Introduction to the American Community Survey Public Use Microdata Sample (PUMS) Files Webinar Transcript

February 14, 2018

Tyson Weister

American Community Survey Office, U.S. Census Bureau

Slide 1 – Introduction to the American Community Survey Public Use Microdata Sample (PUMS) Files

Operator: Welcome and thank you for standing by, at this time all participants are in a listen-only mode until the question and answer session of today's conference. At that time, you may press Star 1 on your phone to ask a question. I would like to inform all parties that today's conference is being recorded. If you have any objections you may disconnect at this time. I would now like to turn the conference over to Mr. Tyson Weister. Thank you, you may begin.

Tyson Weister: Thank you and thank you all for joining us today and celebrating your love of data with the Census Bureau this Valentine's Day afternoon. My name is Tyson Weister. I work in the Census Bureau's American Community Survey Office. Today we're going to talk about an introduction to the American Community Survey Public Use Microdata Sample or PUMS files. Also in the room today is Sirius Fuller, he is a mathematical statistician and technical expert for the PUMS files. So he'll be here as well to help with questions and answer at the end of the webinar today.

I do want to let everyone know upfront that the slides are available on census.gov and I'll show you how to access them at the end so don't feel like you have to rush to write every single URL down or if you miss a detail you definitely have access to all of this information.

Slide 2 – Outline

Here's an overview of what we're going to talk about today. The PUMS files, just as a heads up, you probably know they're more complicated of the Census Bureau products from the American Community Survey. Typically they do involve downloading large datasets on a local computer and then analyzing that with statistical software such as R, SPSS, SAS, or Stata or becoming familiar with the Census Bureau's DataFerrett software.

So with that, there is a lot of work involved. Our goal today is to get you all excited about using the PUMS and the custom analysis that you can do with this data. We think that by giving an overview, talking about the different geographies you can use the data for, showing you how to access the PUMS data and walking through some common questions and speed bumps that people run into, as well as showing you our documentation, that you'll be comfortable getting started with these files. It is an intro to the PUMS webinar so you won't be an expert overnight but you should get a lot of resources out of this that will be helpful for you.

Slide 3 – American Community Survey is Foundational

If you're joining this webinar today, hopefully you are already familiar with the American Community Survey and our data products. We are the flagship survey at the U.S. Census Bureau and we're associated with the Decennial Census Program. But we conduct our survey on an ongoing basis collecting data every month for a total sample of approximately 3.5 million households every year. From that information we, one, have the PUMS files available and we also produce standardized data products as well.

So these products have critical information for small areas and small population groups that you can't get through other federal surveys and programs or private surveys. Hopefully you're already familiar with some of these products that are on American FactFinder. We are pushing out 11 billion statistics every single year, free of charge, for you to access producing over a thousand different tables and repeating those tables for many types of geographies.

If you're not familiar with the ACS we definitely recommend taking a look at census.gov/acs as well as an intro to the ACS webinar to compliment what you're going to see today talking about the PUMS files which is a specific product that comes from the American Community Survey.

Slide 4 – What are PUMS files?

So moving forward, what are the PUMS files? Well, here's a breakdown of the terminology and what we mean by Public Use Microdata Sample that you can see here on the screen.

Essentially the PUMS is when we take a little less than two-thirds of the responses we receive from the American Community Survey, we remove personalized information so you can't identify any particular person or household from the files. We also put a couple of edits in place, again, to protect confidentiality and then we give you that set of individual responses for you to conduct your own custom analysis with the data that's representative of the population. So if you're working in the 1-year PUMS, you'll get a file that has about 1% of all U.S. households. The 5-year file has about 5% of all U.S. households.

Being public use, this data is publically available on census.gov free of charge and it also comes with a suite of data-user guidance to help you work with these resources.

Slide 5 – Summary Data vs. Microdata. What's the Difference?

So let's take a moment here just to visualize this. Hopefully you've seen our data products on American FactFinder in the upper left is screenshot from one of those tables. To produce this table, our statisticians at the Census Bureau have taken a look at the records of all respondents living in Louisiana that indicated a male in their household works in a computer and mathematical occupation. They group them together and weighted them to produce an estimate that is representative for the State of Louisiana. So we can see here in 2016 in Louisiana there were approximately 18,459 males that work in a computer and mathematical occupation.

