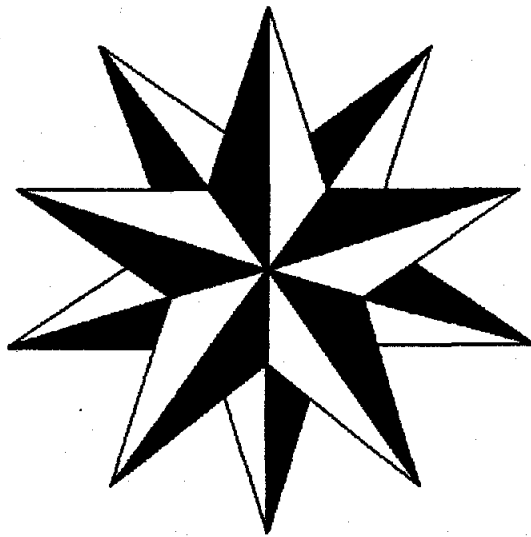


USING SAS/GRAPH TO DISPLAY THE GEOGRAPHIC DISTRIBUTION OF WISAS MEMBERSHIP



**Gary Plazyk
A.T. Kearney, Inc.**

Using SAS/GRAPH¹ to Display the Geographic Distribution of WISAS Membership

Gary F. Plazyk
A.T. Kearney, Inc.
222 S. Riverside Plaza
Chicago, IL 60606
(312) 993-8970

Abstract

The geographic distribution of members of the Wisconsin Illinois SAS Users Group (WISAS) is displayed using graduated circle or bubble maps, at various levels of geographic aggregation: five and three digit zip codes, and state-city.

Some disclaimers: this paper is not a rigorous statistical analysis, nor does it claim to offer one best locational solution. It is a demonstration of methods that may be used to provide insight into a geographic problem, and point the way to more mathematically sophisticated analyses.

Introduction

One topic of discussion at our last WISAS meeting dealt with the suitability of our current meeting location at the College of Lake County. Among the issues raised were: adequacy of the facilities, cost of renting the space, and quality of food. But the major concern dealt with the geographic location of the site, and the resulting travel time for people coming to the WISAS meetings.

The SAS and SAS/GRAPH products offer a number of tools for the analysis and display of geographic data. I felt that SAS could be used to investigate the geographic distribution of WISAS members. This application might provide a suitable case study to illustrate some of the features of SAS, and provide insight for an approach to other similar geographic problems.

The Issues

The purpose of this analysis is to answer three questions:

- * Where are the WISAS members?
- * Where is the meeting location?
- * Are they a good compromise? Is the existing meeting location about equally convenient (or inconvenient) to most of the members?

Reprint Notice: This paper was previously published in *WISAS Proceedings, Volume 2, June Issue*. Edited by LeRoy Bessler. Fox Point, Wis.: Wisconsin Illinois SAS Users Group, 1990.

The Data

The data used in this analysis consist of:

- * Locations of the WISAS members. The five digit zip code of each person on the WISAS mailing list was provided as a flat ASCII file on an MS-DOS diskette.² No personal data was provided; the file contained only five digit zip codes, one for each person.
- * A proprietary cross-reference table of five digit zip code to latitude and longitude. This table is in the form of a SAS dataset developed from the "Geographic Data File"³ product licensed from List Processing Company, Glen Ellyn, Illinois. The latitude and longitude are for the location of the main post office, or for the centroid of the zip code if no main post office exists. Other products similar to this file exist. As a last resort, locations could be developed by inspection or by looking up latitude and longitude in a good world atlas.
- * A proprietary cross-reference table of the principle city and its latitude and longitude in each three digit zip code, developed by A.T. Kearney, Inc. The latitude and longitude are for the location of the main post office of the principle city. This table was developed by inspecting Rand McNally maps of three digit zip codes,^{4 5} and arbitrarily selecting the biggest or most important city in each three digit zip code area. Note that several three digit zip codes may have the same principle city.
- * The unprojected US map outline dataset⁶ provided by SAS Institute, Inc. with the SAS/GRAPH product.

Preliminary Analysis

The five digit zip codes were read into a SAS dataset, and state abbreviations were derived from the five digit zip codes using the SAS function ZIPSTATE. This function is accurate for most zip codes, although there are pathological cases where a zip code may be just across a border in another state. Figure 1 is a horizontal bar chart of the state abbreviations, showing frequency and percent of total membership in each state.

Zip codes were aggregated using PROC SUMMARY, producing a dataset of five digit zip codes and the number of occurrences of each zip code. This file was then merged with the five digit zip code cross-reference file, adding latitude, longitude, and state abbreviation to the aggregated zip codes.

Three digit zip codes were aggregated using PROC SUMMARY, producing a dataset of three digit zip codes and the number of occurrences of each zip code. This file was then matched against the three digit cross-reference file, adding latitude, longitude, state abbreviation, and city to the aggregated zip codes.

One problem with using the three digit zip code aggregation file is that several three digit zip codes may be assigned to the same city. For example, Milwaukee is the primary city for three digit zip codes 530 through 534. To solve this problem, the three digit zip code file was further summarized by state-city. The state-city pair distribution is shown in alphabetic order in Figure 2, and in descending frequency order in Figure 3.

Map Basics

The United States Postal Service Zoning Improvement Plan (zip) code⁷ is a hierarchical geocoding scheme⁸ in common use in all sectors of business and government. The coding scheme consists of a five digit code, and an optional four digit extension. (We will not consider the ZIP-Plus-4 codes here.) The first digit of the zip code designates a major region of the United States. The first two digits designate a state or a major portion of a state. For example, Illinois is divided into two-digit zip codes of 60, 61, and 62. The first three digits of a zip code correspond to a smaller area, on the order of the size of a county. More than one three digit zip code can apply to the same area. The next level of coding is the five digit zip code. Although maps outline datasets are available for three and five digit zip code areas,⁹ they are expensive and I did not have access to them.

The maps in this paper were generated using only points whose looked-up zip code resulted in a state of IL, WI, or IN. The map outlines are the states of Illinois and Wisconsin, selected from the SAS/GRAPH MAPS.US dataset, and subset to points with DENSITY less than or equal to 4. DENSITY is a value associated with each point in a map outline dataset. A value of 0 is the most coarse resolution, and a value of 6 is the most detailed resolution.¹⁰

SAS/GRAPH Maps

PROC GMAP offers several map types.¹¹ CHOROPLETH maps are commonly used when a value can be applied to an entire area. BLOCK (or "skyscraper") maps are misleading, because the height of the block is ordinal, rather than in proportion to the value. For example, if three points have values of 1, 2, and 1000, the generated blocks are of heights 1, 2, and 3. I feel this can be so misleading that I don't use block maps. PRISM maps and SURFACE maps are showy, but high values near the front of the map hide lower values further back on the map. This can be avoided to some degree by changing the viewing point of the map with the TILT and ROTATE options, but this is of limited value. A comparison of the SAS/GRAPH map types was presented by J. Jacob Wind at both SUGI 14¹² and SUGI 15.¹³

Dot Maps

Dot maps depict location of phenomena using points applied to a reference outline called a base map.¹⁴ Figures 4 and 5 are dot maps showing the locations of five digit zip codes and three digit zip codes with square dot symbols, and the location of our current meeting location with a diamond symbol. Dot maps can be generated at least two different ways in SAS.

