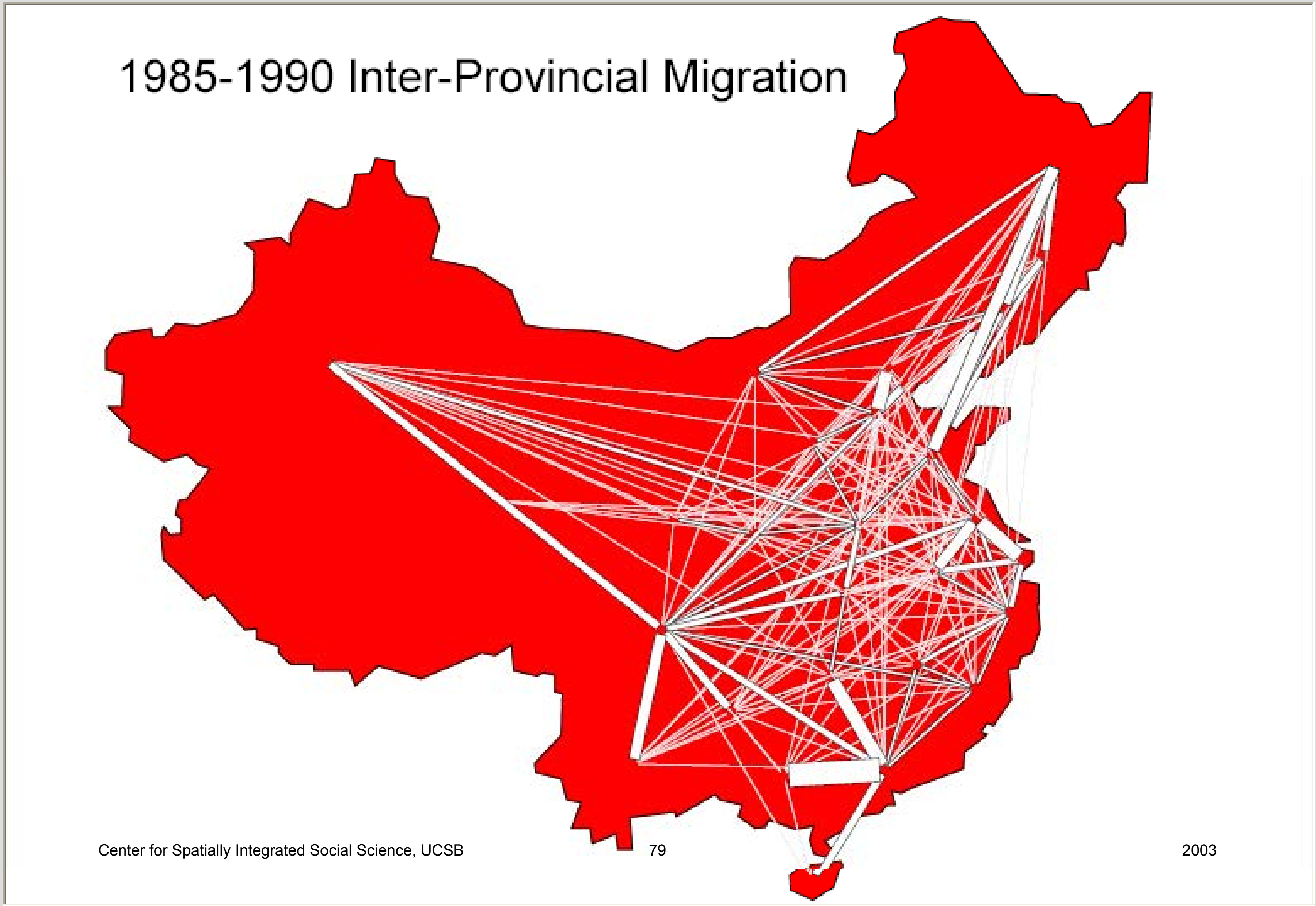
# Two-way, Total (Gross), and Net Migration



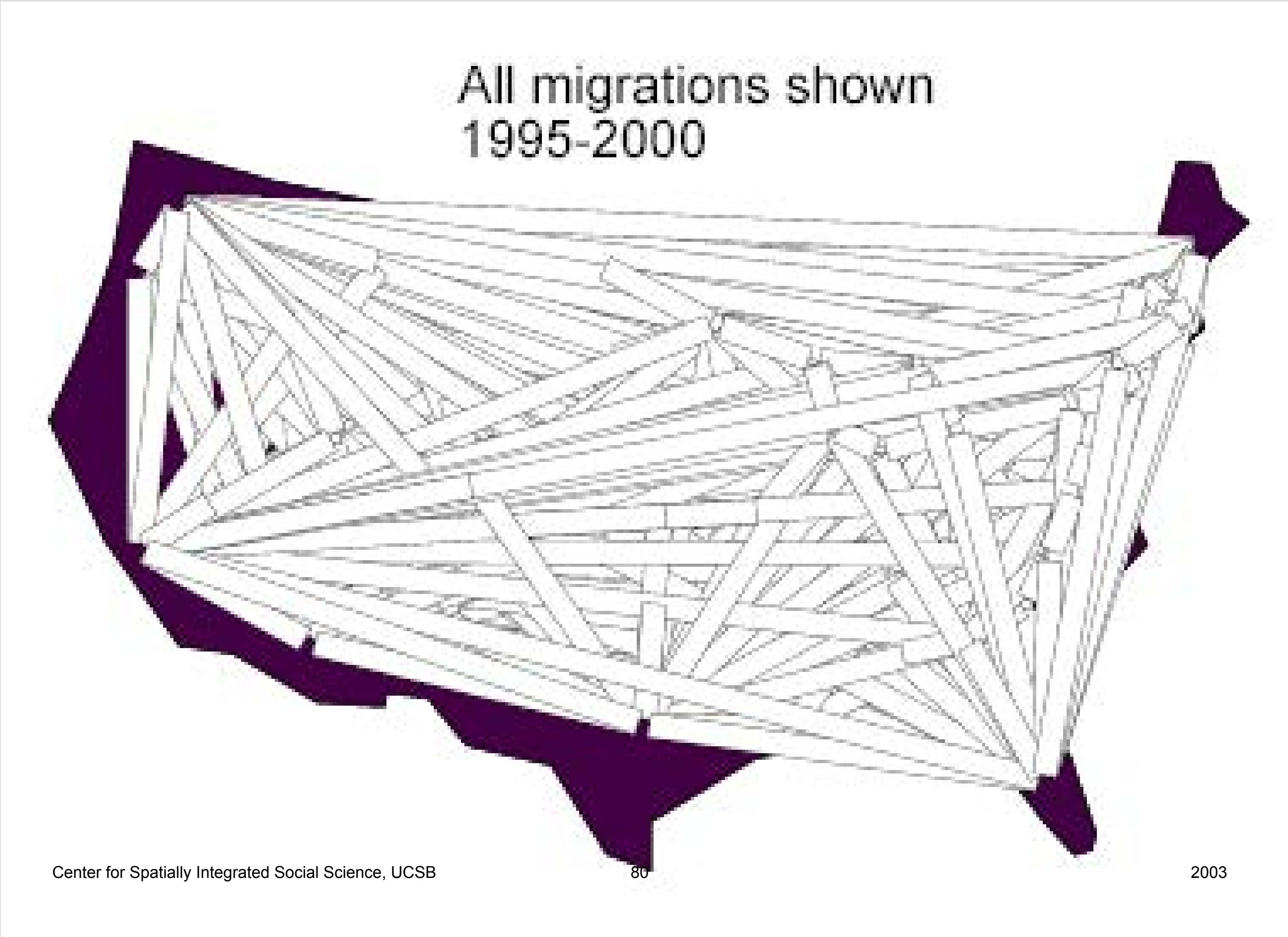

# Showing the majority of inter-provincial moves in China

## Using the flow mapper program



# Showing 2256 flows

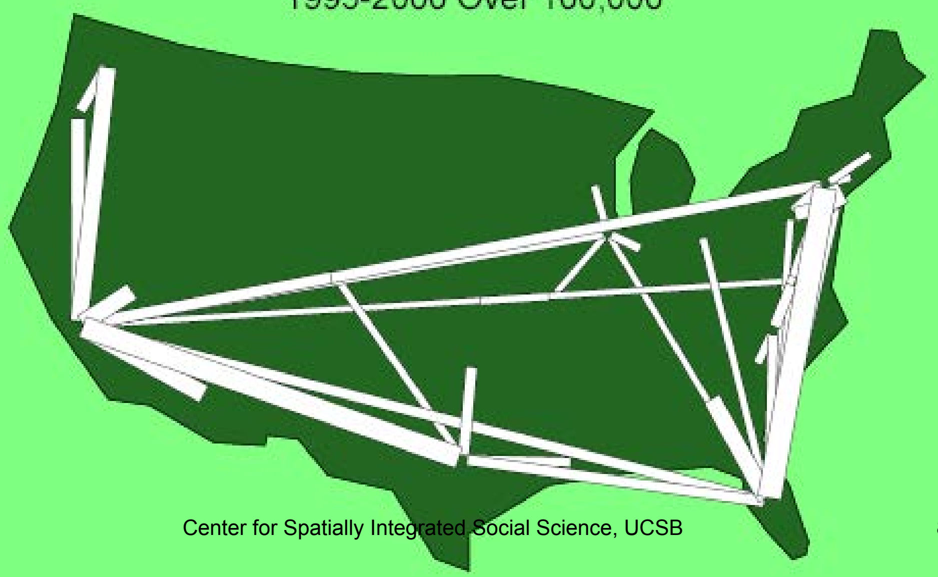
from 48 by 48 table, with constant width bands. Not very useful.



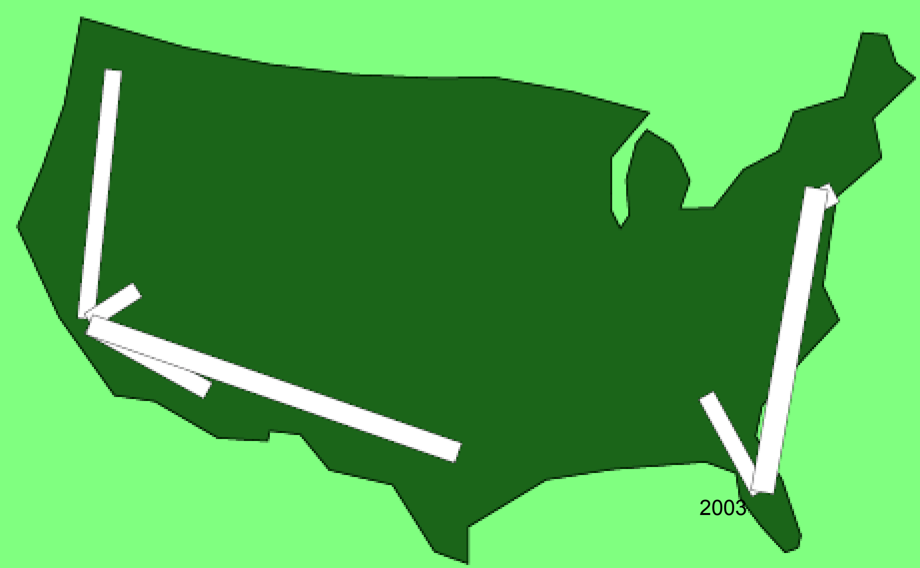
# 1995-2000 Total Migration

## Variable width bands, and parsing by quantity.

1995-2000 Over 100,000

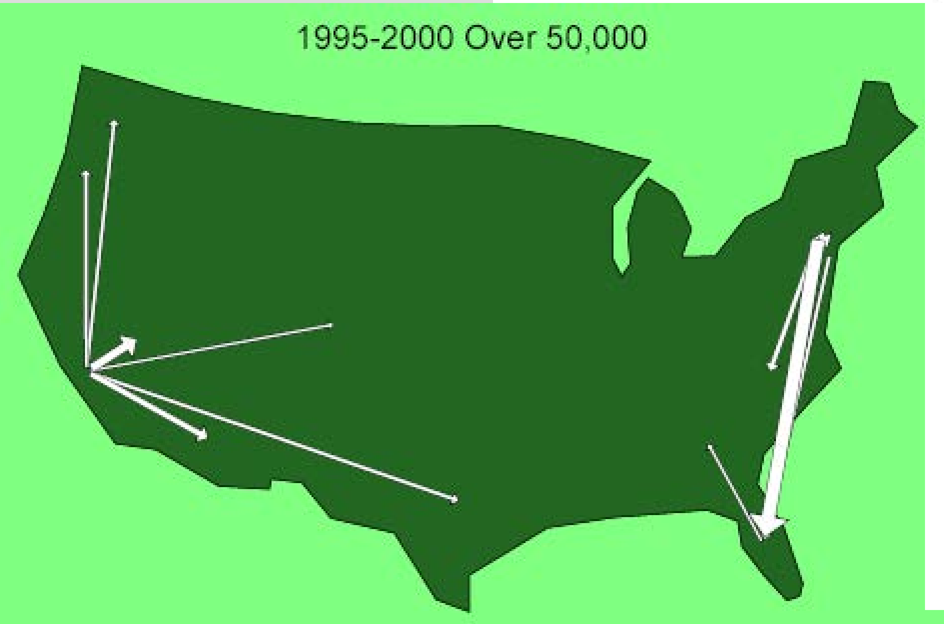


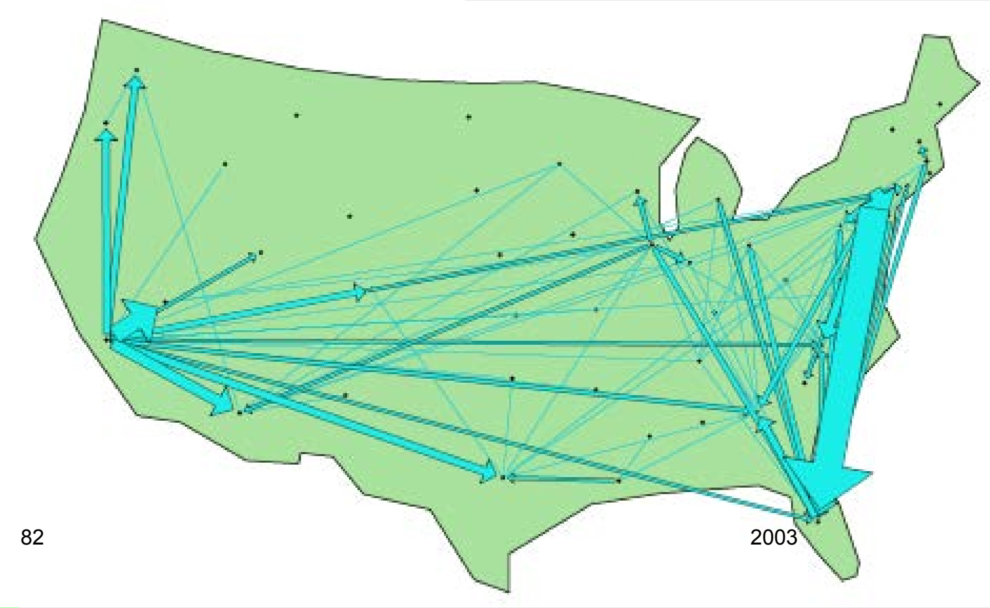
1995-2000 Over 200,000



# 1995-2000 Net Migration

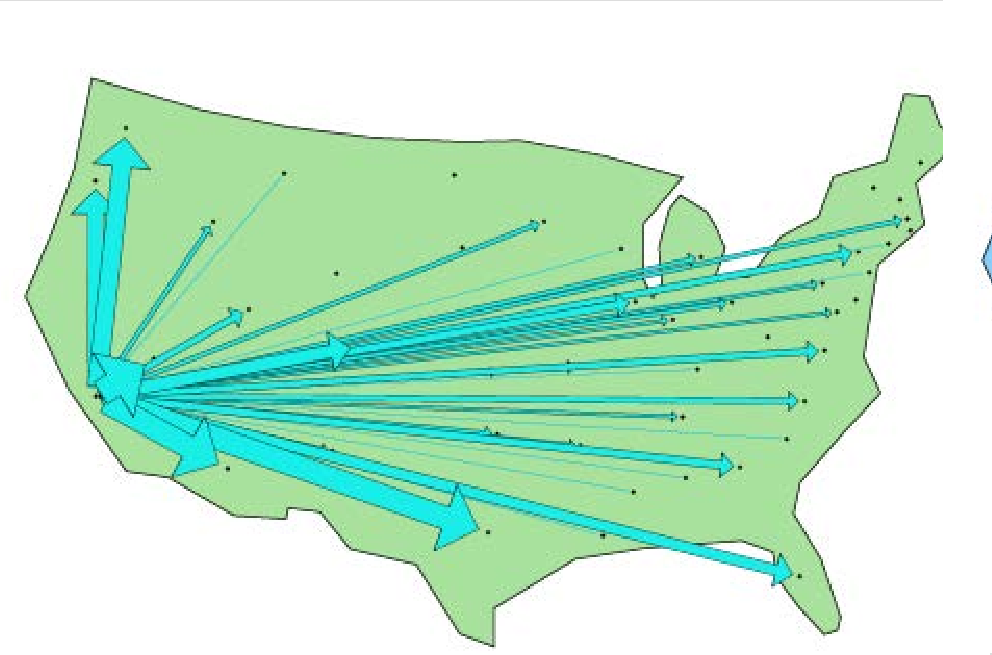
## Complete and simplified.



