Presenting Spatial Data: the Statistical Map As a New Practice

by Michal Peleg[1]
Social Sciences Data Archive
The Hebrew University of Jerusalem

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Washington, D.C., U.S.A. on May 26-29, 1988

Introduction

What are the demographic characteristics of Washington, D.C., and how do they differ from those of Baltimore? What is the migration balance in the downtown area? Where do out-migrants tend to move to and what are the characteristics of mobile households? Is there a correlation between voting patterns and income level and what is the spatial distribution of this relationship? These questions represent statistical data on spatial relationships that govern much of our daily lives and occupy a variety of research fields in the social sciences.

The use of maps in presenting statistics, and the advantages and disadvantages of this method in comparison with more traditional methods will be discussed. I will present and analyse a series of maps of Jerusalem as a case study, and suggest a role for mapping and GIS (geographic information systems) in the data center environment. The function of the computer in producing maps will be mentioned briefly since this subject, which deserves a separate presentation, has already been discussed.

Why Produce A Statistical Map?

The impact of an article, proposal or report is strongly affected by the way the data is presented. Long tables of frequencies and percentages tend to wear the reader out after a few pages, whereas numbers are more easily interpreted when given visual support such as color or pattern graphics. Since the bulk of the information being provided is in the main statistical information, we can chose to provide it by traditional means: tables of frequencies, means, rates; or, if we are more sophisticated, we may instead prepare charts: pie, bar, line or three-dimensional graphs. Statistical charts and graphs are very effective in creating interest and in appealing to the attention of a reader or an audience. If one variable in the distribution is spatial — county, town, district, census tract etc. — one has the additional option of mapping as a display technique. The other presentation techniques mentioned above lack any reference to spatial distributions, to a hierarchy of geographic divisions and the spatial relations among them.

A statistical map is merely another type of graphical presentation that takes advantage of a spatial variable and uses it in an unsymbolic way. The basis of the idea lies in the capacity of the map to act as a remarkably concise summary, to convey a great deal of information, the description and implication of which could otherwise be explained only in many pages of text[2].

A map has the advantage of showing a clear picture of the spatial distribution of a phenomenon and its implications. The following questions can best be answered using statistical maps rather than tabular displays. Do neighboring geographical units tend to have similar attributes? Can we identify the demographic, economic or urban effect of one area upon its neighbors over the years? What neighborhood dominates the quarter's average value and where is it located?

Maps that present these kinds of data are based on the existence of well defined geographic units which have some national, urban or ethnic homogeneity of population. Such units may be countries, cities, census tracts, blocks, regions, etc. While physical maps present a two-dimensional picture of the ground, statistical maps show mainly quantitative relations among the spatial units. Data suitable for representation on such a map are rates of marriage, divorce, birth, mortality and crime, rates of professionals or unemployed in the labour force, etc.

The Procedure of Map Making

The procedure of preparing a statistical map involves several stages and operations. It requires certain types of input, a process of method selection, map design, selection of equipment (hardware and software) and usually, several trial and error runs until the product meets the user's or the cartographer's demands. This procedure was used to produce a few statistical maps of Jerusalem. The object of these maps is to show the spatial distribution of two population characteristics: internal migration balance and socio-economic levels. Tables 1-3 in the appendix display the numeric data.

Migration balance was calculated from the 'Records of changes' by subtracting the number of out-migrants from the number of in-migrants, including first settlement of immigrants. Table 1 contains data for 3 consecutive years and the totals for those years. As can be seen, the net migration balance in Jerusalem is positive, though very small. Detailed examination of the data indicates a rather high variance among neighbourhoods. While the old neighborhoods, both Jewish and Arabic, do not show much movement, the new suburbs absorb a considerable amount of new population. Where are these spatial units located, what areas do they cover? How do these facts influence the urban texture of the city, its transportation system, location of new industrial plants, etc.?

