

DATA TAPES



SOCIAL SCIENCE DATA ARCHIVE / Survey Research Laboratory / University of Illinois

No. 4

September, 1972

Fourth Count Population and Housing Data Now Available

The Fourth Count data from the 1970 Census, for which so many of you have been waiting, are now *finally* available from the Social Science Data Archive (SSDA) of the University's Survey Research Laboratory. A complete computer retrieval system has been developed for operating on the population data, and special purpose programs will be written as required to retrieve the housing data.

The Fourth Count contains 30 population variables and 35 housing variables. A list of these items is presented in Appendix A. Unlike the first three counts, which were complete count or 100-percent data, the Fourth Count data are weighted sample data based on 15- and 5-percent samples. These two samples have several items in common that can be used to generate a 20-percent sample. In general, items from the 100-percent sample are also included for cross-classification purposes.

The population and housing data can be obtained for (a) census tracts, (b) minor civil divisions (townships) and census county divisions, and (c) the following selected census areas:

1. State
2. Total urbanized areas (UAs) within the state
3. Central cities of UAs
4. Urban fringe of UAs
5. Urban places with 10,000 or more population outside of UAs
6. Urban places with 2,500-10,000 population outside UAs
7. Total standard metropolitan statistical areas (SMSAs) in state
8. Central cities of SMSAs
9. Other territory of SMSAs
10. Non-SMSA territory in state
11. Standard consolidated areas (SCAs)
12. SMSAs
13. Places with 2,500 or more population
14. Counties

It should be remembered that the computer tapes contain more data items (cross-tabulations) for tracts and other small areas reported than are found in the printed reports.

The tabulations and cross-tabulations contained in the Fourth Count population data, arranged by actual tabulation number, are listed in Appendix B. The user can request any or all of the population tabulations for any of the census geographic divisions listed above.

The user can also obtain up to 200 tabulations and cross-tabulations on housing data. The user interested in housing data should contact SSDA to obtain a list of these tabulations and make arrangements to have a program written to retrieve the data he requires.

Fourth Count Workshop

SSDA plans to offer one Thursday evening workshop on each campus (Urbana-Champaign and Chicago Circle) in October for potential users of the Fourth Count data. In planning for the workshops, we need an indication of how much interest there would be. Therefore, if you are interested in attending a Fourth Count workshop, please fill out the form on the last page of this newsletter. You will be notified later of the exact time and place.

Also, please use the same sheet if you would like to continue receiving DATA TAPES on a regular basis. This information will help us formalize our mailing list so that it is more accurate and complete.

No News on Fifth and Sixth Counts

For those of you who are interested in the Fifth or Sixth Count summary tapes, thus far we have received no definite word on their availability.

SPSS Files Based on Fourth Count Data

SSDA is developing the capability of creating Statistical Package for the Social Sciences (SPSS) files from the population variables of the Fourth Count census data. These will be small, special-purpose files constructed on selected variables for selected geographic areas and will be extremely useful for student projects, classroom presentations, and limited research

problems. The files will be constructed so that the user can operate on Fourth Count data that has been placed in a format acceptable to SPSS. The SPSS files will broaden the type of analyses that the user can employ on Fourth Count data and will relieve the user from much of the computational burden of the statistical techniques he selects.

Address Matching with Census Data

Historically, many researchers have been concerned with macroscopic issues and have dealt with large aggregates of individuals on several large geographic levels (the nation, the region, the state, the region within state, SMSA, etc.). Relatively few have done detailed analyses on small areas (city blocks, block groups, census tracts, etc.). The 1970 Census has several interesting innovations that can make small-area analysis a viable research strategy, in particular, "geocoding with the ADMATCH system." This system allows the researcher to create *user-defined* geographic areas of analysis by matching his addressed data records against a geographic reference file.

The ADMATCH system is used to add selected geographic identification codes to a user data file. This is done by matching the address of a record in the user input data file with a reference data file that has been specially prepared to coordinate street addresses with other useful geographic identifications (census tract code, block group code, etc.). For the most part, the ADMATCH system will operate with either census geographic reference file--the Dual Independent Map Encoding (DIME) file or the Address Coding Guide (ACG). One of these geographic reference files is available for every SMSA in the state.

