

**The 1960 Data Restoration Project**

**Kathryn M. Coursolle**

Minnesota Population Center, University of Minnesota

**Rebecca Vick**

Minnesota Population Center, University of Minnesota

**Ryan S. Klein**

Minnesota Population Center, University of Minnesota  
Division of Environmental Health Sciences, University of Minnesota

September 26, 2014

This paper was prepared for submission to the annual meeting of the Population Association of America. Research for this project was supported by the National Institutes of Health (grant number R01HD043392). Opinions expressed herein are those of the authors. Please do not cite or quote without permission of the authors. Address correspondence to Kathryn Coursolle, Minnesota Population Center (MPC), University of Minnesota, 50 Willey Hall, 225 19<sup>th</sup> Avenue South, Minneapolis, MN 55455, e-mail: kathrync@umn.edu.

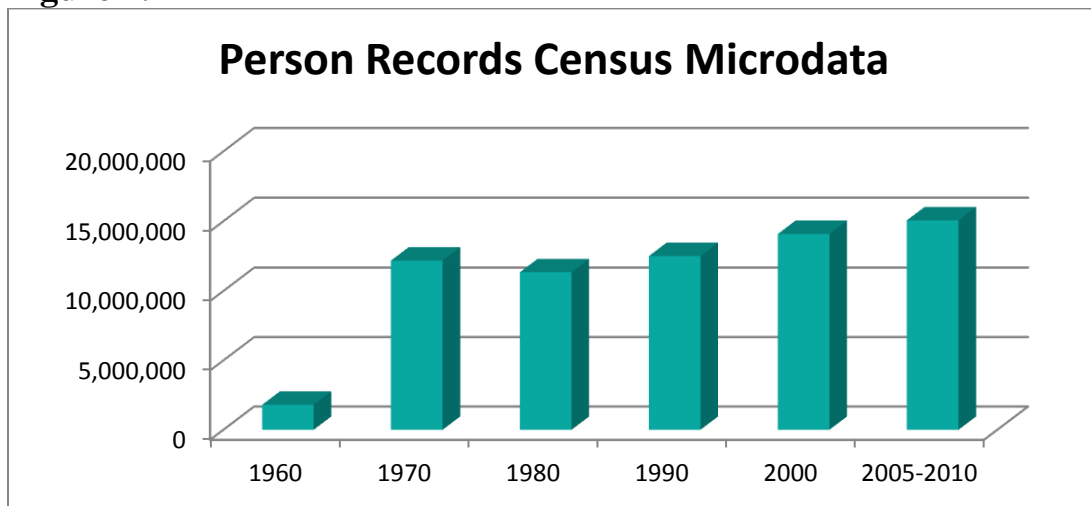
### **Abstract**

The 1960 Data Restoration Project is a collaboration between the Minnesota Population Center and the U.S. Census Bureau to prepare a 5% sample of 1960 Census data, five times denser than the currently available 1% PUMS. This new data set will open new angles of research about the American population not previously possible for this very important historical time point. Not only are there five-times more records than the currently available sample, but the geographic detail in the new sample is greatly improved. To arrive at the 5% sample, the project required restoration of approximately one million records in the Chicago, Illinois area whose digitally recorded data had gone missing. The Minnesota Population Center recovered the original forms, scanned the missing records into digital form, and allocated missing values. In this presentation, we describe the project including methods used to recover and improve the data, publicly-available data products that will be available, and the improved geographic detail in the new 5% PUMS sample.

### Limitations of Current Data

The Census Bureau and the Minnesota Population Center (MPC) are collaborating on a project to recover, integrate, and disseminate data from the 1960 decennial census. Currently, the only nationally-representative microdata to study the 1960 population are the 1% PUMS sample, available through the Integrated Public Use Microdata Series (IPUMS). This data source is problematic because it is a lower density sample than most other PUMS-USA samples. For example, Figure 1 shows the density of the current 1960 PUMS versus other Census decennial and American Community Survey data. It is only a 1% sample (~1.8 million people) while most other decennial samples have a 5% density.

