

# GETTING AND VISUALIZING WORLDWIDE CENSUS AND SURVEY DATA

An introduction to IPUMS

Instructor: Victoria Prieto Rosas

Department of Sociology and Anthropology










[vprieto@clemson.edu](mailto:vprieto@clemson.edu)

## 1. Script

- 1.1. Which are the main IPUMS projects?
- 1.2. What does IPUMS?
- 1.3. Why is it useful for research on Health, Sociology, Economics, Demography, Geography and Environmental Studies?
- 1.4. What kind of research questions might be addressed from IPUMS?
- 1.5. How to use it?
- 1.6. Download data from IPUMS / SPSS for aggregating data / Tableau for visualization

## 2. Which are the main IPUMS projects?

Ipums, a non-profit scientific initiative from the Minnesota Center for Population (University of Minnesota) has nine ongoing projects, all of them available at [www.ipums.org](http://www.ipums.org)

 <p>U.S. Census and American Community Survey microdata from 1950 to the present.</p> <p>VISIT SITE</p>	 <p>Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.</p> <p>VISIT SITE</p>	 <p>Census microdata covering 82 countries from 1960 to the present. <a href="#">IPUMS HARP</a> offers microdata from the 19th and early 20th centuries.</p> <p>VISIT SITE</p>
 <p>Demographic and Health Surveys integrated for analysis across time and space from 1980 to the present.</p> <p>VISIT SITE</p>	 <p>Tabular U.S. Census data and GIS boundary files from 1790 to the present.</p> <p>VISIT SITE</p>	 <p>Integrated data on population and the environment from 1960 to the present.</p> <p>VISIT SITE</p>
 <p>Historical and contemporary time use data from 1965 to the present.</p> <p>VISIT SITE</p>	 <p>Health survey data from the National Health Interview Survey from the 1960s to the present.</p> <p>VISIT SITE</p>	 <p>Survey data on the science and engineering workforce in the U.S. from 1993 to the present.</p> <p>VISIT SITE</p>



### IPUMS USA and IPUMS CPS

Projects for harmonization across time of American Community Surveys and Current Population Surveys

Time span CPS: 1962-2016

Time span US Census: 1850-2000 (every 10 years)

Time span ACS: 2001-2015 (annual)

Geo span: country/state/metro/city

**What can I find at IPUMS CPS/ACS that isn't available at US Census Bureau?** Harmonization and variables that might be comparable to other countries census waves



### Census data (IPUMS International)

Project for the harmonization across time and countries of census data.

Time span CPS: 1960-2010 (not available for all countries)

Countries: 82

Census: 277

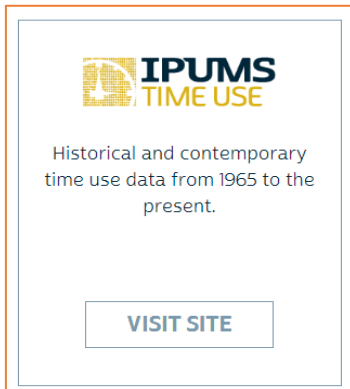
Person records: 614

Data: household/individuals/

Topics: housing, living arrangements, demographics, socioeconomics

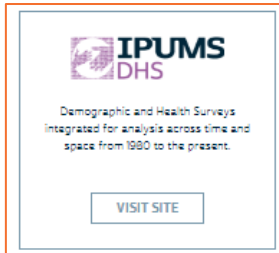
Geography: national/1<sup>st</sup> subnational /2<sup>nd</sup> subnational

### Time Use Surveys Harmonized (IPUMS TUS)



<p>— ATUS —</p> <p>AMERICAN TIME USE SURVEY EXTRACT BUILDER</p> <p>Annual American Time Use Survey (ATUS) data from 2003 forward.</p> <p><a href="#">GET DATA</a></p>	<p>— AHTUS —</p> <p>AMERICAN HERITAGE TIME USE STUDY EXTRACT BUILDER</p> <p>Historical American time use data since 1965 harmonized for comparison over time, including the ATUS samples.</p> <p><a href="#">GET DATA</a></p>	<p>— MTUS —</p> <p>MULTINATIONAL TIME USE STUDY EXTRACT BUILDER</p> <p>MTUS is a project dedicated to making it easy for researchers to use data from around the world.</p> <p><a href="#">GET DATA</a></p>
---	---	---

This project includes individual-level time use data extracted from diaries. The data extract systems make it easy to create data sets containing time use and other variables a user needs, i.e. you can build durations and time allocation on line and add them to the sociodemographic variables of your extracts. This is one of the newest, and therefore least explored projects.



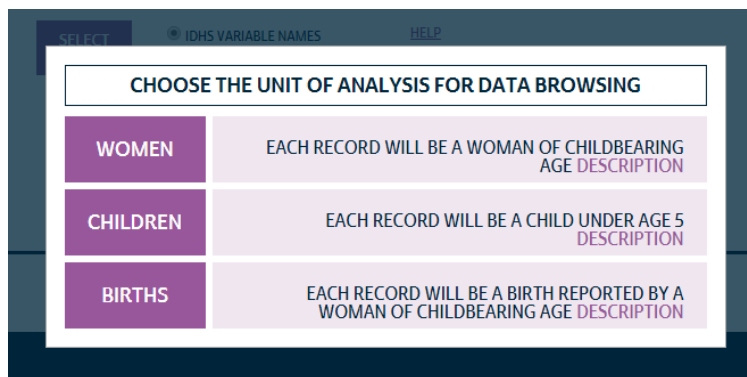
1. Benin
2. Burkina Faso
3. Cameroon
4. Cote d'Ivoire
5. Egypt
6. Ethiopia
7. Ghana
8. Guinea
9. India
10. Kenya
11. Madagascar
12. Malawi
13. Mali
14. Mozambique
15. Niger
16. Nigeria
17. Rwanda
18. Tanzania
19. Uganda
20. Zambia
21. Zimbabwe

**Demographic and Health Surveys (DHS IPUMS)** Harmonization across time and countries of DHS conducted in low- and middle-income countries for over 30 years. Include harmonized thousands of variables on health, well-being of women, children, and births.

You can decide if unit of analysis is women, children of births, and build de extract for those specific groups.

Time span: 1988-2014 (irregular timing by country, with some only having a few samples)

Countries: 21



### IPUMS NHGIS

The National Historical Geographic Information System (NHGIS) provides population, housing, agricultural, and economic data, along with GIS-compatible boundary files, for geographic units in the United States from 1790 to the present (for ACS this covers the shorter period 2011-15). This project is mainly devoted to providing GIS files where several layers of georeferenced data were harmonized across time using a time comparable shapes for the country, state, county, zip code and metro area.

However, IPUMS International also include a session for downloading the country, 1<sup>st</sup> and 2<sup>nd</sup> sub-national level shape data files (**Figure 1**).



#### 4. Why is it useful for research on Health, Sociology, Economics, Demography, Geography and Environmental Studies?

Several research questions demanding international comparison on a broad diversity of socio-economic indicators could be addressed using IPUMS.

Let's take for example scholars at the Department of Parks, Recreation and Tourism Management, which might be interested in studying sedentary behaviors. Now just have a look at some of the papers published under the search "SEDENTARY" that have used IPUMS data:

- Boehmer, Tegan K.; Brownson, Ross C.; Luke, Douglas A.  
2005. "Declining Rates of Physical Activity in the United States: What are the Contributors?" *Annual Review of Public Health* 26: 421-443
- Brownson, Ross C.; Boehmer, Tegan K.  
"Patterns and Trends in Physical Activity, Occupation, Transportation, Land Use, and Sedentary Behaviors." Report for the Transportation Research Board, 2004
- Cockburn, Myles; Mack, Thomas; Hamilton, Ann; Hawkins, Steve  
2004. "Estimate of Physical Activity Prevalence in a Large Population-based Cohort." *Medicine & Science in Sports & Exercise* 36: 253-260
- Mathew Joseph, Nitha,; Bishop, Sheryl, L  
2014. "Self-Reported Physical Activity, Sedentary Behavior and Body Mass Index Among US Asian Indian Women." *Journal of Medical Research and Practice* 3: 63-70
- Pepin, Joanna, R; Sayer, Liana, C; Casper, Lynne, M  
2015. "Marital Status and Mothers Time Use: Child Care, Housework, Leisure, and Sleep." Presented at Population Association of America, San Diego, CA
- Vargas, Andres J.  
2011. "BMI, Physical Activity and Sedentary Behaviors Among Mexican Immigrants to the US: A Time Use Perspective." Texas Tech University Working Paper Series

Figure 2 . IPUMS Bibliography – papers based on IPUMS Data

The screenshot displays the IPUMS Bibliography search interface. On the left, there is a search form with fields for Keywords, Title, Author, Journal, Citation types (set to Journal Article), Year published (set to 2014), Topic (set to Migration and Immigration), and Project (set to IPUMS-International). A 'Search' button is visible at the bottom of the form. On the right, the search results are displayed, showing a list of papers with their authors, years, titles, and journal information. The results are sorted by 'Authors, Primary'. The list includes papers by Abramitzky et al. (2014), Alesynska et al. (2014), Allen et al. (2014), Andersson et al. (2014), Aptekar (2014), Bankston et al. (2014), Bernard et al. (2014), Bernard et al. (2014), and Blumenberg et al. (2014).

