William A. Ericson
Statistical Research Laboratory

Daniel J. Fox
Statistical Research Laboratory

Olivier Zunz
Michigan Society of Fellows

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

Methodological Report #1
of the Collective Action group

and Working Paper #114
of the Center for Research on
Social Organization

September 1974

Copies Available Through:

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

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

The sample described in this paper is the first of a series of samples to be drawn in the future for the study of comparative settlements in the 19th century Atlantic world (Europe and the United States). The study underway, of which the major theme is the urban integration of immigrants from Europe to the U.S., will ultimately compare American and European environments.

The sample for Detroit in 1880 is a first attempt to define a flexible unit of analysis, possibly to be used in other cities or at other points in time. A slightly revised version of this scheme is actually applied for an expansion of the study of Detroit in 1900 and will soon be extended to Baltimore for 1880 and 1900.

Olivier Zunz
Sept. 1974

The two main sources of information for the study are

a) The manuscript schedules of the 1880 U.S. Census for Detroit which give detailed information at the individual level for every inhabitant recorded

b) E. Robinson's atlas of the city of Detroit (1885) which gives detailed information at the block level (parcels, houses and other built structures)

Purpose of the sample

a) to insure a good representation of the population of Detroit in 1880

1) in all areas of the city (without the suburbs which constitute a separate problem to be dealt with in a separate sample)

2) taking into account differences in population density

b) to insure a good representation of the land use pattern in Detroit

Unit of Analysis

Emphasis has been given on the definition of the unit of analysis

a) to study a multiplicity of environments within a given city

b) to have a flexible and relatively small unit of analysis in order to study family life as well as patterns of neighboring

c) but a unit sufficiently large to include a relevant number of people and different types of land use

The unit of analysis which has been chosen can be divided into three independent units which correspond to three different levels of measurement

1 the block front level

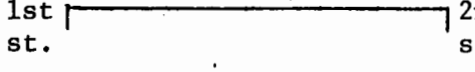
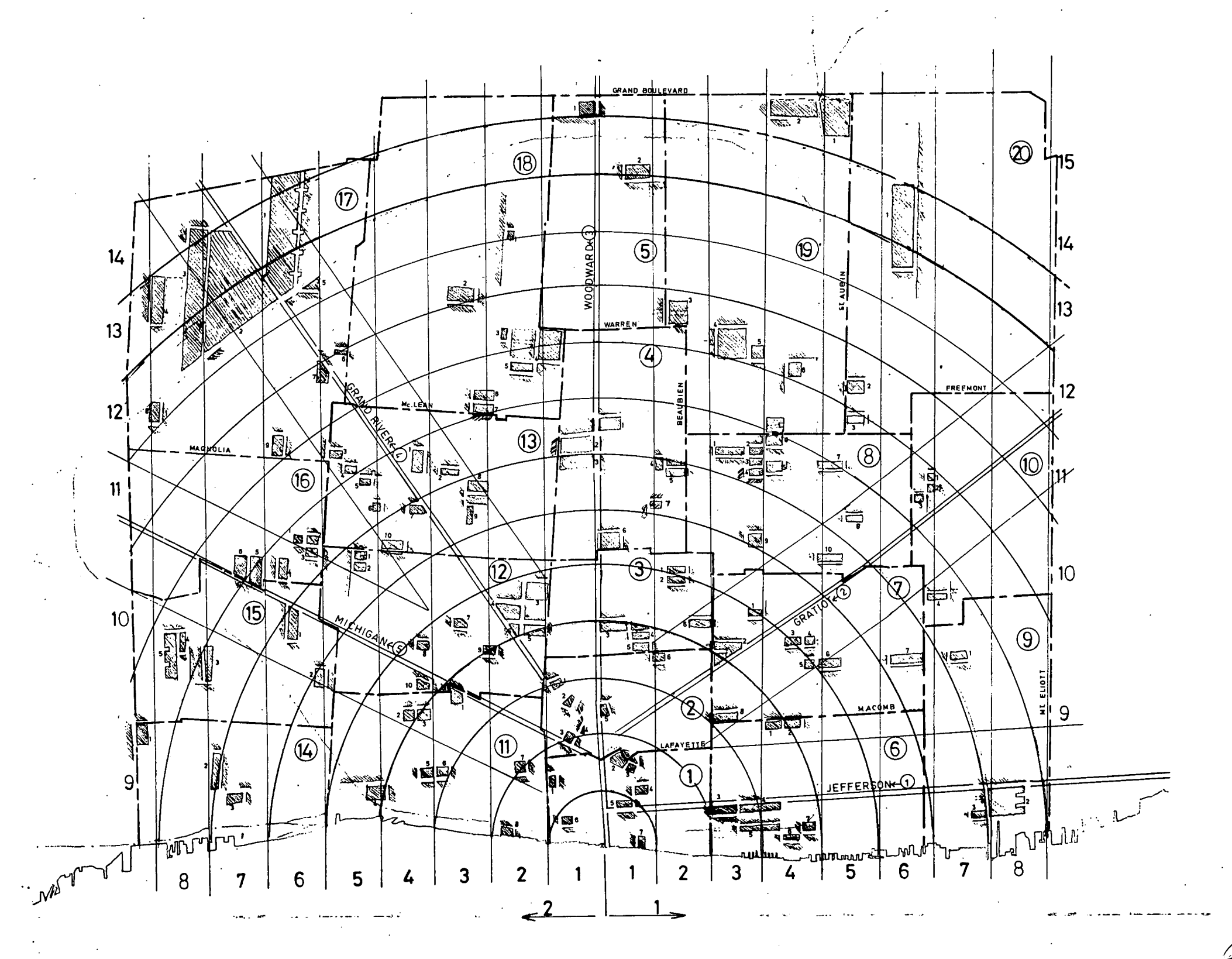


