Methodological Approach to the Study of Urban Ecology of the Jews in the Diaspora

Roberto Bachi

Some of the papers given in this Congress are devoted to the presentation of a joint research project on urban ecology of the Jews in the Diaspora, carried out at the Hebrew University of Jerusalem and at the University of Delaware. This paper illustrates the methodological approach used in the treatment and graphical presentation of the statistical data collected in the course of this work.

A large amount of heterogeneous statistical data was assembled. This heterogeneity had various reasons:

(a) sources of data differed widely, ranging from accurate and detailed modern, official population censuses to historical enumerations found in archives, sample surveys carried out by research institutions of Jewish communities, etc.;

(b) the data related to cities scattered all around the world;

(c) in some cases, data related to one point in time, while in others they spanned over centuries;

(d) data referred to Jewish populations, total populations and selected subpopulations;

(e) the criteria according to which cities were divided into regions differed strongly (e.g. minute classification of the territory of a city into many hundreds of census tracts, subdivision according to historical or administrative quarters or wards, *ad hoc* groupings of urban areas used in surveys by Jewish communities or research bodies, zip code areas, etc.). In cities for which information was available for many dates, regional frameworks often changed in the course of time.

For each city and point of time, the basic statistical information available was generally as follows:

(a) A map showing the territorial boundaries of the city and of each of the regions j into which this territory is divided. As an example, Figure 1 shows the division of the Metropolitan Area of Toronto (Canada) in 1981 into 608 census tracts (which are seen to have different shapes and areas).

(b) A table listing (by name, ordinal number or otherwise) each of the regions j and indicating the number of elements of each of the populations studied (for example, Jews, Protestants, Anglicans, United Church, Catholics, ... , total population) and the number of these elements enumerated in each region j. For example, a table referring to these five religious denominations and the total population of Toronto in 1981 includes 3,648 data specified by regions and six totals for the entire city.

It was desirable to treat this large and complex statistical material by using standard methods which would allow meaningful, straightforward and systematic comparisons of the ecological distributions of the Jews, other relevant subpopulations and total populations at different times and in different cities.

Main Solutions: Methods Used

In order to attain this goal, mainly the following methods were used:

(a) A standard *rational graphical system* to produce automated maps based on strictly scientific principles; such maps are of two main types:

1. *Simple maps* representing separately each distribution (for example, the distribution of the Jewish population or of the total population in a city at a given date). Standardized production allows visual comparison between maps.
2. *Derived maps* in which a comparison between two distributions (e.g., Jewish vs. total population) is worked out by the computer; such maps graphically show the results of the comparison.

The principles on which rational maps are based and the methods used in their construction are explained below. Rational maps enable us to obtain a visual overview of general characteristics of each distribution studied, or of broad similarities and dissimilarities between pairs of distributions, compared by derived maps; and to study, whenever desired, local peculiarities in parts of the distributions which may be of particular interest for ecological research in each city, time and type of population studied.

(b) Methods enabling us to define the *gross* and *net territorial range* over which the elements are distributed for each population distribution studied.

(c) Methods enabling us to indicate or measure, by means of summary geostatistical parameters, position in the plane, absolute and relative dispersion, and shape characteristics of the territorial range over which the studied population is distributed, patterns of distribution for the population netted from influence of area, shape and other characteristics of the territorial range, etc. These parameters allow quantification in the study of population distributions, and quick comparisons between the characteristics of each distribution in different times, of the distributions of different populations or of subpopulations over the same city. Net parameters allow comparisons between characteristics of analogous distributions in different cities.

(d) Calculation of an overall measure of the discrepancy between two distributions (e.g., P = Jewish population distribution, Q = total population distribution), based on the distance to be covered in order to transform, say, P into Q. This method permits very quick comparisons between two distributions over the same territory or between the distributions of a given population at different times. By rendering this measure dimensionless, comparisons may be carried out regarding analogous pairs of distributions in different cities.

Rational Mapping Methods Employed

Simple Rational Maps

To illustrate the mapping methods which have been developed and employed in our project, some graphic examples are given of the Jewish population distribution, of other religious sub-populations and of the total population of the Toronto Metropolitan Area in 1981.

From the available primary information which is detailed according to the 608 census tracts of Figure 1, it is immediately obvious that direct perusal of the detailed data would be of no use: even persons closely familiar with the ecology of Toronto would not be able to associate in their minds the numbers designating each of the 608 tracts with position, size and shape of the tract and frequencies of the religious groups as given by the relevant table. Graphical presentation is therefore a necessary first step in our research.

For this purpose, a mapping system was adopted based on the following principles:

(a) We start by identifying the *gross territorial container*: in our case, this was already defined – being the Metropolitan Area of Toronto.

(b) The gross container is subdivided into a convenient number (in the case of Toronto, about $N = 1{,}000$) of cells i of regular hexagonal shape (see Figure 2) and of constant area $\bar{A} = A/N$ (where A = area of the gross container). In the specific example of Toronto, $\bar{A} = 3.8$ sq. km. Frequencies by census tracts are transformed into frequencies by cells.

(c) In order to facilitate comparisons between the distributions of various population groups, all distributions have been standardized by calculating relative frequencies ${}^{c}f_i = {}^{c}F_i/F$ (where $F = \Sigma_j F_j = \Sigma_i {}^{c}F_i$) and multiplying them by a convenient coefficient K (here $K = 10{,}000$). Values $K^{c}f_i$ are rounded to the nearest integer ($K_i \cong K^{c}f_i$). We thus obtain a distribution *equivalent* to the original, which indicates the number K_i of population elements (out of a total of $K = 10{,}000$) found in each cell i.

(d) Each K_i is represented on the map by a Graphical Rational Pattern (GRP) (Bachi, 1968, 1978b). GRP is a device enabling us to represent any integer $n = 10t + u$ included between 1 and 100 by a symbol formed by u unitary marks of area a (which indicate the units) and t marks of area 10a which indicate the tens (see Figure 3). As the GRP requires little space on the map, GRP representing K_i can be put on the map in correspondence to each cell i.

Figures 4 and 5 give examples of these simple maps, representing respectively the distributions of the total population and Jewish population in Toronto in 1981.

