

NWX-US DEPT OF COMMERCE (US)

Using Public Microdata to Create Custom Tables on data.census.gov

April 2, 2020

(Tyson Weister): Thank you. And thank you all for joining the webinar this afternoon. Today we're going to talk about microdata access on data.census.gov. My name's Tyson Weister, I work on the communication team for data.census.gov. And with today what we wanted to do is give you all a little bit of background information about microdata and what that actually means and what you can get out of this particular tool that you can use as part of the data.census.gov site.

And in order to accomplish that most of what we're going to do today is a live demonstration where I'm going to work through data.census.gov on site and you'll be able to see a couple of examples.

The benefit of using the microdata on data.census.gov is you can create custom tables. So when we talk about customizing a table on data.census.gov in the premade tables what we mean in customizing your table view. But when we talk about creating a custom table using the public use microdata, that's where you really create your own categories of information as to what the table is displaying rather than searching a set of pre-made tables for you.

So to go ahead and get started -- what is public use microdata? Well, when we say public use what we mean first is that of course all of the responses at the Census Bureau, we take extra measures to make sure that privacy and the identity of our respondents is confidential.

So in terms of getting you the public use microdata, we have anonymized the data. We removed personally identifiable information and then we put additional edits in place specifically to protect the confidentiality of respondents. Things like top coding and bottom coding of certain variables as one example

Public use also means it's accessible in a variety of formats, with free information, and a suite of data user guidance to help you through this.

Today we're going to focus on accessing it through data.census.gov/mdata. You can also access this information through the application programming interface or API, and you can continue to download these microdata datasets through our FTP or file transfer protocol site. And then use your own statistical software in order to create your tables.

The nice thing about data.census.gov/mdata is you don't have to use your own statistical software in order to create tables with the microdata and it's pretty user friendly and easy to access the different screens in comparison to certain statistical software.

And then when we say microdata, what that means is that it's a set of individual responses. So in order to make a table out of it, you have to tabulate the data yourself and it needs to be weighted by the user. When you're working through data.census.gov/mdata it's going to automatically pull in the appropriate weight for you.

So just to see this in another way. What do we mean when we say summary data or premade tables vs. microdata? Here's just a quick screenshot that we grabbed from data.census.gov on the premade tables side where we're looking at a table for the state of Louisiana. So we can see this particular table shows that in 2016 in Louisiana, there were approximately 18,459 males who worked in a computer and mathematical occupation.

When we talk about microdata this is where we actually look at a set of edited survey responses. So here in this line we're showing one of the many lines from the records of the microdata and it shows through the different codes, State Code 22, Sex Code 1, and Occupation Code 1030. That all means that this particular response shows a male in Louisiana that's a web developer.

So when you're accessing the premade tables we've taken this set of underlying data, our statisticians have put it in commonly used data tables, they've tabulated and weighted everything for you, where you can grab that summary statistic. On using the microdata, you're being doing that yourself, or using an easy software like data.census.gov/mdat to help you with that.

And why would you go through the trouble to use the public use microdata through microdata access instead of the premade tables that are available to you? The advantages of using the microdata is it lets you get more detail by topic. It also allows you to create custom tables.

So just through ACS alone, the American Community Survey, we push out 11 billion statistics every year, so you'd think we meet every data user need through that statistics, but there's certainly folks that are doing more complex or detailed analysis that need more detailed information than what's available in the premade tables.

And some of the limitations of using the microdata. Although it gives you a lot more detail in topic and allows you that customization, in order to give you that set of detail and still protect the confidentiality of respondents, the downside is you get less detailed information by geography we'll take a look at that here in just a second.

It can be more complex to create your own table rather than to search for the existing table. On the plus side data.census.gov/mdata makes that process a little bit easier for you.

And then the other downside of using the microdata is it can result in less accurate estimates compared to the premade tables.

A couple reasons for this -- there are additional edits that we put in place to the underlying microdata in order to protect the confidentiality of respondents. And in some surveys and programs - the microdata is a subset - of the full set of responses. So as an example from the American Community Survey, only a little less than 2/3 of the responses are used to do the microdata.

So with that information, if you find a premade table on data.census.gov that meets your needs, by all means use that before you go through the trouble to make your own table, but there are definitely advantages of making your own.