Table 2 presents a subset of a study on the characterization and classification of urban geographic units in Israel. This work, carried out by the Central Bureau of Statistics, has been based on a multivariate analysis of 16 demographic and socio-economic variables derived from the 1983 census. By means of factor and cluster analysis, the values of these variables have been combined into a single measure for each geographical unit, then converted to a standard distribution. Several maps of Jerusalem have been prepared to show the above data, to present the distributions and their meaning and to search for a relationship among them. (See maps 1–4 in the appendix).

Choosing the Right Method

The method and type of map must be selected according to the nature of the data, the medium of presentation, the purpose of the map, time and tools available for its preparation, and the audience for whom the map is intended. For any given set of data there is no absolute standard or criterion for selecting a particular type of map. With few exceptions, any set of spatially related statistical data can be portrayed in more than one map.

Non-quantitative maps are probably the simplest in their design. However, aspects of quantity are still the core of statistical mapping. Dickinson[3] divides quantitative statistical maps into 3 main types:

- Those in which quantities occur at a series of points
- Those in which quantities are contained within given areas
- Those in which quantities occur along a series of lines

This paper concentrates on the second type in which data, usually demographic, economic or social, have a spatial distribution that calls for emphasis. That is, a quantitative characteristic distinguishes one area from another and the rates of this differentiation are to be presented. Areal distributions can be presented by several mapping techniques: dot and isoline maps, repeated quantitative symbols, statistical diagrams and other methods. However, one of the more common ways to portray rates, based on clearly delimited areal units, is the choropleth map.

Choropleth maps are more common than any other type, although there are difficulties of a general nature associated with their design and use. The main difficulty lies in representing a quantity that is related to a given boundary-line area. Since the various areas in a map generally are not of equal size, the visual presentation of the areas distorts the distribution which is the subject of the map. Such bias is typical of choropleth maps and there are several techniques with which to deal with it. Another problem arises from a lack of information as to how items are arranged within a boundary-line. The larger the areas are, the greater the bias may be, and vice versa. Selecting smaller and more homogeneous areas (such as block or census tract) can reduce spatial distortion but it also means a larger data-set, an expensive base map, and sometimes results in an over-crowded map.

This method assumes uniform distribution of the subject variable across an area. The main problem is that the varying sizes of area have a substantial effect on the visual perception of quantities and their relationships. This difficulty can be overcome by displaying quantities as density per land unit or as a ratio, percentage, or per capita figure. This explains why a choropleth map is not suitable for representing purely numeric quantities. It can however, by itself or mixed with another technique, represent more than one variable. In such a case, one variable might be displayed by colors or shades and the second variable by crosshatching, repeated patterns, etc. An attempt to present more than 2 variables on a single map may result in a garbled and unclear picture. Obviously, there is a distinct limit to the amount of information which can be incorporated into a single chart and still be appreciated.

The choropleth method was chosen for the Jerusalem series of maps both to present single characteristics and for the bivariate map. The reason for this decision lies in the nature of the data, our users' preference and available software.

Input

Two types of input are essential for creating a statistical map: a data-set that matches statistics to geographic areas and a base map suitable both to the selected method and map type.

Drawing a machine readable base map is one of the first steps in the preparation of any map type. The base map should be designed as a simple, outline map with a minimum of detail so that it can be further used for the production of other statistical maps of the same area. It normally consists of an outline of the contour of the geographical entity which is the subject of the map (city, county, country), and the polygons that represent its division into sub-areas. Such a map may include a few titles, symbols and other background details to make reading easier. However, unnecessary information may confuse the reader, detract attention from the main distributions and obscure the message of the map. Features such as rivers, place names and boundaries should be kept to a minimum, although they may usefully act as a "geographical framework" for the main subject. Extensively used base maps are now commercially available from government mapping agencies as well as from software houses. When the procedure is computer-aided, the base map is a file of digitally coded land and political boundaries, generated either by digitization or by scanning an accurate large map. In such a map each polygon represents an area according to a certain scale. Since one base map is normally used for producing numerous statistical maps, the investment in preparing a good map pays off.

