IPUMS Training and Development: Requesting Terra Data as Microdata

Microdata Exercise

Objective: Use IPUMS Terra to obtain a customized dataset that can be used to answer research questions. This exercise will use microdata and aggregate data for the United States spanning two decades to explore trends in education and employment.

IPUMS Terra: Microdata Extract Overview

RESEARCH QUESTION

Examine the interoperable (microdata and area-level) datasets for United States over three decennial censuses. What trends exist over time in education and unemployment?

OBJECTIVES

- **Create a IPUMS Terra account**
- **Create and download a IPUMS Terra microdata extract**
- **Use IPUMS Terra to attach area-level data to microdata records**

IPUMS TERRA VARIABLES

Microdata variables

* COUNTRY:	Country
EDUCUS:	Educational attainment, United States
EMPSTAT:	Activity status (employment status)
GEOLEV2:	2nd subnational geographic level, world
HHWT:	Household weight
* PERNUM:	Person number
PERSONS:	Number of person records in household
* PERWT:	Person weight
* SAMPLE:	IPUMS sample identifier
SCHOOL:	School attendance
* SERIAL:	Household serial number
* YEAR:	Year

* Denotes preselected variables

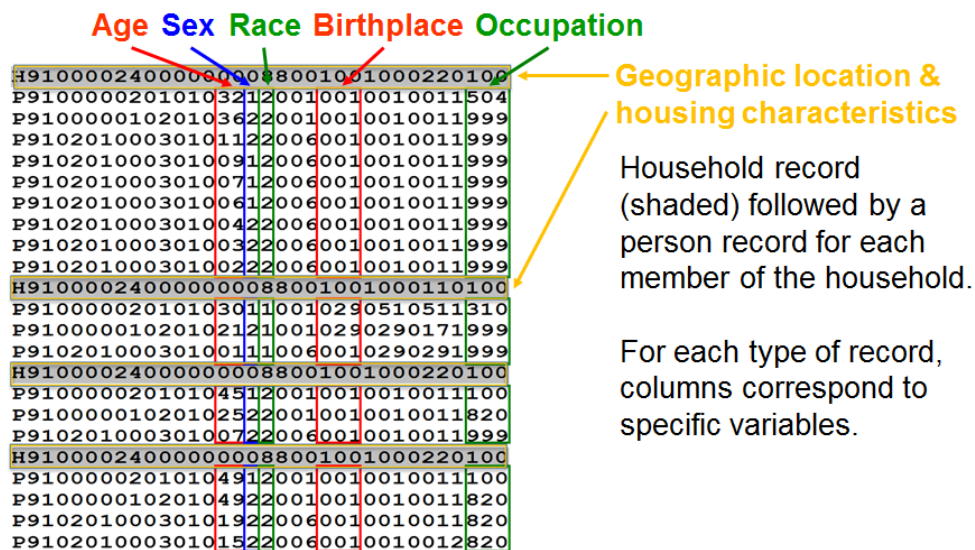### Area-level variables

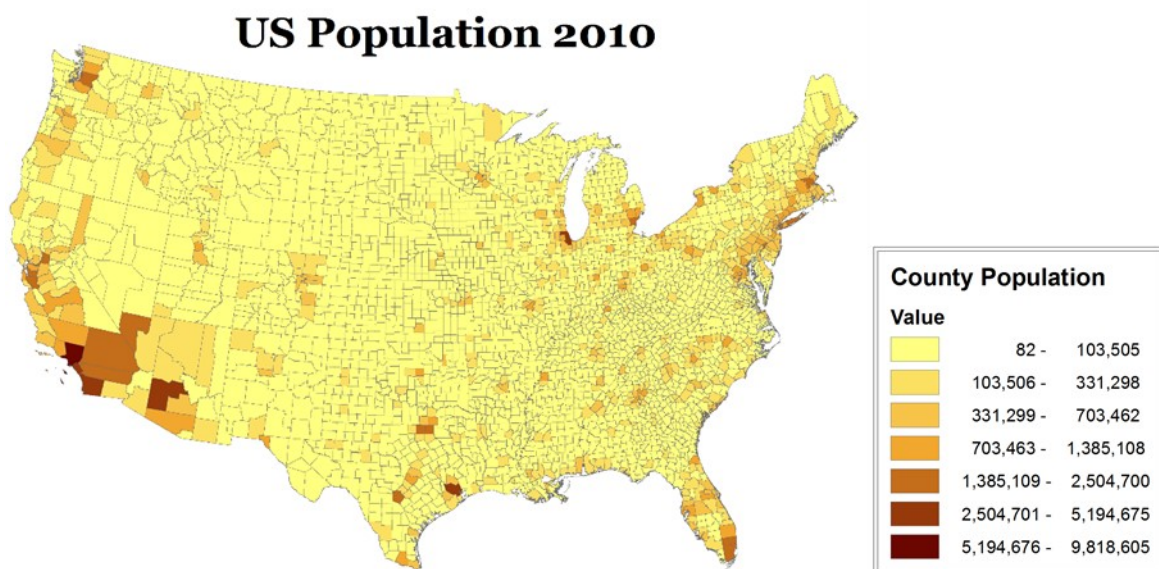|	
UNEMP:	Percent of population unemployed in each census area
PROF:	Percent of population in professional or technical occupation
EDATTAIN:	Percent of population with a specific educational attainment in each census area

DATA TYPE DESCRIPTIONS

Microdata

Microdata provide individual-level information describing characteristics such as age, sex, and occupation. Each microdata record includes the responses to all questions asked in a census or survey for a specific individual. Individual records are nested within households, with full information about the interrelationships of the members of each household. Each dataset is a sample of households from a particular census or survey.

Geographic location & housing characteristics

Household record (shaded) followed by a person record for each member of the household.

For each type of record, columns correspond to specific variables.

Area-level