And then this just shows an example of a data table that you would want to use microdata for that has been used in other studies and research. Occupation, by sex, by marital status, that's an example of something that you could only get through microdata if you wanted to know that particular set.

Census programs in the microdata access tool we have a few different surveys and programs. The American Community Survey and sister Puerto Rico Community Survey dataset. For the ACS 1-Year and ACS 5-Year are in the platform data.census.gov/mdata. We have 1-Year data from 2004 to 2018. And then the 5-Year data from 2009 to 2018.

There are also a couple of additional datasets from the Current Population Survey or CPS. We have the CPS Annual Social and Economic March Supplement from 2014 to 2019. And then CPS Basic Monthly data from 1994 through 2020.

These different surveys and programs have different geographies. We mentioned one of the limitations of microdata is less detail in terms of the geography. For the American Community Survey you can get down to a Public Use Microdata Area at the lowest level of geography. We'll take a look at that in just moment. And for the CPS the lowest level of geography is county.

Public Use Microdata Area or PUMAs as they're called -- if you're not familiar with them already we'll take just a brief moment. They cover the entire country and they're based on population size. So they're areas with about 100,000 people or more. You can use PUMAs or collections of PUMAs to identify most cities that are 100,000 people or more and many metro areas but not all. We'll take an example of that here in just a second.

They have a five-digit unique code that is unique only when you're combining it with the state code and then they cover the entire nation. And they're redefined after each Decennial Census. Right now they're based on census tracts and counties as the building blocks.

So on the left-hand side just to visualize this, here I've taken a screenshot through data.census.gov of our mapping part of the platform there. On the left-hand side what we're showing is the three different public use microdata areas that make up the city of New Orleans. So it's these three PUMAs that are listed out for you. They also have a common name, such as Regional Planning Commission 6, New Orleans City Northeast.

So this is an example where census tracts are used as the building block and we have three PUMAs that cover an entire city and county equivalent.

On the right hand side here we have an example of a single PUMA that spans and covers completely three counties: St. James, St. John the Baptist, and St. Charles Parish, in Louisiana.

It is easy to select your PUMAs on data.census.gov. Here's a screenshot showing the panels to go through the process to select the three PUMAs. We're going to do this on the live site here in just a moment. And if you want to do more research on how PUMA may equate to your area of interest, whether it's a city, county equivalent, or metro area, one of the great resources that isn't the focus of today's webinar but I did want to put out there for you if you're not aware of it already, is our TIGERweb Mapping Application.

So if you visit the link here and then go to TIGERweb on the upper left-hand side you can play around and look at different layers on top of each other. So you can visualize all of the PUMA boundaries and then also turn on the boundaries for your county, place, or metro area and see the overlap if you want to do a little more research on how you can use these for getting data for smaller geographies.

So with that we'll go ahead and move over to the demo side of things here, walking through two different examples for you. The purpose of these examples is first and foremost to show you how to create a custom table. We want you to leave the webinar being able to select, edit, and categorize your variables. Restrict the population that your table represents or the table universe, customize your table layout, and then show you how to download the custom table as well as the underlying microdata.

So to start, we'll go ahead and walk through an example of how to access data on employment status broken out by sex for the Mexican population in the state of Louisiana and then for those three PUMAs that we showed earlier on screen that make up New Orleans.

And again, why would we want to access this through data.census.gov/mdat instead of searching for the premade table that's already available to us? And the reason is that the premade table doesn't have the data broken out for males as we may want to see. So this shows us the closest table that has information where we're looking at data from S0201, we have data for the state, and the Mexican population, and then we can see here it does provide employment status information, but only for the population of 16 years and over.

However, the good news is we have an easy to use tool where you can make this table for yourself.

So when you go to data.census.gov, this is the landing page that you get to and when you scroll down to the bottom, you'll see that there's a whole section on microdata, with a button that says 'Explore Microdata.' So either by clicking that button, or visiting data.census.gov/mdat, it takes you to the landing page for this tool to use the microdata.

On one of the slides we talked about the different datasets that are available. The ACS as well as the Puerto Rico Community Survey, 1 and 5-Year, CPS, ASEC Supplement and then the CPS Basic Monthly. To show you a tour of the site, I'm going to stick with the same dataset today, using the ACS 1-Year estimate, and then there's different vintages, which refer to the year or time period available for that particular dataset.