Before preparing the base map for choropleth mapping, the appropriate geographic units must be selected. They should be small enough to be homogeneous and large enough to be of significance. The level of available data is another factor in selecting the geographical division. Jerusalem is divided into 3 hierarchical levels of areas - quarters, sub-quarters and statistical areas (census tracts). As a map of the quarters would appear to be too heterogeneous, while a map of the 155 statistical areas would be over-crowded, we chose sub-quarters as the unit of analysis. The series of Jerusalem base maps was created by digitizing the statistical area boundaries. Then, two less detailed maps were generated from the original one, one for sub-quarters and the other for quarters.

Interval Selection

In order to show the value of a variable (rates, percentages, or other statistics) for each individual area, the total range of values should be grouped into categories which are distinquished by different colors or shading. Whether the variable is discrete or continuous, the selection of the dividing points among groups considerably affects the significance of the map. The number of groups and their dividing points obviously depends on the range, shape and variance of the distribution. Methods of group creation vary from 'simple' thru automatic quantile groupings of the standard deviation. Maps 1 and 2 use the same intervals selected by quantiles with a modification that takes zero immigration balance as a dividing point. This results in a series of 5 groups: 3 categories for negative migration and 2 for positive values. In map 5, the two variables were grouped into 5 and 6 categories respectively which result in 30 different groups, each with a separate color and graphical presentation. It sounds confusing, but since each variable has a meaningful graphic scale the reader will immediately perceive the patterns and their meanings.

Pattern and Color Selection

Now that the intervals have been established, a suitable range of shading must be selected. Whether cross hatching or color, the principle is to arrange the patterns such that the lightest color or shade corresponds to the smallest magnitude with increasing densities showing successively higher values. In monochrome map production, this means arranging the density of lines in such a way as to give an optical effect from light to dark with respect to color or pattern. Each color or cross hatch pattern indicates an interval in a series of rates or percentages. Generally speaking, it is suggested that one not use too many different colors on any one map but rather one or two bright colors with their slightly muted tones. The use of color enhances the options of choropleth maps, and bright color maps attract the eye more than any other method; it should however be carefully applied.

Maps 1 and 2 show different color selections for the same data set with the same intervals. Map 2 was designed using two different colors. Positive migration is symbolized by the color red and its tones and blue corresponds to the negative migration balance. In Map 1 the intervals are distinguished by gradual green tones. Which map better reveals the facts and delivers the message is not a clear choice. Maps 3 and 4 continue the monochrome shading technique, using six different tones. Map 5 was produced by two consecutive computer runs: the SE score was plotted first, then re-plotted to present the migration balance by graphical patterns.

Trial and Error Process

Drawing a statistical map is a trial and error process. It is apparent by now that the rendering of statistics into pictorial form is far from being a simple procedure. It involves fundamental decisions on methods, equipment, techniques and other parameters of map design and normally entails laborious planning and plotting. Only after producing several drafts and tests does the product satisfy the professional user.

Data Centers, GIS and Statistical Maps

A considerable portion of the collection of a social science data archive consists of geographically oriented data, either microdata or small area statistics from national, regional or local surveys. Furthermore, several data centers have already built and maintain geographic information systems (GIS) for the processing of special purpose data sets. A GIS is normally created to handle the unique characteristics of geographic distributions and it can act as a very sophisticated device to make the most of this kind of data. Such a system is typically based on database management system (DBMS) software, using a hierarchical or a relational model for its schema. A well designed GIS also is capable of performing the conversion and translation of various spatial divisions and boundaries of the same areas. At the same time, we see an increasing effort to improve the presentation of data, especially by R&D and marketing professionals in national and local agencies. The expansion of computer based graphics, mainly LOTUS-like products, stirs the appetite as well as the expectations of these researchers and their audiences. Once they have become acquainted with colored bar and pie charts, they wish to see clear statistical maps when geographical distributions are being displayed. Since statistical maps are a natural output of GIS, they should be produced by data archives as an expansion of data processing services. It is necessary to enhance the GIS to include a coordinate system that corresponds to the geographical units involved. Recently developed software packages can do this, either on a personal computer (PC) or, for large and complicated systems, on a powerful mainframe. However, the success of such an activity depends largely on the co-operation of a professional cartographer willing to share in the enterprise. The project as a whole can enhance both the geographic orientation and the related services offered by the data center.