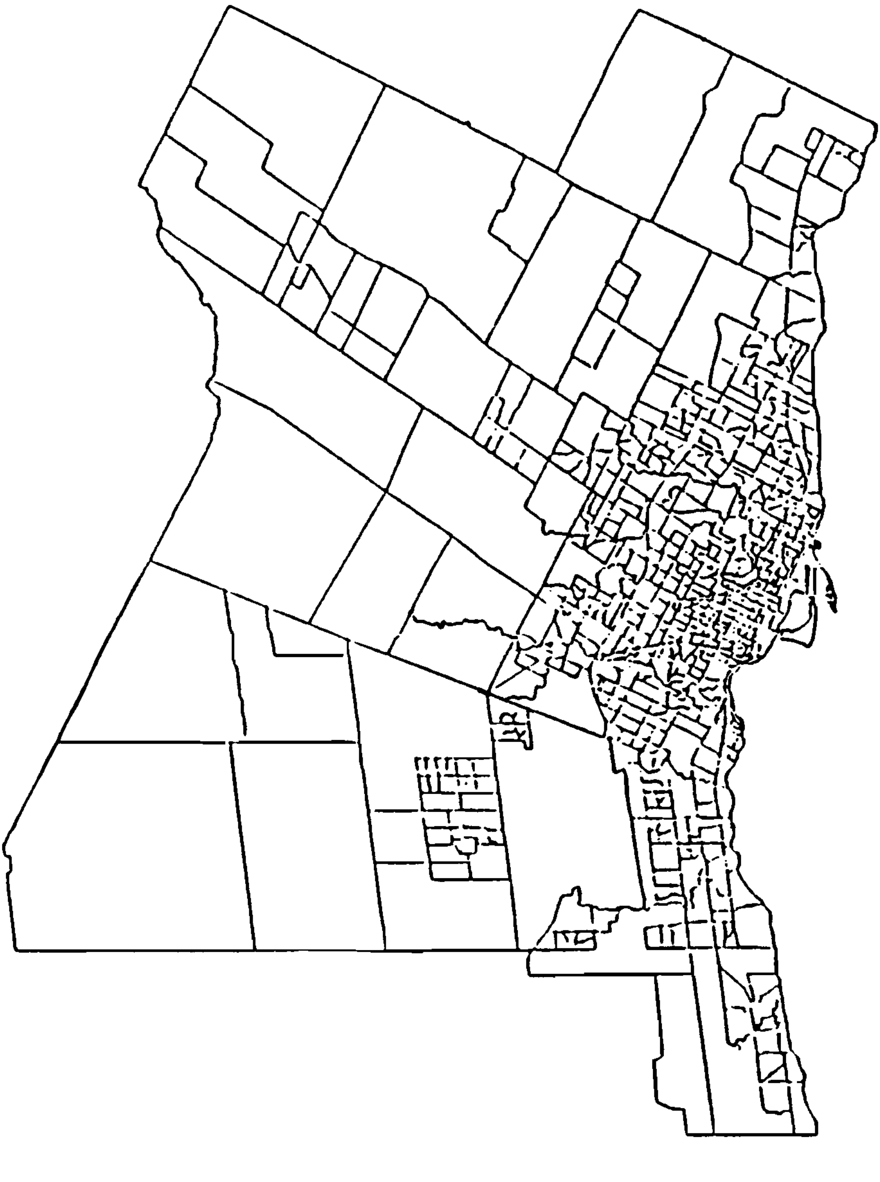


FIGURE 1. TORONTO METROPOLITAN AREA, 1981, DIVIDED INTO 608 CENSUS TRACTS

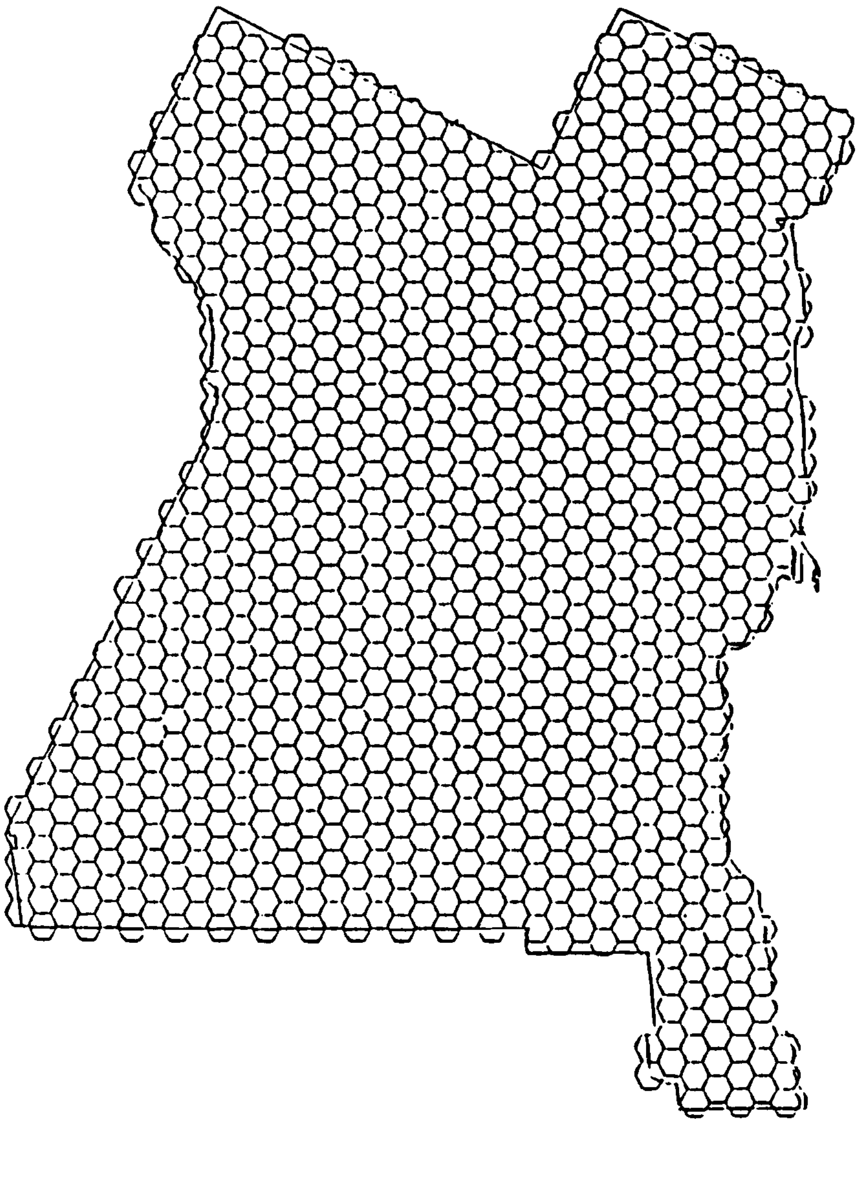
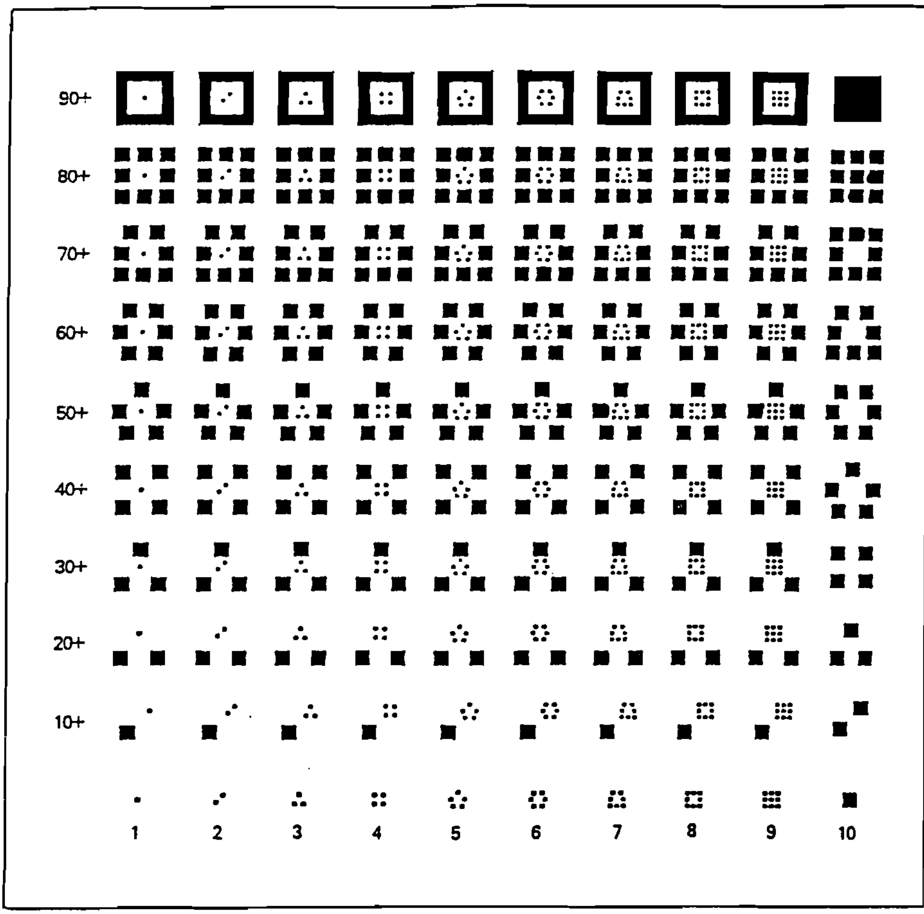