In contrast, the microdata that is used to create this and is also available through the PUMS files, this is an example of someone who responded to the ACS that's a male in the State of Louisiana who is a web developer by occupation. And this is indicated by State Code 22, Sex Code 1, and OCCP Code 1030 as shown here. So from this raw data you could create your own set of estimates like was done in the table in the upper left.

Slide 6 – Why Use PUMS?

You're probably wondering, why would I go through all of that work? Well, let's take a look at an example here of one of our Detailed Tables on American FactFinder, B24126: Detailed Occupation for the Full-Time Year-Round Civilian Employed Female Population 16 Years and Over. It's quite a mouthful and it describes the different questions that are used to create these estimates. So it looks at questions we ask on the survey about occupation,

employment status, weeks worked, hours worked, sex and age; six different questions. So you can get a sense of how detailed these 11 billion estimates in American FactFinder can be.

But every researcher has a unique area of interest and population that they want to look at. So essentially that's why we produced the PUMS files so that you can look at things that aren't included in our set of standardized products.

So as one example, maybe you're interested in looking at occupation broken out by sex and marital status. Maybe you want to know how many married female actuaries there are. Another reason you can use the PUMS is to create new measures with unique combinations of person or household variables. As an example, spouse's occupation. In addition, that microdata allows you to conduct sophisticated statistical analysis to understand relationships between variables. Maybe you want to do a correlation analysis between household income and value. You could use the microdata for that.

Slide 7 – Visualization Using ACS PUMS: Who Marries Whom

So let's take a quick look at the ACS PUMS in action. This is an interactive visualization fitting with today's Valentine's Day theme, Who Marries Whom? It was put together by Bloomberg two years ago. So if you visit the link at the bottom of this screen here you can play with this interactive visualization and see combinations, marital combinations by occupation. In the screenshot that I'm showing today, I hovered over mathematicians and statisticians and it gives you the top spouse combinations for that.

So the heterosexual, we get the top five heterosexual combinations we see mathematicians and statisticians tend to marry physicians and surgeons, software developers, elementary and middle school teachers, registered

nurses, and post-secondary teachers. We also get the top same-sex spouse for mathematicians and statisticians. So we can see that male mathematicians and statisticians tend to marry male public relations specialists, and female mathematicians and statisticians tend to marry female non-retail sales supervisors.

That thick line connecting male mathematicians and statisticians to female registered nurses indicate that's the most common matchup within that occupation. We've shown this as one of the many examples as what's out there and being used on researchers throughout the country looking at the ACS PUMS. This was put together totally by Bloomberg and the Census Bureau wasn't involved in this study. And we haven't evaluated the accuracy of the information but we just wanted to show this as one of the many examples.

Slide 8 – Summary Data vs. Microdata. Pros and Cons.

Hopefully everyone joining the call will have a particular area of interest that you may want to go to the PUMS for. Here's kind of a layout of the pros and cons of using the summary data versus the microdata. There are benefits and limitations for each of those, some of which we've discussed already. So just to recap and lay it out for you all on one page, that summary data, again, these are published tables by Census Bureau analysts with data user input. They've already weighted the estimates, displayed the margins of error so they're easy to use. In addition, they're available for small geographies down to neighborhoods, or as we would say at the Census Bureau, the census tracts and block groups.

On the other hand your microdata of course is a set of individual responses. They allow you to do custom analysis and there's lots of detail in the variables that are available. In the PUMS person file, you can look at each record which

has about 250 variables per record. The PUMS housing file has about 200 variables per record. In addition, we talked about creating new combinations of variables as well so you can really see how much detail is there.

Limitations, of course, with the summary data, what you see is what you get. You're searching for the information that's already there and the content are fixed for your categories, years, and topics. We do push 11 billion estimates out but we can't anticipate or possibly have the resources to meet every single researchers need for the data, sliced and diced how they want it.

On the other hand, some limitations of the microdata, of course being that you have to conduct your own statistical analysis, it's going to be more complex to use. It is based off of a subsample of the ACS responses so other things equal, your margins of error associated with those estimates will be larger using the PUMS versus summary data. The PUMS also has edits to protect privacy. One of those is top coding. So as an example, if someone in the State of Florida filled out the American Community Survey and their record made it into the PUMS, if they indicated they were 105 years old, we would show that in the PUMS record as 95 to mask that extreme value and help protect their confidentiality.