Acknowledgments:

To Amnon Shebo who has prepared the maps for this presentation.

References

1. Bachi, R., Cluff, V. V., and S. Ever-Handani. Geo-statistics in urban ecology - application to Toronto religious sub-populations. -in press April 1988.

2. Cuff, D. J. and M. T. Mattson. Thematic maps: their design and production. New York: Methuen, 1982.

3. Dickinson, G. C. Statistical mapping and the presentation of statistics. London: E. Arnold, 1973.

4. Marble, D. F., Calkins, H. W., and D. J. Peuquet. Basic reading in geographical information systems. Williamsville, NY: SPAD Systems Ltd., 1984.

5. Schmid, C. F. Handbook of graphic presentation. New York, NY: Ronald Press Co., 1979.

Table 1: Migration balance by sub-quarters (1983, 1984, 1985)

Sub-quarter		year			Total
Code	Name	1983	1984	1985	1983-85
	TOTAL	1,557	1,563	2,165	5,285
11	Ramot Eshkol	304	-452	-488	- 636
12	Beit Israel	-593	-720	-485	-1,798
13	Musrara	-284	-209	- 36	- 529
14	Center	-382	-254	-153	- 789
15	Rehavia	- 93	-236	-124	- 453
16	Nahlaot	-925	-541	-515	-1,981
17	Geula	-817	-489	-810	-2,116
21	Romema	- 22	247	121	346
22	Givat Shaul	15	543	1,432	1,990
23	Beit Hakerem	- 70	-354	-392	- 816
24	Givat Ram	50	14	47	111
25	Bait Vegan	75	-241	- 42	- 208
31	Qiryat Hayovel (S)	-404	-315	-139	- 858
32	Qiryat Hayovel (N)	-430	-262	-318	-1,010
33	Ir Ganim	-347	-216	-214	- 777
41	Gonen	-441	-422	-521	-1,384
42	Rassco	- 61	-307	-267	- 635
51	German Colony	-273	-212	-347	- 832
52	Talbieh	-100	-114	-135	- 349
53	Geulim	- 82	-126	82	- 126
54	Talpiot	-247	119	55	- 73
61	Christian Quater	- 14	1	57	44
62	Armenian Quater	13	- 3	43	53
63	Jewish Quater	130	101	-136	95
64	Moslem Quater	-333	- 58	- 50	- 441
71	Shuafat	508	- 23	241	726
72	Ramot Alon	2,836	3,387	2,686	8,909
73	Neve Yaacov	778	136	501	1,415
74	Givat Shapira	- 11	-138	- 11	- 160
75	A-Tur	-215	- 96	-109	- 420
76	Sheikh Jarrah	-121	- 93	- 70	- 284
81	Silwan	158	179	118	455
82	Sur Baher	49	288	56	393
83	East Talpiot	794	759	440	1,993
84	Gilo	3,236	2,200	1,648	7,084