**Figure 1.**



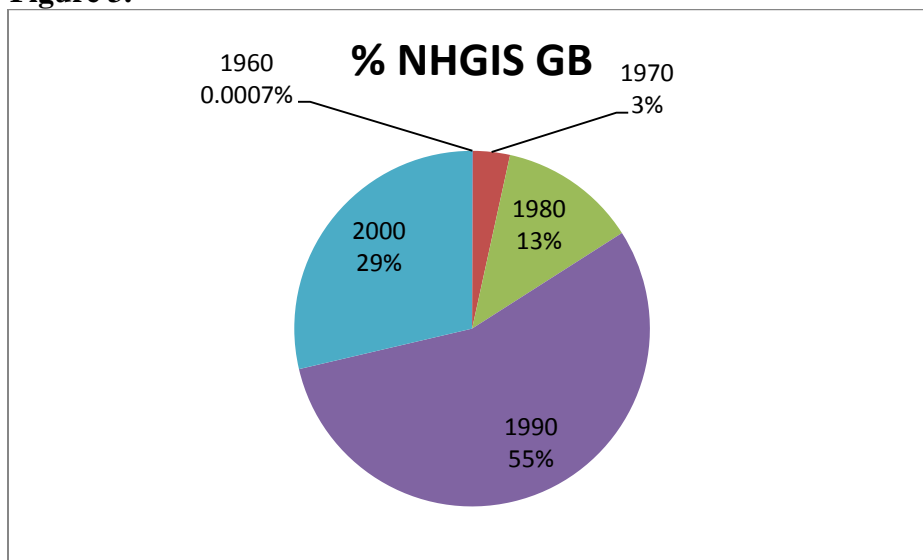
In addition to being low density, the current 1960 PUMS is limited by its lack of geographic detail. The lowest level of geography is at the state level, and the sample does not contain information about place of migration or place of work. Figure 2 shows geography variables that are heavily requested in the IPUMS-USA extract system that are not currently available in the 1% 1960 public use sample.

**Figure 2.**

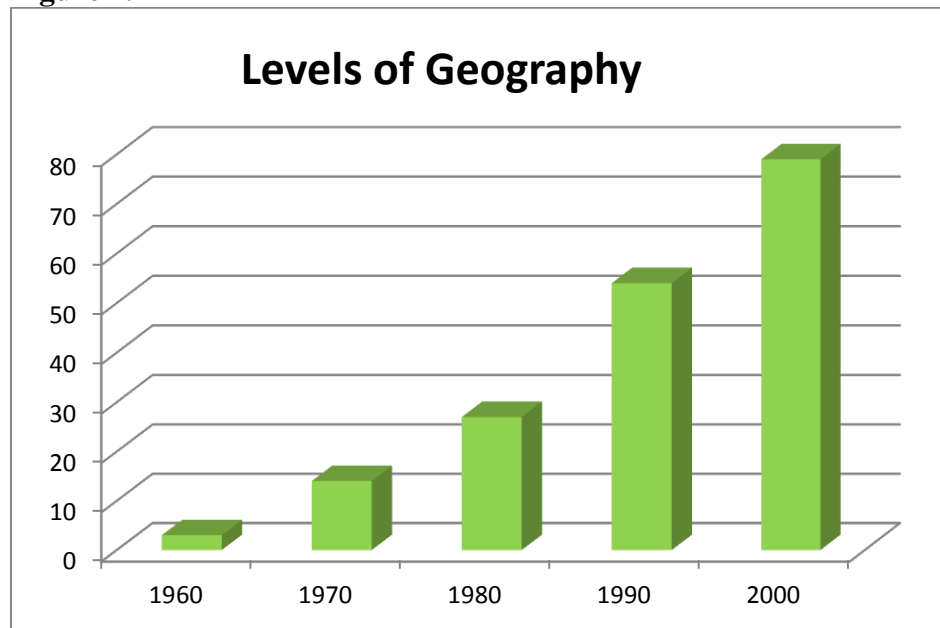
Variables Not Currently Available	
Variable	% of extracts
METAREA	27
COUNTY	22
CITY	20
PUMA	17
PWCITY	7
PWMETRO	6
MIGMET5	4
MIGCITY5	4

Finally, the 1960 aggregate tables have limited comparability to summary data from other years. They have few geographic levels (state, county, tract) and are sparse in terms of content. For example, race is only white and non-white and there are no age by sex cross-tabulations for education and income. See Figures 3 and 4 for more information.