Area-level data describe geographic units defined by boundaries. Geographic units are grouped in sets, such as the counties of the United States or the states of Brazil. In IPUMS Terra, these sets of units are referred to as geographic levels. The data are structured as tables, with a row for each unit and a column for each variable. For example, you may have a table with a row for each county in the United States and columns containing the number of males and the number of females in each county.

US Population 2010

County Population

Value

82 -	103,505
103,506 -	331,298
331,299 -	703,462
703,463 -	1,385,108
1,385,109 -	2,504,700
2,504,701 -	5,194,675
5,194,676 -	9,818,605

Step 1
Sign up

IPUMS uses a **common user management system** covering all IPUMS products. If you have an existing account with any IPUMS product, you will use the same account for IPUMS Terra.

- Go to *https://data.terrapop.org/*
- If you have an existing IPUMS account, click *Login*. After logging in, you will be directed to the registration page for IPUMS Terra.
- If you do not have an IPUMS account, click on *Sign up* to register for access.

Note: All registered IPUMS Terra users have access to U.S. microdata, since IPUMS Terra provides the public version of the microdata made available by the U.S. Census Bureau.

To access microdata for other countries, you must submit an application to IPUMS International. You can do so by clicking the box next to "Check here if you would like access to Microdata" on the IPUMS Terra registration form, and filling out the additional fields that appear.

Approval by IPUMS International may take one or two days. If you already have an IPUMS International account, your IPUMS Terra account will automatically be linked to it, and you will be able to access international microdata through IPUMS Terra. For more help see page 16.

Step 2
Email confirmation and log in

After you have registered with IPUMS Terra, an e-mail will be sent to your account notifying you of approval.
Note: Please be sure to check your trash/spam folders

- Open the e-mail and click on the confirmation link. You will then be logged into IPUMS Terra

Step 3
Start a microdata extract

The extract builder website guides researchers through the workflows for building data extracts. Choose the type of data structure you would like to receive as output for further analysis. In this tutorial, we will be creating a microdata extract with attached area-level variables.