Source: Jerusalem Statistical Yearbook, 1986

Table 2: Migration balance 1985, by sub-quarters - rates per thousand

Sub-quarter Code	Name	Population	Migration Balance	Migration Per Thousand
TOTAL		457,700	2,165	4.73
11	Ramot Eshkol	14,900	- 488	- 32.75
12	Beit Israel	24,000	- 485	- 20.21
13	Musrara	2,100	- 36	- 17.14
14	Center	5,000	- 153	- 30.6
15	Rehavia	7,800	- 124	- 15.9
16	Nahlaot	8,200	- 515	- 62.81
17	Geula	21,500	- 810	- 37.67
21	Romema	14,200	121	8.52
22	Givat Shaul	9,400	1,432	152.34
23	Beit Hakerem	16,900	- 392	- 23.20
24	Givat Ram	2,800	47	16.79
25	Bait Vegan	17,900	- 42	- 2.35
31	Qiryat Hayovel (S)	9,800	- 139	- 14.18
32	Qiryat Hayovel (N)	11,700	- 318	- 27.18
33	Ir Ganim	9,800	- 214	- 21.84
41	Gonen	24,000	- 521	- 21.7
42	Rassco	13,500	- 267	- 19.78
51	German Colony	11,100	- 347	- 31.26
52	Talbieh	3,800	- 135	- 35.53
53	Geulim	9,700	82	8.45
54	Talpiot	10,900	55	5.05
61	Christian Quater	4,500	57	12.67
62	Armenian Quater	2,000	43	21.50
63	Jewish Quater	2,200	- 136	- 61.82
64	Moslem Quater	17,600	- 50	- 2.84
71	Shuafat	32,400	241	7.44
72	Ramot Alon	20,100	2,686	133.63
73	Neve Yaacov	14,800	501	33.85
74	Givat Shapira	9,300	- 11	- 1.18
75	A-Tur	20,800	- 109	- 5.24
76	Sheikh Jarrah	7,500	- 70	- 9.33
81	Silwan	24,000	118	49.17
82	Sur Baher	15,900	56	3.52
83	East Talpiot	11,800	440	37.29
84	Gilo	23,900	1,648	68.95

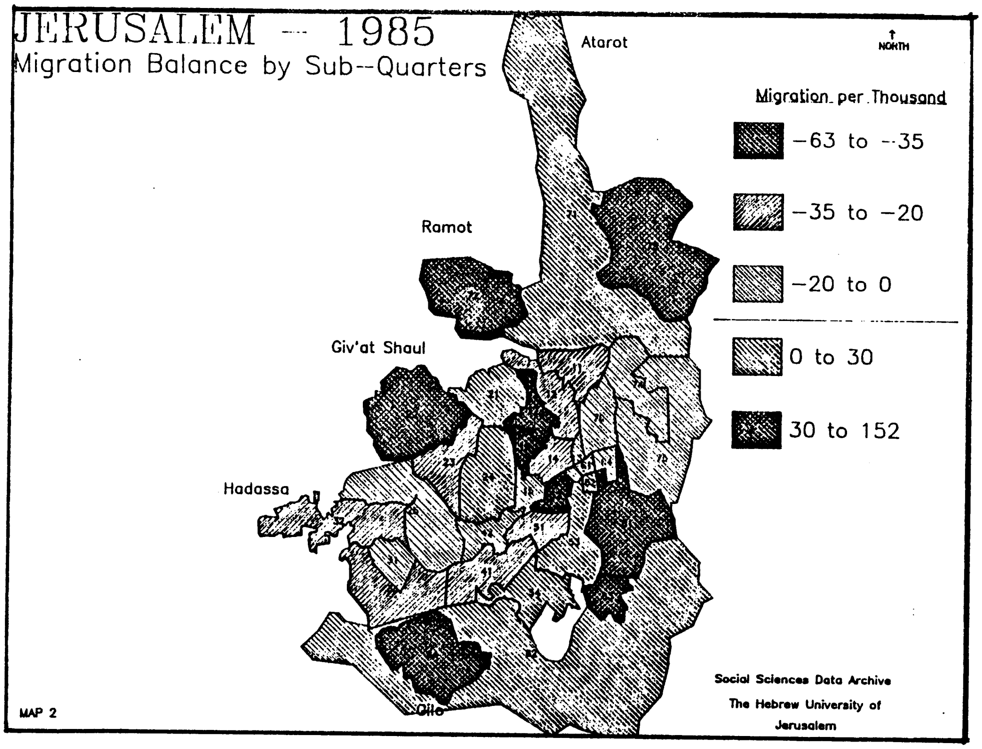
Source : Jerusalem Statistical Yearbook, 1986