FIGURE 2. TORONTO METROPOLITAN AREA, DIVIDED INTO APPROXIMATELY 1,000 HEXAGONAL CELLS

FIGURE 3. SCALE OF COMPUTERIZED GRAPHICAL RATIONAL PATTERNS SHOWING INTEGERS FROM 1 TO 100



Comparative and Cluster Maps

In the comparative study of distributions (such as those of Jews and total population) we may compare our visual impressions of maps (such as Figures 4 and 5). However, we may wish to deepen our analysis and compare systematically – cell by cell – the frequencies of the two populations studied. This might be a complex and tiring operation unless done by the computer.

Let us consider equivalent frequencies K^{P}_{i}, K^{Q}_{i} of two populations P and Q calculated as explained above, with both total frequencies of P and Q being equal to K = 10,000. In the Toronto example of Figures 6 and 7, P = Jews and Q = total population.

The computer calculates for each cell i:

(a) The surpluses $SP_i = (K^{P}_{i} - K^{Q}_{i})$, in cells in which $K^{P}_{i} > K^{Q}_{i}$. Figure 6 represents the surpluses of Jews over total population: these are very concentrated in a central zone.

(b) The surpluses $SQ_i = (K^Q_i - K^P_i)$, in cells in which $K^Q_i > K^P_i$. Figure 7 represents the surpluses of total population over Jews; these are distributed over a much wider area.

(c) The number of elements I_i of P and Q matching (or intersecting) in i: $I_i = \min(K^P_i, K^Q_i)$. A map (not reproduced here) of intersection between P and Q has been compiled showing I_i by means of the appropriate GRPs proportional to each of these variables.

Whenever maps can be printed in colors, we may represent SP_i in GRP in one color and SQ_i in another color, while cells in which $K^Pi = K^Qi = 0$ are left blank.

Obviously, $\Sigma SP_i = SP$ and $\Sigma SQ_i = SQ$ are equal. We can write $SP = SQ = S$, and $S/K = \Delta$, which is Duncan's index of dissimilarity (Duncan and Duncan, 1955). Indicating $I = \Sigma I_i$, it is found that $I + S = K$ and $I/K = 1 - \Delta$.

Maps of surpluses and intersection have proved themselves to be especially useful in dynamic research. Let us compare, for example, the distributions of a population in the same city at two points in time (1, 2), both distributions being reduced to K units. The map of intersections will show the ecological distribution of that part of the population which did not change between 1 and 2 (independently of total changes in the population size between 1 and 2). The maps of surpluses will show the changes which occurred in the distribution in that period.

There is considerable interest in the comparison of the distribution of any population P (reduced to K elements) to the theoretical distribution which would be found if P were distributed uniformly over the N cells; each cell would then have $\bar{K} = K/N$ elements. A map indicating the surpluses $(K_i - \bar{K})$ of actual over average frequencies (for cells in which $K_i > \bar{K}$) attracts attention to areas with high frequencies. As an example, Figure 8 shows surpluses of actual distribution of the total population of Toronto over uniform distribution.

As another example, Figure 9 shows the surpluses of distribution of the United Church population over the uniform distribution. On this map, major clusters are indicated by convex hulls drawn around their cells. Two contiguous cells with $K_i > \bar{K}$ are considered to belong to the same cluster. From this example it is seen that the United Church population has five distinct major clusters, whereas the Jewish population has one major cluster only.

Geostatistical Parameters

General Considerations

With the same information input as the maps, the computer program used in this project enables us to obtain many geostatistical parameters for each of the original population distributions considered, for each of the subsets obtained by the comparison of two sets (such as subsets of intersection or differences), for gross or net territorial containers of each of these distributions, etc.