So the PUMS isn't going to be your source if you want to look at extremes like how many centenarians, millionaires, mansions, or people that have super long commutes to work. In addition, to help protect confidentiality, some of the categories are broader. And then the geographies, the level of geographic detail is less in the PUMS as well. We'll take a look at that more in just a moment.

Slide 9 – PUMS Availability

Another area of note between the summary data and the microdata is that the microdata is released one month after the standard products. As an example, the 2016 ACS 1-year PUMS was released October 2017 and the 2012 to 2016 ACS 5-year PUMS, we released that just last January. However, we do provide early release documentation a week in advance. If you're really eager to get started with the PUMS you can look to that documentation and prepare yourself.

Slide 10 – Multiyear (5-year) PUMS Files

While we're talking about the 5-year PUMS files, I do want to point out that the 5-year PUMS is essentially the 1-year PUMS put together with a couple of enhancements. So as example, the 2012 to 2016 ACS 5-year PUMS is made up of the 1-year PUMS files between 2012 and 2016.

However, we do put some improvements together. One of those are that the new weights are produced for these records so that the weighted population matches the latest population estimates. We also standardize dollar amounts within that file. We use an adjustment factor so they're all standardized to the latest year in the file specifically that way you're comparing apples to apples within a file for those values. And over time sometimes our coding does change from year-to-year. In the 5-year file when possible we crosswalk some of those coding changes so that it's standardized within the 5-year file versus if you put the individual 1-year files together.

Slide 11 - Outline

Now that we have an understanding of what you can do with the PUMS, let's talk about the geographies that you can do an analysis for.

Slide 12 – Limited Geographic Detail

Given that there's 250 person-level variables and 200 housing-level variables, approximately, there isn't quite as much detail as a PUMS user may hope in terms of geography that you can get. We do have to balance the confidentiality concerns with the level of detail that's provided, of course. However, we do provide region, division, state and PUMAs or Public Use Microdata Areas as a geography for you.

PUMAs can be used to identify many cities, counties, metropolitan areas with populations of 100,000 or more but certainly not all of them. We're going to take a look at how to compare those geographies here in just a moment.

Slide 13 – Public Use Microdata Area (PUMA)

But just to give you a little bit more background on PUMA's before we dig into that, they are statistical areas with 100,000 people or more. They are identified with a five-digit unique code within each state and they nest within states or state equivalent entities. They're geographically contiguous and cover the entire nation.

So, on the right side of the screen we can see two maps that are drawn at the same scale with the blue boundaries indicating the PUMAs. Again, these are population-based so as an example in Wyoming which has a population of a little under 600,000, we can see that the PUMAs there cover much larger

geographies but similar population compared to each PUMA in the State of New Jersey which has closer to nine million people.

We redefined these PUMAs after every decennial census and the 2010 Census PUMAs were first used in the 2012 ACS PUMS. They are based on census tracts as a requirement and counties as a recommendation.

So there's many resources out there to compare PUMAs and try to see whether they match up to other geographies that have more meaning to you. One of those resources that we recommend is the Missouri Census Data Center's MABLE/Geocorr tool. You can use this to look at a PUMA and see its allocation factor for another geography such as the county or place. And see the allocation in terms of population size or land area to see whether they do or don't match up relatively well. There's also a link at the bottom to the Census Bureau's reference page for PUMAs that has lots of details as to the history of PUMAs, how the 2010 Census PUMAs were defined, static maps to look at those, as well as PUMA equivalency files. We definitely recommend taking advantage of those resources.

Slide 14 – TIGERweb: Visualize PUMA Boundaries for Your Area

And we do want to take a moment to show you another resource that you can look at to compare how PUMA's match up to your geography that you may be interested in. That tool is TIGERweb and you can access it by visiting tigerweb.geo.census.gov. Again, tigerweb.geo.census.gov. This takes you to the main page and to get started we're going to click on TIGERweb applications, that second option here on the top-hand ribbon. Then in the upper left we're going to click on TIGERweb. That will allow us to compare the 2010 PUMAs to the latest ACS boundaries.

Slide 15 – Select Labels of Interest

All right, so to make this a little more concrete, let's take a look at PUMAs in the city of New Orleans to see how they may or may not match up with that boundary that has more meaning to us. Using TIGERweb we can lay out those different geographies and I'm going to just start zooming in to southeast Louisiana to get a look at that.