**Figure 3.**



**Figure 4.**



To summarize, the 1960 public use sample is too limited for social scientists to utilize the way that researchers typically analyze Census microdata. It is estimated that only about 20% of *Demography* articles using Census microdata could be replicated using 1960 data (Ruggles, Schroeder, Rivers, Alexander, & Gardner, 2011).

### **Background and Importance**

Despite the limitations of the current 1960 public-use data, the 1960 sample remains an important and heavily requested data source. 1960 is a time period in history with important social, political, and family change. The IPUMS-USA 1960 sample ranks 15<sup>th</sup> out of 96 samples in extracts in 2013 (4,366 total extracts in 2013). The sample is downloaded more than any of the 1970 samples and more than any sample prior to 1960. We expect that after releasing a sample five times the density with drastically improved geographic detail, the popularity of 1960 will only increase.

## Project Overview

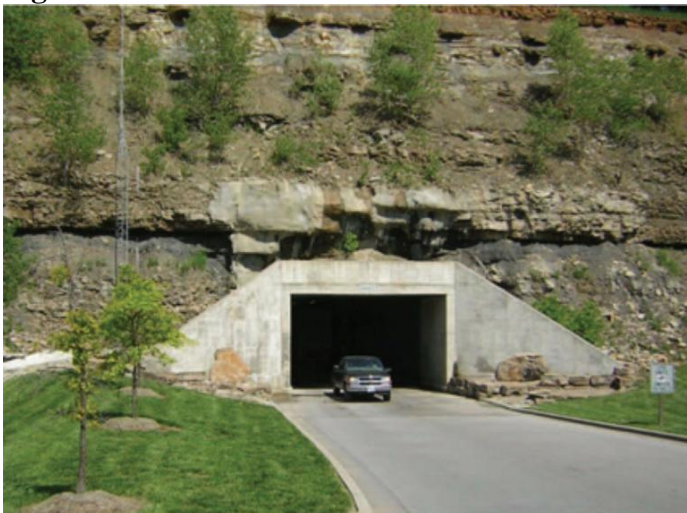
The goal of this presentation is to describe the 1960 Data Restoration Project, including methods used to recover and improve the data. In addition, we plan to discuss project deliverables including new data products and geography variables that will be available.

There are several goals for the 1960 Data Restoration Project. The main deliverables are:

1. A new 5% public-use microdata file
2. New aggregate summary tables
3. A restored 25% internal RDC microdata file

There are several stages of work to complete this project, and we are nearing its conclusion. Critically, the internal 25% long-form data were missing approximately one million people and 300,000 households mainly from the Chicago area of Illinois. The first phase involved restoring the missing cases. Staff at the National Archives and Records Administration (NARA) made digital scans of the original census manuscripts that were stored in a temperature-controlled cave in Lenexa, Kansas (see Figure 5).