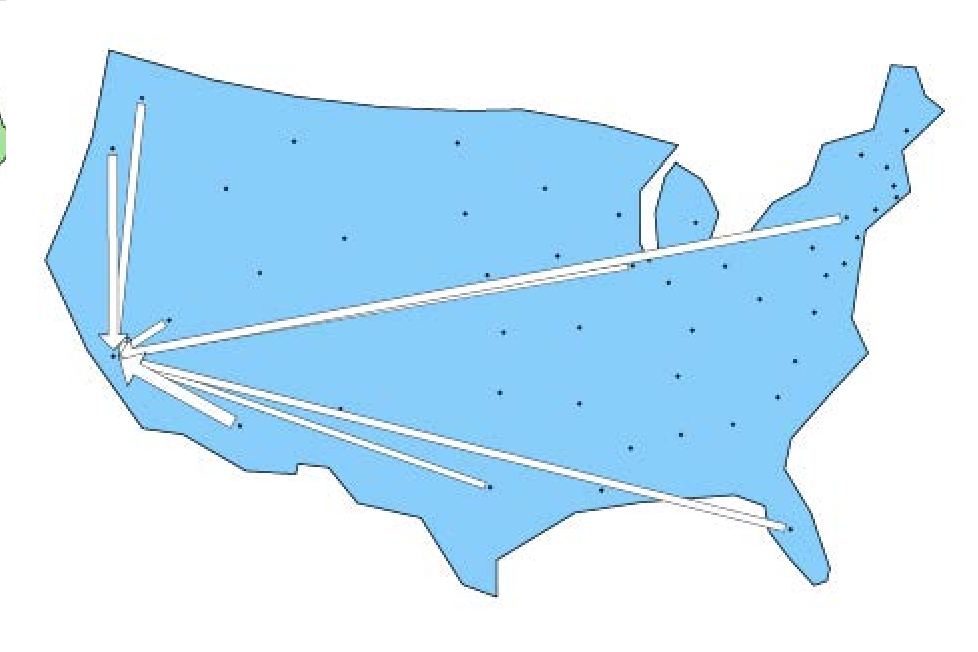



# 1995-2000 Migration from and to California

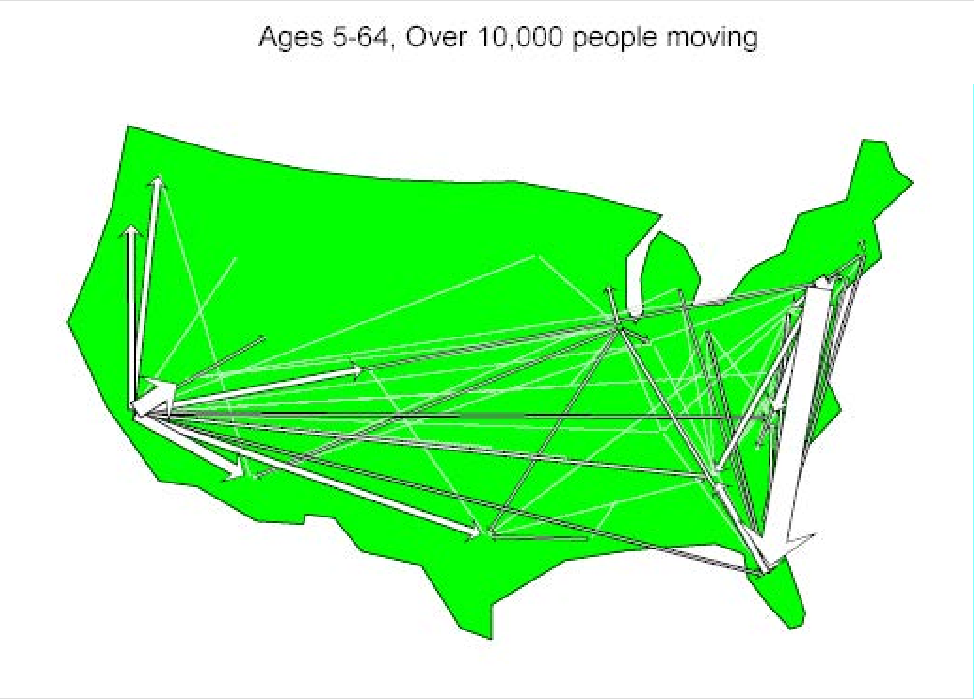
## Flows from CA

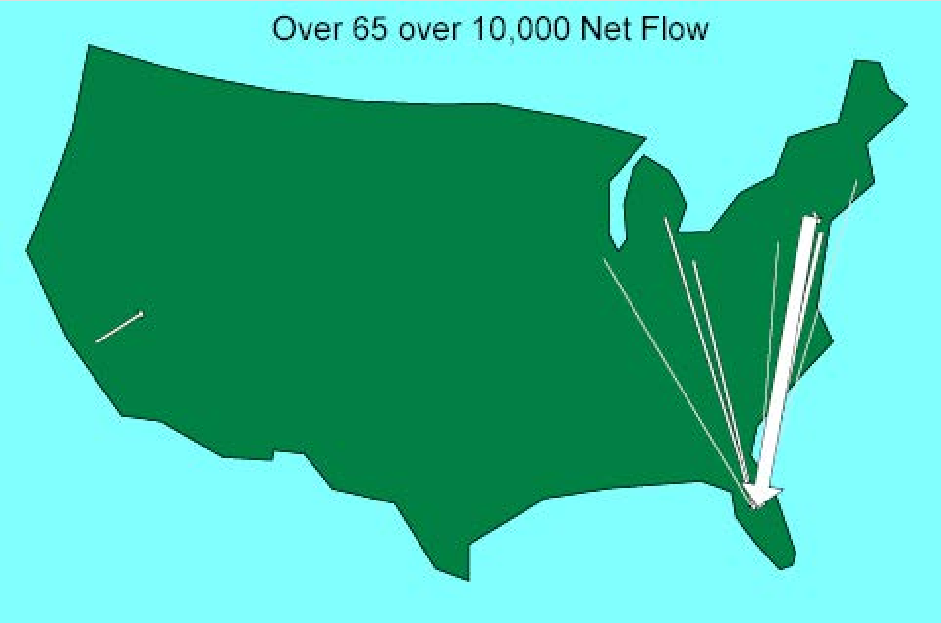


## Major flows to CA

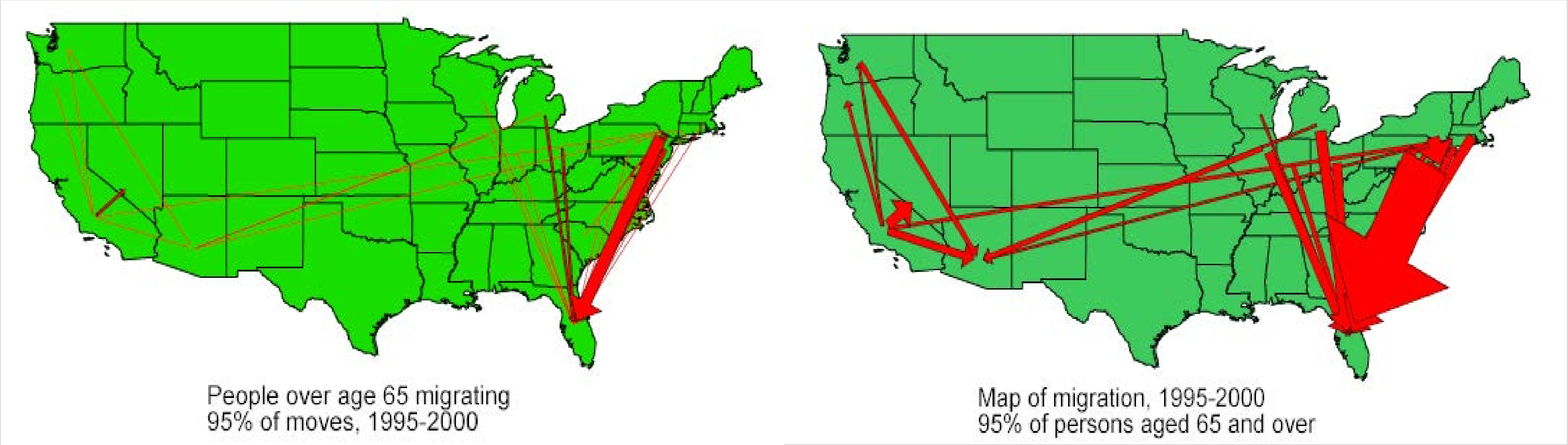


# 1995-2000 Net Migration
## by two age groups, and movement size.



# Two Variants
## Same Data



# Migration Patterns Persist

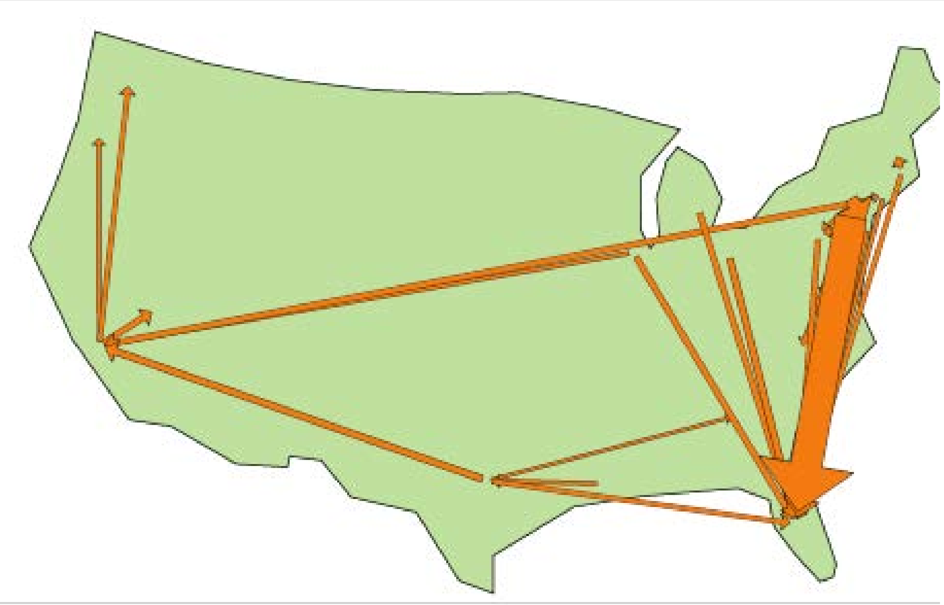
## the Netherlands

1984 1994

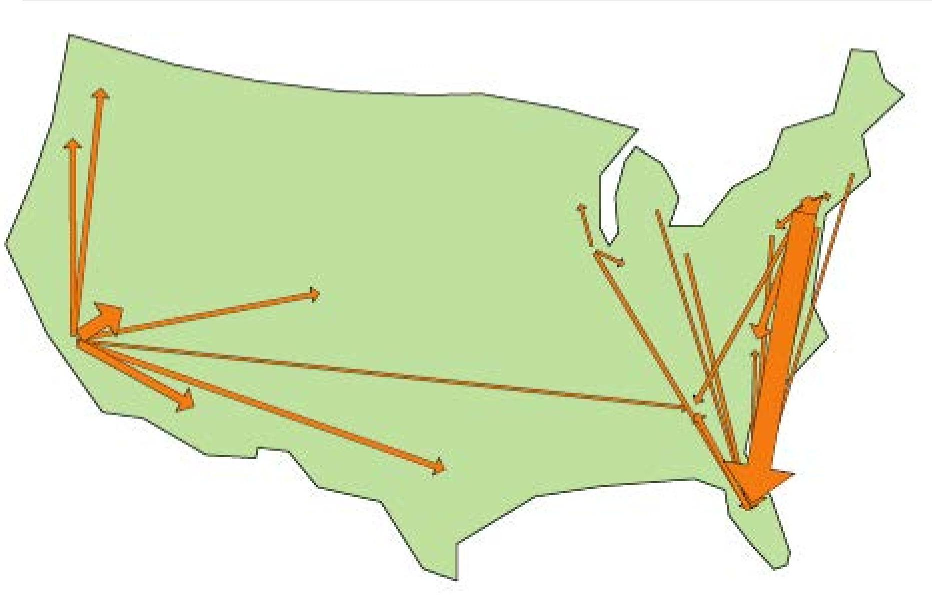
# Net Migration in the United States

US Census Data

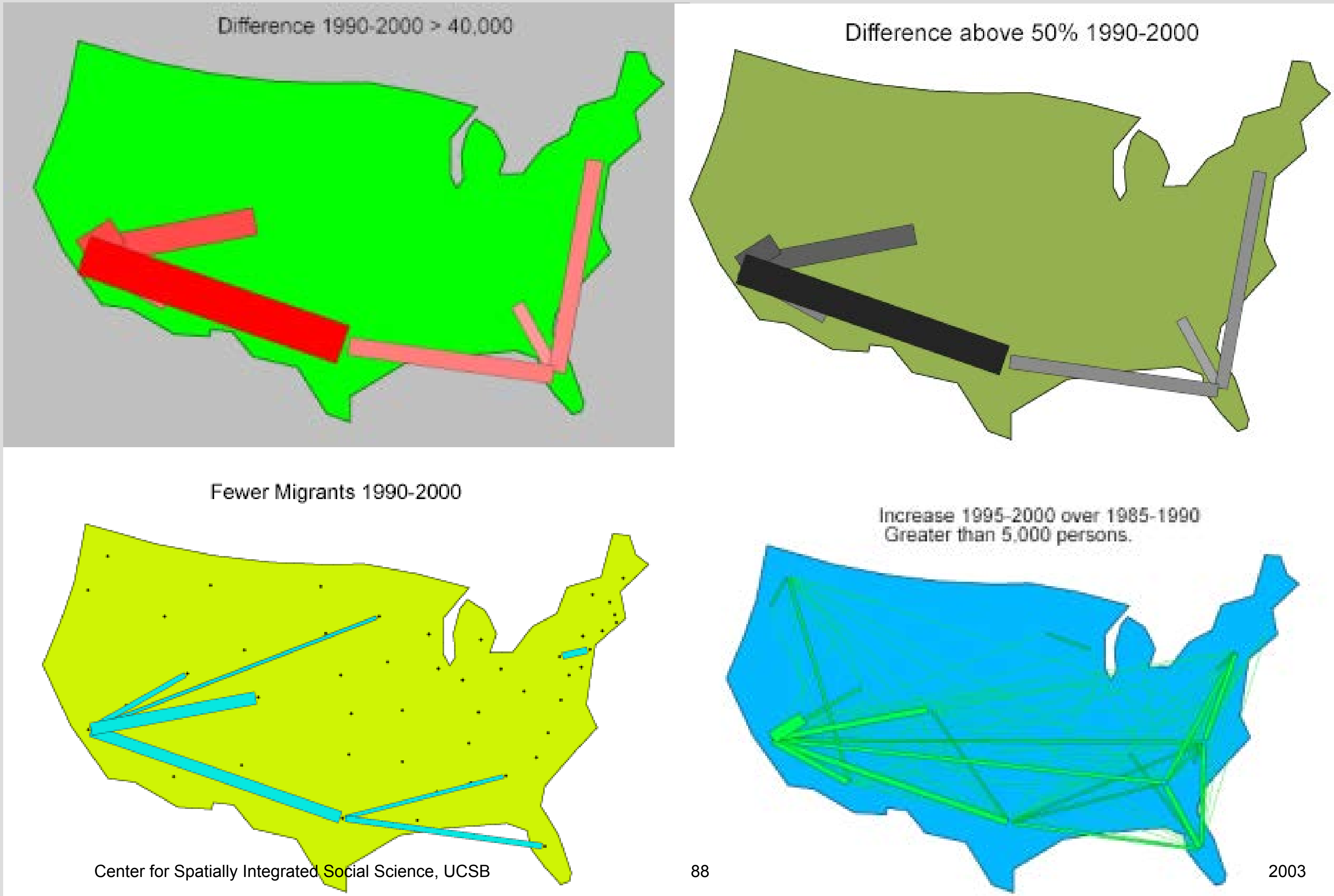
1985-1990



1995-2000



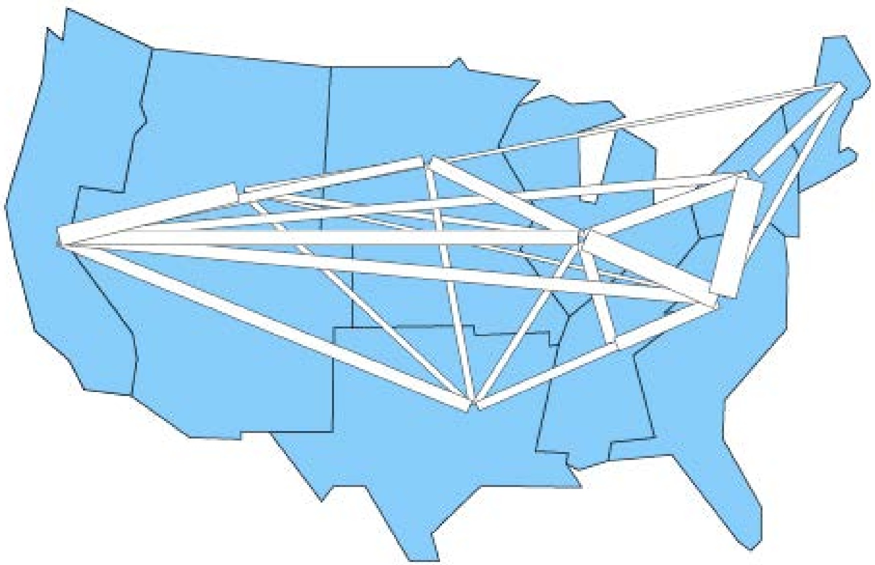
# Difference between 1985-1990 and 1995-2000 Migration
## US Census information

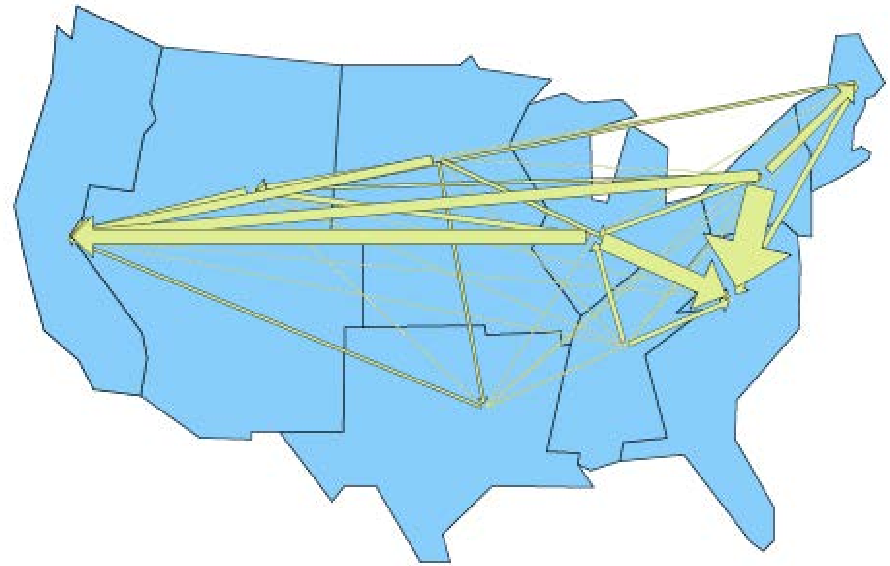


# Migration by Census Divisions

Top: 1965-1970 Migration, Total and Net

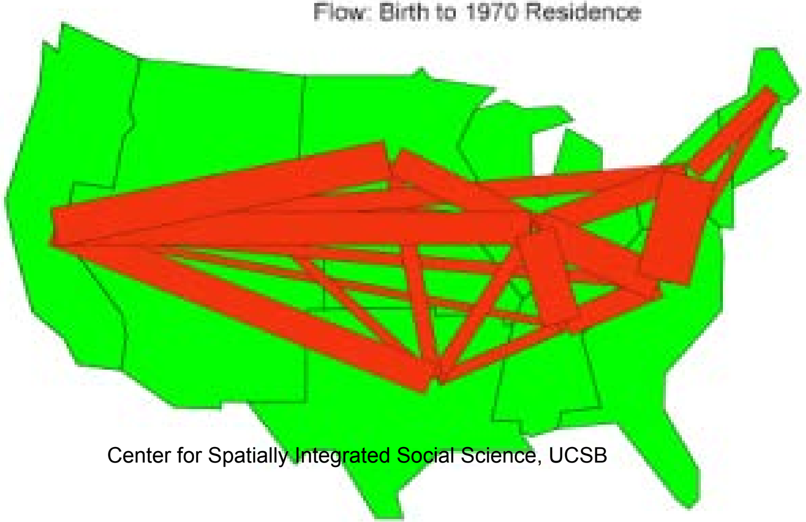
Bottom: Birth to 1970 Residence, Total and Net