So as we're zooming in, we see Orleans Parish here indicated here in the blue and gray outline. And once we get zoomed in pretty far we'll start here on the left. It is an interactive map so your level of zoom does affect the options that are available to you on the left.

We're going to turn on some labels here so we can make sense of what we're looking at. So first I'm going to click that plus sign next to the labels. And then directly underneath of that there's a plus sign for PUMAs, UGAs and ZCTAs. I'm going to drill into that and check the box to the left. And now we see that we've populated a bunch of labels in red and blue. Those correspond to the 2010 Census PUMA's as well as the 2010 zip code tabulation areas. We're not interested in the zip code tabulation area so I'm going to uncheck that to de-clutter my map.

Next let's turn on the label for place. So I'm just under - a couple of options under - it's places and county subdivisions. Click the plus sign. Check the box on the left and then uncheck the ones that you don't want. We're going to leave incorporated places marked and declutter the map by unchecking the others.

So now we see we get a label here in our map for New Orleans. So we don't really know what these labels match up to. We need the boundaries that go

along with the labels. So basically we're going to make the same set of selections.

Slide 16 – Select Boundaries & View Map. Example 1: Three PUMAs Cover City of New Orleans/Orleans Parish

First, we're going to collapse the labels category and then make those selections under the main heading instead of under labels to get the boundaries that correspond with the labels we just put on the map. So once again, I'm going to click where it says PUMAs, next to that plus sign, check the box on the left and uncheck the box to get rid of the boundaries for zip code tabulation areas. I'm going to do the same thing for places and county subdivisions, click the plus sign, check the box to the left and then declutter the map leaving only incorporated places checked.

So now we get the layout of how everything matches up. We see that New Orleans is indicated in this red area here of course ignoring this little area of Gretna and Westwego that bleeds into that red connecting to New Orleans. The division here is along this blue boundary. We can see that there's three PUMAs that match up very neatly with Orleans Parish and the city of New Orleans, those being PUMA 02400, PUMA 02401 and PUMA 02402 in the State of Louisiana.

So if we wanted to do a PUMS analysis for the city of New Orleans we could use data from those three PUMAs to create estimates that represent New Orleans.

Slide 17 – Example 2: PUMA That Spans Counties/County Equivalents

Let's take another quick look here at how some PUMAs look nearby with smaller county or county equivalents. Of course in Louisiana they have Parishes. So we can see that these county or county equivalence with low populations are often grouped together to create a single PUMA. So we see that here in the middle with PUMA 01900 that goes around these three parishes: St. James, St. John the Baptist, and St. Charles. Grouping these together creates a statistical area that meets the 100,000 minimum population threshold for a PUMA. And it's known as the South Central Planning and Development Commission 1, River Parishes, Louisiana PUMA.

So we encourage you to visit TIGERweb, do some similar types of overlays using this tool for your area of interest to take a look at how you can use PUMAs to your advantage. They were created purposefully to match up with other areas with meaning to you. So please don't be turned off by the term PUMA, they really are meant to produce statistics of meaning for you.

Slide 18 – Outline

Going back to the PowerPoint here, let's look at how you can access the PUMS data now that you know what you can do with PUMS and what geographies you can analyze with it.

Slide 19 – PUMS Data on ACS Website

Lots of different ways to get to the PUMS. We recommend going to census.gov/acs as a user friendly interface option. On the left-hand side click where it says Data, then select PUMS Data, and you'll get a list of all of the

different PUMS datasets available to you. In this example, I clicked on the 2012 to 2016 ACS 5-year PUMS.

Slide 20 – American FactFinder

This takes me to another interface, American FactFinder, with, in the upper left of the screen, my search items are already prefilled in for me. And I just have to choose if I want the CSV format or the SAS format.

In this example I clicked in blue where it says 2012 to 2016 ACS 5-year PUMS CSV format.

Slide 21 – American FactFinder (cont'd)

And it takes me to another user friendly interface to download the files for population or housing records by nation and each state.

I clicked here on the Louisiana population records and this opened up the zip file you see in the upper right of the screen, the csv_pla.zip. The very first document in all of these zip files are the ReadMe documents because we do want everybody to read the ReadMe before you dig into the PUMS. It has lots of useful information for you. The next step would be to click on the data, that ss16pla.csv and that opens up your microdata file in the bottom right.

