"Sampling for a Study of the Population and Land Use of Detroit in 1880-1885" *

Olivier Zunz
Michigan Society of Fellows
The University of Michigan

William A. Ericson
Statistical Research Laboratory
The University of Michigan

Daniel J. Fox
Statistical Research Laboratory
The University of Michigan

CRSO Working Paper
124 December 1975
(revision of working
paper # 114)

Copies available through:

The Center for Research
on Social Organization
The University of Michigan
330 Packard Street # 214
Ann Arbor, MI 48104

The design of an efficient sampling scheme for the study of the 19th century urban environment is a challenging problem for American urban history.[1] The historian faces three interesting difficulties. The first is a dilemma shared by most students of geographic variations in social life within large cities. In order to examine the relationship of social life to the general form of the city, a researcher needs observations spread through the urban territory. At this scale, he must ordinarily sacrifice detail to achieve uniform coverage. But to examine the constraints and the routine which are part of everyday experience, he needs close, continuous observations of small populations and small areas. At this scale he must ordinarily sacrifice the attempt to achieve uniform coverage of the city as a whole. Thus the historical study of the American city has often followed two distinct lines of approach: On the one hand, gross patterns of change in urban land use have been investigated to understand aspects of city change such as the dynamics of city growth and the development of suburbanization.[2] On the other hand, intensive studies of the experience of entire neighborhoods or single ethnic or social groups have been conducted.[3]

The second difficulty consists of the many uncertainties in dividing the 19th century city into a set of coherent physical and social areas to be investigated. Our knowledge of the 19th century urban structure is limited. Recent important demographic and socio-economic studies have been conducted for entire cities, rarely taking into account differences among geographic areas and the diversity of the urban territory.[4] The characteristics of archetypal types of urban areas, such as the slum, the ethnic neighborhood, the zone of emergence, the suburb are well known.[5] Yet their interpenetration and their

organization in relation to other areas of the city are still to be explored. Therefore the historian cannot easily use a research program comparable to that of Park and Burgess who divided the Chicago of the twenties into a plausible set of social areas in order to make intensive observations in each.[6]

The third problem arises from the fact that the historical study of the relationship between social divisions and spatial arrangements requires the integration of very different types of information on both the population characteristics and the city's physical structure. Many different levels of information coming from sources usually kept separated have to be collected and organized hierarchically. These are the data on individual city dwellers, the families and households, the ethnic and social groups on the one hand; the dwellings and houses, the streets and the neighborhoods, the larger districts of the city, on the other.[7]

In an attempt to resolve these three difficulties, a sample of 127 areal units was drawn from the entire city of Detroit in 1880. The design permits the historian to study the complex interplay between population characteristics and urban land use patterns at various geographic levels. The sample areal units are representative of the city's micro environment. In them, one can observe the demographic, social, ethnic structure of the population at the level of local life in relation to the urban fabric of the many neighborhoods of the city. Taken together, they provide a meaningful description of the overall urban territory. The combination of sample areas represents the city as a whole. Thus it becomes possible to measure the diversity of local life in the light of a general picture of the city.

Definition of a Sampling Unit

So many complex and often hidden patterns develop in a multi-ethnic city of immigrants that only a fine grain analysis permits one to catch overlapping phenomena, with different boundaries. One of the main difficulties of a geographic sampling procedure is to define a flexible unit of analysis, small enough to permit individual level observations of people and large enough to capture ethnic or socio-economic clustering. However, the grid plan of the American city is a natural sampling frame which can be used to that end. The block front, representing one side of a four sided block, is its smallest geographic component. It is a very flexible unit, easily drawn on a map along one street from one corner of the block to the other. When several frontages are linked, one creates a reconstruction of several urban forms: a block, a street, a small neighborhood. The sampling unit used here consists of a cluster of six fronts, in other words of one block and two opposing fronts (see Fig. 1)

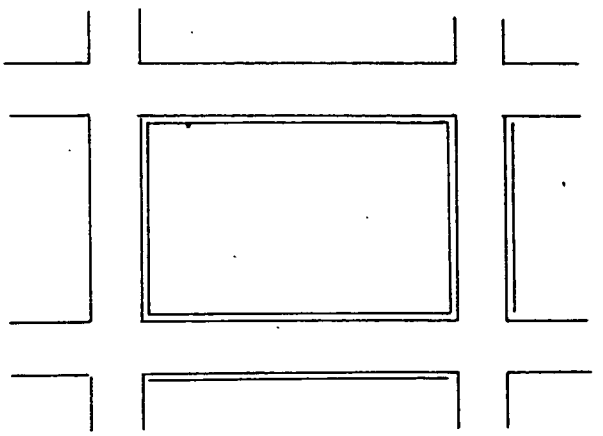


Fig. 1

The sampling unit can be divided into three independent units for purposes of analysis: the front, the block or the larger cluster. Having a multiple unit is necessary. Given the diversity of the urban environment, there is no predetermined best areal unit which would permit one to best study all types of inhabitants and of areas simultaneously. Students of neighborhood activities have long recognized that geographic boundaries vary with different phenomena.[8] Those defining an ethnic cluster are not the same as those of a given social class neighborhood, and in turn overlap with boundaries of non residential areas. Neighborhoods have loose and shifting boundaries which cannot be predetermined for sampling purposes. Rather than artificially delineating them, selecting a geographic unit well fitted to the grid plan increases the chances of measuring changing neighborhood activities.

Broad Areas and Micro Environment

The problem then is to select those units which best represent the variety of the urban environment: its rich structural fabric and its diverse inhabitants. The city is a composite of broad specialized zones, each of which characterized by a dominant type of land use such as residential, commercial, industrial or vacant.[9] However, an area characterized by one primary type of land use may have many other land uses. Predominantly commercial blocks have their residential fronts; residential districts have zones of craftmanship and retail businesses; the city center has its vacant lots. Main accesses and thoroughfares cut across the city and superimpose a city wide logic on local life. Therefore sampling small areas involves recognizing first large areas and then understanding the types of population and land use that interact within them. Unfortunately, no reliable prior

information on small scale patterns of land use and population was available. We therefore decided to investigate the sources available to identify broad zones and then to design the sample of clusters so that it is representative of those zones as well as of the city as a whole.

THE SOURCES

Several sources of information are required for our method which involves drawing a sample of clusters and then to take a census of the population and land usage of the six individual blockfronts of each cluster. First, one needs a detailed set of maps which identifies not only blockfronts but also their lots and buildings and where the houses of each front are numbered. Second, one needs demographic and socio-economic sources where the addresses of the listed individuals are recorded. These two sources have to be matched to draw the sample. In other words, all addresses for a front selected on the map have to be searched in the population listings. When these are available, a fairly accurate picture of the blockfront can be reconstituted: on the one hand, the lots and buildings, on the other, the population. The two basic bodies of data used for this purpose were the Federal Census Manuscript Schedules of 1880 and The Atlas of the City of Detroit published by E. Robinson and R.H. Pidgeon in 1885. The census enumeration of every individual living in the city of Detroit is a rich source for the study of the demographic, ethnic and social structure of the population.[10] The enumeration also lists all addresses. The real estate atlas which maps every block of the city permits one to study in detail the urban form and land use near the time of the census enumeration.[11]

Fronts, Blocks and Clusters in the Atlas

Twenty plates of the Robinson Atlas cover the city of Detroit. For each block, parcels and built structures are designated. Houses are numbered and the atlas contains many indications of non residential buildings. Many fronts in the atlas cannot be found in the census manuscript because they were not enumerated by the census takers. A frontage can be classified as "non enumerable" if it was impossible for a census taker to visit it. This was the case for either totally empty fronts, for fronts where houses face other streets and for fronts which are not residential.

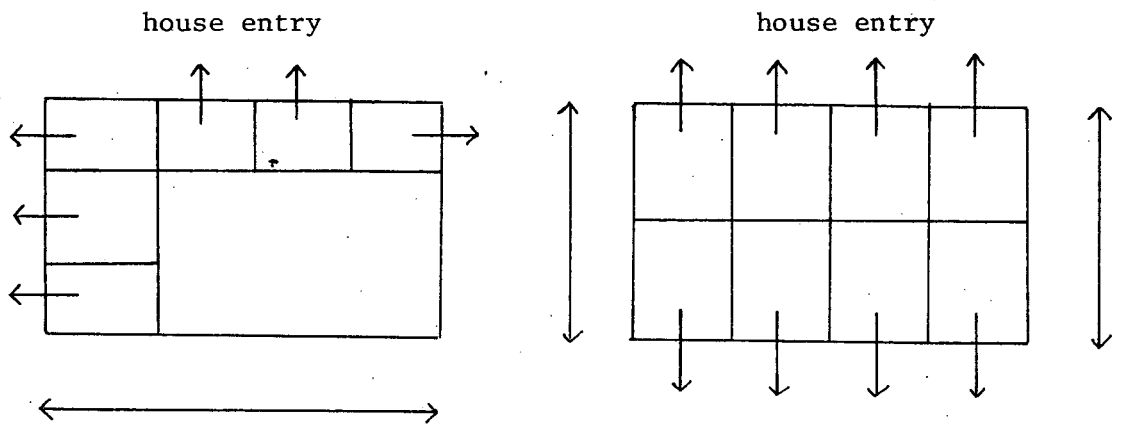
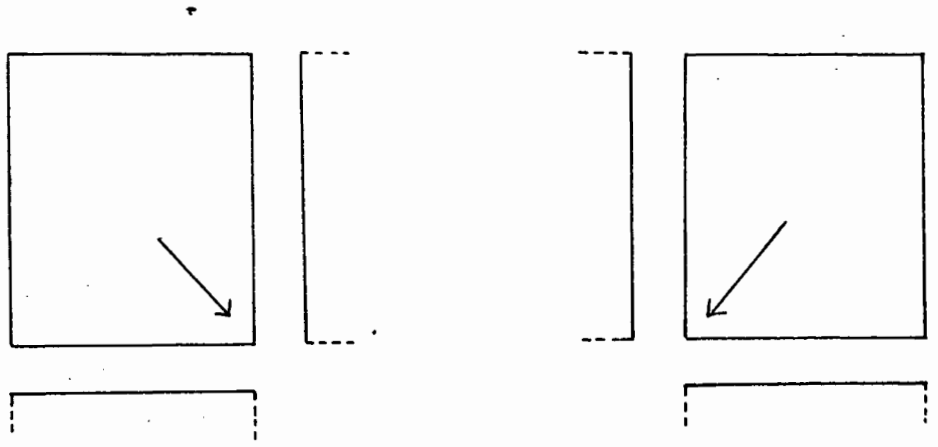


Fig. 2

Blocks and clusters are composed of enumerable and non enumerable frontages (Fig. 2). The ratio of enumerable to non enumerable frontages varies in the city depending on the location, the type of land use and the population density. Knowing this ratio is useful to define broad areas of the city. It is a crude but practical indicator of population

density. The very populated areas are likely to have substantially fewer non enumerable fronts per block than the scarcely populated areas. Population estimates can be computed for different sets of blocks in various areas of the city if one knows the number of inhabited (enumerable) frontages per block and the mean number of inhabitants per enumerable frontage. The atlas therefore contains information allowing us to divide the city into large geographic areas, not only in terms of land use but also of population density. Once these broad zones are defined, sampling clusters of six fronts within them may involve two steps: first randomly selecting blocks, and second randomly selecting a corner of each block to determine the opposing fronts.[12] (Fig. 3)



arrows indicate selected corners in the primary blocks to determine the opposing fronts

Fig. 3

The Fronts in the Census

When selecting clusters of block fronts, we cannot assume that the individual houses of each front would be found in the census manuscripts in the same geographic order as the atlas. We know that many census takers did not follow a strict route, but often shifted from even to odd numbers, or visited parts of streets at one time and revisited the other parts later. This implies that a thorough search in the census, of addresses selected from the atlas, is necessary to collect the information. All atlas addresses must be listed very carefully and every page of the manuscripts scrutinized until the search is exhausted. The Census Bureau's enumeration districts could sometimes be used to narrow down the search in the documents.

Discrepancies between Atlas and Census

Discrepancies always arise in the process of matching sources. In drawing a geographic sample of clusters of the 19th century city, we are using a census taken in 1880 and an atlas published in 1885. Discrepancies between the number of "residential buildings" per front in the atlas and the number of "houses" per front in the census schedules are likely to arise due to the different dates of each survey and underenumeration in the census. Yet this discrepancy can be computed and used as a correction factor for some statistical analysis.

These preliminary observations on the sources lead us to a few unsolved questions. How complete is the census of the population of the clusters reported in the census manuscripts? How do we draw the sample of clusters from the atlas and obtain a good representation of the population? How do we select the clusters so that they accurately

represent the various areas of the city? A pilot sample was drawn from each of the two sources in order to help answer these questions.

THE PILOT SAMPLE

Practically, the pilot sample is designed to give information useful in determining the total number of clusters to sample in different areas of the city, while keeping the search in the census manuscripts to a manageable task. Understandably, there was no systematic prior information available on the demographic, occupational, ethnic and land use composition of fronts, blocks or clusters nor on the geographic differences of these variables across the city. The pilot sample could not be conceived to give prior information on all variables of interest. It was assumed that urban density is the most useful parameter to predict the variability of major population characteristics. Thus the pilot sample was intended to gain information on population and building density across the city. Two independent systematic samples were drawn, one comprising every fourth block on each of the twenty atlas plates and the other a sample of street "runs" from the census manuscripts.

Pilot Sample of Blocks

The atlas contained sufficient detail so that each block could be categorized into one of four classes.

P = Promising (Blocks containing almost all dwelling structures; at least 3/4 of the buildings of the blocks being of the residential type for the downtown area: plates 1, 2, 6, 11; 2/3 elsewhere)

ND = Non-Dwelling (Blocks mainly occupied by non residential establishments (industrial, commercial, etc.): at least 3/4 downtown and at least 2/3 elsewhere).

V = Vacant (Blocks containing no structures: 90 % or more vacant)

O = Other (All other blocks)

The results of this categorization are shown in Table 1.

INSERT TABLE 1

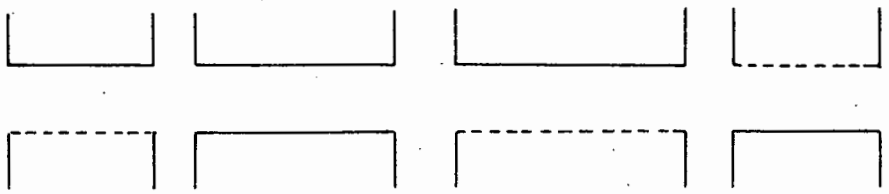
To determine the pilot sample of blocks, every block was numbered; then every fourth block on each of the twenty plates was sampled (even numbered plates: every other even numbered block; odd numbered plates: every other odd numbered blocks). Then for each sampled block the number of non enumerable frontages was determined. The number of non enumerable frontages was subtracted from the total number of frontages for each sample block to yield the number of enumerable frontages per block. The mean numbers of enumerable frontages per type of block are shown in Table 2.

