Center for Demography and Ecology

University of Wisconsin-Madison
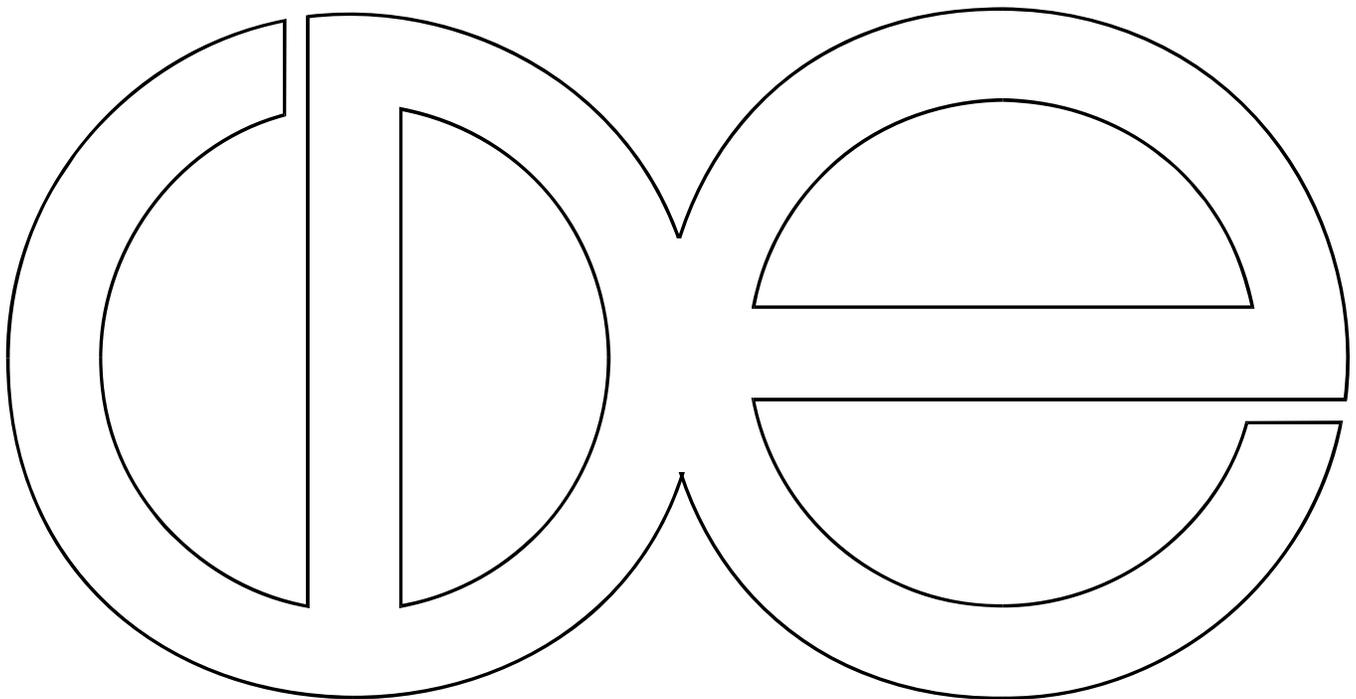
Public Use Samples of 1910 and 1920 Puerto Rico Censuses

Kristen Velyvis

Theresa Thompson-Colón

Halliman Winsborough

Public Use Samples of 1910 and 1920 Puerto Rico Censuses

Abstract
This paper describes new public use samples made from the first and second U.S.
Censuses of Puerto Rico. We made these digital files by sampling and transcribing
information from archival microfilms of the original schedules. Textual information was
transcribed verbatim. Procedures and coding conform to the IPUMS standard for census
samples. Versions of the files are currently available from IPUMS and ICPSR.

By
Kristen Velyvis
Theresa Thompson-Colón
Halliman Winsborough

University of Wisconsin-Madison
January 9, 2005
To be presented at the Latin American Studies Association
2006 International Congress
Session: The Demography of Early Twentieth-Century Puerto Rico
San Juan, Puerto Rico
March 17, 2006

Introduction

The 1910 and 1920 Puerto Rican Census Project consisted of a large-scale data entry activity undertaken at the University of Wisconsin-Madison with the collaboration of the Integrated Public Use Microdata Services (IPUMS) at the University of Minnesota. The project produced two sets of public use sample files of individual and household records drawn from the Puerto Rican censuses of 1910 and 1920. This paper describes how these public use files were produced, why this project was undertaken, and the final data that are now available.

Project Overview

In brief, this project produced two sets of data by sampling and transcribing information from the microfilms of the original schedules of the first and second Censuses of Puerto Rico taken by the U.S. Bureau of the Census. The first public use files are held by the Inter-University Consortium for Political and Social Research (ICPSR) Data Archive, and reflect the data exactly as they were transcribed from the original microfilms. Textual information was transcribed verbatim. These data can be used by historians and others to reflect the complex specificity of the time and place from which it was collected. The second public use files were created using procedures and coding that conform to the IPUMS standard for census samples. They are exactly analogous and strictly comparable to the U.S. public use samples created as part of the IPUMS project (Ruggles and Sobek 1995). These files are most useful for making comparisons across time and place.

This census project was undertaken by Alberto Palloni, Professor of Sociology; Halliman Winsborough, Emeritus Professor of Sociology; and Francisco Scarano, Professor of History, all faculty members at the University of Wisconsin-Madison. The principal investigators (PIs) were each responsible for different facets of the project. Professor Winsborough oversaw the data entry activities and acted as liaison between the PIs and other collaborators on the project. Later, he primarily saw that the documentation was properly prepared and dealt with errors in the data found in the intensive analysis phase. Professor Scarano provided deep historical knowledge of Puerto Rico. He insured the proper interpretation of words and circumstances throughout the process and added substantive weight to analysis. Professor Palloni had primary oversight of the project, but his major contribution is to the analytic part of the research done with the data.

The PIs contracted with the University of Wisconsin Survey Center to establish a data entry and processing operation under the guidance of the Historical Census Project (HCP) staff at the Minnesota Population Center, University of Minnesota. The HCP staff were also sub-contracted to undertake the allocation of missing data so that the files would be comparable to mainland IPUMS files for the same dates. They also undertook the standardization necessary for inclusion of the samples into the IPUMS database. Dr. Ronald Goeken, Research Associate, was the main contact on this project at the Minnesota Population Center.

This project was funded by grant R01 HD41053 from the National Institute for Children and Human Development (NICHD/NIH), under the project title "Public Use Samples of 1910 and 1920 Puerto Rican Censuses."

How the Files Were Produced

The project took about four and a half years to complete starting in January of 2002. There were four main project stages: Pre-data entry, data entry, post-data entry, and analysis. The first six months of the project were spent in the pre-data entry stage, which included acquiring the microfilm of the censuses, buying microfilm readers, setting up the software and hiring data entry clerks. It also included training for the data entry operators in the entry programs and in the interpretation of the manuscript materials.[1]

The data entry stage ensued. Four data entry operators worked full time for 15 months to enter all of the data. Six months into the data-entry stage, the post data-entry processing began. This included data checking and verification, data coding, and format standardization to get the data prepared for IPUMS specifications. The HCP staff also conducted data allocation. Finally, preliminary datasets were delivered in May of 2005. The analysis of the data will continue under the grant until June 2006.