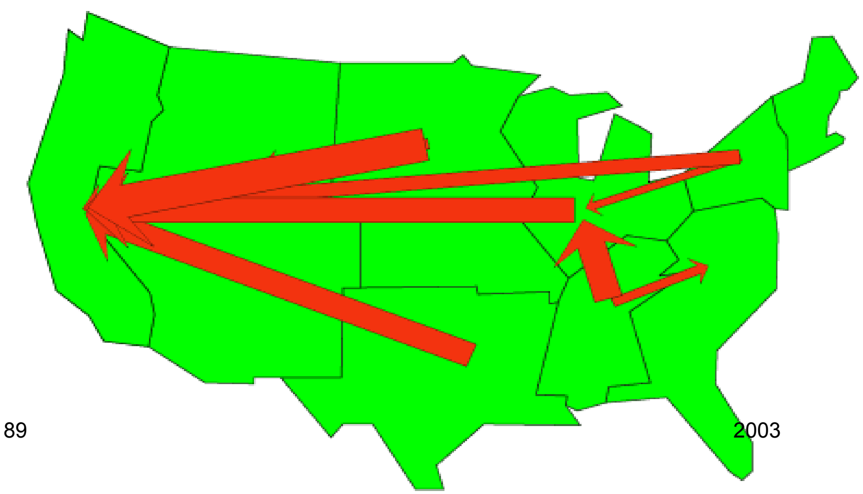




Flow Map: Birth to 1970 Residence

Flow: Birth to 1970 Residence

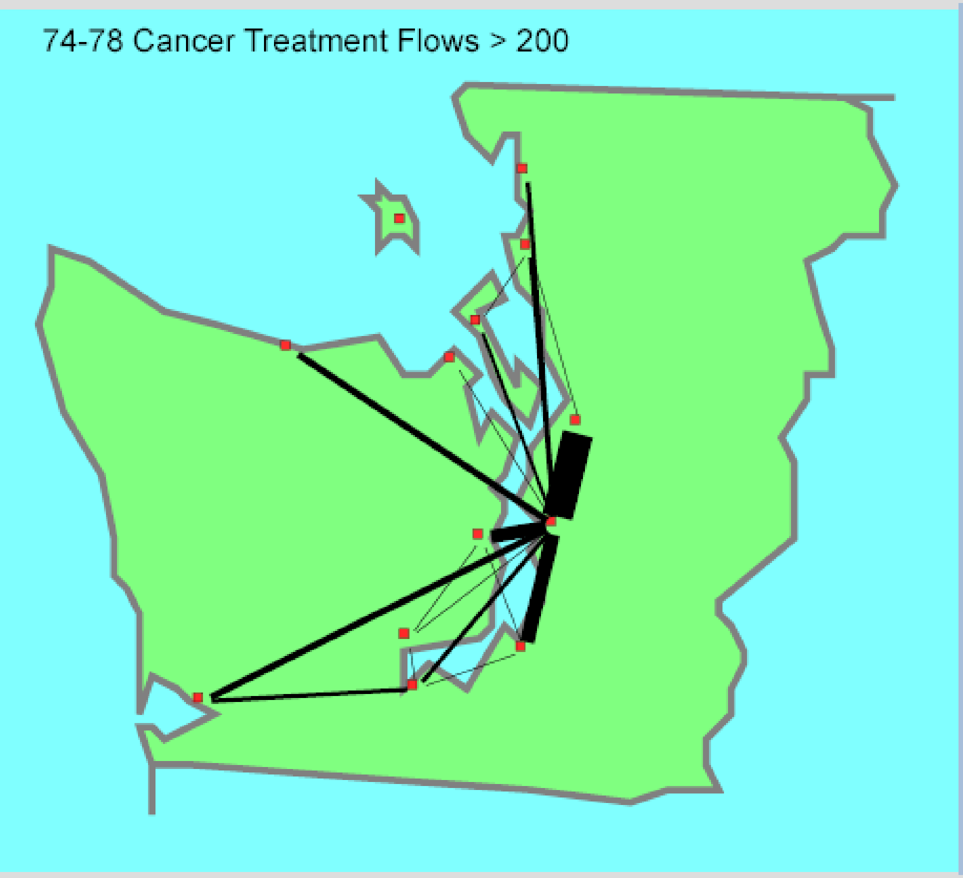


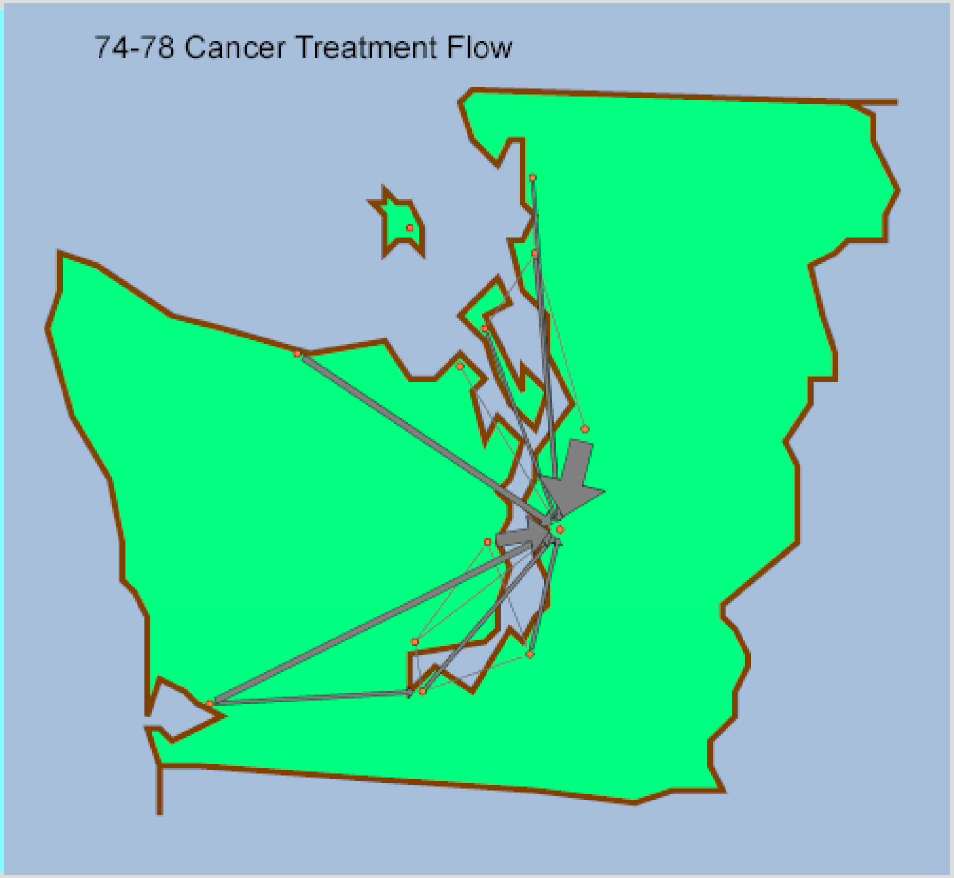


# Gross and Net Moves

$M_{ij} + M_{ji}$ and $|M_{ij} - M_{ji}|$

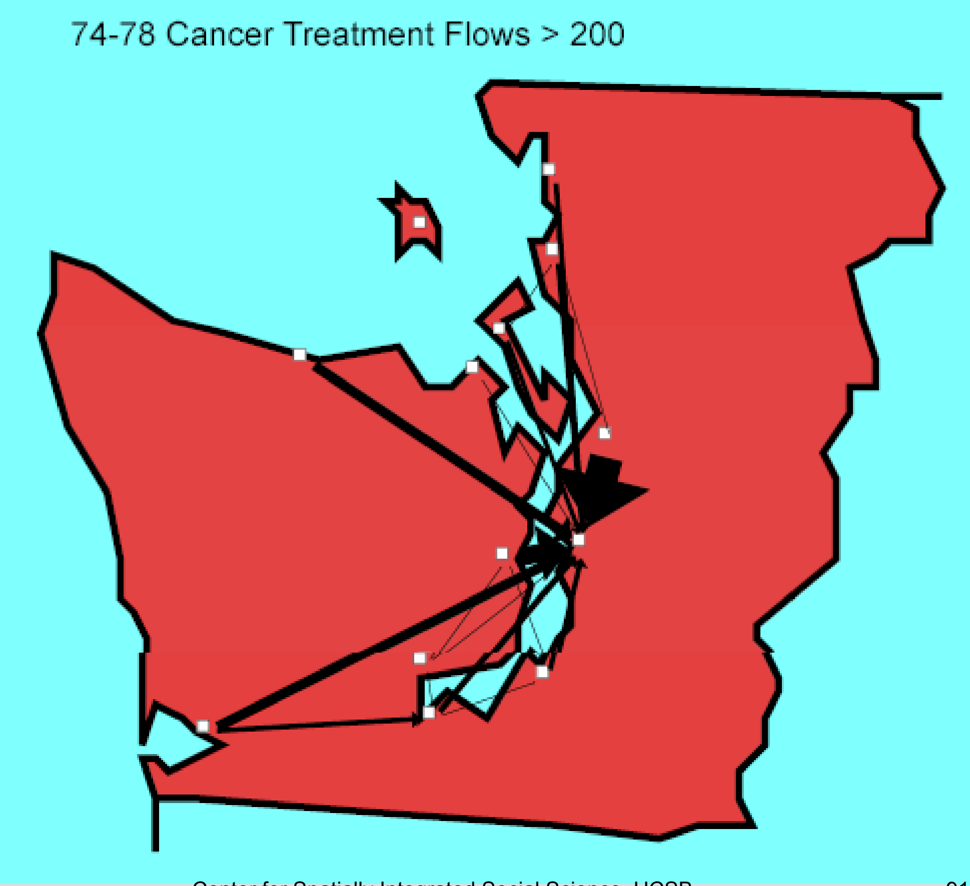
in Western Washington state.

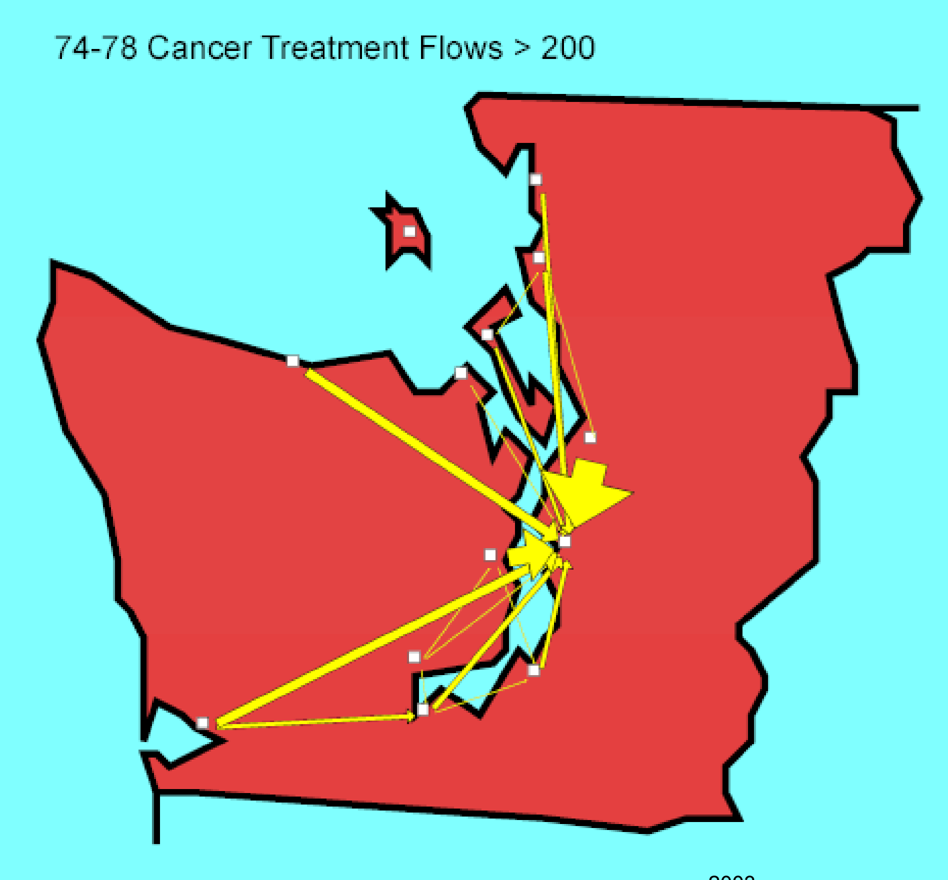


# Variations in style

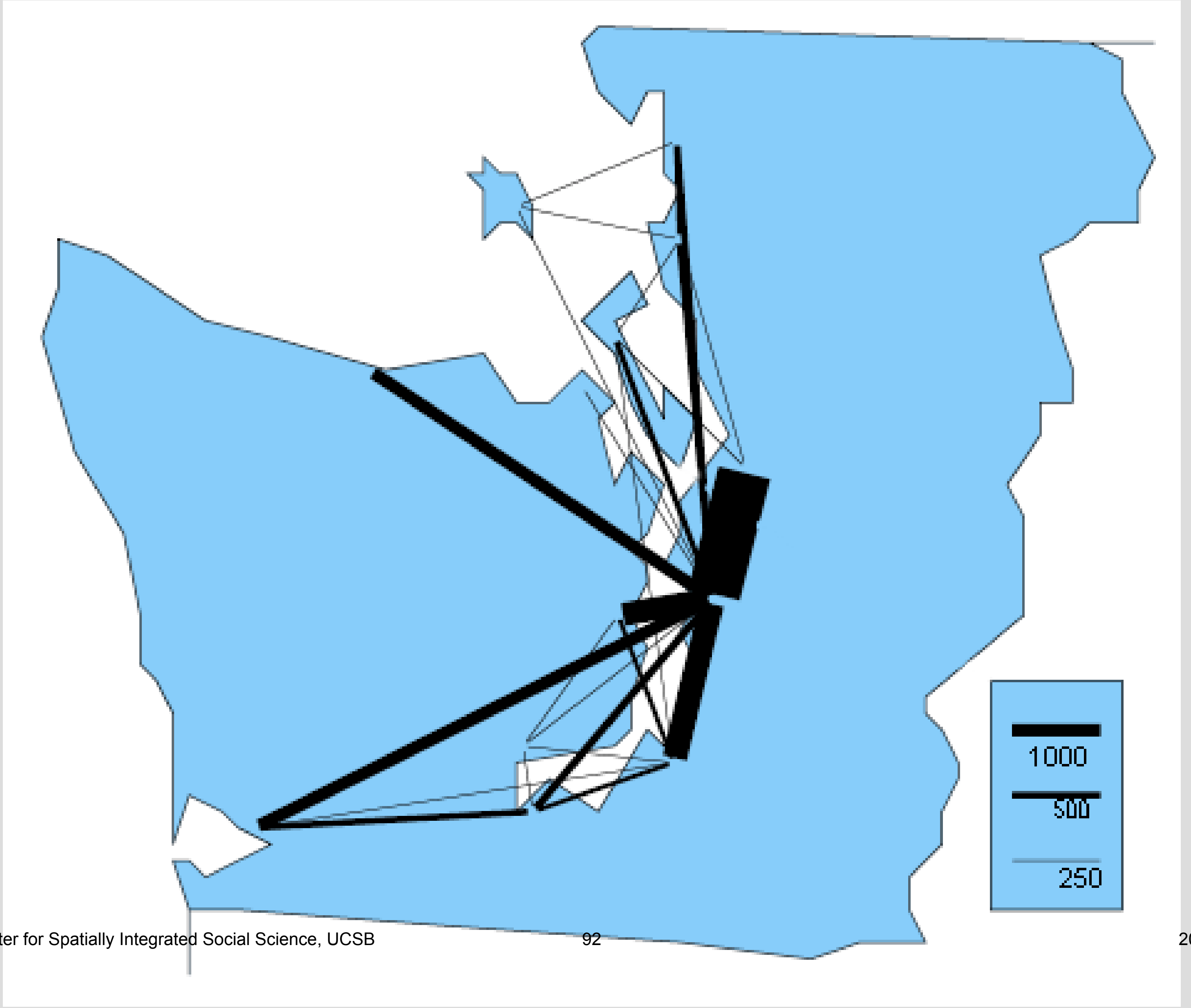
## With islands, showing centroids, and title.



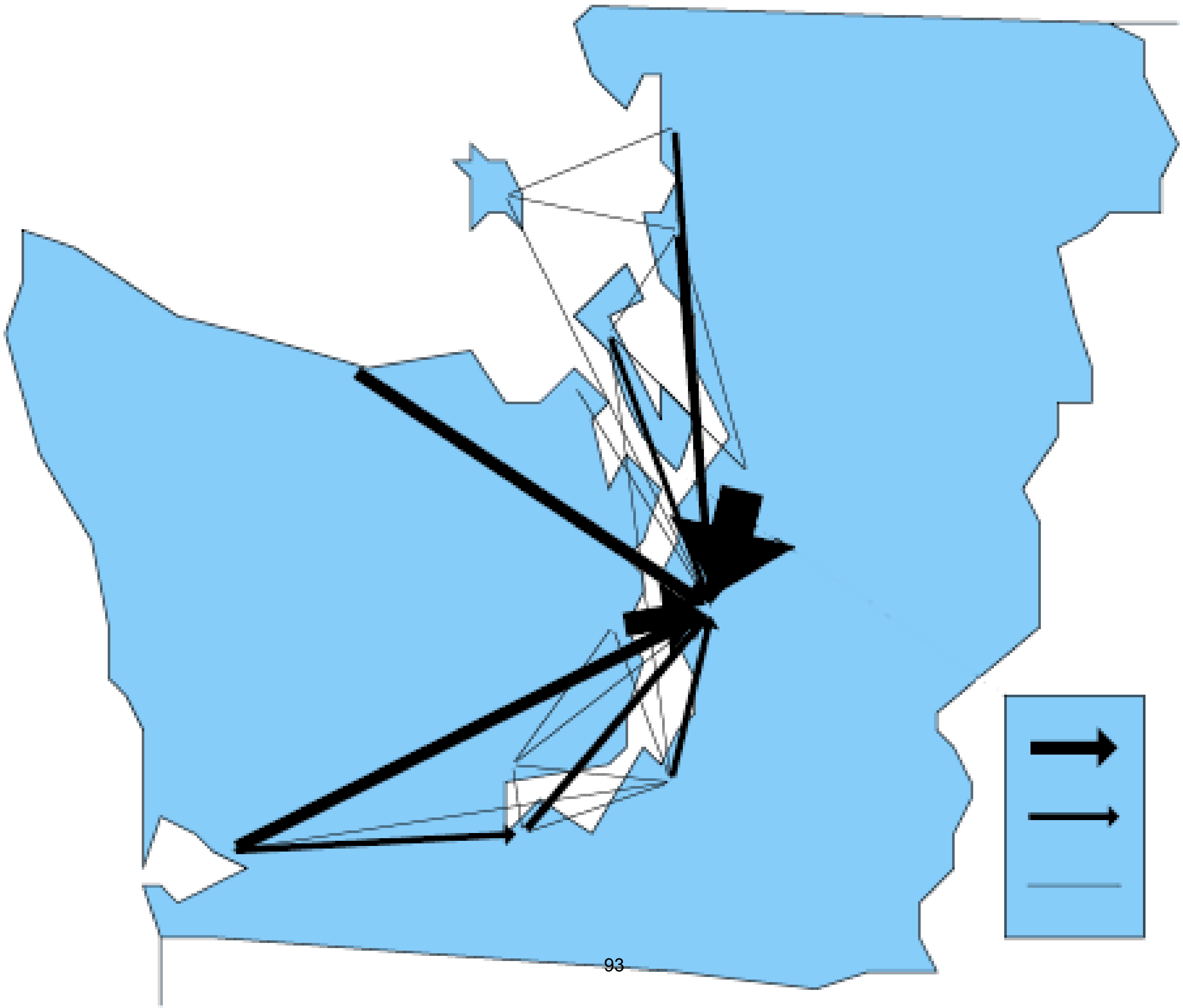



# Legend Box

A legend box (an “island”) with gross moves. Numbers added later.