The first technique for generating a dot map involves synthesizing a "state" outline - in this case, a little square - for each point. This technique was developed before the SAS ANNOTATE facility was available.^{15 16} It still retains an advantage over ANNOTATE point generation because it offers a greater degree of control over the point outline shape, and is conceptually easier to use.¹⁷ It has a disadvantage on maps covering large areas, because the process of producing a projected map distorts the outlines of the synthesized "states". This technique cannot be used to place text on the map.

The second technique for generating a dot map uses the ANNOTATE facility. The points placed by the annotation are not subject to the distortion introduced by map projection, and can be used to place text on a map.¹⁸ But the techniques involved in using ANNOTATE datasets tend to be circuitous, and can be frustrating and time-consuming until mastered.

Graduated Circle or Bubble Maps

Graduated circles have long been used to display quantitative information on maps,¹⁹ and are still in common use.²⁰ Circles of different sizes are used to convey values associated with each point. The most common technique makes the area of the circle proportional to the value of the point.²¹

Circle maps generated with SAS code were described by Evans & Dolan in 1988.²² Also, Robert P. Wederich, Jr., used SAS/GRAPH to generate "bubble" maps for a paper he presented at SUGI 15. This paper will appear in the forthcoming SUGI 15 Proceedings.²³

Figure 6 shows circles synthesized as "states"; note the egg-shaped distortion introduced by the map projection. Figures 7, 8, and 9 show the data with the circles generated using the ANNOTATE facility to place pie charts (with zero slices) at points on the map. Note that true circles are displayed. Some trial-and-error adjustment to the scaling of the circles is necessary to make the final map look neither too "busy" nor too empty.²⁴

The reason for aggregating the data into three digit zip codes and further into cities is apparent from these maps. Figure 7 is simply illegible with data being displayed at the five digit level. Figure 8 exhibits a "bulls-eye" phenomenon caused by mapping multiple three digit zip codes onto the same city. Figures 9, displaying data aggregated by state-city for three digit zip codes, is the most readable.

Bells and Whistles

Additional information may be added to the maps. Figure 10 shows the addition of a grid composed of the lines of latitude from 37 to 48 degrees north, and longitude 87 to 94 west. Figure 11 shows the annotation of state and city name and member count. Note that the labels for Chicago, Evanston, and Gary were flipped to reduce the overprinting of names.

Figure 12 shows the distribution of latitudes, and Figure 13 shows the distribution of longitudes, of the WISAS members in Illinois, Wisconsin, and Indiana. Figure 14 shows a cross-hair surrounded by a box added to the map. The cross-hair is made of lines at the mean latitude and longitude - the "center of gravity" - of the three states' points. The surrounding box is made of lines one standard deviation either side of the mean. The diamond shaped symbol, centered at Grayslake, indicates the current WISAS meeting location at the College of Lake County.

Limitations of This Analysis

It must be noted that this analysis does not claim to locate the "best" location for a meeting site, only the most central to the persons on the mailing list who have addresses in Illinois, Wisconsin, or Indiana. Also, travel times along highways are not explicitly considered; only straight-line "crow-fly" distances are displayed.

A much different result might be obtained based on members that actually attend the meetings. The meeting site may determine who chooses to get onto the mailing list; a meeting site near Wisconsin or Madison would probably attract more members from the surrounding area, and cause the Illinois membership to drop.

Other meeting strategies might be chosen. The Chicago chapter of the Data Processing Management Association (DPMA) alternates its meetings between the suburbs and downtown Chicago.

Studies demonstrate that human perception of different circle areas is not a simple linear relationship. Small circles are over-estimated; large circles are under-estimated.^{25 26} Corrections for these perceptual phenomena are beyond the scope of this paper.

Uses, Implications, and Future Work

Circle maps are useful for visualizing values at various locations on maps. The techniques demonstrated in this paper have been packaged into code to produce circle maps quickly and easily given a file of locations and values. Versions exist for locations stated in terms of five and three digit zip codes, as well as state-city pairs. The only manual intervention required, aside from cleaning up incorrect or misspelled location data, is the input of titles and a scaling factor for circle radius multiplier.

Future work includes developing a program to translate the Postscript output plot file to produce an encapsulated Postscript graphics file. The encapsulated Postscript file could be included directly in an Apple Macintosh document, my company's standard form for proposals and reports.

Summary

We have examined techniques for aggregation of geographic data to a point, and the display of point data using a circle or bubble map. These two techniques together provide a powerful tool for the investigation of geographic statistical data.

This exercise has shown some techniques for using SAS and SAS/GRAPH to help understand geographic data. Although not a real geographic information system (GIS), SAS offers a collection of tools that can, with practice, produce valuable maps and geographic graphics.

I suggest that SAS Institute add a new map type, BUBBLE or CIRCLE, to PROC GMAP. This would work like the BUBBLE plot in PROC GPLOT,²⁷ producing the type of circle maps illustrated here. I think it would be a valuable addition to the SAS/GRAPH mapping capabilities.

With the release of the 1990 census TIGER line files, computer cartography has taken a giant leap forward. As the 1980s were an era of giant strides in computerized business graphics, the 1990s will be the era of great developments in desktop mapping. With SAS Institute's development of a prototype GIS product for the SAS System, demonstrated at SUGI earlier this year,²⁸ I hope to see a new generation of tools for mapping applications.

Acknowledgements

I appreciate the encouragement given me by Walter Strauss, a Principal at A.T. Kearney, and Robert Gremley, a Manager at A.T. Kearney. Their recognition of the value of geographic analytical techniques and insight into real-world consulting applications was most valuable.

I also am indebted to Dr. Karl Kriesel, of the Department of Geography and Environmental Studies at Northeastern Illinois University, for his constructive criticism and suggestions for reference material for cartographic methods.

Processing Environment

This analysis was done using SAS Release 5.18 on an IBM 4381 computer, under the VM/CMS operating system. Graphics were previewed on an MS-DOS PC clone emulating an IBM 3179 graphics terminal. Hardcopy graphics were produced as graphics stream file (GSF) plot files on the mainframe, and downloaded to the PC for printing on an Apple Laserwriter 300 DPI Postscript printer, and an HP 7475A plotter.