How These Data Can Be Useful

These data samples are valuable because of their uniqueness, the interesting time period in Puerto Rico they cover, their comparability with data from other places and their completeness.

Data from Puerto Rico in the early 1900's are rare despite it being a historical time of great interest. The period covered by these data samples is one that includes the origins of one of the most sudden and rapid demographic changes that took place among developing countries in general and Latin American countries in particular. The period is also one of deep transformations in the economy. As the PI's point out in their grant application,

> The case of Puerto Rico is unique. It is the result of a historical experiment that brought together a blend of contrasting economic, social and cultural influences. Up until 1898 the history of Puerto Rico was intimately connected to Spanish colonial domination, and its most important historical features replicated quite closely characteristics encountered in other parts of the Latin American and Caribbean regions under Spanish colonial rule. Starting in 1898, however, Puerto Rico became a territory legally and administratively dependent on the US. The economic infrastructure was completely revamped, with formidable effects on the social class composition and the regional distribution of the population. Furthermore, the ideological and cultural suprastructure experienced an overhaul that led to a contrast between the cultural and social experience of the island and that of the rest of Latin America. More than figuratively, the political incorporation of Puerto Rico into the US at the turn

---

[1] A more detailed description of the data entry processes can be found in the data documentation and project files held with the data at ICPSR.

of the century constituted an extraordinary historical experiment with a number of consequences that historians, sociologists, economists and demographers cannot afford to ignore. The changes experienced and their impact on a number of conditions led also to a social, economic and demographic regime that is in stark contrast to those created in the rest of Latin America. Yet, simultaneously, Puerto Rico was the site of a historical experience quite dissimilar from the one in the US (Palloni, Scarano, and Winsborough 2000: 37).

The data that have been compiled offer detailed individual information and associated community and geographic location identifiers and characteristics from Puerto Rico during this time; this is formidable information for researchers (Palloni et al. 2000). The data can be used to study social and economic conditions in several regions of the island, inform the nature of social and economic processes that took place in the aftermath of Spanish domination, and understand how these affected the composition of the labor force, the organization of the household, the special distribution of the population and the differential growth of cities. The two samples provide very rich data for "a variety of scholars, from demographers interested in population dynamics to social and economic historians who study the economic and social conditions of the population in developing areas in the aftermath of colonial domination" (Palloni et al. 2000: 35-36).

The following sections lay out in more detail how and why these data are especially valuable and useful to demographers and socioeconomic historians in the words the PIs used in their grant application for this project.

Of Interest to Demographers
The data provide "rich materials for the study of demographic transformations that took place at the turn of the century. It … [is] a precious source of information to test theories of fertility decline, the role of public health interventions, and the factors that trigger changes in residential preferences" (Palloni et al. 2000: 35). Some of the most provocative issues that can be explored with these data include describing "levels, patterns and trends of mortality, fertility, nuptiality during the first two decades of the century, and … [comparing] these with those experienced both in the US and in the rest of the Latin American and Caribbean region (Palloni et al. 2000: 36). In addition, "[w]e know of no other detailed source of information at the national level that would enable us to study mortality and fertility processes during a time to dramatic changes, just at the onset of an irreversible demographic transition" (Palloni et al. 2000: 36). In addition, "the 1910 Puerto Rican census, unlike the majority of censuses ever taken contains information of marriage duration. This permits analyses of fertility that are unique and of unusual robustness" (Palloni et al. 2000: 36).

The uniqueness of the Puerto Rican situation and data can also be exploited for looking at issues such as race, which during this period were different in Puerto Rico than in the contemporaneous U.S. Another provocative issue for investigation is the evaluation of

> the nature and determinants of changes in household composition that took place during the period 1910-1920. During this period Puerto Rico

3

experiences the brunt of the impact that results from incorporation into a US dominated territory. First, the foundations of the economy are overhauled and the transition from a coffee-based to a sugar-based economy is completed in less than twenty years. Second, although still heavily dependent on one export, the Puerto Rican economy begins to develop light manufacturing and industry which also lead to important changes in the nature of labor markets. Third, and more importantly, social order and cultural suprastructure are weakened partly through the direct influence of new ideas and values, partly through the influence of migrants, and partly because the material conditions change very rapidly. Modernization theory would have us believe that, as a consequence of these transformations, household and family composition must change as well, and they experience the passage from an extended form to a nuclear form. But, did such transition ever take place at all? Was the household composition at the outset of the process identical to what would be expected according to modernization theory? And, if a change did indeed occur, can we identify the influence of changing demographic constraints and separate it from the influence of changing values? (Palloni et al. 2000: 36).

Of Interest to Social and Economic Historians

We highlight three potential uses of the census micro samples by social and economic historians. First, students of twentieth-century Puerto Rico will, for the first time, be able to answer questions that have dogged the historiography for decades: Beyond the generalities of migration displacements, in what specific ways did the economy of 'American Porto Rico' with its rapid shift away from coffee planting and toward sugar and tobacco, affect residential and family organization of working groups? How did families cope with the revamped labor market and, in particular, with the substantial entry of women into the waged work force? (Acosta-Belén 1986). Second, using complementary landholding and farm tax data now being analyzed, who among the smallholders of the coastal plains took advantage of the American tariff subsidy to participate in the plantation economy as satellite cane farmers? (Scarano 1990). And what effects did this have on their demographic profile? These are two problems connected with much broader issues in the history of the emergence of newer working classes. The focus is on the processes of proletarianization of the peasantry previously engaged in coffee production and with the concomitant transformation into laborers of small land holders.

Third, beyond the easy dismissal by some scholars of certain census categories as "social constructions," what realities of social interaction and family formation lie behind the sharp increase in the percentage of "whites" and the concomitant decline of "blacks" in the early twentieth century?

The answers to these questions will have an influence on new research on Caribbean and Latin American plantation spaces that were quite comparable to Puerto Rico. Economic and social historians of those areas that Wagley aptly called "plantation America" stand to benefit tangibly from the availability of the Puerto Rican materials. It has been a tenet of the comparative literature for some time that in this part of the American tropics the strictures of social and economic life were fundamentally determined by the human ecology associated with one or another of the "after-dinner crops" especially those that were dominant in Puerto Rico: sugar, coffee, and tobacco (Wagley 1957; Steward 1956). In the hands of comparativists, our samples will permit a reexamination of as yet untested

postulates about the comparability of plantation economies and especially of the working populations shaped by them (Palloni et al. 2000: 35).

The Data in More Detail

The samples that were created are available for public use in two forms. The 1910 and 1920 files that reflect the data as they were transcribed verbatim are held in the ICPSR Data Archives. These files include all variables unedited and unallocated for missing values. These versions of the files contain all original input text variables as well. ICPSR is responsible for preserving these versions of the data files and for providing support to data users. These data are available from http://www.icpsr.umich.edu.