Slide 22 – Census Bureau FTP Site

All of these links that I was showing you pull from data that is on the file transfer protocol or FTP site. You can also access those files directly from the FTP site by visiting the link at the bottom. On the left-hand side click the year of the PUMS that you're interested in. And the right-hand side will list out

those files alphabetically for you with the last three characters will be either H or P for housing or person files. And then the last two of course is the state abbreviation.

Slide 23 - DataFerrett

A final way that you can access data for the PUMS is through DataFerrett. You can download the files through DataFerrett but more importantly you can use DataFerrett as a system for your statistical software instead of having to know how to use or have a copy of SAS, Stata, or SPSS on your own. DataFerrett is a tool that searches and retrieves the PUMS data. It can recode variables and create complex custom tabulations. It also pulls in the weights for you automatically so you know that you're creating estimates that are representative of the population overall.

Slide 24 – DataFerrett Assistance

As with any new program that you may want to use, there is of course a time investment and learning curve into getting to know how to use that.

To help with that, we've put together some video tutorials for you at the link on the screen here. In addition, pages 12 through 23 of our ACS Data Users Handbook: What Public Use Microdata Sample Users Need to Know, that has instructions and screenshots to walk you through using DataFerrett.

With that said, we do want to be upfront that some web browsers are starting to discontinue plug-ins that are required for DataFerrett to run. There are technical requirements on the main page, dataferrett.census.gov and some workarounds posted there. So please follow the instructions that you see. In addition if you run into any technical issues, there's contact information on

that page to our DataFerrett technical support. So feel free to give us a call if for any strange reason you are experiencing issues.

In addition, the Census Bureau is in the process of creating a new microdata analysis system that will eventually replace DataFerrett. We hope to have a prototype in place this summer for you all to take a look at the tool, review it, provide comments on the interface and functionality.

Slide 25 – Outline

All right, so now let's say you've downloaded the files and you're ready to dig into the PUMS, what do you do next?

Slide 26 – How Do I Put PUMS Files Together?

In some situations you may need to merge files together. A couple of examples here, the national level files are simply so large that we release them in four different pieces here. So on the screen in the upper right, you see for the U.S. there is an A, B, C and D file, each containing about one-fourth of the population records for the United States. Because they're so large, you download those separately but when you're ready to analyze them, you would piece those files back together for your statistical analysis.

In addition, housing and person level variables are separated into separate files. However, you can merge those together using the serial number variable which references one household and the people who are reported living in that household, the characteristics of the people in there.

There's some example code in the bottom right of the screen for serial number. It's just one of the many examples that you can use to merge these

files together. We do want to just point out a couple of things that you want to keep in mind when merging the files. One is that the housing data does have vacant housing units and of course those vacant units aren't going to have any person data associated with them when you go to merge the files. On the flip side, a select amount of person data has been collected from people living in group quarters such as student housing facilities, prisons, nursing homes. That set of questions is a little bit different from the standard ACS questionnaire and the only data available in the housing unit file for group quarters persons is food stamps. If you want to use that data, you do need to merge those files together.

In addition, when merging these files, keep in mind that most households have multiple people in them. So when you're creating housing unit estimates from a merged file just make sure that you're not double counting any households or multiple counting any households when creating those estimates and doing your weighting. We do encourage you to read the ReadMe document for more detailed information about merging the files.

Slide 27 – Which Weight Should I Apply?

Once you have those files together and are ready to create your estimates, you'll want to weight them. Again, a weight determines how many people or households are represented in one sample case. Most of you all know that you need to weight the data and statistical packages have commands to do this. You just need to know which ones to pull from.

So here's a little overview here. If you want detailed information, of course take a look at the Accuracy of the PUMS document, using the link at the bottom. But as a general overview, if you're creating an estimate that represents a certain number of housing units, you want to apply weight

WGTP, the PUMS household weights. If your estimate represents a certain number of people you want to weight that estimate using PWGTP, the PUMS person weights. And then there's also two sets of replicate weights ending in numbers 1 through 80. If you want to calculate the error associated with your estimates you would use the set of replicate estimate weights. So as an example for a household you would use replicate weights WGTP1 through WGTP80 and we'll talk more about error here in just a moment.

Slide 28 – Why Don’t My PUMS Estimates Match AFF?