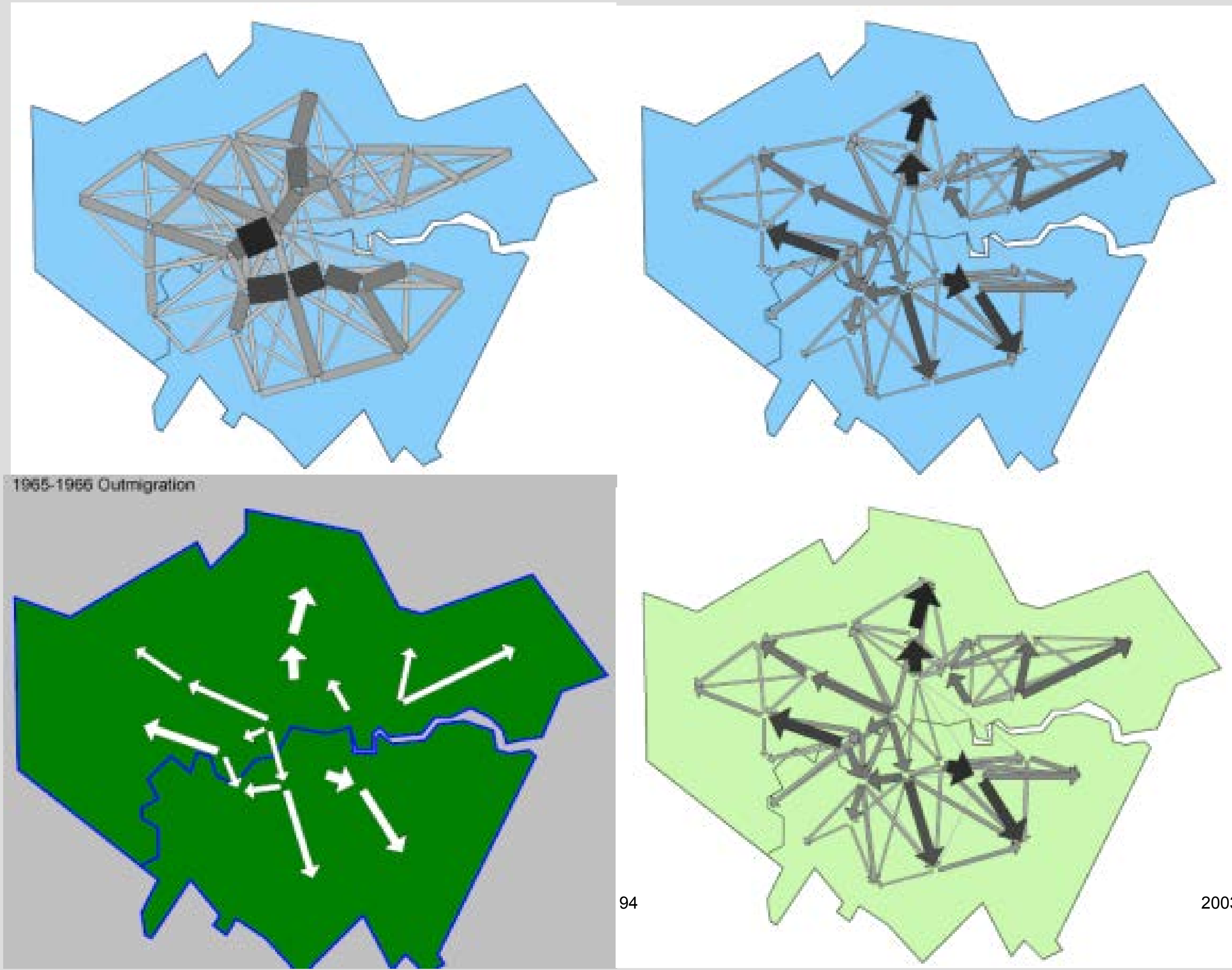
# Legend box for net moves



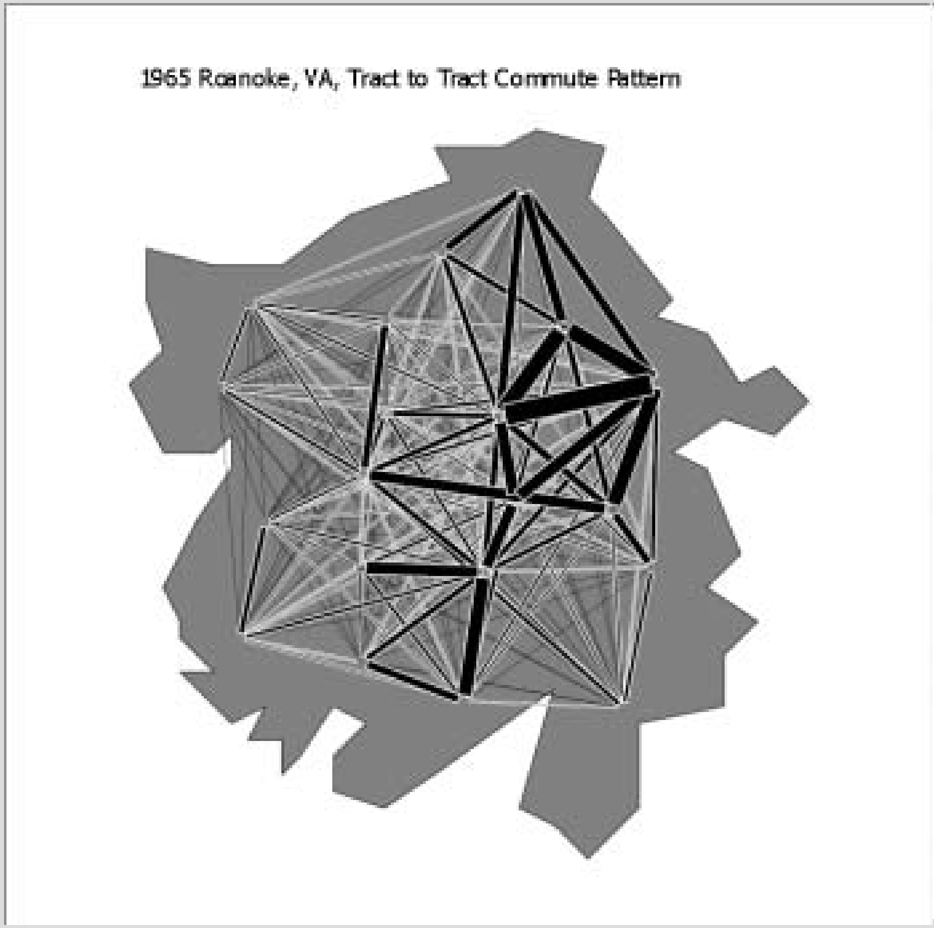
# London 1965-1966 Inter-borough migration

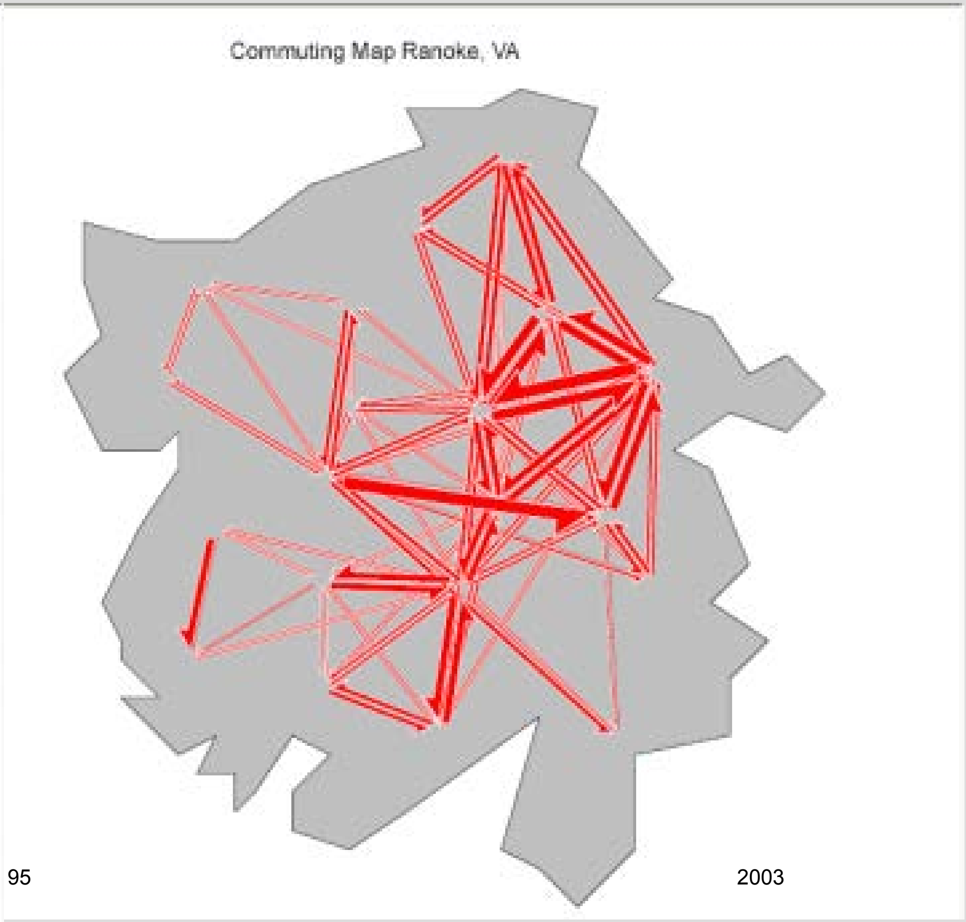
from 33 boroughs.

Exploration of map styles, especially colors,



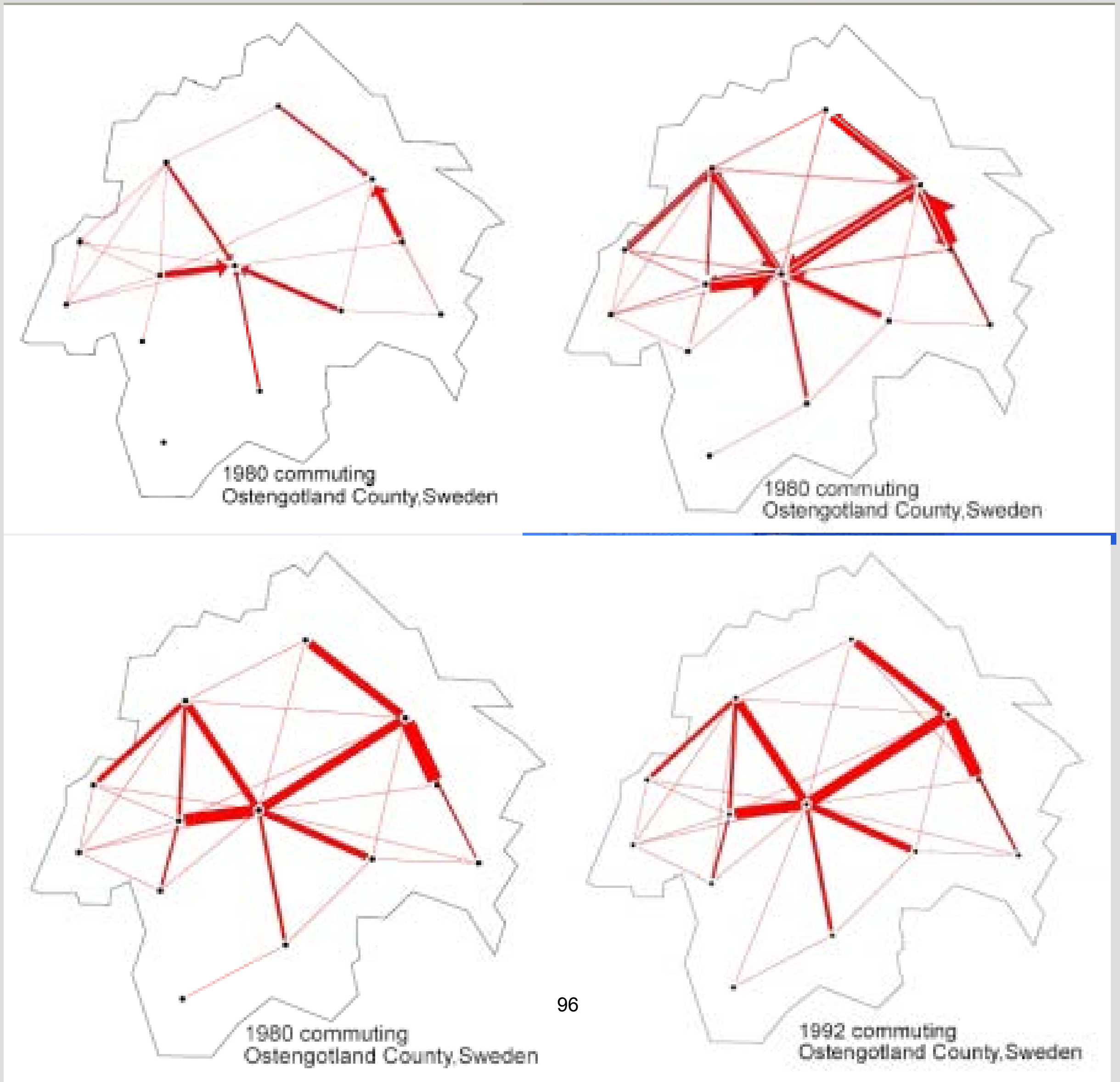
# Commuting Pattern in Roanoke, VA, 1965
## By Census Tract



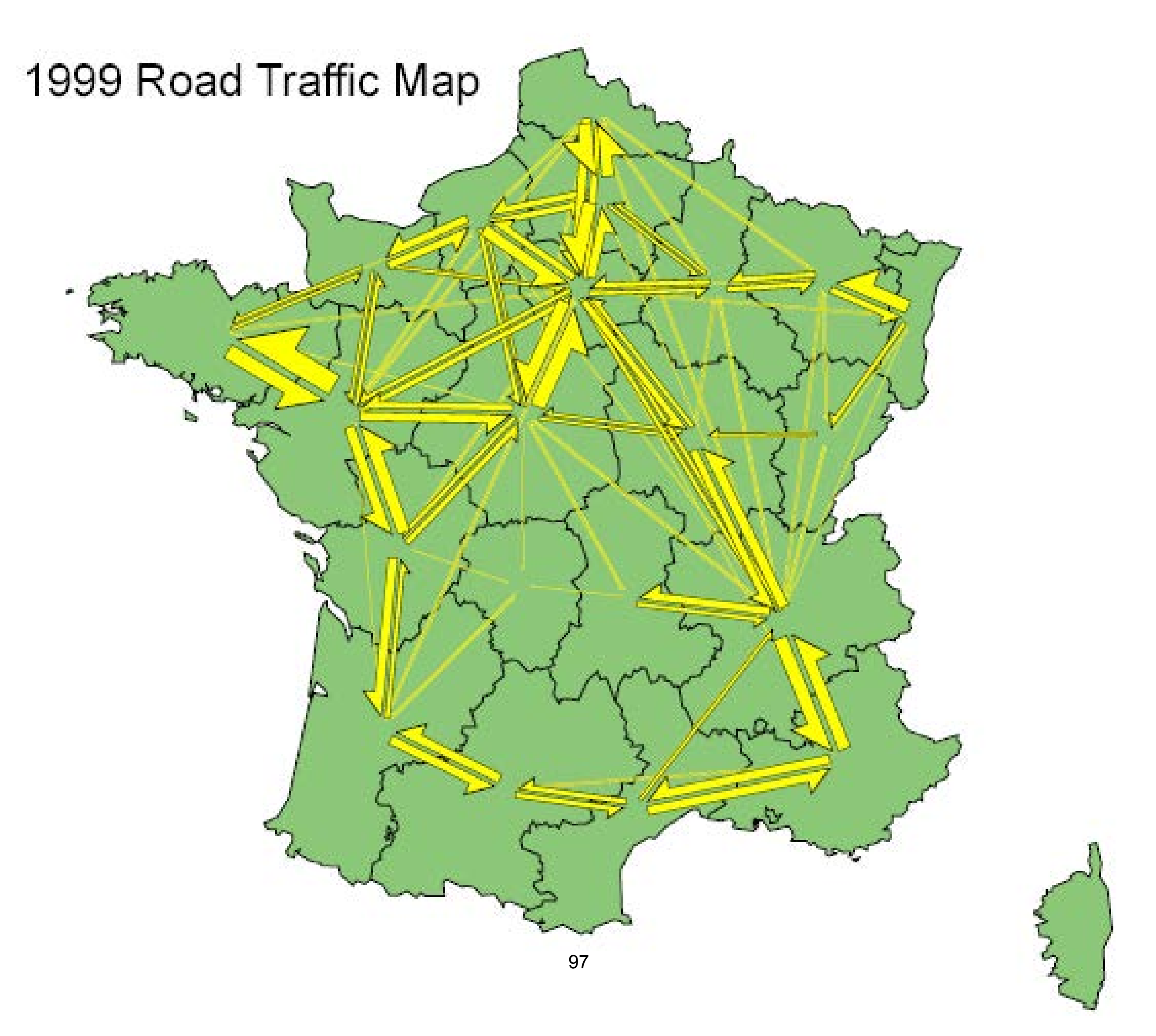
# Commuting in Östengötland, Sweden,

1980 (net, two-way, and total) & 1992 (total)
From P. Åberg (1998)