The second set of 1910 and 1920 files was created in accordance with IPUMS data content and format specifications. The HCP staff reviewed the unedited data and carried out data edits and allocation procedures so that the data conform to their standard for census samples. The IPUMS versions of the samples will be released as part of their public archived collection (http://www.ipums.umn.edu/usa/). IPUMS is responsible for preserving these versions of the data files and for providing support to data users.

Both the ICPSR and the IPUMS versions of the data contain both individual level and household level data. The population covered by the samples includes the entire population of the island of Puerto Rico in 1910 and 1920. The data files consist of 1-in-10 samples for most of the island. The exceptions are the coffee-producing regions and the *municipio,* or county, of Loíza (an African-descent enclave), which were over-sampled at a rate of 1-in-5 from each census year. These very high sampling rates resulted in a total of 298,375 cases being entered for the two censuses combined – 207,807 in the main sample; 84,798 in the coffee-producing regions; and 5,772 in the county of Loíza. Table 1 presents the distribution between the two censuses, the main sample and the over-sample.

Table 1: Distribution of the 1910 and 1920 Puerto Rican Samples

Census Year	Main Sample	Coffee Region Over-Sample	Loíza Over-Sample	Total Sample
1910	95,124	40,092	2,675	**137,891**
1920	112,683	44,706	3,097	**160,484**
Total	**207,807**	**84,798**	**5,772**	**298,375**

A feature of the data that is particularly interesting is the inclusion in the samples of dwelling units clustered in particular areas that are identified as *solares* in the 1910 census and *posteriores* in the 1920 census. These dwelling arrangements existed in Puerto Rico during this time, but did not exist in the mainland U.S. In 1910, a *solar* was defined as a clustering of dwellings in an urban area where a land-tract had been subdivided and had one dwelling built behind another. In 1920, the Census Bureau was more specific in its directions and indicated that if a house was situated behind a main house that was facing the street, this house should be noted as a *posterior*. The sampling

of *solares / posteriores* allows researchers to examine clustering of family relationships within this type of geographic clustering.

Another interesting feature of these data is the complete enumeration of the military and naval institutions in the area of San Juan in 1920. A total of 1,981 individuals who made up the military and naval population in the 1920 census were included in a separate spreadsheet available with the data held by ICPSR.

Sampling Issues
As mentioned, the 1910 and 1920 samples from the Puerto Rican Censuses are probability samples that use both stratification and clustering methods. How one deals with the analytic complications that such methods present depends in great measure on what you want to learn from the data. Here we discuss the issues, their impact on estimates derived from the data, and suggest some ways to cope.

The files are cluster samples in two ways. Here is how we proceeded: A sample line was drawn randomly from every fifty-line enumeration form on the microfilm. If a household head appeared on that line or on any of the four succeeding lines, that person and all members of his/her household were taken into the sample. More than one household could be drawn from the five-line cluster. Thus, lines are clustered on the page and persons are clustered within households.

The files are stratified samples in three ways. First, households in coffee-growing *barrios,* or neighborhoods, and those in a traditionally African-descent *municipio* were sampled at twice the rate of other households in order to yield samples of these relatively rare people that are of sufficient size to permit robust analysis. Thus, in the affected *barrios* nine succeeding lines rather than four are added to the cluster of lines. Second, choosing to draw sample lines from each page of the enumeration forms effectively stratifies the sample by geography since pages cover contiguous households or at least neighborhoods that are proximate one to another. Third, in preparing to draw the sample we discovered that these files include instances of the then to us unfamiliar type of residence called *solares*. *Solares* resemble urban compounds since they include a number of more or less separate households at a single address. The Census Bureau established rules to distinguish the separate households but in our initial inspections it was unclear that they were effective in the field. Early in the project, then, we decided to include everyone in a selected *solares*, whether in the initially selected household or in another unit in the compound without regard to whether that additional unit would fall within the sample or not. We wanted to be prepared to deal with enumerators' problems with the rules if necessary (it turned out not to be much of a problem) and we wanted to include data to permit study of this interesting and rather unique residence type.

In most cases this last stratification is easy to manage: Get rid of the extra households, thus returning to the initially planned sample plan. There is a variable in each record that distinguishes the added households from those that conform to the initial sampling plan. One simply drops the "extra" households at the *solares* address. One

way to accomplish this exclusion is to drop all cases in which that variable, SOLRULE, is equal to 0.

Dealing with the over sampled coffee-growing and African-decent *barrios* is only a bit harder to manage. The variable DATANUM indicates the rate at which the household was sampled. A value of 1 indicates membership in the basic 10% sample group, a value of 2 indicates membership in the 20% sample of the African-descent region, and a value of 3 indicates the 20% sample of the coffee-growing region. Many people deal with stratified samples by constructing weights and using the weights in their data analysis. Here, one might assign a weight of 1 to cases coded 1 on DATANUM and a weight of one-half to cases coded 2 or 3 on that variable. For many purposes a sample so weighted can be regarded as a 10% sample of the population. Thus, the total of the weighted sample size multiplied by 10 yields an estimate of the population total. Using this method, the sample estimates the 1910 population of Puerto Rico as 1,112,360 persons. The Censuses reports the figure of 1,118,012. Thus the estimate is within .5% of the Census count. Is that departure within a reasonable confidence interval? Answering that question requires computing a variance as well as an estimate. We shall address estimating variances below.

Clustering is the hardest to manage if you want to make probability statements about the accuracy of estimates you calculate from these data. Cluster samples can be either more or less powerful than simple random samples depending on the variable considered. If the within-cluster variance of a variable is greater than it would be for randomly drawn clusters from the same population, the sample will be more powerful. If the within-cluster variance is smaller than for randomly drawn clusters, it will be less powerful. Households typically contain variables of both kinds. Age and sex within households is usually more heterogeneous than within randomly drawn clusters while measure of wealth, education, race, and other social status variable are usually more homogeneous within households than within randomly selected clusters. Figuring out the variance of the parameters of a multivariate model estimated from such data is very difficult; sometimes impossible to do analytically. We choose to deal with this problem by providing the user with the possibility of producing an empirical estimate of the variance of any calculation you wish to make. Because the samples for these census records are drawn independently from each page of the enumeration forms we can allocate each page randomly to a separate sample and have several independent samples of the data. If the calculation of interest is made on each sample separately, a variance for the quantity can be estimated empirically and the sampling distribution of the quantity investigated. Then, the resulting variance can perhaps be used for statistical tests and confidence intervals. The variable PAGENUM provides the relevant page number.

As an example of this procedure let us compare the simple random sample confidence interval for the population total with that derived from an empirical estimate of its variance. To accomplish our empirical estimate of the variance, we randomly assigned each enumeration page in the 1910 sample to one of ten separate groups. Each group, then, is a one-in-a-hundred sample of the population of Puerto Rico in 1910. The population size for each group is calculated. Its mean is 11,123.60. The estimated

variance of the population size is 211,820.32 and the standard deviation is 460.24. Were we dealing with a simple random sample with p=.1, the variance would be Np(1-p) or 100,621 and the standard deviation would be 317.21. Thus, the sampling design increases the standard deviation of the population count by about 50%.