These are just some examples taken from the [IPUMS repository of papers](#) published using IPUMS. There you may find an extensive collection of papers addressing the most diverse research questions. The way to use this Bibliographic section is very easy.

The users constantly feed this section of IPUMS since one of the terms of agreements of using IPUMS is to cite the source and to report IPUMS once you have published. This has become a powerful literature browser, such as scholar google of others, so do not hesitate in including it in your research toolbox.

Take the example of someone interested in finding publications on “migration and immigration” published on “2014” as “Journal Article” using data from “any IPUMS projects” (Figure 2).<sup>1</sup>

---

<sup>1</sup> You can select a specific project used for the data of the bibliography you are looking for (**Figure 2**).

---

## 5. How to use IPUMS?

There are two ways of using IPUMS INTERNATIONAL:

- use the online tabulator
- build your own extract of microdata

However, before going any further...

- 1.1. Log in at [www.ipums.org](http://www.ipums.org)
- 1.2. Select the IPUMS project you will be working-on. For today we will pick up IPUMS International

---

### A - Exercise – Using IPUMS on-line tabulator

---

Estimate the share of the foreign-born population by sex, living in every Latin American country 2010 census waves.

- 1- Let's have a quick consider the online tabulator at <https://international.ipums.org/international/sda.shtml>

The screenshot shows the IPUMS International website. On the left is a navigation menu with sections: PROJECT (ABOUT IPUMS-I, HOW TO CITE IPUMS-I, USER REGISTRATION AND LOGIN), DATA (BROWSE AND SELECT DATA, DOWNLOAD YOUR DATA EXTRACT, GEOGRAPHY AND GIS, SUPPLEMENTAL DATA FILES, - ANALYZE DATA ONLINE -), SAMPLES (SAMPLE DESCRIPTIONS, SOURCE DOCUMENTS, VARIANCE ESTIMATION), and RESOURCES (HELP, USER FORUM, BIBLIOGRAPHY, INTERNATIONAL PARTNERS, TEACHING WITH IPUMS). An orange arrow points from the '- ANALYZE DATA ONLINE -' link to the 'GET DATA' button. The main content area has the title 'HARMONIZED INTERNATIONAL CENSUS DATA FOR SOCIAL SCIENCE AND HEALTH RESEARCH'. Below the title is a description of the project and its goals, followed by statistics: '82 countries - 277 censuses - 614 million person records.' and a note that source data are provided by participating National Statistical Offices. At the bottom, there are two buttons: 'CREATE AN EXTRACT' (containing 'GET DATA') and 'CREATE AN ACCOUNT' (containing 'REGISTER'). Below these buttons is a section titled 'WHAT IS IPUMS?' with a brief description of the project's mission.

You can analyze Single-Sample Datasets or Multi-sample/ Select ALL LATIN AMERICAN SAMPLES.

2. **Tabulate on-line** in response to exercise just described. For this exercise, please, select **nativity** from Demographics into a ROW, country sample from GLOBAL GEOGRAPHY in COLUMNS, **year** 2010-11 as FILTER, CONTROL by **sex**, select if you want PERCENTAGES by color row, and check for WEIGHTS. Run table.

Figure 3. IPUMS Tabulator

Variable Selection: [Help](#)

Selected: year

Copy to:

Mode:  Append  Replace

All Latin America and Caribbean

- Household - Technical Household
  - country - Country
  - year - Year
  - sample - IPUMS sample identifier
  - persons - Number of person records in the household
  - hwt - Household weight
  - subsamp - Subsample number
  - hdonate - Donated household
- Household - Group Quarters
- Household - Geography: Global
  - urban - Urban-rural status
  - region - 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]
- Household - Geography: A-L
- Household - Geography: M-Z
- Household - Household Economic
- Household - Utilities
- Household - Appliances, Mechanicals, Other Amenities
- Household - Dwelling Characteristics
- Household - Other Household

SDA Frequencies/Crosstabulation Program  
Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify  
Row: country

OPTIONAL Variable names to specify  
Column: nativity

Control: sex

Selection Filter(s): year(2010-2011) Example: age(18-50)

Weight: pwt - Person weight

TABLE OPTIONS

Percentaging:  
 Column  Row  Total  
 Confidence intervals Level: 95 percent  
 Standard error of each percent

Unweighted  Weighted

Summary statistics   
 Question text  Suppress table   
 Color coding  Show Z-statistic   
 Include missing data values

CHART OPTIONS

Type of chart: (No Chart)

Bar chart options:  
 Orientation:  Vertical  Horizontal  
 Visual Effects:  2-D  3-D

Show percents:  Yes  
 Palette:  Color  Grayscale  
 Size - width: 600 height: 400

Title:

Figure 4. Outcomes by sex for Latin American census 2010-2011

Statistics for sex = 1(Male)					Statistics for sex = 2(Female)						
Cells contain: -Row percent -Weighted N	nativity				ROW TOTAL	Cells contain: -Row percent -Weighted N	nativity				ROW TOTAL
	1 Native-born	2 Foreign-born	9 Unknown/missing				1 Native-born	2 Foreign-born	9 Unknown/missing		
32: Argentina	95.8 18,472,170.0	4.2 808,800.0	.0 .0	.0 .0	100.0 19,280,970.0	32: Argentina	95.4 19,435,580.0	4.6 945,900.0	.0 .0	.0 .0	100.0 20,381,480.0
76: Brazil	99.7 93,113,705.4	.3 319,929.8	.0 .0	.0 .0	100.0 93,433,635.2	76: Brazil	99.7 97,117,230.1	.3 271,884.0	.0 .0	.0 .0	100.0 97,389,114.1
188: Costa Rica	91.0 1,908,740.0	9.0 187,690.0	.0 .0	.0 .0	100.0 2,096,430.0	188: Costa Rica	90.8 2,002,120.0	9.2 202,270.0	.0 .0	.0 .0	100.0 2,204,390.0
214: Dominican Republic	94.9 4,494,370.0	5.1 241,500.0	.0 .0	.0 .0	100.0 4,735,870.0	214: Dominican Republic	96.7 4,547,990.0	3.3 153,980.0	.0 .0	.0 .0	100.0 4,701,970.0
218: Ecuador	98.6 7,078,620.0	1.4 99,580.0	.0 .0	.0 .0	100.0 7,178,200.0	218: Ecuador	98.7 7,208,510.0	1.3 95,620.0	.0 .0	.0 .0	100.0 7,304,130.0
484: Mexico	98.8 53,887,022.0	.9 490,020.0	.3 150,035.0	.0 .0	100.0 54,527,077.0	484: Mexico	98.9 56,801,450.0	.8 478,251.0	.3 153,361.0	.0 .0	100.0 57,433,062.0
591: Panama	95.5 1,636,240.0	4.2 72,370.0	.3 4,470.0	.0 .0	100.0 1,713,080.0	591: Panama	95.6 1,622,570.0	4.2 71,130.0	.3 4,400.0	.0 .0	100.0 1,698,100.0
630: Puerto Rico	91.9 1,640,354.0	8.1 144,910.0	.0 .0	.0 .0	100.0 1,785,264.0	630: Puerto Rico	91.7 1,776,842.0	8.3 160,027.0	.0 .0	.0 .0	100.0 1,936,869.0
858: Uruguay	94.1 1,484,190.0	2.2 34,450.0	3.7 59,060.0	.0 .0	100.0 1,577,700.0	858: Uruguay	94.2 1,607,240.0	2.5 42,130.0	3.4 57,180.0	.0 .0	100.0 1,706,550.0
COL TOTAL	98.6 183,715,411.4	1.3 2,399,249.8	.1 213,565.0	.0 .0	100.0 186,328,226.2	COL TOTAL	98.6 192,119,532.1	1.2 2,421,192.0	.1 214,941.0	.0 .0	100.0 194,755,665.1

Color coding: <-2.0 | <-1.0 | <0.0 | >0.0 | >1.0 | >2.0 | Z

N in each cell: Smaller than expected | Larger than expected

Now, why don't you suggest a question we may answer using the on-line tab?