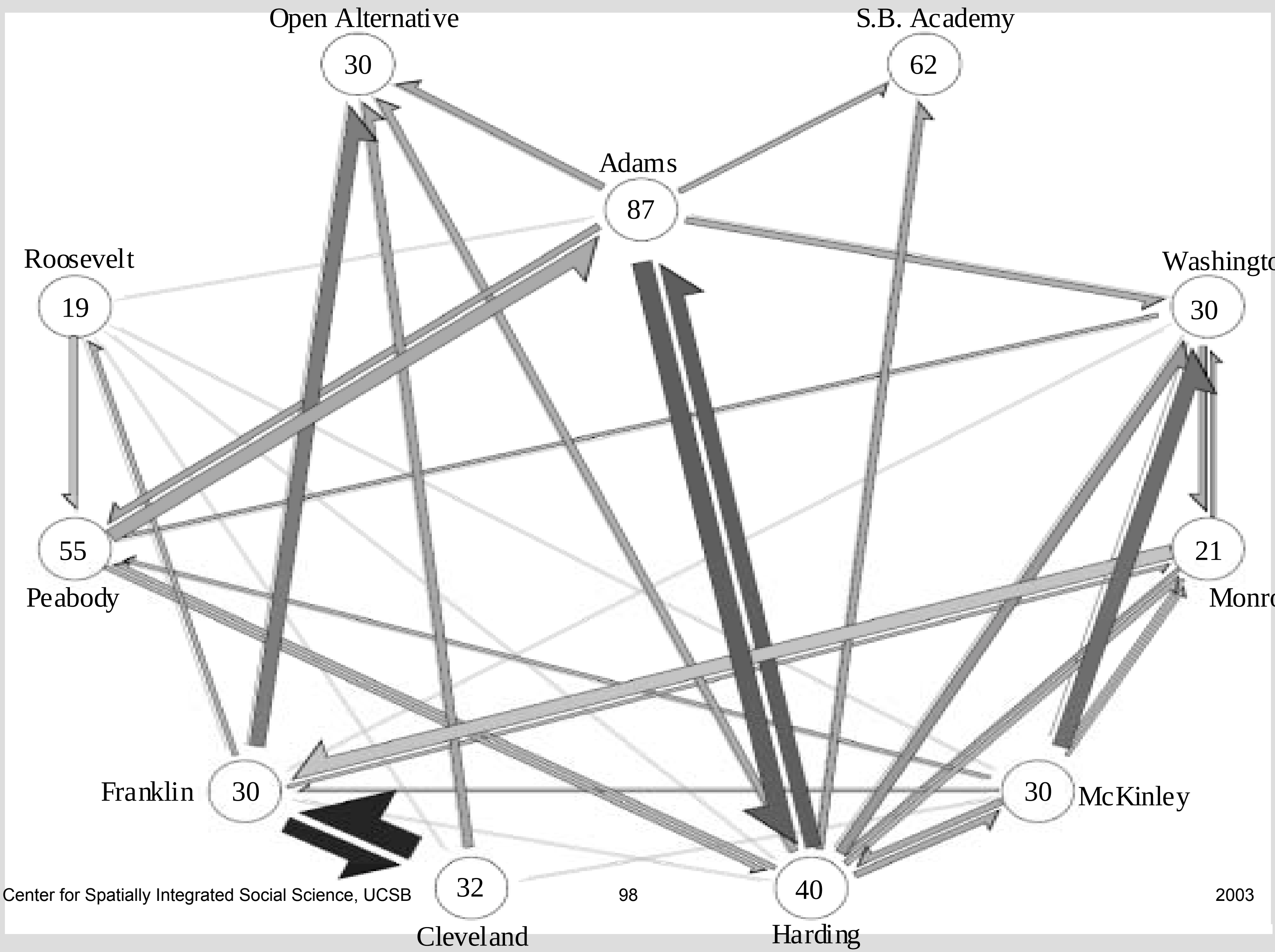
# Movement between French Regions

Data courtesy of Mr. C. Calzada of Paris

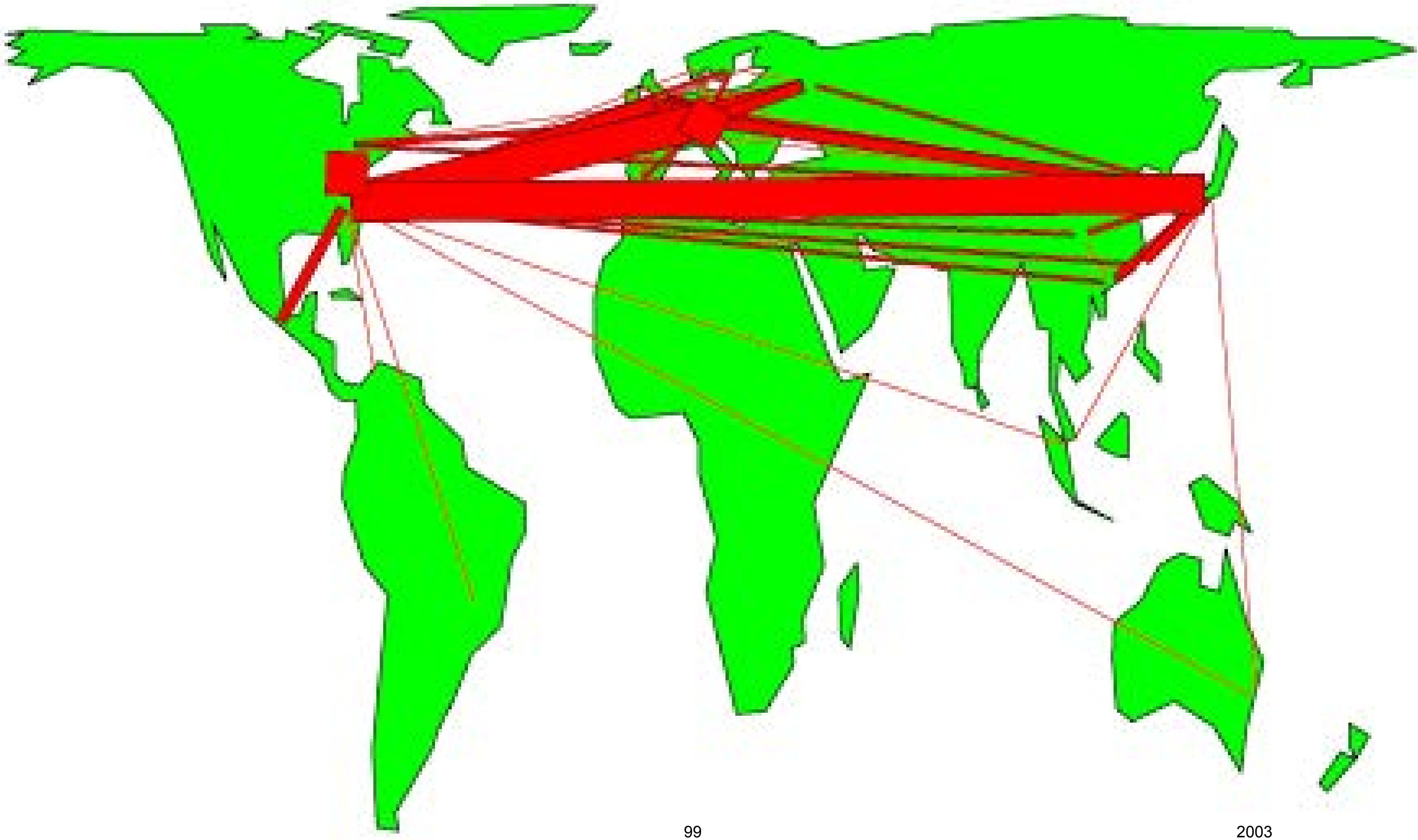


# Transfers between eleven schools in Santa Barbara

School locations adjusted for clarity. Courtesy of Dr. Stuart Sweeney.



# Major World Trade
# 1978 Estimate



99

# The next slide shows a non-geographic map

The diagram is based on a 23 by 23 table of referrals from one scientific field to another from a very large multi-year file of citations. For details see K. Boyack, 2004, *Proc.*, NAS, 101, suppl.1, 5192-5199.

The fields are positioned spatially using an ordination based on the from-to table.

The 'data points' are enlarged to show the labels.

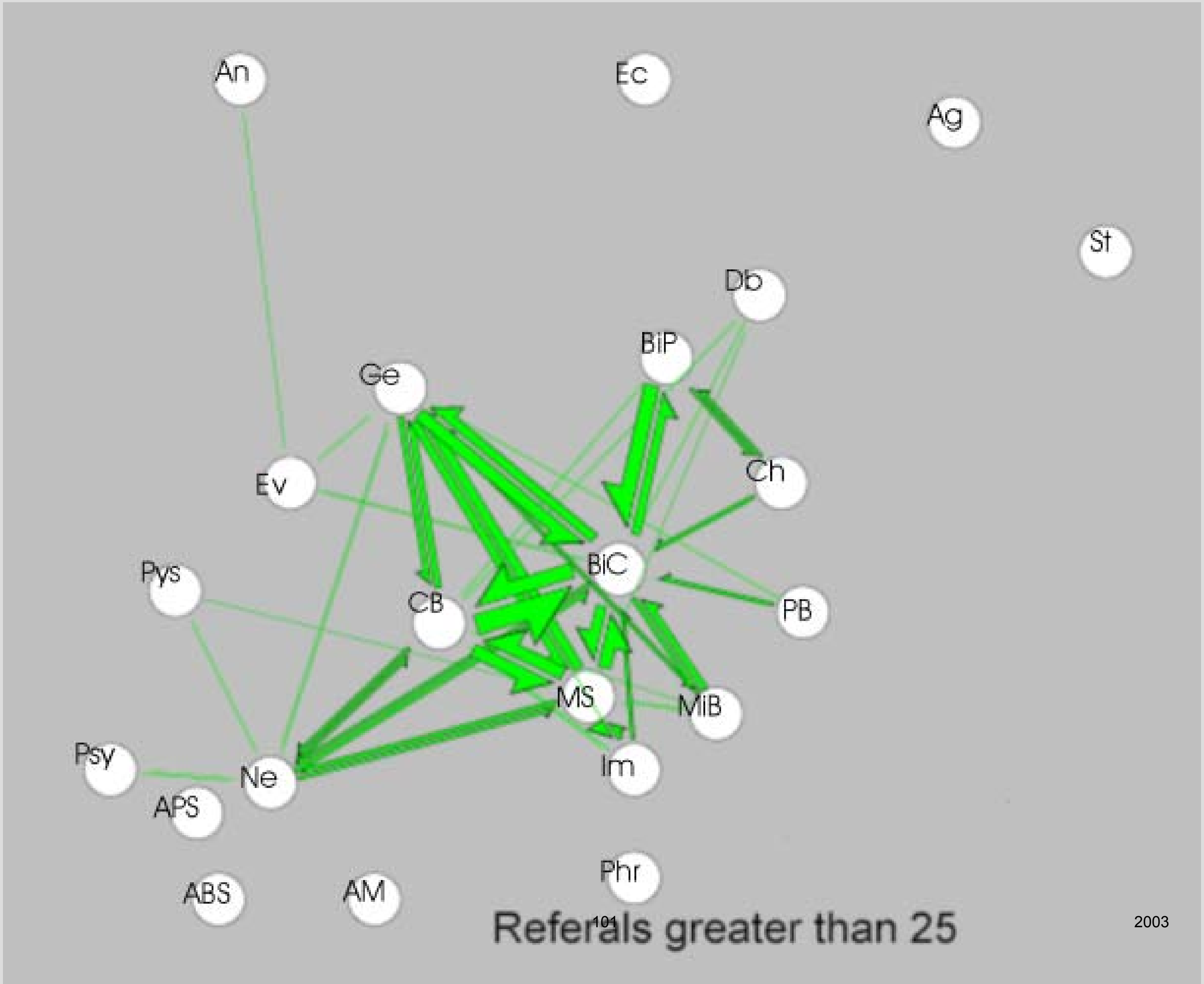
Two-way flows above 25 referrals are shown.

Inter-industry, input-output, or other non-geographic tables, can also be rendered in this fashion.

The fields are:

Ag-Agricultural sciences
An-Anthropology
ABS-Applied Biological Sciences
AM-Applied Mathematics
APS-Applied Physical Sciences
BiC-Biochemistry
BiP- Biphysics
CB-Cell Biology
Ch-Chemistry
DB-Developmental Biology
Ec-Ecology
Ev-Evolution
Ge-Genetics
Im-Immunology
MS-Medical Sciences
Mi-Microbology
Ne-Neurobiology
Phr-Pharmacolgy
Phy-Physiology
PB-Plant Biology
Po-Population Biology
Psy-Psychology
St-Statistics

# Journal to journal referrals between scientific fields



# End of Tutorial

Thank You For Your Attention

NOW experiment with your own data or try some of the files that came with the program in the Data_Sets folder, or repeat part 1 of the tutorial.

Comments or samples of your work done with the flow mapper program are appreciated. Send them to:

Waldo Tobler
Professor Emeritus
Geography department
University of California
Santa Barbara, CA 93106-4060
http://www.geog.ucsb.edu/~tobler

# 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.

<u>The CSISS flow mapping program</u>

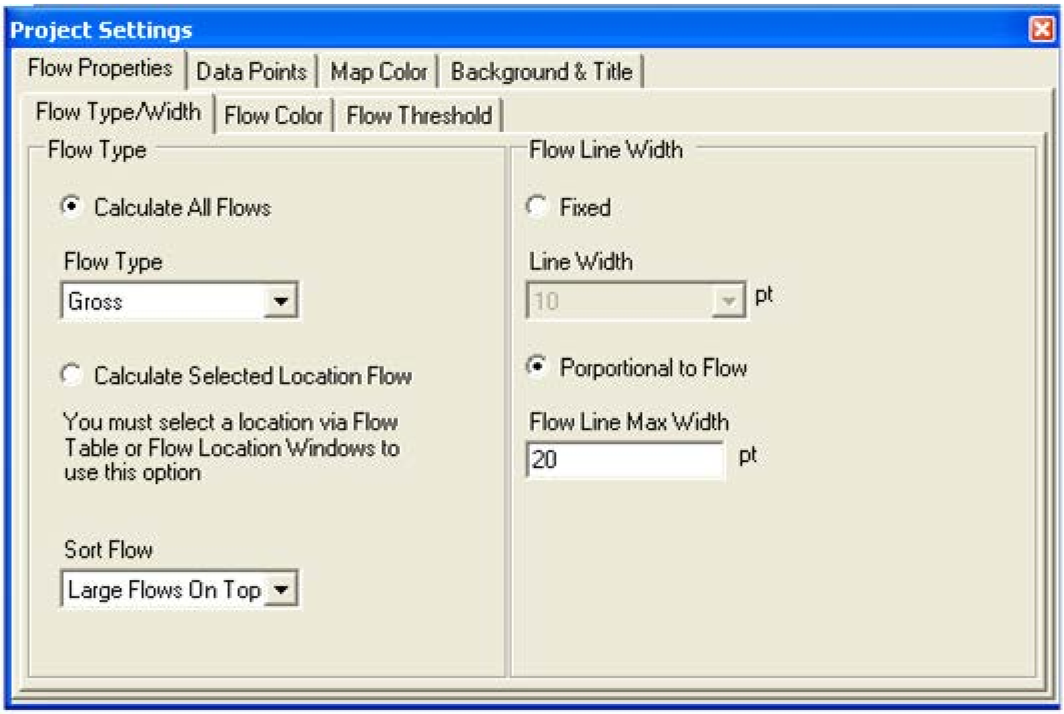
In 2003 the Center for Spatially Integrated Social Science of the Geography Department of the University of California at Santa Barbara supported an effort of somewhat under one-man month to produce an interactive flow mapping program of the discrete display type. This is essentially an updated interactive 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:



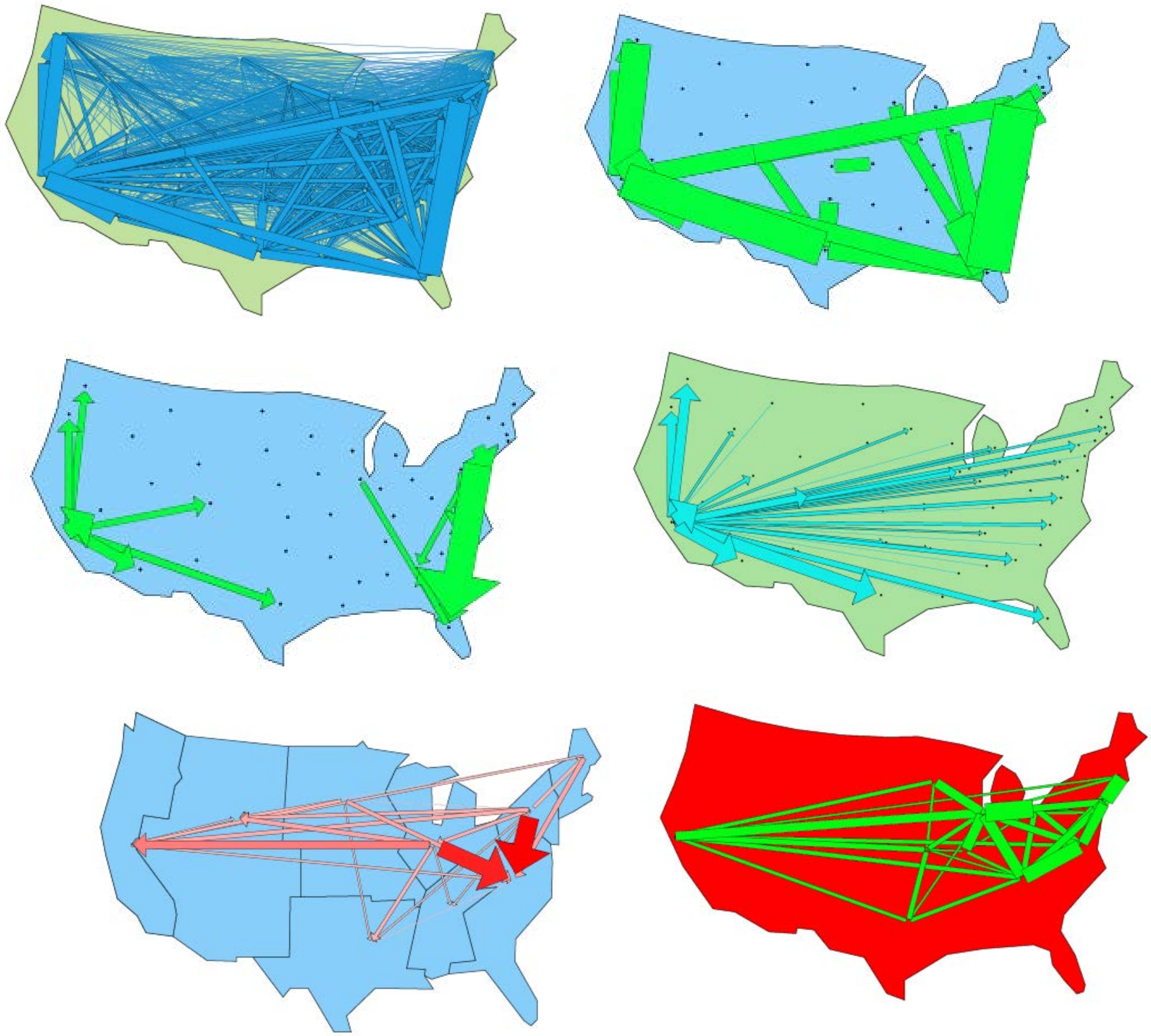

Future Work

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.