The sampling procedures used produced samples that reflect the census populations fairly well. The following tables compare the sample data to the census data in a number of ways. As indicated above, we have used weights to deal with over-sampling and have dropped extra cases entered from *solares* and *posteriores*, that is, extra cases that would not have fallen within the sample under the original sampling plan. The results show that the samples reflect the census populations very well.

First, a comparison by race.

Table 2: Race – 1910

Race	Census Population	Percent of Population	Weighted Sample	Percent of Sample
White	732,555	65.52%	72,167	64.89%
Black	50,245	4.49%	5,025	4.52%
Mulatto	335,192	29.98%	34,028	30.59%
Other	20	0.00%	3	0.00%
Total	1,118,012	100.00%	111,223	100.00%

Census & preliminary ICPSR sample data (4/14/05).

Sample data do not include *solares*.

Table 3: Race – 1920

Race	Census Population	Percent of Population	Weighted Sample	Percent of Sample
White	948,709	72.99%	94,829	74.07%
Black	49,246	3.79%	4,924	3.85%
Mulatto	301,816	23.22%	28,273	22.08%
Other	38	0.00%	1	0.00%
Total	1,299,809	100.00%	128,027	100.00%

Census & preliminary ICPSR sample data (6/6/2005).

Sample data do not include *posteriores*.

Next, we compared the samples to the census data by sex. The 1910 sample consists of 49.93% men compared to 49.85% men in the census. For women there are 50.07% in the sample and 50.15% in the census. In the 1920 data, there are 49.93% men

in the sample and 49.84% in the census. There are 50.07% women in the sample and 50.16% women in the census.

Next we made comparisons by age and sex.

Table 4a: Age x Sex – 1910
Men

Age Groups	Census Population Men	Percent of Population	Weighted Sample	Percent of Weighted Sample
Under 5	94,213	16.91%	9,426.0	16.97%
5 to 9	76,572	13.74%	7,664.5	13.80%
10 to 14	74,280	13.33%	7,326.0	13.19%
15 to 19	53,388	9.58%	5,254.5	9.46%
20 to 24	53,492	9.60%	5,372.5	9.67%
25 to 29	45,836	8.22%	4,597.5	8.28%
30 to 34	35,331	6.34%	3,557.0	6.41%
35 to 39	32,283	5.79%	3,277.0	5.90%
40 to 44	26,539	4.76%	2,619.0	4.72%
45 to 49	18,977	3.41%	1,888.5	3.40%
50 to 54	16,867	3.03%	1,646.5	2.96%
55 to 59	9,241	1.66%	915.0	1.65%
60 to 64	8,996	1.61%	888.5	1.60%
65 to 69	4,612	0.83%	483.0	0.87%
70 to 74	3,156	0.57%	284.0	0.51%
75 to 79	1,525	0.27%	136.5	0.25%
80 to 84	1,139	0.20%	127.0	0.23%
85 +	833	0.15%	71.5	0.13%
Age Unknown	21	0.00%		
Total	557,301	100.00%	55,534.5	100.00%

Puerto Rican Census & preliminary ICPSR sample data (4/14/05).

Sample data do not include *solares*.

Table 4b: Age x Sex – 1910

Women

Age Groups	Census Population Women	Percent of Population	Weighted Sample	Percent of Weighted Sample
Under 5	90,976	16.23%	9,002.5	16.17%
5 to 9	74,651	13.31%	7,319.0	13.14%
10 to 14	69,471	12.39%	6,833.0	12.27%
15 to 19	60,401	10.77%	6,029.0	10.83%
20 to 24	55,016	9.81%	5,452.0	9.79%
25 to 29	48,377	8.63%	4,911.0	8.82%
30 to 34	35,091	6.26%	3,504.0	6.29%
35 to 39	32,005	5.71%	3,145.5	5.65%
40 to 44	25,805	4.60%	2,555.0	4.59%
45 to 49	18,186	3.24%	1,756.0	3.15%
50 to 54	16,484	2.94%	1,693.5	3.04%
55 to 59	9,552	1.70%	968.5	1.74%
60 to 64	10,137	1.81%	1,057.5	1.90%
65 to 69	5,211	0.93%	504.5	0.91%
70 to 74	4,250	0.76%	439.5	0.79%
75 to 79	1,920	0.34%	207.0	0.37%
80 to 84	1,752	0.31%	160.5	0.29%
85 +	1,393	0.25%	145.5	0.26%
Age Unknown	33	0.01%		
Total	560,711	100.00%	55,683.5	100.00%

Puerto Rican Census & preliminary ICPSR sample data (4/14/05).

Sample data do not include *solares*.

Table 5a: Age x Sex – 1920
Men

Age Groups	Census Population Men	Percent of Population	Weighted Sample	Percent of Weighted Sample
Under 5	100,898	15.57%	9,985.5	15.62%
5 to 9	99,150	15.31%	9,705.0	15.19%
10 to 14	85,702	13.23%	8,538.5	13.36%
15 to 19	60,183	9.29%	6,043.0	9.46%
20 to 24	62,114	9.59%	6,142.5	9.61%
25 to 29	44,138	6.81%	4,372.5	6.84%
30 to 34	37,540	5.79%	3,646.5	5.71%
35 to 39	38,438	5.93%	3,768.0	5.90%
40 to 44	31,174	4.81%	3,035.0	4.75%
45 to 49	28,368	4.38%	2,780.0	4.35%
50 to 54	21,878	3.38%	2,137.0	3.34%
55 to 59	11,610	1.79%	1,151.0	1.80%
60 to 64	12,374	1.91%	1,224.0	1.92%
65 to 69	5,415	0.84%	548.0	0.86%
70 to 74	4,138	0.64%	382.0	0.60%
75 to 79	1,829	0.28%	178.5	0.28%
80 to 84	1,622	0.25%	167.5	0.26%
85 +	1,100	0.17%	103.5	0.16%
Age Unknown	154	0.02%		
Total	647,825	100.00%	63,908.0	100.00%

Puerto Rican Census & preliminary ICPSR sample data (6/6/2005).

Sample data do not include *posteriores*.

Table 5b: Age x Sex – 1920
Women

Age Groups	Census Population Women	Percent of Population	Weighted Sample	Percent of Weighted Sample
Under 5	99,357	15.24%	9,773.5	15.26%
5 to 9	95,981	14.72%	9,431.5	14.72%
10 to 14	82,352	12.63%	7,963.0	12.43%
15 to 19	66,065	10.13%	6,458.0	10.08%
20 to 24	66,417	10.19%	6,625.0	10.34%
25 to 29	51,915	7.96%	5,057.5	7.89%
30 to 34	40,465	6.21%	3,896.0	6.08%
35 to 39	38,651	5.93%	3,825.5	5.97%
40 to 44	30,276	4.64%	2,985.0	4.66%
45 to 49	21,905	3.36%	2,205.5	3.44%
50 to 54	19,746	3.03%	1,952.0	3.05%
55 to 59	9,976	1.53%	949.0	1.48%
60 to 64	12,013	1.84%	1,243.0	1.94%
65 to 69	5,547	0.85%	594.5	0.93%
70 to 74	5,018	0.77%	515.5	0.80%
75 to 79	2,137	0.33%	210.5	0.33%
80 to 84	2,213	0.34%	199.5	0.31%
85 +	1,807	0.28%	176.5	0.28%
Age Unknown	143	0.02%		
Total	651,984	100.00%	64,061.0	100.00%

Puerto Rican Census & preliminary ICPSR sample data (6/6/2005).