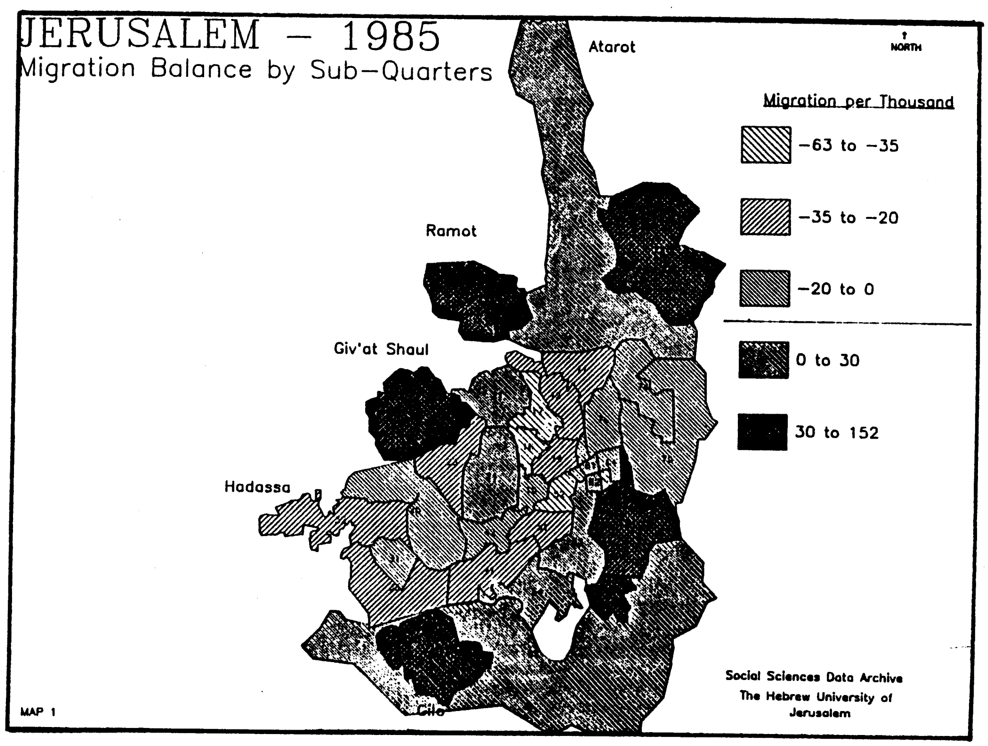
Table 3: Socio–economic scores of sub–quarters, 1983

Sub-quarter Code	Sub-quarter Name	Socio-Economic Score	Population
11	Ramot Eshkol	0.71	14,803
12	Beit Israel	-0.18	23,168
13	Musrara	0.05	2,137
14	Center	0.25	2,367
15	Rehavia	0.84	7,490
16	Nahlaot	0.12	4,497
17	Geula	-0.10	21,417
21	Romema	0.01	12,341
22	Givat Shaul	0.17	6,030
23	Beit Hakerem	0.99	17,522
24	Givat Ram	0.67	1,584
25	Bait Vegan	0.51	17,248
31	Qiryat Hayovel (S)	0.51	9,018
32	Qiryat Hayovel (N)	0.42	11,924
33	Ir Ganim	0.26	9,442
41	Gonen	0.24	24,015
42	Rassco	1.19	13,466
51	German Colony	0.66	11,614
52	Talbieh	0.61	4,189
53	Geulim	0.37	8,945
54	Talpiot	0.52	9,037
61	Christian Quater	-1.01	4,322
62	Armenian Quater	-0.05	2,080
63	Jewish Quater	0.22	2,042
64	Moslem Quater	-2.66	17,102
71	Shuafat	-1.84	29,359
72	Ramot Alon	0.52	11,483
73	Neve Yaacov	0.45	13,111
74	Givat Shapira	1.03	6,876
75	A-Tur	-1.88	19,787
76	Sheikh Jarrah	-0.53	7,602
81	Silwan	-2.66	22,308
82	Sur Baher	-1.72	14,412
83	East Talpiot	0.69	9,419
84	Gilo	0.64	17,486