References

- Edgar, Gary C. 1982. "GEOGRAPHIC DISPLAY OF A MULTI-LOCATION TELECOMMUNICATIONS NETWORK". Proceedings of the Seventh Annual SAS Users Group International Conference. pp. 234-239. Cary, NC: SAS Institute, Inc.
- Evans, Sheila Fitzgerald & Dolan, Robert John. 1988. "SAS/GRAPH SCATTER MAPS" Proceedings of the Thirteenth Annual SAS Users Group International Conference. pp. 723-727. Cary, NC: SAS Institute, Inc.
- Monkhouse, F. J. & Wilkinson, H. R. 1971. Maps and Diagrams: Their Compilation and Construction, Third Edition. London: Methuen & Co. Ltd.
- Muehrcke, Phillip C. 1978. Map Use: Reading, Analysis, and Interpretation. Madison, WI: JP Publications.
- Plazyk, Gary F. 1984. "USING THE SAS AND SAS/GRAPH PRODUCTS FOR TRANSPORTATION ANALYSIS" Proceedings of the Ninth Annual SAS Users Group International Conference. pp. 269-276. Cary, NC: SAS Institute, Inc.

- Rand McNally & Company. 1983. Rand McNally Thematic Maps: Zip Code Map of the U.S. Chicago, IL: Rand McNally & Company.
- Rand McNally & Company. 1990. Sales & Marketing Planning Atlas: 3-Digit Zip Code Data. Chicago, IL: Rand McNally & Company.
- Rand McNally & Company. 1990. Commercial Atlas & Marketing Guide: 121st Edition. Chicago, IL: Rand McNally & Company.
- Robinson, Arthur H. & Sale, Randall D. 1969. Elements of Cartography, Third Edition. New York, NY: John Wiley & Sons, Inc.
- SAS Institute, Inc. 1982. "Drawing Lines and Symbols on SAS/GRAPH Maps". SAS Communications Volume VII, Number 3 (Winter 1982). Cary, NC: SAS Institute, Inc.
- SAS Institute, Inc. 1985. SAS/GRAPH User's Guide, Version 5 Edition. Cary, NC: SAS Institute, Inc.
- U.S. Postal Service. 1989. National Five Digit Zip Code & Post Office Directory, Volumes 1 & 2. Washington, D.C.: National Information Data Center.
- Wederich, Robert P., Jr. 1990. "SELECTING OPTIMAL BRANCH LOCATIONS BASED ON GEOGRAPHIC DISTRIBUTION OF SERVICE CALL VOLUME". Proceedings of the 15th Annual SAS Users Group International Conference. (forthcoming) Cary, NC: SAS Institute, Inc.
- Wind, J. Jacob. 1989. "TAKING THE BEST ADVANTAGE OF SAS/GRAPH SOFTWARE'S GEOGRAPHIC INFORMATION SYSTEM FEATURES". Proceedings of the 14th Annual SAS Users Group International Conference. pp. 214-219. Cary, NC: SAS Institute, Inc.
- Wind, J. Jacob & VanGorder, Jeff. 1990. "USING SAS/GRAPH(R) MAPS AS A NUCLEUS FOR AN INTERACTIVE GEOGRAPHIC INFORMATION SYSTEM". Proceedings of the 15th Annual SAS Users Group International Conference. (forthcoming) Cary, NC: SAS Institute, Inc.

Notes

- ¹ SAS and SAS/GRAPH are the registered trademarks of SAS Institute, Inc., Cary, NC, USA.
- ² WISAS mailing list diskette, March 26, 1990.
- ³ Available from: List Processing Company, Inc., 1200 Roosevelt Road, Glen Ellyn, IL 60137-6098, (708) 932-7000.
- ⁴ Rand McNally three digit zip code map.
- ⁵ Rand McNally Sales & Marketing Planning Atlas: 3-Digit Zip Code Data.

- 6 SAS/GRAPH unprojected and unreduced map data set MAPS.STATES:
SAS/GRAPH User's Guide, p. 442.
- 7 U.S. Postal Service, 1989, p. 1.
- 8 U.S. Postal Service, 1989, p. 11.
- 9 Three and five digit zip code boundary files are available from: Geographic Data Technology, Inc., 13 Dartmouth College Highway, Lyme, New Hampshire, 03768-9713, (603) 795-2183.
- 10 DENSITY subsetting: SAS/GRAPH User's Guide, pp. 319-323.
- 11 SAS/GRAPH map types: SAS/GRAPH User's Guide, pp. 243-275.
- 12 Wind, 1989, pp. 214-219.
- 13 SUGI 15 handout of J. Jacob Wind and Jeff VanGorder paper.
- 14 Description of dot maps: Monkhouse and Wilkinson, pp. 352-353.
- 15 Edgar, 1982, pp. 234-239.
- 16 SAS Communications article, 1982, pp. 22-25.
- 17 Creating various symbols on a map: Plazyk, pp. 269-276.
- 18 Example of using ANNOTATE facility to place text on a map: SAS/GRAPH User's Guide, pp. 316-317.
- 19 Historical use of circles on maps: Robinson & Sale, p 125.
- 20 A nice example of a circle map: Rand McNally 1990 Commercial Atlas & Marketing Guide, 1990, pp 122-123.
- 21 Area of circle proportional to value being represented: Robinson & Sale, 1969, p. 125.
- 22 Evans & Dolan, 1988, pp. 723-727.
- 23 SUGI 15 handout of Robert T. Wederich, Jr. paper.
- 24 Scaling circle sizes for ideal display: Robinson & Sale, 1969, pp. 126-127.
- 25 Compensation for perceived circle sizes: Robinson & Sale, 1969, pp. 125-126.
- 26 Compensation for perceived circle sizes: Muehrcke, 1978, pp. 63-65.
- 27 BUBBLE plots: SAS/GRAPH User's Guide, pp. 284-285, p. 301.
- 28 SAS/GIS demo at SUGI 15, Nashville, TN, April 1-4, 1990.

Figure 1. Zip-Derived State (WICHT01-A)

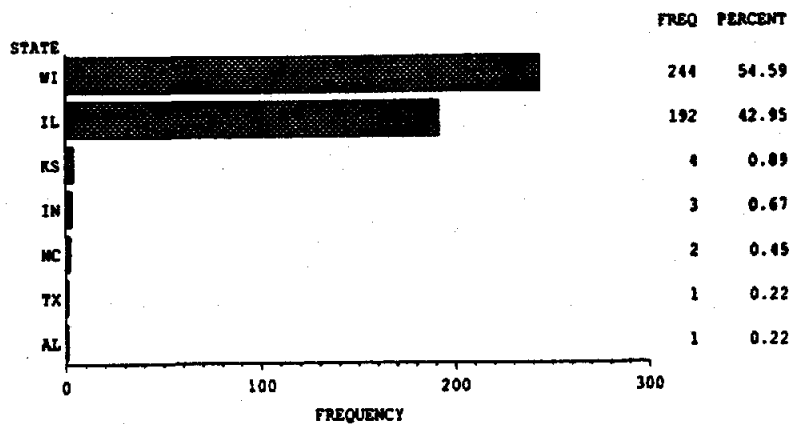


Figure 2. State-City Summary (WICHT01-B)