INSERT TABLE 2

Census Pilot Sample

The census pilot sample gave us some idea of the size of frontages (e.g. numbers of households and persons per frontage) and of the problems of tracing frontages in the census data. Information was recorded for every 80th street in the census manuscript or more exactly for every 80th "run" under the same street name in the manuscript. For each run, we recorded the microfilm reel number, census page, street name, house numbers, number of household and number of people in each household. Every street or "run" recorded in this manner was then traced in the atlas. Using this procedure, we were able to trace 87 frontages

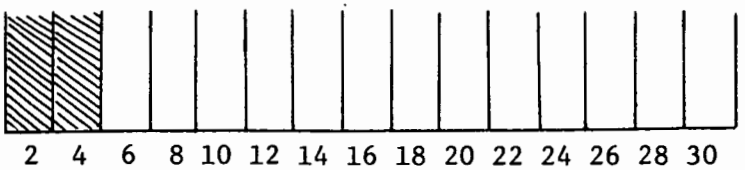
from the census in the atlas. In some cases, under the same street name, the census takers had crossed the street many times, mixing odd and even numbers. For some runs, we actually got several fronts of the atlas (Fig. 4).



- - -: fronts enumerated in one run

Fig. 4

In other cases, under one street name, there were only one or two houses for a given front instead of the 15 or 20 existing in the atlas (Fig. 5).



: house enumerated

Fig. 5

At this point, it was obvious that the census enumeration was even more disorganized than we had expected. It was very rare that a front was fully listed in one run of the manuscript. We knew then that a fair amount of work would be necessary to reorder the manuscript census in order to draw a geographic sample. Yet we had collected enough prior information to estimate gross land use and population characteristics.

ANALYSIS OF PILOT SAMPLE DATA

Population Density in the Census and in the Atlas

It was expected that the most populated fronts of the Atlas (in Promising and Other type blocks) would also be densely populated in the census enumeration; similarly, the least populated fronts of the Atlas (in Non-Dwelling and Vacant type blocks) would be unenumerated in the census. The confirming figures are displayed in Table 3 which gives the Atlas plate number, the street name, the stratum characterizing the sampled frontage, the number of frontages actually sampled in each run and the total number of persons found in the census for those frontages. As expected, the census provided little information about blocks in the Other category and almost none for the Vacant and Non-Dwelling blocks or for plates 5, 14, 17 and 19. Thus, on the basis of the two pilot samples, it was clear that very little of the population was contained in blocks categorized as Vacant or Non Dwelling.

INSERT TABLE 3

Estimating the total population

In order to insure the quality of our pilot matching of the census manuscripts with the Robinson Atlas, we tried to predict the total census population, known to be 116,340,[13] from the data collected from the P and O blocks. We first combined the data from the fronts in P and O type blocks.[14] We then computed the mean number of persons per frontage per street run for each plate. This meant that the data shown in Table 3 were reduced to 33 "independent" frontage observations. The result of this condensation is shown in Table 4.

INSERT TABLE 4

One simple estimate of the total population is given by:

$$\hat{P} = \sum_{i=1}^{20} N_i \bar{X}_i \bar{f}_i , \qquad (1)$$

where N_i is the number of P or O blocks on the ith plate, $\bar{X}_i$ is the mean number of persons per enumerable frontage and $\bar{f}_i$ is the mean number of enumerable frontages per block. For many of the plates, there were few observations on the mean number of persons per frontage and we decided to combine similar plates. This was done on the basis of the types of blocks per plate (see Table 1). Formula (1) was then applied to the eight resulting combinations of plates. The data are shown in Table 5. These combinations lead to 83,088 as an estimate of the total population size, an underestimate by some 33,252.

INSERT TABLE 5

In order to obtain some idea of the sampling error, the following formula was used in computing an estimate of the variance of $\hat{P}$

$$\hat{V}(\hat{P}) = \sum_{i=1}^{8} N_i^2 (\bar{f}_i^2 \, s^2_{\bar{X}_i} + \bar{X}_i^2 s^2_{\bar{f}_i} + s^2_{\bar{X}_i} s^2_{\bar{f}_i}) \quad (2)$$

[15]

where i refers to the ith of the eight plate groups shown in Table 5, $s^2_{\bar{X}_i}$ is the estimated variance of the mean number of persons per frontage, $\bar{X}_i$, and $s^2_{\bar{f}_i}$ is the variance of the mean number of enumerable frontages per block, $\bar{f}_i$. The variances used in determining this estimated value $\hat{V}(\hat{P})$ are shown in Table 6.[16]

INSERT TABLE 6

The application of formula (2) to the values shown in Tables 5 and 6 yielded an estimated standard error of $\hat{P}$ as 10,409. Thus the total population estimate 83,088 was 3.19 estimated standard errors below the true population value of 116,340. This estimate, though quite low, was nevertheless understandable. The factors contributing to it include our exclusion of the ND and V blocks and the small sample used in estimating or computing $\bar{X}_i$ (ranging between 3 and 6). Furthermore the systematic sample of streets in the census was misleading. Many streets appeared to be poorly recorded because information had been located at only one place in the manuscript census when, due to revisitation, the same fronts were very often recorded in several different parts of the census volumes.

Another estimate of 107,782 for the population total or an estimate of .925 (= $\frac{107,782}{116,340}$) was made when considering only the well enumerated fronts in our pilot sample. The basic scheme was to divide the pilot sample into two parts: the first called minor, was where census and atlas were poorly matched, i.e. when the census data contained three or less houses in a front that the atlas showed to have 4 or more residential buildings; the second called major, was where census and atlas were well matched, in other words where most houses on a front had been enumerated in the census on one visit by the census taker. Then formula (1) was applied to the twenty plates using only the data from the major group and estimating the mean number of people per front for the plates without census data.[17]

Geographic Variation of Population Density

At last we grouped the 20 plates into three classes according to population density, on the basis of the estimated number of people per front in the P and O blocks and of the percentages of V and ND type blocks per plate. The first group comprised plates 5, 10, 17, 18, 19 and 20. These seemed to represent sparsely populated parts of the city - each had at least 36 % Vacant or Non Dwelling unit blocks (see Table 1). The next group comprised plates 1, 3, 4, 8, 15 and 16. These plates seemed, on the other hand, to represent the most densely populated areas of the city. Finally the last group consisted of the remaining plates 2, 6, 7, 9, 11, 12, 13 and 14. Some details of the computations leading to this grouping are shown in Table 7. Certain pilot sample summary statistics which characterize the 20 plates are shown in this table together with the classification into the three classes of plates according to population

density.

INSERT TABLE 7

Estimating the Sample Size

The pilot sample provided us with a good idea of the variation of both population and building density across the city. On this basis, it was decided that the final sampling would be more intensive in the class of plates representing the most populated areas of the city. To obtain geographic representation, the sample would be drawn stratifying by plate of the atlas. Independent random samples would be selected for each of the twenty plates. It was felt that drawing at least 100 clusters, with the primary block being classified either Promising or Other, would yield tolerably small standard errors for estimating residential and population characteristics. Such a number would also keep the search in the census manuscripts to a manageable task. In order to estimate other city characteristics, we decided that 25 clusters would be selected from among the Non Dwelling and Vacant blocks.

Given the above design, we estimated the size of the sample in terms of household and persons which would be produced. One difficulty was in estimating the number of enumerable frontages associated with each sampled cluster. The pilot sample provided only estimates by fronts and by blocks. Sampling clusters requires the augmentation of the block by two opposing frontages. Therefore the observed pilot sample number of enumerable frontages per block, $\bar{f}_i$, was adjusted upward in the three groups of plates to reflect the inclusion of the opposing frontages. These are the values $\bar{f}_i^*$ shown in Table 8. These values were quite subjective; in densely populated areas, like group 2, $\bar{f}_i^*$ might be as

large as one and a half times $\bar{f}_i$; for sparsely populated areas with large numbers of V and ND blocks, there might be little difference between $\bar{f}_i$ and $\bar{f}_i$*. The estimated sample number of persons is given using formula (1) replacing N_i by n_i, the sample number of blocks to be selected, and $\bar{f}_i$ by $\bar{f}_i$*. This was computed as 8,930 persons.[18] It was expected that there would be an average of 5 persons per household and thus we estimated that a sample of 100 clusters would contain some 1,786 households. However, even making the assumption that the $\bar{f}_i$* are known without error, the estimated standard error of the estimate of the sampled number of persons is 961.[19] The estimated standard error of the estimate of the sampled number of households is 192.

INSERT TABLE 8

THE FINAL SAMPLING DESIGN

In the end, 102 P and O blocks were selected with their opposing frontages. These included 12 blocks from among the 177 P and O blocks in category 1 - sparsely populated plates - or roughly 6.8 %; 40 of the 425 such blocks in category 2 - densely populated plates - or roughly 9.4 % and 50 of the 650 blocks in category 3 - remaining plates - or about 7.7 %. These ratios were applied separately to the P and O block categories. Also, 25 or about 8.7 % of the 288 ND and V blocks were sampled.

INSERT TABLE 9

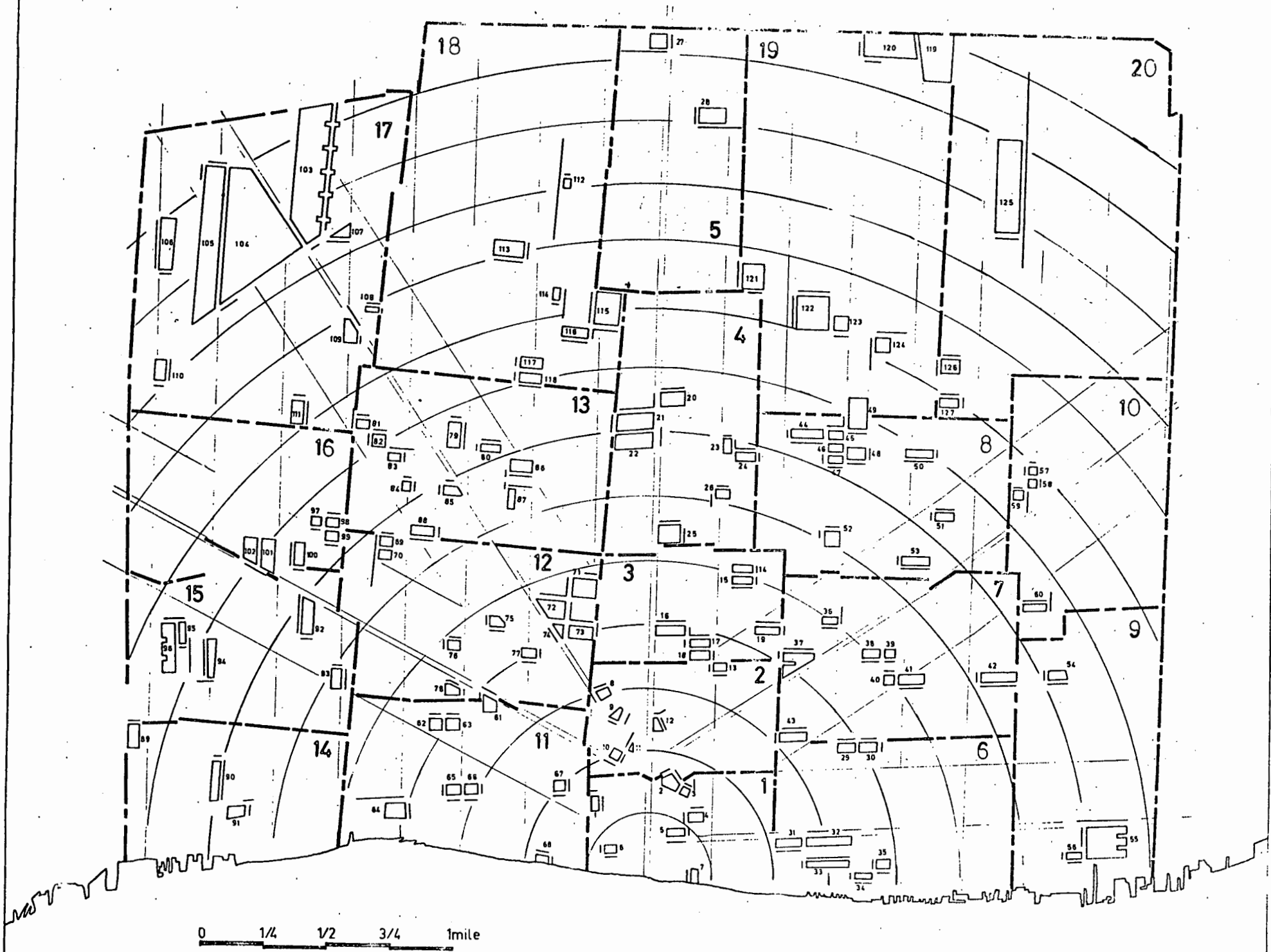
The actual mechanics of drawing the sample were quite straightforward given Table 9 and the tracings of the blocks as shown on the 20 plates

of the atlas. The blocks of each type were serially numbered on the plate and, using random numbers, a simple random sample without replacement of blocks was selected for each of the four block categories within each plate.

Once a block was selected and located on the atlas plate tracing, then the Rand Table of random digits was used to choose one of the, usually four, block corners. The usually two opposing frontages were then included to make up the cluster (see fig. # 3 and footnote # 12). The sampled clusters are displayed in the map below.