fig. 1


GRAND BOULEVARD
WOODWARD
WARREN
ST AUBIN
FREEMONT
BEAUBIEN
McLEAN
GRAND RIVER
MAGNOLIA
MICHIGAN
GRATIOT
MACOMB
LAFAYETTE
JEFFERSON
MT ELLIOTT
14
13
12
11
10
9
15
14
13
12
11
10
9
8 7 6 5 4 3 2 1 1 2 3 4 5 6 7 8
2
1

All households and persons living on that front recorded in the census as well as all characteristics of that front recorded in the atlas

2 the block level

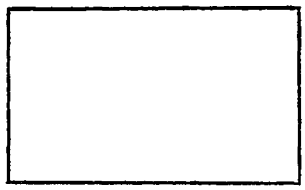


fig. 2

3 the small neighborhood level: 1 block + 2 opposite block fronts.

One major goal of the study was to examine the characteristics of small neighborhoods. These neighborhoods were defined as a city block--all households and persons in the census living on that block--augmented by the households and persons living on several block frontages facing opposite one or more frontages on the block. A typical such neighborhood might be described as shown in the diagram below.

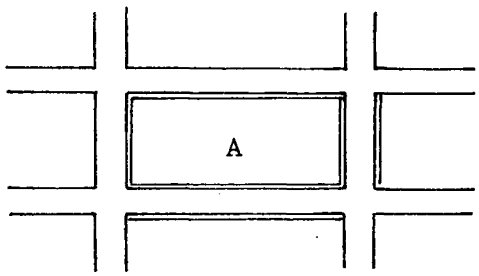


fig. 3

Here if Block A was in the sample all individuals living on its four frontages would be included, as well as, for example, those living across the street opposite two of the block frontages. Thus everyone on the six frontages indicated in double lines would be included in the sample.

The use of these three levels in sampling, data gathering and analysis not only allow to reconstruct several urban forms, e.g., a block, an avenue, a street from block frontages, but to measure the same phenomenon at different scale. For example, it is interesting to see if residential segregation exists at the front level or at the block level or at the small neighborhood level.

Problems of discrepancy between census and atlas

Theoretically and practically, it is possible to locate in the census manuscripts a good number of the inhabitants of the houses drawn in the atlas. The final sample counts 721 block fronts (353 inhabited) and 11,783 people.

a) The census had been taken in 1880 and the atlas published in 1885

20 plates of the Robinson's atlas cover the city of Detroit. In each block, parcels and built structures are designed. Houses are numbered and the atlas contains a good number of indications of non-residential buildings. Whenever a house in the atlas is not found in the census, it might be that this house had been built between 1880 and 1885. This is rarely the case, however, because it has taken a few years to collect information for such a detailed atlas in order to design and publish it. But some new construction or destruction surely happened.

b) Reordering the census is necessary to link atlas and census

The major difficulty of the sampling procedure is that when selecting block fronts, we cannot assume they will be found in the census in the same order as in the atlas. Census takers cross streets, go from even to odd numbers, visit parts of street and revisit the other parts later. This means that a thorough search in the census of the houses chosen in the atlas is necessary to collect the information. The census manuscript must be entirely reordered to collect information. This is possible with the 1880 census manuscripts because they list addresses in cities and thus--with patience--one can reorder information per street to draw a geographical sample.