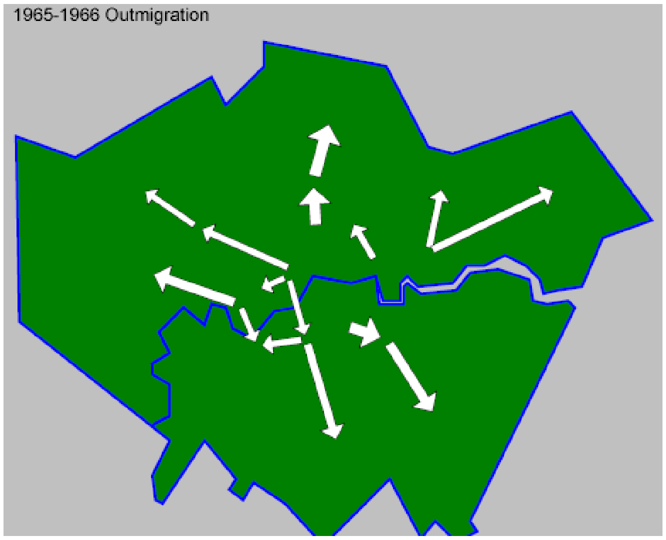
Comments on the attached color maps (it may take a few seconds for the maps to appear on the screen):

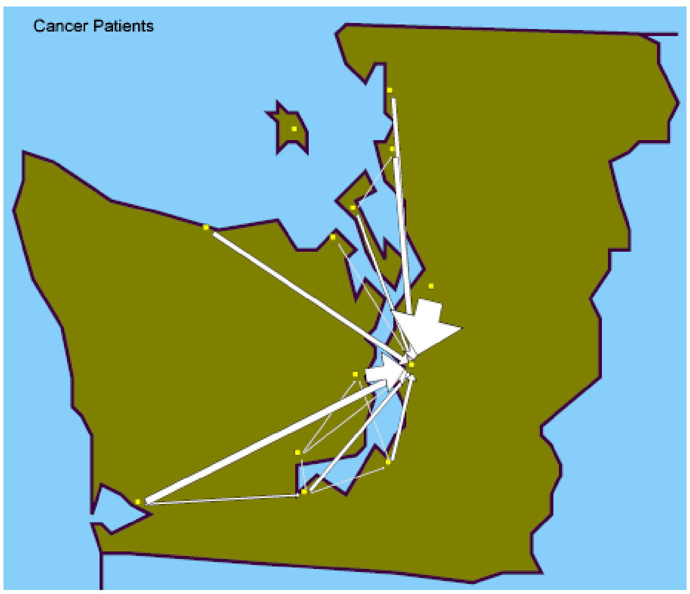
The first four of the colored sample maps use the 2000 US Census Bureau state-to-state tables describing migration for the interval 1995-2000. Alaska, the District of Columbia, and Hawaii are omitted for compatibility with previous studies, as are the diagonal table entries. The 48 by 48 tables yield 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.

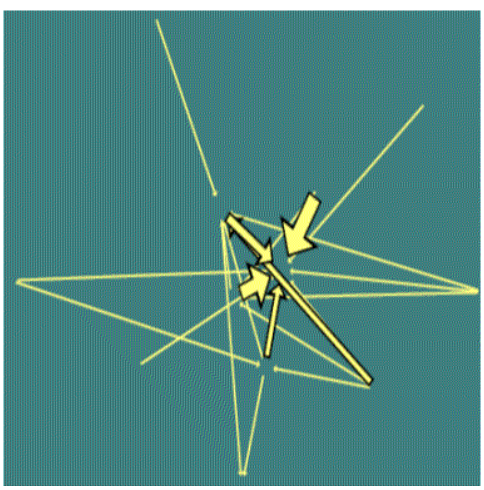
The next page includes a map of net moves between 33 London boroughs; everybody seemed to be leaving the town center in the 1965-1966 time period. In Western Washington State the map shows the 1974-1978 movement of cancer patients from their county of residence to the county of treatment. Both a title 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 maps[1]. 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 dominatingly 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 itineries 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 find 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.

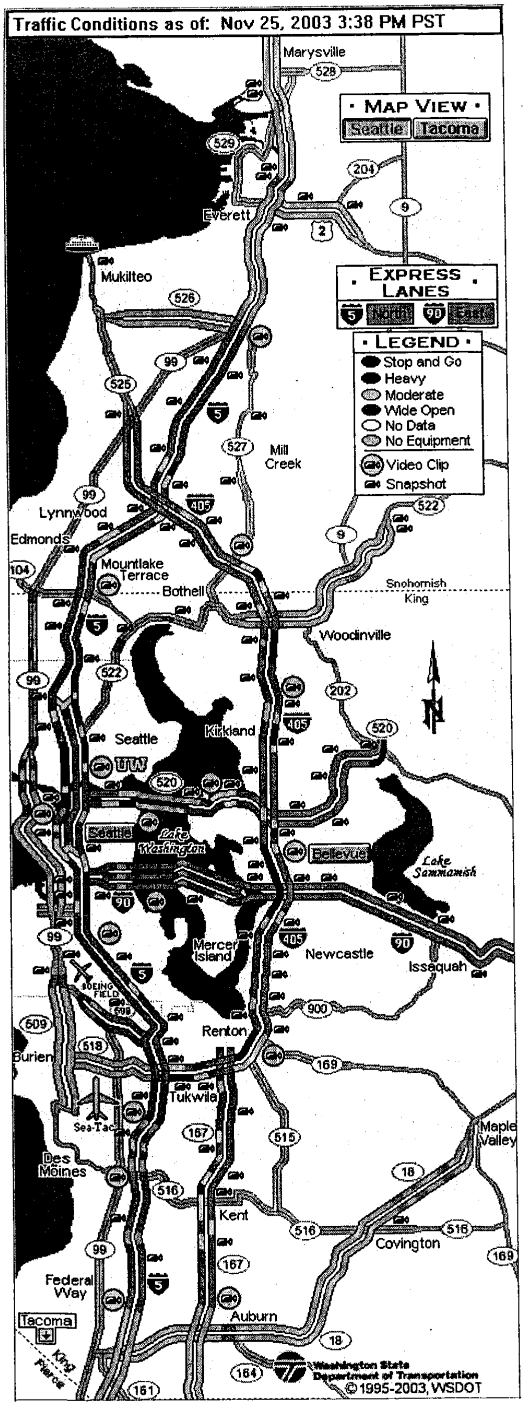
Tobler, W., 1999, "Migration: Ravenstein, Thornthwaite, and Beyond", *Urban Geography*, 14 (4): 327-343.

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

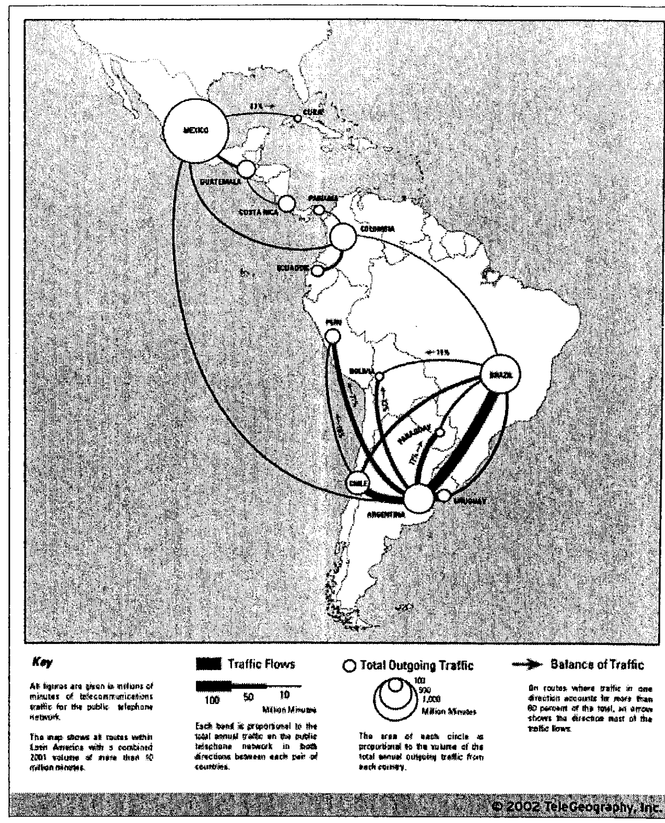
Zhang, J., M. Goodchild, 2002, *Uncertainty in Geographical Information*, Taylor and Francis, New York.

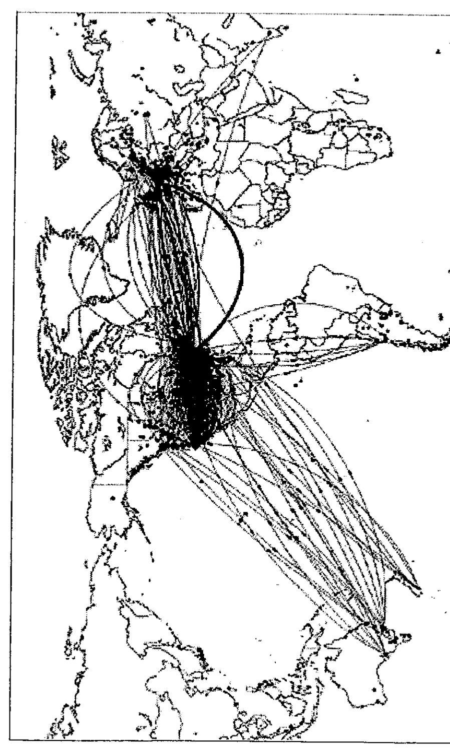
[1] Types of movement maps:

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

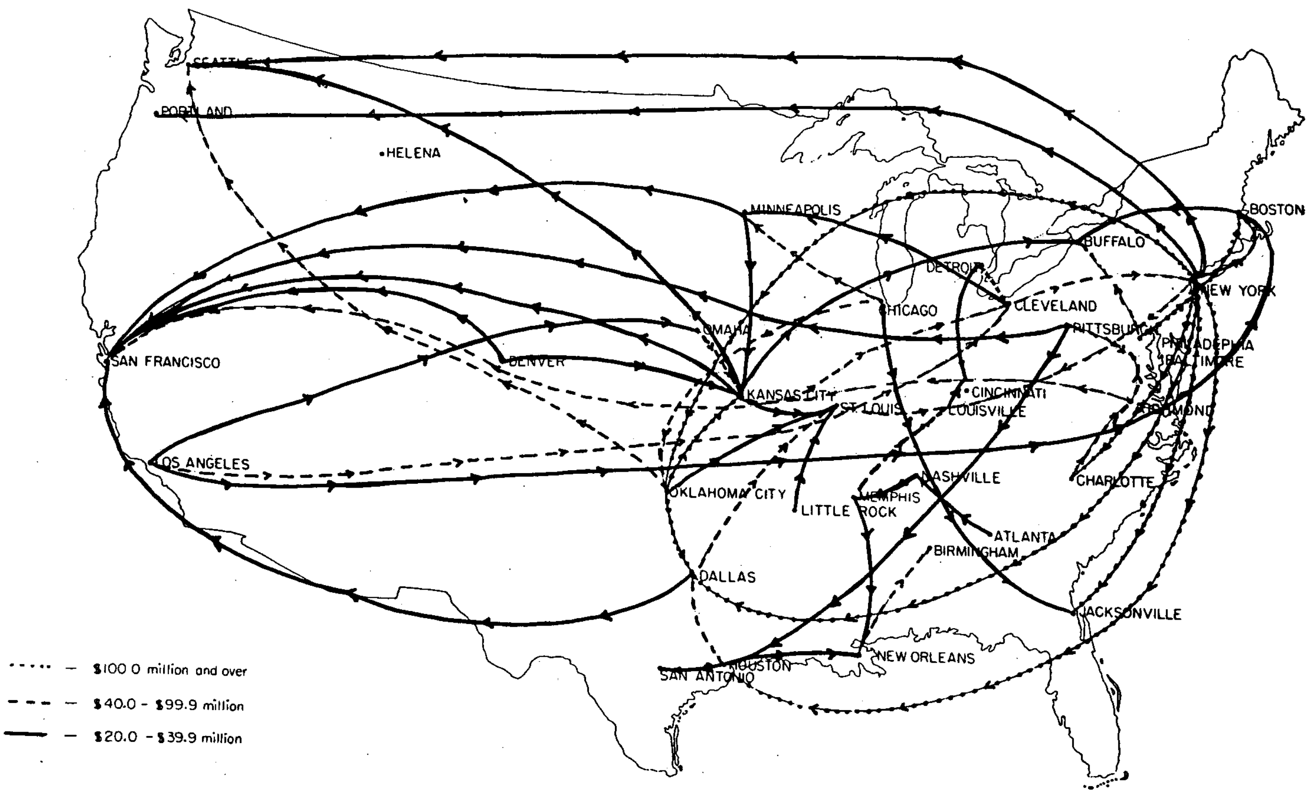



Latin American Telecommunication Traffic Flows, 2001

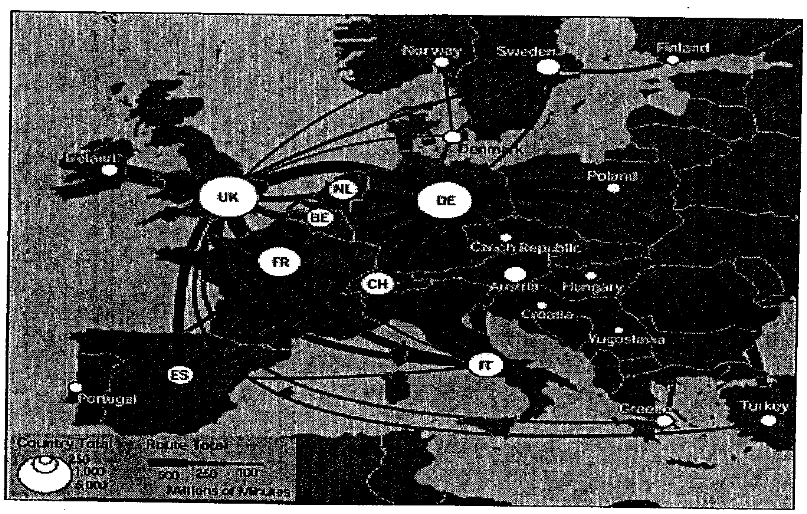


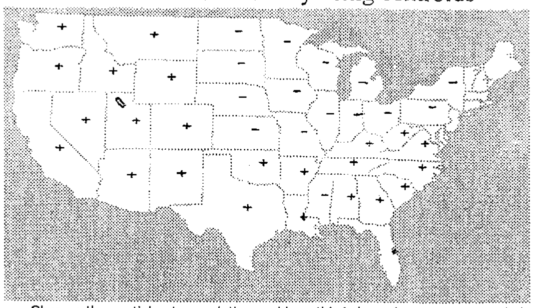
Map of complete aggregate news flow, worldwide, from 13 May 1993. (Courtesy of **Brian Reid**)




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.

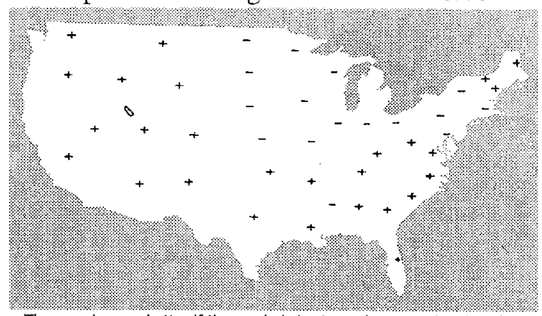


## The Population Change Information Can Be Positioned Locationally using centroids



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

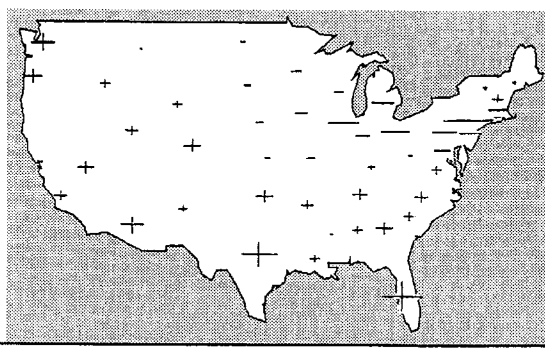
## Population Change at State Centroids



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

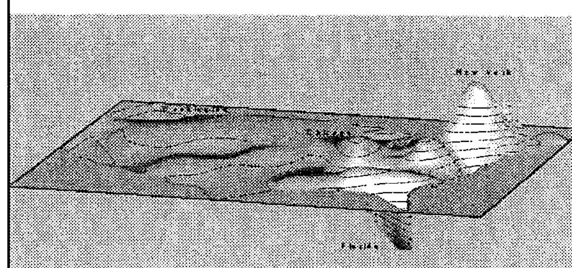
## Gaining and Losing States

Based on the marginals of a 48 by 48 migration table



## The Pressure to Move in the US

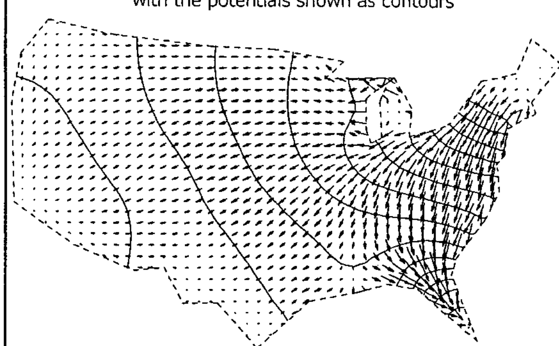
Based on a continuous spatial gravity model



Another view

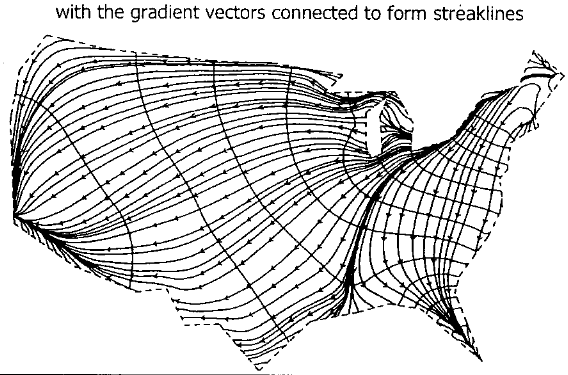
## The Migration Potentials and Gradients

with the potentials shown as contours



## Migration Potentials and Streaklines

with the gradient vectors connected to form streaklines



As an example of geographic movement.