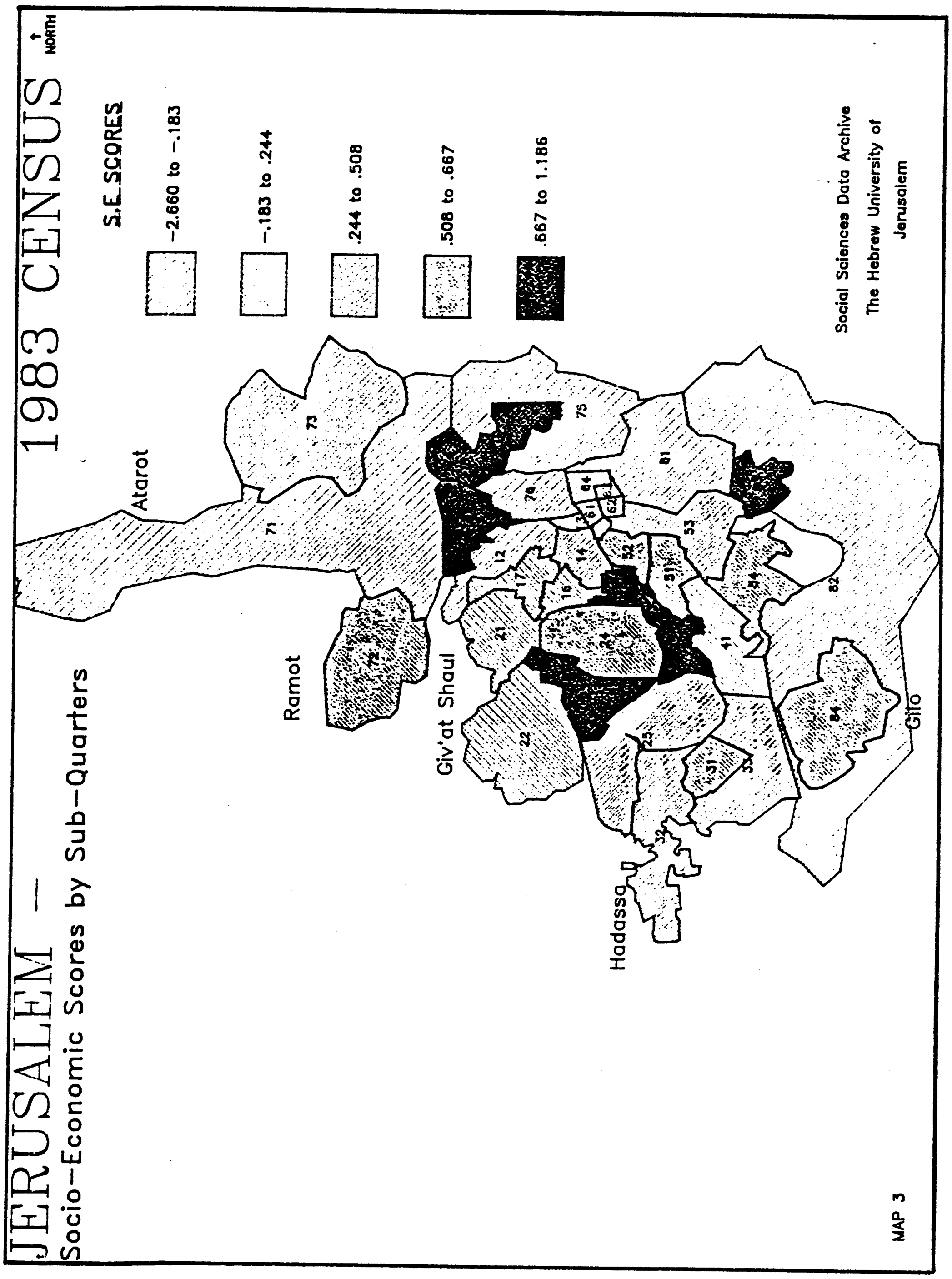
Source : The Central Bureau of Statistics