The ADMATCH system consists of three program phases: preprocessing, matching, and postprocessing.

The preprocessing phase prepares the user file in the fixed format required by the matching phase. To facilitate this operation, two programs are available. The first, the Preprocessor program, interprets the user data file by referencing a series of tables and the address portion of the user data file to produce a 75-character "match key." The tables used by the Preprocessor program allow for alternate spellings of all parts of the address field (including street names, directions, and city and state names). Both the user data file and the reference file must be preprocessed to create this 75-character "match key." With careful modification of the pattern-recognition tables, extraneous information in the data records can be deleted.

The Preprocessor program has also been written to accept house address ranges and fractional addresses. Apartment designations can also be processed. In the case of an address range, the record is interpreted as having a single address--that of the lower bound of the address range. Fractional portions of the address number are truncated. Output from the Preprocessor program consists of a file of accepted records and a file of rejected records.

The second program available for the preprocessing phase is a tape dump program, which allows one to list those records that could not be interpreted. Since the reference file and the data file can be processed in sections, it is not necessary to reprocess the entire file to make corrections and get a higher rate of successful interpretation. One would simply either change the requirements for interpretation of the input file or make corrections in the rejected portion of the file, and then reprocess the file of rejected records.

After both the reference data file and the user input data file have

been preprocessed so that all addresses are in an acceptable format, both files are sorted. The two files must be sorted on exactly the same criterion for the Matcher program to be successful. (Since this program is not designed to read either file backwards, all matches must be made in one pass of both files.)

The sorted records are then processed by the Matcher program. This program operates solely on the 75 characters of geographic identification that were appended by the Preprocessor program. When a match has been made, the program transfers information from the reference data file onto the input data file.

Normally, for a match to occur the Matcher program requires that all address components agree. This is a 100-percent match. However, it is possible to specify a number of different degrees of accuracy in the match level. Thus, a match level of 99 percent would allow records to be matched if there was one blank component. For example, Sheridan would be accepted from the input data file as a street name and could be matched with Sheridan Road. Or, Sheridan Place could be accepted if the data file contained a "match key" for Sheridan. A 99-percent match will not accept a match if there is a discrepancy between two present fields, just if there is one missing field. On the other hand, a 98-percent match level will accept either two blank fields in either file or one disagreement. In this case, a Sheridan Place could be accepted as a Sheridan Road. In the same manner, a 97-percent match level will accept either three blank fields or one disagreement and one blank field.

After the user-defined areas have been geographically coded with the relevant census identifiers, the postprocessing phase provides census data for the user areas and a statistical summary report to describe the social,

demographic, and housing characteristics of each area. For example, a user-defined area may consist of 300 city blocks. After these block numbers have been appended to the user's list of addresses, the user can summarize the data for all 300 blocks to produce a statistical report for his area.

Data Available for Secondary Analysis

SSDA now has 110 data collections available for secondary analysis. These data are from studies conducted by other survey research organizations and by SRL. All studies are being placed on tape (a superior storage medium) and cards will serve only an intermediary function. All studies have complete documentation including detailed codebooks and cumulative and marginal frequency distributions. Sampling and field reports are also available for the SRL studies.

The reorganized archive is planning to increase its collection of data holdings in the near future. An order has already been placed with the Roper Public Opinion Research Center to participate in the National Data Program for the Social Sciences. This program will provide the archive with standardized national level data on important social science variables at yearly intervals so that over a period of five or ten years there will be a basis for studying social trends. The yearly national surveys, which are conducted by the National Opinion Re-

search Center, cover such attitudinal areas as race relations, social control, occupational satisfaction, and sex roles, and include the usual demographics (age, sex, race, education, income, ethnicity, siblings, political party preference, religion, etc.).

SSDA's current activities are only the beginning of a program to increase and broaden the holdings of the archive. If any readers have suggestions concerning data holdings that the archive should and might acquire, please write or call the archive and we shall investigate the matter.