Here, we're going to stick with the most recent 2018 release and then just click next to get started.

As we navigate through this site, we're just going to work through a couple of different tabs here. Going through the process the first step is to select our variable, or variables. We want Employment Status by Sex for Mexicans. A couple of different ways you can go about searching, if you know - as you can see here there's a variable name, and then the label for that particular variable.

These are good places to start your search. So the first thing I'm going to do is type in Hispanic where it says label and then that gives us one variable result. Hispanic recode. And then it has in here the column the number of categories that are part of this variable and that - it shows the type and that it's been recoded or categorized.

The nice thing about the site is you have this detailed button here where you can click and see what do those 24 codes mean and what values are included. So this is a really good way to explore the variables that are available to you. And here once we click into details, we can see that Mexican is one of the results as well as a lot of other detailed Hispanic categories.

In order to select the variable, click the checkbox on the left. Next we want to go ahead and search for the additional variables, employment status and sex.

So we'll go ahead and delete where it says Hispanic in label, that takes us back to all of the 507 variables that are available to us and then we'll go ahead and search for the next variable: sex. Looking into details, it just shows us a value of 1 indicates male and 2 for female. Check the box to add it to our Data Cart here and then the third valuable we wanted was employment.

Again I can go ahead and start typing employment where it says label. You can also click the triangle where it says contains and filter and type something in there as well. So typing in employment there, we do see a couple of different variables that show up in our result. Here we have nine different ones and just reading through the different labels there's Family type and Employment status.

That sounds a little too detailed for what I'm looking for. Employment status of parent. Employment status of recode. And then as I scroll down, not seeing anything on quick glance. So I'm actually intrigued by this employment status recode and I want to know what those recode or categories mean.

So once again click into details and here we see the values of what we can get data for. So it will give us data broken out for people that aren't part of this particular category because they're less than 16 years old, the civilian employed folks who are at work, and many other categories you see here.

So this is clearly the variable we want. I'm going to check the box. And now we see in our data cart there's three indicated there showing our three variables.

Moving on to the next step, where it says select variables in the upper left. Click select geography. You can go ahead and ignore this little note here up at

the top. We do have more geographies available beyond State. I wanted to click the state of Louisiana as well as the three PUMAs for New Orleans.

So starting by clicking States, it's very easy to just work through the checkbox selection here, check the box next to the Louisiana, and you should notice here at the very bottom once you check a box, there is a selection added to the bottom of the screen.

The next step, select the Public Use Microdata Area, which first prompts you for the state, so we click Louisiana and here we get not the five-digit code for each of the PUMAs but the nice label that's associated with each of these PUMAs.

So as we scroll down through this list, we'll get to Regional Planning Commission and we see the label associated with the ones that make up the city of New Orleans -- Regional Planning Commission 6, 7, and 8.

Now that we've checked those boxes, the next step is to click where it says Data Cart. So this is a good place where we can see the different responses for the variables that I selected. If there's something that I selected that I didn't mean to select I can use the icon here for the trashcan to remove it. What I want to know is the employment status broken out by male and female and I wanted that for the Mexican population.

So I have the variable here, HISP, this gives me data for all of the different Hispanic groups as well as folks who are not Hispanic or Latino. But I just want Mexican. So out of these 24 responses I only want one. The easy way to select that is to check the box at the very top which will uncheck all of my selections here and then I can just simply check the one that I want to keep for Mexican.

So in terms of the variables I have Mexican, I have male-female, and I have the different breakouts of employment status. The next step is to move to our table layout.

So this is an optional step where you can look at the layout of the table before you actually start getting data for the table. The default on data.gov/mdat generally is going to cross tabulate all of the variables that you've added into your selection. If you want your table to look like something you would see on the premade tables for the American Community Survey, you can easily change your view.

So what we're looking at right now is kind of showing the geography here on the left, where we have Louisiana, and then we have the different PUMAs underneath.

Usually the premade tables for the ACS, the geography is in the column, so what we can do for selected geography is simply drag and drop for the different areas and see how our tables updated.