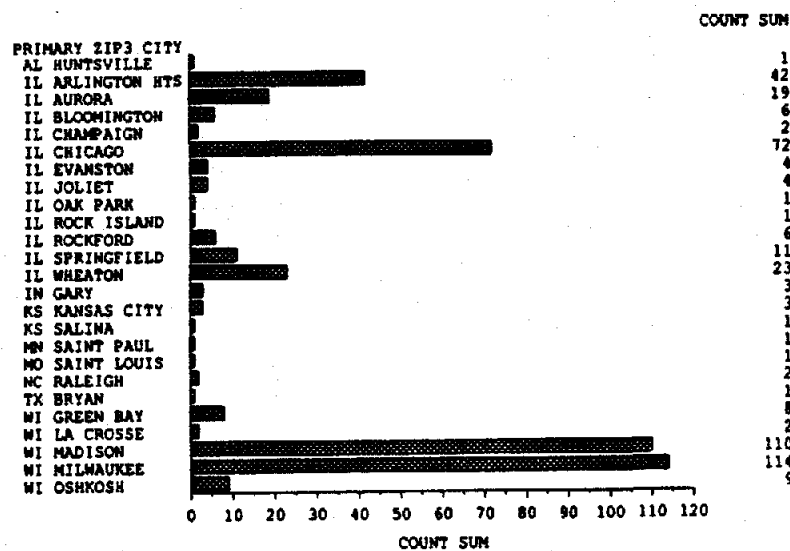


Figure 3. State-City Summary (WICHT01-C)

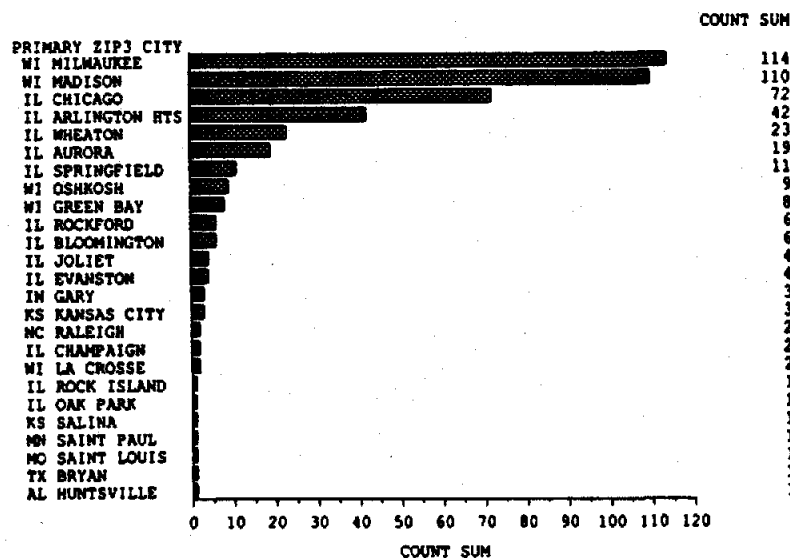


Figure 4. ZIP5 Map (WIMAP01A)

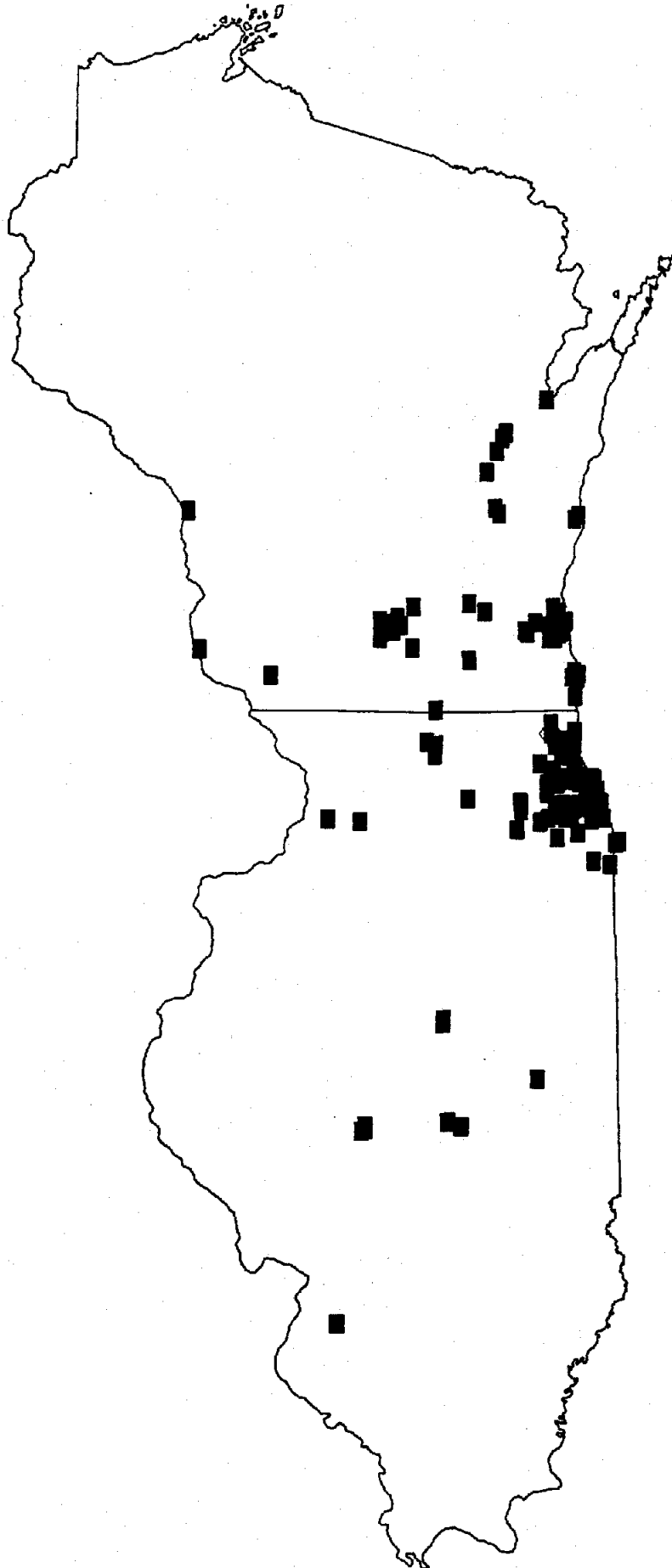


Figure 5. ZIP3 Map (WiMAP01B)