Inquiries

Any inquiries regarding census data acquisition or other data holdings that SSDA has or might obtain should be directed to one of the following persons:

Daniel Amick
Survey Research Laboratory
Box 6905
University of Illinois at Chicago
Circle
Chicago, Illinois 60680
Phone: (312) 996-5304

or

Noel Uri
Survey Research Laboratory
Room 11, David Kinley Hall
University of Illinois at Urbana-
Champaign
Urbana, Illinois 61801
Phone: (217) 333-6572

APPENDIX A

Subject Items Included in the 1970 Census

(Percents indicate size of sample for which the question is asked; 100 percent is complete coverage)

<u>Population items</u>	
100 percent	Relationship to head of household Color or race Age Sex Marital status
20 percent	State or country of birth Years of school completed Number of children ever born Employment status Hours worked last week Weeks worked in 1969 Last year in which worked Occupation, industry, and class of worker Activity 5 years ago Income in 1969
15 percent	Country of birth of parents Mother tongue Year moved into this house Place of residence 5 years ago School or college enrollment (public or private) Veteran status Place of work Means of transportation to work
5 percent	Mexican or Spanish origin or descent Citizenship Year of immigration When married Vocational training completed Presence and duration of disability Occupation-industry 5 years ago
<u>Housing items</u>	
100 percent	Number of units at this address Telephone Private entrance to living quarters Complete kitchen facilities Rooms Water supply Flush toilet Bathtub or shower Basement Tenure Commercial establishment on property Value Contract rent Vacancy status Months vacant
20 percent	Components of gross rent Heating equipment Year structure built Number of units in structure and whether a trailer
15 percent	Farm residence Source of water Sewage disposal Bathrooms Air conditioning Automobiles
5 percent	Stories, elevator in structure Fuel--heating, cooking, water heating Bedrooms Clothes washing machine Clothes dryer Dishwasher Home food freezer Television Radio Second home

APPENDIX B

POPULATION TABULATIONS AND CROSS-TABULATIONS IN THE FOURTH COUNT*

Tabulation Number	
1	AGGREGATE \$ FAMILY INCOME OF FAMILIES
2	AGGREGATE \$ FAMILY INCOME OF FAMILIES WITH FEMALE HEAD
3	AGGREGATE \$ INCOME OF UNRELATED INDIVIDUALS 14 YEARS OLD AND OVER
4	AGGREGATE \$ INCOME OF FEMALE UNRELATED INDIVIDUALS
5	AGGREGATE \$ INCOME OF POPULATION 14 YEARS OLD AND OVER BY SEX
6	AGGREGATE \$ EARNINGS OF MALES 16 YEARS OLD AND OVER IN THE EXPERIENCED CIVILIAN LABOR FORCE BY SELECTED OCCUPATIONS
7	AGGREGATE \$ EARNINGS OF FEMALES 16 YEARS OLD AND OVER IN THE EXPERIENCED CIVILIAN LABOR FORCE BY SELECTED OCCUPATIONS
8	AGGREGATE \$ INCOME BY TYPE FOR FAMILIES AND UNRELATED INDIVIDUALS
9	AGGREGATE \$ INCOME BY TYPE AND SEX
10	AGGREGATE \$ INCOME DEFICIT FOR FAMILIES BELOW POVERTY LEVEL
11	AGGREGATE \$ INCOME DEFICIT FOR UNRELATED INDIVIDUALS 14 YEARS OLD AND OVER BELOW POVERTY LEVEL
12	AGGREGATE \$ INCOME OF SPECIFIED TYPES RECEIVED BY FAMILIES BELOW POVERTY LEVEL
13	AGGREGATE \$ INCOME OF SPECIFIED TYPES RECEIVED BY UNRELATED INDIVIDUALS BELOW POVERTY LEVEL
14	AGGREGATE \$ VALUE OF OWNER-OCCUPIED UNITS BY POVERTY STATUS OF FAMILY OR PRIMARY INDIVIDUAL
15	AGGREGATE \$ GROSS RENT OF RENTER-OCCUPIED UNITS BY POVERTY STATUS OF FAMILY OR PRIMARY INDIVIDUAL
16	AGGREGATE \$ INCOME DEFICIT BETWEEN SPECIFIED POVERTY LEVEL AND TOTAL INCOME FOR FAMILIES AND UNRELATED INDIVIDUALS
17	AGE AND SEX
18	RELATIONSHIP AND SEX
19	FAMILIES BY TYPE, PRESENCE AND AGE OF OWN CHILDREN
20	POPULATION 14 YEARS OLD AND OVER BY MARITAL STATUS AND SEX
21	NATIVITY AND PARENTAGE (15%)
22	COUNTRY OF ORIGIN AND NATIVITY (15%)
23	MOTHER TONGUE AND NATIVITY (15%)
24	SPANISH INDICATORS
25	CITIZENSHIP AND AGE (5%)
26	YEAR OF IMMIGRATION FOR THE FOREIGN-BORN POPULATION (5%)
27	STATE OF BIRTH FOR THE NATIVE POPULATION
28	POPULATION 5 YEARS OLD AND OVER BY RESIDENCE IN 1965 (15%)
29	POPULATION 5 YEARS OLD AND OVER LIVING IN SMSA'S OR ADJACENT TRACTS BY RESIDENCE IN 1965 (15%)
30	POPULATION LIVING IN A DIFFERENT COUNTY BY RESIDENCE AND ACTIVITY STATUS IN 1965 (15%)
31	POPULATION IN ARMED FORCES IN 1970 AND LIVING IN A DIFFERENT COUNTY BY RESIDENCE IN 1965 (15%)