**Figure 5.**

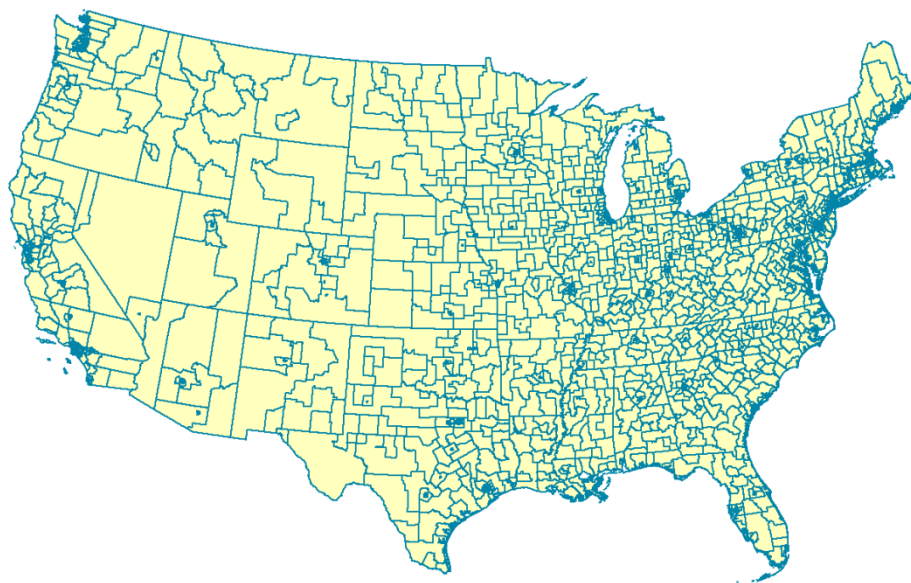


## The 1960 Data Restoration Project

After this was complete, Census Bureau staff at the National Processing Center in Jeffersonville, Indiana converted the digitized images of the microfilmed enumeration forms into machine-readable data. Next, MPC and Census Bureau staff reformatted the restored cases and merged the recovered data with the incomplete internal long form data. MPC staff are currently finishing logical edits and allocating missing data in the restored cases following instructions laid out in the 1960 procedural history.

In addition to restoring the data, MPC staff are making other improvements to the data such as constructing harmonized IPUMS variables and creating new Public Use Microdata Areas (PUMAs). These new “Mini-PUMAs” consist of tracts and untraced counties that are combined to form areas of 50,000 persons or more, as opposed to the 100,000 standard applied to more recent PUMS. See Figure 6 for an illustration of the size of the geographic units.

**Figure 6.**

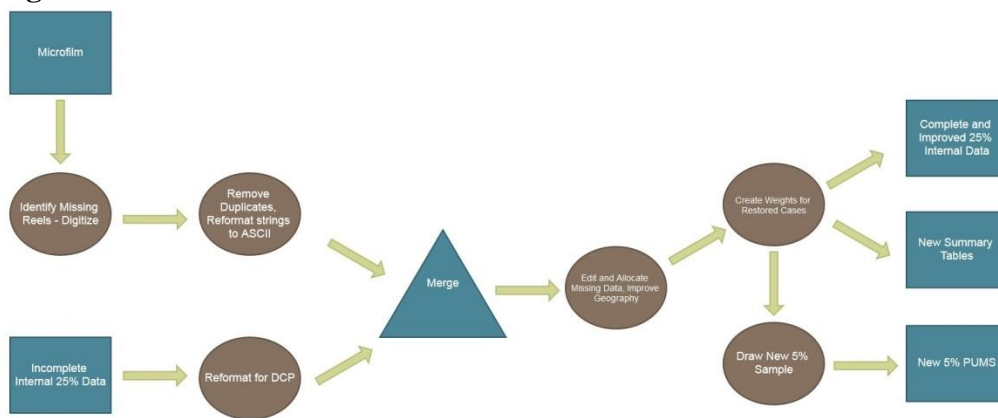


Map made by Ryan Klein

## The 1960 Data Restoration Project

The 1960 units are designed to nest within 2000 PUMA boundaries and be compatible with the new place of migration and place of work geographic identifiers. The new geographic identifiers will drastically improve comparability of 1960 to other data sources. The next steps of the project involve constructing weights for the restored cases and drawing a new 5% PUMS sample to submit for disclosure review. The final stage of the project will include integrating the new 5% public-use microdata file into IPUMS-USA, releasing the restored 25% file for Research Data Center (RDC) researchers, producing new aggregate summary tables to release via the National Historical GIS project (NHGIS), and providing documentation about the data to RDC users and the public. Figure 7 illustrates some of the steps that have been involved in completing this project.

**Figure 7.**



## References

Ruggles, S., Schroeder, M., Rivers, N., Alexander, J. T., & Gardner, T. K. (2011). Frozen Film and FOSDIC Forms: Restoring the 1960 US Census of Population and Housing. *Historical methods*, 44(2), 69-78.