Figure 7. Select variables from INTEGRATED variables (IPUMS)

IPUMS INTERNATIONAL | IPUMS.ORG | SELECT DATA | FAQ | HELP | LOGIN

DATA CART  
YOUR DATA EXTRACT  
3 VARIABLES  
18 SAMPLES  
VIEW CART

SELECT VARIABLES  
HOUSEHOLD PERSON A-Z SEARCH CHANGE SAMPLES

INTEGRATED VARIABLES  
 UNHARMONIZED VARIABLES

HELP  
DISPLAY OPTIONS  
COUNTRY ABBREVIATIONS

AN "X" INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET.

#% DEMOGRAPHIC VARIABLES -- PERSON [TOP]

Add to cart	Variable	Variable Label	Type	AR 2001	AR 2010	BR 2000	BR 2010	CR 2000	CR 2011	DO 2002	DO 2010	MX 2000	MX 2010	PA 2000	PA 2010	PR 2000	PR 2010	US 2000	US 2010	UY 1996	UY 2011	
<input type="checkbox"/>	RELATE	Relationship to household head	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	AGE	Age	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	AGE2	Age, grouped into intervals	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	SEX	Sex	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	MARST	Marital status	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	CONSENS	Consensual union	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	AGEMARR	Age at first marriage or union	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	DURMARR	Duration of current marriage or union	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	MARRNUM	Number of marriages or unions	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	SUBPREL	Relationship to head of subfamily	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	SUBFNUM	Subfamily membership number	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	BIRTHYR	Year of birth	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	BIRTHMO	Month of birth	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

SUPPORTED BY: NATIONAL INSTITUTES OF HEALTH NATIONAL SCIENCE FOUNDATION STAT/TRANSFER AND L

Figure 8. You can also select from UNHARMONIZED variables (original)

IPUMS INTERNATIONAL | IPUMS.ORG | SELECT DATA | FAQ | HELP

SELECT UNHARMONIZED VARIABLES  
HOUSEHOLD PERSON SAMPLE SEARCH CHANGE SAMPLES

INTEGRATED VARIABLES  
 UNHARMONIZED VARIABLES

AN "X" INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET.

← Previous 7 2 Next →

DEMOGRAPHIC VARIABLES -- PERSON (GROUP CONTINUED ON NEXT PAGE...) [TOP]

Add to cart	Variable	Variable Label	Type	Sample
<input type="checkbox"/>	AR2001A_RELATE	Relationship to head	P	Argentina 2001
<input type="checkbox"/>	AR2001A_SEX	Sex	P	Argentina 2001
<input type="checkbox"/>	AR2001A_AGE	Age	P	Argentina 2001
<input type="checkbox"/>	AR2001A_MARST	Marital status	P	Argentina 2001
<input type="checkbox"/>	AR2001A_COHAB	Cohabitation	P	Argentina 2001
<input type="checkbox"/>	AR2001A_MARST	First time union	P	Argentina 2001
<input type="checkbox"/>	AR2001A_MARST	Years in union	P	Argentina 2001
<input type="checkbox"/>	AR2010A_RELATE	Relationship to household head	P	Argentina 2010
<input type="checkbox"/>	AR2010A_SEX	Sex	P	Argentina 2010
<input type="checkbox"/>	AR2010A_AGE	Age	P	Argentina 2010
<input type="checkbox"/>	AR2010A_BIRTHMN	Birth month	P	Argentina 2010
<input type="checkbox"/>	AR2010A_BIRTHYR	Birth year	P	Argentina 2010
<input type="checkbox"/>	BR2000A_SEX	Sex	P	Brazil 2000
<input type="checkbox"/>	BR2000A_RELATE	Relationship to head of household	P	Brazil 2000
<input type="checkbox"/>	BR2000A_RELATE	Relationship to head of family	P	Brazil 2000
<input type="checkbox"/>	BR2000A_AGE	Age	P	Brazil 2000
<input type="checkbox"/>	BR2000A_AGEINMONTH	Age in months	P	Brazil 2000
<input type="checkbox"/>	BR2000A_SPOUSE	Live with spouse or partner	P	Brazil 2000
<input type="checkbox"/>	BR2000A_LSTUNION	Type of last union	P	Brazil 2000
<input type="checkbox"/>	BR2000A_MARST	Marital status	P	Brazil 2000
Add to cart	Variable	Variable Label	Type	Sample
<input type="checkbox"/>	BR2010A_RELATE	Relationship to head of household	P	Brazil 2010
<input type="checkbox"/>	BR2010A_SEX	Sex	P	Brazil 2010
<input type="checkbox"/>	BR2010A_AGE	Age	P	Brazil 2010

Don't miss the distinction between harmonized and un-harmonized variables. In the future if you are working with a single country -not oriented to international comparison- you might be interested in using the original country variables. Harmonization has the pro of comparison but the cons of losing precious information.

For example, Brazil is one of the few countries that include income at the census, or Uruguay asks for same – sex couple. Every country has its uniqueness, and IPUMS takes that into account.

Today we will use **INTEGRATED VARIABLES** as in **Figure 7**

If you click on the variable, you may see the codes description, full definition and even a view of the cases count (following image).

Figure 9. Selecting the variable for EDATTAIN

SELECT VARIABLES			CHANGE SAMPLES	INTEGRATED VARIABLES	UNHARMONIZED VARIABLES	HELP															
HOUSEHOLD	PERSON	A-Z	SEARCH			DISPLAY OPTIONS COUNTRY ABBREVIATIONS															
AN "X" INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET.																					
EDUCATION VARIABLES -- PERSON [TOP]																					
Add to cart	Variable	Variable Label	Type	AR 2001	AR 2010	BR 2000	BR 2010	CR 2000	CR 2011	DO 2000	DO 2010	MX 2000	MX 2010	PA 2000	PA 2010	PR 2000	PR 2010	US 2000	US 1996	UY 2011	
<input type="checkbox"/>	SCHOOL	School attendance	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	LIT	Literacy	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	EDATTAIN	Educational attainment, international recode	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
<input type="checkbox"/>	YRSCHOOL	Years of schooling	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
<input type="checkbox"/>	EDUCAR	Educational attainment, Argentina	P	X	X																
<input type="checkbox"/>	EDUCBR	Educational attainment, Brazil	P		X	X															
<input type="checkbox"/>	EDUCCR	Educational attainment, Costa Rica	P					X	X												
<input type="checkbox"/>	EDUCDO	Educational attainment, Dominican Republic	P							X	X										
<input type="checkbox"/>	EDUCMX	Educational attainment, Mexico	P									X	X								
<input type="checkbox"/>	EDUCPA	Educational attainment, Panama	P											X	X						
<input type="checkbox"/>	EDUCPR	Educational attainment, Puerto Rico	P													X	X				
<input type="checkbox"/>	EDUCUS	Educational attainment, United States	P															X	X		
<input type="checkbox"/>	EDUCUY	Educational attainment, Uruguay	P																	X	
<input type="checkbox"/>	LEFTSCH	Reason for leaving school	P						X		X										

Since we are interested in mapping our results, we need to add geographical data, by selecting HOUSEHOLD/GEOGRAPHY GLOBAL/ and the GEOLEVEL 1 VARIABLES FOR 1<sup>ST</sup> SUB-NATIONAL LEVEL for selected countries and years.