There's not really any right or wrong way to do this. It just changes your view between what you put in the columns vs. the row. So Geography here in the columns, I'll go ahead and move Sex to the row, and Employment Status to the row. And the order of the variables in the row does make a difference in what's displayed on screen.

And the last thing I'm going to do is - I don't really need Hispanic - Mexican repeated across the table. I just want the table population to reflect folks who are Mexican but I don't need a repeating label in my table that says Mexican.

So if you want to play with the layout that way -- there's nothing wrong with keeping it in now, it just gives my table a little bit more of a cluttered look. If I drag HISP to not in table, that's going to restrict the sample universe. So my table's only going to reflect Mexican because I only selected that one response out of all 24, but it's cleaned up my table, I don't have that repeating label here cluttering up my look.

So here we have what we wanted. Our geographies here at the top, there's a line for total. Then we get data for the state of Louisiana in each of the Public Use Microdata Areas as we go across. We get data for the male population, with the different employment status breakouts and then the female population with the different employment status breakouts.

So whenever you're happy, once you've moved the variables into different rows, columns or not on table, the next thing to do is just click on the lower right side of your screen where it says 'View Table.'

So the benefit of playing with the view in the previous layout option is it's not pulling the data each time you make a change, so you're able to play with the view quickly and then ask the system to go and get the data one time. It just helps the site run a little bit faster for you.

So here we have our total information. So with the data just what we looked at the last screen, we can see for the state of Louisiana, the Total population, again we're looking at Mexicans, is 96,366. And then we see the breakout of that within each of the three Public Use Microdata Areas. We see the Total Male Mexican population. And then the Employment status for each of those in the break outs here.

Going down we see this repeated for the female. One of the things that we added here in the latest release is a total column. So this shows you the total across the board which were previously not included in data.census.gov/mdat as of a couple of months ago.

Just scrolling back up to the table it does give you some information here showing that we're pulling from the ACS 1-Year Public Use Microdata Sample, the year is 2018 and we can swap easily back and forth between vintages and notice for Weighting it's already pulled the appropriate weight. So we asked for a number that reflects the total number of people and it pulled in the PUMS person weight.

If we wanted a number that reflected the total number of households, you'd want to make sure to use the housing weight.

And then if you want to switch to unweighted either just the totals from the raw microdata that haven't been tabulated since it's based on a sample, the underlying sample has to be tabulated to reflect the overall population, not just the sum of the individual responses that we took and received.

From here we have the information that we were looking for, just want to show you an additional thing that you can do on the site, under download and share. You get the option to work with the data off the site.

So the first option is to download a view of the table in CSV format. So here we have our table, very similar to how it was displayed on screen where we have our employment status broken out for the total and female population and then the different geographies. Beyond that you can also extract the raw data in .csv format. Here I'm just going to show you that one option. Once you click that and press download underneath, you'll get a zip file.

We do recommend using Google Chrome with data.census.gov. And then when you open up the zip file in that browser you'll have a button on the lower left side of your screen to click on.

I'll choose, 'No, don't unzip the files.' And here we have the different files to choose from. What you want is the .csv file. This is - we had showed you on an earlier slide what microdata looks like. This is all of the records that match what we requested.

So, Mexicans in the state of the Louisiana, including the PUMAs within the state that we requested as well. And we're just getting the variables that we asked for so the PUMS Person weight, the Variable sex that we selected, ESR, Employment Status Recode, and HISP, for Hispanic. As well as the State code and the PUMA code.

Beyond getting the data in .csv, we also have the information for you in a couple of other ways as well. One is copying a bookmark. So whether you choose it from this button or just copy from the address bar, then you can open a new tab and you can share this for yourself for later you can share it with a colleague once they click that.

It does take a second before it loads up. It'll show the landing page for just a moment. But once you scroll down, you'll be able to see the exact table view that you went through the steps to create earlier.

And then the last way you can work with the data offsite is by copying and working with the API or Application Programming Interface. One is the API query, so pasting that - just to show you really quickly here. The first link is to the raw data. So as we had downloaded the .csv file earlier, if you wanted to

access this through the API instead of a .csv file, know that that option is available to you, as well as this last option that is the API query that actually tabulates the data. So this is the API for the raw information, the one I last copied and am pasting now is the data table that we showed on screen on data.census.gov/mdat. This is the full tabulated query from the API.