Aware of these difficulties, we designed the sample.

SAMPLE DESIGN

1. Atlas-Block Description

The city was divided into twenty geographic regions corresponding to the 20 individual plates in this atlas. The atlas contained sufficient detail that each block could be categorized into one of four classes.

P = Promising (blocks containing almost all dwelling structures. 3/4 downtown, 2/3 uptown)

ND = Non-Dwelling (blocks mainly occupied by commercial establishments, etc. and not likely many or any dwelling units. 3/4 downtown, 2/3 uptown)

V = Vacant (blocks containing no structures) 90% vacant

O = Other (blocks containing a mixture of P & ND, P & V, ND & V, 1/3, 2/3 or 1/2, 1/2

The results of this categorization are shown in Table 1.

Since blocks differ in structure--number of frontages, number of inhabitable frontages, etc.--it was decided to look at a pilot sample of blocks. Two independent pilot samples were drawn--one comprising every fourth block on each of the twenty atlas plates and the other a sample from the census file.

2. 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 blocks; odd numbered plates: every other odd numbered blocks) and for each block the number of block frontages and the number of non-enumerable frontages was determined. A frontage was classed as non-enumerable if it was impossible for a census taker to visit it. Thus, it is impossible to find it in the manuscript of the census. It is the case of either totally empty fronts or when houses face other streets or when the front is not residential.

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	

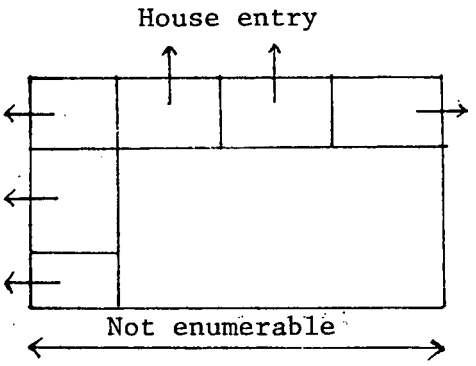


fig. 4

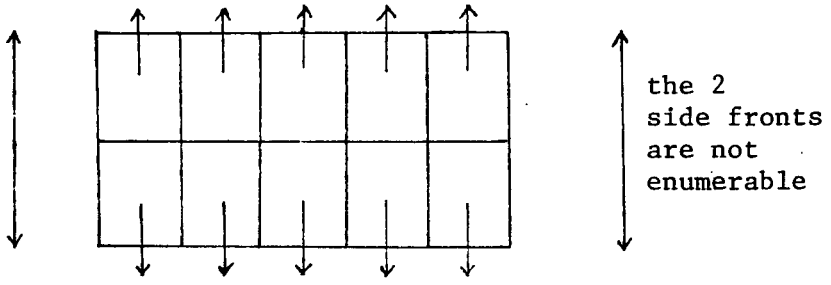


fig. 5

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 block are shown in Table 2.

3. Census Pilot Sample:

In order to get some idea of the size of frontages in terms of numbers of households and persons per frontage as well as determining the difficulty in tracing frontages in the census data, a sample of 87 frontages was selected. This sample was selected by recording information in the census on every 80th street: microfilm reel number, census page, street name, houses numbers, number of households, number of people in each household.

Every street recorded in this manner was then traced 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 (fig. 6). For some streets of

TABLE 2

Pilot Block Sample

Mean Numbers of Enumerable Frontages Per Block

By Typle 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

the random sample, we actually got several block fronts of the atlas

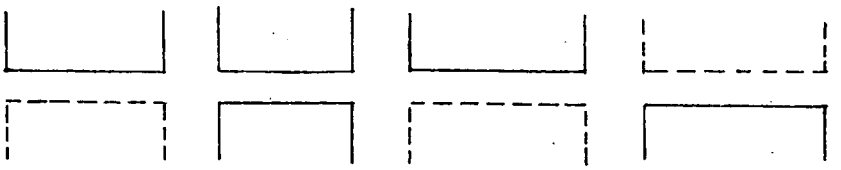


fig. 6

In other cases, under one street name, the random sample in the census mentioned only one or two houses for a given front when the atlas mentioned 15 or 20.