Figure 10. Add geo variables – recall we have already downloaded GIS shape files in Figure 1

SELECT VARIABLES			CHANGE SAMPLES	INTEGRATED VARIABLES	UNHARMONIZED VARIABLES	HELP														
HOUSEHOLD	PERSON	A-Z	SEARCH			DISPLAY OPTIONS COUNTRY ABBREVIATIONS														
AN "X" INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET.																				
GEOGRAPHY: GLOBAL VARIABLES -- HOUSEHOLD [TOP]																				
Add to cart	Variable	Variable Label	Type	AR 2001	AR 2010	BR 2000	BR 2010	CR 2000	CR 2011	DO 2000	DO 2010	MX 2000	MX 2010	PA 2000	PA 2010	PR 2000	PR 2010	US 2000	US 1996	UY 2011
<input type="checkbox"/>	URBAN	Urban-rural status	H	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	REGION	Continent and region of country	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	GEOLEV1	1st subnational geographic level, world [consistent boundaries over time]	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	GEOLEV2	2nd subnational geographic level, world [consistent boundaries over time]	H	X	X			X	X			X	X	X	X					

We make sure to have the ID variables and the WEIGHTS (Figure 11 AND Figure 12) for person and household (recall these are 1-10% samples of the original census). Moreover, once that is added we check out by clicking VIEW CART and click on CREATE DATA EXTRACT.

Figure 11. FROM SECTION HOUSEHOLD ADD TECHNICAL Variables

IPUMS INTERNATIONAL | IPUMS.ORG | SELECT DATA | FAQ | HELP | LOGIN

DATA CART  
YOUR DATA EXTRACT  
6 VARIABLES  
18 SAMPLES  
VIEW CART

SELECT VARIABLES  
HOUSEHOLD PERSON A-Z SEARCH CHANGE SAMPLES

INTEGRATED VARIABLES UNHARMONIZED VARIABLES  
HELP DISPLAY OPTIONS COUNTRY ABBREVIATIONS

AN "X" INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET.

TECHNICAL HOUSEHOLD VARIABLES — HOUSEHOLD [TOP]

Add to cart	Variable	Variable Label	Type	AR 2001	AR 2010	BR 2000	BR 2010	CR 2000	CR 2011	DO 2002	DO 2010	MX 2000	MX 2010	PA 2000	PA 2010	PR 2000	PR 2010	US 2000	US 2010	UY 1996	UY 2011
<input type="checkbox"/>	RECTYPE	Record type	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	COUNTRY	Country [preselected]	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	YEAR	Year [preselected]	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	SAMPLE	IPUMS sample identifier [preselected]	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	SERIAL	Household serial number [preselected]	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	PERSONS	Number of person records in the household	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	HIWT	Household weight	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	SUBSAMP	Subsample number	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input type="checkbox"/>	HDONATE	Donated household	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<input type="checkbox"/>	STRATA	Strata identifier	H	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Figure 12. FROM SECTION PERSON ADD TECHNICAL Variables

IPUMS INTERNATIONAL | IPUMS.ORG | SELECT DATA | FAQ | HELP | LOGIN

DATA CART  
YOUR DATA EXTRACT  
6 VARIABLES  
18 SAMPLES  
VIEW CART

SELECT VARIABLES  
HOUSEHOLD PERSON A-Z SEARCH CHANGE SAMPLES

INTEGRATED VARIABLES UNHARMONIZED VARIABLES  
HELP DISPLAY OPTIONS COUNTRY ABBREVIATIONS

AN "X" INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET.

TECHNICAL PERSON VARIABLES — PERSON [TOP]

Add to cart	Variable	Variable Label	Type	AR 2001	AR 2010	BR 2000	BR 2010	CR 2000	CR 2011	DO 2002	DO 2010	MX 2000	MX 2010	PA 2000	PA 2010	PR 2000	PR 2010	US 2000	US 2010	UY 1996	UY 2011
<input checked="" type="checkbox"/>	PERNUM	Person number [preselected]	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<input checked="" type="checkbox"/>	PERWT	Person weight [preselected]	P	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

## 2. Submit extract and prepare to download the data

Finally, we name the extract and customize or filter cases in case we are interested in, and we SUBMIT EXTRACT Figure 13. However, here you can select a particular group of cases, attach additional characteristics and reduce the size of the file.

Figure 13. Finish and submit extract

IPUMS INTERNATIONAL

EXTRACT REQUEST (HELP)

SAMPLES: 8 (show) Change

VARIABLES: 14 (show) Change

DATA FORMAT: Default (fixed-width text) Change

STRUCTURE: Rectangular (person) Change

ESTIMATED SIZE: 1M3.4 MB (How to reduce extract size)

OPTIONS

SELECT CASES: Include only specified cases (for example, persons age 60 and older).

ATTACH CHARACTERISTICS: Attach data from mother, father, spouse or household head as a new variable (for example, education of mother).

CUSTOMIZE SAMPLE SIZES: Specify the number of cases to include from each sample in your extract.

Describe your extract  
Revision of (COURSE\_12APRIL2017)

SUBMIT EXTRACT

Once the extract is sent, you will immediately have access to the syntax for SAS STATA SPSS that enables opening the extract (txt FORMAT). Please select the syntax for SPSS for today's workshop. A few minutes later an e-mail will let you know that the zip data file is ready to be download and you will find it right below where it states "Data"

Figure 14 – Download the data

Extract Number	Date	Formatted Data	Fixed-width Text Files					Revise Extract	Resubmit Extract	Description (click to edit)	Hide selections	
			Data	SPSS	SAS	STATA	Codebook					
64	2017-04-10	--	Data	SPSS	SAS	STATA	Basic	DDI	revise	resubmit	COURSE_12APRIL2017	<input type="checkbox"/>
63	2017-03-09								revise	resubmit	gis_exer	<input type="checkbox"/>
62	2017-02-06								revise	resubmit	Revision of (articulo6)	<input type="checkbox"/>
61	2017-02-06								revise	resubmit	articulo6	<input type="checkbox"/>

Also check you'll always be able to revise and resubmit your extract from the past.

### 3. Now open SPSS to visualize the microdata for your extract

Go to File/Open/Syntax, and get the SPSS syntax downloaded from your IPUMS extract "COURSE\_12APRIL2017". You will have to add the path to the data file into the first line of the code letting know SPSS that the data should be open from that directory and that it should open it by using all the labels and values coded below (Figure 15). Remember to unzip the data files downloaded from IPUMS. In this case, the .txt data is actually in file 'ipumsi\_00064.dat'.

Figure 15.a – Open the data in Spss

```

1 *Encoding: UTF-8.
2 * NOTE: You need to edit the 'cd' command to specify the path to the directory
3 * where the data file is located. For example: "C:\ipums_directory".
4 .
5 .
6 cd "D:\Getting_and_Visualizing_Worldwide_Census_Data".
7 .
8 data list file = "ipumsi_00064.dat" /
9 COUNTRY 1-3
10 YEAR 4-7
11 SAMPLE 8-16
12 SERIAL 17-26
13 GEOLV1 27-32
14 PERNUM 33-35
15 PERWT 36-43 (2)
16 AGE2 44-45
17 SEX 46-46
18 MARST 47-47
19 MARST 48-50
20 NATIVITY 51-51
21 EDATTAIN 52-52
22 EDATTAIN 53-55
23 .
24 .
25 variable labels
26 COUNTRY "Country"
27 YEAR "Year"
28 SAMPLE "IPUMS sample identifier"
29 SERIAL "Household serial number"

```

This will take a while, while data is getting extracted. Later check your e-mail, and you will see the data you have just ordered.