Moving on, just want to show you a second example to work through as you're on the site and you're ready to go back to the landing page to start fresh.

In the upper left, where it says Microdata, click that. That's going to wipe everything out and start you fresh. Here we can select the Dataset and Vintage. Just for ease of use, I want to show you how to use this site. So, I'm just going to stick with the same dataset, and then just as before the 2018 ACS 1-Year Public Use Microdata Sample.

In this case, just wanting to walk through the second example. That was to pull the Hispanic by Age. So, really, we wanted to pull Hispanic 50-plus in Louisiana. Again, why would we go through the trouble to access this using the microdata? It's not in the pre-made table.

So, this shows you one of the pre-made tables for Sex by Age. You can see this table breaks out information for 45-to-54 years old and so on. We don't have the option in this table to get the age category of interest, but we can easily create it ourselves on data.census.gov/mdat.

So, we already know from the last time that we want data for Hispanic, and the variable name for that is HISP. If you click that, check the box, it's added it to our Data Cart here.

Then we'll search for our next variable. First delete what we put in variable before. And then we do see Age here up at the top, if you didn't notice that, and you started typing in Age, it's doing a text search for everything that's in the variable label.

Notice here that, for the first time, under Type, we're seeing Estimate instead of Recode. And when we click into Details, it's showing us a value of 1 — Age 1 to Age 99, and then a separate category for folks with the value of 00. That represents folks under the age of 1.

So, because this variable is not recoded, it's continuous; once we check the box, you're going to see a note here up at the top, the variable's continuous, and you have to create your own category group or recode in order to actually use it.

Moving on to the next step, under Select Geography, clicking on State, and here we can choose Louisiana. Then we're ready to go to our Data Cart and show you how to create that recode.

So, one of the things we have is AGEP. And what you need to do in order to actually use this in your table, is tell the system what categories you want to use. So, under Create Custom Group, what you want to do is make sure the boxes are checked for everything that you want to create a category for.

I want to create a category for Age 0 to 49, and then for Age 50 plus, so I want to make sure that I have both of these boxes checked, so I'm grabbing Age 0, and then I'm grabbing all of the available ages in the continuous category of 1 to 99 that I can create my own custom grouping for. So, we're good there. I click Create Custom Group.

So, for this first group we can go ahead and create our own label. I'm just going to put Age 0 to 49. Then I'm actually going to tell the system that I want the age group, first by clicking the box under 1 to 99 and then where it says 1 to 99, I want to change that to 49, and then I do want under the year of 1 to be included in the categories, so I'm going to check the box to pull in that as well.

So, it looks good here. We have 0 and 1 to 49. We saved the group. And then we have this under Age recode, our very first recode or category that we created. Everything else is going to be lumped into Not Elsewhere Classified. So, here we can click Edit Group. And we could continue this process to create additional categories.

I'm fine with the category as-is, but I just want to update the label. So, I'm going to put in Age 50 plus. We can check the box, and then show that it's telling the system to include in this category values of age between 50 and the top-coded category of 99.

Quick Save Group, and now we have our two categories that we are interested in, ready to place into our table. And notice on the left, it's created a separate variable for that, AGEP_ Recode 1.

Last thing I need to do is review the Hispanic variables. So, we wanted this for all of the Hispanic groups. This particular variable provides 24 different values — one that reflects Not Hispanic or Latino. So, I don't want to include them in my table. I want data for Hispanic or Latino, so by unchecking the Not Hispanic or Latino box, the remaining 23 out of the 24 responses all are for detailed Hispanic groups, including an All Other category.

So, we have the Hispanic population captured in these 23 groups. With that, we're ready to move to the Table Layout. Because we have a continuous variable, it's always going to show in the values in Table Cells option. You can just leave it there. But what you want to make sure is that the recoded variable shows on the table where you'd like it to be.

So, just looking here again, just so I have something familiar to look at, I like the selected Geography to appear in the column, just so it looks kind of similar to the pre-made tables I'm used to seeing on data.census.gov. Here we see Louisiana is now in the column.

