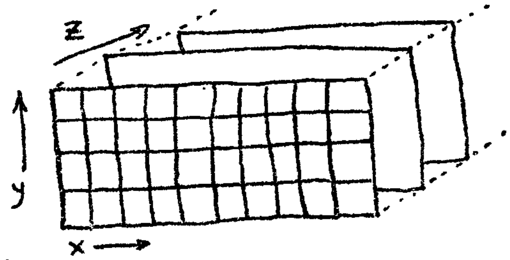
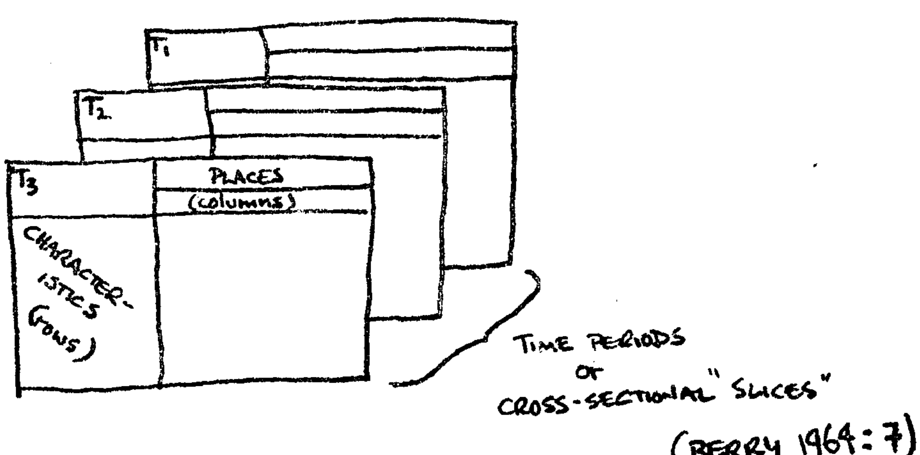
Methodological appendix: data structures and data analysis

The fundamental data structure to be employed in the proposed research consists of a temporally ordered set of planimetric maps of the region, which may be arrayed into a 3-dimensional matrix xyz:



The matrix has associated with it an array of characteristics or variables (defined by the theoretical concerns of the particular research problem), such that values on a variable represent events which can be uniquely specified by spatial (xy) and temporal (z) location. Using this framework, a given variable may be examined as an areal distribution of a class of events at a particular time (ie, a landscape), and as a distribution of that class over time (ie, a trajectory or 'timescape'); it is useful to visualize landscapes as analytical surfaces, and trajectories as changes in such surfaces.

Berry (1964) presents a generalized matrix formulation for geographical analysis in which rows represent characteristics (my 'associated array' of system-elements, above) and columns represent places (my x and y dimensions):



He then discusses eight basic and two derived "modes of geographical analysis":

> These ten modes fall into three series. The first includes (a) studies of the nature of single spatial distributions, (b) of the covariance of distributions at the same period of time or of the distribution of the same phenomenon at different periods of time, and (c) of the covariance of different distributions through time. A similar series of three levels characterizes the second series, which spans (a) locational inventories, (b) studies of areal differentiation and of sequent occupance, and (c) investigations of changing areal differentiation. The third series involves, at its simplest, (a) the cross-sectional interplay of studies of spatial distributions and associations, locational inventories and areal differentiation, and (b) at its more elaborate level the interplay of all nine of the earlier analytic modes.
>
> Berry (1964:7-8)

Operations upon this matrix thus provide a means to fulfilment of "the object of regional geography (which) is to find the essential characteristics of a particular region --its 'regional character' based upon the localized associations of variables in place-- by examining a wide range of variables over a limited number of places" (Berry 1964:9).

Neft (1966) provides a comprehensive guide to elementary statistics for areal distributions, basing his discussion on the concept of areal moments, which are average measures on areal frequency distributions: "the calculation of a moment distribution transforms a microgeographic collection of individual items into a macrogeographic, areally continuous variable" (Neft 1966:20). Thus, one can find the arithmetic mean center (S_c) of an areal distribution --its center of gravity-- by locating the point at which the value of the 2nd areal moment is at a minimum. Comparison of S_c at a number of points in time will thus indicate the general trends in the pattern of an areal distribution, and calculation of corresponding standard distance deviations (S_r) (where the square root of the second areal moment is at a minimum) will provide an index of the expansion or contraction of the distribution over time. Plotting of S_c at five-year intervals for any areal distribution will indicate the direction and rates of general trends; applied to the proposed research, it is anticipated that trajectories of S_c, when plotted for such variables as land value, population density, and retail sales, will provide aquantitative measures of the increasing centrality of Kentville.