DETROIT
SAMPLE

From E. ROBINSON & R.H. PIDGEON: ATLAS OF THE CITY OF DETROIT 1885



THE SAMPLE

A Census of 127 Clusters

Based on an exhaustive search in the manuscript census, the final sample includes 12,185 people.[20] This comprises 2,410 households on 721 frontages or 127 clusters. 353 frontages are actually inhabited. The others are non residential or more generally non enumerable. The mean household size was 5.58, the mean number of people per front was 33.47 (min = 1; max = 152).[21] A full land use and population census has been taken of each unit. From the atlas and other sources, all the physical characteristics of each front have been recorded.[22]

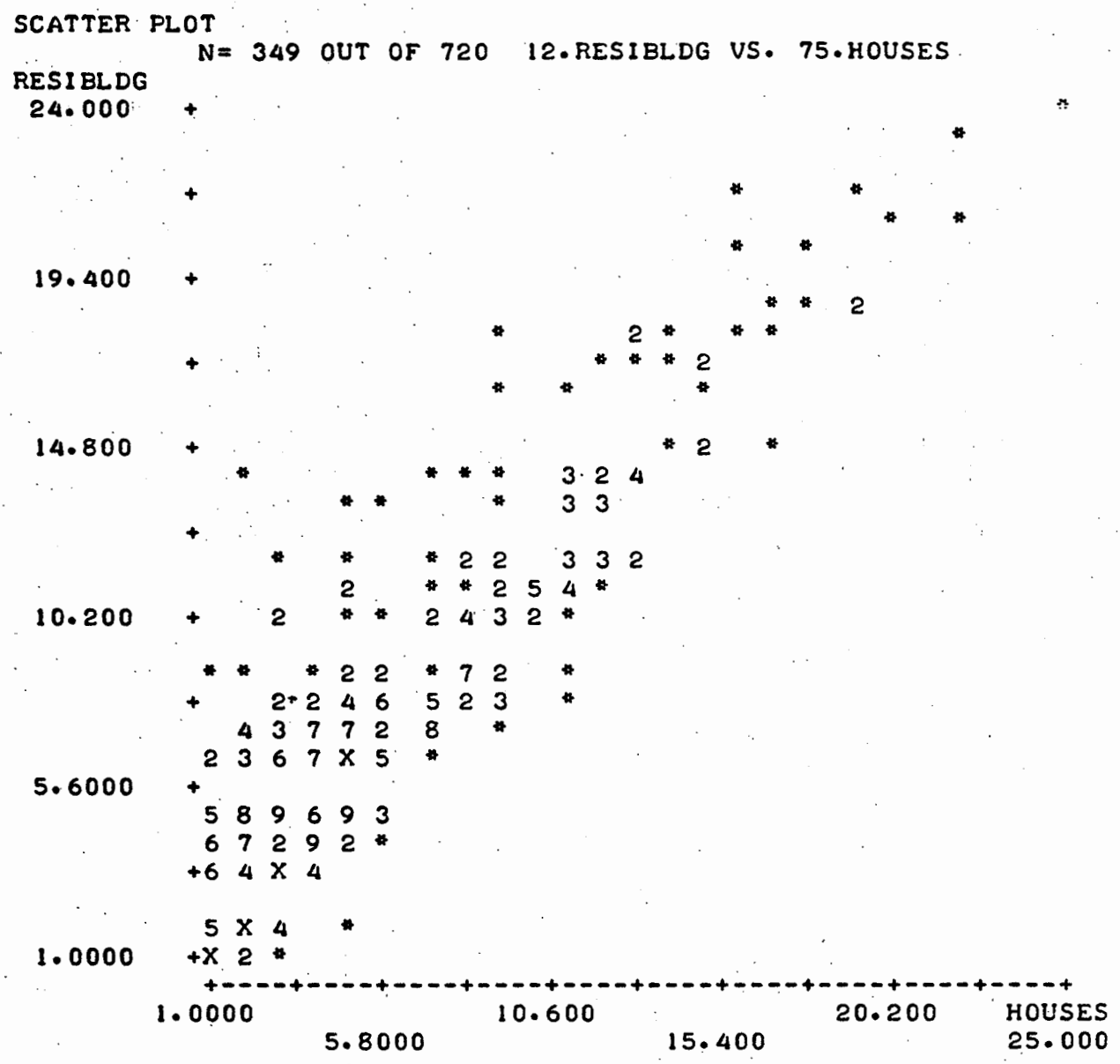
A detailed land use survey has been conducted to record the number and types of parcels, the number and types of buildings, (residential, non residential, small unnumbered buildings, stables), and the specific type of occupancy in each lot and built structure: types of residences, of manufacture, of craftmanship, of commerce, of business and professional services, of public and quasi public services; types of transportation, communication and utilities equipment; amount of unused space. Each area land use system is thus fully described and can be correlated with population characteristics.

From the census manuscripts, the demographic, occupational and ethnic characteristics of every individual living on the fronts have been coded. This total enumeration of the clusters is essential to study problems of density, household structure, family size, family composition and demographic behavior, as well as for the study of complex ethnic and social distribution patterns. The final sample includes a variety of data-sets, each corresponding to one level of data collection and/or analysis: the individual file of 12,185 people, the family

file (2,410 cases), the front file (720 cases), the block file (127 cases), the cluster file (127 cases).[23]

The Discrepancy Rate between the Sources

A simple measure of discrepancy between the two sources - the Atlas and the Census - is given by subtracting the number of houses found in the census from the number of residential buildings pictured in the atlas for each front. A scatter plot of the number of residential buildings per front against the number of houses in the census shows that the discrepancy rate is low and regular (Fig. 6). The discrepancy itself varies from -3 to 12 (Fig. 7). The mean discrepancy is only 1.62 houses per front. It is accounted for by houses built between 1880 and 1885 (e.g. houses shown in the atlas but not found in the census) and in some cases, by houses found in the census at addresses in between atlas addresses.



Number of Residential Buildings per Front in the Atlas against Number of Houses per Front in the Census Manuscript. There are 349 inhabited fronts (out of 364 listed in the census) with information in both sources.

Fig. 6

HISTOGRAM/FREQUENCIES

MIDPOINT	HIST%	COUNT	FOR 80.DISCREPA (EACH X= 3)
-3.0000	.6	2	+X
-2.0000	1.4	5	+XX
-1.0000	7.2	25	+XXXXXXXXX
0.	26.1	91	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
1.0000	19.8	69	+XXXXXXXXXXXXXXXXXXXXXX
2.0000	16.9	59	+XXXXXXXXXXXXXXXXXX
3.0000	12.3	43	+XXXXXXXXXXXXXX
4.0000	6.6	23	+XXXXXXXX
5.0000	4.6	16	+XXXXXX
6.0000	1.1	4	+XX
7.0000	2.0	7	+XXX
8.0000	.6	2	+X
9.0000	.6	2	+X
10.000	0.	0	+
11.000	0.	0	+
12.000	.3	1	+X
MISSING		371	
TOTAL		720	(INTERVAL WIDTH= 1.0000)

Discrepancy between the two sources: number of residential building per front in the atlas - number of houses per front in the census.
For example, there are 91 fronts or 26.1% of the inhabited fronts where the number of houses in the census is the same than the number of residential buildings in the atlas.

Fig. 7

The Standard Errors of Estimates

Some aspects of the representativeness of the sample can be judged by checking to see how known population characteristics (from published census figures) are estimated from our sample. Many of these characteristics are proportions. For example the proportions of inhabitants who were whites, the proportion born in Canada, the proportions of females in the working population. In each of these cases the estimated proportion is a ratio where both numerator and denominator are subject to sampling variability.

Suppose that the population proportion to be estimated from the sample is designated by

$$R = X/Y$$

where Y is the population number of persons and X is the number of those having the attribute. In such a case R is estimated by

$$r = x/y$$

where x and y are sample estimates of X and Y respectively.

For the sampling design used, x and y are given by

$$x = \sum_{h=1}^{H} w_h \sum_{\alpha=1}^{a_h} x_{h\alpha}$$

and

$$y = \sum_{h=1}^{H} w_h \sum_{\alpha=1}^{a_h} y_{h\alpha}$$

where H is the number of strata, a_h is the number of clusters in stratum h,[24] w_h is a weight associated with the h-th stratum and is proportioned to N_h/n_h, the ratio of the population number of cluster to the sample number of clusters in the h-th stratum, and $x_{h\alpha}$ and $y_{h\alpha}$ are the sample totals for the α-th cluster within the h-th stratum.[25]

Due to the stratification by plate of the atlas which permitted us to achieve geographic coverage, some plates have only one sampled cluster per stratum. In computing r, strata having one (or less) sampled clusters present no difficulty. However in estimating the variance of r (or its square root, the standard error of r), one needs two or more sample observations or cluster per stratum. In computing the standard error of r, denoted by SE(r), the 80 strata were reduced to 27 by combining similar strata so that each combined stratum had at least two sampled clusters.

In figure 8 we display the sample estimate, r, for each of ten population ratios. A range of one standard error above and below the sample estimate is also indicated.[26] Finally the true population value from the published census is indicated by an "X".[27]

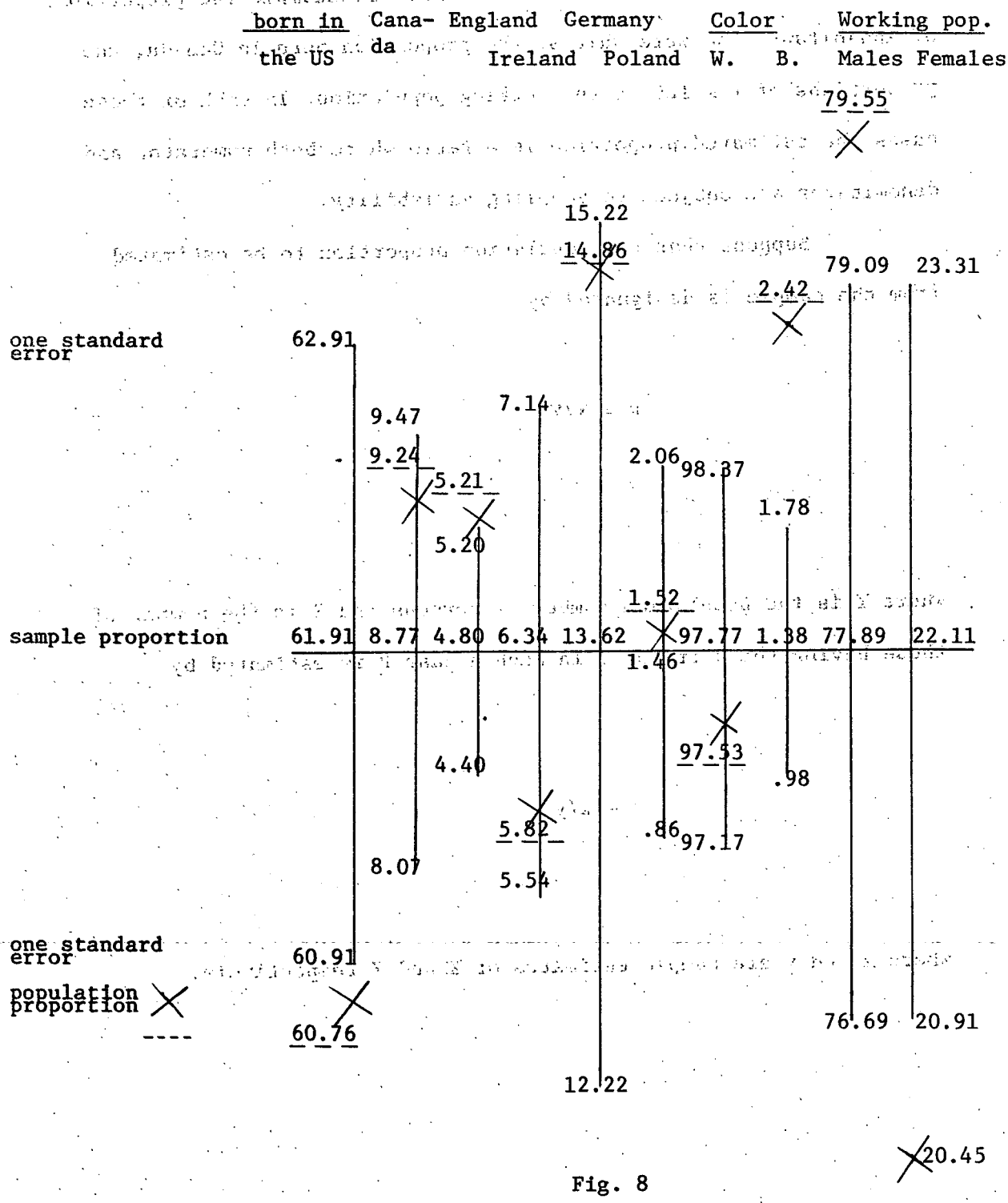


Fig. 8

CONCLUSION

This sampling procedure is only one way to sample data for the study of differential processes of settlement in a city. It is best to concentrate the analysis on the relationship between spatial and social organization in small areas.[28] The 1880 sample has been duplicated for 1900 and enlarged to include clusters within the expanded city limits. This permits us to study the evolution of the 1880 sampled clusters over time in relation to the change in the city structure, size and population. Other methods should be explored, with different emphasis, such that a variety of possible sampling schemes can be developed for the historical study of the urban environment. The success of quantitative urban history calls for the use of sampling methods which can cope with the rich texture of the urban phenomenon.

TABLE 1

Block Categorization by Atlas Plate

Number of Blocks

Plate #	Promising	Non-Dwelling	Vacant	Other	Total	Proportion V or ND
1	29	15	0	28	72	.21
2	63	10	0	9	82	.12
3	68	0	0	0	68	-0-
4	35	0	3	30	68	.04
5	5	0	11	12	28	.39
6	69	9	0	8	86	.10
7	94	3	0	12	109	.03
8	85	0	4	25	114	.04
9	28	4	3	13	48	.15
10	21	0	20	14	55	.36
11	73	11	2	14	100	.13
12	100	7	1	11	119	.07
13	86	1	6	29	122	.06
14	23	3	2	18	46	.11
15	42	3	1	18	64	.06
16	49	0	2	16	67	.03
17	13	1	59	30	103	.58
18	18	0	42	24	84	.50
19	6	0	41	24	71	.58
20	5	0	24	5	34	.71
Totals	912	67	221	340	1540	

TABLE 2

Pilot Block Sample

Mean Numbers of Enumerable Frontages Per Block

By Type of Block

(Sample Size in Parentheses)

Plate	Promising	Non-Dwelling	Vacant	Other	Total Sample Size
1	3.0000 (8)	4.0000 (1)	--- (0)	3.3333 (9)	18
2	3.4375 (16)	1.7500 (4)	--- (0)	4.0000 (1)	21
3	3.7059 (17)	--- (0)	--- (0)	--- (0)	17
4	3.0909 (11)	--- (0)	.5000 (2)	3.250 (4)	17
5	2.5000 (2)	--- (0)	1.0000 (2)	2.0000 (3)	7
6	3.3125 (16)	1.0000 (1)	--- (0)	2.0000 (4)	21
7	3.5000 (24)	--- (0)	--- (0)	2.5000 (4)	28
8	3.3750 (24)	--- (0)	1.5000 (2)	3.0000 (3)	29
9	3.1667 (6)	--- (0)	2.0000 (1)	3.0000 (5)	12
10	3.2857 (7)	--- (0)	1.0000 (4)	2.3333 (3)	14
11	3.1579 (19)	0.0000 (2)	1.5000 (2)	3.0000 (2)	25
12	3.2692 (26)	0.0000 (2)	0.0000 (1)	3.0000 (1)	30
13	3.2609 (23)	--- (0)	--- (0)	2.3750 (8)	31
14	3.0000 (4)	0.0000 (1)	0.0000 (1)	2.3333 (6)	12
15	2.9000 (10)	1.0000 (1)	--- (0)	2.6000 (5)	16
16	2.8000 (15)	--- (0)	--- (0)	2.5000 (2)	17
17	2.2500 (4)	--- (0)	.3333 (15)	2.1429 (7)	26
18	2.6667 (3)	--- (0)	.6364 (11)	2.1429 (7)	21
19	--- (0)	--- (0)	.9231 (13)	2.600 (5)	18
20	4.0000 (1)	--- (0)	1.1667 (6)	2.5000 (2)	9
					389

TABLE 3

Plate #	Street Name	Stratum	No. of Frontages	No. of People
1	Shelby	ND	1	6
1	Larned	P	6	386
2	Baubian	P	1	24
2	Clifford	P and O	5	56
3	High St. E.	P	4	134
3	Henry St.	P	1	10
4	Woodward	P and O	2	62
6	Hastings	P	1	3
6	Larned	P	5	120
6	Crogham	P	5	114
7	Dequinder	P	3	41
7	Prospect	P	4	49
7	Macomb	P	1	4
8	St. Antoine	P	5	64
8	Kentucky	P	1	32
8	Arndt	P	1	45
9	Adair	O	1	24
10	Arndt	P	3	68
11	Front	P	1	25
12	Locust	O	5	50
12	Huron	P	1	18
12	Wabash	P	3	16
13	Myrtle	P	1	6
13	23rd Ave.	P	5	78
13	Noble	P	3	35
15	Michigan	O	1	8

TABLE 3 (Cont.)