Figure 15.b - Once the data is ready the microdata file would look like this

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	COUNTRY	Numeric	3	0	Country	{32, Argenti...	None	9	Right	Unknown	Input
2	YEAR	Numeric	4	0	Year	{1960, 1960...	None	6	Right	Unknown	Input
3	SAMPLE	Numeric	9	0	IPUMS sample ...	{32197001, ...	None	11	Right	Unknown	Input
4	SERIAL	Numeric	10	0	Household seri...	None	None	12	Right	Unknown	Input
5	GEOLEV1	Numeric	6	0	1st subnational...	{2002, City...	None	9	Right	Unknown	Input
6	PERNUM	Numeric	3	0	Person number	None	None	8	Right	Unknown	Input
7	PERWT	Numeric	9	2	Person weight	None	None	11	Right	Unknown	Input
8	AGE2	Numeric	2	0	Age, grouped in...	{1, 0 to 4}...	None	6	Right	Unknown	Input
9	SEX	Numeric	1	0	Sex	{1, Male}...	None	5	Right	Unknown	Input
10	MARST	Numeric	1	0	Marital status [...	{0, NIU (not ...	None	7	Right	Unknown	Input
11	MARSTD	Numeric	3	0	Marital status [...	{0, NIU (not ...	None	8	Right	Unknown	Input
12	NATIVITY	Numeric	1	0	Nativity status	{0, NIU (not ...	None	10	Right	Unknown	Input
13	EDATTAIN	Numeric	1	0	Educational att...	{0, NIU (not ...	None	10	Right	Unknown	Input
14	EDATTAIND	Numeric	3	0	Educational att...	{0, NIU (not ...	None	11	Right	Unknown	Input
15											
16											

Now we will proceed to the construction of aggregated variables to map and visualize IPUMS data. Estimate and plot the share of female 20-29 living in a union by subnational level 1 in the Americas, first for 2000, second for 2010

#### 4. Create a new syntax file to preliminary steps for aggregation of data

There we will be coding to create the set of new dummy variables that will be useful to work with rates to be plotted in maps

Figure 16 –Create a new syntax file

	SERIAL	GEOLEV1	PERNUM	PERWT	AGE2	SEX	MARST	MARSTD	NATIVITY	EDATTAIN
200101	1000	32002	1	10.00	26	1	4	400	1	1
200101	2000	32002	1	10.00	18	2	1	100	2	2
200101	3000	32002	1	10.00	9	1	2	210	1	4
200101	4000	32002	1	10.00	18	2	1	100	1	3
200101	5000	32002	1	10.00	11	1	2	210	2	4
200101	6000	32002	1	10.00	9	2	1	100	2	4
200101	7000	32002	1	10.00	13	1	1	100	1	4
200101	8000	32002	1	10.00	13	2	1	100	2	3
200101	9000	32002	1	10.00	9	2	1	100	1	3
200101	10000	32002	1	10.00	12	1	1	100	2	4
200101	11000	32002	1	10.00	12	1	1	100	1	4
200101	12000	32002	1	10.00	12	1	1	100	2	3
200101	13000	32002	1	10.00	11	2	2	210	2	2
200101	14000	32002	1	10.00	28	2	4	400	1	3
200101	15000	32002	1	10.00	8	2	1	100	2	3
200101	16000	32002	1	10.00	15	1	2	210	2	4
200101	17000	32002	1	10.00	14	1	3	300	1	3
200101	18000	32002	1	10.00	15	1	4	400	2	2
200101	19000	32002	1	10.00	12	1	2	210	2	4
200101	20000	32002	1	10.00	9	1	4	400	2	4
200101	21000	32002	1	10.00	16	1	2	210	2	3
200101	22000	32002	1	10.00	17	1	3	300	2	4

On this new file, we will be copy-paste the following code (only sections within orange border lines).

## 5. Creation of the first set of dummy variables

---

\*Creation of variable female where women will take value 1 (former 2), and males value 0 (previous 1). The SYSMIS in SPSS stands for coding as missing values “.”

This and the following variables will be used in exercise 1.

---

```
RECODE SEX (1=0) (2=1) (9=SYSMIS) INTO fem.  
VARIABLE LABELS fem 'female'.  
EXECUTE.
```

---

\*Creation of an age group 20-29 as dummy variable. Since we have selected AGE in groups we have to use the labels for the ages while re-codifying into this new dummy AGE2029 variable.

---

```
RECODE AGE2 (0 thru 7=0) (8 thru 9=1) (10 thru 21=0) (99=sysmis) INTO age2029.  
VARIABLE LABELS age2029 'age2029'.  
EXECUTE.
```

---

\*Creation of a dummy for identifying the women that are or haven't ever been in any union (marriage or cohabitation). This is the key variable for Exercise 1: Have cohabitation been expanding from the 2000s to 2010s? The increase of cohabitation among female has been interpreted as one of the signs for the Second Demographic

Transition, which involves: more divorces, more cohabitation, fewer marriages, lower and later fertility, among other indicators of family change.

---

```
RECODE MARST (9=SYSMIS) (0=SYSMIS) (1=1) (2 thru 4=0) INTO nonunion.  
VARIABLE LABELS nonunion 'never union'.  
EXECUTE.
```

---

\*Creation of a dummy for identifying the highly-educated people

We are creating this variable to cross it in the future with non-in-union women, to explore if the speed of the cohabitation expansion in the Americas was larger among the most educated women than for the total female population aged 20-29.

---

```
RECODE EDATTAIN (9=SYSMIS) (0=SYSMIS) (4=1) (1 thru 3=0) INTO university.  
VARIABLE LABELS university 'university completed'.  
EXECUTE.
```

---

## 6. The importance of weights

Although IPUMS International provides data on CENSUS, we are working with 1%, 5% or 10% samples. Therefore, we need to weight the data. Weights in IPUMS are identified with suffix WT and work for both weighting and expanding. The command for weighting in SPSS follows. Recall using it for every frequency tabulation and check weights are on before aggregating data.

---

```
*weight the dataset.  
WEIGHT BY PERWT.
```

---

**\*Asking for a crosstab.**

CROSSTABS

/TABLES=university BY COUNTRY

/FORMAT=AVALUE TABLES

/CELLS=COUNT

/COUNT ROUND CELL.

Compare outcomes after weight is on and with weight off. Recall that SPSS shows if the data is weighted or not at the bottom right of the data view browser.

**university completed \* Country Crosstabulation – UNWEIGHTED**

Count

		Country				
		Costa Rica	Dominican Republic	Panama	Puerto Rico	Total
university completed	.00	342361	833053	279992	29233	1484639
	1.00	53831	55497	31270	5760	146358
Total		396192	888550	311262	34993	1630997

**university completed \* Country Crosstabulation – WEIGHTED**

Count

		Country				
		Costa Rica	Dominican Republic	Panama	Puerto Rico	Total
university completed	.00	3423610	8330530	2799920	3026041	17580101
	1.00	538310	554970	312700	566779	1972759
Total		3961920	8885500	3112620	3592820	19552860



## 7. Creation of the second set of variables by combining the recently created dummy variables

In this way, we will be preparing the numerator and denominator of our future shares of singlehood and immigration.

\*numerator.

```
COMPUTE nonunion_female_2029=fem = 1 & age2029 = 1 & nonunion = 1.
VARIABLE LABELS nonunion_female_2029 'Female2029_NonUnion'.
EXECUTE.
```

\*denominator.

```
COMPUTE female_2029=fem = 1 & age2029 = 1.
VARIABLE LABELS female_2029 'Female2029'.
EXECUTE.
```

\*numerator by education.

```
COMPUTE nonunion_female_2029_univ=fem = 1 & age2029 = 1 & university = 1 &
nonunion = 1.
VARIABLE LABELS nonunion_female_2029_univ 'Female2029university'.
EXECUTE.
```

\*denominator by education.

```
COMPUTE female_2029_univ=fem = 1 & age2029 = 1 & university = 1.
VARIABLE LABELS female_2029_univ 'Female2029university'.
EXECUTE.
```


## 8. Transform data from long to wide / Aggregate command in SPSS

This will be the most time-consuming part of your work if you are working with SPSS.

At this point, we have data in long format, but we would need to transform it into wide to respond our research questions (share of female with certain characteristics and proportion of immigrants by country, sub-national level, and year). Also, most of the GIS analysis is based on this format.

*Figure 17. Just an example. From long microdata to wide aggregated by date*

	A	B	C
1	ID	Date	Return
2	100500	01/01/2000	8
3	100500	02/01/2000	6
4	100500	03/01/2000	4
5	100500	04/01/2000	9
6	222622	02/01/2000	2
7	222622	03/01/2000	1
8	222622	04/01/2000	3
9	222622	05/01/2000	0



	A	B	C
1	date	return100500	return222622
2	01/01/2000	8	
3	02/01/2000	6	2
4	03/01/2000	4	1
5	04/01/2000	9	3
6	05/01/2000		0

Therefore, the transformation from microdata to aggregate data (long to wide) is necessary.

The idea is that you have several rows for each subject (country/county/state) and you want to transform these observations into a single line. In SPSS this process is called aggregate.