- **Click on the *Start Extract* button for Microdata Output**

Microdata Output
characteristics of individual people with attached contextual variables derived from area-level and/or raster data
Read more
Start Extract

Area-level Output
characteristics of geographic units including aggregate population data and/or summaries from raster data
Read more
Start Extract

Raster Data Output
data in spatial grids potentially derived from area-level data
Read more
Start Extract

| **Step 4**
Examine the
IPUMS Terra
interface | **The IPUMS Terra interface for the first step of the workflow consists of the following elements:**

Navigation Bar: Shows the major steps in the workflow, the sub-steps of the current step, and your progress through the workflow. The navigation bar steps will reflect the workflow you select. The step and sub-step you are currently on are highlighted in green.

Data Cart: Provides a summary of the data you have selected to include in your extract. The data cart is updated as you make selections throughout the workflow.

Availability Grid: Shows the availability of variables by dataset and enables selection of variables and datasets.

Variables Panel: Lists topics for which microdata variables are available. Clicking on a topic will populate the rows of the availability grid with the variables in that topic.

Datasets Panel: Lists countries in the IPUMS Terra system, and provides options to filter by time and hide countries without microdata. Clicking on a continent will list the countries in the continent. Clicking on a country will populate the columns of the availability grid with the datasets available for that country. You may add all countries in a continent to the grid by clicking the "Browse All" line. |

Step 5

Browse datasets for the United States

- **Click on _North America_ in the Browse Datasets panel.**

The listing 'United States (8)' indicates that 8 years of data are available for the US.

- **Click on _United States._**

Step 6

Select datasets

The available datasets for United States appear as columns in the availability grid.

- **Check the boxes to select years: _2000, 2010._**

Step 7
Select microdata variables

Browse Variables

Household 🛒6   Person 🛒3

Technical Household 🛒 6

Group Quarters

Geography: Global

Geography: A-E

To see available variables, choose a topic in the Browse Variables panel. Topics are organized into two categories, Household and Person, accessible via the tabs in the Browse Variables panel.

- **Under the Household tab, choose the variable topic *Technical Household*.**

The availability grid will be updated with

Six of the technical household variables are preselected, as indicated by the cart icon on the topic name and check marks in the availability grid. These variables are usually important for analysis and are preselected for convenience. These variables can be removed if you deem they are unnecessary.

Step 8
Select household variables

- **Remove the variable *FORMTYPE* since it is not available for our selected datasets.**

Select Microdata

Select Data What is this?

Technical Household Variables		1960	1970	1980	1990	2000	2005	2010	2015
☐ RECTYPE	Record type	●	●	●	●	●	●	●	●
☑ COUNTRY *Preselected*	Country	●	●	●	●	✔	●	✔	●
☑ YEAR *Preselected*	Year	●	●	●	●	✔	●	✔	●
☑ SAMPLE *Preselected*	IPUMS sample identifier	●	●	●	●	✔	●	✔	●
☑ SERIAL *Preselected*	Household serial number	●	●	●	●	✔	●	✔	●
☐ PERSONS	Number of person records in the household	●	●	●	●	●	●	●	●
☑ HHWT *Preselected*	Household weight	●	●	●	●	✔	●	✔	●
☐ SUBSAMP	Subsample number	●	●	●	●	●	●	●	●
☐ HHDONATE	Donated household								
☑ FORMTYPE *Preselected*	Questionnaire type received by household								
☐ STRATA	Strata identifier					●			

The green dots for unselected variables indicate they are available in the indicated samples.

- **Choose the additional variable *PERSONS***

PERSONS will be added to your Data Cart.

Step 9

Select household variables

- **Select the Household topic *Geography Global***

You can simplify the availability grid to show only the variables available in the datasets you are browsing.

- **Uncheck "*Display unavailable variables*"**
- **Select *GEOLEV2***

Select Microdata

Select Data What is this?