Plate #	Street Name	Stratum	No. of Frontages	No. of People
15	Wabash	P	1	21
16	Michigan	P	1	7
16	Sullivan	P	1	74
16	Wabash	P	5	50
18	Second Ave.	P	2	18
20	Chene	P	2	21

TABLE 4

Plate Number	No. of Effectively Indep. Frontages	# Persons/ Frontage
1	1	64.33
2	2	24.00, 11.20
3	2	33.50, 10.00
4	1	31.00
5	0	---
6	3	3.00, 24.00, 22.80
7	3	13.67, 12.25, 4.00
8	3	12.80, 32.00, 45.00
9	1	24.00
10	1	22.67
11	1	25.00
12	3	10.00, 18.00, 5.33
13	5	6.00, 15.60, 11.67, 8.00, 15.75
14	0	---
15	2	48.00, 21.00
16	3	7.00, 74.00, 10.00
17	0	---
18	1	9.00
19	0	---
20	1	10.50

TABLE 5

Plate numbers	# of P & O Blocks	$\bar{X}_i$ mean # persons/frontage	$\bar{f}_i$ mean # of enumerable frontages	$N_i \bar{X}_i \bar{f}_i$
5,10, 17, 18, 19, 20	177	14.00	2.500	6,195
1, 9, 14, 15	199	39.25	2.9434	22,990
3, 4, 16	198	27.50	3.2041	17,446
2, 6, 11	236	18.33	3.2069	13,873
7	106	10.00	3.3571	3,559
8	110	30.00	3.3333	11,000
12	111	11.00	3.2593	3,980
13	115	11.60	3.0323	4,045
	1252			83,088

TABLE 6

Plate numbers	$s^2_{\bar{x}_i}$	$s^2_{\bar{f}_i}$
5, 10, 17, 18, 19, 20	20.33	.01955
1, 9, 14, 15	104.56	.01119
3, 4, 16	107.92	.01273
2, 6, 11	13.98	.01498
7	9.33	.01644
8	86.33	.01424
12	14.33	.02733
13	4.16	.02900

Table 7

Plate #	Average # Persons per Frontage	#P or 0 Blocks	Proportion ND or V Blocks	No. Enum. Frontages per Block	Estimate of Population Size	Estimate of Population per Block	Block Categorization
1	64.33	57	.21	3.1765	11,648	204.3	2
2	17.60	72	.12	3.4706	4,398	61.1	3
3	21.75	68	.00	3.7059	5,481	80.6	2
4	31.00	65	.04	3.1333	6,314	97.1	2
5	--	17	.39	2.2000	--	--	1
6	16.60	77	.10	3.0500	3,899	50.6	3
7	9.97	106	.03	3.3571	3,548	33.5	3
8	29.93	110	.04	3.3333	10.974	99.8	2
9	24.00	41	.15	3.0909	3,041	74.2	3
10	22.67	35	.36	3.0000	2,380	68.0	1
11	25.00	87	.13	3.1429	6,836	78.6	3
12	11.11	111	.07	3.2592	4,019	36.2	3
13	11.40	115	.06	3.0323	3,975	34.6	3
14	--	41	.11	2.6000	--	--	3
15	34.50	60	.06	2.8000	5,796	96.6	2
16	30.33	65	.03	2.7647	5,450	83.9	2
17	--	43	.58	2.1818	--	--	1
18	9.00	42	.50	2.3000	869	20.7	1
19	--	30	.58	2.6000	--	--	1
20	10.50	10	.71	3.0000	315	31.5	1

TABLE 8

Category	Plate Group	Pilot Sample Mean # persons/front, $\bar{X}_i$	Sample Size P and O Blocks, n_i	Pilot Sample # enum. front/block, $\bar{f}_i$	f_i^*
1	5, 10, 17-20	14.0567	12	2.5000	2.75
2	1, 3, 4, 8, 15, 16	32.3858	40	3.1759	4.30
3	2, 6, 7, 9, 11-14	14.1261	50	3.1636	4.10

Table 9

SAMPLE DESIGN

	Population Size				Sample Size			
Plate #	P	O	ND	V	P	O	ND	V
5	5	12	0	11	0	1	0	1
10	21	14	0	20	1	1	0	2
17	13	30	1	59	1	2	0	6
18	18	24	0	42	1	2	0	4
19	6	24	0	41	0	2	0	4
20	5	5	0	24	1	0	0	2
Total Group 1	68	109	1	197	4	8	0	19
1	29	28	15	0	3	3	1	0
3	68	0	0	0	6	0	0	0
4	35	30	0	3	4	3	0	0
8	85	25	0	4	8	2	0	0
15	42	18	3	1	4	1	0	0
16	49	16	0	2	5	1	0	0
Total Group 2	308	117	18	10	30	10	1	0
2	63	9	10	0	4	1	1	0
6	69	8	9	0	5	1	1	0
7	94	12	3	0	7	1	0	0
9	28	13	4	3	2	1	0	0
11	73	14	11	2	6	1	1	0
12	100	11	7	1	8	1	1	0
13	86	29	1	6	7	2	0	1
14	23	18	3	2	2	1	0	0
Total Group 3	536	114	48	14	41	9	4	1
Grand Total	912	340	67	221	75	27	5	20

* The authors wish to thank Karol P. Krótki for his assistance in computing the standard errors of estimate and Charles Tilly for his comments on the manuscript. The design of the sample and the data collection have been made possible by grants from the Michigan Society of Fellows and of the Population Development Fund of the Ford Foundation.

1 - In addition to W.G. Cochran, Sampling Techniques New York, 1963, the historian may wish to consult R.S. Schofield "Sampling in Historical Research" in E.A. Wrigley, Nineteenth Century Society Cambridge, 1972: 146-190.

2 - S. Bass Warner, The Private City, Philadelphia in Three Periods of its Growth Philadelphia, 1968

- K.T. Jackson, "Urban Deconcentration in the Nineteenth Century: A Statistical Inquiry" in L.F. Schnore, ed. The New Urban History Princeton, 1975: 110-142

3 - M. Rischin, The Promised City: New York's Jews, 1870-1914 Cambridge, 1962

4 - S. Thernstrom, The Other Bostonians Cambridge, 1973

5 - R. Woods and A. Kennedy, The Zone of Emergence, Observations on the Lower, Middle and Upper Working Class Communities of Boston, 1905-1914 Cambridge, 1962 (original manuscript, 1904-1914)

6 - See one of the first essays of the Chicago ecological school, R.E. Park, E.W. Burgess and R.D. McKenzie, The City Chicago, 1925

7 - Problems of record linkage have usually been studied in terms of nominal linkages in various sources rather than of people and areas. See E.A. Wrigley, ed. Identifying People in the Past London, 1973

8 - R.D. McKenzie, The Neighborhood: a Study of Local Life in the City of Columbus, Ohio Chicago, 1923

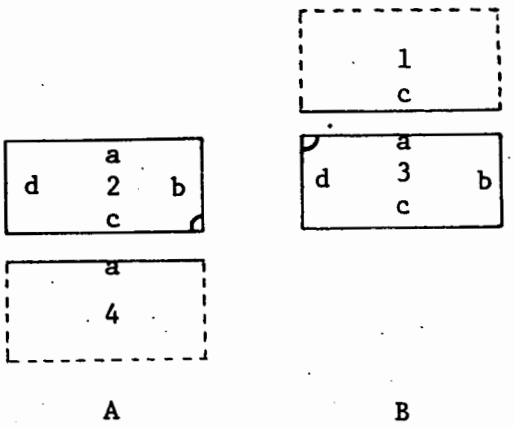
- S. Keller, The Urban Neighborhood: A Sociological Perspective New York, 1968

9 - H. Bartholomew, Land Uses in American Cities Cambridge, 1965
- F. Stuart Chapin, Urban Land Use Planning, New York, 1957

10 - For selected references on the use of the manuscript census by historians, see J.B. Sharpless and R.M. Shortridge, "Biased Underenumeration in Census Manuscripts: Methodological Implications." Journal of Urban History, Vol. I, number 4, August 1975: 409-439.

11 - E. Robinson and R.H. Pidgeon, Atlas of the City of Detroit and Suburbs Embracing Portions of Hamtramck, Springwells and Greenfield Townships New York, 1885

12 - If the blocks are selected without replacement and if any of the 4 corners can then be independently chosen for each block, there is a possibility of overlapping fronts between clusters if two selected primary blocks happen to be contiguous.



13 - Statistics of the Population of the United States at the 10th Census-June 1, 1880 Washington D.C., 1883: 420

14 - Ignoring the one N.D. frontage of plate 1

15 - $\hat{P}$ involves the sum of products of random variables. Formulae for variances of such products may be found in L.A. Goodman "On the Exact Variance of Products." Journal of the American Statistical Association 55 (1960): 708-713

16 - Variances shown in table 6 are computed with the following formula:

$$s^2_{\bar{X}} = \frac{\Sigma(X_i - \bar{X})^2}{n(n-1)}$$

17 - For the 5 plates without census data (5, 14, 17-19), we estimated the mean number of people per front. A scatter plot of the number of people per front against the % of P and O blocks per plate for the 15 plates with census data, led us to classify them into three groups:

1) few people per front (< 21) and small proportion of P and O blocks (< 65 %): plates 20 and 10

2) many people per front (> 60) and large proportion of P and O blocks (> 70 %): plates 2, 6, 7, 8, 9, 11, 12, 13, 15, 16

3) middle group: less than 42 people per front and above 70 % of P and O blocks in the plate: plates 1,3 and 4

On the basis of the proportion of P and O blocks, plates 5, 17, 18 and 19 were assigned to group 1 and plate 14 to group 3. At this stage, these plates lacking census data were assigned the mean number of people per front for their assigned group.

18 -

$$\hat{p} = \sum_{i=1}^{3} n_i \bar{f}_i * \bar{X}_i$$

19 -

$$\hat{V}(\hat{p}) = \sum_{i=1}^{3} n_i^2 \bar{f}_i^2 \left[s^2_{\bar{X}_i} \right]$$

20 - Including a research trip to the University of Pittsburgh library, where the original manuscripts are deposited, to read the few unreadable pages of the microfilm of the National Archives.

21 - Our prediction was most inaccurate for $\bar{X}_i$, the mean number of people per front in category 3. We had corrected some of this expected underestimation by generally overestimating $\bar{f}_i$*, the mean number of enumerable fronts per cluster:

category	mean number of people/front		mean number of enumerable fronts/cluster	
	Pilot sample	Final sample	Estimated	Final sample
1	14.05	17.68	2.75	1.81
2	32.38	32.46	4.30	3.70
3	14.12	35.69	4.10	3.69

22 - The 1880 City Directory of Detroit, 966 p. J.W. Weeks and co. Detroit, 1880.

23 - These are the most commonly used. The routines for match merging the data-sets and computing summary statistics at various levels of analysis are available in the Michigan Interactive Data Analysis System. See D.J. Fox and K.E. Guire, Documentation for MIDAS, revised edition August 1974, Statistical Research Laboratory of The University of Michigan, Ann Arbor, Michigan 48104.

24 - The stratum of a cluster of 6 fronts is identified by its primary block.

25 - In the original sample design there were 80 strata - twenty plates, each with four types of blocks. However, only four different sampling fractions were used, and thus there are only four different values of w_h. These correspond effectively to P and O blocks for each of the 3 groups of plates shown in Table 9 and to the ND and V blocks over all plates. The actual weights used were taken as being proportional to N_i/n_i where the constant of proportionality was chosen so that the weighted sum of the sample number of persons in each of these four categories reproduced the total sample size of 12,185 persons. Details are shown below.

Group	N_i	n_i	N_i/n_i	w_i	sample # of persons
1	177	12	14.75	1.2162	433
2	425	40	10.625	.8761	4597
3	650	50	13.000	1.0719	6823
4	288	25	11.52	.9499	332

26 - The formula for the estimated variance of r is given by

$$\text{Var } (r) = \frac{1}{\bar{x}^2}\left[\text{Var } (y) + r^2 \text{ var } (x) - 2r \text{ Cov } (x,y)\right]$$

where

$$V(y) = \sum_{h=1}^{H} \sum_{\alpha=1}^{a_h} (y'_{h\alpha} - \bar{y}'_h)^2,$$

$$V(x) = \sum_{h=1}^{H} \sum_{\alpha=1}^{a_h} (x'_{h\alpha} - \bar{x}'_h)^2$$

and

$$\text{Cov}(x,y) = \sum_{h=1}^{H} \sum_{\alpha=1}^{a_h} (x'_{h\alpha} - \bar{x}'_h)(y'_{h\alpha} - \bar{y}'_h)$$

Also,

$$x'_{h\alpha} = w_h x_{h\alpha},$$

$$y'_{h\alpha} = w_h y_{h\alpha}$$

and

$$\bar{y}'_h = \frac{1}{a_h} \sum_{\alpha=1}^{a_h} y'_{h\alpha},$$

$$\bar{x}'_h = \frac{1}{a_h} \sum_{\alpha=1}^{a_h} x'_{h\alpha}$$

Finally the standard error of r is given by

$$SE(r) = \sqrt{Var(r)}$$

see Leslie Kish, Survey Sampling, New York, 1965, chap. 6: "unequal clusters"

27 - Statistics of the Population of the United States at the 10th Census, op. cit.: 536-541, 420, 876

28 - See Olivier Zunz "Detroit en 1880 : espace et ségrégation" Center for Research on Social Organization of The University of Michigan, Working Paper # 121, August 1975. Forthcoming in Annales E.S.C.