But what I wanted data for is the total Hispanic population Age 50 and over. However, I'm seeing Age_Recode 1 not in-table. I want to drag that into the row, so I actually see the breakout of 0 to 49 and 50 plus on my screen.

And then from here, we have the option to take the table as-is. If I wanted this information across all of the 23 detailed groups, I could keep it as-is. But if I just want the information for the total Hispanic population, I can move HISP into Not on Table. Again, it's restricting the universe, so since I unchecked the box for Not Hispanic or Latino, it's removing them from the data that I'm going to see.

And then we can go ahead and click View Table in the lower right. Scrolling down, we start to see the table, but notice here we're seeing the first option, 22.99. All we need to do is where it says the dropdown Average of Age — that was the default that it put us in, since we had a continuous variable this time, we actually want to change it to Count.

So, here we get our data that the total Hispanic population in Louisiana, and then we see the age breakout for 0 to 49 versus 50 plus. So, there's

approximately 48,844 Hispanic population in Louisiana that are over the age of 50.

Moving on, just want to wrap up with a couple things and make sure that we have time for questions. We'll go ahead and start the Operator allowing folks to queue up their questions.

If you all do have questions that you're not able to ask during the Webinar, we do have our contact information here listed for the Public Information Office, if you're a member of the media, please route your question through PIO@Census.gov, and then for the public, you can ask your questions and then give us comments on how we can make the site work better by emailing us at cedsci.feedback@census.gov.

Operator, can we open it up for questions.

Coordinator: Yes. If you would like to ask a question, please press Star-1, unmute your line, record your name clearly. Your name is required to introduce your question. If you wish to withdraw your question, please press Star-2. Again, to ask a question, please press Star-1. Allow a few moments for questions to come through.

(Tyson Weister): Great. Thank you. And as we're waiting for a question to come through, one of the things that I wanted to make sure that we mentioned and really hit home is the development of this site does depend on your feedback. So, this is a site that we're continuing to actively work in, do continuous improvement, and push those updates live to you every two months based on your feedback.

So, please tell us how we can make it work better for you. Email us at cedsci.feedback@census.gov.

In addition to that, we have a suite of educational materials for data.census.gov across the board. That information is listed here on the left. And when you visit the link in the upper left for what is data.census.gov and scroll down on the page a little bit, there is a section devoted to microdata access — a couple of PDFs with step-by-step instructions that walk you through this site.

Coordinator: We have a question in-queue from Julie — I'm sorry — (Judy). The line is open.

(Judy): Hi, yes, can you hear me?

(Tyson Weister): Yes.

(Judy): My question is for CPS data, say, can you pull in several months, like you can in DataFerrett? And it will give you, like a rolling average?

(Tyson Weister): That's a great question. We currently do not have that functionality on MDAT — data.census.gov/mdat. We've received a lot of comments about it, so it will definitely go into the overall feedback and something that we can put on our wish list to give to you.

It's not on the site now, so the only thing you can do to get information historically is to create a separate table for each month, and there's not functionality that gives you the rolling average as of yet.

(Judy): Thank you.

(Tyson Weister): You're welcome.

Coordinator: As a reminder, if you would like to ask a question, please press Star-1, unmute your line and record your name clearly. Thank you.

(Tyson Weister): And as we're waiting for questions to come through, I did see a couple pop-ups in the Chat. One was, "What geography is used if I don't select a geography?"

Just like on data.census.gov, when you're doing microdata and you don't select a geography, the default is the United States.

And then one other question was how to cross-tabulate variables.

When you're on data.census.gov/mdat, everything's going to be cross-tabulated by default. So, everything that you put into your rows and your columns is cross-tabulating and giving you the data.

You can reposition variables between rows and columns to give you a different view, but it's giving you the same estimates, just in a different layout. If you don't want your variables cross-tabulated, what you want to do is create separate tables with only an individual variable or only the variables that you want cross-tabulated together.

Do we have any questions queued up, Operator?

Coordinator: I have no questions in-queue at this moment.

(Tyson Weister): Okay. One thing we did want to make sure we share is the Webinar evaluation. This really helps us to make sure that we put out Webinars and continue to improve them and make them useful to you. Tell us what you like;

what you didn't like. And we'll give a little bit more time for questions to come through.