Sample data do not include *posteriores*.

Finally we look at both families and individuals per *municipio*. Cases marked in red indicate *municipios* where the comparison between the sample and population data results in a chi sq higher than 10 with one degree of freedom.

Table 6a: Population per *Municipio* – 1910

Municipio	Number of Families per *Municipio*				Number of Persons per *Municipio*			
	Census Popltn	% of Popltn	Wt'ed Sample	% of Sample	Census Popltn	% of Popltn	Wt'ed Sample	% of Sample
Adjuntas	3,250	1.48%	326.0	1.45%	16,954	1.52%	1,694.0	1.52%
Aguada	2,326	1.06%	246.0	1.10%	11,587	1.04%	1,183.0	1.06%
Aguadilla	4,218	1.92%	431.0	1.92%	21,419	1.92%	2,111.0	1.90%
Aguas Buenas	1,463	0.67%	142.0	0.63%	8,292	0.74%	836.0	0.75%
Aibonito	2,068	0.94%	194.0	0.86%	10,815	0.97%	995.0	0.89%
Añasco	2,934	1.33%	287.0	1.28%	14,407	1.29%	1,443.5	1.30%
Arecibo	8,453	3.85%	865.0	3.85%	42,429	3.80%	4,122.0	3.71%
Arroyo	1,576	0.72%	151.0	0.67%	6,940	0.62%	664.0	0.60%
Barceloneta	2,240	1.02%	213.0	0.95%	11,644	1.04%	1,069.0	0.96%
Barranquitas	1,818	0.83%	195.0	0.87%	10,503	0.94%	1,071.0	0.96%
Barros	2,578	1.17%	275.0	1.22%	15,028	1.34%	1,659.0	1.49%
Bayamón	6,062	2.76%	562.0	2.50%	29,986	2.68%	2,920.0	2.63%
Cabo Rojo	3,713	1.69%	380.0	1.69%	19,562	1.75%	1,971.0	1.77%
Caguas	5,271	2.40%	515.0	2.29%	27,160	2.43%	2,658.0	2.39%
Camuy	2,137	0.97%	221.0	0.98%	11,342	1.01%	1,143.0	1.03%
Carolina	2,849	1.30%	292.0	1.30%	15,327	1.37%	1,623.0	1.46%
Cayey	3,562	1.62%	343.0	1.53%	17,711	1.58%	1,642.0	1.48%
Ciales	3,228	1.47%	325.5	1.45%	18,398	1.65%	1,848.0	1.66%
Cidra	1,998	0.91%	209.0	0.93%	10,595	0.95%	1,130.0	1.02%
Coamo	3,047	1.39%	302.0	1.34%	17,129	1.53%	1,727.0	1.55%
Comerio	1,759	0.80%	186.0	0.83%	11,170	1.00%	1,133.0	1.02%
Corozal	2,242	1.02%	218.0	0.97%	12,978	1.16%	1,272.0	1.14%
Culebra	228	0.10%	41.0	0.18%	1,315	0.12%	132.0	0.12%
Dorado	946	0.43%	90.0	0.40%	4,885	0.44%	491.0	0.44%
Fajardo	4,298	1.96%	449.0	2.00%	21,135	1.89%	2,113.0	1.90%
Guayama	3,874	1.76%	391.0	1.74%	17,379	1.55%	1,731.0	1.56%
Guayanilla	2,096	0.95%	214.0	0.95%	10,354	0.93%	987.0	0.89%
Gurabo	1,955	0.89%	191.0	0.85%	11,139	1.00%	1,047.0	0.94%
Hatillo	1,918	0.87%	206.0	0.92%	10,630	0.95%	1,163.0	1.05%
Humacao	5,167	2.35%	547.0	2.44%	26,678	2.39%	2,658.0	2.39%
Isabela	3,131	1.42%	328.0	1.46%	16,852	1.51%	1,773.0	1.59%
Juana Díaz	5,935	2.70%	609.5	2.71%	29,157	2.61%	2,881.5	2.59%
Juncos	2,225	1.01%	214.0	0.95%	11,692	1.05%	1,188.0	1.07%
Lajas	2,057	0.94%	218.0	0.97%	11,071	0.99%	1,178.0	1.06%
Lares	4,303	1.96%	435.0	1.94%	22,650	2.03%	2,333.5	2.10%
Las Marías	2,132	0.97%	219.5	0.98%	10,046	0.90%	994.5	0.89%
Loíza	2,307	1.05%	231.0	1.03%	13,317	1.19%	1,337.5	1.20%

13

Manatí	3,307	1.50%	332.0	1.48%	17,240	1.54%	1,693.0	1.52%
Maricao	1,604	0.73%	161.0	0.72%	7,158	0.64%	718.0	0.65%
Maunabo	1,455	0.66%	153.0	0.68%	7,106	0.64%	754.0	0.68%
Mayagüez	9,338	4.25%	932.5	4.15%	42,429	3.80%	4,110.0	3.69%
Moca	2,702	1.23%	275.0	1.22%	13,640	1.22%	1,388.0	1.25%
Morovis	2,121	0.96%	201.0	0.90%	12,446	1.11%	1,157.0	1.04%
Naguabo	2,652	1.21%	253.0	1.13%	14,365	1.28%	1,445.0	1.30%
Naranjito	1,480	0.67%	150.0	0.67%	8,876	0.79%	956.0	0.86%
Patillas	2,890	1.31%	292.0	1.30%	14,448	1.29%	1,475.0	1.33%
Peñuelas	2,355	1.07%	229.0	1.02%	11,991	1.07%	1,209.0	1.09%
Ponce	14,814	6.74%	1,515.5	6.75%	63,444	5.67%	6,076.5	5.46%
Quebradillas	1,492	0.68%	159.0	0.71%	8,152	0.73%	822.0	0.74%
Rincón	1,346	0.61%	134.0	0.60%	7,275	0.65%	747.0	0.67%
Río Grande	2,497	1.14%	257.0	1.14%	13,948	1.25%	1,457.0	1.31%
Río Piedras	3,488	1.59%	336.0	1.50%	18,880	1.69%	1,798.0	1.62%
Sabana Grande	2,157	0.98%	213.0	0.95%	11,523	1.03%	1,158.0	1.04%
Salinas	2,375	1.08%	214.0	0.95%	11,403	1.02%	984.0	0.88%
San German	4,516	2.05%	434.0	1.93%	22,143	1.98%	2,127.0	1.91%
San Juan	10,293	4.68%	1,325.0	5.90%	48,716	4.36%	4,833.0	4.34%
San Lorenzo	2,634	1.20%	282.0	1.26%	14,278	1.28%	1,513.0	1.36%
San Sebastian	3,638	1.65%	349.5	1.56%	18,904	1.69%	1,849.5	1.66%
Santa Isabel	1,508	0.69%	159.0	0.71%	6,959	0.62%	656.0	0.59%
Toa Alta	1,585	0.72%	163.0	0.73%	9,127	0.82%	887.0	0.80%
Toa Baja	1,167	0.53%	133.0	0.59%	6,254	0.56%	690.0	0.62%
Trujillo Alto	1,056	0.48%	114.0	0.51%	6,345	0.57%	722.0	0.65%
Utuado	7,631	3.47%	776.5	3.46%	41,054	3.67%	4,105.5	3.69%
Vega Alta	1,547	0.70%	145.0	0.65%	8,134	0.73%	813.0	0.73%
Vega Baja	2,305	1.05%	240.0	1.07%	12,831	1.15%	1,320.0	1.19%
Vieques	2,539	1.15%	265.0	1.18%	10,425	0.93%	954.0	0.86%
Yabucoa	3,456	1.57%	344.0	1.53%	17,338	1.55%	1,817.0	1.63%
Yauco	6,515	2.96%	658.0	2.93%	31,504	2.82%	3,078.5	2.77%
Total	**219,827**	**100.00%**	**22,457.5**	**100.00%**	**1,118,012**	**100.00%**	**111,235.5**	**100.00%**