"Sampling for a Study of the Population and Land Use of Detroit in 1880-1885" *

Olivier Zunz
Michigan Society of Fellows
The University of Michigan

William A. Ericson
Statistical Research Laboratory
The University of Michigan

Daniel J. Fox
Statistical Research Laboratory
The University of Michigan

CRSO Working Paper
124 December 1975
Revised June 1976

Copies available through:

The Center for Research
on Social Organization
The University of Michigan
330 Packard Street # 214
Ann Arbor, MI 48104

The design of an efficient sampling scheme for the study of population and space in the 19th century is a challenging problem for historians.[1] To examine the relationship of social life to the general form of the city, a researcher must have in his sample observations that cover the whole territory. Working on that scale however, he ordinarily sacrifices detail to achieve coverage. But to examine the constraints and the routine which are part of everyday experience, he needs that very detail--intensive observations of small populations and small areas. When he has that detail, he ordinarily sacrifices the attempt to achieve uniform coverage of the city as a whole. The two goals have seemed mutually exclusive in any single sampling design. Thus the historical study of the American city has often followed two distinct lines of approach: on the one hand, gross patterns in urban land use have been investigated to understand aspects of the city's change, its dynamics of growth, the development of suburbanization, for example.[2] On the other hand, intensive studies of the experience of neighborhoods or single ethnic or social groups have been conducted.[3]

The sampling scheme that we present here was conceived to study the city of Detroit in the late 19th century, both in the diversity of its neighborhoods and in its entirety. It is an areal sample of 127 geographic units drawn from the entire city. We capture Detroit in a snapshot in 1880-1885 when it was still a medium size city of 116,340 inhabitants. The sample was conceived to meet four criteria: to represent the whole city population in terms of demography, ethnicity and occupation which are the three most important sets of variables detailed in the U.S. census records; to represent geographic clustering in small neighborhoods so as to study how various categories of people were collected in the urban environment and the forms and intensities of their clustering;

to represent the interplay between population characteristics and land use patterns; and finally, to provide a baseline for studying the city's later development and transformation when intensive urbanization and changes in the immigration patterns, economic geography and community organization overtook it.

In attempting to satisfy these conditions, we faced three basic difficulties. The first difficulty came from uncertainties as to how to divide the 19th century city into coherent physical and social areas for investigation. Knowledge of the 19th century urban structure is incomplete. The late 19th century city, it is said, was in transition from the "walking city" of the ante bellum period to the segregated metropolis of the 20's. It had ceased to be the commercial city of the 1850's, small in scale with many different types of people and activities juxtaposed despite important socio-ethnic cleavages. But it was not yet the giant industrial metropolis,with neat patterns of residential segregation. The characteristics of archetypal urban areas, such as the slum, the ethnic neighborhood, the zone of emergence, and the suburb are well known.[4] Yet their interpenetration and their organization in relation to other areas of the city have been little explored. The city in transition has been defined more by what it was not rather than by what it was. In order to define it more clearly, we decided to sample all types of areas in the city: residential, non residential, even vacant, so as to fully locate social space within urban space.

The second problem arose from the fact that to study the relationship between social divisions and spatial arrangements requires the integration of very different types of information on both the population characteristics and the city's physical structure. Many different levels

of information coming from sources usually kept separated have to be collected and organized hierarchically. These are the data on individual city dwellers, the families and households, the ethnic and social groups on the one hand; the dwellings and houses, the streets and the neighborhoods, the larger districts of the city, on the other.[5] Matching different types of information for each sample unit requires of course that good historical sources be available. Fortunately such is the case for Detroit: the first comprehensive real estate atlas of the city of Detroit (the Robinson and Pidgeon atlas) is a detailed land use survey and appeared in 1885;[6] the manuscript of the 1880 U.S. census lists addresses for all individuals enumerated.[7]

The third problem, the most complex, was to decide the form--size and shape-of the areal units to be selected. The sample areal units were to be representative of the city's micro environment. In them, one should be able to observe the population's demographic, social, and ethnic life and structure at the local level. Taken together, they should provide a meaningful description of the overall urban territory; the combination of sample areas should represent the city as a whole. Thus it becomes possible to measure the diversity of local life in the light of a general picture of the city.

The following presentation describes the sampling unit and the sources of the study; it also presents the analysis of a pilot sample chosen to alleviate some of our uncertainties before we drew the final sample.

Definition of a Sampling Unit

So many complex and often hidden patterns develop in a multi-ethnic city of immigrants that only a fine grain analysis permits one to

catch overlapping phenomena, with different boundaries. One of the main difficulties of a geographic sampling procedure is to define a flexible unit of analysis, small enough to permit individual level observations of people and large enough to capture ethnic or socio-economic clustering. However, the grid plan of the American city is a natural sampling frame which can be used to that end. The block front, representing one side of a four sided block, is its smallest geographic component. It is a very flexible unit, easily drawn on a map along one street from one corner of the block to the other. When several frontages are linked, one creates a reconstruction of several urban forms: a block, a street, a small neighborhood. The sampling unit used here consists of a cluster of six fronts, in other words of one block and two opposing fronts (see Fig. 1).

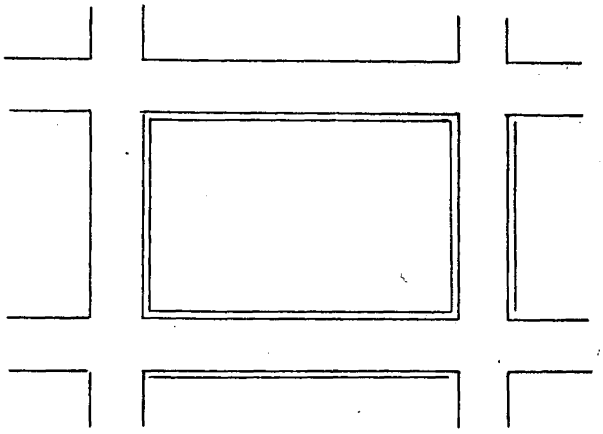


Fig. 1

The sampling unit can be divided into three units for purposes of analysis: the front, the block or the larger cluster. The block is the simplest one; it is geometrically well defined. But it is well known that city blocks often cover widely different realities. On the same block, the front on the main street is often different from the fronts on the

small streets, or the outside of the block differs from the inside. We chose the triple unit of front, block and cluster of six fronts to maintain a fine level of resolution to catch subtle patterns of concentration and scattered patterns of dominance. The addition to the primary block of two randomly selected opposing fronts permits us to represent streets on both sides without including all opposing fronts. The triple unit gives the possibility of changing levels with different analysis and/or comparing the same analysis at several geographic levels. The front is to be used for a fine survey of the housing-pattern. The block is more appropriate for questions related to urban densities. The cluster is a large enough unit to capture clustering patterns at a small neighborhood level (inhabited clusters have a mean of 123 inhabitants in the 1880 final sample). Using the three levels consecutively permits the measuring of geographic variations in clustering patterns of different ethnicity, social classes, age groups and other characteristics of the population; or in the contiguity of residences and non residential activities. Having a multiple unit is necessary. Given the diversity of the urban environment, there is no predetermined areal unit which would best permit one to study all types of inhabitants and areas simultaneously. Students of neighborhood activities have long recognized that geographic boundaries vary with different phenomena.[8] Those defining an ethnic cluster are not the same as those of a given social class neighborhood, and in turn overlap with boundaries of non residential areas. Neighborhoods have loose and shifting boundaries which cannot be predetermined for sampling purposes. Rather than artificially delineating them, selecting a geographic unit well fitted to the grid plan and sampling a large number of units increases the

chances of measuring changing neighborhood activities. The analysis of the same phenomenon at different geographic levels, the analysis of sets of units in specific parts of the city and the addition of all units together would permit us to study many questions related to the geography of the city and the distribution of demographic, ethnic and social patterns.

Given the source materials available and the overall aims of the study it was decided to draw some probability sample of the six-front clusters (using the atlas as a sampling frame) and then collect data (from the census manuscripts and the atlas) on all persons living in the selected clusters and on all buildings. The considerations leading up to the final design of the probability sample depended on the information available, the atlas and census manuscripts, tempered by economic and other feasibility constraints. We discuss some important aspects of these data sources before presenting the sample design.

The Sources

The two basic bodies of data used for sampling were "The Atlas of the City of Detroit published by E. Robinson and R.H. Pidgeon in 1885 and the Federal Census Manuscript Schedules of 1880. The atlas provided a detailed set of maps which identified not only blockfronts but also their lots and buildings and the house numbers. The Census Schedules provided demographic and socio-economic data for the listed individuals with their addresses. To draw the sample these two sources had to be matched. All addresses for any front selected had to be searched in the population listings. Thus a fairly accurate picture of the blockfront was reconstituted: on the one hand, the lots and buildings, on the other, the population.

Fronts, Blocks and Clusters in the Atlas

The Robinson Atlas covers the city of Detroit in twenty plates. For each block, parcels and buildings are designated. Houses are numbered and the atlas contains many indications of non residential buildings. Many fronts in the atlas cannot be found in the census manuscript because they were not enumerated by the census takers. A frontage was classified as "non enumerable" if it was impossible for a census taker to visit it. This was the case for either totally empty fronts, for fronts where houses faced other streets and for fronts which were not residential.

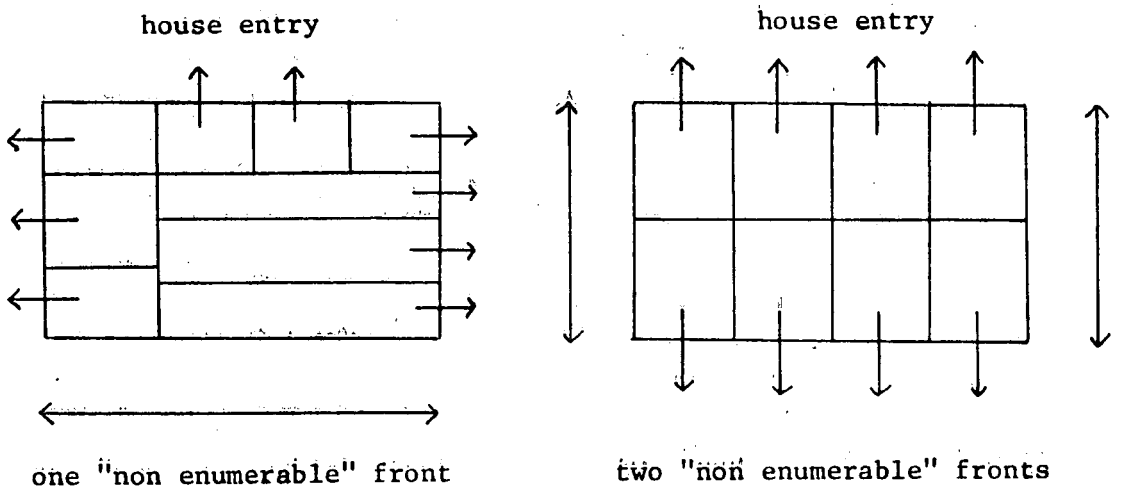
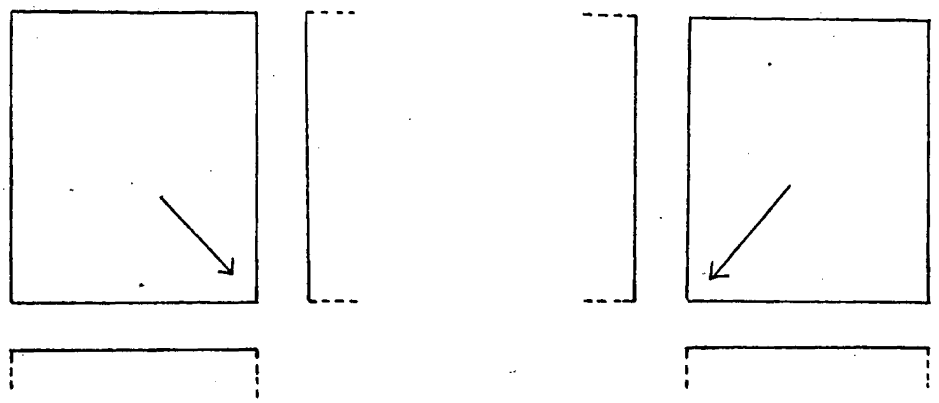


Fig. 2

Blocks and clusters are composed of enumerable and non enumerable frontages (Fig. 2). The ratio of enumerable to non enumerable frontages varies in the city depending on the location, the type of land use and the population density. Knowing this ratio is useful as a means of defining areas of the city. It is a crude but practical indicator of population density.

The very populated areas are likely to have substantially fewer non enumerable fronts per block than the scarcely populated areas. Population estimates can be computed for different sets of blocks in various areas of the city if one knows the number of inhabited (enumerable) frontages per block and the mean number of inhabitants per enumerable frontage. The atlas therefore contained information allowing us to divide the city into large geographic areas, not only in terms of land use but also of population density. Once these broad zones are defined, sampling clusters of six fronts within them may involve two steps: first randomly selecting blocks, and second randomly selecting a corner of each block to determine the opposing fronts.[9] (Fig. 3)



arrows indicate selected corners in the primary blocks to determine the opposing fronts

Fig. 3

Discrepancies between Atlas and Census

When selecting clusters of block fronts, we cannot assume that the individual houses of each front would be found in the census manuscripts in

the same geographic order as the atlas. We know that many census takers did not follow a strict route, but often shifted from even to odd numbers, or visited parts of streets at one time and revisited the other parts later. This implies that a thorough search in the census, for addresses selected from the atlas, is necessary to collect the information. All atlas addresses must be listed very carefully and every page of the manuscripts scrutinized until the search is exhausted.[10] Obviously, discrepancies always arise in the process of matching sources. In drawing a geographic sample of clusters from the 19th century city, we were using a census taken in 1880 and an atlas published in 1885. Discrepancies between the number of "residential buildings" per front in the atlas and the number of "houses" per front in the census schedules are likely to arise due to the different dates of each survey and underenumeration in the census. Yet this discrepancy could be computed and might be used as a correction factor for some statistical analysis.

These preliminary observations on the sources led us to a few unsolved questions. How complete for our clusters was the enumeration of the population reported in the census manuscripts? How were we to draw the sample of clusters from the atlas and obtain a good representation of the population? How to select the clusters so that they accurately represent the various areas of the city? A pilot sample was drawn from each of the two sources in order to help answer these questions and to give information useful in determining the total number of clusters to sample in different areas of the city, while also keeping manageable the search in the census manuscripts.