*All tabulations not otherwise indicated represent 20% sample. Those indicated as 20% tabulations for Spanish Americans are tabulated on a 15% basis.

Tabulation
Number

32	INMATES OF INSTITUTIONS IN 1970 AND LIVING IN A DIFFERENT COUNTY BY RESIDENCE IN 1965 (15%)
33	POPULATION ATTENDING COLLEGE IN 1970 AND LIVING IN A DIFFERENT COUNTY BY RESIDENCE IN 1965 AND RESIDENCE WITH PARENTS IN 1970 (15%)
34	YEAR MOVED INTO UNIT (15%)
35	POPULATION AT WORK DURING THE CENSUS WEEK BY PLACE OF WORK (15%)
36	POPULATION AT WORK DURING THE CENSUS WEEK BY MEANS OF TRANSPORTATION TO WORK (15%)
37	CIVILIAN MALES 16 YEARS OLD AND OVER BY VETERAN STATUS (15%)
38	POPULATION 3 YEARS OLD AND OVER ENROLLED IN SCHOOL BY LEVEL AND TYPE OF SCHOOL (15%)
39	POPULATION 3-34 YEARS OLD ENROLLED IN SCHOOL BY AGE (15%)
40	POPULATION 18-24 YEARS OLD BY COMPLETION OF HIGH SCHOOL AND COLLEGE
41	POPULATION 16-21 YEARS OLD NOT ENROLLED IN SCHOOL BY YEARS OF SCHOOL COMPLETED, EMPLOYMENT STATUS, AND SEX
42	POPULATION 25 YEARS OLD AND OVER BY YEARS OF SCHOOL COMPLETED AND SEX
43	POPULATION 16-64 YEARS OLD WITH LESS THAN 3 YEARS OF COLLEGE COMPLETED BY VOCATIONAL TRAINING AND SEX (5%)
44	FEMALES 15-44 YEARS OLD BY MARITAL STATUS, NUMBER OF CHILDREN AND AGE
45	FAMILIES BY PRESENCE OF SONS/DAUGHTERS OF SELECTED AGE LEVELS
46	SUBFAMILIES BY SUBFAMILY TYPE AND NUMBER OF OWN CHILDREN UNDER 18 YEARS OLD
47	MARRIED COUPLES BY SUBFAMILY MEMBERSHIP
48	MARRIED COUPLES BY PRESENCE AND AGE OF OWN CHILDREN AND AGE OF HUSBAND
49	POPULATION UNDER 18 YEARS OLD BY RESIDENCE WITH PARENTS
50	EVER MARRIED POPULATION 14-54 YEARS OLD BY MARITAL HISTORY (5%)
51	POPULATION IN GROUP QUARTERS BY TYPE OF GROUP QUARTERS
52	POPULATION 16-64 YEARS OLD NOT INMATES AND NOT ATTENDING SCHOOL BY DISABILITY, EMPLOYMENT STATUS, AND SEX (5%)
53	MALES 20-49 YEARS OLD AND FEMALES 15-44 YEARS OLD BY YEARS OF SCHOOL COMPLETED
54	POPULATION 16 YEARS OLD AND OVER BY LABOR FORCE STATUS, SELECTED CHARACTERISTICS, AND SEX
55	POPULATION 16 YEARS OLD AND OVER IN THE LABOR FORCE BY AGE AND SEX
56	POPULATION 14-15 YEARS OLD BY LABOR FORCE STATUS AND SEX
57	PRESENCE AND AGE OF OWN CHILDREN UNDER 18 AND LABOR FORCE STATUS OF FEMALES 16 YEARS OLD AND OVER BY MARITAL STATUS
58	EMPLOYED POPULATION 16 YEARS OLD AND OVER BY OCCUPATION
59	EMPLOYED FEMALES 16 YEARS OLD AND OVER BY OCCUPATION
60	EMPLOYED MALES 14-15 YEARS OLD BY OCCUPATION
61	EMPLOYED FEMALES 14-15 YEARS OLD BY OCCUPATION
62	EMPLOYED POPULATION 16 YEARS OLD AND OVER BY INDUSTRY AND SEX
63	EMPLOYED POPULATION 14-15 YEARS OLD BY INDUSTRY AND SEX
64	MALES 16 YEARS OLD AND OVER BY WEEKS WORKED IN 1969 AND AGE
65	FEMALES 16 YEARS OLD AND OVER BY WEEKS WORKED IN 1969 AND AGE
66	POPULATION 14-15 YEARS OLD BY WEEKS WORKED IN 1969