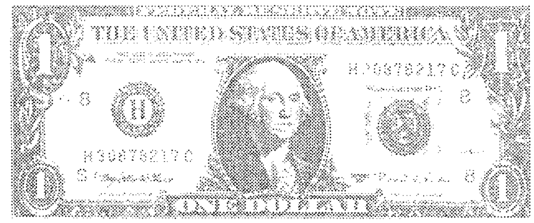
## In the United States the Currency Indicates Where It Was Issued

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

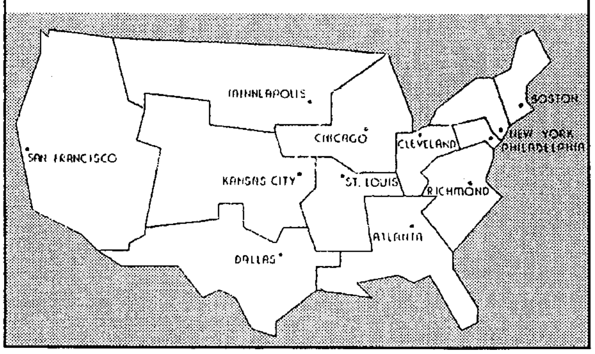
(Federal Reserve Note)




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

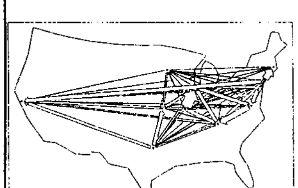
between Federal Reserve Districts, in hundreds, Feb. 1976

| From: \ To: | B | NY | P | Cl | R | A | Ch | SL | M | K | D | SF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Boston | 2040 | 239 | [illegible] | 52 | 137 | 118 | 90 | 10 | 16 | 15 | 13 | 38 |
| New York | 602 | 1980 | 231 | 309 | 336 | 307 | 286 | 15 | 45 | 26 | 18 | 261 |
| Philadelphia | 143 | [illegible] | 560 | 84 | 342 | 130 | 134 | 8 | 25 | 10 | 10 | 80 |
| Cleveland | 68 | 192 | 27 | 1296 | 171 | 177 | 603 | 16 | 44 | 43 | 19 | 31 |
| Richmond | 150 | 266 | 158 | 226 | 3899 | 573 | 295 | 20 | 62 | 54 | 22 | 152 |
| Atlanta | 122 | 159 | 57 | 186 | 319 | 3741 | 439 | 30 | 51 | 78 | 102 | 139 |
| Chicago | 97 | 155 | 39 | 496 | 143 | 266 | 5630 | 74 | 278 | 100 | 46 | 290 |
| St. Louis | 31 | 56 | 13 | 142 | 80 | 201 | 573 | 342 | 46 | 128 | 47 | 109 |
| Minneapolis | 17 | 26 | 11 | 32 | 29 | 47 | 295 | 10 | 1438 | 51 | 14 | 38 |
| Kansas City | 20 | 41 | 5 | 55 | 40 | 72 | 215 | 33 | 120 | 501 | 36 | 247 |
| Dallas | 31 | 41 | 8 | 38 | 46 | 165 | 175 | 20 | 37 | 253 | 788 | 203 |
| San Francisco | 52 | 87 | 22 | 54 | 114 | 106 | 251 | 22 | 137 | 128 | 43 | 5380 |

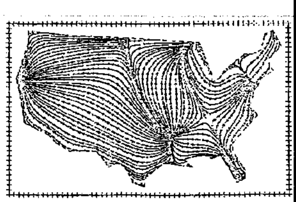
## Dollar Bill Movement in the US

net flow

Discrete representation



Continuous pattern representation



Based on computer programs from the 1980s

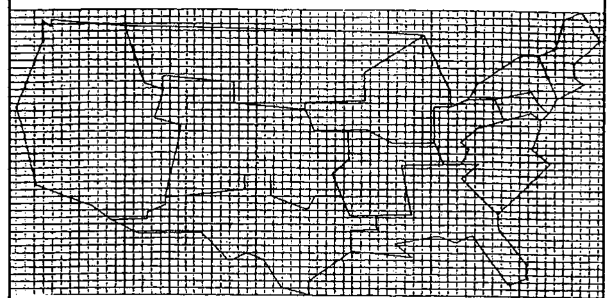
## The Table of Dollar Bill Movements

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

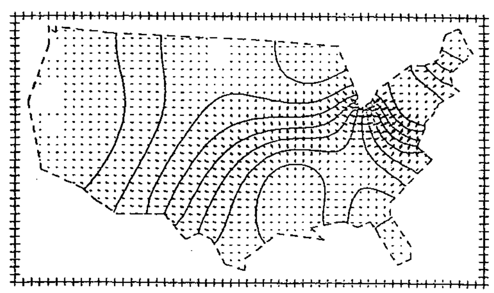
The result is a system of partial differential equations solved by a finite difference iteration to obtain the potential field.

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.

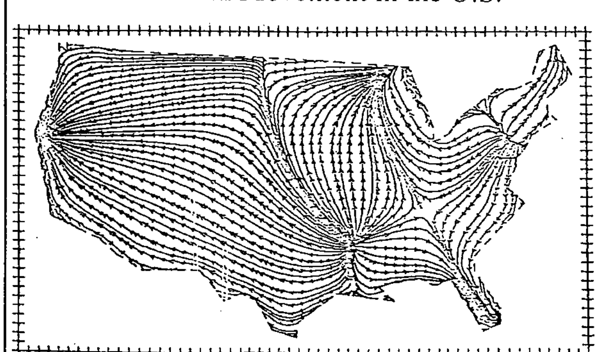
## Solving the Equations Yields the Potential

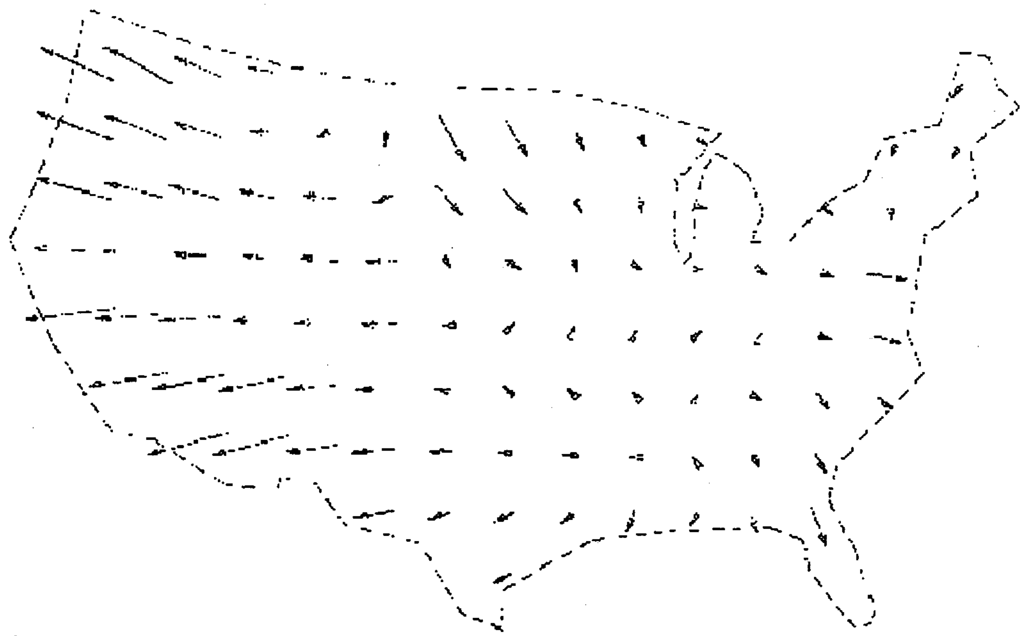
Shown here by contours



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.

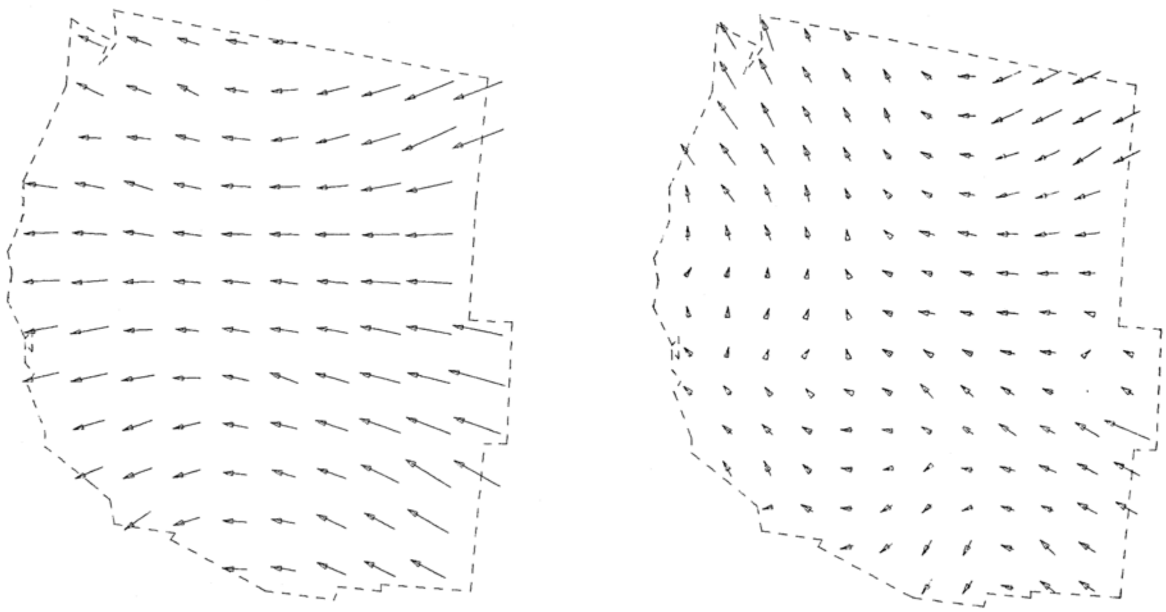




1935-1940 White Migration

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.



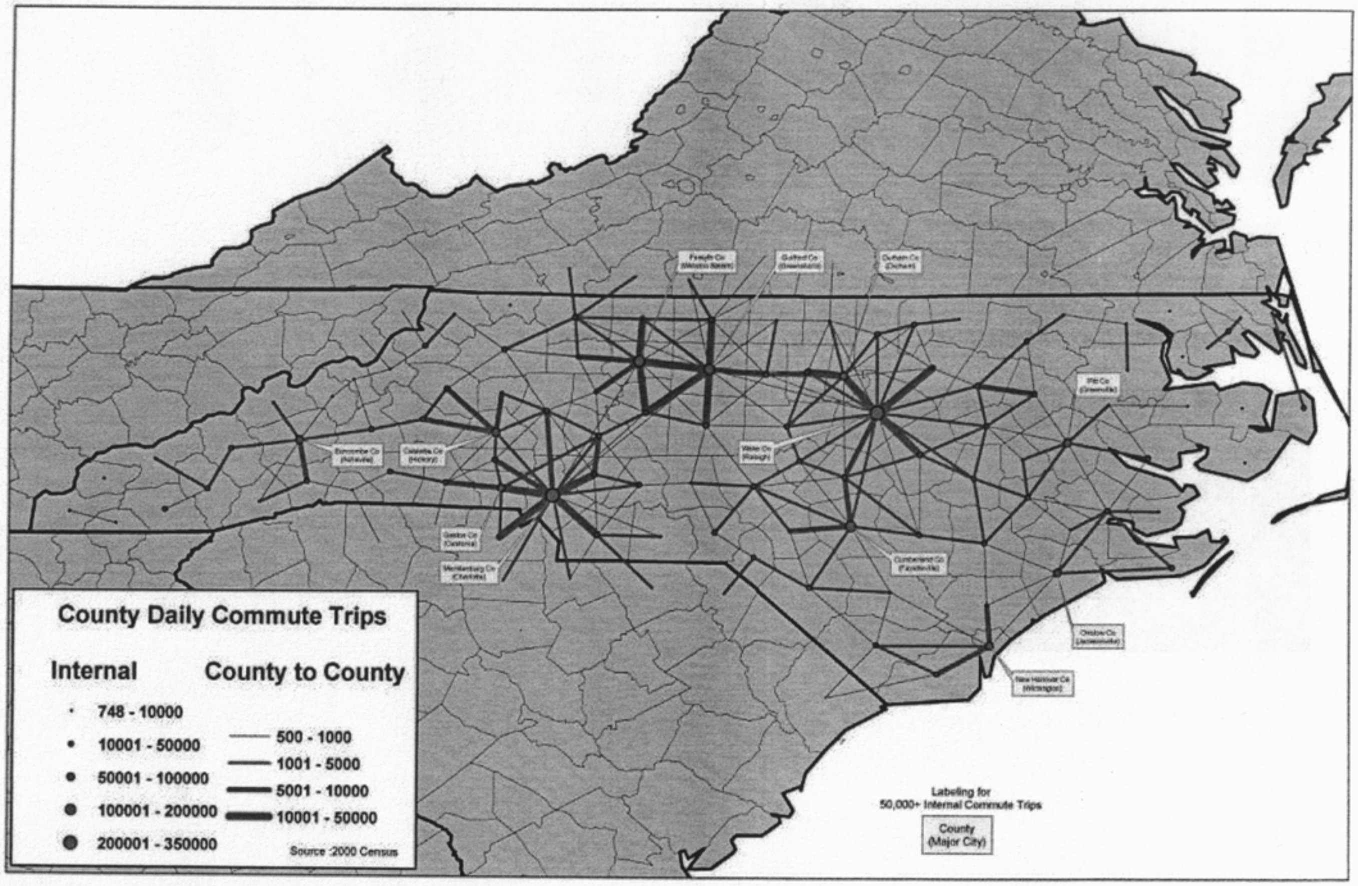



EXHIBIT A – North Carolina County-to-County Commuting

# 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/).**

**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.**

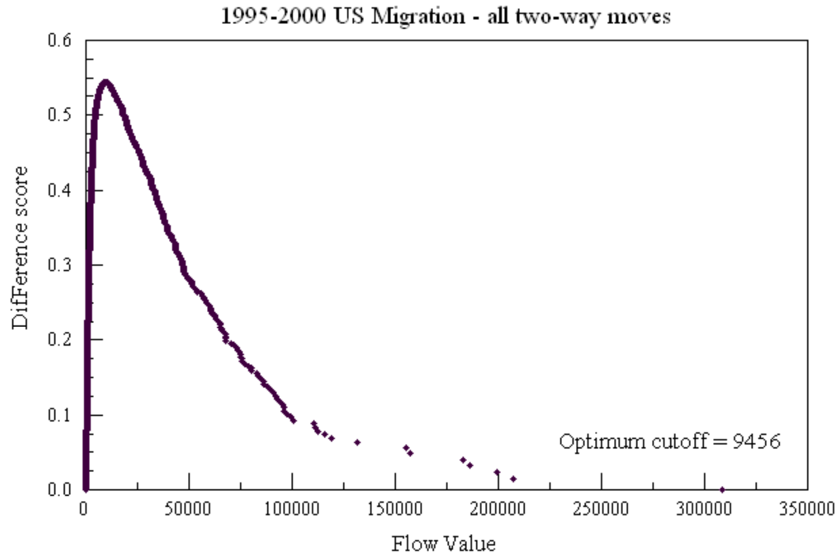
**Masser, I., 1976, “The design of spatial systems for internal migration systems”, *Regional Studies*, 10, 39-52.**

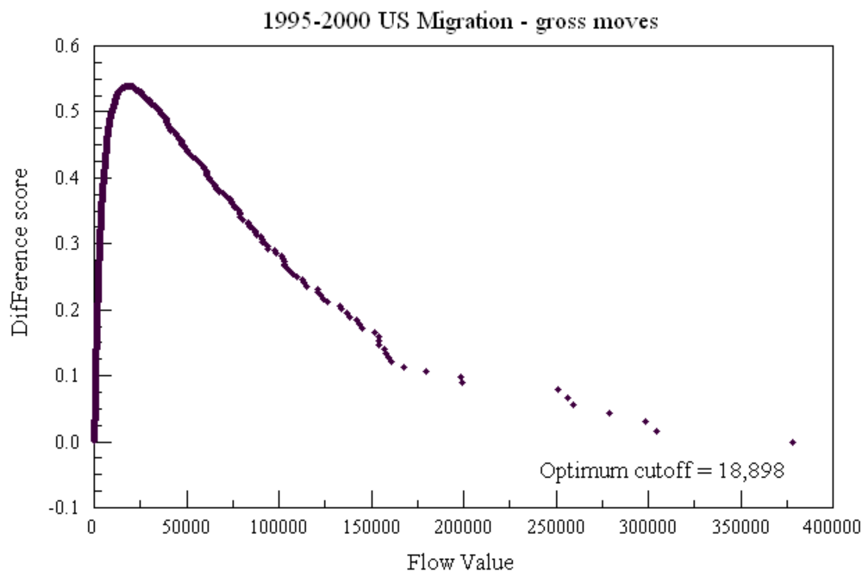
**Masser, I., and S. Scheurwater, 1980, "Functional Regionalization of Spatial interaction data…",** ***Environment and Planning*****, A, 12: 1357-1382.**

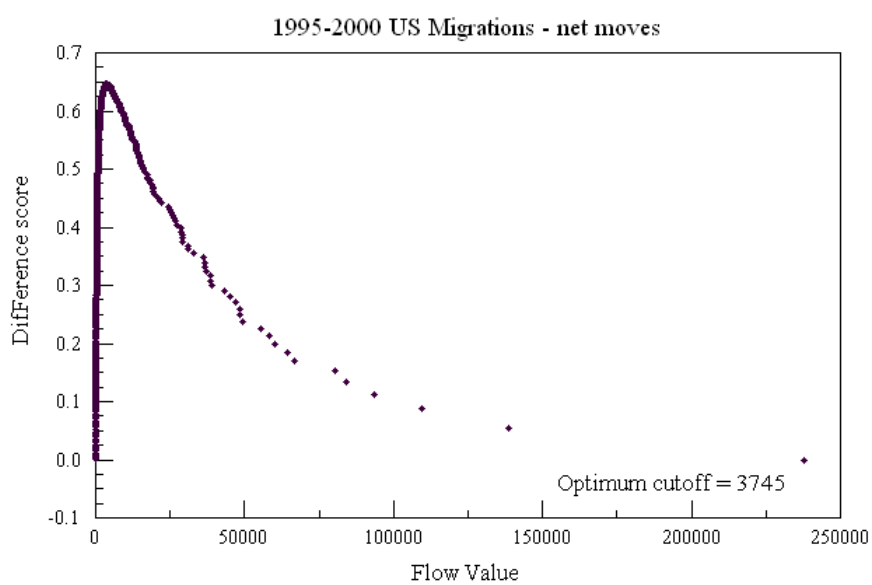
**Openshaw, S., 1977, "Optimal zoning systems for spatial interaction models",** ***Environment and Planning*****, A, 9: 169-184.**

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

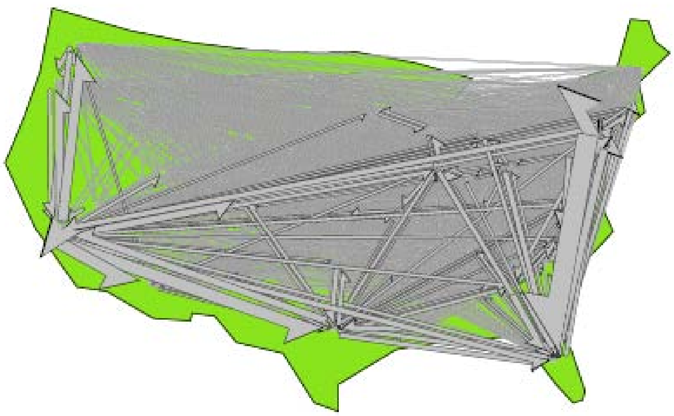
**(The computer program used to perform the threshold calculation can be obtained from the author)**


1995-2000 US Migration - all two-way moves
DifFerence score
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Optimum cutoff = 9456
0
50000
100000
150000
200000
250000
300000
350000
Flow Value