Once you've calculated your estimate, you probably want to know if you've done it correctly. A common question we get is “Why don’t my PUMS estimates match American FactFinder?”

Well the purpose of using the PUMS is to create things that aren't available in FactFinder, not to recreate what's already there. And those estimates, if you are recreating one that's in FactFinder, it will differ. One of the reasons is that the PUMS, as we talked about, is based off of less than two thirds of the cases that were used to produce the estimates on American FactFinder. In addition, we include additional edits. We talked about broader categories, top and bottom coding.

So if you're trying to see if you're calculating estimates correctly, we've put together a document to help assist you with that. It's called the PUMS Estimates for User Verification. So what you would do is pull up that document. Here we're showing for the State of Louisiana we could calculate a PUMS estimate from the 2016 ACS 1-year PUMS file: How many people age 0 through 4 live in Louisiana? After we apply our weights we should get a result of 306,215.

Again, this probably isn't the reason why I would want to go to the PUMS to calculate specifically that variable. But the idea is that using this example you could check some person variables and some household variables and if you're matching up to the estimates for verification, you should feel very confident that you're applying your weights correctly for your other PUMS estimates as well that aren't on this spreadsheet for you to match up.

Slide 29 – How Accurate Are My PUMS Estimates? Margins of Error (MOEs) Explained.

If you know you're calculating your PUMS estimates correctly, you also probably want to know how reliable that estimate is. First we're going to take a look at our standard data products, how this is represented which is by a margin of error. So a margin of error is the measure of the possible variation of the estimate around the population value. At a given confidence level, it's expected that the estimate and actual population value will differ by no more than the margin of error. And the Census Bureau standard is the 90% confidence level.

So we've worked out an example here in the bottom right. We see for the State of Wyoming, the number of males under five years old is about 19,649 plus or minus a margin of error of 310. We can use that to calculate the lower bound and upper bound which will tell us, essentially, that the data tells us we are 90% confident the true number of males under age 5 in Wyoming falls between 19,339 and 19,959.

Slide 30 – How Accurate Are My PUMS Estimates? (continued)

We recommend that you calculate the error associated with your PUMS estimates as well. And we have two different ways for you to do that outlined

in the Accuracy document. One is to approximate it using the design factor which uses a generalized variance method. If you're a more advanced user, we talked about those replicate weights earlier on a previous slide. That uses the successive difference method. This method will give you a more accurate margin of error but involves more work. These are both referenced in the Accuracy of the PUMS. In addition, the replicate weights has more documentation in the Variance Replicate Estimate Tables documentation. This is written for the standard data products but the concepts and the formulas will transfer over to calculating standard errors for your PUMS estimates.

Once you calculate the standard error and get a sense of how reliable that estimate is, you may want to compare one PUMS estimate to another. And the Census Bureau recommends that you take into account that error to see if the differences between the estimates are statistically different or if they're just likely to have occurred by chance, given that the ACS is a survey and we collect our data by sampling just certain amount of the population rather than the full population like the decennial census.

You can use the formula on the left or you can easily transition your standard error into a margin of error. Once you have that you can plug and chug your PUMS estimates and margins of error into the Statistical Testing Tool and it will tell you whether those results are statistically different or not.

Slide 31 – Use Caution…

Some final words of caution, again, please do calculate your margins of error. There are about three million person level records in the PUMS. It sounds like a lot but once you start slicing and dicing to small populations with smaller level geographies such as PUMAs, it's really easy to start trying to create estimates with not a good number of sample cases. In that instance if you're

getting high standard errors or margins of errors, you may want to consider using multiyear files or a larger geographic area or combining PUMAs together.

We talked about extreme values being masked to avoid disclosure and protect confidentiality of respondents. The variables listed here are particularly affected by top and bottom coding so they're generally fine to use unless you're trying to calculate an extreme value.

And if you're using past PUMS data, past multiyear PUMS, they may have dual-vintage PUMA codes. We're not going to get into detail right now about that but do read the corresponding ReadMe file. Records in the 2012 to 2016 ACS 5-year PUMS are all based on the same set of PUMAs from the 2010 Census. You don’t have to worry about that for a while if you're only interested in working with current data.

Slide 32 – Outline

All right, so we talked about a lot of documentation as we've went along today.