A related set of statistics includes the harmonic mean center (H_c) and harmonic mean distance deviation (H_r); H_c is the point at which the value of the inverse of the first areal moment is at a minimum, and corresponds to the peak of the potential surface of the distribution, and H_r provides a measure of dispersion around H_c. Potential incorporates the concepts of mass (eg, frequency at a point) and distance, and has been widely applied in the study of commodity marketing and prices (Warntz 1959) and population (Stewart 1948, cited by Warntz):

> Potential at a point may be thought of as a measure of the proximity of that point to all other places in the system, or as a measure of aggregate accessibility of the point to all the other points in a region.
>
> Abler et al (1971:216)

Thus, potentials calculated for a set of points define a potential surface which changes over time as the values of the points change. In the proposed research, it is anticipated that Kentville will occupy H_c for many variables (eg, population, real estate value, retail sales), and that a decline in the value of H_r during the period 1946-1971 will further indicate the increasing primacy of Kentville.

In addition to the elementary summary statistics above, geographers have made use of trend-surface mapping. This procedure is analogous to regression analysis in linear statistics, and involves the fitting of a plane (or quadratic or higher-order surface) to an areal distribution. Mapping of residuals from a given trend-surface identifies both the location and intensity of deviations (positive and negative) from the overall trend. An extended discussion of methods and applications is given in Chorley and Haggett 1965, who note that

> trend-surface mapping represents an attempt to build up some generalized picture of areal variability in order to test some process-response model, in which an attempt is made to explain distributions in terms of sets of process factors...trend-surface analysis has already been used as an adjunct to forms of areal multivariate analysis (ie, multiple regression) in attempts to apportion out 'explanation for residuals' and to explain the regional variability of one areal feature by means of corresponding areal variations in assumed controlling factors.
>
> Chorley and Haggett (1965:196,212)

As Berry (1964) suggests, areal distributions of variables can be compared with one another by mapping of correlations between surfaces and residuals from regression; such maps

> portray the undulations of statistical surface(s) at some level of detail by delineating irregularities of various magnitudes consisting of elevations, depressions, and gradients, uniquely organized and variously positioned and oriented. Whether or not the phenomena are in any way causally interconnected, the two surfaces necessarily bear some relationship to one another; the recognition, description, and analysis of this are central aims in geographical research.
>
> Robinson (1962:414)

Further discussion of methods for regression analysis is found in Thomas (1968) and Thomas and Anderson (1965).

All of the above statistics can be easily performed upon a data matrix of the form of the xyz space-time coordinate system discussed above; much of the time in the field will be spent in mapping relevant variables into this data matrix, which can be stored and examined using computer facilities at Acadia University, in Wolfville. Fundamental to the coding operation is the 1:50,000 topographical map series issued by the Canadian Department of Mines and Technical Surveys, which is provided with a 1000-metre grid. It is assumed that this graticule-size will yield sufficient detail for study of phenomena of regional significance: the Annapolis Valley is contained within a 50x140 matrix of 1-km^2 (ie, 100 hectares or 247 acres) cells. Of the 7000 potential data-points thus defined, some 60% will be empty for most purposes because they are over water or undeveloped forest. Accordingly, data on the region can be contained on one or more decks of some 2800 cards (ie, one card for each 1-km^2 cell), where each column contains the value of a single variable at a single point in time (the coding task can be further reduced by entering only non-zero values). Data coded in thsi way may be thought of as providing an ordered 25-year series of 'snapshots' of various aspects of the region's development.

The 1:50,000 base map shows all roads, buildings, orchards and wooded areas. Specialized maps (eg, land use, soils, hydrology as discussed above) are available at less-detailed scales, and relevant information can be read from these maps with the use of 1-km^2 grids aligned with those of the 1:50,000 map. Acetate overlays will be used to plot distributions extracted from comparison of aerial photographs (land use, building activity, vegetation).

Aerial photographic coverage of the entire region is available for several points in time, and provides an extremely important potential data source. A conventional aerial photograph records everything which reflects light within the 0.3- 0.7 micron band of the spectrum (within the limitations of lens and film resolution, both of which exceed that of the eye), and other remote sensor systems (infrared, magnetometry, radar, etc) can further extend this range of potentially continuously available data. The practical problems of interpreting and utilizing the resources of aerial photography have rarely been discussed in sources readily accessible to anthropologists, and few are aware of their potential; although archaeologists have long made considerable use of these resources for reconaisance surveys, it is only very recently (with interest in ecological and settlement-pattern approaches) that the patterns shown on aerial photographs have themselves been considered as data. The greatest development of aerial photographic analysis has taken place within geography and geology (and military science), and it is to this literature that one must turn for guides to the interpretation of aerial photographic data (eg, ASP 1960, Stone 1964, Estes 1966, Latham 1963, Hammond 1964). It is important to note that the efficient use of aerial photographs requires detailed ground-truth data, in the form of maps, areal statistics, and on-site groundchecks. Stone 1964 provides a comprehensive guide to interpretation which will be followed in this research. A photomosaic of the most recent complete coverage will be constructed (on special order from the Department of Energy, Mines, and Resources, at a cost of about $800) to serve as a basis for comparison with previous coverage, which will be ordered in the form of individual (and thus stereoscopically readable) frames.