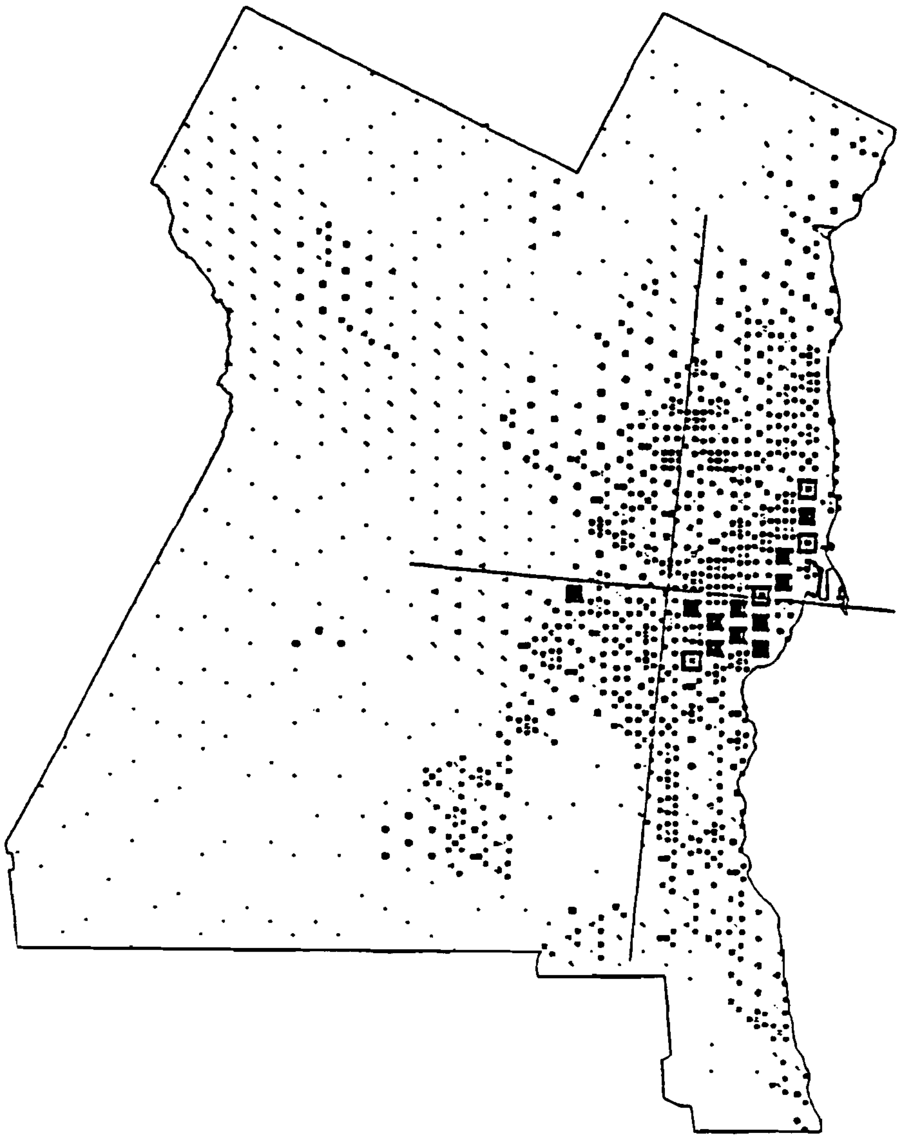


FIGURE 4. EQUIVALENT DISTRIBUTION OF TOTAL POPULATION, TORONTO, 1981

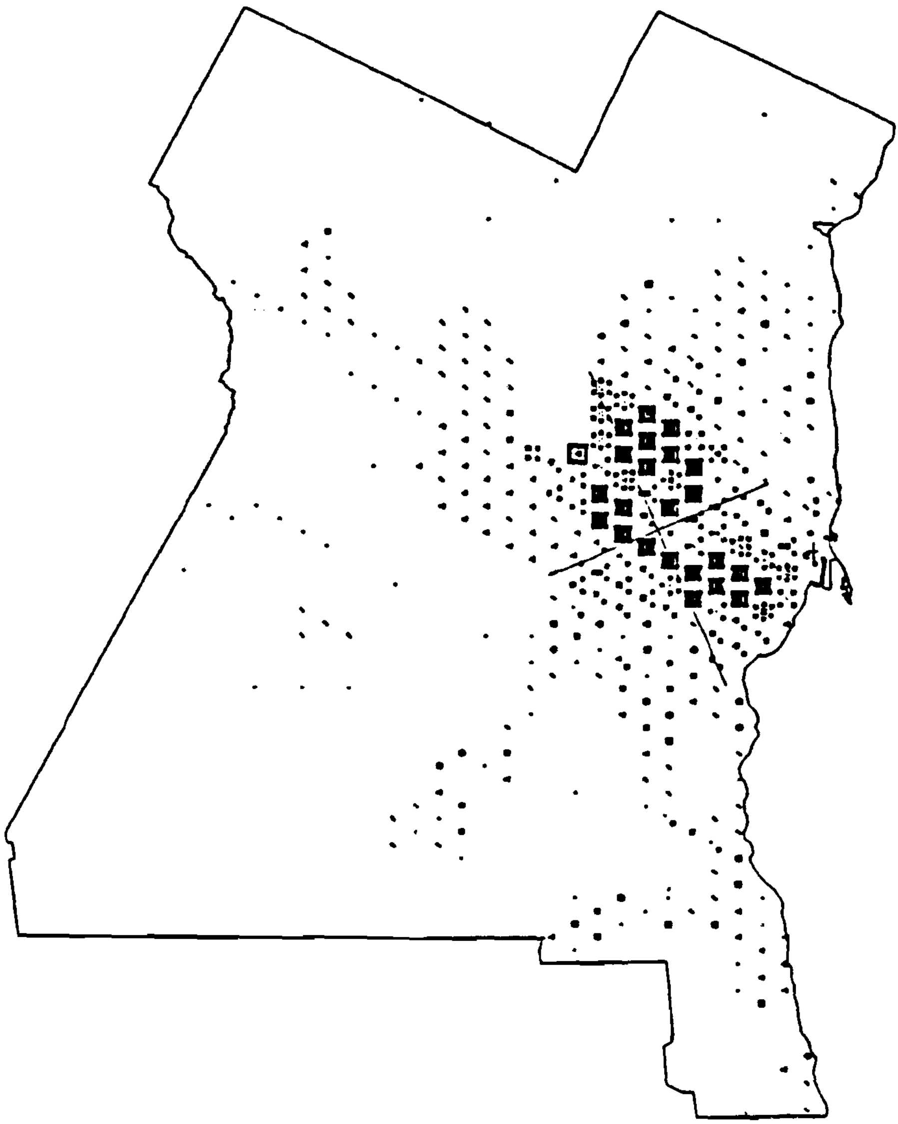


FIGURE 5. EQUIVALENT DISTRIBUTION OF JEWISH POPULATION, TORONTO, 1981

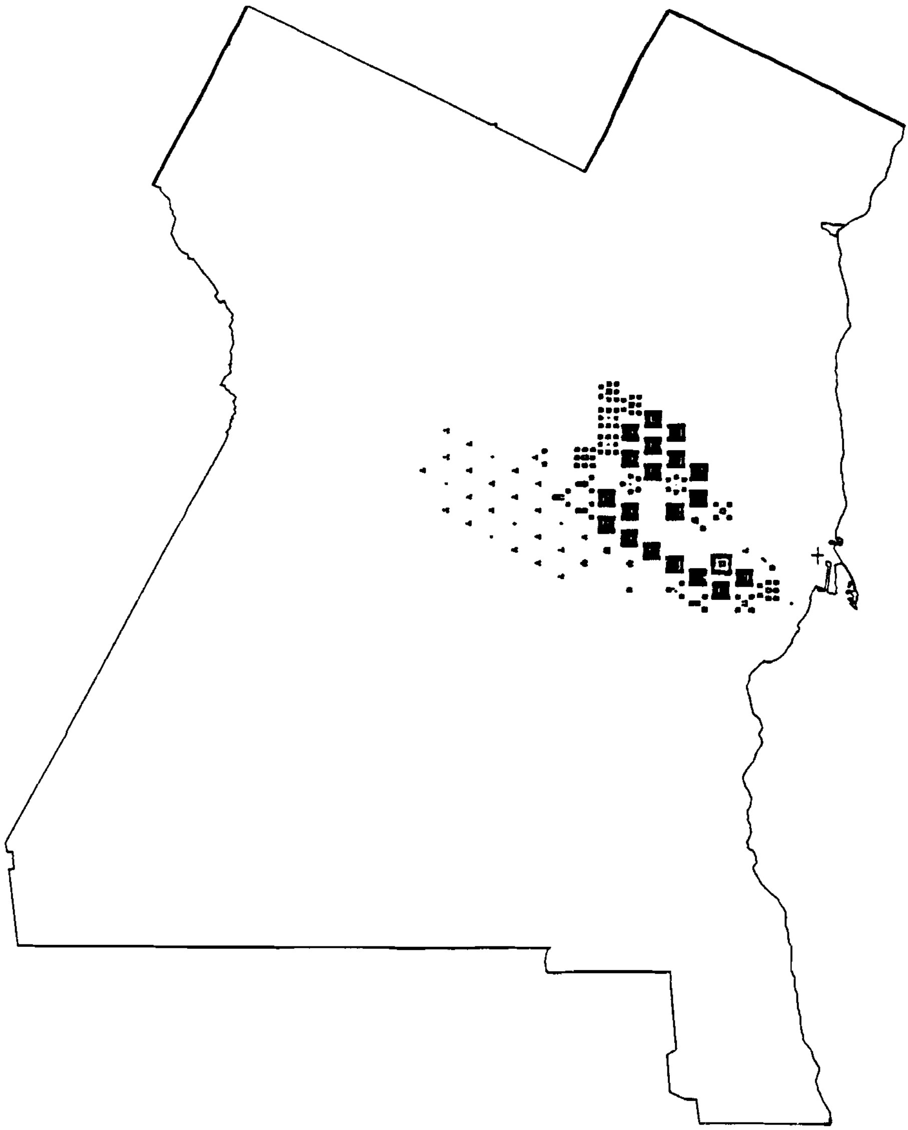


FIGURE 6. SURPLUS OF JEWISH POPULATION OVER TOTAL POPULATION, TORONTO, 1981

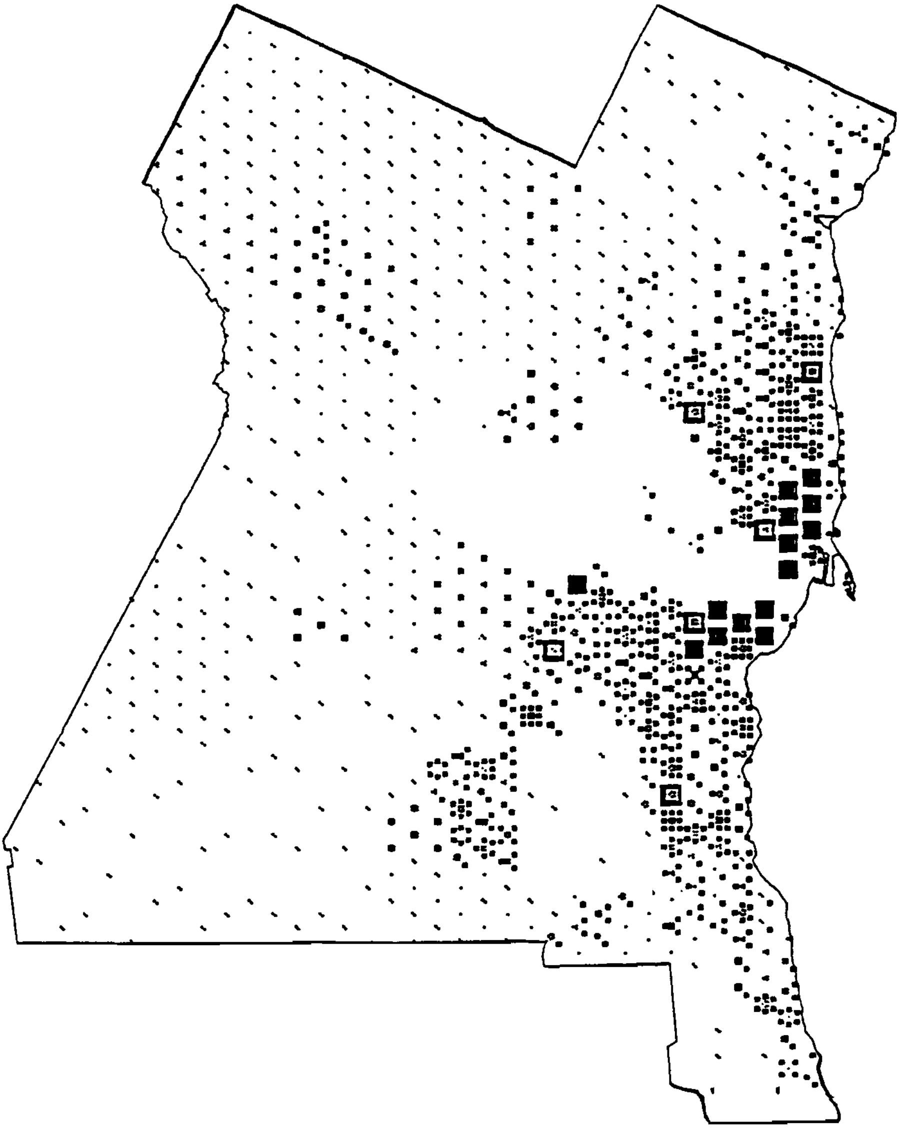


FIGURE 7. SURPLUS OF TOTAL POPULATION OVER JEWISH POPULATION, TORONTO, 1981

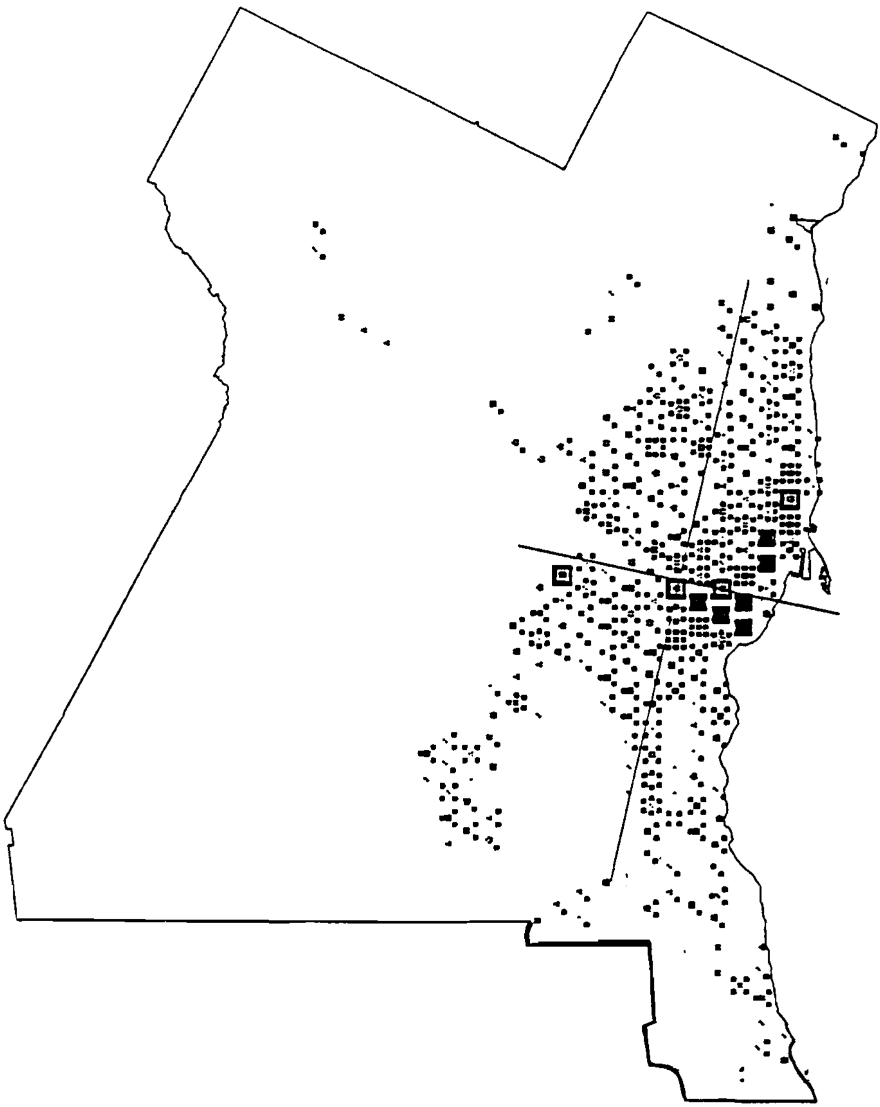


FIGURE 8. SURPLUS OF TOTAL POPULATION OVER UNIFORM DISTRIBUTION, TORONTO, 1981