Slide 33 – ACS Main Page

You can access those files by visiting census.gov/acs. On the left-hand side click on Technical Documentation.

Slide 34 – PUMS Technical Documentation

Then click on PUMS Documentation and PUMS Technical Documentation. You'll see the ReadMe file as well as a list of the broad subjects included in the PUMS. The Data Dictionary will be there as well. This includes detailed

variables and information about how each variable is coded. To go along with that for certain variables that have a large number of coded responses, there's another code list that explains in a little more detail what those codes are for things like ancestry and occupation. The top and bottom coded values are also available for each state, as well as the Accuracy of PUMS and the PUMS Estimates for User Verification documents.

Slide 35 – PUMS Data Dictionary

All right, we've talked about a lot of those files so far but haven't got a chance to talk about the Data Dictionary yet. It is important to look at this for every variable and we'll give you just one example. Recently we got a question about the PUMS and why there was someone who was not showing in the labor force in their PUMS record but they had an occupation coded. Great question, why would there be an occupation for someone not in the labor force. If you go to the PUMS data dictionary, the values indicated in the letter b correlates to a blank value in the PUMS record. So this tells you that for occupation, the records should be blank if the person is less than 16 years old or if they're not in the labor force and have never worked or they last worked five years ago or more.

So for people who are recently unemployed we would code their most recent occupation as detailed in the Data Dictionary. Even things that may seem obvious like age, you might want to consult the data dictionary to find out what an age of zero means.

Slide 36 – Source Us!

So when you're using our data, we do love if you can source us. It's very easy to do. Just specify the American Community Survey years, specify if you're

using the one or five year Public Use Microdata Sample. It helps us spread the word. It helps us find out how our data is being used and inspire other people to use it.

Slide 37 – Roses are Red. Violets are Blue. We Love #ACSdata. Do You Love it Too?

We'd also like to share a special invitation with you all to connect with us on social media: Roses are red. Violets are blue. We love #ACS data! Do you love it too? Let us know! We're on Facebook, Twitter, YouTube, Instagram, Pinterest and LinkedIn. And in the lower left of the slide you will see contact information for your general ACS and PUMS questions. We have telephone numbers and emails as well as that technical support for DataFerrett.

Slide 38 – American Community Survey Data Users Group

Final resource is our American Community Survey Data Users Group. This is an online community of interested people using ACS data including the PUMS. There's about 2,000 people in this online community that you can join and share your expert knowledge with each other on the topic. There's a subgroup specifically for people using the ACS Public Use Microdata Sample. You can visit the link at the bottom to join the group.

Slide 39 – Need Local Stats?

Finally we do have data dissemination specialists located throughout the country that speak data and communicate in plain language. They would be happy to give a follow-up presentation, local workshop, webinars, anything that you may want feel free to reach out and ask for that. You can contact at the 1-844-ASK-DATA or census.askdata@census.gov. Thank you all for

joining today, we will go ahead and turn it back over to the operator to open it up for questions.

Operator: Thank you, we will now begin the question and answer session. If you would like to ask a question, please press Star 1, unmute your phone and record your name clearly. Your name is required to introduce your question. If you need to withdrawal your question, press Star 2.

Again, to ask a question, please press Star 1. It will take a few moments for questions to come through, please standby.

Tyson Weister: Thank you, Operator. And while we are waiting for questions to come in, we do want to show you how to access the slides for this presentation. If you visit census.gov/acs this will take you to our main page here. Go ahead and click on the events tab here on the main page and you'll see the Intro to the ACS Public Use Microdata Sample Files Webinar. Click into that link and this morning we've uploaded the slides for the presentation. So do know that that resource is live and available to you now.

Operator: We do have a question on the phone lines. Our question comes from (Dave Armstrong), your line is open.

(Dave Armstrong): Hello, I have one question about PUMA zones. Given that I'm in the New York City area and am doing work that's related to transportation I'm interested in not only the PUMA resident zone but also the work PUMAs and if I remember correctly, the work PUMA for somebody working in Manhattan would be all of Manhattan unless there's some other breakdown of that. I'm interested to know whether given the size of Manhattan whether statistically there's a possibility of breaking that out into some size subdivision, not

necessarily the PUMA for the - not necessarily the same as the residents PUMA for work.