Puerto Rican Census & preliminary ICPSR sample data (4/14/05).

Sample data do not include *solares*.

Red **cases indicate where chi squares are higher than 10 with 1 degree of freedom.**

Population per *Municipio* – 1920

Municipio	Number of Families per *Municipio*				Number of Persons per *Municipio*			
	Census Popltn	% of Popltn	Wt'ed Sample	% of Sample	Census Popltn	% of Popltn	Wt'ed Sample	% of Sample
Adjuntas	3,310	1.32%	336.5	1.33%	17,988	1.38%	1,858.0	1.45%
Aguada	2,604	1.04%	264.0	1.05%	12,981	1.00%	1,351.0	1.05%
Aguadilla	4,605	1.83%	448.0	1.77%	24,287	1.87%	2,390.0	1.87%
Aguas Buenas	1,834	0.73%	187.0	0.74%	10,741	0.83%	1,063.0	0.83%
Aibonito	2,306	0.92%	230.0	0.91%	13,264	1.02%	1,242.0	0.97%
Añasco	2,833	1.13%	283.0	1.12%	13,834	1.06%	1,357.5	1.06%
Arecibo	9,560	3.81%	956.0	3.78%	46,578	3.58%	4,702.0	3.67%
Arroyo	1,630	0.65%	160.0	0.63%	7,074	0.54%	717.0	0.56%
Barceloneta	2,609	1.04%	260.0	1.03%	13,442	1.03%	1,385.0	1.08%
Barranquitas	1,942	0.77%	194.0	0.77%	11,600	0.89%	1,128.0	0.88%
Barros	2,637	1.05%	251.0	0.99%	15,758	1.21%	1,560.0	1.22%
Bayamon	6,212	2.47%	638.0	2.53%	30,739	2.36%	3,142.0	2.45%
Cabo Rojo	4,334	1.73%	471.0	1.86%	22,412	1.72%	2,284.0	1.78%
Caguas	7,055	2.81%	695.0	2.75%	35,920	2.76%	3,584.0	2.80%
Camuy	2,690	1.07%	258.0	1.02%	14,228	1.09%	1,386.0	1.08%
Carolina	2,776	1.11%	295.0	1.17%	15,563	1.20%	1,620.0	1.26%
Cayey	4,420	1.76%	403.0	1.60%	23,618	1.82%	2,088.0	1.63%
Ciales	3,564	1.42%	353.0	1.40%	20,730	1.59%	2,046.0	1.60%
Cidra	2,682	1.07%	261.0	1.03%	14,789	1.14%	1,429.0	1.12%
Coamo	3,153	1.26%	323.0	1.28%	17,749	1.37%	1,809.0	1.41%
Comerio	2,527	1.01%	266.0	1.05%	14,708	1.13%	1,509.0	1.18%
Corozal	2,375	0.95%	224.0	0.89%	14,369	1.11%	1,338.0	1.04%
Culebra	146	0.06%	12.0	0.05%	839	0.06%	57.0	0.04%
Dorado	1,139	0.45%	119.0	0.47%	5,842	0.45%	607.0	0.47%
Fajardo	2,940	1.17%	273.0	1.08%	14,302	1.10%	1,319.0	1.03%
Guayama	4,376	1.74%	452.0	1.79%	19,192	1.48%	1,883.0	1.47%
Guayanilla	2,360	0.94%	233.0	0.92%	12,083	0.93%	1,199.0	0.94%
Gurabo	2,206	0.88%	237.0	0.94%	12,882	0.99%	1,406.0	1.10%
Hatillo	2,466	0.98%	237.0	0.94%	13,979	1.08%	1,360.0	1.06%
Humacao	3,898	1.55%	401.0	1.59%	20,229	1.56%	2,050.0	1.60%
Isabela	3,612	1.44%	353.0	1.40%	19,809	1.52%	1,869.0	1.46%
Juana Diaz	3,915	1.56%	388.0	1.54%	18,529	1.43%	1,774.5	1.39%
Juncos	2,485	0.99%	239.0	0.95%	13,151	1.01%	1,124.0	0.88%
Lajas	2,228	0.89%	231.0	0.91%	11,908	0.92%	1,177.0	0.92%
Lares	4,535	1.81%	461.5	1.83%	25,197	1.94%	2,542.5	1.98%
Las Marias	2,044	0.81%	221.0	0.87%	10,736	0.83%	1,106.0	0.86%
Loiza	2,682	1.07%	265.0	1.05%	15,804	1.22%	1,548.5	1.21%
Manati	3,844	1.53%	370.0	1.46%	20,100	1.55%	1,952.0	1.52%
Maricao	1,655	0.66%	177.0	0.70%	8,291	0.64%	838.5	0.65%