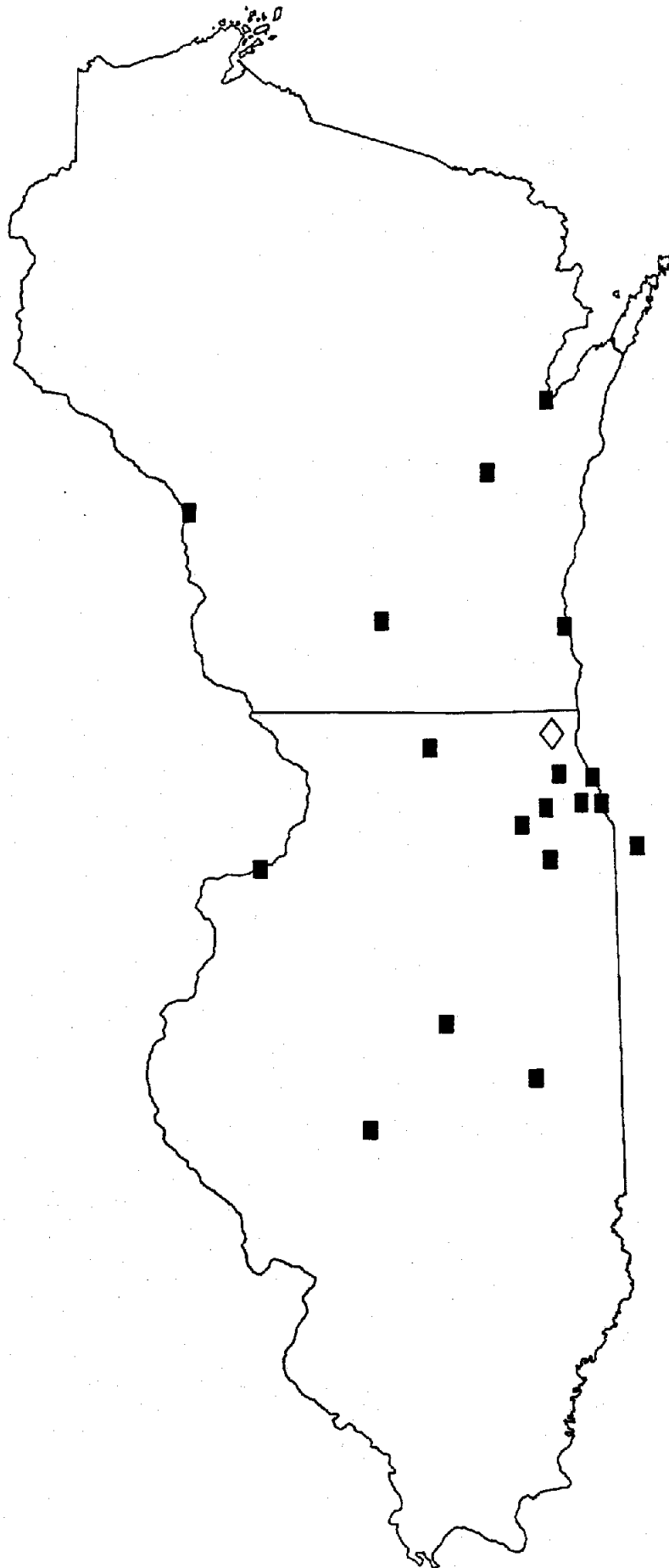


Figure 6. ZIP5 Map (WIMAP02)

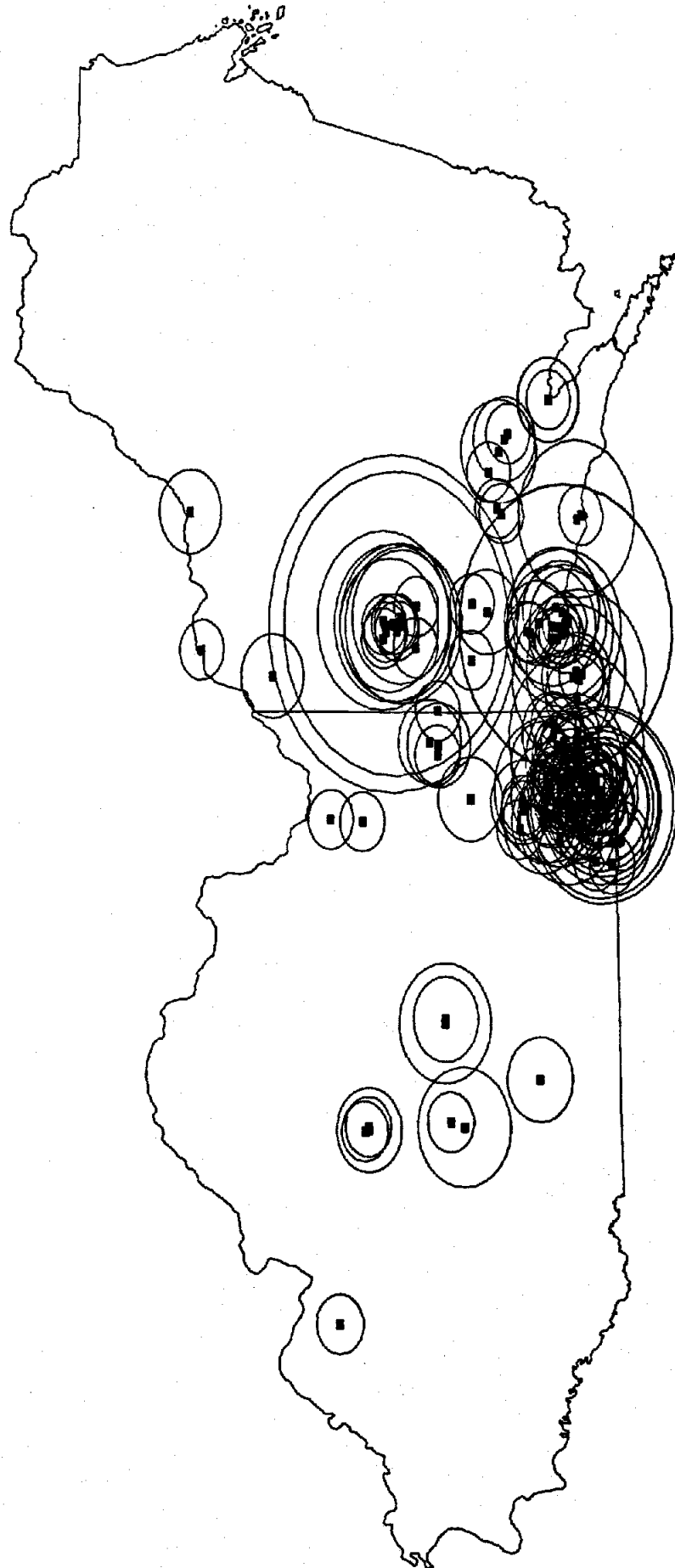


Figure 7. ZIP5 Map (WIMAP02A)