Show only selected variables ❶

Show only selected datasets ❶

Geography: Global Variables

☐ Display unavailable variables ❶

		Datasets							
		United States							
		1960 ☐	1970 ☐	1980 ☐	1990 ☐	2000 ☑	2005 ☐	2010 ☑	2015 ☐
☐ URBAN	Urban-rural status	●			●				
☐ REGIONW	Continent and region of country	●	●	●	●	●	●	●	
☐ GEOLEV1	1st subnational geographic level, world [consistent boundaries over time]	●	●	●	●	●	●	●	●
☑ GEOLEV2	2nd subnational geographic level, world [consistent boundaries over time]					✔	●	✔	●

Step 10

Select person variables

Browse Variables

Household 🛒6   Person 🛒3

Technical Person 🛒 3

Constructed Family Interrelationship

Demographic

Fertility and Mortality

- **Under the Person tab, select the variable topic *Technical Person*.**

Three variables in this category are preselected.

- **Unselect the variable *RESIDENT* as the data is unavailable (no green dots) for selected datasets. (You may need to check "Display unavailable variables")**

Continue adding variables to your extract by selecting the following Person variables:

- Education → *EDUCUS*
- Education → *SCHOOL*
- Work → *EMPSTAT*

| **Step 11**	When you have selected both datasets and variables, the NEXT button will become active and turn green, allowing you to move on to the next step.

Step 11

Attach

area-level

data

When you have selected both datasets and variables, the NEXT button will become active and turn green, allowing you to move on to the next step.

The navigation bar indicates that the next step will be to Attach Area-Level Data.

- **Click *NEXT* to move to the Area-Level data screen.**

Note: Attaching area-level data will add contextual variables to your microdata records. The contextual variables contain information about the geographic unit in which the individual lives.
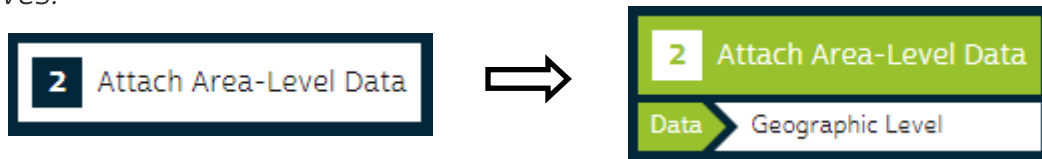
Step 12

Select

area-level

variables

The datasets for which you selected microdata are pre-selected for area-level data.

- **In the Browse Variables panel, click on *Education*. The availability grid will be updated with available education variables.**

Note: Area-level variables are organized in variable groups, which are similar to the columns in aggregate data tables. To see the individual variables within a group and access their metadata, use the arrow to expand the group.

- **Choose the variable group *EDATTAIN*, by checking the multi-select box. EDATTAIN is available for both selected datasets.**

Note: All four variables within the EDATTAIN group will be added to your Data Cart.

Select the following variables from the topic *Employment*
- *UNEMP*
- *PROF*

- **Click *NEXT* to move to geographic selection**

Step 13

Select a harmonized geographic level

Our research question involves examining how education and employment have changed between 2000 and 2010. In order to study change over time, it is important to use geographic units with boundaries that are consistent over the time frame under study. In IPUMS Terra, such units are referred to as harmonized. If harmonized units are not used, apparent changes may be due to changing boundaries rather than actual changes in population characteristics.

- **Leave *Harmonized (consistent)* selected**
- **Click *NEXT***

Step 14
Skip attaching raster data

You have the option of selecting raster data to create additional contextual variables. The raster grid cells in each geographic unit are summarized to produce a value for the geographic unit, which is then attached to microdata records for the individuals in the unit.
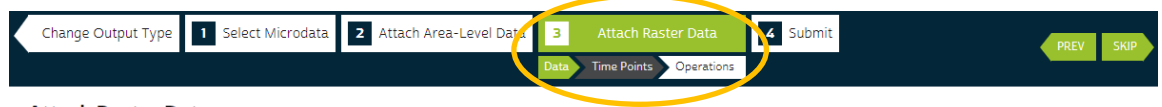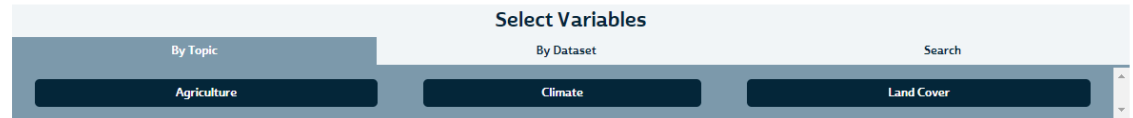
This exercise does not involve raster data so we can just skip this step.