Maunabo	1,522	0.61%	139.0	0.55%	7,973	0.61%	732.0	0.57%
Mayaguez	9,191	3.66%	913.5	3.62%	41,612	3.20%	4,049.0	3.16%
Moca	2,981	1.19%	272.0	1.08%	15,791	1.21%	1,385.0	1.08%
Morovis	2,395	0.95%	241.0	0.95%	14,660	1.13%	1,463.0	1.14%
Naguabo	2,886	1.15%	303.0	1.20%	15,788	1.21%	1,689.0	1.32%
Naranjito	1,704	0.68%	177.0	0.70%	10,503	0.81%	1,109.0	0.87%
Patillas	2,778	1.11%	275.0	1.09%	14,284	1.10%	1,403.0	1.10%
Peñuelas	2,554	1.02%	248.0	0.98%	13,598	1.05%	1,336.0	1.04%
Ponce	16,344	6.51%	1,628.5	6.45%	71,426	5.50%	6,945.0	5.42%
Quebradillas	1,673	0.67%	188.0	0.74%	9,404	0.72%	974.0	0.76%
Rincon	1,605	0.64%	152.0	0.60%	8,476	0.65%	832.0	0.65%
Rio Grande	2,244	0.89%	218.0	0.86%	13,247	1.02%	1,274.0	0.99%
Rio Piedras	4,359	1.74%	430.0	1.70%	23,035	1.77%	2,202.0	1.72%
Sabana Grande	2,317	0.92%	235.0	0.93%	12,305	0.95%	1,235.0	0.96%
Salinas	2,921	1.16%	306.0	1.21%	12,971	1.00%	1,339.0	1.05%
San German	4,808	1.92%	477.0	1.89%	23,848	1.83%	2,377.0	1.86%
San Juan	14,853	5.92%	1,652.0	6.54%	71,443	5.50%	6,515.0	5.09%
San Lorenzo	3,213	1.28%	315.0	1.25%	18,136	1.40%	1,808.0	1.41%
San Sebastian	4,025	1.60%	407.5	1.61%	22,049	1.70%	2,258.0	1.76%
Santa Isabel	1,826	0.73%	168.0	0.67%	7,257	0.56%	700.0	0.55%
Toa Alta	1,768	0.70%	169.0	0.67%	10,505	0.81%	955.0	0.75%
Toa Baja	1,378	0.55%	167.0	0.66%	7,121	0.55%	818.0	0.64%
Trujillo Alto	1,226	0.49%	128.0	0.51%	7,470	0.57%	835.0	0.65%
Utuado	6,339	2.53%	634.5	2.51%	35,135	2.70%	3,521.0	2.75%
Vega Alta	1,789	0.71%	194.0	0.77%	9,970	0.77%	1,008.0	0.79%
Vega Baja	2,878	1.15%	302.0	1.20%	15,756	1.21%	1,725.0	1.35%
Vieques	2,619	1.04%	267.0	1.06%	11,651	0.90%	1,191.0	0.93%
Yabucoa	3,768	1.50%	357.0	1.41%	19,623	1.51%	1,865.0	1.46%
Yauco	5,096	2.03%	500.5	1.98%	25,848	1.99%	2,554.0	1.99%
Ceiba	1,161	0.46%	122.5	0.48%	5,973	0.46%	595.0	0.46%
Guanica	2,029	0.81%	187.0	0.74%	9,948	0.77%	902.0	0.70%
Guayanabo	1,874	0.75%	181.0	0.72%	10,800	0.83%	1,076.0	0.84%
Hormigueros	936	0.37%	97.0	0.38%	4,584	0.35%	494.0	0.39%
Jayuya	2,253	0.90%	233.5	0.92%	12,463	0.96%	1,305.5	1.02%
Las Piedras	1,880	0.75%	186.0	0.74%	10,620	0.82%	994.0	0.78%
Luquillo	1,214	0.48%	117.0	0.46%	6,251	0.48%	637.0	0.50%
Villalba	2,401	0.96%	243.5	0.96%	13,040	1.00%	1,252.0	0.98%
Total	**250,997**	**100.00%**	**25,258.0**	**100.00%**	**1,299,809**	**100.00%**	**128,154.0**	**100.04%**

Puerto Rican Census & preliminary ICPSR sample data (6/6/2005).
Sample data do not include *posteriores*.
Red cases indicate where chi squares are higher than 10 with 1 degree of freedom.

Quality of the Data

The data-entry procedure for the 1910 and 1920 samples emphasized accuracy and efficiency, rather than speed. Data checking and data verification were also scrupulously undertaken. The team at the UW Survey Center followed up on their data entry by first checking the data. This entailed running the raw entered data through a logical check program, which highlighted potential data entry errors based on the relationship between different fields. For example, if sex was male but relationship corresponded to a female, the team compared the flagged data to the original microfilm data and corrected the information.

After the checking process was completed, the cleaned data were run through the verifying process. For quality control, this entailed double-entering one-tenth of the cases, comparing the cleaned data to the re-entered file (entered by a second data entry operator) and again reconciling the cases. Reconciling entailed comparing cases that did not match to the original microfilm and choosing the correct value. At this point, the data checker assigned "errors." Error reports listed the number of differences or discrepancies found in the cleaned data. Overall, the data entry error rates were .12 percent for both the 1910 data and the 1920 data, however, if one calculates the error rates including the null cells, which a data entry operator must recognize, only .002 percent of the actual entries in the 1910 sample were erroneous and only .007 percent of the entries in the 1920 sample were erroneous.

The quality of the data was also checked and improved during the process of allocation of missing variables. Further logical checks were done, and the data were looked at more closely at the case level to determine whether there were errors. These continuing processes indicate to us that the final data available at ICPSR and IPUMS will of very good quality.

Some Comments on the Taking of a U.S. Census in Puerto Rico in 1910

The U.S. Census has always been designed to collect and provide data that many kinds of contemporary users will find meaningful and useful. By 1910, experience, thought, and trial and error had produced a method of census taking that seemed to lead to that end for users in the United States. Of course these methods were culture laden depending, as they did, on shared cultural practices and shared meanings. To make this assertion is not to say that the Census was narrow in its acceptance of the diverse practices and meaning that it encountered. It was fairly well adapted to such diverse groups as immigrant communities in New York, cotton plantations in Louisiana, dairy farms in the Midwest, large ranches in Montana, Indian reservations, somewhat nomadic groups in Alaska and multiracial individuals in Hawaii. Alaska and Hawaii were the only non-contiguous "territories" enumerated prior to 1910. Nonetheless, assumptions about culture abounded.

The U.S. Census counts the population and assays its characteristics by first figuring out where people live. People are assumed to live at a single location at a certain point in time, one that can be reliably situated on a map and placed within the politically and socially relevant communities for which the census will publish data. Further, people are expected to live in recognizable dwelling units and be grouped therein by households

made up of mostly related individuals. One person in the household, usually male, was designated "head" of the household and an individual's "family" membership was determined in relationship to one's kinship with the head. Thus, membership in a family is dependent on co-residence and hence on the definition and boundaries of a household. Further, some relationships were recognized as kin and some were not. For example, a step-child is kin but a god-child is not.

When the Census Bureau recognizes that a group of people do not fit well within its expectations, definitions, and rules, it typically creates a special procedure or rule to deal with them. Census day is set for April 1, a time when a minimum number of people are not in their usual location. People who are away from home on the census day are enumerated on a special "transit" night. College students have sometimes been reallocated back to their place of usual residence. A considerable effort is made to search out hobos and homeless people on transit night. Over time elaborate rules for defining the extent and boundaries of a household have come into being but in 1910 there were still comparatively primitive.

The Census Bureau did its first enumeration in Puerto Rico in 1910. Since Puerto Rico had not previously contributed to the development of Census methodology there was a certain culture clash in this first encounter that is as revealing about the United States as about Puerto Rico. Many of the definitions, rules, and procedures imbedded in a census seem transparent to contemporaries because they are so well articulated with the contemporary culture of the time and place. Only a change makes them more visible. Applying the well established procedure of the U.S. Census to Puerto Rico in 1910 made some interesting things visible. We were prepared for some of them but others were discovered serendipitously as we proceeded with the boring tasks of checking for errors. Four arenas proved most interesting and important: Race, family, geography, and ownership.