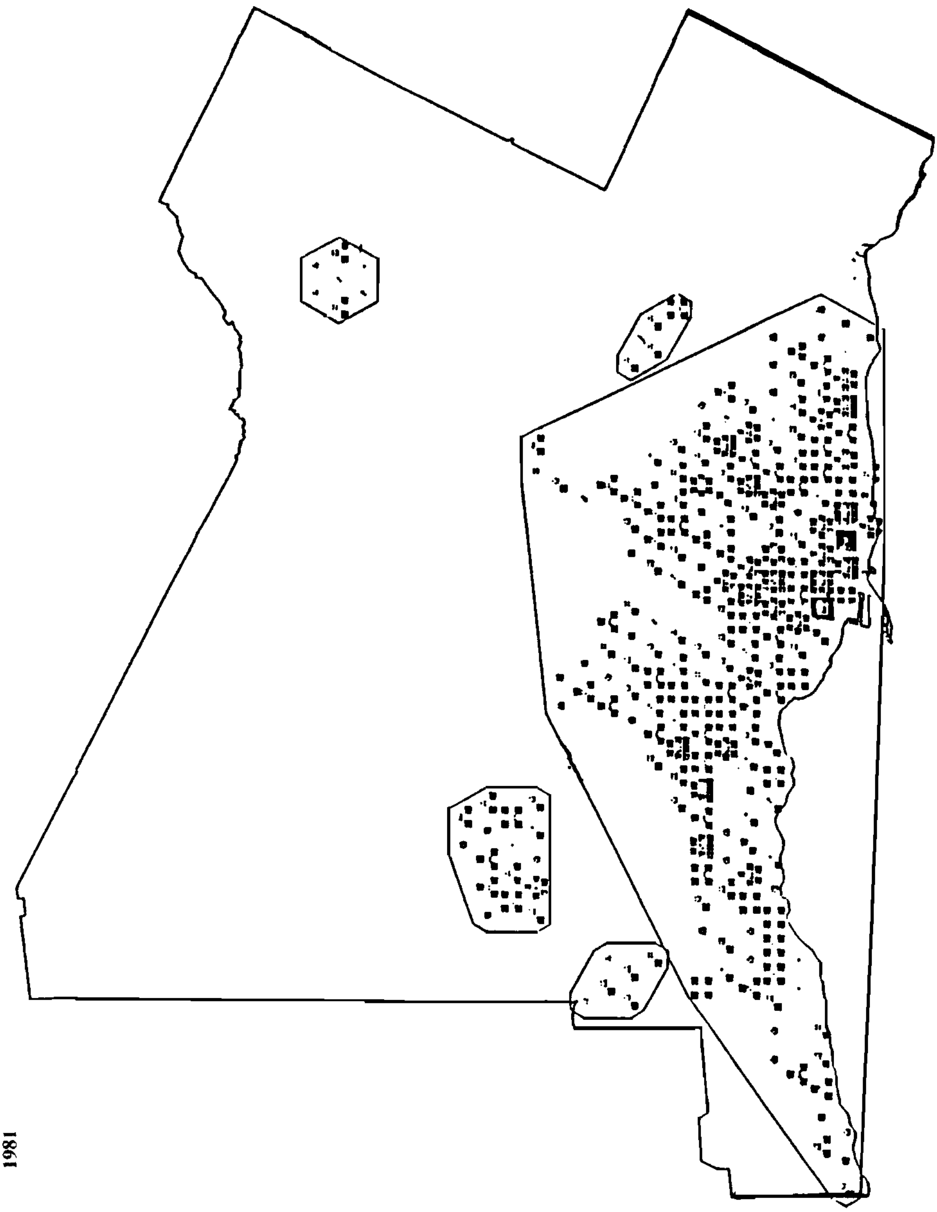


FIGURE 9. SURPLUSES OF UNITED CHURCH POPULATION OVER UNIFORM DISTRIBUTION INCLUDED IN MAJOR CLUSTERS, TORONTO, 1981

To simplify the presentation, we assume that the distributions studied have been transformed into equivalent distributions over hexagonal cells i. Let the proportion of population elements or territorial elements included in cell i out of the total be indicated by the relative weight w_i of cell i. ($\Sigma w_i = 1$.) This weight is attached to the center (X_i, Y_i) of cell i. In our discussion of geostatistical parameters, we shall start with the more commonly used measures:

(a) the *centers* which indicate the average location of the population or of the territorial elements considered;

(b) some parameters of *extent and directions of territorial dispersion* and;

(c) *dimensionless parameters of characteristics of territorial shape and patterns of population distributions.*

A summary parameter measuring *discrepancy or segregation between two distributions* is discussed separately.