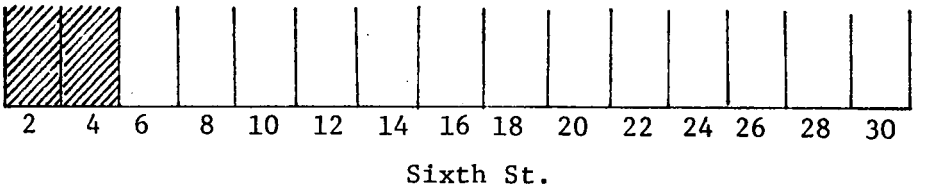


fig. 7

Despite this intriguing variablity we recorded for every street in the random sample:

street name

block number and stratum in the atlas

# of dwelling structures in the atlas per front

number of houses in the census per front

number of people in the census per front

Some of the data from this sample are displayed in Table 3 which gives the atlas plate number, street name, the stratum characterizing the sampled frontages, the number of frontages sampled and the total number of persons found in the census for those frontages.

This sample provided little or no information about blocks in the O, V, and ND strata or for plates 5, 14, 17, and 19. Also since single streets were traced and it was expected that adjacent frontages along the same street would be similar; it was decided to compute the mean number of persons per frontage per street run for each plate. This meant that the

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

data above were reduced to 33 "independent" observations. The results of this condensation are shown in Table 4. These data combine the P and O type blocks and ignore the one ND frontage in Plate 1.

On the basis of these pilot samples and the atlas information it seemed that very little of the population was contained in blocks categorized V or ND.

4. Some Analysis of Pilot Sample Data:

We were interested in seeing how the sample data above could do in estimating the total census population known to be 116,340.* One simple estimate was the following:

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

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. We also wanted to estimate the standard error of the estimate and, since there were so few observations on the mean numbers of persons per frontage for many of the plates--it was decided to lump similar plates. Formula (1) was then applied to the eight resulting combinations of plates. The data are shown in Table 5.

These results lead to 83,088 as an estimate of the total population size--an underestimate by some 116,340 - 83,088 = 33,252. There are a fair number of factors contributing to this--ignoring the ND and V blocks, the small sample used in estimating or computing $\bar{X}_i$ (ranging between 3 and 6). Furthermore the random sample of streets in the census was misleading. Many streets appeared to be poorly recorded because information had been

*1880 published census

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

located in the census at only one point when, due to revisitation, the same fronts are very often recorded in several different parts of the census volumes.

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

$$\hat{V}(\hat{P}) = \sum_{i=1}^{8} N_i^2 (\bar{f}_i^2 s_{\bar{x}_i}^2 + \bar{x}_i^2 s_{\bar{f}_i}^2 + s_{\bar{x}_i}^2 s_{\bar{f}_i}^2), \qquad (2)$$

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

The application of formula (2) to the values shown in Tables 5 and 6 yielded an estimated standard deviation of $\hat{P}$ as 10,409. Thus the total population estimate 83,088 was 3.19 standard error below the true population value of 116,340.

OTHER ANALYSES OF PILOT SAMPLE DATA*

Another estimate of 107,782 for the population total or an estimate of .925 (= $\frac{107,782}{116,340}$) was made. The basic scheme used was to divide the sample into two parts

1. # houses in census = 1, 2, or 3 and # dwellings in atlas $\geq$ 4 + # houses
2. all others $\doteq$ MAJOR

and then to obtain an estimate based only upon the data from the MAJOR GROUP where most households in a front were enumerated on the first visit.

*Another presentation of this estimate is actually being studied by Daniel Fox using the new cluster command in MIDAS (Michigan Interactive Data Analysis System)

TABLE 5

Plate numbers	N_i # of Per 10 Blocks	$\bar{X}_i$ mean # persons/frontage	$\bar{f}_i$ mean # per enumerable frontages	$N_i \bar{X}_i \bar{f}_i$
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

5. THE MAIN SAMPLE DESIGN

Since the neighborhoods or clusters of block frontages were to form a major unit of analysis, it was decided to include at least 100 such neighborhoods. It was expected that very few, if any, such neighborhoods would result from sampling ND or V blocks, thus the 100 clusters would come from the P and O blocks. Such a number, it was felt would yield tolerably small standard errors for estimating neighborhood characteristics. Actually 102 such blocks were selected with their opposing frontages. As a check, and in order to estimate city characteristics 25 neighborhoods were selected from among the ND and V blocks. In order to obtain geographic representation the sample was selected stratifying by plate, independent samples were selected from each plate. The sampling was slightly more intensive among those plates seeming, on the basis of the pilot sample, to have a higher population concentration.