## 5.1. Let's go to DATA / AGGREGATE

Once this window appears, enter the variables that will be your new units of analysis into the "Break Variable" box, where you are expected to add the variables you will aggregate by. For the purpose of this exercise include YEAR, COUNTRY and GEOLEVEL1.

Then define your "Summaries of Variable" Our summary variables are in this case we will use This mean we will end up with a file with as many rows as years x country x geolevel1, where the variables would be the SUM of our selected indicators created from microdata. Remember you can always change the default summary operation on "Function" (MIN, MAX, MEAN, DEV, SUM, COUNT, etc.) according to your needs and the characteristics of the variables.

Figure 18. These are the steps to aggregate in SPSS using the menu

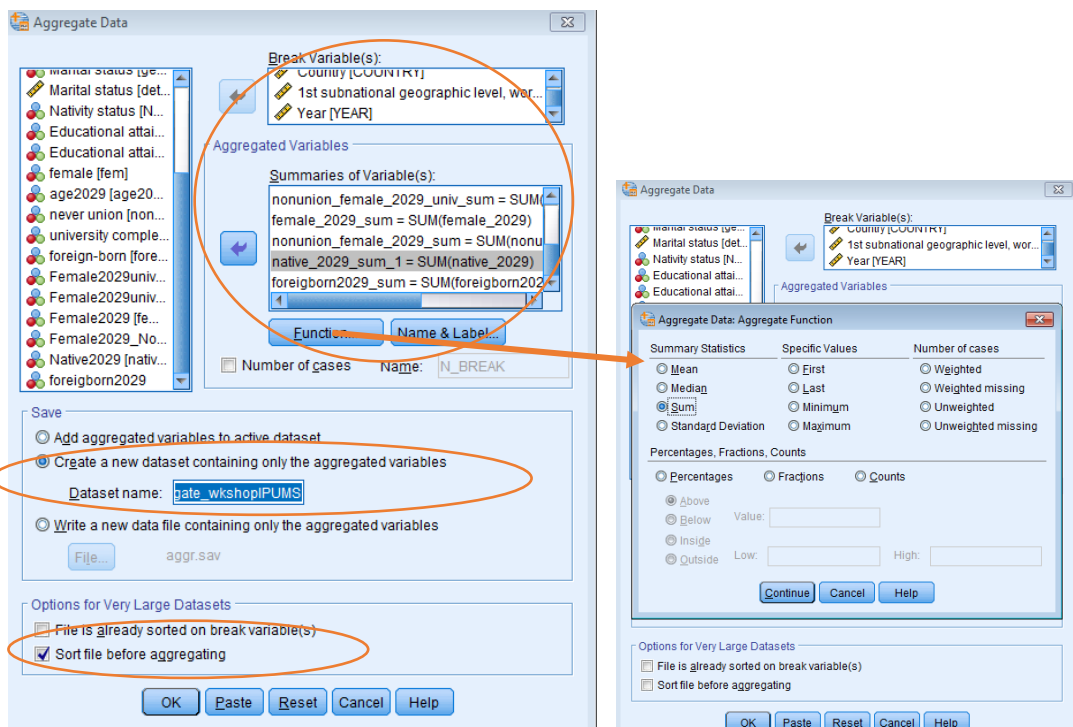


Figure 19. And this is the way your wide-data will look like after aggregating

	COUNTRY	GEOLEV1	YEAR	female_sum	age2029_sum	nonunion_sum	university_sum	foreign_sum	Female_2029_univ_sum	nonunion_female_2029_sum	female_2029
187	484	484025	2000	1277224.00	443630.00	1343755.00	133650.00	7183.00	23856.00	15313.00	22
188	484	484026	2000	1028426.00	378127.00	1158631.00	103756.00	17216.00	18367.00	13422.00	20
189	484	484027	2000	503720.00	345204.00	1823446.00	75035.00	1050.00	14288.00	8765.00	18
190	484	484028	2000	1306423.00	518378.00	1429169.00	147870.00	37133.00	25534.00	15847.00	26
191	484	484029	2000	489642.00	171588.00	523849.00	38726.00	884.00	5367.00	5293.00	3
192	484	484030	2000	3546720.00	1128220.00	3655153.00	252377.00	6161.00	48286.00	30560.00	61
193	484	484031	2000	833833.00	298284.00	878072.00	61022.00	3487.00	18124.00	6154.00	15
194	484	484032	2000	694912.00	222381.00	748875.00	43776.00	3941.00	7429.00	4773.00	12
195	484	484091	2010	620736.00	281965.00	613432.00	85253.00	11823.00	14828.00	8913.00	18
196	484	484002	2010	1558838.00	447948.00	1653209.00	203986.00	124532.00	29966.00	18147.00	27
197	484	484003	2010	317867.00	114558.00	308481.00	58467.00	13849.00	7933.00	4883.00	5

Compare these rows to the ones we had at the beginning in Figure 16.a.

In case you rather to work on syntax, here is what you should be typing.

---

\*SYNTAX FOR aggregate by year, country, 1st subnational level.

```
DATASET DECLARE aggregate_wkshopIPUMS.  
SORT CASES BY COUNTRY GEOLEV1 YEAR.  
AGGREGATE  
  /OUTFILE='aggregate_wkshopIPUMS'  
  /PRESORTED  
  /BREAK=COUNTRY GEOLEV1 YEAR  
  /fem_sum=SUM(fem)  
  /age2029_sum=SUM(age2029)  
  /nonunion_sum=SUM(nonunion)  
  /university_sum=SUM(university)  
  /foreign_sum=SUM(foreign)  
  /female_2029_univ_sum=SUM(female_2029_univ)  
  /nonunion_female_2029_univ_sum=SUM(nonunion_female_2029_univ)  
  /female_2029_sum=SUM(female_2029)  
  /nonunion_female_2029_sum=SUM(nonunion_female_2029)  
  /native_2029_sum_1=SUM(native_2029)  
/foreignborn2029_sum=SUM(foreignborn2029).
```

---

## 9. Create the rates we will be using for visualization later

\*share of women 20-29 that are not in a union

---

```
DATASET ACTIVATE aggregate_wkshopIPUMS.  
COMPUTE nonunionrate=nonunion_female_2029_sum / female_2029_sum.  
EXECUTE.
```

\*share of highly educated women 20-29 that are not in a union

---

```
COMPUTE nonunionrate_univ=nonunion_female_2029_univ_sum /  
female_2029_univ_sum.  
EXECUTE.
```

\*share of immigration in total population

---

```
COMPUTE nonunionrate=foreignborn2029_sum / (native_2029_sum_1 +  
foreignborn2029_sum).  
EXECUTE.
```

---

\*For some playground on SPSS using aggregated data.

\*some playground on tabulating visualizing aggregated data in STATA

---

```
SORT CASES BY COUNTRY.  
SPLIT FILE LAYERED BY COUNTRY.  
MEANS TABLES=nonunionrate_univ BY YEAR  
  /CELLS=MEAN COUNT STDDEV.
```

```
SPLIT FILE OFF.
```

---

Just a quick comparison of means by year and country for the indicator non-union rate created for female 20-29.

## Report

nonunionrate

Country	Year	Mean	N	Std. Deviation
Costa Rica	2011	46.4812	7	9.29016
	Total	46.4812	7	9.29016
Dominican Republic	2010	22.3959	25	5.02065
	Total	22.3959	25	5.02065
Panama	2010	29.6261	7	7.52232
	Total	29.6261	7	7.52232
Puerto Rico	2010	77.5835	6	5.11122
	Total	77.5835	6	5.11122

### 10. You can visualize sub-national level data in many GIS software. Today we will try Tableau

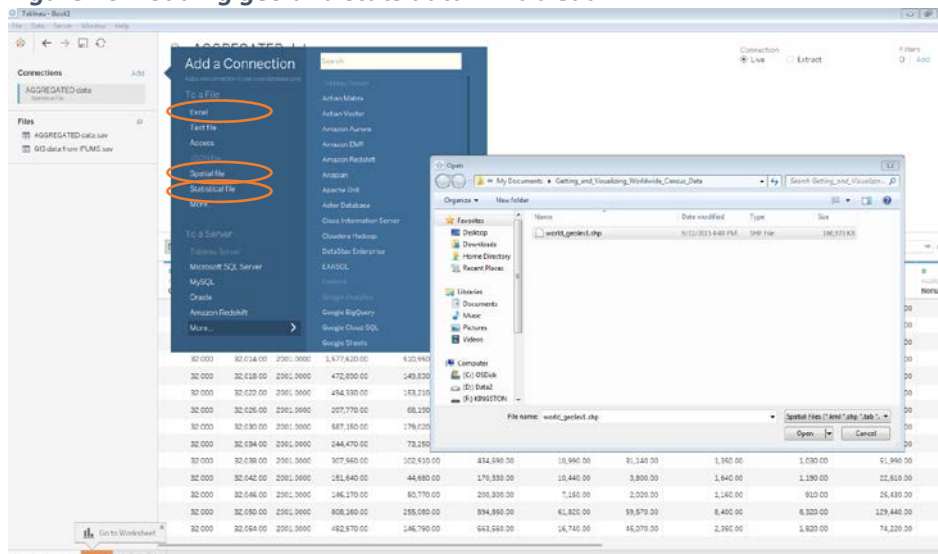
First, save the data in SPSS to keep the work we have done. Second, go to File/Export and save the data into EXCEL file. In this way, we will have an Excel version of the data aggregated in SPSS. That data will work as our Excel (statistical information) for Tableau.