Mean Centers

The mean center is the point from which the average of the squares of the distances from centers of cells i is minimal. The coordinates of the mean center $(\bar{X}, \bar{Y})$ are calculated simply as the weighted averages of the coordinates of the centers of the cells i:

$$\bar{X} = \Sigma w_i X_i; \qquad \bar{Y} = \Sigma w_i Y_i.$$

It has been found that centers of any two distributions (P,Q) and of their derived subsets SP, SQ, I are connected by the following general laws:

(a) the centers of SP, P, I are collinear;

(b) the centers of SQ, Q, I are collinear;

(c) indicating by C a distance between two centers, $C_{PQ} = \Delta\, C_{SP.SQ}$;

(d) $(C_{I.P})/(C_{I.SP}) = (C_{I.Q})/(C_{I.SQ}) = \Delta$;

(e) $(C_{P.SP})/(C_{I.SP}) = (C_{Q.SP})/(C_{I.SQ}) = 1 - \Delta$.

Figure 10 shows a magnified example of these relationships, for Jews and total population in Toronto. Here lines between centers of SP and SQ, P and Q, run parallel from west to east, with a slight shift to the north. The experience gained in geostatistical research in many cities and for different distributions over countries is that these lines connecting population centers are very useful for analytical and comparative purposes.

Extent and Direction of Absolute Territorial Dispersion

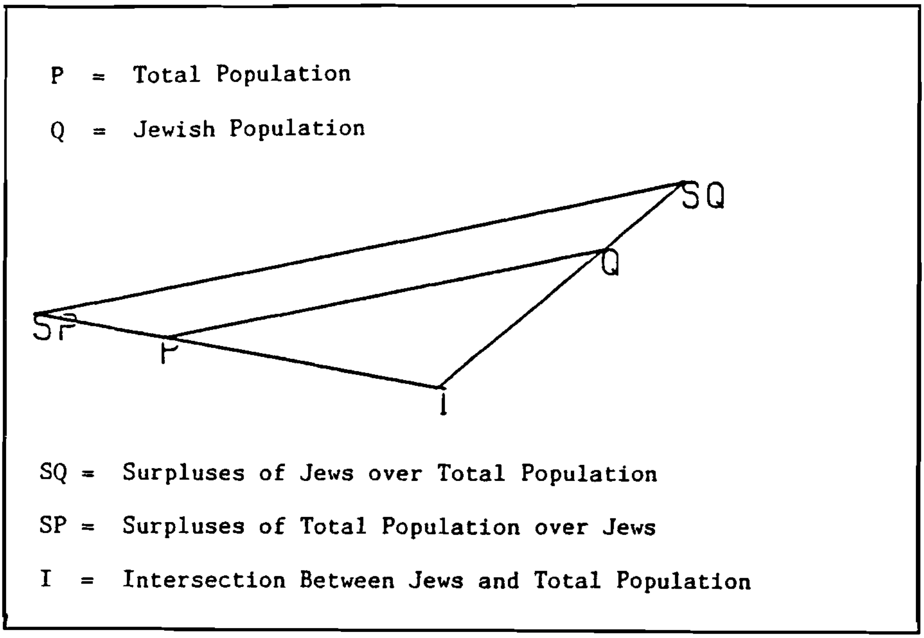
Among various parameters of dispersion suggested in the literature, the following have been mainly used in our project:

(a) To measure dispersion in the north-south, east-west directions, use of standard deviations σ_x, σ_y along conventional axes may be justified:

$$\sigma_x = \sqrt{\Sigma_i w_i (X_i - \bar{X})^2}; \qquad \sigma_y = \sqrt{\Sigma_i w_i (Y_i - \bar{Y})^2}.$$

(b) However, those standard deviations can be regarded as just one of an infinite

FIGURE 10. MAGNIFIED MAP OF CENTERS, TORONTO, 1981



set of pairs of standard deviations which could be calculated along orthogonal axes passing through the center.

Among these pairs of axes, the *principal axes* (M,m) are of special interest. Coordinates along those axes are uncorrelated, and one of these standard deviations (σ_m) is minimal, while the other (σ_M) is maximal in comparison to standard deviations calculated along any other axes through the center.

The identification of the major axis (M) is of peculiar descriptive importance because we can regard it as an indicator of the principal direction of the distribution. It is in fact the line from which the sum of squares of distances of the elements is minimal. Standard deviations along these axes measure separately territorial dispersion in two selected orthogonal directions. However, we may be interested in expressing territorial dispersion by means of one summary measure. Various solutions are available for solving this problem, such as the following:

(c) Let us indicate by:

$$d_i = \sqrt{(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2}$$

the distance between each element i of the set considered and the center. Two simple *linear* measures of dispersion can be obtained by taking the arithmetic average of the distances from the center ($d = \Sigma w_i d_i$) or the quadratic average of these distances: $d = \sqrt{w_i d_i^2}$. Here we shall use d, and term it *standard distance* (other names have also been suggested in the literature).

(d) Sometimes *areal* measures may be preferred, as we deal with dispersion over two-dimensional territory. A first solution is to take:

$$d^2 = \sigma x^2 + \sigma y^2 = \sigma_M^2 + \sigma_m^2$$

which can be termed as *distance variance.*

Another parameter of areal dispersion can be obtained by considering the area ($\sigma_M.\sigma_m$) of the rectangle having as sides respectively the largest and the shortest standard deviation (σ_M,σ_m). This product is known to be minimal in comparison to any other product between standard deviations along two orthogonal axes through the center.

Consider the relationship between d^2 and the double of this measure:

$$d^2 = 2\sigma_M\sigma_m + (\sigma_M - \sigma_m)^2$$

Generically, we call *oblongity* the property of a distribution of having a different dispersion along its principal axes. Oblongity can be measured in various ways. We consider here $(\sigma_M - \sigma_m)^2$ as a squared measure of absolute oblongity. We can take $2\sigma_M\sigma_m$ as a parameter of areal dispersion netted from effects of oblongity, while d^2 is a parameter of areal dispersion affected by oblongity. We call $s = 2\sigma_m\sigma_m$ by the arbitrary name of *scatter.* The ratio $d^2/2\sigma_x'\sigma_y'$ measures *relative oblongity.*

Center, angle between conventional axes and principal axes, and dispersion along principal axes can be represented graphically in a simple way by means of a cross, as shown in many of the maps given in this paper. The center of the cross is placed at the mean center of the set considered. Its hands indicate the principal axes; the length of the hands of the cross is respectively proportional to $2\sigma_M$ and $2\ \sigma_m$. Generally, the crosses display a good fit to the distribution represented.

The crosses help in the quick appraisal of important features of each distribution and in comparing the various maps. They also help to create a bridge between detailed analysis of maps and synthetical overview through use of parameters.

Parameters of Shape and Patterns Independent of Dimension and Position of Population Distribution

In our project, we compare characteristics of the distributions of the Jewish and other populations and of the territories in which these populations are found, in different cities and different times. For this purpose, parameters measuring the main characteristics of shape of territories and of patterns of population distributions are employed which are independent of position and dimensions of the cities and of the distributions studied. Detailed presentation of these parameters would require discussion of technicalities which cannot be given in this paper. Therefore, we limit ourselves to listing the main parameters employed.