For the purpose of selecting the sample the 20 plates were grouped into three classes. One was that comprised of 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 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 groups of plates described above.

Table 7

Plate #	Average # Persons per Block	#P or 0 Blocks	Proportion N, D, or V Blocks	No. Envm. 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

It was decided to sample 12 blocks from among the 177 P and O blocks in Category 1 or 6.8% roughly; 40 of the 425 such blocks in Category 2 or 9.4% roughly and 50 of the 650 blocks in Category 3 or about 7.7%. These rates 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.

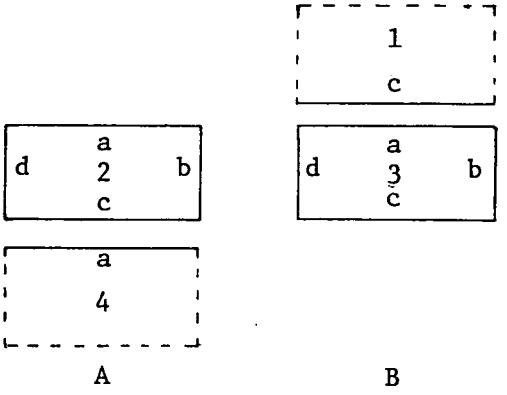
The final sample design--numbers of each type of block to include for each plate is shown in Table 8.

The actual mechanics of drawing the sample were quite straightforward given Table 8 and 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 using the Rand Table of random digits one of the (usually 4) block corners was randomly chosen and the (usually 2) opposing block frontages were then also included to make up the neighborhood. See Figure 3 on page 3; this figure illustrates the resulting sample neighborhood resulting from drawing Block A and then choosing the corner numbered 4.

This method creates some complications for the analysis since the same front can appear several times on different neighborhoods (fig. 8).

fig. 8



Front 2b is in neighborhoods A & B

Front 3d is also in neighborhoods A & B

Table 8

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

6. Estimated Sample Number of Households and Persons

Given the design above and, in fact as part of the choice of size of sample, it was of some interest to estimate the size of the sample which would be produced in terms of households and persons. This was needed in order to judge the magnitude of the task in recording the data from the census files.

In order to get a better estimate of the number of persons per frontage the plates were grouped into the three classes as used in the design. It was expected that the 25 V and ND blocks would contribute little to the number of persons found. In the analysis below these are ignored. One difficulty was in estimating the number of enumerable frontages associated with each sampled block because of the augmentation by opposing frontages--this also leads to the possibility of including some frontages several times as opposing frontages for different block selections. Thus the observed pilot sample number of enumerable frontages per block, $\bar{f}_i$, had to be adjusted upward to reflect the inclusion of several opposing frontages. These are the values $\bar{f}_i^*$ shown in Table 9. These values were quite subjective--in densely populated areas, like group 2, $\bar{f}_i^*$ might be as large as one and one-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 $\bar{f}_i^*$. These values, as subjectively guessed, are as shown.

The estimated sample number of persons is given using formula (1) replacing N_i by n_i and $\bar{f}_i$ by $\bar{f}_i^*$. This was computed as 8,930 persons. It was expected that there would be an average of 5 persons per household and thus we expected the sample to contain some 1,786 households. However, even making the assumption that the $\bar{f}_i^*$ are known without error, the

TABLE 9

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

standard error of the estimate of the sample number of persons is 961 or of the sample number of households is 192.

We did not take into consideration, for this estimate, the population estimate of 107,782 suggested by the analysis of the pilot sample data divided into two groups--Minor and Major. The final sample counts 11,783 people on 721 frontages. 353 frontages are actually inhabited. The others are non enumerable or non residential. The mean household size is exactly 5 (5.022), the mean number of people per front is 33.380 (min = 1; max = 152). Most houses located in the atlas have been found in the mansucript census. The rate of success for this search is extremely high. The computation of success and failure in the search will be done soon to evaluate the precision of the enumeration in the census.

W.A. Ericson
D.J. Fox
O. Zunz