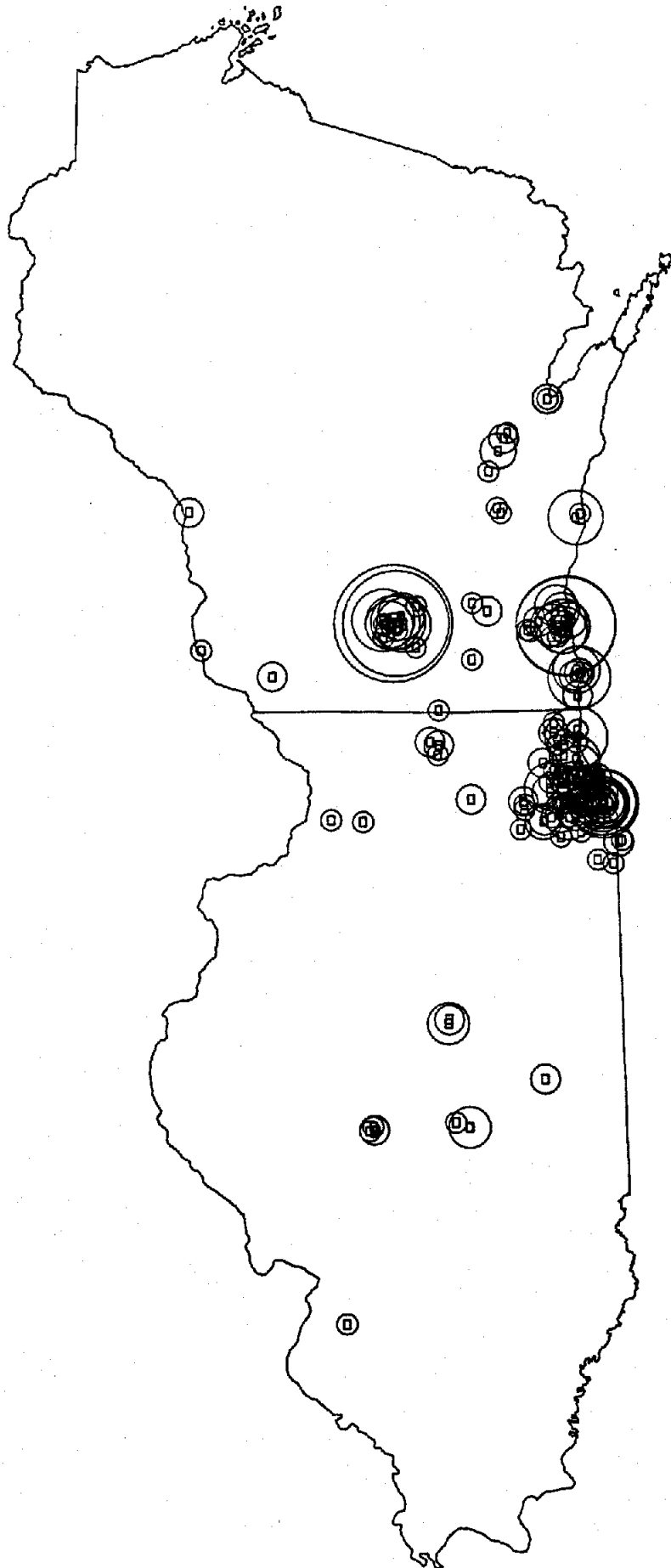


Figure 8. ZIP3 Map (WIMAP02B)

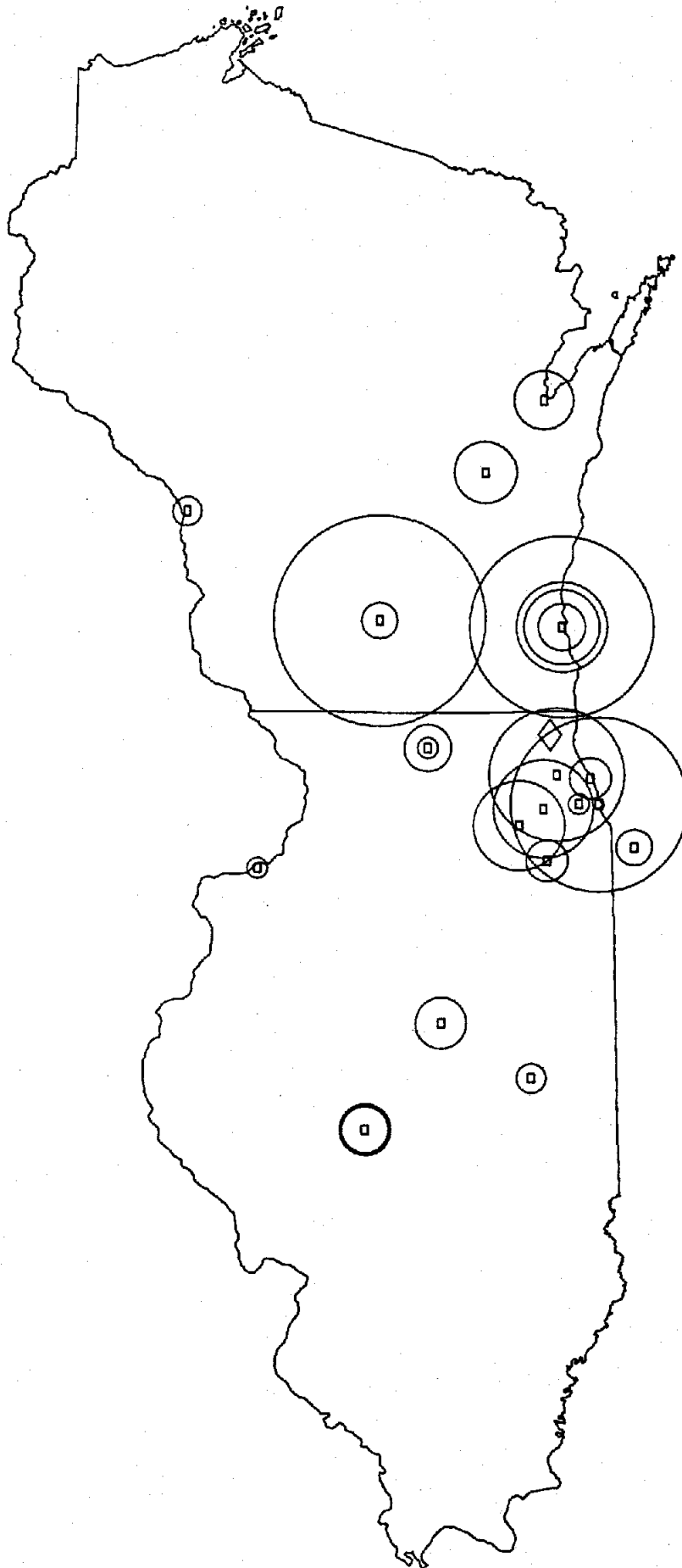


Figure 9. State-City Map (WIMAP02C)