- **Click *SKIP* to go to the Submit step**

Step 15
Check data Cart

Before submitting your extract request, you should check your data cart. Two sections of the cart should be populated.

- **Microdata should have 12 variables, 2 datasets**
- **Area-level type should have 6 variables, 2 datasets.**

You can click the number of variables or datasets for each data type to see detailed lists of your selections.

Microdata Extract Cancel

1 | Select Microdata ▼
    Variables   **Datasets**
    12   2
2 | Attach Area-level Data ▼
    Variables   **Datasets**
    6   2

Geographic Levels

Countries
United States (2000, 2010)

Time Frame
Harmonized (Consistent)

Administrative Level
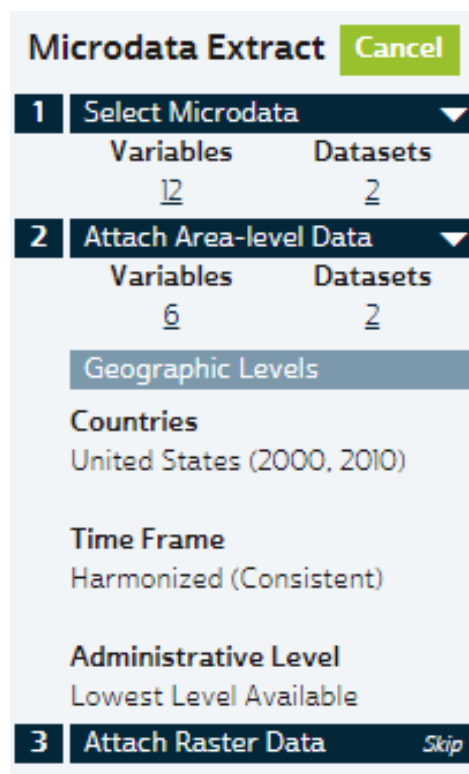Lowest Level Available
3 | Attach Raster Data *Skip*
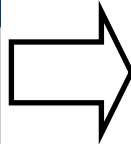
Microdata

Variables

COUNTRY	Country
YEAR	Year
SAMPLE	IPUMS sample identifier
SERIAL	Household serial number
PERSONS	Number of person records in the household
HHWT	Household weight
GEOLEV2	2nd subnational geographic level, world [consistent boundaries over time]
PERNUM	Person number
PERWT	Person weight
SCHOOL	School attendance
EDUCUS	Educational attainment, United States
EMPSTAT	Activity status (employment status)

Datasets

United States	2000
United States	2010

Microdata Extract Cancel

1 | Select Microdata ▼
    Variables   **Datasets**
    12   2
2 | Attach Area-level Data ▼
    Variables   **Datasets**
    6   2
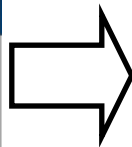
Geographic Levels

Countries
United States (2000, 2010)

Time Frame
Harmonized (Consistent)
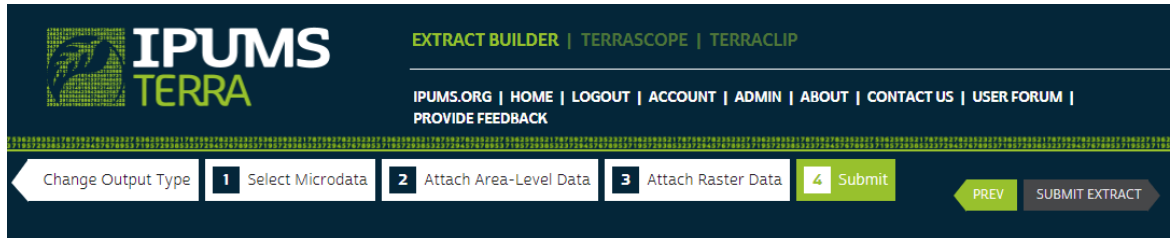
Administrative Level
Lowest Level Available
3 | Attach Raster Data *Skip*