Race
Race is perhaps the most interesting issue we encountered. Early on in transcribing the information, data entry operators noted that there were a great many changes made on the original schedules in the race column. The enumerator had written in one race and someone had changed it in subsequent editing, probably someone in Washington, D.C. The volume of the changes seemed odd. Up until 1960 when self-enumeration began, the Census Bureau enumerated race as a social category defined by the local community. The enumerator, as a representative of the community, simply assigned each person's race. Given this "social" definition, the volume of changes seemed unusual.

The rules for editing race, however, depended on the instructions to enumerators and their "logical" extensions. The instructions to enumerators for 1910 said,

> For census purposes, the term "black" (B) includes all persons who are evidently full-blooded negroes, while the term "mulatto" (Mu) includes all other persons having some proportion or perceptible trace of negro blood.

To the people in Washington writing the editing rules, the implications of the "blood" concept in this statement are that one black or mulatto parent and one white

parent must produce a mulatto child and further that two parents of a single race must produce a child of that race. Under this "blood" concept, race is very much a family thing. Insisting on those implications reduces the "social definition" component - and hence the enumerators choice in the assignment - of race to persons living without their putative parents in the family.

Apparently Puerto Rican enumerators classified people phenotypically rather than according to the implications of the "blood" concept. A person was black if they looked black, white if they looked white and mulatto if they looked mulatto without regards to what their parents were. We suspect the enumerators are reflecting Puerto Rican cultural practices in so doing. Therefore, in transcribing the data we decided to maintain the enumerators original race assignment where possible and to note the fact of an ignored change by Census editors in the "comments" variable for race. The result of this decision is to produce more mulattoes in our sample than are found in the census publications.

Family
Several aspects of the Census definition of family seem especially likely to be contingent on cultural practices. First is the dependence of family membership on co-residence. Given that the historic census was unable to deal with relationships across space except by aggregation, *i.e*., by saying that two persons live in the same household, town, county, state, *etc*., making the break at the household seems as reasonable as anything else. The problem is that it makes family membership dependent on the boundaries drawn around a household. Clearly there can be two households in a single building. Whether there can be several buildings for a single household is more tricky. To some extent it depends on how "addresses" are assigned, which is a real mare's nest. In Puerto Rico in 1910 the Census found that some kinds of urban compounds – *solares* - contained several structures and several households at a single address. The Bureau constructed and ah hoc procedure to indicate the separate structures and households at the same address. Our initial experience with the forms resulting from enumeration of these units suggested that the ad hoc rules were poorly followed. We decided, therefore, to include all persons living in a sampled *solares*, indicating which household would have been included had we not taken them all. This decision means that there are data for the investigation of this interesting housing arrangement.

A second issue is the dependence of family membership on each person's relationship to a single "household head." Here again, the problem is not the method but the definitions. Kinship is a complex and extensive web. Tracing the socially acknowledged relationships of a group of people to a single reference person is a manageable way to capture some part of the web. The choice of the reference person is, however, fairly consequential. Whether the oldest male in the household is a good choice for all groups is questionable. Another problem is that kinship terms are notably various across cultures. In some cases, there simply is not an English term for recognized kin words in Spanish. For example the Spanish term, "*concuñada*" can only be translated as "wife of one's spouse's brother." Kinship is defined as a relationship by "blood" or marriage, but an adopted child is kin while a god-child is not. Here again, meanings are different between the mainland and Puerto Rico.

Geography

Typically the U.S. Census activity begins by collecting recent maps of the area to be enumerated. Boundaries are determined for each of the areal units for which tabulations will be made in the census. Sometimes these units are neatly nested one within another but often they are not. From the overlap of the boundaries of these "official" areal units, ad hoc units are constructed such that any official unit for which a tabulation is desired can be produced by aggregation of the ad hoc units. Each of these ad hoc units becomes one or more enumeration districts. The larger ones become several enumeration districts. Each enumeration district is assigned to a single enumerator to survey. Tallies of persons by characteristics are first calculated for these enumeration districts. These enumeration district tallies are then aggregated to yield tables for various kinds of areas. If one is lucky, maps of these enumeration districts survive. For Puerto Rico in 1910 we have been unable to uncover them. Neither have we been able to uncover reliable maps for the boundaries of some official units. The census publication for Puerto Rico in 1910 contains a map showing boundaries for *municipios* but there is no map for *barrios*. Indeed, *barrio* was not a political unit in 1910 and its boundaries may have had no legal description. This has been a problem for us as we have tried to discover why our counts for some *barrios* are at considerable variance with the published ones. If, in the tallying process, enumeration districts were shifted, intentionally or accidentally, between official units, we would have no way of uncovering the fact.

Ownership

In 1910 capitalist institutions were less developed in Puerto Rico than in the United States. Questions of ownership of property and whether the property was mortgaged were answered less well than other questions. Perhaps the most dramatic issue in this arena arose, however, when the editing and allocation team at IPUMS wanted to infer that all owned property for which the mortgage question was blank were mortgaged. We insisted that if these blanks were to be filled in it should be by the hot-deck procedure whereby one looks for a nearby, similar household with both variables answered and substitutes its value for the missing one.

Summary of Comments on the Taking of a U.S. Census in Puerto Rico in 1910

The point of these observations is that the application of a practiced census to a new area can be revealing about both the new and the old areas. Another implication is also worth noting. A professional census bureau will learn a lot in its first trial in a new area. Comparison of the first and second instances, for example the 1910 and 1920 U.S. Censuses of Puerto Rico, are likely to confound social change and bureaucratic learning.

Conclusion

All of us involved in this project have labored to accomplish two somewhat contradictory goals simultaneously; to produce a modern statistical file that is as true to the time and place as possible and also to create a resource which permits cross-time and cross-place comparisons. The degree to which we have succeeded will only be told as scholars use this new resource in subsequent research.

References

Acosta-Belén, Edna. 1986. *The Puerto Rican Woman: Perspectives on Culture, History, and Society*, 2nd ed. New York: Praeger.

Palloni, Alberto, Francisco Scarano, and Halliman Winsborough. 2000. "Public Use Samples of 1910 & 1920 Puerto Rican Censuses." Grant Application to the Department of Health and Human Services Public Health Service.

Ruggles, Steven and Russell R. Menard. 1994. *Public Use Microdata Sample of the 1880 United States Census of Population: User's Guide and Technical Documentation.* Inter-University Consortium for Political and Social Research.

Ruggles, Steven and Russell R. Menard. 1995. "The Minnesota Historical Census Projects." *Historical Methods* 28: 6-10.

Ruggles, Steven and Matthew Sobek. 1995. "Integrated Public Use Microdata Series: User's Guide." Social History Research Laboratory.

Scarano, Francisco A. 1990. "Estructuras de la Plantacion Azucarera Esclavista: El Modelo Clasico y Sus Variaciones." *Del Caribe (Santiago de Cuba)* VI(16-17): 6-14.

Steward, Julian H. 1956. *The People of Puerto Rico: A Study in Social Anthropology.* Urbana, IL: University of Illinois Press.

Wagley, Charles. 1957. "Plantation America: A Culture Sphere." In *Caribbean Studies. A Symposium.* Ed. Vera Rubin. Mona, Jamaica, University College of the West Indies, Institute of Social and Economic Research, in association with Columbia University Program for the Study of Man in the Tropics [1957].