Rex, were there other questions through the Chat that maybe were noteworthy? You got – you wanted to share with the group so we could share the response?

(Rex): Sure. So, there's a question from Andy. He asks, "Are there any types of data that are not obtainable from the Census tables from microdata?"

(Tyson Weister): So, right now on this site, in terms of the microdata, what we have available is just the American Community Survey, and then, two of the Current Population Survey datasets. So, you only have access to those three datasets.

It gives you a lot more detail in terms of what's available to the topics, but you don't get the actual response that was given. So, sometimes things have already been categorized.

But just to show you some additional information that may help answer the question, if you're coming at this from ACS, just as an example, I'll give you information for them, but I do encourage you to browse the different technical documentation for each survey and program to complement what you're able to access and see on data.census.gov/mdata.

On the left-hand side for ACS where it says, "Technical Documentation," there's a whole section devoted to Public Use Microdata Sample or PUMS. And then once we load this screen, there's a section for PUMS Technical Documentation. And here's where you can browse what's available.

So, we have showed you on this site how to search for a variable. Click on the button that says Details and explore the values for that particular variable. If you'd like to see this whole list for the entire dataset — the PUMS Data Dictionary, I'm just going to open it for the 2018 ACS 1-Year, because that's what we were using. It's 132 pages, so it has each variable name.

As an example, we looked at ESR or Employment Status Recode. I'm just doing a Control F on the document. We see the variable name listed here. But the value for that particular variable will only be one character long, so label is Employment Status Recode, and then these are the values associated with the label.

One good use for this document — I searched on this site for Hispanic. However, if I had searched for Mexican and wasn't really sure where to go from there, this document is useful in terms of searching for the individual values that you see will come up in your results here, and it's a good way to explore what detail, in terms of the topic, are available to you.

Operator, do we have questions queued up?

Coordinator: There are no questions in-queue at this moment.

As a reminder, if you would like to ask a question, please press Star-1, unmute your lines and record your name clearly.

(Tyson Weister): Okay, Rex. Is there another question or two that we'll take?

(Rex): Yes.

((Crosstalk))

(Tyson Weister): Sure.

(Rex): Yes, so this question is from (Christina). She asked, “Is there any international data available on census.gov?”

(Tyson Weister): We do have some international information, kind of outside of the expertise for this particular Webinar, but since we have some extra time, I will share the little bit that I know about it.

So, when you go to Census.gov — and they have changed our site a little bit over time. Under Surveys and Programs, I believe it’s under All. Actually, sorry, I’m going to click under Population Projection. I’m not able to quickly find it. We do have some international information. If you go to Surveys Programs and click All Surveys and Programs, you can explore the different datasets; we do 130 different ones at the Census Bureau.

Oh, here’s International Programs. But you can also write in your question to cedsci.feedback@census.gov and that would give us a little bit of time to work with this offline. I think what I had in mind was the International Database, and this gives you some information and population statistics demographics at the international level. That’s the one I was thinking of. But if you want to start there, and then, if you have follow-up questions, we can certainly route you to the appropriate folks.

Coordinator: We have a question in-queue. One moment.

(Tyson Weister): And Rex, was there one more question that we could share with the group?

(Rex): Sure. So, someone asked, “Are we able to create margin of error for PUMS data?”

(Tyson Weister): Oh, that’s a great question that we get a lot. And we’ve gotten a lot of feedback on. There currently is not a way on data.census.gov/mdata. So, the microdata access on our free site doesn’t have that capability.

However, you can calculate it on your own if you’re using your own statistical software. So, we hope that in the future we’ll be able to expand and add more capabilities to microdata access on our site, but in the meantime, if you wanted that information, the PUMS Technical Documentation walks through that process.

And then the PUMS estimates for user verification, since you do have to do your own weighting — I’m just going to open this up really quickly here.

If you’re doing your own weighting, this gives you basically some exercises that you could work through on your own. So, if you were – even if your end goal wasn’t to calculate for Alabama the total number of people between the age of 0 and 4, if you want to make sure that, you know, how to download the files, apply the correct weights, and work with things correctly, you could go through this exercise on your own, do that math and statistics, and then what you’ll get is the PUMS estimate.