Tabulation

Number

67	EMPLOYED POPULATION 16 YEARS OLD AND OVER BY INDUSTRY, CLASS OF WORKER, AND SEX
68	EMPLOYED POPULATION 14 YEARS OLD AND OVER BY OCCUPATION AND SEX
69	EMPLOYED POPULATION 14 YEARS OLD AND OVER BY INDUSTRY AND SEX
70	EXPERIENCED UNEMPLOYED POPULATION 16 YEARS OLD AND OVER BY OCCUPATION AND SEX
71	POPULATION 14-15 YEARS OLD WHO WORKED IN 1969 BY OCCUPATION AND SEX
72	MALES 30-49 YEARS OLD BY WORK STATUS IN 1965 AND 1970
73	FEMALES 21-49 YEARS OLD WITH NO OWN CHILDREN 5-10 YEARS OLD BY WORK STATUS IN 1965, IN 1970, AND PRESENCE OF OWN CHILDREN UNDER 5 YEARS
74	FEMALES 21-49 YEARS OLD WITH OWN CHILDREN 6-10 YEARS OLD AND NO CHILDREN UNDER 6
75	FAMILY INCOME
76	INCOME OF UNRELATED INDIVIDUALS
77	POPULATION 14 YEARS OLD AND OVER BY INCOME AND SEX
78	MALES 16 YEARS OLD AND OVER IN THE EXPERIENCED CIVILIAN LABOR FORCE BY EARNINGS AND OCCUPATION
79	FEMALES 16 YEARS OLD AND OVER IN THE EXPERIENCED CIVILIAN LABOR FORCE BY EARNINGS AND OCCUPATION
80	TYPE OF INCOME AND FAMILY STATUS
81	TYPE OF INCOME AND SEX
82	RATIO OF FAMILY INCOME TO POVERTY LEVEL
83	AGGREGATE NUMBER OF PERSONS IN FAMILIES BELOW POVERTY LEVEL
84	FAMILIES BY PRESENCE OF RELATED CHILDREN UNDER 18, TYPE OF FAMILY, AND POVERTY STATUS
85	RELATED CHILDREN UNDER 18 BY TYPE OF FAMILY, POVERTY STATUS, AND AGE
86	FAMILY HEADS 65 YEARS OLD AND OVER BY POVERTY STATUS
87	CIVILIAN MALE FAMILY HEADS 14-64 YEARS OLD BY LABOR FORCE AND POVERTY STATUS
88	FEMALE FAMILY HEADS IN THE LABOR FORCE WITH RELATED CHILDREN UNDER 6 YEARS OLD BY POVERTY STATUS
89	UNRELATED INDIVIDUALS BY AGE AND POVERTY STATUS
90	POPULATION 65 YEARS OLD AND OVER BY POVERTY STATUS
91	RELATED CHILDREN 5-17 YEARS OLD IN FAMILIES WITH INCOMES BELOW \$3,000
92	RELATED CHILDREN UNDER 18 YEARS OLD IN FAMILIES BELOW POVERTY LEVEL BY PRESENCE OF PARENTS
93	POPULATION RECEIVING SOCIAL SECURITY OR RAILROAD RETIREMENT BY AGE AND POVERTY STATUS
94	NUMBER OF FAMILIES BELOW POVERTY LEVEL RECEIVING INCOME OF SELECTED TYPES
95	NUMBER OF UNRELATED INDIVIDUALS BELOW POVERTY LEVEL RECEIVING INCOME OF SELECTED TYPES
96	FAMILIES AND PRIMARY INDIVIDUALS IN HOUSING UNITS BY TENURE AND POVERTY STATUS
97	FAMILIES AND PRIMARY INDIVIDUALS IN HOUSING UNITS LACKING ONE OR MORE PLUMBING FACILITIES BY POVERTY STATUS
98	SPECIFIED POVERTY LEVEL, TYPE OF FAMILY, AND SELECTED FAMILY STATUS
99	SPECIFIED POVERTY LEVEL AND AGE