Area-level Data

Variables

EDUCLESSPRIM	Percent of persons age 25+ with less than primary education
EDUCPRIMARY	Percent of persons age 25+ who completed primary education
EDUCSECOND	Percent of persons age 25+ who completed secondary education
EDUCTERTIARY	Percent of persons age 25+ who completed tertiary education
OCCPROF	Percent of workers in professional or technical occupations
UNEMPLOY	Unemployment rate

Datasets

United States	2000
United States	2010

Step 16
Submit

your extract

- Give your extract a short, descriptive Extract Title, maybe, "U.S. education and employment microdata." The Extract Title will appear in your Extract History.
- If you want the GIS shapefiles for the geographic levels used in your extract, check the box to *Include boundary files.* (Not required for this exercise.)
- Click *Submit Extract*
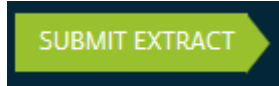
You will receive an email when the extract is ready.

Step 17
Download

your extract

ipums@umn.edu

to me ▾

Your IPUMS Terra extract 'U.S. education and employment microdata' is ready.

To retrieve your data, codebook, and command files, for Extract #3518, go to the link below.

https://data.terrapop.org/user/extract_requests

Thank you for your support.

Sincerely,
The IPUMS Terra Team

- **To download the data, follow the link in the e-mail, which will take you to your account's "Extract History" page as shown below.**

This page can also be accessed by clicking *ACCOUNT* in the page header

The data will be delivered in a compressed format, make sure you have software available to extract the files.

Optional IPUMS International Registration

IPUMS Terra users are able to access U.S. microdata without registering with IPUMS International. However, if you have international research interests the IPUMS International microdata collection will be of interest.

USAGE LICENSES

IPUMS TERRA USAGE LICENSE (AREA, RASTER AND BOUNDARY DATA)

By completing this registration, you agree to the following terms of use.

☐ **Redistribution: You will not redistribute the data without permission.** (Required)

> For IPUMS Terra area, raster and boundary data, you may publish a subset of the data to meet journal requirements for accessing data related to a particular publication. Contact us for permission for any other redistribution of this data.

☐ **Citation: Cite the IPUMS Terra data appropriately** (Required)

> For information on proper citation refer to citation and use. Publications and research reports making use of IPUMS Terra should be added to our Bibliography.

IPUMS INTERNATIONAL USAGE LICENSE (MICRODATA)

Please check all of the following boxes to indicate that you have read about the limitations of the IPUMS-International data and you agree to abide by the conditions of use. The purpose of this license is to specify the terms and conditions under which integrated microdata samples disseminated by IPUMS-International of the University of Minnesota may be used.

☐ **Redistribution: You will not redistribute the data.** (Required)

> All data extracted from the IPUMS-International database are intended solely for the use of the licensee. Under IPUMS-International agreements with collaborating agencies, redistribution of the data to third parties is prohibited. Each member of a research team using the data must apply for access and be licensed individually.

☐ **Usage: You agree to use IPUMS-International data for scholarly research and educational purposes only. Commercial use is strictly prohibited.** (Required)

> These microdata extracts are provided for the exclusive purposes of teaching and scholarly research, and may not be used for any other purposes without explicit written approval from the relevant official statistical authority. Users are prohibited from using IPUMS-International data or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise. The publishing of research results based on IPUMS-International microdata is permitted in communications such as scholarly papers, journals, research reports and the like.

☐ **Confidentiality: You will not use the data to identify individuals.** (Required)

> Users will maintain the confidentiality of persons and households. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data. Any attempt to ascertain the identity of persons or households from the microdata is prohibited. Alleging that a person or household has been identified in