You should get a match for this, that 2-8-5-4-5-5, and then you can calculate the standard error and the margin of error and then make sure to match it up against the results that you see here.

And once you’ve done that successfully in the ones that are available here — at least maybe try a couple of them — you’ll know that you’re applying the

weights correctly and you have the correct process, and then you can use that process to the estimate and corresponding margin of error that you actually want to create.

Coordinator: We do have...

((Crosstalk))

(Tyson Weister): Operator, do we have questions in queue?

Coordinator: Yes. We have two questions in queue. First question — not really sure how to pronounce the name — but your line is open.

Caller? If you...

(Peter): Hi. My name is Peter. I would like to know all the data depending on what type of database engine.

(Tyson Weister): I'm sorry, Peter, would you repeat the question?

(Peter): The data, you're showing everything located in on-premise and Cloud, or which database, so Oracle or SQL or Azure, (Seradata)? what type of database?

(Tyson Weister): Okay. That's a great question. So, everything that's feeding into the interface on data.census.gov including data.census.gov microdata access comes from our public API, or application programming interface.

So, that's where the data are in...

(Peter): It's a relational database and not SQL?

(Tyson Weister): You'd have to email the question in. I can't remember offhand if it's Oracle or My SQL.

(Peter): Okay.

(Tyson Weister): If you email us at cedsci.feedback, we can confirm that with you.

And the next question, Operator?

Coordinator: Yes, the next question's from Nancy. Your line is open.

(Nancy): Hi. Can you hear me?

(Tyson Weister): Yes.

(Nancy): Okay. Thank you. In Cal — I work at the California State Data Center. So, in California, we have set up our PUMA codes so that the first three digits are the county codes for the PUMA area.

How — or if — can we adjust a variable such as the PUMA so that we only read the first three digits, for example, rather than the whole five digits of the PUMA if we wanted to roll these estimates up instead of individual PUMAs, to roll those PUMAs up into like a county level? So, I want to look at just the first three variables of the PUMA variable. Just the first three digits.

(Tyson Weister): That's a good question. I am not sure that you are able to do that currently, at least through this interface. I'm just going to check really quickly.

So, PUMAs, when you download the files, there will be a variable for PUMA, and you'll have to work with it just like you work with all of the rest of the variables.

On data.census.gov, PUMA is only selectable through the Select Geography portion of the site, and as you can see, when I work through the panel here for PUMA and get to California, it is nice that it's put together by county. So, if you are trying to get a certain county, it's in alphabetical order; it looks like everything is county-based, as you were describing.

The only downside is that it could involve clicking several times. We don't have a way to mass-click or mass-select. But it is by county here, and it is alphabetized for you, so that's the best recommendation if you're trying to work on this site and get PUMAs for a single county or even multiple counties.

And then, if you are working with the data off this site and downloading the files, you would be able to manipulate on your own to do what you're describing.

(Nancy): If we do that, is this – the output – similar to the way that it is with the ACS on the tabulated where you start adding up variables, then your margin of error really starts to get wacky. Isn't that — and that's a technical term, wacky — isn't that kind of — is that going to happen with what I do here, creating estimates and then summing them up? Is that going to have the same problem?

(Tyson Weister): So, in terms of the estimates here do have margins of error. You just can't see them on data.census.gov in that we're calculate them on this particular microdata access tool.

You can do that separately if you download the files, work with it offline, and then, in terms of adding the margin of error when you're trying to add it for derived estimates, there is guidance on how to do that. I don't really have an answer for you. I'm not an expert on margin of error, but if you're not familiar, there is some information on the American Community Survey website under guidance for data users.

(Nancy): No, I'm familiar with it, but I'm just saying that — that's okay. I will put a comment in about it would be nice to be able to do that, so.

(Tyson Weister): Okay.

(Nancy): Okay. I'll add that as a suggestion. Thank you.

(Tyson Weister): Thank you.

Coordinator: There are no questions in-queue at this moment.

(Tyson Weister): All right. Well, great. Thank you all again for joining the Webinar. Just want to send another reminder out that we would love to continue to put on these educational materials and Webinars and know how to make them more useful to you, so if you could let us know what you liked and what you'd like to see in the future, we're really appreciate that.

Thank you all for joining this afternoon.

END