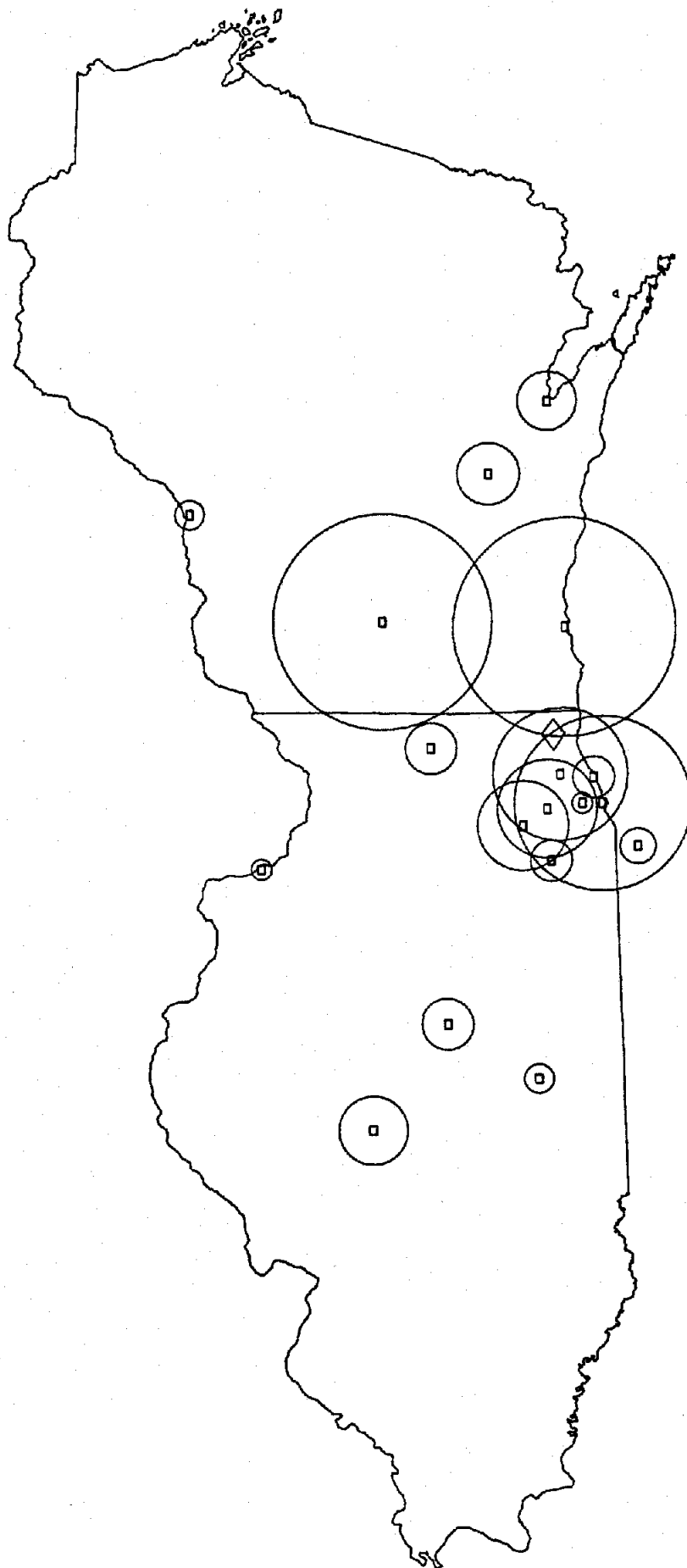


Figure 11. State-City Map (WIMAP04C)

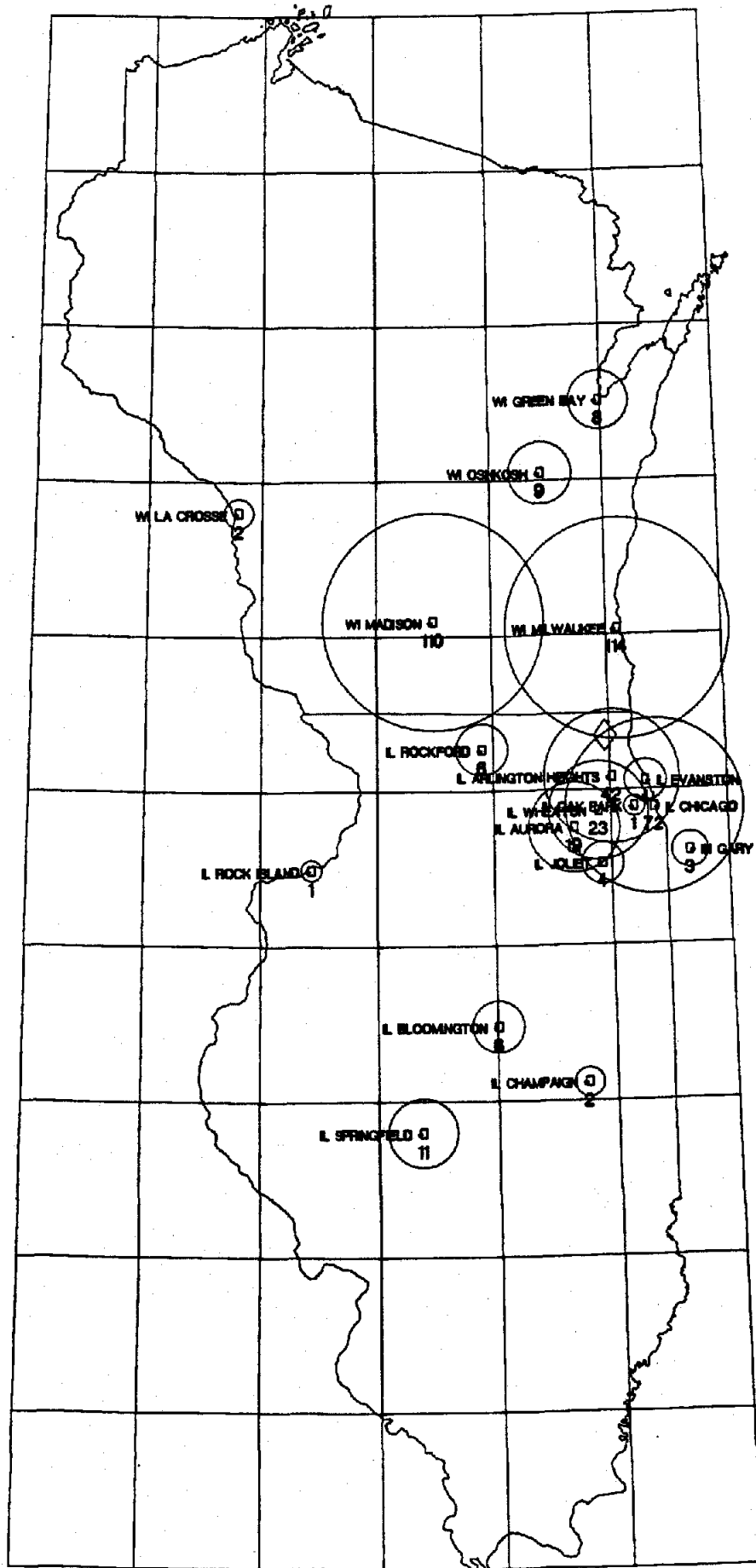


Figure 12. Lat-Lon Summary (WICHT02-A)