Tyson Weister: Thank you for your question here. The work PUMA's or place of work PUMAs, also known as POWPUMAs that you're referencing, those are based on counties or county equivalents. The reason is that given the topic they are a little bit - they have a little bit more strict rules for disclosure avoidance and to protect confidentiality of respondents. I believe Manhattan is the smallest area that you can get the place of work PUMA's for.

(Dave Armstrong): Right.

Tyson Weister: There's more details about like how they're defined on the PUMA reference page on the slide. Unfortunately for disclosure avoidance reasons that is the geography that's available for the place of work PUMAs.

(Dave Armstrong): Right, my thought is that some counties are quite small and some like Manhattan are quite large. So from the statistical point of view, there might be some basis for subdividing a large county like New York or Los Angeles or something like that. But I guess what you're telling me is at least that hasn't been done so far, I don’t know whether that's considered as a possibility for the future?

Tyson Weister: Yes, you could definitely email your feedback in to the ACSO support email that we have here. Let me go back to the slide, and we could definitely pass that along to the area that is responsible. And we'd be happy to pass along your feedback.

(Dave Armstrong): Okay, thank you.

Operator: We show no further questions at this time. But, again, as a reminder, you can press Star 1 on your phone or you can press Star 1 on your phone and record your name if you have a question. One moment please.

We do have questions coming in. One moment. The next question is from (Stephanie Gallow), your line is open.

(Stephanie Gallow): Hi, this is (Stephanie) with the Dallas Fed, I was just wondering whether you have a sense of whether it's better to use the PUMS file from the Census versus IPUMS, whether there's any particular advantages to using the files directly from the Census? Thanks.

Tyson Weister: Hi, thank you for your question. I'm not personally an experienced user in the IPUMS available from the Minnesota Population Center. I do know that they tend to put standardizations in place, so if you're doing something over time you may want to consider looking into that resource a little bit more. Other than that, I'm not aware off the top of my head of any particular advantages of using the IPUMS versus what's directly available through the Census Bureau's website.

(Stephanie Gallow): Okay, thanks.

Operator: The next question comes from (R.J. Grow), your line is open.

(R.J. Grow): Hi, thank you so much for doing this webinar. I have actually a complex question if you will. In particular, we're interested in doing time series analysis for household information from the 1950s, 1980s, as well as current. I know that IPUMS is a predecessor of ACS. Is there any challenges that you can think of that I might face by using or merging the data? And my second

question is, is that it's a different interface for IPUMS, any recommendations there?

Tyson Weister: Can you please repeat the second part of your question?

(R.J. Grow): Oh, I know that the IPUMS data interface or the data house of it is different than the way that ACS would present it. Any foreseen issues with merging the two or also pulling it down from the data house?

(Sirius Fuller): This is Sirius Fuller again and I work with the Tyson. So for the first part, if you're merging together 1950s and 1980s and then more current data, I assume you're using the decennial PUMS data. You'll have to be aware that the geographies change over time and so you might have to contact somebody in the decennial office to see potential pitfalls if you're merging the data how the boundaries below the state would change over time. And it's - I'm not sure, but it might be in the TIGERweb as well to compare.

In terms of the second question, you're correct, the way the IPUMS works, they're independent of Census, how they house the data and they have a lot more than just the PUMS data. So, you know, they format the way they do and you have to contact them. I don't - I haven't used it extensively but I think you - there might be some work if you're merging it but they do have the PUMS data. But as Tyson said they formalize some of their variables or say they add extra variables to the PUMS data. But the underlying data, my understanding is it's the same.

(R.J. Grow): If I'm not mistaken, they ask the same set of questions, it's just the matter of the sample has grown over time. At least that's from my experience. Thank you very much.

Tyson Weister: And also to follow-up with that, if you are interested in tracking the history of a particular PUMA over time, I have put up here the contact information for geographic customer service at the 301-763-1128 and geo.geography@census.gov.

(R.J. Grow): Thank you, I'll write it down.

Operator: And, again, as a reminder, if you would like to ask a question, you can press Star 1 on your phone. One moment please for any additional questions. We are showing now further questions at this time.

Tyson Weister: Great, thank you, Operator. Well thank you all for joining the webinar this afternoon. We definitely appreciate you taking the time to tune in today. If you do have any additional questions, feel free to reach out to us, our contact information is included here on the slide and we look forward to working with you soon.

Operator: That does conclude today's conference, thank you for participating. You may disconnect at this time.

END