The Pilot Samples

No systematic information for the demographic, occupational, ethnic

and land use composition of fronts, blocks or cluster, or for the geographic differences of these variables across the city had ever been assembled. The pilot samples could not provide information on all variables of interest. It was assumed that urban density is the most useful parameter for predicting the variability of major population characteristics. Thus the pilot samples were intended to gain information on population and building density across the city. Two independent systematic samples were drawn, one comprising every fourth block on each of the twenty atlas plates and the other a sample of street "runs"--that is of subsets of the population listings under the same street name from the census manuscripts.

Pilot Sample of Blocks

The atlas contained sufficient detail so that each of the 1540 city blocks could be placed in one of four classes.

P = Promising: Blocks containing almost all dwelling structures.

ND = Non Dwelling: Blocks mainly occupied by non residential establishments (industrial, commercial, etc.)

V = Vacant: Blocks containing no structures, 90% or more vacant.

O = All other blocks.

For P and ND blocks, we treated the downtown area a little differently than the rest of the city in order to take into account the density decline from the center to the periphery. This decline was expected and indeed visible in the atlas. Plates 1,2,6, and 11 roughly comprised all the area within one mile of the city center. In those plates, 3/4 of the dwelling structures of a block had to be residential for that block to qualify for P, and 3/4 ND to qualify for ND. Elsewhere a ratio of 2/3 was considered sufficient. The results of this categorization are shown in table 1.

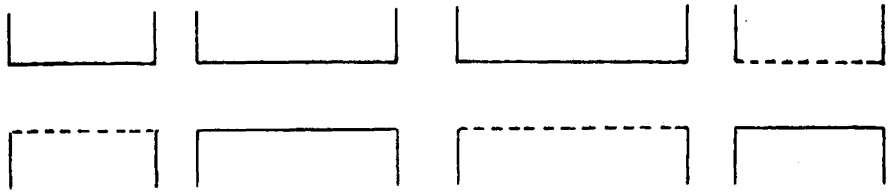
INSERT TABLE 1

To determine the pilot sample of blocks, every block was numbered; then every fourth block on each of the twenty plates was sampled (even numbered plates: every other even numbered block; odd numbered plates: every other odd numbered block). Then for each sampled block the number of non-enumerable frontages was determined. The number of non enumerable frontages was subtracted from the total number of frontages for each sample block to yield the number of enumerable frontages per block. The mean numbers of enumerable frontages per type of block are shown in Table 2.

INSERT TABLE 2

Census Pilot Sample

The census pilot sample gave us some idea of the size of frontages (e.g. numbers of households and persons per frontage) and of the problems of tracing frontages in the census data. Information was recorded for every 80th street in the census manuscript, or more exactly for every 80th "run" of consecutive persons listed as living on the same street in the manuscript. For each run, we recorded the microfilm reel number, census page, street name, house numbers, number of households and number of people in each household. Every street or "run" recorded in this manner was then traced in the atlas. Using this procedure, we were able to locate 87 frontages in the atlas. In some cases, under the same street name, the census takers had crossed the street many times, mixing odd and even numbers. For some runs, we actually got several fronts of the atlas (Fig.4).



- - -: fronts enumerated in one run

Fig. 4

In other cases, under one street name, there were only one or two houses for a given front instead of the 15 or 20 existing in the atlas (Fig. 5).

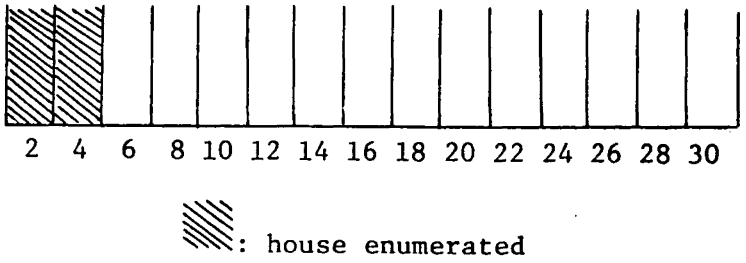


Fig. 5

At this point, it was obvious that the census enumeration was even more disorganized than we had expected. It was very rare that a front was fully listed in one run of the manuscript. Yet we had collected enough information to estimate gross land use and population characteristics.

ANALYSIS OF PILOT SAMPLE DATA

Population Size and Density: The Census and the Atlas

As was expected, the most populated fronts of the Atlas (in Promising and Other type blocks) were also densely populated in the census enumeration; similarly, the least populated fronts of the Atlas (in Non-Dwelling and Vacant type blocks) were unenumerated in the census. The census provided relatively little information about blocks in the Other category and almost none for the Vacant and Non-Dwelling blocks or for plates 5, 14, 17 and 19. In order to insure the quality of our pilot matching of the census manuscripts with the Robinson Atlas, we tried to predict the total census population, known to be 116,340,[11] from the data collected from the P and O type blocks. We first combined the data from the fronts in P and O type blocks. We then computed the mean number of persons per frontage per street run for each plate. This means that the original 87 frontages (or segment of frontage) observations--see fig. 4 and 5-- were consolidated into 33 "independent" frontage observations for the 20 plates. One single

estimate of the total population is given by:

$$\hat{P} = \sum_{i=1}^{20} N_i \bar{X}_i \bar{f}_i , \qquad (1)$$

where N_i is the number of P or O blocks on the ith plate, $\bar{X}$ is the mean number of persons per enumerable frontage and $\bar{f}_i$ is the mean number of enumerable frontages per block. Since four plates were without sample data(see table 3), and seven contained only one observation, we decided to combine plates on the basis of the type of blocks in them (see table 1). Plates 7, 8,12, 13 were the most dense plates with more than 100 blocks. Each had at least 3 frontage observations; thus we kept each of them separated. Then we lumped plates 5,10, 17-20. These plates comprehended peripheral areas with at least 45% ND and V type blocks. The last groups comprised plates 1,9,14,15 (Mixed land use but a very large number of O type blocks), 2, 6,11 (mixed land use but a very large number of P type blocks), 3,14,16, (no ND at all). Formula (1) was then applied to the eight resulting combinations of plates. The data are shown in table 3. These combinations lead to 83,088 as an estimate of the total population size, an underestimate by 33,252.

INSERT TABLE 3

In order to obtain some idea of the sampling error, the following formula was used in computing an estimate of the variance of P:

$$\hat{V}(\hat{P}) = \sum_{i=1}^{8} N_i^2 \left(\bar{f}_i^2 \, s^2_{\bar{X}_i} + \bar{X}_i^2 s^2_{\bar{f}_i} + s^2_{\bar{X}_i} s^2_{\bar{f}_i}\right)^{12} \qquad (2)$$

where i refers to the ith of the eight plate groups shown in table 5, $s^2_{\bar{x}}$ is the estimated variance of the mean number of persons per frontage, $\bar{X}_i$, and $s^2_{\bar{f}}$ is the variance of the mean number of enumerable frontages per block, $\bar{f}_i$. The variances used in determining this estimated value $\hat{V}(\hat{P})$ are shown in table 3.[13]

The application of formula (2) to the values shown in table 3 yielded an estimated standard error of $\hat{P}$ as 10,409. Thus the total population estimate 83,088 was 3.19 estimated standard errors below the true population value of 116,340. This estimate, though quite low, was nevertheless understandable. The factors contributing to it include our exclusion of the ND and V blocks and the small sample used in estimating or computing $\bar{X}_i$ and $\hat{V}(\hat{P})$ (ranging between 3 and 6 frontage observations per group of plates). Furthermore the systematic sample of street runs in the census was misleading. Many actual block fronts had been poorly recorded in our pilot sample because information had been located at only one place in the manuscript census when, due to revisitation, the same fronts were most likely recorded in several different parts of the census volumes.

Another estimate of 107,782 for the population total or an estimate of .925 (= $\frac{107,782}{116,340}$) was made when considering only the well enumerated fronts in our pilot sample. The basic scheme was to divide the pilot

sample into two parts: the first, called minor, was where census and atlas were poorly matched, i.e. when the census data contained three or less houses in a front that the atlas showed to have 4 or more residential buildings; the second, called major, was where census and atlas were well matched, in other words where most houses on a front had been enumerated in the census on one visit by the census taker. Then formula (1) was applied to the twenty plates using only the data from the major group and estimating the mean number of people per front for the plates without census data in the pilot sample.[14]

At this stage in the analysis of the pilot sample data, we knew that much work was necessary to reorder th manuscript census in order to locate persons living in a sampled front. Yet we felt that a thorough search in the manuscript census was manageable and would suffice in order to match the two sources. We had to gain some idea of geographic variation on density patterns before deciding on how to stratify the final sample.

Geographic Variation of Density Patterns

Building and population densities reflect the variety of the urban scene that we wanted to capture. To be sure, our pilot sample was composed of two very small systematic samples. Population estimates that we derived from it were not fully reliable. Yet we felt that we now knew enough to group the twenty plates of the atlas into several density classes. By analyzing the estimated population per front and the types of blocks per plate, a pattern emerged. The peripheral plates--5,10,17-20--were all more than two miles from the city center. As we had already noticed, they were almost vacant. The remaining plates could be divided into two other groups on the basis of estimated population per block. Plates 1,3,4,8,15 and 16 seemed to represent the mostdensely populated areas of the city with at least 80 people per city block. The last group consisted of the remaining plates 2,6,7,9,

11,12, 13, and 14. The computations leading to the classification into the three classes of plates according to population density are shown on table 4.

INSERT TABLE 4

Estimating the Sample Size

The analysis of the pilot sample data helped us to achieve two results: it acquainted us with the procedure of matching the sources to draw the cluster sample; it allowed us to divide the city into three large density areas. Given our analysis of population estimates and types of blocks, we felt that drawing at least 100 clusters, the primary block being classified either Promising or Other, would yield tolerably small standard errors for estimating residential and population characteristics. Such a number would also keep the search in the census manuscripts to a manageable task. In order to estimate other city characteristics, we decided that 25 clusters would be selected from the Non Dwelling and Vacant blocks.

Following this design, we then estimated the size of the sample, that is the number of households and persons it would produce. One difficulty was in estimating the number of enumerable frontages associated with each sampled cluster. The pilot sample provided estimates only by blocks. The sampling of clusters, it will be remembered, require augmenting the block by two opposing frontages. Therefore the observed pilot sample number of enumerable frontages per block, $\bar{f}_i$, observed in the pilot sample, was adjusted upward in each of the three groups of plates to reflect the inclusion of the opposing frontages. These are the values $\bar{f}_i^*$ shown in table 5. These values were determined quite subjectively; in densely populated areas, like group 2, $\bar{f}_i^*$ is as large as one and a half times $\bar{f}_i$; for sparsely populated areas with large numbers of V and ND blocks, there is little difference between $\bar{f}_i$ and $\bar{f}_i^*$. The estimated sample number of persons is given using formula (1) replacing N_i by n_i, the sample number of blocks to be selected, and $\bar{f}_i$ by $\bar{f}_i^*$. This was computed as 8,930 persons.[15] It was expected that there would be an average of 5 persons per household and thus we estimated that a sample

of 100 clusters would contain some 1,786 households. However, even making the assumption that the $\bar{f}_i^*$ are known without error, the estimated standard error of the estimate of the sampled number of persons is 961.[16] The estimated standard error of the estimate of the sampled number of households is 192.

INSERT TABLE 5

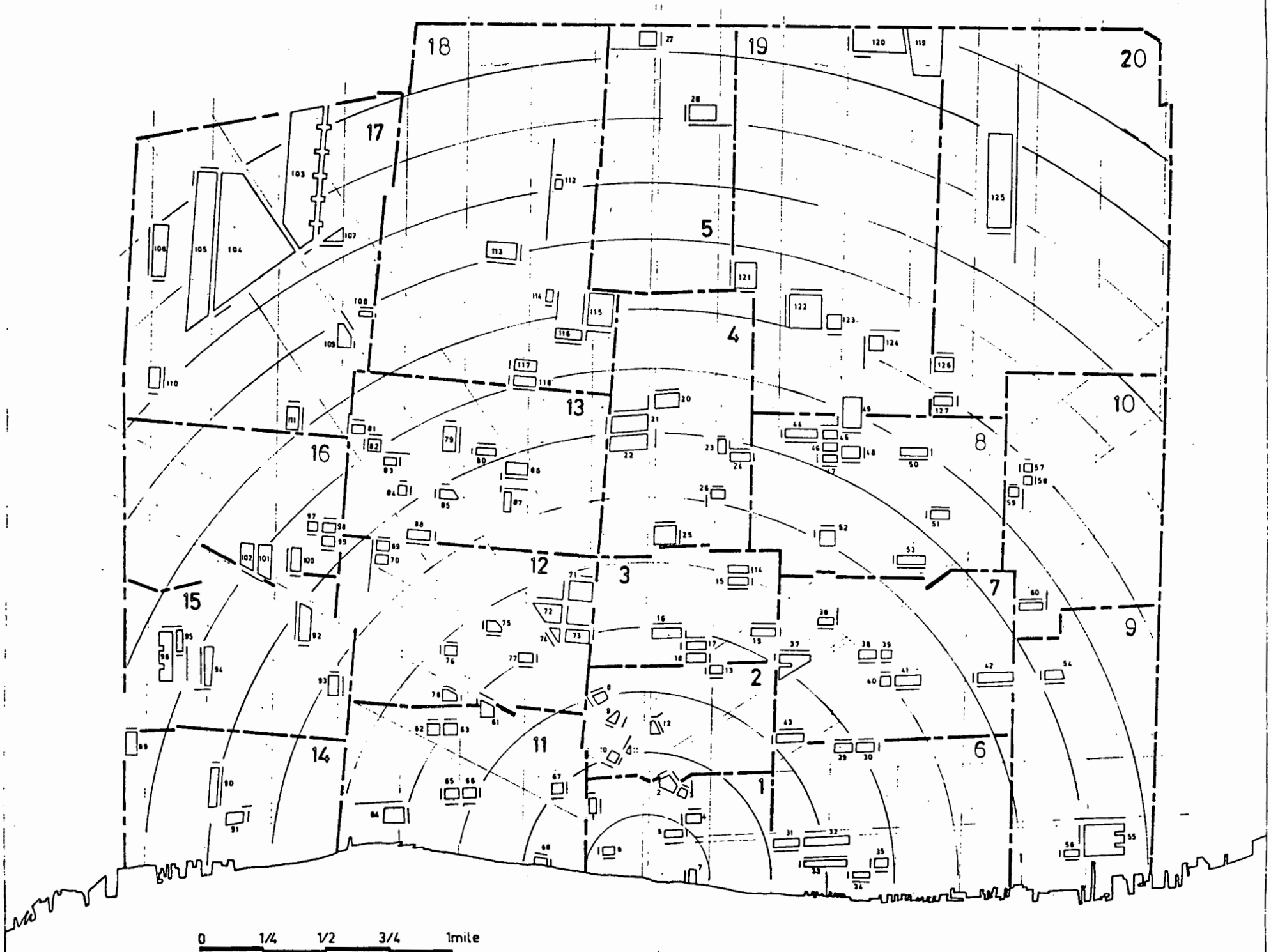
In the end 102 P and O blocks were selected with their opposing frontages. We sampled rather intensely in the most populated plates (Category 2),9.4% of the P and O blocks, that is 40 of the 425 such blocks. Applying the same ratio to category one and three would have yielded too many clusters, more than needed at at too great expense at the time of the data collection. We therefore reduced it to roughly 6.8% or 12 blocks from among the 177 P and O blocks in category one--sparsely populated plates; and about 7.7% or 50 of the 650 blocks in category 3--remaining plates. These ratios were applied separately to the P and O block categories. Also 25 or about 8.7% of the 288 ND and V blocks were sampled, yielding a total sample of 127 geographic units.