Now, open Tableau.

Then, in ADD A CONNECTION click on SPATIAL FILE (Figure 21). We will be bringing the shape file containing GEOLEVEL1 variable (IPUMS identificatory for the 1st sub-national level). Remember that in Figure 1 we got the "world1.geolevel1.shp", when we downloaded from IPUMS GIS the "world\_geolevel1" zipped folder. Now is the time to use that data: the .shp file contained in that folder.

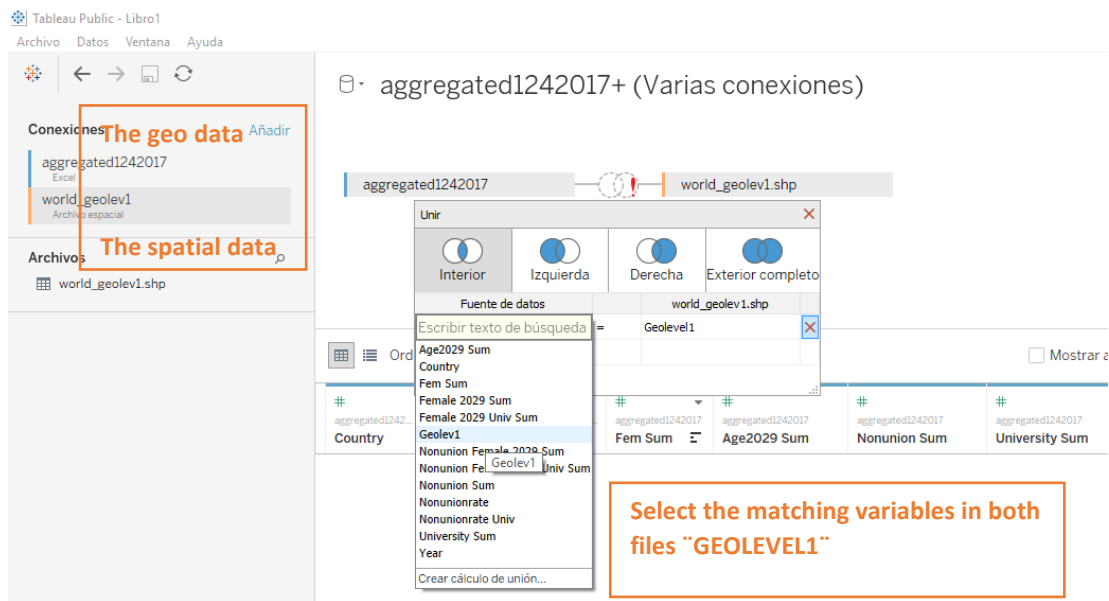
Then you will see at the right top of the recently added file the label "Add", click on it and insert the Excel we exported from SPSS.

**Figure 20. Loading geo and stats data in Tableau**

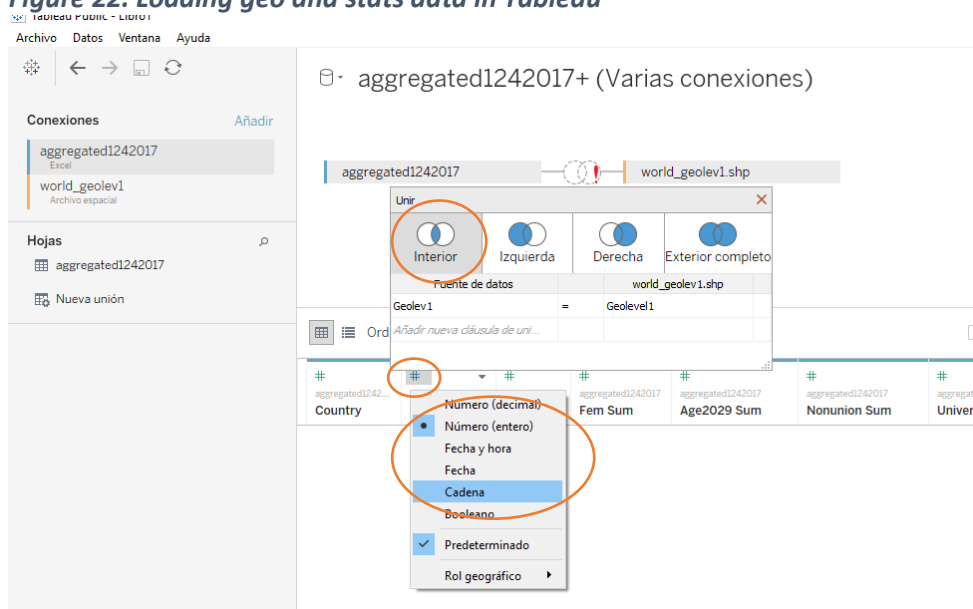


Then you will have in blue the display of geographical data and in orange a display of the statistical variables. However, there is one more step before getting both files matched. You have to select the geolevel1 variable in both files (that is the variable that will enable the match since it has the same ID for the 1<sup>st</sup> sub-national units). To get the joint done you need to see that the red exclamation mark on the top of the options “interior/left/right/full” is gone. To remove this, it will be necessary that you check that both geolevel variables, one in EXCEL FILE IMPORTED, other in GEOGRAPHY IMPORTED are in string format.