With regard to characteristics of the shape of territories, *compactness* is regarded in the literature as very important. Among the many methods proposed to measure it, *relative distance variance* has been found to be the most suitable (Maceachren, 1985). This parameter is obtained by taking the ratio between (a) a measure of dispersion (the distance variance: see above) of the territory studied per unit of area, and (b) the minimal area of dispersion per unit of area, which is found for the most compact shape, the circle. This parameter – which actually measures the inverse of compactness, viz. relative dispersion – is constant for similar territories (having the same shape and different areas).

Relative distance variance can be split into two factors:

(a) relative oblongity (see above);

(b) looseness due to the structure of the territory. This can be measured by the *relative scatter* (ratio of scatter per unit of area to its minimum, found in ellipses). The relative scatter is constant for territories with same structure (viz. affinely equivalent).

These concepts and methods can be extended to the study of patterns of distribution of populations. This study can moreover be enhanced by other comparisons between geostatistical parameters of the population and those of the territory over which it is distributed.

Summary Measures of Segregation: Duncan's Δ and Minimal Distance

We have considered above many methods to compare the distribution of two populations (P,Q) over the same territory. This variety of approaches responds to the complexity and importance of the questions which can be asked in comparing distributions of P and Q. However, we may be interested as well in calculating synthetic measures of discrepancy between distributions of P and Q or *segregation* between P and Q.

A method often employed for measuring segregation is the index of dissimilarity (Δ) (Duncan and Duncan 1955; Duncan, Cuzzort and Duncan, 1961). Indicating by (f^P_j, f^Q_j) respectively the proportion of elements P and Q in each region j, $\Delta = 1/2|f^P_j - f^Q_j|$. The hypothesis behind the calculation is that if we wish to equalize the two regional distributions, Δ is the proportion of elements to be removed from the regions with $f^P_j > f^Q_j$ and to be transferred into the regions with $f^Q_j > f^P_j$. The index Δ has been criticized for various weaknesses and especially because (a) it is not sensitive to geographical location of the elements to be shifted and (b) it is variant with changes in the network of regions used (White, 1986).

To avoid these weaknesses, the *minimal distance* approach has been suggested (Bachi, 1963; 1981). Let us again suppose that distributions P and Q have the same size, that they are distributed over the same network of cells and that we wish to render the distribution P (origins) equal to that of Q (destinations). We suppose that each element of P is associated with an element of Q in such a way that the total of the distances between associated origins and destinations is minimal. For this purpose, the program utilizes a standard algorithm used for solution of the common transport problem (Syslow, Deo and Kowalik, 1983). Elements of P intersecting in the same cell i with elements of Q are considered to cover a distance 0. Therefore, the calculations are limited to the surpluses of P as origins and the surpluses of Q as destinations of movers.

The algorithm enables one to calculate the average distance $\bar{D}_s = \bar{D}_{sP,sQ}$ between associated origins and destinations of *moving elements*. As non-movers have a proportion $(1 - \Delta)$ and distance 0, the average distance $\bar{D}_{P,Q}$ between P and Q is obtained by:

$$\bar{D}_{P,Q} = (1 - \Delta)0 + \Delta\bar{D}_s = \Delta\bar{D}_s.$$

Minimal distance $\bar{D}_{P,Q}$ can thus be related to Δ. However, while Δ measures only the proportion of movers, $\bar{D}_s$ measures the average distance covered by them, and $\bar{D}_{P,Q}$ takes into consideration both aspects.

Dividing $\bar{D}_{P,Q}$ by a convenient measure of dispersion of P, Q or both, it is possible to obtain dimensionless values of distance between pairs of distributions. In our comparative study of Jewish populations, dimensionless measures of distances between Jews and total population in different cities and times were. These measures are found to be of considerable interest. $\bar{D}_{P,Q}$ has various properties (among which is being a metric, having $\bar{D}_{P,P} = \bar{D}_{Q,Q} = 0$; $\bar{D}_{P,Q} = \bar{D}_{Q,P}$; $\bar{D}_{P,R} + \bar{D}_{R,Q} \geq \bar{D}_{P,Q}$), but they cannot be discussed here. $\bar{D}^2_{P,Q}$ and $\bar{D}^2_S$ can be split into components indicating respectively the effects of:

(a) squared distance between the centers of the two sets;
(b) squared difference between standard distances; and
(c) residuals.

Effects of the position of P and Q and of their areas and oblongities over $\bar{D}_{P,Q}$ can also be ascertained.

Acknowledgements

This paper was revised after the completion of the project. The project was supported by the U.S.-Israel Bi-National Science Foundation. The principal investigators were R. Bachi, S. DellaPergola and V. Klaff. Thanks are given to S. Ever-Hadani for the complex programming work carried out for this project.

References

Bachi, R. (1963). "Standard Distance Measures for Spatial Analysis". *Regional Science Association Papers*, Vol. X. pp. 83–132.

Bachi, R. (1968). *Graphical Rational Patterns. A New Approach to Graphical Presentation of Statistics.* Israel University Press, Jerusalem and Press Transactions, Rutgers, The State University, Brunswick, N.J.

Bachi, R. (1973). "Geostatistical Analysis of Territories". *Bulletin of the International Statistical Institute*, Vol. XLV, Book 1. pp. 121–133.

Bachi, R. (1978). "Proposals for the Development of Selected Graphical Methods", in: *Graphical Presentation of Statistical Information.* U.S. Bureau of the Census, Washington. (Technical Paper). pp. 23–67.

Bachi, R. (1981). "Mapping the Main Characteristics of Distribution of Populations over Territories". Invited Paper. *Bulletin of the International Statistical Institute*, Vol. XLIX, Book 2. pp. 1003–1026.

Duncan, O.D. and Duncan, B. (1955). "A Methodological Analysis of Segregation Indexes". *American Sociological Review*, Vol. 20. pp. 210–217.

Duncan, O.D., Cuzzort, R.P., Duncan, B. (1961). *Statistical Geography. Problems in Analyzing Areal Data.* The Free Press, Glencoe, Ill.

Galvani, L. (1933). "Sulla determinazione del centro di gravita e del centro mediano di una popolazione". *Metron*, Vol. XI, no. 1. pp. 17–48.

Maceachren, A.M. (1985). "Compactness and Geographic Shape: Comparison and Evaluation Measures". *Geografiska Annaler*, Vol. 67B, no. 1. pp. 53–67.

Schmid, C.F. (1983). *Statistical Graphics*. Wiley, N.Y.

Syslow, M.M., Deo, N., Kowalik, J.S. (1983). "Transportation Problem". *Discrete Optimization Algorithms with Pascal Programs*. Prentice Hall, Englewood Cliffs. pp. 54–71.

White, M.J. (1986). "Segregation and Diversity Measures in Population Distribution". *Population Index*, Vol. 52, no. 2. pp. 198–221.