INSERT TABLE 6

The actual mechanics of drawing the sample were quite straightforward given table 6 and the tracings of the blocks as shown on the 20 plates of the atlas. The blocks of each type were serially numbered on the plate and, using random numbers, a simple random sample without replacement of blocks was selected for each of the four block categories within each plate. Once a block was selected and located on the atlas plate tracing, then the Rand Table of random digits was used to choose one of the, usually four, block corners. The usually two opposing frontages were then included to make up the cluster (see fig. #3 and footnote #9). The sampled clusters are displayed in the map below. The final design was thus a stratified sample--by plate and block type--of clusters.

DETROIT
SAMPLE

From E.ROBINSON & R.H.PIDGEON:ATLAS OF THE CITY OF DETROIT 1885



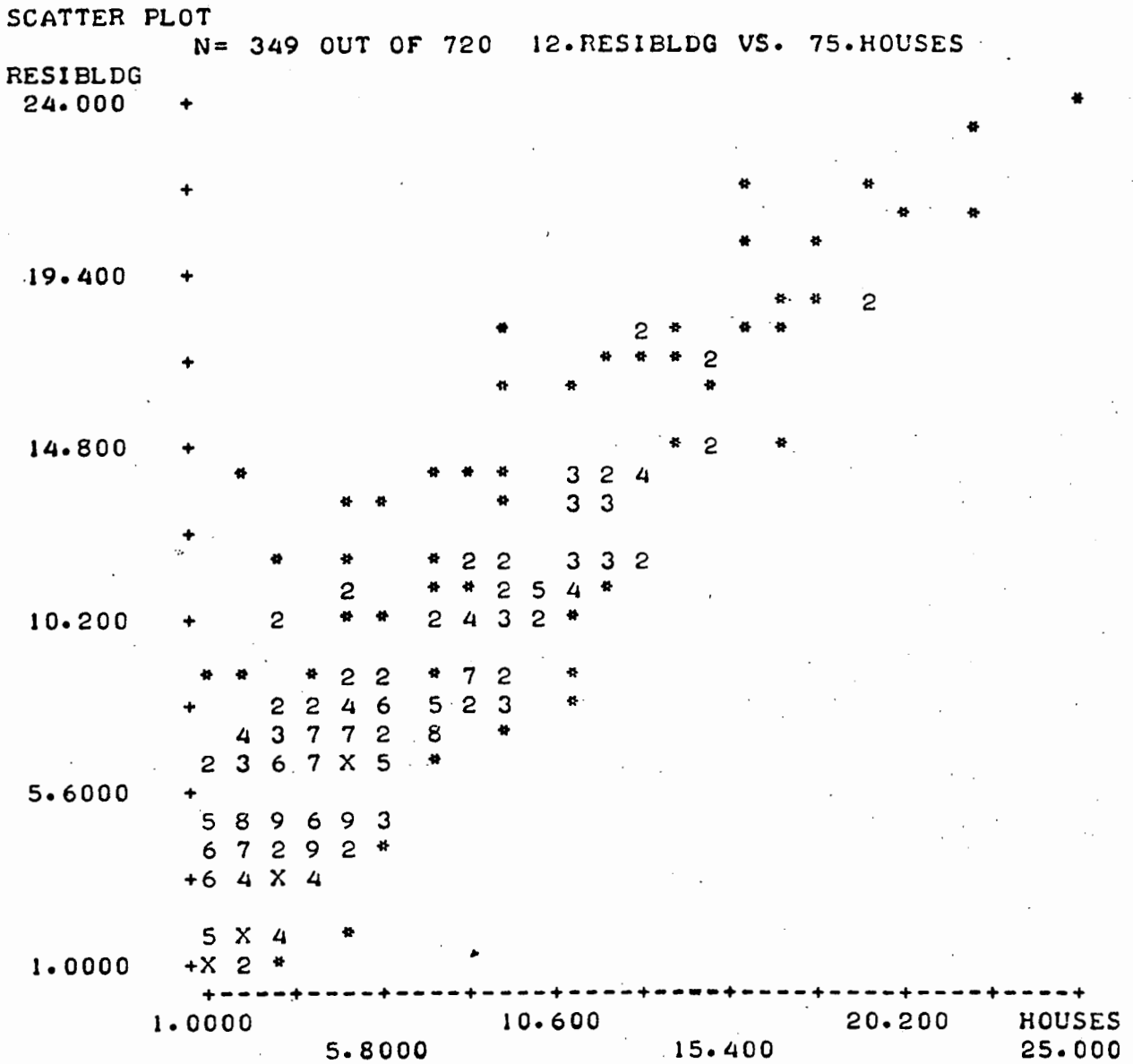
THE SAMPLE

A Census of 127 Clusters

Based on an exhaustive search in the manuscript census, the final sample includes 12,185 people.[17] This comprises 2,410 households on 721 frontages or 127 clusters. Of these,353 frontages were actually inhabited. The others were non residential or more generally non enumerable. The mean household size was 5.58, the mean number of people per front was 33.47 (min = 1; max = 152).[18] A full land use and population census has been taken of each unit. From the atlas and other sources, all the physical characteristics of each front have been recorded.[19] A detailed land use survey has been conducted to record the number and types of parcels, the number and types of buildings, (residential, non residential, small unnumbered buildings, stables), and the specific type of occupancy in each lot and built structure: types of residences, of manufacture, of craftmanship, of commerce, of business and professional services, of public and quasi public services; types of transportation, communication and utilities equipment; amount of unused space. Each area land use system is thus fully described and can be correlated with population characteristics. From the census manuscripts, the demographic, occupational and ethnic characteristics of every individual living on the fronts have been coded. This total enumeration of the clusters is essential to study the problems of density, household structure, family size, family composition and demographic behavior, as well as for the study of complex ethnic and social distribution patterns. The final sample resulted in a variety of data-sets, each corresponding to one level of data collection and/or analysis: the individual file of 12,185 people, the family file (2,410 cases), the front file (720 cases), the block file (127 cases), the cluster file (127 cases).[20]

The Discrepancy Rate between the Sources

A simple measure of discrepancy between the two sources-- the Atlas and the Census--wasgiven by subtracting the number of houses found in the census from the number of residential buildings pictured in the atlas for each front. A scatter plot of the number of residential buildings per front against the number of houses in the census shows that the discrepancy rate is low and regular (Fig.6). The discrepancy itself varies from -3 to 12. The mean discrepancy is only 1.62 houses per front. It is accounted for by houses built between 1880 and 1885 (i.e. houses shown in the atlas but not found in the census), and in some cases, by houses found in the census at addresses in between atlas addresses.



Number of Residential Buildings per Front in the Atlas against Number of Houses per Front in the Census Manuscript. There are 349 inhabited fronts (out of 364 listed in the census) with information in both sources.

Fig. 6

The Standard Errors of Estimates

Some aspects of the representativeness of the sample can be judged by checking to see how known population characteristics (from published census figures) are estimated from our sample. Many of these characteristics are proportions. For example the proportions of inhabitants who were whites, the proportion born in Canada, the proportions of females in the working population. In each of these cases the estimated proportion is a ratio where both numerator and denominator are subject to sampling variability.

Suppose that the population proportion to be estimated from the sample is designated by

$$R = X/Y$$

where Y is the population number of all persons in the population and X is the number of those having the attribute. In such a case R is estimated by

$$r = x/y$$

where x and y are sample estimates of X and Y respectively.

For the sampling design used, x and y are given by

$$x = \sum_{h=1}^{H} w_h \sum_{\alpha=1}^{a_h} x_{h\alpha}$$

and

$$y = \sum_{h=1}^{H} w_h \sum_{\alpha=1}^{a_h} y_{h\alpha}$$

where H is the number of strata, a_h is the number of clusters in stratum h,[21] w_h is a weight associated with the h-th stratum and is proportioned to N_h/n_h, the ratio of the population number of cluster to the sample number of clusters in the h-th stratum, and $x_{h\alpha}$ and $y_{h\alpha}$ are the sample totals for the α-th cluster within the h-th stratum.[22]

Owing to the stratification by plate of the atlas which permitted us to achieve geographic coverage, some plates have only one sampled cluster per stratum. In computing r, strata having one (or less) sampled clusters presented no difficulty. However in estimating the variance of r (or its square root, the standard error of r), one needs two or more sample observations or cluster per stratum. In computing the standard error of r, denoted by SE(r), the 80 strata were reduced to 27 by combining similar strata so that each combined stratum had at least two sampled clusters.

In figure 7 we display the sample estimate, r, for each of ten population ratios. A range of two standard errors above and below the sample estimate is also indicated.[23] Finally the true population value from the published census is indicated by an " X ".[24]

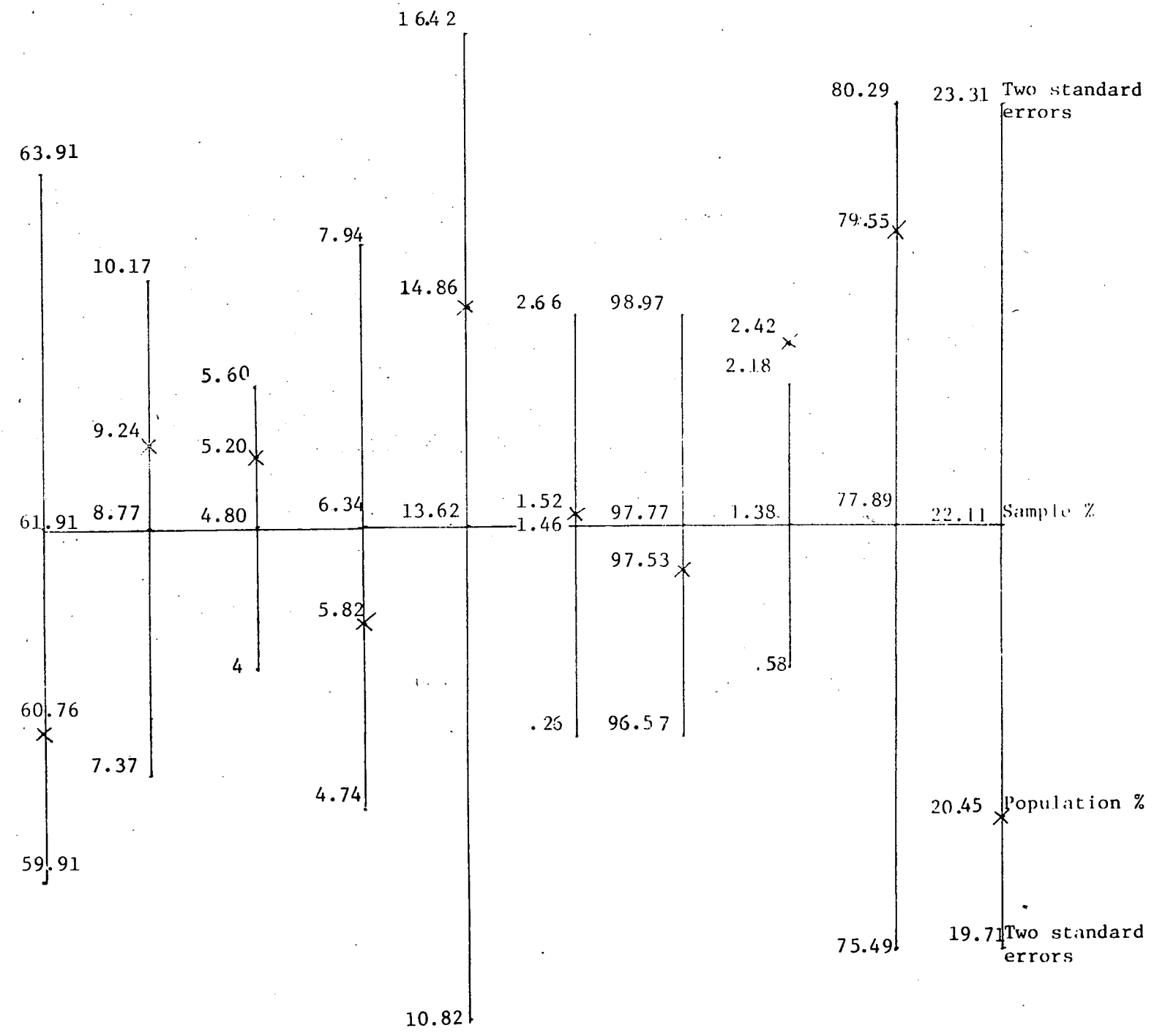


Fig. 7

Conclusion

This sampling procedure is only one way to sample data for the study of differential processes of settlement in a city. It is best to concentrate the analysis on the relationship between spatial and social organization in small areas. For instance, the 1880 sample has been used to investigate the geographic distribution of the ethnic, occupational and socio-ethnic groups in the city. We determined the forms of grouping and the type of spatial segregation that was occurring. To take only one problem, ethnic segregation, the sample proved to be highly sensitive to the subtler forms of clustering that would have been overlooked, had a simpler, more traditional design, been used.[25] Analysis concerning the demography of the neighborhoods, the organization of local life in the city and the interplay between work and residence is in progress. The 1880 sample has been duplicated for 1900 and enlarged to include clusters within the expanded city limits. This permits us to study the evolution of the 1880 sampled clusters over time in relation to the change in the city structure, size and population. Other methods should be explored, with different emphasis, such that a variety of possible sampling schemes can be developed for the historical study of the urban environment. The success of quantitative urban history calls for the use of sampling methods which can cope with the rich texture of the urban phenomenon.