**Figure 21. Selecting the matching variable to do the joint**

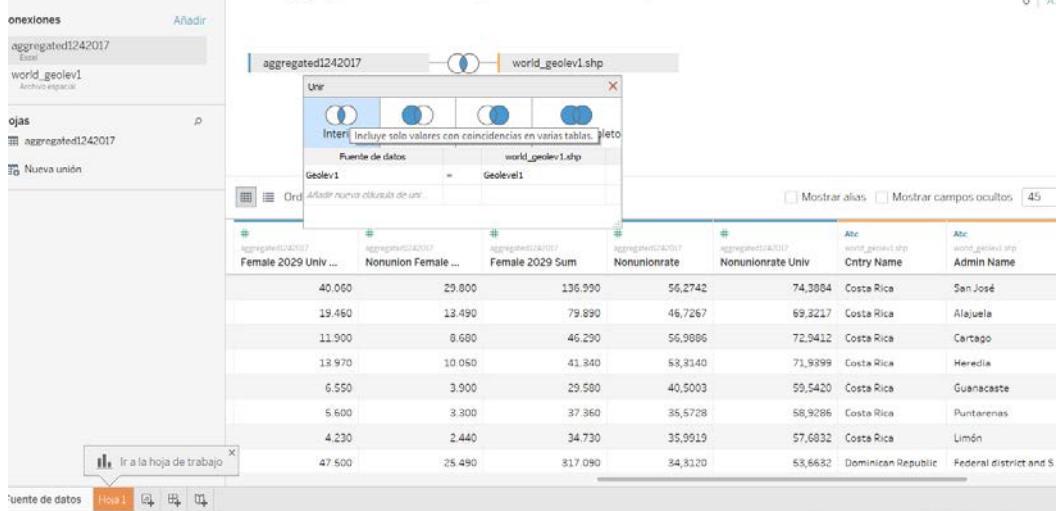


**Figure 22. Loading geo and stats data in Tableau**



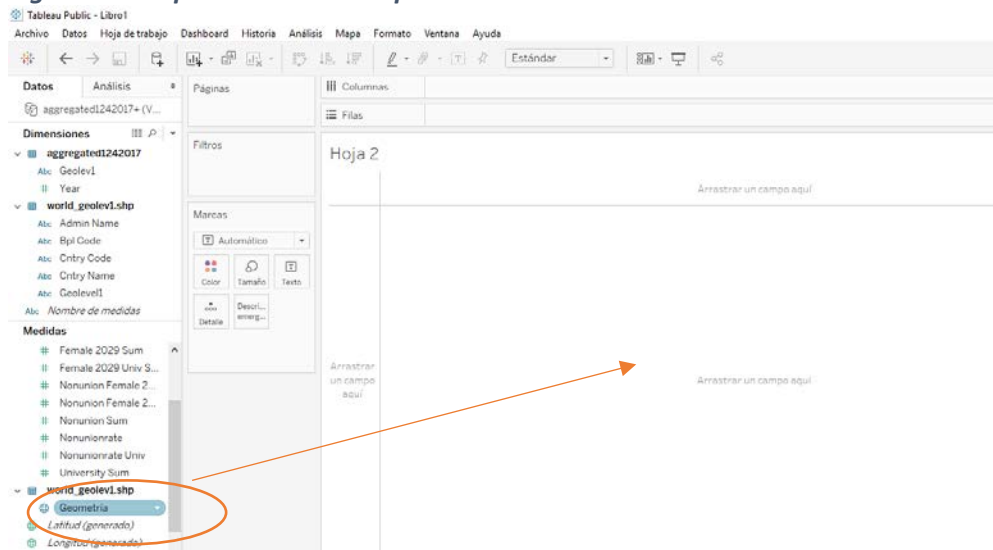
Click on the numeral sign # on left top of the label for “geolevel1” at the section for data “aggregated1242017”, and change numeric into a string. Immediately after, you will get the match done. Select interior or inner match.

**Figure 23. Final appearance of the data**

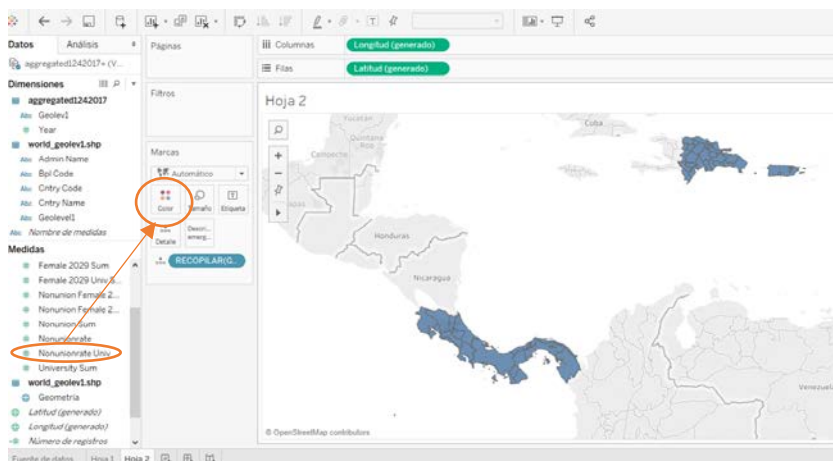


In case you want to see a map, select a new worksheet, and grab the geography towards the central area of the worksheet

**Figure 24. Step 1 to create a map**

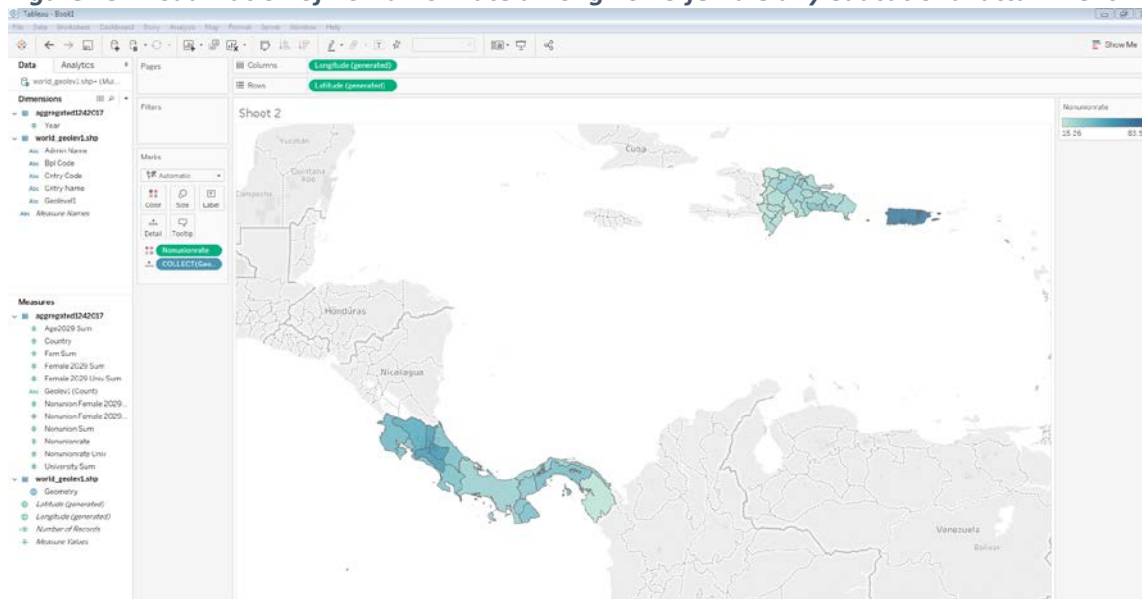


This is what you will get, but to build the map representing your data you will need to add the selected indicator into the area where a set of colorful circles are shown.



Finally, you just need to select the function of the selected variable and switch it from SUM to DIMENSION, and a hatch map with gradient colors according to the rate intensity will appear. The legend will be at the top right side of the map.

**Figure 25. Visualization of non-union rate among 20-29 female any educational attainment**



**Figure 26. Visualization of non-union rate among 20-29 female highly educated**

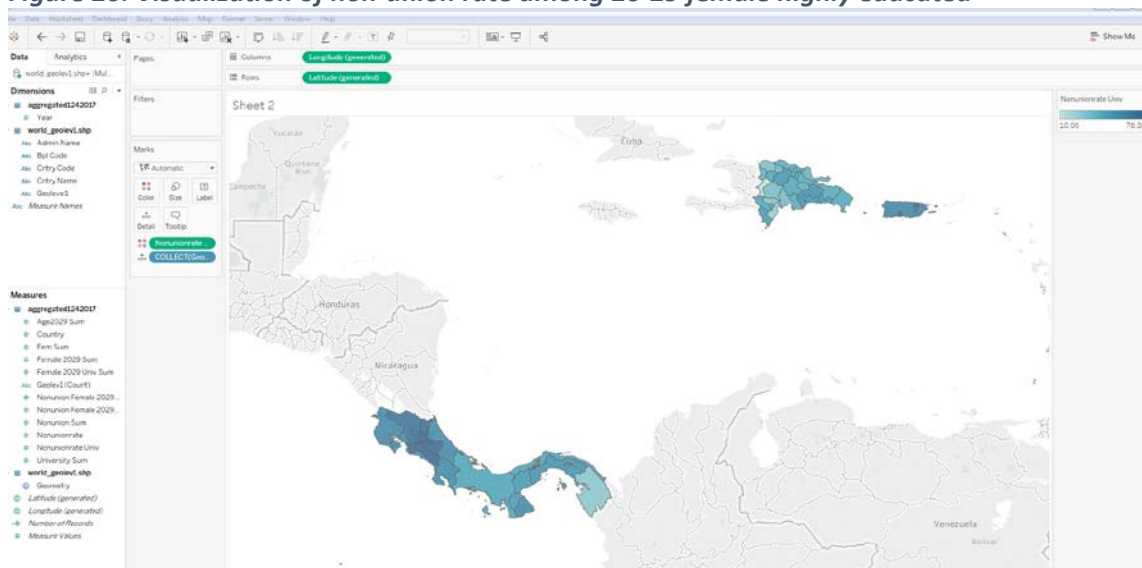


Figure 27. Tableau also enables visualization by bars. In this case the median by country of the non-union rate among 20-29 female. This median is estimated from the 1<sup>st</sup> country sub-national level values