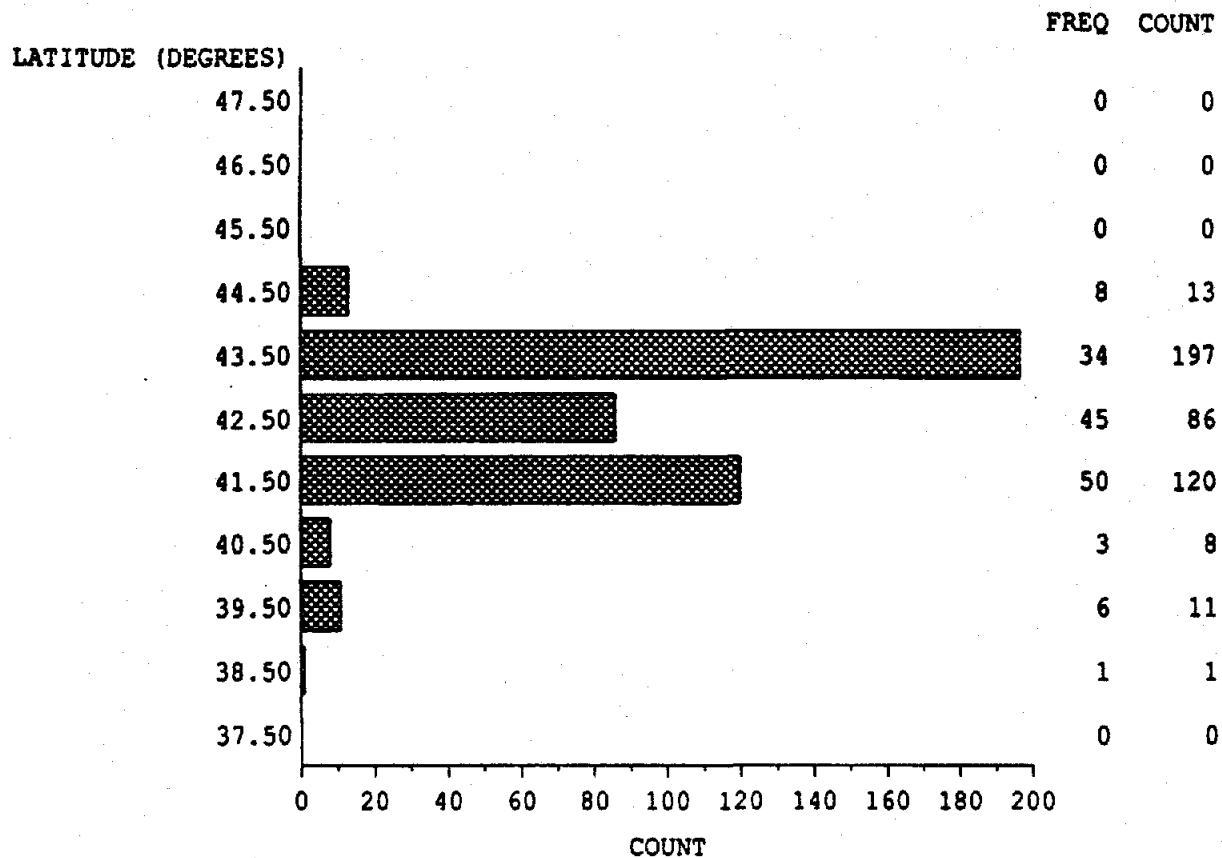


Figure 13. Lat-Lon Summary (WICHT02-B)

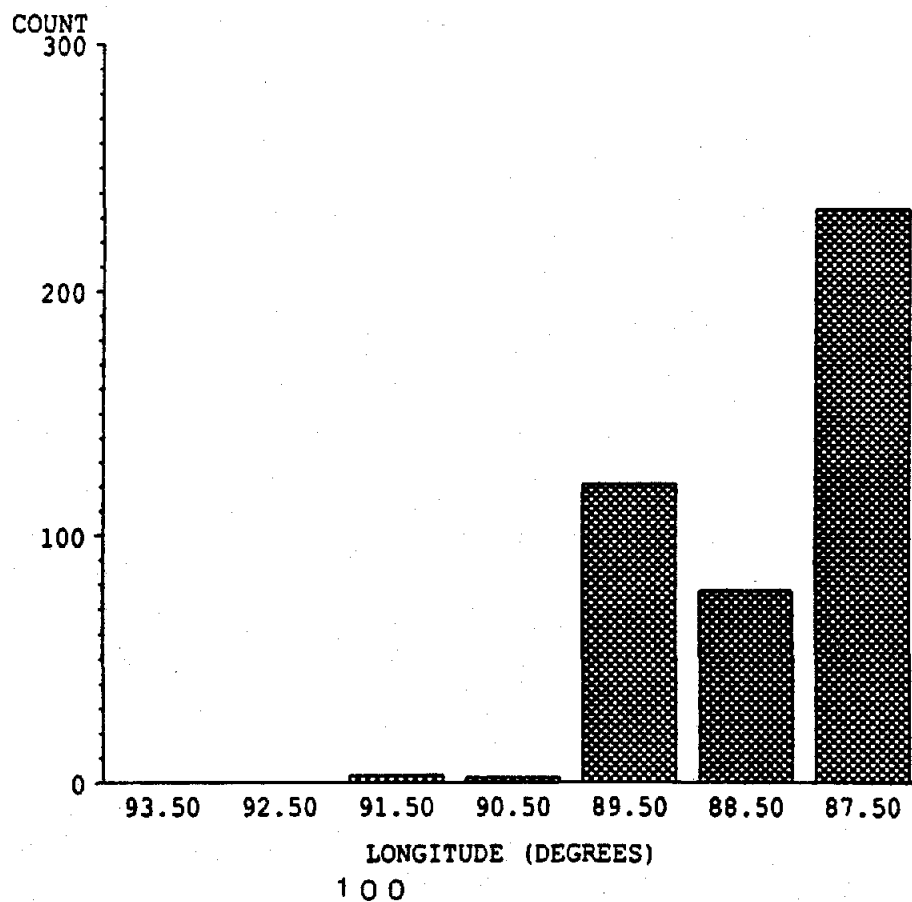


Figure 14. IL, WI, IN Center of Gravity Map (WIMAP03D)