Maps 1 – 2: Migration Balance by Sub–Quarters, 1985

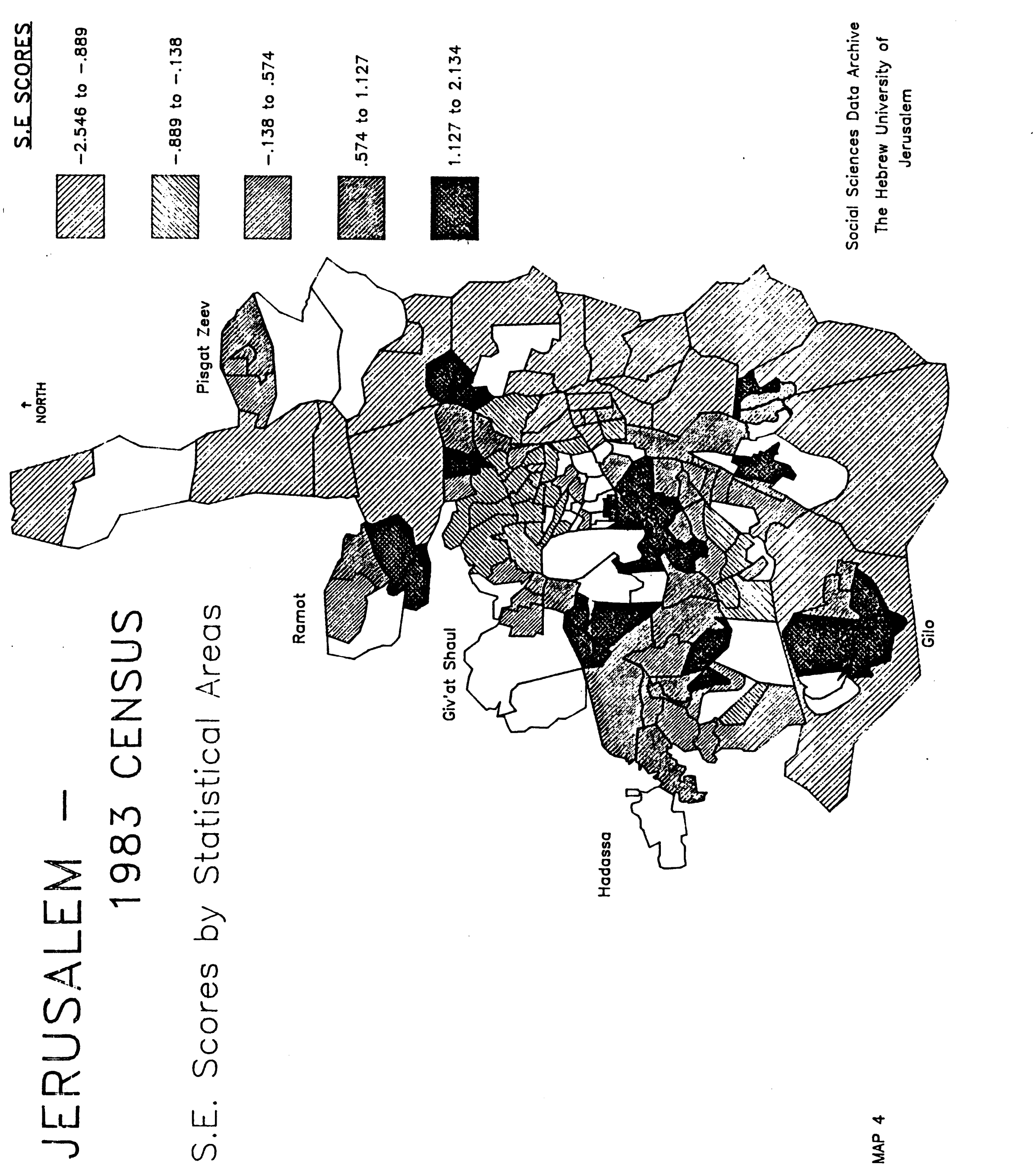




Map 3: Socio-Economic Scores by Sub-Quarters, 1983



Map 4: Socio-Economic Scores by Statistical Areas, 1983



Maps 5: Migration Balance and Socio-Economic Scores by Sub-Quarters