Tabulation
Number

100	COUNT OF PERSONS SUBSTITUTED PLUS PERSONS WITH LESS THAN TWO SAMPLE CHARACTERISTICS REPORTED
101	POPULATION ALLOCATIONS AND SUBSTITUTIONS
102	COUNT OF HOUSEHOLDS SUBSTITUTED
103	FARM RESIDENCE
104	SEX
105	RACE
106	AGE
107	NATIVITY
108	NATIVITY OF PARENTS (15%)
109	YEAR MOVED INTO UNIT (15%)
110	RELATIONSHIP
111	POPULATION 14 YEARS OLD AND OVER BY MARITAL STATUS
112	POPULATION 3-34 YEARS OLD BY ENROLLMENT STATUS AND AGE (15%)
113	POPULATION 3-34 YEARS OLD ENROLLED IN SCHOOL BY LEVEL OF SCHOOL (15%)
114	POPULATION 25 YEARS OLD AND OVER BY YEARS OF SCHOOL COMPLETED
115	POPULATION 16 YEARS OLD AND OVER BY EMPLOYMENT STATUS
116	EMPLOYED POPULATION 16 YEARS OLD AND OVER BY OCCUPATION
117	EMPLOYED POPULATION 16 YEARS OLD AND OVER BY INDUSTRY
118	POPULATION 16 YEARS OLD AND OVER BY WEEKS WORKED IN 1969
119	POPULATION 16 YEARS OLD AND OVER BY CLASS OF WORKER
120	POPULATION 14 YEARS OLD AND OVER BY INCOME AND SEX
121	FAMILIES AND UNRELATED INDIVIDUALS BY INCOME
122	POPULATION 14 YEARS OLD AND OVER BY INCOME STATUS, TYPE OF INCOME, AND SEX
123	POPULATION 14 YEARS OLD AND OVER BY AGGREGATE INCOME AND SEX
124	UNWEIGHTED SAMPLE POPULATION COUNT
125	UNWEIGHTED SAMPLE HOUSING COUNT
126	100% POPULATION COUNT
127	100% HOUSING COUNT

- I am interested in attending a Fourth Count workshop
- on the Urbana-Champaign campus
 - on the Chicago Circle campus

- I would like to receive DATA TAPES on a regular basis

NAME:

ADDRESS:

CITY:

Return to:

Mrs. Mary A. Spaeth, Editor
DATA TAPES
Survey Research Laboratory
414 David Kinley Hall
University of Illinois
Urbana, Illinois 61801