TABLE 1

Block Categorization by Atlas Plate

Number of Blocks

Plate #	Promising	Non-Dwelling	Vacant	Other	Total	Proportion V or ND
1	29	15	0	28	72	.21
2	63	10	0	9	82	.12
3	68	0	0	0	68	-0-
4	35	0	3	30	68	.04
5	5	0	11	12	28	.39
6	69	9	0	8	86	.10
7	94	3	0	12	109	.03
8	85	0	4	25	114	.04
9	28	4	3	13	48	.15
10	21	0	20	14	55	.36
11	73	11	2	14	100	.13
12	100	7	1	11	119	.07
13	86	1	6	29	122	.06
14	23	3	2	18	46	.11
15	42	3	1	18	64	.06
16	49	0	2	16	67	.03
17	13	1	59	30	103	.58
18	18	0	42	24	84	.50
19	6	0	41	24	71	.58
20	5	0	24	5	34	.71
Totals	912	67	221	340	1540	

TABLE 2

Pilot Block Sample

Mean Numbers of Enumerable Frontages Per Block

By Type of Block

(Sample Size in Parentheses)

Plate	Promising	Non-Dwelling	Vacant	Other	Total Sample Size
1	3.0000 (8)	4.0000 (1)	--- (0)	3.3333 (9)	18
2	3.4375 (16)	1.7500 (4)	--- (0)	4.0000 (1)	21
3	3.7059 (17)	--- (0)	--- (0)	--- (0)	17
4	3.0909 (11)	--- (0)	.5000 (2)	3.250 (4)	17
5	2.5000 (2)	--- (0)	1.0000 (2)	2.0000 (3)	7
6	3.3125 (16)	1.0000 (1)	--- (0)	2.0000 (4)	21
7	3.5000 (24)	--- (0)	--- (0)	2.5000 (4)	28
8	3.3750 (24)	--- (0)	1.5000 (2)	3.0000 (3)	29
9	3.1667 (6)	--- (0)	2.0000 (1)	3.0000 (5)	12
10	3.2857 (7)	--- (0)	1.0000 (4)	2.3333 (3)	14
11	3.1579 (19)	0.0000 (2)	1.5000 (2)	3.0000 (2)	25
12	3.2692 (26)	0.0000 (2)	0.0000 (1)	3.0000 (1)	30
13	3.2609 (23)	--- (0)	--- (0)	2.3750 (8)	31
14	3.0000 (4)	0.0000 (1)	0.0000 (1)	2.3333 (6)	12
15	2.9000 (10)	1.0000 (1)	--- (0)	2.6000 (5)	16
16	2.8000 (15)	--- (0)	--- (0)	2.5000 (2)	17
17	2.2500 (4)	--- (0)	.3333 (15)	2.1429 (7)	26
18	2.6667 (3)	--- (0)	.6364 (11)	2.1429 (7)	21
19	--- (0)	--- (0)	.9231 (13)	2.600 (5)	18
20	4.0000 (1)	--- (0)	1.1667 (6)	2.5000 (2)	9
					389

TABLE 3

Plate Numbers # of effectively independent frontages drawn from census pilot sample ( ).	# of P & O Blocks	$\bar{X}_i$ mean # persons/frontage	$S^2_{\bar{X}_i}$	$\bar{f}_i$ mean # of enumerable frontages	$S^2_{\bar{f}_i}$	$N_i \bar{X}_i \bar{f}_i$
5(0), 10(1), 17(0), 18(1), 19(0), 20(1)	177	14.00	20.33	2.500	.01955	6,195
1(1), 9(1), 14(0), 15(2)	199	39.25	104.56	2.9434	.01119	22,990
3(2), 4(1), 16(3)	198	27.50	107.92	3.2041	.01273	17,446
2(2), 6(3), 11(1)	236	18.33	13.98	3.2069	.01498	13,873
7(3)	106	10.00	9.33	3.3571	.01644	3,559
8(3)	110	30.00	86.33	3.3333	.01424	11,000
12(3)	111	11.00	14.33	3.2593	.02733	3,980
13(5)	115	11.60	4.16	3.0323	.02900	4,045
	1252					83,088

Table 4

Plate #	Average # Persons per Frontage	#P or 0 Blocks	Proportion ND or V Blocks	No. Enum. Frontages per Block	Estimate of Population Size	Estimate of Population per Block	Block Categorization
1	64.33	57	.21	3.1765	11,648	204.3	2
2	17.60	72	.12	3.4706	4,398	61.1	3
3	21.75	68	.00	3.7059	5,481	80.6	2
4	31.00	65	.04	3.1333	6,314	97.1	2
5	--	17	.39	2.2000	--	--	1
6	16.60	77	.10	3.0500	3,899	50.6	3
7	9.97	106	.03	3.3571	3,548	33.5	3
8	29.93	110	.04	3.3333	10.974	99.8	2
9	24.00	41	.15	3.0909	3,041	74.2	3
10	22.67	35	.36	3.0000	2,380	68.0	1
11	25.00	87	.13	3.1429	6,836	78.6	3
12	11.11	111	.07	3.2592	4,019	36.2	3
13	11.40	115	.06	3.0323	3,975	34.6	3
14	--	41	.11	2.6000	--	--	3
15	34.50	60	.06	2.8000	5,796	96.6	2
16	30.33	65	.03	2.7647	5,450	83.9	2
17	--	43	.58	2.1818	--	--	1
18	9.00	42	.50	2.3000	869	20.7	1
19	--	30	.58	2.6000	--	--	1
20	10.50	10	.71	3.0000	315	31.5	1

TABLE 5

Category	Plate Group	Pilot Sample Mean # persons/front, $\bar{X}_i$	Sample Size P and O Blocks, n_i	Pilot Sample # enum. front/block, $\bar{f}_i$	f_i*
1	5, 10, 17-20	14.0567	12	2.5000	2.75
2	1, 3, 4, 8, 15, 16	32.3858	40	3.1759	4.30
3	2, 6, 7, 9, 11-14	14.1261	50	3.1636	4.10

Population Size

Sample Design

Table 6

Table 6

SAMPLE DESIGN

	Population Size				Sample Size			
Plate #	P	O	ND	V	P	O	ND	V
5	5	12	0	11	0	1	0	1
10	21	14	0	20	1	1	0	2
17	13	30	1	59	1	2	0	6
18	18	24	0	42	1	2	0	4
19	6	24	0	41	0	2	0	4
20	5	5	0	24	1	0	0	2
Total Group 1	68	109	1	197	4	8	0	19
1	29	28	15	0	3	3	1	0
3	68	0	0	0	6	0	0	0
4	35	30	0	3	4	3	0	0
8	85	25	0	4	8	2	0	0
15	42	18	3	1	4	1	0	0
16	49	16	0	2	5	1	0	0
Total Group 2	308	117	18	10	30	10	1	0
2	63	9	10	0	4	1	1	0
6	69	8	9	0	5	1	1	0
7	94	12	3	0	7	1	0	0
9	28	13	4	3	2	1	0	0
11	73	14	11	2	6	1	1	0
12	100	11	7	1	8	1	1	0
13	86	29	1	6	7	2	0	1
14	23	18	3	2	2	1	0	0
Total Group 3	536	114	48	14	41	9	4	1
Grand Total	912	340	67	221	75	27	5	20

The authors wish to thank Karol P. Krotki for his assistance in computing the standard errors of estimate, Charles Tilly and David Bien for their comments on the manuscript. The design of the sample and the data collection have been made possible by grants from the Michigan Society of Fellows and of the Population Development Fund of the Ford Foundation.

1 - In addition to W.G. Cochran, Sampling Techniques New York, 1963, the historian may wish to consult R. S. Schofield "Sampling in Historical Research" in E.A. Wrigley, Nineteenth Century Society Cambridge, 1972: 146-190.

2 - S. Bass Warner, The Private City, Philadelphia in Three Periods of its Growth Philadelphia, 1968.

- K.T. Jackson, "Urban Deconcentration in the Nineteenth Century: A Statistical Inquiry" in L.F. Schnore, ed. The New Urban History Princeton, 1975: 110-142.

3 - M. Rischin, The Promised City: New York's Jews, 1870-1914 Cambridge, 1962.

4 - R. Woods and A. Kennedy, The Zone of Emergence, Observations on the Lower, Middle and Upper Working Class Communities of Boston, 1905-1914 Cambridge, 1962 (original manuscript, 1904-1914).

5 - Problems of record linkage have usually been studied in terms of nominal linkages in various sources rather than of people and areas. See E.A. Wrigley, ed. Identifying People in the Past London, 1973.

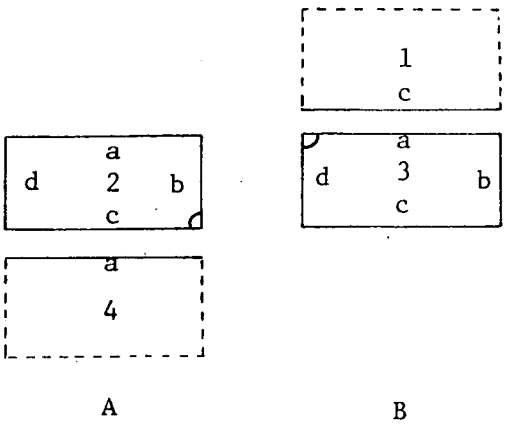
6 - E. Robinson and R.H. Pidgeon, Atlas of the City of Detroit and Suburbs Embracing Portions of Hamtramck, Springwells and Greenfield Townships New York, 1885.

7 - For selected references on the use of the manuscript census by historians, see J.B. Sharpless and R.M. Shortridge, "Biased Underenumeration in Census Manuscripts: Methodological Implications." Journal of Urban History, Vol. I, number 4, August 1975: 409-439.

8 - R.D. McKenzie, The Neighborhood: a Study of Local Life in the City of Columbus, Ohio Chicago, 1923.

- S. Keller, The Urban Neighborhood: A Sociological Perspective New York, 1968.

9 - If the blocks are selected without replacement and if any of the 4 corners can then be independently chosen for each block, there is a possibility of overlapping fronts between clusters if two selected primary blocks happen to be contiguous.



Front 2b is in clusters A & B

Front 3b is also in clusters A & B

10 - The Census Bureau's enumeration districts boundaries are useful to narrow down the search in the documents.

11 - Statistics of the Population of the United States at the 10th Census - June 1, 1880 Washington, D.C., 1883: 420.

12 - $\hat{P}$ involves the sum of products of random variables. Formulae for variances of such products may be found in L.A. Goodman "On the Exact Variance of Products." Journal of the American Statistical Association 55 (1960): 708-713.

13 - Variances shown in table 6 are computed with the following formula:

$$s^2_{\bar{X}} = \frac{\Sigma(X_i - \bar{X})^2}{n(n-1)}$$

14 - For the 5 plates without census data (5, 14, 17-19), we estimated the mean number of people per front. A scatter plot of the number of people per front against the % of P and O blocks per plate for the 15 plates with census data, led us to classify them into three groups:

1) few people per front (<21) and small proportion of P and O blocks (<65%): plates 20 and 10

2) many people per front (>60) and large proportion of P and O blocks (>70%): plates 2, 6, 7, 8, 9, 11, 12, 13, 15, 16

3) middle group: less than 42 people per front and above 70% of P and O blocks in the plate: plates 1, 3 and 4.

On the basis of the proportion of P and O blocks, plates 5, 17, 18 and 19 were assigned to group 1 and plate 14 to group 3. At this stage, these plates lacking census data were assigned the mean number of people per front for their assigned group.

15 - $$\hat{p} = \sum_{i=1}^{3} n_i \bar{f}_i^* \bar{X}_i$$

16 - $$\hat{V}(\hat{p}) = \sum_{i=1}^{3} n_i^2 \bar{f}_i^2 \left[s^2_{\bar{X}_i} \right]$$

17 - Including a research trip to the University of Pittsburgh library, where the original manuscripts are deposited, to read the few unreadable pages of the microfilm of the National Archives.

18 - Our prediction was most inaccurate for $\bar{X}_i$, the mean number of people per front in category 3. We had corrected some of this expected underestimation by generally overestimating $\bar{f}_i^*$, the mean number of enumerable fronts per cluster:

category	mean number of people/front		mean number of enumerable fronts/clusters	
	Pilot sample	Final sample	Estimated	Final sample
1	14.5	17.68	2.75	1.81
2	32.38	32.46	4.30	3.70
3	14.12	35.69	4.10	3.69

19 - The 1880 City Directory of Detroit, 966 p. J.W. Weeks and co. Detroit, 1880.

20 - These are the most commonly used. The routines for match merging the data-sets and computing summary statistics at various levels of analysis are available in the Michigan Interactive Data Analysis System. See D.J. Fox and K.E. Guire, Documentation for MIDAS, revised edition August 1974, Statistical Research Laboratory of The University of Michigan, Ann Arbor, Michigan 48104.

21 - The stratum of a cluster of 6 fronts is identified by its primary block.

22 - In the original sample design there were 80 strata - twenty plates, each with four types of blocks. However, only four different sampling fractions were used, and thus there are only four different values of w_h. These correspond effectively to P and O blocks for each of the 3 groups of plates shown in Table 9 and to the ND and V blocks over all plates. The actual weights used were taken as being proportional to N_i/n_i where the constant of proportionality was chosen so that the weighted sum of the sample number of persons in each of these four categories reproduced the total sample size of 12,185 persons. Details are shown below.

Group	N_i	n_i	N_i/n_i	w_i	sample # of persons
1	177	12	14.75	1.2162	433
2	425	40	10.625	.8761	4597
3	650	50	13.000	1.0719	6823
4	288	25	11.52	.9499	332

23 - The formula for the estimated variance of r is given by

$$\text{Var } (r) = \frac{1}{x^2} \left[\text{Var } (y) + r^2 \text{ var } (x) - 2r \text{ Cov } (x,y) \right]$$

where

$$V(y) = \sum_{h=1}^{H} \sum_{\alpha=1}^{a_h} (y'_{h\alpha} - \bar{y}'_h)^2,$$

$$V(x) = \sum_{h=1}^{H} \sum_{\alpha=1}^{a_h} (x'_{h\alpha} - \bar{x}'_h)^2$$

and

$$\text{Cov}(x,y) = \sum_{h=1}^{H} \sum_{\alpha=1}^{a_h} (x'_{h\alpha} - \bar{x}'_h)(y'_{h\alpha} - \bar{y}'_h)$$

Also,

$$x'_{h\alpha} = w_h x_{h\alpha},$$

$$y'_{h\alpha} = w_h y_{h\alpha}$$

and

$$\bar{y}'_h = \frac{1}{a_h} \sum_{\alpha=1}^{a_h} y'_{h\alpha},$$

$$\bar{x}'_h = \frac{1}{a_h} \sum_{\alpha=1}^{a_h} x'_{h\alpha}$$

Finally the standard error of r is given by

$$SE(r) = \sqrt{Var(r)}$$

see Leslie Kish, Survey Sampling, New York, 1965, chap. 6: "unequal clusters"

24 - Statistics of the Population of the United States at the 10th Census, op. cit.: 536-541, 420, 876

25 - See Olivier Zunz "Detroit en 1880 : espace et ségrégation" Center for Research on Social Organization of The University of Michigan, Working Paper # 121, August 1975. Forthcoming in Annales E.S.C.