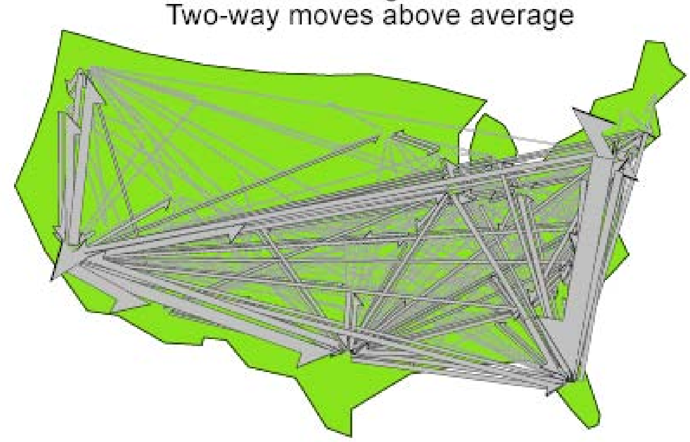

1995-2000 US Migration - gross moves
DifFerence score
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
Optimum cutoff = 18,898
0
50000
100000
150000
200000
250000
300000
350000
400000
Flow Value


1995-2000 US Migrations - net moves
DifFerence score
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
Optimum cutoff = 3745
0
50000
100000
150000
200000
250000
Flow Value

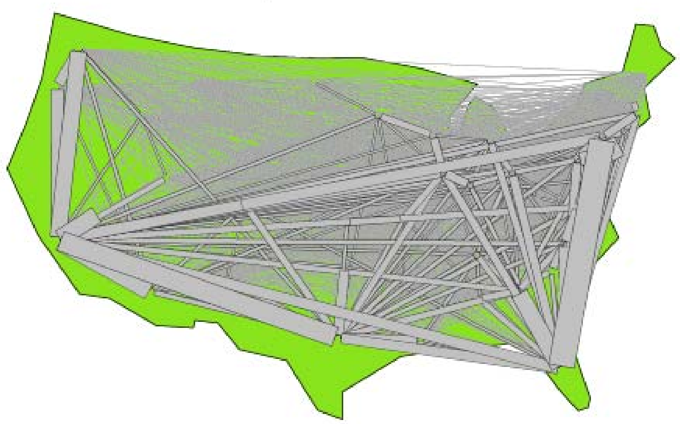
1995-2000 US Migration
All two-way moves



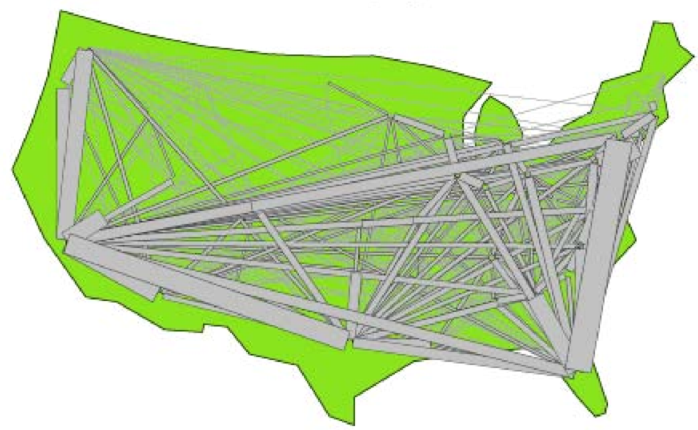
1995-2000 US Migration
Two-way moves above average



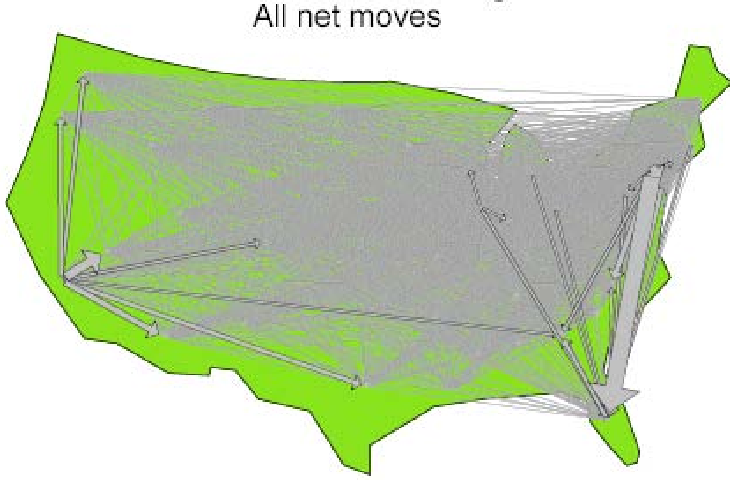
1995-2000 US Migration
All gross moves



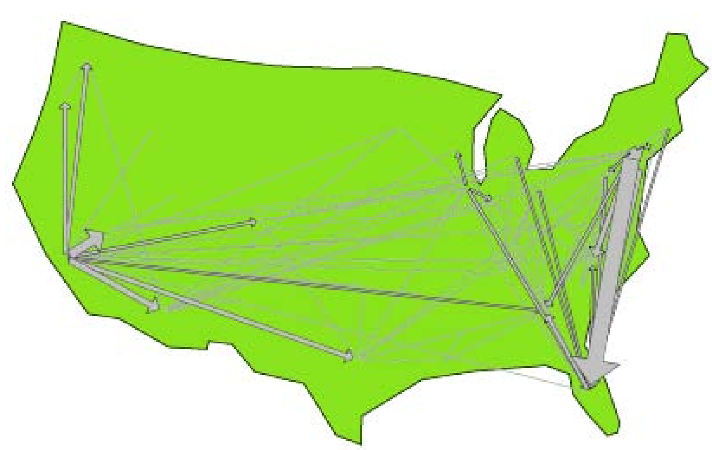
1995-2000 US Migration
Above average gross moves



1995-2000 US Migration
All net moves



1995-2000 US Migration
Above averge net moves



# 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 x (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 x (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 x (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 S0C77-0O191.

**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.

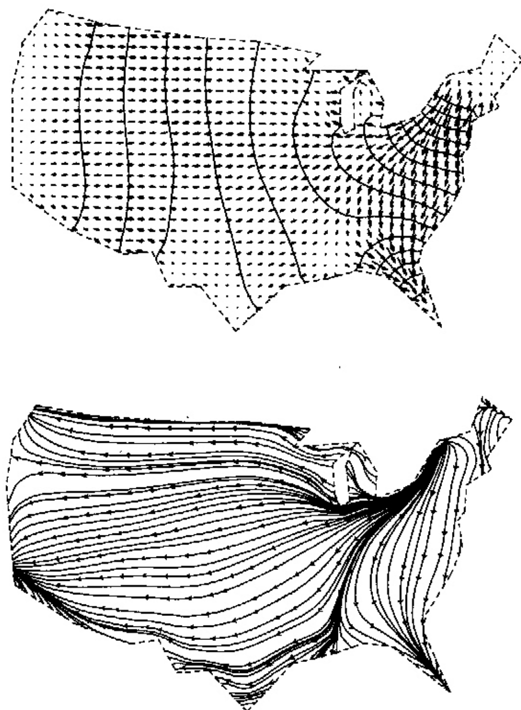


Figure 2. Cartographatron display of 9,931,000 desire line traces of personal trips in Chicago. *Chicago Area Transportation Study*, 1959; fig. 23, p. 46.

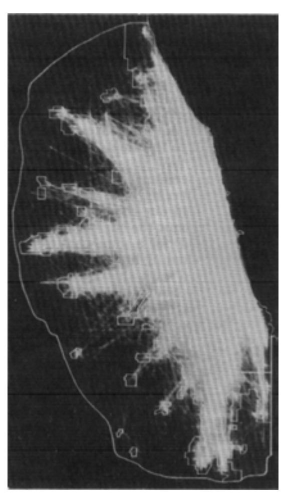


Figure 3. Arrow and shading types.

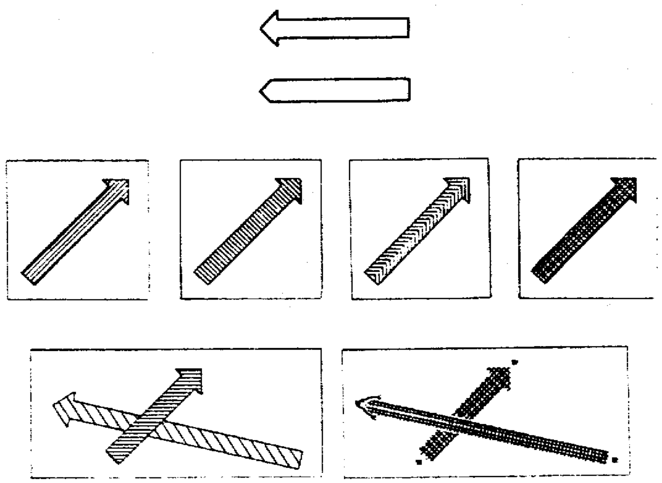


Figure 4. Net migration between states 1965-1970, showing the effect of varying arrow widths. Minor flows deleted.

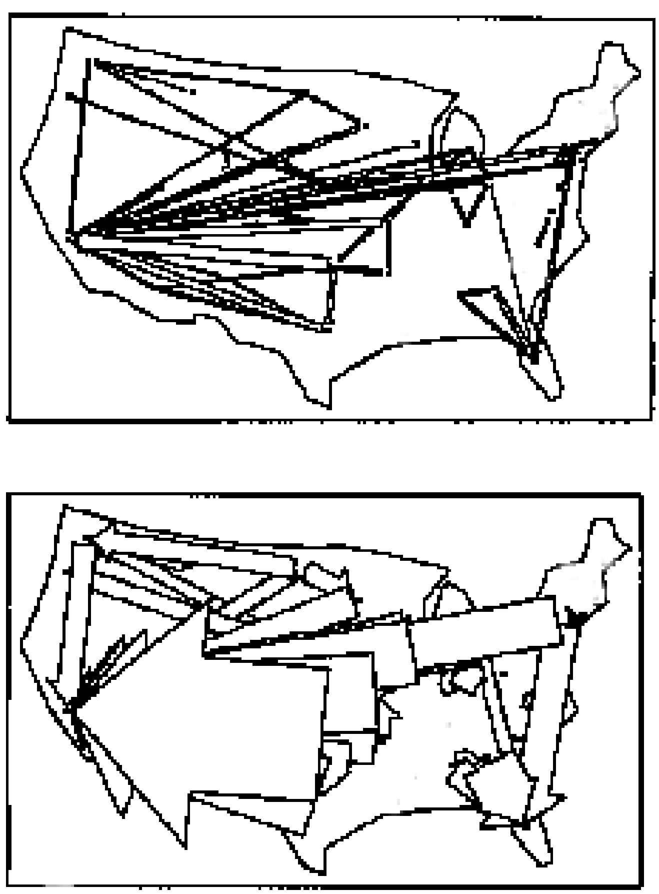


Figure 5. Bi-directional arrows.

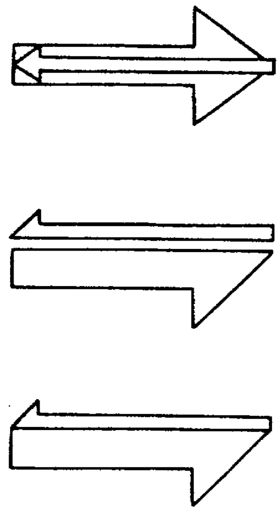


Figure 6. Gross migration to and from California, 1965-1970, with constant density shading.

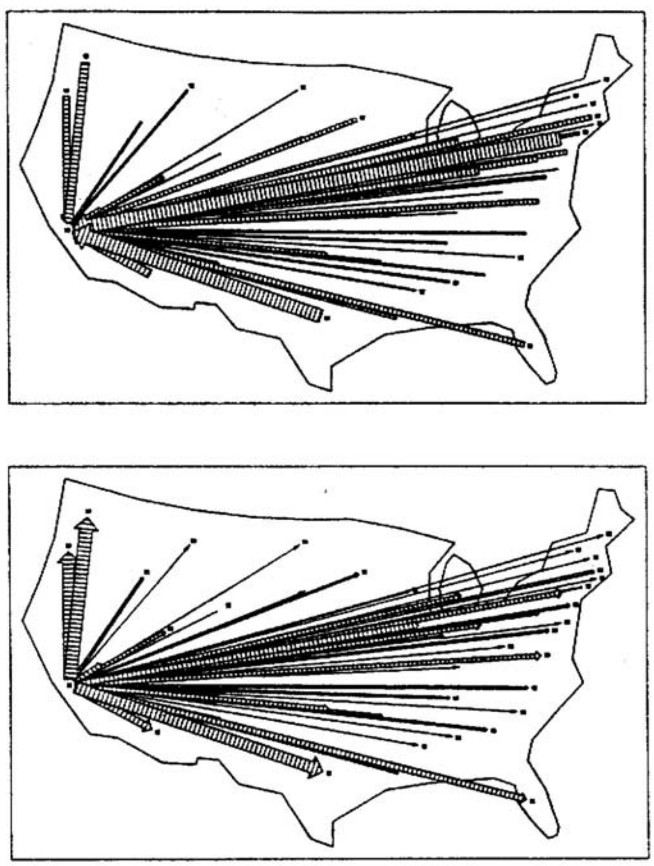


Figure 7. Net migration 1965-1970, with all volumes below the average deleted. Less than 60 of the possible 1128 (48 x 47 / 2) arrows need be shown to represent the majority of the migration.

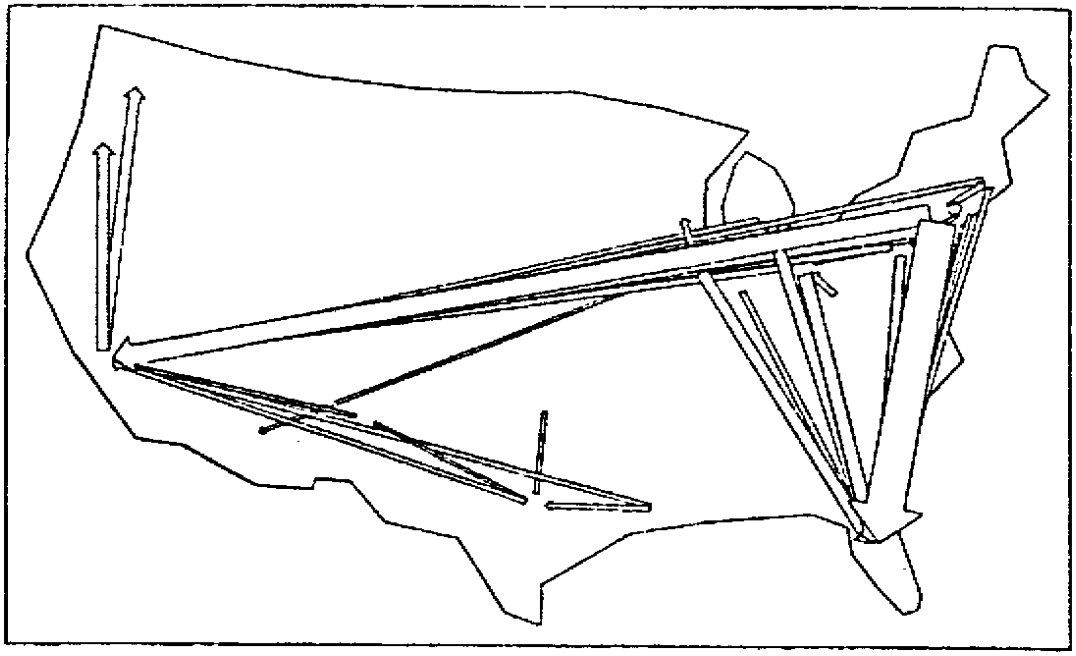


Figure 8. State to state migration, 1965-1970

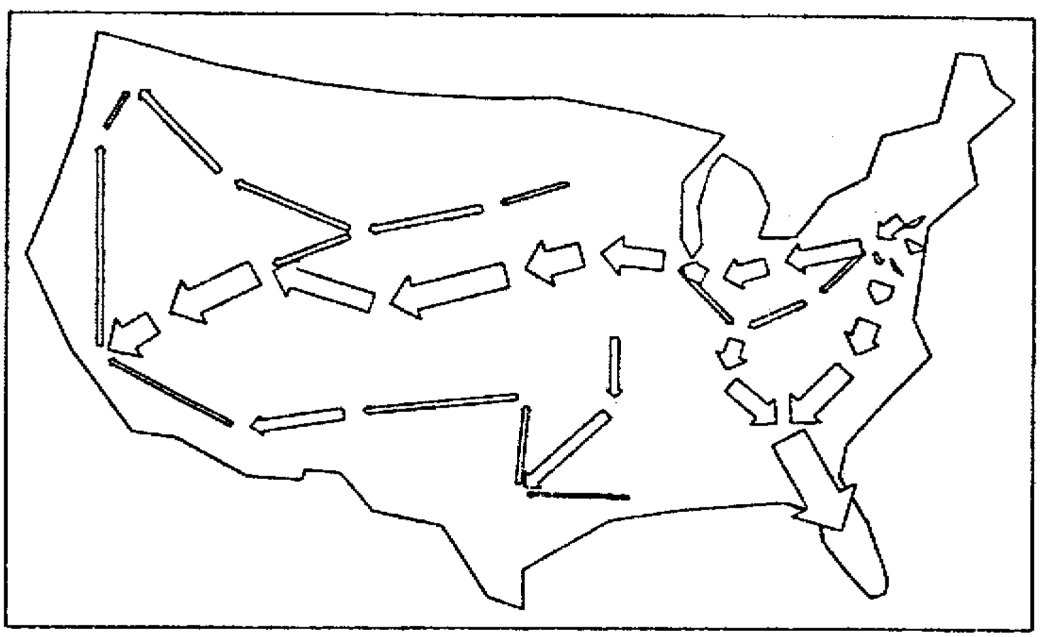


Figure 9. Ravenstein (1885), Map 5, page 183.

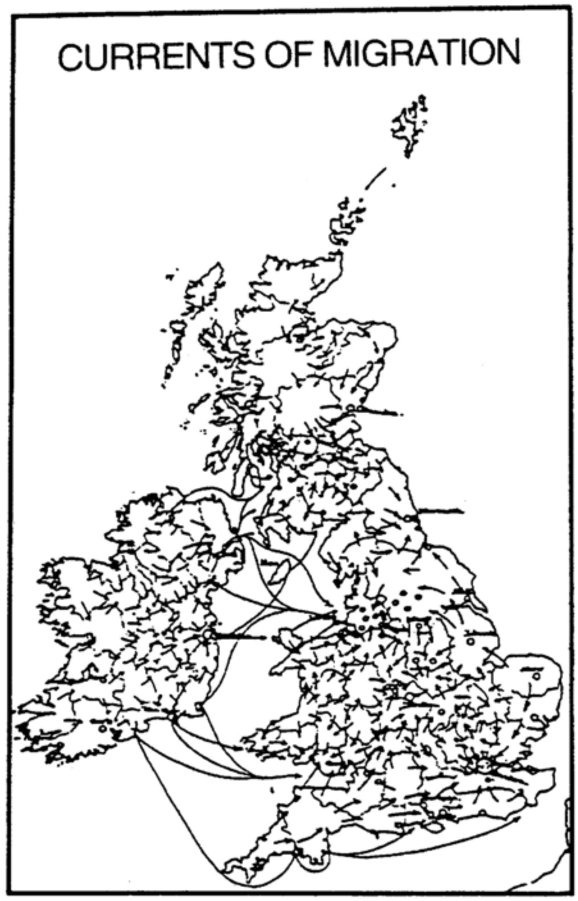



Table 1.

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.